

Interactive comment on “Global trends in marine nitrate N isotopes from observations and a neural network-based climatology by Patrick A. Rafter et al.
Anonymous Referee #1
Received and published: 1 February 2019

The original comment is in bold font. The response to the comment is in regular font.

Overview: The paper targets a useful goal – providing a map of 15N-NO3 estimates for the global ocean for use in biogeochemical studies. To do this, it uses a neural network to obtain a relationship between sparse observed 15N-NO3 and World Ocean Atlas (WOA) values of temperature, salinity, oxygen, phosphate and nitrate, and then maps the derived 15N-NO3 estimates. The utility of the approach is assessed via correlation statistics between the estimates and the observations. There are areas where the estimates and observations agree well and others where they agree poorly. The latter are ascribed to temporal offsets between the WOA data collection and the 15N-NO3 observations.

To be clear, our interpretation of the observation-model comparison is that the model estimates the mean values quite well, but does not include temporal variability and therefore will not capture temporal variability. (beginning LINE 381; revised manuscript)

As far as it goes, the paper is sound, but it doesn’t go very far (as an aside it does provide clear and well-constructed descriptions of possible mechanistic causes of the spatial variations in the 15N-NO3 observations, although these do not really derive from or depend on the mapping exercise). It could be improved by addressing the following issues: 1. Is the neural network (NN) approach demonstrably better than a multiple linear regression (MLR) to the same input variables? Assessing this would be useful for two reasons: a. The MLR has the advantage that it provides a simple equation that all can use with their local and future input variable observations [(notably MLR approaches are becoming widely used for nitrate in the context of BGC-Argo observations; Carter et al. 2017, <https://doi.org/10.1002/lom3.10232>] b. Determining whether and in which parts of the ocean the non-linear NN approach out-performs the linear MLR approach is likely to shed light on the processes that drive 15N-NO3 variations.

Great comment. To address this we built a single global Multiple Linear Regression (MLR) model using all the same predictors used in the Ensemble Array of Neural Networks (EANN). We found that the MLR performs much worse than the EANN at predicting nitrate $\delta^{15}\text{N}$. The coefficient of determination for each method and each ocean basin’s upper 1000 m is shown in the table below.

	Atlantic	Pacific	Indian	Southern Ocean
MLR R2	0.04	0.49	0.51	0.34
EANN R2	0.53	0.78	0.76	0.68

The reason for this worse performance is likely that the MLR approach assumes the training parameters are independent of each other, but also dependent on nitrate $\delta^{15}\text{N}$. This is not the case and so the EANN approach performs noticeably better.

2. Are there other metrics that could assess possible causes of the quality of the matches and mismatches between estimates and observations, to go beyond simply ascribing them to temporal offsets? For example since some of the 15N-NO₃ estimates were probably collected synchronously with the WOA data, do these points show closer agreement?

We do not ascribe differences between model and observations to temporal offsets. We suggest that the model predicts an annual climatology of nitrate $\delta^{15}\text{N}$, while the observations measure the instantaneous $\delta^{15}\text{N}$. There is no temporal component in the ANN. The WOA data that we are using are the annual climatologies – there are no corresponding observations of $\delta^{15}\text{N}$.

Can agreement with mechanistic understanding be assessed – for example in regions where single processes largely dominate 15N- NO₃ variations (e.g. nitrate assimilation in Southern Ocean surface waters) does the NN approach produce sensible correlations between [nitrate] and 15N-NO₃ ?

This is a good suggestion, but we find that adding an additional analysis of the regional model estimates is beyond the scope of this paper. In fact, we are already using the EANN results to examine global nitrate uptake patterns in a current study that will be outlined in a dedicated manuscript.

Details: Line 63: ammonia assimilation is also a significant determinant of the 15N of organic matter.

We revised the manuscript to clarify that these sentences refer to organic matter production by the assimilation of nitrate. Good comment. (LINE 60; revised manuscript)

Line 370: meaning of sentence beginning “Equivalent processes... was opaque.

The revised manuscript clarifies this sentence. It refers to how the model nitrate $\delta^{15}\text{N}$ predicts that intermediate water nitrate $\delta^{15}\text{N}$ in the Indian Ocean has a similar value as the corresponding waters in the Pacific. We argue that this is likely because “equivalent processes” established the pre-formed characteristics of both water masses (i.e., partial nitrate assimilation in the Southern Ocean surface). (LINE 556; revised manuscript)

Lines 384-395: This discussion of separating nitrification from denitrification influences on deep water 15N-NO₃ values would benefit from recognition that relationships with O₂ and nitrate have opposite signs.

Good comment. The well-known south-to-north lowering of deep Pacific O₂ and increase in nitrate concentrations is consistent with the remineralization of organic matter and not the lateral advection of nitrate from ODZ regions. This will be added to the revised manuscript. (LINE 633; revised manuscript)

Line 403: The estimate low sinking organic matter $\delta^{15}\text{N}$ estimate of +1.5 should be compared to published results in Lourey et al., 2003, which show good

agreement.

We have added and refer to this citation's results in the revised manuscript.

Interactive comment on “Global trends in marine nitrate N isotopes from observations and a neural network-based climatology” by Patrick A. Rafter et al.

Anonymous Referee #2

Received and published: 16 February 2019

There are many detailed responses to Reviewer #2’s comments. We have stated where these responses translate into revised text in the manuscript. Please let us know if there are any comments that should also drive a revision of manuscript text.

The original comment is in bold font. The response to the comment is in regular font.

The nitrate isotope database and gridded product generated by the authors has the potential to be extremely valuable for studies of the marine nitrogen cycle. I commend them for undertaking this important task, which will benefit researchers broadly. Because it does have such strong potential utility, I would really like to see the paper describe a bit more clearly what was actually done here, and how it compares with other methods of data gridding.

In particular, I think the authors should further explain and reference the neural network model used to generate the gridded product. There’s only one paper in the references, from 1996, that seems to relate at all to the methods they applied. More detail should be given here so that the results could be reproduced, or extended as additional nitrate isotope data become available.

Next, the discussion and conclusions about the marine nitrogen cycle were largely confirmatory of earlier studies, but also almost beside the point of this particular manuscript. I would have found it more interesting, in the context of what was done here, to see how this kind of approach to data binning compares to alternative methods. Are there significant difference between this neural network approach, and a World Ocean Atlas approach of data interpolation? What are the implications of some of the choices made in building the model?

Specific comments are given below.

Lines 106-111: How does this neural network actually work? Does it use learning based on surrounding data to inform the values of unknown points? Where are the equations that go into the model? What is/are the function(s) that produces d15N values from the gridded T, S, NO3-, O2, and PO43- data?

Our neural network has no explicit spatial component. We do not use latitude, longitude, or sampling depth as inputs to the model. Instead our model is purely a nonlinear function of physical and biological ocean parameters such as T, S NO3, etc. that all have implicit spatial characteristics. The model learns the relationship between d15N and these parameters for the locations where there are d15N

observations and, since we are using fields from the World Ocean Atlas (WOA) that have data everywhere, the model can estimate d15N for the locations where there are no observations using the nonlinear relationship it has learned. The function that models the relationship between d15N and training inputs is

$$d15N = a(a(I*W1+B1)*W2+B2)$$

where a is our activation function, which in this case is the hyperbolic tangent, I (size 7,170 binned observations by 6 input parameters) is our array of inputs [T S NO3 O2 ...], and $W1$ (size 6 by 25), $W2$ (size 25 by 1), $B1$ (size 25 by 1), and $B2$ (size 1 by 1) are our adjustable free parameters.

As a simple example, let us assume our only inputs (I) are T and S and they connect to a single node in the hidden layer. In this case, there are three total weights. One weight connects T to the hidden layer, one connects S , and another weight connects the hidden layer to the predicted d15N value. Let us also assume our activation function (a) is linear so we do not need to normalize our input data, and our bias weights ($B1$, $B2$) are zero. This simplifies the above equation to

$$d15N = (I*W1)*W2 = (T*w_{11}+S*w_{12})*w_{21}$$

For a single temperature and salinity pair (25 °C, 33 PSU) and initial weights $w_{11} = 0.5$ °C⁻¹, $w_{12} = 0.5$ PSU⁻¹, and $w_{21} = 0.2$ permil

$d15N = (25 * 0.5 + 33 * 0.5) * 0.2 = 5.8$ permil. This is a predicted value. If our target value were 6 permil only small adjustments to the value of the weights would be necessary to match that observation. This works for a single observation. In reality, we have thousands of observations we would like to optimally match our predictions to, while at the same time not overfitting.

Lines 116-119: Please clarify the description of depth binning.

An observation is binned to the depth layer closest to its sampling depth. Observations with sampling depths at the midpoint between layers in the model grid are binned to the shallower layer. We have updated the manuscript accordingly. (LINE 117; new manuscript)

Lines 122-123: Why were whole ship tracks used for validation, rather than a more random selection?

Our rationale for using whole ship tracks will be more clearly detailed in the revised manuscript and will be similar to the following response.

Note that this comment refers to our external validation, which is in addition to an internal validation that uses randomly selected data.

Imagine that we have a dataset that is made up of many cruises and we use a randomly selected 20% of this dataset for internal validation and another randomly selected 10% of this data to perform an external validation. Despite being randomly

selected, the external validating dataset will be from the same cruises as the wider data. In other words, despite being randomly selected, the validating dataset will be highly correlated geographically.

Instead, we have selected several cruises where none of the data was used to teach the model. These cruises are in areas where the model has not “learned” anything about nitrate and these data therefore provide a more difficult test of the model. (LINE 254; revised manuscript)

Line 131: How was the daily chlorophyll used in an otherwise annual gridded product?

We have updated the manuscript to clarify that daily chlorophyll data from the specified time period is not only binned to the model grid but also averaged to produce an annual field. (LINE 133; revised manuscript)

Section 2.2 needs more references, especially 2.2.3 (lines 137-151). There is a lot of terminology here that is not defined or referenced, such as hidden layer, node, activation function, which should be defined and explained further. Also, it is not clear what you are applying weights to in the model.

We have updated the text to provide a brief description of the neural network workflow, including defining some of the terms used and including a few additional citations. Weights form a linear system using inputs from the prior layer to produce values for the nodes in the next layer, as defined in a previous response. Using an activation function transforms this linear system to a nonlinear system. The hidden layer acts as intermediary between the input features and the target variable. Each of its nodes act as targets for the input layer and inputs for the final target layer. This increases the amount of learning the model can achieve by adding additional free parameters in the form of connections between nodes in one layer and nodes in the next. (LINE 106; revised manuscript)

Line 158: It says that 10% of the observations were withheld to validate the networks. How were these chosen? More generally, how were the data for training, test, and validation chosen?

We specify that 10% of the data is withheld randomly, but we updated the manuscript to clarify that EACH individual network has a random 10 percent withheld. This means each individual network sees a somewhat different training and test set. Some of the training data for one might be test data for another, and vice versa. Our final external validation set contains data that no individual network had available to it for training and is used to test the performance of the ensemble mean. (LINE 266; revised manuscript)

Line 165: What are the implications of using whole cruise tracks for external validation rather than randomly chosen stations or grid cells?

We responded to this above and will update the manuscript accordingly.

Lines 179-180: Could this be shown (that the ensemble performs better than any single member of the ensemble) using your results, or is this a general feature? Does it apply here?

This is a general feature noted by Breiman (1996) that applies to certain machine learning methods such as EANNs. As our method uses EANNs, it applies here as well and the R^2 values of the internal validation sets versus the ensemble mean is greater than the R^2 value of each individual ensemble member because the ensemble mean incorporates members that saw different data during training. This does not necessarily apply to the external validation set, as that is comprised of data that no member has seen. However, the ensemble mean performs better than 19 out of 25 of the ensemble members on the external validation set in terms of a greater R^2 value and lower RMSE. Recall also that, since we curated ensemble members by first using the internal validation sets, these members are all higher performers, so the odds of roughly 1 in 5 of picking an ensemble member that does better on this particular external validation set is an overestimate of the actual odds if members were not curated. This is something that will be clarified in the updated manuscript. (LINE 330; revised manuscript)

Discussion section:

How does the discussion stem from their results from the neural network model? Most of the discussion seems to focus on general features discussed in the original papers about the data used to generate the product. It would be more satisfying for this reviewer to read about how some of the choices they made in producing the model impacted the results.

In order to reply to previous comments, the revised manuscript will necessarily have much more information on the inner workings of the model and how these choices impact the results. Hopefully these will address the immediate concerns of the Reviewer. (beginning SECTION 2.2; revised manuscript)

However, speaking as an observationalist (this is Rafter writing), I believe the most logical discussion of these modeling results requires an examination of how they fit with the published literature. As such, the Discussion section uses the model results to provide insight to marine nitrate $\delta^{15}\text{N}$ that was previously hampered by poor geographic coverage.

For example, 1) Is there only one way to produce the neural network model?

1. A neural network model is a very general method, so there are many different ways to set up the architecture of the network, including number of hidden layers, size of hidden layers, how nodes in the hidden layer are activated, the type and number of input features we choose to include or not include, and the training algorithm among others. Aspects of these are covered by Rumelhart et al. (1986), Hornik et al. (1989), Weigand et al. (1990), and Thimm and Fiesler (1997).

2) How were choices made? What tradeoffs were tolerated? What are the implications?

2. The rationale for some of these choices were explicitly stated in section 2.2.3 of the manuscript, such as using only one hidden layer with 25 nodes in order to keep the number of weights (free parameters) low relative to the number of training data, or our use of a hyperbolic tangent activation function.

Other choices were not explicitly stated and will be in the revised manuscript. For instance, the specific choice of our input features was dictated by our desire to achieve the best possible R^2 value on our internal validation sets. Additional inputs besides those we included, such as latitude, longitude, silicate, euphotic depth, or sampling depth either did not improve the R^2 value or degraded it, indicating that they are not essential parameters for characterizing this system.

Every choice was made for model simplicity, accuracy or a combination of the two. The inclusion of larger networks in terms of more input parameters resulted in models that did not generalize as well to new data, as indicated by their degraded performance on test data. Larger networks in terms of hidden layers and nodes increase each individual network's ability to learn on training data by virtue of there being more free parameters, but there is a general rule of how large a network should be relative to the amount of training data, as discussed by Weigand et al. (1990), and we tried to stay well within it.

One potential tradeoff is that other combinations of input features might better apply to certain regions than others. We opted to use the set of input features that yielded the best results globally, but on a regional scale other combinations of inputs may be better.

Having created a globally optimized, annual d15N climatology, there are several implications to consider. While, our external validation set demonstrates our model generalizes well to certain regions, it is clear that our model does not perform equally well everywhere. We opted for overall accuracy in our model, so for regions with relatively poor fit it is unclear whether this is due to our chosen combination of input features not working as well for a specific region or whether it is due to training data that is not representative of the mean state of d15N in that region.

3) How does this approach compare with other methods for gridding?

3. Standard interpolation techniques such as objective mapping would not work here, especially at 1-degree resolution and 33 vertical depth levels, due to the sparseness of the d15N data. Ocean parameters from the WOA, for

instance, have much greater sampling density in order to create the interpolated fields. The EANN approach is more appropriate for sparse data, as it forms a relationship with more highly sampled ocean parameters to estimate d15N. There are many possible methods to model the relationship between these parameters and d15N, but simpler methods lack the complexity to adequately match the training data, let alone extrapolate well to new data. As an example, we built a single global Multiple Linear Regression (MLR) model using all the same predictors used in the Ensemble Array of Neural Networks (EANN). We found that the MLR performs much worse than the EANN at predicting nitrate $\delta^{15}\text{N}$. The coefficient of determination for each method and each ocean basin's upper 1000 m is shown in the table below.

	Atlantic	Pacific	Indian	Southern Ocean
MLR R2	0.04	0.49	0.51	0.34
EANN R2	0.53	0.78	0.76	0.68

4) Are there particular nodes that performed well in some locations vs. others?

4. Because we randomly sampled from available observations to create the training data for each network, this sampling is pretty evenly distributed spatially. The same applies to test data. Since each network had to pass the same criteria on the test set in order to be admitted into the ensemble the individual networks do not greatly differ in their performance in regions where there is data, especially given that we optimized our combination of input parameters for a global analysis and did not consider different combinations that might lead to better regional performance.

There are certain fairly large areas of the ocean where no observational data was available for this analysis. In these areas the individual ensemble members generate a larger range of estimates, as there is higher uncertainty about what the “truth” is. In these cases, the ensemble mean can be seen as splitting the difference or taking the most likely scenario of the estimates of d15N in these regions. That is the benefit of using the ensemble, as it provides the best general fit for the global ocean. The uncertainties of the EANN predictions are illustrated in Figure 5.

Lines 415-423: It's not clear how the authors 'easily dismiss' an explanation about lateral advection of elevated nitrate d15N from ODZ regions. I think this section should be clarified. The way they set it up (seeing an increase in the Pacific but not Atlantic) does not seem to further the argument they are trying to make since the largest ODZ regions are in the Pacific, not the Atlantic.

This discussion (which will be revised in the new manuscript) refers to deep Pacific nitrate $\delta^{15}\text{N}$, which increases from the Southern to Northern hemisphere. Similarly, deep Pacific waters originate at the Southern Ocean surface and move from the southern to northern hemisphere. An important addition to this discussion (suggested by Reviewer 1) is that while deep Pacific nitrate $\delta^{15}\text{N}$ increases from south-to-north, dissolved oxygen concentrations DECREASE and nitrate concentrations INCREASE. Grouping these observations together we have: (1) abyssal Pacific circulation moves from south-to-north, (2) oxygen decreases, (3) nitrate concentration increases, and (4) nitrate $\delta^{15}\text{N}$ increases. Taken together, these known changes in deep Pacific waters are a persuasive argument that the change in deep Pacific nitrate $\delta^{15}\text{N}$ originates from the remineralization of sinking organic matter (i.e., ammonification and nitrification of organic matter N).

