Dual oxygen and temperature luminescence learning sensor with parallel inference

Francesca Venturini[a,1], Umberto Michelucci[a], Michael Baumgartner[b]

[a]*TOELT llc, Birchlenstr. 25, 8600 Dübendorf*
[b]*Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences,
Technikumstrasse 9, 8401 Winterthur, Switzerland*

Abstract

A well-known approach to the optical measure of oxygen is based on the quenching of luminescence by molecular oxygen. The main challenge for this measuring method is the development of an accurate mathematical model. Typically, this is overcome by using an approximate empirical model where these effects are parametrized ad hoc. The complexity increases further if multiple parameters (like oxygen concentration and temperature) need to be extracted, particularly if they are cross interfering. The common solution is to measure the different parameters separately, for example, with different sensors, and correct for the cross interferences. In this work, we propose a new approach based on a learning sensor with parallel inference. We show how it is possible to extract multiple parameters from a single set of optical measurements, without the need of any *a priori* mathematical model, and with unprecedented accuracy. The proposed approach is not limited to oxygen and temperature sensing. It can be applied to the sensing with multiple luminophores, whenever the underlying mathematical model is not known or too complex.

Keywords: Spectroscopy, Fluorescence, Luminescence, Neural networks, Remote sensing, Sensors

1. Introduction

The simultaneous determination of multiple physical quantities can be very advantageous in many sensor applications, for example, when an in-situ or a remote acquisition is required. If the physical effect on which the measurement method is based presents cross-interference of more than one

quantity, their simultaneous determination becomes a necessity. Optical luminescence sensing is particularly attractive for multiple sensing. Using the same measuring principle, several optical elements, like optical fibers and detectors, can be shared in the setup for the detection of more than one parameter, thus allowing a compact and simple sensor design.

The typical approaches to multiple sensing are based on either the use of a single luminescence indicator (luminophore), whose luminescence is sensitive to more than one physical quantity, or the use of several luminophores, one for each quantity, embedded in a substrate and placed in close physical proximity [1, 2, 3, 4, 5, 6]. To be able to determine each quantity separately, it may be necessary to determine more than one optical property (e.g., absorption spectrum, emission spectrum, luminescence intensity, decay time). Another possibility is to measure one single optical property using special detection schemes that take advantage of the emission properties of the used luminophores [7, 4, 8, 9, 6, 10].

The problem of dual sensing is particularly relevant in applications that involve oxygen sensing. The determination of oxygen partial pressure is of great interest in numerous fields, like medicine, biotechnology, environmental monitoring, or chemistry since oxygen plays an important role in many processes [11, 4]. One of the most used optical measuring approaches uses the effect of the dynamical luminescence quenching by oxygen molecules. The measuring principle is based on the measurement of the luminescence of a specific luminophore, whose intensity and decay time are reduced due to collisions with molecular oxygen [12].

Sensors based on this principle must rely on approximated empirical models to parametrize the dependence of the measured sensing quantity (e.g., luminescence intensity or decay time) on influencing factors. Among these, the temperature is the factor with the strongest influence since both the luminescence and the quenching phenomena are strongly temperature-dependent. Therefore, in any optical oxygen sensor, the temperature must be continuously monitored, most frequently with a separate sensor, and used to correct the calculated oxygen concentration [13]. This task can be difficult in practical implementation and may become a significant source of error in sensors based on luminescence sensing. Another disadvantage of this approach is that the parametrization of the sensor response with temperature is system-specific since it depends on how the sensing element was fabricated and on the sensor itself [14, 15, 16, 17, 18, 19].

In this work, we propose a revolutionary approach based on neural net-

works for parallel inference. The method enables accurate dual-sensing, using one single luminophore, and measuring a single quantity. Instead of describing the response of the sensor as a function of the relevant parameters through an analytical model, a neural network was designed and trained to predict both oxygen concentration and temperature simultaneously. This new approach is based on multi-task learning (MTL) neural network architectures. These are characterized by common hidden layers, whose output is then the input of multiple branches of task-specific hidden layers. MTL architectures were chosen because they can learn correlated tasks [20, 21, 22, 23, 24, 25] and are flexible enough to be usable in multi-dimensional regressions [26].

The collection of the large amount of data that is needed for the training and test of neural networks cannot be performed by hand due to the big amount of measurements needed to obtain good results from the neural network. Therefore in this paper, a fully automated data collection setup was developed, which controls both the instruments for adjusting the gas concentration and temperature, and collects the sensor response. This allowed us to collect enough measurements to train the neural network on real data and to test the sensor performance on other unseen real data. This was necessary since our previous work [REF1] was a purely theoretical study that used only synthetic data that was obtained by an analytical mathematical model based on a parametrized Stern-Volmer equation.

REF1: [26]

This work proposes a paradigm shift from the classical description of the response of a sensor through an approximate model to the use of MTL sensor learning thanks to neural networks. These will learn the complex inter-parameter dependencies and sensor-specific response characteristics from a large amount of data automatically collected. This new method will enable to build sensors even if the response of the system to the physical quantities is too complex to be comfortably described by a mathematical model.

In this work a real physical optical sensor have been built, was characterized and used for the acquisition of data needed for the training. The results described here demonstrate a real-sensor performance which would not be possible with a standard approach. Additionally, in this work there is much more. The necessary data acquisition procedure to be used efficiently with neural networks is described and demonstrated. This is essential to a sensor based on machine learning.

2. Methods

2.1. Luminescence Quenching for Oxygen Determination

Luminescence-based oxygen sensors usually consist of a luminophore whose luminescence intensity and decay time decrease for increasing O_2 concentrations. This reduction is due to collisions of the excited luminophore with molecular oxygen, which thus provides 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 [12]

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

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 Eq. (1). This deviation can be due, for example, to heterogeneities of the micro-environment of the luminophore, or to the presence of static quenching [4]. A proposed scenario describes this non-linear behavior as due to the presence in the substrate of two or more environments, in which the luminescence is quenched at different rates [27, 28]. This multi-site model describes the SV curve as the sum of n contributions as

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

where f_i's are the fractions of the total emission for each component under unquenched conditions, and K_{SVi}'s are the associated effective Stern–Volmer constants. Depending on the luminophore and on the substrate material, the models proposed in the literature may be even more complex [28, 29, 30].

In most industrial and commercial sensors, the decay time τ is frequently preferred to intensity measurement because of its higher reliability and robustness [31]. 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 additional advantage of allowing very simple and low-cost implementation.

