

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

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

C1

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 $\delta^{15}\text{N}$ values from the gridded T, S, NO_3^- , O_2 , and PO_4^{3-} data?

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

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

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

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?

C2

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

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?

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.

For example, 1) Is there only one way to produce the neural network model? 2) How were choices made? What tradeoffs were tolerated? What are the implications? 3) How does this approach compare with other methods for gridding? 4) Are there particular nodes that performed well in some locations vs. others?

Lines 415-423: It's not clear how the authors 'easily dismiss' an explanation about lateral advection of elevated nitrate $\delta^{15}\text{N}$ 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.

Figures— Figure 2—How many different 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?

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

C3

the training set? Or validation set?

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 $\delta^{15}\text{N}$ from the ODZs in the eastern tropical Pacific.

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

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?

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2018-525>, 2019.

C4