The confusing part of this discussion (pointed out by the reviewer) is that this south-to-north elevation of deep Pacific nitrate $\delta^{15}\text{N}$ cannot be explained by the lateral advection (i.e., along isopycnal) transport of high nitrate $\delta^{15}\text{N}$ from the upper Pacific ODZ regions. This is because this explanation predicts that the highest nitrate $\delta^{15}\text{N}$ would be found where shallow Pacific waters are first converted into deep Pacific waters in the deep South Pacific. Because this is the opposite of what we observe, it cannot explain the data. (LINE 633; revised manuscript)

Figure 2. How many different selections of training, test, and validation sets did the authors test in the neural network model? What was the rationale behind the choice of the whole cruise tracks that were used for validation?

This was answered above and new text will be available in the revised manuscript.

Figure 3. Panel C was helpful. Panels A and B were also useful, but the choice of the non-linear color scale bar, where most of the data points were off scale, was unusual. In panel A, also please clarify whether this includes all of the data, or just those from the training set? Or validation set?

We have adjusted the color bar in the revised manuscript (and see below). This figure includes all of the data where there are model results.

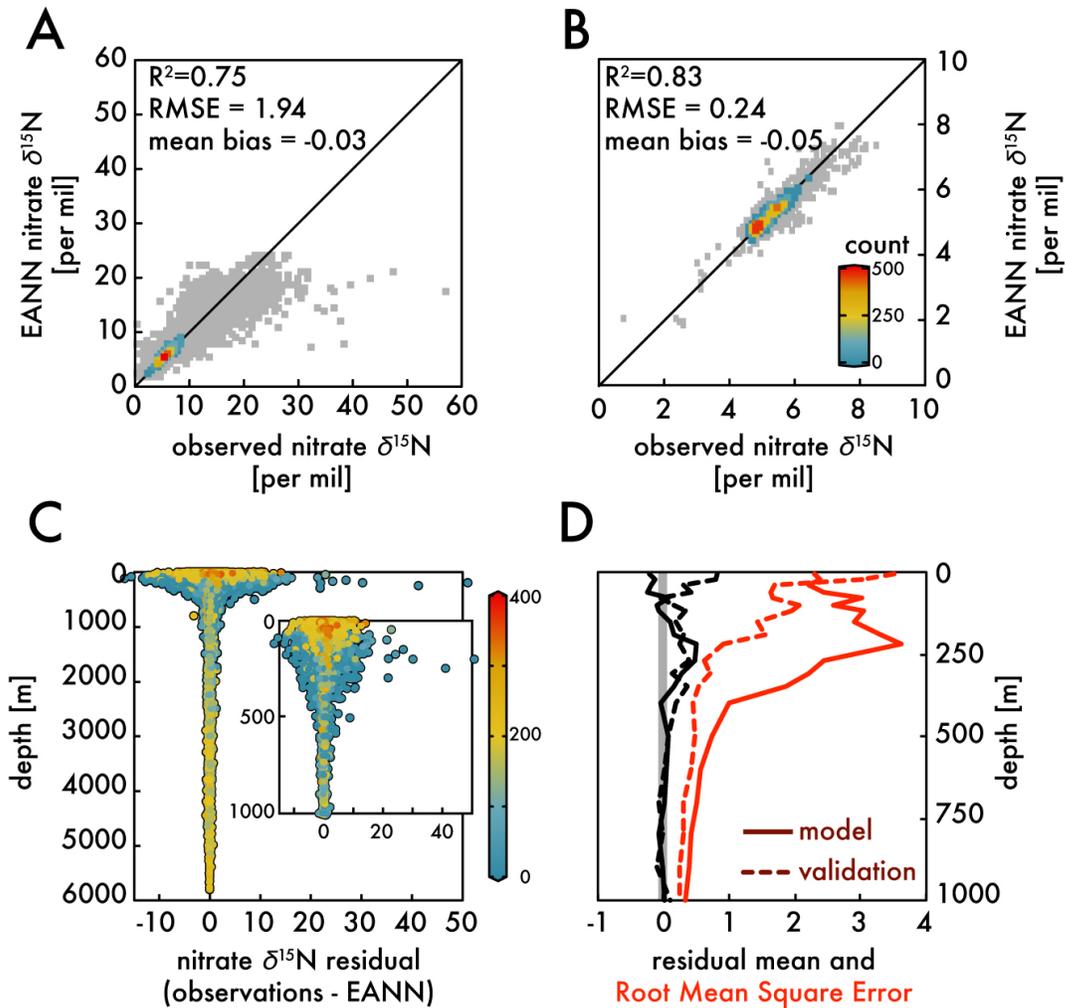


Figure 4. The statistics for the different zonally averaged sections were useful, but I question the utility of the zonally averaged Pacific, given some of the large zonal gradients in d15N from the ODZs in the eastern tropical Pacific.

We agree that they obscure the strong zonal gradients that occur in the lower latitude upper Pacific. But we also find them to be useful sources of discussion (for example the trends in deep Pacific nitrate $\delta^{15}\text{N}$). We will highlight the limitations of zonally-averaged view in the revised manuscript.

Figure 5. The contours were extremely difficult to read, and the panels on the right hand side (E-H) were not particularly helpful. I also wondered how much of the patchiness, especially in panel A, is driven by the distribution of available d15N data?

The revised Figure 5 can be seen below. We have discretized the color bar to more clearly indicate the contour value and use a color bar instead of black and white contours to show the standard deviation (right).

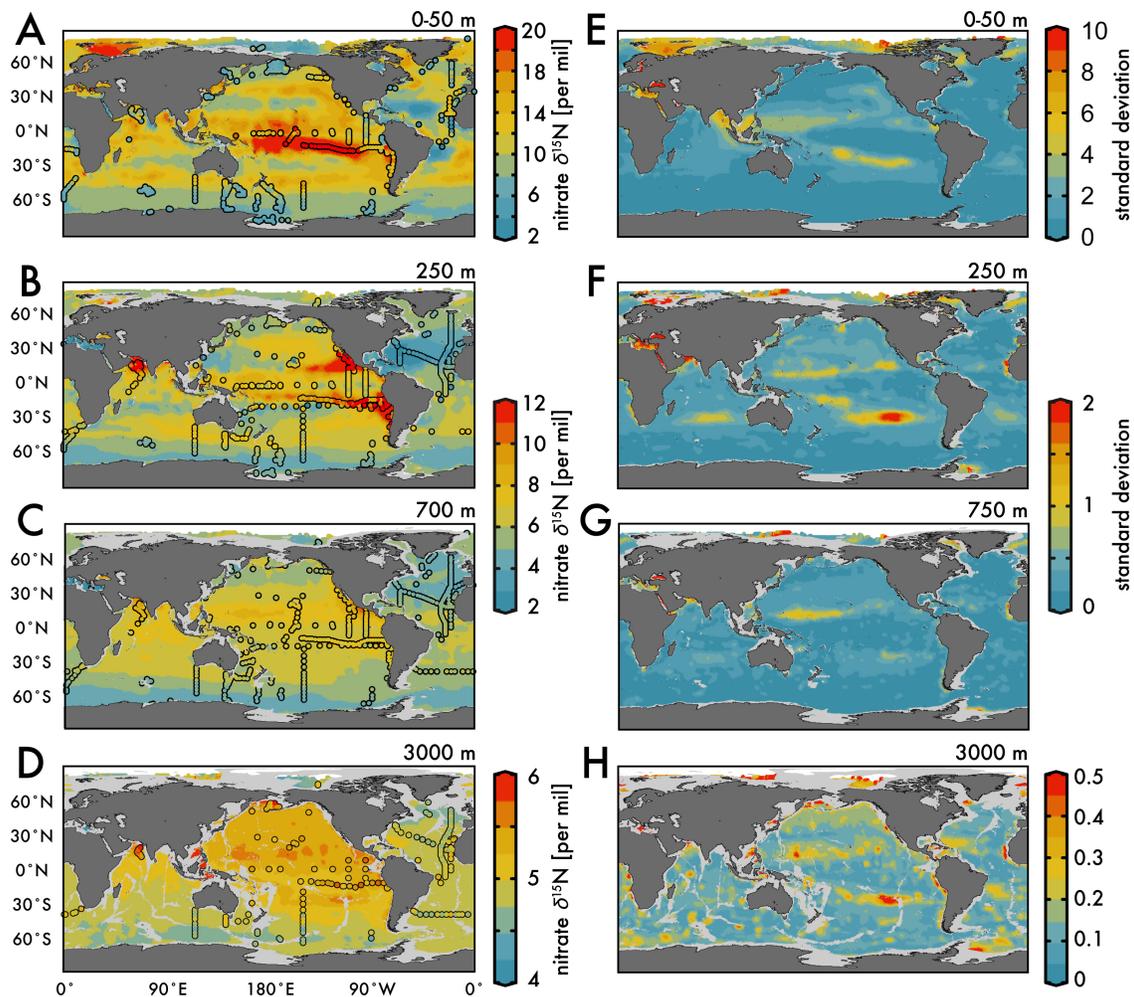
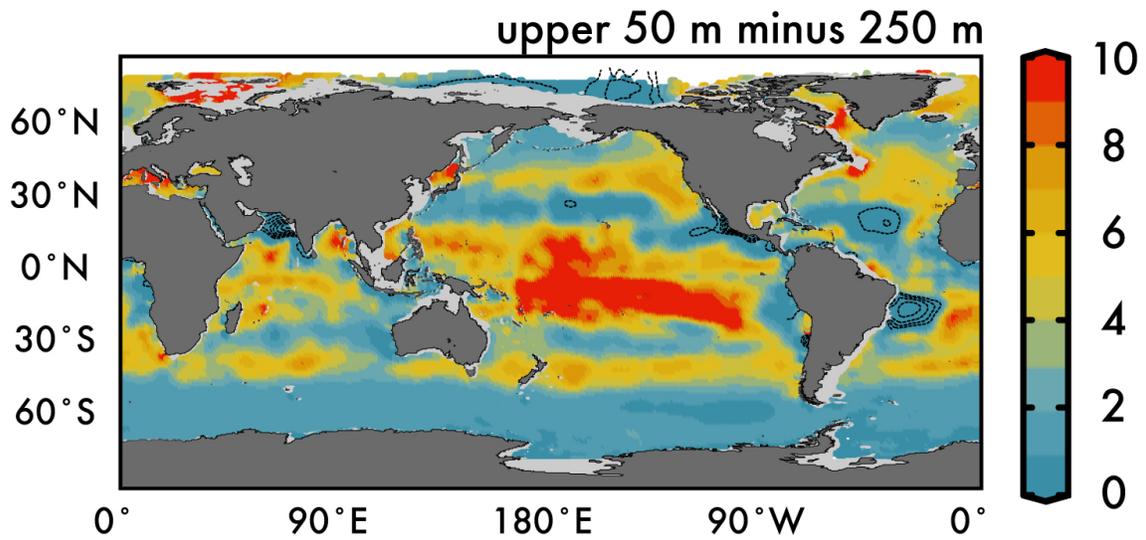


Figure 6. Again, the contours are difficult to see. Can you indicate negative numbers with a different color, or allow the color bar to include negative numbers?

The revised Figure 6 can be seen below. Once again we have discretized the color bar to more clearly illustrate the variability. We identify negative values by the dashed contour lines.



Main document changes and comments

Page 2: Inserted Patrick Rafter 2/17/19 2:06 PM

from nitrate assimilation

Page 2: Deleted Patrick Rafter 2/17/19 2:06 PM

of

Page 2: Inserted Patrick Rafter 2/17/19 2:06 PM

formed from nitrate assimilation

Page 3: Inserted Patrick Rafter 3/6/19 9:34 PM

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 $\delta^{15}\text{N}$ data were withheld from binning to be used as an independent validation set (see section 2.2.4).

Page 3: Deleted Patrick Rafter 3/6/19 9:34 PM

When binning vertically, we use the midpoint between the depth values of one layer and the next as the partition between bins (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.). A point that lies right at the midpoint between depth intervals is binned to the shallower interval. 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 $\delta^{15}\text{N}$ data were withheld from binning to be used as an independent validation set (see section 2.2.4).

Page 3: Deleted Microsoft Office User 3/8/19 3:35 PM

These include

Page 3: Inserted Microsoft Office User 3/8/19 3:35 PM

We have six input features including

Page 3: Inserted Patrick Rafter 3/6/19 9:35 PM

averaged and

Page 3: Inserted Patrick Rafter 3/6/19 9:35 PM

n annual

Page 3: Deleted Patrick Rafter 3/6/19 9:35 PM

Page 3: Inserted Aaron Bagnell 3/9/19 7:00 PM

Page 4: Inserted Microsoft Office User 3/8/19 11:39 AM

The choice of these specific input features was dictated by our desire to achieve the best possible R^2 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 R^2 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 sparsity

Page 4: Formatted Microsoft Office User 3/8/19 11:41 AM

Font:(Default) Cambria

Page 4: Formatted Microsoft Office User 3/8/19 11:41 AM

Font:(Default) Cambria

Page 4: Formatted Microsoft Office User 3/8/19 11:41 AM

Font:(Default) Cambria

Page 4: Formatted Microsoft Office User 3/8/19 11:41 AM

Font:(Default) Cambria

Page 4: Formatted Microsoft Office User 3/8/19 11:41 AM

Font:(Default) Cambria

Page 4: Deleted Patrick Rafter 3/14/19 4:46 PM

sparsity

Page 4: Inserted Patrick Rafter 3/14/19 4:46 PM

scarcity

Page 4: Inserted Microsoft Office User 3/8/19 11:43 AM

of $\delta^{15}\text{N}$ data in some regions, it is not possible to ascribe the impact of a specific combination of input features versus the impact of available $\delta^{15}\text{N}$ data, which may not be representative of the region's climatological state, to the relative model performance in these regions.

Page 4: Formatted Microsoft Office User 3/8/19 11:39 AM

Font:(Default) Cambria

Page 4: Deleted Microsoft Office User 3/8/19 1:53 PM

Page 4: Inserted Patrick Rafter 3/6/19 9:35 PM

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.

Page 4: Inserted Microsoft Office User 3/8/19 3:33 PM

we have

Page 4: Deleted	Microsoft Office User	3/8/19 11:15 AM
≈		
Page 4: Inserted	Microsoft Office User	3/8/19 11:15 AM
170		
Page 4: Deleted	Microsoft Office User	3/8/19 11:15 AM
000		
Page 4: Inserted	Microsoft Office User	3/8/19 3:33 PM
we have		
Page 4: Inserted	Microsoft Office User	3/8/19 11:30 AM
01		
Page 4: Deleted	Microsoft Office User	3/8/19 11:30 AM
00		
Page 4: Inserted	Patrick Rafter	3/6/19 9:36 PM
(Weigend et al., 1990)		
Page 4: Inserted	Microsoft Office User	3/8/19 1:56 PM
To create a nonlinear system, a		
Page 4: Deleted	Microsoft Office User	3/8/19 1:56 PM
A		
Page 4: Inserted	Patrick Rafter	3/6/19 9:37 PM
An activation function introduces the nonlinearity, transform		
Page 4: Deleted	Microsoft Office User	3/8/19 1:56 PM
introduces the nonlinearity,		
Page 4: Inserted	Microsoft Office User	3/8/19 1:57 PM
s		
Page 4: Deleted	Microsoft Office User	3/8/19 1:57 PM
ing		
Page 4: Inserted	Patrick Rafter	3/6/19 9:37 PM
ing the product of the weights and input features and creat		
Page 4: Inserted	Microsoft Office User	3/8/19 1:57 PM
es		
Page 4: Deleted	Microsoft Office User	3/8/19 1:57 PM
ing		
Page 4: Inserted	Patrick Rafter	3/6/19 9:37 PM
ing the values assigned to nodes in the hidden layer		
Page 4: Inserted	Microsoft Office User	3/8/19 1:57 PM
. These		
Page 4: Deleted	Microsoft Office User	3/8/19 1:57 PM

that

Page 4: Inserted Patrick Rafter 3/6/19 9:37 PM

that act as new features for estimating the target $\delta^{15}\text{N}$ data. Our model utilizes the hyperbolic tangent as its activation function between the input and hidden layer as well as between the hidden and output layer due to its relative speed and general performance (Thimm and Fiesler, 1997).

Page 4: Deleted Microsoft Office User 3/8/19 3:33 PM

relative

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

The values of nodes in the hidden layer (H) can be defined as

$$H = a(I \cdot W_1 + b_1)$$

where H is an array containing the values of the hidden nodes, a is the activation function

Page 4: Inserted Microsoft Office User 3/8/19 3:38 PM

(here, the hyperbolic tangent)

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

, I is a

Page 4: Inserted Microsoft Office User 3/8/19 9:20 PM

7170x6

Page 4: Deleted Microsoft Office User 3/8/19 9:20 PM

n

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

n array containing the

Page 4: Inserted Microsoft Office User 3/8/19 3:38 PM

values of the

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

input features

Page 4: Inserted Microsoft Office User 3/8/19 3:38 PM

at the locations of the binned observations (there are 7170 binned observations and 6 input parameters)

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

, W_1 is a

Page 4: Inserted Microsoft Office User 3/8/19 9:20 PM

6x25

Page 4: Deleted Microsoft Office User 3/8/19 9:20 PM

matrix

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM
matrix

Page 4: Inserted Microsoft Office User 3/8/19 9:20 PM
array

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM
of weights that connect input features to hidden nodes, and b_1

Page 4: Deleted Microsoft Office User 3/8/19 9:21 PM

Page 4: Inserted Microsoft Office User 3/8/19 11:17 AM

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM
is a

Page 4: Inserted Microsoft Office User 3/8/19 9:20 PM
 7170×25

Page 4: Deleted Microsoft Office User 3/8/19 9:20 PM
 n

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM
 n array of weights

Page 4: Inserted Microsoft Office User 3/11/19 12:23 PM
(25 unique values repeated 7170 times)

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM
that connects a bias node to the hidden nodes.

