Deep Learning Algorithms for Chemometric Analysis in MIR Gas Spectroscopy

Sanghoon Chin, Jérôme Van Zaen, Enric Muntané, Séverine Denis, Steve Lecomte and Laurent Balet
Centre Suisse d'Electronique et de Microtechnique SA (CSEM), CH-2002 Neuchâtel, Switzerland
sanghoon.chin@csem.ch

Abstract: We have implemented machine learning techniques into a mid-infrared gas spectrometer for two specific goals: the improvement of chemometric analysis using artificial neural networks and geostatistical analysis over a geographic area using Kriging.

1. Introduction

Photonic sensing systems in the mid-infrared (MIR) wavelength range have attracted substantial attention to the optical community thanks to its outstanding capability to detect minute concentration of molecules [1]. Most of these MIR spectroscopies for multispecies monitoring have been realized using broadband light sources such as supercontinuum (SC) lasers since their spectrum can cover the fingerprint of rotational and vibrational resonances of the most interesting gas molecules. Despite recent advancements in trace gas monitoring performance, the sensing systems are still seeking for solutions to overcome a few challenges. For instance, when many gas species co-exist in a gas mixture sample, the accuracy of gas concentration retrieval degrades due to the spectral cross-interference between different species and nonlinear noise of lasers, and the blended light absorption spectrum makes it difficult to classify gas species for better interpretation for air quality monitoring [2]. In recent decades, the interdisciplinary research combining the sensing technology and machine learning technique has been intensively investigated as an efficient and robust methodology to improve the reliability of gas spectroscopy. In this paper, we will present two successful demonstrations of machine learning algorithms applied to improve the performance of the MIR spectroscopy.

2. Deep Learning for Chemometric Analysis

During the development of a scanning grating-based MIR trace gas sensing system using a SC light source in our laboratory [3], detrimental spectral shift in the measured gas absorption spectrum is observed as shown in Figure 1(a), which is mainly attributed to the temperature variation inside a multipass gas cell. More specifically, mirrors integrated in the system suffer from thermo-mechanical displacement, resulting in unwanted beam steering. This phenomenon essentially causes the adverse spectral shift in the measurement, severely impairing the sensing system.

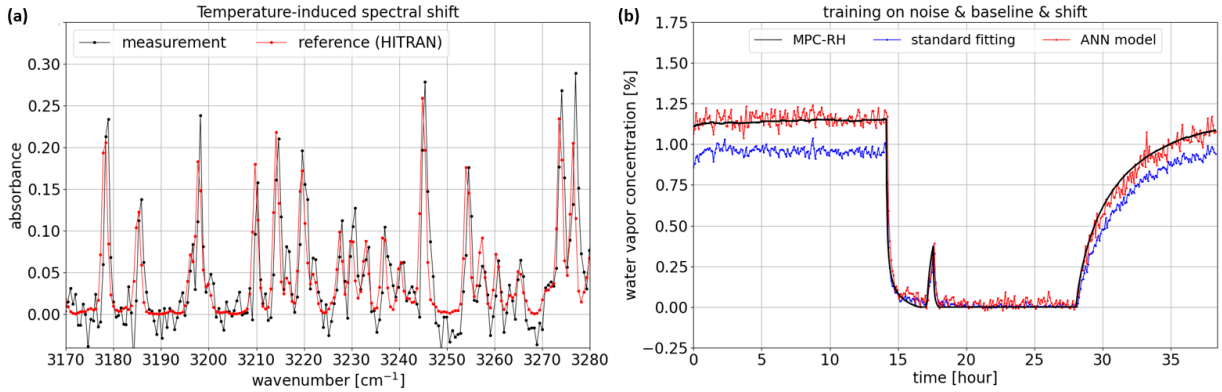

Figure 1: (a) Spectral shift of 0.41cm^{-1} in the water vapor absorbance spectrum, caused by temperature change of 1°C inside the MPC. (b) H2O concentration estimated by the MLP model trained on noise, baseline and spectral shift.

To effectively overcome this problem, an artificial neural network (ANN) is implemented to accurately compute the gas concentration from the distorted gas absorption spectrum. The measurement dataset collected by our MIR gas sensing system is small due to the lack of operation time, so this dataset is used to characterize the sensing system in terms of fluctuation of spectral density profile (referred to as baseline in the nonlinear least-squares fitting algorithm) and relative intensity noise. Using the results of system characterization, a large synthetic dataset of absorbance spectra is generated based on HITRAN database. Each absorbance spectrum in this synthetic dataset includes random baseline, noise, and gas concentrations. A multilayer perceptron (MLP) model is then trained on these synthetic absorbance

spectra. Finally, the model is evaluated on the small dataset of real absorbance spectra, as shown in Figure 1(b). Due to the presence of spectral shift, the standard least-square fitting algorithm results in a non-negligible amount of error in the gas concentration computation. On the other hand, the water vapor concentration has been accurately retrieved by the MLP model, showing a very good agreement with the reference concentration measured by an electrical relative humidity sensor embedded inside the gas cell.

