sensors

MDPI

*Type of the Paper (Article, Review, Communication, etc.)*

# Artificial Neural Network to Eliminate the Detrimental Spectral Shift on Mid-Infrared Gas Spectroscopy

**Sanghoon Chin[1*], Jérôme Van Zaen[1], Séverine Denis[1], Enric Muntané[1], Stephan Schröder[2], Hans Martin[2], Laurent Balet[1] and Steve Lecomte[1]**

[1] *Centre Suisse d'Electronique et de Microtechnique SA (CSEM), CH-2002 Neuchâtel, Switzerland*
[2] SenseAir AB, 82060, Delsbo, Sweden
* Correspondence: sanghoon.chin@csem.ch

Abstract: We demonstrate the successful implementation of an artificial neural network (ANN) to eliminate detrimental spectral shifts imposed in the measurement of laser absorption spectrometers (LAS). Since the principle of LAS relies on the analysis of spectral characteristics of biological and chemical molecules, the accuracy and precision of the spectrometer is essentially prone to the presence of unwanted spectral shift in the measured molecular absorption spectrum over the reference spectrum. In this paper, an ANN was applied to a scanning grating-based mid-infrared trace gas sensing system, which suffers from temperature-induced spectral shifts. Using the HITRAN database, we generated synthetic gas absorbance spectra with random spectral shifts for training and validation. The ANN was trained with these synthetic spectra to identify the occurrence of spectral shifts. Our experimental verification shows that such an ANN can be an excellent tool to accurately retrieve the gas concentration from imprecise or distorted spectrum of gas absorption.

**Keywords:** spectral analysis; artificial neural network; quantitative gas analysis; trace gas sensing; mid-infrared; absorption spectroscopy; supercontinuum source;

## 1. Introduction

Photonic sensing systems in the mid-infrared (MIR) wavelength range have recently attracted substantial attention to the optical community due to its outstanding capability to detect minute traces of molecules in complex gas mixtures. The reason lies in the fact that the fingerprint of rotational and vibrational resonances of gas molecules originating from biological and chemical activities along this spectral window is a few orders of magnitude stronger than other spectral range. Therefore, MIR spectroscopy systems have been extensively employed for various applications such as air quality monitoring, health diagnostics and scientific research [1-6]. In this context, the generation of MIR light sources emitting in the wavelength range of 2-10 μm has been effectively demonstrated, using various techniques such as quantum cascade lasers [7,8], optical parametric oscillators [9,10] and supercontinuum lasers [11-14]. Among them, the high-brightness supercontinuum light source, associated with a high-resolution diffractive grating spectrometer shows a high performance of accurate detection on minute traces of multispecies gas molecules [5].

However, we observed that any presence of thermal fluctuation of the gaseous analyte can lead to detrimental thermally-induced mechanical stresses to optical components in the system, causing beam steering. Therefore, unwanted spectral shifts may occur in the measured absorption spectrum which, in turn, result in a non-negligible amount of error for the gas concentration computation, traditionally performed by least-square fitting process between the measured spectrum and the database reference spectrum [11]. Such a spectral calibration issue has been already addressed in a similar grating-based gas

**Citation:** To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

sensing system [14] and has been solved by cross-correlation calculation between the measured absorbance spectrum and the reference spectrum to estimate the exact amount of the spectral shift caused by mechanical disturbances in the system. Yet, this solution requires a heavy mathematical computation that slows down the measurement speed and the clear presence of known gas species to precisely compare with the reference, which might limit the sensing performance for ambient air quality monitoring.

Over the last few years, ANNs have been applied to multiple scientific research domains and engineering applications, due to their relevance for complex nonlinear problems. Recently, an ANN has been applied to a MIR gas spectrometer to accelerate the estimation of gas concentration in complex gas matrix conditions, showing promising results [15-17]. In this paper, we propose to use a multilayer perceptron (MLP) to provide an accurate prediction of the gas concentration from gas absorption spectra with a random spectral shift for the first time to the best of our knowledge.