Although the multi-site model was introduced for luminescence intensities, it is frequently also used to describe the oxygen dependence of the decay times [28, 32]. Therefore, in the simplest case of a two-sites scenario, the model can be rewritten in terms of phase shift as [33]

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

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 ω is the angular modulation frequency. It is to be noted that the quantities θ_0, f, K_{SV1}, and K_{SV2} are all non-linearly temperature dependent [34, 35, 36]. Additionally, if the modulation frequency is varied, they may show a frequency dependence, an artifact due to the approximate nature of the model. Finally, Eq. (3) needs to be inverted to determine $[O_2]$ from the measured quantity θ.

The proposed approach not only solves the difficulties of finding an approximate mathematical model for a complex system, but also allows the determination of multiple quantities simultaneously. Even if it is an approximate description, however, the structure of Eq. (3), remains relevant to understand the structure of the data and optimize the architecture of the neural network.

2.2. Experimental Procedure

The optical setup used in this work for the luminescence measurements is shown schematically in Fig. 1. To be able to acquire a large number of data, the program for both the instrument control and the data acquisition was written using the software LabVIEW by National Instruments. The acquisition procedure is described in detail in Section 2.2.2.

2.2.1. Experimental Setup

The sample used for the characterization and test is a commercially available Pt-TFPP-based oxygen sensor spot (PSt3, PreSens Precision Sensing).
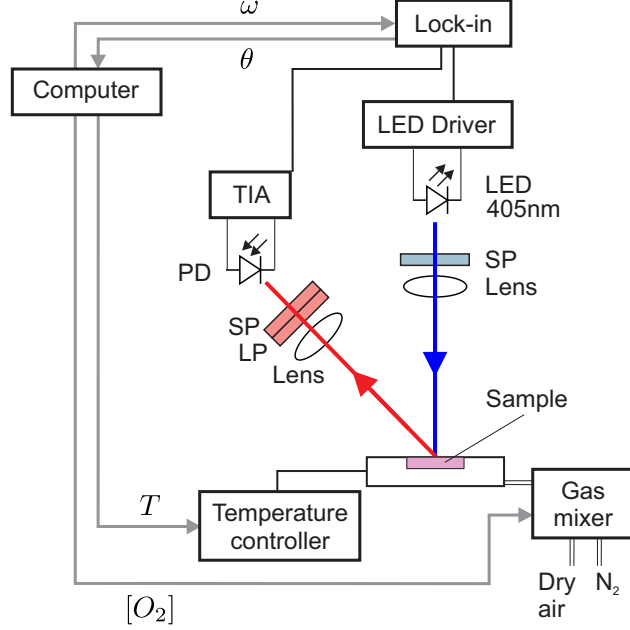
Figure 1: Schematic diagram of the experimental setup. Blue indicates the excitation optical path, red the luminescence one. SP: shortpass filter; LP: longpass filter PD: photodiode; TIA: trans-impedance amplifier.

To control its temperature, the sample was placed in good thermal contact with a copper plate, set in a thermally insulated chamber. The temperature of this plate was adjusted and stabilized using a Peltier element 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 was obtained by mixing 20 % dry air with 80 % nitrogen and therefore corresponds to 4 % vol O_2. The absolute error on the oxygen concentration adjusted with the gas mixing device is estimated to be below 1 % air.

The excitation light was provided by a 405 nm LED (VAOL-5EUV0T4, VCC Visual Communications Company LLC), filtered by a shortpass (SP) filter with cut-off at 498 nm (498 SP BrightLine HC Shortpass Filter, Semrock) and focused on the surface of the samples with a collimation lens. The luminescence was focused by a lens and collected by a photodiode (SFH 213,

6

Osram). To suppress stray light and light reflected by the sample surface, the emission channel was equipped with a longpass filter with cut-off at 594 nm (594 LP Edge Basic Longpass Filter, Semrock) and a shortpass filter with cut-off at 682 nm (682 SP BrightLine HC Shortpass Filter, Semrock). 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.

2.2.2. Automated Data Acquisition

The large amount of data needed for the training and the test of the neural network was acquired using an automated acquisition program which followed the flow-chart shown in Fig. 2. First, the program fixed the temperature and concentration. Then, the phase shift was measured for 50 modulation frequencies between 200 Hz and 15 kHz. This measurement was repeated 20 times. Next, keeping the temperature fixed, the program changed the oxygen concentration and the entire frequency-loop was repeated. The oxygen concentration was varied between 0 % air and 100 % air in 5 % air steps. Finally, the temperature was changed, and then the oxygen and frequency loops where repeated. The temperature was varied between 5 °C and 45 °C in 5 °C steps. The total number of measurements was thus 50 (frequencies) x 20 (loops) x 21 (oxygen concentrations) x 9 (temperatures) = 189'000, which required a total acquisition time of approximately 65 hours. This number of measurements was chosen as a compromise between maximizing the number of data and avoiding photodegradation, which naturally occurs when the sample is subjected to illumination. At the end of the session, a minimal change in the phase shift was observed.

2.3. Neural Network Approach

The software component of this new sensor type is based on a neural network model (NNM). A NNM is made of three components [37]: a neural network architecture (that includes how neurons are connected, the activation functions and all the hyperparameters), a loss function (here indicated with L) and an optimizer algorithm. In this section, those three components are described in detail.

2.3.1. Neural Network Architecture

The neural network used in this work has a multi-task-learning architecture and is depicted schematically in Fig. 3. It consists of three *common*

7

hidden layers with 50 neurons each, which generates as output a "shared representation". The name shared representation comes from the fact that the output of common hidden layers is used to predict both $[O_2]$ and T. These layers are followed by three branches, one without additional layers to predict $[O_2]$ and T at the same time, and two with each two additional
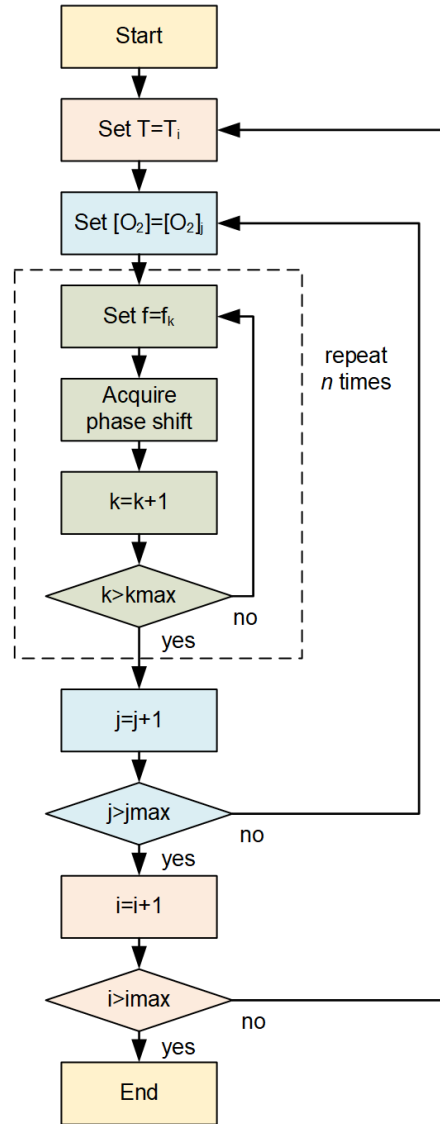
Figure 2: Flow-chart of the automated data acquisition program.

task-specific hidden layers to predict respectively $[O_2]$ and T. 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 features specific to each output and therefore improve the prediction accuracy. The number of neurons of each task-specific hidden layer used in this work is five. The activation function is the sigmoid function for all the neurons. A study of which network architecture works best with this kind of data can be found in [26].

