**Model-free dual oxygen and temperature luminescence sensor based on neural networks**

**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*](mailto:francesca.venturini@zhaw.ch)

*Compiled March 27, 2020*

**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 oxy- gen 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 for the sensing quantity, including the influencing effects of the sensor components due to their complexity and their interactions. Typically, this is overcome by using an empirical approximate model where these effect 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 neural networks that is completely model-free. We show how it is possible to extract multiple parameters, in this work the oxygen concentration and temperature, from a single set of optical measurements, without the need of any** *a priori* **mathematical model with an unprecedented accuracy. The results show that the neural network achieves predictions of both parameters, which are comparable to the accuracy of commercial senors and therefore demon- strate the feasibility of this approach. Furthermore, the proposed approach is not limited to oxy- gen and temperature sensing but can be applied to the luminescence of multiple luminophores, whenever the underlying mathematical model is not known or too complex to derive the desired quantities from a single measurement.**

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](https://doi.org/10.1364/OA_License_v1)

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

## INTRODUCTION

The simultaneous determination of multiple physical quanti- ties 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. Optical lumi- nescence 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 ei- ther 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, em- bedded in a substrate and placed in close physical proximity [[1](#_bookmark20)–[6](#_bookmark22)]. To be able to determine each quantity separately it may be necessary to determine more than one optical property (e.g., ab- sorption 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 used luminophores [[4](#_bookmark21), [6](#_bookmark22)–[10](#_bookmark23)].

The problem of dual sensing is particularly relevant in appli- cations that involve oxygen sensing. The determination of oxy- gen 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](#_bookmark21), [11](#_bookmark24)]. 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

## METHODS

* 1. **Luminescence Quenching for Oxygen Determination**

Luminescence-based oxygen sensors usually consist of a dye molecule (luminophore) whose luminescence intensity and de- cay time decrease for increasing O2 concentrations. This re- duction is due to collisions of the excited luminophore with molecular oxygen, which thus provides a radiationless deactiva- tion process (collisional quenching). In the case of homogeneous media characterized by an intensity decay which is a single expo- nential, the decrease in intensity and lifetime are both described by the Stern-Volmer (SV) equation [[12](#_bookmark25)]

measurement of the luminescence of a specific luminophore, whose intensity and decay time are reduced due to collisions with molecular oxygen [[12](#_bookmark25)].

*τ*

*I*0 =

*I*

*τ*0 = 1 + *KSV* · [*O*2] (1)

Sensors based on this principle must rely on approximated empirical models to parametrize the dependence of the mea- sured sensing quantity (e.g., luminescence intensity or decay time) on influencing factors. Among these, temperature is the parameter with the strongest influence since both the lumines- cence 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 oxy- gen concentration [[13](#_bookmark26)]. 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 sens-

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, *KSV* the Stern–Volmer constant and [*O*2] indicates the oxygen concentration.

For practical applications, the luminophore needs to be em-

bedded in a supporting substrate, frequently a polymer. As a result, the SV curve deviates from the linear behavior of equation ([1](#_bookmark1)). This deviation can be due, for example, to heterogeneities of the micro-environment of the luminophore, or to the presence of static quenching [[4](#_bookmark21)]. A scenario that describes this non-linear behavior involves the presence in the substrate of two or more environments, in which the lumineschence is quenched at differ- ent rates [[27](#_bookmark32), [28](#_bookmark33)]. This multi-site model describes the SV curve as the sum of *n* contributions as

∑ 1 + *KSV*,*i* · [*O*2]

ing element was fabricated and on the sensor itself [[14](#_bookmark27)–[19](#_bookmark28)].

*i*=1

(2)

*I*0 *n*

*I* =

*fi* l−1

In this work, we propose a revolutionary approach based on neural networks. 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 neu- ral network was designed and trained to predict both oxygen concentration and temperature simultaneously. This new ap- proach 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](#_bookmark29)–[25](#_bookmark30)] and are flexible enough to be usable in multi-dimensional regressions [[26](#_bookmark31)].

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

where *fi* ’s are the fractions of the total emission for each com- ponent under unquenched conditions, and *KSVi* ’s are the as- sociated 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](#_bookmark33)–[30](#_bookmark34)].

In most industrial and commercial sensors, the decay time *τ* is frequently preferred to intensity measurement because of its higher reliability and robustness [[31](#_bookmark35)]. The determination of the decay time is done most easily in the frequency domain by mod- ulating 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 lumines- cence intensities, it is frequently also used to describe the oxygen dependence of the decay times [[28](#_bookmark33), [32](#_bookmark36)]. Therefore, in the sim- plest case of a two-sites scenario, the model can be rewritten in terms of phase shift as [[33](#_bookmark37)]

tan *θ*0(*ω*, *T*) =( *f* (*ω*, *T*) +

test the sensor performance on other unseen real data.

This work proposes a paradigm shift from the classical de- scription 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

tan *θ*(*ω*, *T*, [*O*2])

1 + *KSV*1(*ω*, *T*) · [*O*2]

1 − *f* (*ω*, *T*) −1 (3)

1

1 + *KSV*2(*ω*, *T*) · [*O*2]

−

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.

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, *KSV*1 and *KSV*2 are the associated Stern–Volmer constants for

# 

N2

air

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 Bright- Line 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, Os- ram). 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.



LED

405nm

PD

SP

Lens

SP

LP

Lens

Sample

Dry

Temperature controller

Gas mixer

Computer

LED Driver

TIA

Lock-in

***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

**Fig. 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.

each component, and [*O*2] indicates the oxygen concentration. It is to be noted that the quantities *θ*0, *f* , *KSV*1, and *KSV*2 are all non-linearly temperature dependent [[34](#_bookmark38)–[36](#_bookmark39)] and may result frequency dependent, an artifact of the approximation of the model. Finally, Eq. [3](#_bookmark2) 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](#_bookmark2) remains relevant to understand the structure of the data and optimize the architecture of the neural network.