## 2. Development of MIR gas spectrometer and spectral shift issue

Figure 1 depicts the simplified schematic diagram of the scanning grating-based MIR trace gas sensing system that we have developed in our laboratory, using a supercontinuum (SC) light source that is spectrally broadened from 2 μm to ~4.5 μm. The SC light is sent to a 10 m-long multi-pass cell (MPC), where the light interacts with the gas analyte, and the wavelength-specific absorption occurs. Then, the light emerging from the cell is directed to a blazed grating with 450 lines/mm. The 1$^{st}$ order diffracted light was focused on a single pixel detector for the gas absorption analysis. The spectrum of the SC light was readily resolved by rotating the grating that is mounted on a motorized rotational stage. Under the following conditions: a free space beam path length of 36 cm between the grating and the detector and the grating dispersion of 27.9 μm/°, a grating rotation increment by 0.01 ° induced a rise in a geometrical beam steering of 60 μm. It corresponds to a spectral shift of 0.51 cm$^{-1}$ in wavenumber. Based on these parameters, a 50 μm slit was placed in front of the detector to precisely record the spectral power density of the light source, resulting in a spectral resolution of 1.15 cm$^{-1}$ for the sensing system. The grating was then scanned from 327° to 330° by steps of 0.01 ° to obtain the partial absorption spectrum of water vapor contained in the ambient air while the ambient air was pumped into the cell. Notice that the free space path length from the light source to the detector is ~90 cm excluding the beam path through the cell; hence, proving that the light absorption inside the cell would be dominant compared to the light absorption occurring in the open space portions.

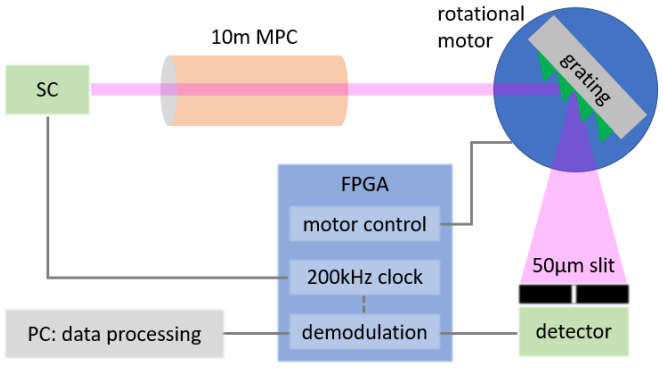


Figure 1: Simplified schematic diagram of the scanning grating-based MIR SC gas spectrometer

From the measured $H_2O$ absorbance spectra as shown in Figure 2(a), a spectral shift of water absorption peaks over time is clearly observed. The amount of the spectral shift was then compared to the change of temperature inside the MPC, showing a strong correlation as shown in Figure 2(b). The temperature inside the MPC was simultaneously measured by an electrical temperature sensor (BME280, BOSCH) embedded inside the

MPC. For this reason, we believe that the spectral shift can be mainly attributed to the variation of temperature inside the MPC. We suspect that thermally-induced movements of the optical components in the system lead to such detrimental beam steering, causing the spectral shift in the measurement. However, we did not investigate further to determine the cause of the beam discrepancies.

According to our measurements, the spectral shift of the absorption spectrum seems to have complex response to the temperature change, since the slope efficiency defined as the ratio of spectral shift to temperature change varies, as shown in Figure 2(d). However, the coefficient was estimated to be in the order of 0.01 °/K, implying that the measured gas absorption profile will be spectrally shifted with a coefficient of 0.51 $cm^{-1}$/K. Figure 2(c) illustrates the water absorbance spectra of the first and last measurements, explicitly showing a spectral shift by 0.03° while the temperature was changed by 2.6°C between these two measurements. On the other hand, the measured absorbance spectrum was compared to the HITRAN reference absorbance to calculate the amount of the spectral shift, resulting in a spectral shift of ~1.55$cm^{-1}$, showing a good agreement with our estimation of spectral shift of 0.51$cm^{-1}$ by grating rotation of 0.01°. Furthermore, when we performed the typical least-square fitting algorithm to retrieve the gas concentration, we obtained 11125 ppm (or 1.113%) and -1435 ppm (or -0.144%) for the first and the last measurements, respectively. Therefore, the sensing system turns to be severely impaired since the thermally-induced spectral shift invalidates the predefined background transmission profile and the reference gas absorption profile used for the fitting algorithm. Moreover, due to the complex response of spectral shift to the temperature change, a fixed slope coefficient cannot precisely compensate for the adverse thermal effect for practical applications. More details will be discussed later.