The network was trained with two types of input to test its effectiveness. In the first case, each observation consists of a vector of 50 values defined as

$$\boldsymbol{\theta}_s = \left(\frac{\theta(w_1)}{90}, \frac{\theta(w_2)}{90}, ..., \frac{\theta(w_{50})}{90} \right) \tag{4}$$

where w_i are the 50 values of the angular modulation frequency of the excitation light (see Sec. 2.2). This choice of dividing by 90 have been made since, during testing, have been observed that such a normalisation delivers the best results and avoid problems like divergence of the neural network model. In the second case, each observation is

$$\boldsymbol{\theta}_n = \left(\frac{\theta(w_1)}{\theta_0(w_1)}, \frac{\theta(w_2)}{\theta_0(w_2)}, ..., \frac{\theta(w_{50})}{\theta_0(w_{50})} \right) \tag{5}$$

where $\theta_0(w_i)$ is the value of the phase shift without oxygen quenching at the angular modulation frequency w_i.

2.3.2. Loss Function

The task-specific loss functions for each branch i are indicated with L_i and is the mean square error (MSE) defined as

$$L_i = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{d_i} (y_{k,i}^{[j]} - \hat{y}_{k,i}^{[j]})^2, \quad i = 1, 2, 3 \tag{6}$$

where n is the number of observations in the input dataset; $\boldsymbol{y}_i^{[j]} \in \mathbb{R}^{d_i}$ is the measured value of the desired quantity for the jth observation, with $j = 1, ..., n$ and d_i is the dimension of the neural network branch output. In
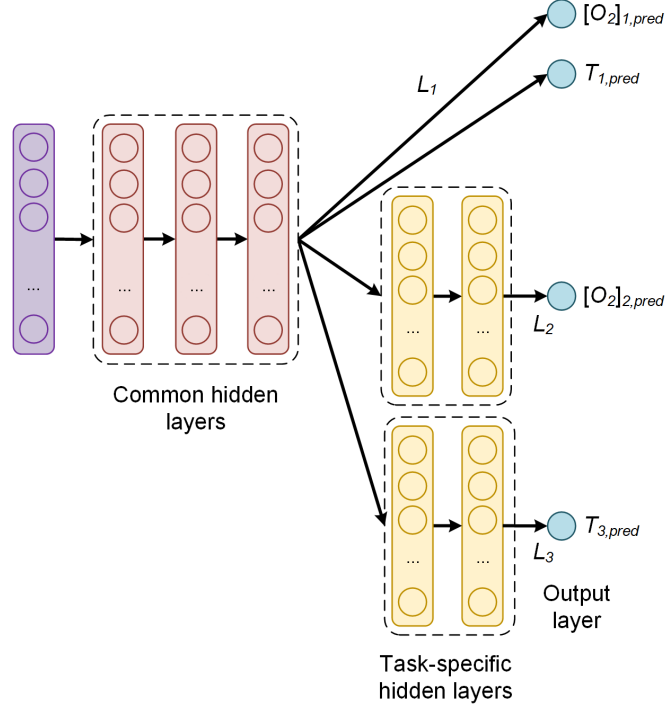
Figure 3: Architecture of the multi-task learning neural network used in this paper. The common hidden layers generate a "shared representation" as output, that is used as input to task specific branches that learn specific features to each quantity and therefore improve the prediction accuracy. L_i are the task-specific loss functions; $[O_2]_{i,pred}$ and $T_{i,pred}$ are the oxygen concentration and temperature predictions of the corresponding branch i. Note that branch 2 and 3 have only one output.

this case $d_1 = 2, d_2 = 1$ and $d_3 = 1$. $\hat{\boldsymbol{y}}_i^{[j]} \in \mathbb{R}^{d_i}$ is the output of the network branch i, when evaluated on the jth observation. Since there are multiple branches, a global loss function L is defined as a linear combination of the task-specific loss functions with weights α_i

$$L = \sum_{i=1}^{n_T} \alpha_i L_i. \tag{7}$$

The parameters α_i have to be determined during the hyper-parameter tuning phase to optimize the network predictions. In this paper, being the loss

10

function the MSE (Eq. (6)), the global loss function is

$$L = \sum_{i=1}^{3} \alpha_i \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{d_i} (y_{i,k}^{[j]} - \hat{y}_{i,k}^{[j]})^2 \qquad (8)$$

The global loss function weights used for this work were $\alpha_1 = 0.3$, $\alpha_2 = 5$ and $\alpha_3 = 1$. These parameters are the result of a hyper-parameter tuning for this architecture [26].

2.3.3. Optimiser Algorithm

The loss function was minimized using the optimizer Adaptive Moment Estimation (Adam) [38, 37]. The implementation was performed using the TensorFlow^{TM} library. The training was performed with a starting learning rate of 10^{-3}. Two types of training were investigated to compare the training efficiency and performance of the network. *No-batch training*: with this method all the training data are used to perform an update of the weights and to evaluate the loss function. The loss function used is given by Eq. (6) where n is the total number of observations available. *Mini-batch training*: with this method the weights update is performed after the network has seen 32 observations. In this case Eq. (6) is used with $n = 32$. For each update of the weights, 32 random observations are chosen from the training dataset without repetitions until all the training data are fed to the network. The authors have tested many mini-batch sizes. 32 was a good balance between performance (MSE values) results and time duration of training. Note that generally speaking a neural network model with small mini-batches will need less epochs to reach a certain performance value 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.

No-batch training has the advantage of stability and speed since it performs one weight update using the entire training dataset. Mini-batch training is normally much more effective, allowing to reach smaller values of the loss function in less epochs, but usually takes more time [37]. In our experiments for $20 \cdot 10^3$ epochs no-batch training took roughly five minutes on a modern MacBook Pro, while mini-batch training with $b = 32$ took approximately 1 hour, thus resulting ca. 12 times slower.

