

「地下ダムの水質汚濁・渇水を同時予測する人工知能の創生」

神戸大学大学院農学研究科

井上 一哉

1 研究の背景と目的

南西諸島は、九州南部の大隅諸島から台湾にかけて連なる島嶼群の総称であり、東シナ海と太平洋の境界に位置している。地質学的にも日本本土とは異なり、喜界島や沖縄諸島中南部、宮古諸島等は更新世のサンゴ礁が隆起して形成された琉球石灰岩に覆われている。多孔質で高透水性な地質特性を有することから、琉球石灰岩地帯では、蒸発分を除くと、雨水の大半は地下へ浸透するため、河川や湖沼等の地表水はほとんど見られない。また、降雨のほとんどは梅雨や台風期に集中するため、降雨量の年較差が大きく、南西諸島はたびたび干ばつ被害を被ってきた。

地下に浸透した水は琉球石灰岩と不透水性基盤である島尻泥岩層の境界面を流下し、島の周囲から海へ流出するため、水利用はガーと呼ばれる海岸断崖沿いの湧水や自然洞穴泉に依存しており、農業は農業用水を天水に頼る不安定な農業形態であった。そのため、南西諸島では 1970 年代以降、地下ダム事業を積極的に展開してきた。地下ダムは、地下の帯水層中に人工的に難透水性の止水壁を築造し、地下水流れを堰き止めて地下水を貯留する農業水利施設である。地下ダムの導入により、基幹作物であるサトウキビの収量増加のみならず、マンゴーをはじめとする熱帯果樹や野菜、園芸作物といった高収益作物への転換が進んでいる。地下ダム建設により、以前と比べて渇水の可能性は激減したものの、南西諸島では無降水日の日数も長期的には増加しており、持続的な農業活動を展開する上でダム貯水量と水質の維持は最重要課題といえる。

貯水量の変動は降水量や蒸発散量といった気象要素に加え、農業活動等の土地利用の影響が複雑に絡み合っており、未だ解明されていない。また、地下ダムサイトにおける貯水量の変動に関する研究例は少なく、特に機械学習の導入はほとんどなされていない。そこで本研究では、農家の渇水に対する不安を取り除き、適切な水管理による農業用水の安定供給を実現するため、機械学習を用いて地下ダム流域における将来の貯水量を予測すると同時に、硝酸態窒素汚染対策として水質の将来動向についても予測の対象とするシステムを構築する。

2 研究方法・研究内容

2. 1 研究対象サイト

研究対象の喜界島は、鹿児島県から南南西に約 380 km に位置する面積 56.9 km² の離島であり、奄美群島の 1 つである。本島はサンゴ礁が隆起してできた島であり、琉球石灰岩地帯に位置することから地表水はほとんど存在せず、自然状態においては基幹作物であるサトウキビの新植・生育期を襲う干ばつにより、農業は度々被害を受けてきた。そこで、農業用水源として新たに地下水を開発する目的で地下ダムおよび集水井・揚水機場などの取水施設を建設し、基幹的農業水利施設の整備にて、農業用水の安定的確保・供給が実現している。図 1 に喜界島の全体図、ならびに喜界地下ダム流域の様子を示す。

総貯水容量 180 万 m³ を有する喜界地下ダムは、図 1 の赤線にて示すダム軸に沿って水位計測のための観測孔が設置されており、地下水揚水のための集水井が 8 箇所存在する。地下水観測は、2000 年 4 月 1 日から 1 か月間隔で地下水位が観測されており、一例として、図 2 に観測孔 1A、30A および 8 号集水井の地下水位の経時変化を示す。ダム満水位は、20 m の位置であり、観測孔ごとに固有の変動特性を有していることがわかる。また、集水井では地下水揚水の影響を直接受けることから、時期によっては大きく水位低下することもあり、ダム湖の水量や水質予測にはこのような変動特性を効率よく学習する必