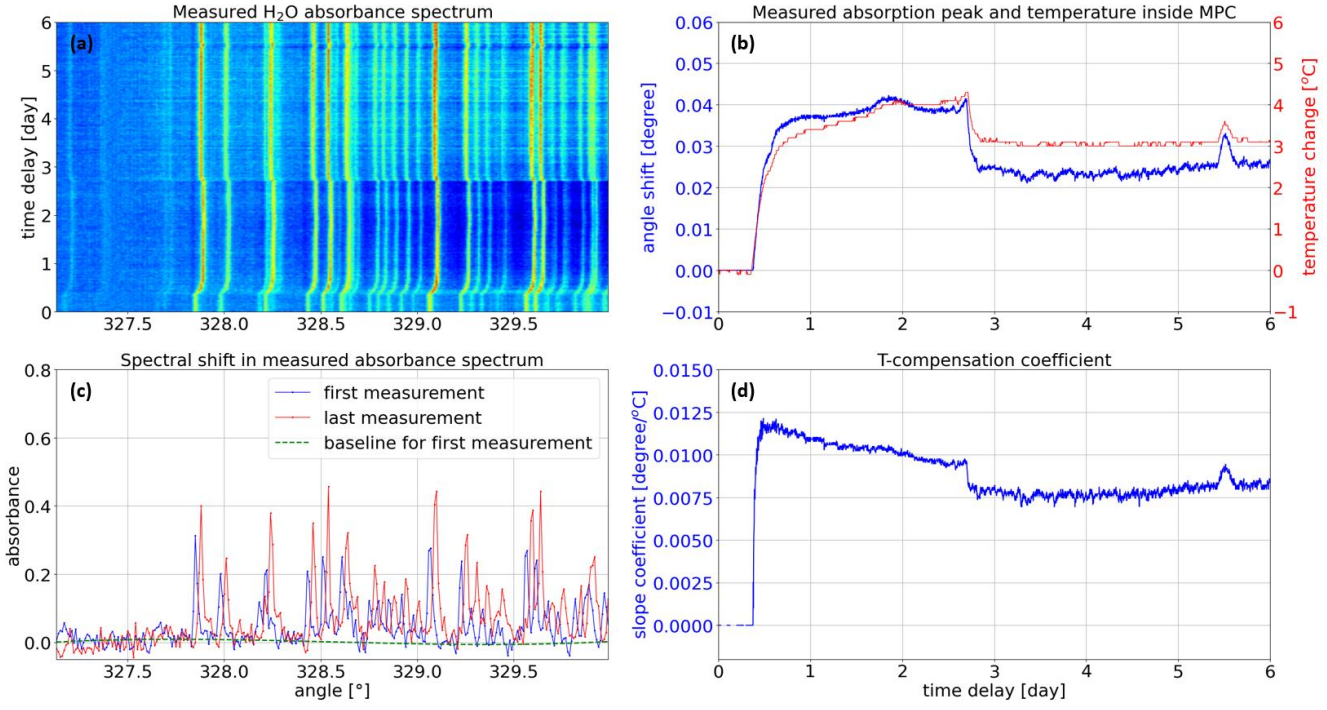


Figure 2: (a) Continuous measurement of $H_2O$ absorption profile while the gas cell temperature changes. (b) slope coefficient calculated by the ratio of the amount of spectral shift to temperature change. (c) $H_2O$ absorbance spectra for the first and last measurements, showing a clear spectral shift due to the temperature change inside the MPC. (d) Measured slope coefficient defined as the ratio of the spectral shift to temperature change.

