

Estimated nitrate d15N modeled using an ensemble of artificial neural networks (EANNs)

Website: <https://www.bco-dmo.org/dataset/768655>

Data Type: model results

Version: 1

Version Date: 2019-05-28

Project

» [Collaborative research: Combining models and observations to constrain the marine iron cycle](#) (Fe Cycle Models and Observations)

Contributors	Affiliation	Role
Rafter, Patrick	University of California-Irvine (UC Irvine)	Principal Investigator, Contact
Bagnell, Aaron	University of California-Santa Barbara (UCSB)	Co-Principal Investigator
DeVries, Timothy	University of California-Santa Barbara (UCSB)	Co-Principal Investigator
Marconi, Dario	Princeton University	Co-Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

We utilize an ensemble of artificial neural networks (EANNs) to interpolate our global ocean nitrate d15N database, producing complete 3D maps of the data. By utilizing an artificial neural network (ANN), a machine learning approach that effectively identifies nonlinear relationships between a target variable (the isotopic dataset) and a set of input features (other available ocean datasets), we can fill holes in our data sampling coverage of nitrate d15N.

Table of Contents

- [Coverage](#)
- [Dataset Description](#)
 - [Methods & Sampling](#)
 - [Data Processing Description](#)
- [Data Files](#)
- [Supplemental Files](#)
- [Related Publications](#)
- [Parameters](#)
- [Project Information](#)
- [Funding](#)

Coverage

Spatial Extent: N:83.5 E:180 S:-79.5 W:-180

Methods & Sampling

For complete methodology, refer to Rafter et al. (2019). In summary:

Data Compilation: Nitrate d15N observations were compiled from studies dating from 1975 to 2018. This global ocean nitrate d15N database was interpolated using an ensemble of artificial neural networks (EANNs). For the compiled observed global ocean nitrate d15N data, see the related dataset: <https://www.bco-dmo.org/dataset/768627>

Building the neural network model: We utilize an ensemble of artificial neural networks (EANNs) to interpolate our global ocean nitrate d15N database, producing complete 3D maps of the data. By utilizing an artificial neural network (ANN), a machine learning approach that effectively identifies nonlinear relationships

between a target variable (the isotopic dataset) and a set of input features (other available ocean datasets), we can fill holes in our data sampling coverage of nitrate d15N.

Binning target variables (Step 1): We binned the nitrate d15N observations to the World Ocean Atlas 2009 (WOA09) grid with a 1-degree spatial resolution and 33 vertical depth layers (0-5500 m). When binning vertically, we use the depth layer whose value is closest to the observation's sampling depth (e.g. the first depth layer has a value of 0 m, the second of 10 m, and the third of 20 m, so all nitrate isotopic data sampled between 0-5 m fall in the 0 m bin; between 5-15 m they fall in the 10 m bin, etc.). An observation with a sampling depth that lies right at the midpoint between depth layers is binned to the shallower layer. If more than one raw data point falls in a grid cell we take the average of all those points as the value for that grid cell. Certain whole ship tracks of nitrate d15N data were withheld from binning to be used as an independent validation set.

Obtaining input features (Step 2): Our input dataset contains a set of climatological values for physical and biogeochemical ocean parameters that form a non-linear relationship with the target data. We have six input features including objectively analyzed annual-mean fields for temperature, salinity, nitrate, oxygen, and phosphate taken from the WOA09 (<https://www.nodc.noaa.gov/OC5/WOA09/woa09data.html>) at 1-degree resolution. Additionally, daily chlorophyll data from Modis Aqua for the period Jan-1-2003 through Dec-31-2012 is averaged and binned to the WOA09 grid (as described in Step 1) to produce an annual climatological field of chlorophyll values, which we then log transform to reduce their dynamic range.

The choice of these specific input features was dictated by our desire to achieve the best possible R² value on our internal validation sets (Step 4). Additional inputs besides those we included, such as latitude, longitude, silicate, euphotic depth, or sampling depth either did not improve the R² value on the validation dataset or degraded it, indicating that they are not essential parameters for characterizing this system globally. By opting to use the set of input features that yielded the best results for the global oceans, we potentially overlooked combinations of inputs that perform better at regional scales. However, given the scarcity of d15N data in some regions, it is not possible to ascribe the impact of a specific combination of input features versus the impact of available d15N data, which may not be representative of the region's climatological state, to the relative model performance in these regions.