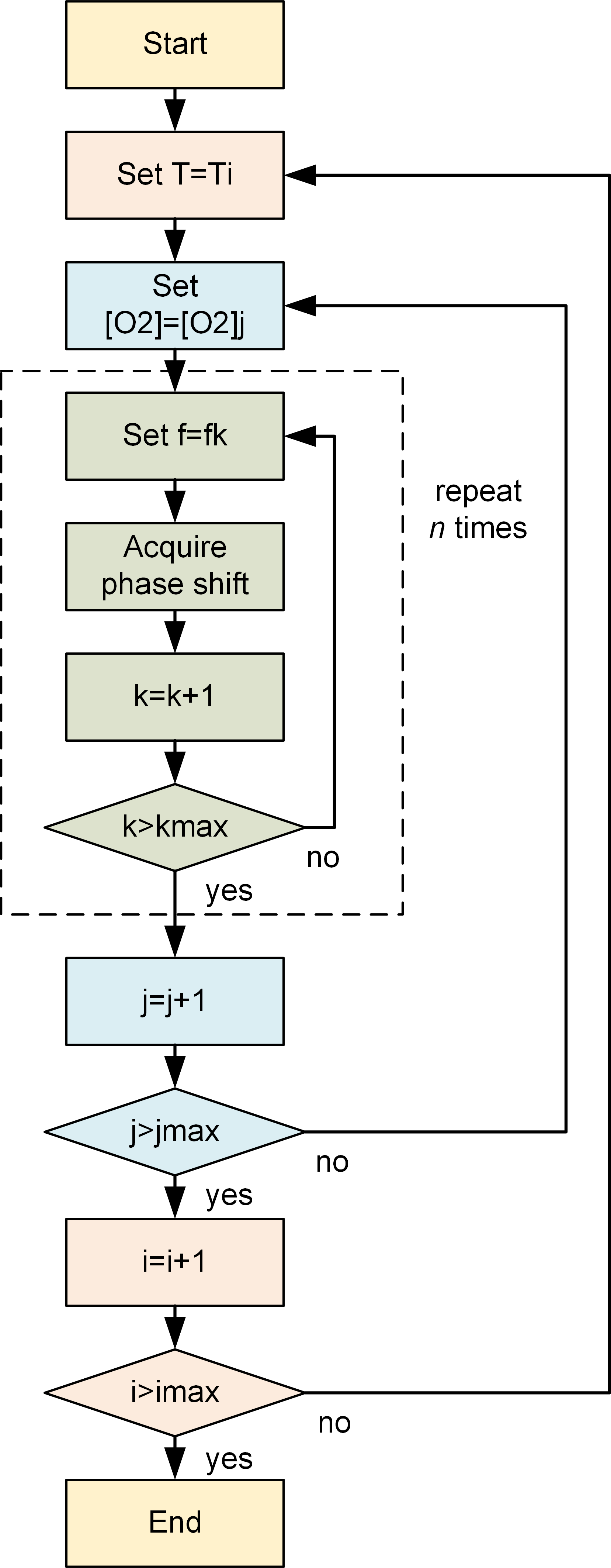
* 1. **Experimental Procedure**

The optical setup used in this work for the luminescence mea- surements is shown schematically in Fig. [1](#_bookmark3). To be able to acquire a large number of data, both the software for the instrument con- trol 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](#_bookmark4).

* + 1. ***Experimental Setup***

The sample used for the characterization and test is a commer- cially 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 concen- tration 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



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

program which followed the flow-chart shown in Fig. [2](#_bookmark6). 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.

* 1. **Neural Network Approach**

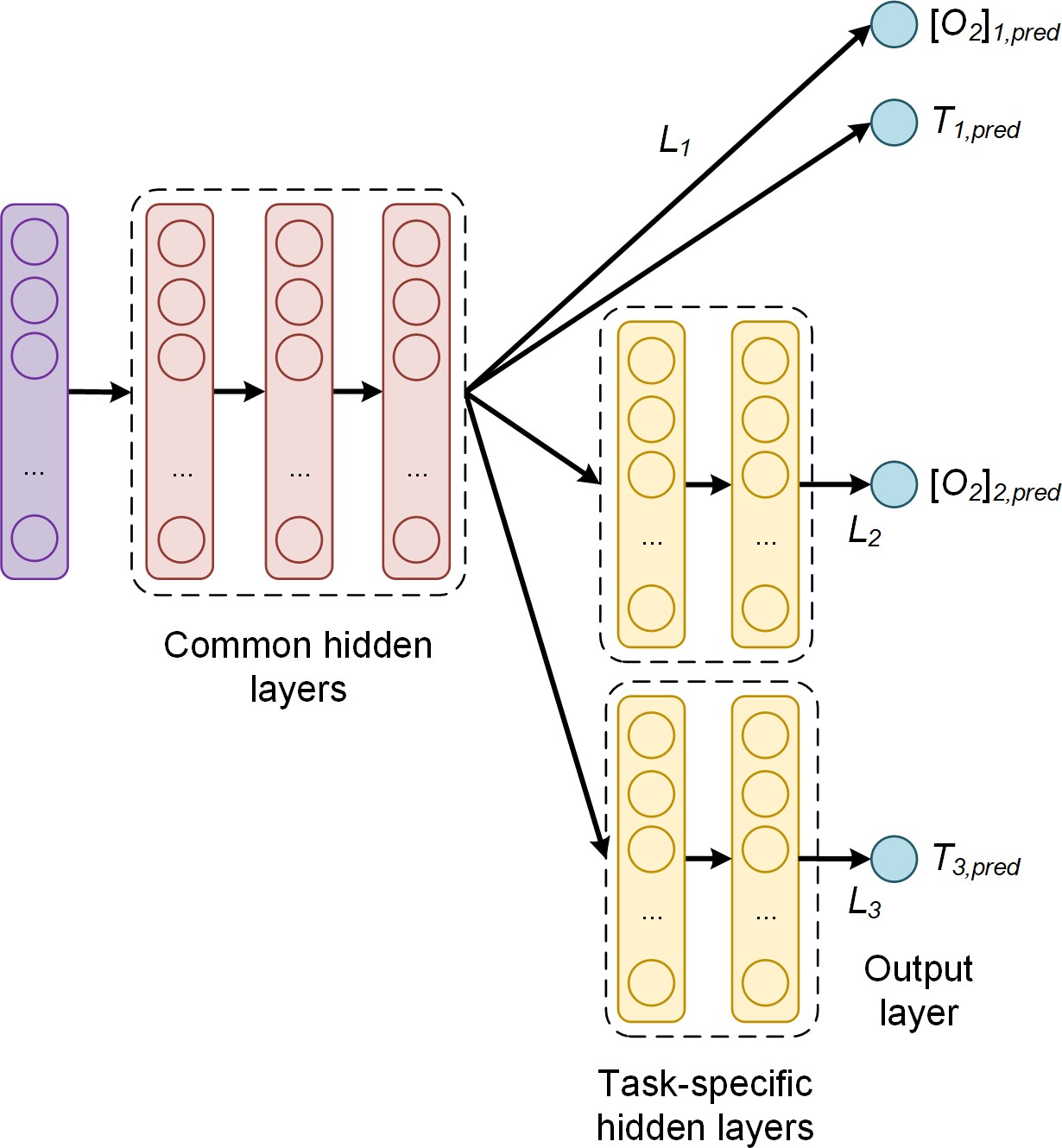
The software component of this new sensor type is based on a neural network model (NNM). A NNM is made of three com- ponents [[37](#_bookmark40)]: a neural network architecture (that includes how neurons are connected, the activation functions and all the hy- perparameters), a loss function (here indicated with *L*) and an optimizer algorithm. In this section, those three components are described in detail.