## 3. Development of artificial neural network

To effectively overcome this problem, a MLP was applied to accurately estimate the gas concentration in the presence of such inevitable spectral shift. Moreover, to validate the proof-of-concept we decided to focus on the measurement of water vapor concentration since a capacitive relative humidity (RH) sensor embedded inside the multipass cell could be used as reference. Due to the limited number of available measured absorbance spectra, our neural network was trained with synthetic absorbance spectra. First, we extracted the molecular absorption coefficients for water vapor ($H_2O$) and methane ($CH_4$) with a Lorentzian profile from the HITRAN database. As the goal was to accurately estimate $H_2O$ concentration, $CH_4$ was used as a perturbation since the atmospheric methane absorption is certainly present in the set scanning spectral range. The next step to generate an absorbance spectrum was to sample a $CH_4$ concentration between 0 and 100 ppm and a $H_2O$ concentration between 0 and 10 % from a scaled beta distribution with parameters a = 0.5 and b = 1. In fact, the water vapor concentration of 10% corresponds to the 100% relative humidity under environmental conditions of 46°C and atmospheric pressure, which considers harsh environments such as gaseous pollutants monitoring generated from the incineration plant. The distribution parameters were selected to favor lower gas concentrations where estimation errors should be lower. In turn, these sampled concentrations were utilized to generate a transmittance spectrum with a 10 m optical path length and a convolution with a Gaussian kernel and a spectral resolution of 1.15 $cm^{-1}$. Finally, an absorbance spectrum was computed by taking the logarithm of the transmittance spectrum. Each synthetic absorbance spectrum spans the band ranging from 3010 $cm^{-1}$ to 3290 $cm^{-1}$ in 0.05 $cm^{-1}$ steps and includes 5601 samples.

To mimic the practical signal noise and nonlinear spectral behavior, we applied three perturbations to the generated synthetic spectrum. Firstly, using our measurements, the level of root-mean-squared noise within a finite spectral window, at which the water vapor absorption is negligible, was analyzed. More specifically, the window in the vicinity of grating rotation angle of 327.5°, corresponding to 3159.4$cm^{-1}$ in wavenumber as shown in Figure 2(c) was selected for our analysis. As a result, an additive white Gaussian noise with a standard deviation of 0.025 in absorbance unit was imposed onto the synthetic absorbance spectrum. Secondly, a random baseline generated as Legendre polynomials of degree 4 with coefficients sampled uniformly was also added since SC light sources typically suffer from intrinsic peak-to-peak random fluctuations in their intensity and spectral power density. Such an adverse noise is mainly attributed to the mechanism of the incoherent nonlinear spectral broadening process during the supercontinuum generation [18] and it varies the baseline of the measured absorbance spectrum as shown in Figure 2(c), which leads to a non-negligible amount of error in the computed gas concentration. After analyzing the temporal variation of the baseline over the whole measurements the range of each coefficient of the $4^{th}$ order polynomial function was thoroughly determined to generate a random baseline. Thirdly, we added a random spectral shift that is sampled uniformly between -10 $cm^{-1}$ and 10 $cm^{-1}$, covering the temperature change of ±19.6°C. The last step of synthetic data generation was to scale the absorbance spectra between 0 and 1 to facilitate the training. The scaling factor was determined from the maximum possible concentrations of $CH_4$ and $H_2O$ without any perturbation.

Figure 3 illustrates the considered MLP architecture. A MLP is a fully-connected feed-forward ANN, where each neuron is connected to all neurons in neighboring layers. In our work, the MLP is composed of three layers, namely: two hidden layers with 256 neurons and ReLU (Rectified Linear Unit) activation and one output layer with a single unit and linear activation. Our network takes an absorbance spectrum as input and estimates the $H_2O$ concentration as follows:

$$x_1 = \mathrm{ReLU}\{W_1 x_0 + b_1\}$$
$$x_2 = \mathrm{ReLU}\{W_2 x_1 + b_2\}$$
$$x_3 = W_3 x_2 + b_3$$

where $x_0 \in \mathbb{R}^{5601}$ is the input spectrum, $x_1 \in \mathbb{R}^{256}$, $x_2 \in \mathbb{R}^{256}$, and $x_3 \in \mathbb{R}$ are the outputs of each layer, $W_1 \in \mathbb{R}^{256\times5601}$, $W_2 \in \mathbb{R}^{256\times256}$, and $W_3 \in \mathbb{R}^{1\times256}$ are the layer