11

2.4. Performance Evaluation

To evaluate the performance of the sensor, different metrics were analyzed. These are discussed in the next sections. The measured dataset S have been divided in two parts: one containing 80% of randomly chosen observations (indicated with S_{train}), and one containing the remaining 20% of the data (indicated with S_{test}). The results shown in the following sections have been obtained by measuring the different metrics on the S_{test} dataset.

2.4.1. Absolute Error on the Prediction

The metric used to compare predictions from expected values is the absolute error (AE) defined as the absolute value of the difference between the predicted and the expected value for a given observation. Note that in the architecture described in the previous sections, only branch 1 and 2 can predict $[O_2]$, while only branch 1 and 3 can predict T. The AE for the oxygen concentration for the jth observation $[O_2]^{[j]}$ is

$$AE_{[O_2]}^{[j]} = |[O_2]_{pred}^{[j]} - [O_2]_{meas}^{[j]}|. \tag{9}$$

where $[O_2]_{pred}^{[j]}$ and $[O_2]_{meas}^{[j]}$ are respectively the $[O_2]$ network prediction and measured value. The further quantity used to analyse the performance of the network is the mean absolute error (MAE), defined as the average of the AE. For example, for the oxygen prediction using the training dataset S_{train}, the $MAE_{[O_2]}$ is defined as

$$MAE_{[O_2]}(S_{train}) = \frac{1}{|S_{train}|} \sum_{j \in S_{train}} |[O_2]_{pred}^{[j]} - [O_2]_{real}^{[j]}| \tag{10}$$

where $|S_{train}|$ is the size (or cardinality) of the training dataset. AE_T and MAE_T are similarly defined, using the prediction and the measured temperature values.

2.4.2. Kernel Density Estimation

A fundamental quantity to study the performance of the network is the prediction distribution of the AEs. This metrics carries information on the probability of the network to predict the expected value. To better illustrate this distribution, the kernel density estimate (KDE) of the distributions of the AEs was also calculated for both the oxygen concentration and the temperature. KDE is a non-parametric algorithm to estimate the probability

density function of a random variable by inferring the population distribution based on a finite data sample [39]. In this work a Gaussian Kernel and a Scott bandwidth adaptive estimation [40] using the seaborn Python package [41] were used.

2.4.3. Error Limited Accuracy η

Generally, in a commercial sensor, the accuracy quantifies the performance of the sensor and helps to decide if the chosen device is appropriate for the application of interest. The above-defined metrics (AE, MAE and KDE) are useful to compare the performance of different NNMs but do not help quantify which error the neural network senor will ultimately have in practice. For this reason, in this work we introduce a new metric, called Error Limited Accuracy (ELA) and indicated with η.

Definition. *In a regression problem, given the metric AE, and a chosen value of it $\hat{AE}$, the ELA η limited by the error $\hat{AE}$ is defined as the number of predictions $\hat{y}$ of the NNM that lie in the range $|\hat{y} - y| \leq \hat{AE}$, with y the expected value, divided by the total number of observations. It will be indicated with $\eta(\hat{AE})$. In more mathematical terms, given the set*

$$E(\hat{AE}) = \{\hat{y}^{[i]} \ with \ i = 1, ..., n \ | \ |\hat{y}^{[i]} - y^{[i]}| \leq \hat{AE}\} \tag{11}$$

$\eta(\hat{AE})$ is defined as

$$\eta(\hat{AE}) = \frac{|E(\hat{AE})|}{n} \tag{12}$$

where $|E(\hat{AE})|$ is the cardinality of the set $E(\hat{AE})$ or in other words, the number of its elements.

This metric allows interpreting the regression problem as a classification one. $\eta(\hat{AE})$ simply describes how many observations are predicted by the NNM within a given value of the absolute error. In other words, it represents the percentage of predictions that are within a certain error $\hat{AE}$ from the expected values. Finally, if we take $\hat{AE}$ big enough, all the predictions will be classified perfectly, so $\eta(\hat{AE})$ is expected to approach 1. The smaller $\hat{AE}$ is, the smaller will be the number of predictions correctly classified. We finally define $\overline{AE}$ as the value for which $\eta(\overline{AE}) = 1$, so the value of the absolute error for which the network predicts all the observations correctly. This value ($\overline{AE}$) will give us the biggest error in the sensor predictions.

3. Results and Discussion

3.1. Luminescence Experimental Results

As described in Section 2.1, the phase shift non-linearly depends on the oxygen concentration according to the Stern-Volmer equation. It depends also on the temperature, which influences the luminescence and the collision mechanisms, and on the modulation frequency of the excitation light, as described in Eq. (3). The measured behaviour of the phase shift for variations of these three quantities is shown in the Figs. 4 to 6.

Fig. 4 shows the measured phase shifts as a function of the oxygen concentration at a constant modulation frequency of 6 kHz and for increasing temperatures. For clarity, the results at only few selected temperatures are shown. The decrease of the phase shift due to the collisional quenching is clearly visible in all curves. The phase shift is, as expected, also strongly temperature-dependent. For $[O_2] = 0$, in the absence of oxygen, the reduction of the phase shift with increasing T is due to temperature quenching; the influence of temperature becomes stronger at higher oxygen concentration, as a result of the increase of the diffusion rates of oxygen through the sample.

For a given oxygen concentration, the phase shift is strongly dependent on the modulation frequency, as it can be seen in Fig. 5, where the shape of the frequency response is determined by the distribution of decay times of the sample. From the figure it is visible that the reduction of the phase shift
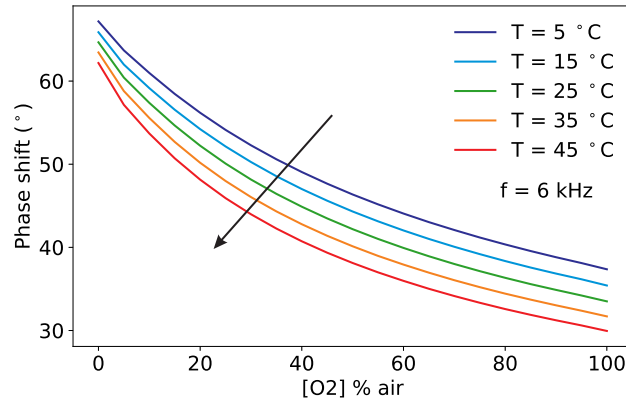

