

Advanced Analysis Techniques and Deep Learning for Atmospheric Measurements

Dissertation

zur Erlangung des Grades eines
'Doktor rerum naturalium (Dr. rer. nat.)'
der Fachbereiche

- 08 - Physik, Mathematik und Informatik,
- 09 - Chemie, Pharmazie und Geowissenschaften,
- 10 – Biologie,
- Universitätsmedizin

der Johannes Gutenberg-Universität Mainz

angefertigt am Max-Planck-Institut für Chemie

Max Planck Graduate Center

vorgelegt von

Lenard Lukas Röder

geboren am 24.11.1996 in Wiesbaden

Mainz, 11.12.2023

1. Gutachter: Prof. Dr. Jos Lelieveld

2. Gutachter: Prof. Dr. Patrick Windpassinger

Tag der mündlichen Prüfung:
19.02.2024

Declaration

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg-University Mainz to insure standard scientific practice.

A handwritten signature in black ink, appearing to read 'Lenard Röder', written in a cursive style.

Lenard Lukas Röder
Mainz, 11.12.2023

Acknowledgements

This section has been removed from the online version due to data protection measures.

Abstract

This work explores a wide range of data analysis and signal processing methods for different possible applications in atmospheric measurements. While these methods and applications span a wide area of disciplines, the evaluation of applicability and limitations and the results of this evaluation show many similarities.

In the first study, a new framework for the temporal characterization of airborne atmospheric measurement instruments is provided. Allan-Werle-plots are applied to quantify dominant noise structures present in the time series. Their effects on the drift correction capabilities and measurement uncertainty estimation can be evaluated via simulation. This framework is applied to test flights of an airborne field campaign and reveals an appropriate interval between calibration measurements of 30 minutes. During ground operation, the drift correction is able to reduce the measurement uncertainty from 1.1 % to 0.2 %. Additional short-term disturbances during airborne operation increase the measurement uncertainty to 1.5 %.

In the second study, the applicability and limitations of several noise reduction methods are tested for different background characteristics. The increase in signal-noise-ratio and the added bias strongly depend on the background structure. Individual regions of applicability show almost no overlap for the different noise reduction methods.

In the third study, a fast and versatile Bayesian method called sequential Monte Carlo filter is explored for several applications in atmospheric field experiments. This algorithm combines information provided via the measurements with prior information from the dominant chemical reactions. Under most conditions the method shows potential for precision enhancement, data coverage increase and extrapolation. Limitations are observed that can be analyzed via the entropy measure and improvements are achieved via the extension by an additional activity parameter.

In the final study, state-of-the-art neural network architectures and appropriate data representations are used to reduce the effect of interference fringes in absorption spectroscopy. Using the neural network models as an alternative to linear fitting yields a large bias which renders the model approach not applicable. On the task of background interpolation the neural network approach shows robust de-noising behavior and is shown to be transferable to a different absorption spectrometer setup. Application of the interpolation to the test set lowers the detection limit by 52 %.

This work highlights the importance of in-depth analysis of the effects and limitations of advanced data analysis techniques to prevent biases and data artifacts and to determine the expected data quality improvements. An elaboration of the limitations is particularly important for deep learning applications. All presented studies show great potential for further applications in atmospheric measurements.

Zusammenfassung

In dieser Doktorarbeit wird eine breite Palette von Datenanalyse- und Signalverarbeitungsmethoden für verschiedene mögliche Anwendungen für atmosphärische Messungen untersucht. Während diese Methoden und Anwendungsfälle verschiedene Disziplinen umfassen, lassen sich viele Gemeinsamkeiten in der Analyse von Anwendbarkeit und Limitierungen sowie den Ergebnissen dieser Analyse aufzeigen.

Ein neues Verfahren für die zeitliche Charakterisierung von flugzeuggetragenen atmosphärischen Messungen wird in der ersten Studie definiert. Mithilfe von Allan-Varianz Graphen werden dominante Hintergrundstrukturen der Zeitreihe quantifiziert. In Simulationen können die Effekte dieser Strukturen auf die Driftkorrektur und Schätzung der Messunsicherheit erforscht werden. Das neue Verfahren wurde auf Daten von Testflügen einer Flugmesskampagne angewandt und bestätigte ein ausreichendes Intervall von Kalibrierungen von 30 Minuten. Während des Bodenbetriebs konnte durch die Driftkorrektur die Messunsicherheit von 1.1 % auf 0.2 % gesenkt werden. Zusätzliche kurzweilige Störungen während des Flugbetriebs erhöhten die Messunsicherheit auf 1.5 %.

In einer zweiten Studie wurden die Anwendbarkeit und die Limitierungen verschiedener Methoden zur Rauschunterdrückung für unterschiedliche Hintergrundstrukturen getestet. Die Verbesserung des Signal-Rausch-Verhältnisses und das Bias zeigen dabei eine starke Abhängigkeit von der Hintergrundstruktur. Bereiche der Anwendbarkeit für die individuellen Methoden zeigen kaum Überlappungen.

Der sequentielle Monte Carlo Filter, eine schnelle und vielseitige Methode aus der Bayes'schen Statistik, wird in einer dritten Studie für mehrere Anwendungen in atmosphärischen Feldmessungen untersucht. Der Algorithmus vereint Informationen aus den Messungen mit bekannten chemischen Reaktionen. In den meisten Fällen zeigt diese Methode großes Potential für die Verbesserung von Präzision, Datenabdeckung und Extrapolation. Mithilfe der Entropie können Limitierungen der Methode untersucht werden und die Einführung eines zusätzlichen Parameters führt zu verbesserter Robustheit.

In der letzten Studie werden Architekturen von neuronalen Netzwerken auf dem neuesten Stand der Technik und entsprechende Repräsentierung von Daten verwendet, um die Effekte von Etalon Interferenzen auf Absorptionsspektren zu verringern. Als Alternative für lineare Regression wird wegen eines hohen Bias keine gute Anwendbarkeit erzielt. Als Alternative zur linearen Interpolation von Hintergrundspektren zeigen die neuronalen Netzwerke robuste Ergebnisse zur Rauschunterdrückung. Mit geringem Trainingsaufwand können diese Netzwerke auch auf ein anderes Spektrometer angewandt werden. Im Test-Datensatz wurde das Detektionslimit um 52 % gesenkt.

Zusammenfassung

Die Resultate dieser Doktorarbeit heben hervor, wie wichtig eine detaillierte Analyse der Effekte und Limitierungen von Datenanalysemethoden ist, um Artefakte und Verzerrungen zu verhindern und die Verbesserungen der Datenqualität abzuschätzen. Die Analyse der Limitierungen ist besonders wichtig in Anwendungen von tiefen neuronalen Netzwerken. Alle Studien zeigen darüber hinaus großes Potential für zukünftige Anwendungen in atmosphärischen Messungen.

Contents

Abstract	v
Zusammenfassung	vii
1 Introduction	1
2 Theory and Methods	5
2.1 Atmospheric Measurements	5
2.2 Infrared absorption spectroscopy	8
2.3 Deep Learning	13
3 Temporal Stability and Instrument Performance of an airborne QCLAS	19
3.1 Introduction	20
3.2 The TRISTAR instrument	20
3.3 Ground operation inside the HALO aircraft	23
3.4 Airborne operation	26
3.5 Summary and Conclusion	30
4 Applicability of Advanced Noise Reduction Methods	33
4.1 Introduction	33
4.2 Noise reduction methods	34
4.3 Simulation	38
4.4 Simulation results	40
4.5 Conclusion	45
5 Data Quality Enhancement via Sequential Monte Carlo Filters	47
5.1 Introduction	48
5.2 Sequential Monte Carlo	49
5.3 Chemical reactions and measurements	52
5.4 Experiments	55
5.5 Conclusions	62
6 Neural Network Architectures for Absorption Spectroscopy	65
6.1 Introduction	66
6.2 Neural Network Architectures	68
6.3 Experimentation and Network Training	72
6.4 Results and Discussion	74
6.5 Summary and Conclusion	80

Contents

7 Discussion and Outlook	83
Appendix	87
1 Chemical reactions and formulas	87
2 Supplement to Chapter 3	87
3 Supplement to Chapter 4	89
4 Supplement to Chapter 5	89
5 Supplement to Chapter 6	95
Bibliography	99

1 Introduction

The earth's atmosphere features complex interactions between trace gases, aerosol particles, clouds and sunlight. Understanding these processes is necessary to develop highly accurate models that can predict the sources and effects of climate change and air pollution. Insights into the complex chemical system can be gained via coordinated field measurements of several chemical and meteorological parameters.

Atmospheric chemistry measurements feature a wide range of different experimental conditions. The requirements to temporal resolution, temporal stability, precision and sensitivity depend greatly on the particular species and circumstances and can range over several orders of magnitude. One method capable of providing the necessary requirements for several trace gases is laser absorption spectroscopy. This particular method has been significantly improved on a hardware basis during the last decades [1, 2, 3, 4, 5], but the software data analysis is mostly based on conventional and basic mathematics and statistics. Many new frameworks for data analysis including artificial neural networks have been established recently. Some have already covered applications to laser absorption spectroscopy [6, 7, 8], but the data analysis of state-of-the-art measurements still relies on conventional approaches [9].

An important milestone in environmental data science is the work by C. Torrence and G.P. Compo, "A practical guide to wavelet analysis", published 1998 in the Bulletin of the American Meteorological Society [10]. This article reports the application of the advanced data analysis technique called continuous wavelet transform (CWT) to the oceanography field El Nino-Southern Oscillation. Their article features a step-by-step practical guide to the analysis done along with important remarks, extensions and limitations. This work has been recognized by over 10,000 citations¹ in research articles that apply wavelet transforms in various disciplines.

What separates the work by Torrence and Compo from the numerous studies developed during the last decades? This characteristic can be explored in a simple mathematical consideration:

Definition. An experiment E records data D to gain information about an observable Y . The data is prone to various noise sources and experimental limitations. A possible reconstruction X of the target observable can be obtained from the recorded data via signal processing C . The variable X follows some probability distribution that depends on the observable, the experiment and the signal processing. Performance metrics M

¹Web of Science, 24th July 2023

1 Introduction

can be defined that quantify the capability of X to reconstruct Y .

$$\begin{aligned} E &: Y \rightarrow D \\ C &: D \rightarrow X \\ X &\sim P(Y, E, C) \\ 0 &\leq M(Y, X) \end{aligned}$$

Example. The QCLAS-instrument (quantum cascade laser absorption spectroscopy) TRISTAR² [11] E records Wavelength Modulation Spectroscopy (WMS) spectra D of a trace gas mixing ratio Y . Fluctuations in laser power and frequency, temperature and pressure inside the absorption cell, optical noise, detector noise, interference with other absorbing molecules and data acquisition electronics induce noise to the acquired spectra D . The mixing ratio is reconstructed via linear fitting against a calibration spectrum to obtain X . The probability distribution can be mostly assumed Gaussian with combined relative and absolute noise effects that may be expressed by the precision P and the detection limit DL of the instrument:

$$\sigma_X^2 = DL_{1\sigma}^2 + X^2 P^2$$

Definition. (Continued). An advanced analysis technique C^* provides an alternative reconstruction X^* of the observable. The performance compared to the classical approach C can be expressed by comparison of metrics M . In contrast to the classical approach, advanced analysis techniques are often accompanied by parameters λ that have to be chosen accordingly to provide the desired denoising properties: $C^* = C_\lambda^*$. Since X and X^* depend on E , the performance increase of C^* may also depend on E . Under certain experimental conditions there even might not exist a set of parameters λ that provide a performance increase.

Therefore, a research article that reports a performance increase of a specific set of C^* , λ and E does not show that C^* is applicable for any experiment with similar conditions. The strengths and limitations of C^* have to be elaborated to provide this information. Most of the studies that implement advanced analysis techniques focused on their specific experimental setup. This may impact the significance of the applicability and the limitations of the method presented since only a small portion of the possible search space is explored. This may have an effect on the of reproducibility in other experiments and is therefore associated to the number of citations and the benefit for the scientific community.

A special tool in data scientific tasks is the subfield of machine learning called deep learning, that has lead to many breakthroughs in artificial intelligence research during the last decades and started a new AI hype. Applications have been spreading to many disciplines and are ubiquitous in the public media [12]. Considering the definitions above, a challenge for the application of neural networks in natural sciences arises. In deep learning, the search space of λ is very large. The parameters include

²Tracer in-situ TDLAS (tunable diode laser absorption spectrometer) for atmospheric research

model topology, size, optimizer choice and learning rate. Therefore it is not possible to prove that for a certain scenario E there exists no neural network C_λ^* that provides a performance increase compared to the classical approach. In the literature, only positive examples of neural network applications are reported, as it can never be excluded that a different setup of λ might achieve the desired outcome. Furthermore, due to the black box properties of neural networks, the explainability and trustworthiness of neural networks is widely discussed [13, 14]. Thus, critical discussion of the limitations is mandatory.

The goal of this dissertation is the qualitative and quantitative evaluation of the applicability of recent analysis methods to atmospheric measurements. A primary focus is pointed to the data analysis of spectra produced by the WMS instrument TRISTAR. Designed and fine-tuned methods can also be evaluated using a different spectrometer and multiple trace gases. The study will put special emphasis on advanced analysis methods based on neural networks. The thesis is organized as follows: Chapter 2 covers the relevant theory and methods of atmospheric measurements and deep learning. Chapter 3 describes the TRISTAR instrument in detail and applies data scientific methods to the characterization of its temporal stability. Chapter 4 consists of a published article which aims to compare different advanced analysis techniques under different experimental conditions. Chapter 5 consists of a published article that introduces a novel statistical method to improve data quality in atmospheric measurements. In Chapter 6 neural network architectures are explored as an alternative to conventional fitting and interpolation in absorption spectroscopy applications. Chapter 7 discusses the key results of the individual studies and provides an outlook regarding future applications and opportunities.

2 Theory and Methods

2.1 Atmospheric Measurements

Atmospheric measurements provide insights into mixing ratios and number concentration of trace gases and aerosol particles, respectively, photolysis frequencies and meteorological properties, e.g., wind velocity, ambient temperature and pressure, cloud coverage and precipitation. Typical measurement modes are remote sensing techniques aboard satellites, continuous ground based measurements and field campaigns at ground stations or aboard moving platforms, e.g., vehicles, aircraft, weather balloons, drones and ships. The goals of atmospheric measurements are the investigation of hazardous air pollution and the improvement of our understanding of chemical processes, furthermore the validation of weather and climate models for better weather forecasting and climate modeling, particularly modeling of global warming and climate change [15]. Measurement techniques feature optical spectroscopy approaches [16], mass spectrometry [17], chemiluminescence [18], fluorimetry [19] and many others. Depending on the measured species and measurement platform, different requirements are needed regarding measurement precision, accuracy, time resolution, selectivity, in-situ or sample collection, autonomous or manual operation [15].

This section features a short description of trace gases and their dynamics in the troposphere as well as a description of measurement uncertainty and its composition. Sect. 2.2 provides a detailed description of Infrared Absorption Spectroscopy techniques. Details about other measurement techniques as well as further processes in the atmosphere, e.g., heterogeneous reactions, particle formation and aerosol interactions are not discussed in this thesis. Suggestions for the interested reader include Heard et al. (2006) [20] that features an extensive elaboration of common and state-of-the-art atmospheric measurement techniques and their challenges, and Seinfeld and Pandis (1998) [21], who present the dynamics of trace gases and aerosol particles throughout the atmosphere.

2.1.1 Trace gases in the troposphere

The mixing ratio χ defines the relative contribution to the total air volume of a particular gas species and can be related to the concentration c through the ideal gas law:

$$\chi = \frac{V}{V_{\text{air}}} = \frac{n}{V_{\text{air}}} \frac{RT}{p} = \frac{RT}{p} c \quad (2.1)$$

The earth's atmosphere mainly consists of nitrogen (N_2) and oxygen (O_2), with $\chi = 0.79$ and $\chi = 0.20$, respectively. However, local and global properties of the

2 Theory and Methods

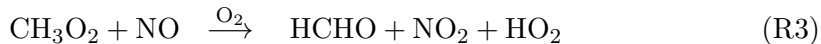
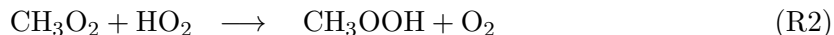
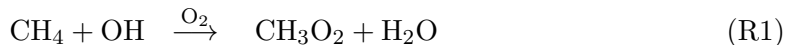
atmosphere, e.g., chemical reactions, particle formation, radiative forcing and air pollution are driven by trace gases which are much less abundant. The greenhouse effect is dominated by tropospheric carbon dioxide (CO₂), methane (CH₄), ozone (O₃), nitrous oxide (N₂O) and water vapour (H₂O). Some of these compounds only contribute to the total air in the parts per billion volume (ppbv) range. Atmospheric chemistry in the troposphere is driven by chemical reaction with radicals with typical mixing ratios in the parts per trillion volume (pptv) range [21].

Junge (1974) [22] reported an anti-proportionality between the atmospheric variability of a trace gas and its residence time in the troposphere. This highlights how studying the mixing ratios of low abundant gases can shed light on the dominant dynamical processes in the troposphere. The temporal change of the concentration of a trace gas g can be described by several contributions, with the (photo-)chemical production P , the (photo-)chemical loss L , transport T , emission E and deposition D :

$$\frac{d}{dt}c_g = P(g) - L(g) + E(g) - D(g) \pm T(g) \quad (2.2)$$

By assuming a steady state and calculating the dominant production and loss contributions via (photo-)chemical reactions, deviations from the steady state due to other chemical reactions, emission, deposition and transport can be investigated [23]. Minimizing the uncertainty of these processes has an impact on the quality and accuracy of global climate and chemistry model forecasts.

The variety of different reactive species enables a large number of possible chemical reactions. This creates several competing pathways in the cycles and catalysis of trace gases which dictate the net effect of the interplay of these reactions. An example of this phenomenon is visible in the oxidation of CH₄ and volatile organic compounds (VOC). CH₄ forms the methyl peroxy radical CH₃O₂ through reaction with the OH radical and subsequent combination with O₂. This intermediate either reacts with NO or HO₂ to form HCHO or methylhydroperoxide (CH₃OOH), respectively [21]. The former reaction yields an HO₂ radical which can be recycled to OH by reaction with NO. Ozone is then formed via photolysis of NO₂. The dominant pathway dictates the net ozone production of the system [24]. The described reactions are summarized in reactions (R1) through (R5).



Ozone itself is a greenhouse gas, an air pollutant and the primary source of the OH radical in the troposphere [25]. Thus, measuring an ensemble of trace gases participating in these processes is required to understand the dominant chemistry and its effects on air pollution and global climate [21].

2.1.2 Measurement Uncertainty in Atmospheric Measurements

Measurement devices always feature some uncertainties that emerge from statistical and systematic effects of the measurement procedure. The total measurement uncertainty is a composition of several different types of errors that each contribute differently. A very important distinction to be made are the separation of statistical and systematic errors, which is equivalent to the distinction between precision and accuracy as in the picture of a marksman and a shooting target described in Heard et al. [15]. Statistical error, or variance, or precision, refers to a spread around the mean value of an ensemble of measurements X . Systematic error, or bias, or accuracy, refers to the deviation of the mean value from the true value Y . Often the mean squared error (MSE) is defined as the sum of variance and squared bias:

$$\text{MSE} = \text{E} \left[(X - Y)^2 \right] = \text{Var}(X) + (\bar{X} - Y)^2 \quad (2.3)$$

Signal processing techniques try to minimize the MSE to accomplish both objectives, low variance and low bias, at once. However, if further data analysis and processing is intended, low bias is preferred over low variance since further averaging has no effect on the bias. Therefore, biased data might lead to wrong conclusions. Thus, signal correction can be useful that lowers the bias while increasing the statistical error σ .

Example. A measurement device aboard an aircraft detects the mixing ratio through the concentration of a trace gas inside a measurement cell operated at cabin pressure. However, the pressure directly influences the conversion from concentration into mixing ratio according to the ideal gas law. Thus, a bias is introduced into the measurement that correlates cabin pressure to mixing ratio. For pressure correction, a pressure sensor is attached to the measurement cell. The statistical error of the pressure sensor contributes to the total statistical uncertainty in order to decrease the systematic error.

Typically the precision P is defined as the relative statistical error. This implies (2.4a), but other power dependencies of σ_X from X are possible. In a counting experiment, for example, the noise is scaled through (2.4b). In most applications, the major contributions can be split into relative and absolute noise, where the detection limit (DL) (2.4c) refers to the noise present at zero mixing ratio and therefore refers to the capability of the measurement device to distinguish a signal from background noise. McManus et al. [26] refer to light and dark noise, and proportional noise, with respect to absolute and relative noise sources, respectively.

$$\sigma_X = PX \quad (2.4a)$$

$$\sigma_X \sim \sqrt{X} \quad (2.4b)$$

$$\sigma_X \stackrel{X \approx 0}{=} \text{DL}_{1\sigma} \quad (2.4c)$$

The detection limit DL is often given with an additional factor of 2

$$\text{DL}_{2\sigma} = 2 \cdot \text{DL}_{1\sigma},$$

2 Theory and Methods

which allows for an alternative interpretation of the detection limit from the perspective of hypothesis testing. A measured data point X reported from the instrument is tested against the null hypothesis H_0 that the true value is 0 and X is just a random fluctuation of the instrument. With a significance threshold of approximately 5 %, values that are less than $DL_{2\sigma}$ away from 0 cannot reject H_0 .

Another important property of uncertainties is the time dependency. Variations in ambient parameters that influence the instrument sensitivity lead to drifts. These drifts in sensitivity, if uncorrected, worsen the accuracy of the measurement. Regular measurements of calibration gas enables linear correction of these drifts, exchanging the systematic deviation with additional statistical uncertainty. This additional uncertainty, however, cannot be lowered through further averaging and sets a limit to the statistical error. Thus, the relative measurement uncertainty depends on the integration time τ by

$$\frac{\sigma_X(\tau)^2}{X^2} = \frac{P^2}{\tau} + R^2, \quad (2.5)$$

where R refers to the reproducibility of the calibrations. The same logic applies to absolute measurement uncertainty: Drifts in the instrument offset have to be corrected via background or zero measurements. These corrections enable more accurate low-signal measurements but determine the final detection limit.

Thus, a full description of the instrument performance for an atmospheric measurement device has to contain the following metrics:

- The time-independent relative error R
- The time-dependent relative error P at a certain τ
- The accuracy (e.g. due to provided measurement error of the primary calibration gas source)
- The time-independent and time-dependent parts of the detection limit DL

2.2 Infrared absorption spectroscopy

A method proven capable of measuring trace gases in the atmosphere is laser absorption spectroscopy. It has been described in detail, for instance, in [1, 5, 27].

Absorption spectroscopy makes use of quantum-mechanical transitions in specific energy levels of molecules. While atomic transitions typically lie in the ultraviolet and visible region of the spectrum of light, molecules can be identified from rotation-vibrational transitions that span from near infrared (NIR) to mid infrared (MIR). Each transition is broadened through natural broadening due to the natural lifetime of the excited state, yielding a Lorentzian profile. The transition line is further broadened through Doppler broadening and pressure broadening, that yield a mixture of a Gaussian and a Lorentzian profile, respectively. The absorption line corresponds to

2.2 Infrared absorption spectroscopy

a convolution of the Gaussian profile α_G and the Lorentzian profile α_L and is called Voigt profile α_V .

$$\alpha_G(\lambda) \sim \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}\right) \quad (2.6)$$

$$\alpha_L(\lambda) \sim \frac{1}{1 + \left(\frac{\lambda - \lambda_0}{\gamma}\right)^2} \quad (2.7)$$

$$\alpha_V(\lambda) = \alpha_G * \alpha_L \quad (2.8)$$

In absorption spectroscopy the concentration of a molecule of interest is measured from the absorption at the transition wavelength λ_0 . According to the Lambert-Beer-law (2.9), the strength of the absorption can be approximated linearly if the optical depth D is small enough. The optical depth is the product of the absorption strength S , the concentration c in molec/cm³ of the absorber along the beam path, the length of the beam path L and the beam width A .

$$\begin{aligned} I(\lambda) &= I_0 \exp(-\alpha_V(\lambda)D) \approx I_0 (1 - \alpha_V(\lambda)D) \\ \Rightarrow \Delta I &\simeq I_0 \alpha_V(\lambda)D \end{aligned} \quad (2.9)$$

For a high precision measurement it is crucial to maximize the optical depth as it directly translates to the signal amplitude. The absorption strength can be optimized via careful line selection. In principle, picking the transition with the highest available absorption strength is optimal, but interferences with other transition lines nearby have to be considered as well as the availability of a light source in the corresponding frequency region. The beam intensity per area depends on the light source used. Picking a higher power laser source yields an increased signal strength. The length of the beam path can be increased by geometric design of the absorption cell. However, a very big absorption cell is impractical in most applications and requires a high gas flow to ensure a reasonable temporal resolution. Special mirror alignments enable multiple passes through the absorption cell without beam interference. These multipass absorption cells (MPCs) range from the White cell [2] and the astigmatic Herriott cell [3, 28] to more complicated setups [29]. However, a high quality coating of the mirrors is mandatory that maximizes the reflectivity R of the surface for the target wavelength. Otherwise, the reduction of the beam intensity due to an increased number of reflections N mitigates the positive effect of an enhanced path length:

$$\Delta I \simeq I_0 R^N \alpha_V(\lambda) c S A L N \sim R^N N \quad (2.10)$$

In atmospheric sciences, two major modes of measurement operation are dominant. The first approach uses direct absorption techniques. This way, determination of the gas mixing ratio is possible without the need of calibration gas, since both the intensity I_0 and the intensity difference ΔI are measured which enables the determination of the optical depth. Challenges of this operation mode are interferences of abundant gases in the open optical path and baseline corrections [30]. The second approach

2 Theory and Methods

uses wavelength modulation to measure a harmonic of the absorption profile instead. Wavelength modulation suppresses baseline drift and open optical path absorption but requires a more complicated operation and makes calibration against a gas standard necessary [9].