* + 1. ***Neural Network Architecture***

The neural network used in this work has a multi-task-learning architecture and is depicted schematically in Fig. [3](#_bookmark7). 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 addi- tional *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



**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. *Li* are the task-

specific loss functions; [*O*2]*i*,*pred* and *Ti*,*pred* are the oxygen con-

centration and temperature predictions of the corresponding

branch *i*. Note that branch 2 and 3 have only one output.

* + 1. ***Loss Function***

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

1 *n di*

architecture uses the common hidden layers to find common

*Li* = ∑ ∑(*y*[*j*] − *y*ˆ[*j*] )2, *i* = 1, 2, 3 (6)

features beneficial to each of the two tasks. During the training phase, learning to predict [*O*2] will influence the common hid-

*n j*=1 *k*=1

*k*,*i*

*k*,*i*

den layers and, therefore, the prediction of *T*, and vice-versa.

where *n* is the number of observations in the input dataset;

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

[*j*]

*i*

***y***

∈

the *j*

R*di* is the measured value of the desired quantity for

*th* observation, with *j* = 1, ..., *n* and *di* is the dimension of

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](#_bookmark31)].

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

***θ****s* = ( *θ*(*w*1) , *θ*(*w*2) , ..., *θ*(*w*50) 1 (4)

90

90

90

the neural network branch output. In this case *d*1 = 2, *d*2 = 1 and *d*3 = 1. ***y***ˆ[*j*] ∈ R*di* 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 *αi*

*i*

*nT*

*L* = ∑ *αi Li* . (7)

*i*=1

where *wi* are the 50 values of the modulation frequency of the excitation light (see Sec. [B](#_bookmark5)). In the second case, each observation

is

***θ****n* = ( *θ*(*w*1) , *θ*(*w*2) , ..., *θ*(*w*50) 1 (5)

*θ*0(*w*1)

*θ*0(*w*2)

*θ*0(*w*50)

The parameters *αi* have to be determined during the hyper- parameter tuning phase to optimize the network predictions. In this paper, being the loss function the MSE (Eq. [6](#_bookmark8)), the global loss function is

where *θ* (*w* ) is the value of the phase shift without oxygen

3 1 *n di*

[*j*]

[*j*] 2

0 *i L* = ∑ *αi n* ∑ ∑(*yi*,*k* − *y*ˆ*i*,*k* )

(8)

quenching at the modulation frequency *wi* .

*i*=1

*j*=1 *k*=1

The global loss function weights used for this work were *α*1 = 0.3, *α*2 = 5 and *α*3 = 1. These parameters are the result of a hyper-parameter tuning for this architecture [[26](#_bookmark31)].

* + 1. ***Optimiser Algorithm***

The loss function was minimized using the optimizer Adap- tive Moment Estimation (Adam) [[37](#_bookmark40), [38](#_bookmark41)]. 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](#_bookmark8)) where *n* is the total number of obser- vations 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](#_bookmark8)) 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, al- lowing to reach smaller values of the loss function in less epochs, but usually takes more time [[37](#_bookmark40)]. In our experiments for 20 103 epochs No-batch training took roughly five minutes on a mod- ern 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.

* 1. ***Kernel Density Estimation***

A fundamental quantity to study the performance of the net- work 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 con- centration 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 dis- tribution based on a finite data sample [[39](#_bookmark42)]. In this work a Gaussian Kernel and a Scott bandwidth adaptive estimation [[40](#_bookmark43)] using the seaborn Python package [[41](#_bookmark44)] were used.

* 1. ***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 per- formance 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, which we will call Error Limited Accuracy (ELA) and that we indicate with *η*.

**Definition.** *In a regression problem, given the metric AE, and a certain value of it A*ˆ*E, the ELA η limited by the error A*ˆ*E is defined as the number of predictions y*ˆ *of the NNM that lie in the range y*ˆ *y A*ˆ*E, with y the expected value, divided by the total number*

| − | ≤

*of observations. It will be indicated with η*(*A*ˆ*E*)*. In more mathematical*

*terms, given the set*

*E*(*A*ˆ*E*) = {*y*ˆ[*i*] *with i* = 1, ..., *n* | |*y*ˆ[*i*] − *y*[*i*]| ≤ *A*ˆ*E*} (11)

*η*(*A*ˆ*E*) *is defined as*

***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 dif-

*η*(*A*ˆ*E*) =

|*E*(*A*ˆ*E*)| (12)

*n*

ference between the predicted and the expected value for a given observation. Note that in the architecture described in the previ- ous sections, only branch 1 and 2 can predict [*O*2], while *T* can be predicted by branch 1 and 3. In what follows, the predicted

[*O*2] will always be from branch 2, while the predicted *T* from branch 3. For the oxygen concentration for the *jth* observation [*O*2][*j*] the *AE* is

*where E*(*A*ˆ*E*) *is the cardinality of the set E*(*A*ˆ*E*) *or in other words, the number of its elements.*

This metric allows interpreting the regression problem as a classification one. *η*(*A*ˆ*E*) simply describes how many observa- tions are predicted by the NNM within a given value of the

| |

absolute error. In other words it is simply the percentage of predictions that are within a certain error *A*ˆ*E* from the expected

values. Finally, if we take *A*ˆ*E* big enough, all the predictions will

*pred*

[*j*]

*AE*

[*O*2]

= |[*O*2][*j*]

[*j*]

[ 2]*meas*

− *O*

|. (9)

be classified perfectly, so *η*(*A*ˆ*E*) is expected to approach 1 for large *A*ˆ*E* values. The smaller *A*ˆ*E* is, the smaller will be the num-

Where [*O*2][*j*] and [*O*2][*j*] 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 pre-

*pred*

*meas*

diction using the training dataset *Strain*, the *MAE*[*O*2 ] is defined as

ber of predictions correctly classified. *η*(*A*ˆ*E*). A special case is the value *AE* = *AE* for which *η*(*AE*) = 1 because within *AE* the network would predict all the observations correctly. This value

(*AE*) will give us the biggest error in the sensor predictions.

## RESULTS AND DISCUSSION

* 1. **Luminescence Experimental Results**

As described in Section [A](#_bookmark0), the phase-shift non-linearly depends

*MAE*

1 [*j*]

[*j*]

on the oxygen concentration according to the Stern-Volmer equa-

[*O*2](*Strain* ) = *S* ∑

| |*train*

*j*∈*Strain*

|[*O*2]*pred* − [*O*2]*real* | (10)

tion. Additionally, the phase shift depends on the temperature, which influences the luminescence and the collision mechanisms, and on the modulation frequency of the excitation light as de-

where *Strain* is the size (or cardinality) of the training dataset.

| |

*AET* and *MAET* are similarly defined, using *T* instead of [*O*2].

scribed in Eq. [3](#_bookmark2). The measured behaviour of the phase shift for variations of these three quantities are shown in the Figs. [4](#_bookmark13) to [6](#_bookmark15).

