Journal of Quantitative Spectroscopy & Radiative Transfer 211 (2018) 115–122

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Notes

Applicability of neural networks to etalon fringe filtering in laser spectrometers

J.M. Nicely[a,b,*], T.F. Hanisco[b], H. Riris[b]

[a] *Universities Space Research Association, Columbia, MD, USA*
[b] *Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA*

ARTICLE INFO

Article history:
Received 19 January 2018
Revised 5 March 2018
Accepted 5 March 2018
Available online 10 March 2018

Keywords:
Neural networks
Diode laser spectroscopy
Etalon fringes
Atmospheric retrievals

ABSTRACT

We present a neural network algorithm for spectroscopic retrievals of concentrations of trace gases. Using synthetic data we demonstrate that a neural network is well suited for filtering etalon fringes and provides superior performance to conventional least squares minimization techniques. This novel method can improve the accuracy of atmospheric retrievals and minimize biases.

1. Introduction

High resolution Tunable Laser Absorption Spectroscopy (TLAS) is a sensitive spectroscopic technique that uses absorption of laser light to measure concentrations of trace gases at a specific frequency or wavelength using Beer's law [1]. TLAS offers high sensitivity and relative immunity to interferences from other gases as long as a suitable spectral "fingerprint" of the gas of interest is available, along with a suitable laser source and detector. The invention of the diode laser has enabled different variants of TLAS, such as frequency or wavelength modulation spectroscopy [2–5], photo-acoustic spectroscopy [6,7], and cavity ring-down spectroscopy [8,9]. These techniques are used in diverse applications ranging from explosives detection [10], industrial process control [11], greenhouse gas monitoring [12,13] and the search for life on Mars [14]. Recent advances in fiber amplifier technology and other high power laser sources have enabled new applications from airborne platforms [15–20]. Scaling the laser power will enable new remote sensing instruments that measure concentrations of greenhouse gases on Earth or biogenic gases on Mars from space.

The first active (laser) space mission using Integrated Path Differential Absorption (IPDA), is a collaboration between the French Centre National d' Etudes Spatiales (CNES) with the German Aerospace Centre (DLR) to measure methane in the near infrared. The mission, called MERLIN (Methane Remote Sensing Lidar Mission), scheduled for launch in 2021 uses a two-wavelength optical parametric oscillator to measure the integrated path differential absorption of a methane line at 1645.5 nm [21]. IPDA measures the differential absorption of a gas column at two or more wavelengths and estimates its concentration. In the United States, the 2007 National Research Council Decadal Survey for Earth Science [22] has directed NASA to implement an active laser mission called ASCENDS (Active Sensing of CO_2 Emissions over Nights, Days, and Seasons) to measure CO_2 from space [23].

The high accuracy and precision needed for these space based measurements pose significant challenges for instrument design and signal processing. All spectroscopic instruments (ground-based, space, or airborne) regardless of their specific measurement approach have both systematic and random error sources that limit their accuracy and precision. The instruments must have high signal to noise ratio (SNR) to meet the precision requirements, and very low systematic errors, sometimes also referred to as "bias", "stability", or "drift", to meet the accuracy requirements. High SNR is typically accomplished by scaling the laser power and using sensitive low noise detectors and electronics to achieve shot noise limited performance. However, a high SNR does not necessarily lead to low bias measurements over long time periods. "Low bias" measurements are much harder to achieve and the SNR is not a good metric in measuring the accuracy or "drift" of the instrument. Instead, the concept of Allan variance, first introduced to assess the stability of frequency standards, has been successfully applied by Werle to TLAS and spectroscopic instruments to evaluate their accuracy and precision [24,25].

---

[*] Corresponding author at: Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

E-mail address: julie.m.nicely@nasa.gov (J.M. Nicely).

https://doi.org/10.1016/j.jqsrt.2018.03.004

Terms, like “drift” or “bias”, describing the accuracy and precision of an instrument over a certain time interval are often used interchangeably and in a rather qualitative fashion. Eisenhart, in a seminal paper in 1963, provided quantitative definitions of these concepts for instrument calibration, which have been adopted by the National Institute of Standards and Technology (NIST) and applied to several industries [26]. For a random process, precision usually represents the width or standard deviation of a distribution of successive measurements of a specific variable. The distribution is often assumed to be normal (Gaussian) and the mean or average of the distribution should converge to the “true” or expected value. Ideally, averaging successive measurements in a random process should reduce the width of the distribution and improve the precision. In a real instrument, even though the precision or standard deviation may improve by increasing the number of measurements, the actual measured mean value may converge to a different value than the “true” or expected value. The difference between the “true” value and the actual measured mean value is typically called “bias”, “drift”, or “systematic error” and is a measure of the accuracy. More importantly, this bias may change over time in a non-linear and non-periodic fashion. Increasing the number of measurements and/or averaging period will not improve the accuracy of the measurement and depending on the time constant involved, it may actually degrade it.

