PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Dual oxygen and temperature sensing with single indicator using multi-task-learning neural networks

Francesca Venturini, Umberto Michelucci, Michael Baumgartner

Francesca Venturini, Umberto Michelucci, Michael Baumgartner, "Dual oxygen and temperature sensing with single indicator using multi-task-learning neural networks," Proc. SPIE 11354, Optical Sensing and Detection VI, 113541C (1 April 2020); doi: 10.1117/12.2554941

SPIE.

Event: SPIE Photonics Europe, 2020, Online Only

Invited Paper

Dual oxygen and temperature sensing with single indicator using multi-task-learning neural networks

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

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

ABSTRACT

The optical determination of oxygen partial pressure is of great interest in numerous areas, like medicine, biotechnology, and chemistry. A well-known optical measuring approach is based on the quenching of luminescence by the oxygen molecules. The conventional approach consists in measuring the intensity decay time and relate it to the oxygen concentration through a multi-parametric model (Stern–Volmer equation). The parameters of this equation 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. This work proposes a new approach based on a multi-task learning (MTL) neural network. Using the luminescence data of one single indicator, which is sensitive to both oxygen and temperature, the neural network achieves predictions of both parameters which are comparable to the accuracy of commercial senors. The impact of the new proposed approach is however not limited to dual oxygen and temperature sensing, but can be applied to all those cases in which the sensor response is too complex, to be comfortably described by a mathematical model.

Keywords: luminescence sensor; oxygen sensor; temperature sensor; dual sensing; luminescence quenching; artificial intelligence; neural network; machine learning

1. INTRODUCTION

The determination of oxygen partial pressure is of great interest in numerous areas, like medicine, biotechnology, and chemistry. Among the different methods used to determine the oxygen concentration, optical techniques are particularly attractive because they do not consume oxygen, have a fast response time, allow a good precision and accuracy. A well-known optical measuring approach is the quenching of luminescence by the oxygen molecules. The measuring principle is based on the measurement of the luminescence of a specific molecule, called luminophore, whose intensity and the decay time are reduced due to collisions with oxygen.[1,2]

Sensors based on this principle rely on approximate empirical models to parametrise the dependence of the luminescence intensity or the decay time on influencing factors. The most relevant factor is the temperature since both the luminescence and the quenching phenomena are strongly temperature-dependent. The conventional approach consists in measuring the intensity decay time and relate it to the oxygen concentration through a multi-parametric model (Stern–Volmer equation). The parameters of this equation are, however, all temperature-dependent.[3–5] Therefore the temperature needs to be known to correct the oxygen concentration and is therefore measured separately, either optically or with a completely different sensor.

The relevant parameters when using luminescence quenching are extracted normally with non-linear regression, where the mathematical model of the Stern-Volmer equation describes the functional dependence of the dependent variable from a set of independent variables. A new approach to solve these difficulties is to use feed-forward neural networks to predict the desired variables.[6] Unfortunately, feed-forward neural networks usually perform less efficiently when applied to multi-dimensional regression problems. In other words, this type of networks being less flexible max have difficulties in predicting simultaneously multiple variables that depend from the input dataset in fundamentally different ways. Multi-task learning (MTL) neural network architectures,

Further author information: (Send correspondence to F.V.)
F.V.: E-mail: francesca.venturini@zhaw.ch

Optical Sensing and Detection VI, edited by Francis Berghmans, Anna G. Mignani, Proc. of SPIE
Vol. 11354, 113541C · © 2020 SPIE · CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2554941

on the other hand, learn commonalities and differences across tasks by using multiple branches of task-specific layers, which have as input the output of a common set of layers. Depending on how information is encoded in the data, these architectures may results in improved learning efficiency and prediction accuracy.[7–10]

In this paper, we propose a new multi-task learning (MTL) neural network approach applied to luminescence sensing, where the luminescence data of one single indicator are dependent on two quantities, oxygen concentration and temperature, which are otherwise hard to extract separately. However, the proposed approach is not limited to dual oxygen and temperature sensing but can be applied to all those cases in which the system is too complex, to be comfortably described by a mathematical model.

2. THEORETICAL MODEL FOR OXYGEN LUMINESCENCE QUENCHING

Oxygen-quenching luminescence sensors are based on the measurement of the decrease of the luminescence intensity and decay time of the luminophore due to collisional quenching with molecular oxygen. In the case of homogeneous media characterised by an intensity decay which is a single exponential, the decrease in intensity and lifetime are both described by the Stern-Volmer equation[1]

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

where I_0 and I are the luminescence intensities in the absence and presence of oxygen respectively, 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.

