**Reviewer 1**

In principle, the manuscript provides an interesting approach of how to evaluate data and how to deal with interfering parameters. It is definitely worth to follow the concept although there are few sections that need to be elaborated further.

The proposed method sounds like a supervised pattern-recognition algorithm to me, where the underlying correlation between the analyte and the measurement parameter / indicator response is unknown. Scanning over the whole frequency range either the oxygen concentration or the temperature is changed continuously to establish a calibration matrix of the sensor.

First of all, it must be noted that the correlation between the two parameters and the indicator response is one of the best characterized. However, the approach is interesting for other parameter interdependencies and indicator correlations that are still unknown.

Unfortunately, the manuscript failed in explaining how the sensor is interesting for the user. Is the calibration reproducible / transferable or must the user calibrate the sensor each time investing 65 hours for the calibration? How is the procedure to evaluate an unknown sample? Is it possible to apply the method online for real-time sensing? How well would the algorithm perform at low oxygen concentrations and high temperatures or vice versa, meaning on the extrema? What is the dynamic range of the method and the maximal error then?

A similar procedure has already been published using an indicator based on thermally activated delayed fluorescence and should be addressed here (https://doi.org/10.1021/acssensors.9b02512).

Despite all the comments, the approach is an interesting one that has a high potential for other optical sensors, where the correlation is not as well known as it is the case for oxygen and temperature.

Specific comments:

Line 15-16: Within the manuscript, the method is applied to a single indicator and not to a multi-indicator system. Multi-indicator systems are more complex as inherent parameter interactions might occur that are not predicable leading to an error propagation throughout the evaluation procedure. As the method proposed here bases on a single indicator, I doubt whether the method can directly be transferred to a multi-indicator system without any adjustments.

Line 25-27: unclear

Line 43: Stern-Volmer is not an empirical model and T-dependencies neither.

Linear 51: the manuscript describes a new approach on data evaluation not a new sensor.

Line 66-68: Are all parameter-dependencies and especially all sensor-specific response characteristics learnable? What about the photo-degradation over time? Is it something the method might be able to compensate?

Line 72: Even though decay time-based approaches are more robust, might it be possible to use also an intensity-based approach?

Line 76: The Temperature dependency depends on the chosen indicator and range. There are multiple systems known where the T-dependency can be described as a linear correlation with a negligible error propagation.

Line 85: What is the expected dynamic (in terms of decay time) of the indicators regarding the oxygen and the temperature dependency?

Line 87: Even though the setup was already described somewhere else, it would be great to include it in the SI. Please, add also the reference sensors that are needed to ensure that the target parameters are matched.

Line 91: Have you experienced any background effects at low modulation frequencies (200Hz)?

Line 100: The section 2.3 requires elaboration. It might be nice to get a better understanding on how the data are analyzed. How are the (unknown) sample data analyzed after the training? Maybe you could use one sample as explanation and attach it to the SI. What about the visualization? Can you show how the final matrix pattern, identified by the algorithm, looks like?

Line 104-105: You submit your paper in a sensor journal who might be not familiar with the network architecture. Please, revise this sentence for a reader-oriented outreach and include more information on how the network architecture is supposed to work.

Line 113: Elaborate in more detail, why 32 is the optimal number. How did you optimize the batch size?

According to your literature, a special tuning is needed here.

Line 123: What is the CPU and how many cores are needed? Please, provide concrete information focusing on the facts required here and avoid woolly formulations.

Line 130: Why did you use the mean of the absolute error instead of the residuals?

Line 141: Why did you choose accuracy as the only performance parameter? To my knowledge, the accuracy can be biased and shouldn’t be the only figure of merit, when it comes to describe the performance of an algorithm (http://eecs.wsu.edu/~holder/courses/cse6363/spr04/pubs/Provost98.pdf)

Line 180: Have you thought about reproducibility / calibration-transfer? How is the sensor intended to be used by the user? Does the user have to make the whole calibration by themselves again?

Line 198: Single indicator sensors are unusual for optical chemical sensors? How would the approach perform in multi-layer systems and deal with their inherent indicator-interaction that possibly occur?

Line 199: A calibration curve of the final matrix would help to understand how the user can then analyze its unknown samples.

Line 200: Please, provide more (concrete) information on the comparison. What is the typical error of commercial sensors? Which sensors did you use for comparison?

**Reviewer 2**

It is well written manuscript.  However, there are something which may be better to revise.

1) There is an approach which detect all parameter of interferences simultaneously on the solid phase absorbance detection for gas analysis. This may be cited.  (Analytical Chemistry, 81, 4183-4191 (2019).)

2) It is not clear about the temperature.  Is this for sample gas, sensor , or whole system include sample gas?

3) Relative humidity may also be effects to the response. Did you concern about RH effects?

4) "AE" is described in the main text.  But it was not explained in figure caption. It is better the figure can be understand separately.

5) It is better to put the basic sensor performance after the optimization. LODs, reproducibility, and response speed?

6) Also the sensor resolution on the concentration should be shown. Because the tiny difference of the response by temperature was discussed.