

機械学習を用いた農業環境の時空間変動解析

Machine learning-based spatiotemporal modelling of agro-environmental systems

○福田信二¹

○FUKUDA Shinji

1. はじめに

農業生産の場となる水田や河川・水路等の水環境は、気候変動や人為的土地利用の影響を受けて複雑かつ地域特異的な特徴を有している。特に、気象・水文・水質・土壌などの環境要素の変動は、空間的にも時間的にも不均一であり、その特性を適切に計測し、高度に解析することは、持続可能な農業環境管理の基盤となる。近年、センサネットワークやリモートセンシング技術の発展により、多変量かつ高頻度な環境データの取得が可能となっており、それらを有効活用するための解析手法や可視化技術の開発が求められている。そこで本研究では、代表的な機械学習であるランダムフォレスト (Breiman, 2001) を用いた解析例として、コロンビアにおける水田圃場の土壌水分の時空間変動解析、ベトナム・メコンデルタ河口域における塩分時系列再構築等の事例を紹介し、生物・物理プロセスとの関連性について議論する。

2. 方法

コロンビアでは、水田圃場 (3地点, 4圃場) において、灌漑後12日間以上の期間に、異なる深度 (12 cm, 20 cm) を調査可能なGPS付TDRセンサを使用し、各圃場100地点以上の体積含水率を毎日計測した。取得したデータには、位置情報と気象情報 (気温, 風速, 湿度などの観測地; イバゲ空港) を付加し、ランダムフォレストによる土壌水分量の時空間変動モデルを構築した。メコンデルタでは、ベンチェ省に位置する4地点 (An Thuan, Ben Trai, Binh Dai, Vam Kanh) における水位の連続観測値 (2時間間隔) と塩分データは断続的観測地 (2時間間隔; 約2週間ごと) を使用した。水位の時系列データ (現在から10時間前まで6時点データ) を入力とし、現在の塩分を出力とする時系列回帰モデルを構築し、各観測地点でのモデルの汎化性能を評価した。解析には、統計ソフトR (R Core Team, 2015) のパッケージ “randomForest” (Liaw and Wiener, 2002) を使用し、変数の重要度および応答曲線に基づいて、対象現象の影響要因を可視化した。再現性の評価には、平均二乗誤差 (MSE) およびNash-Sutcliffe Efficiency (NSE) を使用し、50回の独立した試行によって、モデルの構造のばらつきを評価した。

3. 結果と考察

コロンビアでの解析では、12cm深のモデル (RF_{12cm}) が特に高い再現性を示し、灌漑日からの経過日数や圃場内の位置が土壌水分動態に大きな影響を与えていることが変数重要度から明らかになった。応答曲線 (Fig. 1) では、観測初期における急激な土壌乾燥や緯度経度に応じた土壌水分動態など、非線形的な挙動が確認された。

メコンデルタでの解析では、各地点におけるランダムフォレストモデルの予測精度は高く

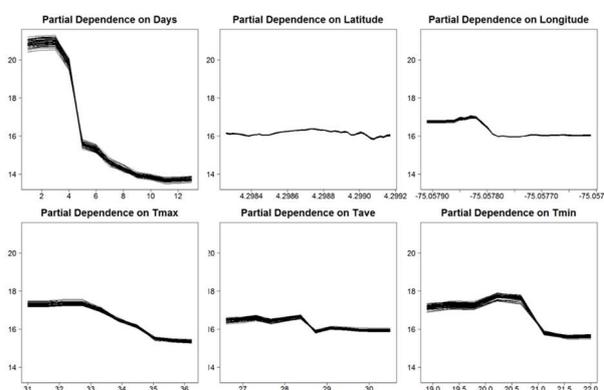


Fig. 1 Partial dependence plots for soil moisture content (%) modelled in Potorerito (RF_{12cm})

¹ 東京農工大学大学院農学研究院 Institute of Agriculture, Tokyo University of Agriculture and Technology

(NSE > 0.9), 水位時系列に基づく塩分濃度の再現は良好であった (Fig. 2). 一方で, ある地点で構築したモデルを別地点に適用した交差検証では性能の著しい低下が確認され (NSE: -0.86~0.27), 水位の変動パターンが類似していても, 塩分応答は河道地形や潮汐影響等の地域特性に強く依存することが示唆された. 応答曲線や変数重要度 (Fig. 3) の解析により, 地点ごとの塩分動態の差異やモデル誤差の季節的の傾向を可視化することができた. これは, 長期的な水位-塩分関係を理解する上で有用な情報である.