要がある。そこで本研究では、土地改良区の定期計測データに加えて、観測孔と集水井にて水質調査を実施することで水量・水質予測システムの学習情報を増やすこととした。

2. 2 機械学習アルゴリズム

本研究における予測システムの軸となる学習アルゴリズムを選定すべく、複数のアルゴリズムにて一連の予測システムを構築した。具体的には、決定木系のアルゴリズムとして採用頻度の高い **Random forest** (以下, **RF** と記す)を基準にして, **RF** の拡張版である **Extremely randomized trees** (以下, **ERT** と記す), ならびに **Gradient boosting decision trees** (以下, **GBDT** と記す)を適用した。**RF** は、複数の決定木を作成し、それらの予測を組み合わせる最終的な結果を得るアンサンブル学習手法である。各決定木はブートストラップ法でサンプリングしたデータで学習し、特徴量のランダム選択を行い、平均値により予測する。**ERT** は、**RF** に似たアンサンブル学習手法であり、決定木の構築時に分岐の基準となる特徴量と閾値をランダムに選択する。これによりモデルの多様性が増し、訓練(学習)が高速化される同時に、過学習を防ぐ効果がある。複数の木の結果を平均して予測する。**GBDT** は、決定木をベースにしたアンサンブル学習法の一つであり、ひとつ前の決定木が間違えた部分を補足するように新しい決定木の作成を繰り返し、予測モデルの精度を高めていく手法である。目的変数とそれまでに作成した決定木による予測値の差を学習し、誤差少なくなるよう決定木を直列に追加することで誤差が小さくなることを期待できる。

決定木系のみの適用では、システムの良否について十分に判断できないことから、系の異なるアルゴリズムとして、**Support vector regression** (以下, **SVR** と記す)も対象の1つに加えた。**SVR** は、サポートベクターマシン (**SVM**) を基にした回帰モデルであり、与えられたデータに対して許容誤差範囲内で最も平坦な関数(回帰直線)を見つける手法である。特徴的な点は、誤差が設定した閾値内に収まるように最適な境界(サポートベクトル)を選ぶ点であり、非線形な関係にも対応できるカーネル関数を使用することで高精度な予測が可能である。

2. 3 観測データ群

貯水量/水質予測システムにてアルゴリズムの学習情報となる観測データ群は、観測孔での計測が中心となる。本研究では、ダム軸沿いの観測孔および集水井にて計測した地下水位・水質データをもとに将来の地下水位と硝酸態窒素濃度を予測した。地下水位は2000年4月から2024年2月まで、約1か月、または約2か月おきに観測されている。加えて、気象データとして、喜界島のアメダスにて計測されている降水量、平均気温、最高気温、平均風速、最大風速の月別値を利用した。一例として、気温と降水量の経時変化を