If the luminophore is embedded in a matrix or substrate, which is mostly the case when building a sensor, the Stern-Volmer curve deviates from the linear behavior of Eq. (1).[2] This behavior is attributed, for example, to heterogeneities of the micro-environment of the luminescent indicator and can be described by several different proposed models. One simple scenario involves the presence of at least different environments, in which the luminophore is quenched at different rates, often referred to as the multi-site model.[11,12] In this case, the Stern-Volmer curve is written as sum contributions

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

where f_i's are the fractions of the total emission for each component under unquenched conditions, K_{SVi}'s are the associated effective Stern–Volmer constants. The sum is performed over the number of sites.

Rather than measuring the intensity, the decay time is a preferred quantity to determine the oxygen concentration. The decay time is easily measured in the frequency domain by modulating the intensity of the excitation. The emitted luminescence is also modulated but shows a phase shift θ, which depends on the decay time. It was shown that the phase shift θ is related to the oxygen concentration $[O_2]$ and temperature T for a two-site model through the approximate equation[6]

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

where θ_0 and θ, respectively, are the phase shifts in the absence and presence of oxygen, ω is the angular frequency of the modulation of the excitation light. Since both the luminescence and the quenching of luminescence are strongly influenced by the temperature, the parameters θ_0, K_{SV1}, K_{SV2}, and f need to be modeled through different temperature dependencies. Finally, in the sensor firmware the Eq. (3) must be inverted to obtain $[O_2]$ from θ, once T and ω are known.

Differently from the conventional approach, which relies on the measurement of the temperature, then used to correct the parameters of the analytical model, this paper describes an approach which overcomes these difficulties by using a neural network which learns to extract these inter-dependencies. To train the network, a large enough amount of data is needed, where θ is measured for varying T and ω.

3. EXPERIMENTAL SETUP

Fig. 1 presents a schematic diagram of the experimental setup used for the acquisition of the luminescence data. The sample used for the characterization and test is a commercially available Pt-TFPP-based oxygen sensor spot (PSt3, PreSens Precision Sensing GmbH). To control the temperature of the sample, this were placed in good thermal contact with a copper plate, placed in a thermally insulated chamber. The temperature of this plate was adjusted between 0 °C and 45 °C using a Peltier element and stabilized with a temperature controller (PTC10, Stanford Research Systems). The oxygen concentration was adjusted using a self-made gas-mixing apparatus which mixed dry air and N_2. In the following, the concentration of oxygen will be given in % of the oxygen concentration of the dry air and indicated with % air.

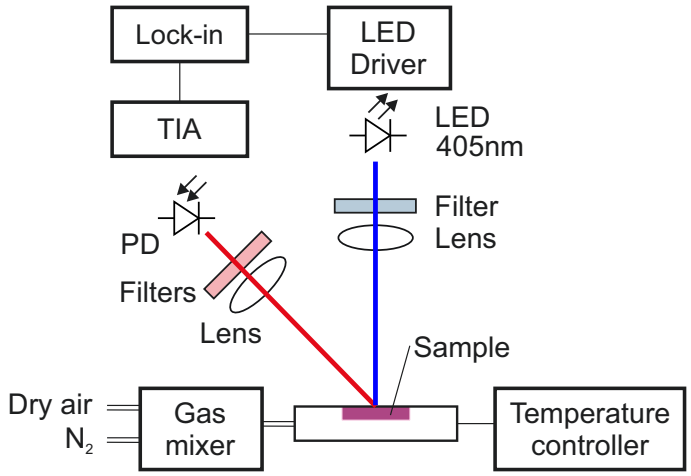

Figure 1. Schematic diagram of the optical experimental setup. Blue is the excitation, red the luminescence optical path. PD: photodiode; TIA: trans-impedance amplifier.

The excitation light was provided by a 405 nm LED (VAOL-5EUV0T4, VCC Visual Communications Company,), filtered by a short pass 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 focused by a lens was collected by a photodiode (SFH 213 Osram) after filtering with a long pass filter with cut-off at 594 nm (Semrock 594 LP Edge Basic long pass) and a short pass filter with cut-off at 682 nm (Semrock 682 SP Bright Line HC short pass). The driver for the LED and the trans-impedance amplifier (TIA) are self-made. For the frequency generation and the phase detection a two-phase lock-in amplifier (SR830, Stanford Research Inc.) was used. The modulation frequency was varied between 200 Hz and 15 kHz.