Figure 4: Measured phase shift as a function of the oxygen concentration for selected temperatures at a fixed modulation frequency of 6 kHz. The arrow marks increasing temperatures.
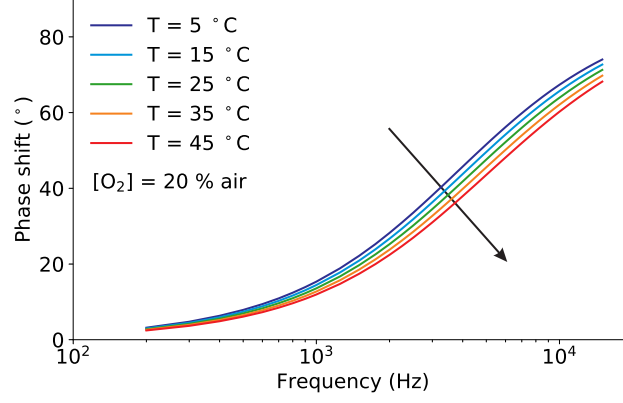
Figure 5: Measured phase shift as a function of the modulation frequency for selected temperatures at a fixed oxygen concentration of $[O_2] = 20$ % air. The arrow marks increasing temperatures.
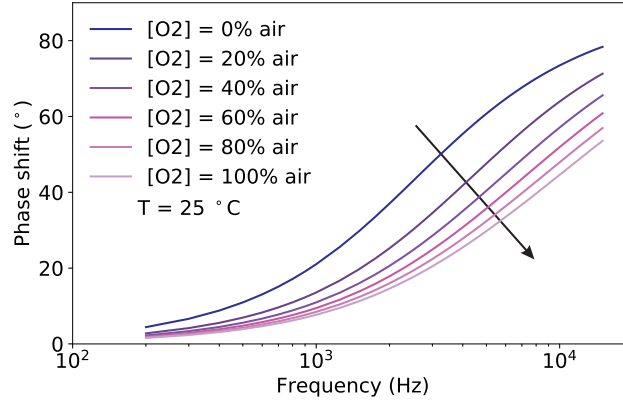

Figure 6: Measured phase shift as a function of the modulation frequency for selected oxygen concentrations at a fixed temperature of $T = 25$ °C. The arrow marks increasing oxygen concentrations.

with increasing temperatures is not constant but depends on the modulation frequency.

For completeness, the effect of the oxygen concentration on the frequency response at a fixed temperature is shown in Fig. 6. Compared to Fig. 5, the frequency response of the sample is affected more strongly by the oxygen concentration than by temperature. In other words, the sample has a higher sensitivity to oxygen than to temperature.

The measurements of Figs. 4 to 6 show how similar the curves of the phase

15

shift may look for different values of oxygen, temperature and modulation frequency. This helps to understand why it is not possible from the measurement of the phase shift, or even of the phase shift for varying modulation frequencies, to simultaneously determine both the oxygen concentration and the temperature using Eq. (3). The temperature must be known in advance and used to compute the oxygen concentration. This is no longer the case with the neural network approach, as it will be shown in the next section.

3.2. Sensor performance

First, the effect of the training on the sensor performance was investigated. As described in Section 2.3.3, the neural network was trained with no-batches and with mini-batches. For this comparison the network was trained for 20'000 epochs using the input observations $\boldsymbol{\theta}_s$ as defined in Eq. (4). The results for $AE_{[O_2]}$ and AE_T are shown in Fig. 7(A) and 7(B), respectively. The blue histogram shows the AE distribution when using no-batch, the gray when using mini-batches of size 32. The KDE profiles help illustrating the features of the histogram. The effect of introducing mini-batches on the performance is significant. The predictions distributions get much narrower, the mean average errors decrease from $MAE_{[O_2]} = 2.4$ % air and $MAE_T = 3.6$ °C to $MAE_{[O_2]} = 1.4$ % air and $MAE_T = 1.6$ °C. Although the performance is significantly improved, from Fig. 7(A) and 7(B) it can also be clearly seen that errors as high as approximately 5 % air for $[O_2]$ or 12 °C for T are possible. While deciding what the right mini-batch size is, training time must be taken into account. Decreasing the mini-batch size has the side effect of increasing the training time. While training the network without batches requires just a few minutes, reducing the mini-batch size to 32 increases the training time to approximately one hour.

Fig. 7(C) and 7(D) show the comparison between prediction distributions with 20'000 and 100'000 epochs (always using a mini-batch of size 32), using the input observations $\boldsymbol{\theta}_s$ as defined in Eq. (4). The effect of longer training is a dramatic improvement in the performance. When the network was trained for 100'000 epochs the mean average errors are reduced to only $MAE_{[O_2]} = 0.22$ % air and $MAE_T = 0.27$ °C. Additionally, all the predictions for $[O_2]$ lie below 0.94 % air, and for T lie below 2.1 °C.

