

Analysis and Predictive Modelling of Oceanic Variables in the Atlantic Ocean Using Multivariate Statistical Methods

Mfon David Umoh

mfonslago20@gmail.com

Maritime Academy of Nigeria

Udoh Felix Evans

University of Uyo

John O. Esin

Maritime Academy of Nigeria

Research Article

Keywords: multivariate regression, sea surface temperature, sea surface salinity, ocean currents, predictive modelling, Atlantic Ocean, Copernicus Marine Service

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11206483/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

The interrelationships among sea surface temperature (SST), sea surface salinity, and horizontal ocean current velocity govern how heat and salt are redistributed in the Atlantic Ocean, yet the directional structure of these relationships is rarely quantified systematically. This study analyses four oceanic variables together, ocean potential temperature, sea water salinity, eastward current velocity, and northward current velocity, using oceanographic observations from the Atlantic Ocean.

Methods

Each variable was modelled in turn as the dependent variable against the remaining three as predictors, yielding 28 regression models (seven per dependent variable, spanning one-, two-, and three-predictor specifications). The models were compared on six metrics: the correlation coefficient (R), the coefficient of determination (R^2), root mean squared error (RMSE), mean squared error (MSE), mean absolute percentage error (MAPE), and mean absolute bias error (MABE).

Results

The three-predictor models were optimal for all four dependent variables. Eastward velocity was the best-predicted variable ($R^2 = 0.771$, adjusted $R^2 = 0.733$), with salinity ($\beta = -0.723$, $p < .001$), temperature ($\beta = -0.246$, $p = .045$), and northward velocity ($\beta = -0.281$, $p = .024$) all contributing significantly. Salinity was predicted nearly as well ($R^2 = 0.697$), driven almost entirely by eastward velocity ($\beta = -0.954$, $p < .001$). The models for temperature ($R^2 = 0.227$) and northward velocity ($R^2 = 0.271$) explained less variance, with eastward velocity the only significant predictor in each.

Conclusions

The predictive structure of the variable system is asymmetric: zonal circulation occupies a central position, acting at once as the strongest driver of salinity and temperature variability and as the variable most coherently explained by the thermohaline state. The findings offer a quantitative, equation-level framework for predictive modelling of upper-ocean variables and can inform the design of data-driven ocean forecasting systems.

Background

Sea surface temperature, salinity, and ocean current velocity are the core state variables of the upper ocean. Between them they control the density structure of the water column, the transport and redistribution of heat and salt, air–sea exchange, and the climate-relevant circulation patterns of the Atlantic basin [1,2]. Before one can estimate ocean state, predict climate, or design an operational forecasting system, it helps to know how these variables influence one another, both directionally and quantitatively.

The Atlantic is a particularly consequential place to ask such questions. Its heat transport is dominated by the Atlantic Meridional Overturning Circulation (AMOC), whose variability directly shapes North Atlantic SST patterns [3,4] and whose stability depends on the density gradients that temperature and salinity jointly determine [5]. At low latitudes, zonal equatorial currents modulate SST anomalies through advection and mixing [6,7], while the salinity field reflects the balance of evaporation, precipitation, and transport by the circulation [8]. The result is a tangle of bidirectional relationships that pairwise analyses cannot fully untangle.

Multivariate statistical methods are the natural framework for this problem. By modelling each variable in turn as the dependent variable against the remaining variables, one can quantify the full directional structure of the system: which variables make strong predictors, which are best predicted, and where predictive information concentrates [9,10]. Multiple linear regression remains the workhorse of such analyses in the geosciences, provided its assumptions are respected and its coefficients are interpreted with an eye on predictor intercorrelation [11,12]. Consistent, quality-controlled reanalysis products from the Copernicus Marine Service [13] also make such analyses reproducible and easy to compare across studies.

Empirically, however, a gap remains. Existing studies of Atlantic Ocean variables tend to examine isolated pairs of variables, whether SST and salinity [14], currents and SST [7], or circulation and climate [4], or they focus on specific current systems [15]. To our knowledge, no one has yet reported a systematic evaluation in which all four core upper-ocean variables are alternately modelled as dependent variables, with model selection carried out over multiple competing specifications and evaluation metrics.

The study had four objectives: (1) to characterize the statistical properties of SST, salinity, eastward velocity, and northward velocity in the study area; (2) to develop a family of 28 regression models covering all one-, two-, and three-predictor specifications for each of the four dependent variables; (3) to identify the best-performing predictive model for each dependent variable using a multi-metric evaluation framework; and (4) to pull the full set of results into a coherent account of the predictive structure of the Atlantic upper-ocean variable system, with implications for data-driven ocean modelling [16,17].

SST variability and its mechanisms

SST varies across temporal and spatial scales under the combined influence of surface heat fluxes, advection, mixing, and coupled ocean–atmosphere feedbacks. Deser et al. [1] synthesized the patterns and mechanisms of SST variability, and advection by ocean currents stands out in their account as a first-order contributor to regional SST anomalies. In the tropical Atlantic, the Atlantic Niño and related modes organize SST variability by coupling wind stress, thermocline depth, and current transport [18,19]. One implication is that current velocity fields carry genuine predictive information for SST, and the present study tests that premise directly.

Salinity in the ocean–climate system

Sea surface salinity is set by the freshwater flux (evaporation, precipitation, runoff) and redistributed by the circulation [8]. Because salinity, together with temperature, determines seawater density, it feeds back on circulation and stratification. Li et al. [20] documented the progressive stratification of the world ocean over the past half-century, a trend in which salinity changes participate. At the basin scale, Kido et al. [14] characterized the scale-dependent covariability of SST and sea surface salinity and showed that the thermohaline relationship is mediated by mixed-layer processes. These results argue for including salinity both as predictor and as predictand in the present design.

