

# Dual oxygen and temperature sensing with automatic data acquisition for machine learning

FRANCESCA VENTURINI<sup>1,2,\*</sup>, UMBERTO MICHELUCCI<sup>2</sup>, AND MICHAEL BAUMGARTNER<sup>1</sup>

<sup>1</sup>Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401 Winterthur, Switzerland

<sup>2</sup>TOELT LLC; Birchlenstr. 25, 8600 Dübendorf, Switzerland

\*Corresponding author: francesca.venturini@zhaw.ch

Compiled February 4, 2020

A well-known optical approach to measure oxygen partial pressure is the quenching of luminescence by the oxygen molecules. Sensors based on this principle typically rely on approximate empirical models to parametrise the dependence of the sensing quantity on influencing factors, like the temperature. In this work, we propose a new multi-task learning (MTL) neural network approach which allows the extraction of both the oxygen concentration and the temperature using one single indicator and measuring a single quantity, namely the decay time. The results demonstrate that firstly, using neural networks it is possible to extract both the oxygen concentration and the temperature from the measurement of one single quantity and using one single indicator; secondly, that the use of the proposed MTL networks allow more accurate and stable predictions for both the parameters. Furthermore, the proposed MTL approach is not limited to luminescence quenching but paves the way for applications where the mathematical model is not known, too complex or not really of interest.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

<http://dx.doi.org/10.1364/optica.XX.XXXXXX>

## 1. INTRODUCTION

The determination of oxygen partial pressure is of great interest in numerous areas, like medicine, biotechnology, and chemistry. Among the different methods used to determine oxygen concentration, optical techniques are particularly attractive because they do not consume oxygen, have a fast response time, allow a good precision and accuracy. Further advantages of optical sensors include that they can be used for both in-situ measurement and for remote acquisition, via optical fibers. The most used optical measuring approach uses the effect of the quenching of luminescence by the oxygen molecules [1]. The measuring principle is based on the measurement of the luminescence of a specific molecule or luminophore, whose intensity and decay time are reduced due to collisions with molecular oxygen [2].

Sensors based on this principle typically rely on approximate empirical models to parametrise the dependence of the sensing quantity on influencing factors. A new approach is to use feed-forward neural networks to predict the desired variables. Unfortunately, those kind of networks usually perform

less efficiently when applied to multi-dimensional regression problems. In this work, we propose a new multi-task learning (MTL) neural network approach which allows the extraction of both the oxygen concentration and the temperature using one single indicator and measuring a single quantity, namely the decay time. The results demonstrate that firstly, using neural networks it is in principle possible to extract both the oxygen concentration and the temperature from the measurement of one single quantity and using one single indicator; secondly, that the use of the proposed MTL networks allow more accurate and stable predictions for both the parameters. Furthermore, the proposed MTL approach is not limited to luminescence quenching but may be of particular relevance in all those cases, where the mathematical model is not known, too complex or not really of interest and the only goal of the regression problem is to build a system that is able to determine as accurately as possible..



nm (Semrock 594 LP Edge Basic long pass) and a short pass filter with cut-off at 682 nm (Semrock 682 SP Bright Line HC short pass). The driver for the LED and the trans-impedance amplifier (TIA) are self-made. For the frequency generation and the phase detection a two-phase lock-in amplifier (SR830, Stanford Research Inc.) was used. The modulation frequency was varied between 200 Hz and 15000 kHz.

### C. Automated Data Acquisition

The series of measurements were carried out following the flow shown in Fig. 2. First the acquisition program fixed the temperature and concentration. Then the phase-shift was measured for varying the modulation frequency. This measurement was repeated 20 times. Next, keeping the temperature fixed, the program changed the oxygen concentration and the entire frequency-loop was repeated. Finally, the temperature is changed and then the oxygen and frequency loops were repeated. The total number of measurements is  $50 \times 20 \times 21 \times 9 = 189000$  which lasted ca. 65 hours. The number of measurement was chosen as a compromise between maximal number of measurements and minimal change of the sample due to photobleaching.

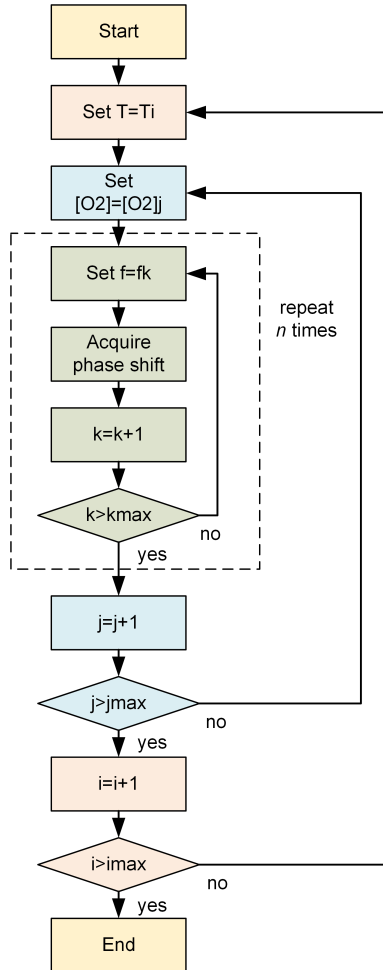


Fig. 2. Flow-chart of the automatic measurement.