weights, and $b_1 \in \mathbb{R}^{256}$, $b_2 \in \mathbb{R}^{256}$, and $b_3 \in \mathbb{R}$ are the layer biases. The MLP parameters were trained through back-propagation by minimizing the mean squared error (MSE) with the Adam optimizer; hence, optimizing the weights and bias for each neuron. Finally, the output layer provided the result of the hidden layers as the gas concentration. The training was performed with a learning rate of 0.001 for 500 epochs of 1000 batches, where each batch was composed of 100 synthetic absorbance spectra. It means that the MLP was presented with 50 million different spectra over the whole training procedure. At each epoch, the MLP was evaluated with a validation set of the same size as the training set. In turn, MLP parameters corresponding to the epoch with the lowest MSE on the validation set were selected for the evaluation on the real absorbance spectra.

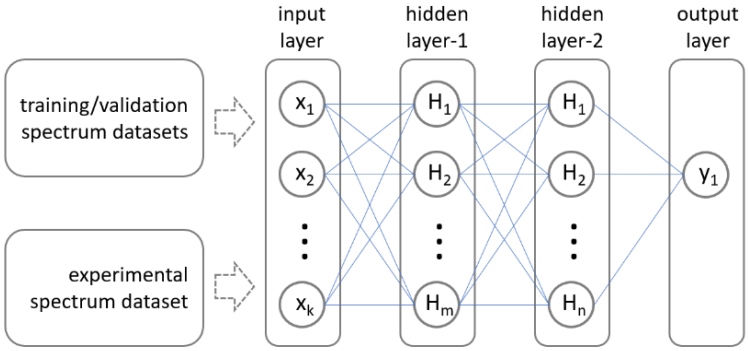


Figure 3: MLP architecture, consisting of input and output layers and 2 hidden layers with 256 units and ReLU activation.

To evaluate the reliability of the fully trained MLP algorithm, a test dataset consisting of experimentally measured absorbance spectra was applied to the network. To prepare the test dataset, we used two individual gas bottles filled with a calibration mixture of the gases of interest: one bottle of 100 % nitrogen and the other of mixed gas of 99.995 % nitrogen and 0.005 % (equivalent to 50 ppm) methane. The two bottles were combined before being fed into the gas cell. The gas flowrate of each bottle was controlled independently to vary the water vapor concentration between 0 % and an atmospheric level of ~1% with different levels of $CH_4$ perturbation. The sensing system was operated over 38 hours with a measurement time of 3 minutes to continuously acquire the gas absorption spectra under different environmental conditions. The measurement started while the two bottles were closed, and the MPC was under normal atmospheric condition. Then, during the time window between time 14 hours and 28 hours the methane concentration was set to 5 different levels: 0, 0.5, 1, 1.9 and 50 ppm and each concentration remained constantly over 1-2 hours.