Zonal and meridional currents in the Atlantic

The equatorial and tropical Atlantic host strong zonal currents whose variability bears directly on the SST field. Brandt et al. [6] documented the zonal current system of the equatorial Atlantic, and Tuchen et al. [7] showed that the modulation of equatorial currents and Tropical Instability Waves during the 2021 Atlantic Niño controlled how SST anomalies developed. Meridional currents are generally weaker in the tropics, but they transport heat across latitude bands and link the tropical and subtropical gyres [15,21]. Hsu et al. [22] demonstrated the tight coupling between surface currents and oceanic fronts, where temperature gradients concentrate. The finding in this study that eastward velocity is the strongest predictor among the analyzed variables fits well with this literature's emphasis on zonal advection.

The AMOC and thermohaline circulation

The AMOC connects the variable system to the large-scale climate. Cunningham et al. [3] established the continuous observational record of the AMOC at 26.5°N, and Buckley and Marshall [4] reviewed its mechanisms, emphasizing its sensitivity to buoyancy forcing. Rahmstorf [5] showed that the thermohaline circulation has multiple stable states, with transitions governed by density gradients that depend jointly on temperature and salinity. These results give the statistical relationships quantified here some physical weight, particularly the strong bidirectional salinity–velocity coupling.

Multivariate statistical methods in oceanography

Multivariate approaches have a mature methodological basis in the marine sciences. Allen and Somerfield [9] formalized multivariate skill assessment for marine model outputs; Shaw [10] codified the application of multivariate statistics to environmental data; and Thomson and Emery [2] provide the standard treatment of descriptive statistics and regression in physical oceanography. On the regression side specifically, Richter and Stavn [11] addressed functional relationship estimation in

oceanographic systems where predictors and predictands are both measured with error, and Geladi et al. [12] documented the hazards of nonlinearity and predictor collinearity in environmental regression, which is the context needed to make sense of the coefficient sign reversals observed here.

Model evaluation metrics

Which metrics one reports conditions the conclusions one draws. Willmott and Matsuura [23] and Chai and Draxler [24] debated the relative merits of MAE and RMSE and ended up on the same page: the two are complementary, with RMSE penalizing large errors and MAE robust to outliers. Chicco et al. [25] showed that R^2 , MSE/RMSE, MAE, and MAPE probe different aspects of regression performance and should be reported jointly. This study follows that practice, selecting best models by consistent superiority across R , R^2 , RMSE, MSE, MAPE, and MABE.

Data sources and data-driven ocean modelling

The Copernicus Marine Service provides the reanalysis products on which quantitative Atlantic studies increasingly rely [13,26]. Current velocity fields from such products descend from the measurement tradition of acoustic Doppler current profiling [27]. Beyond classical regression, data-driven methods, including neural networks for SST prediction [17,28] and machine learning more broadly [16], are growing complements to dynamical modelling. Regression frameworks of the kind developed here supply the baselines and predictor-selection evidence those methods need.

Ecological and climatic stakes

Variations in the analyzed variables eventually propagate to marine ecosystems and climate. González-Pola et al. [29] documented the decadal covariance of physical variables and ecosystem state in the North Atlantic, and SST anomalies define marine heatwaves, which have intensified globally (mechanisms reviewed in [1]). Predictive models of the upper-ocean variable system therefore serve purposes well beyond physical oceanography.

Gap addressed

To our knowledge, no prior study has reported a symmetric multivariate regression analysis of Atlantic Ocean reanalysis data in which each of the four core upper-ocean variables is alternately modelled against the other three, with systematic model selection across 28 specifications and six evaluation metrics. This study fills that gap.

Methods

Data

The analysis uses oceanographic observations from the Atlantic Ocean, comprising four variables: ocean potential temperature (SST, °C), sea water salinity (so, PSU), eastward ocean current velocity (u_o , m s^{-1}), and northward ocean current velocity (v_o , m s^{-1}). The data cover January 2024 to October 2025 and were analysed as monthly averages. The variable naming follows the Copernicus Marine Service convention, and the data are consistent with the GLORYS reanalysis ecosystem [13,26]. The dataset contains 22 observations for the temperature and velocity variables and 47 observations for salinity. The study area is the South Atlantic Ocean, roughly bounded by 60°S–0° latitude and 70°W–20°E longitude. Data came from the Copernicus Marine Data Store Global Ocean Physics Reanalysis product, which has a horizontal spatial resolution of about $0.083^\circ \times 0.083^\circ$ (roughly 8 km) and offers daily and monthly temporal resolutions.

Statistical framework

The analysis proceeded in two stages, following the methodological standards of Thomson and Emery [2] and Shaw [10].

Stage 1: Descriptive statistics. For each variable: count, mean, standard deviation, minimum, maximum, median, 25th and 75th percentiles, and range.

Stage 2: Multivariate regression. Each of the four variables served in turn as the dependent variable. For each dependent variable, seven models were fitted: three single-predictor models, three two-predictor models, and one full three-predictor model, for 28

models in total. Model selection favoured the specification with the highest R and R² and the lowest RMSE, MSE, MAPE, and MABE, following the multi-metric selection practice supported by Chai and Draxler [24] and Chicco et al. [25]. For each best model, unstandardized coefficients (B), standardized coefficients (β), standard errors, t-statistics, and significance levels were evaluated; coefficient interpretation follows the cautionary framework for collinear environmental predictors of Geladi et al. [12] and Richter and Stavn [11].

Results

Descriptive statistics

Table 1. Descriptive statistics for the four oceanic variables.

Statistic	SST (°C)	uo (m/s)	vo (m/s)	so (PSU)
Count	22	22	22	47
Mean	26.586	0.142	0.041	35.997
Std. dev.	1.837	0.074	0.033	0.234
Min	24.078	0.035	0.0004	35.490
Q1	24.718	0.083	0.020	35.806
Median	26.806	0.143	0.032	36.006
Q3	28.066	0.179	0.048	36.214
Max	29.519	0.282	0.135	36.321
Range	5.441	0.247	0.135	0.831

The SST mean of 26.59 °C is characteristic of warm tropical–subtropical Atlantic waters, and the range of 5.44 °C indicates substantial thermal variability. Eastward velocity is persistently positive (mean 0.142 m s⁻¹, all values above 0), pointing to stable zonal flow, whereas northward velocity is weak (mean 0.041 m s⁻¹) with a near-zero minimum (0.0004 m s⁻¹), indicating episodes of near-stagnant meridional flow. Salinity (mean 35.997 PSU) sits within the 35–37 PSU band typical of tropical open-ocean waters, and its small standard deviation (0.234) and narrow IQR (about 0.41 PSU) suggest a stable haline structure.