Additionally to a scanning sweep, the laser frequency ν is modulated with a higher frequency sine-wave, typically in the range 1 - 100 kHz. The modulation depth a can be expressed by the dimensionless parameter m and the line width γ .

$$\nu = \nu_0 + a \cos(\Omega t) = \nu_0 + m\gamma \cos(\Omega t) \quad (2.11)$$

The detector signal is then demodulated with a synchronized sine-wave with n times the modulation frequency Ω . Integration or low-pass filtering yields the n -th harmonic of the absorption line H_n . Wavelength modulation spectroscopy is sometimes also called harmonic detection and derivative spectroscopy, since the n -th harmonic is closely related to the n -th derivative of the absorption profile [5].

$$H_n(\nu) = \int_0^{\frac{\pi}{\Omega}} \alpha_V(\nu + m\gamma \cos(\Omega t)) \cos(n\Omega t) dt \quad (2.12)$$

The order n governs the shape of the harmonic, signal strength and the bandwidth. A decent trade-off between bandwidth and signal strength is achieved via $n = 2$ and called $2f$ -spectroscopy. As a result of the experimental realization of the wavelength modulation in tunable diode lasers, an additional amplitude modulation of the laser intensity comes into effect that slightly distorts the harmonics up to the order 2. This phenomenon is called residual amplitude modulation [27]. In industrial gas sensing, the need for a calibration standard can be discarded through the use of $nf/1f$ techniques that adjust for laser intensity changes [31].

2.2.1 Optical etalon interference fringes

Relative uncertainty in infrared absorption spectroscopy is dominated by effects that linearly scale the intensity difference (2.10). These include variations in cell pressure and temperature, laser intensity and detector sensitivity. Small variations caused by the physical expansion of gas due to cell pressure and temperature can be easily corrected. This requires negligible influence to line broadening and line strength. Experimental setups such as $nf/1f$ spectroscopy enable correction for laser intensity and detector sensitivity variations [31]. Fast variations can be suppressed through higher integration times. Slow drifts can be suppressed via regular calibrations with the reference gas.

Absolute uncertainty limits the minimal optical depth an infrared absorption spectrometer is able to distinguish from background noise. While noise sources such as detector noise, laser shot noise and electronic noise are small compared to the absorption signals achieved by commonly used laser, detector and MPC setups, the main absolute limitation is caused by optical etalon interference fringes. These emerge from standing waves in the optical path between two reflecting surfaces with distance ℓ and

refractive index n of the medium and cause repeating Lorentzian peaks with a free spectral range (FSR)

$$\Delta\nu_{\text{FSR}} = \frac{c}{2n\ell} \quad (2.13)$$

and finesse

$$\mathcal{F} = \frac{\Delta\nu_{\text{FSR}}}{\Delta\nu_c}, \quad (2.14)$$

where $\Delta\nu_c$ refers to the width of the Lorentzian peak. In the low finesse limit, the fringes contribute an ensemble of sinusoidal interferences whose FSR may occupy the same frequency range as the absorption spectrum. The origin of fringes has been theoretically discussed [30, 32, 26, 33] and several studies have explored experimental techniques [34, 5, 27] and signal processing [35, 36, 37] to reduce the impact of fringes. Under some experimental conditions, there may be the possibility to filter the fringe if the FSR is much lower than the contributing frequencies to the absorption profile [36]. If the fringe pattern can easily be split into a small number of additive sine-waves, a Levenberg-Marquardt fit might approximate the background to reasonable precision for a fringe reduction [8]. However, in most ambient measurements, the fringes interfere and produce a rather complicated background pattern [32, 33]. Here, a fitting routine becomes unstable and filtering techniques also fail [8]. A common counteraction are regular measurements of absorption-free signals by flooding the measurement cell with purified gas and background subtraction from ambient measurements [38, 9]. This method requires temporal stability of the fringe pattern between consecutive background measurements and lowers the overall duty cycle.

Since the FSR depends on the distance between the reflecting surfaces, tiny changes due to ambient changes in temperature, pressure, or acceleration can influence the shape and position of the fringes. These ambient changes are ubiquitous in field experiments, especially aboard moving platforms, e.g. aircraft. Top-performing instruments feature temperature stabilization [11, 38, 9] and pressure stabilization [9] of the optical setup to reduce influence on fringe stability and allow for a lower detection limit and duty cycle.

2.2.2 Allan-Werle-Plots

Werle et al. [39] have introduced Allan-Werle-plots as a method for investigating time dependencies in spectroscopic measurements. Many publications [38, 29] report the temporal stability of their setup using such Allan-Werle-plots. However, many more insights can be gained as reported in [32].

The Allan variance or two-sample-variance consists of the mean squared difference between adjacent data points for different integration times τ :

$$\sigma_x(\tau)^2 = \frac{1}{2} \sum_{k=0}^{M-2} (x_{k+1}(\tau) - x_k(\tau))^2 \quad (2.15)$$

$$x_k(\tau) = \frac{1}{\tau} \int_{t=k\tau}^{(k+1)\tau} x(t) dt \quad (2.16)$$

2 Theory and Methods

The Allan variance of uncorrelated, white noise decreases proportional to τ^{-1} . Brownian noise, or random walk noise, increases with τ . The latter refers to random drifts in the data. When plotting the Allan variance versus the integration time, these are the major contributions that limit the maximum integration time [39]. However, other noise contributions also influence the structure of the Allan-Werle-plot. As shown by Werle [32], flicker noise or $1/f$ noise contributes a constant noise floor. Sinusoidal variations in the data lead to the aliasing pattern

$$\sigma^2 \sim \frac{\sin(\pi\tau f)^4}{(\pi\tau f)^2}, \quad (2.17)$$

where f refers to the frequency of the sine wave. Figure 2.1 shows an exemplary Allan-Werle-plot of data simulated via the following setup:

$$\begin{aligned} h_t &= h_{t-1} + \epsilon_B \\ f_t &= \mathcal{F}^{-1}\left(\frac{\epsilon_F}{2\ln(2)\omega}\right) \\ x_t &= 100 + h_t + \epsilon_W + \sin(2\pi t f) + f_t \\ &\text{with } \epsilon_B \sim \mathcal{N}(0, 0.04^2), \epsilon_W \sim \mathcal{N}(0, 2^2), f = 512^{-1}, \epsilon_F = 0.4 \end{aligned}$$

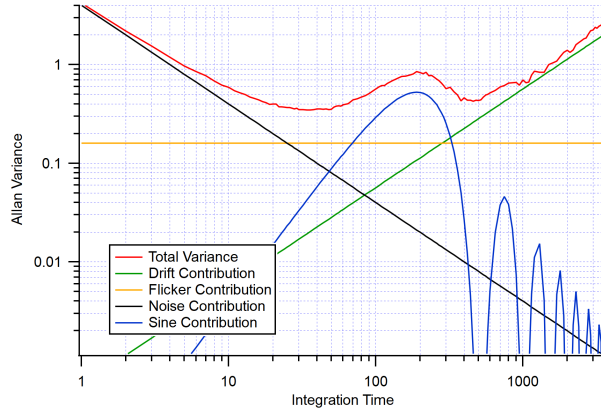


Figure 2.1: Allan-Werle-plot of simulated data containing white noise, random walk, flicker and sinusoidal contributions. The individual contributions are drawn separately in black (white noise), green (random walk), yellow (flicker) and blue (sinusoidal). The total Allan variance (red) is the sum of the individual variances. At a particular integration time, the Allan variance is dominated by specific contributions. In this case, at low integration time the white noise dominates. At intermediate integration times the Allan variance is dominated by flicker and sinusoidal contributions. For high integration times the random walk drifts take over.

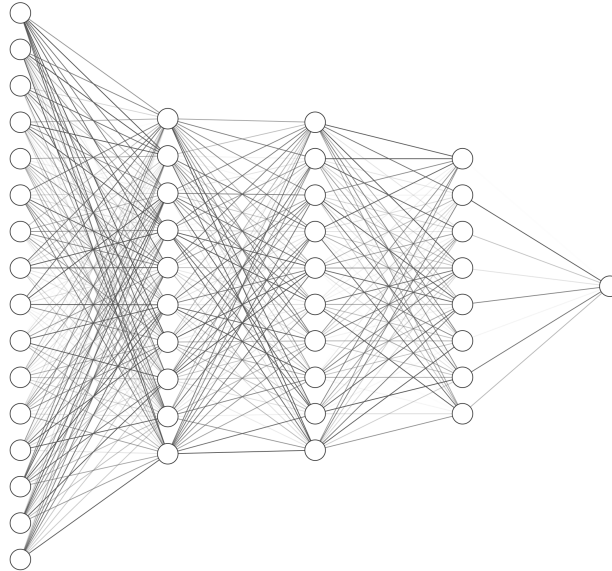


Figure 2.2: Schematic of a fully connected neural network. The nodes of each layer are drawn as circles, the weights are drawn as straight lines between the nodes. The opacity of the lines encodes the weight values. The shown FCNN consists of an input layer with size 16, hidden layers of sizes 10, 10 and 8, and a single output neuron. The image was created using [40].

2.3 Deep Learning

Machine learning is the field of data-driven optimization of complicated functions that can be applied in classification and artificial intelligence tasks. Deep learning is a novel subspace of that field which focuses on deep artificial neural networks (DNN). The word deep refers both to the depth of the network due to its number of layers, and to the depth of the representation that is built onto several layers of abstraction. Neural networks have a long and interesting history in neuroscience and computer science. Some of the key concepts of state-of-the-art deep learning have been developed 40 to 60 years ago. A comprehensive description of the field and its history is, e.g., given in the first chapter of the book by Goodfellow et al. (2016) [41]. The field of deep learning contains many advanced tools and practices that have been evaluated in practice and allow for an efficient training procedure and high inference performance [42, 43, 44, 45, 46, 47, 48]. In this quickly evolving field the best practices and network architectures are constantly changing. A comprehensive explanation of commonly used model design techniques will not be given in this section as it is beyond the scope of this thesis. This section will give a brief introduction on the most recent wave of neural networks and discuss common implementations and applications relevant for the field of this thesis.

The feed-forward multi-layer-perception (MLP) is the most primitive building block of

2 Theory and Methods

neural networks. It consists of iterated linear transformations represented via matrix multiplications W_k and a following non-linearity σ called activation function. The outputs of a layer y_k are inputs to the subsequent layer y_{k+1} . A schematic is shown in Figure 2.2.

$$y_{k+1} = \sigma(W_k y_k) \quad (2.18)$$

In the MLP architecture each node (neuron) of a layer is connected to each node of the next layer, thus, it is also called fully-connected neural network (FCNN). Since the weight matrices scale up with the product of the adjacent number of neurons, complex FCNNs become harder to train and are prone to overfitting. Overfitting is the phenomenon when the machine learning algorithm does not generalize the target distribution but instead adapts to specific features of the training samples [42]. It is analogous to memorizing facts by heart versus understanding the underlying structure.

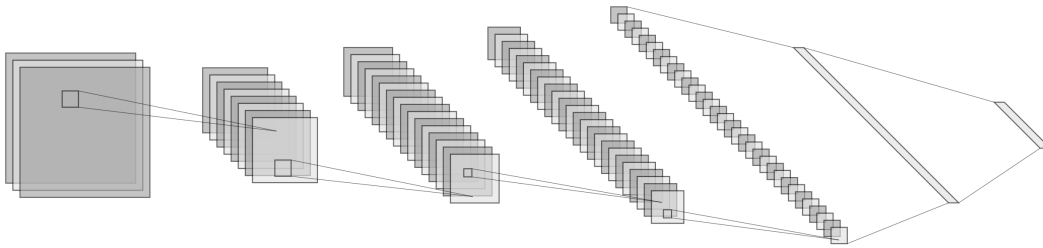


Figure 2.3: Schematic of convolutional neural network. The size of the rectangles encodes the width and height of the image. The stacked rectangles refer to the color maps and feature maps, respectively. In the beginning of the CNN the small squares show the effect of filters acting at each position in the images individually through equivariant operations, e.g., convolution filters and local pooling operations. At the end of the CNN a global pooling operation reduces the feature maps to a one-dimensional vector and a fully connected layer is applied to obtain the output. The image was created using [40].

Convolutional neural networks (CNN) replace the full matrix multiplications with convolutions of filters which kernel sizes are small compared to the dimension of the input data. These convolution filters share the same weights at each position in the image. This leads to a much lower number of weights and incorporates a translational symmetry. The layer output is equivariant under translations of the input image position. CNNs greatly improved the performance of image processing tasks [49, 50, 51, 52]. A schematic of a CNN is shown in Figure 2.3.

Recurrent neural networks (RNNs) include layers with hidden vectors that are modified and fed back into the layer for each input in a sequence. This enables time series and sequence processing and has been state of the art in natural language processing (NLP) before transformers were introduced. Several types of RNN building blocks were proposed. A major contribution was the long-short-term memory cell (LSTM) proposed by Hochreiter et al. [53]. It features a second hidden vector for longer correlation times. Recurrent neural networks require a lot of memory for training since the whole sequence is unfolded in the training process.

The attention mechanism was introduced by Vaswani et al. [54]. There a query sequence gates the actual input sequence to focus on important parts of the sequence at a time. This development quickly became state-of-the-art in NLP and led to similar architectures which use self-attention called transformer [55, 56, 57]. Transformer architectures dominate generative artificial intelligence applications like GPT (generative pre-trained transformer) [58].

2.3.1 Applications of neural networks

Neural network applications, inspired by the human brain's structure, have revolutionized technology and science, driving breakthroughs in diverse fields such as image recognition, natural language processing, and autonomous systems.

– ChatGPT¹ [59]

Neural networks feature many ground-breaking applications in various fields, e.g., image recognition [50], artificial intelligence [60] and – as demonstrated by OpenAI's chatbot ChatGPT – natural language processing [58].

Image recognition tasks range from classification [50, 57] over image segmentation [44, 61] to generation of images [47]. Most recent advancements in image generation enable high resolution image generation and image modification from text prompts [62]. Figure 2.4 shows example pictures generated from a diffusion model². Image recognition is useful for various applications, e.g., medical diagnosis [44, 63] and autonomous driving [64].

A famous application that has impacted the media and industry in early 2023 is natural language processing. Transformer architectures [55] based on the attention mechanism [54] sparked the era of large language models that can generate seemingly human-like text and are able to hold conversations [58, 12]. Chatbots similar to ChatGPT have been integrated into industry applications, e.g. personal assistants and customer support [65, 12], and are highly discussed in education [66].

2.3.2 Applications in atmospheric sciences and spectroscopy

Several studies have also explored applications of neural networks to spectroscopy tasks and atmospheric sciences. These applications can be grouped into different categories, dependent on the objective of the neural network used.

A typical application of machine learning in general is classification. Liu et al. [67] explored shallow CNN architectures for detection of extreme weather phenomena from global climate reanalysis datasets. Pham et al. [68] applied a neural network to pixel-based estimation of landslide susceptibility. Acquarelli et al. [69] applied CNNs to

¹ChatGPT (GPT-3.5) response to prompt "Write a 1-2 sentence paragraph that introduces into neural network applications"

²<https://stablediffusionweb.com/>

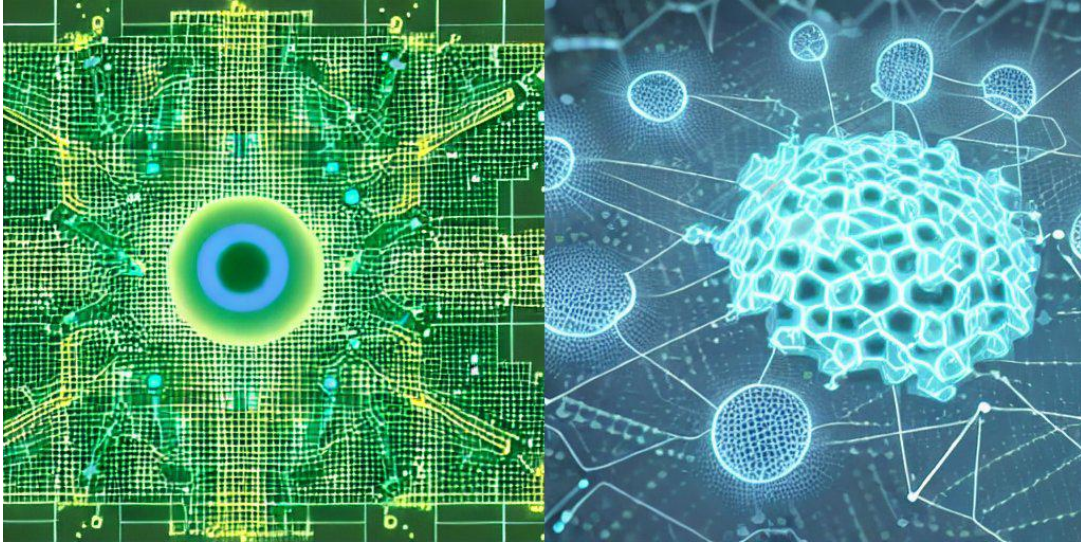


Figure 2.4: Example images generated from a diffusion model² [62] from the prompt "introduction to neural network applications".

classify spectroscopic data acquired from Fourier-Transform Infrared (FTIR), Near-Infrared (NIR) and Raman spectroscopy. Xia et al. [70] also used CNNs to distinguish NIR and Mid-Infrared (MIR) spectra obtained from several plastics.

Another application category in atmospheric sciences is forecasting. Ong et al. [71] applied recurrent neural networks to the forecasting of $\text{PM}_{2.5}$ in Japan from environmental data. Li et al. [72] used the more advanced LSTM architecture to forecast $\text{PM}_{2.5}$ in downtown Beijing. Gil et al. [73] studied the ozone formation from HONO measured via QCL absorption spectroscopy. In their study they experimented with concentration estimation of HONO via artificial neural networks from meteorological data and nitrogen oxides, NO and NO_2 .

Special applications in quantitative tasks are *surrogate models* and inversions. These refer to simple neural networks trained from synthetic data generated from expensive and lengthy calculations to speed up the process. An important example is the reconstruction in tomography absorption spectroscopy as reported, e.g., in [74, 75]. Reconstruction of the target is a complex non-linear problem requiring computationally expensive calculations. While the training of neural networks requires a lot of computation, the inference is comparably light-weight. This way, a surrogate model capable to deliver the output to a complex calculation with desired precision, but far less computation time, is beneficial in time-critical tasks. Rankine et al. [76] use a similar approach to rapidly predict X-ray spectra.

Some studies even experiment with neural networks to directly fit absorption spectra to gas concentration. Nicely et al. [8] conducted a simulation study that used a shallow neural network for fringe reduction in direct absorption spectroscopy. They synthesized a mixture of a CO_2 absorption line and a discrete number of pre-defined etalon fringe patterns. In their proof-of-concept study they successfully showed a de-

noising behavior but also discussed a limiting bias introduced by the network. Other studies, e.g., Tian et al. [77], report a good linearity of a direct fit performed by a neural network for high SNR input spectra.

2.3.3 Limitations of neural networks

Since the inner structure of neural networks is completely determined by the training process, it is challenging to reason how a network functions. The performance is therefore determined by the quality of the training data. Misbehavior of neural networks on unseen data may originate from poor training data quality or poor training objective.

During training process, the distribution of training data is mapped to an output that minimizes the loss function. If inference data is distributed differently, the network may output unexpected results. This phenomenon is called out-of-distribution error. In classification tasks, wrong labels may be returned with high confidence despite the sample originating from a different distribution [78, 79, 80]. The training data may also contain undesired biases which will effect the network behavior. Sexism and discrimination towards certain groups and populations have been reported in machine learning applications due to biased training examples [81, 14]. Thus, careful preparation of the training data is necessary to handle these challenges. Elaboration of possible inference data encountered is necessary to evaluate the risk of out-of-distribution errors.

The loss function or the training objective may be poorly chosen or may not be appropriate for the task. Macroscopic objectives, e.g., generalization capability and other objectives that cannot be described by analytic functions, cannot be ensured via good training performance. An example is the notorious overfit: While the researcher aims for a classifier with high generalization abilities to determine the different input classes, the network may instead memorize features of the training samples to achieve the best possible classification accuracy but renders itself unable to achieve similar performance on unseen data [42]. Vague objectives that cannot easily be described analytically are a major challenge in AI safety [80]. Before deploying neural networks in applications, deviations from the true objective and the minimized loss function need to be evaluated.

Another concerning phenomenon that limits the trustworthiness of neural networks are adversarial attacks. The inner structure of a neural network can be exploited to influence the output class by applying structured, but tiny variations to the input [79]. Several studies have covered defense mechanisms against this vulnerability [46, 82].

3 Temporal Stability and Instrument Performance of an airborne QCLAS

This chapter was uploaded to the preprint server Research Square as "Determination of Temporal Stability and Instrument Performance of an airborne QCLAS via Allan-Werle-plots" [83] under the terms of the Creative Commons CC BY license: <https://creativecommons.org/licenses/by/4.0/>.

How to cite: Lenard L. Röder, Linda M. Ort, Jos Lelieveld and Horst Fischer. Determination of Temporal Stability and Instrument Performance of an airborne QCLAS via Allan-Werle-plots, 19 November 2023, PREPRINT (Version 1) available at Research Square <https://doi.org/10.21203/rs.3.rs-3619758/v1>

The preprint is currently under review for publication in Applied Physics B: Lasers and Optics.

My contribution to the study is the design of the study, carrying out the data analysis, drafting text and figures and literature research.

Abstract

Allan-Werle-plots are an established tool in infrared absorption spectroscopy to quantify temporal stability, maximum integration time and best achievable precision of a measurement instrument. In field measurements aboard a moving platform, however, long integration times reduce time resolution and smooth atmospheric variability. A high accuracy and time resolution are necessary as well as an appropriate estimate of the measurement uncertainty. In this study, Allan-Werle-plots of calibration gas measurements are studied to analyze the temporal characteristics of a Quantum Cascade Laser Absorption Spectrometer (QCLAS) instrument for airborne operation. Via least-squares fitting the individual noise contributions can be quantified and different dominant regimes can be identified.

Through simulation of data according to the characteristics from the Allan-Werle-plot, the effects of selected intervals between in-flight calibrations can be analyzed. An interval of 30 min is found sufficient for successful drift correction during ground operation. The linear interpolation of the sensitivity increases the accuracy and lowers the measurement uncertainty from 1.1 % to 0.2 %. Airborne operation yields similar results during segments of stable flight but additional flicker and sinusoidal contributions. Simulations verify an appropriate interval of 30 min in airborne operation. The expected airborne measurement uncertainty is in the range 0.8 % to 2.4 %.

3.1 Introduction

Allan-Werle-plots have become a major tool in infrared spectroscopy applications for the characterization of instrumental stability and performance for field experiments [84, 38, 29, 85, 86, 87, 88]. Werle et al. [39] originally applied the two-sample-variance, or Allan Variance, to tunable diode laser absorption spectroscopy (TDLAS) to determine the maximum integration time and minimum achievable precision at this integration time.

However, a large maximum integration time is not necessarily needed in field experiments. For many species, the atmospheric variability is much higher than the instrument uncertainty [22]. In this case, averaging the data to a lower time resolution filters atmospheric features of the time series. In field experiments aboard a moving platform, e.g. aircraft-borne or drone-borne field experiments, a high temporal resolution is needed to enable localization of the measurements [89] and further data-driven analysis, e.g. calculation of fluxes via eddy covariance and wavelet transforms [90, 91, 92]. Furthermore, determination of the instrumental properties in the laboratory or in a stationary environment may not be transferable to in-flight performance [9].

According to Werle (2011) [32] many more features of the instrumental temporal dependencies can be extracted from an Allan-Werle-plot. Apart from uncorrelated, white noise, and random walk drifts, other noise contributions, e.g. flicker noise and sinusoidal interferences can be determined. This study will characterize instrumental temporal dependencies of the TRISTAR (tracer in-situ TDLAS for atmospheric research) instrument described in [11, 93, 38, 94] and investigate differences in the performance during ground and airborne operation. Using the information provided via Allan-Werle-plots, suitable repetition times of in-flight calibrations and their implication to measurement uncertainty estimation will be discussed.

3.2 The TRISTAR instrument

The TRISTAR instrument has been originally designed by Wienhold et al. (1998) [11] as a three-channel TDLAS that accumulates three laser beam paths via beam splitters [11] and in a later design via pneumatic moving mirrors [93]. The combined beam path is guided through an astigmatic White cell [2, 11] with a base length of 0.5 m and 128 passes. The outgoing beam is split into a signal and reference beam and both are focused to individual detectors. A reference cell filled with a high concentration of the target gas is inserted into the reference beam path to enable line determination and line locking via laser current offset adaption. The instrument has been successfully deployed in various aircraft campaigns [95, 96, 38, 97] aboard a Learjet 35A and the High Altitude Long-range (HALO) Gulfstream G550 research aircraft. During these campaigns a range of trace gases has been measured including CO, CH₄, HCHO, NO₂ and N₂O. Measurement uncertainties of 0.5 %, 0.5 % and 0.7 % have been achieved for CO, N₂O and CH₄, respectively [38].

During 25 years of operation, various modifications have been applied to the original setup, including a major revision since the last reported experimental setup by Tadic et al. (2017) [94]. The cryogenic cooling system has been abolished by replacing individual components that were cooled via liquid nitrogen, i.e. lasers and detectors, by thermoelectrically cooled components. The three-laser setup has been replaced by single-laser operation which increases data coverage of the target species. While these measures might decrease the optimal performance and limit the diversity of data output, a simpler instrument maintenance and more autonomous operation can be achieved. The latter is enabled by substituting the V25 electronics and DOS PC by a National Instruments real time controller paired with an ARK Windows PC running a LabView real time application and graphical user interface. Establishing a PC-based user interface enables remote operation via internet and reduces the required training time for airborne operators in aircraft measurement campaigns. Autonomous and remote operation has been demonstrated during the ground-based field experiment at the German Taunus Observatory in 2021 [98], requiring only maintenance and calibration gas source exchange as on-site activities.

3.2.1 Optics and gas flow

A room temperature quantum cascade laser (RT-QCLAS) inside a high heat load (HHL) housing (AlpesLasers, Switzerland) is driven in cw-mode and emits a divergent beam. The beam is captured and focused onto a first focal point via a three-mirror telescopic optics (IPM Freiburg). Via a pair of an off-axis parabolic (OAP) and elliptic (OAE) mirror the beam is converted into a parallel beam with a width of approximately 14 mm. The large beam waist makes the setup less sensitive to aerosol particles and imperfections on the mirrors and turbulence in the optical path.