### D. Artificial Neural Network Design

The artificial network used in this work has a multi-task-learning architecture and is depicted in Figure 3. It consists of three

common hidden layers with 50 neurons each which generate as output a "shared representation". The name comes from the fact that the output of those layers is used to evaluate both  $[O_2]$  and  $T$ . These are followed by three branches, two with each two additional "task-specific hidden layers" to predict respectively  $[O_2]$  and  $T$ , and then one without additional layers to predict  $[O_2]$  and  $T$  at the same time. The shared representation is the input of two "task-specific hidden layers", that learn how to predict  $[O_2]$  and  $T$  better. This architecture uses the common hidden layers to find common features beneficial to each of the two tasks. During the training phase, learning to predict  $[O_2]$  will influence the common hidden layers and therefore, the prediction of  $T$ , and vice-versa. The further task-specific hidden layers learn specific features to each output and therefore improve the prediction accuracy. The number of neurons of each task-specific hidden layer is 5.

As activation function the sigmoid activation functions was used for all the neurons. A common choice for the cost function in regression problems is the mean square error (MSE), which is defined as

$$MSE = \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^d (y_k^{[j]} - \hat{y}_k^{[j]})^2 \quad (4)$$

where  $n$  is the number of observations in the input dataset;  $y^{[j]} \in \mathbb{R}^d$  is the measured value of the desired quantity for the  $j^{th}$  observation (indicated as a superscript between square brackets), with  $j = 1, \dots, n$ ;  $\hat{y}^{[j]} \in \mathbb{R}^d$  is the output of the network, when evaluated on the  $j^{th}$  observation. Since there are multiple branches, a global cost function  $L$  needs to be defined as a linear combination of the task-specific cost functions with weights  $\alpha_i$  will be minimized

$$L = \sum_{i=1}^{n_T} \alpha_i L_i. \quad (5)$$

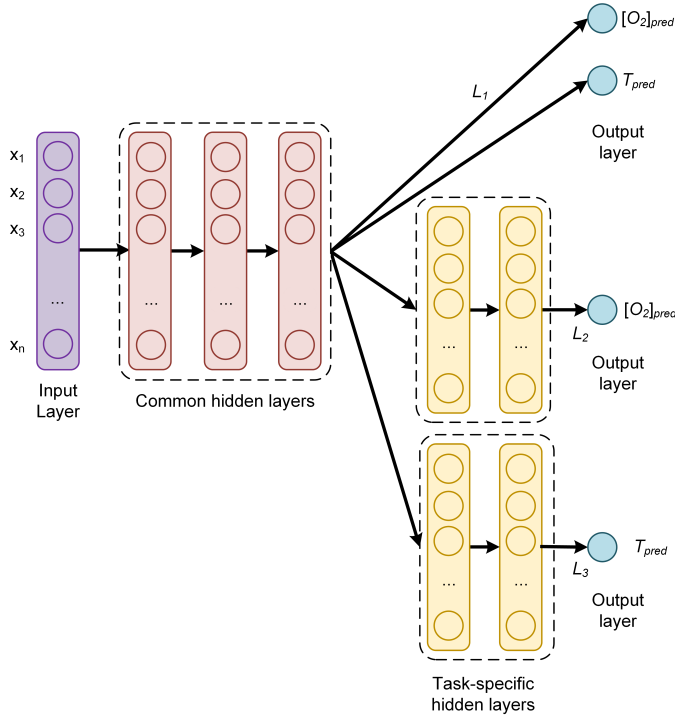
The parameters  $\alpha_i$  have to be determined during the hyper-parameter tuning phase to optimize the network predictions. In this paper, being the cost function the MSE (Equation 4), the global cost function of Equation 5 is

$$L = \sum_{i=1}^{n_T} \alpha_i \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^d (y_k^{[j]} - \hat{y}_k^{[j]})^2 \quad (6)$$

where  $n_T$  is the number of tasks;  $n$  is the number of observations in the input dataset;  $y^{[j]} \in \mathbb{R}^d$  is the measured value of the desired quantity for observation  $j$ , with  $j = 1, \dots, n$ ;  $\hat{y}^{[j]} \in \mathbb{R}^d$  is the output of the network, when evaluated on the  $j^{th}$  observation. The global cost function weights used for the plots were  $\alpha_1 = 0.3$ ,  $\alpha_2 = 5$  and  $\alpha_3 = 1$ . Those values were chosen because they result in the lowest MAEs (see discussion in Section 3).