Predictive models for SST

Seven models were fitted with SST as the dependent variable. The three-predictor model came out best:

Table 2. Evaluation metrics, SST as dependent variable.

Predictors	No.	R	R ²	RMSE	MSE	MAPE (%)	MABE
so	1	0.124	0.015	1.781	3.173	6.131	1.619
uo	1	0.371	0.138	1.667	2.778	5.565	1.467
vo	1	0.125	0.016	1.781	3.172	6.068	1.601
so, uo	2	0.460	0.211	1.594	2.541	5.181	1.362
so, vo	2	0.167	0.028	1.770	3.133	6.098	1.609
uo, vo	2	0.372	0.138	1.666	2.776	5.547	1.463
so, uo, vo	3	0.476	0.227	1.578	2.491	5.070	1.335

Table 3. Coefficients of the best SST model: SST = 182.056 – 4.231(so) – 20.658(uo) – 8.036(vo). (Adj. R² = 0.098; SE = 1.745.)

Predictor	B	SE	β	T	p
(Constant)	182.056	107.362	—	1.696	.107
uo	-20.658	9.594	-0.831	-2.153	.045
vo	-8.036	13.332	-0.145	-0.603	.554
so	-4.231	2.948	-0.512	-1.435	.168

Only eastward velocity contributed significantly ($\beta = -0.831$, $p = .045$). Salinity ($p = .168$) and northward velocity ($p = .554$) were not individually significant. Overall, the model explains 22.7% of the variance in SST. The observed and predicted values of sea surface temperature based on the best model are shown in fig 1.

Predictive models for eastward velocity

The three-predictor model was again best, and by a wide margin:

Table 4. Evaluation metrics, uo as dependent variable.

Predictors	No.	R	R ²	RMSE	MSE	MAPE (%)	MABE
so	1	0.786	0.618	0.0447	0.00199	37.78	0.0361
SST	1	0.371	0.138	0.0671	0.00450	58.51	0.0574
vo	1	0.396	0.157	0.0663	0.00440	61.74	0.0561
so, SST	2	0.833	0.694	0.0400	0.00160	33.51	0.0333
so, vo	2	0.844	0.712	0.0388	0.00150	34.25	0.0336
SST, vo	2	0.512	0.262	0.0620	0.00385	55.85	0.0544
so, SST, vo	3	0.878	0.771	0.0346	0.00120	27.82	0.0285

Table 5. Coefficients of the best uo model: $uo = 9.072 - 0.240(so) - 0.010(SST) - 0.628(vo)$. (Adj. $R^2 = 0.733$; $SE = 0.038$.)

Predictor	B	SE	β	t	P
(Constant)	9.072	1.358	—	6.681	.000
so	-0.240	0.038	-0.723	-6.320	.000
SST	-0.010	0.005	-0.246	-2.153	.045
vo	-0.628	0.255	-0.281	-2.459	.024

All three predictors were individually significant. Salinity dominates ($\beta = -0.723$, $p < .001$), followed by northward velocity ($\beta = -0.281$, $p = .024$) and temperature ($\beta = -0.246$, $p = .045$). The model explains 77.1% of the variance in eastward velocity, the strongest result in the study. The observed and predicted values of eastward sea velocity based on the best model are shown in fig 2.

Predictive models for northward velocity

Table 6. Evaluation metrics, vo as dependent variable (best model bolded).

Predictors	No.	R	R ²	RMSE	MSE	MAPE (%)	MABE
so	1	0.117	0.014	0.0321	0.00103	507.96	0.0233
SST	1	0.125	0.016	0.0321	0.00103	514.67	0.0228
uo	1	0.396	0.157	0.0297	0.00088	254.81	0.0202
so, SST	2	0.161	0.026	0.0319	0.00102	487.02	0.0227
so, uo	2	0.506	0.256	0.0279	0.00078	116.43	0.0185
SST, uo	2	0.397	0.158	0.0297	0.00088	253.95	0.0204
so, SST, uo	3	0.520	0.271	0.0276	0.00076	94.75	0.0191

Table 7. Coefficients of the best vo model: $vo = 3.212 - 0.085(so) - 0.002(SST) - 0.401(uo)$. (Adj. $R^2 = 0.149$; SE = 0.031.)

Predictor	B	SE	β	t	P
(Constant)	3.212	1.877	—	1.711	.104
so	-0.085	0.051	-0.569	-1.672	.112
SST	-0.002	0.004	-0.137	-0.603	.554
uo	-0.401	0.163	-0.894	-2.459	.024

Only eastward velocity was individually significant ($\beta = -0.894$, $p = .024$). The model explains 27.1% of the variance in northward velocity. The large MAPE values reflect the small magnitudes of vo, where modest absolute errors translate into large relative errors. The observed and predicted values of sea northward sea velocity based on the best model are shown in fig 3.

Predictive models for salinity

Table 8. Evaluation metrics, salinity as dependent variable.

Predictors	No.	R	R ²	RMSE	MSE	MAPE (%)	MABE
SST	1	0.124	0.015	0.216	0.0465	0.476	0.171
uo	1	0.786	0.618	0.134	0.0180	0.317	0.114
vo	1	0.117	0.014	0.216	0.0466	0.491	0.177
SST, uo	2	0.806	0.650	0.128	0.0165	0.303	0.109
SST, vo	2	0.161	0.026	0.214	0.0460	0.482	0.173
uo, vo	2	0.814	0.663	0.126	0.0159	0.299	0.108
SST, uo, vo	3	0.835	0.697	0.120	0.0143	0.277	0.0995

Table 9. Coefficients of the best salinity model: $so = 37.096 - 0.024(SST) - 2.871(uo) - 1.586(vo)$. (Adj. $R^2 = 0.647$; SE = 0.132.)

