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

Francesca Venturini[1,2,*], Umberto Michelucci[2], and Michael Baumgartner[1]

[1] *Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401 Winterthur, Switzerland*
[2] *TOELT LLC; Birchlenstr. 25, 8600 Dübendorf, Switzerland*
[*] *Corresponding author: francesca.venturini@zhaw.ch*

Compiled February 5, 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 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.

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 measuremnt 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 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..

2. METHODS

A. Luminescence Quenching for Oxygen Determination

Luminescence-based oxygen sensors usually consist of a dye molecule (luminophore) whose luminescence intensity and decay time decrease depending on the O_2 concentration. This reduction is due to collisions of the excited luminophore with molecular oxygen, which thus provide a radiationless deactivation process (collisional quenching). In the case of homogeneous media characterized by an intensity decay which is a single exponential, the decrease in intensity and lifetime are both described by the Stern-Volmer (SV) equation [2]

$$\frac{I_0}{I} = \frac{\tau_0}{\tau} = 1 + K_{SV} \cdot [O_2] \tag{1}$$

where I_0 and I, respectively, are the luminescence intensities in the absence and presence of oxygen, τ_0 and τ the decay times in the absence and presence of oxygen, K_{SV} the Stern–Volmer constant and $[O_2]$ indicates the oxygen concentration.

For practical applications, the luminophore needs to be embedded in a supporting substrate, frequently a polymer. As a result, the SV curve deviates from the linear behavior of equation (1). This deviation can be due, for example, to heterogeneities of the micro-environment of the luminescent indicator, or the presence of static quenching [1]. A scenario which describes this non-linear behavior involves the presence in the substrate of at two or more environments, in which the indicator is quenched at different rates [3, 4]. This multi-site model describes the SV curve as the sum of two or more

$$\frac{I_0}{I} = \left[\frac{f_i}{1 + K_{SV,i} \cdot [O_2]}\right]^{-1} \tag{2}$$

where I_0 and I, respectively, are the luminescence intensities in the absence and presence of oxygen, f_i's are the fractions of the total emission for each component under unquenched conditions, K_{SVi}'s are the associated effective Stern–Volmer constants, and $[O_2]$ indicates the oxygen concentration. A simplification of this model is that one of the sites is not quenched and therefore the constant K_{SV2} is zero. Depending on the luminophore and on the substrate material, the proposed model may be even more complex [4–6].

Since in most industrial and commercial sensor applications, the decay time τ is frequently preferred to intensity measurement because of its higher reliability and robustness. The determination of the decay time is done most easily in the frequency domain by modulating the intensity of the excitation. As a result, the emitted luminescence is also modulated but shows a phase shift θ due to the finite lifetime of the excited state. This method, has the advantage of allowing very simple and low-cost realization and is widely used in commercial applications.

Although the multi-site model was introduced for luminescence intensities, it is frequently also used to describe the oxygen dependence of the decay times [4, 7]. Therefore, in the simplest case of a two-sites scenario, the model can be rewritten in terms of phase shift as [8]

$$\frac{\tan\theta(\omega,T,[O_2])}{\tan\theta_0(\omega,T)} = \left(\frac{f(\omega,T)}{1 + K_{SV1}(\omega,T)\cdot[O_2]} + \frac{1-f(\omega,T)}{1 + K_{SV2}(\omega,T)\cdot[O_2]}\right)^{-1} \tag{3}$$

where θ_0 and θ, respectively, are the phase shifts in the absence and presence of oxygen, f and $1-f$ are the fractions of the total emission for each component under unquenched conditions, K_{SV1} and K_{SV2} are the associated Stern–Volmer constants for each component, and $[O_2]$ indicates the oxygen concentration. It is to be noted that the quantities θ_0, f, K_{SV1}, and K_{SV2} are all non-linearly temperature dependent and may result frequency dependent, an artifact of the approximation of the model. Finally, Eq. [**?**] needs to be inverted to determine $[O_2]$ from the measured quantity θ.

B. Experimental Setup

The sample used for the characterization and test is a commercially available Pt-TFPP-based oxygen sensor spot (PSt3, PreSens Precision Sensing GmbH). To control the temperature of the samples, these were placed in good thermal contact with a copper plate, placed in a thermally insulated chamber. The temperature of this plate was adjusted at a fixed value between 5 °C and 45 °C in 5 °C steps using a Peltier element and stabilized with a temperature controller (PTC10, Stanford Research Systems). The thermally insulated chamber was connected to a self-made gas-mixing apparatus which enabled to vary the oxygen concentration between 0 % and 20 % vol O_2 by mixing nitrogen and dry air from two bottles. In the following, the concentration of oxygen will be given in % of the oxygen concentration of dry air and indicated with % air. This means, for example, that 20 % air corresponds to 4 % vol O_2 and 100 % air corresponds to 20 % vol O_2. The absolute error on the oxygen concentration adjusted with the gas mixing device is estimated to be below 1 % air. The oxygen concentration was varied between 0 % air and 100 % air in 5 % steps.

