https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Open Access Earth System Science Data Discussions

Barium in seawater:

Dissolved distribution, relationship to silicon, and barite saturation state determined using machine learning

Öykü Z. Mete[1,2,3,4,*], Adam V. Subhas[2], Heather H. Kim[2], Ann G. Dunlea[2], Laura M. Whitmore[5], Alan M. Shiller[6], Melissa Gilbert[6], William D. Leavitt[3,7], and Tristan J. Horner[1,2,*]

[1]NIRVANA Laboratories; [2]Department of Marine Chemistry & Geochemistry; Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA; [3]Department of Earth Sciences, Dartmouth College, Hanover, NH 03755, USA; [4]Now at: Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA; [5]International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK 99775, USA; [6]School of Ocean Science and Engineering, University of Southern Mississippi, Stennis Space Center, MS 39529, USA; [7]Department of Chemistry, Dartmouth College, Hanover, NH 03755, USA

*Corresponding author: omete@fas.harvard.edu or Tristan.Horner@whoi.edu

Abstract

Barium is widely used as a proxy for dissolved nutrients and particulate organic carbon fluxes in seawater. However, these proxy applications are limited by insufficient knowledge of the dissolved distribution of Ba ([Ba]). For example, there is significant spatial variability in the Ba–Si relationship, and ocean chemistry may influence sedimentary Ba preservation. To help address these issues, we developed 4,095 models for predicting [Ba] using Gaussian Progress Regression Machine Learning. These models were trained to predict [Ba] from standard oceanographic observations using GEOTRACES data from the Arctic, Atlantic, Pacific, and Southern Oceans. Trained models were then validated by comparing predictions against withheld [Ba] data from the Indian Ocean. We find that a model using depth, T, S, [O_2], [PO_4], and [NO_3] as predictors can accurately predict [Ba] in the Indian Ocean with a mean absolute percentage deviation of 6.3 %. We use this model to simulate [Ba] on a global basis using these same six predictors in the World Ocean Atlas. The resulting [Ba] distribution constrains the total Ba budget of the ocean to 122±8 $\times 10^{12}$ mol and clarifies the global relationship between dissolved Ba and Si. We also calculate the saturation state of seawater with respect to barite, revealing that the ocean below 1,000 m is, on average, at or near saturation. We describe a number of possible applications for our model output, ranging from use in biogeochemical models to paleoproxy calibration. Our approach could be extended to other trace elements with relatively minor adjustments and demonstrates the utility of machine learning to accurately simulate the distributions of tracers in the sea.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

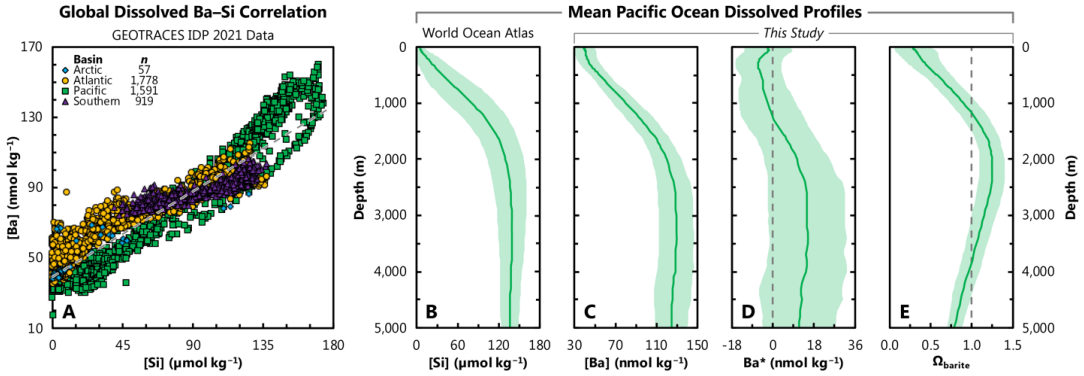
Open Access Earth System Science Data Discussions

1. Introduction

Barium (Ba) is a Group II trace metal that is widely applied in studies of modern and ancient marine biogeochemistry, despite lacking a recognized biochemical function (e.g., Horner & Crockford, 2021). These applications of Ba are based on two empirical correlations relating to its dissolved and particulate cycles. The first correlation relates to the dissolved concentration of Ba, hereafter [Ba], which is strongly correlated with that of the algal nutrient silicon (Si; as dissolved silicic acid; Fig. 1; Chan et al., 1977). Unlike [Si], ambient [Ba] concentrations are faithfully recorded by a number of marine carbonates, such as planktonic (e.g., Hönisch et al., 2011) and benthic foraminifera (e.g., Lea & Boyle, 1990), surface- (e.g., Gonneea et al., 2017) and deep-sea corals (e.g., Anagnostou et al., 2011; LaVigne et al., 2011), and mollusks (e.g., Komagoe et al., 2018). Preservation of these signals means that the Ba content of carbonates can be related to the Ba content of seawater and, by extension, that of Si. Accordingly, the Ba–Si proxy has been applied to understand ocean nutrient dynamics on decadal (e.g., Lea et al., 1989) to millennial timescales (e.g., Stewart et al., 2021).

The nutrient-like distribution of dissolved Ba in seawater is thought to be sustained by the second empirical correlation, relating to cycling of particulate Ba. Particulate Ba in seawater occurs mostly in the form of discrete, micron-sized crystals of the mineral barite ($BaSO_4$(s), barium sulfate; e.g., Dehairs et al., 1980; Stroobants et al., 1991). Pelagic $BaSO_4$ is an ubiquitous component of marine particulate matter (e.g., Light & Norris, 2021) and constitutes the principal removal flux of dissolved Ba from seawater (Paytan & Kastner, 1996). Pelagic $BaSO_4$ is thought to precipitate within ephemeral particle-associated microenvironments that develop during the microbial oxidation of sinking organic matter (e.g., Chow & Goldberg, 1960; Bishop, 1988). The flux of particulate $BaSO_4$ to the seafloor is correlated with the flux of exported organic matter (e.g., Dymond et al., 1992; Eagle et al., 2003; Serno et al., 2014; Hayes et al., 2021). This correlation means that the accumulation rate of sedimentary $BaSO_4$—or its main constituent, Ba—can be used to trace patterns of past organic matter export on timescales ranging from millenia to millions of years (e.g., Bains et al., 2000; Paytan & Griffith, 2007; Schmitz, 1987; Schroeder et al., 1997).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023



59 **Figure 1. Distribution of barium in seawater. A.** Property–property plot showing the 4,345 co-located,
60 core-feature complete dissolved data used in ML model training (Sect. 2). Sample locations shown in Figure
61 2. Dashed line shows best-fit linear regression through these data, whereby [Ba] = 0.54·[Si] + 39.3. Panels
62 **B., C., D.,** and **E.** show average Pacific Ocean dissolved depth profiles of [Si], [Ba], Ba*, and Ω_{barite},
63 respectively. Solid line denotes the arithmetic mean and the shaded region encompasses one standard
64 deviation either side of the mean.

65 While the Ba-based proxies are valuable, their applications are potentially limited by insufficient
66 knowledge of the distribution of [Ba]. For example, there is significant vertical and spatial
67 variability in the Ba–Si relationship (Sect. 3.3.; Fig.), which we quantify using Ba* (e.g., Horner
68 et al., 2015):

69 $$Ba^* = [Ba]_{in\ situ} - [Ba]_{predicted} \quad \text{[Eq. 1]}$$

70 where $[Ba]_{predicted}$ is based on the Ba–Si linear regression (Fig. 1):

71 $$[Ba]_{predicted} = 0.54 \cdot [Si]_{in\ situ} + 39.3 \quad \text{[Eq. 2]}$$

72 Here, $[Si]_{in\ situ}$ has units of µmol kg^{-1} and $[Ba]_{predicted}$ nmol kg^{-1}; therefore, Ba* also has units of
73 nmol kg^{-1}. The vertical profile of Ba* is rarely conservative (Fig. 1) and these variations could
74 introduce uncertainty in the reconstruction of [Si] using Ba.

75 The relationship between sedimentary $BaSO_4$ accumulation rates and productivity also contains a
76 significant degree of scatter (e.g., Serno et al., 2014; Hayes et al., 2021). Some of this scatter may
77 relate to variability in $BaSO_4$ preservation, which is at least partially sensitive to ambient saturation

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

78 state, Ω_{barite} (e.g., Schenau et al., 2001; Fig. 1). The saturation state of a parcel of water with respect
79 to $BaSO_4$ is defined as:

80 $$\Omega_{barite} = Q / K_{sp} \quad \text{[Eq. 3]}$$

81 where Q is the Ba and sulfate ion product and K_{sp} is the *in situ* $BaSO_4$ solubility product. Discerning
82 the importance of Ω_{barite} on $BaSO_4$ preservation has hitherto been challenging owing to the sparsity
83 of *in situ* [Ba] measurements. Accurately determining the global distribution of [Ba] would be
84 valuable for geochemists and oceanographers, and would enable a more thorough investigation of
85 the effects of preservation on $BaSO_4$ fluxes and refinement of the Ba–Si nutrient proxy.

86 A powerful way of interrogating oceanic element distributions is through modeling. Broadly, there
87 are two modeling approaches relevant for simulating [Ba]: mechanistic (i.e., theory driven) and
88 statistical modeling (i.e., data driven; e.g., Glover et al., 2011). Mechanistic or process-based
89 modeling is generally viewed as the gold-standard approach; model outputs are derived from sets
90 of underlying equations that are based on fundamental theory. As such, mechanistic model outputs
91 can be interrogated to obtain understanding of processes and their sensitivities. However, creating
92 a mechanistic model of the marine Ba cycle requires embedding a biogeochemical model of $BaSO_4$
93 cycling within a computationally expensive global circulation model. Although the computational
94 cost associated with building mechanistic models has been reduced by the development of ocean
95 circulation inverse models (e.g., DeVries, 2014; John et al., 2020), this approach still requires
96 detailed parametrizations of the marine Ba cycle, which do not currently exist. In contrast,
97 statistical models are based on extracting patterns from existing data and using those relationships
98 to make predictions. Statistical models encompass a wide variety of approaches ranging from
99 regression analysis to machine learning (ML). Of particular interest to our study are ML models,
100 which can make predictions without any explicit parameterizations of causal relationships.
101 Machine learning models are computationally efficient and can be highly accurate, though they
102 offer limited interpretability. Machine learning is increasingly being used to solve problems in
103 Earth and environmental sciences, including simulating the dissolved distribution of tracers in the
104 sea (e.g., cadmium, Roshan & DeVries, 2021; iodine, Sherwen et al. 2019; nitrogen isotopes of
105 nitrate, Rafter et al., 2019).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

106 The goal of this study is to obtain an accurate simulation of [Ba], which ML makes possible even
107 in the absence of a process-level understanding of the marine Ba cycle. We tested thousands of
108 ML models that were trained using quality-controlled GEOTRACES data from the Arctic,
109 Atlantic, Pacific, and Southern Oceans, supplemented by Argo, satellite chlorophyll, and
110 bathymetry data products (Sect. 2.). Models were tested for their accuracy by simulating [Ba] in
111 the Indian Ocean and comparing predictions against observations made between 1977–2013.
112 Importantly, no Indian Ocean data were seen by any of the models during training (Sect. 2.). From
113 this, we identify the optimal set of predictor variables that results in the most accurate estimates of
114 [Ba], calculate model uncertainties, and simulate [Ba], Ba*, and Ω_{barite} on a global basis (Sect. 5.).
115 This result will be valuable for researchers interested in marine Ba cycling, and demonstrates the
116 utility of ML to tackle problems in marine biogeochemistry.

117 2. Training and testing data

118 Machine learning algorithms are adept at making accurate predictions of a target variable by
119 identifying relationships between variables within large data sets. However, making accurate
120 predictions first requires that a ML algorithm is trained on existing observations of that variable
121 alongside a number of other parameters. These other parameters, hereafter termed features, are an
122 important part of model training; features should encode information that may help the ML
123 algorithm predict [Ba], otherwise their inclusion may diminish model performance. Features
124 should also be well characterized in the global ocean, which allows ML models to make predictions
125 in regions beyond the initial training dataset. We selected 12 model features by considering the
126 tradeoff between feature availability and presumed predictive power (Table 1). While testing more
127 features may have resulted in a more accurate final model, we found that many observations of
128 [Ba] did not have corresponding data for several features. Thus, including more features would
129 have meant fewer training data. In subsequent sections, we find that only 4–7 features are needed
130 to accurately predict [Ba]. As such, we did not evaluate the predictive power of other predictors
131 beyond the initial feature set.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Table 1. List of oceanographic parameters chosen as model features. The features tested were selected based on their presumed predictive power and geospatial coverage.

#	Parameter Name	Abbreviation	Units	Coverage*
1	Latitude	Lat.	degrees north (°N)	–
2	Longitude	Long.	degrees east (°E)	–
3	Sample collection depth	z	meters (m)	–
4	Temperature	T	degrees Celsius (°C)	97.44%
5	Salinity	S	unitless, but often written in 'units' of PSU or PSS	97.44%
6	Dissolved oxygen	$[O_2]$	μmol kg^{-1}	97.44%
7	Dissolved nitrate	$[NO_3^-]$	μmol kg^{-1}	97.44%
8	Dissolved phosphate	$[PO_4^{3-}]$	μmol kg^{-1}	97.44%
9	Dissolved silicon (as silicic acid)	$[Si(OH)_4]$	μmol kg^{-1}	97.44%
10	Maximum monthly mean mixed-layer depth	MLD	meters (m)	88.20%
11	Mean average annual surface chlorophyll	Chl *a*	mg m^{-3}	93.95%
12	Bathymetry	Bathy.	meters (m)	100%

*Coverage values represent the percentage of data points within the World Ocean Atlas 2018 grid that have available data for a given parameter. Latitude, longitude, and depth have 100 % coverage as these features define the grid itself.

