Dual oxygen and temperature luminescence sensor based on artificial intelligence

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*

The optical determination of oxygen partial pressure is of great interest in numerous areas, like medicine, biotechnology, and chemistry. A well-known approach to the optical measure of oxygen is based on the quenching of luminescence by molecular oxygen. Sensors based on this principle typically rely on approximate empirical models to characterize the dependence of the luminescence intensity and decay time on the oxygen concentration through a multi-parametric model (Stern–Volmer equation), whose parameters are, however, all temperature-dependent. Therefore, the temperature needs to be known to determine the oxygen concentration and is measured separately, either optically or with a completely different sensor. In this work, we propose a new artificial intelligence approach, which allows the extraction of both the oxygen concentration and the temperature using one single indicator, which is sensitive to both oxygen and temperature, and measuring only the decay time. The results show that the neural network achieves predictions of both parameters, which are comparable to the accuracy of commercial senors and therefore demonstrate the feasibility of this approach. Furthermore, the proposed artificial intelligence approach is not limited to oxygen and temperature sensing but can be applied to the luminescence of multiple indicators, whenever the underlying mathematical model is not known or too complex to derive the desired quantities from a single measurement.

1. INTRODUCTION

The simultaneous determination of multiple physical quantities can be very advantageous in many sensor applications, for example, when an in-situ or remote acquisition is required. If the physical effect on which the measurement method is based presents cross-interference between the desired quantities, their simultaneous determination becomes a necessity. In particular 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 easy 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 quantities or the use of several luminophores, one for each quantity, embedded in a substrate and placed in close physical proximity [1–6]. To be able to determine each quantity separately it is usually necessary to determine more than one optical property (e.g., absorption spectrum, emission spectrum, luminescence intensity, decay time) or to use special detection schemes to take advantage of the differences between optical properties of the used luminophores [4, 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 chem-

istry since oxygen plays an important role in many processes [4, 11]. 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 of the environment where the indicator is placed. Among these, temperature is the parameter with the strongest influence since both the luminescence and the quenching phenomena are strongly temperature-dependent. Therefore, today 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 be a major 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 [14–19].

In this work, we propose a revolutionary approach based on artificial intelligence. The method enables accurate dual-sensing, using one single indicator, 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, in facts, can learn correlated tasks [20–24] and are flexible enough to be usable in multi-dimensional regressions [25].

The collection of the large amount of data that is needed for the training and test of neural networks cannot be performed by hand. Therefore, a fully automated setup was developed, which both controls the instruments for adjusting the sensing element environment, medium gas concentration and temperature, and collects the sensor response. This allowed to collect enough measurements to train the neural network on real data and to test the sensor performance on other unseen real data.

This work proposes a paradigm shift from the classical description of the response of the sensor through an approximate empirical parametric model to the use of MTL 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 they are too complex to be comfortably described by a mathematical model.

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 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 equation (1). This deviation can be due, for example, to heterogeneities of the micro-environment of the luminescent indicator, or to the presence of static quenching [4]. A scenario that describes this non-linear behavior involves the presence in the substrate of two or more environments, in which the indicator is quenched at different rates [26, 27]. This multi-site model describes the SV curve as the sum of n contributions as

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

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 [27–29].

In most industrial and commercial sensors, the decay time τ is frequently preferred to intensity measurement because of its higher reliability and robustness [30]. 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 implementation 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 [27, 31]. Therefore, in the simplest case of a two-sites scenario, the model can be rewritten in terms of phase shift as [32]

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

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 [33–35] and may result frequency dependent, an artifact of the approximation of the model. Finally, Eq. 3 needs to be inverted to determine $[O_2]$ from the measured quantity θ. The proposed approach defies the need for the mathematical model through the use of neural network approach. However, the structure of Eq. 3 remains relevant to understand the structure of the data and optimize the architecture of the neural network.

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

B. 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, both the software for the instrument control and for the data acquisition was written using the software LabVIEW by National Instruments. The acquisition procedure is described in detail in Sect. B.2.