Several planar mirrors guide the path to the entrance of the White cell. There the parallel beam is focused to a second focal point using another pair of an OAE and OAP mirror. Then the divergent beam is injected through wedged CaF₂ windows into the double corner cube White cell, where it gets reflected 128 times.

The White cell has a length of 0.5 m and holds a volume of 3 l. The cell is drained by a scroll vacuum pump set to a constant gas flow of 3 slm (standard liters per minute). On the top of the cell a movable valve operated by a step motor regulates the cell pressure to 50 hPa. The inlet is connected to a T-junction, where calibration gas can be injected by a mass flow controller (Bronkhorst). Excess calibration gas exits through the ambient air inlet. A detailed schematic of the tubing is given in Tadic et al. (2017) [94], Figure 2.

The outgoing beam is focused on a third focal point and converted back into a parallel beam using another pair of OAE and OAP mirrors. The parallel beam is split via a CaF₂ beam splitter (50:50) into signal and reference beam. The reference beam is further passed through a reference cell filled with high concentration of the target gas. Both parallel beams are focused via coated CaF₂ lenses with focal length of 40 mm onto room-temperature photovoltaic infrared detectors (VIGO systems, Poland). All optical components are mounted on a rigid aluminum plate that is electrically heated

3 Temporal Stability and Instrument Performance of an airborne QCLAS

to ensure thermal stability. This makes the optical setup less sensitive to distance variations caused by thermal fluctuations.

3.2.2 Electronics and data processing

The instrument runs autonomously inside a NI LabView real time application on the real time controller (NI cRIO 9038). The software controls pump, inlet and MFC and writes data to disk. Through the FPGA module the scanning ramp and modulation is set at a clock frequency of 250 kHz with a modulation frequency of 31.25 kHz and a scanning frequency of 60 ramps per second. The ramp and modulation are input to the laser controller that operates the QCL at a set temperature and laser base current to obtain a laser frequency near the absorption central frequency.

The detector output is digitally demodulated at twice the modulation frequency for $2f$ wavelength modulation spectroscopy (WMS). The raw signal is down-sampled by a CIC-stage to reduce the amount of data. Individual spectra of the increasing (up) and the decreasing (down) part of the triangular scanning frequency are averaged separately to obtain a time resolution of 1 Hz. The averaged up and down spectra are saved to disk and further processed for online operation.

The described configuration of modulation frequency, scanning frequency and time resolution can be adjusted to the needs of the experimental requirements.

Throughout the experiments carried out in this study, the QCL is operated near 2190 cm^{-1} , where a strong line pair of CO and N₂O is located [99]. The scanning range is optimized to enable simultaneous measurements of both species.

The saved data is post-processed using custom code written in IGOR Pro (Wavemetrics). The raw signal is filtered using a IIR low-pass filter, the line centers and widths are determined from the reference signal and the spectra are linearly fit against the most stable absorption spectrum obtained during a calibration measurement.

3.2.3 Instrumental performance measures

The total measurement uncertainty is composed of several metrics that each contribute differently to the limitation of the measurement device. The systematic error is defined by the accuracy and describes the overall deviation of all data points from the true value. Since the TRISTAR device uses the relative measurement mode of WMS, the uncertainty of the conversion factor from arbitrary units to mixing ratio contributes to the accuracy. The accuracy depends on the calibration gas mixing ratio uncertainty as determined in the laboratory or provided by the gas standard supplier.

The systematic deviations caused from drifts can be mitigated by regular calibrations. Considering changes in relative sensitivity, the drifts can be corrected via linear interpolation of the reciprocal mixing ratios. This interpolation scheme requires that the drift speed is not faster than the interval between calibrations. The reproducibility R refers to the relative spread of the calibrations over a complete measurement. This uncertainty cannot be lowered by further averaging.

3.3 Ground operation inside the HALO aircraft

The precision P refers to the relative statistical error at a given integration time after drift correction. It is obtained from the relative standard deviation of the calibration measurements. The total relative error at native time resolution is a combination of P and R:

$$\frac{\sigma_\chi}{\chi} = \sqrt{P^2 + R^2} \quad (3.1)$$

The detection limit refers to the lowest mixing ratio distinguishable from noise. It is estimated from the standard deviation of absorption-free gas injected into the absorption cell. By interpreting the precision as pure absolute error, the precision at the mixing ratio of the calibration gas gives an upper boundary and a worst case estimate to the detection limit.

Werle et al. [39] define the Allan variance as the mean squared difference between adjacent data batches for different integration times τ :

$$\sigma_\chi(\tau)^2 = \frac{1}{2} \sum_{k=0}^{M-2} (\chi_{k+1}(\tau) - \chi_k(\tau))^2 \quad (3.2)$$

$$\chi_k(\tau) = \frac{1}{\tau} \int_{t=k\tau}^{(k+1)\tau} \chi(t) dt \quad (3.3)$$

As reported in [32], plotting the Allan variance as a function of the integration time (Allan-Werle-plot) provides insight into the temporal characteristics of an absorption spectroscopy measurement. The maximum integration time τ_{\max} refers to the minimum in the Allan-Werle-plot where the trade-off between the decrease from the integration of white noise and the build-up from integration of drifts is optimal. Considering only uncorrelated white noise σ_W and random walk noise σ_B the Allan-variance σ^2 and τ_{\max} can be modeled as

$$\sigma^2(\tau) = \frac{\sigma_W^2}{\tau} + \frac{\sigma_B^2}{2} \tau \quad \text{and} \quad (3.4)$$

$$\tau_{\max} = \sqrt{2} \frac{\sigma_W}{\sigma_B}, \quad (3.5)$$

respectively.

3.3 Ground operation inside the HALO aircraft

The TRISTAR instrument is part of the TRIHOP rack that also carries the hydrogen peroxide and higher organic peroxide monitor HYPHOP [19]. The rack is mounted into the research aircraft HALO. The inlet is connected to a T-junction that may provide scrubbed cabin air during zero cycles. Scrubbing is shut off during CO measurements. The calibration gas tube is connected to a 6l composite bottle (AUER). During ground operation the measurement devices run on external power while the aircraft auxiliary power unit and main engine remain shut off. As a consequence the integrated air conditioning of the aircraft also remains off and air conditioning has to be provided externally by ventilation. This lead to a slowly increasing temperature inside the aircraft over a whole day of ground operation up to 35 °C cabin temperature.

3.3.1 Performance of calibration gas measurement

Several functional tests were carried out during ground operation. Determining the precision and long-term stability of the instrument is important to estimate the expected performance during the research flights. In order to collect suitable data the cell was flushed with calibration gas regularly and also for longer time periods. Extended measurements of calibration gas provide a good estimate of the statistical measurement uncertainty directly from the standard deviation. A time series and corresponding Allan-Werle-plot of such an extended period of calibration gas measurement is shown in Figure 3.1. Calibration gas was injected over a period of 2 hours. The absolute and relative standard deviation of the measurement are 1.7 ppbv and 1.1 %, respectively. The Allan-variance $\sigma^2(\tau)$ can successfully be modeled by pure noise σ_W and drift σ_B contributions. The optimal fit parameters are given in the figure caption of Figure 3.1 and the maximum integration time results to 20s according to (3.5).

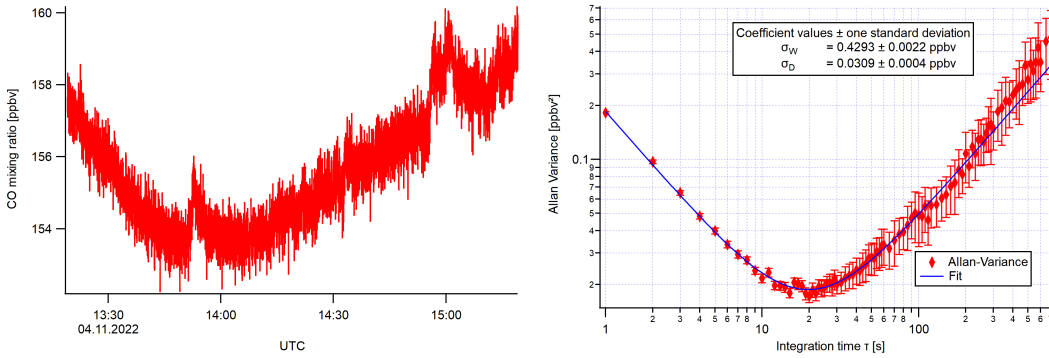


Figure 3.1: Calibration gas measurement during ground operation inside the HALO aircraft. Left: time series of the 2 hour segment of calibration gas. Right: Corresponding Allan-Werle-plot with fit to (3.4). The optimal fit parameters are $\sigma_W = (0.4293 \pm 0.0022)$ ppbv and $\sigma_B = (0.0309 \pm 0.0004)$ ppbv.

3.3.2 Performance of ambient gas measurement

Determining the measurement uncertainty as indicated in the previous section is not possible during ambient measurements, however, since the standard deviation is a mixture of atmospheric variability and instrument uncertainty. Instrument drifts are also hard to distinguish from changes in the mixing ratios of the sampled air masses. Determining an appropriate interval of calibrations to minimize the effect of drifts and estimate the measurement uncertainty is a non-trivial task during field experiments and airborne field experiments, especially.

Using the Allan-Werle-plot in Figure 3.1, similar data can be simulated that shares the temporal characteristics obtained from the measurement. This can be achieved with the following simulation setup, where χ_0 refers to the mixing ratio of the calibration

3.3 Ground operation inside the HALO aircraft

gas.

$$h_t = h_{t-1} + \epsilon_B \quad (3.6)$$

$$\chi_t = \chi_0 + h_t + \epsilon_W \quad (3.7)$$

$$\text{with } \epsilon_B \sim \mathcal{N}(0, \sigma_B^2),$$

$$\epsilon_W \sim \mathcal{N}(0, \sigma_W^2)$$

A resulting time series can be found in Figure 3.2. Starting from this simulated data, points can be artificially flagged as calibration and ambient to simulate the effect of different intervals between calibrations during a real field measurement scenario. The effects on the correction of the drifts and the estimation of the true measurement uncertainty can be studied. For each individual simulation, R is calculated from the spread of the means of the flagged calibrations. The reciprocal means are interpolated to estimate the sensitivity drift. The interpolated sensitivity is multiplied to the amplitudes to obtain the drift-corrected amplitude. P is obtained from the relative standard deviation of the drift-corrected calibration amplitudes. The relative standard deviation of the corrected amplitudes which are flagged ambient, on the other hand, gives a good estimate of the residual measurement uncertainty (RMU) due to drifts and noise present in the data. An optimal interval between calibrations will decrease the RMU until P is reached. The estimated measurement uncertainty (EMU) as defined in (3.1) should optimally be close to the true relative standard deviation of the simulated dataset.

Figure 3.3 shows the described metrics for different intervals between calibrations. The observable features match the expectations: If the calibration frequency is too low, the drifts cannot be successfully corrected. Thus, the RMU is close to the original measurement uncertainty. Furthermore, the EMU is very unstable since there are too few calibration measurements in total. With more calibrations both of these effects are reduced. The EMU approximates the original measurement uncertainty better and the RMU decreases. In principle, more calibrations are always advantageous. However, more calibrations also result in less ambient measurements and a high calibration gas consumption. An empirical choice is made that subdivides the calibration frequencies into three regimes: Too many calibrations only contribute diminishing returns while decreasing the duty cycle, too few calibrations cause the effects described at the beginning of this paragraph, and the appropriate choice lies in the middle. The appropriate region is chosen within $10\tau_{\max}$ and $100\tau_{\max}$, where τ_{\max} refers to the maximum integration time derived from the Allan-Werle-plot.

Accordingly a calibration interval of 30 min lies within the desired regime. On another day of ground operation, regular calibrations were carried out at that interval. The resulting values for R and P are 1.0 % and 0.2 %, respectively, and accurately estimate the total measurement uncertainty. Furthermore, a successful correction for the random walk drifts can be assumed considering the results of the simulation shown in Figure 3.3. The residual measurement uncertainty of the ambient data can therefore be estimated with the precision P, reducing the uncertainty from a value of 1.1 % to 0.2 % as a consequence of regular calibrations.

3 Temporal Stability and Instrument Performance of an airborne QCLAS

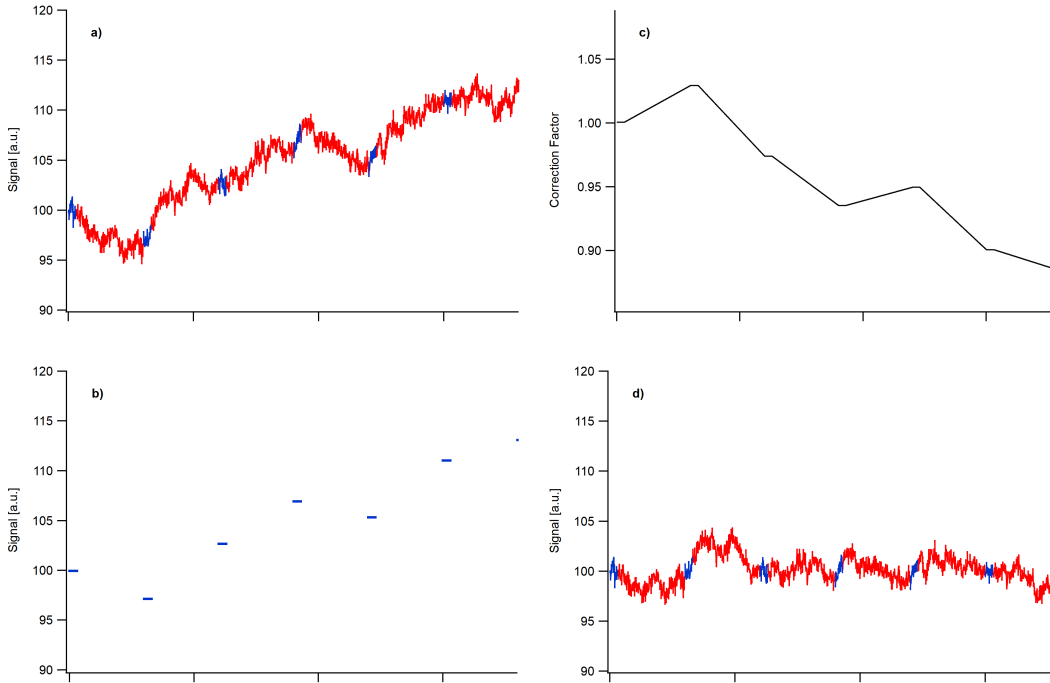


Figure 3.2: Example simulation process carried out to retrieve the metrics P, R, RMU and EMU. Top left: Simulated time series with flagged ambient (red) and calibration (blue) segments. Bottom left: individual means of flagged calibrations. The standard deviation of these values is used to calculate R. Top right: Interpolated sensitivity that is connected to the reciprocal of the calibration means. Bottom right: Drift-corrected time series. The standard deviation of the data flagged calibration yields P, the standard deviation of the data flagged ambient yields RMU.

3.4 Airborne operation

In the previous section, the instrument has been characterized during ground operation aboard the HALO aircraft. An appropriate interval for calibrations has been chosen that effectively corrects for random walk drifts and allows for a reduction of the measurement uncertainty. Now the question arises: Is the ground-based performance transferable to airborne operation?

In order to study the instrument response to disturbances caused by airborne operation, e.g., vibration, acceleration, orientation and pressure, a test flight was performed where TRISTAR was constantly measuring calibration gas. The data acquired can be used to qualitatively and quantitatively analyze the effects of this new environment to the instrumental performance.

Figure 3.4 shows the time series and the corresponding Allan-Werle-plot of this test flight. The time series indicates different environmental conditions throughout the flight. During takeoff and landing periods, rapid accelerations impact the optical

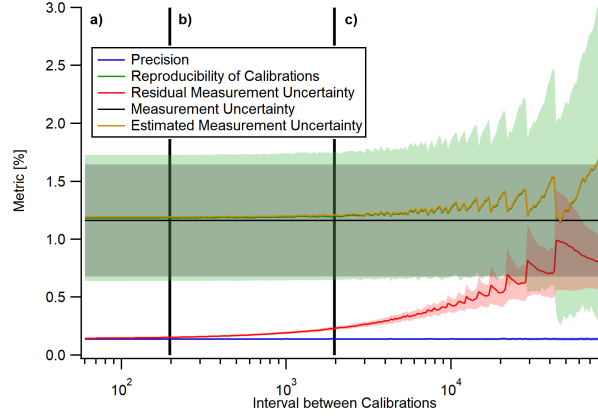


Figure 3.3: Performance metrics for different intervals between calibrations. The simulation was repeated 100 times. The solid lines indicate the mean values and the shaded areas indicate the standard deviations of each metric for the 100 repetitions. The metrics shown are R (green), P (blue), RMU (red) and EMU (yellow). The black horizontal line shows the total measurement uncertainty of the simulated time series. The black vertical lines separate the three regimes a): too many calibrations, b): appropriate interval between calibrations and c): too few calibrations. The spikes in the results originate from discrete changes in the number of calibrations that fall inside the total number of data points in the time series for each given calibration interval.

setup and lead to distortions in the signal. At high altitudes, the pressure regulation of the inlet starts to oscillate and results in oscillations of the gas concentration. In post processing these oscillations can be corrected by assuming a linear dependence on the cell pressure. On intermediate flight levels the system is more stable. Therefore, calculating the Allan variance for the complete data set is no longer meaningful, as the experimental conditions and the underlying distribution is changing. Thus, the Allan variance is calculated individually for the different conditions derived from the time series and metadata.

The Allan variance of the total data set is a combination of the individual variances. This obstructs the information extractable in the individual Allan-Werle-plots. The takeoff and landing phase is dominated by a strong sine wave and drift that are caused by the mechanical system reacting and recovering from the strong changes in acceleration. The high altitude section is dominated by a strong sine wave that matches the oscillations caused by the pressure regulation. The flight segments on intermediate flight levels are quite stable. During these segments a series of left and right turns and various altitude changes were carried out to study the instrument performance.

Again, the Allan variances are fit against theoretical models to quantify the temporal features. This time the composition is more complex, so pure white noise and random walk is not sufficient to reproduce the measured features. Flicker noise σ_F is added that acts as a noise floor according to [32]. Sinusoidal variations contribute a term proportional to A_{sin} and dependent on the period T_{sin} . Additionally, constrained

3 Temporal Stability and Instrument Performance of an airborne QCLAS

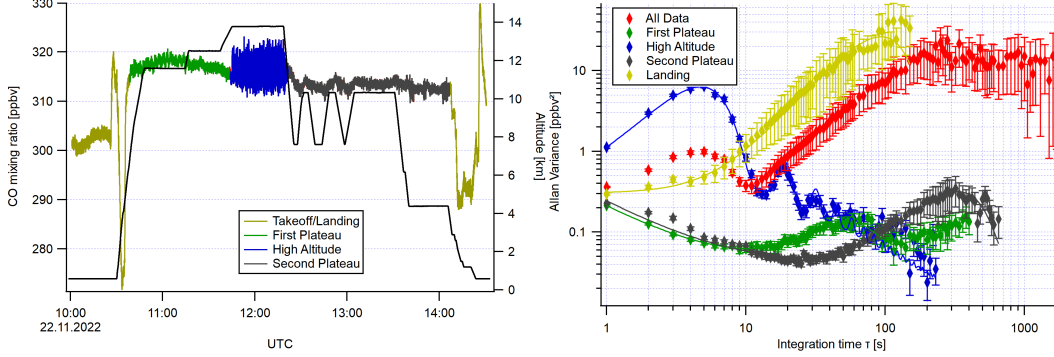


Figure 3.4: Calibration gas measurement during test flight #01 on the 22nd of November, 2022. The data is segmented according to changes in the environmental conditions into the takeoff and landing phase (yellow), first plateau segment (green), high altitude segment (blue), second plateau segment (grey), complete data set (red). Left: time series of the 5 hour flight, separated into these segments, plotted against the left axis and GPS altitude (black) plotted against the right axis. Right: Corresponding Allan-Werle-plots with fits to (3.8). The optimal fit parameters are summarized in Table 3.1.

random walk noise that behaves like white noise for longer time spans is introduced with amplitude A_{MH} and transition time T_{MH} . The theoretical model results in (3.8), and the most appropriate fit is chosen with the least number of parameters. The resulting fit parameters are summarized in Table 3.1.

$$\sigma^2(\tau) = \frac{\sigma_W^2}{\tau} + \frac{\sigma_B^2}{2}\tau + \sigma_F^2 + A_{\sin}^2 \frac{\sin\left(\pi\frac{\tau}{T_{\sin}}\right)^4}{\left(\pi\frac{\tau}{T_{\sin}}\right)^2} + \frac{A_{MH}^2}{\sqrt{2}} \left(\sqrt{\frac{T_{MH}-1}{\tau}} + \sqrt{\frac{\tau}{T_{MH}}} \right)^{-2} \quad (3.8)$$

Parameter	First Plateau	Second Plateau	High Altitude	Landing
σ_W [ppbv]	0.421 ± 0.007	0.456 ± 0.005	–	–
σ_B [ppbv]	0.0230 ± 0.0013	–	–	–
σ_F [ppbv]	0.189 ± 0.006	0.1918 ± 0.0025	–	0.552 ± 0.008
A_{\sin} [ppbv]	0.385 ± 0.014	0.730 ± 0.022	3.31 ± 0.04	7.03 ± 0.37
T_{\sin} [s]	144 ± 5	897 ± 26	12.82 ± 0.06	251 ± 17
A_{MH} [ppbv]	–	–	1.68 ± 0.04	–
T_{MH} [s]	–	–	3.07 ± 0.17	–

Table 3.1: Optimal fit parameters of each Allan-Werle-plot shown in Figure 3.4 when fitting against (3.8). Parameters marked – are neglected since they do not contribute major features of the Allan variance.

During the stable segments the values extracted from the Allan-Werle-plots for the white noise and random walk contributions are similar to the values obtained from the Allan-Werle-plot of the ground operation. This indicates similar effects influencing the

instrumental performance on both modes of operation. Features exclusive to airborne operation are manifested through additional flicker and sinusoidal contributions during the stable segments. More noticeable features are visible in the time series and enable segmentation of the time series.

Similar to the analysis carried out in the previous section, data was simulated that follows the temporal characteristics of the two stable segments of the flight. The results of the simulation procedure is shown in Figure 3.5. Since the Allan Variance in these cases contains flicker and sinusoidal components, the results differ from the simulation results of the ground operation. The flicker and sinusoidal components contribute further short-term uncertainties that increase the RMU on shorter periods even with high frequency of calibrations. Thus, the measurement uncertainty on ambient data can no longer be estimated based on the precision only but needs to be approximated by the EMU (3.1). Accurate estimation of the EMU is again possible within the appropriate regime of calibration intervals. This regime can again be calculated from (3.5) and lies inbetween $10\tau_{\max}$ and $100\tau_{\max}$. For the first plateau segment, τ_{\max} is equal to 26 s. For the second plateau segment, no value for σ_B could be extracted from the Allan-Werle plot. A time interval of 30 min again lies inside the desired range.

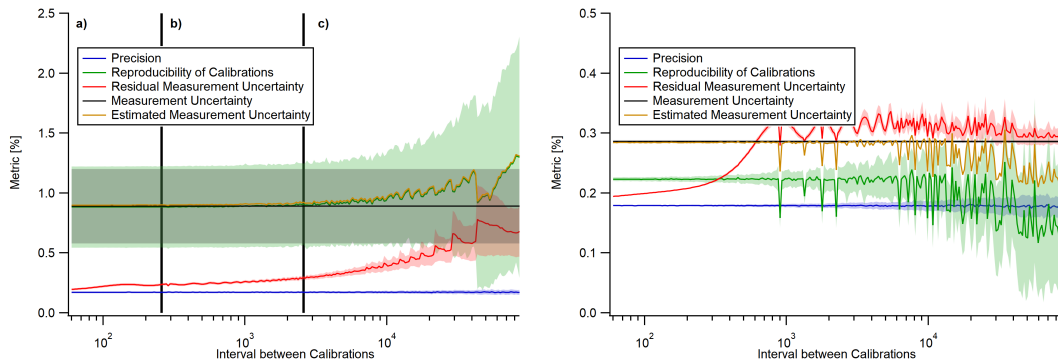


Figure 3.5: Performance metrics for different intervals between calibrations. The simulation was repeated 100 times. Left: First plateau segment of Testflight #1. Right: Second plateau segment of Testflight #1. The solid lines indicate the mean values and the shaded areas indicate the standard deviations of each metric for the 100 repetitions. The metrics shown are R (green), P (blue), RMU (red) and EMU (yellow). The black horizontal line shows the total measurement uncertainty of the simulated time series. The black vertical lines (only applies to left plot) separate the three regimes a): too many calibrations, b): appropriate interval between calibrations and c): too few calibrations. The spikes in the results originate from discrete changes in the number of calibrations that fall inside the total number of data points in the time series for each given calibration interval.

During a second test flight, the operation of the TRISTAR instrument was switched to an operation scheme used in real research flights. This way the instrumental behavior during gas switching and pressure regulation at different altitudes were tested. Ambient gas measurements were carried out and regular calibration measurements were performed for drift correction and measurement uncertainty determination.

3 Temporal Stability and Instrument Performance of an airborne QCLAS

The time series of Testflight #2 is shown in Figure 3.6. The precision P and reproducibility R as evaluated from the repeated calibrations are equal to 0.35 % and 1.50 %, respectively. The estimated measurement uncertainty lies between the relative standard deviation of the total Testflight #1 and the reduced data set, where takeoff and landing phase are discarded. These values are equal to 1.5 %, 2.4 % and 0.76 %, respectively.