Page 4: Inserted Microsoft Office User 3/8/19 3:39 PM
The factor of 25 represents the number of nodes in the hidden layer, and is chosen

Page 4: Deleted Aaron Bagnell 3/9/19 7:59 PM
and is

Page 4: Deleted Aaron Bagnell 3/9/19 7:20 PM

Page 4: Inserted Aaron Bagnell 3/9/19 7:20 PM
by experimentation to find the maximum number of effective parameters (Foresee and Hagan 1997), i.e. where adding new parameters no longer improves performance on

Page 4: Deleted Microsoft Office User 3/11/19 1:48 PM

Page 4: Inserted Microsoft Office User 3/11/19 1:48 PM

Page 4: Inserted Aaron Bagnell 3/9/19 8:04 PM

an internal validation set (Step 4).

Page 4: Deleted Microsoft Office User 3/10/19 11:51 AM

Page 4: Deleted Aaron Bagnell 3/9/19 7:20 PM

because ?

Page 4: Inserted Microsoft Office User 3/8/19 9:28 PM

because ?

Page 4: Inserted Patrick Rafter 3/6/19 9:39 PM

The bias node acts as an offset term, similar to a constant term in a linear function, and has a value that is always 1.

At the output layer, the network produces a prediction of the target nitrate isotopic data (

Page 5: Inserted Microsoft Office User 3/8/19 9:32 PM

$\delta^{15}\text{N}$

Page 5: Deleted Aaron Bagnell 3/9/19 7:48 PM

Page 5: Deleted Microsoft Office User 3/8/19 9:32 PM

t

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

t_{pred}). Similar to how nodes in the hidden layer are a function of the inputs and a set of weights,

Page 5: Inserted Microsoft Office User 3/8/19 9:32 PM

$\delta^{15}\text{N}$

Page 5: Deleted Aaron Bagnell 3/9/19 7:48 PM

Page 5: Deleted Microsoft Office User 3/8/19 9:32 PM

t

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

t_{pred} is a function of the hidden nodes and an additional set of weights. The predicted values (t_{pred}) can be defined as

$$\langle \text{DELTA} \rangle \delta^{15}\text{N } t_{\text{pred}} = a(H \cdot W_2 + b_2)$$

Page 5: Deleted Microsoft Office User 3/8/19 9:32 PM

(t_{pred})

Page 5: Deleted Patrick Rafter 3/14/19 3:20 PM

□

Page 5: Deleted Aaron Bagnell 3/9/19 7:48 PM

Page 5: Deleted Microsoft Office User 3/8/19 9:32 PM

t

Page 5: Inserted Microsoft Office User 3/8/19 11:31 AM

where H (size 7170×25) has been previously defined, W_2 (size 25×1) is a matrix of weights that connect features in the hidden layer to nodes in the output layer, and b_2 (size 7170×1) is an array of weights (all of the same value) that connects a bias node to the output layer.

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

The ANN learns by comparing

Page 5: Inserted Microsoft Office User 3/8/19 9:33 PM

$\delta^{15}\text{N}$

Page 5: Deleted Aaron Bagnell 3/9/19 7:48 PM

Page 5: Deleted Microsoft Office User 3/8/19 9:33 PM

t

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

t_{pred} to the actual $\delta^{15}\text{N}$ data (

Page 5: Inserted Microsoft Office User 3/8/19 9:33 PM

$\delta^{15}\text{N}$

Page 5: Deleted Aaron Bagnell 3/9/19 7:48 PM

Page 5: Deleted Microsoft Office User 3/8/19 9:33 PM

t

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

t_{data}), attempting to minimize the value of the cost function

Page 5: Deleted Patrick Rafter 3/6/19 9:40 PM

A nonlinear activation function transforms the product of the weights and input features, creating the values assigned to nodes in the hidden layer. Our model utilizes the hyperbolic tangent as its activation function between the input and hidden layer as well as between the hidden and output layer due to its relative speed and general performance (Thimm and Fiesler, 1997). At the output layer, the network produces a prediction of the target nitrate isotopic data (t_{pred}), which it then compares to the actual values of that dataset (t_{data}). The ANN attempts to minimize the value of the cost function

Page 5: Inserted Patrick Rafter 3/6/19 9:40 PM

as a way of backpropagating

Page 5: Deleted	Aaron Bagnell	3/9/19 7:53 PM
back		
Page 5: Inserted	Aaron Bagnell	3/9/19 7:42 PM
the		
Page 5: Inserted	Patrick Rafter	3/6/19 9:40 PM
errors		
Page 5: Inserted	Aaron Bagnell	3/9/19 7:45 PM
between $\delta^{15}\text{N}_{\text{pred}}$ and $\delta^{15}\text{N}_{\text{data}}$ backwards through the network (Rumelhart et al., 1986)		
Page 5: Inserted	Patrick Rafter	3/6/19 9:41 PM
(Rumelhart et al., 1986)		
Page 5: Inserted	Patrick Rafter	3/6/19 9:42 PM
To ensure good generalization of		
Page 5: Formatted	Patrick Rafter	3/14/19 4:31 PM
No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers		
Page 5: Inserted	Microsoft Office User	3/10/19 11:54 AM
the		
Page 5: Deleted	Microsoft Office User	3/10/19 11:54 AM
each		
Page 5: Inserted	Patrick Rafter	3/6/19 9:42 PM
each trained ANN to novel data, we randomly withhold 10% of the		
Page 5: Deleted	Microsoft Office User	3/10/19 11:54 AM
to novel data		
Page 5: Inserted	Microsoft Office User	3/8/19 9:44 PM
$\delta^{15}\text{N}$		
Page 5: Deleted	Microsoft Office User	3/8/19 9:44 PM
target isotopic		
Page 5: Inserted	Patrick Rafter	3/6/19 9:42 PM
target isotopic data (t_{data}) to be used as a		
Page 5: Deleted	Microsoft Office User	3/10/19 11:53 AM
(t_{data})		
Page 5: Inserted	Microsoft Office User	3/10/19 1:43 PM
n		
Page 5: Deleted	Microsoft Office User	3/10/19 1:43 PM
different		
Page 5: Inserted	Patrick Rafter	3/6/19 9:42 PM

different internal validation set for each network. This is data that the individual 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

Page 5: Deleted **Microsoft Office User** **3/10/19 1:47 PM**

individual

Page 5: Inserted **Microsoft Office User** **3/10/19 1:47 PM**

(see Step 5)

Page 5: Deleted **Microsoft Office User** **3/10/19 1:45 PM**

. Our pass criterion is an R^2 value greater than 0.81 To ensure good generalization of the trained ANN to novel data, we randomly withhold 10% of the target isotopic data (t_{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. Our pass criterion is an R^2 value greater than 0.9 between the ANN's predicted value and the actual values of the validation set

Page 5: Inserted **Patrick Rafter** **3/6/19 9:42 PM**

. Our pass criterion is an R^2 value greater than 0.81

Page 5: Inserted **Patrick Rafter** **3/14/19 4:20 PM**

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.

Page 5: Deleted **Patrick Rafter** **3/14/19 4:44 PM**

This independent validation set is never used in the process of developing our ensemble of ANNs.

Page 5: Inserted **Microsoft Office User** **3/10/19 1:42 PM**

(using a different random 10% validation set)

Page 5: Inserted **Microsoft Office User** **3/10/19 1:44 PM**

A network is admitted into the ensemble if it yields an R^2 value greater than 0.81 on the validation dataset.

Page 6: Inserted **Microsoft Office User** **3/10/19 1:48 PM**

E

Page 6: Deleted **Microsoft Office User** **3/10/19 1:48 PM**

E

Page 6: Inserted	Microsoft Office User	3/10/19 1:49 PM
------------------	-----------------------	-----------------

to

Page 6: Inserted	Microsoft Office User	3/8/19 11:24 AM
------------------	-----------------------	-----------------

on average

Page 6: Inserted	Microsoft Office User	3/8/19 11:23 AM
------------------	-----------------------	-----------------

compared to a single randomly generated ensemble member. Compared to each of its members,

Page 6: Deleted	Microsoft Office User	3/8/19 11:23 AM
-----------------	-----------------------	-----------------

, as demonstrated by the

Page 6: Inserted	Microsoft Office User	3/8/19 11:24 AM
------------------	-----------------------	-----------------

our ensemble mean sees improved performance on all internal validation sets and has a higher R^2 and lower root mean square error on the independent validation set compared to 19 of the 25 members

Page 6: Deleted	Microsoft Office User	3/8/19 11:24 AM
-----------------	-----------------------	-----------------

improved performance of the ensemble versus any single member on the independent validation set

Page 6: Inserted	Patrick Rafter	3/14/19 3:26 PM
------------------	----------------	-----------------

raw / unbinned

Page 7: Deleted	Microsoft Office User	3/10/19 2:04 PM
-----------------	-----------------------	-----------------

negligible

Page 7: Inserted	Microsoft Office User	3/10/19 2:04 PM
------------------	-----------------------	-----------------

smaller

Page 7: Deleted	Microsoft Office User	3/10/19 2:08 PM
-----------------	-----------------------	-----------------

mean marine

Page 7: Inserted	Microsoft Office User	3/10/19 2:08 PM
------------------	-----------------------	-----------------

climatological

Page 7: Deleted	Microsoft Office User	3/10/19 2:16 PM
-----------------	-----------------------	-----------------

can

Page 7: Inserted	Microsoft Office User	3/10/19 2:16 PM
------------------	-----------------------	-----------------

will

Page 7: Deleted	Microsoft Office User	3/10/19 2:17 PM
-----------------	-----------------------	-----------------

all areas of research using this widely used geochemical measurement

Page 7: Inserted	Microsoft Office User	3/10/19 2:17 PM
------------------	-----------------------	-----------------

studies of the marine nitrogen cycle

Page 7: Deleted	Microsoft Office User	3/10/19 2:18 PM
-----------------	-----------------------	-----------------

indicated

Page 7: Inserted Microsoft Office User 3/10/19 2:18 PM

show

Page 7: Deleted Microsoft Office User 3/10/19 2:20 PM

range

Page 7: Inserted Microsoft Office User 3/10/19 2:20 PM

variability

Page 7: Deleted Microsoft Office User 3/10/19 2:20 PM

variability

Page 7: Deleted Microsoft Office User 3/10/19 2:20 PM

also

Page 7: Inserted Microsoft Office User 3/10/19 2:29 PM

.

Page 7: Deleted Microsoft Office User 3/10/19 2:29 PM

.

Page 7: Deleted Microsoft Office User 3/10/19 2:28 PM

A notable difference between the EANN and a 3D circulation model nitrate $\delta^{15}\text{N}$ is that the EANN does not overestimate values for the Bay of Bengal and underestimate it in the Arabian Sea (Somes et al., 2010).

Page 7: Inserted Patrick Rafter 3/6/19 10:09 PM

A notable difference between the EANN and a 3D circulation model nitrate $\delta^{15}\text{N}$ is that the EANN does not overestimate values for the Bay of Bengal and underestimate it in the Arabian Sea

Page 7: Inserted Patrick Rafter 3/6/19 10:11 PM

.

Page 8: Inserted Microsoft Office User 3/10/19 2:28 PM

A notable difference between the EANN and a previous biogeochemical model estimate of nitrate $\delta^{15}\text{N}$ (Somes et al., 2010) is that the EANN correctly captures the higher nitrate $\delta^{15}\text{N}$ in the Arabian Sea compared to the Bay of Bengal.

Page 8: Deleted Patrick Rafter 3/7/19 11:12 AM

has

Page 8: Inserted Patrick Rafter 3/7/19 11:12 AM

produces organic matter with

Page 9: Deleted Microsoft Office User 3/10/19 2:34 PM

and in Fig. 1B

Page 9: Deleted Microsoft Office User 3/10/19 2:35 PM

ing

Page 9: Deleted Microsoft Office User 3/10/19 2:35 PM

of the

Page 9: Inserted	Microsoft Office User	3/10/19 2:36 PM
------------------	-----------------------	-----------------

at at

Page 9: Deleted	Microsoft Office User	3/10/19 2:36 PM
-----------------	-----------------------	-----------------

e

Page 9: Deleted	Microsoft Office User	3/10/19 2:36 PM
-----------------	-----------------------	-----------------

depths

Page 9: Inserted	Patrick Rafter	3/6/19 9:05 PM
------------------	----------------	----------------

wintertime

Page 10: Inserted	Patrick Rafter	3/6/19 9:53 PM
-------------------	----------------	----------------

as far as the tropics

Page 10: Inserted	Patrick Rafter	2/17/19 1:06 PM
-------------------	----------------	-----------------

The South Indian Ocean is one region particularly devoid of published nitrate $\delta^{15}\text{N}$ observations (Fig. 2), but the EANN makes specific predictions about its distribution. For example, the modeled nitrate $\delta^{15}\text{N}$ predicts that intermediate-depth Indian Ocean nitrate is similarly elevated in $\delta^{15}\text{N}$ to the intermediate-depth South Pacific (Fig. 5C). Considering that both intermediate-depth water masses are formed from Southern Ocean surface waters, it is reasonable to propose that nitrate $\delta^{15}\text{N}$ are similarly elevated by partial nitrate consumption. The EANN therefore provides testable predictions for nitrate $\delta^{15}\text{N}$ observations throughout the Indian Ocean.

Page 10: Deleted	Patrick Rafter	2/17/19 1:13 PM
------------------	----------------	-----------------

Equivalent processes must drive the $\delta^{15}\text{N}$ in the intermediate-depth Indian Ocean, which is similarly elevated in the EANN, although direct observations are needed in order to confirm how well the EANN extrapolates in this region.

Page 10: Deleted	Microsoft Office User	3/10/19 8:12 PM
------------------	-----------------------	-----------------

from the surface to 5500

Page 10: Inserted	Microsoft Office User	3/10/19 8:12 PM
-------------------	-----------------------	-----------------

with depth

Page 10: Deleted	Microsoft Office User	3/10/19 8:12 PM
------------------	-----------------------	-----------------

m

Page 10: Inserted	Patrick Rafter	3/7/19 11:17 AM
-------------------	----------------	-----------------

(here defined as 3000 m and below)

Page 10: Inserted	Microsoft Office User	3/10/19 8:12 PM
-------------------	-----------------------	-----------------

inter-

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

-scale

Page 10: Deleted	Microsoft Office User	3/10/19 8:12 PM
------------------	-----------------------	-----------------

model

Page 10: Inserted	Microsoft Office User	3/10/19 8:12 PM
-------------------	-----------------------	-----------------

EANN

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

differences

Page 10: Inserted	Microsoft Office User	3/10/19 8:13 PM
-------------------	-----------------------	-----------------

gradients

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

shown

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

even

Page 10: Inserted	Microsoft Office User	3/10/19 8:13 PM
-------------------	-----------------------	-----------------

smaller than the corresponding

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

larger for the nitrate

Page 10: Inserted	Microsoft Office User	3/10/19 8:13 PM
-------------------	-----------------------	-----------------

inter-basin gradients in observed

Page 10: Inserted	Microsoft Office User	3/10/19 8:14 PM
-------------------	-----------------------	-----------------

,

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

observations

Page 10: Deleted	Microsoft Office User	3/10/19 8:13 PM
------------------	-----------------------	-----------------

measurements

Page 10: Inserted	Microsoft Office User	3/10/19 8:13 PM
-------------------	-----------------------	-----------------

observations

Page 11: Inserted	Patrick Rafter	2/17/19 1:59 PM
-------------------	----------------	-----------------

, which

Page 11: Inserted	Patrick Rafter	2/17/19 1:59 PM
-------------------	----------------	-----------------

is consistent with annually-averaged sinking organic matter $\delta^{15}\text{N}$ of ≈ 0.9 to 1.6‰

Page 11: Inserted	Patrick Rafter	3/14/19 3:48 PM
-------------------	----------------	-----------------

Page 11: Deleted	Patrick Rafter	3/7/19 11:04 AM
------------------	----------------	-----------------

numbers in

Page 11: Inserted	Patrick Rafter	3/14/19 3:49 PM
-------------------	----------------	-----------------

This is consistent with the known increase in nitrate concentrations and lowering of deep oxygen concentrations from the deep South to Tropical and North Pacific (e.g., see Fig. 4E in (Rafter et al., 2013)).

Page 12: Inserted	Microsoft Office User	3/10/19 8:17 PM
D		
Page 12: Deleted	Microsoft Office User	3/10/19 8:17 PM
Vertical, d		
Page 12: Inserted	Patrick Rafter	3/14/19 4:21 PM
J. J. Becker,		
Page 12: Inserted	Patrick Rafter	3/7/19 11:15 AM
and M. Kienast		
Page 12: Inserted	Patrick Rafter	3/7/19 9:29 AM
Many figures were made using Ocean Data View software (Schlitzer, 2002).		
Header and footer changes		
Text box changes		
Header and footer text box changes		
Footnote changes		
Endnote changes		

1 Global trends in marine nitrate N isotopes from observations and a neural network- 2 based climatology

3
4 Patrick A. Rafter¹, Aaron Bagnell², Dario Marconi³, and Timothy DeVries²

5
6 1: University of California, Irvine; 2 University of California, Santa Barbara; 3: Princeton
7 University

8 9 **Abstract**

10 Nitrate is a critical ingredient for life in the ocean because, as the most abundant form of
11 fixed nitrogen in the ocean, it is an essential nutrient for primary production. The
12 availability of marine nitrate is principally determined by biological processes, each having
13 a distinct influence on the N isotopic composition of nitrate (nitrate $\delta^{15}\text{N}$)—a property that
14 informs much of our understanding of the marine N cycle as well as marine ecology,
15 fisheries, and past ocean conditions. However, the sparse spatial distribution of nitrate $\delta^{15}\text{N}$
16 observations makes it difficult to apply this useful property in global studies, or to facilitate
17 robust model-data comparisons. Here, we use a compilation of published nitrate $\delta^{15}\text{N}$
18 measurements ($n = 12277$) and climatological maps of physical and biogeochemical tracers
19 to create a surface-to-seafloor, 1° resolution map of nitrate $\delta^{15}\text{N}$ using an Ensemble of
20 Artificial Neural Networks (EANN). The strong correlation ($R_2 > 0.87$) and small mean
21 difference ($<0.05\text{‰}$) between EANN-estimated and observed nitrate $\delta^{15}\text{N}$ indicates that
22 the EANN provides a good estimate of climatological nitrate $\delta^{15}\text{N}$ without a significant bias.
23 The magnitude of observation-model residuals is consistent with the magnitude of
24 seasonal-decadal changes in observed nitrate $\delta^{15}\text{N}$ that are not captured by our
25 climatological model. As such, these observation-constrained results provide a globally-
26 resolved map of mean nitrate $\delta^{15}\text{N}$ for observational and modeling studies of marine
27 biogeochemistry, paleoceanography, and marine ecology.