4. ARTIFICIAL INTELLIGENCE APPROACH

The architecture of the multi-task neural network is shown in Fig. 2. The idea behind this MTL architecture is that the *common hidden layers* generate as output a shared representation, that is then used as input to *task-specific hidden layers*. While the common hidden layers learn features that are common to both parameters, oxygen concentration and temperature, the two task-specific branches learn features specific to each quantity and therefore improve the prediction accuracy. In this work, a network with three common hidden layers of 50 neurons each and two task-specific hidden layers of five neurons each was used. The sigmoid activation functions was used for all the neurons. This architecture was chosen because it was previously proved on synthetic data to be working optimally for this specific problem.[13] The input to the network is a vector of 50 values of $\theta(w_i)/\theta_0(w_i)$ where w_i with $i = 1, ..., 50$ are the modulation frequencies.

Multiple cost functions L_i with $i = 1, ..., n_T$, with n_T the number of tasks, are required to use this network architecture (see Fig. 2). The cost functions L_i are the mean square error. The global cost function L is defined

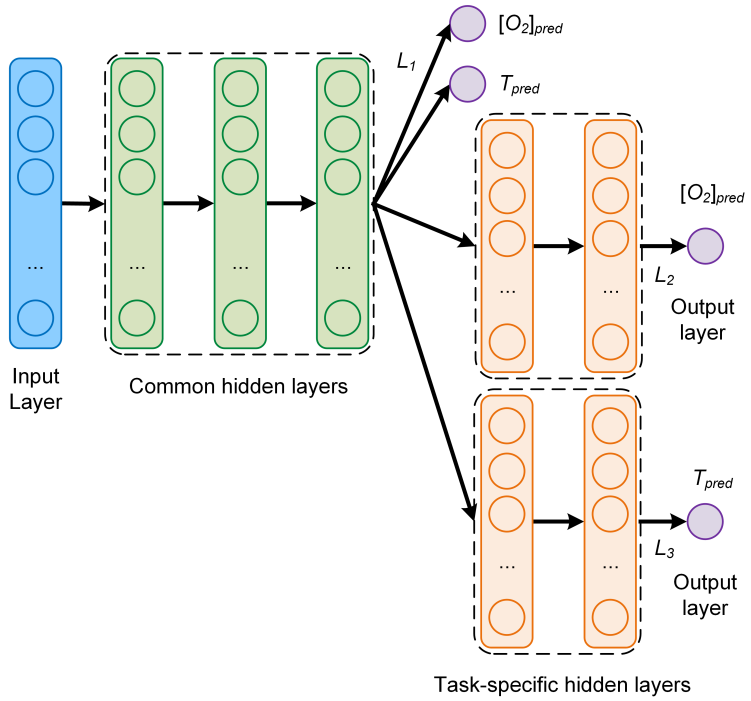

Figure 2. Architecture of the multi-task learning neural network. L_i indicate the cost functions.

as a linear combination of the task-specific cost functions with weights α_i

$$L = \sum_{i=1}^{n_T} \alpha_i \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{d} (y_{i,k}^{[j]} - \hat{y}_{i,k}^{[j]})^2 \tag{4}$$

where n is the number of observations in the input dataset; $\mathbf{y_i}^{[j]} \in \mathbb{R}^{d_i}$ is the measured value of the desired quantity for observation j for branch i, with $j = 1, ..., n$; $\hat{\mathbf{y}}_i^{[j]} \in \mathbb{R}^{d_i}$ is the output of the branch i, when evaluated on the j^{th} observation. The global cost function weights used here were $\alpha_1 = 0.3$, $\alpha_2 = 5$ and $\alpha_3 = 1$.[13]

The training was performed with a learning rate of 10^{-3} without using any batch. To minimize the cost function, the optimizer Adaptive Moment Estimation (Adam)[14,15] was used. The implementation was performed using the TensorFlow™ library. All the results shown in this paper have been obtained by training the model for $2 \cdot 10^4$ epochs.

To compare results from different network models, the absolute error (AE) defined as the absolute value of the difference between the predicted and the expected value for a given observation was calculated. In this paper all the predictions of the network are for $[O_2]$ the output of branch 2, and for T the output of branch 3. For example, the AE of the j^{th} oxygen concentration observation $[O_2]^{[j]}$ is given by[13]

$$AE_{[O_2]}^{[j]} = |[O_2]_{pred}^{[j]} - [O_2]_{meas}^{[j]}|. \tag{5}$$

The further quantity used to characterize the performance of the network is the mean absolute error (MAE), defined as the average of the absolute value AE of Eq. 5. Again, for the oxygen prediction, using the training dataset S_{train}, the $MAE_{[O_2]}$ is defined as

$$MAE_{[O_2]}(S_{train}) = \frac{1}{|S_{train}|} \sum_{j \in S_{train}} |[O_2]_{pred}^{[j]} - [O_2]_{real}^{[j]}| \tag{6}$$

where $|S_{train}|$ is the size (or cardinality) of the training dataset.