The results of Fig. 7(C) and 7(D) demonstrate two new findings: 1) with the proposed approach, it is possible to predict both $[O_2]$ and T at the same time from the phase shift using a single luminophore; 2) the prediction has an expected error which is comparable or below the typical accuracy of
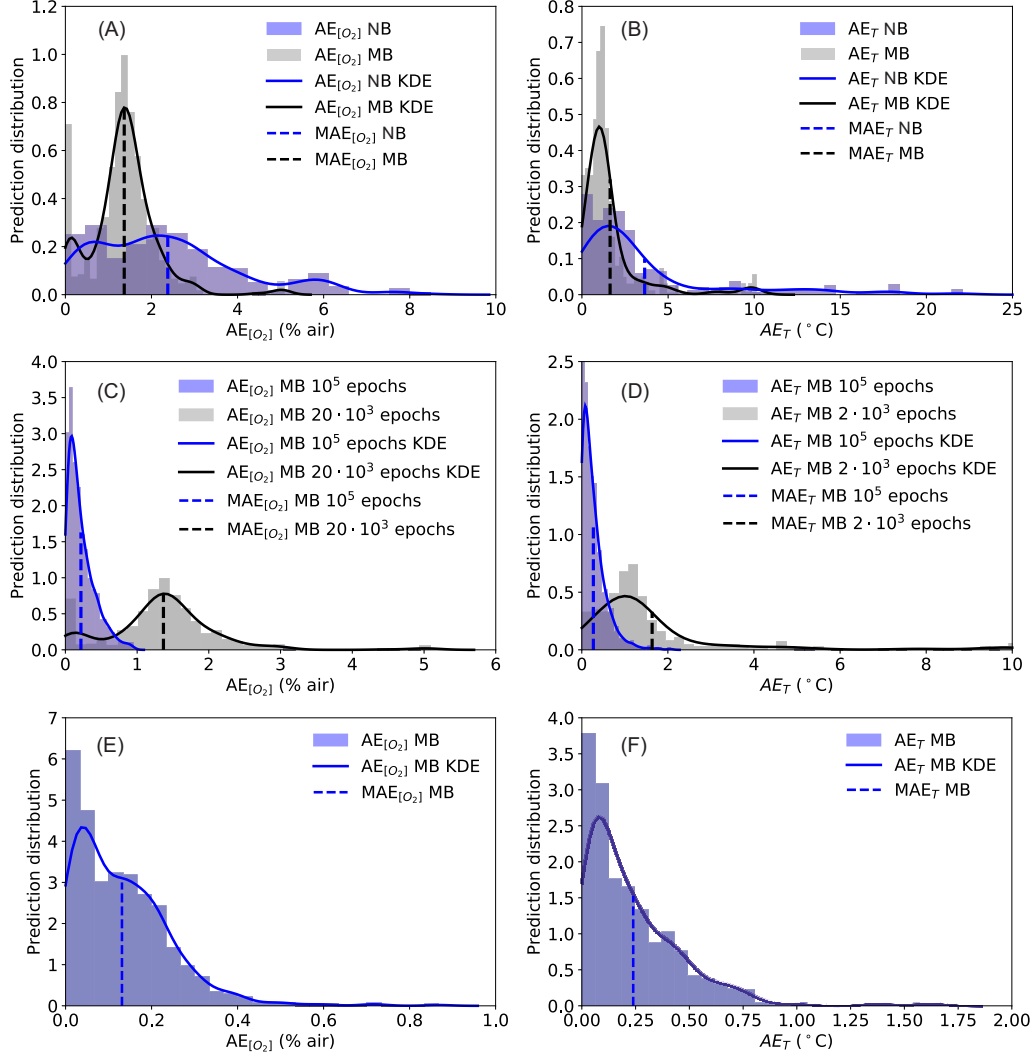
Figure 7: Distributions of the neural network predictions for the oxygen concentration (panels (A), (C) and (E)) and for the temperature (panels (B), (D) and (F)). In all panels the normalized prediction distribution histogram (columns), the kernel density estimate (KDE) of the distribution of the AEs (solid line), and MAE (dashed vertical line) are shown. Panels (A) and (B): Comparison between training using no batches (NB) and using mini-batches (MB) with a batch size of 32 for 20'000 epochs; the input of the network is $\boldsymbol{\theta}_s$. Panels (C) and (D): Comparison between training using mini-batches (MB) with a batch size of 32 for 100'000 and 20'000 epochs; the input of the network is $\boldsymbol{\theta}_s$. Panels (E) and (F): training using mini-batches (MB) with a batch size of 32 for 20'000 epochs; the input of the network is $\boldsymbol{\theta}_n$.
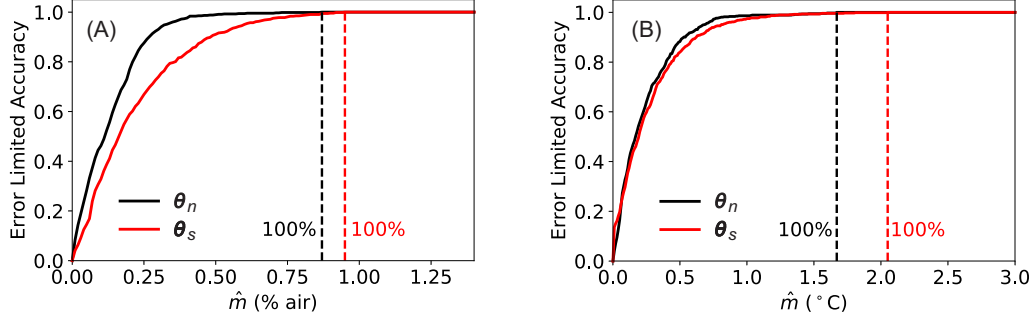
Figure 8: Comparison of the ELA η: Panel (A) oxygen prediction, panel (B) temperature prediction. The black lines are the results obtained with a network that was trained with $\boldsymbol{\theta}_n$ as input for 20'000 epochs with mini-batchs of size 32, while the red ones with $\boldsymbol{\theta}_s$ as input for 100'000 epochs with mini-batchs of size 32. The dashed lines indicates the values of the $\overline{AE}$ for which the predictions would give $\eta = 1$.

commercial sensors. The possibility of dual sensing paves the road to the development of a completely new generation of sensors. The price to pay is that the training of a network for 100'000 epochs requires approximately 5 hours on a modern laptop.

To investigate if the training can be performed more efficiently, the normalized phase shift $\boldsymbol{\theta}_n$ defined in Eq. (5) is used as input to the network. The performance of the network in this case, with a mini-batch size of 32 and a training of 20'000 epochs is shown in Fig. 7(E) and 7(F). The performance is further improved: even if the number of epochs is only 20'000 the mean average errors are better than what obtained with $\boldsymbol{\theta}_s$ and a training of 100'000 epochs, achieving $MAE_{[O_2]} = 0.13$ % air and $MAE_T = 0.24$ °C. The distributions are also narrower, particularly for the temperature. Additionally, all the $AE_{[O_2]}$ lie below 0.87 % air, and AE_T below 1.7 °C. This type of training is clearly more efficient. The reason may lie in the additional information which is fed to the network when using the input $\boldsymbol{\theta}_n$ and in the simplified functional behavior of $\boldsymbol{\theta}_n$ compared to $\boldsymbol{\theta}_s$ as it may be expected by Eq. (3).

The performance of the different neural networks is summarized in Table 1.

3.3. Error Limited Accuracy

The metrics discussed in the previous sections are useful to compare the network performance and to measure how good the predictions are. However,

Table 1: **Summary of the performance for neural network models**

Input	Epochs / Batch size	$MAE_{[O_2]}$	$MAE_{[T]}$
$\boldsymbol{\theta}_s$	20'000 / no batch	2.4 % air	3.6 °C
$\boldsymbol{\theta}_s$	20'000 / 32	1.4% air	1.6 °C
$\boldsymbol{\theta}_s$	100'000 / 32	0.22 % air	0.27 °C
$\boldsymbol{\theta}_n$	20'000 / 32	0.13 % air	0.24 °C

Table 2: **Summary of the values of $\overline{AE}$ for the cases shown in Fig. 8(A) and 8(B).**

Input	Epochs / Batch size	$\overline{AE}_{[O_2]}$	$\overline{AE}_T$
$\boldsymbol{\theta}_s$	100'000 / 32	0.95 % air	2.1 °C
$\boldsymbol{\theta}_n$	20'000 / 32	0.87% air	1.7 °C

they do not offer an understanding on what a sensor built with such a model could achieve. For practical applications, the relevant question is rather what is the maximum error which the sensor will have in predicting the oxygen concentration and temperature. To answer this question, the ELA (η) defined in Section 2.4.3 can be used. As explained previously, η is defined depending on the chosen metric m. In this section, the metric chosen is $m = AE_{[O_2]}$ for the oxygen concentration and $m = AE_T$ for the temperature. This new metrics will allow the determination of the maximum error of the sensor.