B.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). To control the temperature of the samples, these were placed in good thermal contact with a copper plate, set in a thermally insulated chamber. The temperature of this plate was adjusted 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 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 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. The luminescence was focused by a lens 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.

B.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
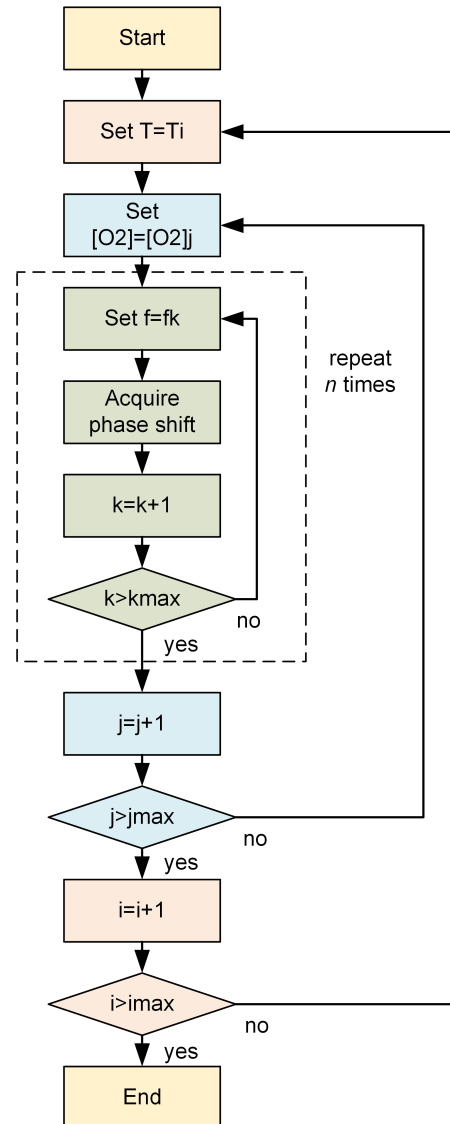

Fig. 2. Flow-chart of the automated data acquisition program.

in 5 °C steps. The total number of measurements was thus 50 x 20 x 21 x 9 = 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.

C. 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 [36]: 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.

C.1. Neural Network Architecture

The artificial network used in this work has a multi-task-learning architecture and is depicted schematically in Fig. 3. It consists of three *common 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, and two with each two additional *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 [25].

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 modulation frequency of the excitation light (see Sec. B). 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 modulation frequency w_i.

C.2. Loss Function

The task-specific loss functions L_i are the mean square error (MSE) defined as

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

where n is the number of observations in the input dataset; $\boldsymbol{y}^{[j]} \in \mathbb{R}^{d_i}$ is the measured value of the desired quantity for the jth observation (indicated as a superscript between square brackets),
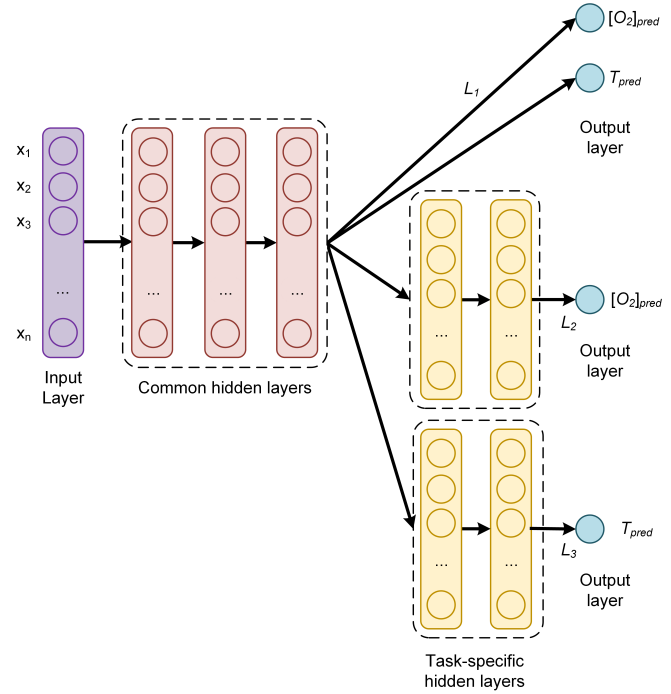