The 12 features used to predict [Ba] and their associated data sources are summarized in Table 1 and described below. The first three features (latitude, longitude, depth) record geospatial information that defines the location of an observation in three-dimensional space. To avoid numerical discontinuities, latitude and longitude were introduced into the model as a hyperparameter consisting of the cosine and sine of their respective values (in radians). Data for features 1–3 were included in the sample metadata. Features 4–9 encode physical (temperature, salinity) and chemical (oxygen, nutrients) information that is routinely measured alongside [Ba]. These data were generally available for the same bottle as the [Ba] measurements; however, when that was not the case, nutrient data were taken from the corresponding location during a separate

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

cast, or, in the case of oxygen, from linearly interpolated sensor data. The final three features are independent of depth, meaning that all samples within a given vertical profile exhibit the same value for MLD (mixed-layer depth), sea-surface chlorophyll *a*, and bathymetry. Features 10–12 were drawn from several data sources. A climatology of MLD (feature 10) was compiled using the Argo database (Holte et al., 2017). We selected maximum monthly mean MLD as the feature of interest, as this appears to be the spatiotemporal scale most relevant for influencing [Ba] distributions (Bates et al., 2017). Feature 11 represents a blended SeaWiFS and MODIS climatology of chlorophyll *a* that was obtained from the Copernicus Marine Environment Monitoring Service (CMEMS, 2021). We calculated the mean annual chlorophyll *a* for each grid cell in the data product and log transformed the data to reduce parameter weighting (e.g., Rafter et al., 2019). Data for MLD and chlorophyll *a* were extracted at the location of [Ba] observations using nearest-neighbor interpolation and their values logged in the master record. Bathymetric information (feature 12) was extracted from one of two sources. Our preferred source was the sample metadata, which generally included a value for bathymetry. For samples lacking bathymetric information, we used nearest-neighbor interpolation to extract a value from the *ETOPO5* Global Relief Model (National Geophysical Data Center, 1993). Occasionally, the *ETOPO5*-extracted bathymetry was shallower than the deepest observation of [Ba] in a given vertical profile. In such cases, the bathymetry logged in the master record was set to 1.01 times the depth of the deepest observation in that profile.

The [Ba] data from the Indian Ocean were collected from a multitude of primarily pre-GEOTRACES sources (Table 2). As such, these data were generally incomplete for the 12 features used to train the ML models. Rather than using a mixture of *in situ* and interpolated data, we decided to interpolate all Indian Ocean data for parameters 4–12. Data for parameters 4–9 were linearly interpolated from the nearest vertical profile in the World Ocean Atlas 2018 (WOA; Boyer et al., 2018; García et al., 2018a; 2018b; Locarnini et al., 2018; Zweng et al., 2018) and values for MLD and chlorophyll *a* were extracted from the aforementioned data products using nearest-neighbor interpolation. Bathymetric information was obtained from either the WOA or *ETOPO5*. For the vast majority of most samples, bathymetry was taken as the arithmetic mean of the maximum depth of the nearest vertical profile in the WOA and the depth at the standard level below. For example, if the maximum depth at a station was 950 m, the bathymetry was recorded as 975 m, which is the mean of levels 46 (950 m) and 47 (1,000 m). For profiles with a maximum

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

174 depth of 5,500 m—level 102, the lowest in the WOA—bathymetry was recorded as either 5,550
175 m or the nearest-neighbor interpolated value from *ETOPO5*, whichever was deeper.

176 **Table 2. Data sources.** Information regarding the source of [Ba] incorporated into the master record.

Purpose	Region	Expedition ID	Data source	Data Originators (if unpublished)
Model training	South Atlantic (Meridional)	GA02	GEOTRACES IDP 2017 (Schlitzer et al., 2018)	Jose M. Godoy
	North Atlantic (Zonal)	GA03	Rahman et al., 2022	
	South Atlantic (Zonal)	GA10	Horner et al., 2015; Bates et al., 2017; Hsieh & Henderson, 2017; Bridgestock et al., 2018	
	Southern Ocean (Meridional)	GIPY04	GEOTRACES IDP 2017 (Schlitzer et al., 2018)	Frank Dehairs
	Southern Ocean (Zonal)	GIPY05	Hoppema et al., 2010	
	Arctic	GIPY11	Roeske et al., 2012	
		GN01	Whitmore et al., 2022	
	Pacific (Meridional)	GP15	GEOTRACES IDP 2021 (GEOTRACES IDP Group, 2021)	Laura Whitmore, Melissa Gilbert, Emilie Le Roy, Tristan Horner, Alan Shiller
	Subtropical South Pacific (Zonal)	GP16	Rahman et al., 2022	
Model validation	Indian Ocean	GEOSECS	Craig & Turekian (1980)	
		INDIGO 1	Jeandel et al. (1996)	
		INDIGO 2		
		INDIGO 3		
		SR3	Jacquet et al. (2004)	
		SS259	Singh et al. (2013)	

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

This data ingestion process resulted in a master record containing 5,502 observations of [Ba] that also contained a corresponding value for all 12 of the features listed in Table 1. The record was then split into a Pareto partition: the first partition was used for ML model training (4,345 observations, 79 % of data; Fig. 1A) and the second for model testing (1,157 data; 21 %). This partitioning was determined based on the basin from which the sample was collected; data from the Arctic, Atlantic, Pacific, and Southern Oceans were used in model training, whereas the 1,157 [Ba] data from the Indian Ocean were reserved for model testing (Table 2; Fig. 2). This location-based separation of training and testing data was chosen to minimize overfitting, which can occur when the training–testing separation is randomly assigned (see e.g., Rafter et al, 2019).

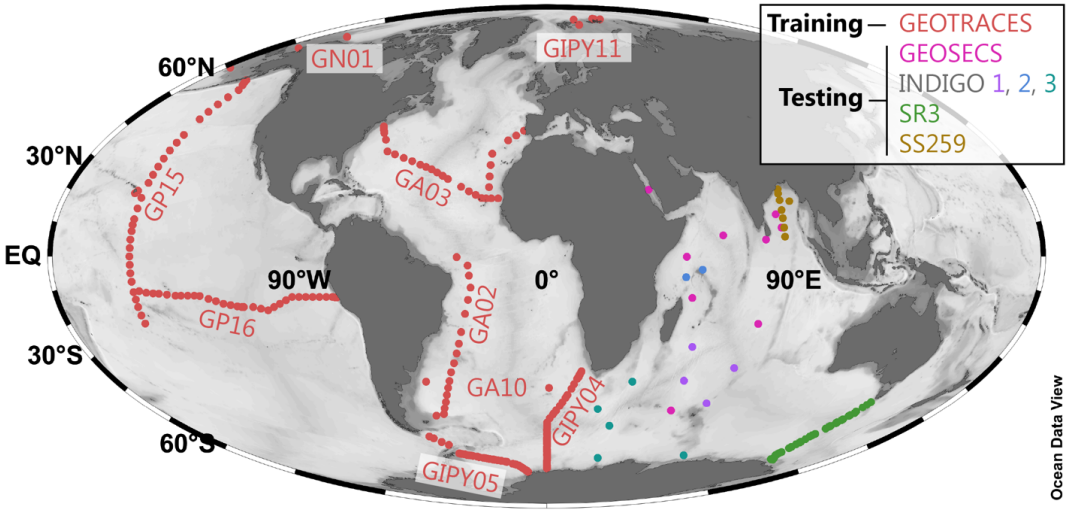


Figure 2. Geographical distribution of the training and testing data. The 4,345 core-feature complete training data (red; Fig. 1) are from the GEOTRACES 2021 Intermediate Data Product (GEOTRACES IDP Group, 2021); GEOTRACES expedition identifiers are noted next to each section. Testing data from the Indian Ocean are color-coded by expedition (see key; *n* = 1,157); data sources are listed in Table 2.

3. Methods

In the following subsections we discuss details of the specific ML algorithm that was used for model development (Sect. 3.1.), explain the model training and testing process (Sect. 3.2.), and describe how a global prediction of [Ba] was obtained and interrogated (Sect. 3.3.).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

3.1. Algorithm selection

We opted for supervised ML using a Gaussian Process Regression learner, implemented in MATLAB. This particular ML algorithm is non-parametric, kernel-based, and probabilistic. This type of algorithm is ideal when working with continuous data that also contains a certain level of noise, such as from measurement uncertainty or oceanographic variation. The basis and kernel function parameters were chosen as constant and exponential, respectively, as this combination was found to produce the most accurate predictions. All predictors were normalized and standardized to have a mean of zero and a standard deviation of unity. This process placed all parameters on the same relative range and was intended to diminish scale dependencies.

A significant problem in supervised ML algorithms is overfitting: the tendency to produce highly precise fits to the training data that cannot then be generalized to new domains or environments. We attempted to minimize overfitting by performing cross-validation during model training and during model testing. First, we used holdout cross-folding during model training. Data were randomly split into two folds, one containing 80 % of the data for model training and the other 20 % withheld for model validation (i.e., the holdout fold). This holdout process was intended to eliminate models that could only generate arbitrary fits to specific subsets of the training data. In the second stage of cross-validation, we evaluated the performance of trained models by comparing predictions against a set of withheld [Ba] observations from the Indian Ocean. None of the Indian Ocean data were seen by the models during training. This withholding was intended to help identify models that were generalizable to new environments and therefore the entire ocean (see Sect. 3.2).

3.2. Model training and testing

The training partition of the master record was used to train 4,095 different machine learning models with the goal of finding a model that could accurately simulate the global distribution of [Ba]. The number of models tested derives from the number of features investigated; each model uses a unique combination of the 12 features in Table 1 and our testing followed a factorial design

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

whereby each feature was either enabled or disabled. This design yields a total of 2^{12} unique feature combinations (i.e., $levels^{features}$). Since a model with no features enabled cannot be trained, the final number of unique, trainable, ML models with ≥1 features was 2^{12}–1=4,095. The full experiment list is provided in Section 6. Each of the 4,095 models was trained using the same 4,345 input data and with the same function parameters described above (Sect. 3.1.). Testing every possible feature combination allowed us to select for models with the highest predictive power while minimizing overfitting.

In the second stage of cross validation, trained models were used to predict [Ba] for the withheld data from the Indian Ocean. Each of the 4,095 trained models were provided with the feature information that that particular model required to simulate [Ba] and the predictions recorded. The accuracy of the models was assessed by comparing ML model predictions against observed [Ba] for the Indian Ocean data (n = 1,157) and calculating the mean absolute deviation (MAD) and mean absolute percentage deviation (MAPD). The MAD is defined as:

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n} \left| [Ba]_{predicted} - [Ba]_{observed} \right|}{n} \quad \text{[Eq. 4]}$$

and MAPD as:

$$\mathrm{MAPD} = \frac{100\ \%}{n} \sum_{i=1}^{n} \left| \frac{[Ba]_{predicted} - [Ba]_{observed}}{[Ba]_{observed}} \right| \quad \text{[Eq. 5]}$$

Models with lower accuracy exhibit higher MAD and MAPD, whereas models with high accuracy will have lower MAD and MAPD. For reference, the Ba–Si linear regression predicts Indian Ocean [Ba] with a MAD and MAPD of 6.8 nmol kg^{-1} and 9.7 %, respectively (Fig. 3). These values can be considered as benchmarks for the ML models.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

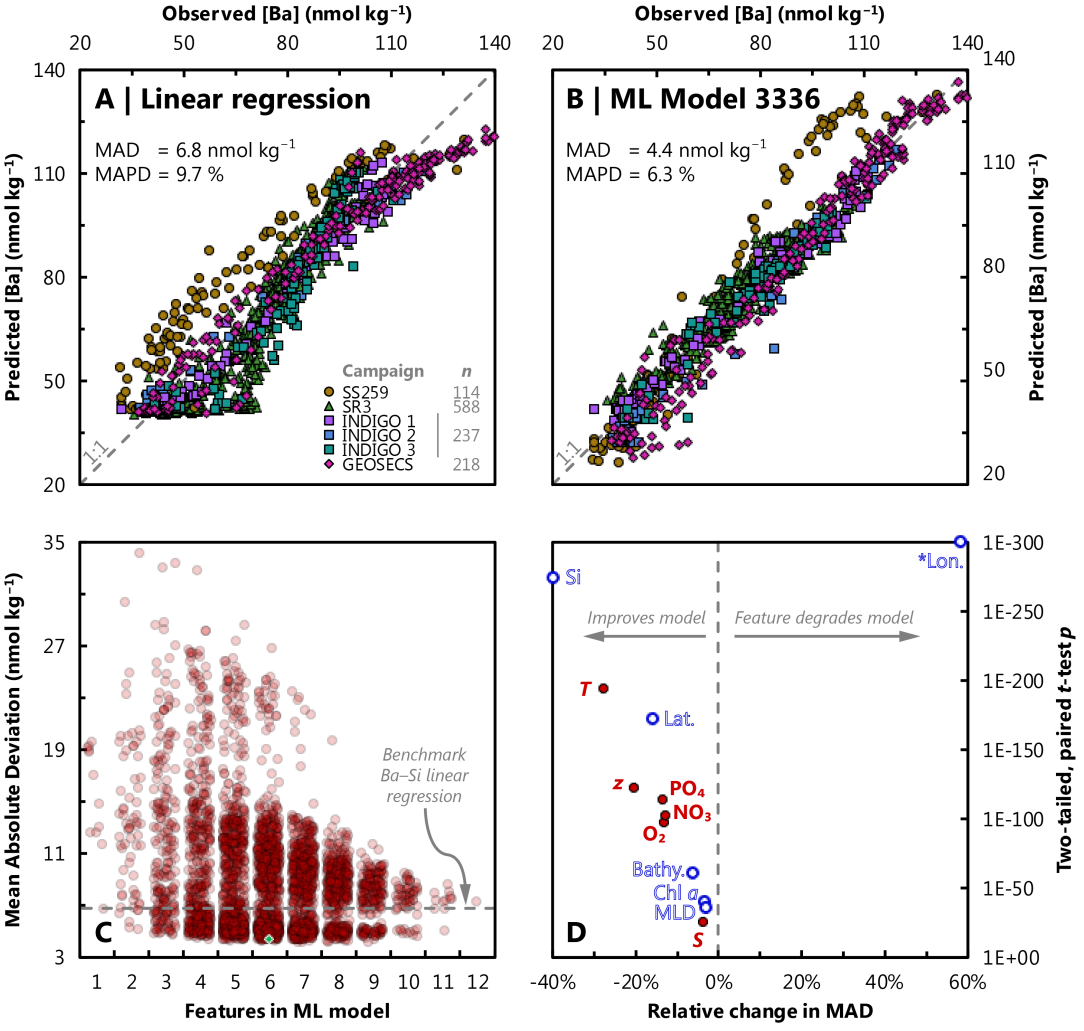


Figure 3. Evaluation of ML model performance and optimal predictor combination. Top: Crossplots of model-predicted [Ba] against observed [Ba] from the Indian Ocean for two predictor models: The benchmark Ba–Si linear regression (**A**) and our favored predictor model, #3336 (**B**). Data locations and sources in Fig. 2 and Table 2, respectively; *n* refers to the number of testing data for each campaign. Mean Absolute Deviation (MAD) and Mean Absolute Percentage Deviation (MAPD) are noted for both models. **Bottom:** Trained model MAD against number of features in the ML model (**C**); diamond indicates Model #3336 (Sect. 5.1.). Points are plotted with 80 % transparency to illustrate data density. Plot of *p*-values against relative change in MAD for each feature (**D**). Lower *p* values occur toward the top of the figure. Open and closed symbols indicate features that were excluded and included in model #3336, respectively. *Longitude is plotted with *p* = 1E-300, though the actual value is less than the smallest allowable positive number in our plotting software (~2E-308).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

3.3. Global predictions

A select number of models with low MAD and MAPD were used to simulate [Ba] on a global basis. The process by which we selected these models is described in Section 5.1. Global simulations were performed on the same grid as the WOA, which was also used as the data source for features 1–9 (Boyer et al., 2018). The WOA is a 1°×1° resolution data product with around 41,000 stations that contain up to 102 depth levels spanning 0–5,500 m in 5, 25, 50, or 100 m increments. Data for features 10–12 (MLD, chlorophyll *a*, and bathymetry) were also resampled to the WOA grid using the same sources and interpolation methods as described for the Indian Ocean testing data in Section 2. Model outputs were visualized using Ocean Data View software (ODV; Figs. 4–7; Schlitzer, 2023).