If the source of the bias can be identified and adequately modeled (for example the temperature dependence of a detector, improved knowledge of spectroscopic or atmospheric parameters, surface reflectance, etc.), a time-dependent “bias correction” may be introduced to calibrate and remove the bias and improve the accuracy of the measurement. Often even if the source of the bias is not explicitly identified, the bias correction may still be implemented to improve the accuracy of an instrument. In addition, many in-situ and ground-based TLAS instruments use frequent calibrations with an external, NIST-traceable standard to correct biases and verify the performance of the instrument. For a space or airborne instrument however, a verifiable external calibration standard is not always possible.

These issues of course, are not new, have been recognized for decades, are common to virtually all spectroscopic instruments, both passive and active, and considerable effort has been devoted to addressing them. However, the high spectral purity of lasers in TLAS and IPDA instruments presents some unique challenges. Unlike grating and Fourier spectrometers, TLAS instruments have much higher resolution and the laser linewidth is essentially a delta function compared to most atmospheric lineshapes. This is advantageous since it provides specificity, relative immunity to interferences from other gases and there is no need for instrument slit function calibrations. However, because of the high spectral resolution TLAS is more susceptible to interference (etalon) fringes. Etalon fringes are unwanted optical interference patterns that arise from multiple weak reflections from each optical surface in the optical path. They manifest themselves as a varying baseline structure in the absorption lineshape and they could, in principle, be subtracted or filtered out, if they remained stable in time. However, their phase, period and amplitude are a function of small path-length changes due to opto-mechanical shifts, vibration, temperature and pressure changes and changes in the index of refraction of various optical elements. As a result, etalon fringes introduce a time-dependent, non-stationary, background structure that is superimposed and is often indistinguishable from the signal (absorption lineshape) and cannot be filtered out without affecting the signal. Identifying the optical elements involved in etalon fringes using their free spectral range is possible only if there are only one or two dominant and easily identifiable fringes that can be adequately sampled. In practice, in a realistic system, there can be multiple fringes with different frequency components with varying amplitude and phase, making them impossible to adequately sample and identify.

Etalon fringes have been shown to limit the precision, accuracy and averaging time of TLAS spectrometers. Ways to mitigate them and reduce their impact include careful opto-mechanical design using wedged and anti-reflection coated optics, the use of reflective instead of transmissive optics, frequent calibrations using a reference cell with a calibrated gas mixture and/or background (null) gas, mechanical or electronic dithering methods and various digital signal processing (DSP) techniques. Some of these methods, like a reference cell with a calibration and/or background gas, are mostly applicable to in-situ spectrometers but are not easily adaptable to airborne or space measurements. Others, like mechanical dithering, may not be desirable in a space instrument where the use of mechanisms that can fail in orbit is not considered good engineering practice. Signal processing techniques however, are inexpensive, easily adaptable to various instrument configurations, and can be changed in orbit.

Kalman [27,28] and adaptive filtering [29,30] have shown some promise in suppressing fringes. Singular Value Decomposition (SVD) and principal component analysis (PCA) algorithms have also shown considerable promise [31] in removing fringe-induced bias. However, some etalon fringes cannot be filtered with DSP without affecting the signal because they have free spectral ranges and frequency components similar to those of the signal (absorption lineshape). Standard methods of atmospheric retrievals using optimal estimation (OE) and linear least squares minimization techniques [32] that fit the lineshape to an atmospheric transmittance model suffer from a similar limitation. They work extremely well if the noise is randomly distributed and etalon fringes have a different spectral decomposition than the signal. Most of the optimal estimation approaches generally try to minimize some “cost function” and often assume random or near random stationary noise. However, fringes are neither random, nor stationary. As a result, least squares minimization will invariably try to fit the fringes in addition to the absorption lineshape and the measurement will have a time-dependent bias that changes as the fringes change.

For large amplitude fringes, it can be obvious that the fit is incorrect but when long-term, high-precision and -accuracy measurements are required, these changes in background structure are very difficult or impossible to discern. Even if high precision (high SNR) is achieved in an instrument, a varying bias introduced into the measurement as the fringes change over time can be erroneously attributed to real changes in the absorption lineshape and consequently the trace gas concentration. Repeated calibrations or other statistical metrics such as the root mean square error of the fit are useful in identifying bias but cannot necessarily remove it. So the question arises: is there another signal processing approach that can be used to identify and perhaps correct bias? In this paper, we will examine the application of a neural network to simulated absorption lineshapes and show it can potentially provide better performance than optimal estimation approaches.

2. Neural networks

Artificial neural networks are a machine learning technique capable of modeling non-linear, high dimensional systems. They accomplish this by mimicking the massively parallel processing that occurs within biological neural networks, mathematically linking a set of inputs to some output via a system of interconnected nodes [33]. Neural networks have been applied to a diverse range of problems including image and speech recognition, stock market prediction, and pattern classification [34]. Within the physical sciences, neural networks have been used to forecast smog events [33], optimize aerosol optical depth [35] and cloud-top height [36] retrieved by satellite, estimate dissociation energies and rate