Fig. 3. Architecture of the multi-task learning neural network used in this paper. The common hidden layers generate as output a "shared representation" that is used as input to task specific branches that learn specific features to each quantity and therefore improve the prediction accuracy. $x_1...x_n$ are the input observations; L_i are the task-specific loss functions; $[O_2]_{pred}$ and T_{pred} are the oxygen concentration and temperature predictions of the network respectively.

with $j = 1, ..., n$ for the ith branch and d_i is the dimension of the neural network branch output. In this case $d_1 = 2, d_2 = 1$ and $d_3 = 1$. $\hat{y}^{[j]} \in \mathbb{R}^{d_i}$ is the output of the network branch i, when evaluated on the jth observation. Since there are multiple branches, a global loss function L needs to be 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 hyperparameter tuning phase to optimize the network predictions. In this paper, being the loss function the MSE (Eq. 6), the global loss function results

$$L = \sum_{i=1}^{n_T} \alpha_i \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{d_i} (y_k^{[j]} - \hat{y}_k^{[j]})^2 \tag{8}$$

where n_T is the number of tasks. The global loss function weights used for the plots were $\alpha_1 = 0.3, \alpha_2 = 5$ and $\alpha_3 = 1$ [25].

C.3. Optimiser Algorithm

The loss function was minimized using the optimizer Adaptive Moment Estimation (Adam) [36, 37]. 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 (3780). *Mini-batch training*: with this method the weight update is performed after the network has seen 32 observations. In this case Eq. (6) is used with $n = 32$. For each weight update, 32 random observations are chosen from the training dataset without repetitions until all the training data were fed to the network.

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 [36]. 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. The implementation was performed using the TensorFlow™ library.

D. Performance Evaluation

To evaluate the performance of the sensor, different quantities or metrics were analyzed and are discussed in the next sections.

D.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 T can be predicted by branch 1 and 3. In all the results shown in the following sections, the predicted $[O_2]$ is from branch 2, while the predicted T is from branch 3. For the oxygen concentration for the jth observation $[O_2]^{[j]}$ the *AE* is

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

Where $[O_2]^{[j]}_{pred}$ and $[O_2]^{[j]}_{meas}$ are respectively the $[O_2]$ network prediction and expected value. The further quantity used to analyse 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}, the $MAE_{[O_2]}$ is defined as

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

where $|S_{train}|$ is the size (or cardinality) of the training dataset. AE_T and MAE_T are similarly defined, using T instead of $[O_2]$.

D.2. Kernel Density Estimation

A fundamental quantity to study the performance of the network is the prediction distribution of the *AE*s. This metrics carries information on the probability of the network to predict the expected value. Additionally, the kernel density estimate (KDE) of the distributions of the *AE*s for both the oxygen concentration and the temperature was also calculated. 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 [38]. In this work a Gaussian Kernel and a Scott bandwidth adaptive estimation [39] using the seaborn Python package [40] were used.

D.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* and *MAE*) are useful to compare the performance of different NNM 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, which we will call Error Limited Accuracy (ELA) and that we indicate with η.

Definition.   *In a regression problem, given a metric m, and a certain value of it $\hat{m}$, the ELA for this metric η limited by the error $\hat{m}$ is defined as the number of predictions $\hat{y}$ of the NNM that lie in the range $[m(\hat{y}) - \hat{m}, m(\hat{y}) + \hat{m}]$ divided by the total number of observations. It will be indicated with $\eta(\hat{m})$. Therefore, given a set*

$$E(\hat{m}) = \{\hat{y}^{[i]} \text{ with } i = 1, ..., n \mid |m(\hat{y}^{[i]}) - m(y^{[i]})| \le \hat{m}\} \tag{11}$$