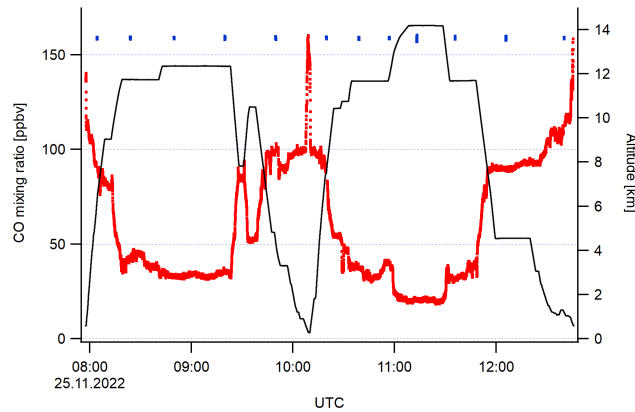


Figure 3.6: Ambient and calibration gas measurement during Testflight #02 on the 25th of November, 2022. The plot shows the ambient mixing ratio (red), calibration gas mixing ratio (blue) plotted against the left axis and GPS altitude (black) plotted against the right axis.

3.5 Summary and Conclusion

Allan-Werle-plots have become a key tool in infrared absorption spectroscopy to investigate the temporal stability of a measurement instrument. However, these plots are commonly used to determine the maximum integration time and the best achievable precision. In this study, Allan-Werle-plots were used to analyze the temporal characteristics of the newest iteration of the TRISTAR instrument both for ground operation inside the research aircraft and for airborne operation. Determination of the instrumental stability through the maximum integration time is not beneficial due to the desired time resolution during research flights. Obviously, such unstable measurement environment like a moving aircraft can lead to strong drifts.

Instead, the Allan-Werle-plots were used to identify the dominant noise types present during the measurement. Via least-squares fitting the contributions of white noise, random walk noise, flicker noise, meta-stable random walks and sinusoidal contributions were obtained for the different measurement conditions.

Simulation of data that recreates the characteristics observed in the Allan-Werle-plots allowed for an extensive analysis of the choice of the interval between calibrations. Several metrics were defined to identify the capability of the procedure to correct for drifts and estimate the true measurement uncertainty. An interval of 30 min was found

3.5 Summary and Conclusion

sufficient for complete drift correction during ground operation. Thus, estimation of the measurement uncertainty with only the precision was enabled. The expectations were verified using a second data set with the appropriate calibration frequency.

The same analysis was carried out using data from two test flights. Segmentation of the time series according to the sudden changes in the measurement conditions enabled resolution of the dominant noise types for each individual segment during the flight. Dominant sinusoidal contributions were observed during a high altitude segment due to oscillations in the absorption cell pressure regulation. Low-frequency contributions dominated the takeoff and landing phases. Investigation of stable segments during the flight showed similar characteristics to ground operation. This designated sufficient intervals between calibrations of 30 min during research flights for measurement uncertainty estimation and drift correction. However, apparent flicker and sinusoidal contributions lead to a lowered accuracy for reasonable calibration frequencies. Therefore, the measurement uncertainty had to be estimated considering the reproducibility of calibrations. A second test flight verified the estimated measurement uncertainty for the chosen calibration frequency.

The results of this study show yet another great application of Allan-Werle-plots to increase the understanding and the accuracy of infrared absorption spectrometers. The implications for drift correction and measurement uncertainty are also transferable to other measurement instruments for trace gas measurements.

4 Applicability of Advanced Noise Reduction Methods

This chapter was published as "Theoretical Investigation of Applicability and Limitations of Advanced Noise Reduction Methods for Wavelength Modulation Spectroscopy" by Springer Nature, Dec 29, 2021 in Applied Physics B: Lasers and Optics [100] under the terms of the Creative Commons CC BY license: <https://creativecommons.org/licenses/by/4.0/>.

How to cite: Röder, L.L., Fischer, H. Theoretical investigation of applicability and limitations of advanced noise reduction methods for wavelength modulation spectroscopy. *Appl. Phys. B* **128**, 10 (2022).

<https://doi.org/10.1007/s00340-021-07737-z>

My contribution to the study is the design of the study, carrying out the simulation and data analysis, drafting text and figures and literature research.

Abstract

In this study, the applicability and limitations of several statistical and mathematical methods for noise reduction in wavelength modulation spectroscopy are analyzed. Background noise is simulated for different frequencies and frequency confinement. The noise is added to an absorption line of varying amplitude. The noise reduction methods (NRMs) are applied to the simulated signals and their performances are analyzed and compared. All NRMs show great increase in signal to noise ratio (SNR) while keeping bias low under certain conditions of the simulated signal. For each NRM the subspace of best performance is summarized and highlighted. Little to no overlap is found between these subspaces. Therefore, the optimal NRM strongly depends on measurement conditions and NRM quality cannot be compared in a general context.

4.1 Introduction

Tunable diode laser or quantum cascade laser absorption spectroscopy (TDLAS / QCLAS) using wavelength modulation spectroscopy (WMS) have become major tools for molecule detection in various disciplines due to their high flexibility, time resolution, sensitivity, precision and selectivity. Main applications are atmospheric measurements of trace gases and combustion process controls in industry [101, 27, 31]. Advancements in laser technology like the quantum cascade laser (QCL) [4] and the implementation of multi pass absorption cells [2, 3] have lead to major improvements

4 Applicability of Advanced Noise Reduction Methods

of the technique in the last decades. The main limitations for precision and minimum detectable mixing ratios are laser disturbances, power or frequency drifts and optical noise patterns [1, 39, 7]. These noise patterns often have an underlying dominant frequency and a varying phase. The frequency can lie in the same region as the absorption signal, depending on the optical setup.

Many new data analysis techniques for noise reduction have been adapted to absorption spectroscopy experiments during the last decades and were reviewed several times in a qualitative manner [7, 6]. These methods range from signal analysis and denoising via time-frequency transformations [36, 102] to adaptive statistical methods [35, 37]. However, most recent publications of state-of-the-art experiments still rely on basic data analysis established in the last century [31, 9]. This may originate from the specific design of these analysis methods for their target experiment; thus, leaving questions unanswered about the experimental requirements and limitations. This paper attempts to compare a few prominent and also lesser known noise reduction methods (NRM) via simulation of a wide array of possible noise shapes. The goal to reach is a quantitative statement about the experimental conditions necessary for the given NRM to increase the signal-to-noise-ratio (SNR), while simultaneously maintaining a low bias.

This paper is organized as follows: In Sect. 4.2 an overview of the analyzed NRMs is given and small modifications to ensure compatibility are explained. In Sect. 4.3 the simulation procedure is shown. Finally the simulation results are analyzed and interpreted.

4.2 Noise reduction methods

First, the theory of WMS is described in short. According to the Lambert-Beer law the transmitted intensity I through an absorbing medium for small optical density can be approximated by:

$$I(\omega) \approx I_0 (1 - \alpha(\omega)) \quad (4.1)$$

where I_0 is the transmitted intensity without absorption and $\alpha(\omega)$ is the absorption profile. α approaches a Gaussian profile for low pressure and a Lorentzian or Cauchy profile for high pressure. In intermediate cases the profile can be expressed as a convolution of both, a so called Voigt profile:

$$\begin{aligned} \alpha_V(\omega) &= \alpha_G(\omega) * \alpha_L(\omega) \\ &\sim \exp\left(-\frac{(\omega - \omega_0)^2}{2\sigma^2}\right) * \frac{1}{1 + \left(\frac{\omega - \omega_0}{\gamma}\right)^2} \end{aligned} \quad (4.2)$$

In WMS the frequency ω is modulated with a modulation depth a and modulation frequency Ω and the signal is subsequently demodulated using an analog or digital lock-in amplifier set to an integer multiple of the modulation frequency $n\Omega$. Therefore the retrieved signal S is proportional to the n th harmonic H_n of the cosine series of

the absorption profile:

$$S \sim H_n(\omega) = \int \alpha(\omega + a \cos(\Omega t)) \cos(n\Omega t) dt \quad (4.3)$$

More detailed derivations have been given several times in the literature and can be found for example in [27, 1, 6]. In the course of this paper the second harmonic is used as most methods considered in this study were originally designed for second harmonic detection. It is depicted in Figure 4.1. In principle this study can be conducted using other harmonics, normalization schemes [31] or direct absorption, as well.

4.2.1 Wiener filter

Werle et al.[35] have described the implementation of an adaptive finite impulse response filter (FIR), the Wiener filter. It is calculated for each measurement spectrum X by a least squares algorithm, minimizing the mean square error between the measurement spectrum and a synthesized version by convolution of the filter with a calibration spectrum C . Thus, knowledge about the target signal is used to optimize the FIR.

$$X_i = \sum_{k=-m}^m \gamma_k C_{i+k} \quad (4.4)$$

A free parameter of this method is the filter size m . The appropriate value has to be chosen by a measurement of known mixing ratio.

4.2.2 Singular value decomposition

A simple statistical approach was proposed by Mappé-Fogaing et al.[37] that assumes a stable and dominant absorption signal compromised by an unstable background. By performing a singular value decomposition (SVD) on a batch of measured spectra the signal is split into eigenvectors. Under the right circumstances some of the eigenvectors contain signal information, while the rest contains only noise.

The singular value decomposition decomposes a matrix X with dimension $N \times M$ into the basis of normalized wavelets V , corresponding singular values Σ and the propagation matrix U .

$$X = U\Sigma V^T \quad (4.5)$$

In their paper a decent initial signal to noise ratio is given so the signal information is stored in the first eigenvectors. They choose the appropriate effective dimension of the data set empirically. Naturally this becomes a very difficult task when SNR is low. To adapt to the simulation conducted in this paper, Pearson correlation coefficients between each normalized wavelet and a reference spectrum are calculated and only the first m eigenvectors, sorted by their correlation coefficients, are considered during reconstruction. This procedure extends the empirical approach from the original paper but requires optimization of the free parameter m . The dimension has to be chosen during the calibration phase.

4.2.3 Discrete wavelet transform

Denoising techniques based on the discrete wavelet transform (DWT) have become widely used in signal processing, feature extraction and image denoising [103, 104]. The continuous wavelet transform is a time-frequency transformation similar to the Fourier transform [105]. It is defined by a convolution of the signal x with modified versions of a special function called *mother wavelet* ψ , that is scaled by the parameter a and translated by the parameter b :

$$\mathcal{W}(a, b) = \int \psi^* \left(\frac{t - b}{a} \right) x(t) dt \quad (4.6)$$

Employing the redundancy of this procedure, Daubechies [106] carefully designed special wavelets and evaluated only discrete scales. This way the transform no longer has a two-dimensional result but has the same dimension as the input. With Mallats algorithm this procedure now becomes an iterative convolution with discrete filter banks to decompose the signal into approximation and detail coefficients [107]. Based on this representation of the original signal several studies have shown applications for removal of trends and suppression of noise for a given signal, also for TDLAS [108, 109, 6]. A study by Li et al. [36] has successfully applied the noise reduction based on Steins unbiased risk estimator (SURE) [110] with the non-negative Garrote threshold [111] on WMS. In their study they compared several wavelets and maximum decomposition levels and retrieved the optimal choice for their experimental setup. Hence, the choice of wavelet and maximum decomposition level is a free parameter and has to be chosen during calibration.

4.2.4 Fourier domain analysis

Another approach utilizing time-frequency transformation has been proposed by Hartmann et al. [102]. They take advantage of the linear decrease of the logarithmic power spectrum of a Cauchy profile and interpolate discrete disturbances to this slope caused by sinusoidal background, effectively removing the fringe patterns. They emphasize, however, that the technique required fine tuning of signal and noise for this procedure to work that easily.

A number of adjustments have to be made for this method to be used fully automatically on second harmonic Voigt profiles. The absorbance of a measured signal is calculated by performing a masked linear fit of the power spectrum of the measured signal to the power spectrum of a second harmonic Voigt, acquiring the parameters from a reference spectrum. Alternatively it can be fit directly to the power spectrum of the measured reference. The mask is a free parameter with left and right boundary index. Depending on the background structure the background is more or less confined in the frequency domain. This way the mask would be able to ignore points in the power spectrum influenced by the background and only consider noise free parts of the signal power spectrum.

Regarding the emphasis on the special requirements by the authors it can be seen here

that improper choice of the mask or a background that is not at all confined to a small bandwidth can lead to unforeseen consequences if this method is applied blindly.

4.2.5 Empirical mode decomposition

The empirical mode decomposition (EMD) is an adaptive decomposition scheme that splits a signal into its intrinsic mode functions (IMF) with the same number of zero crossings as extreme points [112]. Meng et al. [113] have applied Savitzky-Golay filters to the IMFs and reconstructed the signal using the Pearson correlation coefficients of the filtered IMF and the original signal as a weight. They have shown in their study that this method called EMD-FCR (Filter-Correlation-Reconstruction) outperforms Wiener filters and DWT noise reduction for their particular experimental setup. A drawback of this method is the high computational effort. In addition, the order and window size of the Savitzky-Golay filter have to be optimized during the calibration phase as they are free parameters.

The FCR part was modified slightly for this simulation since the Pearson correlation with the original signal enhances noise instead of suppressing it in cases of very low SNR. As a modification the correlation coefficients are evaluated against a reference spectrum instead, similar to the SVD method.

4.2.6 Artificial neural network

A feed-forward neural network consists of layers of affine transformations that are separated by a non-linear function called activation. This can be expressed by a matrix multiplication of the input x_i with a weight Matrix W_i and addition of a bias vector b_i , where i is the layer number. The result of this operation is then pointwise evaluated using the activation function σ :

$$y_i = \sigma(W_i x_i + b_i) \quad (4.7)$$

During the training phase the output is compared to the target value via a metric called loss function. The weight matrix parameters $w_{i,jk}$ and biases $b_{i,k}$ are optimized by propagation of the gradient of the loss backwards through the network. Many sophisticated improvements to model design, loss functions, optimization, learning and prevention of overfitting have been made in the last few years. This paper will not cover any more details on this topic since this is not the main focus of this study. Here we refer to reviews and books about this field [41, 52].

A simple artificial neural network has been proposed by Nicely et al. [8] as a proof of concept for optical fringe reduction. The model design of the authors is kept very simple as they emphasize their goal being only to proof the concept of applicability. Clearly the model is not optimally chosen for the given task and does not learn very fast.

For the simulation in this study eight identical networks were trained on different subspaces of the target simulation and the choice of one specific network is set as a free parameter to be optimized during calibration. Each model is trained on synthetic data for 250 epochs with 2^{14} iterations each.

4 Applicability of Advanced Noise Reduction Methods

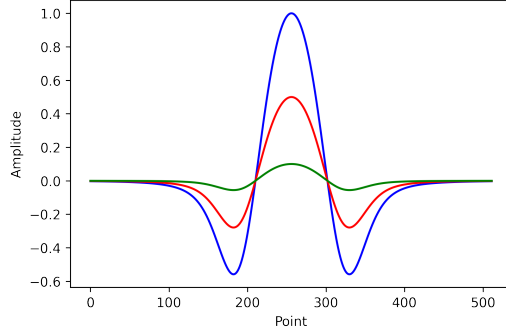


Figure 4.1: Reference $2f$ wavelength modulated spectrum of Voigt profile of carbon monoxide transition at wave number 2190.02 cm^{-1} for a pressure of 50 hPa, a temperature of 293 K and modulation depth $2.2\cdot\text{HWHM}$, scaled to a maximum amplitude of 1 (blue), 0.5 (red) and 0.1 (green).

4.3 Simulation

For the simulation a measurement scheme is chosen that mimics real experimental conditions and allows for optimization of the free parameters of the NRMs. The measurement scheme requires three different kinds of input data:

- **Reference:** A noise-free absorption spectrum. In experiments this could be either a very high concentration measurement or a simulation using the HITRAN database [114]. However, deviations of the used spectrum from the real underlying noise-free spectrum of the experiment can lead to undesired effects. In the simulation the reference is used to initialize some of the NRMs and also to calculate the absorbance.
- **Calibration:** A noisy spectrum with known mixing ratio. This could be a measurement used for instrument characterization or a calibration during field experiments. This data will be used to optimize free parameters of the NRMs for the given background signal.
- **Ambient:** Target noisy spectrum to be denoised.

As reference a Voigt profile is calculated using data from the HITRAN database [114]. The chosen absorption profile is an idealized absorption spectrum of carbon monoxide at wave number 2190.02 cm^{-1} for a pressure of 50 hPa and a temperature of 293 K [115]. The number of points is set to $2^9 = 512$, the line is centered at 2^8 and the scale is adjusted so that the Gaussian width σ caused by Doppler Broadening is set to 16. The modulation depth is set to $2.2\cdot\text{HWHM}$ according to the studies of Reid and Labrie [116]. The resulting $2f$ modulated spectrum is numerically calculated using (4.3) and is shown in Figure 4.1 for several amplitudes.

The noisy part of the spectrum \tilde{n}_{fc} is generated as a function of two parameters: frequency f and noise complexity c . The Fourier power spectrum is set to a Gaussian of

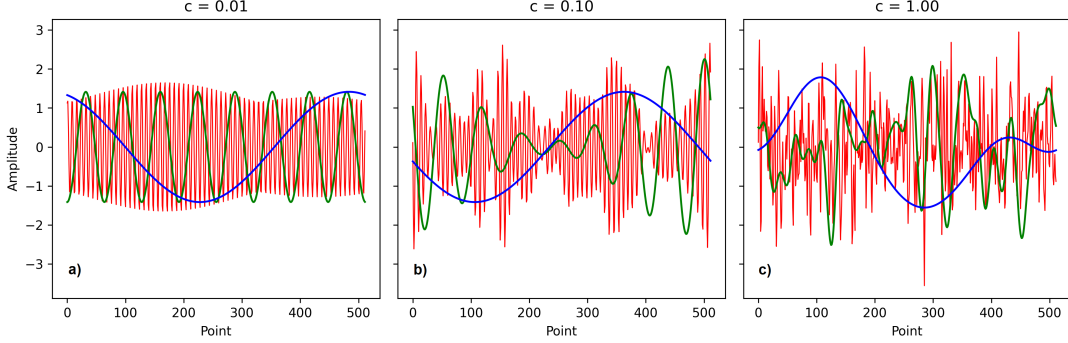


Figure 4.2: Example spectra for different noise parameters. Complexity c is set to 0.01, 0.1 and 1 in plot a), b) and c), respectively. In all plots the blue line, green line and red line corresponds to a frequency of 2^0 , 2^3 and 2^6 , respectively.

mean f and width $f \cdot c$ and the positive angles are sampled from a uniform distribution between $-\pi$ and π . The negative frequency angles are set to the negation of the positive angles to acquire a purely real signal after inverse Fourier transform. The complexity parameter c is a measure of the contribution to the noise by neighboring frequencies. For high values of c and f the noise becomes almost white, e.g. $c > 4$ and $f > 512$.

The resulting time signal is then normalized to $\sigma = 1$. Afterwards the reference spectrum is weighted with the third simulation parameter, the amplitude s , and added to the signal. Examples with $s = 0$ are shown in Figure 4.2.

$$\tilde{n}_{fc} = \mathcal{F}_f^{-1}(t) \left[\exp \left(\frac{-(f' - f)^2}{2f^2c^2} \right) \cdot e^{i\phi} \right] \quad (4.8)$$

$$\text{with } \phi \in \mathcal{U}(-\pi, \pi)$$

$$n_{sfc} = \left(\frac{\tilde{n}_{fc}}{\sqrt{\text{Var}[\tilde{n}_{fc}]}} \right) + s \cdot x \quad (4.9)$$

Now the three parameters are discretized along logarithmic scales to form a 64^3 cube consisting of $2^{18} = 262,144$ individual triplets. The signal amplitude, frequency and complexity range from 0.01 to 100, 2^0 to $M/2 = 2^8$ and 0.01 to 1, respectively. For each triplet 32 calibration spectra are generated for free parameter optimization and 1024 ambient spectra are simulated. A simple grid search is used to optimize the free parameters. In possible future simulations more sophisticated approaches can be applied to speed up the procedure; e.g. some publications propose algorithms to efficiently choose the optimum window of a Savitzky-Golay filter which is used in the EMD-FCR [117, 118]. The NRM is applied to the N ambient spectra using the optimized parameters. The absorbance a_i is retrieved by a final linear fit to the reference spectrum. These absorbances are then compared to the true amplitude s and the mean squared residual (MSR) and also the mean residual squared (MRS) are returned. The MSR and MRS are defined as follows:

$$\text{MSR} = \frac{1}{N} \sum_i (s - a_i)^2 \quad (4.10)$$

$$\text{MRS} = \left(s - \frac{1}{N} \sum_i a_i \right)^2 \quad (4.11)$$

This procedure is repeated for each NRM and also run once without noise reduction as a baseline. This is important since a least squares linear fit already shows great noise reduction properties under certain background conditions.

4.4 Simulation results

For more clarity the resulting values from the simulation MSR and MRS can be expressed by two different measures:

$$\Delta S = 10 \log_{10} \left(\frac{\text{MSR}_0}{\text{MSR}} \right) \quad (4.12)$$

$$\beta = \frac{\sqrt{\text{MRS}}}{s} \cdot 100 \% \quad (4.13)$$

MSR_0 is the baseline MSR without noise reduction. Now ΔS describes the increase in SNR in the unit of decibels (dB) and the bias β describes the relative deviation of the sample mean from the true value in percent. Note that in extreme cases the SNR increase can be negative and the bias can reach values above 100 %. These performance measures are displayed by individual animations, where each frame shows an image plot of the fc -plane at a certain signal amplitude s which is increased from frame to frame from 0.01 to 100. The animations can be found in the electronic supplementary material¹. As an example the Fourier Domain Analysis method (FDA) is evaluated explicitly in the following subsection, as it shows a variety of different characteristics. Then the performances of all NRMs is summarized and compared.

4.4.1 Case study: Fourier domain analysis

The resulting SNR increase and corresponding bias are shown in Figure 4.3 for several signal amplitudes s . In the first image at $s = 0.01$ the SNR increase gradually grows from 30 dB at highest frequency to about 60 dB at a frequency of 2^5 for complexity values below 0.2. There ΔS evolves into a stripe pattern of several overlapping discrete lines. Here ΔS ranges up to over 70 dB. Below frequency 2^3 the SNR increase of the stripes drops slightly and the pattern is overlaid with a smooth 40 dB region that tapers down at higher complexity. Finally ΔS increases again at lowest frequency. Meanwhile β is mostly close to zero for low complexities. The bias in the overlying smooth region at low frequency ranges from 20 % to 50 %, however. The bias is greatly

¹Electronic supplementary material is available in the online version of [100].

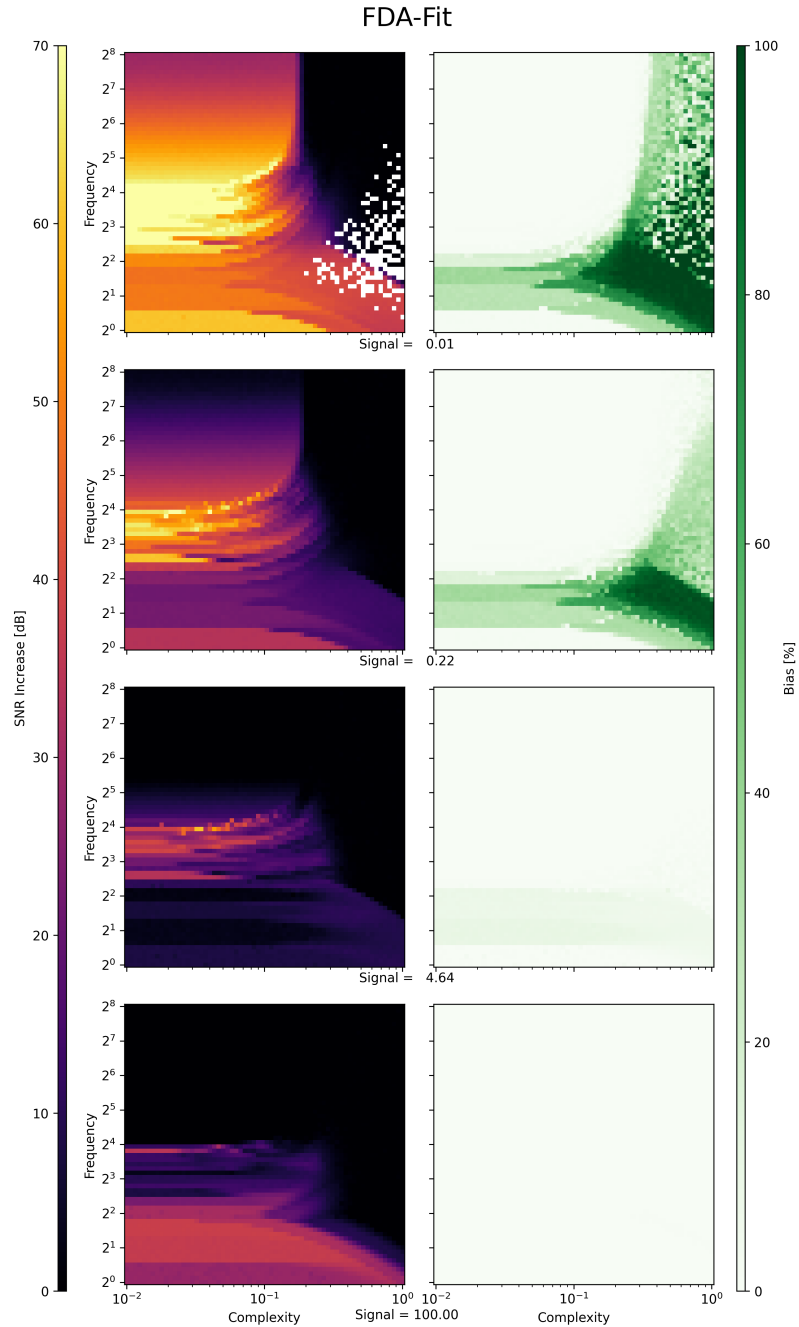


Figure 4.3: Sequence of image plots of SNR increase (left) and bias (right) of the FDA algorithm in the fc -plane for different signal amplitudes, increasing from top to bottom. Each plot sequence has a shared color map ranging from 0 dB to 70 dB and 0 to 100 %, respectively. White pixels in the SNR increase plot indicate points, where the bias is 100 % or above. Depending on the experimental requirements the method should only be applied when high SNR increase is achieved while maintaining a low bias.

