The authors thank reviewer #1 for useful and constructive feedback, and have thoroughly edited the manuscript accordingly. Our response to specific suggestions are listed below following each comment and line reference.

In this study, the use of Artificial Neural Networks for the association of the data provided by low cost agrometeorological stations with crop Evapotranspiration obtained by eddy covariance with main target the estimation of crop ET based on low cost agrometeorological stations is investigated. The main idea is that eddy covariance stations could be located near low cost agrometeorological stations for short periods in order to train an ANN that could provide crop ET from the low-cost station for the rest of the period. The idea is very interesting, even if it does not seem to be practical to relocate the eddy covariance stations every little while and to train ANNs for operational purposes. Though, the scientific questions are really interesting and the expected results could provide useful information and be the basis for further study. The applied methodology seems also to be sound and the obtained results very interesting.

Regarding the practicality of relocating the eddy covariance system, we have addressed more thoroughly in the text the reasons why doing so would be feasible and even cost effective for applied management. In summary, methods such as eddy covariance (and lysimetry) are currently impractical for applied on-farm water management precisely because of their cost. However, the equipment required for eddy covariance is readily installed and re-located, given adequate technical expertise is available. Dedicating such equipment to one long term monitoring site is the typical use for such equipment (such as in the Ameriflux network), but what we proposed here was to evaluate if sharing the cost of equipment and labor among many sites would make the data from an eddy-covariance system more accessible and useful to more end-users. In this way, the additional effort and cost from relocating an eddy-covariance system would be distributed among those users. We haven't explicitly called out any additional cost because in our experience, the effort to set up and relocate such a temporary system is much less than the fixed costs for the system (equipment, employment of the trained personnel, negotiating access to sites, etc.)

On the other hand, the main weaknesses of this study are related to the presentation of the methodology and of the results. I found it really hard to follow the methodology and to distinguish what is tested in each case, what is compared with what, and what were the final results and the final conclusions of each part of the results. While the objectives of the study seem to be clear, I had to read the entire manuscript (methodology, conclusions, results and discussion) to more or less understand how these objectives were met. There were also some additional objectives presented in the methodology or the results sections that makes things a little more confusing.

My opinion I that the manuscript should be carefully revised in order to present more clearly the objectives, the methodology used to achieve them, the obtained results and the final conclusions.

I acknowledge the confusing lack of organization in the first submitted manuscript, and we have thoroughly edited to address clarity of study objectives, disambiguating the methodology, results, and conclusions, and have fixed some unclear terminology to help clarify the presentation.

As regards the objectives, some more attention should be given at the last part of the introduction (lines 115-124) where objectives are mixed with results. You may avoid presenting the results in the introduction and try to clearly present the main objectives and any other additional objectives.

This part of the introduction has been re-written according to the reviewer's suggestions. We clarified all of the objectives, and moved results to the correct section of the manuscript.

Concerning the methodology, the most problematic part seems to be the section 2.2. The first part (lines 221-224) is very confusing and it is not easy to understand what is calculated with what data and for what purpose. Then there is a presentation of the ANN and then again, a presentation of what is calculated.

We have revised the methodology and added some organizational detail to clarify what we have done. The specific section at lines 221-224 has been revised.

One possible way to make this easier to understand would be to have a separate section presenting the ANN (e.g. 2.2.). The following section (e.g. 2.3.) could present in detail and clearly what is examined and which is the
corresponding methodology for this. As I can understand the main parameters examined are: a) the best set of sensors for low cost stations to estimate ETc and b) the optimum training period and the optimum strategy. However, I also saw some additional objectives in the results (e.g. effect of Irregular Irrigation on Actual ET and some other), if they are important please explain them as well and include them in the general context. It should be clearly presented what is calculated and with what data, the periods that were used for training and for validation, specifically what is compared with what, and the general strategy with specific steps. A graph / table / diagram explaining the methodology could possibly help.

Because we did not develop a new programmatic neural network approach, we relied on citations from the literature to explain the actual use of the ANN training and verification. We did not add an additional section as suggested, but more clearly delineated the field study and analysis methods, primarily by explaining in more detail the study objectives (optimizing the sensor array and minimizing training time), and connecting these objectives to the main goal of the project (to evaluate the feasibility of ANNs to make eddy covariance and low cost sensors more directly useful in irrigation management).

The result section is also confusing similarly to the methodology. It is organized in sections without specific reasoning. For example, the sections titles are: “Optimization of ANN Training”, then “Effect of Irregular Irrigation on Actual ET at Site 1”, “Monitoring Cumulative Crop Water Use at Site 2” and finally “Actual ET from Fully Irrigated Alfalfa”. The sections of the results should be inline with the main scientific questions and / or the main methodological steps. In the current organization the main questions are mixed together and partly answered here and there. The results are also fragmentally presented and it is very difficult to understand the entire picture. A better organization of the results section is needed (which is actually Results and Discussion).