η *is defined as*

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

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

This metric allows interpreting the regression problem as a classification one. $\eta(\hat{m})$ simply describes how many observations are predicted by the NNM within a given value of the chosen metric $\hat{m}$. If m is the *AE* of the predictions, $\eta(AE)$ has a very intuitive and useful interpretation: it is simply the percentage of predictions that are within a certain error $\hat{m}$ from the expected values. Finally, if we take $\hat{m}$ big enough, all the predictions will be classified perfectly, so η is expected to approach 1 for large $\hat{m}$ values. The smaller $\hat{m}$ is, the smaller will be the number of predictions correctly classified. $\eta(\hat{m})$ with $\hat{m} = \hat{AE}$ A special case is the value $m = \overline{m}$ for which $\eta(\overline{m}) = 1$ because within $\overline{m}$ the network would predict all the observations correctly. This value ($\overline{m}$) will give us the biggest error in the sensor predictions.

3. RESULTS AND DISCUSSION

A. Luminescence Experimental Results

As described in Section A, the phase-shift non-linearly depends on the oxygen concentration according to the Stern-Volmer equation. Additionally, the phase shift depends on the temperature,
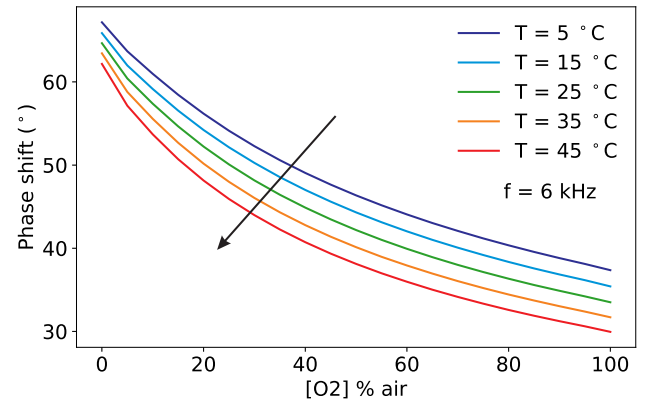

Fig. 4. Measured phase shift as a function of the oxygen concentration for selected temperatures at a fixed modulation frequency of 6 kHz. The arrow marks increasing temperatures.
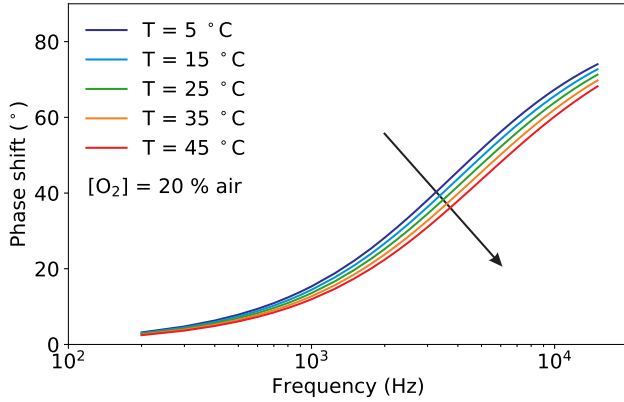
Fig. 5. Measured phase shift as a function of the modulation frequency for selected temperatures at a fixed oxygen concentration of $[O_2] = 20\%$. The arrow marks increasing temperatures.

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

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 are 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 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 in the sample (substrate?).

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. Again, the reduction of the phase shift 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 much stronger affected 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 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 artificial intelligence approach, as it will be shown in the next section.