4 Applicability of Advanced Noise Reduction Methods

enhanced at high complexity, especially at low frequency, where it reaches values above 100 %.

Obviously at high complexity the part of the Fourier spectrum containing the signal information is dominated by noise. The bias is reduced for low complexity but still significant for low frequencies. Here the noise interferes with the first peak of the Fourier power spectrum. At frequencies above 2^3 the algorithm effectively ignores parts of the power spectrum that differ from the underlying signal. The stripy pattern corresponds to discrete changes to the boundary indices of the mask chosen in the calibration phase. The gradual decrease for higher frequencies probably does not originate from the NRM but from the baseline performance of the least squares linear fit, which acts as a low pass filter and therefore performs better on higher frequency noise.

In the second plot at $s = 0.22$ the SNR increase has shrunk uniformly while the described patterns stay the same. This also follows from the baseline performance increasing with higher signal, hence lowering the ratio between absolute SNR and baseline SNR. The bias starts to tend to zero at high frequency and high complexity. The higher frequencies have lower impact on the method since all signal information is contained in the lower frequency components.

As displayed in the next image at $s = 4.64$ this process continues. The SNR increase is now constrained to the $2^2 - 2^4$ frequency region while β is close to zero almost everywhere. Only the overlying smooth region at low frequency and low complexity shows bias values up to 10 %. At the same time the SNR increase starts to rise at this region.

The final image at $s = 100$ supports this new characteristic. While bias and SNR increase have gone down to zero almost everywhere, SNR increase grows to about 40 dB in the former biased region at low frequency and low complexity. A cause could be an increased capability to reconstruct the low frequency part of the power spectrum from the higher frequency features due to the high signal. This implies that up to $s \approx 5$ only the first peak in the power spectrum may have been used by the algorithm. As a summary the FDA method performs extraordinary well at complexity below 0.2 and in the frequency range $2^3 - 2^5$ and also at lower frequencies for high input SNR. The described features are visualized in Figure 4.4. Here the method is applied to example spectra for three different settings.

4.4.2 Performance summary

Interpreting the information from the animations given in the electronic supplementary material, the other NRMs are evaluated in a similar manner as the case study from the previous section. These results will not be discussed in such detail at this point but are instead summarized to limit the extent of this paper.

The Wiener filter achieves an SNR increase up to 19 dB for frequency between 2^4 and 2^6 and complexity below 0.2, independent of signal amplitude s . Similar to the gradual decline of ΔS for higher frequencies due to baseline performance, this method performs better for lower frequencies inside the described region. The bias differs from

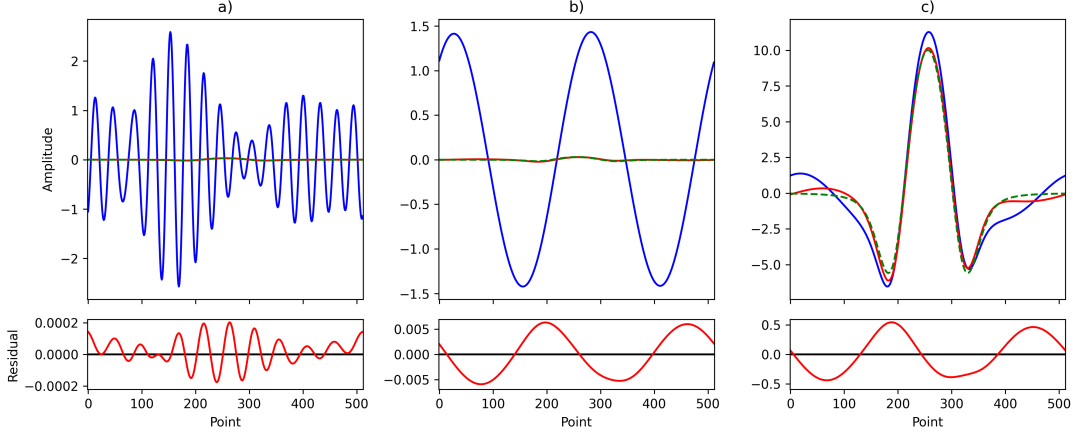


Figure 4.4: Examples of simulated signal (dashed green), noise (solid blue) and result of the FDA NRM (solid red) for three different variable settings. The lower subplots each show the residual (reconstruction minus true spectrum). The chosen values of the variables s , f , and c and the resulting SNR increase ΔS and β are given as follows: **a)** $s = 0.03$, $f = 16$, $c = 0.1$, $\Delta S = 47$ dB, $\beta = 4.7 \cdot 10^{-3}$ %; **b)** $s = 0.03$, $f = 2$, $c = 0.1$, $\Delta S = 39$ dB, $\beta = 32$ %; **c)** $s = 10$, $f = 2$, $c = 0.1$, $\Delta S = 19$ dB, $\beta = 0.34$ %. As mentioned in the text ΔS and β strongly depend on variable setting.

zero significantly only at very low frequency or high complexity.

The NRM based on SVD performs best at signals below 1, complexity below 0.08 and two different frequency ranges. At frequencies higher than 2^3 the method shows no significant performance gain compared to the baseline, between 1.4 and 2^2 β grows to values of 60 %. In this intermediate region the noise is too close to the signal to be distinguished.

The Discrete Wavelet Transform with Stein Threshold shows noisy SNR increase independent from input signal s for complexity below 0.2. Some broad bands emerge at frequencies equal to a power of two above 2^5 where ΔS reaches values above 90 dB. In this region the bias is always close to zero. The noisy SNR increase corresponds to a noisy choice of wavelet family and order in the calibration phase.

The method called EMD-FCR shows high bias for almost all parameter combinations. Obviously the output is lowered drastically if the noise free spectrum does not match with an IMF since the filtered IMFs are multiplied by the correlation coefficient. In the simulation even at high signals s the absorption spectrum is split into two or more IMFs very often. Here a thresholding scheme instead of a weighted sum could lead to better performance. However, at signal amplitude between 0.02 and 0.2 a narrow region of low bias and ΔS above 10 dB propagates along a curved path through the fc -plane. Here it seems the EMD separates signal from background and maps the absorption signal into a single IMF.

The final NRM analyzed is the combined set of simple artificial neural networks (ANN). Although the base model is primitively designed and was not optimized for this special task, the results show an average SNR increase of 15 dB and average bias

4 Applicability of Advanced Noise Reduction Methods

below 1% for signal amplitude between 1.4 and 17, frequency below 2^2 and complexity below 0.26. The performance coincides with a particular network choice during calibration phase.

Figure 4.5 shows an overview of the best performance for each NRM in the fc -plane. The corresponding mean values for ΔS and β are given in the figure caption.

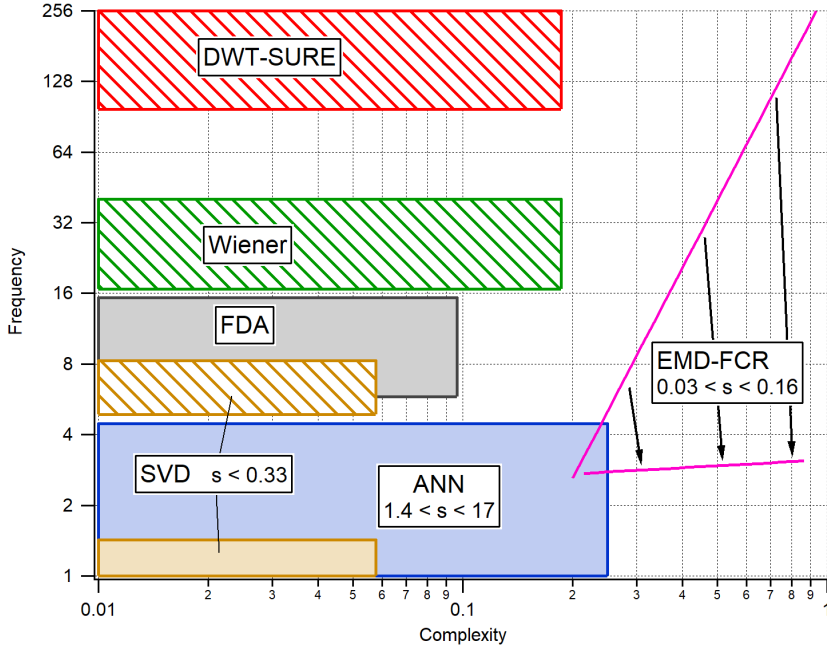


Figure 4.5: Highlights of the best performances of the analyzed NRMs in the fc -plane. The mean SNR increase and bias is given for the rectangular regions or the line trajectory, respectively: **Wiener:** $\overline{\Delta S} = 7.3$ dB, $\overline{\beta} = 8.7 \cdot 10^{-4}$ %; **SVD:** high frequency region: $\overline{\Delta S} = 24.9$ dB, $\overline{\beta} = 2.0$ %; low frequency region: $\overline{\Delta S} = 40.9$ dB, $\overline{\beta} = 5.6$ %; **DWT-SURE:** $\overline{\Delta S} = 10.6$ dB, $\overline{\beta} = 1.2 \cdot 10^{-5}$ %; **FDA:** $\overline{\Delta S} = 38.2$ dB, $\overline{\beta} = 7.1 \cdot 10^{-3}$ %; **EMD-FCR:** $\overline{\Delta S} = 14.3$ dB, $\overline{\beta} = 4.9$ %; **ANN:** $\overline{\Delta S} = 15.0$ dB, $\overline{\beta} = 1.0$ %. Rectangles filled with a stripe pattern indicate negative values of ΔS .

4.4.3 Remarks

Now some flaws of the simulation will be mentioned which influence the results of this paper.

- The reference spectrum and added signal is assumed to be noise-free and stable over time. Distortions and drifts in real experiments can influence the performance of some of the NRMs. Prominent effects that can cause distortions in experimental conditions that have not been considered are residual amplitude modulation [27], the phase shift between wavelength modulation and intensity modulation [119] and the non-linearity of (4.1).

- Although the simulated noise covers a big range of noise characteristics, it is only an idealization of real background structures. Only fringe-like noises with a dominant frequency have been considered. Thus, the performance of some NRMs may vary depending on the background structure.
- Since the number of simulated ambient spectra is finite, this paper only gives statistical estimates of expected results.
- Since the number of simulated calibration spectra is finite, the grid search not always chooses the optimal free parameters. This could result in an underestimation of performance.

4.5 Conclusion

There is no best noise reduction method for wavelength modulation spectroscopy. Parameters like the dominant noise frequency, noise power spectrum confinement and input signal strength can vastly influence the performance of some methods and lead to distortions of the output. A proper algorithm has to be carefully chosen according to the experimental setup. The results from this paper can help predict possible performances given the noise parameters for a particular setup. Each analyzed method showed different behavior for the simulation parameters. Although some NRMs were altered slightly to fit the simulation conditions, all of them were able to outperform the least squares linear fit by at least 10 dB without introducing large biases. The best performances were highlighted and little to no overlap of these best performance ranges were found. This result could explain the odd phenomenon that the methods outperformed other established techniques in the studies in which they were proposed, sometimes leading to apparent contradictions between different studies.

While a quantitative comparison for a large space of noisy input shapes was given, many questions arose or were left unanswered. There are infinitely many more noise shapes to be tested. The effect of distortions to the noise free spectrum were ignored. The choice of free parameters and small adaptations to the methods could be improved. Finally there is still a long way to be able to identify and evaluate the best noise reduction method under experimental conditions, but this paper tries to take a first big step.

A follow-up study is planned to further analyze and quantify the results given in this paper under several experimental conditions. The methods will be tested for two different absorption spectrometers (different noise patterns) and several different trace gases (varying initial signal to noise ratios).

5 Data Quality Enhancement via Sequential Monte Carlo Filters

This chapter was published as "Data Quality Enhancement for Atmospheric Chemistry Field Experiments via Sequential Monte Carlo Filters" by Copernicus, Mar 07, 2023, in Atmospheric Measurement Techniques [120] under the terms of the Creative Commons CC BY license: <https://creativecommons.org/licenses/by/4.0/>.

How to cite: Röder, L. L., Dewald, P., Nussbaumer, C. M., Schuladen, J., Crowley, J. N., Lelieveld, J., and Fischer, H.: Data quality enhancement for field experiments in atmospheric chemistry via sequential Monte Carlo filters, *Atmos. Meas. Tech.*, 16, 1167–1178, <https://doi.org/10.5194/amt-16-1167-2023>, 2023.

My contribution to the study is the design of the study, carrying out the data analysis, drafting text and figures and literature research.

Abstract

In this study, we explore the applications and limitations of sequential Monte Carlo (SMC) filters to field experiments in atmospheric chemistry. The proposed algorithm is simple, fast, versatile and returns a complete probability distribution. It combines information from measurements with known system dynamics to decrease the uncertainty of measured variables. The method shows high potential to increase data coverage, precision and even possibilities to infer unmeasured variables. We extend the original SMC algorithm with an activity variable that gates the proposed reactions. This extension makes the algorithm more robust when dynamical processes not considered in the calculation dominate and the information provided via measurements is limited. The activity variable also provides a quantitative measure of the dominant processes. Free parameters of the algorithm and their effect on the SMC result are analyzed. The algorithm reacts very sensitively to the estimated speed of stochastic variation. We provide a scheme to choose this value appropriately. In a simulation study O_3 , NO , NO_2 and j_{NO_2} are tested for interpolation and de-noising using measurement data of a field campaign. Generally, the SMC method performs well under most conditions, with some dependence on the particular variable being analyzed.

5.1 Introduction

Insight into the complex chemical system of the atmosphere is often achieved by conducting coordinated field experiments where an ensemble of trace gases, meteorological variables, physical properties and aerosol compositions are measured that comprehensively characterize the sampled air masses [121, 122, 123]. Field campaigns track variables along a spatiotemporal trajectory and are prone to local and temporal events. These are not resolvable by satellite measurements or chemical-transport models.

Quantitative analysis of data from field campaigns is often hindered by low data quality and insufficient data coverage of all parameters needed at each time step. The latter may result from poor instrumental time resolution, sporadic instrument failures, measurement duty cycles or instrument calibration. Assuming uncorrelated data loss of just 10 % per instrument, a field experiment with 10 different measurement instruments would lose 65 % of simultaneous data.

The reconstruction or enhancement of time steps with lost data or poor data quality is not easily achievable. Linear interpolation and moving average filters act as low pass filters that dampen high frequency variations of the measured variables. Thus, the main advantage of the field measurement compared to remote sensing is suppressed. The calculation of missing data with photostationary state (PSS) calculations works for many species, but introduces a bias as all other processes are disregarded without an estimate of reconstruction error [124]. Using the outputs of chemical-transport models as replacements lowers the local and temporal resolution. The latter approach also contradicts the goal of some field campaigns that try to evaluate model predictions [125].

Sequential Monte Carlo (SMC) methods have become a useful tool in combining prior knowledge of a dynamical system with noisy measurements. Originally applied to trajectory reconstruction [126, 127], this method has become a major tool in meteorology for data assimilation to a theoretical model [128, 129]. In these fields, the SMC model is applied to enhance the performance of a trusted model with additional information provided by noisy measurements.

Ensemble Monte Carlo methods have been used in relation to atmospheric chemistry measurements where they enabled the estimation of dynamics [130], reactions [131] or emission sources [132, 133]. Other novel applications cover enhancements of neural networks and machine learning methods [134, 135].

The goal of this work is to explore the SMC method in the enhancement of data quality, data coverage and in the augmentation of data to include unmeasured species in a system of measured atmospheric variables that are connected via known chemical reactions. The focus is shifted from the enhancement of model outputs as in most recent studies [129] towards data quality enhancement as originally intended [136]. Enhanced data quality enables a more comprehensive data analysis of field campaign measurement data. Several applications will be tested on measurements taken in July 2021 at the Taunus Observatory, Kleiner Feldberg, Germany [98]. The study will focus on the chemical system of ozone (O_3), nitric oxide (NO), nitrogen dioxide (NO_2) and

the photolysis frequency of NO₂ (j_{NO_2}).

In the following section, the basic theory of the SMC method will be explained. In Sect. 5.3, the underlying chemistry of the considered system and the measurement techniques used to derive the dataset will be described. In Sect. 5.4, several experiments using the measured data and the SMC method will be conducted and discussed.

5.2 Sequential Monte Carlo

The N -dimensional state vector of a system is defined as $\mathbf{x}_{t_n} =: \mathbf{x}_n \in \mathbb{R}^N$ at time steps t_n ($n \in \{1..L\}$). This vector contains all unknown or hidden true values in the system. This state vector evolves partly deterministically according to a transition function f_n and partly stochastically by addition of some noise w_n . The counterpart of \mathbf{x}_n is the measurement vector $\mathbf{y}_n \in \mathbb{R}^M$ that contains all available measurements at time step t_n . The elements of the state vector and the measurement vector are connected via an auxiliary function h_n that depends on the state vector and measurement noise v_n . The functions f and h can also depend on auxiliary parameters \mathbf{u} that are considered to be exact.

5.2.1 Basic procedure

The implementation of an SMC algorithm requires a known conditional probability distribution function (PDF) $p(\mathbf{y}_n|\mathbf{x}_n)$ that can be easily calculated and a procedure to sample from the prior PDF [137]:

$$p(\hat{\mathbf{x}}_n) = p(\mathbf{x}_n|\mathbf{x}_{n-1}). \quad (5.1)$$

where the distribution $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ contains prior information about the dynamics of the state vector and $p(\mathbf{y}_n|\mathbf{x}_n)$ describes the probability of a measurement given a particular realization of the state vector \mathbf{x}_n . The latter probability distribution encodes the uncertainty of the measurement instruments.

Applying Bayesian theory, the posterior PDF results from the calculation of the expression:

$$p(\mathbf{x}_n) = p(\mathbf{x}_n|\mathbf{y}_n) = \frac{p(\mathbf{y}_n|\hat{\mathbf{x}}_n)p(\hat{\mathbf{x}}_n)}{p(\mathbf{y}_n|\mathbf{y}_{0:n-1})}, \quad (5.2)$$

where $p(\mathbf{y}_n|\mathbf{y}_{0:n-1})$ depends on all previous information. Now the process can be considered Markovian as the dependence on all previous information can be described by direct dependence on the most recent time step only: $p(\mathbf{y}_n|\mathbf{y}_{0:n-1}) = p(\mathbf{y}_n|\mathbf{y}_{n-1})$. This distribution is most likely not a retractable expression [138]. An exception is the Kalman filter [139] that requires the transition function f and the measurement function h to be linear with purely Gaussian-distributed state and measurement noise. These conditions are not met in a system of chemical reactions as the reactions can be highly nonlinear and abundances of molecules follow a probability distribution function (PDF) that is 0 for negative abundances.

5 Data Quality Enhancement via Sequential Monte Carlo Filters

The main idea to overcoming this numerical limitation in SMC filters is the approximation of the PDF of \mathbf{x}_n with a finite number of samples $\mathbf{x}_n^{(i)}$ (particles). The PDF of the state vector is approximated by the empirical distribution:

$$p(\mathbf{x}_n) \approx \frac{1}{K} \sum_{i=1}^K \delta(\mathbf{x}_n - \mathbf{x}_n^{(i)}), \quad (5.3)$$

where K represents the number of particles and δ denotes the Dirac measure. If the particles are sampled from the true PDF, the empirical PDF approximates the true PDF for $K \rightarrow \infty$. Given that the initial particles were sampled according to $p(\mathbf{x}_0)$ this can be ensured by sequential updating of the particles followed by a bootstrap filter. The particles are updated by application of the transition function to form the prior PDF:

$$\hat{\mathbf{x}}_n^{(i)} = f_n(\mathbf{x}_{n-1}^{(i)}, w_n^{(i)}) \sim p(\mathbf{x}_n | \mathbf{x}_{n-1}). \quad (5.4)$$

Then a weight is calculated for each particle that is related to the distance of the particle to the measurement:

$$q^{(i)} = \frac{p(\mathbf{y}_n | \hat{\mathbf{x}}_n^{(i)})}{\sum_{j=1}^K p(\mathbf{y}_n | \hat{\mathbf{x}}_n^{(j)})}. \quad (5.5)$$

The posterior distribution is then approximated by bootstrap resampling [137, 138] from the prior particles according to their individual weight $q^{(i)}$. This is implemented by calculation of the cumulative sum of all weights and choosing the k th particle where a uniform random variable $u^{(i)}$ is less or equal to the cumulative sum up to the k th particle:

$$u^{(i)} \sim \mathcal{U}(0, 1), \quad (5.6)$$

$$\mathbf{x}_n^{(i)} = \hat{\mathbf{x}}_n^{(k)} \quad \text{where } u^{(i)} \leq \sum_{j=1}^k q^{(j)}, \quad (5.7)$$

$$p(\mathbf{x}_n) \approx \frac{1}{K} \sum_{i=1}^K \delta(\mathbf{x}_n - \mathbf{x}_n^{(i)}). \quad (5.8)$$

5.2.2 Auxiliary particle filter

This approximation assumes that K is large enough. The main challenge of the method is the possibility of the particles to collapse into a single mode [140]. It is possible that a single particle carries a weight very close to 1 while all other particles carry weights close to 0. In these cases, the posterior approaches a δ -distribution without any statistics. In the literature, there are many approaches to counter this problem, maintaining similar weights for all particles [129]. This is especially important for high-dimensional problems such as data assimilation, as the number of particles has

to grow exponentially with the size of the measurement vector M [140]. A common approach [138, 129] suggests sampling from a proposal distribution $q(\mathbf{x}_n|\mathbf{y}_{0:n})$ instead of the prior, that nudges the particles into the direction of the posterior before applying the bootstrap filter. Pitt and Shephard [127] described a method they called *Auxiliary Particle Filter* where they define the proposal PDF as

$$q(\mathbf{x}_n^{(i)}|\mathbf{y}_{0:n}) = p(\boldsymbol{\mu}_n^{(i)}|\mathbf{y}_n), \quad (5.9)$$

where $\boldsymbol{\mu}$ is a likely draw from the prior PDF. This can be achieved by a simplified version of the transition function f_n without a stochastic part. For each particle an intermediate weight $\lambda^{(i)}$ is calculated and afterwards $R > K$ samples are drawn from the particles where

$$p(j = i) \propto \lambda^{(i)} \quad \text{with } j \in \{1..R\}. \quad (5.10)$$

The prior is then constructed from this mixture prior analogous to Eq. (5.4). For the posterior evaluation the weights (5.5) have to be rescaled by the first stage weights (5.10) to compensate the introduced bias before applying the bootstrap filter (5.7). This method can still lead to weight collapse but increases the statistics and efficiency of the SMC method as the particles of the posterior empirical PDF are less likely to be degenerate by construction [129]. Recently Fearnhead and Künsch [141] and van Leeuwen et al. [129] reviewed several novel approaches to counter this problem for high dimensional systems such as weather forecasting; Pulido and van Leeuwen [142] proposed a particle flow formalism that completely counters weight collapse [143]. These methods will however not be considered in more detail, since they are typically dealing with $N \sim 10^9$ and $M \sim 10^7$, while a measurement field campaign lies in the range $N, M < 100$. Due to the high dimension of the former systems, it is very likely that several measurements differ from the prediction by many standard deviations. As mentioned above, the goal in those cases is to optimize a trusted model with the information provided via noisy measurements. In our case, the centerpiece of the system is the observation. The results from our SMC algorithm should never disagree with the measurements but should rather assist the observations, finding a more precise estimate of a variable that is similar to a weighted average of several measurements.

Further weight-maintaining adaptations to the SMC method can be considered for future applications. For now, weight collapse will be tracked throughout the experiments as a metric. In this study, the following entropy will be considered:

$$H(\hat{\mathbf{x}}_n) = - \sum_{i=1}^K \lambda_n^{(i)} \log \left(\lambda_n^{(i)} \right), \quad (5.11)$$

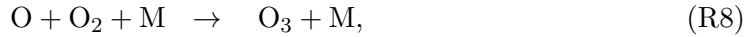
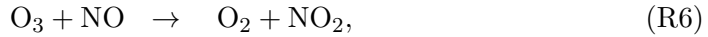
$$H(\mathbf{x}_n) = - \sum_{i=1}^K q_n^{(i)} \log \left(q_n^{(i)} \right), \quad (5.12)$$

where H is close to its maximum value $\log(K)$ or $\log(R)$, respectively, when all particles share similar weights. The maximum value is reached if and only if the measurement does not contribute any additional information. The effective dimension R^* of

a posterior can be approximated by $\exp(H)$ where a total collapse to a single particle corresponds to $H \rightarrow 0$ and $R^* \rightarrow 1$. Low entropy is not necessarily a tracer for poor performance of the SMC method but might indicate vast deviations of the actual chemical system from the considered model. An example might be sudden emission of relevant trace gases, changes in wind direction or other local effects.

5.3 Chemical reactions and measurements

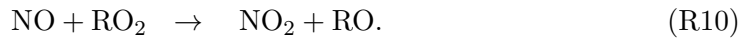
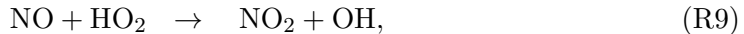
This study focuses on the interplay between tropospheric O_3 , NO and NO_2 . According to Leighton [144] and Nicolet [145] the concentrations of these trace gases reach a steady state for a few minutes during daytime. The relevant reactions are



where Reaction (R8) can be considered fast compared to Reaction (R7). The reaction coefficient $k_{O_3,NO} =: k_1$ is taken from [146]. The photolysis frequency j_{NO_2} varies between ca. 0 at night and several $10^{-3}s^{-1}$. The photostationary state is reached when

$$k_1 [O_3] [NO] = j_{NO_2} [NO_2]. \quad (5.13)$$

Under atmospheric conditions, this photostationary state is additionally affected by peroxy radicals predominantly originating from the oxidation of volatile organic compounds (VOCs) by e.g. OH or O_3 . Both hydroperoxy (HO_2) and organic peroxy radicals (RO_2) convert NO to NO_2 (Reactions R9 and R10). In addition, further chemical reactions, direct emission, deposition and transport processes influence this steady state [147, 148, 124]:



The coordinated measurements (TO21 campaign) took place in July and August 2021 on the Kleiner Feldberg mountain (826m, $50^\circ 13' 18''$ N, $8^\circ 26' 45''$ E), Germany, located in a rural, forested region under anthropogenic influence from several large cities within a radius of ca. 35km. This site has previously been used for field campaigns and is described in more detail in [149] and [150]. NO and NO_2 were measured via a photolysis-chemiluminescence detector described in [151] and [18]. j_{NO_2} was calculated from actinic flux measurements by a spectral radiance detector (MetCon GmbH) [152]. Ozone was measured by two commercial UV absorption monitors (2B Technologies). Several other trace gases and chemical variables were measured during the campaign that will not be considered in this study. An in-depth discussion of all measurements in this campaign can be found in [98]. Meteorological data were provided by a weather station of the German Weather Service (DWD) on the summit.