To minimize the cost function, the optimizer Adaptive Moment Estimation (Adam) [9, 10] was used. The training was performed with a starting learning rate of  $10^{-3}$  and using batch-learning, which means that the weights were updated only after the entire training dataset has been fed to the network. Batch-learning was chosen because of its stability and speed since it reduces the training time of a few orders of magnitude in comparison to, for example, stochastic gradient descent [10]. Therefore it makes experimenting with different networks a feasible endeavor. The implementation was performed using the TensorFlow library.

The metric used to compare results from different network models is the absolute error (AE) defined as the absolute value of the difference between the predicted and the expected value



**Fig. 3.** Architecture of the multi-task learning architecture.

for a given observation. For the oxygen concentration of the  $j^{th}$  observation  $[O_2]^{[j]}$  the AE is

$$AE_{[O_2]}^{[j]} = |[O_2]_{pred}^{[j]} - [O_2]_{meas}^{[j]}|. \quad (7)$$

The further quantity used to analyze the performance of the network is the mean absolute error (MAE), defined as the average of the absolute value of the difference between the predicted and the expected oxygen concentration or temperature. For example, for the oxygen prediction using the training dataset  $S_{train}$ ,  $MAE_{[O_2]}$  is defined as

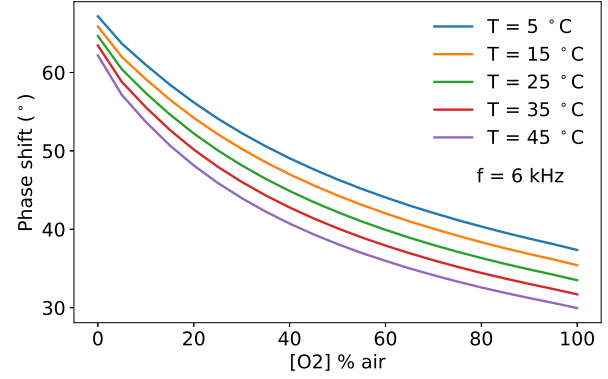
$$MAE_{[O_2]}(S_{train}) = \frac{1}{|S_{train}|} \sum_{j \in S_{train}} |[O_2]_{pred}^{[j]} - [O_2]_{real}^{[j]}| \quad (8)$$

where  $|S_{train}|$  is the size (or cardinality) of the training dataset. For example, in this work  $|S_{train}|=20000$ . The  $AE_T$  and  $MAE_T$  are similarly defined.

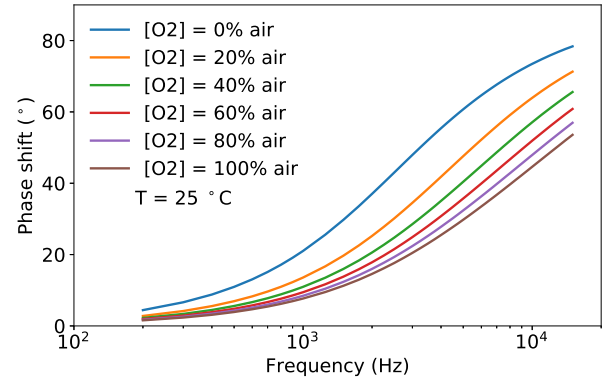
### 3. RESULTS AND DISCUSSION

As described in Section A, the phase-shift depends not only on the oxygen concentration, but also on the temperature and on the modulation frequency of the excitation light. This is shown in Fig. 5 and Fig. 6.

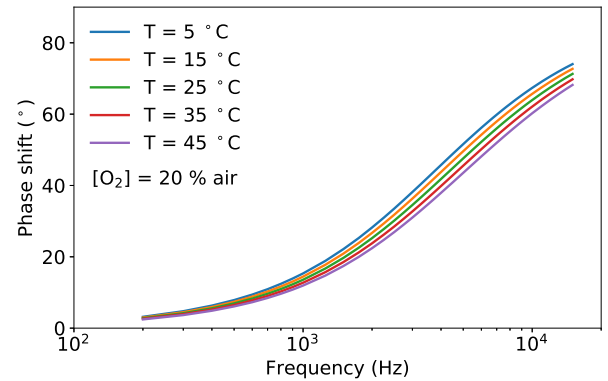
Fig. 5 shows the phase-shift of the luminescence at a fixed temperature for different oxygen concentrations as a function of the modulation frequency. Fig. 6 gives the complementary information of the phase-shift at a fixed oxygen concentrations for different temperatures as a function of the modulation frequency. Comparing the curves of the two picture it is clear why it is not possible, when measuring one single phase-shift or even a curve of the phase-shift dependent from the modulation frequency to simultaneously extract both the oxygen concentration and the temperature.