Predictor	B	SE	β	t	P
(Constant)	37.096	0.482	—	77.034	.000
SST	-0.024	0.017	-0.200	-1.435	.168
uo	-2.871	0.454	-0.954	-6.320	.000
vo	-1.586	0.949	-0.236	-1.672	.112

Eastward velocity was the only individually significant predictor of salinity ($\beta = -0.954, p < .001$). The model explains 69.7% of the variance in salinity and achieves an exceptionally low MAPE of 0.277%. The observed and predicted values of salinity based on the best model are shown in fig 4.

Summary of best models

Table 10. Best-performing predictive models.

Dependent variable	Best model	R	R ²	Adj. R ²	RMSE	Significant predictors
SST	so, uo, vo	0.476	0.227	0.098	1.578	uo (p = .045)
uo	so, SST, vo	0.878	0.771	0.733	0.035	so (p < .001), SST (p = .045), vo (p = .024)
vo	so, SST, uo	0.520	0.271	0.149	0.028	uo (p = .024)
Salinity	SST, uo, vo	0.835	0.697	0.647	0.120	uo (p < .001)

The three-predictor specification was optimal for all four dependent variables, and every one of the six metrics selected it consistently.

Discussion

A hierarchical predictive structure

The results reveal a pronounced asymmetry in how predictive information is distributed across the variable system. Eastward velocity is at once the best-predicted variable ($R^2 = 0.771$, with all predictors significant) and the strongest individual predictor of the other three, significant in the SST, vo, and salinity models. This dual centrality marks zonal circulation as the structural hub of the upper-ocean variable system: it is the variable most coherently determined by the thermohaline state, and the one whose variations most strongly propagate to salinity and temperature.

The hierarchy makes physical sense. The dominance of salinity in predicting eastward velocity ($\beta = -0.723, p < .001$) echoes the density-driven framework of Rahmstorf [5]: salinity gradients, through their effect on density, modulate the pressure fields that drive zonal flow. The strong reverse pathway, eastward velocity predicting salinity ($\beta = -0.954, p < .001$), reflects advective control of the haline field, consistent with the satellite-derived tropical Atlantic salinity variability documented by Tzortzi et al. [8]. Taken together, the two pathways form a salinity–velocity feedback loop: density gradients drive the flow, and the flow redistributes the density-determining fields [4, 5].

Interpreting the SST and vo models: partial predictability

The modest R^2 values for SST (0.227) and northward velocity (0.271) are informative rather than disappointing. Deser et al. [1] emphasized that SST is governed by processes beyond the ocean interior, including air–sea heat fluxes, atmospheric forcing, and remote teleconnections [18, 19], none of which appear among the three predictors. Meridional velocity in the tropical Atlantic is likewise strongly wind-forced and controlled by basin-scale pressure gradients [15, 21]. In both cases the significant residual association with eastward velocity ($\beta = -0.831$ and $\beta = -0.894$, respectively) confirms that zonal advection carries a genuine predictive signal [6, 7], while the unexplained variance quantifies the share of these fields that processes external to the analyzed variable set control.

The adjusted R^2 of the SST model (0.098) calls for particular caution. With three predictors and 22 observations, the fit is modest, and the model is best read as evidence that uo carries SST-relevant information, not as a complete predictive account of SST. Reporting F-statistics and formal collinearity diagnostics (VIF) for each model would strengthen the inferential basis further.

Collinearity and sign reversals

A distinctive feature of the results is the consistent negative sign of predictor coefficients across models. Higher salinity, higher eastward velocity, and higher northward velocity are all associated with lower predicted values of the dependent variables in the

fitted models, even though the underlying physical system contains well-known positive covariations (e.g., the T–S relationship [14]). This pattern is a classical signature of strong predictor intercorrelation, in which coefficients absorb the joint variance structure and individual signs become unstable [12]. The sign reversals should therefore be read jointly with the bivariate structure rather than causally; Model II regression approaches that account for error in all variables [11] are a natural robustness extension.

Evaluation metrics: magnitudes and their interpretation

The multi-metric selection framework performed as intended: the three-predictor models were selected unanimously by all six metrics for all four dependent variables, so the architecture conclusions hold regardless of metric choice [24, 25]. The metric magnitudes themselves deserve a careful reading. The salinity model's MAPE of 0.277% reflects both a strongly determined predictand and a very low-variability field ($SD = 0.234$ PSU). The v_o model's very large MAPE (94.7% at best) is a small-denominator effect: with a mean v_o of 0.041 m s^{-1} , percentage errors are intrinsically large, so it is not evidence of a worse fit in absolute terms (its RMSE, 0.0276 m s^{-1} , is small). This contrast is exactly why MAE/RMSE and MAPE should be reported jointly rather than substituted for one another [24].

Relation to data-driven ocean modelling

The hierarchy identified here has direct design implications for data-driven and hybrid ocean models [16]. Predictor selection matters. Models predicting SST should include current velocity, but the present results suggest that thermohaline variables alone explain only a quarter of its variance, so SST data-driven systems need atmospheric predictors to close the gap [17, 28]. Eastward velocity, conversely, is an attractive predictand for reanalysis-based learning systems: 77% of its variance is determined by the other three variables, which makes it a strong candidate for gap-filling and quality-control applications in reanalysis pipelines [13].

Broader stakes

Because the analyzed variables jointly define the buoyancy and advective state of the upper ocean, their predictive structure bears on downstream applications ranging from climate variability assessment to ecosystem state [20, 29]. The equation-level results of this study can be used directly as statistical transfer functions where full dynamical models are unavailable.

Limitations

Five limitations qualify these findings. The sample sizes (22 for the temperature/velocity models; 47 for salinity) limit statistical power, particularly for the SST model, where adjusted R^2 is near zero. The study is observational and correlational, so the fitted equations describe predictive association, not causal mechanism. Formal collinearity diagnostics (VIFs) and ANOVA F-tests were not part of the reported output and remain a priority for follow-up work. Only linear specifications were fitted, and non-linear relationships documented in the oceanographic literature [7, 12] may raise the achievable predictive ceiling. Finally, study-area and period metadata must be supplied to situate the findings geographically and seasonally.