The optical setup used in this work for the luminescence measurements is shown schematically in Fig. 1.

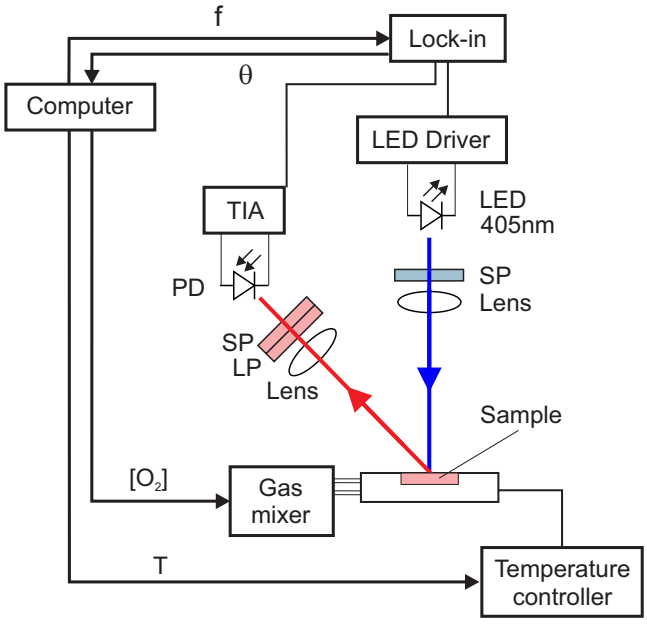

Fig. 1. Scheme of the optical experimental setup. Blue is the excitation, red the luminescence optical path. SP: short pass filter; LP: long pass filter PD: photodiode; TIA: trans-impedance amplifier.

The excitation light was provided by a 405 nm LED (VAOL-5EUV0T4, VCC Visual Communications Company LLC), filtered

by a an short pass (SP) filter with cut-off at 498 nm (Semrock 498 SP Bright Line HC short pass) and focused on the surface of the samples with a collimation lens (EO43987, Edmund Optics). The luminescence was focussed by a lens (G063020000, LINOS) and collected by a photodiode (SFH 213 Osram). To suppress stray light and light reflected by the sample surface, the emission channel was equipped with a long pass filter with cut-off at 594 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 where repeated. The total number of measurements is 50x20x21x9=189000 which lasted ca. 65 hours. The number of measurement was chosen as a compromise between maximizing the number of measurements and avoiding photodegradation. A the end of the session a minimal change in the phase shift was observed (QUANTIFY?)

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

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

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

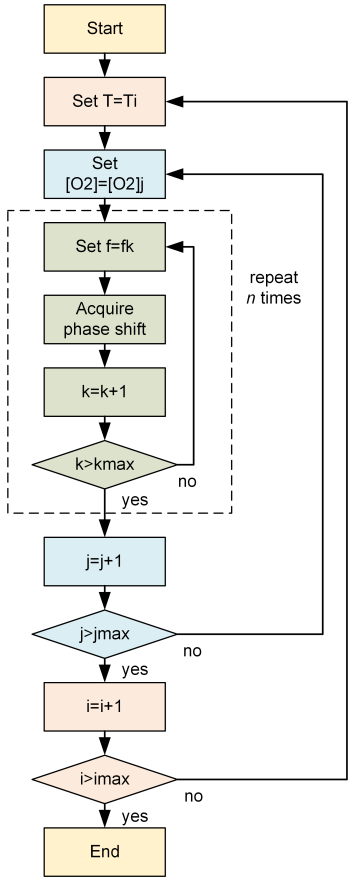

Fig. 2. Flow-chart of the automatic measurement.

combination of the task-specific cost functions with weights α_i will be minimized

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

The parameters α_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; $\boldsymbol{y}^{[j]} \in \mathbb{R}^d$ is the measured value of the desired quantity for observation j, with $j = 1, ..., n$; $\hat{\boldsymbol{y}}^{[j]} \in \mathbb{R}^d$ is the output of the network, when evaluated on the j^{th} observation. he 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).

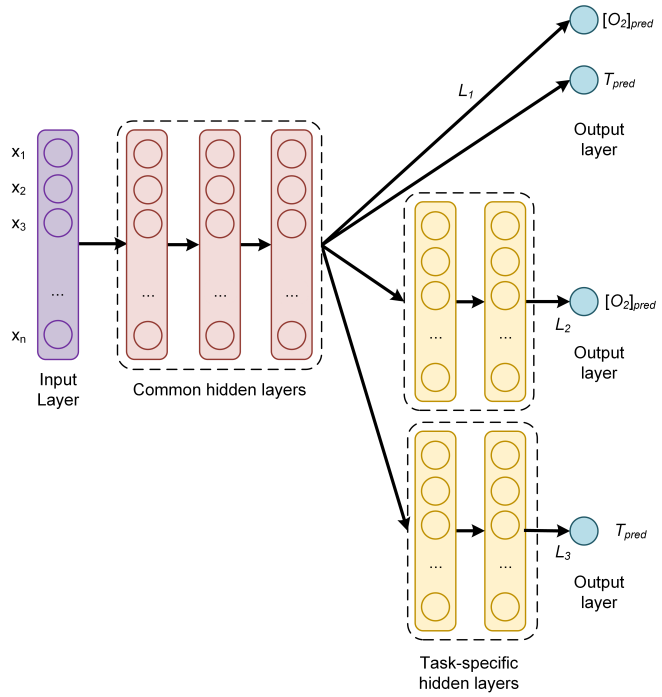

Fig. 3. Architecture of the multi-task learning architecture.

