Biogeosciences Discuss., 10, C4322–C4328, 2013
www.biogeosciences-discuss.net/10/C4322/2013/

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.

S. Dengel et al.

sigrid.dengel@helsinki.fi

Received and published: 15 August 2013

Reply to referee#1 (D. Papale): (Corresponding author: Sigrid Dengel) (The answers appear after each of the referee's comment)

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.

Reply: Wind direction as an input variable has been decomposed into its horizontal (along wind/u) and perpendicular (cross-wind/v) components in order to avoid the problem described by the referee prior to applying the artificial neural networks to some of the data series in the revised manuscript.

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?).

Reply: Section 2.1 (Methane flux and meteorological data) does now also include a description of the u* filtering applied to the Lompolojänkkä and Barrow data and how the threshold has been chosen for the Lompolojänkkä data (for example).

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)

Reply: Section 2.1 (Methane flux and meteorological data) does now also include a short description of the reasoning behind filtering Kytalyk data according to atmospheric stability and why no classical u* filtering was carried out at this particular site.

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.

Reply: Additional information has been added in section 2.2 (Artificial Neural Networks) regarding the representativeness of training data. A training dataset should be representative of the overall dataset, meaning the whole range of meteorological and flux variability (including emission events so that the network can learn such conditions). Furthermore, we also draw attention to the cross-correlative (cross-dependency) nature of climatic variables that can occur when choosing input variables as some only add little extra information to the network. Only two datasets out of six were u* filtered so that the night time problem has not been touched in the current manuscript.

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

Reply: The applied neural networks have been now modified by including three different approaches: one including solar radiation as an indicator of time of day and three fuzzy sets representing the seasonal variation. A second approach has been carried out by removing solar radiation and replacing it with the four fuzzy sets representing time of day. A third approach is tested by incorporating the lagged effects of precipitation and water table depth.

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.

Reply: The sentence mentioned above in section (now) 2.5 (Statistical Analysis) has been rephrased for clarity. Regarding missing fluxes and simultaneously missing u* data, I (corresponding author) do disagree. A malfunctioning gas analyser (be it CO2, or CH4) or drizzle on an open path gas analyser (causing loss of data) does not cause gaps in sonic anemometer data. This means that missing flux data does not automatically imply missing u* data, hence this issue has not been addressed in the manuscript.

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.

Reply: The lagged effect of precipitation and water table depth on methane fluxes has been further extended in the current manuscript. These lagged effects have also been included as input variables in our third test, a more thermo-hydrological approach in order to test their applicability as an input variable. The performance of the relevant networks remained very high with correlation coefficients of 0.94 – 0.97 (when validating the network performance) for the two sites where WTD was applied. In order to include the lagged effect we adjusted the WTD and precipitation data according to confirmed lag times found in the methane flux literature. The method suggested by the referee appeared too speculative, hence the choice of lag times reported in the literature.

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.

Reply: This section appeared out of context and did not add any extra information to the neural network study and has been removed from the current manuscript.

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?

Reply: The performance of the mixed scenarios (represents the most realistic gap distribution), originally chosen by Moffat et al. (2007) as a crosscheck on the other 4 (very short to long gap length) gap scenarios and also included in the current study show the same, in some cases even better results than some of the individual scenarios themselves, resulting in very low mean root square error values, justifying their choice of scenario to be used in the current gap-filling study.

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

Reply: This sentence has now been corrected.

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.

Reply: Respective graphs have been replaced and annotations corrected appropriately.

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.

Reply: This graph has been removed and replaced with two graphs showing only the goodness of fit for the validation datasets and a description of which scenario has been illustrated.

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.

Reply: The applied neural networks have been now modified by including three different approaches. Two include the same input variable for all sites as a standardised method across all sites: one including solar radiation as an indicator of time of day and three fuzzy sets representing the seasonal variation and the second approach that has been carried out by removing solar radiation and replacing it with the four fuzzy sets representing time of day. A third approach in which different input variables were chosen at each site is tested by also incorporating the lagged effects of precipitation and water table depth.

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.

Reply: Several sentences have been added in section 4 (Discussion) and 5 (Challenges and recommendations) addressing further the issue of data representativeness, input variable choice, flux dynamics and lag effects of hydrological properties and recommendations for future studies.

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.

Reply: 90% of the data has been used for training of the artificial neural networks, as applied in Moffat et al. (2007). In order to avoid over-fitting, we applied the early stopping method (see section 2.4) implemented in the neuralnet package in R Statistical Language. The training process stops when all partial derivatives of the error function (summed square error) reach the pre-specified threshold value of 0.01 (1%). 10% of the data, representing the artificial gaps were used to validate the network performance.

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 allowing a shift of the artificial gap of +/- half of the length in order to maximize the number of data points removed.

Reply: The appropriate sentence in the text has been modified to read correctly as the original text appeared unclear and misleading. Moffat et al. (2007) presented 10 scenarios per gap length. In the current study the three scenarios chosen, were selected in such a way that the maximal existing data coverage was given. Nevertheless, three long gap scenario could not be applied to each site due to lack of enough data coverage or overlap between existing data and artificial gaps. Furthermore, none of the datasets included an entire 12-days (long gap) gap free period to act as a classical “long gap” test data set.

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.

Reply: These graphs have been removed in the revised manuscript and have been replaced with graphs showing only the output from the neural networks using 4 neurons only. This gave us the possibility to present the data with better clarity, summarising the correlation coefficients and the RMSE (in true physical units now) for a better overview and comparison.

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