28 29 **1 Introduction**

30 In contrast to other marine nutrients (e.g., phosphate and silicate), the inventory of nitrate
31 (NO_3^-) is mediated by biological processes, where the main source is N_2 fixation by
32 diazotrophic phytoplankton and the main sink is denitrification (via a microbial
33 consortium in oxygen deficient waters and sediments) (Codispoti and Christensen, 1985).
34 Biological processes also determine the distribution of marine nitrate throughout the water
35 column, with phytoplankton assimilating nitrate / lowering nitrate concentrations in the
36 surface ocean and the microbially-mediated degradation of organic matter in the
37 subsurface. (The latter involving the multi-step process of ammonification (organic matter
38 $\rightarrow \text{NH}_4^+$) and nitrification ($\text{NH}_4^+ \rightarrow \text{NO}_2^- \rightarrow \text{NO}_3^-$)).) By regulating the global inventory and
39 distribution of marine nitrate, these N cycling processes control global net primary
40 productivity, the transfer of nutrients to higher trophic levels such as fishes, and the
41 strength of the ocean's biological carbon pump (Dugdale and Goering, 1967).

42
43 Each of these biologically mediated N transformations affects the N isotopic composition of
44 nitrate in unique ways (Figs 1A & 1B and see Section 2), adjusting the relative abundance
45 of ^{15}N and ^{14}N in oceanic nitrate relative to the atmosphere. $\delta^{15}\text{N} = (^{15}\text{N}/^{14}\text{N}_{\text{sample}} /$

46 $^{15}\text{N}/^{14}\text{N}$ standard) – 1), multiplied by 1000 to give units of per mil (‰); see (Sigman and
47 Casciotti, 2001) for simplified equations from (Mariotti et al., 1981). Nitrate $\delta^{15}\text{N}$
48 measurements have become a powerful tool for understanding the ‘biogeochemical history’
49 of marine nitrate, which includes nitrate assimilation by phytoplankton (Miyake and Wada,
50 1967; Wada and Hattori, 1978), nitrogen fixation (Carpenter et al., 1997; Hoering and Ford,
51 1960), denitrification (Liu, 1979), and nitrification (Casciotti et al., 2013). For example, the
52 consumption of nitrate by denitrification (red line in Fig. 1A) has a larger impact on the
53 residual nitrate $\delta^{15}\text{N}$ than does partial nitrate assimilation by phytoplankton (yellow line in
54 Fig. 1), and thus very high $\delta^{15}\text{N}$ values serve as a fingerprint of denitrification. Nitrate $\delta^{15}\text{N}$
55 is also influenced by the addition of nitrate via remineralization of organic matter. The
56 exact influence of remineralization depends on the isotopic composition of the organic
57 matter, and could result in both higher or lower nitrate $\delta^{15}\text{N}$ (Fig. 1A). Nitrate introduced
58 into the water column by the remineralization of organic matter formed by N_2 -fixing
59 phytoplankton has an isotopic composition close to that of air (0-1‰), and serves to lower
60 the mean ocean $\delta^{15}\text{N}$ (Fig. 1B). On the other hand, organic matter formed **from nitrate**
61 **assimilation** in regions where the plankton use most of the available nitrate can be
62 isotopically heavy, and its remineralization will increase the $\delta^{15}\text{N}$ of ambient nitrate (Fig.
63 1B). The actual value of organic matter $\delta^{15}\text{N}$ **formed from nitrate assimilation** is mostly
64 determined by: (1) the $\delta^{15}\text{N}$ of nitrate delivered to the euphotic zone (the subsurface
65 source), which in turn is dependent on the degree of water-column denitrification and (2)
66 the degree of nitrate consumption at the ocean surface, with heavier values associated with
67 greater nitrate consumption (Fig. 1B). Accordingly, changes in organic matter $\delta^{15}\text{N}$ (and
68 therefore sediment $\delta^{15}\text{N}$ used for paleoceanographic work) can reflect variability of the
69 source nitrate $\delta^{15}\text{N}$ and/or variability of the degree of nitrate consumption (e.g., see (Rafter
70 and Charles, 2012)).

71
72 Because of nitrate’s place at the base of the marine ecosystem, nitrate $\delta^{15}\text{N}$ is also useful for
73 understanding the lifecycles of higher trophic level organisms such as fish (Graham et al.,
74 2007; Tawa et al., 2017) and fishery productivity (Finney et al., 2002, 2000). The $\delta^{15}\text{N}$ of
75 whole sediment and microfossils provides insight *by proxy* of past ocean nitrate
76 transformations (Altabet and Francois, 1994a; Kienast et al., 2008; Ren et al., 2009;
77 Robinson et al., 2004; Sigman et al., 1999b)—work that places important constraints on
78 modern ocean N cycling (Altabet, 2007; Eugster et al., 2013; Ren et al., 2017). With an
79 understanding of the N transformations described above and their influences on the N
80 isotopic composition of nitrate, we can begin using nitrate $\delta^{15}\text{N}$ measurements to trace the
81 integrated biogeochemical history of marine nitrate. However, identifying basin- and
82 global-scale trends in nitrate $\delta^{15}\text{N}$ is challenged by the limited spatial extent of nitrate $\delta^{15}\text{N}$
83 observations (Fig. 2). Here, we compile a global database of nitrate $\delta^{15}\text{N}$ measurements
84 (Fig. 2) and use an Ensemble Artificial Neural Network (EANN) to produce a map of the
85 global nitrate $\delta^{15}\text{N}$ distribution at 1-degree spatial resolution. We find that the mapped
86 nitrate $\delta^{15}\text{N}$ climatology matches the observations well and should be a valuable tool for
87 estimating mean conditions and for constraining predictive nitrate $\delta^{15}\text{N}$ models (Somes et
88 al., 2010; Yang and Gruber, 2016). Below we briefly discuss how the EANN was used to
89 produce global maps of nitrate $\delta^{15}\text{N}$ (Section 2), address the ability of the EANN to match

Patrick Rafter 2/17/2019 2:06 PM

Deleted: of

91 the measured $\delta^{15}\text{N}$ (Section 3), and examine the EANN-mapped $\delta^{15}\text{N}$ climatology and global
92 compilation of nitrate $\delta^{15}\text{N}$ in the context of published work (Section 4).

94 2 Methods

95 2.1 Data Compilation

96 Nitrate $\delta^{15}\text{N}$ observations (Fig. 2; references in Table 3) were compiled from studies dating
97 from 1975 (Cline and Kaplan, 1975) to 2018 (Fripiat et al., 2018), including data from the
98 GEOTRACES Intermediate Data Product (Schlitzer et al., 2018). Whenever possible, the
99 data was acquired via the original author, but in other cases the data was estimated from
100 the publication directly. All observations were treated equally, although the failure to
101 remove nitrite when using the “denitrifier method” may bias the nitrate $\delta^{15}\text{N}$ to low values
102 (Rafter et al., 2013). These measurements have been identified as “nitrate+nitrite” in the
103 dataset to acknowledge this potential biasing, which predominantly affects observations in
104 the upper 100 m (Kemeny et al., 2016; Rafter et al., 2013).

106 2.2 Building the neural network model

107 We utilize an ensemble of artificial neural networks (EANNs) to interpolate our global
108 ocean nitrate $\delta^{15}\text{N}$ database (Fig. 2), producing complete 3D maps of the data. By utilizing
109 an artificial neural network (ANN), a machine learning approach that effectively identifies
110 nonlinear relationships between a target variable (the isotopic dataset) and a set of input
111 features (other available ocean datasets), we can fill holes in our data sampling coverage of
112 nitrate $\delta^{15}\text{N}$.

114 2.2.1 Binning target variables (Step 1)

115 We binned the nitrate $\delta^{15}\text{N}$ observations (red symbols in Fig. 2) to the World Ocean Atlas
116 2009 (WOA09) grid with a 1-degree spatial resolution and 33 vertical depth layers (0-5500
117 m) (Garcia et al., 2010). When binning vertically, we use the depth layer whose value is
118 closest to the observation’s sampling depth (e.g. the first depth layer has a value of 0 m, the
119 second of 10 m, and the third of 20 m, so all nitrate isotopic data sampled between 0-5 m
120 fall in the 0 m bin; between 5-15 m they fall in the 10 m bin, etc.). An observation with a
121 sampling depth that lies right at the midpoint between depth layers is binned to the
122 shallower layer. If more than one raw data point falls in a grid cell we take the average of all
123 those points as the value for that grid cell. Certain whole ship tracks of nitrate $\delta^{15}\text{N}$ data
124 were withheld from binning to be used as an independent validation set (see section 2.2.4).

126 2.2.2 Obtaining input features (Step 2)

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

Patrick Rafter 3/6/2019 9:34 PM

Deleted: When binning vertically, we use the midpoint between the depth values of one layer and the next as the partition between bins (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.). A point that lies right at the midpoint between depth intervals is binned to the shallower interval. 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 $\delta^{15}\text{N}$ data were withheld from binning to be used as an independent validation set (see section 2.2.4).

Microsoft Office User 3/8/2019 3:35 PM

Deleted: These include

Patrick Rafter 3/6/2019 9:35 PM

Deleted:

155
156 The choice of these specific input features was dictated by our desire to achieve the best
157 possible R² value on our internal validation sets (Step 4). Additional inputs besides those
158 we included, such as latitude, longitude, silicate, euphotic depth, or sampling depth either
159 did not improve the R² value on the validation dataset or degraded it, indicating that they
160 are not essential parameters for characterizing this system globally. By opting to use the
161 set of input features that yielded the best results for the global oceans, we potentially
162 overlooked combinations of inputs that perform better at regional scales. However, given
163 the scarcity of δ¹⁵N data in some regions, it is not possible to ascribe the impact of a specific
164 combination of input features versus the impact of available δ¹⁵N data, which may not be
165 representative of the region's climatological state, to the relative model performance in
166 these regions.

Microsoft Office User 3/8/2019 11:41 AM
Formatted ... [1]

Patrick Rafter 3/14/2019 4:46 PM
Deleted: sparsity

Microsoft Office User 3/8/2019 11:39 AM
Formatted: Font:(Default) Cambria

Microsoft Office User 3/8/2019 1:53 PM
Deleted: .

2.2.3 Training the ANN (Step 3)

169 The architecture of our ANN consists of a single hidden layer, containing 25 nodes, that
170 connects the biological and physical input features (discussed in Step 2) to the target
171 nitrate isotopic variable (as discussed in Step 1). The role of the hidden layer is to
172 transform input features into new features contained in the nodes. These are given to the
173 output layer to estimate the target variable, introducing nonlinearities via an activation
174 function. The number of nodes in this hidden layer, as well as the number of input features,
175 determines the number of adjustable weights (the free parameters) in the network.
176 Because there is a danger of over-fitting the model, which occurs when the ANN is over-
177 trained on a dataset so that it cannot generalize well when presented with new data, it is a
178 good practice to have a large number of training data (we have 7170 binned data points)
179 relative to the number of weights (we have 201 free parameters) (Weigend et al., 1990). To
180 create a nonlinear system, an activation function transforms the product of the weights and
181 input features and creates the values assigned to nodes in the hidden layer. These act as
182 new features for estimating the target δ¹⁵N data. Our model utilizes the hyperbolic tangent
183 as its activation function between the input and hidden layer as well as between the hidden
184 and output layer due to its speed and general performance (Thimm and Fiesler, 1997).

Microsoft Office User 3/8/2019 11:15 AM
Deleted: ≈...170000...binned data p... [2]

185
186 The values of nodes in the hidden layer (H) can be defined as

$$H = a(I \cdot W_1 + b_1)$$

188
189 where H is an array containing the values of the hidden nodes, a is the activation function
190 (here, the hyperbolic tangent), I is a 7170x6 array containing the values of the input
191 features at the locations of the binned observations (there are 7170 binned observations
192 and 6 input parameters), W₁ is a 6x25 array of weights that connect input features to
193 hidden nodes, and b₁ is a 7170x25 array of weights (25 unique values repeated 7170 times)
194 that connects a bias node to the hidden nodes. The factor of 25 represents the number of
195 nodes in the hidden layer, chosen by experimentation to find the maximum number of
196 effective parameters (Foresee and Hagan 1997), i.e. where adding new parameters no
197 longer improves performance on an internal validation set (Step 4). The bias node acts as

Microsoft Office User 3/8/2019 9:20 PM
Deleted: n...array containing the val... [3]

Aaron Bagnell 3/9/2019 7:59 PM
Deleted: and is...hosen ... [4]

Microsoft Office User 3/11/2019 1:48 PM
Deleted: ...an internal validation se... [5]

Aaron Bagnell 3/9/2019 7:20 PM
Deleted: because ?

225 an offset term, similar to a constant term in a linear function, and has a value that is always
226 1.

227
228 At the output layer, the network produces a prediction of the target nitrate isotopic data
229 ($\delta^{15}\text{N}_{\text{pred}}$). Similar to how nodes in the hidden layer are a function of the inputs and a set of
230 weights, $\delta^{15}\text{N}_{\text{pred}}$ is a function of the hidden nodes and an additional set of weights. The
231 predicted values can be defined as
232

$$\langle \text{DELTA} \rangle \delta^{15}\text{N}_{\text{pred}} = a(H \cdot W_2 + b_2)$$

233
234 where H (size 7170x25) has been previously defined, W_2 (size 25x1) is a matrix of weights
235 that connect features in the hidden layer to nodes in the output layer, and b_2 (size 7170x1)
236 is an array of weights (all of the same value) that connects a bias node to the output layer.
237

238 The ANN learns by comparing $\delta^{15}\text{N}_{\text{pred}}$ to the actual $\delta^{15}\text{N}$ data ($\delta^{15}\text{N}_{\text{data}}$), attempting to
239 minimize the value of the cost function
240

$$\text{cost} = \frac{\sum_{i=1}^n (\langle \text{DELTA} \rangle \delta^{15}\text{N}_{\text{pred}}^i - \langle \text{DELTA} \rangle \delta^{15}\text{N}_{\text{data}}^i)^2}{n}$$

241
242 by iteratively adjusting the weights using the Levenberg-Marquardt algorithm (Marquardt,
243 1963) as a way of propagating the errors between $\delta^{15}\text{N}_{\text{pred}}$ and $\delta^{15}\text{N}_{\text{data}}$ backwards through
244 the network (Rumelhart et al., 1986).
245

2.2.4 Validating the ANN (Step 4)

247 To ensure good generalization of the trained ANN, we randomly withhold 10% of the $\delta^{15}\text{N}$
248 data to be used as an internal validation set for each network. This is data that the network
249 never sees, meaning it does not factor into the cost function, so it works as a test of the
250 ANN's ability to generalize. This internal validation set acts as a gatekeeper to prevent poor
251 models from being accepted into the ensemble of trained networks (see Step 5). A second,
252 independent or 'external' validation set (blue symbols in Fig. 2), composed of complete ship
253 transects from the high and low latitude ocean were omitted from binning in Step 1 and
254 used to establish the performance of the entire ensemble. Our rationale for using complete
255 ship transects is the following. If we randomly chose 10% of observations to perform an
256 external validation, this dataset will be from the same cruises as the wider data. In other
257 words, despite being randomly selected, the validating observational dataset will be highly
258 correlated geographically. Contrast this with validating the EANN results with observations
259 from whole research cruises in unique geographic regions—areas where the model has not
260 "learned" anything about nitrate. We therefore argue that these observations from whole
261 ship tracks therefore provide a more difficult test of the model.
262

2.2.5 Forming the Ensemble (Step 5)

263 The ensemble is formed by repeating Steps 3 to 4 (using a different random 10% validation
264 set) until we obtain 25 trained networks for the nitrate $\delta^{15}\text{N}$ dataset. A network is admitted
265 into the ensemble if it yields an R^2 value greater than 0.81 on the validation dataset. Using
266

Aaron Bagnell 3/9/2019 7:48 PM
Deleted:

Microsoft Office User 3/8/2019 9:32 PM
Deleted: t

Aaron Bagnell 3/9/2019 7:48 PM
Deleted:

Microsoft Office User 3/8/2019 9:32 PM
Deleted: t

Microsoft Office User 3/8/2019 9:32 PM
Deleted: (t_{pred})

Patrick Rafter 3/14/2019 3:20 PM
Deleted: []

Aaron Bagnell 3/9/2019 7:48 PM
Deleted:

Microsoft Office User 3/8/2019 9:32 PM

Aaron Bagnell 3/9/2019 7:48 PM
Deleted:

Microsoft Office User 3/8/2019 9:33 PM
Deleted: t

Aaron Bagnell 3/9/2019 7:48 PM
Deleted:

Microsoft Office User 3/8/2019 9:33 PM
Deleted: t

Patrick Rafter 3/6/2019 9:40 PM
Deleted: A nonlinear activation function transforms the product of the weights and input features, creating the values assigned to nodes in the hidden layer. Our mod... [6]

Aaron Bagnell 3/9/2019 7:53 PM
Deleted: back

Microsoft Office User 3/10/2019 11:54 AM
Deleted: each

Microsoft Office User 3/10/2019 11:54 AM
Deleted: to novel data

Patrick Rafter 3/14/2019 4:31 PM
Formatted ... [7]

Microsoft Office User 3/8/2019 9:44 PM
Deleted: target isotopic

Microsoft Office User 3/10/2019 11:53 AM
Deleted: (t_{data})

Microsoft Office User 3/10/2019 1:43 PM
Deleted: different

Microsoft Office User 3/10/2019 1:47 PM
Deleted: individual

Microsoft Office User 3/10/2019 1:45 PM
Deleted: . Our pass criterion is an R²... [8]

Patrick Rafter 3/14/2019 4:44 PM
Deleted: This independent validatio... [9]

326 | an EANN instead of any single network provides several advantages. For example, the
327 | random initialization of the weight values in each network as well as differences in the
328 | training and internal validation sets used across members make it possible for many
329 | different networks to achieve similar performance on their respective validation set while
330 | generalizing to areas with no data coverage differently. By performing this type of data
331 | subsampling and taking an ensemble average, similar to bootstrap aggregating (Breiman,
332 | 1996) this approach on average improves the robustness of the generalization in areas
333 | without data coverage compared to a single randomly generated ensemble member.
334 | Compared to each of its members, our ensemble mean sees improved performance on all
335 | internal validation sets and has a higher R² and lower root mean square error on the
336 | independent validation set compared to 19 of the 25 members. The range of values given
337 | by the ensemble also provides a measure of the uncertainty for our estimations of $\delta^{15}\text{N}$.