Fig. 8 displays the ELA $\eta(\widehat{AE})$ for oxygen concentration (A) and for the temperature (B). In each panel, the results obtained using the input $\boldsymbol{\theta}_n$ and a training for 20'000 epochs are shown in black, and the results obtained using the input $\boldsymbol{\theta}_s$ and a training for 100'000 epochs in red. In both cases, the training was performed with mini-batches of size 32. The dashed lines indicate the values of the $\overline{AE}_{[O_2]}$ and $\overline{AE}_T$ for which the error limited accuracy η equals 1. In other words, all the predictions will have an error equal or smaller than $\overline{AE}$.

Fig. 8(A) shows that, for the network trained with $\boldsymbol{\theta}_s$ as input, the model would predict perfectly all the oxygen concentrations within 0.95 % air error. For the network trained with $\boldsymbol{\theta}_n$ this value is futher reduced to 0.87 % air. $\overline{AE}_{[O_2]}$ can be interpreted as the accuracy a sensor based on this NNM would have. Fig. 8(B) shows the results of the same analysis for the temperature measurement. The interpretation is similar to the one given above for the oxygen concentration. For the network trained with $\boldsymbol{\theta}_s$ as input, the model

19

would predict perfectly all the temperature values within $\overline{AE}_T = 2.1$ °C error. For the network trained with $\boldsymbol{\theta}_n$ this value would be $\overline{AE}_T = 1.7$ °C. The values of $\overline{AE}_{[O_2]}$ and $\overline{AE}_T$ are summarized in Table 2.

4. Conclusions

In this work, a new sensor learning approach to luminescence sensing is presented. The proposed method allows parallel inference, or the extraction of multiple physical quantities simultaneously, from a single dataset without any *a priori* mathematical model, even in presence of cross interferences. Classical approaches to this type of problems in physics can be challenging or impossible to solve if the mathematical models describing the functional dependencies are too complex or even unknown.

The approach is demonstrated by a luminescence sensor, which uses a single luminophore and a single measuring channel, and can measure simultaneously both the oxygen concentration and the temperature of a medium. This is achieved using a multi-task learning neural network model, which was trained on a large dataset. The results in the prediction of the oxygen concentration and temperature show unprecedented accuracy for both parameters, demonstrating that this approach can make a new generation of dual- or even multiple-parameter sensors possible. To be able to estimate the expected error or accuracy of a sensor build with a given NNM, a new metric, the ELA ($\eta(AE)$), is proposed. The ELA enables to estimate how many predicted values lie within a certain absolute error from the expected measurement. This new metric allows giving a maximum measurement error of the NNM results, effectively transforming a regression problem in a classification one.

The ability in predicting both $[O_2]$ and T at the same time, from the a single set of data obtained with a single indicator, has profound implication for the development of luminescence sensors. Sensors will become easier and cheaper to build since no separate temperature measurements are necessary anymore. Similarly, the effect of interferenes can be learned by the neural network and do not need to be corrected for in the data processing. This work opens the road to complete new optical sensing approaches for future generations of sensors. Those sensors will be able to extract multiple physical quantities from a common set of data at the same time to achieve consistent results that are both accurate and stable. The described approach is relevant

for many practical applications in sensor science and demonstrates that this model-free approach has the potential of revolutionizing optical sensing.

Additionally as mentioned in this paper, to gather the data and train the model 65 hours (of training data collection) plus 5 hours (for training the model) are required. The question arises if this amount of time is needed if one wants to train the same neural network with a new sensor. Fortunately first tests seem to indicate that this is not necessary. In machine learning it is possible to use what is called transfer learning REF to fine-tune an almost completely trained network on a new dataset with a much-reduced amount of data (whose collection will require much less than 65 hours) that would be measured on the new sensor. The authors are working in this direction and will investigate the extent of new data needed in a future publication. The first tests are very positive and indicates that such a transfer is indeed possible and very efficient. [42]

Disclosures. The authors declare no conflicts of interest.

[1] M. I. Stich, L. H. Fischer, O. S. Wolfbeis, Multiple fluorescent chemical sensing and imaging, Chemical Society Reviews 39 (8) (2010) 3102–3114.

[2] S. M. Borisov, R. Seifner, I. Klimant, A novel planar optical sensor for simultaneous monitoring of oxygen, carbon dioxide, pH and temperature, Analytical and bioanalytical chemistry 400 (8) (2011) 2463–2474.

[3] T. Kameya, Y. Matsuda, Y. Egami, H. Yamaguchi, T. Niimi, Dual luminescent arrays sensor fabricated by inkjet-printing of pressure-and temperature-sensitive paints, Sensors and Actuators B: Chemical 190 (2014) 70–77.

[4] X.-d. Wang, O. S. Wolfbeis, Optical methods for sensing and imaging oxygen: materials, spectroscopies and applications, Chemical Society Reviews 43 (10) (2014) 3666–3761.

[5] S. Santoro, A. Moro, C. Portugal, J. Crespo, I. Coelhoso, J. Lima, Development of oxygen and temperature sensitive membranes using molecular probes as ratiometric sensor, Journal of Membrane Science 514 (2016) 467–475.

[6] S. Biring, A. S. Sadhu, M. Deb, An Effective Optical Dual Gas Sensor for Simultaneous Detection of Oxygen and Ammonia, Sensors 19 (23) (2019) 5124.

[7] B. B. Collier, M. J. McShane, Time-resolved measurements of luminescence, Journal of Luminescence 144 (2013) 180–190.

[8] C. Stehning, G. A. Holst, Addressing multiple indicators on a single optical fiber-digital signal processing approach for temperature compensated oxygen sensing, IEEE Sensors Journal 4 (1) (2004) 153–159.

[9] P. Jorge, C. Maule, A. Silva, R. Benrashid, J. Santos, F. Farahi, Dual sensing of oxygen and temperature using quantum dots and a ruthenium complex, Analytica chimica acta 606 (2) (2008) 223–229.

[10] J. P. Moore, C. Higgins, O. McGaughey, B. G. Lawless, B. D. MacCraith, Exploiting sensor cross sensitivity: achieving temperature compensation via a dual-element optical oxygen sensor, in: Advanced Environmental, Chemical, and Biological Sensing Technologies IV, vol. 6377, International Society for Optics and Photonics, 63770I, 2006.