Here we have matched the results and the sub-sections more clearly to the objectives which have been more clearly defined.

A table (or tables) summarizing the results (e.g. performance) for each site, each set of sensors, each training duration or any other categorization appropriate could make much easier for the readers to understand what is going on.

We have added this table.

The Figures are very good and very informative. Some minor improvements are needed. E.g. in Figures 3, 4, and 8 the slopes and intercepts of the fitted lines should be also presented. It can be indicated, but you should also show the x=y line. Please also clarify if R2 and RMSE are related to the line fit or to the comparison between observed-estimated values. Units should be added in Figs 7 and 12 and please check the legend of Fig. 10.

We have updated the captions, axes labels, and legends of several figures to improve clarity of what is presented. Slopes and intercepts were not added to figures 3 and 4, primarily because the line slope (analogous to a crop coefficient) would mis-represent what was measured. I.e., the large scatter of our results (inclusive of periods with inhomogeneous and non-stationary conditions) means that the linear regression is not a reliable indicator of a uniform coefficient. Instead, these plots attempt to characterize the robustness of the modelled 30 minute flux, and evaluate a change in this robustness from a non-adaptive application of the Penman-Monteith equation. Figure 10 was removed (see our general comments on the revision). Figure 12 has also been removed, but the vertical axis in Figure 7 does not have units as what is shown is a ratio (actual ET divided by reference ET).

Finally, in the last section which is mostly conclusions, the key general results of the study should be also clearly and specifically presented along with their importance for practical applications and possible future research. For example, any general results / guidelines concerning the required set of sensors for the “low-cost” stations or the possible strategies that could be followed for the practical application of this research.

We have added more discussion of the key general results, along with our expectations of the practical applications and guidelines of the required sensors and methods. We have suggested a few areas of further study to pursue.
Here I would like to mention that it would be interesting to consider the possible cost and the practicality (purchase, installation, maintenance etc.) of the various sets of sensors evaluated. Actually, the setup of the “low-cost” station (lines 156-161) is hardly that of a typical agro-meteorological and a low-cost station. According to the concept it would be interesting to test if a typical agro-meteorological station can provide adequate ETa results with the help of the trained ANN.

Here in particular, we would like to point out that our set-up was never intended to demonstrate a recommended suite of low-cost sensors, but rather be inclusive of as many possible such sensors so that we could optimize the relevant set. What we end up recommending is much more limited in cost and complexity. Future studies underway intend to evaluate the viability of the method using actual existing agro-meteorology stations, which potentially would be useful for backcasting data and comparing to decision support tools such as remote sensing.

Another general comment is that the suggested strategy (lines 482-487) seems not to be very practical (many times relocation of equipment, very high skills required for dealing constantly with equipment, eddy, ANN etc.). This should be discussed.

A final general comment is that the study didn't resulted in a specific low-cost configuration for all purposes as well as a specific training procedure. These should be clarified in order for this study to be useful for practical applications. Though, I would like to state that the obtained results are still very interesting and provide a good base for further study.

Hopefully our responses above have satisfied these concerns.

Two short specific comments: a) Figure 9. Is the ANN performance better than the simple correlation e.g. a single ratio? b) Lines 346-349, soil moisture spatial variability could be also a problem (see for example Soulis K.X. and Elmaloglou S., 2016, Optimum soil water content sensors placement in drip irrigation scheduling systems: the concept of Time Stable Representative Positions, Journal of Irrigation and Drainage Engineering, ASCE, 10.1061/(ASCE)IR.1943-4774.0001093, 04016054. or Soulis K.X. and Elmaloglou S., 2018. Optimum soil water content sensors placement for surface drip irrigation scheduling in layered soils. Computers and Electronics in Agriculture, 152, 1-8. doi:10.1016/j.compag.2018.06.052)

a) In regards to a simple ratio, please see our comment above regarding the regression line. In general, I believe the scatter of the resulting flux estimates (taken naively across all field conditions), compared with many complex methods presented in the ASCE Manual, demonstrate that there are opportunities for this type of machine learning to identify patterns of crop response that are not easily described by crop coefficients in applied irrigation management.

b) spatial variability is absolutely a confounding factor in using it. We did discuss the temporal decoupling of time series, as we used time series data (and treated records as independent samples). Future work with machine learning might potentially be used to evaluate spatial processes as well, but the instrumentation we were afforded for this study did not allow us to conduct a meaningful analysis of spatial patterns.

Based on the above concerns and comments I believe that the work presented in this study is very good, very laborious, scientifically sound, and generally very interesting. Accordingly, it worth to be published; however, a lot of work (a thorough revision) is still needed in order to make it easier to read and to understand.

We appreciate this thorough review, and feel that our manuscript has been very much improved as a result of the constructive comments.