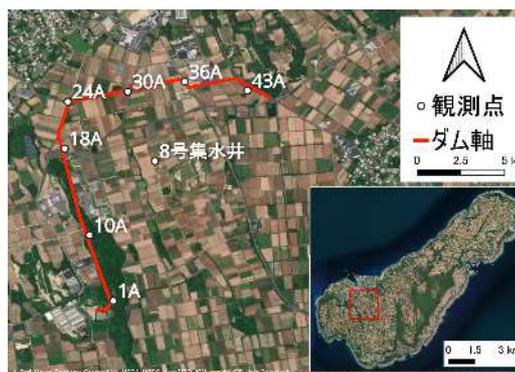


図 1：喜界島の対象流域ならびに代表的な観測点位置

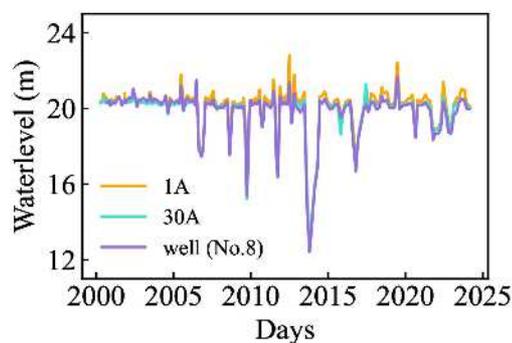


図 2：観測データの一例。観測孔 1A と 30A, 8 号集水井における水位変化。

図 3 に示す。喜界島は亜熱帯性気候に属しており、年間を通して温暖な気候となっている。

本研究にて対象とする予測日は 2023 年 1 月 (以下, 3 か月後予測日と記す), 2023 年 4 月 (以下, 6 か月後予測日と記す) とした。これは観測データを予測値と見なすことによって, 開発したシステムの性能を評価するためである。本モデルによる予測対象項目は 3 か月後, 6 か月後の地下ダム貯水量である。そのためにまず, 予測日直前までの観測データ群を学習して地下水位を観測点ごとに予測した。次に, 喜界島の地質構成を参考にして, 貯水位と貯水量の関係 (以下, $H-V$ 式と記す) を求め, 本システムにて水位を予測すると $H-V$ 式を経由して貯水量が予測される流れである。観測水位から貯水量への変換式を図 4 に示す。

水質予測システムでは地下水の硝酸態窒素汚染の早期発見や予防に向けた予測が目的である。そのため, 30 日後や 60 日後の硝酸態窒素濃度の予測を達成すべく, 2023 年 12 月から 2024 年 11 月までに得られた観測データに基づく予測システムを構築した。観測データ群としては, 硝酸態窒素, 亜硝酸態窒素, カルシウム, マグネシウム, pH, 水温, TDS, EC であり, これらをすべて学習情報として活用した。2024 年 11 月の観測データを予測値と比べることで性能評価とした。

3 研究成果

3.1 貯水量予測

学習アルゴリズムを選定するため, 3 つのアルゴリズムに対して代表的な観測孔と集水井にて予測した水位の結果を図 5 に示す。図中, 予測日の実測データを併記しており, ERT の精度の良さがわかる。これは ERT の学習過程では分岐選択にランダム性を取り入れていることが要因と推量される。ERT により予測した予測日の貯水量について, 3 か月後予測, 6 か月後予測の結果を図 6 に示す。図中, 100 回の予測計算に基づく予測の分散を示す。本モデルでは, 6 か月後までの貯水量を良好に予測する一方で, 予測期間が長くなるにつれて予測の分散は増大している。予測すべき期間が長くなればなるほど, 不規則な変動形態などの不確実要因の蓄積により, 予測精度が低下することは自然な現象である。