Conclusions

This study conducted a symmetric multivariate regression analysis of four core upper-ocean variables in the Atlantic Ocean, namely SST, salinity, and the eastward and northward components of current velocity, fitting 28 regression models and selecting the best specification for each dependent variable through a six-metric framework. Four principal conclusions follow.

1. **The predictive structure is hierarchical, with eastward velocity at its center.** Zonal velocity is the best-predicted variable ($R^2 = 0.771$, all predictors significant) and the only significant predictor in the SST, northward velocity, and salinity models. It acts at once as the principal conduit of predictability and as the variable most coherently governed by the thermohaline state.
2. **Salinity and eastward velocity form a bidirectional coupling.** Salinity is the strongest predictor of u_o ($\beta = -0.723$, $p < .001$) and u_o the dominant predictor of salinity ($\beta = -0.954$, $p < .001$), a feedback loop consistent with density-driven circulation theory [5] and advective haline redistribution [8].
3. **SST and meridional velocity are only partially determined by the internal variable set.** With $R^2 = 0.227$ and 0.271 respectively, the models quantify the share of these fields governed by processes external to the analyzed system, such as atmospheric forcing and basin-scale dynamics.
4. **Multi-metric, specification-sweeping model selection is essential practice for oceanographic regression.** The unanimous selection of three-predictor models across all six metrics demonstrates robustness, while the widely varying MAPE magnitudes

(0.28%–94.7%) show that single-metric selection would be misleading.

Future work should extend the analysis with collinearity diagnostics and ANOVA reporting, larger and spatially distributed samples, non-linear and machine learning specifications [16, 17], and the inclusion of atmospheric predictors to close the SST predictive gap. The resulting equation-level framework can serve as a baseline for data-driven Atlantic Ocean forecasting systems and as a transfer-function toolkit for reanalysis gap-filling applications [13, 28].

Abbreviations

AMOC: Atlantic Meridional Overturning Circulation; ANOVA: analysis of variance; GLORYS: Global Ocean Reanalysis and Simulation; IQR: interquartile range; MABE: mean absolute bias error; MAE: mean absolute error; MAPE: mean absolute percentage error; MSE: mean squared error; PSU: practical salinity unit; R: correlation coefficient; R²: coefficient of determination; RMSE: root mean squared error; SD: standard deviation; SE: standard error; so: sea water salinity; SST: sea surface temperature; T–S: temperature–salinity; uo: eastward ocean current velocity; vo: northward ocean current velocity; VIF: variance inflation factor.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in the Copernicus Marine Data Store repository (Global Ocean Physics Reanalysis, GLORYS12V1), https://data.marine.copernicus.eu/products/GLOBAL_MULTIYEAR_PHY_001_030.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

MDU conceived and designed the study, carried out the analysis, and drafted the manuscript. UFE and JOE contributed to the interpretation of the results and revised the manuscript critically for intellectual content. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Deser C, Alexander MA, Xie S-P, Phillips AS. Sea surface temperature variability: patterns and mechanisms. *Annu Rev Mar Sci*. 2010;2(1):115-143. <https://doi.org/10.1146/annurev-marine-120408-151453><https://doi.org/10.1016/j.jmarsys.2008.05.009>