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 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

A. Luminescence Results

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.

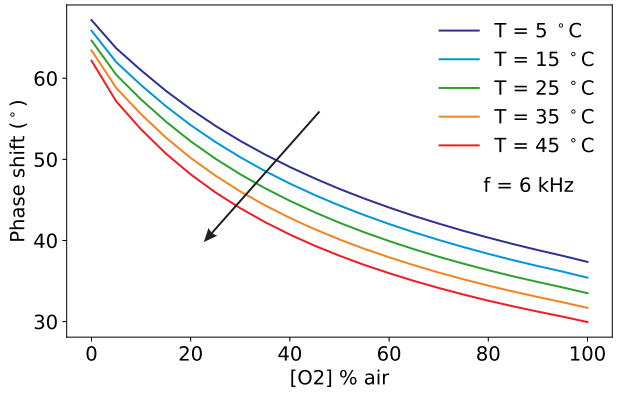

Fig. 4. Phase-shift as a function of the oxygen concentration, for selected temperatures at a fixed modulation frequency of 6kHz.

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.

B. Neural network performance

4. CONCLUSIONS

DISCLOSURES

Disclosures. The authors declare no conflicts of interest.

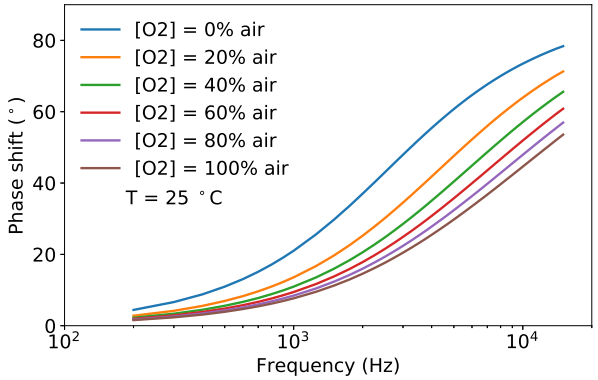

Fig. 5. Phase-shift for selected oxygen concentrations as a function of the modulation frequency at a fixed temperature $T = 25°$.

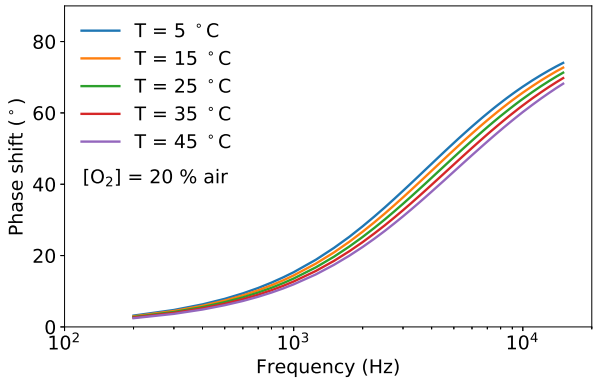

Fig. 6. Phase-shift for selected oxygen concentrations as a function of the modulation frequency at a fixed oxygen concentration $[O_2] = 20\%$ at selected temperatures.

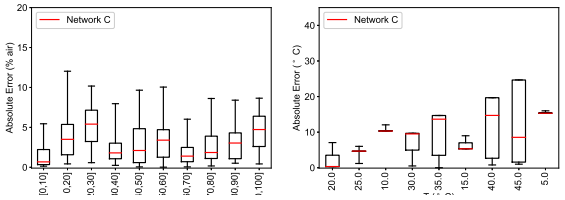

Fig. 7. Phase-shift at a fixed temperature $T = 25°$ for selected oxygen concentrations as a function of the modulation frequency of the excitation light.

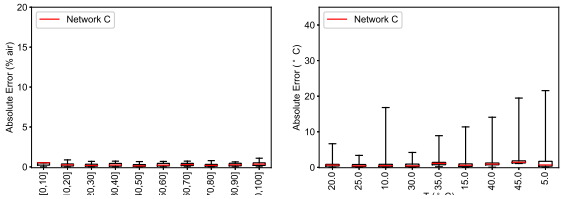

Fig. 8. Phase-shift at a fixed temperature $T = 25°$ for selected oxygen concentrations as a function of the modulation frequency of the excitation light.

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, "Photophysics 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 τoor 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).

~