3. Kriging for Geostatistical Analysis

An approach based on Kriging is tested to validate the feasibility of gas concentration estimation over a geographic area from the limited set of MIR gas sensing systems, which enables to predict gas concentrations in zones without any close sensors. However, due to insufficient number of datasets to train a Kriging model, a public dataset is used since generating synthetic data is impossible in this case. Temperature measured at 40 different stations in France (See white dots in Figure 2(b)) in January 2020 is then used for training and validating the model. Ordinary Kriging is applied to the dataset to estimate the value of interest at locations without sensors. Kriging is basically an interpolation method based on Gaussian processes. A crucial advantage of Kriging is that it also estimates the variance of a prediction, which can be used to determine the prediction reliability. To evaluate ordinary Kriging for temperature estimation, a model is trained by data from 39 stations while excluding one station. Then, the temperature at the excluded station is predicted by the trained model and compared to the true value as shown in Figure 2(a), showing a good agreement. However, it must be noticed that predicted temperatures are under-estimated for stations located on the coast, which is attributed to two reasons: the ocean climate and fewer neighbouring stations, which reduces the estimation accuracy. By contrast, stations, where temperatures are over-estimated are located at higher altitudes than other stations. Therefore, over-estimated temperatures can be explicitly explained since temperature usually decreases with altitude and the model is trained using the station location as inputs but not their altitude. Bear in mind that it is possible to add the altitude information in the Kriging framework to improve estimation performance.

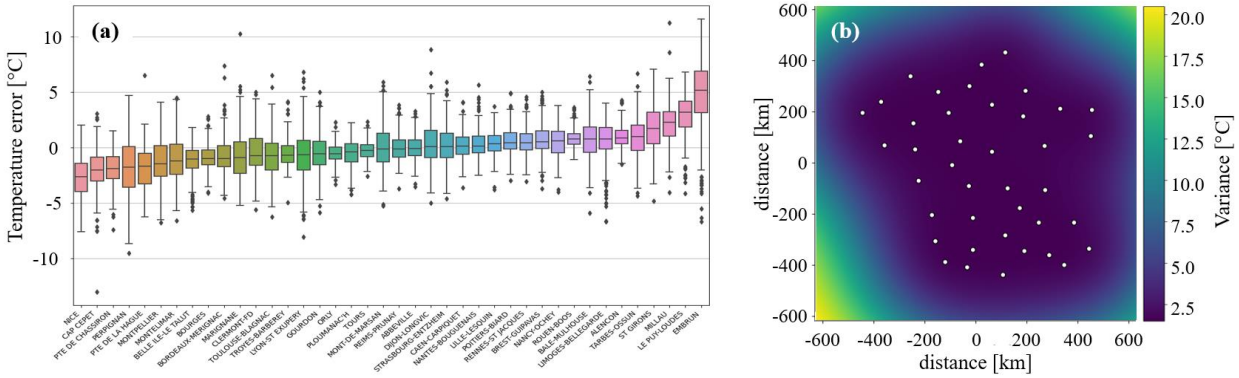

Figure 2: As a result of the Kriging prediction by a Gaussian covariance model, (a) temperature estimation errors for each station computed with a Gaussian covariance model and (b) estimated variance while white dots denote measurement stations.

4. Conclusions

We have demonstrated successful use cases of machine learning and Kriging for the MIR gas spectroscopy. Deep learning algorithms can accurately compute the gas concentration from noisy absorbance spectra through random spectral shifts. Moreover, Kriging interpolation appear to be a promising solution to precisely estimate the gas concentrations over a geographical area as demonstrated here with temperatures from a set of weather stations. Overall, the interdisciplinary approach of ML methods and sensor technologies can benefit other research fields.

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

5. References

[1] Z. Du *et al*, "Mid-Infrared Tunable Laser-based Broadband Fingerprint Absorption Spectroscopy for Trace Gas Sensing: A Review", Appl. Sci. **9**, 338 (2019).

[2] M.A. Abbas *et al*, "Fourier transform spectrometer based on high-repetition-rate mid-infrared supercontinuum sources for trace gas detection", Opt. Express **29**, 22315 (2021).

[3] M.A.Z. Chowdhury *et al*, "Evaluation of machine learning methods for classification of rotational absorption spectra for gases in the 220-330 GHz range", Appl. Phys. B **127:34** (2021).

[4] S. Chin *et al*, "Artificial Neural Network to Eliminate Detrimental Spectral Shift on Mid-Infrared Gas Spectroscopy", Sensors **23**, 8232 (2023).