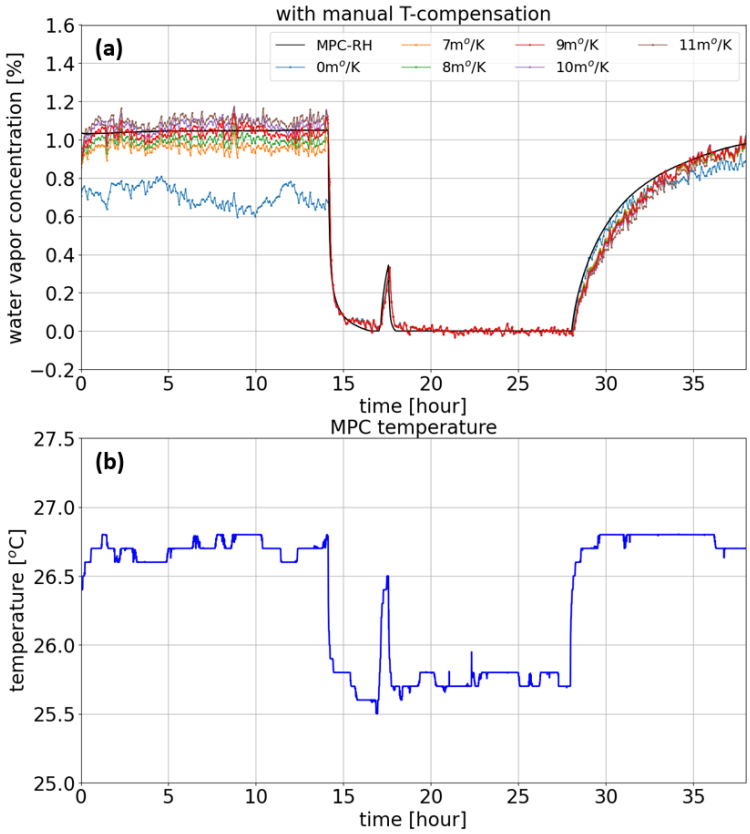


Figure 4: (a) Reference H2O concentration profile in black and retrieved profiles resulting from the manual temperature compensation for different coefficients. (b) Measured temperature inside the MPC.

The water vapor concentration measured by the relative humidity shows the atmospheric concentration of 1.05 %, as shown in Figure 4(a), which corresponds to 27.75% RH. However, when the MPC was fed by the sample gas the water concentration was abruptly decreased to 0 % due to the nitrogen purging effect in the gas cell. Next, the gas flow was stopped after 28 hours and the atmospheric water vapor started to enter the gas cell, increasing to about 1 % at a measurement time of 38 hours. It is important to mention that during this test period the temperature inside the gas cell varied from 25.50 °C to 26.80 °C, as shown in Figure 4(b) while the frequency calibration of the grating angle was performed at 25.75 °C. Thereby, we expect that our sensing system would be impaired by the grating angle deviation of ~0.015°, equivalent to the spectral shift of ~0.76cm$^{-1}$ in wavenumber. It's noticeable that the water concentration profile shows a perfect matching with the temperature variation profile of the gas cell. Overall, as expected, due to the adverse spectral shift the water concentration retrieved by the typical least-mean-square fitting algorithm shows a ~33% error for the atmospheric air sample in the beginning of the test. To evaluate the performance of the manual temperature compensation, the original absorption spectrum was spectrally corrected using 5 different slope coefficients at 0.007-0.011°/K, as shown in Figure 5(a). We found that the slope coefficient of 0.009 °/K results in the best agreement to the reference water concentration profile while the other coefficients indicate either underestimation or overestimation.

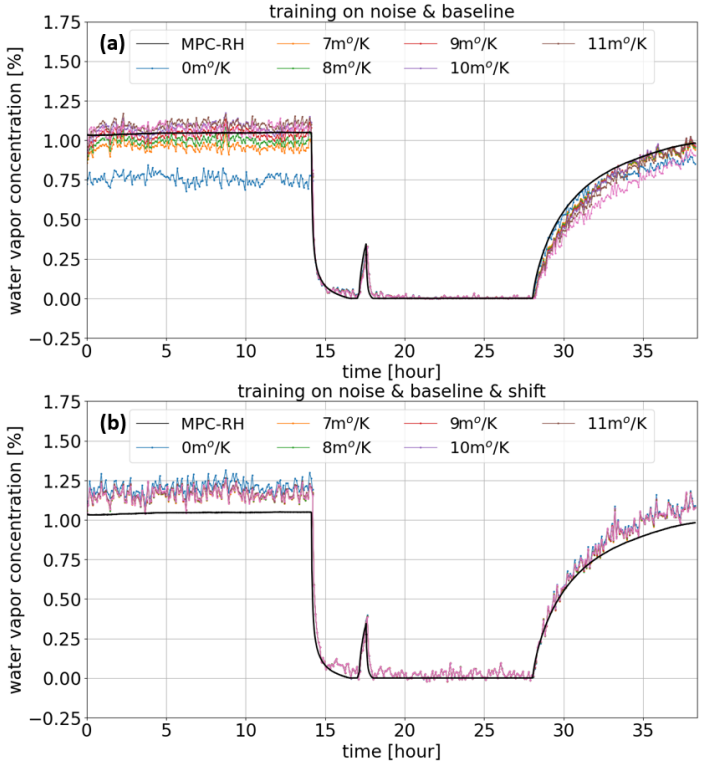


Figure 5: $H_2O$ concentration estimated by the ANN (a) when the MLP training relied on the noise and baseline and (b) when the MLP training relied on the noise, baseline and spectral shift.

