Response to referee #1

23.1 2019

We thank the anonymous referee #1 for the comprehensive and sound comments which certainly helped to improve and clarify the paper. Our responses follow the order of the comments.

A revised version of the paper can be found in the supplement.

p1,3-7: we agree that data availability is a problem for the standard regression as well as the ANN method; however, the data sets we used were considerably larger than the ones used previously and were checked for consistency with MOST. The ANN method seems more flexible than standard multivariate regression because no assumptions about the functional form of the relationship are made. Having explicit formulas instead of ANNs would only be an advantage if there were some physics behind the formulas, which is not (yet) the case.

p2, 1: we agree

p2, 7: rephrased more explicitely

p2, 10: corrected

p2, 11-12: text changed

p2, 14-15: text adapted. We do not suggest anything, we just quote the paper. Overall best model is not mentioned in the paper.

p2, 16-17: text changed

p2, 22.25: text changed

p2, 8-25: text extended

p3,8: we do not agree here: the quantities mentioned are not stability parameters. We used potential temperature, therefore no index v.

p3, 15: done

p4, 1-7: moved to introduction and rephrased

p4, 12-13: rephrased

p4, 14: done

p4, 15: done

p4, 20: yes; rephrased

p4, 25: we used the superscript Omega to make clear that N refers to the output ("Omega") layer

p5, 1-9: done

Table 1: we rephrased and extended the text

p5, 18-20: we rephrased the description. DE-Keh was left out in training and validation in the second experiment

Table 2: done

Table 2 and A1: done

Figure 1: replaced

p5, 21: done

p5, 27-28: done

Sec 2.3&2.5: Moved "Cross-validation" after "Data"

Ch 2: batch is the whole training set

p7, 16: rephrased in sec 2.5.

Fig 3: done

Fig 4b: done

p7, 27: done

p7, 28: which statement do you mean?

p8, 1: rephrased

p8, 3: done

p8, 7: done

p8, 8: done

Table 3: table rearranged

p8, 9-23: this is not generally true. We added a sentence "Networks with seven inputs have in our case no substantial advantage over networks with six inputs."

p8, 25-26: we don't get what you mean here

Table 4/5: done

p8, 28-29: done

p9, 2-3: rephrased

p9, 7: we do not agree here because: a) the training data set we used was quite large, and b) using even more and even better training data would probably also improve the results of the simpler nets, so cost/benefit might not change.

Sec 3.2: brief summary added

p9, 13-15: rephrased

p9, 19: input was every 30 min, corrected

p9, 23: data were new in the sense that time periods were used which had not been used for training and validation; the DE-Tha site had not been used at all before, because a) the sites selected in sec 2.4 were more consistent with MO than DE-Tha and b) the DE-Tha time series covered only one year. For running the LSM for the DE-Tha site, a more comprehensive input data set (including e.g. radiation, precipitation) was required, which was only available for the year 1998. The years for the two sites were mixed up in the paper: it should be 2011 for the DE-Fal site and 1998 for the DE-Tha site. (not 2011, as the paper says erroneously).

Fig 7: done

p10, 10: replaced "training method" with "data sampling method"

p10, 14-15: done

p10, 26-29:

grammar/typos:

overall: we use nonlinear

p1, 3: done

p1, 18: done

p1, 19 : done

p2, 2: done

p2, 5-25: changed to present tense consistently

p2, 20: sentence rephrased

p2, 28: we prefer to leave it as it is

p4, 16: done

p4, eqn#: done

p5, 17: done

p6, 2: done

p6, 9: done

Figs 3 and 4: done

Fig 4, caption: done

p7, 21: done

p7, 23: done

p7, 25: done

p8, 4: done

Table 3: done

Sec 3.2: changed to active form

p8, 30: done

p9, 16-18: done

Sec 4: changed to active form

Updated Reply to referee #2:

Our reply is structured like this: we quote the essential part of the referee's comment in inverted commas, followed by our reply. Changes in the manuscript are highlighted there.

"... BC is entirely empirically specified ...":

the BC (or rather u* and T*) is derived on the basis of MOST, and as we state in sec 2.1, on this basis our goal is to determine u* and T* from known quantities, which are in our case modelled or observed wind and temperature gradients in the surface layer. So it's not "entirely empirical".

"Little insight into the data ...":

data have been described and checked carefully for compatibility with MOST (sec. 2.3). Underlying physics is MOST, described in sec. 2.1. Performance of the algorithms is discussed in secs. 3 and 4.

"... and the authors' own results suggest a much simpler model would fit their data equally well. A path they do not investigate.":

We have added an analysis using the same data, but with a multivariate linear regression model (MLR).

"... is that of more-or-less uncritically applying ANNs to the data-set, without examining their suitability, and to what extent the data can be explained by simpler models":

we gave the reasons why we tried ANNs: see remarks above and secs. 1 and 2.1 of the paper. Simpler model: see MLR above.

"In fact you show a 1-layer, 1-neuron "network" performs basically the same as 1-layer, 12-neurons, or a deep network with 2-layers with 7-7 neurons (7 inputs).":

this is not the case. Especially in fig. 3 one can see a substantial trend that a network with one single hidden neuron is outperformed by networks with several hidden neurons. Also fig. 4 shows this trend in an attenuated pattern.

"This result strongly suggests that almost all the predictive power of ANN for this data is contained in a linear fit.":

As the results of the MLR show, a linear fit does not work well.

"Given this, it seems redundant and unnecessarily complicated to use the heavymachinery of ANNs, with its associated costs":

we don't think ANNs can be considered heavy machinery nowadays; the difficult parts of the work are a) obtaining and filtering data, and b) validation and testing - this has to be done for all kinds of regression methods.

"... an extremely simple main relationship between features and output:"

the task is indeed simple: approximate a single valued function of six variables. The essential physics is captured in the Monin-Obukhov length and the dimensionless gradients (i.e. stability functions).

"Indeed the difference between these figures indicates that the more complex networks are overfitting the data from the available towers.":

we discuss our use of the (less complex) 6-3-2 ANN in secs. 3 and 4.

"... comparison with simpler models ...":

See MLR above. Also, a comparison is done in secs. 3 and 4 with the regression (not physics!) based functions in the literature.

"Only if ANNs do significantly better than linear models is the current work worth publishing.": why? In our case, the ANNs do markedly better than MLR.

"I'm not convinced by the assertion that there is a significant computational speed advantage to be gained by replacing MOST with an ANN ...":

we are not sure here either - but we would like to try. This was the first step - next step will be implementation in a regional climate model (RCM).

"3d LES simulation": we do not intend to do LES simulations.

"Please explain what is special about your models that causes this situation to be reversed. Please quantify the time spend by your code in various parts of the calculation":

The situation is not reversed. Climate models, especially RCMs, are very expensive to run (climatologically relevant multidecadal simulations at high resolution can take several tens of weeks on a high performance system), so every saving is valuable, especially in view of the other advantages. We hope to save around five percent (i.e. about one week), taking into account parallelisation. Text has been rewritten

"I would appreciate in Section 2 an enumeration of all assumptions made, perhaps with some comment on their validity and their role in simplifying the MOST model.":

We have reformulated the text in section 2.1

"Are these fluxes measured directly": at all sites used, fluxes are measured by the eddy covariance method, from which u*and T* are derived with the formulas from sec. 2. We have added this in the data description.

"What modelling assumptions are inherent to your ANN approach?": this is explained in sec. 2.1; we have reformulated the description.

"Title should be "turbulence fluxes" not "turbulent fluxes".":

we would like to stick to the terminology used in the boundary layer meteorology community, which is "turbulent fluxes" (see e.g. Arya's book).