**Biogeosciences
Discussions**

Interactive comment on "Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network" *by* M. Telszewski et al.

Y. Liu (Referee)

yliu18@gmail.com

Received and published: 14 April 2009

General comments:

This manuscript presents an estimate of the basin-wide monthly pCO2 distribution from the sparsely measured pCO2 and the basin-wide SST, CHL and MLD maps by using the SOM neural network. The topic is within the scope of Biogeosciences, and the method is of interest to an even wider scope of readers. As a novel and powerful data analysis tool, the SOM is becoming popular in various disciplines. I believe this timely work will have an immediate impact in biogeosciences community.

The description of the method is clear. The assumptions and the implementation of

the experiments (SOM analysis) are acceptable, and the results are better than those obtained by others. Thus, I recommend it to be published in Biogeosciences after very minor revisions on the interpretation of the SOM results, processing of the data, and citation of a new paper. The paper can be more concise, and English writing also needs to be improved in general. As a non-expert of the bioscience, I reserve my comments on the biogeoscience part of the manuscript.

Specific comments:

(1) P3385 L20, "some blooms are not resolved by SOM", and P3396 L8-9, "Discrepancies identified in the eastern subtropics reveal the method's tendency to 'smooth' highest and lowest values". Another explanation of the smoothed patterns is the shape of the neighborhood function. The Gaussian neighborhood function was chosen in the SOM training. According to a sensitivity study on the choice of neighborhood function in extracting the known patterns of progressive sine waves, the Gaussian neighborhood function returns the smoothest SOM patterns, while the Epanechikov (ep) neighborhood function gives the most accurate patterns (see Figure 5 of Liu et al., 2006b). This difference is also seen from the real geophysical data (moored time series of coastal ocean currents). The patterns of strong currents due to the hurricane and tropical storm activities were extracted using the "ep" neighborhood function (Liu et al., 2006b), however, these extreme current patterns were not seen from the SOM patterns using the Gaussian neighborhood function (Liu and Weisberg, 2005). So, if the "ep" function is used instead, the SOM patterns might be less smoothed. A repetition of the analysis with a different neighborhood function is not suggested here, however, a few sentences need to be added in the text to clarify the reason why some strong blooms are smoothed by the SOM. This would be indicative to other SOM users as well. Both papers mentioned above (Liu and Weisberg, 2005; Liu et al. 2006b) have already been cited in the manuscript.

(2) P3380 L14-15, "The SST analyses were done weekly and interpolated linearly to daily values", and later in P3381, L4, "All parameters were re-gridded onto weekly

frequency ..." It seems that the interpolation to daily time series is not necessary, because the SST product is already weekly. An interpolation in time may be necessary if the two weekly time stamps are different. Even though the extra linear interpolation may not substantially modify the SST, the less manipulation of the original data, the better.

(3) In a recent paper (Friedrich and Oschlies, 2009), basin-wide monthly maps of pCO2 in 2005 were derived from model results and satellite SST and CHL using the SOM (they called it "KFM" instead). I suggest the authors to cite that paper and compare this work with theirs. As I see, the data sets are different (there are some overlaps, though), and the results are much better in the present analysis according to the reduced RMS values ($\sim$12 versus $\sim$20 μatm).

Technical corrections:

(1) P3375 L16, "cover" should be "coverage". This change should also apply to many other places in the text, e.g., P3376 L15&27, P3381 L1&3, P3390 L25, P3393 L2, Table 1.

(2) P3384 L2, "Many more" should be "More".

(3) P3384 L7, "few" should be "less".

(4) P3386 L12-13, the sentence "Most neurons ... of the data" needs rewriting.

(5) P3388 L10, "daily" should be "day".

(6) P3399 L9, "Weisberg R." should be "Weisberg R. H."

(7) P3399 L11, "Liu, Y. and Weisberg R." should be "Liu, Y., Weisberg, R. H., and He, R."

(8) P3399 L13, "Weisberg R." should be "Weisberg R. H."

References:

Friedrich, T., and A. Oschlies (2009), Neural network-based estimates of North Atlantic surface pCO2 from satellite data: A methodological study, J. Geophys. Res., 114, C03020, doi:10.1029/2007JC004646.

Liu, Y., and R. H. Weisberg (2005), Patterns of ocean current variability on the West Florida Shelf using the self-organizing map, J. Geophys. Res., 110, C06003, doi:10.1029/2004JC002786.

Liu, Y., R. H. Weisberg, and C. N. K. Mooers (2006b), Performance evaluation of the self-organizing map for feature extraction, J. Geophys. Res., 111, C05018, doi:10.1029/2005JC003117.