Training the ANN (Step 3): The architecture of our ANN consists of a single hidden layer, containing 25 nodes, that connects the biological and physical input features (discussed in Step 2) to the target nitrate isotopic variable (as discussed in Step 1). The role of the hidden layer is to transform input features into new features contained in the nodes. These are given to the output layer to estimate the target variable, introducing nonlinearities via an activation function. The number of nodes in this hidden layer, as well as the number of input features, determines the number of adjustable weights (the free parameters) in the network. For complete information, refer to Rafter et al. (2019).

Validating the ANN (Step 4): To ensure good generalization of the trained ANN, we randomly withhold 10% of the d15N data to be used as an internal validation set for each network. This is data that the network never sees, meaning it does not factor into the cost function, so it works as a test of the ANN's ability to generalize. This internal validation set acts as a gatekeeper to prevent poor models from being accepted into the ensemble of trained networks (see Step 5). A second, independent or 'external' validation set, composed of complete ship transects from the high and low latitude ocean were omitted from binning in Step 1 and used to establish the performance of the entire ensemble. Our rationale for using complete ship transects is the following. If we randomly chose 10% of observations to perform an external validation, this dataset will be from the same cruises as the wider data. In other words, despite being randomly selected, the validating observational dataset will be highly correlated geographically. Contrast this with validating the EANN results with observations from whole research cruises in unique geographic regions—areas where the model has not "learned" anything about nitrate. We therefore argue that these observations from whole ship tracks therefore provide a more difficult test of the model.

Forming the Ensemble (Step 5): The ensemble is formed by repeating Steps 3 to 4 (using a different random 10% validation set) until we obtain 25 trained networks for the nitrate d15N dataset. A network is admitted into the ensemble if it yields an R² value greater than 0.81 on the validation dataset. For complete information, refer to Rafter et al. (2019).

Data Processing Description

See supplemental MatLab files:

[d15NO3_ANN.m](#) - This script creates an ensemble of Artificial Neural Networks (ANNs) to interpolate delta15N

of nitrate observations globally using ocean datasets from the World Ocean Atlas 2009.

[PR19_d15NO3_ANN_anclim_inputs](#) - input data for d15NO3_ANN.m

[PR19_d15NO3_ANN_anclim_outputs](#) - output of d15NO3_ANN.m

[[table of contents](#) | [back to top](#)]

Data Files

File
d15N_MODEL.csv (Comma Separated Values (.csv), 30.76 MB) MD5:f939fe2154877eaf069593da1afdd7a1
Primary data file for dataset ID 768655

[[table of contents](#) | [back to top](#)]

Supplemental Files

File	
d15NO3_ANN.m	(MATLAB Programming Script (.m), 2.93 KB) MD5:81acc3254a181d73c836834d668d777f
This script creates an ensemble of Artificial Neural Networks (ANNs) to interpolate delta15N of nitrate observations globally using ocean datasets from the World Ocean Atlas 2009.	
PR19_d15NO3_ANN_anclim_inputs.mat	(MATLAB Programming Script (.m), 42.53 MB) MD5:6c315ee3007cd4c38e9d8e0ffa61ad7a
Input data to script d15NO3_ANN.m	
PR19_d15NO3_ANN_anclim_outputs.mat	(MATLAB Programming Script (.m), 16.76 MB) MD5:76305301995ee71c0f97d7ab9bbe5eee
Output of script d15NO3_ANN.m	

[[table of contents](#) | [back to top](#)]

Related Publications

Rafter, P. A., & Sigman, D. M. (2015). Spatial distribution and temporal variation of nitrate nitrogen and oxygen isotopes in the upper equatorial Pacific Ocean. *Limnology and Oceanography*, 61(1), 14–31.

doi:[10.1002/lno.10152](https://doi.org/10.1002/lno.10152)

General

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 Discussions*, 1–31. doi:[10.5194/bg-2018-525](https://doi.org/10.5194/bg-2018-525)