Fig. [4](#_bookmark13) 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 reduc- tion 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](#_bookmark14), 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](#_bookmark15). Compared to Fig. [5](#_bookmark14), 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](#_bookmark13) to [6](#_bookmark15) show how similar the curves of the phase shift may look for different values of oxygen, tem- perature 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](#_bookmark2)). 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.

* 1. **Sensor performance**

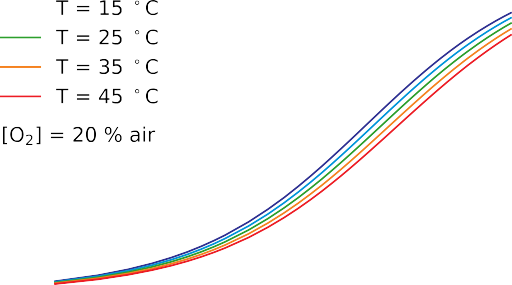
First, the effect of the training on the sensor performance was investigated. As described in Sect. [C.3](#_bookmark11), the neural network was trained with no-batches and with mini-batches. For this compar- ison the network was trained for 20’000 epochs using the input observations ***θ****s* as defined in Eq. ([4](#_bookmark9)). The results for *AE*[*O*2 ] and

*AET* are shown in Fig. [7](#_bookmark16)(A) and [7](#_bookmark16)(B), respectively. The blue

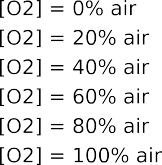
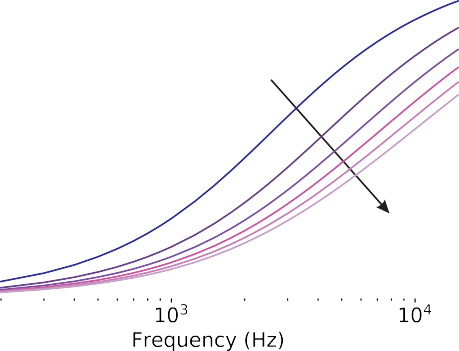
histogram shows the *AE* distribution when using no-batch, the gray when using mini-batches of size 32. The KDE profiles help



**Fig. 4.** Measured phase shift as a function of the oxygen con- centration 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 concen- tration of [*O*2] = 20%. The arrow marks increasing tempera- tures.



**Fig. 6.** Measured phase shift as a function of the modulation frequency for selected oxygen concentrations at a fixed tem- perature of *T* = 25◦C. The arrow marks increasing oxygen

concentrations.

illustrating the features of the histogram. The effect of intro- ducing mini-batches on the performance is extraordinary. The predictions distributions get much narrower, the mean average

errors decrease from *MAE*[*O*2 ] = 2.4 % air and *MAET* = 3.6◦C to *MAE*[*O*2 ] = 1.4 % air and *MAET* = 1.6◦C. Although the per- formance is significantly improved, from Fig. [7](#_bookmark16)(A) and [7](#_bookmark16)(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.

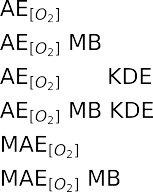
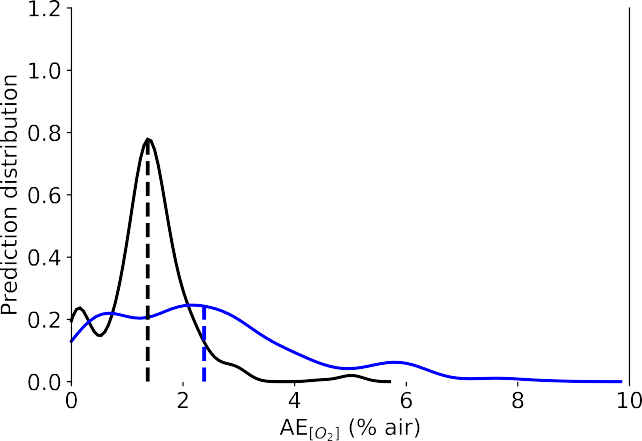
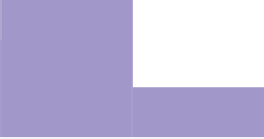
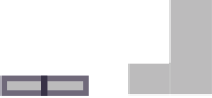
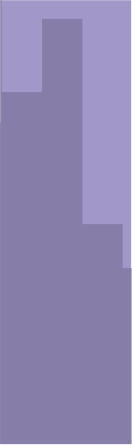
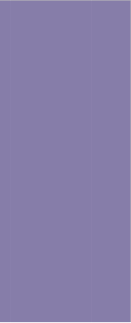
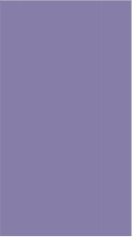
Fig. [7](#_bookmark16)(C) and [7](#_bookmark16)(D) show the comparison between prediction distributions with 20’000 and 100’000 epochs (always using a mini-batch of size 32) using the input observations ***θ****s* as defined in Eq. ([4](#_bookmark9)). The effect of a longer training is a dramatic improve- ment 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 *MAET* = 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](#_bookmark16)(C) and [7](#_bookmark16)(D) demonstrate two new find-

