**Answer to reviewers for Paper 395175**

**“Dual oxygen and temperature luminescence learning sensor with parallel inference”**

**Please provide a detailed response to the review report(s), if applicable. Make sure that all points are addressed in a clear and concise manner. You may enter the response in the text box below or upload a file.**

**Reply to Reviewer 1**

The first reviewer poses the question about the difference between the manuscript and a previously published work and comments that the parameters for the neural network are previously described and “This largely undermined the contributions of this manuscript.”

The criticism is not founded. Our previous work of reference 26 is a purely theoretical study on multi-dimensional regression problems to demonstrate that they can be solved using multi-tasking neural network architectures and which architectures are most appropriate for this type of problems. As example synthetic data were generated using the classical Stern-Vollmer equation which is a well-known and best accepted analytical model for this type of sensors.

The manuscript submitted to Optica simply uses the architecture, which was previously found to work best. It is typical of scientific work to start from previous findings, without redoing the same work again. Similarly, for the excitation of the luminophore we use the wavelength which is known to work best with this type of luminophore.

However, the work which we have submitted is not which architecture work best for a theoretical type of problem. We built a real physical optical sensor, we characterized and used for the acquisition of data needed for the training (that was taken experimentally with a new setup also described in the paper) and we demonstrated a real-sensor performance which would not be possible with a standard approach. Secondly, the paper submitted here for publication does much more. It defines how the data acquisition should be carried out and how the real data should be gathered to be used efficiently with neural networks, which is essential to a sensor based on machine learning.

Furthermore, it introduces a new metric, the Error Limited Accuracy, which bridges the gap between two fields: physics and computer sciences. We strongly believe this metric is necessary for the characterization and, thus, the spreading of a new generation of sensors into applications.

The experimental result per se (Figs. 4 to 6) are new, unpublished and, to the best of the authors’ knowledge, of unprecedented detail and quality as compared to any previous publications of this type of sensor since span thoroughly in a four-dimensional space (oxygen concentration, temperature, modulation frequency and phase shift).

In conclusion, this work demonstrates that a new generation of sensor which are based on the proposed approach is not only possible but would work better than the conventional ones. This work thus represents a paradigm shift. This is the reason why we strongly think the work is of great relevance for the public of Optica and for the optical community in general and should be published on Optica.

.

**Reply to reviewer 2**

The authors thank the reviewer for the positive feedback and constructive comments. If accepted, we will revise the paper to take into accounts all the very good points.

Please find below the reply to each point.

**1. In section C.1, the authors described two types of input that were used to train the network. Please explain why the phase shifts of the first input (theta\_s) are divided by 90.**

The reason is that when working with neural networks, the initialisation of the input observations is a key factor in getting good results. The authors have found, during testing, that such a normalisation delivers the best results and avoid problems like divergence of the neural network model.

**2. Is there any particular reason why the mini-batch of 32 samples was used?**

The authors have tested many mini-batch sizes. We found that 32 was a good balance between performance (MSE values) results and duration of training. Note that generally speaking a neural network model with small mini-batches will need less epochs to reach a certain performance but will take quite a long time for each epoch. On the opposite side of the spectrum a neural network model with a big mini-batch will take many more epochs to reach the same performance and will take less time for each epoch.

**3. Are the results in the Sensor Performance Section from the training data set or the testing data set (new data that never been seen by the model)? If it is from the testing data set, could the authors describe its data collection?**

The results in the mentioned section are, as the reviewer correctly suggested, from the test dataset (a dataset that the model have never seen). The original dataset have been split in the classical 80%/20% split (using stratified sampling), and the 20% have been used to test for overfitting and to get results on an unseen dataset.

**4. To generate the model, it requires 65 hours of training data collection and an additional 5 hours for training the model. Could the authors discuss whether this process should be repeated every time a new sensor is manufactured? Or is there any way to translate the generated model into a new sensor?**

This is a very appropriate and interesting questions. In machine learning it is possible to use transfer learning to adjust an almost completely trained network with a much-reduced amount of data. The authors are working in this direction and will investigate the extent of new data needed in the future. The first tests are very positive and indicates that such a transfer is indeed possible and very efficient.