Then, all gas absorption spectra, including original and spectrally shifted by the manual temperature compensation, were applied to the trained network. In a previous work [17], the ANN was trained mainly on two features of the system noise and the baseline of the light transmission spectral profile for the rapid gas concentration analysis. Therefore, for comparison reasons our first network was also trained on these two features with the random parameter range as previously explained. But, as shown in Figure 5(a) the trained network demonstrated nearly the same results as the typical fitting result, meaning that the capacity of this network was insufficient to overcome the complex spectral shift problem. Accordingly, our MLP network was trained again, including the feature of the temperature-induced spectral shift and the water vapor concentration was predicted by the network, as shown in Figure 5(b). All temperature-compensated spectra with a different slope coefficient converged to predict an identical pattern of concentration variation in time, following the reference profile unlike the case of the previous network. It proves the large potential of completely overcoming such unpredictable spectral shifts. However, it must be mentioned that this MLP architecture tends to overestimate the water concentration by ~11% with respect to the RH sensor, while preserving the sensing sensitivity, which was determined by the standard deviation of the concentration fluctuation over the first 12 hours measurement. Such error in quantitative analysis might result from any possible bias in the relative humidity measurement. Nevertheless, the performance of the ANNs for this task could be improved by investigating more complex architectures including convolutional and pooling layers.

## 4. Conclusions

In conclusion, we have identified a temperature-induced spectral shift issue in our scanning grating-based MIR gas sensing system and have proposed a promising solution to deal with such a detrimental thermal effect. According to our successful experimental demonstration, an ANN is a promising approach to unambiguously overcome any presence of complex spectral distortion imposed on any spectroscopic data. The degradation

of the sensing performance by such a spectral shift can occur to any spectroscopic sensing systems which require an accurate calibration of frequency in the measurement. So, we can emphasize that the proposed solution has a large potential to make trace gas sensing systems based on gas absorption spectrum more robust and immune to any variation of environmental conditions. In addition, we believe that machine learning will also be powerful on any types of spectral distortions once they are properly characterized.

**Author Contributions:** Conceptualization, S.C and L.B.; methodology, S.C., L.B. and S.L.; software, J.V.Z, S.D. and E.M.; mechanical design and integration, L.B, S.C., S.S. and H.M; validation, S.C., L.B. and J.V.Z; writing—original draft, S.C. and J.V.Z; writing—review and editing, S.C., J.V.Z, S.D., E.M., S.S., H.M., L.B. and S.L

**Funding:** This project has received funding from Horizon 2020, the European Union's Framework Program for Research and Innovation, under grant agreement No.101015825 (TRIAGE).