A selection of the most accurate models of [Ba] were then used to simulate Ba* and Ω_{barite}. The calculation of Ba* is shown in Equations 1 and 2. The coefficients in Equation 2 are based on data from the GEOTRACES 2021 Intermediate Data Product and specifically the subset of these data shown in Figure 1. These coefficients differ from previous formulations of Ba* that were based primarily on [Ba] and [Si] data from the Southern and Atlantic Oceans (e.g., Horner et al., 2015; Bates et al., 2017). Calculation of ML–model-derived Ba* used values of $[Si]_{in\ situ}$ from the WOA 2018 (García et al., 2018b) and $[Ba]_{in\ situ}$ from ML model output. Values of Ω_{barite} were computed using the method described by Rushdi et al. (2000), summarized in Equation 3. In this formulation, sulfate is assumed to be conservative with respect to salinity and thus this method cannot be used to predict Ω_{barite} in restricted basins, such as the Black Sea or Caspian Sea. As with calculation of Ba*, values of $[Ba]_{in\ situ}$ were obtained from ML models and co-located *T*, *S*, and pressure data were extracted from the WOA (Locarnini et al., 2018; Zweng et al., 2018).

Output from the most accurate ML models was then used to calculate mean [Ba] and Ω_{barite} for each basin, for a series of prescribed depth bins, and for the global ocean. This calculation was performed by weighting each cell in the model output by its volume, which ensures a fair comparison between any two points in the model output. We then subdivided the global ocean into five sub-basins: Arctic, Atlantic, Indian, Pacific, and Southern. Basin boundaries were defined as per Eakins & Sharman (2010), though we merged the Mediterranean and Baltic Seas into the Atlantic and considered the South China Sea as part of the Pacific Ocean. Neither [Ba] nor Ω_{barite}

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

were simulated in the Black or Caspian Seas and thus these regions are not included in the global mean calculations.

4. Results

4.1. Quantifying accuracy

Here we examine model accuracy and assess the role of different features in setting model performance. Accuracy was assessed using the mean absolute deviation (MAD; Eq. 3), which is a measure of the correspondence between predicted and observed [Ba] for the n = 1,157 data from the Indian Ocean. This correspondence is illustrated for [Ba] predicted using WOA-interpolated [Si] and the Ba–Si linear regression (Fig. 3A) and for ML model #3336 (Fig. 3B; Sect. 5.1.). The calculation of MAD was repeated for all 4,095 trained models and the results are summarized in Figure 3C. Of these models, 1,687 (41 %) achieve a superior MAD in the Indian Ocean compared to the Ba–Si linear regression benchmark of 6.8 nmol kg^{-1}. In general, ML models with fewer features tend to exhibit higher MAD than models with many features. However, adding more features to a model can also degrade its performance. Binned by the number of features, the MAD of the median model decreases from 15.8 to 7.1 nmol kg^{-1} as the number of features is increased from one to five. Beyond five features, the median-model MAD plateaus; the median MAD of the 2,510 models with ≥6 features is 7.8 nmol kg^{-1}. If considering only the most-accurate ML model within each bin, MAD monotonically decreases from 6.6 to 4.0 nmol kg^{-1} as the number of features is increased from one to eight. As the number of features is increased from nine to twelve, the MAD of the most-accurate ML model within each bin monotonically increases from 4.1 to 7.2 nmol kg^{-1} (Fig. 3C). Thus, the number of features necessary to accurately predict [Ba] in the Indian Ocean appears to be between five and eight.

We then quantified the importance of different features to model performance through a feature addition analysis (Fig. 3D). For example, model #3352 contains five features: z, S, [O_2], [PO_4], and [NO_3] and achieves a MAD of 4.6 nmol kg^{-1}. Adding T to this model increases the number of features to six and reduces the MAD to 4.4 nmol kg^{-1} (Fig. 3B). Since we used a full factorial experiment design, we were able to perform analogous pairwise comparisons between the 2^{11}

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

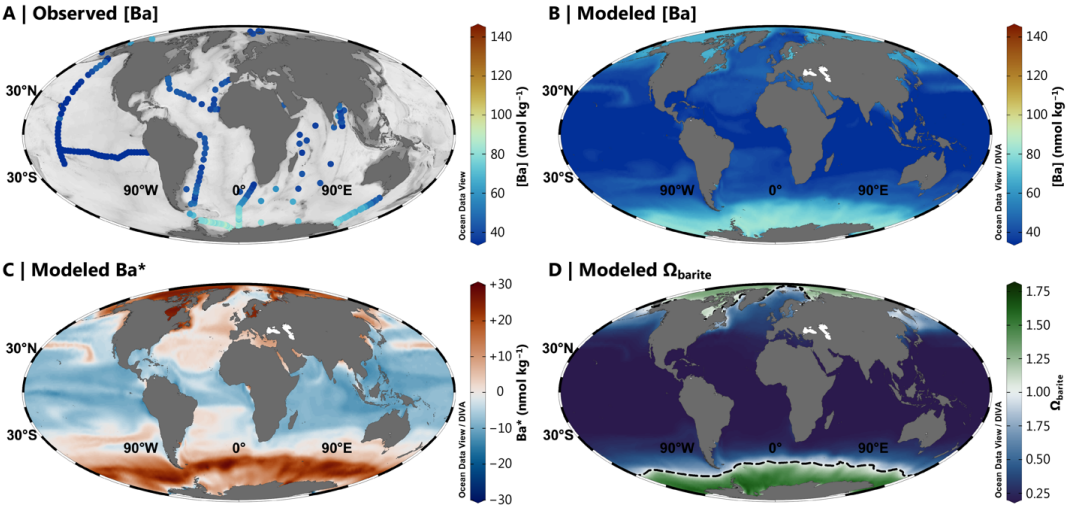
models that contained a certain feature, such as *T*, and the 2^{11} that did not. We quantified the effect of adding a feature by comparing the percentage change in MAD relative to the mean MAD of the two models. The likelihood that the inclusion of a given feature affected the MAD of the models was then quantified using a two-tailed, paired *t*-test. Lower *p* values indicate a higher likelihood that the addition of a feature significantly changed the MAD. This analysis reveals that the addition of any one of 11 features will, on average, improve an ML model. Silicate (–40 %), *T* (–27 %), and *z* (–20 %) improved the models the most and *S*, chlorophyll *a*, and MLD the least (all –3 %). Latitude, [PO_4], [O_2], [NO_3], and bathymetry improved the models by –16 % to –6 %. Longitude was the only feature found to degrade model performance, with a mean change in MAD of +59 %. The largest *p* value associated with these comparisons was 5E–25, indicating that these relationships were highly significant (Fig. 3D).

4.2. Model outputs

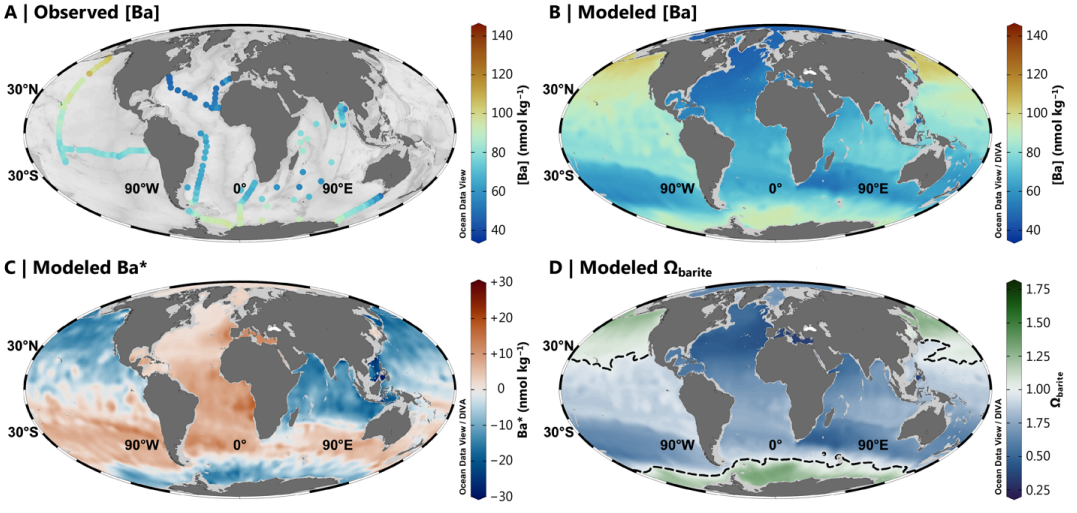
Almost 1,700 models achieved superior accuracy compared to the Ba–Si linear regression benchmark of 6.8 nmol kg^{-1}. We winnow this list to a single model, #3336, in the next section. We henceforth refer to model #3336 as our favored predictor model, which achieves a MAD of 4.4 nmol kg^{-1} using *z*, *T*, *S*, [O_2], [PO_4], and [NO_3] as predictors (Fig. 3B). Model #3336 is used to simulate [Ba], Ba*, and Ω_{barite} on a global basis and to calculate whole-ocean averages. Surface plots showing the model outputs for the sea surface, 1,000 m, 2,000 m, and 4,000 m are shown in Figures 4, 5, 6, and 7, respectively.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Open Access Earth System Science Data Discussions



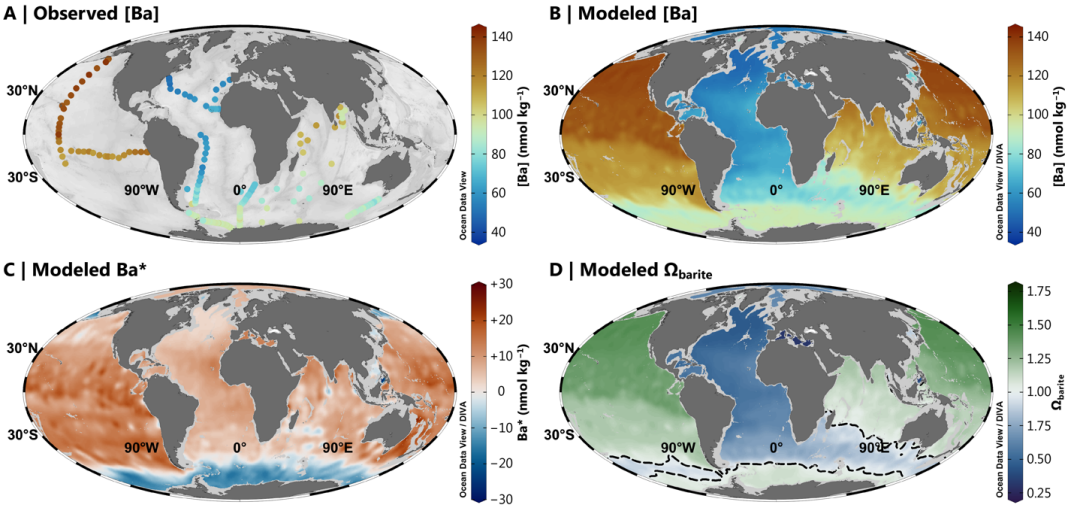
326 **Figure 4. Barium at the sea surface.** Observed [Ba] between 0–50 m (**A**); Model 3336 [Ba] (**B**), Ba* (**C**),
327 and Ω_{barite} (**D**). The dashed line in Panel D indicates the $BaSO_4$ saturation horizon (i.e., Ω_{barite} = 1.0). Panels
328 A and B use the *roma* color map, whereas Panels C and D use *vik* and *cork*, respectively (Crameri, 2018).
329 Color palettes and parameter ranges are the same for the respective panels in Figure 5–7.



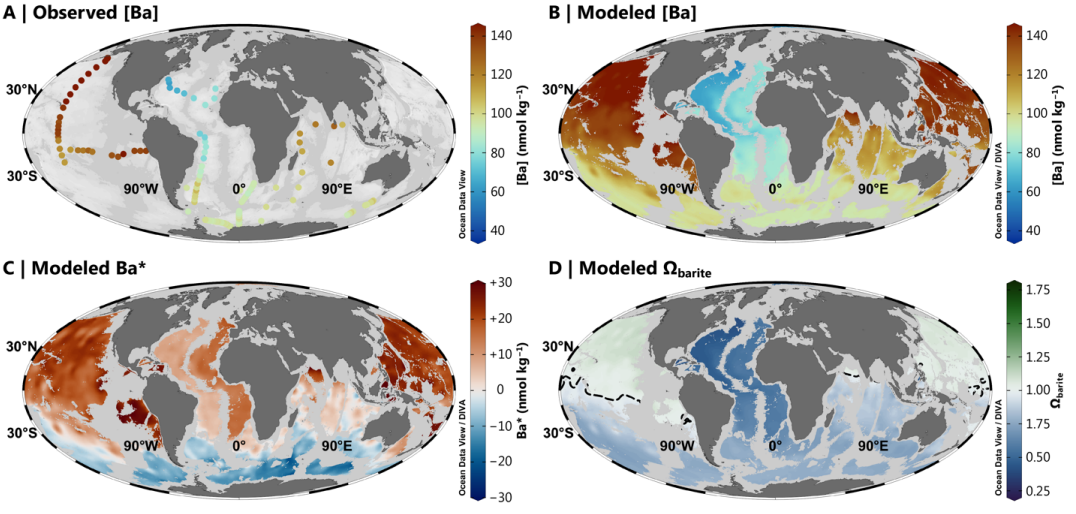
330 **Figure 5. Barium at 1,000 m.** Observed [Ba] (**A**); Model 3336 [Ba] (**B**), Ba* (**C**), and Ω_{barite} (**D**). The dashed
331 line in Panel D indicates the $BaSO_4$ saturation horizon.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Open Access Earth System Science Data Discussions



332 **Figure 6. Barium at 2,000 m.** Observed [Ba] (**A**); Model 3336 [Ba] (**B**), Ba* (**C**), and Ω_{barite} (**D**). The dashed
333 line in Panel D indicates the $BaSO_4$ saturation horizon.