# 

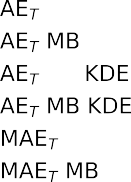
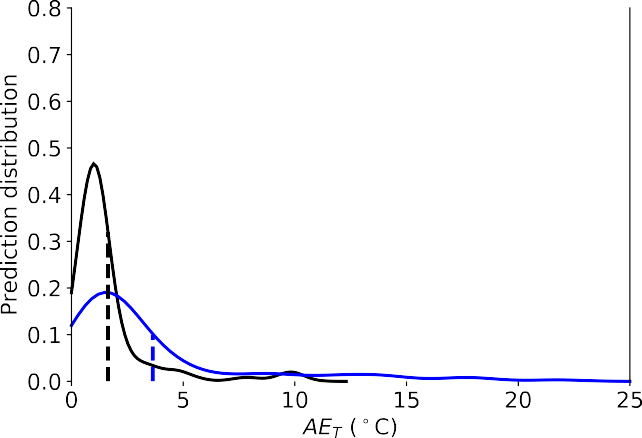
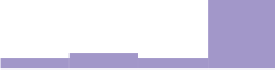
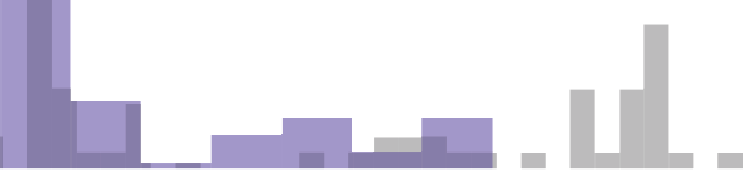
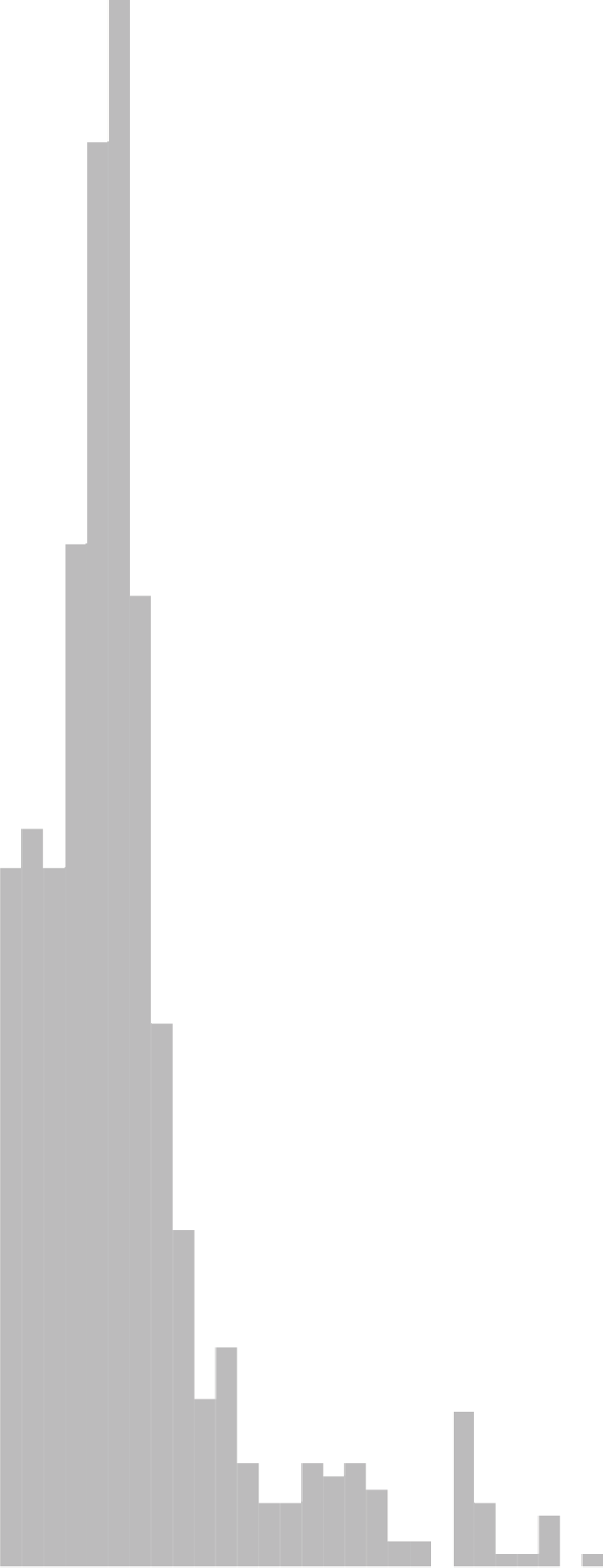


(A)

N

N

N

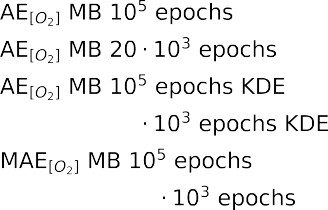
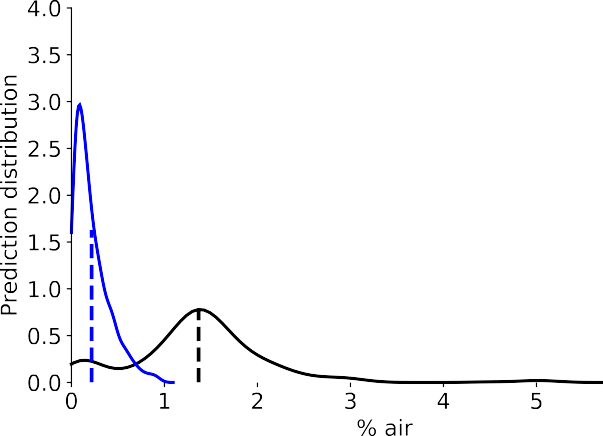
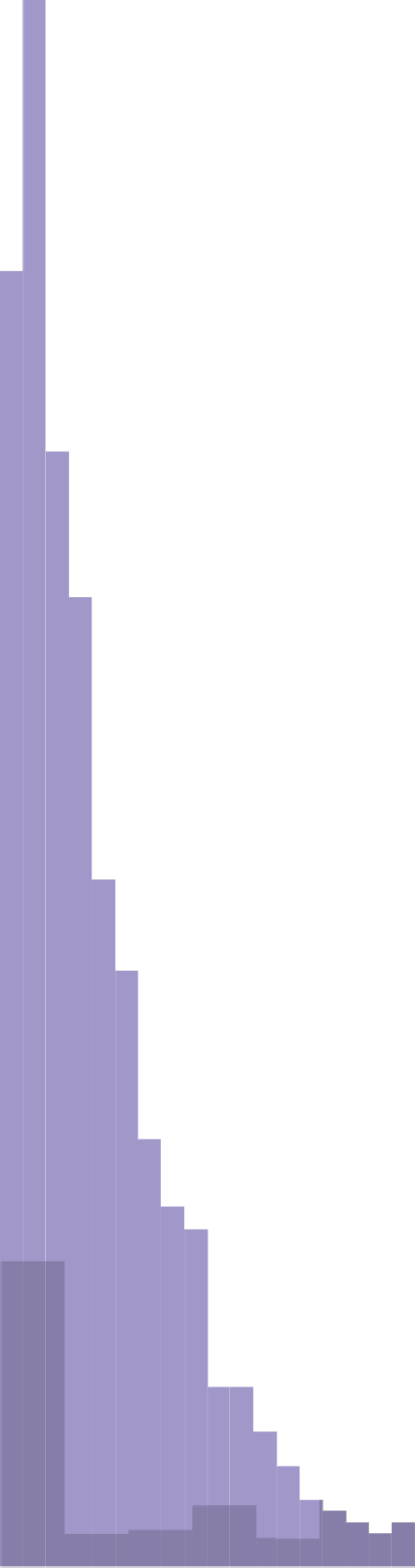
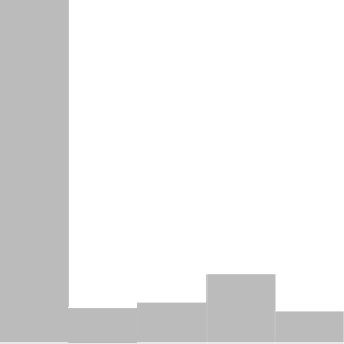