B. Sensor performance

First, the effect of the training on the sensor performance was investigated. As described in Sect. C.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
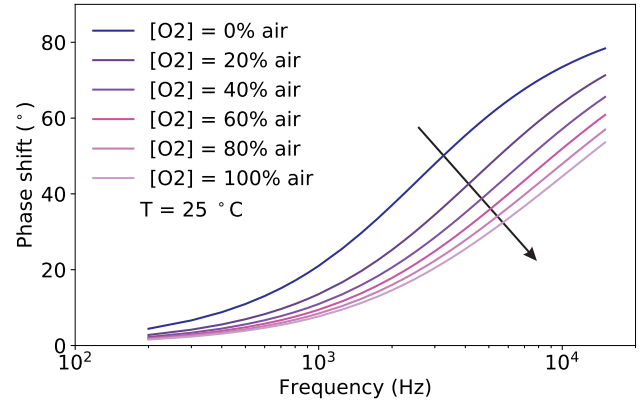
observations θ_s as defined in Eq. (4). The results for $AE_{[O_2]}$ and AE_T are shown in Fig. 7(A) and 7(B), respectively. The blue histogram shows the AE distribution when using no-batch, the gray when using mini-batches of size 32. The KDE profiles help illustrating the features of the histogram. The effect of introducing mini-batches on the performance is extraordinary. The predictions distributions get much narrower, the mean average errors decrease from $MAE_{[O_2]} = 2.38$ % air and $MAE_T = 3.63°$ C to $MAE_{[O_2]} = 1.36$ % air and $MAE_T = 1.63°$ 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 roughly one hour.

The performance can be improved by simply training the network for many more epochs. xxx IS IT EXPECTED? xxx 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 θ_s as defined in Eq. (4). The effect of a 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 a the phase shift using a single indicator; 2) that the prediction have expected errors which are comparable or below the typical accuracy of 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 to train a network for so many epochs requires, on a modern laptop, roughly 5 hours.

To investigate if the training can be performed more efficiently, the normalized phase shift θ_n defined in Eq. (5) where 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 training is only 20'000 the mean average er-
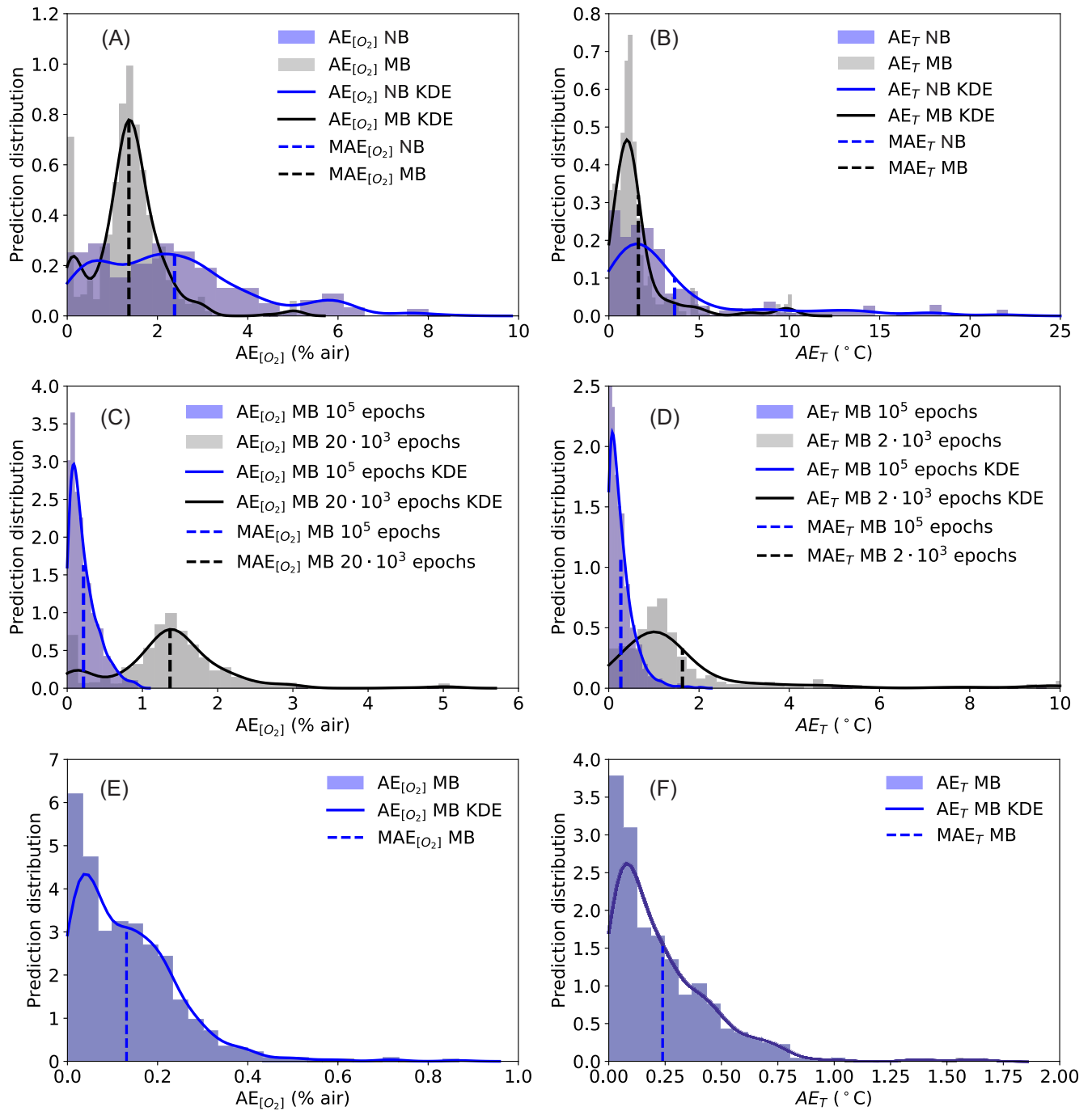
Fig. 7. Performance of the neural network for the oxygen (panels (A), (C) and (E)) and for the temperature (panels (B), (D) and (F)) predictions. In all panels the normalized prediction distribution histogram (columns), the kernel density estimate (KDE) of the distribution of the AEs (solid line), and 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 $\boldsymbol{\theta}_n$.