5.3.1 SMC setup

The setup used in this study is based on the following definition: the state vector \mathbf{x} and the measurement vector \mathbf{y} are both four-dimensional and encode the mixing ratio of O₃, NO and NO₂ in units of parts per billion volume (ppbv) and the photolysis frequency j_{NO_2} in s⁻¹. Therefore, the auxiliary function h simplifies to the identity. The transition function f is composed of an initial randomization of each dimension that follows a lognormal distribution and numerical integration of the differential equation resulting from the chemical reactions. The parameters of the distribution are chosen so that the mean and standard deviation are equal to the current value and a given standard deviation σ_0 , respectively. The choice of a lognormal distribution for chemical systems has been discussed, e.g., in [153], and solves the problem of otherwise possible negative values for the abundances. With the scheme proposed here, the lognormal distribution approximates a Gaussian distribution as the standard deviation becomes smaller than the mean. Reactions (R6) and (R7) result in the differential equation:

$$\frac{d}{dt}\mathbf{x} = \begin{pmatrix} -k_1 [\text{O}_3] [\text{NO}] \frac{p}{k_B T} + j_{\text{NO}_2} [\text{NO}_2] \\ -k_1 [\text{O}_3] [\text{NO}] \frac{p}{k_B T} + j_{\text{NO}_2} [\text{NO}_2] \\ -j_{\text{NO}_2} [\text{NO}_2] + k_1 [\text{O}_3] [\text{NO}] \frac{p}{k_B T} \\ 0 \end{pmatrix}, \quad (5.14)$$

where k_B is Boltzmann's constant that converts the reaction coefficient to units of parts per billion by volume per second ppbv⁻¹s⁻¹ at given pressure p and temperature T . These values will be input into the calculation as auxiliary variables for each time step. The parameters p , T and k_1 will be set as fixed values and their uncertainties will not be considered. However, any deviations from the true values can be compensated via the initial randomization. The function used to calculate the weights as in Eq. (5.5) is defined as the product of Gaussian kernels:

$$p(\mathbf{y}_n | \hat{\mathbf{x}}_n) \propto \prod_{m=1}^M \exp\left(-\frac{(\hat{\mathbf{x}}_{n,m} - \mathbf{y}_{n,m})^2}{2\sigma_{n,m}^2}\right), \quad (5.15)$$

where $\sigma_{n,m}$ is constructed by

$$\sigma_{n,m}^2 = \text{DL}_m^2 + (\text{P}_m \hat{\mathbf{x}}_{n,m})^2, \quad (5.16)$$

using the characteristic detection limit DL and precision P for each instrument. In cases of missing measurements, the factor is set to 1. To ensure numerical stability, Eq. (5.15) in the actual algorithm is replaced by

$$\log(p(\mathbf{y}_n | \hat{\mathbf{x}}_n)) = \sum_{m=1}^M -\frac{(\hat{\mathbf{x}}_{n,m} - \mathbf{y}_{n,m})^2}{2\sigma_{n,m}^2}. \quad (5.17)$$

The model output is constructed from full Bayesian inference to convert the approximate probability distribution to an estimate for the state and an estimate of the

error:

$$\bar{\mathbf{x}}_n = \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n)} [\mathbf{x}_n] \approx \frac{1}{K} \sum_{i=1}^K \mathbf{x}_n^{(i)}, \quad (5.18)$$

$$\Delta \mathbf{x}_n = \sqrt{\mathbb{V}_{\mathbf{x}_n \sim p(\mathbf{x}_n)} [\mathbf{x}_n]} \approx \left(\frac{1}{K-1} \sum_{i=1}^K (\mathbf{x}_n^{(i)} - \bar{\mathbf{x}}_n)^2 \right)^{\frac{1}{2}}. \quad (5.19)$$

5.3.2 Model extension

In each iteration the individual particles will evolve towards the photostationary state where Eq. (5.14) equals 0. This state corresponds to the photostationary state for a given $\text{NO}_x = \text{NO} + \text{NO}_2$, $\text{O}_x = \text{O}_3 + \text{NO}_2$ and j . Particles which are close to the measurement then have high probability to be sampled. If no measurement is available, all particles have an equal chance of being sampled. Therefore, the posterior will shift towards the photostationary state in the unsupervised state. This behavior can cause high biases in combination with low uncertainty on the prediction if no measurements are available and, at the same time, the chemistry is dominated by processes not regarded in the algorithm. During nighttime, the photolysis frequency of NO_2 is 0 so other sources of NO, e.g., emissions from soil or plants play a dominant role [154]. An example of this effect can be found in the supplement.

Thus, we extend the state vector described in the previous section with an additional variable $\eta \in \{0, 1\}$ so that $\mathbf{x}^* = (\mathbf{x}, \eta)$. This variable will be called *activity* and gates the differential equation according to

$$\frac{d}{dt} \mathbf{x}^* = \eta \frac{d}{dt} \mathbf{x}. \quad (5.20)$$

Now each particle can be either *active* or *passive* if $\eta = 1$ or 0, respectively. If the chemical reactions incorporated into Eq. (5.14) dominate the chemistry, it is more likely that *active* particles survive and vice versa. A small probability p_η to switch activity is included into the randomization phase to prevent mode collapse. This way the algorithm can turn chemical processes on and off, whichever is more likely according to the measurements. Additionally, the mean value of η can give insights into the relative importance of the chemical processes incorporated.

Figure 5.1 shows the diel profile of η for the complete dataset in comparison with the diel average of j_{NO_2} , along with a box-whisker plot. For low photolysis frequencies, the activity lies close to 0 and sharply increases with higher actinic flux. At $j \approx 0.003 \text{ s}^{-1}$, the median activity rises above 50% and later saturates at around 90%. The 10% and 25% quantiles suggest a very skewed distribution at noon. This indicates deviations from the PSS calculation due to other processes, e.g., Reaction (R9).

5.3.3 Comparison with constrained box-model calculations

Similar calculations have been conducted using observationally constrained box-models [155, 23, 98] in which a selection of measured parameters (e.g., trace gas mixing

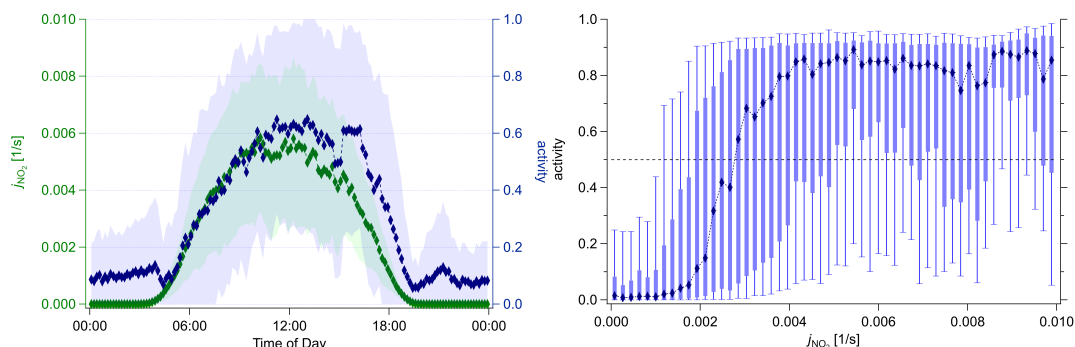


Figure 5.1: Left: diel profile of photolysis frequency j_{NO_2} (green) and activity η (blue) averaged for each minute interval. Time is in UTC. Right: Box-Whisker plot of activity as a function of photolysis frequency. The markers and dotted line mark the median, boxes range from the 25 % quantile to 75 % quantile, the whiskers mark the 10 % and 90 % quantiles, respectively. The black dashed line marks the transition from passive to active regime $\eta > 0.5$. A clear nonlinear correlation is visible. Although the correlation is nonlinear the Pearson correlation coefficient equals 0.67.

ratios and photolysis rates) are used as time-dependent inputs for a detailed chemical reaction scheme. Physical effects, e.g., deposition and uptake can be adjusted to best replicate measured outputs [98]. Sensitivity studies can be conducted by a variation of reaction rates and other parameters [23]. Typically, these model calculations have runtimes in the order of seconds for a full dataset, dependent on the number of reactions.

From a qualitative perspective, these calculations have some similarities with the SMC method. However, in our case, the choice of appropriate constraints is based on Bayesian theory and the quantitative measurement uncertainty. Sensitivity studies are automatically obtained due to the description of the state as a probability distribution. Unconsidered effects can be compensated via stochastic variability. Also, measurement errors are not directly propagated to the output since the measurement vector is separated from the state vector. In the limit of low constraint uncertainties and full chemical description of the system, the outputs of SMC and box-model calculations converge. In other cases, the latter may be used to prepare a full SMC run, benefiting from its low runtime, and enables detailed chemical investigation of the system [23].

5.4 Experiments

In order to study the effect of the SMC method on time series of chemical systems, several experiments were conducted on the measurement. The capability of the method to interpolate missing data points was tested by artificially discarding data and comparing the reconstruction of the SMC algorithm with the original measurement. The result is evaluated using the mean square error (MSE), and the squared error divided

by the standard deviation (χ^2):

$$\text{MSE} = \frac{1}{L} \sum_n (\bar{\mathbf{x}}_n - \mathbf{y}_n)^2, \quad (5.21)$$

$$\chi^2 = \frac{1}{L} \sum_n \left(\frac{\bar{\mathbf{x}}_n - \mathbf{y}_n}{\Delta \mathbf{x}_n} \right)^2. \quad (5.22)$$

This will be described in more detail in Sect. 5.4.1 and 5.4.2. Another possible application is enhancement of the precision of a measurement within a system. For this test, white noise is added to the observed data. The reconstruction is also analyzed in terms of MSE and χ^2 . The model is performing well if the MSE is close to the uncertainty of the measurement and χ^2 is close to 1. Finally we discuss the possibility to augment the dataset to include unmeasured variables.

We give a depiction of the algorithm used in Sect. 5.4.1 and 5.4.2 in Algorithm 1.

Algorithm 1 Auxiliary particle filter in a O_3 , NO , NO_2 , j_{NO_2} , η system.

Require: $\mathbf{x} = (\text{O}_3, \text{NO}, \text{NO}_2, j_{\text{NO}_2}, \eta)$

for all time steps t_n **do**

Auxiliary phase

$\boldsymbol{\mu}^{(i)} \leftarrow \mathbf{x}_{n-1}^{(i)} + \dot{\mathbf{x}}_{n-1}^{(i)} \Delta t$ using (5.14), (5.20)

 Calculate $\lambda^{(i)} \leftarrow p(\mathbf{y}_n | \boldsymbol{\mu}^{(i)})$ using (5.17)

 Sample R random particles $\mathbf{x}^{(j)}$ from $\mathbf{x}_{n-1}^{(i)}$ weighted by $\lambda^{(i)}$

Randomization phase

for all particles $\mathbf{x}^{(j)}$ **do**

for all $\xi \in \{\text{O}_3, \text{NO}, \text{NO}_2, j_{\text{NO}_2}\}$ **do**

 Resample ξ from lognormal distribution with mean ξ and standard deviation

$\sigma_{0,m}$

end for

 Switch η to $1 - \eta$ with probability $p_\eta = 0.025$

end for

$\mathbf{x}^{(j)} \leftarrow \mathbf{x}^{(j)} + \dot{\mathbf{x}}^{(j)} \Delta t$ using (5.14), (5.20)

 Calculate $q^{(j)} \leftarrow p(\mathbf{y}_n | \mathbf{x}^{(j)})$ using (5.17)

 Rescale by auxiliary weights $q^{(j)} \leftarrow \frac{q^{(j)}}{\lambda^{(i)}}$

 Sample K random particles $\mathbf{x}^{(i)}$ from $\mathbf{x}^{(j)}$ weighted by $q^{(j)}$

$\mathbf{x}_n^{(i)} \leftarrow \mathbf{x}^{(i)}$

end for

5.4.1 Interpolation

The SMC method is tested as an alternative to interpolation of missing data by randomly discarding sections of data with interval size T . This process is repeated for each dimension of the state vector i.e. each molecule and the photolysis frequency. Then the missing data are reconstructed.

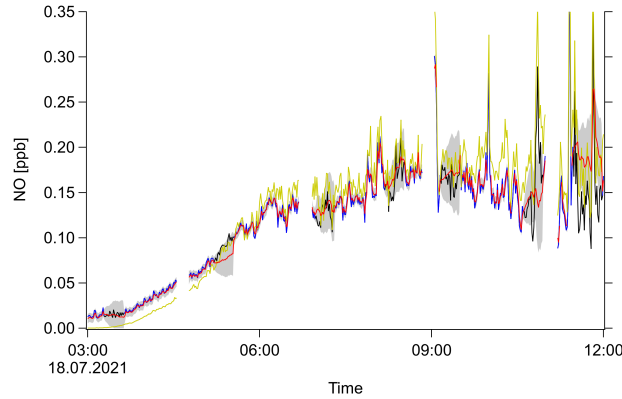


Figure 5.2: Example result of SMC used for interpolation. NO mixing ratio as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial gaps (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded gray region).

The algorithm described at the beginning of this section is applied to the whole dataset. Missing data in each dimension are automatically interpolated since the algorithm returns a value for \boldsymbol{x} at all time steps. The uncertainty is given by the standard deviation of the ensemble of particles (5.19). If data are missing in some dimension, the likelihood of the particles is less sparse. This leads to survival of more particles and a higher spread of the posterior. Hence, the standard deviation and the entropy increase. Once measurement data are available again, only data points close to the measurements will be sampled. Entropy and standard deviation decrease again. If the mean of the distribution strongly deviates from the measurement at this point, only a few particles survive and both entropy and standard deviation become very small. The resulting standard deviation underestimates the uncertainty of the model at this point. The requirements for the approximation of the posterior with the finite sample of particles does not hold anymore. Therefore, data points where the entropy is small will be discarded. Since there are effectively less particles left to sample from than samples to be drawn, this threshold is set to

$$H(\boldsymbol{x}_n) < \log(K) \Leftrightarrow R^* < K \quad (5.23)$$

After a low entropy incident, the ensemble may require a few iterations to converge again. Considering this effect and discarding additional points may increase the data accuracy while lowering data coverage. Throughout this analysis no additional points were discarded. In applications, the number of low entropy events may be reduced using an increased ensemble size which requires a longer runtime.

Figure 5.2 shows an example plot of NO with random artificial data gaps of 30 min that have been interpolated by the SMC method. Depending on activity, the SMC ensemble mean either tends towards the PSS equilibrium or stays approximately constant, while the ensemble spreads and increases the standard deviation. This spreading happens fast in the beginning of a data gap and slows down afterwards. This might follow the

5 Data Quality Enhancement via Sequential Monte Carlo Filters

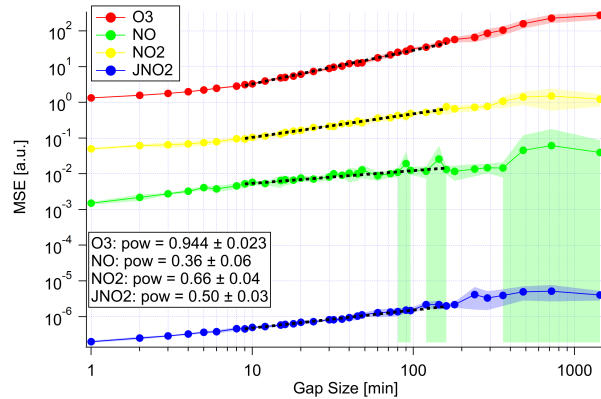


Figure 5.3: MSE of the SMC estimation as a function of artificial data gap size to study the interpolation capabilities. Mean MSE as lines and markers and standard deviation as shaded region for the ensemble of repetitions. The plot shows the results of all variables: ozone (red), NO (green), NO₂ (yellow) and j_{NO_2} (blue). Note that the unit of MSE is arbitrary to fit all variables in one plot. The unit is parts per billion by volume squared (ppbv²) for the MSE of trace gases and inverse seconds squared (s⁻²) for the MSE of the photolysis frequency. The dashed black lines show power-law fits ($y = Ax^{\text{pow}}$) fitted to the intermediate regions. The fit estimate for pow is given in the annotation.

\sqrt{N} behavior of a sum of normally distributed variables. In this example plot, the spreading speed matches the behavior of the system so that the measurement at the end of a data gap lies within the $\pm 1\sigma$ interval.

This procedure was repeated for each species and a wide range of data gaps between 1 min and 1 d. The data gaps were shuffled eight times for each setup to achieve better statistics. Figure 5.3 shows the resulting MSE. For low gap sizes, the MSE stays constant. This constant corresponds to the base deviation of the SMC estimate from the unaltered measurement. This value is expected to be larger than 0, since the model combines the prior knowledge with the measurement and therefore introduces a small bias. We will call this bias the *intrinsic model bias*.

With increasing gap size, the MSE starts to increase. In Figure 5.3, it is clearly visible that the slope varies strongly with different variables. This slope corresponds to the power-law coefficient of MSE with larger gap size. In the uninformative case of linear interpolation and Brownian noise, this slope is equal to 1. A lower power-law coefficient indicates an effective contribution of information through the remaining measurements considered. This coefficient is close to 1 for ozone. Therefore, this method is not capable of estimating the course of ozone in cases of instrument failure considering this particular system. This is not a strong limitation since ozone can be measured precisely enough with commercial instruments. However, one should keep this limit in mind in other systems where a species cannot be effectively described by the chemistry of the remaining variables in the system.

At very high gap sizes, the MSE jumps to higher values and the standard deviation also increases. This indicates a higher sensitivity to the particular data gap position.

In the limit, the SMC estimate approaches the PSS calculation since no additional information can be provided via measurements. For the variables that can be estimated by PSS reasonably well, the MSE does not increase anymore at the largest gap sizes. The second performance measure χ^2 also increases with larger gap size, but the slope is more sensitive to the actual dynamics of the dataset. If the PSS gives a reasonable estimate of a value but at the same time conflicts with another important process, the SMC estimate follows the PSS and predicts a very low standard deviation. The actual measurement can be multiple standard deviations away from the SMC estimate, thus a high value of χ^2 is reached. For NO and j_{NO_2} this is most likely the case due to an additional NO source from soil and an additional NO sink via Reaction (R9). Here χ^2 reaches high values before the 2 h mark. A plot can be found in the Supplement.

5.4.2 Precision enhancement

In this section, the SMC method is applied to artificially noised measurements to test the capability of reconstructing the original signal. The SMC method combines the prior knowledge given by the system dynamics and the precisely measured variables with the remaining information provided by the noisy measurement. If the prior overlaps with the likelihood, the result will be a more precise estimate of the noised variable. If the prior is far away from the measurement due to another process dominating the system, e. g., during the night, the posterior will be close to the likelihood. An example plot is shown in Figure 5.4. Here, normally distributed noise is applied to the NO₂ measurement. The expected different effects at daytime and nighttime are clearly visible. During the night, the SMC result follows the structure of the measurement while PSS outputs unrealistic values. The noise is reduced only by regularization of the variability through σ_0 . This effect will be discussed in more detail later. At daytime, the SMC estimate lies between measurement and PSS calculation with a strong tendency towards the photostationary state.

Algorithm 1 is applied again to the noised datasets to obtain values for MSE, χ^2 and also the baseline MSE. The latter value in fact equals the square of the noise added. Figure 5.5 shows the results of this experiment. In all cases, the MSE is constant for low noise. It is dominated by the intrinsic model bias. With increasing noise, the MSE starts to rise once the baseline MSE reaches the intrinsic model bias. At this point though, the MSE increases less steeply than the baseline MSE. Therefore, the additional information provided by the system dynamics successfully decreased the noise. At the same time, the value for χ^2 starts to increase. The SMC estimate becomes overly confident as the prediction according to the dynamics can no longer be falsified by measurement accuracy. The MSE and χ^2 start to saturate when the limit of extrapolation is reached.

A similar plot showing the resulting values of χ^2 is shown in Figure 3 in the Supplement. The value of χ^2 of the photolysis frequency decreases at the beginning, until the added noise gets close to the detection limit. Up to this point, the increased uncertainty during the night influences the uncertainty of the SMC estimate which decreases χ^2 . For O₃, the system starts to diverge at high uncertainty. Again, this

5 Data Quality Enhancement via Sequential Monte Carlo Filters

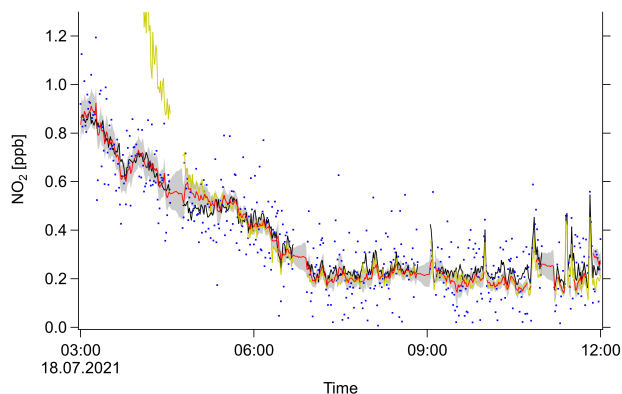


Figure 5.4: Example result of the SMC method used for de-noising. NO_2 mixing ratio as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial noise (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded gray region).

indicates that ozone cannot be completely reconstructed from the PSS calculation using only the considered molecules. Thus, the application of the method for de-noising is limited when other processes dominate.

5.4.3 Extrapolation

The state vector can also be appended with an unmeasured variable. If this variable is strongly coupled to measured variables through the system dynamics, the SMC calculation can give reasonable estimates. This problem can also be interpreted as the limit of infinitely large data gaps or measurements with infinite uncertainty.

Figure 5.6 shows an example plot of j_{NO_2} . The SMC result corresponds with the photostationary state calculation, as expected. Additionally, the SMC method is aware of the actual speed of the chemical reactions and is regularized via σ_0 with regards to the speed of unconsidered effects. In the example plot, the estimation shows only moderate agreement with the actual measurement. This discrepancy can be explained by other effects interfering with the system, e.g., NO emission from soil during the nighttime or other sinks of NO such as Reactions (R9) and (R10).

One has to be careful, however, if multiple unknown variables are coupled. In the case that a small variation in one can be compensated by variation of another variable, the system is singular and will most likely diverge to unrealistic values within a few iterations.

5.4.4 Free system parameters

The performance of the SMC method can change under variation of important free parameters. The most basic parameter is the measurement error $\sigma_{n,m}$ that relates to the detection limit DL and precision P via Eq. (5.16). This parameter governs how

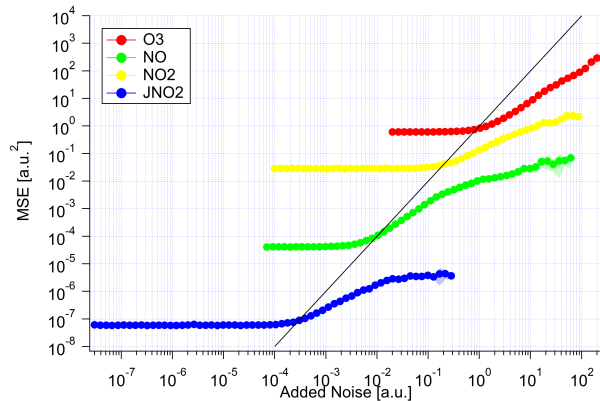


Figure 5.5: MSE of the SMC estimation as a function of artificial noise to study the precision enhancement abilities. The MSE units are represented by lines and markers and the standard deviation by the shaded region for the ensemble of repetitions. The plot shows the results of all variables: ozone (red), NO (green), NO₂ (yellow) and j_{NO_2} (blue). Note that the units of MSE and artificial noise are arbitrary to display all variables in one plot. The unit of the noise is ppbv for the trace gases and inverse seconds s⁻¹ for the photolysis frequency. The unit of MSE is the square, respectively. The solid black line indicates the baseline MSE (MSE = noise²)

far particles are allowed to spread from the measurement. An overestimation of this error will bias the algorithm output towards the prior estimate, an underestimation will bias the output towards the measurement. However, $\sigma_{n,m}$ can easily be chosen appropriately if the values of DL and P match the actual performance of the instrument during the measurement.

The switching probability p_η has to be tuned for a reasonable performance. If p_η is too high, activity is not dominated by inheritance but by random switching. The algorithm will output values similar to the system where activity is set to 1 but with slightly slower dynamics enabled. This also leads to unstable behavior. If p_η is too small, it is hard to switch from one state to another. Therefore, the algorithm will become unstable if the environmental conditions switch too fast. Examples are given in the Supplement for the interpolation of the photolysis frequency.

The standard deviation of the prior σ_0 has to be chosen carefully. It encodes the expected speed of variation of the system due to stochastic and unconsidered effects. This regularizes the resulting time series to low frequencies. If an appropriate value is chosen for this standard deviation, the algorithm already shows nice de-noising characteristics for each individual variable in \mathbf{x} , even if no system dynamic is considered at all. This effect has been reported, e.g., by Riris et al. [156] and Leleux et al. [157], for applications of a simple Kalman filter to mixing ratios of trace gases. Therefore, one might encounter seemingly good performance of this algorithm when in fact the result is just dominated by the expected speed of variations.