334 **Figure 7. Barium at 4,000 m.** Observed [Ba] (**A**); Model 3336 [Ba] (**B**), Ba* (**C**), and Ω_{barite} (**D**). The dashed
335 line in Panel D indicates the $BaSO_4$ saturation horizon.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Model #3336 contains 3,305,505 predictions for each of [Ba], Ba*, and Ω_{barite}. Values of [Ba] range from 23.3–158.0 nmol kg^{-1}, with an unweighted mean of 71.8 nmol kg^{-1}. Based on our formulation of Ba* (Eqs. 1, 2), varies from –102.7 to +51.3 nmol kg^{-1} and possesses an unweighted mean of +2.2 nmol kg^{-1}. Values of Ω_{barite} vary from 0.11 to 1.76 and exhibit an unweighted mean of 0.75. To account for the different volumes represented by each grid cell in the model, we constructed a volume-weighted mean of [Ba] and Ω_{barite} for the ocean as a whole, for each ocean basin, and for a series of prescribed depth bins (Fig. 8). Look at the ocean as a whole, the probability density function of [Ba] roughly resembles a uniform distribution, with a mean ocean [Ba] of 89 nmol kg^{-1} (Fig. 8A). Within this mean is considerable spatial and vertical variation. For example, the Arctic Ocean exhibits the lowest volume-weighted mean [Ba] of 55 nmol kg^{-1}, whereas mean Pacific [Ba] = 106 nmol kg^{-1}. Likewise, [Ba] exceeding 100 nmol kg^{-1} rarely occurs above 1,000 m and values <45 nmol kg^{-1} are virtually absent below 1,000 m (Fig. 8B).

The probability density function of volume-weighted Ω_{barite} is closer to a normal distribution, possessing a mean value of 0.82. The Arctic, Atlantic, and Indian Oceans are, on average, undersaturated with respect to $BaSO_4$, all exhibiting $\Omega_{barite} \leq 0.81$. In contrast, the Pacific Ocean is close to saturation (Ω_{barite} = 0.97), and the Southern Ocean slightly exceeds it (Ω_{barite} = 1.04; Fig. 8C). Values of Ω_{barite} <0.25 are only found above 1,000 m, whilst values of Ω_{barite} exceeding 1.45 are exceptionally rare and are found only in the upper 1,000 m of the Southern Ocean. Lastly, Ω_{barite} tends to increase between the 0–250 m, 250–1,000 m, and 1,000–2,000 m depth bins, increasing from 0.42, to 0.63, and 0.96, respectively. Average Ω_{barite} in the deepest bin (2,000–5,500 m) is slightly lower, with a mean value of 0.92 (Fig. 8D).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

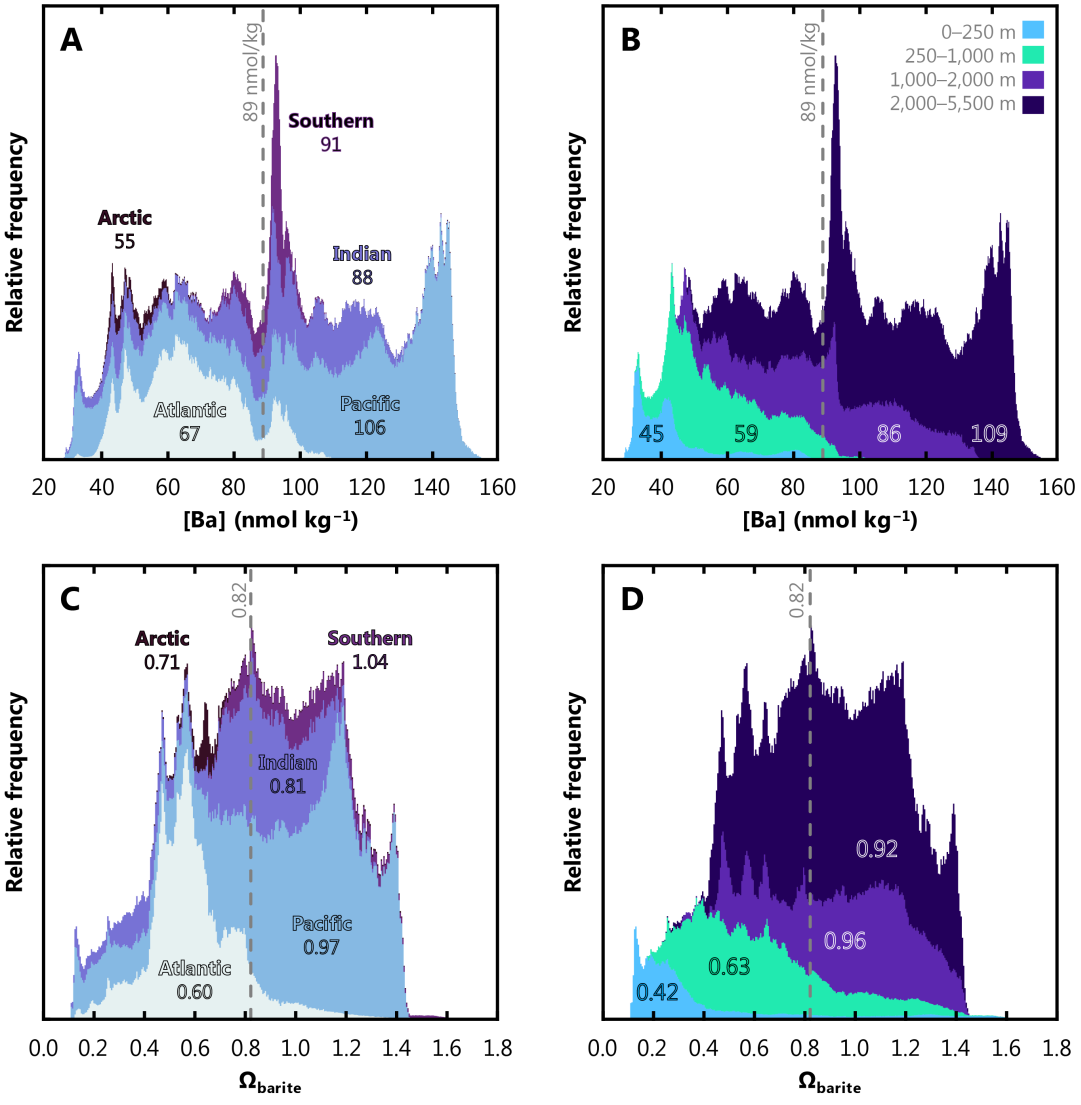


Figure 8. Stacked, volume-weighted histograms showing the relative frequency distribution of dissolved [Ba] (A, B) and Ω_{barite} (C, D) in the global ocean. The left column shows data grouped by basin, whereas the right column is grouped by a prescribed depth bin (key in B). Numbers in each panel display the mean property value for that bin. Dashed line shows the global mean.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

5. Discussion

5.1. Identification of the optimal predictor model

Our results show that 1,687 of the 4,095 ML models (41 %) produce more accurate predictions than the benchmark. In this case, the benchmark is predicting Indian Ocean [Ba] based on the Ba–Si linear regression and using WOA [Si] as the only predictor (Figs. 1, 3C). However, choosing *the* optimal feature combination is challenging given the sheer number of skillful ML models. Here, we winnow the list from 1,687 to a single model (#3336) by eliminating models based on the number and information content of various features. First, longitude was found to be the only feature that consistently degraded the performance of trained models (Fig. 3D). We therefore disabled this feature, eliminating 53 models. Having removed longitude, we also decided to eliminate models utilizing latitude; this decision was guided by the principle that a generalizable model should be able to predict [Ba] using only physical and/or chemical predictors, independent of where a sample is located in x–y space. This removed a further 853 models, winnowing the list to 781. After longitude, the features offering the least improvement to ML model performance were MLD and Chl. *a*, which improved trained models by around –3 % (Fig. 3D). Indeed, the median model was degraded by 2 % if Chl. *a* was included. Eliminating models containing either of MLD (397) or chlorophyll *a* (194 models) reduced the number of models exceeding the Ba–Si benchmark of 6.8 nmol kg^{-1} to 190, of which 122 utilized [Si] as a predictor. While [Si] was amongst the strongest overall predictors of [Ba] (Fig. 3D), incorporating [Si] into a trained model introduces potential circularity into the calculation of Ba* (see Eqs. 1, 2). Moreover, the four-most accurate surviving models containing [Si], models 3144, 3268, 3716, and 3732, achieve a similar MAD of 4.2 nmol kg^{-1} to our ultimately favored model #3336 (4.4 nmol kg^{-1}).

Eliminating models containing [Si] reduced the list of models to 68. Of these, 39 were eliminated as they contained ≤4 features, noting that our earlier analysis showed that the optimal number of features needed to accurately predict [Ba] in the Indian Ocean was between five and eight (Sect. 4.1.). Models with fewer than four features likely do not contain sufficient information to make accurate predictions in the full range of environments encountered in the Indian Ocean. Likewise, beyond eight features, trained models tended to produce worse fits to observations. We suspect that the reason is overfitting; feature-rich models can be overfit to the training data and are unable to generalize when presented with completely new environments. Thus, this winnowing process

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

reduced the number of candidate ML models to 29. All of these models exhibited superior accuracy compared to the Ba–Si linear correlation (MAD = 6.8 nmol kg^{-1}); were not trained using longitude, latitude, MLD, chlorophyll *a*, or [Si]; and possessed between five and seven features. The most accurate amongst these 29 models, #3336, utilizes six features—*z*, *T*, *S*, [O_2], [PO_4], and [NO_3]—and achieves a MAD and MAPD in the Indian Ocean of 4.4 nmol kg^{-1} and 6.3 %, respectively (Fig. 3B). This level of accuracy represents at least a 35 % improvement compared to predicting [Ba] in seawater using the Ba–Si linear regression. We consider model #3336 as the optimal configuration for predicting [Ba] in this study, and use this model to simulate [Ba], Ba*, and Ω_{barite} in Figure 4–8.

5.2. Model validation

We now explore the validity of model #3336 in terms of its oceanographic consistency, the sources of uncertainty that affect its accuracy, and potential limitations of the model output. We find that model #3336 reproduces the major known features of the marine [Ba] distribution and makes testable predictions for regions that are yet to be sampled.

5.2.1. Visual inspection of model output

Visual inspection of model output is an important component of data analysis considering the limits of statistical tests (see e.g., Anscombe, 1973). Models may produce statistically satisfactory fits to the testing data, but the oceanic realism of the output is also important to consider. Modeled [Ba] should display patterns consistent with related oceanographic properties and exhibit smooth vertical and spatial variations (Boyle & Edmond, 1975). Predicted [Ba] from model #3336 does indeed show smooth and systematic spatial and vertical variations that also resembles sparse observations (Figs. 3–7).

There are, however, several sharp gradients in modeled [Ba], particularly at the sea surface (Fig. 3). These variations generally show an increase in [Ba] close to land and especially near the mouths of major rivers. This is reassuring given that elevated sea-surface [Ba] close to rivers is both widely reported and is one of the major proxy applications of Ba: reconstructing spatiotemporal patterns

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

of terrestrial runoff by measuring the Ba:Ca ratio of carbonates (e.g., Sinclair & McCulloch, 2004; LaVigne et al., 2016). Model #3336 correctly identifies elevated [Ba] near the Ganges–Brahmaputra (Singh et al., 2013) and Río de la Plata outflows (GEOTRACES IDP Group, 2021). Model #3336 also predicts elevated surface [Ba] in the Gulf of Guinea (Niger and Volta Rivers) as well as Gulf of St. Lawrence (St. Lawrence River), though these latter predictions await corroboration. Interestingly, model #3336 does not predict elevated surface [Ba] at all major outflows; neither the Mississippi nor Amazon Rivers are associated with significant increases in surface [Ba]. The lack of elevated [Ba] at these rivers may reflect seasonal variations in Ba release that are not captured by our mean annual model (e.g., Joung & Shiller, 2014), or it may indicate that these rivers are not major *net* sources of Ba to the surface ocean (e.g., Coffey et al., 1997).

Overall, model #3336 makes accurate, oceanographically consistent predictions of [Ba] in the Indian Ocean using input data from the WOA. Model #3336 also makes a number of testable predictions of [Ba] in regions lacking direct observations. Given that these predictions were made using the same model and the same WOA inputs, we believe that it is reasonable to assume that model #3336 output is an accurate representation of mean annual global [Ba].

5.2.2. Quantifying uncertainties

We now describe and, where possible, quantify two possible sources of uncertainty to our ML model output. Before doing so, we describe how uncertainty is quantified as well as the uncertainty of existing approaches. Certain ML models, such as Gaussian Process Regression, offer low interpretability, meaning it is not possible to assess uncertainty using a conventional error propagation. Thus, all model uncertainties are assessed *post hoc*, by comparing predictions against observations. Our preferred metrics are MAD and MAPD—mean absolute deviation and mean absolute percentage deviation, defined in Equations 4 and 5, respectively. Existing approaches for estimating [Ba] result in a wide range of uncertainties. At the low end, the uncertainty associated with measuring [Ba] in seawater represents a fundamental limit to the accuracy of any model. A number of analysts report relative [Ba] uncertainties in the range of 1–2 % (e.g., Pyle et al., 2018; Cao et al., 2020). This level of intra-laboratory uncertainty is typical for [Ba] data obtained using isotope dilution–inductively coupled plasma mass spectrometry, and applies to GEOTRACES-era

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

445 datasets and to much of the training data from the Indian Ocean. However, intra-laboratory
446 uncertainty is typically much smaller than inter-laboratory uncertainty, which is often between 6–
447 9 % (e.g., Hathorne et al., 2013). At the upper end, the benchmark Ba–Si linear regression achieves
448 a MAPD of 9.7 % in the Indian Ocean (Fig. 3A). Thus, useful ML models of [Ba] should achieve
449 uncertainties between 1–10 %. Indeed, our favored predictor model, #3336, achieves a MAPD of
450 6.3 %.