Results

Rafter, P. A., Sigman, D. M., Charles, C. D., Kaiser, J., & Haug, G. H. (2012). Subsurface tropical Pacific nitrogen isotopic composition of nitrate: Biogeochemical signals and their transport. *Global Biogeochemical Cycles*, 26(1), n/a–n/a. doi:10.1029/2010gb003979 <https://doi.org/10.1029/2010GB003979>

General

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
latitude	Latitude in degrees north	degrees North
longitude	Longitude in degrees East	degrees East
depth	Depth	meters (m)
d15N	modeled nitrate d15N	per mil
d15N_stdev	standard deviation	per mil

[[table of contents](#) | [back to top](#)]

Project Information

Collaborative research: Combining models and observations to constrain the marine iron cycle (Fe Cycle Models and Observations)

NSF Award Abstract:

Tiny marine organisms called phytoplankton play a critical role in Earth's climate, by absorbing carbon dioxide from the atmosphere. In order to grow, these phytoplankton require nutrients that are dissolved in seawater. One of the rarest and most important of these nutrients is iron. Even though it is a critical life-sustaining nutrient, oceanographers still do not know much about how iron gets into the ocean, or how it is removed from seawater. In the past few years, scientists have made many thousands of measurements of the amount of dissolved iron in seawater, in environments ranging from the deep sea, to the Arctic, to the tropical oceans. They found that the amount of iron in seawater varies dramatically from place to place. Can this data tell us about how iron gets into the ocean, and how it is ultimately removed? Yes. In this project, scientists working on making measurements of iron in seawater will come together with scientists who are working on computer models of iron inputs and removal in the ocean. The goal is to work together to create a program that allows our computer models to "learn" from the data, much like an Artificial Intelligence program. This program will develop a "best estimate" of where and how much iron is coming into the ocean, how long it stays in the ocean, and ultimately how it gets removed. This will lead to a better understanding of how climate change will impact the delivery of iron to the ocean, and how phytoplankton will respond to climate change. With better climate models, society can make more informed decisions about how to respond to climate change. The study will also benefit a future generation of scientists, by training graduate students in a unique collaboration between scientists making seawater measurements, and those using computer models to interpret those measurements. Finally, the project aims to increase the participation of minority and low-income students in STEM (Science, Technology, Engineering, and Mathematics) research, through targeted outreach programs.

Iron (Fe) is an important micronutrient for marine phytoplankton that limits primary productivity over much of the ocean; however, the major fluxes in the marine Fe cycle remain poorly quantified. Ocean models that attempt to synthesize our understanding of Fe biogeochemistry predict widely different Fe inputs to the ocean, and are often unable to capture first-order features of the Fe distribution. The proposed work aims to resolve these problems using data assimilation (inverse) methods to "teach" the widely used Biogeochemical Elemental Cycling (BEC) model how to better represent Fe sources, sinks, and cycling processes. This will be achieved by implementing BEC in the efficient Ocean Circulation Inverse Model and expanding it to simulate the cycling of additional tracers that constrain unique aspects of the Fe cycle, including aluminum, thorium, helium and Fe isotopes. In this framework, the inverse model can rapidly explore alternative representations of Fe-cycling processes, guided by new high-quality observations made possible in large part by the GEOTRACES program. The work will be the most concerted effort to date to synthesize these rich datasets into a realistic and mechanistic model of the marine Fe cycle. In addition, it will lead to a stronger consensus on the magnitude of fluxes in the marine Fe budget, and their relative importance in controlling Fe limitation of marine ecosystems,

which are areas of active debate. It will guide future observational efforts, by identifying factors that are still poorly constrained, or regions of the ocean where new data will dramatically reduce remaining uncertainties and allow new robust predictions of Fe cycling under future climate change scenarios to be made, ultimately improving climate change predictions. A broader impact of this work on the scientific community will be the development of a fast, portable, and flexible global model of trace element cycling, designed to allow non-modelers to test hypotheses and visualize the effects of different processes on trace metal distributions. The research will also support the training of graduate students, and outreach to low-income and minority students in local school districts.

[[table of contents](#) | [back to top](#)]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1658392

[[table of contents](#) | [back to top](#)]