2. Thomson RE, Emery WJ. Data analysis methods in physical oceanography. 3rd ed. Amsterdam: Elsevier Science; 2014.<https://doi.org/10.1029/2007JC004435><https://doi.org/10.1029/2007JC004435>
3. Cunningham SA, Kanzow T, Rayner D, Baringer MO, Johns WE, Marotzke J, et al. Temporal variability of the Atlantic Meridional Overturning Circulation at 26.5°N. *Science*. 2007;317(5840):935-938.
<https://doi.org/10.1126/science.1141304><https://doi.org/10.1002/2015RG000493><https://doi.org/10.1002/2015RG000493>
4. Buckley MW, Marshall J. Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: a review. *Rev Geophys*. 2016;54(1):5-63. <https://doi.org/10.1002/2015RG000493><https://doi.org/10.5194/gmd-7-1247-2014><https://doi.org/10.5194/gmd-7-1247-2014>
5. Rahmstorf S. Ocean circulation and climate during the past 120,000 years. *Nature*. 2002;419(6903):207-214.
<https://doi.org/10.1038/nature01090><https://doi.org/10.7717/peerj-cs.623><https://doi.org/10.7717/peerj-cs.623>
6. Brandt P, Hormann V, Bourlès B, Fischer J, Schott FA, Stramma L, et al. Oxygen tongues and zonal currents in the equatorial Atlantic. *J Geophys Res Oceans*. 2008;113(4):C04012.
<https://doi.org/10.1029/2007JC004435><https://doi.org/10.1016/j.dsr2.2023.105262><https://doi.org/10.1016/j.dsr2.2023.105262>
7. Tuchen FP, Brandt P, Hahn J, Hummels R, Dengler M. Modulation of equatorial currents and Tropical Instability Waves during the 2021 Atlantic Niño. *J Geophys Res Oceans*. 2023;128(6):e2022JC020011.
<https://doi.org/10.1029/2022JC020011><https://doi.org/10.48670/moi-00021><https://doi.org/10.48670/moi-00021>
8. Tzortzi E, Josey SA, Srokosz M, Gommenginger C. Tropical Atlantic salinity variability: new insights from SMOS. *Geophys Res Lett*. 2013;40(10):2143-2147.
<https://doi.org/10.1002/grl.50225><https://doi.org/10.1126/science.1141304><https://doi.org/10.1126/science.1141304>
9. Allen JI, Somerfield PJ. A multivariate approach to model skill assessment. *J Mar Syst*. 2009;76(1-2):83-94.
<https://doi.org/10.1016/j.jmarsys.2008.05.009><https://doi.org/10.1146/annurev-marine-120408-151453><https://doi.org/10.1146/annurev-marine-120408-151453>
10. Shaw PJA. Multivariate statistics for the environmental sciences. London: Hodder Arnold; 2003.[https://doi.org/10.1016/S0169-7439\(98\)00204-4](https://doi.org/10.1016/S0169-7439(98)00204-4)[https://doi.org/10.1016/S0169-7439\(98\)00204-4](https://doi.org/10.1016/S0169-7439(98)00204-4)
11. Richter SJ, Stavn RH. Determining functional relations in multivariate oceanographic systems: Model II multiple linear regression. *J Atmos Oceanic Technol*. 2014;31(7):1663-1672. <https://doi.org/10.1175/JTECH-D-13-00210.1><https://doi.org/10.1093/icesjms/fsae029><https://doi.org/10.1093/icesjms/fsae029>
12. Geladi P, Hadjiiski L, Hopke P. Multiple regression for environmental data: nonlinearities and prediction bias. *Chemom Intell Lab Syst*. 1999;47(2):165-173. [https://doi.org/10.1016/S0169-7439\(98\)00204-4](https://doi.org/10.1016/S0169-7439(98)00204-4)[https://doi.org/10.1016/S0169-7439\(98\)00204-4](https://doi.org/10.1016/S0169-7439(98)00204-4)
13. Lellouche J-M, Greiner E, Le Galloudec O, Garric G, Regnier C, Dreviron M, et al. Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system. *Ocean Sci*. 2018;14(5):1093-1126.
<https://doi.org/10.5194/os-14-1093-2018><https://doi.org/10.1029/2021JC017373><https://doi.org/10.1029/2021JC017373>
14. Kido S, Nonaka M, Tanimoto Y. Impacts of salinity variation on the mixed-layer processes and sea surface temperature-salinity covariability. *Geophys Res Lett*. 2021;48(24):e2021GL096010.
<https://doi.org/10.1029/2021GL096010><https://doi.org/10.1029/2021GL096010><https://doi.org/10.1029/2021GL096010>
15. Schott FA, McCreary JP, Johnson GC. Shallow pathways of the Atlantic subtropical cells. *J Phys Oceanogr*. 2004;34(4):1783-1797. <https://doi.org/10.1175/2408.1><https://doi.org/10.5194/os-14-1093-2018><https://doi.org/10.5194/os-14-1093-2018>
16. Sonnewald M, Lguensat R, Jones DC, Dueben PD, Brajard J, Balaji V. Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environ Res Lett*. 2021;16(7):073008. <https://doi.org/10.1088/1748-9326/ac0eb0><https://doi.org/10.1038/s41558-020-00918-2><https://doi.org/10.1038/s41558-020-00918-2>
17. Patil K, Deo MC. Basin-scale prediction of sea surface temperature with artificial neural networks. *J Atmos Oceanic Technol*. 2018;35(7):1441-1455. <https://doi.org/10.1175/JTECH-D-17-0217.1><https://doi.org/10.1175/JTECH-D-17-0217.1><https://doi.org/10.1175/JTECH-D-17-0217.1>
18. Xie S-P, Carton JA. Tropical Atlantic variability: patterns, mechanisms, and climate impacts. In: Wang C, Xie S-P, Carton JA, editors. *Earth's climate: the ocean-atmosphere system*. Washington: American Geophysical Union; 2004. p. 121-142. (Geophysical Monograph Series, vol 147). <https://doi.org/10.1029/147GM07><https://doi.org/10.1007/s00382-013-1968-5><https://doi.org/10.1007/s00382-013-1968-5>

19. Wang C. Atlantic climate variability and its associated atmospheric circulation cells. *J Clim*. 2002;15(13):1516-1536. [https://doi.org/10.1175/1520-0442\(2002\)015<1516:ACVAIA2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1516:ACVAIA2.0.CO;2)<https://doi.org/10.1038/nature01090><https://doi.org/10.1038/nature01090>
20. Li G, Cheng L, Zhu J, Trenberth KE, Mann ME, Abraham JP. Increasing ocean stratification over the past half-century. *Nat Clim Change*. 2020;10(12):1116-1123. <https://doi.org/10.1038/s41558-020-00918-2><https://doi.org/10.1175/JTECH-D-13-00210.1><https://doi.org/10.1175/JTECH-D-13-00210.1>
21. Perez RC, Hormann V, Lumpkin R, Brandt P, Bourlès B, Araujo M, et al. Mean meridional currents in the central and eastern equatorial Atlantic. *Clim Dyn*. 2014;43(11):2943-2962. <https://doi.org/10.1007/s00382-013-1968-5><https://doi.org/10.1175/2408.1><https://doi.org/10.1175/2408.1>
22. Hsu P-C, Centurioni L, Shao H. Surface current variations and oceanic fronts in the Southern East China Sea: drifter experiments, coastal radar applications, and satellite observations. *J Geophys Res Oceans*. 2021;126(10):e2021JC017373. <https://doi.org/10.1029/2021JC017373>
23. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean squared error (RMSE) in assessing average model performance. *Clim Res*. 2005;30(1):79-82. <https://doi.org/10.3354/cr030079><https://doi.org/10.1088/1748-9326/ac0eb0><https://doi.org/10.1088/1748-9326/ac0eb0>
24. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci Model Dev*. 2014;7(3):1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014><https://doi.org/10.1016/C2010-0-66512-1><https://doi.org/10.1016/C2010-0-66512-1>
25. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021;7:e623. <https://doi.org/10.7717/peerj-cs.623><https://doi.org/10.1029/2022JC020011><https://doi.org/10.1029/2022JC020011>
26. Copernicus Marine Service. Global Ocean Physics Reanalysis (GLORYS12V1), Version 1 [dataset]. Toulouse: Mercator Ocean International; 2023. <https://doi.org/10.48670/moi-00021><https://doi.org/10.1002/grl.50225><https://doi.org/10.1002/grl.50225>
27. Gordon RL. Acoustic Doppler current profiler principles of operation: a practical primer. 2nd ed. San Diego: RD Instruments; 1996.[https://doi.org/10.1175/1520-0442\(2002\)015](https://doi.org/10.1175/1520-0442(2002)015)[https://doi.org/10.1175/1520-0442\(2002\)015](https://doi.org/10.1175/1520-0442(2002)015)
28. Choi H-M, Kim M-K, Yang H. Deep-learning model for sea surface temperature prediction near the Korean Peninsula. *Deep Sea Res Part II Top Stud Oceanogr*. 2023;208:105262. <https://doi.org/10.1016/j.dsr2.2023.105262><https://doi.org/10.1029/147GM07><https://doi.org/10.1029/147GM07>
29. González-Pola C, Fratantoni P, Larsen KMH, Beszczynska-Möller A, Birchenough SNR. Decadal (2010-2019) variability in the marine ecosystems of the North Atlantic. *ICES J Mar Sci*. 2024;81(3):505-520. <https://doi.org/10.1093/icesjms/fsae029>