451 Now we consider two factors that contribute to the observed 6.3 % uncertainty: realization
452 uncertainty and uncertainties in the training data. The realization uncertainty stems from the fact
453 that two models trained on the same training dataset—even with the exact same subset of model
454 features—will produce slightly different predictions. This is due to the holdout cross-folding
455 process used during model training, which partitions the training dataset into subsets (see Sect.
456 3.1.). Because the partitioning is random, the training process results in a slightly different trained
457 model each time the model is realized. We quantified the realization uncertainty by training model
458 #3336 100 times and calculating the relative standard deviation of the different predictions of [Ba]
459 for all 3,305,505 values in the output. This uncertainty is small; the median, mean, and maximum
460 realization uncertainty was 0.03 %, 0.04 %, and 0.32 % variability in modeled [Ba].

461 Next we consider uncertainties in the training data. As noted above, many labs report uncertainties
462 on [Ba] measurements of 1–2 %, while inter-laboratory differences may be up to a factor of five
463 larger. However, this does not consider any uncertainties associated with the other physical and
464 chemical features used to predict [Ba]. In general, these uncertainties should be small since all
465 overboard sensors are regularly calibrated and biogeochemical properties are determined using
466 established methods that are based on GO-SHIP best practices (Hood et al., 2010). Moreover, all
467 GEOTRACES sections include crossover stations that are intended to facilitate intercalibration of
468 all parameters, including those used here to predict [Ba] (Fig. 2; Cutter, 2013). The WOA, MLD,
469 Chl. *a*, and bathymetry data products are similarly subjected to stringent quality review and so we
470 consider it unlikely that these data contribute systematic biases. We believe that the most likely
471 source of uncertainty relates to the fact that all predictor information used for model testing in the
472 Indian Ocean was derived from time-averaged data products, whereas [Ba] was derived from *in*
473 *situ* measurements. We made this decision because the *in situ* data were incomplete for all 12 core
474 features (Table 1), and this would have necessitated interpolation for some features and not others.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

475 Since all models were tested using the same predictor information, the comparison process should
476 avoid systematic errors, though this does not preclude temporal variability, described next.

477 *5.2.3. Other considerations*

478 We now consider four other factors that potentially contribute to the uncertainty of the model
479 output: short- and long-term temporal variations, limitations of ML, and uncertainties regarding
480 the thermodynamic properties of $BaSO_4$. Short-timescale variability in [Ba] may affect how
481 models were evaluated, though this effect is difficult to quantify. In principle, the trained models
482 should be able to resolve seasonal variations in [Ba] since they were trained on *in situ* physical and
483 chemical data. In contrast, model predictions in the Indian Ocean were made using annual average
484 physical and chemical conditions and then evaluated by comparing these predictions against *in*
485 *situ* [Ba]. The temporal mismatch between Indian Ocean observations and predictions is unlikely
486 to be significant in the deep ocean, where seasonal variations are minor and the Ba residence time
487 is longest (e.g., Hayes et al., 2018). Seasonal variations are, however, likely to matter more for the
488 surface ocean. We were able to minimize some of the impact of these uncertainties by using long-
489 term averages of Chl. *a* and the maximum monthly mean MLD during model training and testing.
490 Significant seasonal mismatches for other parameters are unavoidable given that [Ba] data are too
491 sparse to develop a time-resolved model. We suspect that these variations are most likely to be
492 significant for boundary sources rather than biogeochemical cycling of Ba; significant
493 biogeochemical drawdown of surface [Ba] over seasonal timescales appears to be rare (e.g., Esser
494 & Volpe, 2002), whereas there are large seasonal variations in river discharge that impact near-
495 shore [Ba] (e.g., Samanta & Dalai, 2016). These suspicions could be tested using a model with
496 better than 1×1° spatial resolution, which—in theory—is possible with model #3336, so long as
497 similarly high-resolution data are provided for the six predictors utilized by this model (z, T, S,
498 $[O_2]$, $[PO_4]$, and $[NO_3]$). While it is challenging to precisely quantify seasonal uncertainties, we
499 note that model #3336 performs well at low [Ba], which is found mostly near the surface, where
500 seasonal variations should be largest (Figs. 3B, 8B). Likewise, seasonal variations will have only
501 a minor effect on our calculations of global mean [Ba] or Ω_{barite} (Fig. 8).

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Long-term variability in [Ba] may also influence model performance, since the testing data from the Indian Ocean were collected between 1977 (GEOSECS) and 2008 (SS259). If secular changes in Indian Ocean [Ba] were occurring, we might expect models to make accurate predictions for some datasets at the expense of others. In contrast, we note that model #3336 reproduces all testing datasets similarly well, with the exception of a subset of samples from SS259 in the deep Bay of Bengal (Fig. 3C). Here we observe that model #3336 tends to predict ~9 % higher [Ba] than observed by Singh et al. (2013), particularly around 2,000 m (Fig. 6A, B). However, model #3336 correctly predicts [Ba] at nearby GEOSECS stations 445 and 446, also in the Bay of Bengal, sampled some 31 years prior. The origin of the model–data discrepancy in this region is uncertain; we speculate that it may reflect differences in how *in situ* [Ba] was measured, noting Singh et al. (2013) opted for standard addition over isotope dilution. Alternatively, it could reflect a ~9 % decrease in [Ba] in the deep Bay of Bengal since the 1970's.

A third factor concerns the limitations of ML itself. We note that no trained model was able to achieve a MAPD better than ~6 %. This value may represent one of three things. First, it may point toward an intrinsic limitation of Gaussian Process Regression. Other types of ML, such as Decision Trees or Artificial Neural Networks, may be able to achieve superior accuracy, though this was not investigated. Second, it may indicate that the 12 features investigated provide insufficient information about [Ba] to achieve higher accuracy. We view this as unlikely given that our earlier analysis showed that only 5–8 features were needed to accurately simulate [Ba] and that the 12 features investigated have proved useful in other studies simulating dissolved tracer distributions (e.g., Rafter et al., 2019; Sherwen et al., 2019; Roshan & DeVries, 2021). However, this does not rule out the existence of other features beyond the 12 that we tested that are more useful for predicting [Ba], only that we did not investigate them. Third, it is possible that the lowest MAPD of ~6 % reflects the current limit of inter-laboratory uncertainty in determining [Ba]. We note that inter-laboratory uncertainties of 6–9 % were reported for the measurement of Ba:Ca in carbonates (n = 10 labs; Hathorne et al., 2013). If the ~6 % MAPD derives from inter-laboratory uncertainty, it is unlikely that further model refinements will improve the accuracy of [Ba] predictions: the fundamental limitation is the data, not the model.

A final source of uncertainty concerns the solubility coefficients used to calculate Ω_{barite}. We calculated Ω_{barite} based on the computation described by Rushdi et al. (2000). This calculation is

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

532 based on $BaSO_4$ solubility data from Raju & Atkinson (1988), who note good agreement with the
533 thermodynamic data of Blount (1977). These solubility data were obtained based on
534 experimentation with lab-made, coarse-grained $BaSO_4$, which is unlikely to be wholly
535 representative of the microcrystalline $BaSO_4$ precipitates found in seawater. Thus, the absolute
536 values of Ω_{barite} calculated here may be subject to eventual revision; however, the vertical (Fig. 1),
537 spatial (Figs. 3–7), and whole-ocean (Fig. 8) trends in Ω_{barite} are robust. Should new
538 thermodynamic data for marine-relevant micron-sized pelagic $BaSO_4$ become available, updated
539 maps of Ω_{barite} could be easily recalculated using existing model #3336-derived [Ba] data.

540 **5.3. Model applications**

541 Here we provide an overview of the main model features in [Ba], Ba* and Ω_{barite}, then outline four
542 possible applications of the model output. Predictions from model #3336 shows several interesting
543 features in [Ba] (Figs. 4–7). Model #3336 reproduces the nutrient-like distribution of [Ba] (Fig.
544 1C) and shows a general increase in [Ba] along the Meridional Overturning Circulation; mean [Ba]
545 increases from 67 to 88 to 106 nmol kg^{-1} from the Atlantic to Indian to the Pacific Ocean,
546 respectively. The model also predicts some variation in shallow [Ba] that follows major surface-
547 water currents, such as a region of elevated [Ba] associated with the North Pacific Current, as well
548 as low [Ba] in the western North Atlantic associated with the Gulf Stream (Fig. 4B; Talley et al.,
549 2011). Taking the ocean as a whole, we can also use our model output to calculate the total Ba
550 inventory of the ocean. Using the mean oceanic [Ba] of 89 nmol kg^{-1} and multiplying by the mass
551 of seawater (1.37×10^{21} kg) yields a total inventory of 122±8 Tmol Ba, whereby the range is based
552 on the MAPD of model #3336 (6.3 %). This estimate of the total oceanic Ba inventory is
553 approximately 10–21 % lower than previous estimates of 145 Tmol Ba (Dickens et al., 2003;
554 Carter et al., 2020). Thus, given current estimates of global marine Ba fluxes of between 18 (Paytan
555 & Kastner, 1996) and 44 Gmol Ba yr^{-1} (Rahman et al., 2022), the mean residence time of Ba in
556 seawater is likely between 2,700–7,200 years.

557 Next we consider the main features in Ba* and Ω_{barite}. In the surface ocean, patterns of Ba*
558 resemble those of [Ba] (Fig. 4). This is likely because in large parts of the ocean, surface [Si]
559 approaches 0 µmol kg^{-1}; thus, variations in Ba* derive mostly from variations in [Ba]. Barium-star

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

560 is, however, strongly positive in the top 200 m of the Southern Ocean, even though [Si] is in the
561 10's of μmol kg^{-1} range. At 1,000 m, the North Pacific, Southern, and Indian Oceans exhibit
562 negative values between –10 to –20 nmol kg^{-1}, whereas the Atlantic and South Pacific Oceans are
563 positive around +10 nmol kg^{-1} (Fig. 5). Below, 1,000 m, the Southern and the Atlantic Oceans do
564 not exhibit significant gradients in Ba* and are negative and positive down to the seafloor,
565 respectively. In contrast, the Indian and North Pacific Oceans exhibit positive Ba* around 2,000
566 m, between +5 and +15 nmol kg^{-1} (Fig. 6). The most positive Ba* values, between +20 and +30
567 nmol kg^{-1}, are observed at 4,000 m in the Pacific, specifically in the Peru and Chile Basins as well
568 as the Philippine Sea (Fig. 7). Vertical profiles of Ω_{barite} are in agreement with published values;
569 comparisons are provided in the Appendix (e.g., Jeandel et al., 1996; Monnin et al., 1999; Rushdi
570 et al., 2000). Excepting the high latitudes, the surface ocean is undersaturated with respect to
571 $BaSO_4$ (i.e., Ω_{barite} <1) and the lowest values of Ω_{barite} in the open ocean are observed in the cores
572 of the Subtropical Gyres (Ω_{barite} between 0.1–0.2; Figs. 4D, 8D). Whereas the Southern Ocean
573 remains supersaturated (i.e., Ω_{barite} >1) to around 2,000 m, the Arctic Ocean switches to
574 undersaturated conditions below ~250 m. Below 1,000 m, most of the North Pacific is
575 supersaturated with respect to $BaSO_4$ and by 2,000 m almost all of the ocean exhibits Ω_{barite} >1,
576 excepting the Atlantic Ocean, which is undersaturated at all depths (Fig. 6D). The South Pacific
577 and Indian Oceans return to undersaturated conditions by 4,000 m, whereas the majority of the
578 North Pacific exhibits Ω_{barite} >1 to the seafloor (Fig. 7D). From a volumetric perspective, the ocean
579 below 1,000 m exhibits a mean $\Omega_{barite} \geq 0.92$, which implies that much of the deep ocean is close to
580 saturation with respect to $BaSO_4$ (Fig. 8D).

581 Lastly, in the spirit of maximizing model utility, we suggest three possible uses for these data.
582 First, the outputs can be used for model intercomparison and intercalibration. For example, a
583 number of statistical models, such as Optimum Multiparameter Optimization, have been
584 successfully used to study Ba cycling in the North Atlantic (Le Roy et al., 2018; Rahman et al.,
585 2022), Southeast Pacific (Rahman et al., 2022), and Mediterranean Sea (Jullion et al., 2017). These
586 models can apportion the relative contributions of *in situ* biogeochemical cycling and conservative
587 mixing to observed [Ba]; however, accurate quantification of these processes requires *a priori*
588 knowledge of end-member water mass [Ba], which model #3336 can provide. Our model could
589 also be used to benchmark output from process-based models, such as the Ocean Circulation
590 Inverse Model (e.g., Roshan & DeVries, 2021). Second, the output can be used for interpolation

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

purposes. For example, many groups investigated Ba partitioning into various types of marine carbonates (see Sect. 1 for examples); however, these investigations are sometimes performed without a co-located measurement of [Ba]. In these cases output from model #3336 could be used to help calibrate specific substrates, such as deep-sea corals or benthic forams. This also avoids the potential for circular reasoning whereby [Si] is used to estimate [Ba], which is then reconstructed from the Ba:Ca ratio of carbonates to estimate [Si]. Third, the model output makes testable predictions for regions of the ocean that have yet to be sampled by GEOTRACES-style surveys. Several of these regions, such as the Southern Ocean, exhibit with sharp lateral and vertical gradients in [Ba], Ba*, and Ω_{barite}. Such gradients should be considered prime targets for future process-oriented studies of marine Ba cycling.

6. Data availability

Data described in this manuscript can be accessed at the *Biological and Chemical Oceanography Data Management Office* under data doi:10.26008/1912/bco-dmo.885506.1 (Horner & Mete, 2023).