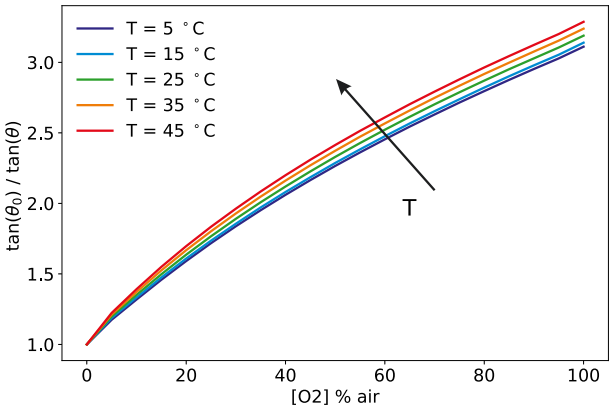

Figure 3. Measured Stern-Volmer curve for selected temperatures at a fixed modulation frequency of 6 kHz. The arrow marks increasing temperatures.

Finally, the kernel density estimate (KDE) of the distributions of the AEs was investigated to study the distribution of the prediction probability distribution for both the oxygen concentration and the temperature. The 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.[16] In this work, a Gaussian Kernel and a Scott bandwidth adaptive estimation[17] using the seaborn Python package[18] were used.

The experimental dataset was divided randomly in a training dataset containing 80 % of the data, used to train the network, and a development dataset containing 20 % of the data, used to test the ability of the network to generalize when applied to unseen data.

5. RESULTS

The experimentally measured Stern-Volmer curve is shown in Fig. 3 for selected temperatures. The curve is clearly non-linear, as expected from Eq. 3. With increasing temperatures the quenching of the luminescence increases, as it is expected for collisional quenching due to faster diffusion.[1]

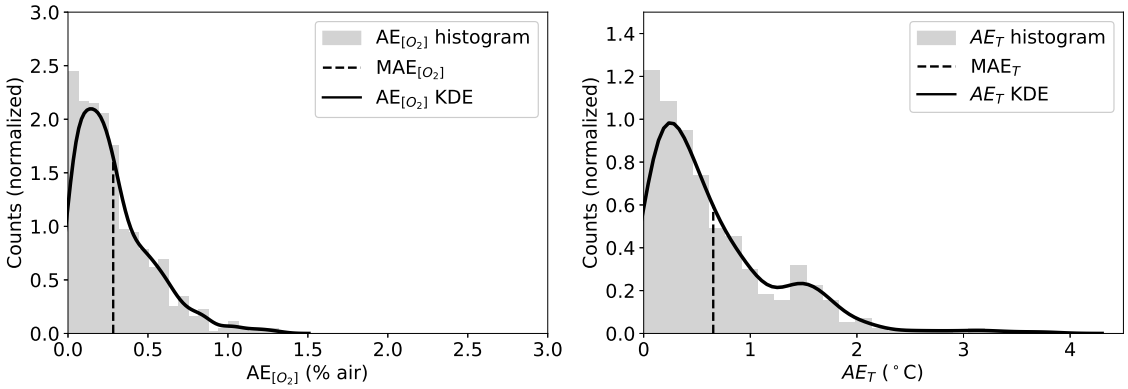

Figure 4. Performance of the neural network for the oxygen (left plot) and the temperature (right plot) predictions: Normalized prediction distribution histogram (columns) and kernel density estimate (KDE) of the distribution of the AEs (solid line). The corresponding MAE is also shown as a dashed line in each diagram.

The performance of the neural network in predicting the oxygen concentration and the temperature is analyzed by investigating the distribution of the absolute error for each parameter $AE_{[O_2]}$ and AE_T, as defined in Eq. 5. The results are summarized in Fig. 4, where the histogram of the AE, their average over all the values MAE and the KDE are plotted. As shown in Fig. 4, the prediction of the oxygen concentration across all the temperatures is practically always below an $AE_{[O_2]}$ of 1 % air with a $MAE_{[O_2]}$ of only 0.3% air. The prediction

of the temperature is broader, but the error AE_T is almost always below 2 °C. The MAE_T is 0.7 °C. These findings shoe that the neural network learns to predict both parameters simultaneously exceptionally well.