(B)

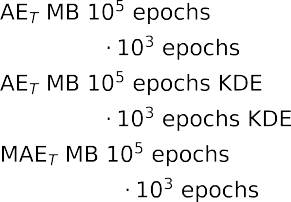
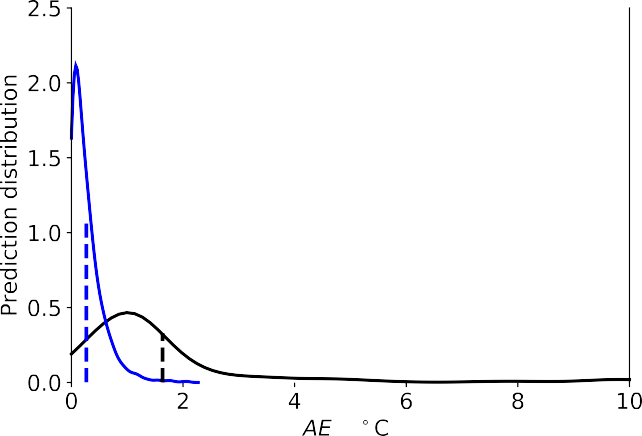
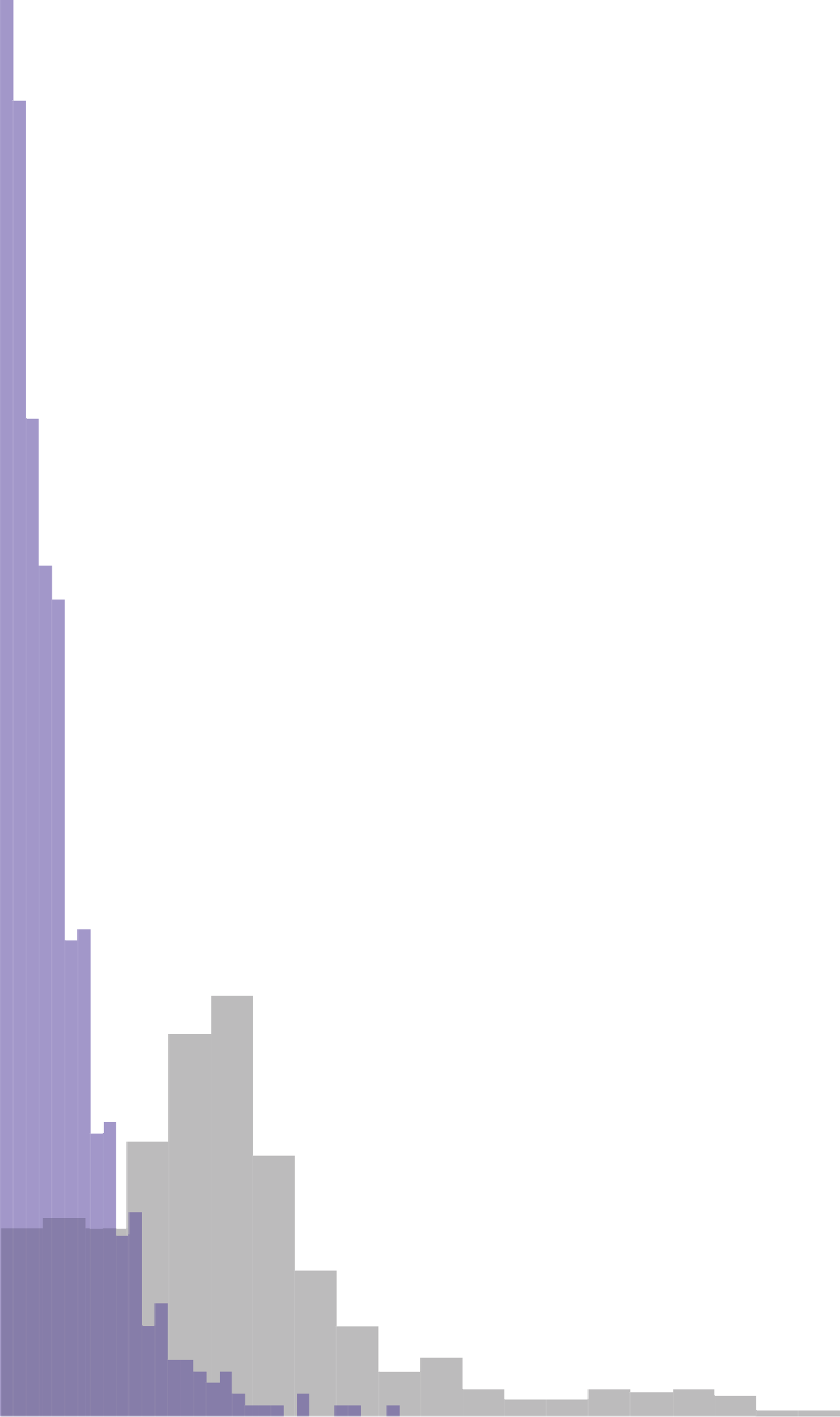
N

N

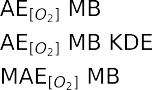
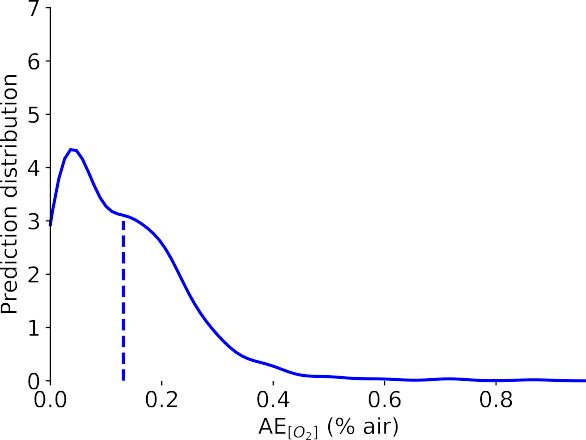
N