ただし, ランダムフォレストは, データの入出力関係を捉える手法であり, 物理的プロセスを直接的に反映するものではない. したがって, 必要に応じて物理プロセスを反映したモデル構造について検討する必要がある. 例えば, 解析期間の始期および終期付近では塩分推定誤差の傾向が異なり, 乾季から雨季への季節変化の影響だと推察される. 精度向上のためには, 流量変動などの物理プロセスの反映が重要であると考えられる.

以上のように, データ駆動解析は, 対象とする現象・システムに影響を及ぼす主要な環境要因の特定や時空間的な非線形関係の可視化を通して, 観測データに基づく現象理解の深化に寄与するものである. また, その知見は物理モデルのパラメータ設定や構造的検討にも有用であり, プロセスベースモデルへのフィードバック手段としての可能性も示唆された. 例えば, 近年注目される Physics-Informed Neural Networks (PINNs) のように, 物理プロセスを明示的に損失関数に組み込んだ学習の枠組みを参考にすれば, ランダムフォレスト等の手法においても, 物理的な整合性を担保した解析への展開が可能かもしれない.

4. おわりに

本報では, ランダムフォレストを用いた農業環境における時空間変動のデータ駆動解析を通じて, 物理モデルでは捉えきれない非線形関係の抽出や変動パターンの可視化が有効であることを例示した. 一方で, 従来型のデータ駆動解析では, モデルの汎用性や因果解釈に限界があることから, 今後は物理モデルとのハイブリッド化や物理プロセスを考慮したモデル最適化が必要である. また, 生物多様性指標や環境負荷評価といった他のデータとの融合による応用展開も視野に入れ, 農業と環境の調和に資する予測モデルの構築に向けて, データ駆動モデルとプロセスベースモデルの相補的あるいは階層的な活用が重要になると考えられる.

参考文献

- Breiman, L. (2001): Random forests. *Machine Learning* 45, 5–32.
 Liaw, A. and Wiener, M. (2002): Classification and regression by random forest. *R News* 2(3), 18–22.
 R Core Team (2015): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

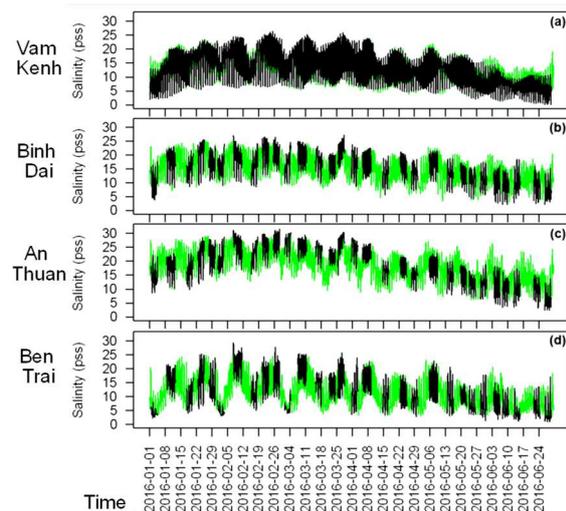


Fig. 2 Observed and reconstructed salinity time-series (Black: observed; Green: reconstructed).

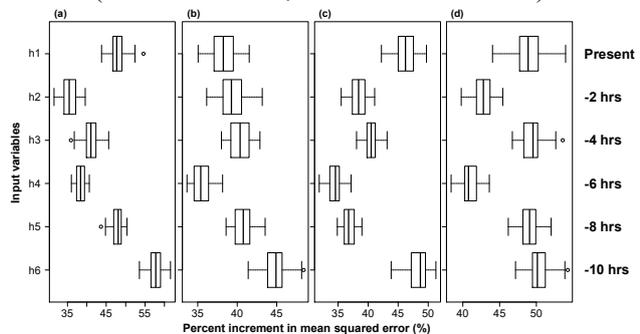


Fig. 3 Variable importance computed by RF for salinity time-series reconstruction