The results described above were obtained using an input vector of $\theta(\omega_i)/\theta_0(\omega_i)$ for all 50 modulation frequencies w_i. A relevant question for practical implementation is which frequencies of the entire range used here carry the most relevant information about the oxygen concentration and temperature dependence. Therefore, an analysis of the dependence of the $MAE_{[O_2]}$ of the modulation frequencies was carried out. The training and the tests of the network were performed using an increasingly smaller input size vector. Specifically, starting from the lowest modulation frequencies, an increasing number of values of $\theta(\omega_i)/\theta_0(\omega_i)$ were removed. The results are shown in Fig. 5, where the $MAE_{[O_2]}$ is plotted versus the lowest frequency used for the input vector.

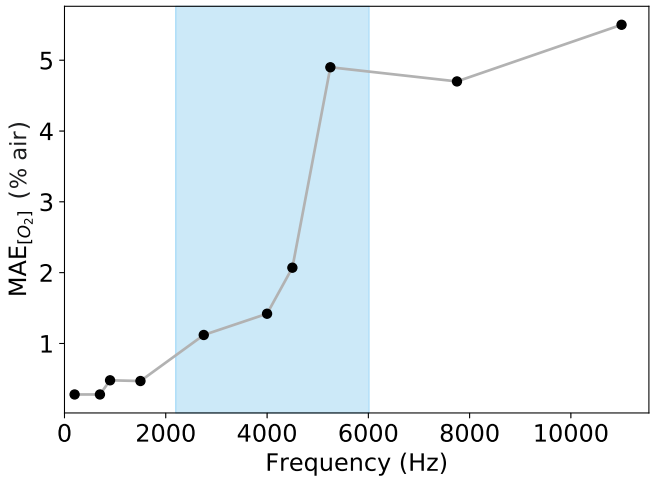

Figure 5. Development of the $MAE_{[O_2]}$ when reducing the range of the frequency sweep. The marked area indicates the frequency range where the most useful physical information are contained.

Fig. 5 shows, that by removing points at low frequency, up to ca. 2000 Hz, the $MAE_{[O_2]}$ is only marginally affected, therefore indicating that the data at low frequency carry little useful information for the task of the sensor. However, by further reducing the input vector size, particularly removing values of $\theta(\omega_i)/\theta_0(\omega_i)$ for $w_i > 4000$ Hz, the $MAE_{[O_2]}$ jumps from an absolute error below 1 % air to 5 % air. By removing additional points there is not an additional increase in the $MAE_{[O_2]}$, indicating that the crucial information is contained in data taken with a modulation frequency in the range between 2000 and 6000 Hz. These findings are consistent with observation that the maximum of the difference of the phase shift without quenching (in nitrogen) and with the strongest quenching (in air) of the sample is around 3900 Hz. The results of this analysis indicate that the sensor is expected to perform similarly well for a frequency sweep over a much reduced range. This relaxes the requirements on the sensor electronics for a practical implementation. It is to be noted, that the useful frequency range depends on the luminescence decay time and therefore, is determined by the choice of the luminophore.

6. CONCLUSIONS

In this work, the authors present a new approach to multiple luminescence sensing and demonstrate it with a dual oxygen and temperature sensor. The sensor is based uses a single luminophore, a single excitation and one detection channel. Thanks to a MTL neural network can predict the oxygen concentration and the temperature with an mean absolute of error below 0.3% air and below 0.7 °C respectively. These findings, therefore, prove that, to build an accurate oxygen sensor is not necessary to implement a complex non-linear pseudo-physical model describing the dependence of the measured quantity, here the phase shift, from the many influencing parameters. A further advantage is that sensor-specific deviations from a simple Stern-Volmer model, arising from the method for the immobilization of the indicators in the substrate, or even by sensor components, like absorption filters or glues, are not relevant and will be learned from the network automatically. The price to pay is the acquisition of a large amount of data, which however could be performed automatically by a setup, not requiring manual intervention and human supervision.

This work paves the road to a new and completely different approach in sensor development. Using artificial intelligence is not necessary to define mathematical models to capture all influencing factors and to perfect the hardware of the sensor to minimize undesired factors which cause interferences or cross-sensitivity. Once the sensor hardware is given, that is all the optical, electronic, mechanical, and chemical components are assembled, the neural network can be trained to learn to predict the quantities of interests.

REFERENCES

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

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

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

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

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

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

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

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

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

[10] Argyriou, A., Evgeniou, T., and Pontil, M., "Multi-task feature learning," in [*Advances in neural information processing systems*], 41–48 (2007).

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

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

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

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

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

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

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

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