

勾配ブースティングに基づく海面水温の予測 Sea-Surface Temperature Forecasting using Gradient Boosting

○榎本 剛・齋藤亜紀・松本涼太・中下早織
○Takeshi ENOMOTO, Aki SAITO, Ryota MATSUMOTO, Saori NAKASHITA

Machine learning (ML) techniques enable fast and accurate data-driven forecasting. This study attempts to predict the sea-surface temperature (SST) solely from data. Instead of neural networks, a gradient-boosting tree is adopted to inherently handle missing values, discontinuities and non-linearity. This study utilizes XGBoost, a popular gradient-boosting implementation, to SST prediction (XGBST). Preliminary experiments using HIMSST as input showed that XGBST outperformed the persistence forecasts, as evidenced by lower root-mean-square errors (RMSE). Although learning processes and feature selection require refinement to improve horizontal distribution, the proposed method can handle SST products with missing values and is extensible to incorporate other features such as atmospheric reanalysis and ocean surface currents. (108 words).

1. はじめに

海面水温 (SST: sea-surface temperature) は、大気下端の境界条件として、海面におけるエンタルピーフラックスを通じて、大気に影響を与えていた。海面水温の異常は、線状降水帯や、豪雪、熱波など災害をもたらしうる顕著現象の重要な要因となるため、その実態の把握と予測は防災の観点から極めて重要である。

海面水温の予測は従来、海洋物理モデルを用いて行われてきたが、時間変動が激しい大気の影響を受け、海洋内部よりも短い時間スケールで変化するため不確実性が大きい。近年では、機械学習を用いたデータ駆動による予測も試みられている (Zrira et al. 2024)。

本研究では、勾配ブースティングを用いて、数日先の SST の予測を試みた。勾配ブースティングは結果の解釈可能性が高く、学習や推論が高速であり、表形式データに対する精度は最新の機械学習手法に匹敵する。本研究では、勾配ブースティングの効率的な実装である XGBoost (Chen and Guestrin 2016) を用いた。

2. 実験設定

入力は、気象庁 HIMSST (解像度 0.1°) あるいは MGDSST (0.25°) の前日、当日の SST、その東西及び南北勾配、9 点平均、日付、経度、緯度、出力は 5 日先の SST である。2022 年 6 月一か月分の日別 SST から乱択した 200 点でのデータを学習に用い、2023 年 6 月のデータで検証を行った。

3. 結果

HIMSST を学習し、2023 年 6 月 15 日から 5 日先の 20 日を予測した例 (Fig. 1) では、亜熱帯を中心に検証日の HIMSST との差は、±0.5 K 程度であった。一方、日本海や長江河口で -2 K 超の低温バイアス、本州の南海上 (133°E, 28°N) で 2.5 K 超の高温バイアスが見られ、黒潮続流域の渦の再現性が著しく悪かった (Fig. 2)。

対象期間中、東シナ海から日本海、亜熱帯にかけては 1 K 程度上昇し、本州南海上では 0.5 K 程度下降していた (Fig. 3)。XGBoost による予測は、亜熱帯の変化は捉えられたものの、日本海や本州南海上での急激な変化が予測できなかった。

1 か月間の統計では、領域平均誤差 (RMSE) が持続予報の 1.05 K から 0.947 K に低減した。特徴量の重要度 (feature importance) を確認すると、SST の 9 点平均が地点 SST の倍近いスコアとなり、最も寄与が高かった。

次に MGDSST を用いた実験では、日本海や本州南海上、黒潮続流域での誤差が縮小し、領域平均誤差が持続予報の 0.957 K から 0.69 K に低減した (Fig. 4)。この場合、特徴量の重要度は、地点 SST が 9 点平均を大きく上回った。

4.まとめと今後の展望

XGBoost を用いて軽量な海面水温予測モデルを構築した。統計的には持続予報を上回り精度が得られたが、特定の海域で大きな誤差が残る課題も

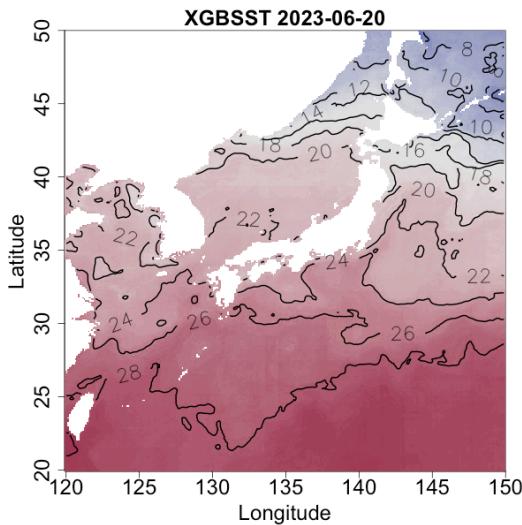


Fig. 1 5-day forecast of sea-surface temperature starting from 15 June 2023 using HIMSST and XGBoost,