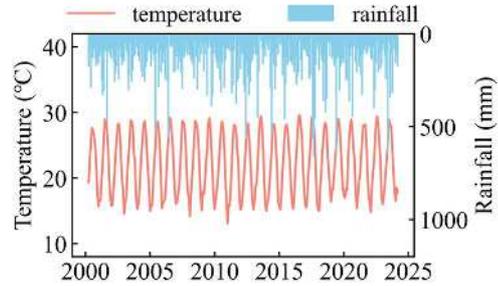


図 3: 気温と降水量の経時変化

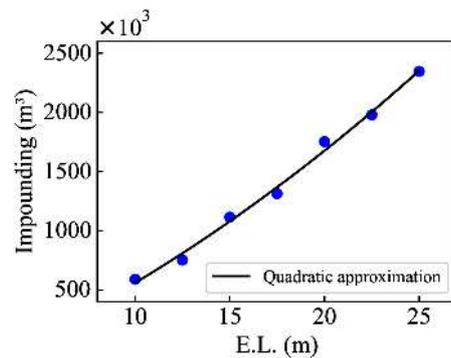


図 4: 観測水位から貯水量への変換

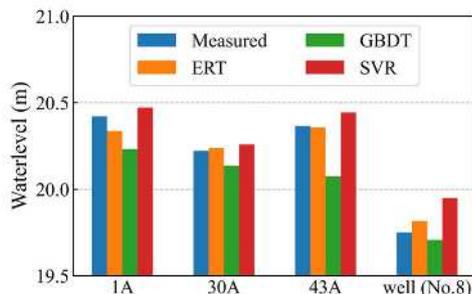


図 5: 学習アルゴリズムの精度の比較

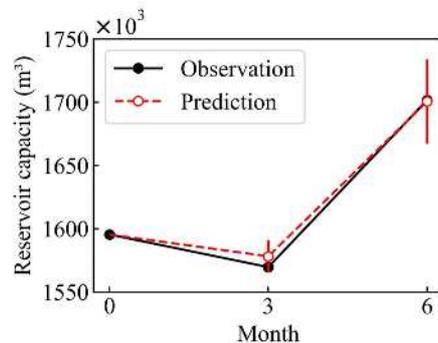


図 6: 貯水量予測結果と予測の分散

観測孔単独で貯水量を予測した結果，ならびに各観測孔の値を平均した予測貯水量を明示するため，図7に3か月後と6か月後の予測結果を描く．図中，貯水量をダムの総貯水容量で除した貯水率として，実測の貯水率とともに数値を記している．単一の観測孔より予測される値は実測値の前後でばらつく結果になる．これは図2に示すように，観測孔ごとに固有の水位変動幅を有していることから，観測孔単独で貯水量を予測すると数値にばらつきが生じるためである．開発した貯水量予測モデルは各観測孔の水位変動特性を平滑化するモデルであり，良好な予測値を導くことができる．

3. 2 水質予測

図8に学習情報の1つである硝酸態窒素濃度の推移を示す．1日単位で直接計測することは困難であることから，年間5度の実計測結果を基にして，観測点の間はトレンドを取った上で，1%のホワイトノイズを付加することにより，学習データの量を見かけ上増やしている．同様のノイズ付加を他の観測項目についても実施することで1年間の観測結果より30日後や60日後の硝酸態窒素濃度について予測の可否を判断した．

学習アルゴリズムは貯水量予測にて優秀な成果を導いたERTを採用しており，予測結果を可視化できるように，観測孔単独で硝酸態窒素濃度を予測した結果，ならびに各観測孔の値を平均した予測濃度を図9に明示する．30日後，60日後ともに実測濃度に近い予測濃度を推定できており，本研究にて開発した水質予測システムは60日までの予測であれば，良好な結果を導くといえる．

4 結論

4. 1 結論

本研究では，喜界地下ダムを対象に機械学習を用いた貯水量予測システムと水質予測システムを開発した．まず，本研究にて扱うデータ群と学習器の相性について，学習アルゴリズムとして **Extremely randomized trees (ERT)** の性能の高さが示された．また，貯水量予測システムは6か月後まで，水質予測システムは60日後までを良好に予測できる能力を有していた．両システムは学習データの観測期間が大幅に異なるため，予測期間に相違が生じているものの，継続的な観測により，さらなる予測精度の向上が期待される．