7. Conclusions

This study presents a spatially and vertically resolved global model of [Ba] determined using Gaussian Process Regression machine learning. The model reproduces several known features of the marine [Ba] distribution and makes testable predictions in regions that are yet to be sampled. Analysis of the model output reveals the mean oceanic [Ba] is 89 nmol kg^{-1}, implying a total marine Ba inventory of 122±8 Tmol. Using predictors from the World Ocean Atlas, we also estimate the global distribution of Ba* and Ω_{barite}. Both properties exhibit significant gradients that can be systematically investigated in future studies. The mean oceanic Ω_{barite} is 0.82, though between 1,000–5,500 m the mean is ≥0.92, implying that the deep ocean is close to saturation with respect to $BaSO_4$. Our model output should prove valuable in studies of Ba biogeochemistry, specifically for statistical- and process-based model validation, calibrating sedimentary archives,

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

and for identifying promising regions for further study. More broadly, our study demonstrates the utility of using machine learning to accurately simulate the distributions of trace elements in seawater. With minor adjustments, our approach could be employed to make predictions for other dissolved tracers in the sea.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Open Access Earth System Science Data Discussions

Appendix

Here we compare our results with published profiles of Ω_{barite}. Our results were calculated using the thermodynamic model of Rusdi et al. (2000), model #3336 [Ba], and WOA *T*, *S*, and pressure. Literature profiles of Ω_{barite} were calculated using one of three different thermodynamic models and *in situ* observations of [Ba], *T*, *S*, and pressure. In general, there is strong agreement between modeled and *in situ* Ω_{barite} whereby our model reproduces the shape of published profiles (Fig. A1). There are, however, some small systematic offsets between the various approaches, and we suspect that these derive from differences in the underlying thermodynamic models.

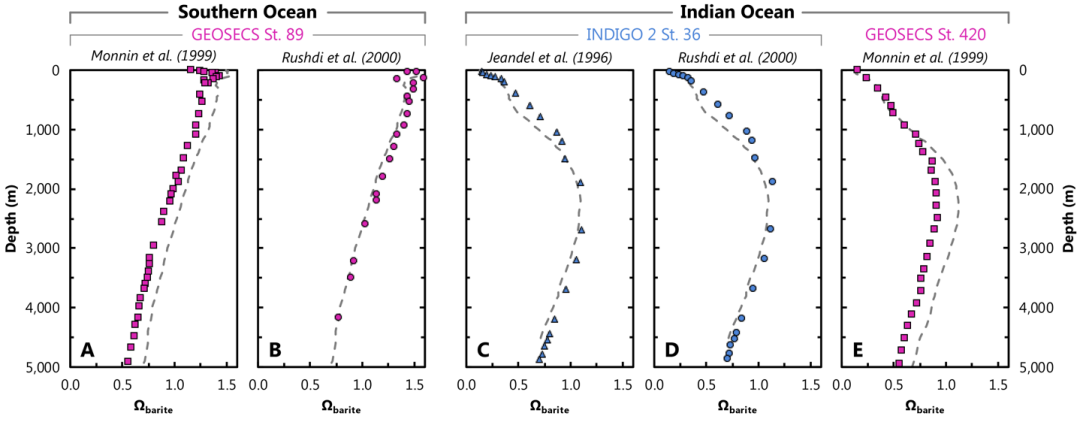


Figure A1. Comparison of literature- (symbols) and Model #3336-derived (dashed line) estimates of Ω_{barite}. Panels **A** and **B** show profiles of Ω_{barite} at GEOSECS St. 89 (60°0' S, 0°2' E). The other panels are from the Indian Ocean; **C** and **D** are from INDIGO 2 St. 36 (6°9' S, 50°55' E) and **E** from GEOSECS St. 420 (0°3' S, 50°55' E), some ≈675 km north of INDIGO 2 St. 36.

We compare our model output with literature data at two stations (Fig. A1). These stations were selected for comparing Ω_{barite} because at both locations at least two studies calculated profiles of Ω_{barite} using the same underlying *in situ* data for [Ba], *T*, *S*, and pressure. This ensures a fair comparison between studies, since any differences in modeled Ω_{barite} derive from the thermodynamic model and not the input data. Likewise, the literature profiles at these locations were based on calculations for pure, rather than strontian, $BaSO_4$, as in our study. Published profiles of Ω_{barite} were extracted graphically from each study using *WebPlotDigitizer* (Rohatgi, 2022). This extraction process may introduce some minor scatter in the literature data, though this is relatively minor relative to the range of variation in Ω_{barite}.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023


Open Access Earth System Science Data Discussions

641 First, we examine profiles of Ω_{barite} reported for GEOSECS St. 89 in the Southern Ocean (Fig. A1;
642 Monnin et al., 1999; Rushdi et al., 2000). Modeled and published profiles show supersaturation in
643 the surface ocean and undersaturation below 2,000–2,500 m. Profiles from Rushdi et al. (2000)
644 show excellent agreement with Ω_{barite} calculated from model #3336 [Ba] and WOA *T*, *S*, and
645 pressure, with our output slightly negatively offset by a MAD of 0.06 (n = 22). Given that we use
646 the same thermodynamic model as Rushdi et al. (2000), the overall excellent agreement with their
647 study is not surprising. However, the result is nonetheless reassuring since our study uses mean
648 annual values for the various inputs, whereas Rushdi et al. (2000) utilized *in situ* data. There is a
649 slightly larger offset between our profile of Ω_{barite} and that calculated by Monnin et al. (1999), with
650 our profile offset to higher Ω_{barite} by a MAD of 0.13 (n = 41). This most likely reflects differences
651 in the underlying thermodynamic model and not the *in situ* data since our model reproduces the
652 same overall profile shape as Monnin et al. (1999). Likewise, both Monnin et al. (1999) and Rushdi
653 et al. (2000) used the same *in situ* input data and their results are highly comparable, albeit with
654 an offset similar to that between our results and Monnin et al. (1999).

655 Next we examine profiles of Ω_{barite} in the Indian Ocean for samples from INDIGO 2 St. 36 (Fig.
656 A1; Jeandel et al., 1996; Rushdi et al., 2000). Profiles of Ω_{barite} show undersaturation at the surface,
657 moderate supersaturation between 2,000–3,500 m, then return to undersaturated conditions down
658 to the seafloor. Our profile shows overall excellent agreement with that of Jeandel et al. (1996),
659 whereby our data are offset to slightly lower Ω_{barite} with a MAD of 0.06 (n = 21). The largest
660 differences are observed between 594–1,042 m depth, where the MAD is 0.15 (n = 3). Our profile
661 shows similarly good agreement with that of Rushdi et al. (2000), whereby our data are offset to
662 lower Ω_{barite} with a MAD of 0.07 (n = 20). As with the data of Jeandel et al. (1996), we observe a
663 larger offset between modeled Ω_{barite} and the data of Rushdi et al. (2000) between 594–1,042 m,
664 equivalent to a MAD of 0.17 (n = 3). We consider these larger mesopelagic offsets in Ω_{barite} to be
665 the result of differences in the predictors (i.e., [Ba], *T*, *S*), rather than the thermodynamic model;
666 Jeandel et al. (1996) and Rushdi et al. (2000) use the same *in situ* predictor data and yield similar
667 Ω_{barite}, despite using different thermodynamic models.

668 We also compared our results with data from St. 420 of GEOSECS (Monnin et al., 1999), which
669 is located ≈675 km north of INDIGO 2 St. 36 (Fig. 2). As with data from the Southern Ocean
670 (GEOSECS St. 89), our profile data are offset to higher Ω_{barite} than those of Monnin et al. (1999)

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023


Open Access Earth System Science Data Discussions

by a MAD of 0.12 (n = 29). However, our modeled Ω_{barite} is generally in much closer agreement with Monnin et al. (1999) above 1,250 m than below, equivalent to a MAD of 0.03 (n = 9) and 0.16 (n = 20), respectively. In this case it is more challenging to ascribe a unique cause of the differences in calculated Ω_{barite}; these offsets could relate to differences in the predictors or the thermodynamic model.

Overall, our ML-derived profiles of Ω_{barite} show excellent agreement with *in situ* data, both in terms of profile shape and absolute values of Ω_{barite} to within 0.1. Should a revised thermodynamic model and/or improved $BaSO_4$ solubility coefficients become available, a new grid of Ω_{barite} could be calculated using Model #3336 [Ba] and WOA T, S, and pressure data.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Author contributions

Project conceptualization and funding acquisition by T.J.H. Data curation, formal analysis, investigation, and methodology by O.Z.M., A.V.S., and T.J.H. Data visualization by A.V.S. and T.J.H. Software provided by O.Z.M., A.V.S., H.H.K., and A.G.D. Writing (original draft) by O.Z.M. and T.J.H.; review and editing by H.H.K., A.G.D., L.M.W., A.M.S., M.G., and W.D.L.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We are thankful to the many data originators who contributed—and the funding agencies that enabled the generation of—quality-controlled dissolved Ba data to the 2021 GEOTRACES Intermediate Data Product. The GEOTRACES IDP represents an international collaboration and is endorsed by the Scientific Committee on Oceanic Research. We are especially grateful to Frank Dehairs, who provided comments on an early draft of the text and shared additional testing data from the Indian Ocean, as well as Karen Grissom, who provided laboratory assistance to A.M.S. We kindly acknowledge use of the *Discovery* high-performance compute nodes at Dartmouth College Research Computing.

Financial support

O.Z.M was supported by WHOI's Academic Programs Office through a *Summer Student Fellowship*. A.M.S. acknowledges support from the U.S. National Science Foundation (OCE-0927951, OCE-1137851, OCE-1261214, OCE-1436312, and OCE-1737024), as does T.J.H. (OCE-1736949 and OCE-2023456). T.J.H. was further supported by *The Andrew W. Mellon Foundation Endowed Fund for Innovative Research* and *The Breene M. Kerr Early Career Scientist Endowment Fund*.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

References

Anagnostou, E., Sherrell, R. M., Gagnon, A., LaVigne, M., Field, M. P., & McDonough, W. F. (2011). Seawater nutrient and carbonate ion concentrations recorded as P/Ca, Ba/Ca, and U/Ca in the deep-sea coral Desmophyllum dianthus. *Geochimica et Cosmochimica Acta*, *75*(9), 2529–2543. https://doi.org/10.1016/j.gca.2011.02.019

Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician, 27*(1), 17-21. https://doi.org/10.1080/00031305.1973.10478966

Bains, S., Norris, R. D., Corfield, R. M., & Faul, K. L. (2000). Termination of global warmth at the Palaeocene/Eocene boundary through productivity feedback. *Nature*, *407*(6801), 171–174. https://doi.org/10.1038/35025035

Bates, S. L., Hendry, K. R., Pryer, H. V., Kinsley, C. W., Pyle, K. M., Woodward, E. M. S., & Horner, T. J. (2017). Barium isotopes reveal role of ocean circulation on barium cycling in the Atlantic. *Geochimica et Cosmochimica Acta*, *204*, 286–299. https://doi.org/10.1016/j.gca.2017.01.043

Boyer, Tim P.; García, Hernán E.; Locarnini, Ricardo A.; Zweng, Melissa M.; Mishonov, Alexey V.; Reagan, James R.; Weathers, Katharine A.; Baranova, Olga K.; Paver, Christopher R.; Seidov, Dan; Smolyar, Igor V. (2018). World Ocean Atlas 2018. NOAA National Centers for Environmental Information. Dataset. https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18

Bishop, J. K. (1988). The barite–opal–organic carbon association in oceanic particulate matter. *Nature*, *332*(6162), 341-343.

Blount, C. W. (1977). Barite solubilities and thermodynamic quantities up to 300 degrees C and 1400 bars. *American Mineralogist*, *62*(9–10), 942–957.

Boyle, E., & Edmond, J. M. (1975). Copper in surface waters south of New Zealand. *Nature*, *253*(5487), 107–109. https://doi.org/10.1038/253107a0

Bridgestock, L., Hsieh, Y.-T., Porcelli, D., Homoky, W. B., Bryan, A., & Henderson, G. M. (2018). Controls on the barium isotope compositions of marine sediments. *Earth and Planetary Science Letters*, *481*, 101–110. https://doi.org/10.1016/j.epsl.2017.10.019

Cao, Z., Li, Y., Rao, X., Yu, Y., Hathorne, E. C., Siebert, C., Dai, M., & Frank, M. (2020). Constraining barium isotope fractionation in the upper water column of the South China Sea. *Geochimica et Cosmochimica Acta*, *288*, 120–137. https://doi.org/10.1016/j.gca.2020.08.008

Carter, S. C., Paytan, A., & Griffith, E. M. (2020). Toward an Improved Understanding of the Marine Barium Cycle and the Application of Marine Barite as a Paleoproductivity Proxy. *Minerals*, *10*(5), 421. https://doi.org/10.3390/min10050421

Chan, L. H., Drummond, D., Edmond, J. M., & Grant, B. (1977). On the barium data from the Atlantic GEOSECS expedition. *Deep Sea Research*, *24*(7), 613–649. https://doi.org/10.1016/0146-6291(77)90505-7

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

740 Chow, T. J., & Goldberg, E. D. (1960). On the marine geochemistry of barium. *Geochimica et*
741 *Cosmochimica Acta*, *20*(3), 192–198. https://doi.org/10.1016/0016-7037(60)90073-9

742 Coffey, M., Dehairs, F., Collette, O., Luther, G., Church, T., & Jickells, T. (1997). The Behaviour
743 of Dissolved Barium in Estuaries. *Estuarine, Coastal and Shelf Science*, *45*(1), 113–121.
744 https://doi.org/10.1006/ecss.1996.0157

745 Copernicus Marine Environment Monitoring Service. (2021). *Global Ocean Chlorophyll, PP and*
746 *PFT (Copernicus-GlobColour) from Satellite Observations: Monthly and Daily Interpolated*
747 *(Reprocessed from 1997)* [Data set]. Mercator Ocean International.
748 https://doi.org/10.48670/MOI-00100

749 Craig, H., & Turekian, K. K. (1980). The GEOSECS program: 1976-1979. Earth and Planetary
750 Science Letters, 49(2), 263-265. https://doi.org/10.1016/0012-821X(76)90062-5.