Figures

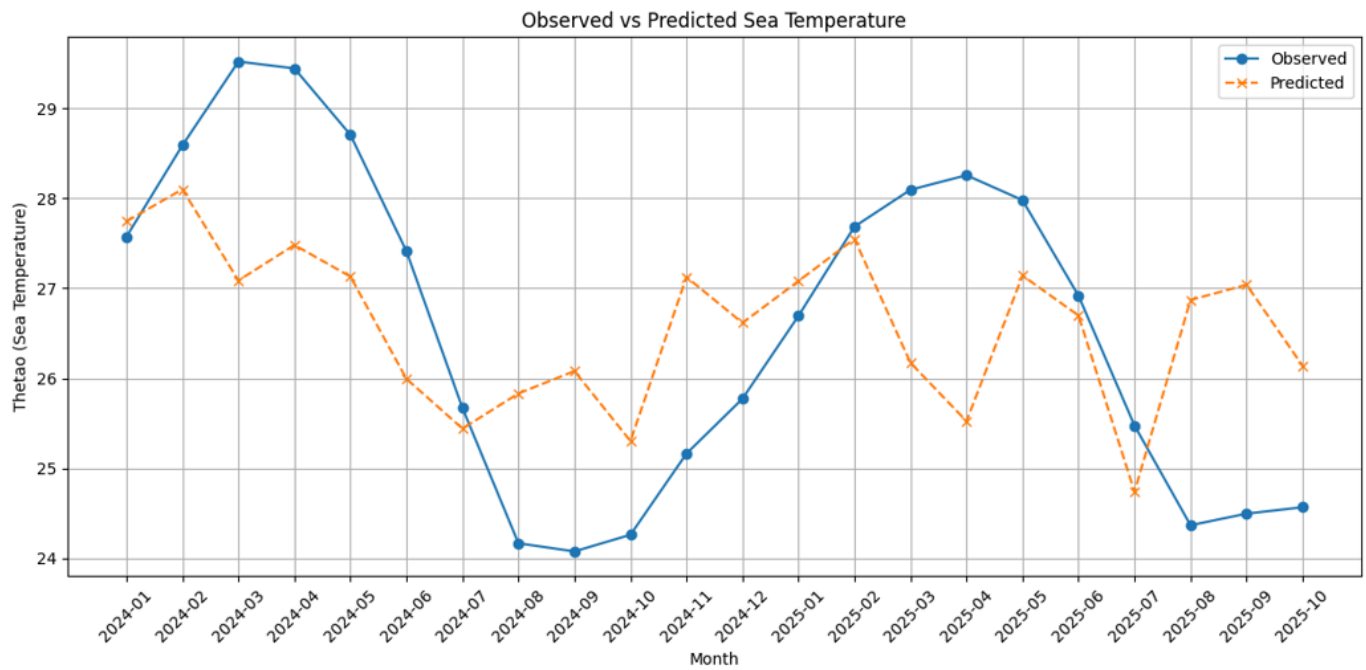


Figure 1

Observed and predicted values for sea surface temperature.

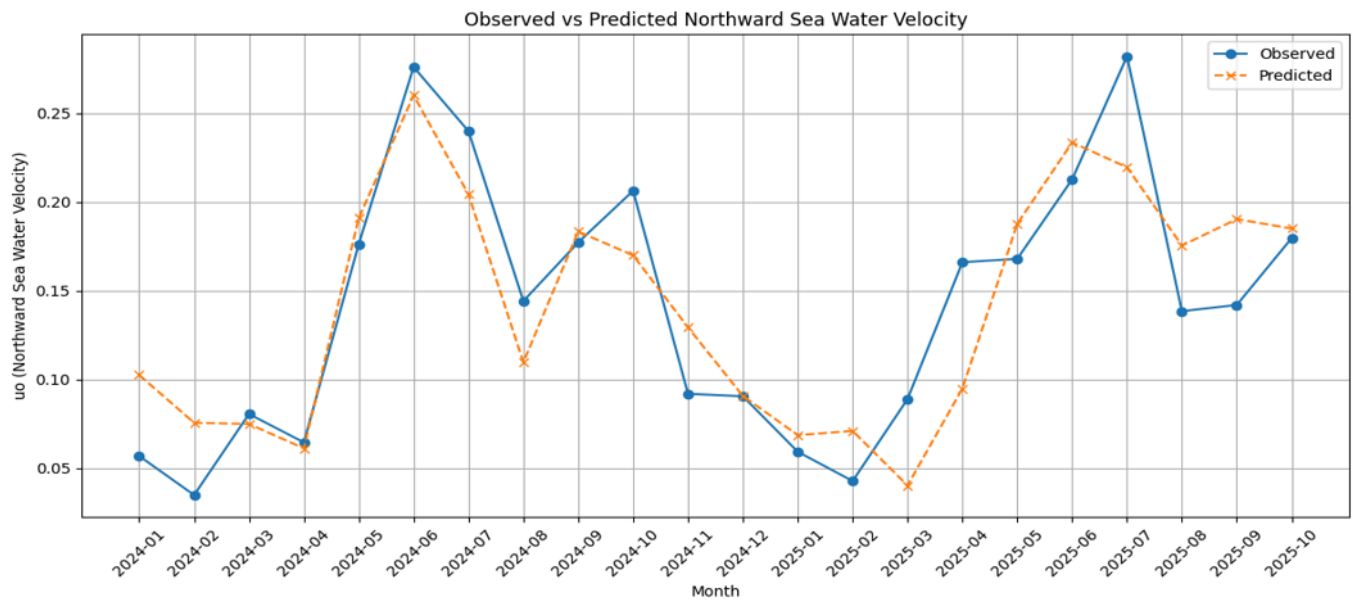


Figure 2

Observed and predicted values for eastward ocean current velocity (u_o).