**Acknowledgments:** The authors thank the EU Horizon 2020 for financially supporting the research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Z. Du, S. Zhang, J. Li, N. Gao and K. Tong, "Mid-Infrared Tunable Laser-based Broadband Fingerprint Absorption Spectroscopy for Trace Gas Sensing: A Review",Appl. Sci. **9**, 338 (2019).
2. H. Huszár, A. Pogány, Z. Bozóki, Á. Mohácsi, L. Horváth and G. Szabó, "Ammonia monitoring at ppb level using photoacoustic spectroscopy for environmental application", Sens. Actuator B **134**, 1027-1033 (2008).
3. R. Ghorbani and F. M. Schmidt, "Real-time breath gas analysis of CO and CO2 using an EC-QCL", Appl. Phys. B **123** (2017).
4. L. Liu, B. Xiong, Y. Yan, J. Li and Z. Du, "Hollow Waveguide-Enhanced Mid-Infrared Sensor for Real-Time Exhaled Methane Detection", Photon. Technol. Lett. **28**, 1613-1616 (2016).
5. J. Ropcke, S. Welzel, N. Lang, F. Hempel, L. Gatilova, O. Guaitella, A. Rousseau and P.B. Davies, "Diagnostic studies of molecular plasmas using mid-infrared semiconductor lasers", Appl. Phys. B **92**, 335-341 (2008).
6. M.A. Bolshov, Yu.A. Kuritsyn and Yu.V. Romanovskii, "Tunable diode laser spectroscopy as a technique for combustion diagnostics", Spectroc. Acta Part B-Atom. Spectrosc. **106**, 45-66 (2015).
7. E. Normand, M. McCulloch, G. Duxbury and N. Langford, "Fast, real-time spectrometer based on a pulsed quantum-cascade laser", Opt. Lett. **28**, 16 (2003).
8. S. Chin, V. Mitev, E. Giraud, R. Maulini, S. Blaser and D. L. Boiko, "Electrically driven frequency blue-chirped emission in Fabry-Perot cavity quantum cascade laser at room temperature", Appl. Phys. Lett. **118**, 021108 (2021).
9. D.D. Arslanov, M Spuneli, J. Mandon, S.M. Cristescu, S.T. Persijn and F.J.M Harren, "Continuous-wave optical parametric oscillator based infrared spectroscopy for sensitive molecular gas sensing", Laser Photonics Rev. **7**, 188-206 (2013).
10. O. Kara, F. Sweeney, M. Rutkauskas, C. Farrell, C.G. Leburn and D.T. Reid, "Sensitive multi-species trace gas sensor based on a high repetition rate mid-infrared supercontinuum source", Opt. Express **15**, 21358-21366 (2019).
11. M.A. Abbas, K.E. Jahromi, M. Nematollahi, R. Krebbers, N. Liu, G. Woyessa, O. Bang, L. Huot, F.J.M. Harren and A. Khodabkhsh, "Fourier transform spectrometer based on high-repetition-rate mid-infrared supercontinuum sources for trace gas detection", Opt. Express **29**, 22315 (2021).
12. K.E. Jahromi, M. Nematollahi, Q. Pan, M.A. Abbas, S.M. Cristescu, F.J.M. Harren and A. Khodabakhsh, "Sensitive multi-species trace gas sensor based on a high repetition rate mid-infrared supercontinuum source", Opt. Express **28**, 26091-26101 (2020).
13. V.V. Goncharov and G.E. Hall, "Supercontinuum Fourier transform spectrometry with balanced detection on a single photodiode", J. Chem. Phys. **145**, 084201 (2016).
14. K.E. Jahromi, Q. Pan, A. Khodabakhsh, C. Sikkens, P. Assman, S.M. Cristescu, P.M. Moselund, M. Janssens, B.E. Verlinden and F.J.M. Harren, "A broadband mid-infrared trace gas sensor using supercontinuum light source: applications for real-time quality control for fruit storage", Sensors **19**, 2334 (2019).
15. T. Quyang, C. Wang, Z. Yu, R. Stach, B. Mizaikoff, B. Liedberg, G.B. Huang and Q.J. Wang, "Quantitative Analysis of Gas Phase IR Spectra Based on Extreme Learning Machine Regression Model", Sensors **19**, 5535 (2019).
16. T. Voumard, T. Wildi, V. Brasch, R.G. Alvarez, G.V. Ogando and T. Herr, "AI-enabled real-time dual-comb molecular fingerprint imaging", Opt. Lett. **45**, 6583-8586 (2020).
17. J. Goldschmidt, L. Nitzsche, S. Wolf, A. Lambrecht and J. Wollenstein, "Rapid Quantitative Analysis of IR Absorption Spectra for Trace Gas Detection by Artificial Neural Networks Trained with Synthetic Data", Sensors **22**, 857 (2022).
18. K. Kwarkye, M. Jensen, R.D. Engelsholm, M.K. Dasa, D. Jain, P. Bown, P.M. Moselund, C.R. Petersen and O. Bang, "In-amplifier and cascaded mid-infrared supercontinuum sources with low noise through gain-induced soliton spectral alignment", Sci. Rep. **10**, 8230 (2020).