Biogeosciences Discuss., 10, C2892–C2895, 2013
www.biogeosciences-discuss.net/10/C2892/2013/

Biogeosciences
Discussions
Open Access

Interactive comment on "Testing the applicability of neural networks as a gap-filling method using CH_4 flux data from high latitude wetlands" *by* S. Dengel et al.

D. Papale (Referee)

darpap@unitus.it

Received and published: 19 June 2013

The paper presents an application of Artificial Neural Network to fill gaps in CH4 fluxes measurements collected using the eddy covariance technique. The analysis is promising and results interesting but I think that additional analysis/test are needed to correctly interpret the results and try to improve the results that at the moment are questionable.

P7733 L17: WD is a variable that should not be used as it is. This because a WD of 1 degree and 359 degrees are ecologically identical but would have a completely different weight in the ANN. Generally these variables are decomposed as sin and cos of the angle to avoid this problem.

P7733 L17-18: the ustar filtering is an important aspect that deserve some more details. In particular I would like to see explained how the threshold has been calculated (is the CO2 threshold used?).

P7733 L19: why the Kytalyk site has not been filtered by ustar? It is relevant for the paper because under the assumption that ustar filtering remove data where fluxes are underestimated, keeping these data in the dataset would interfere with the ANN training (there could be halfhours with the same meteo conditions but different fluxes due to advection)

P7734 L2: the representativeness of the dataset is important. You should explain what you mean by representative (of the fluxes? Of the driver? Of the time?) and how this has been implemented in the training. How did you ensured this? For example that the ANN has been not over-trained with daytime data just because there are more example since the ustar cut more at night. I don't like self citation but the concept I'm referring to is described in the Aubinet 2012 eddy covariance book, section 6.3.3.3.

P7735 L13-16: I would have done the opposite, because the diurnal cycle is given somehow by the incoming radiation. Did you test both?

P7736 L13-15: the sentence is not very clear, I suggest to reformulate it. In addition I would add that since you are proposing a method for gapfilling, in general when the fluxes are missing also ustar is missing.

P7736 L19: the lagged effect is important and interesting. Could be an important added value to the paper the test of these variables. For example the cumulated precipitation or temperature of previous hours/days, or also an interpolation of the water table depth.

P7736 L20-P7737 L9: It is not clear why this analysis has been done. It would have been justified if used to select the drivers for the ANN, but with a fix list defined it is somehow out of context (not needed for the ANN application). May be a better explanation would help. However the figure are impossible to read because too small.

In you want to keep this analysis in please better define the context, think about different figures and give a general introduction about how to interpret the results.

P7737 L20-21: what does it mean that the mix scenario (so a distribution of artificial gaps added to the dataset) has been chosen to gap-fill the measurements?

P7740 L22: "For each neuron added to the hidden layer. . ."

P7740 L23-24: the Pearson Coeff is reported in black for training and grey for the test set (as reported in the legend), so opposite to what reported in the text.

P7741 L7-12: this figure is not needed. However would be better with only the real validation data (no training data). It is important also to specify which artificial gap scenario are representing.

P7741 L25-27: it would be interesting also to test different input combinations. It is not needed to use the same drivers for all the sites (sites are different so drivers could have different importance, in some site could be missing and this is a problem in the gapfilling). I suggest to add this analysis that is relevant and useful, together with the lagged variables.

P7742 L20-26: I don't think that the insufficient results obtained in the testing/validation are due to a problem of local minimum, that should be limited increasing the number of initializations. Dataset representativeness, input used and missing, difficulties in the fluxes dynamic are probably the reasons to investigate and to try to solve.

P7742 L23; L29: the use of the term testing/validation is confusing, Which method did you use to avoid the overfitting? The Early stopping? If so, this should be explained and the dataset used in this process identified as "test" dataset. Then the 10% excluded with the artificial gaps are the "validation" data. Otherwise explain what you did and what are validation and testing datasets.

P7743 L1-2: the fact that artificial gaps can include pre-existing gaps is a problem in the results interpretation. For example an artificial gap of 12days with only one day of

measurements, should not be interpreted as result for the large-gaps-scenario. I would suggest to allow a shift of the artificial gap of +/- half of the length in order to maximize the number of data points removed.

Figures 4 and 5: The figure is dominated by the black dots that can only increase (r2) or decrease (RMSE) because the training do this (reduce the error) and for this reason are not significant. What is more important is the performances on the test data. For this reason I suggest to remove the training data results, or at least to make them light grey and the test data in black.

Dario Papale

Interactive comment on Biogeosciences Discuss., 10, 7727, 2013.