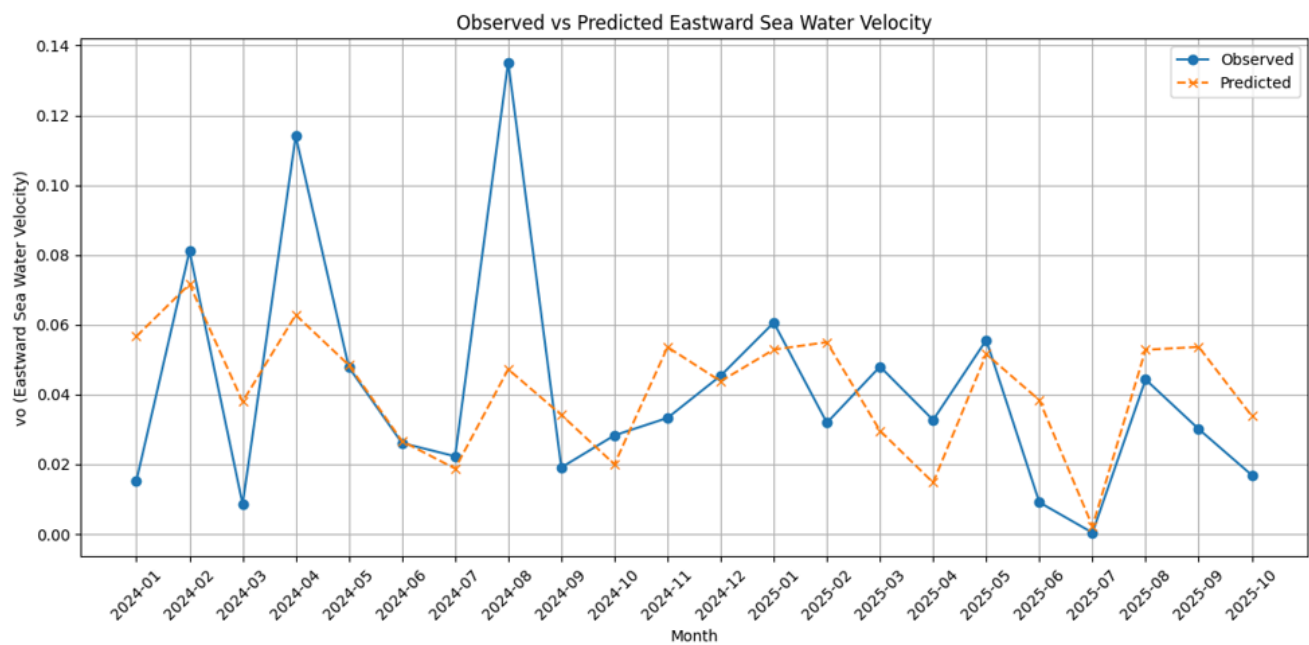


Figure 3

Observed and predicted values for northward ocean current velocity (v_o).

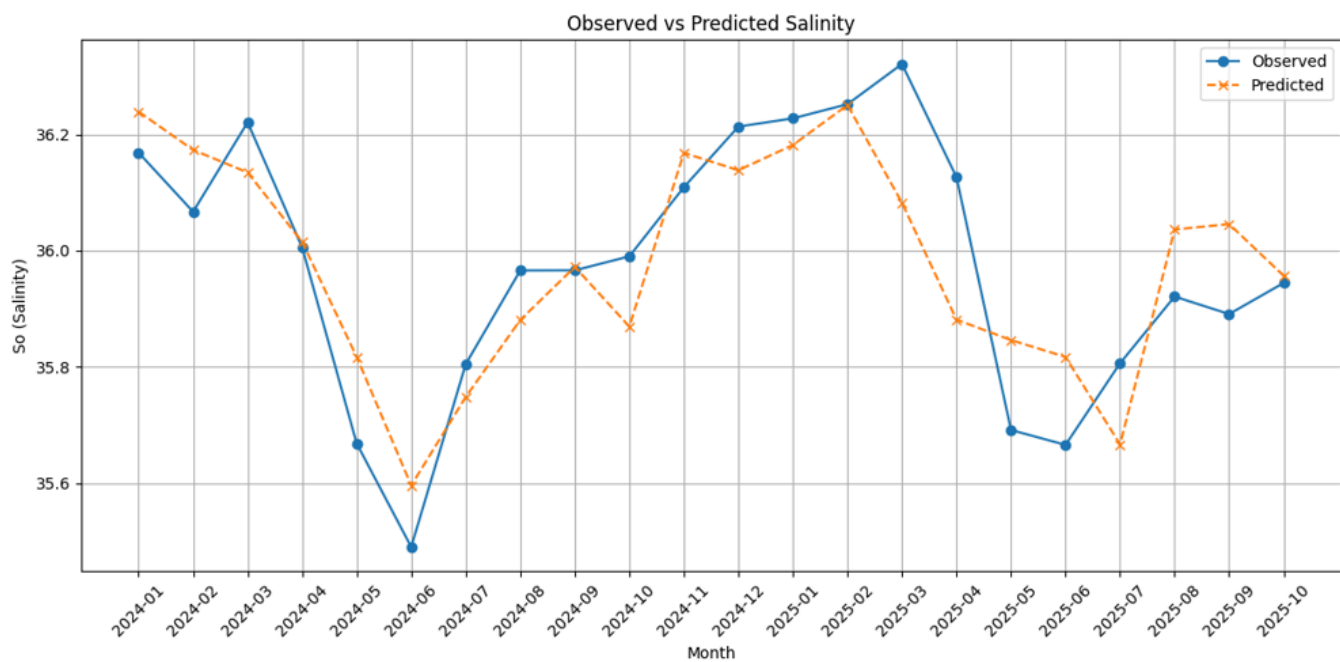


Figure 4

Observed and predicted values for sea surface salinity.