[11] D. B. Papkovsky, R. I. Dmitriev, Biological detection by optical oxygen sensing, Chemical Society Reviews 42 (22) (2013) 8700–8732.

[12] J. R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd ed., Springer, New York, ISBN 978-0-387-31278-1, 2006.

[13] F. Li, Y. Wei, Y. Chen, D. Li, X. Zhang, An intelligent optical dissolved oxygen measurement method based on a fluorescent quenching mechanism, Sensors 15 (12) (2015) 30913–30926.

[14] W. Xu, R. C. McDonough, B. Langsdorf, J. Demas, B. DeGraff, Oxygen sensors based on luminescence quenching: interactions of metal complexes with the polymer supports, Analytical Chemistry 66 (23) (1994) 4133–4141.

[15] S. Draxler, M. E. Lippitsch, I. Klimant, H. Kraus, O. S. Wolfbeis, Effects of polymer matrixes on the time-resolved luminescence of a ruthenium complex quenched by oxygen, The Journal of Physical Chemistry 99 (10) (1995) 3162–3167.

[16] P. Hartmann, W. Trettnak, Effects of polymer matrices on calibration functions of luminescent oxygen sensors based on porphyrin ketone complexes, Analytical chemistry 68 (15) (1996) 2615–2620.

[17] A. Mills, Controlling the sensitivity of optical oxygen sensors, Sensors and Actuators B: Chemical 51 (1-3) (1998) 60–68.

[18] D. Badocco, A. Mondin, P. Pastore, S. Voltolina, S. Gross, Dependence of calibration sensitivity of a polysulfone/Ru (II)-Tris (4, 7-diphenyl-1, 10-phenanthroline)-based oxygen optical sensor on its structural parameters, Analytica chimica acta 627 (2) (2008) 239–246.

[19] F. Dini, E. Martinelli, R. Paolesse, D. Filippini, A. D'Amico, I. Lundström, C. Di Natale, Polymer matrices effects on the sensitivity and the selectivity of optical chemical sensors, Sensors and Actuators B: Chemical 154 (2) (2011) 220–225.

[20] P. M. Argyriou A., Evgeniou T., Multi-task feature learning., In Proceedings of the 19th International Conference on Neural Information Processing Systems (NIPS'06) .

[21] S. Thrun, Is learning the n-th thing any easier than learning the first?, in: Advances in neural information processing systems, 640–646, 1996.

[22] R. Caruana, Multitask learning, Machine learning 28 (1) (1997) 41–75.

[23] Y. Zhang, Q. Yang, A survey on multi-task learning, arXiv preprint arXiv:1707.08114 .

[24] J. Baxter, A model of inductive bias learning, Journal of artificial intelligence research 12 (2000) 149–198.

[25] K.-H. Thung, C.-Y. Wee, A brief review on multi-task learning, Multimedia Tools and Applications 77 (22) (2018) 29705–29725.

[26] U. Michelucci, F. Venturini, Multi-task learning for multi-dimensional regression: application to luminescence sensing, Applied Sciences 9 (22) (2019) 4748.

[27] E. Carraway, J. Demas, B. DeGraff, J. Bacon, Photophysics and photochemistry of oxygen sensors based on luminescent transition-metal complexes, Analytical chemistry 63 (4) (1991) 337–342.

[28] J. N. Demas, B. DeGraff, W. Xu, Modeling of luminescence quenching-based sensors: comparison of multisite and nonlinear gas solubility models, Analytical Chemistry 67 (8) (1995) 1377–1380.

[29] P. Hartmann, M. J. Leiner, M. E. Lippitsch, Luminescence quenching behavior of an oxygen sensor based on a Ru (II) complex dissolved in polystyrene, Analytical Chemistry 67 (1) (1995) 88–93.

[30] A. Mills, Response characteristics of optical sensors for oxygen: models based on a distribution in τoor kq, Analyst 124 (9) (1999) 1301–1307.

[31] Y. Wei, Y. Jiao, D. An, D. Li, W. Li, Q. Wei, Review of Dissolved Oxygen Detection Technology: From Laboratory Analysis to Online Intelligent Detection, Sensors 19 (18) (2019) 3995.

[32] M. Quaranta, S. M. Borisov, I. Klimant, Indicators for optical oxygen sensors, Bioanalytical reviews 4 (2-4) (2012) 115–157.

[33] U. Michelucci, M. Baumgartner, F. Venturini, Optical oxygen sensing with artificial intelligence, Sensors 19 (4) (2019) 777.

[34] V. I. Ogurtsov, D. B. Papkovsky, Modelling of phase-fluorometric oxygen sensors: Consideration of temperature effects and operational requirements, Sensors and Actuators B: Chemical 113 (2) (2006) 917–929.

[35] Y.-L. Lo, C.-S. Chu, J.-P. Yur, Y.-C. Chang, Temperature compensation of fluorescence intensity-based fiber-optic oxygen sensors using modified Stern–Volmer model, Sensors and Actuators B: Chemical 131 (2) (2008) 479–488.

[36] N. Zaitsev, P. Melnikov, V. Alferov, A. Kopytin, K. German, Stable optical oxygen sensing material based on perfluorinated polymer and fluorinated platinum (II) and palladium (II) porphyrins, Procedia Engineering 168 (2016) 309–312.

[37] U. Michelucci, Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks, APRESS Media, LLC, ISBN 9780198520115, 2018.

[38] J. A. Kingma, D.P.; Ba, Adam: A method for stochastic optimization. In Proceedings of 3rd., In Proceedings of 3rd International Conference on Learning Representations, ICLR 2015 (2015) 1–15.

[39] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media, 2009.

[40] S. R. Sain, D. W. Scott, On locally adaptive density estimation, Journal of the American Statistical Association 91 (436) (1996) 1525–1534.

[41] M. Waskom, O. Botvinnik, J. Ostblom, S. Lukauskas, P. Hobson, MaozGelbart, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, C. Swain, A. Miles, T. Brunner, D. O'Kane, T. Yarkoni, M. L. Williams, C. Evans, mwaskom/seaborn: v0.10.0 (January 2020), doi:10.5281/zenodo.3629446, URL https://doi.org/10.5281/zenodo.3629446, 2020.

[42] K. Weiss, T. M. Khoshgoftaar, D. Wang, A survey of transfer learning, Journal of Big Data 3 (1) (2016) 9, doi:10.1186/s40537-016-0043-6, URL https://doi.org/10.1186/s40537-016-0043-6.
~