751 Crameri, F. (2018). Scientific colour maps. *Zenodo*. https://doi.org/10.5281/zenodo.5501399.

752 Crockford, P. W., Wing, B. A., Paytan, A., Hodgskiss, M. S. W., Mayfield, K. K., Hayles, J. A.,
753 Middleton, J. E., Ahm, A.-S. C., Johnston, D. T., Caxito, F., Uhlein, G., Halverson, G. P.,
754 Eickmann, B., Torres, M., & Horner, T. J. (2019). Barium-isotopic constraints on the origin
755 of post-Marinoan barites. *Earth and Planetary Science Letters*, *519*, 234–244.
756 https://doi.org/10.1016/j.epsl.2019.05.018.

757 Cutter, G. A. (2013). Intercalibration in chemical oceanography—getting the right number.
758 *Limnology and Oceanography: Methods*, **11**(7), 418-424.

759 Dehairs, F., Chesselet, R., & Jedwab, J. (1980). Discrete suspended particles of barite and the
760 barium cycle in the open ocean. *Earth and Planetary Science Letters*, *49*(2), 528–550.
761 https://doi.org/10.1016/0012-821X(80)90094-1

762 DeVries, T. (2014). The oceanic anthropogenic CO_2 sink: Storage, air-sea fluxes, and transports
763 over the industrial era. *Global Biogeochemical Cycles*, 28(7), 631-647.
764 https://doi.org/10.1002/2013GB004739

765 Dickens, G. R., Fewless, T., Thomas, E., & Bralower, T. J. (2003). Excess barite accumulation
766 during the Paleocene-Eocene thermal Maximum: Massive input of dissolved barium from
767 seafloor gas hydrate reservoirs. In S. L. Wing, P. D. Gingerich, B. Schmitz, & E. Thomas,
768 *Causes and consequences of globally warm climates in the early Paleogene*. Geological
769 Society of America. https://doi.org/10.1130/0-8137-2369-8.11

770 Dymond, J., Suess, E., & Lyle, M. (1992). Barium in Deep-Sea Sediment: A Geochemical Proxy
771 for Paleoproductivity. *Paleoceanography*, *7*(2), 163–181.
772 https://doi.org/10.1029/92PA00181.

773 Eagle, M., Paytan, A., Arrigo, K. R., van Dijken, G., & Murray, R. W. (2003). A comparison
774 between excess barium and barite as indicators of carbon export. *Paleoceanography*, 18(1).
775 https://doi.org/10.1029/2002PA000793

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

776 Eakins, B.W., & Sharman, G.F. (2010). Volumes of the World's Oceans from ETOPO1. *NOAA*
777 *National Geophysical Data Center, Boulder, CO.*
778 https://www.ngdc.noaa.gov/mgg/global/etopo1_ocean_volumes.html

779 Esser, B. K., & Volpe, A. M. (2002). At-sea high-resolution chemical mapping: Extreme barium
780 depletion in North Pacific surface water. *Marine Chemistry*, *79*(2), 67–79.
781 https://doi.org/10.1016/S0304-4203(02)00037-3

782 García H. E., K.W. Weathers, C.R. Paver, I. Smolyar, T.P. Boyer, R.A. Locarnini, M.M. Zweng,
783 A.V. Mishonov, O.K. Baranova, D. Seidov, and J.R. Reagan (2018a). World Ocean Atlas
784 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Dissolved Oxygen
785 Saturation. A. Mishonov Technical Editor. NOAA Atlas NESDIS 83, 38pp.
786 http://www.nodc.noaa.gov/OC5/indprod.html

787 García H.E., K.W. Weathers, C.R. Paver, I. Smolyar, T.P. Boyer, R.A. Locarnini, M.M. Zweng,
788 A.V. Mishonov, O.K. Baranova, D. Seidov, and J.R. Reagan (2018b). World Ocean Atlas
789 2018. Vol. 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate).
790 A. Mishonov Technical Editor, NOAA Atlas NESDIS 84, 35pp.
791 http://www.nodc.noaa.gov/OC5/indprod.html

792 GEOTRACES Intermediate Data Product Group (2021). The GEOTRACES Intermediate Data
793 Product 2021 (IDP2021). NERC EDS British Oceanographic Data Centre NOC.
794 doi:10.5285/cf2d9ba9-d51d-3b7c-e053-8486abc0f5fd.

795 Glover, D., Jenkins, W., & Doney, S. (2011). *Modeling Methods for Marine Science*. Cambridge:
796 Cambridge University Press. doi:10.1017/CBO9780511975721

797 Gonneea, M. E., Cohen, A. L., DeCarlo, T. M., & Charette, M. A. (2017). Relationship between
798 water and aragonite barium concentrations in aquaria reared juvenile corals. *Geochimica et*
799 *Cosmochimica Acta*, 209, 123-134.

800 Griffith, E. M., Fantle, M. S., Eisenhauer, A., Paytan, A., & Bullen, T. D. (2015). Effects of ocean
801 acidification on the marine calcium isotope record at the Paleocene–Eocene Thermal
802 Maximum. *Earth and Planetary Science Letters*, *419*, 81–92.
803 https://doi.org/10.1016/j.epsl.2015.03.010

804 Hathorne, E. C., Gagnon, A., Felis, T., Adkins, J., Asami, R., Boer, W., Caillon, N., Case, D.,
805 Cobb, K. M., Douville, E., deMenocal, P., Eisenhauer, A., Garbe-Schönberg, D., Geibert, W.,
806 Goldstein, S., Hughen, K., Inoue, M., Kawahata, H., Kölling, M., Cornec, F. L., Linsley, B.
807 K., McGregor, H. V., Montagna, P., Nurhati, I. S., Quinn, T. M., Raddatz, J., Rebaubier, H.,
808 Robinson, L., Sadekov, A., Sherrell, R., Sinclair, D., Tudhope, A. W., Wei, G., Wong, H.,
809 Wu, H. C., You, C.-F. (2013). Interlaboratory study for coral Sr/Ca and other element/Ca ratio
810 measurements. *Geochemistry, Geophysics, Geosystems*, *14*(9), 3730–3750.
811 https://doi.org/10.1002/ggge.20230

812 Hayes, C. T., Anderson, R. F., Cheng, H., Conway, T. M., Edwards, R. L., Fleisher, M. Q., Ho, P.,
813 Huang, K.-F., John, S. G., Landing, W. M., Little, S. H., Lu, Y., Morton, P. L., Moran, S. B.,
814 Robinson, L. F., Shelley, R. U., Shiller, A. M., & Zheng, X.-Y. (2018). Replacement Times

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

of a Spectrum of Elements in the North Atlantic Based on Thorium Supply. *Global Biogeochemical Cycles*, *32*(9), 1294–1311. https://doi.org/10.1029/2017GB005839

Hayes, C. T., Costa, K. M., Anderson, R. F., Calvo, E., Chase, Z., Demina, L. L., Dutay, J.-C., German, C. R., Heimbürger-Boavida, L.-E., Jaccard, S. L., Jacobel, A., Kohfeld, K. E., Kravchishina, M. D., Lippold, J., Mekik, F., Missiaen, L., Pavia, F. J., Paytan, A., Pedrosa-Pamies, R., Petrova, M. V., Rahman, S., Robinson, L. F., Roy-Barman, M., Sanchez-Vidal, A., Shiller, A., Tagliabue, A., Tessin, A. C., van Hulten, M., Zhang, J. (2021). Global Ocean Sediment Composition and Burial Flux in the Deep Sea. *Global Biogeochemical Cycles*, *35*(4), e2020GB006769. https://doi.org/10.1029/2020GB006769

Holte, J., Talley, L. D., Gilson, J., & Roemmich, D. (2017). An Argo mixed layer climatology and database. *Geophysical Research Letters*, *44*(11), 5618–5626. https://doi.org/10.1002/2017GL073426

Hood, E.M., C.L. Sabine, and B.M. Sloyan, eds. (2010). The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines. *IOCCP Report Number 14*, ICPO Publication Series Number 134. Available online at http://www.go-ship.org/HydroMan.html.

Hönisch, B., Allen, K. A., Russell, A. D., Eggins, S. M., Bijma, J., Spero, H. J., Lea, D. W., & Yu, J. (2011). Planktic foraminifers as recorders of seawater Ba/Ca. *Marine Micropaleontology*, *79*(1–2), 52–57. https://doi.org/10.1016/j.marmicro.2011.01.003

Hoppema, M., Dehairs, F., Navez, J., Monnin, C., Jeandel, C., Fahrbach, E., & de Baar, H. J. W. (2010). Distribution of barium in the Weddell Gyre: Impact of circulation and biogeochemical processes. *Marine Chemistry*, *122*(1), 118–129. https://doi.org/10.1016/j.marchem.2010.07.005

Horner, T. J., & Crockford, P. W. (2021). *Barium Isotopes: Drivers, Dependencies, and Distributions through Space and Time* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108865845.

Horner, T. J., Kinsley, C. W., & Nielsen, S. G. (2015). Barium-isotopic fractionation in seawater mediated by barite cycling and oceanic circulation. *Earth and Planetary Science Letters*, *430*, 511–522. https://doi.org/10.1016/j.epsl.2015.07.027

Hsieh, Y.-T., & Henderson, G. M. (2017). Barium stable isotopes in the global ocean: Tracer of Ba inputs and utilization. *Earth and Planetary Science Letters*, *473*, 269–278. https://doi.org/10.1016/j.epsl.2017.06.024

Horner, T. J., Mete, O. Z. (2023) A spatially and vertically resolved global grid of dissolved barium concentrations in seawater determined using Gaussian Process Regression machine learning. [Data Set]. *Biological and Chemical Oceanography Data Management Office* (BCO-DMO). (Version 1) Version Date 2023-02-22. https://www.bco-dmo.org/dataset/885506; doi:10.26008/1912/bco-dmo.885506.1.

Jacquet, S. H. M., Dehairs, F., & Rintoul, S. (2004). A high resolution transect of dissolved barium in the Southern Ocean. *Geophysical Research Letters*, 31(14). https://doi.org/10.1029/2004GL020016

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

854 Jeandel, C., Dupré, B., Lebaron, G., Monnin, C., & Minster, J.-F. (1996). Longitudinal
855 distributions of dissolved barium, silica and alkalinity in the western and southern Indian
856 Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 43(1), 1–31.
857 https://doi.org/10.1016/0967-0637(95)00098-4

858 John, S. G., Liang, H., Weber, T., DeVries, T., Primeau, F., Moore, K., Holzer, M., Mahowald,
859 N., Gardner, W., Mishonov, A., Richardson, M. J., Faugere, Y., & Taburet, G. (2020).
860 AWESOME OCIM: A simple, flexible, and powerful tool for modeling elemental cycling in
861 the oceans. *Chemical Geology*, 533, 119403. https://doi.org/10.1016/j.chemgeo.2019.119403

862 Joung, D., & Shiller, A. M. (2014). Dissolved barium behavior in Louisiana Shelf waters affected
863 by the Mississippi/Atchafalaya River mixing zone. *Geochimica et Cosmochimica Acta*, *141*,
864 303–313. https://doi.org/10.1016/j.gca.2014.06.021

865 Jullion, L., Jacquet, S. H. M., & Tanhua, T. (2017). Untangling biogeochemical processes from
866 the impact of ocean circulation: First insight on the Mediterranean dissolved barium
867 dynamics. *Global Biogeochemical Cycles*, *31*(8), 1256–1270.
868 https://doi.org/10.1002/2016GB005489

869 Komagoe, T., Watanabe, T., Shirai, K., Yamazaki, A., & Uematu, M. (2018). Geochemical and
870 Microstructural Signals in Giant Clam Tridacna Maxima Recorded Typhoon Events at
871 Okinotori Island, Japan. *Journal of Geophysical Research: Biogeosciences*, *123*(5), 1460–
872 1474. https://doi.org/10.1029/2017JG004082

873 LaVigne, M., Grottoli, A. G., Palardy, J. E., & Sherrell, R. M. (2016). Multi-colony calibrations
874 of coral Ba/Ca with a contemporaneous in situ seawater barium record. *Geochimica et*
875 *Cosmochimica Acta*, *179*, 203–216. https://doi.org/10.1016/j.gca.2015.12.038

876 LaVigne, M., Hill, T. M., Spero, H. J., & Guilderson, T. P. (2011). Bamboo coral Ba/Ca:
877 Calibration of a new deep ocean refractory nutrient proxy. *Earth and Planetary Science*
878 *Letters*, *312*(3), 506–515. https://doi.org/10.1016/j.epsl.2011.10.013

879 Lea, D. W., & Boyle, E. A. (1990). Foraminiferal reconstruction of barium distributions in water
880 masses of the glacial oceans. *Paleoceanography*, *5*(5), 719–742.
881 https://doi.org/10.1029/PA005i005p00719

882 Lea, D. W., Shen, G. T., & Boyle, E. A. (1989). Coralline barium records temporal variability in
883 equatorial Pacific upwelling. *Nature*, *340*(6232), 373–376. https://doi.org/10.1038/340373a0

884 Le Roy, E., Sanial, V., Charette, M. A., van Beek, P., Lacan, F., Jacquet, S. H. M., Henderson, P.
885 B., Souhaut, M., García-Ibáñez, M. I., Jeandel, C., Pérez, F. F., & Sarthou, G. (2018). The
886 ^{226}Ra–Ba relationship in the North Atlantic during GEOTRACES-GA01. *Biogeosciences*,
887 *15*(9), 3027–3048. https://doi.org/10.5194/bg-15-3027-2018

888 Light, T., & Norris, R. (2021). Quantitative visual analysis of marine barite microcrystals: Insights
889 into precipitation and dissolution dynamics. *Limnology and Oceanography*, *66*(10), 3619–
890 3629. https://doi.org/10.1002/lno.11902

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

Locarnini, R.A., A.V. Mishonov, O.K. Baranova, T.P. Boyer, M.M. Zweng, H.E. Garcia, J.R. Reagan, D. Seidov, K.W. Weathers, C.R. Paver, and I.V. Smolyar (2018). World Ocean Atlas 2018, Volume 1: Temperature. A. Mishonov, Technical Editor. NOAA Atlas NESDIS 81, 52pp. https://www.nodc.noaa.gov/OC5/indprod.html .

Monnin, C., Jeandel, C., Cattaldo, T., & Dehairs, F. (1999). The marine barite saturation state of the world's oceans. *Marine Chemistry*, *65*(3), 253–261. https://doi.org/10.1016/S0304-4203(99)00016-X

National Geophysical Data Center (1993). 5-minute Gridded Global Relief Data (ETOPO5). National Geophysical Data Center, NOAA. https://doi.org/10.7289/V5D798BF.