Microsoft Office User 3/10/2019 1:48 PM

Deleted: E

Microsoft Office User 3/8/2019 11:23 AM

Deleted: , as demonstrated by the

Microsoft Office User 3/8/2019 11:24 AM

Deleted: improved performance of the ensemble versus any single member on the independent validation set

3 Results

3.1 Global nitrate $\delta^{15}\text{N}$ observations

341 | The global compilation of nitrate $\delta^{15}\text{N}$ includes 1180 stations from all major ocean basins
342 | and some minor seas (Fig. 2) giving a total of 12277 nitrate $\delta^{15}\text{N}$ measurements. Within
343 | this dataset, 1197 nitrate $\delta^{15}\text{N}$ measurements were withheld from the EANN and used to
344 | validate the EANN results to ensure good extrapolation (the ‘external’ validation dataset;
345 | blue symbols in Fig. 2, see Section 2). With observations from the surface to as deep as
346 | 6002 m (Rafter et al., 2012), we find that nitrate $\delta^{15}\text{N}$ ranges from $\approx 1\text{‰}$ in the North
347 | Atlantic (e.g., Marconi et al., (2015)) to 68.7‰ in the Eastern Tropical South Pacific
348 | (Bourbonnais et al., 2015). Nitrate $\delta^{15}\text{N}$ of $\approx 1\text{‰}$ was also irregularly observed in the
349 | shallow North and South Pacific (Liu et al., 1996; Yoshikawa et al., 2015). These latter
350 | observations were included in the training dataset, although we should note that the
351 | measurements using the ‘Devarda’s Alloy’ method (Liu et al., 1996) is thought to be biased
352 | low (Altabet and Francois, 2001). Similarly, the inclusion of nitrite for ‘denitrifier method’
353 | nitrate $\delta^{15}\text{N}$ can bias the measurement to lower values (Kemeny et al., 2016; Rafter et al.,
354 | 2013).

3.2 Marine nitrate $\delta^{15}\text{N}$ observations-model comparison

357 | The observed and EANN-predicted nitrate $\delta^{15}\text{N}$ measurements are distributed around a 1:1
358 | line in Fig. 3A (all data), with considerably less scatter for the deeper values (data >1000 m;
359 | Fig. 3B). The correlation coefficient of determination for the observations versus the model
360 | nitrate $\delta^{15}\text{N}$ gives an $R^2=0.75$ for the raw / unbinned observations used to train the EANN
361 | and an R^2 of 0.78 for the validation dataset. We can also examine the performance of the
362 | EANN with the nitrate $\delta^{15}\text{N}$ “residual” or the difference between observed and modeled
363 | $\delta^{15}\text{N}$, which indicates a mean residual or ‘mean bias’ value of -0.03‰ for the entire dataset
364 | and +0.18‰ for the validation dataset.

366 | Examining the observation-EANN residuals via the Root Mean Square Error (RMSE), we
367 | find an RMSE of 1.94‰ for the data used to train the EANN and an RMSE of 1.26‰ for the
368 | external validation dataset. There is a clear relationship between RMSE and depth, with a
369 | significantly higher RMSE for the upper 500 m (Figs. 3C and 3D). Comparing these residual
370 | values with dissolved oxygen concentrations (color in Fig. 3C), we find that >2‰ RMSE for

376 the surface is associated with high oxygen while $>2.7\text{‰}$ RMSE at ≈ 250 m is associated with
377 the lowest oxygen. Furthermore, the RMSE of the observation-EANN residuals differs
378 between the datasets used to train the model (solid red line in Fig. 3D) and validate the
379 model (dashed line in Fig. 3D).

380
381 The RMSE patterns in Figs. 3C and 3D are to be expected given the natural variability in
382 nitrate $\delta^{15}\text{N}$ driven by assimilation in the upper ocean and denitrification in the shallow
383 sub-surface—variability which is not captured by the climatological EANN. Rafter and
384 Sigman, (2016), presented a 5-year time-series of nitrate $\delta^{15}\text{N}$ from the eastern equatorial
385 Pacific, which showed that variability of nitrate assimilation produces seasonal-to-
386 interannual deviations of $\delta^{15}\text{N}$ of $\pm 2.5\text{‰}$, which is similar to the magnitude of the RMSE in
387 the surface ocean (2.2‰). Although there are no nitrate $\delta^{15}\text{N}$ time-series measurements
388 from the subsurface Oxygen Deficient Zone (ODZ) waters where denitrification occurs,
389 nitrate $\delta^{15}\text{N}$ in ODZs presumably have similar seasonal-to-interannual (or longer timescale)
390 variability due to changes in the rate and extent of water column denitrification (Deutsch et
391 al., 2011; Yang et al., 2017). For example, a larger degree of nitrate undergoing water
392 column denitrification would explain the extreme $\delta^{15}\text{N}$ values at the bottom right of Fig.
393 3A—observations that all come from the ODZ waters of the Eastern Tropical South Pacific
394 (Bourbonnais et al., 2015; Casciotti et al., 2013; Rafter et al., 2012; Ryabenko et al., 2012).
395 Some of these very high nitrate $\delta^{15}\text{N}$ values are associated with nitrate concentrations <1
396 $\mu\text{mol kg}^{-1}$ (Bourbonnais et al., 2015), values much lower than within our climatology for the
397 subsurface Eastern Tropical South Pacific. These values thus represent episodic
398 denitrification events that the EANN will not be able to capture because it is trained on
399 climatological data. In the deep ocean where temporal variability is *smaller*, the
400 observation-EANN residuals of 0.2‰ are the same magnitude as the $\delta^{15}\text{N}$ analytical errors,
401 further emphasizing the ability of the model to match climatological average conditions.

402 4 Discussion

403
404 The EANN's skillful estimate of *climatological* nitrate $\delta^{15}\text{N}$ *will* be useful for *studies of the*
405 *marine nitrogen cycle*. The zonal average view of EANN nitrate $\delta^{15}\text{N}$ for each major ocean
406 basin (Fig. 4) includes statistics comparing the observations versus EANN results above
407 and below 1000 m. These region-specific statistics *show* a weaker correlation between
408 EANN and observed nitrate $\delta^{15}\text{N}$ in the deep Atlantic and Southern Ocean, despite low
409 RMSE and negligible mean bias. This weak correlation likely derives from the limited
410 *variability* of deep nitrate $\delta^{15}\text{N}$ ($\pm 0.1\text{‰}$) in these basins (see Fig. 5D).

411
412 The nitrate $\delta^{15}\text{N}$ sections in Fig. 4 *show* elevated values for the low latitude, upper
413 mesopelagic Pacific (Fig. 4A) and Indian Oceans (Fig. 4D) where water column
414 denitrification raises the residual nitrate $\delta^{15}\text{N}$ (Fig. 1A). Viewing this elevated nitrate $\delta^{15}\text{N}$
415 at the 250 m depth horizon (Fig. 5) better reveals the spatial heterogeneity of the
416 observations and EANN results. (It is because of this intra-basin heterogeneity, and the fact
417 that many observations are biased towards the areas of denitrification, that we did not plot
418 the observed nitrate $\delta^{15}\text{N}$ within the zonally-averaged Fig. 4 views.) The EANN error for the
419 Fig. 5 depth intervals (Figs. 5E-5H) is the standard deviation of the 25 ensemble members
420 of the EANN and shows a decrease in ensemble variability with depth—a trend that is

- Microsoft Office User 3/10/2019 2:04 PM
Deleted: negligible
- Microsoft Office User 3/10/2019 2:08 PM
Deleted: mean marine
- Microsoft Office User 3/10/2019 2:16 PM
Deleted: can
- Microsoft Office User 3/10/2019 2:17 PM
Deleted: all areas of research using this widely used geochemical measurement
- Microsoft Office User 3/10/2019 2:18 PM
Deleted: indicated
- Microsoft Office User 3/10/2019 2:20 PM
Deleted: range
- Microsoft Office User 3/10/2019 2:20 PM
Deleted: variability
- Microsoft Office User 3/10/2019 2:20 PM
Deleted: also
- Microsoft Office User 3/10/2019 2:29 PM
Deleted: .
- Microsoft Office User 3/10/2019 2:28 PM
Deleted: A notable difference between the EANN and a 3D circulation model nitrate $\delta^{15}\text{N}$ is that the EANN does not overestimate values for the Bay of Bengal and underestimate it in the Arabian Sea (Somes et al., 2010).

437 consistent with the overall decrease in observed nitrate $\delta^{15}\text{N}$ variability with depth (Figs. 4
438 & 5).

439
440 Below we inspect the observed and EANN-predicted nitrate $\delta^{15}\text{N}$ and discuss the
441 consistency of these results with our understanding of published work. This analysis begins
442 with the spatial distribution of nitrate delivered to the upper ocean. We then discuss the
443 impacts of upper ocean nitrate assimilation on organic matter $\delta^{15}\text{N}$ and consider the
444 influence of organic matter remineralization on sub-surface nitrate.

445

446 **4.1 Subsurface and surface nitrate $\delta^{15}\text{N}$**

447 The nitrate $\delta^{15}\text{N}$ distribution at 250 m depth (Fig. 5B) offers a view of nitrate at a depth
448 that is deeper than source waters in many ocean regions (e.g., 100 to 150 m in the
449 equatorial Pacific (Rafter and Sigman, 2016)), but is negligibly influenced by nitrate
450 assimilation, and therefore provides a qualitative view of spatial trends in nitrate delivered
451 to the surface ocean. Nitrate $\delta^{15}\text{N}$ at this depth is highest in the North and South Eastern
452 Tropical Pacific and Arabian Seas (Fig. 5B), due to the influence of water column
453 denitrification in the ODZs in these regions (Altabet et al., 2012; Bourbonnais et al., 2015;
454 Ryabenko et al., 2012), which preferentially uses the light isotope and leaves the residual
455 nitrate enriched in ^{15}N . [A notable difference between the EANN and a previous
456 biogeochemical model estimate of nitrate \$\delta^{15}\text{N}\$ \(Somes et al., 2010\) is that the EANN
457 correctly captures the higher nitrate \$\delta^{15}\text{N}\$ in the Arabian Sea compared to the Bay of
458 Bengal.](#)

459

460 Lowest $\delta^{15}\text{N}$ values of sub-surface nitrate are found in the Southern Ocean and in the North
461 Atlantic. The North Atlantic subtropical gyre in particular has the lowest $\delta^{15}\text{N}$ values in any
462 basin (Fig. 5B; also see (Fawcett et al., 2011; Knapp et al., 2005, 2008)), which can be
463 attributed to the remineralization of low- $\delta^{15}\text{N}$ organic matter originating from N_2 -fixation,
464 which [produces organic matter with](#) a $\delta^{15}\text{N}$ between 0 and -1‰ (similar to atmospheric N_2 ;
465 see Fig. 1B (Carpenter et al., 1997; Hoering and Ford, 1960)). Prior work argues that this
466 nitrate $\delta^{15}\text{N}$ lowering requires the bulk of Atlantic N_2 -fixation ($\approx 90\%$) to occur in the
467 tropics (Marconi et al., 2017) followed by the advection of remineralized nitrate to the
468 North Atlantic. This contrasts with numerical models arguing for high N_2 -fixation rates in
469 the North Atlantic (Ko et al., 2018). Similar local minima of sub-surface $\delta^{15}\text{N}$ appear in all
470 the sub-tropical gyres (Fig. 5B), which is consistent with observations (Casciotti et al.,
471 2008; Yoshikawa et al., 2015) and presumably indicates the importance of N_2 -fixation in
472 these regions (Ko et al., (2018) and others). The N_2 -fixation $\delta^{15}\text{N}$ signal in the Pacific Ocean
473 is counteracted by the influence of water-column denitrification in that basin, which
474 imparts a high $\delta^{15}\text{N}$ signal, but a local minimum in $\delta^{15}\text{N}$ can still be seen in the Pacific
475 subtropical gyres (Fig. 4A).

476

477 Nitrate assimilation by phytoplankton in the upper ocean is influenced by both the
478 subsurface source nitrate $\delta^{15}\text{N}$ and the degree of nitrate assimilation (Miyake and Wada,
479 1967; Wada and Hattori, 1978) (Fig. 1B). This gives the expectation that average nitrate
480 $\delta^{15}\text{N}$ values for the upper 50 m (Fig. 5A) will be consistently higher than those at 250 m
481 (Fig. 5B). However, the highest values in the upper 50 m are not found above the ODZ

Patrick Rafter 3/7/2019 11:12 AM

Deleted: has

483 regions, but are on the edges of high nitrate concentration upwelling zones in the Southern
484 Ocean, equatorial Pacific, and subarctic gyres (contours in Fig. 2). Circulation in these ‘edge’
485 regions allows for nitrate to be advected along the surface, lengthening its time in the
486 surface ocean and allowing more utilization to elevate the residual nitrate $\delta^{15}\text{N}$ pool. In
487 other words, the degree of nitrate utilization appears to play a more important role in
488 determining surface nitrate $\delta^{15}\text{N}$ than the initial value. (This is not the case for the organic
489 matter $\delta^{15}\text{N}$ produced from this nitrate, which will be discussed more below.)

490
491 Despite our expectation of higher nitrate $\delta^{15}\text{N}$ in the upper 50 m versus 250 m (Figs. 5A vs.
492 5B), we identify two types of regions where this difference is negative (Fig. 6): above ODZ
493 waters and in subtropical gyres. The explanation for the negative values above the ODZ
494 regions is that the nitrate $\delta^{15}\text{N}$ at 250 m must be much higher than the nitrate $\delta^{15}\text{N}$
495 upwelled to the surface. This is consistent with elevated ODZ nitrate $\delta^{15}\text{N}$ having an
496 indirect path to waters outside of ODZ regions (Peters et al., 2017; Rafter et al., 2013). The
497 subtropical gyres also have modeled nitrate $\delta^{15}\text{N}$ in the upper 50 m that is less than 250 m,
498 but this finding is difficult to test with observations because of low nitrate concentrations.
499 That said, the model predicts a lower nitrate $\delta^{15}\text{N}$ in the upper ocean relative to that at 250
500 m, which is consistent with N_2 -fixation in these regions.

501
502 Our discussion above highlights the difficulty of distinguishing between the competing
503 influences of the subsurface source nitrate $\delta^{15}\text{N}$ and the degree of nitrate utilization on
504 residual nitrate $\delta^{15}\text{N}$. Clearly a static depth does not reflect the subsurface source of nitrate
505 delivered to the surface and a more robust method for estimating this subsurface source
506 needs to be developed. However, some generalizations can be made regarding the organic
507 matter $\delta^{15}\text{N}$ produced in these regions and its potential influence (via remineralization) on
508 subsurface nitrate throughout the water column via the export and remineralization of
509 organic matter (Sigman et al., 2009a). For example, a local minimum in $\delta^{15}\text{N}$ is visible at
510 250 m depth in the Eastern Equatorial Pacific (Fig. 5B; also discussed in several studies
511 (Rafter et al., 2012; Rafter and Sigman, 2016)) is caused by the remineralization of organic
512 matter with a low $\delta^{15}\text{N}$ due to partial nitrate consumption at the surface. Below we discuss
513 these and other influences on intermediate-depth nitrate $\delta^{15}\text{N}$.