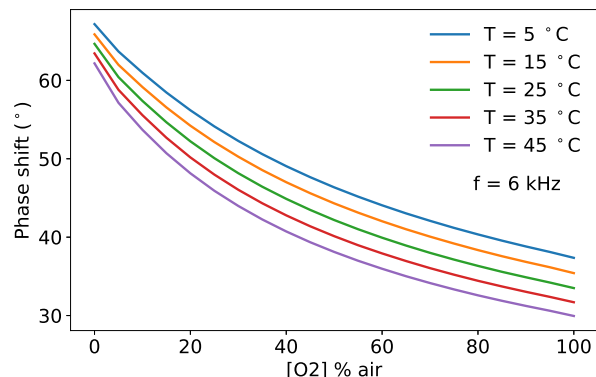
**Fig. 4.** Phase-shift as a function of the oxygen concentration, for selected temperatures at a fixed modulation frequency of 6kHz.



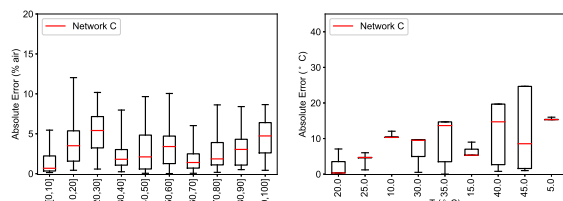
**Fig. 5.** Phase-shift for selected oxygen concentrations as a function of the modulation frequency at a fixed temperature  $T = 25^\circ$ .



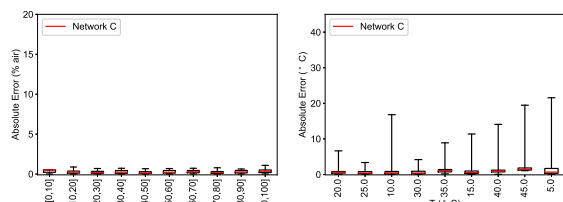
**Fig. 6.** Phase-shift for selected oxygen concentrations as a function of the modulation frequency at a fixed oxygen concentration  $[O_2] = 20\%$  at selected temperatures.



**Fig. 7.** Phase-shift at a fixed oxygen concentration  $[O_2] = 20\%$  air for selected temperatures as a function of the modulation frequency of the excitation light.



**Fig. 8.** Phase-shift at a fixed temperature  $T = 25^\circ$  for selected oxygen concentrations as a function of the modulation frequency of the excitation light.



**Fig. 9.** Phase-shift at a fixed temperature  $T = 25^\circ$  for selected oxygen concentrations as a function of the modulation frequency of the excitation light.

## 4. CONCLUSIONS

## DISCLOSURES

**Disclosures.** The authors declare no conflicts of interest.

## REFERENCES

1. X.-d. Wang and O. S. Wolfbeis, "Optical methods for sensing and imaging oxygen: materials, spectroscopies and applications," *Chem. Soc. Rev.* **43**, 3666–3761 (2014).
2. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd ed. (Springer, New York, 2006).
3. E. Carraway, J. Demas, B. DeGraff, and J. Bacon, "Photo-physics and photochemistry of oxygen sensors based on luminescent transition-metal complexes," *Anal. chemistry* **63**, 337–342 (1991).
4. J. N. Demas, B. DeGraff, and W. Xu, "Modeling of luminescence quenching-based sensors: comparison of multisite and nonlinear gas solubility models," *Anal. Chem.* **67**, 1377–1380 (1995).
5. P. Hartmann, M. J. Leiner, and M. E. Lippitsch, "Luminescence quenching behavior of an oxygen sensor based on a ru (ii) complex dissolved in polystyrene," *Anal. Chem.* **67**, 88–93 (1995).
6. A. Mills, "Response characteristics of optical sensors for oxygen: models based on a distribution in  $\tau_{\text{oor}}$  kq," *Analyst* **124**, 1301–1307 (1999).
7. M. Quaranta, S. M. Borisov, and I. Klimant, "Indicators for optical oxygen sensors," *Bioanal. reviews* **4**, 115–157 (2012).
8. U. Michelucci, M. Baumgartner, and F. Venturini, "Optical oxygen sensing with artificial intelligence," *Sensors* **19**, 777 (2019).
9. J. A. Kingma, D.P.; Ba, "Adam: A method for stochastic optimization. in proceedings of 3rd." In Proc. 3rd Int. Conf. on Learn. Represent. ICLR 2015 pp. 1–15 (2015).
10. U. Michelucci, *Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks* (APRESS Media, LLC, 2018).

~