Paytan, A., & Griffith, E. M. (2007). Marine barite: Recorder of variations in ocean export productivity. *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*(5), 687–705. https://doi.org/10.1016/j.dsr2.2007.01.007

Paytan, A., Griffith, E. M., Eisenhauer, A., Hain, M. P., Wallmann, K., & Ridgwell, A. (2021). A 35-million-year record of seawater stable Sr isotopes reveals a fluctuating global carbon cycle. *Science*, *371*(6536), 1346–1350. https://doi.org/10.1126/science.aaz9266

Paytan, A., & Kastner, M. (1996). Benthic Ba fluxes in the central Equatorial Pacific, implications for the oceanic Ba cycle. *Earth and Planetary Science Letters*, *142*(3), 439–450. https://doi.org/10.1016/0012-821X(96)00120-3

Paytan, A., Kastner, M., Campbell, D., & Thiemens, M. H. (1998). Sulfur Isotopic Composition of Cenozoic Seawater Sulfate. *Science*, *282*(5393), 1459–1462. https://doi.org/10.1126/science.282.5393.1459

Paytan, A., Kastner, M., Martin, E. E., Macdougall, J. D., & Herbert, T. (1993). Marine barite as a monitor of seawater strontium isotope composition. *Nature*, *366*(6454), 445–449. https://doi.org/10.1038/366445a0

Pyle, K. M., Hendry, K. R., Sherrell, R. M., Legge, O., Hind, A. J., Bakker, D., Venables, H., & Meredith, M. P. (2018). Oceanic fronts control the distribution of dissolved barium in the Southern Ocean. *Marine Chemistry*, *204*, 95–106. https://doi.org/10.1016/j.marchem.2018.07.002

Rafter, P. A., Bagnell, A., Marconi, D., & DeVries, T. (2019). Global trends in marine nitrate N isotopes from observations and a neural network-based climatology. *Biogeosciences*, *16*(13), 2617–2633. https://doi.org/10.5194/bg-16-2617-2019

Rahman, S., Shiller, A. M., Anderson, R. F., Charette, M. A., Hayes, C. T., Gilbert, M., Grissom, K. R., Lam, P. J., Ohnemus, D. C., Pavia, F. J., Twining, B. S., & Vivancos, S. M. (2022). Dissolved and particulate barium distributions along the US GEOTRACES North Atlantic and East Pacific Zonal Transects (GA03 and GP16): Global implications for the marine barium cycle. *Global Biogeochemical Cycles*, *36*(6), e2022GB007330. https://doi.org/10.1029/2022GB007330.

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

928 Raju, K., & Atkinson, G. (1988). Thermodynamics of "scale" mineral solubilities. 1. Barium
929 sulfate(s) in water and aqueous sodium chloride. *Journal of Chemical & Engineering Data*,
930 *33*(4), 490–495. https://doi.org/10.1021/je00054a029

931 Roeske, T., Loeff, M. R. vd, Middag, R., & Bakker, K. (2012). Deep water circulation and
932 composition in the Arctic Ocean by dissolved barium, aluminium and silicate. *Marine*
933 *Chemistry*, *132–133*, 56–67. https://doi.org/10.1016/j.marchem.2012.02.001.

934 Rohatgi, A. (2022). WebPlotDigitizer. Version: 4.6, Pacifica, CA, USA.
935 https://automeris.io/WebPlotDigitizer.

936 Roshan, S., & DeVries, T. (2021). Global Contrasts Between Oceanic Cycling of Cadmium and
937 Phosphate. *Global Biogeochemical Cycles*, *35*(6). https://doi.org/10.1029/2021GB006952

938 Rushdi, A. I., McManus, J., & Collier, R. W. (2000). Marine barite and celestite saturation in
939 seawater. *Marine Chemistry*, *69*(1–2), 19–31. https://doi.org/10.1016/S0304-4203(99)00089-
940 4

941 Samanta, S., & Dalai, T. K. (2016). Dissolved and particulate Barium in the Ganga (Hooghly)
942 River estuary, India: Solute-particle interactions and the enhanced dissolved flux to the
943 oceans. *Geochimica et Cosmochimica Acta*, *195*, 1–28.
944 https://doi.org/10.1016/j.gca.2016.09.005

945 Schenau, S. J., Prins, M. A., De Lange, G. J., & Monnin, C. (2001). Barium accumulation in the
946 Arabian Sea: Controls on barite preservation in marine sediments. *Geochimica et*
947 *Cosmochimica Acta*, *65*(10), 1545–1556. https://doi.org/10.1016/S0016-7037(01)00547-6.

948 Schlitzer, R. (2023). Ocean Data View, https://odv.awi.de.

949 Schlitzer, R., Anderson, R. F., Dodas, E. M., Lohan, M., Geibert, W., Tagliabue, A., Bowie, A.,
950 Jeandel, C., Maldonado, M. T., Landing, W. M., Cockwell, D., Abadie, C., Abouchami, W.,
951 Achterberg, E. P., Agather, A., Aguliar-Islas, A., van Aken, H. M., Andersen, M., Archer, C.,
952 Auro M., de Baar H. J., Baars O., Baker A. R., Bakker K., Basak C., Baskaran M., Bates N.
953 R., Bauch D., van Beek P., Behrens M. K., Black E., Bluhm K., Bopp L., Bouman H.,
954 Bowman K., Bown J., Boyd P., Boye M., Boyle E. A., Branellec P., Bridgestock L., Brissebrat
955 G., Browning T., Bruland K. W., Brumsack H.J., Brzezinski M., Buck C. S., Buck K. N.,
956 Buesseler K., Bull A., Butler E., Cai P., Mor P. C., Cardinal D., Carlson C., Carrasco G.,
957 Casacuberta N., Casciotti K. L., Castrillejo M., Chamizo E., Chance R., Charette M. A.,
958 Chaves J. E., Cheng H., Chever F., Christl M., Church T. M., Closset I., Colman A., Conway
959 T. M., Cossa D., Croot P., Cullen J. T., Cutter G. A., Daniels C., Dehairs F., Deng F., Dieu
960 H. T., Duggan B., Dulaquais G., Dumousseaud C., EchegoyenSanz Y., Edwards R. L.,
961 Ellwood M., Fahrbach E., Fitzsimmons J. N., Flegal A. R., Fleisher M. Q., van de Flierdt T.,
962 Frank M., Friedrich J., Fripiat F., Fröllje H., Galer S. J. G., Gamo T., Ganeshram R. S.,
963 GarciaOrellana J., GarciaSolsona E., GaultRingold M., George E., Gerringa L. J. A., Gilbert
964 M., Godoy J. M., Goldstein S. L., Gonzalez S. R., Grissom K., Hammerschmidt C., Hartman
965 A., Hassler C. S., Hathorne E. C., Hatta M., Hawco N., Hayes C. T., Heimbürger L.E., Helgoe
966 J., Heller M., Henderson G. M., Henderson P. B., van Heuven S., Ho P., Horner T. J., Hsieh
967 Y.T., Huang K.F., Humphreys M. P., Isshiki K., Jacquot J. E., Janssen D. J., Jenkins W. J.,
968 John S., Jones E. M., Jones J. L., Kadko D. C., Kayser R., Kenna T. C., Khondoker R., Kim

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

T., Kipp L., Klar J. K., Klunder M., Kretschmer S., Kumamoto Y., Laan P., Labatut M., Lacan F., Lam P. J., Lambelet M., Lamborg C. H., Le Moigne F. A. C., Le Roy E., Lechtenfeld O. J., Lee J.M., Lherminier P., Little S., LópezLora M., Lü Y., Masqué P., Mawji E., McClain C. R., Measures C., Mehic S., Barraqueta J.L. M., van der Merwe P., Middag R., Mieruch S., Milne A., Minami T., Moffett J. W., Moncoiffe G., Moore W. S., Morris P. J., Morton P. L., Nakaguchi Y., Nakayama N., Niedermiller J., Nishioka J., Nishiuchi A., Noble A., Obata H., Ober S., Ohnemus D. C., van Ooijen J., OSullivan J., Owens S., Pahnke K., Paul M., Pavia F., Pena L. D., Peters B., Planchon F., Planquette H., Pradoux C., Puigcorbe V., Quay P., Queroue F., Radic A., Rauschenberg S., Rehkämper M., Rember R., Remenyi T., Resing J. A., Rickli J., Rigaud S., Rijkenberg M. J. A., Rintoul S., Robinson L. F., Roca-Martí M., Rodellas V., Roeske T., Rolison J. M., Rosenberg M., Roshan S., van der Loeff M. M. R., Ryabenko E., Saito M. A., Salt L. A., Sanial V., Sarthou G., Schallenberg C., Schauer U., Scher H., Schlosser C., Schnetger B., Scott P., Sedwick P. N., Semiletov I., Shelley R., Sherrell R. M., Shiller A. M., Sigman D. M., Singh S. K., Slagter H. A., Slater E., Smethie W. M., Snaith H., Sohrin Y., Sohst B., Sonke J. E., Speich S., Steinfeldt R., Stewart G., Stichel T., Stirling C. H., Stutsman J., Swarr G. J., Swift J. H., Thomas A., Thorne K., Till C. P., Till R., Townsend A. T., Townsend E., Tuerena R., Twining B. S., Vance D., Velazquez S., Venchiarutti C., VillaAlfageme M., Vivancos S. M., Voelker A. H. L., Wake B., Warner M. J., Watson R., van Weerlee E., Weigand M. A., Weinstein Y., Weiss D., Wisotzki A., Woodward E. M. S., Wu J., Wu Y., Wuttig K., Wyatt N., Xiang Y., Xie R. C., Xue Z., Yoshikawa H., Zhang J., Zhang P., Zhao Y., Zheng L., Zheng X.Y., Zieringer M., Zimmer L. A., Ziveri P., Zunino P. & Zurbrick C. (2018). The GEOTRACES Intermediate Data Product 2017. *Chemical Geology*, *493*, 210–223. https://doi.org/10.1016/j.chemgeo.2018.05.040

Schmitz. (1987). Barium, equatorial high productivity, and the northward wandering of the Indian continent. Paleoceanography, 2(1), 63–77. https://doi.org/10.1029/PA002i001p00063

Schroeder, J. O., Murray, R. W., Leinen, M., Pflaum, R. C., & Janecek, T. R. (1997). Barium in equatorial Pacific carbonate sediment: Terrigenous, oxide, and biogenic associations. *Paleoceanography*, *12*(1), 125–146. https://doi.org/10.1029/96PA02736.

Serno, S., Winckler, G., Anderson, R. F., Hayes, C. T., Ren, H., Gersonde, R., & Haug, G. H. (2014). Using the natural spatial pattern of marine productivity in the Subarctic North Pacific to evaluate paleoproductivity proxies. *Paleoceanography*, 29(5), 438-453. https://doi.org/10.1002/2013PA002594

Sherwen, T., Chance, R. J., Tinel, L., Ellis, D., Evans, M. J., & Carpenter, L. J. (2019). A machine-learning-based global sea-surface iodide distribution. *Earth System Science Data*, *11*(3), 1239–1262. https://doi.org/10.5194/essd-11-1239-2019

Sinclair, D. J., & McCulloch, M. T. (2004). Corals record low mobile barium concentrations in the Burdekin River during the 1974 flood: Evidence for limited Ba supply to rivers? *Palaeogeography, Palaeoclimatology, Palaeoecology*, *214*(1), 155–174. https://doi.org/10.1016/j.palaeo.2004.07.028

Singh, S. P., Singh, S. K., & Bhushan, R. (2013). Internal cycling of dissolved barium in water column of the Bay of Bengal. *Marine Chemistry*, *154*, 12–23. https://doi.org/10.1016/j.marchem.2013.04.013

https://doi.org/10.5194/essd-2023-67
Preprint. Discussion started: 2 March 2023

1011 Stewart, J. A., Li, T., Spooner, P. T., Burke, A., Chen, T., Roberts, J., Rae, J. W. B., Peck, V.,
1012 Kender, S., Liu, Q., & Robinson, L. F. (2021). Productivity and Dissolved Oxygen Controls
1013 on the Southern Ocean Deep-Sea Benthos During the Antarctic Cold Reversal.
1014 *Paleoceanography and Paleoclimatology*, *36*(10), e2021PA004288.
1015 https://doi.org/10.1029/2021PA004288

1016 Stroobants, N., Dehairs, F., Goeyens, L., Vanderheijden, N., & Van Grieken, R. (1991). Barite
1017 formation in the Southern Ocean water column. *Marine Chemistry*, *35*(1), 411–421.
1018 https://doi.org/10.1016/S0304-4203(09)90033-0

1019 Talley, L. D., Pickard, G. L., & Emery, W. J. (Eds.). (2011). *Descriptive physical oceanography:*
1020 *An introduction* (6th ed). Academic Press.

1021 Waldeck, A. R., Hemingway, J. D., Yao, W., Paytan, A., & Johnston, D. T. (2022). The triple
1022 oxygen isotope composition of marine sulfate and 130 million years of microbial control.
1023 *Proceedings of the National Academy of Sciences*, *119*(31), e2202018119.
1024 https://doi.org/10.1073/pnas.2202018119

1025 Whitmore, L. M., Shiller, A. M., Horner, T. J., Xiang, Y., Auro, M. E., Bauch, D., Dehairs, F.,
1026 Lam, P. J., Li, J., Maldonado, M. T., Mears, C., Newton, R., Pasqualini, A., Planquette, H.,
1027 Rember, R., & Thomas, H. (2022). Strong Margin Influence on the Arctic Ocean Barium
1028 Cycle Revealed by Pan-Arctic Synthesis. *Journal of Geophysical Research: Oceans*, *127*(4),
1029 e2021JC017417. https://doi.org/10.1029/2021JC017417

1030 Zweng, M.M, J.R. Reagan, D. Seidov, T.P. Boyer, R.A. Locarnini, H.E. Garcia, A.V. Mishonov,
1031 O.K. Baranova, K.W. Weathers, C.R. Paver, and I.V. Smolyar (2018). World Ocean Atlas
1032 2018, Volume 2: Salinity. A. Mishonov, Technical Editor, NOAA Atlas NESDIS 82, 50pp.
1033 http://www.nodc.noaa.gov/OC5/indprod.html