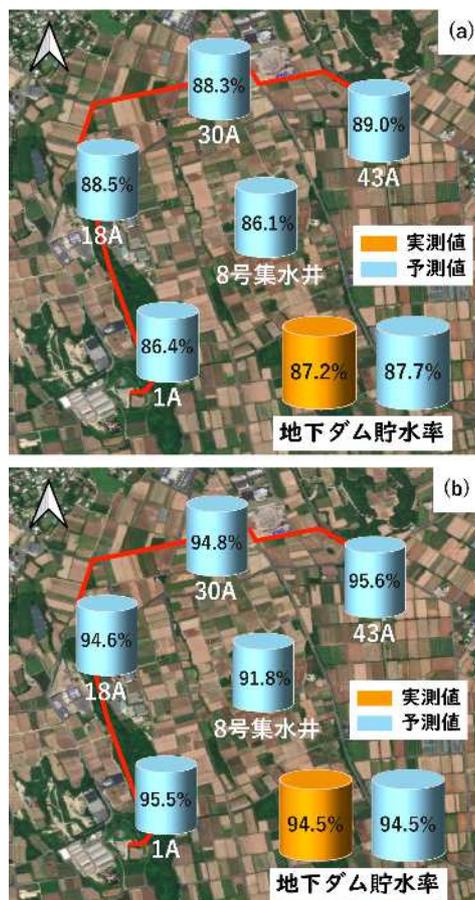


図7：地下ダム貯水量の予測結果：(a) 3か月後，(b) 6か月後．

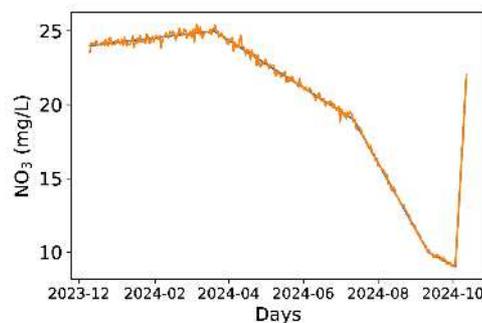


図8：硝酸態窒素の推移と学習データ

4. 2 生活や産業への貢献と波及効果

生活や産業への貢献としては、安全・安心な水利用の達成にある。水量・水質ともに良好な状態ある水の与える生活・農業活動へ恩恵は計り知れず、研究成果の他地区・他分野への波及効果も期待できる。AI 研究はあらゆる学術分野へ浸透してきている一方で、水管理と AI の連携は世界的に手薄である。兵庫県内では防災重点ため池を筆頭に、健全な水利用として監視すべきため池が膨大に存在することから、将来的には、低コスト・省力化・迅速なため池管理へ本研究の成果を応用する。これらの研究成果が情報工学を越えて防災科学や土木工学、農学など幅広い分野において利用される波及効果に期待したい。

参考文献

- 1) Geurts, P., Ernst, D. and Wehenke l, L.: Extremely randomized trees, Machine Learning, 63, pp.3-42, 2006.
- 2) Friedman, J. H.: Stochastic gradient boosting, Computational Statistics & Data Analysis, 38(4), pp.367-378, 2002.
- 3) Gunn, S.: Support Vector Machines for classification and regression, ISIS technical report, pp.31-39, 1998.

謝辞

本研究の遂行には、ひょうご科学技術協会学術研究助成を受けました。記して、感謝の意を表します。

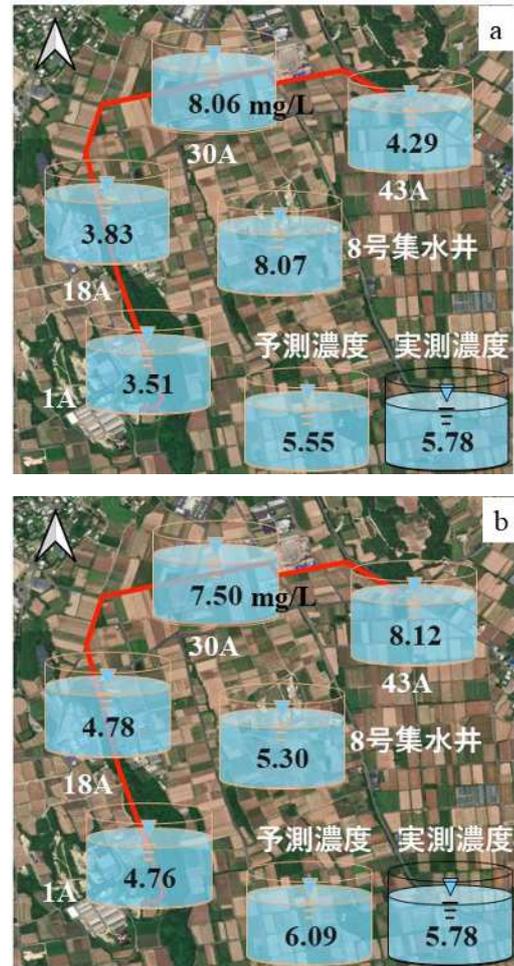


図 9：地下ダムの硝酸態窒素濃度の予測結果：(a) 30 日後、(b) 60 日後。