rors are better than what obtained with $\boldsymbol{\theta}_s$ as input and a training of 100'000 epochs: $MAE_{[O_2]} = 0.13$ % 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. xxxEXPLAIN BETTERxxxx

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

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	2.38 % air	3.63 °C
$\boldsymbol{\theta}_s$	20'000 / 32	1.36% air	1.63 °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

C. Error Limited Accuracy Plots

The discussion in the previous sections helps in measuring how the network behave and how good the predictions are, but makes understanding what sensor we could build with such model almost impossible. For a sensors one has to be able to estimate the maximum error for the measurements. To achieve this we developed a completely new metric: the ELA (η), as defined in section D.3, that will allow us to characterise the sensor with a maximum error. As explained previously, η is defined only when one chooses a metric m. In this section the metric chosen is $m = AE_{[O_2]}$ for the oxygen concentration and $m = AE_T$ for the temperature.

In Fig. 8(A) and 8(B) you can see a comparison of $\eta(AE_{[O_2]})$ and $\eta(AE_T)$ for network trained with mini-batches of size 32 and for 20'000 and 100'000 epochs. The black line are the results for the network that were trained with $\boldsymbol{\theta}_n$ as input for 20'000 epochs, while the red one with $\boldsymbol{\theta}_s$ as input for 100'000 epochs for the oxygen concentration.

The dashed lines indicates the values of the $\overline{AE_{[O_2]}}$ for which the predictions would give a 100 accuracy within that error. For example 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 would be 0.87 % air.

In Fig. 8(B) the same analysis can be found for $m = \eta(AE_T)$. The interpretation is similar to the one given above for the oxygen concentration predictions. For the network trained with $\boldsymbol{\theta}_s$ as input, the model would predict perfectly all the T values within $\overline{AE_T} = 2.05°C$ error. For the network trained with $\boldsymbol{\theta}_n$ this value would be $\overline{AE_T} = 1.67°C$. Those results are summarised in Table 2.

4. CONCLUSIONS

The results in the prediction of the oxygen concentration and temperature show unprecedented accuracy for both parameters, demonstrating that this approach will make a new generation of

Table 2. Summary of the values of m_1 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	2.05 °C
$\boldsymbol{\theta}_n$	20'000 / 32	0.87% air	1.67 °C

sensors possible for dual or even multiple sensing. The unprecedented accuracy in predicting both $AE_{[O_2]}$ and T at the same time, from the same set of data, will allow sensors to become easier to build, since no separate temperature measurements are necessary anymore. To be able to estimate what kind of sensor one could build with a given NNM, in this paper a new metric, the ELA ($\eta(\hat{m})$), was proposed to be able to estimate how many predictions like within a certain value of the chosen metric from the expected values.