514 515 **4.2 Intermediate-depth nitrate $\delta^{15}\text{N}$ variability**

516 Waters at “intermediate” depths (here shown as the 750 m surface in Fig. 5C) are important
517 because they are part of a large-scale circulation that initially upwells in the Southern
518 Ocean and ultimately resupplies nutrients to the low latitude thermocline (Palter et al.,
519 2010; Sarmiento et al., 2004; Toggweiler et al., 1991; Toggweiler and Carson, 1995). Within
520 the context of this overturning, the nitrate upwelling in the Southern Ocean is initially
521 $\approx 5\text{‰}$ (Figs. 4C & 5C) and the $\delta^{15}\text{N}$ is elevated $\approx 2\text{‰}$ by partial nitrate assimilation in
522 surface waters as they are advected equatorward (see Figs. 5A and 6). Deep [wintertime](#)
523 mixing in the Subantarctic Pacific converts these surface waters into mode and
524 intermediate waters (Herraiz-Borreguero and Rintoul, 2011), introducing nitrate with a
525 “pre-formed” $\delta^{15}\text{N}$ of $\approx 6\text{‰}$ into the intermediate-depth South Pacific and South Atlantic
526 (Rafter et al., 2012, 2013; Tuerena et al., 2015) at depths between ≈ 600 -1200 m. The

Microsoft Office User 3/10/2019 2:34 PM
Deleted: and in Fig. 1B

Microsoft Office User 3/10/2019 2:35 PM
Deleted: ing

Microsoft Office User 3/10/2019 2:35 PM
Deleted: of the

Microsoft Office User 3/10/2019 2:36 PM
Deleted: e

Microsoft Office User 3/10/2019 2:36 PM
Deleted: depths

532 penetration of this pre-formed signal (nitrate $\geq 6\text{‰}$) into the interior can be clearly seen in
533 the Atlantic Ocean between $\approx 40^\circ\text{S}$ to 20°N (Fig. 4B).

534
535 The same signal is carried with Southern Ocean mode and intermediate waters into the
536 Pacific basin as far as the tropics (Lehmann et al., 2018; Rafter et al., 2013), although it is
537 difficult to distinguish in the model results against the higher background $\delta^{15}\text{N}$ in this
538 basins (Figs. 4A, 4D, 5C). The same process presumably introduces elevated nitrate $\delta^{15}\text{N}$ to
539 the Indian Ocean, which has similar values at this depth. Nitrate $\delta^{15}\text{N}$ increases from the
540 Southern Ocean toward the equator in the Pacific and Indian Oceans, but not in the Atlantic
541 (Fig. 5C). Organic matter has a lower $\delta^{15}\text{N}$ in the Atlantic than in the Pacific and Indian
542 because of a lack of water-column denitrification supplying high- $\delta^{15}\text{N}$ water to the surface,
543 and because of the high rates of N_2 -fixation which supply isotopically light N to organic
544 matter (Marconi et al., 2017; Tuerena et al., 2015). This contrast in intermediate-depth
545 nitrate $\delta^{15}\text{N}$ can be traced to the lower $\delta^{15}\text{N}$ of organic matter remineralized in this
546 region—an explanation that is also consistent with enhanced N_2 fixation in the tropical
547 Atlantic (Marconi et al., 2017). The increase in intermediate-depth nitrate $\delta^{15}\text{N}$ from the
548 Subantarctic to the tropical Pacific appears to result from the remineralization of organic
549 matter with a $\delta^{15}\text{N}$ elevated by high source nitrate $\delta^{15}\text{N}$ (near the ODZ) or extreme
550 elevation of residual nitrate $\delta^{15}\text{N}$ (advected along the surface away from the equator; see
551 high surface nitrate $\delta^{15}\text{N}$ in Fig. 5A). Previous work suggests that direct mixing with
552 denitrified waters represents only a small fraction of the change from the pre-formed high
553 latitude value ($\approx 6\text{‰}$) to tropical nitrate $\delta^{15}\text{N}$ of $\approx 7\text{‰}$ (Peters et al., 2017; Rafter et al.,
554 2012, 2013).

555
556 The South Indian Ocean is one region particularly devoid of published nitrate $\delta^{15}\text{N}$
557 observations (Fig. 2), but the EANN makes specific predictions about its distribution. For
558 example, the modeled nitrate $\delta^{15}\text{N}$ predicts that intermediate-depth Indian Ocean nitrate is
559 similarly elevated in $\delta^{15}\text{N}$ to the intermediate-depth South Pacific (Fig. 5C). Considering
560 that both intermediate-depth water masses are formed from Southern Ocean surface
561 waters, it is reasonable to propose that nitrate $\delta^{15}\text{N}$ are similarly elevated by partial nitrate
562 consumption. The EANN therefore provides testable predictions for nitrate $\delta^{15}\text{N}$
563 observations throughout the Indian Ocean.

564 4.4 Deep-sea nitrate $\delta^{15}\text{N}$ trends

565 Our discussion above suggests that the basin-scale balance of N_2 -fixation and water-column
566 denitrification is a major contributor to inter-basin nitrate $\delta^{15}\text{N}$ gradients in the upper
567 ocean, lowering values in the Atlantic Oceans compared to the Pacific and Indian Oceans.
568 Averaging EANN nitrate $\delta^{15}\text{N}$ with depth for each ocean basin (Fig. 7), we find that these
569 basin-scale nitrate $\delta^{15}\text{N}$ differences also persist into the deep-sea (here defined as 3000 m
570 and below). (Note that the inter-basin EANN nitrate $\delta^{15}\text{N}$ gradients in Fig. 7 are smaller
571 than the corresponding inter-basin gradients in observed $\delta^{15}\text{N}$, because the observations
572 are spatially biased towards areas of water column denitrification in the Pacific and Indian
573 Oceans (see Fig. 2).)

574
575

Patrick Rafter 2/17/2019 1:13 PM

Deleted: Equivalent processes must drive the $\delta^{15}\text{N}$ in the intermediate-depth Indian Ocean, which is similarly elevated in the EANN, although direct observations are needed in order to confirm how well the EANN extrapolates in this region.

Microsoft Office User 3/10/2019 8:12 PM

Deleted: from the surface to 5500

Microsoft Office User 3/10/2019 8:12 PM

Deleted: m

Microsoft Office User 3/10/2019 8:13 PM

Deleted: -scale

Microsoft Office User 3/10/2019 8:12 PM

Deleted: model

Microsoft Office User 3/10/2019 8:13 PM

Deleted: differences

Microsoft Office User 3/10/2019 8:13 PM

Deleted: shown

Microsoft Office User 3/10/2019 8:13 PM

Deleted: even

Microsoft Office User 3/10/2019 8:13 PM

Deleted: larger for the nitrate

Microsoft Office User 3/10/2019 8:13 PM

Deleted: observations

Microsoft Office User 3/10/2019 8:13 PM

Deleted: measurements

592 The remineralization of organic matter is one process that can—and has been used to—
593 explain both the elevation of deep Pacific nitrate $\delta^{15}\text{N}$ (Peters et al., 2017; Rafter et al.,
594 2013; Sigman et al., 2009a)(Peters et al., 2017; Rafter et al., 2013; Sigman et al., 2009) and
595 lowering of deep Atlantic nitrate $\delta^{15}\text{N}$ (Knapp et al., 2008; Marconi et al., 2017; Tuerena et
596 al., 2015) relative to the deep ocean mean. Here we provide two additional pieces of
597 evidence that argue for the remineralization of organic matter as the key driver of these
598 deep-sea nitrate $\delta^{15}\text{N}$ differences. Our first piece of evidence is that the average subsurface
599 source of nitrate to the Pacific and Indian Ocean surface has a significantly higher $\delta^{15}\text{N}$ (by
600 2‰ at the 250 m depth surface) than the Atlantic and Southern Oceans (Figs. 5B and 7).
601 Nitrate $\delta^{15}\text{N}$ at 250 m is an admittedly imprecise estimate for the nitrate upwelled to the
602 surface, but even a slight elevation in Pacific source nitrate $\delta^{15}\text{N}$ and near complete nitrate
603 utilization at the surface will translate into higher sinking organic matter $\delta^{15}\text{N}$ (i.e., see Fig.
604 1B).

605
606 Our second piece of evidence that the export and remineralization of organic matter drives
607 the inter-basin differences in deep nitrate $\delta^{15}\text{N}$ comes from sediment trap measurements.
608 Averaging published sediment trap organic matter $\delta^{15}\text{N}$ from the subtropical and tropical
609 Pacific gives a value of $8.5 \pm 2.9\text{‰}$ (Knapp et al., 2016; Robinson et al., 2012), which is
610 significantly higher than measured in traps from the Atlantic ($4.5 \pm 1.5\text{‰}$) (Freudenthal et
611 al., 2001; Holmes et al., 2002; Lavik, 2000; Thunell et al., 2004). Given observed Southern
612 Ocean nitrate characteristics (Rafter et al., 2013), we estimate an even lower typical sinking
613 organic matter $\delta^{15}\text{N}$ of $+1.5\text{‰}$ for this region, which assumes initial nitrate values equal the
614 Upper Circumpolar Deep Water and final values from the Open Antarctic Zone. This value is
615 consistent with annually-averaged sinking organic matter $\delta^{15}\text{N}$ of ≈ 0.9 to 1.6‰ (Lourey et
616 al., 2003), although published results from the iron-fertilized Kerguelen Plateau region are
617 predictably higher (Trull et al., 2008). The much lower Southern Ocean sinking organic
618 matter $\delta^{15}\text{N}$ is consistent with partial consumption of nitrate at the surface (see Fig. 1B)
619 and the entrainment of this nitrate in equatorward-moving intermediate waters acts to
620 export nitrate with elevated $\delta^{15}\text{N}$ to intermediate waters throughout the Southern
621 Hemisphere (see discussion above). Based on this evidence, it appears that global patterns
622 of sinking organic matter $\delta^{15}\text{N}$ are consistent with the remineralization of this organic
623 matter driving subtle, but significant differences in deep-sea nitrate $\delta^{15}\text{N}$.

624
625 An alternative explanation for the deep-sea nitrate $\delta^{15}\text{N}$ differences in Fig. 7 is that they
626 reflect the lateral (along isopycnal) advection of elevated nitrate $\delta^{15}\text{N}$ from ODZ regions.
627 However, we can easily dismiss this explanation by looking at the meridional trends in
628 deep-sea nitrate $\delta^{15}\text{N}$ —following the deep waters from their entrance in the south and
629 movement northward. What we find is that deep EANN nitrate $\delta^{15}\text{N}$ (Fig. 5D) is lowest in
630 the Southern Ocean and increases equatorward in the Pacific. Average observed nitrate
631 $\delta^{15}\text{N}$ below 2500 m increases from $4.7 \pm 0.1\text{‰}$ in the Pacific sector of the Southern Ocean to
632 $4.9 \pm 0.2\text{‰}$ in the deep South Pacific, $5.4 \pm 0.2\text{‰}$ in the deep tropical Pacific, and $5.2 \pm 0.2\text{‰}$
633 in the deep North Pacific. This is consistent with the known increase in nitrate
634 concentrations and lowering of deep oxygen concentrations from the deep South to
635 Tropical and North Pacific (e.g., see Fig. 4E in (Rafter et al., 2013)). This contrasts with no
636 significant change in deep Atlantic nitrate $\delta^{15}\text{N}$, despite the export of slightly elevated

Patrick Rafter 3/7/2019 11:04 AM

Deleted: numbers in

638 nitrate $\delta^{15}\text{N}$ into intermediate-depth Atlantic (see above and (Tuerena et al., 2015)) and the
639 introduction of a different deep water mass (North Atlantic Deep Water) in the North
640 Atlantic. The distribution of deep Pacific nitrate $\delta^{15}\text{N}$ is coherent with elevated organic
641 matter $\delta^{15}\text{N}$ being produced and exported from the lower latitude surface and
642 remineralized at depth. In other words, inter-basin differences sinking organic matter $\delta^{15}\text{N}$
643 best explains the inter-basin differences in deep EANN and observed nitrate $\delta^{15}\text{N}$.
644 | Diapycnal mixing from the low latitude Pacific ODZ regions may also play a role in the
645 south-to-north elevation of deep Pacific nitrate $\delta^{15}\text{N}$, but we cannot quantify the magnitude
646 of that influence without a circulation model. Future work should look into this issue.

647 **5 Conclusions**

648 We find that an Ensemble of Artificial Neural Networks (EANN) can be trained on
649 climatological distributions of physical and biogeochemical tracers to reproduce a global
650 database of nitrate $\delta^{15}\text{N}$ observations (Fig. 2) with good fidelity (Fig. 3). We used the EANN
651 to produce global climatological maps of nitrate $\delta^{15}\text{N}$ at a 1 degree-resolution from the
652 surface to the seafloor. These results help identify spatial patterns (Figs. 4-6) and quantify
653 regional and basin-average oceanic values of nitrate $\delta^{15}\text{N}$ (Fig. 7). Major differences
654 between the observed and EANN-predicted nitrate $\delta^{15}\text{N}$ appear to be caused by temporal
655 variability of nitrate $\delta^{15}\text{N}$ in the upper ocean and in ODZs associated with variable nitrate
656 uptake and denitrification rates. Additional measurements of nitrate $\delta^{15}\text{N}$ will help to
657 develop seasonally-resolved maps that can improve upon the climatological mean map
658 provided here.
659

660
661 | Acknowledgments: M. Altabet, K. Casciotti, A. Santoro, B. Pasquier, [J.J. Becker](#), two
662 anonymous reviewers [and M. Kienast](#), as well as J. Granger and D. M. Sigman for (at-the-
663 time) unpublished data. A complete list of references can be found in the Appendix. The
664 compiled data set and data product is available in several online databases (BCO-DMO.org,
665 pangaea.de, and webodv.awi.de). [Many figures were made using Ocean Data View software](#)
666 [\(Schlitzer, 2002\)](#). Custom made color palettes and are available via [www.prafter.com](#).
667

Microsoft Office User 3/10/2019 8:17 PM

Deleted: Vertical, d

669 **References**

- 670 Altabet, M. A.: Constraints on oceanic N balance/imbalance from sedimentary N-15 records,
671 *Biogeosciences*, 4(1), 75–86, 2007.
- 672 Altabet, M. A. and Francois, R.: Sedimentary nitrogen isotopic ratio as a recorder for surface
673 ocean nitrate utilization, *Glob. Biogeochem. Cycles*, 8(1), 103–116, 1994a.
- 674 Altabet, M. A. and Francois, R.: The use of nitrogen isotopic ratio for reconstruction of past
675 changes in surface ocean nutrient utilization, in *Carbon Cycling in the Glacial Ocean:
676 Constraints on the Ocean's Role in Global Change*, vol. 117, pp. 281–306, Springer-Verlag
677 Berlin Heidelberg, 1994b.
- 678 Altabet, M. A. and Francois, R.: Nitrogen isotope biogeochemistry of the antarctic polar
679 frontal zone at 170 degrees W, *Deep-Sea Res. Part Ii-Top. Stud. Oceanogr.*, 48(19–20),
680 4247–4273, 2001.
- 681 Altabet, M. A., Murray, D. W. and Prell, W. L.: Climatically linked oscillations in Arabian Sea
682 denitrification over the past 1m.y.: Implications for the marine N cycle, *Paleoceanography*,
683 14(6), 732–743, 1999.
- 684 Altabet, M. A., Ryabenko, E., Stramma, L., Wallace, D. W. R., Frank, M., Grasse, P. and Lavik,
685 G.: An eddy-stimulated hotspot for fixed nitrogen-loss from the Peru oxygen minimum
686 zone, *Biogeosciences*, 9(12), 4897–4908, doi:10.5194/bg-9-4897-2012, 2012.
- 687 Altabet, M. A., Pilskaln, C. .. Thunell, R. .. Pride, C. .. Sigman, D. .. Chavez, F. .. Francois, R.: The
688 nitrogen isotope biogeochemistry of sinking particles from the margin of the Eastern North
689 Pacific, *Deep-Sea Res. Part -Oceanogr. Res. Pap.*, 46(4), 655–679, 1999.
- 690 Bourbonnais, A., Lehmann, M. F., Waniek, J. J. and Schulz-Bull, D. E.: Nitrate isotope
691 anomalies reflect N₂ fixation in the Azores Front region (subtropical NE Atlantic), *J.
692 Geophys. Res.*, 114(C3), doi:10.1029/2007JC004617, 2009.
- 693 Bourbonnais, A., Altabet, M. A., Charoenpong, C. N., Larkum, J., Hu, H., Bange, H. W. and
694 Stramma, L.: N-loss isotope effects in the Peru oxygen minimum zone studied using a
695 mesoscale eddy as a natural tracer experiment, *Glob. Biogeochem. Cycles*, 29(6), 793–811,
696 doi:10.1002/2014GB005001, 2015.
- 697 Brandes, J. A., Devol, A. H., Yoshinari, T., Jayakumar, D. A. and Naqvi, S. W. A.: Isotopic
698 composition of nitrate in the central Arabian Sea and eastern tropical North Pacific: A
699 tracer for mixing and nitrogen cycles, *Limnol. Oceanogr.*, 43(7), 1680–1689, 1998.
- 700 Breiman, L.: Bagging predictors, *Mach. Learn.*, 24(2), 123–140, doi:10.1007/BF00058655,
701 1996.
- 702 Carpenter, E. J., Harvey, H. R., Fry, B. and Capone, D. G.: Biogeochemical tracers of the
703 marine cyanobacterium *Trichodesmium*, *Deep-Sea Res. Part -Oceanogr. Res. Pap.*, 44(1),
704 27–38, doi:10.1016/s0967-0637(96)00091-x, 1997.