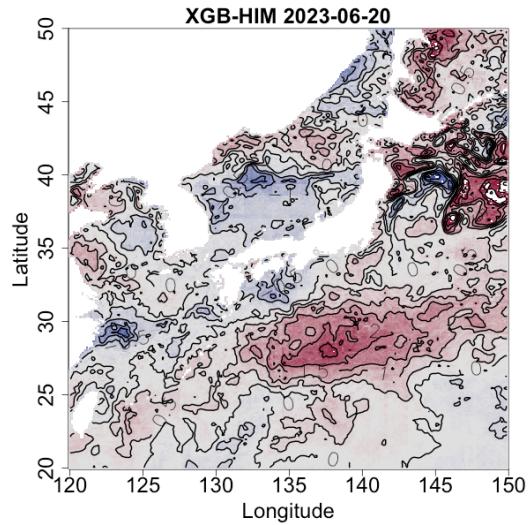


Fig. 2 Difference between the 5-day forecast and HIMSST on 20 June 2023.

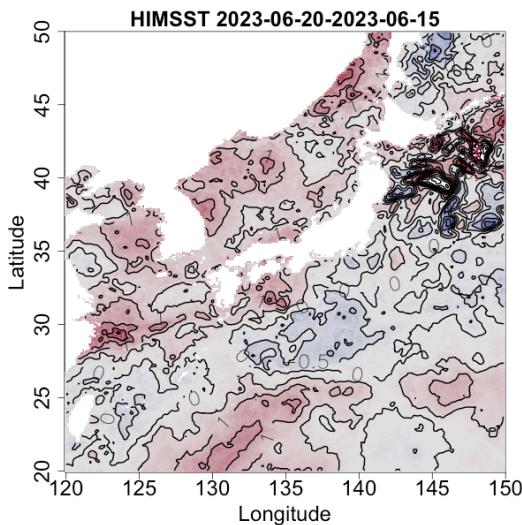


Fig. 3 Observed 5-day change in HIMSST (20 June minus 15 June 2023)

明らかになった。このような結果は、学習データの地点数が少ないことが一因であると考えられる。解像度が高い HIMSST の学習で 9 点平均が、解像度の低い MGDSST では地点 SST が特徴量として最も重要であったことは、HIMSST にノイズが含まれることを示唆している。

今後は、学習期間や地点数、季節に対する依存性の調査に加え、大気再解析を用いた大気海洋相互作用の考慮や、空間パターンを取り入れた学習手法の導入に取り組む予定である。

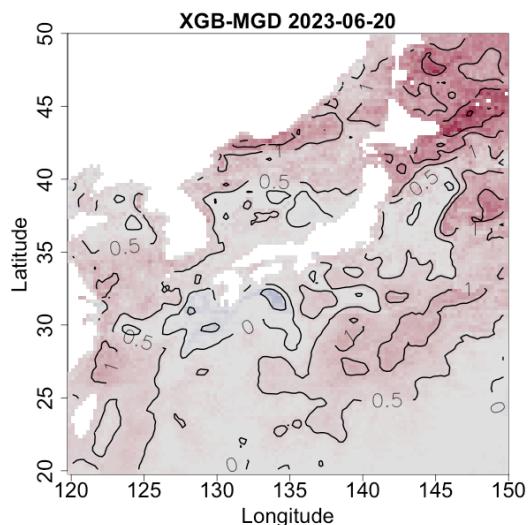


Fig. 4 As in Fig. 2, but using MGDSST as input instead of HIMSST.

参考文献

- Chen, T., and C. Guestrin, 2016: XGBoost: A Scalable Tree Boosting System, [doi:10.48550/arXiv.1603.02754](https://doi.org/10.48550/arXiv.1603.02754).
- Zrira, N., A. Kamal-Idrissi, R. Farssi, and H. A. Khan, 2024: Time series prediction of sea surface temperature based on BiLSTM model with attention mechanism. *J. Sea Res.*, **198**, 102472, [doi:j.seares.2024.102472](https://doi.org/10.1016/j.seares.2024.102472).