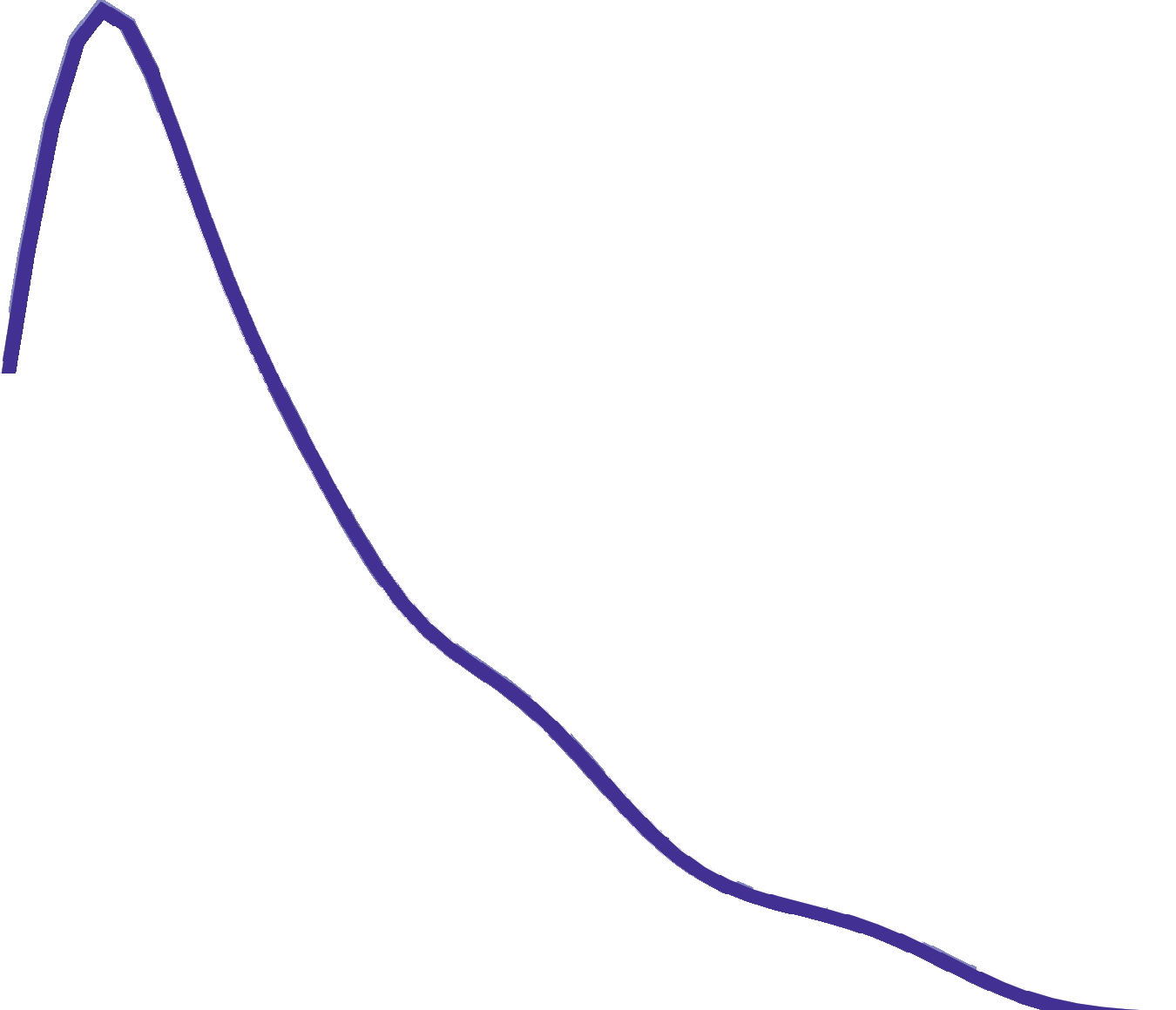
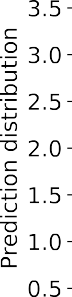
(C)



(D)



(E)



(F)

**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 *AE*s (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 ***θ****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 ***θ****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 ***θ****n*.

ings: 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 luminophore; 2) that the prediction have expected errors which are comparable or below the typical accuracy of com-

### Table 2. Summary of the values of *AE* for the cases shown in Fig. [8](#_bookmark19)(A) and [8](#_bookmark19)(B).

Input Epochs / Batch size *AE*[*O*2 ] *AET*

mercial 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, approximately 5 hours.

To investigate if the training can be performed more effi- ciently, the normalized phase shift ***θ****n* defined in Eq. ([5](#_bookmark10)) 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](#_bookmark16)(E) and [7](#_bookmark16)(F). The performance is further improved; even if the training is only 20’000 the mean average er- rors are better than what obtained with ***θ****s* as input and a training

of 100’000 epochs: *MAE*[*O*2 ] = 0.13 % air and *MAET* = 0.24◦C.

The distributions are also narrower, particularly for the tem-

perature. Additionally all the *AE*[*O*2 ] lie below 0.87 % air, and

*AET* 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 ***θ****n* and in the simplified functional behavior of ***θ****n* compared to ***θ****s* as it may be expected by Eq. [3](#_bookmark2). xxxEXPLAIN BETTERxxxx

The performance of the different neural networks is summa- rized in Table [1](#_bookmark18).

### Table 1. Summary of the performance for neural network models

|  |  |  |  |
| --- | --- | --- | --- |
| Input Epochs / Batch size | | *MAE*[*O*2 ] | *MAE*[*T*] |
| ***θ****s* 20’000 / no batch | | 2.4 % air | 3.6 ◦ *C* |
| ***θ****s* | 20’000 / 32 | 1.4% air | 1.6 ◦ *C* |
| ***θ****s* | 100’000 / 32 | 0.22 % air | 0.27 ◦ *C* |
| ***θ****n* | 20’000 / 32 | 0.13 % air | 0.24 ◦ *C* |

* 1. **Error Limited Accuracy Plots**

The metrics discussed in the previous sections are useful to com- pare the network performance and to measure how good the predictions are. However they do not offer an understanding on what a sensor build with such a model could achieve. For practical applications, the relevant question is what is the maxi- mum error which the sensor will have in predicting the oxygen concentration and temperature. To answer this question we in- troduce a completely new metric: the ELA (*η*), as defined in Sect.

[D.3](#_bookmark12). 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* = *AET* for the temperature.

This new metrics will allow the determination of the maximum

error of the sensor.

Fig. [8](#_bookmark19) displays the ELA *η*(*AE*) for oxygen concentration (A) and for the temperature (B). In each panel the results obtained using the input ***θ****n* and a training for 20’000 epochs are shown in black, and the results obtained using the input ***θ****s* and a training for 100’000 epochs in red. In both cases the training was per- formed with mini-batches of size 32. The dashed lines indicate

the values of the *AE*[*O*2 ] and *AET* for which the error limited accuracy *η* equals 1. In other words, all the predictions will have

an error equal or smaller than *AE*. The values of *AE*[*O*2 ] and

*AET* are summarized in Table [2](#_bookmark17).