If the value chosen is too small, the system cannot reproduce rapid changes that do not originate from the chosen dynamic. Figure 5.7 shows an example plot for the

5 Data Quality Enhancement via Sequential Monte Carlo Filters

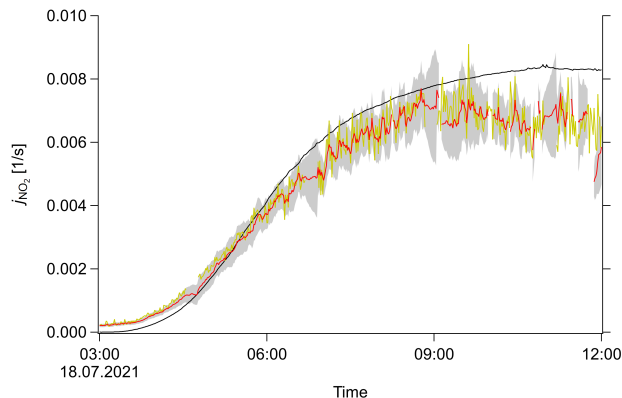


Figure 5.6: Example result of SMC method used for inference. Photolysis frequency j as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded gray region).

photolysis frequency. The system cannot catch up with the speed of sunrise and sunset and decouples from the measurement. If the value is chosen too high, the standard deviation is increased. This can lead to a flattened out probability distribution that effectively reduces the statistics of appropriate particles and therefore can also lead to unstable behavior.

Here, we propose the analysis of the entropy as a measure. If σ_0 is too small, the distribution is very condensed and all particles get similar weights. The entropy approaches a constant value. If the value is too high, the distribution spreads out and particles at the edge of the distribution get much lower weights than particles at the center. The entropy decreases with increased σ_0 . A proper choice of this parameter lies in the transition region from constant entropy to decreasing entropy. An example plot can be found in Figure 10.

Throughout this study, each σ_0 is calculated for each step from a constant and a linear contribution similar to Eq. (5.16), where $\sigma_{0,\text{const}}$ and $\sigma_{0,\text{rel}}$ are obtained from a linear fit of the measured difference between consecutive samples vs the measurement itself. This choice falls into the transition region mentioned before. Additional elaborations with regards to this parameter as well as the used σ_0 values can be found in the Supplement.

5.5 Conclusions

In this study, we demonstrate that the SMC method is a very versatile method that can effectively enhance data quality of atmospheric field measurements. We have shown satisfactory results when applied to data coverage increase, precision enhancement and inference of unmeasured variables. The algorithm is composed of simple steps and only introduces simplified chemical dynamics into a system of measurements. This way,

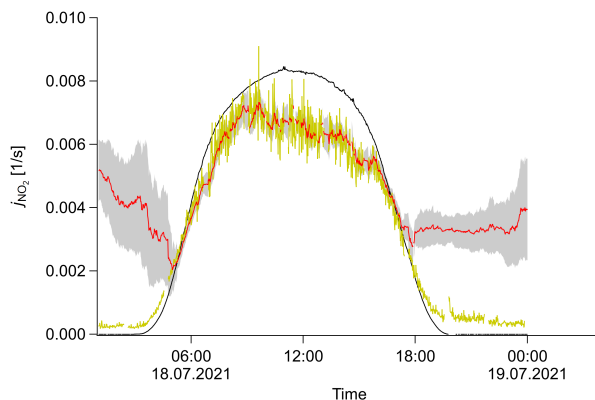


Figure 5.7: Variation of the free parameter σ_0 and its exemplary effect on the inference of j_{NO_2} . $\sigma_{0,\text{rel}}$ is decreased by a factor of 3. Photolysis frequency j as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded gray region).

the data quality can be enhanced without precise knowledge of complex reactions and processes such as emission, uptake, deposition or mixing with other air masses. The algorithm automatically detects deviations from the proposed simple dynamics by switching from the active state to the passive state. This ensures stability and gives quantitative insights about the underlying dominant processes. Furthermore, the entropy value encodes the information gained through the measurement and therefore the missing information in the prior estimate.

Along with several benefits over other approaches, we also explored the limitations of this method. Without the model extension by the activity variable η , the algorithm can produce unrealistic estimates when the system dynamics deviate from the proposed reactions. Variables that follow the proposed dynamics quite well and only differ slightly will lead to an underestimation of the standard deviation. Variables that do not follow the proposed dynamics at all do not benefit from the system dynamics but will be regularized with regards to the speed of possible variations. In this case, the algorithm is very sensitive to the proposed values of σ_0 . This free parameter has to be chosen very carefully. However, a proper value can be chosen by the analysis of observed variations and entropy.

The proposed method should not be seen as a replacement for PSS calculations, box-model calculations, model estimates or actual measurements, but it is an extension to the arsenal of numerical analysis for measurements in atmospheric chemistry. It provides many desirable properties as it is very simple, and returns salvageable higher moments of the estimated distribution while requiring a low runtime. A single run with the described setup and the whole 32 d dataset took 18 min of runtime on an 8-core desktop PC.

An open question is the stability of the algorithm when applied to a more complicated system with a higher dimension. Repeating the technical procedure of this study using a higher-dimensional system is restricted by data coverage and data quality in existing

5 Data Quality Enhancement via Sequential Monte Carlo Filters

datasets. We suggest that many applications of this method for different chemical systems are necessary in the future to fully rate the potential of the SMC method in the analysis of atmospheric chemistry field experimental data.

In general, we emphasize the versatility and high potential of this algorithm. Under the right circumstances, the SMC method can be utilized to enhance data quality and data coverage to allow for a more comprehensive data analysis of field campaign measurement data. However, we suggest conducting similar experiments when applied to a new system of variables. In particular, if the method is applied to a system of precise measurements along with a single imprecise, irregular or nonexistent measurement, the latter variable should be analyzed with regards to interpolation capability, precision enhancement ability and sensitivity to hyper parameters before conclusions can be drawn from the SMC result. These tests could be conducted on modeled data or on a different dataset where the same variables were measured.

6 Neural Network Architectures for Absorption Spectroscopy

An earlier version of this chapter was uploaded as "Neural Network Architectures for Absorption Spectroscopy" to the preprint server Research Square [158] under the terms of the Creative Commons CC BY license: <https://creativecommons.org/licenses/by/4.0/>.

How to cite: Lenard L. Röder. Neural Network Architectures for Absorption Spectroscopy, 20 April 2023, PREPRINT (Version 1) available at Research Square <https://doi.org/10.21203/rs.3.rs-2832856/v1>

This revised version is currently under Review for publication in Applied Physics B.

Abstract

State-of-the-art neural network architectures in image classification and natural language processing were applied to interference fringe reduction in absorption spectroscopy by interpreting the data structure accordingly. A model was designed for temporal interpolation of background spectra and a different model was created for gas concentration fitting. The networks were trained on experimental data provided by a wavelength modulation spectroscopy instrument and the best performing architectures were analyzed further to evaluate generalization performance, robustness and transferability. A BERT-styled fitter achieved the best performance on the validation set and reduced the mean squared error of fitted amplitude by 99.5%. However, analysis of the de-noising behavior showed large biases. A U-Net styled convolutional neural network reduced the mean squared error of the interpolation by 93.2%. Evaluation on a test set provided evidence that the combination of model interpolation and linear fitting was robust and the detection limit was improved by 52.4%.

Transferring the trained interpolator model to a different spectrometer setup showed no chaotic out-of-distribution effects. Additional fine-tuning further increased the performance.

Neural network architectures cannot be generally applied to all absorption spectroscopy tasks. However, given the right task and the data representation, robust performance increase is achievable.

6.1 Introduction

Artificial neural networks and deep learning have contributed to major breakthroughs in several applications like image classification, segmentation, generation of images and text, natural language processing (NLP) and many more. These new frameworks outperform conventional machine learning algorithms in many disciplines [52, 41, 44, 47, 55].

However, they cannot be trivially applied to regression tasks in natural sciences. In contrast to most conventional data analysis methods, hyperparameter optimization often needs complete retraining of the neural network and is therefore associated with high computational effort. Error estimation of the output is a research field on its own [159] and the model prediction can react chaotically to tiny deviations of the input [79]. Deep learning applications often differ from regression tasks with respect to their goal, constraints and requirements [41].

Nonetheless, several studies have successfully applied neural networks to spectroscopy tasks and reported performance increases compared to conventional approaches. Applications feature classification of spectroscopic data obtained from Raman spectroscopy and other spectroscopic techniques [69, 70], speeding up expensive calculations via surrogate models for nonlinear tomography [74, 75] and spectrum prediction [76], and concentration estimation [8, 160, 77]. Nicely et al. [8] used a shallow neural network for fringe reduction in direct absorption spectroscopy using simulated data. Tian et al. [77] report a good linearity of a direct fit performed by a neural network for high SNR input spectra. In our recent study we observed good performance of a neural network based noise reduction scheme for a specific noise structure where all tested conventional methods fell short [100].

This study will focus on absorption spectra obtained via wavelength modulation spectroscopy but the main considerations should be transferable to direct detection methods or other data acquisition schemes, as well. The main noise sources of wavelength modulation spectroscopy instruments can be split into relative and absolute contributions. The relative contributions cause disturbances proportional to the concentration of the measured species in the cell, e. g. variations in pressure, temperature, laser power and detector sensitivity. The main absolute limitation is often caused by notorious etalon fringe patterns that emerge from reflective surfaces of the optics [39, 7]. Other possible noise sources can be interference with absorption of other molecules, laser and detector stability or signal processing electronics. This study will, however, focus mainly on the reduction of noise resulting from interference fringe patterns. The procedures described may also be able to remove other noise types, as long as the requirements in subsection 6.2.3 are fulfilled.

6.1.1 Neural Networks Overview

The basic architecture of an artificial neural network (ANN) consists of iterated layers y_k of linear transformations implemented via matrix multiplications W_k and a

following non-linearity σ [41]:

$$y_{k+1} = \sigma(W_k y_k) \quad (6.1)$$

Theoretically this architecture can approximate any function using only a single intermediate (“hidden”) layer [161, 162, 163]. Advanced architectures try to optimize the ability of the model to learn and to generalize while being very efficient in time and memory.

A first breakthrough in image recognition tasks was achieved via convolutional neural networks (CNN) [164]. The matrix multiplications are replaced with convolutions of small filters whose weights are shared between all positions in the image [51]. This operation also ensures translation equivariance of the model output with regards to the input image [41]. Together with normalization schemes to overcome internal covariate shift [43], shortcut paths to decrease gradient decay [45] and randomization patterns to lower the chance of overfitting [42], models were constructed that outperformed human predictions for image classification tasks.

Vaswani et al. [54] introduced the transformer, a new architecture for NLP that builds on a mechanism called attention, where a query sequence gates the input sequence to focus on important parts. This method led to other similar architectures that have become state-of-the-art in NLP, such as the bidirectional encoder representation from transformers (BERT) [55]. Others adjusted the transformer-based architectures for image classification and achieved comparable results to state-of-the-art convolutional architectures [56, 57].

The field of deep learning is fast evolving and state-of-the-art candidates are constantly changing. Researchers seeking to apply deep learning to their discipline may not need to design whole neural network topologies themselves, but rather adjust already tested state-of-the-art networks to their need. Results may benefit from focusing on finding the appropriate already available state-of-the-art model. A drawback of this approach, though, is the vast size of state-of-the-art models.

6.1.2 Opportunities for Neural Networks

These fringe patterns can interfere with the frequency region of the signal. In our previous study we reported different behavior and performance of several noise reduction methods depending on this interference. Many conventional methods are based on frequency separation of signal and background and therefore only show great performance if this interference is small. Otherwise the method will produce a high bias. We reported, however, an improved performance of a neural network based approach in a region of high interference [100].

This behavior can be motivated in an example: Suppose an absorption spectrum is obstructed by a pure sinusoidal background modulation and some white noise. When fitted, the resulting spectrum will show high variations depending on the phase of the sine background. Conventional numerical noise reduction schemes that act on the frequency domain will not be able to reconstruct the original signal but will most likely dampen the result towards zero. However, if the structure of the background is known

a priori, noise reduction is very efficient. In this example the structure of the background can be extracted from the edge of the spectrum by a curve fit and the obtained sine can be subtracted from the noised spectrum. Of course, such a simple example of background can simply be removed by applying established experimental techniques, e.g., lock-in amplification. However, the background structure in a real experiment is often more complex. It cannot be easily removed via experimental techniques and the fitting parameters cannot be easily fit. A machine learning algorithm, though, could learn the distribution of background structures and infer the interference with the signal. This approach requires the background structure to have a sparse distribution and be stable over the whole absorption spectrum.

Another conventional approach to reduce the impact of etalon fringe patterns is regular determination of background spectra and an interpolation scheme [38, 9]. If the background structure changes point-wise and slow enough, this approach will give precise estimates of the underlying background structure of a measured spectrum. The interpolated spectrum can be subtracted from the measurement and will yield great noise reduction. In a second example a problem of this approach will be discussed: Assume the background structure does not change point-wise but along the frequency axis. If e.g. the phase changes in the order of half a cycle between two background measurements, the result of the interpolation will vastly differ from the true background structure. Given *a priori* knowledge about the speed of the phase shift, the background can be reconstructed again with high precision. In practice obtaining this *a priori* knowledge can be very hard as a realistic background has a much more complex structure and could consist of several phase changing periodic structures or beat interferences. Also the true shape of the fringe pattern could be obscured by aliasing effects [33]. Again a machine learning algorithm could learn the distribution of these structures and yield a better interpolation scheme to reconstruct the background.

In section 6.2 possible interpretations of the data in order to motivate appropriate model architectures are discussed. In section 6.3 the spectrometer is described in more detail and the training data acquisition and processing is presented. In section 6.4 the performances are discussed for the training process, validation set evaluation, test set evaluation and further transfer applications.

6.2 Neural Network Architectures

As motivated in the introduction, two different machine learning objectives will be considered: concentration fitting (6.2) and temporal interpolation (6.3), with spectrum size N , number of spectra for each measurement C and sequence interval T . The fitter F directly fits an absorption spectrum to obtain the trace gas concentration. The interpolator I gets a temporal sequence that contains regular background measurements and reconstructs the evolution of the background in the inbetween region.

$$F : Z \longrightarrow A \quad Z \in R^{N \times C} \quad A \in R^1 \quad (6.2)$$

$$I : X \longrightarrow Y \quad X \in R^{T \times N \times C} \quad Y \in R^{T \times N \times C} \quad (6.3)$$

6.2.1 Possible Fitting Architectures

Although the input to the fitting task is one-dimensional (disregarding the feature axis) increasing the dimension to two by a redundant linear transformation might be beneficial. This enables the application of models that have proven themselves on popular image classification benchmarks. In this study the continuous wavelet transform (CWT) with the DOG2 wavelet is used to scale up the spectra to $R^{N \times J \times C}$ with $J = 128$. The continuous wavelet transform has been very successful in environmental sciences [10, 103], but similar behavior can be expected for other linear transformations. An example plot of a CWT-transformed spectrum can be found in Figure 6.1.

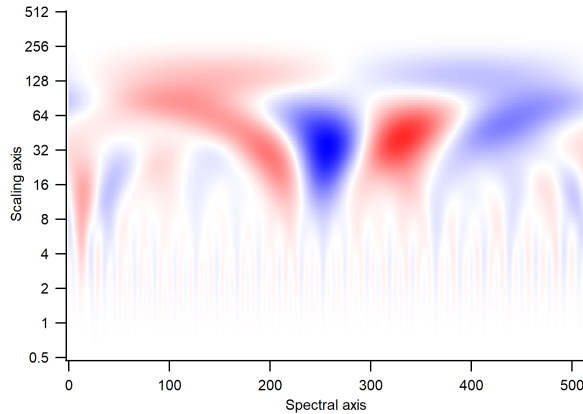


Figure 6.1: Example plot of a CWT-transformed spectrum, taken from the validation set. Values smaller than zero are colored red, values larger than zero are colored blue. The plot suggests how to interpret the data as image-like. The anisotropy of the image is visible, with high frequency noise at the lower end of the scaling axis, signal at the higher end of the scaling axis and left-right-centered along the spectral axis, and obstructing fringes in the same scaling range but isotropic along the spectral axis. The SNR of the underlying spectrum is 3.2.

Now, the data can be interpreted in different ways: a $N \times J$ image of the whole wavelet spectrum or an N -dimensional sequence of J -dimensional vectors.

If data is considered as image-like, a CNN architecture can be used. A novel family of efficient and effective CNNs is EfficientNetV2 [48], from which the smallest iteration EfficientNetV2B0 (EffV2) will be used for performance reasons. A CNN architecture iteratively extracts features of the image y_k and scales down using convolutional (Conv) blocks that contain pooling operations and strided convolutions.

$$y_{k+1} = \text{Conv}(y_k) \quad (6.4)$$

6 Neural Network Architectures for Absorption Spectroscopy

In classification tasks the final layer is globally reduced and fed through a shallow fully connected neural network (FCNN). The result is then softmaxed to retrieve the probability for each class. In this case, though, the output needs to represent the concentration of the trace gas, therefore the final output is not passed through an activation layer. This way the output can be any real number.

However, as mentioned earlier, a CNN is translation invariant with respect to its input - in image processing the algorithm should not depend on the global position of the object. In the data described there is an anisotropy: The signal is left-right-centered and occupies a distinct part of the wavelet scaling, while noise is isotropic and can appear differently throughout the scaling dimension. The vision transformer (ViT) [57] decomposes a picture of shape $R^{N \times J \times C}$ into patches of shape $R^{P \times P}$ and flattens the pixels to obtain data of shape $R^{N/P \cdot J/P \times P^2 \cdot C}$. These patches are embedded via a linear transformation. The patch embedding is then added to a learned positional encoding that represents the position of the patches with regards to the complete image. Afterwards the architecture is very similar to BERT [55] and contains of iterated transformer (Trans) blocks containing attention layers:

$$y_{k+1} = \text{Trans}(y_k) \quad (6.5)$$

With the attention mechanism, the model can learn global dependencies and structure of the data due to a higher receptive field [56]. The final layer of the transformer can again be globally reduced and fed into the FCNN.

Additionally, a hybrid architecture will be considered composed of an EffV2 backbone consisting of the first 5 blocks of the model, whose output is linearly transformed and directly fed to a ViT as patch embeddings. Dosovitskiy et al. reported similar performance of both approaches and found a similarity in function of the first transformer layers to the CNN backbone. The backbone will extract local features of the image that can be processed globally by the transformer [57].

If the data is considered as sequence-like, the transformer architecture can be applied directly. In this case the data is simply reshaped to $R^{N \times (J \cdot C)}$ and fed into a BERT-styled architecture [55]. Table 6.1 summarizes these approaches and gives a small overview of the chosen size.

Task	Model	Data Interpretation	Mechanism	Parameters
fitter	EffV2 [48]	image	Conv	6.3 M
	ViT [57]	image	Trans	10.2 M
	Hybrid [57]	image	Conv/Trans	5.2 M
	BERT [55]	sequence	Trans	10.1 M
interpolator	U-Net [44, 48]	image	Conv	11.6 M
	ViT [57]	image	Trans	9.9 M
	BERT [55]	sequence	Trans	10.1 M

Table 6.1: Summary of utilized model architectures and number of parameters.

6.2.2 Possible Interpolation Architectures

The interpolator also gets 2D inputs: temporal sequences of spectra. The spectra at the beginning and the end and also at regular distances throughout the sequence are considered pure background and are point-wise interpolated to fill the intermediate temporal region. Also the spectral edges of this intermediate region is kept as no absorption signal is to be expected here. An example can be found in Figure 6.2. Again the data can be interpreted image-like or sequence-like but this time with a different objective.

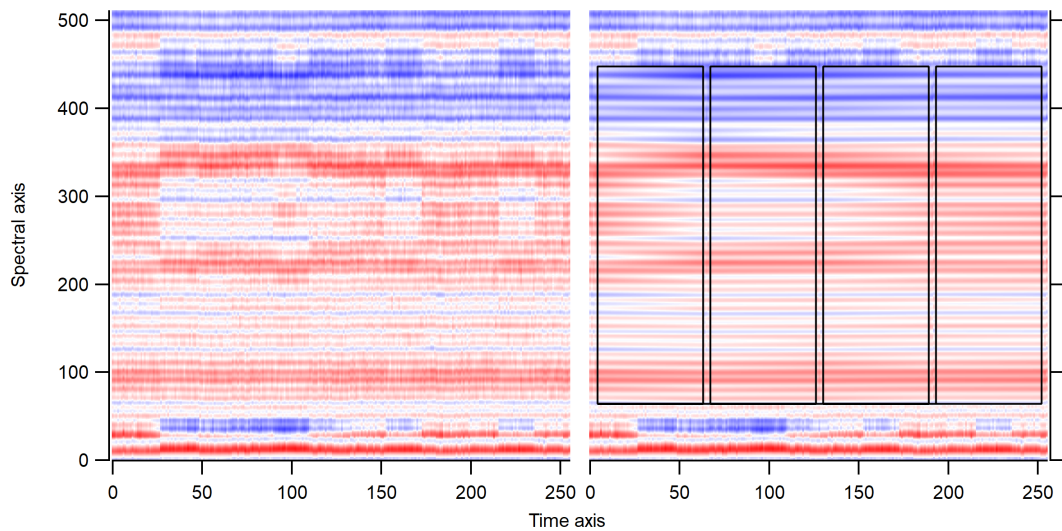


Figure 6.2: Example plot of a temporal background spectrum sequence (left) and estimation via linear interpolation along the time axis (right). The black boxes mark interpolated regions. The temporal interpolation anchors and spectral edges are left unchanged. Values smaller than zero are colored red, values larger than zero are colored blue. The plot suggests how to interpret the task similar to image reconstruction. The objective is to reconstruct the original sequence (left) from the linear interpolation (right).

If data is considered as image-like, the U-Net variant can be used, that has been utilized for segmentation tasks [44] or image noise reduction [165]. The down-scaling part of the U-Net is constructed from a subset of the EffV2 and the inverse of the initial EffV2. In the inverse part every downscaling operation is replaced by an upscaling operation to obtain outputs with the original dimension.

Another approach would be again a ViT styled segmentation followed by a transformer architecture. The final patches can be linearly transformed and concatenated back to the original shape of the picture. A hybrid architecture is not considered here since the U-Net architecture already requires a lot of memory.

If data is again considered sequence-like, a linear transformation can be applied to each $N \times C$ vector. In contrast to the CWT a learnable linear transformation is used here since the dimension does not need to be increased in this case. This linear embedding can then be input to a BERT styled model. The output is again fed

to a linear transformation and gets reshaped to match the desired output. In their original paper the authors introduced a masked language model (MLM) where words were randomly replaced by a *missing token* and their model was pre-trained in an unsupervised fashion to learn natural language structure [55]. In the example given here most of the data is missing, but the target output can be slowly transformed to the desired input by increasing the number of interpolated spectra in the intermediate region to speed up the initial learning period. This procedure is also applied to the image-like representations.

6.2.3 Remarks

In summary several ways to interpret the data have been considered and appropriate established neural network architectures have been chosen for each interpretation. It is important to emphasize here that different performances of these models do not indicate advantages of one model architecture or data interpretation over the other. No proofs or evidence of a specific model type to be favored can be given, as these are only single random examples. The performance depends very much on the choice of hyper-parameters like optimizer setup, learning rate value and schedule or number of parameters. Taking into account several ways to understand data may lead to good performance without sensitive variations of hyper-parameters and emerging biases as a consequence.

All procedures motivated in this section and in the introduction require several characteristics of the underlying background structure for the neural network approach to work accordingly. All of these requirements are fulfilled for the specific spectrometer setup used in this study with regards to interference fringe patterns as a main noise source. The procedures might show similar performance on other noise sources, if these requirements are also met there:

- **Sparse noise distribution.** All possible noise shapes must follow a sparse probability distribution compared to independent white noise. Otherwise no prior information about noise structure can be extracted from background measurements.
- **Local stability.** The noise structure needs to be locally stable so reconstruction is possible from the absorption-free parts of the data.
- **Global stability.** The noise structure needs to be stable over time, otherwise the prior information needed for reconstruction changes and the network needs to be retrained. If the structure changes too fast, retraining is needed before desirable performance can be achieved.

6.3 Experimentation and Network Training

The experiment conducted for this study is based on the instrument TRISTAR reported in [11, 38]. It is driven by a room temperature quantum cascade laser [4] from

6.3 Experimentation and Network Training

Alpes Lasers operated near the formaldehyde (HCHO) transition at 1759.72 cm^{-1} [166]. The laser beam is guided into a 50 cm long White Cell [2] where it gets reflected 128 times to yield an effective path length of 64 m. The beam exits the cell and is split by a beam splitter into two separate paths. One beam is guided through a reference cell filled with the substance of interest at high concentration. Both beams are focused on identical infrared detectors from VIGO Systems. The laser is modulated by a slow triangle wave that scans through the absorption spectrum and a fast sine wave that is demodulated at twice frequency by the data acquisition FGPA. The resulting spectrum is similar to the second derivative of the absorption profile, depending on the modulation depth. In the experiment each data point consists of $C = 2$ spectra, one from the increasing part of the triangle wave and one from the decreasing part, with $N = 512$ points each.

Experimental data of the instrument is gathered for 14 days. The instrument inlet is connected to an air purifier to remove the substance of interest, in this case HCHO. The reference cell is filled with a high concentration of HCHO. The detector channel without reference cell detects absorption-free spectra that only consist of the interference fringe background structure. Spectra acquired at this detector will be called BGD. The detector channel with inserted reference cell can be assumed fringe-free due to the high concentration signal. Spectra acquired at this detector will be called REF. Only the last 82% of this data is used for training.

Training data for the interpolator is created using sequences of BGD. A sequence X is input to a network that gets 4 seconds of spectra every 63 seconds and does point-wise linear interpolation in the intermediate region. The interpolation anchors and 64 points at both edges of the spectrum are left unchanged. Then this array of shape $R^{T \times N \times C}$ is normalized for each subarray in C while mean and standard deviation are stored. This prepared matrix is fed through the interpolator and the output is rescaled by the standard deviation and translated according to the mean that was stored before. The loss is calculated as mean-squared-error (MSE) loss between this final matrix Y and the input sequence X .

Training data for the fitter is created using random pairs of BGD and REF. Two log-uniform distributed values A , S are drawn for the concentration and the signal-noise-ratio (SNR). The input to the network Z is a linear combination of BGD and REF with

$$Z = A \text{ REF} + \frac{A}{S} \text{ BGD} \quad (6.6)$$

First this input is normalized and the standard deviation is stored. Then the CWT is performed that transforms to a dimension of $R^{N \times J \times C}$. The CWT result is fed through the fitter to obtain a single value. This value is rescaled by the standard deviation stored. The loss is calculated as MSE loss between this rescaled output value and the value A . For additional investigation two instances of fitters are trained for different SNR ranges.