DISCLOSURES

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. M. I. Stich, L. H. Fischer, and O. S. Wolfbeis, "Multiple fluorescent chemical sensing and imaging," Chem. Soc. Rev. **39**, 3102–3114 (2010).
2. S. M. Borisov, R. Seifner, and I. Klimant, "A novel planar optical sensor for simultaneous monitoring of oxygen, carbon dioxide, ph and temperature," Anal. bioanalytical chemistry **400**, 2463–2474 (2011).
3. T. Kameya, Y. Matsuda, Y. Egami, H. Yamaguchi, and T. Niimi, "Dual luminescent arrays sensor fabricated by inkjet-printing of pressure-and temperature-sensitive paints," Sensors Actuators B: Chem. **190**, 70–77 (2014).
4. 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).
5. S. Santoro, A. Moro, C. Portugal, J. Crespo, I. Coelhoso, and J. Lima, "Development of oxygen and temperature sensitive membranes using molecular probes as ratiometric sensor," J. Membr. Sci. **514**, 467–475 (2016).
6. S. Biring, A. S. Sadhu, and M. Deb, "An effective optical dual gas sensor for simultaneous detection of oxygen and ammonia," Sensors. **19**, 5124 (2019).
7. B. B. Collier and M. J. McShane, "Time-resolved measurements of luminescence," J. Lumin. **144**, 180–190 (2013).
8. C. Stehning and G. A. Holst, "Addressing multiple indicators on a single optical fiber-digital signal processing approach for temperature compensated oxygen sensing," IEEE Sensors J. **4**, 153–159 (2004).
9. P. Jorge, C. Maule, A. Silva, R. Benrashid, J. Santos, and F. Farahi, "Dual sensing of oxygen and temperature using quantum dots and a ruthenium complex," Anal. chimica acta **606**, 223–229 (2008).
10. J. P. Moore, C. Higgins, O. McGaughey, B. G. Lawless, and 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, 2006), p. 63770I.
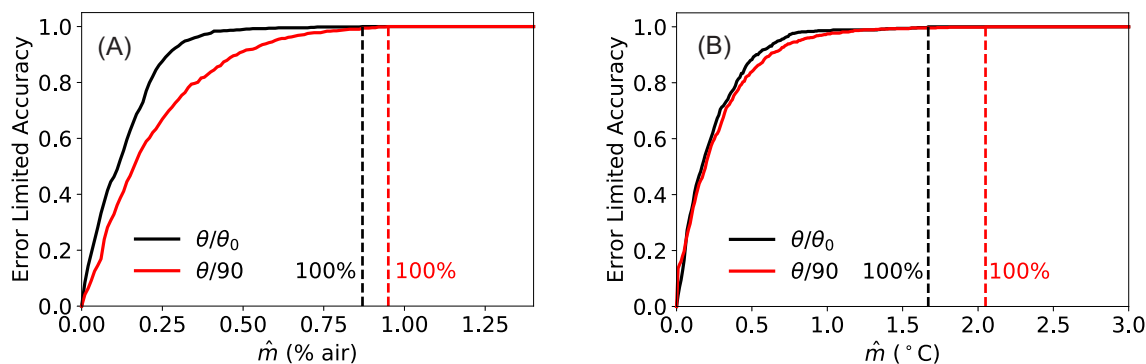
Fig. 8. Comparison of η versus $AE_{[O_2]}$ for network trained with mini-batches of size 32 and for 20'000 and 100'000 epochs. The black line are the results for the network that were trained with $\boldsymbol{\theta}_n$ as input for 20'000 epochs, while the red one with $\boldsymbol{\theta}_s$ as input for 100'000 epochs. The dashed lines indicates the values of the $\overline{AE_{[O_2]}}$ for which the predictions would give a 100 accuracy within that error.