***θ****s* 100’000 / 32 0.95 % air 2.1 ◦ *C*

***θ****n* 20’000 / 32 0.87% air 1.7 ◦ *C*

Fig. [8](#_bookmark19)(A) shows that, for the network trained with ***θ****s* as input, the model would predict perfectly all the oxygen concentrations within 0.95 % air error. For the network trained with ***θ****n* this value is futher reduced to 0.87 % air. This can be interpreted as the accuracy a sensor based on this NNM would have. Fig. [8](#_bookmark19)(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 predictions. For the network trained with ***θ****s* as input, the model would predict perfectly

all the temperature values within *AET* = 2.1◦C error. For the

network trained with ***θ****n* this value would be *AET* = 1.7◦C.

1. **CONCLUSIONS**

In this work, a new model-free approach to optical sensing was described. This have the power of solving the problem of extracting multiple separate physical quantities at the same time from a single dataset without any *a priori* mathematical model of the measuring effect or sensor components. This type of problems in physics can be challenging or impossible to solve if the mathematical models describing the functional dependence of the dependent variable from a set of independent variables are too complex or unknown.

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 sensors possible for dual or even multiple sensing. The unprece-

dented 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 and cheaper to build since no separate temperature mea- surements are necessary anymore. To be able to estimate what kind of sensor one could build with a given NNM a new metric, the ELA (*η*(*AE*)), was proposed to be able to estimate how many predictions lies within a certain value of the absolute error from the expected values. This new metric allows to give a maximum

measurement error of the NNM results, effectively transforming a regression problem in a classification one.

This work opens the road to a complete new optical sens- ing methodology that will allow a new generation of sensors to be built. Those sensors will be able to extract multiple phys- ical 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 have the potential of revolutionising optical sensing.

## DISCLOSURES

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

# 



(A)

*n*

*s*



(B)

*n*

*s*

**Fig. 8.** Comparison of the ELA *η*: Panel (A) oxygen prediction, panel (B) temperature prediction. The black line are the results for the network that were trained with ***θ****n* as input for 20’000 epochs with mini-batchs of size 32, while the red one with ***θ****s* as input for 100’000 epochs with mini-batchs of size 32. The dashed lines indicates the values of the *AE* for which the predictions would give *η* = 1.



## 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 pla- nar 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. Ni- imi, “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 sens- ing 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 measure- ments of luminescence,” J. Lumin. **144**, 180–190 (2013).
  8. C. Stehning and G. A. Holst, “Addressing multiple indi- cators 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).

* 1. 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, Chemi- cal, and Biological Sensing Technologies IV,* vol. 6377 (Interna- tional Society for Optics and Photonics, 2006), p. 63770I.
  2. D. B. Papkovsky and R. I. Dmitriev, “Biological detection by optical oxygen sensing,” Chem. Soc. Rev. **42**, 8700–8732 (2013).
  3. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy, 3rd ed.*

(Springer, New York, 2006).

* 1. 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).
  2. W. Xu, R. C. McDonough, B. Langsdorf, J. Demas, and

B. DeGraff, “Oxygen sensors based on luminescence quenching: interactions of metal complexes with the poly- mer supports,” Anal. Chem. **66**, 4133–4141 (1994).

* 1. 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 oxy- gen,” The J. Phys. Chem. **99**, 3162–3167 (1995).
  2. 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).
  3. A. Mills, “Controlling the sensitivity of optical oxygen sen- sors,” Sensors Actuators B: Chem. **51**, 60–68 (1998).
  4. D. Badocco, A. Mondin, P. Pastore, S. Voltolina, and

S. Gross, “Dependence of calibration sensitivity of a poly- sulfone/ru (ii)-tris (4, 7-diphenyl-1, 10-phenanthroline)- based oxygen optical sensor on its structural parameters,” Anal. chimica acta **627**, 239–246 (2008).

* 1. F. Dini, E. Martinelli, R. Paolesse, D. Filippini, A. D’Amico,
     1. 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).
  2. P. M. Argyriou A., Evgeniou T., “Multi-task feature learn- ing.” In Proc. 19th Int. Conf. on Neural Inf. Process. Syst. (NIPS’06) (2006).
  3. S. Thrun, “Is learning the n-th thing any easier than learn- ing the first?” in *Advances in neural information processing systems,* (1996), pp. 640–646.
  4. R. Caruana, “Multitask learning,” Mach. learning **28**, 41–75 (1997).
  5. Y. Zhang and Q. Yang, “A survey on multi-task learning,” arXiv preprint arXiv:1707.08114 (2017).
  6. J. Baxter, “A model of inductive bias learning,” J. artificial intelligence research **12**, 149–198 (2000).
  7. K.-H. Thung and C.-Y. Wee, “A brief review on multi-task learning,” Multimed. Tools Appl. **77**, 29705–29725 (2018).
  8. U. Michelucci and F. Venturini, “Multi-task learning for multi-dimensional regression: application to luminescence sensing,” Appl. Sci. **9**, 4748 (2019).
  9. E. Carraway, J. Demas, B. DeGraff, and J. Bacon, “Photo- physics and photochemistry of oxygen sensors based on luminescent transition-metal complexes,” Anal. chemistry **63**, 337–342 (1991).
  10. J. N. Demas, B. DeGraff, and W. Xu, “Modeling of lumines- cence quenching-based sensors: comparison of multisite and nonlinear gas solubility models,” Anal. Chem. **67**, 1377– 1380 (1995).
  11. P. Hartmann, M. J. Leiner, and M. E. Lippitsch, “Lumines- cence quenching behavior of an oxygen sensor based on a ru (ii) complex dissolved in polystyrene,” Anal. Chem. **67**, 88–93 (1995).
  12. A. Mills, “Response characteristics of optical sensors for oxygen: models based on a distribution in *τ*oor kq,” Ana- lyst. **124**, 1301–1307 (1999).
  13. 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).
  14. M. Quaranta, S. M. Borisov, and I. Klimant, “Indicators for optical oxygen sensors,” Bioanal. reviews **4**, 115–157 (2012).
  15. U. Michelucci, M. Baumgartner, and F. Venturini, “Optical oxygen sensing with artificial intelligence,” Sensors **19**, 777 (2019).
  16. 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).
  17. 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,” Sen- sors Actuators B: Chem. **131**, 479–488 (2008).
  18. N. Zaitsev, P. Melnikov, V. Alferov, A. Kopytin, and K. Ger- man, “Stable optical oxygen sensing material based on per- fluorinated polymer and fluorinated platinum (ii) and pal- ladium (ii) porphyrins,” Procedia Eng. **168**, 309–312 (2016).
  19. U. Michelucci, *Applied Deep Learning - A Case-Based Ap- proach to Understanding Deep Neural Networks* (APRESS Me- dia, LLC, 2018).
  20. 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).
  21. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*(Springer Science & Business Media, 2009).
  22. S. R. Sain and D. W. Scott, “On locally adaptive density estimation,” J. Am. Stat. Assoc. **91**, 1525–1534 (1996).
  23. 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).

~