- 705 Casciotti, K. L. and McIlvin, M. R.: Isotopic analyses of nitrate and nitrite from reference
706 mixtures and application to Eastern Tropical North Pacific waters, *Mar. Chem.*, 107(2),
707 184–201, doi:10.1016/j.marchem.2007.06.021, 2007.
- 708 Casciotti, K. L., Trull, T. W., Glover, D. M. and Davies, D.: Constraints on nitrogen cycling at
709 the subtropical North Pacific Station ALOHA from isotopic measurements of nitrate and
710 particulate nitrogen, *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 55(14–15), 1661–1672,
711 doi:10.1016/j.dsr2.2008.04.017, 2008.
- 712 Casciotti, K. L., Buchwald, C. and McIlvin, M.: Implications of nitrate and nitrite isotopic
713 measurements for the mechanisms of nitrogen cycling in the Peru oxygen deficient zone,
714 *Deep Sea Res. Part Oceanogr. Res. Pap.*, 80, 78–93, doi:10.1016/j.dsr.2013.05.017, 2013.
- 715 Cline, J. D. and Kaplan, I. R.: Isotopic fractionation of dissolved nitrate during denitrification
716 in the eastern tropical North Pacific Ocean, *Mar. Chem.*, 3(4), 271–299, doi:10.1016/0304-
717 4203(75)90009-2, 1975.
- 718 Codispoti, L. and Christensen, J.: Nitrification, denitrification and nitrous oxide cycling in
719 the eastern tropical South Pacific ocean, *Mar. Chem.*, 16(4), 277–300, doi:10.1016/0304-
720 4203(85)90051-9, 1985.
- 721 De Pol-Holz, R., Robinson, R. S., Hebbeln, D., Sigman, D. M. and Ulloa, O.: Controls on
722 sedimentary nitrogen isotopes along the Chile margin, *Deep-Sea Res. Part II-Top. Stud.*
723 *Oceanogr.*, 56(16), 1100–1112, doi:10.1016/j.dsr2.2008.09.014, 2009.
- 724 Dehairs, F., Fripiat, F., Cavagna, A.-J., Trull, T. W., Fernandez, C., Davies, D., Roukaerts, A.,
725 Fonseca Batista, D., Planchon, F. and Elskens, M.: Nitrogen cycling in the Southern Ocean
726 Kerguelen Plateau area: evidence for significant surface nitrification from nitrate isotopic
727 compositions, *Biogeosciences*, 12(5), 1459–1482, doi:10.5194/bg-12-1459-2015, 2015.
- 728 Deutsch, C., Brix, H., Ito, T., Frenzel, H. and Thompson, L.: Climate-Forced Variability of
729 Ocean Hypoxia, *Science*, 333(6040), 336–339, doi:10.1126/science.1202422, 2011.
- 730 DeVries, T., Deutsch, C., Rafter, P. A. and Primeau, F.: Marine denitrification rates
731 determined from a global 3-dimensional inverse model, *Biogeosciences Discuss.*, 9(10),
732 14013–14052, doi:10.5194/bgd-9-14013-2012, 2012.
- 733 DiFiore, P. J., Sigman, D. M., Trull, T. W., Lourey, M. J., Karsh, K., Cane, G. and Ho, R.: Nitrogen
734 isotope constraints on subantarctic biogeochemistry, *J. Geophys. Res.-Oceans*, 111(C8),
735 doi:10.1029/2005jc003216, 2006.
- 736 Dugdale, R. C. and Goering, J. J.: Uptake of new and regenerated forms of nitrogen in
737 primary production, *Limnol. Oceanogr.*, 12(2), 196–206, 1967.
- 738 Eugster, O., Gruber, N., Deutsch, C., Jaccard, S. L. and Payne, M. R.: The dynamics of the
739 marine nitrogen cycle across the last deglaciation, *Paleoceanography*, 28(1), 116–129,
740 doi:10.1002/palo.20020, 2013.

- 741 Fawcett, S. E., Lomas, M., Casey, J. R., Ward, B. B. and Sigman, D. M.: Assimilation of upwelled
742 nitrate by small eukaryotes in the Sargasso Sea, *Nat. Geosci.*, 4(10), 717–722,
743 doi:10.1038/ngeo1265, 2011.
- 744 Finney, B. P., Gregory-Eaves, I., Douglas, M. S. V. and Smol, J. P.: Fisheries productivity in the
745 northeastern Pacific Ocean over the past 2,200 years, *Nature*, 416(6882), 729–733,
746 doi:10.1038/416729a, 2002.
- 747 Finney, B. P. et al.: Impacts of Climatic Change and Fishing on Pacific Salmon Abundance
748 Over the Past 300 Years, *Science*, 290, 795–799., 2000.
- 749 Freudenthal, T., Neuer, S., Meggers, H., Davenport, R. and Wefer, G.: Influence of lateral
750 particle advection and organic matter degradation on sediment accumulation and stable
751 nitrogen isotope ratios along a productivity gradient in the Canary Islands region, *Mar.
752 Geol.*, 17, 2001.
- 753 Fripiat, F., Declercq, M., Sapart, C. J., Anderson, L. G., Bruechert, V., Deman, F., Fonseca-
754 Batista, D., Humborg, C., Roukaerts, A., Semiletov, I. P. and Dehairs, F.: Influence of the
755 bordering shelves on nutrient distribution in the Arctic halocline inferred from water
756 column nitrate isotopes, *Limnol. Oceanogr.*, 63(5), 2018.
- 757 Galbraith, E. D.: Interactions between climate and the marine nitrogen cycle on glacial-
758 interglacial timescales, University of British Columbia, Vancouver., 2007.
- 759 Garcia, H. E., Locarnini, T. P., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K. and
760 Johnson, D. R.: Volume 4: Nutrients (phosphate, nitrate, and silicate), in *World Ocean Atlas*
761 2009, pp. 1–44, U.S. Government Printing Office., 2010.
- 762 Gaye, B., Nagel, B., Dähnke, K., Rixen, T. and Emeis, K.-C.: Evidence of parallel denitrification
763 and nitrite oxidation in the ODZ of the Arabian Sea from paired stable isotopes of nitrate
764 and nitrite, *Glob. Biogeochem. Cycles*, 27(4), 1059–1071, doi:10.1002/2011GB004115,
765 2013.
- 766 Graham, B. S., Grubbs, D., Holland, K. and Popp, B. N.: A rapid ontogenetic shift in the diet of
767 juvenile yellowfin tuna from Hawaii, *Mar. Biol.*, 150(4), 647–658, doi:10.1007/s00227-006-
768 0360-y, 2007.
- 769 Granger, J., Prokopenko, M. G., Sigman, D. M., Mordy, C. W., Morse, Z. M., Morales, L. V.,
770 Sambrotto, R. N. and Plessen, B.: Coupled nitrification-denitrification in sediment of the
771 eastern Bering Sea shelf leads to (15)N enrichment of fixed N in shelf waters, *J. Geophys.
772 Res.-Oceans*, 116, doi:10.1029/2010jc006751, 2011.
- 773 Granger, J., Prokopenko, M. G., Mordy, C. W. and Sigman, D. M.: The proportion of
774 remineralized nitrate on the ice-covered eastern Bering Sea shelf evidenced from the
775 oxygen isotope ratio of nitrate, *Glob. Biogeochem. Cycles*, 27(3), 962–971,
776 doi:10.1002/gbc.20075, 2013.

- 777 Herraiz-Borreguero, L. and Rintoul, S. R.: Subantarctic mode water: distribution and
778 circulation, *Ocean Dyn.*, 61(1), 103–126, doi:10.1007/s10236-010-0352-9, 2011.
- 779 Hoering, T. C. and Ford, H. T.: The isotope effect in the fixation of nitrogen by azotobacter, *J.*
780 *Am. Chem. Soc.*, 82(2), 376–378, doi:10.1021/ja01487a031, 1960.
- 781 Holmes, E., Lavik, G., Fischer, G., Segl, M., Ruhland, G. and Wefer, G.: Seasonal variability of
782 $\delta^{15}\text{N}$ in sinking particles in the Benguela upwelling region, , 18, 2002.
- 783 Karsh, K. L., Trull, T. W., Lourey, A. J. and Sigman, D. M.: Relationship of nitrogen isotope
784 fractionation to phytoplankton size and iron availability during the Southern Ocean Iron
785 RElease Experiment (SOIREE), *Limnol. Oceanogr.*, 48(3), 1058–1068, 2003.
- 786 Kemeny, P. C., Weigand, M. A., Zhang, R., Carter, B. R., Karsh, K. L., Fawcett, S. E. and Sigman,
787 D. M.: Enzyme-level interconversion of nitrate and nitrite in the fall mixed layer of the
788 Antarctic Ocean: Antarctic Fall Nitrate Isotopes, *Glob. Biogeochem. Cycles*, 30(7), 1069–
789 1085, doi:10.1002/2015GB005350, 2016.
- 790 Kienast, M., Lehmann, M. F., Timmermann, A., Galbraith, E., Bolliet, T., Holbourn, A.,
791 Normandeau, C. and Laj, C.: A mid-Holocene transition in the nitrogen dynamics of the
792 western equatorial Pacific: Evidence of a deepening thermocline?, *Geophys. Res. Lett.*,
793 35(23), doi:10.1029/2008gl035464, 2008.
- 794 Knapp, A. N., Sigman, D. M. and Lipschultz, F.: N isotopic composition of dissolved organic
795 nitrogen and nitrate at the Bermuda Atlantic time-series study site, *Glob. Biogeochem.*
796 *Cycles*, 19(1), doi:10.1029/2004gb002320, 2005.
- 797 Knapp, A. N., DiFiore, P. J., Deutsch, C., Sigman, D. M. and Lipschultz, F.: Nitrate isotopic
798 composition between Bermuda and Puerto Rico: Implications for N_2 fixation in the
799 Atlantic Ocean, *Glob. Biogeochem. Cycles*, 22(3), doi:10.1029/2007gb003107, 2008.
- 800 Knapp, A. N., Sigman, D. M., Lipschultz, F., Kustka, A. B. and Capone, D. G.: Interbasin isotopic
801 correspondence between upper-ocean bulk DON and subsurface nitrate and its
802 implications for marine nitrogen cycling, *Glob. Biogeochem. Cycles*, 25,
803 doi:10.1029/2010gb003878, 2011.
- 804 Knapp, A. N., Casciotti, K. L., Berelson, W. M., Prokopenko, M. G. and Capone, D. G.: Low rates
805 of nitrogen fixation in eastern tropical South Pacific surface waters, *Proc. Natl. Acad. Sci.*,
806 113(16), 4398–4403, doi:10.1073/pnas.1515641113, 2016.
- 807 Ko, Y. H., Lee, K., Takahashi, T., Karl, D. M., Kang, S.-H. and Lee, E.: Carbon-Based Estimate of
808 Nitrogen Fixation-Derived Net Community Production in N-Depleted Ocean Gyres, *Glob.*
809 *Biogeochem. Cycles*, doi:10.1029/2017GB005634, 2018.
- 810 Lavik, G.: Nitrogen isotopes of sinking matter and sediments in the South Atlantic,
811 Universität Bremen, Bremen, Deutschland., 2000.

- 812 Lehmann, M. F., Sigman, D. M., McCorkle, D. C., Brunelle, B. G., Hoffmann, S., Kienast, M.,
813 Cane, G. and Clement, J.: Origin of the deep Bering Sea nitrate deficit: Constraints from the
814 nitrogen and oxygen isotopic composition of water column nitrate and benthic nitrate
815 fluxes, *Glob. Biogeochem. Cycles*, 19(4), doi:10.1029/2005gb002508, 2005.
- 816 Lehmann, N., Granger, J., Kienast, M., Brown, K. S., Rafter, P. A., Martínez-Méndez, G. and
817 Mohtadi, M.: Isotopic Evidence for the Evolution of Subsurface Nitrate in the Western
818 Equatorial Pacific, *J. Geophys. Res. Oceans*, doi:10.1002/2017JC013527, 2018.
- 819 Liu, K. K.: Geochemistry of inorganic nitrogen compounds in two marine environments: The
820 Santa Barbara Basin and the ocean off Peru, University of Southern California, Los Angeles.,
821 1979.
- 822 Liu, K. K., Su, M. J., Hsueh, C. R. and Gong, G. C.: The nitrogen isotopic composition of nitrate
823 in the Kuroshio Water northeast of Taiwan: Evidence for nitrogen fixation as a source of
824 isotopically light nitrate, *Mar. Chem.*, 54(3-4), 273-292, doi:10.1016/0304-
825 4203(96)00034-5, 1996.
- 826 Lourey, M. J., Trull, T. W. and Sigman, D. M.: Sensitivity of delta N-15 of nitrate, surface
827 suspended and deep sinking particulate nitrogen to seasonal nitrate depletion in the
828 Southern Ocean, *Glob. Biogeochem. Cycles*, 17(3), doi:10.1029/2002gb001973, 2003.
- 829 Marconi, D., Alexandra Weigand, M., Rafter, P. A., McIlvin, M. R., Forbes, M., Casciotti, K. L.
830 and Sigman, D. M.: Nitrate isotope distributions on the US GEOTRACES North Atlantic
831 cross-basin section: Signals of polar nitrate sources and low latitude nitrogen cycling, *Mar.*
832 *Chem.*, 177, 143-156, doi:10.1016/j.marchem.2015.06.007, 2015.
- 833 Marconi, D., Sigman, D. M., Casciotti, K. L., Campbell, E. C., Alexandra Weigand, M., Fawcett,
834 S. E., Knapp, A. N., Rafter, P. A., Ward, B. B. and Haug, G. H.: Tropical Dominance of N₂
835 Fixation in the North Atlantic Ocean: Tropical Lead of Atlantic N₂ fixation, *Glob.*
836 *Biogeochem. Cycles*, 31(10), 1608-1623, doi:10.1002/2016GB005613, 2017.
- 837 Mariotti, A., Germon, J. C., Hubert, P., Kaiser, P., Letolle, R., Tardieux, A. and Tardieux, P.:
838 Experimental determination of nitrogen kinetic isotope fractionation—some principles—
839 illustration for the denitrification and nitrification processes, *Plant Soil*, 62(3), 413-430,
840 1981.
- 841 Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *J.*
842 *Soc. Ind. Appl. Math.*, 11(2), 431-441, doi:10.1137/0111030, 1963.
- 843 Martin, T. S. and Casciotti, K. L.: Paired N and O isotopic analysis of nitrate and nitrite in the
844 Arabian Sea oxygen deficient zone, *Deep Sea Res. Part Oceanogr. Res. Pap.*, 121, 121-131,
845 doi:10.1016/j.dsr.2017.01.002, 2017.
- 846 Miyake, Y. and Wada, E.: The Abundance Ratio of ¹⁵N/¹⁴N in Marine Environments, *Rec.*
847 *Oceanogr. Works Jpn.*, 9(1), 1967.

848 Palter, J. B., Sarmiento, J. L., Gnanadesikan, A., Simeon, J. and Slater, R. D.: Fueling export
849 production: nutrient return pathways from the deep ocean and their dependence on the
850 Meridional Overturning Circulation, *Biogeosciences*, 7(11), 3549–3568, doi:10.5194/bg-7-
851 3549-2010, 2010.

852 Pantoja, S., Repeta, D. J., Sachs, J. P. and Sigman, D. M.: Stable isotope constraints on the
853 nitrogen cycle of the Mediterranean Sea water column, *Deep-Sea Res. Part -Oceanogr. Res.*
854 *Pap.*, 49(9), 1609–1621, doi:10.1016/s0967-0637(02)00066-3, 2002.

855 Peters, B. D., Lam, P. J. and Casciotti, K. L.: Nitrogen and oxygen isotope measurements of
856 nitrate along the US GEOTRACES Eastern Pacific Zonal Transect (GP16) yield insights into
857 nitrate supply, remineralization, and water mass transport, *Mar. Chem.*,
858 doi:10.1016/j.marchem.2017.09.009, 2017.

859 Rafter, P. A. and Charles, C. D.: Pleistocene equatorial Pacific dynamics inferred from the
860 zonal asymmetry in sedimentary nitrogen isotopes, *Paleoceanography*, 27,
861 doi:10.1029/2012pa002367, 2012.

862 Rafter, P. A. and Sigman, D. M.: Spatial distribution and temporal variation of nitrate
863 nitrogen and oxygen isotopes in the upper equatorial Pacific Ocean, *Limnol. Oceanogr.*,
864 61(1), 14–31, doi:10.1002/lno.10152, 2016.

865 Rafter, P. A., Sigman, D. M., Charles, C. D., Kaiser, J. and Haug, G. H.: Subsurface tropical
866 Pacific nitrogen isotopic composition of nitrate: Biogeochemical signals and their transport,
867 *Glob. Biogeochem. Cycles*, 26, doi:10.1029/2010gb003979, 2012.

868 Rafter, P. A., DiFiore, P. J. and Sigman, D. M.: Coupled nitrate nitrogen and oxygen isotopes
869 and organic matter remineralization in the Southern and Pacific Oceans, *J. Geophys. Res.*
870 *Oceans*, 118, 1–14, doi:10.1002/jgrc.20316, 2013.

871 Ren, H., Sigman, D. M., Meckler, A. N., Plessen, B., Robinson, R. S., Rosenthal, Y. and Haug, G.
872 H.: Foraminiferal Isotope Evidence of Reduced Nitrogen Fixation in the Ice Age Atlantic
873 Ocean, *Science*, 323(5911), 244–248, doi:10.1126/science.1165787, 2009.

874 Ren, H., Chen, Y.-C., Wang, X. T., Wong, G. T. F., Cohen, A. L., DeCarlo, T. M., Weigand, M. A.,
875 Mii, H.-S. and Sigman, D. M.: 21st-century rise in anthropogenic nitrogen deposition on a
876 remote coral reef, *Science*, 356(6339), 749–752, doi:10.1126/science.aal3869, 2017.

877 Robinson, R. S., Brunelle, B. G. and Sigman, D. M.: Revisiting nutrient utilization in the glacial
878 Antarctic: Evidence from a new method for diatom-bound N isotopic analysis,
879 *Paleoceanography*, 19(3), doi:10.1029/2003pa000996, 2004.