11. D. B. Papkovsky and R. I. Dmitriev, "Biological detection by optical oxygen sensing," Chem. Soc. Rev. **42**, 8700–8732 (2013).

12. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy, 3rd ed.* (Springer, New York, 2006).

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

14. W. Xu, R. C. McDonough, B. Langsdorf, J. Demas, and B. DeGraff, "Oxygen sensors based on luminescence quenching: interactions of metal complexes with the polymer supports," Anal. Chem. **66**, 4133–4141 (1994).

15. S. Draxler, M. E. Lippitsch, I. Klimant, H. Kraus, and O. S. Wolfbeis, "Effects of polymer matrixes on the time-resolved luminescence of a ruthenium complex quenched by oxygen," The J. Phys. Chem. **99**, 3162–3167 (1995).

16. P. Hartmann and W. Trettnak, "Effects of polymer matrices on calibration functions of luminescent oxygen sensors based on porphyrin ketone complexes," Anal. chemistry **68**, 2615–2620 (1996).

17. A. Mills, "Controlling the sensitivity of optical oxygen sensors," Sensors Actuators B: Chem. **51**, 60–68 (1998).

18. D. Badocco, A. Mondin, P. Pastore, S. Voltolina, and 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," Anal. chimica acta **627**, 239–246 (2008).

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

20. S. Thrun, "Is learning the n-th thing any easier than learning the first?" in *Advances in neural information processing systems,* (1996), pp. 640–646.

21. R. Caruana, "Multitask learning," Mach. learning **28**, 41–75 (1997).

22. Y. Zhang and Q. Yang, "A survey on multi-task learning," arXiv preprint arXiv:1707.08114 (2017).

23. J. Baxter, "A model of inductive bias learning," J. artificial intelligence research **12**, 149–198 (2000).

24. K.-H. Thung and C.-Y. Wee, "A brief review on multi-task learning," Multimed. Tools Appl. **77**, 29705–29725 (2018).

25. U. Michelucci and F. Venturini, "Multi-task learning for multi-dimensional regression: application to luminescence sensing," Appl. Sci. **9**, 4748 (2019).

26. 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).

27. 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).

28. 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).

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

30. Y. Wei, Y. Jiao, D. An, D. Li, W. Li, and Q. Wei, "Review of dissolved oxygen detection technology: From laboratory analysis to online intelligent detection," Sensors. **19**, 3995 (2019).

31. M. Quaranta, S. M. Borisov, and I. Klimant, "Indicators for optical oxygen sensors," Bioanal. reviews **4**, 115–157 (2012).

32. U. Michelucci, M. Baumgartner, and F. Venturini, "Optical oxygen sensing with artificial intelligence," Sensors **19**, 777 (2019).

33. V. I. Ogurtsov and D. B. Papkovsky, "Modelling of phase-fluorometric oxygen sensors: Consideration of temperature effects and operational requirements," Sensors Actuators B: Chem. **113**, 917–929 (2006).

34. Y.-L. Lo, C.-S. Chu, J.-P. Yur, and Y.-C. Chang, "Temperature compensation of fluorescence intensity-based fiber-optic oxygen sensors using modified stern–volmer model," Sensors Actuators B: Chem. **131**, 479–488 (2008).

35. N. Zaitsev, P. Melnikov, V. Alferov, A. Kopytin, and K. German, "Stable optical oxygen sensing material based on perfluorinated polymer and fluorinated platinum (ii) and palladium (ii) porphyrins," Procedia Eng. **168**, 309–312 (2016).

36. U. Michelucci, *Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks* (APRESS Media, LLC, 2018).

37. 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).

38. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction* (Springer Science & Business Media, 2009).

39. S. R. Sain and D. W. Scott, "On locally adaptive density estimation," J. Am. Stat. Assoc. **91**, 1525–1534 (1996).

40. 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, and C. Evans, "mwaskom/seaborn: v0.10.0 (january 2020)," (2020).

~