880 Robinson, R. S., Kienast, M., Albuquerque, A. L., Altabet, M., Contreras, S., Holz, R. D., Dubois,
881 N., Francois, R., Galbraith, E., Hsu, T. C., Ivanochko, T., Jaccard, S., Kao, S. J., Kiefer, T., Kienast,
882 S., Lehmann, M. F., Martinez, P., McCarthy, M., Mobius, J., Pedersen, T., Quan, T. M.,
883 Ryabenko, E., Schmittner, A., Schneider, R., Schneider-Mor, A., Shigemitsu, M., Sinclair, D.,

- 884 Somes, C., Studer, A., Thunell, R. and Yang, J. Y.: A review of nitrogen isotopic alteration in
885 marine sediments, *Paleoceanography*, 27, doi:10.1029/2012pa002321, 2012.
- 886 Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning Representations by Back-
887 Propagating Errors, *Nature*, 323(6088), 533–536, doi:10.1038/323533a0, 1986.
- 888 Ryabenko, E., Kock, A., Bange, H. W., Altabet, M. A. and Wallace, D. W. R.: Contrasting
889 biogeochemistry of nitrogen in the Atlantic and Pacific Oxygen Minimum Zones,
890 *Biogeosciences*, 9(1), 203–215, doi:10.5194/bg-9-203-2012, 2012.
- 891 Sachs, J. P., Repeta D. J.: Oligotrophy and nitrogen fixation during eastern Mediterranean
892 sapropel events, *Science*, 286(5449), 2485–2488, 1999.
- 893 Sarmiento, J. L., Gruber, N., Brzezinski, M. A. and Dunne, J. P.: High-latitude controls of
894 thermocline nutrients and low latitude biological productivity, *Nature*, 427(6969), 56–60,
895 doi:10.1038/nature02127, 2004.
- 896 Schlitzer, R.: Ocean Data View. [online] Available from: [http://www.awi-](http://www.awi-bremerhaven.de/GEO/ODV)
897 [bremerhaven.de/GEO/ODV](http://www.awi-bremerhaven.de/GEO/ODV), 2002.
- 898 Schlitzer, R., Anderson, R. F., Dodas, E. M., Lohan, M., Geibert, W., Tagliabue, A., Bowie, A.,
899 Jeandel, C., Maldonado, M. T., Landing, W. M., Cockwell, D., Abadie, C., Abouchami, W.,
900 Achterberg, E. P., Agather, A., Aguiar-Isas, A., van Aken, H. M., Andersen, M., Archer, C.,
901 Auro, M., de Baar, H. J., Baars, O., Baker, A. R., Bakker, K., Basak, C., Baskaran, M., Bates, N. R.,
902 Bauch, D., van Beek, P., Behrens, M. K., Black, E., Bluhm, K., Bopp, L., Bouman, H., Bowman,
903 K., Bown, J., Boyd, P., Boye, M., Boyle, E. A., Branellec, P., Bridgestock, L., Brissebrat, G.,
904 Browning, T., Bruland, K. W., Brumsack, H.-J., Brzezinski, M., Buck, C. S., Buck, K. N.,
905 Buesseler, K., Bull, A., Butler, E., Cai, P., Mor, P. C., Cardinal, D., Carlson, C., Carrasco, G.,
906 Casacuberta, N., Casciotti, K. L., Castrillejo, M., Chamizo, E., Chance, R., Charette, M. A.,
907 Chaves, J. E., Cheng, H., Chever, F., Christl, M., Church, T. M., Closset, I., Colman, A., Conway,
908 T. M., Cossa, D., Croot, P., Cullen, J. T., Cutter, G. A., Daniels, C., Dehairs, F., Deng, F., Dieu, H.
909 T., Duggan, B., Dulaquais, G., Dumousseaud, C., Echegoyen-Sanz, Y., Edwards, R. L., Ellwood,
910 M., Fahrbach, E., Fitzsimmons, J. N., Russell Flegal, A., Fleisher, M. Q., van de Flierdt, T.,
911 Frank, M., Friedrich, J., Fripiat, F., Fröllje, H., Galer, S. J. G., Gamo, T., Ganeshram, R. S., Garcia-
912 Orellana, J., Garcia-Solsona, E., Gault-Ringold, M., et al.: The GEOTRACES Intermediate Data
913 Product 2017, *Chem. Geol.*, doi:10.1016/j.chemgeo.2018.05.040, 2018.
- 914 Sigman, D. M. and Casciotti, K. L.: Nitrogen Isotopes in the Ocean, in *Encyclopedia of Ocean*
915 *Science*, pp. 1884–1894, Academic Press., 2001.
- 916 Sigman, D. M., Altabet, M. A., McCorkle, D. C., Francois, R. and Fischer, G.: The delta N-15 of
917 nitrate in the Southern Ocean: Consumption of nitrate in surface waters, *Glob. Biogeochem.*
918 *Cycles*, 13(4), 1149–1166, 1999a.
- 919 Sigman, D. M., Altabet, M. A., Francois, R., McCorkle, D. C. and Gaillard, J. F.: The isotopic
920 composition of diatom-bound nitrogen in Southern Ocean sediments, *Paleoceanography*,
921 14(2), 118–134, 1999b.

- 922 Sigman, D. M., Granger, J., DiFiore, P. J., Lehmann, M. M., Ho, R., Cane, G. and van Geen, A.:
923 Coupled nitrogen and oxygen isotope measurements of nitrate along the eastern North
924 Pacific margin, *Glob. Biogeochem. Cycles*, 19(4), doi:10.1029/2005gb002458, 2005.
- 925 Sigman, D. M., DiFiore, P. J., Hain, M. P., Deutsch, C. and Karl, D. M.: Sinking organic matter
926 spreads the nitrogen isotope signal of pelagic denitrification in the North Pacific, *Geophys.*
927 *Res. Lett.*, 36, doi:10.1029/2008gl035784, 2009a.
- 928 Sigman, D. M., DiFiore, P. J., Hain, M. P., Deutsch, C., Wang, Y., Karl, D. M., Knapp, A. N.,
929 Lehmann, M. F. and Pantoja, S.: The dual isotopes of deep nitrate as a constraint on the
930 cycle and budget of oceanic fixed nitrogen, *Deep-Sea Res. Part -Oceanogr. Res. Pap.*, 56(9),
931 1419–1439, doi:10.1016/j.dsr.2009.04.007, 2009b.
- 932 Smart, S. M., Fawcett, S. E., Thomalla, S. J., Weigand, M. A., Reason, C. J. C. and Sigman, D. M.:
933 Isotopic evidence for nitrification in the Antarctic winter mixed layer, *Glob. Biogeochem.*
934 *Cycles*, 29(4), 427–445, doi:10.1002/2014GB005013, 2015.
- 935 Somes, C. J., Schmittner, A. and Altabet, M. A.: Nitrogen isotope simulations show the
936 importance of atmospheric iron deposition for nitrogen fixation across the Pacific Ocean,
937 *Geophys. Res. Lett.*, 37, doi:10.1029/2010gl044537, 2010.
- 938 Tawa, A., Ishihara, T., Uematsu, Y., Ono, T. and Ohshimo, S.: Evidence of westward
939 transoceanic migration of Pacific bluefin tuna in the Sea of Japan based on stable isotope
940 analysis, *Mar. Biol.*, 164(4), doi:10.1007/s00227-017-3127-8, 2017.
- 941 Thimm, G. and Fiesler, E.: Optimal Setting of Weights, Learning Rate, and Gain, , 15, 1997.
- 942 Thunell, R. C., Sigman, D. M., Muller-Karger, F., Astor, Y. and Varela, R.: Nitrogen isotope
943 dynamics of the Cariaco Basin, Venezuela, *Glob. Biogeochem. Cycles*, 18(3),
944 doi:10.1029/2003gb002185, 2004.
- 945 Toggweiler, J. R. and Carson, S.: What Are Upwelling Systems Contributing to the Ocean's
946 Carbon and Nutrient Budgets?, in *Upwelling in the Ocean: Modern Processes and Ancient*
947 *Records*, edited by K.-C. . meis C.P. Summerhayes M. V. Angel, R. L. Smith, and B. Zeizschcl,
948 John Wiley & Sons Ltd., 1995.
- 949 Toggweiler, J. R., Dixon, K. and Broecker, W. S.: The Peru upwelling and the ventilation of
950 the South-Pacific thermocline, *J. Geophys. Res.-Oceans*, 96(C11), 20467–20497,
951 doi:10.1029/91jc02063, 1991.
- 952 Trull, T. W., Davies, D. and Casciotti, K.: Insights into nutrient assimilation and export in
953 naturally iron-fertilized waters of the Southern Ocean from nitrogen, carbon and oxygen
954 isotopes, *Deep Sea Res. Part II Top. Stud. Oceanogr.*, 55(5–7), 820–840,
955 doi:10.1016/j.dsr2.2007.12.035, 2008.
- 956 Tuerena, R. E., Ganeshram, R. S., Geibert, W., Fallick, A. E., Dougans, J., Tait, A., Henley, S. F.
957 and Woodward, E. M. S.: Nutrient cycling in the Atlantic basin: The evolution of nitrate

- 958 isotope signatures in water masses, *Glob. Biogeochem. Cycles*, 29(10), 1830–1844,
959 doi:10.1002/2015GB005164, 2015.
- 960 Umezawa, Y., Yamaguchi, A., Ishizaka, J., Hasegawa, T., Yoshimizu, C., Tayasu, I., Yoshimura,
961 H., Morii, Y., Aoshima, T. and Yamawaki, N.: Seasonal shifts in the contributions of the
962 Changjiang River and the Kuroshio Current to nitrate dynamics in the continental shelf of
963 the northern East China Sea based on a nitrate dual isotopic composition approach,
964 *Biogeosciences*, 11(4), 1297–1317, doi:10.5194/bg-11-1297-2014, 2014.
- 965 Voss, M.: Räumliche und zeitliche Verteilung stabiler Isotope (d15N, d13C) in
966 suspendierten und sedimentierten Partikeln im Nördlichen Nordatlantik, Christian-
967 Albrechts-Universität zu Kiel, 1991.
- 968 Voss, M., Dippner, J. W. and Montoya, J. P.: Nitrogen isotope patterns in the oxygen-deficient
969 waters of the Eastern Tropical North Pacific Ocean, *Deep-Sea Res. Part -Oceanogr. Res. Pap.*,
970 48(8), 1905–1921, doi:10.1016/s0967-0637(00)00110-2, 2001.
- 971 Wada, E.: Nitrogen Isotope Fractionation and Its Significance in Biogeochemical Processes
972 Occurring in Marine Environments, in *Isotope Marine Chemistry*, pp. 375–398., 1980.
- 973 Wada, E. and Hattori, A.: Nitrogen isotope effects in the assimilation of inorganic
974 nitrogenous compounds by marine diatoms, *Geomicrobiol. J.*, 1(1), 85–101, 1978.
- 975 Weigend, A. S., Huberman, B. A. and Rumelhart, D. E.: Predicting The Future: A
976 Connectionist Approach, *Int. J. Neural Syst.*, 1(3), 193–209,
977 doi:10.1142/s0129065790000102, 1990.
- 978 Wong, G. T. F., Chung, S.-W., Shiah, F.-K., Chen, C.-C., Wen, L.-S. and Liu, K.-K.: Nitrate
979 anomaly in the upper nutricline in the northern South China Sea - Evidence for nitrogen
980 fixation, *Geophys. Res. Lett.*, 29(23), 12-1-12-4, doi:10.1029/2002GL015796, 2002.
- 981 Wu, J., Calvert, S. E. and Wong, C. S.: Nitrogen isotope variations in the subarctic northeast
982 Pacific: relationships to nitrate utilization and trophic structure, *Deep Sea Res. Part*
983 *Oceanogr. Res. Pap.*, 44(2), 287–314, 1997.
- 984 Yang, S. and Gruber, N.: The anthropogenic perturbation of the marine nitrogen cycle by
985 atmospheric deposition: Nitrogen cycle feedbacks and the ¹⁵N Haber-Bosch effect, *Glob.*
986 *Biogeochem. Cycles*, 30(10), 1418–1440, doi:10.1002/2016GB005421, 2016.
- 987 Yang, S., Gruber, N., Long, M. C. and Vogt, M.: ENSO-Driven Variability of Denitrification and
988 Suboxia in the Eastern Tropical Pacific Ocean, *Glob. Biogeochem. Cycles*, 31(10), 1470–
989 1487, doi:10.1002/2016GB005596, 2017.
- 990 Yoshikawa, C., Nakatsuka, T. and Wakatsuchi, M.: Distribution of N* in the Sea of Okhotsk
991 and its use as a biogeochemical tracer of the Okhotsk Sea Intermediate Water formation
992 process, *J. Mar. Syst.*, 63(1–2), 49–62, doi:10.1016/j.jmarsys.2006.05.008, 2006.

993 Yoshikawa, C., Makabe, A., Shiozaki, T., Toyoda, S., Yoshida, O., Furuya, K. and Yoshida, N.:
994 Nitrogen isotope ratios of nitrate and N* anomalies in the subtropical South Pacific,
995 *Geochem. Geophys. Geosystems*, 16(5), 1439–1448, doi:10.1002/2014GC005678, 2015.

996
997

998 **Appendix: References for this version of seawater nitrate $\delta^{15}\text{N}$ compilation**

region	year of sampling	month of sampling	reference
Pacific North – Subarctic	Unknown sampling date	<i>na</i>	(Altabet and Francois, 1994b)
Indian – Arabian Sea	1995	8	(Altabet et al., 1999)
Southern Ocean - Pacific	1996-1998	1-4,8-11	(Altabet and Francois, 2001)
Pacific North – Gulf of California	1990	6	(Altabet, 1999)
Pacific North – Subarctic	1971	7	(Wada, 1980)
Indian – Arabian Sea	1994	4	(Brandes et al., 1998)
Pacific North – ETNP	1993	12	(Brandes et al., 1998)
Pacific North - Kuroshio	1992 & 1994	3 & 4	(Liu et al., 1996)
Pacific North – Tropical	1997	10 & 11	(Voss et al., 2001)
Pacific North – Subarctic	2003	2	(Galbraith, 2007)
Atlantic North	2004	5	(Bourbonnais et al., 2009)
Mediterranean	1996	5	(Sachs, 1999)
Mediterranean	1998	1	(Pantoja et al., 2002)
Pacific North – Subarctic	2002	6	(Lehmann et al., 2005)
Pacific South – Tropical	1977	6	(Liu, 1979)
Pacific South – Tropical	2002 & 2004	4 & 5	(De Pol-Holz et al., 2009)
Pacific North – Okhotsk	1998, 1999, 2000	6 & 9	(Yoshikawa et al., 2006)
Pacific Tropical	2006	6	(Kienast et al., 2008)
Southern Ocean – Indian	2005	1 & 2	(Trull et al., 2008)
Pacific South – Tropical	2008 & 2009	1, 2, & 12	(Ryabenko et al., 2012)
Indian South	2011	10 & 11	(Dehairs et al., 2015)
Pacific South – Tropical	2012	11	(Bourbonnais et al., 2015)
Indian North	2007	9	(Gaye et al., 2013)
Atlantic South	2010 & 2012	10 & 1	(Tuerena et al., 2015)
Pacific South	2009	6	(Yoshikawa et al., 2015)

Pacific North –SCS	1997	4	(Wong et al., 2002)
Pacific North – Bering Sea	2008 & 2007	4 & 5	(Granger et al., 2011, 2013)
Arctic – Beaufort	2009	9	Granger unpublished
Atlantic North	2010	10 & 11	Jenkins et al. Unpublished GEOTRACES (Knorr_199_leg4.pdf)
Atlantic Tropical	2010	2 & 3	Frank et al. Unpublished GEOTRACES (meteor81_1.pdf)
Pacific Tropical	2013	5 & 6	(Lehmann et al., 2018)
Pacific North	2008	7	Granger Unpublished
Pacific North	2009 & 2011	2 & 7	(Umezawa et al., 2014)
Pacific South – Tropical	2010 & 2011	2-4	(Knapp et al., 2016)
Atlantic North – Subarctic	1989	6	(Voss, 1991)
Southern Ocean – Pacific	1995	4	(Sigman et al., 1999a)
Southern Ocean – Indian	1995	1	(Sigman et al., 1999a)
Southern Ocean – Pacific	2016		(Kemeny et al., 2016)
Pacific North – Tropical	2003	10	(Sigman et al., 2005)
Pacific North – ALOHA	2000	11	(Sigman et al., 2009b)
Atlantic – North	2001-2002	1-12	(Knapp et al., 2005)
Atlantic – North	2002	10	(Knapp et al., 2008)
Pacific – North	2003	7 & 8	(Knapp et al., 2011)
Indian - South	1999	1 & 2	(Karsh et al., 2003)
Southern – Indian	1998 & 1999	2,3,4,9, & 12	(DiFiore et al., 2006)
Southern – Atlantic	2012	7	(Smart et al., 2015)
Atlantic – North	2011	10 & 11	(Marconi et al., 2015)
Pacific – North ALOHA	2004	7	(Casciotti et al., 2008)
Pacific – South Tropical	2005	11	(Casciotti et al., 2013)
Pacific – North Tropical	2003	11	(Casciotti and Mcllvain, 2007)
Indian – Arabian	2007	9	(Martin and Casciotti, 2017)
Pacific – Tropical	2004-2007	3-12	(Rafter et al., 2012; Rafter and Sigman, 2016)

Indian – Arabian Sea	2007	10	(DeVries et al., 2012) or Rafter and Sigman Unpublished
Pacific South – Tasman Sea	2010	1 & 2	Rafter and Sigman Unpublished
Pacific South – Tropical	2010	2	(Rafter et al., 2012)
Atlantic North – Subarctic	2010	4	Rafter and Sigman Unpublished
Pacific South	2005	1	(Rafter et al., 2013)
Pacific South – Tropical	2013	10-12	(Peters et al., 2017)
Atlantic North – Subarctic	2013	8	(Marconi et al., 2017)
Pacific North – Subarctic	1993	5	(Wu et al., 1997)
Arctic	2014	7 & 8	(Fripiat et al., 2018)

999
1000