For validation, sequences of BGD and pairs of BGD and REF along with pre-determined random values are taken from the first 9-18% of the measurement data. The best iteration of each interpolator or fitter, respectively, is then applied to the test

set. The test set consists of a sequence of BGD and injections of calibration gas into the measurement cell from the first 9% of the measurement time. Variation and point-wise accuracy of background and calibration signal can then be determined and compared to a conventional approach.

Each network is trained on a HPC cluster hosting several NVIDIA V100 GPUs. The training process is parallelized through the distributed learning scheme Horovod [167].

6.4 Results and Discussion

In this section the performance of the described model architectures will be discussed.

6.4.1 Training Performance

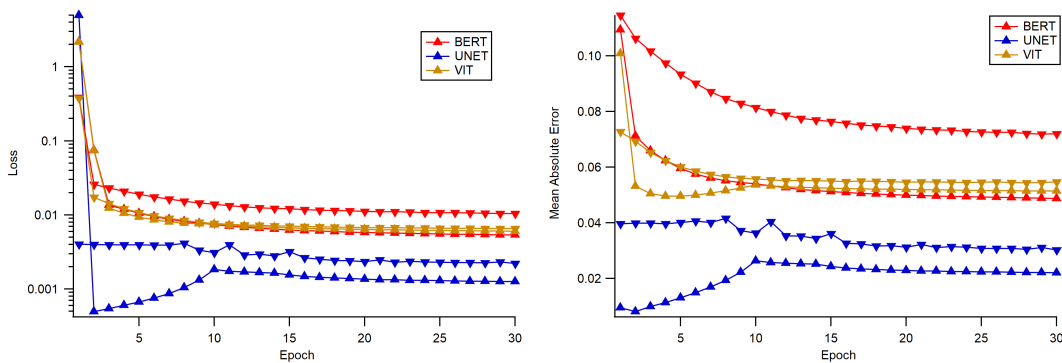


Figure 6.3: Training metric (upward triangle) and validation metric (downward triangle) of interpolator models based on BERT (red), VIT (yellow) and UNET (blue) architecture: Left: Loss (mean squared error); Right: MAE (mean absolute error)

As already mentioned in Section 6.2, the interpolator was trained on an easier task during the first 10 epochs for a more efficient training process. A random portion of spectra in the input sequence were exchanged by the corresponding spectra in the target sequence. This procedure reinterprets the MLM from BERT or an autoencoder application to enable faster learning of the underlying distribution. The amount of exchanged spectra was linearly decreased towards zero in epoch 10. All model architectures show decreasing training loss and error metrics while validation metrics do not show major indications for overfitting. Figure 6.3 shows losses and mean absolute error of the interpolator models. Additional plots can be found in the appendix. For further investigation of the robustness, several ablation studies [57] were performed where the number of parameters of each model was reduced by scaling down the feature axes. The scaled down models achieved faster convergence speeds while resulting in similar final metrics. Example plots can also be found in the appendix.

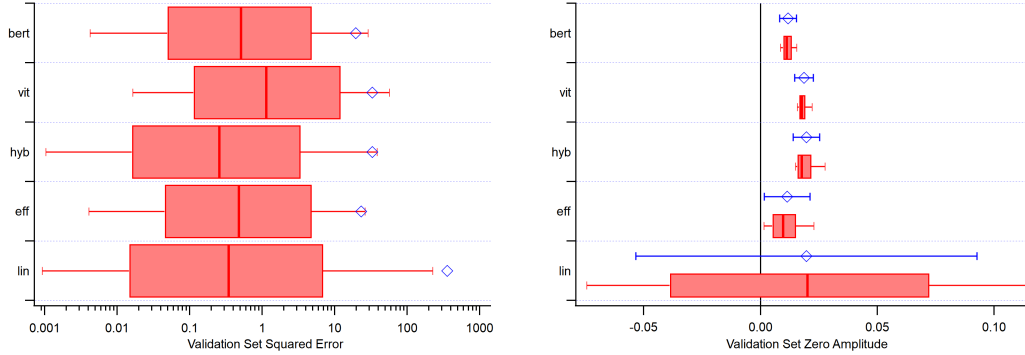


Figure 6.4: Evaluation of fitter models trained on low SNR range and comparison to linear fit. Red boxes indicate 25%-quantile, median and 75%-quantile; red whiskers indicate 10% and 90% quantile. Blue diamonds show the mean, blue whiskers (if provided) show one standard deviation. Left: absolute squared error between prediction and true value. Maxima are in the order 10^4 . Right: predicted amplitude of pure background spectra.

6.4.2 Validation Set Performance

Although different performances during training might suggest the most suitable model choice for each application, these metrics only provide guidance during the training process. Model performance should always be derived from experiments that more closely reassemble real applications. Thus, the models are evaluated in more detail using the validation sets.

The fitter model is evaluated by calculating the absolute squared error between model prediction and true signal amplitude of each spectrum in the validation set. For the investigation of small signal behavior the pure background spectra are also input to the model to obtain predicted zero amplitudes.

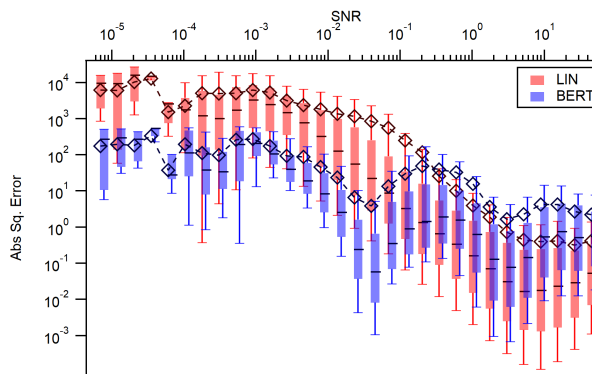


Figure 6.5: Fitter squared error as a function of input SNR. Linear fit in red, low SNR trained BERT variant prediction in blue. Boxes indicate 25%-quantile, median and 75%-quantile; whiskers indicate 10% and 90% quantile. Diamonds and lines show the mean. For better visualization the BERT plot is slightly offset along the x -axis.

Figure 6.4 shows an overview of the evaluation results for the fitters trained on a low SNR range. The BERT-styled variant shows the least combination of bias and variance for the zero amplitude while performing similar to the EffV2 and Hybrid variant overall. The mean squared error over the validation set is reduced by 99.5% and the mean squared error on pure background spectra by 97.4%. The absolute squared error can be further investigated depending on the input SNR. This is depicted in Figure 6.5. The BERT variant outperforms linear fitting in the low SNR regime by 1-2 orders of magnitude but falls short in the high SNR range.

The fitters trained on a high SNR range fail to outperform linear fitting and perform poorly for SNR values below their training data. A large bias is introduced when fitting pure background spectra due to the lack of low signal examples in their training distribution. Detailed plots similar to the ones shown in the low SNR example can be found in the appendix.

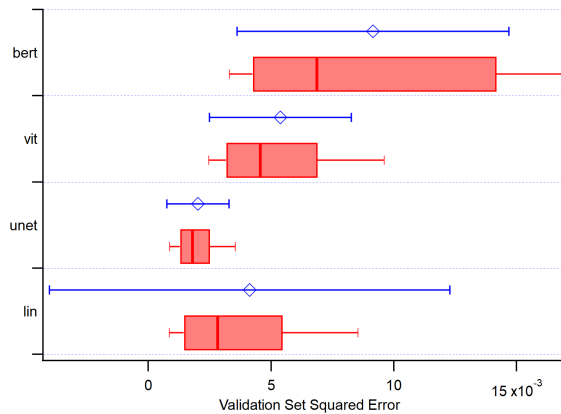


Figure 6.6: Point-wise squared error between prediction and true spectra of interpolator models and comparison to linear interpolation. Red boxes indicate 25%-quantile, median and 75%-quantile; red whiskers indicate 10% and 90% quantile. Blue diamonds show the mean, blue whiskers show one standard deviation.

The interpolator models are evaluated by calculating point-wise squared errors between predicted background spectra and measured background. The evaluation result is shown in Figure 6.6. Here the U-Net variant clearly shows the best performance. The mean squared error is reduced by 93.2%. This does not necessarily indicate an advantage of this exact model architecture compared to a transformer type, since further hyper-parameter optimizations can result in increased training efficiency and model performance. The behavior of the U-Net variant is analyzed in more detail by calculating the point-wise squared errors in dependence of the distance of the spectrum from the nearest interpolation anchor. This relation can be found in Figure 6.7. In the case of a stable, slowly changing background, the linear interpolation would give best estimations near the anchors, which is the case in this dataset. The model prediction shows no dependency on the distance to the nearest anchor and reduces the error evenly over the sequence.

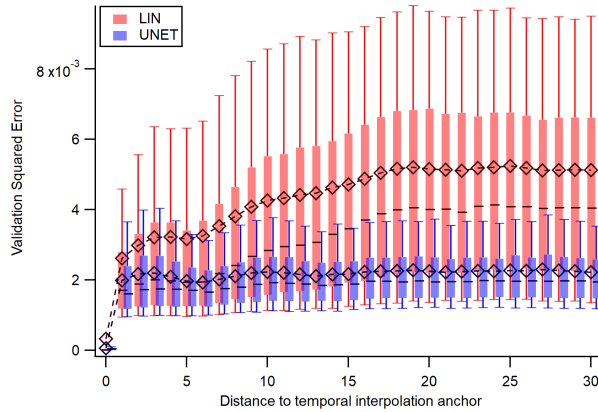


Figure 6.7: Point-wise squared error as a function of distance to nearest interpolation anchor. Linear interpolation in red, U-Net variant prediction in blue. Boxes indicate 25%-quantile, median and 75%-quantile; whiskers indicate 10% and 90% quantile. Diamonds and lines show the mean. For better visualization the U-Net plot is slightly offset along the x -axis.

6.4.3 Denoising Behavior

Despite the extraordinary mean squared error reduction from application of the fitter model, the denoising behavior shows undesirable properties that introduces disadvantages compared to classical linear fitting. In the low SNR limit the variance of the linear method is very large compared to the signal amplitude. It is shown in Figure 6.8 that the fitter model introduces a small bias while substantially reducing the variance. This leads to an effective reduction of the MSE. However, the bias is comparable to the signal amplitude which renders the method almost useless in applications.

A possible explanation of this behavior is a poorly chosen loss function. While the loss is reduced, the actual objective is not reached. Alternative explanations are non-optimal model and optimizing schemes or an impossible task. For the investigation of the loss function, a variation of the experiment is conducted where instead of estimating the absolute concentration, the final layer is activated using the sigmoid function and the model is trained to distinguish pure background spectra from absorption spectra with an SNR of 0.01. Figure 6.9 shows the training performance of this model. The accuracy of the validation set shows no significant performance increase compared to random guessing. Thus, either the combination of model architecture and optimizing scheme is not optimal or the distinction between background and low SNR signal is not possible. The latter is the case if the data does not match the requirements given in Sect. 6.2.3. More specifically, the noise structure might not be locally stable or not sparse enough for the fitter task.

The interpolator shows desirable properties in the denoising behavior. In Figure 6.10 the deviation from the true value of both approaches is depicted. The example from Section 6.2 is used. While coarse structures are visible in the difference between linear interpolation and original spectra, the residual structure of the model interpolation is

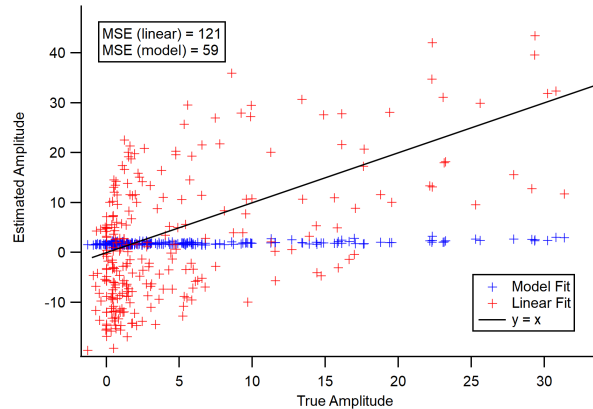


Figure 6.8: Denoising behavior of the fitter model. Estimated amplitude against true value for the linear fit (red) and the low-SNR fitter model output using the BERT-styled variant (blue) for a low SNR example batch with 256 spectra. The black solid line refers to a perfect fit $y = x$. The MSE values for this example batch are 121 and 59 for the linear fit and the fitter model, respectively. Despite the improvement in MSE, the fitter model produces a large bias which makes the estimation less meaningful.

much less correlated. This indicates effective learning of the underlying distribution and successful reconstruction.

For the rest of this chapter, only the interpolator will be considered.

6.4.4 Test Set Performance

The test set is evaluated using experimental performance metrics instead of model loss. This allows further quantification of the performance and generalization ability of each method. In absorption spectroscopy experiments, key parameters that limit the instrument performance are the reproducibility of calibrations due to temporal drifts, the relative precision, and the detection limit. The reproducibility of calibrations is retrieved by first averaging individual calibration intervals and calculating the standard deviation of these averages. It is a measure for long-term drifts and stability. The precision is the relative standard deviation of all calibration amplitudes after correction of the long-term drifts. The detection limit is two times the standard deviation of background amplitudes. Since only absorption-free gas was measured during the experiment, the detection limit can be evaluated using spectra marked as background as well as spectra marked as ambient.

Figure 6.11 gives an overview of the resulting experimental performance metrics. Similar to the validation set results the model-based approach achieve a better detection limit than the linear approach. Applying the interpolator model and a linear fit achieves a very robust reduction of detection limit as both the absolute and the relative errors are reduced.

Example time series of background and calibration amplitudes are shown in Figure 6.12. Again, very robust behavior can be observed for the combination of inter-

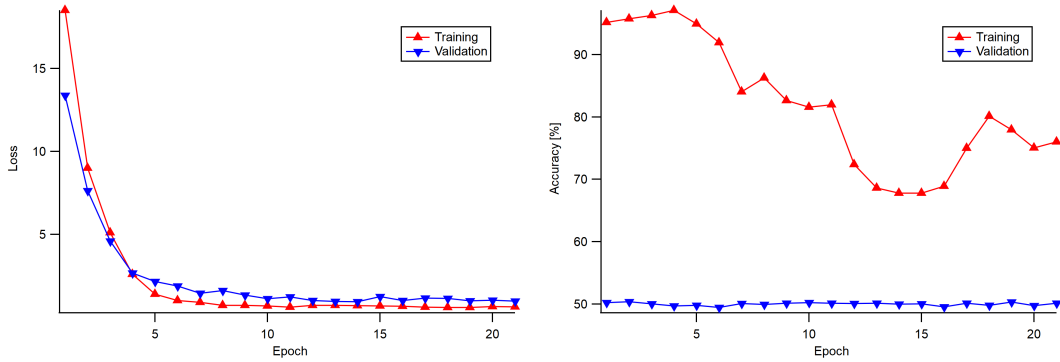


Figure 6.9: Training metric (red upward triangle) and validation metric (blue downward triangle) of classifier model based on EffNet architecture. Left: Loss (binary cross-entropy); Right: accuracy

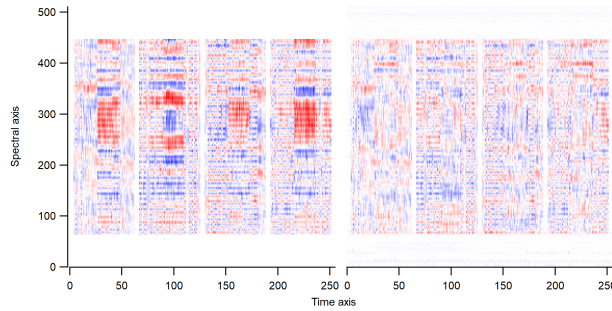


Figure 6.10: Denoising behavior of the interpolator model. Difference between the example of the temporal background spectrum sequence shown in Figure 6.2 and the linear interpolation along the time axis (left) and the interpolator model output using the U-Net variant (right).

polator model and linear fit. This indicates evidence that the proposed interpolator architecture provides robust noise reduction.

6.4.5 Transferability

An important question left unanswered is the transferability of the networks. What kind of output is created if a different spectrometer is used? Is a similar performance increase possible? Will the result contain chaotic artifacts due to out-of-distribution problems?

To answer these questions, the trained model is applied to a different QCL absorption spectrometer which utilizes a 20 cm long Herriott Cell configuration with 182 passes [3, 28] and is operated at the carbon monoxide (CO) transition at 2190.02 cm^{-1} [99]. Data acquisition and detectors are similar, but the experiment is driven at a different modulation and sweeping frequency.

Evaluation data for the interpolator was gathered by flooding the absorption cell with

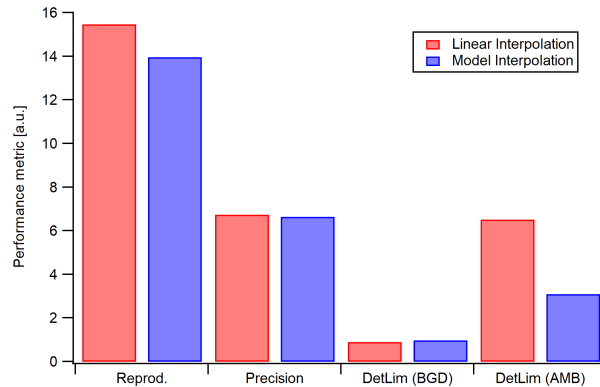


Figure 6.11: Experimental performance metrics for the different interpolation techniques. Reprod: reproducibility of calibrations; DetLim (BGD): detection limit (2σ) evaluated at spectra marked background; DetLim (AMB): detection limit (2σ) evaluated at spectra marked ambient.

nitrogen. Zero gas was measured for 70 hours. Application of the pre-trained version directly resulted in similar performance than linear interpolation. This indicates that the transferred model may require additional fine-tuning to achieve a good performance, while it does not result in chaotic artifacts. For fine-tuning, the first 7 hours of nitrogen measurement were taken as new training and validation data and the pre-trained interpolator model was trained further for 10 epochs, which results in 0.8% of fine-tuning iterations compared to pre-training iterations. A comparison of the evaluation is shown in Figure 6.13.

6.5 Summary and Conclusion

In this study several possible applications of neural networks to absorption spectroscopy experiments were tested by interpreting the data structure in a way that several state-of-the-art neural network architectures can achieve good performance. These architectures were chosen from image classification and natural language processing tasks. A model for interpolation of background spectra and a model for gas concentration fitting of absorption spectra were created. Each neural network was trained on measured data and showed good generalization performance. The best performing instance of each task was further evaluated using test data and data from a different type of absorption spectrometer.

Fitters trained on a high SNR range did not outperform linear fitting. The best performing fitter trained on a low SNR range was of the BERT-type. It decreased the mean squared error on the validation set by 99.5% and the mean squared error on pure background spectra by 97.4%. However, undesirable denoising behavior was observed that rendered the method unusable. Training a classifier with the same architecture showed that this behavior was not caused by a poor choice of loss function but is caused

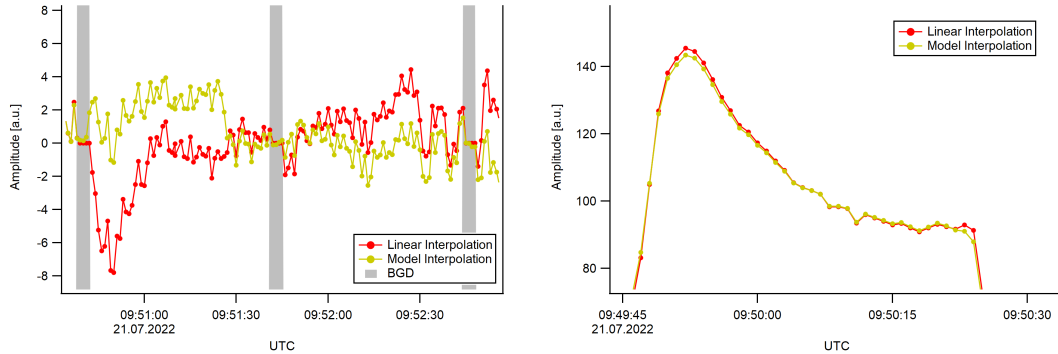


Figure 6.12: Examples of amplitude results for different interpolation techniques. Top: Example result during zero gas measurement. Grey regions indicate declaration of the spectra as background. Bottom: Example result during calibration gas injection.

either by insufficient architecture and optimizing scheme or an impossible objective. Considering the human-level performance in image recognition and natural language processing tasks of the chosen architectures, the objective might not be possible due to the strong interference of signal and background. However, due to the large search space and choice of hyper-parameters this can only be speculated.

The best performing interpolator was of the U-Net type and reduced the mean squared error of the validation set by 93.2%. It showed less dependence on the distance from the nearest interpolation anchor than the linear interpolation. The combination of model interpolation and linear fitting showed very robust behavior and decreased the relative error by 8.2% and the detection limit by 52.4% on the test set.

It was shown that the interpolator model can be transferred to a different spectrometer without chaotic out-of-distribution effects. However, the performance of the pre-trained model on the different setup does not match the performance on the original spectrometer setup and may become worse than conventional approaches. The performance can be enhanced via fine-tuning on new data. Using just 0.8% fine-tuning iterations in relation to initial training iterations, the interpolator mean squared error was reduced by 36.3% compared to the conventional approach.

In conclusion, using state-of-the-art architectures is no guarantee to obtain a well performing neural network if the task is not appropriate. But, interpreting the task in multiple ways to include many state-of-the-art architectures can make the application less sensitive to specific properties of a chosen network and speed up the architecture design significantly.

In this study only $2f$ -wavelength modulation spectroscopy was considered, but the concept should also work for other absorption spectroscopy data acquisition schemes due to the similarities of the spectral features and the dominant noise sources. More fine-tuning may be required in this case.

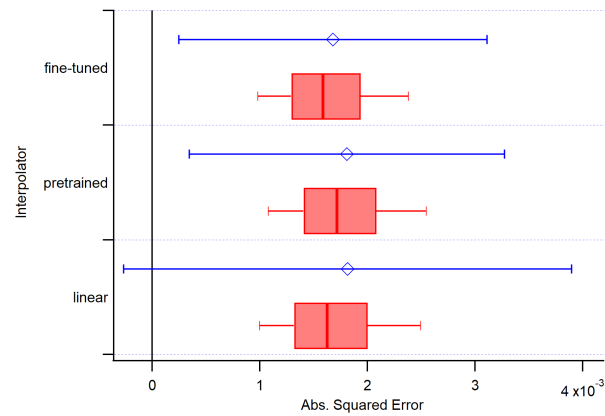


Figure 6.13: Performance comparison of linear approaches to interpolator model transferred to the alternative spectrometer. Point-wise squared error between background spectra and interpolation reconstruction with linear interpolation, pre-trained variant and fine-tuned variant. Red boxes indicate 25%-quantile, median and 75%-quantile; red whiskers indicate 10% and 90% quantile. Blue diamonds show the mean, blue whiskers show one standard deviation.

7 Discussion and Outlook

This PhD study explored a wide range of data analysis and signal processing methods for different possible applications in atmospheric measurements. The analysis methods featured time-frequency analysis, Bayesian statistics and machine learning. The investigated applications featured time series processing of an atmospheric dataset, data quality estimation of a measurement instrument and absorption spectrum processing of infrared absorption spectroscopy. This selection of various data processing tasks and analysis methods spans several scientific disciplines, but many global properties of the evaluation scheme and the implications for atmospheric measurements were observed.

Before applying advanced signal processing techniques to a measurement instrument, the characteristics limiting the data precision and accuracy have to be determined. In the first study (Chapter 3), studying the data analysis with Allan-Werle-plots revealed quantitative properties of the dominant noise structures present in the time series and their effects on the drift correction capabilities and measurement uncertainty estimation. Limitations of the Allan-Werle-plot became obvious when the data set was not segmented into individual sections that matched the changes in environmental setup. Applying Allan-Werle-plots directly to dynamically influenced time series decreases the interpretability of the analysis and cannot resolve the dominant noise structures. Elaboration of these noise structures can determine the limiting properties of the measurement device.

The TRISTAR instrument is limited by interference fringe patterns that contribute sinusoidal variations of the time series and cannot be completely reduced via interpolation of calibration measurements. Possible improvements in data quality may be achieved by either applying advanced signal processing to the absorption spectra before calculating the mixing ratios, or by post-processing of the time series to reduce the deviation from the true value.

In the second study (Chapter 4), noise reduction methods that act on the absorption spectrum were tested. Despite the numerous publications that report performance increases from such advanced signal processing methods, the actual performance strongly depends on the properties of the signal and background spectrum. Studying the effects of various noise reduction methods for different background spectrum characteristics revealed confined regions of applicability that showed little overlap. Only considering the reduction of MSE lead to wrong assumptions of applicability if a large bias is introduced in return. The properties of a noise reduction method have to be evaluated on the target background structure before application is possible. Constraints with respect to the bias of the method have to be considered for the target application.

7 Discussion and Outlook

The dominant free spectral range of the fringes present in the TRISTAR device obstruct the frequency range of the absorption spectrum. Thus, noise reduction via classical approaches such as statistical or time-frequency analysis are not possible. Only the application of neural networks could provide a performance increase by inducing learned prior information about the noise structure to the signal processing. However, this method is expected to introduce a bias into the processed data.

In the third study (Chapter 5), post-processing of time series of atmospheric measurements was tested. The statistical and numerical method called Sequential Monte Carlo filter was explored as a new method for time series processing. In a Bayesian way the information provided by the chemical reactions of a system of variables can be used to form a prior estimation of the state. The measurements with their corresponding uncertainties provide the likelihoods to refine this prior and obtain a corrected time series that considers the chemical reactions and weights the different uncertainties accordingly. Experiments for interpolation, noise reduction and extrapolation were conducted that each showed reasonable performances, but limitations were also observable. A former iteration of the algorithm resulted in divergence of some variables when the proposed dynamics do not match the dominant chemical processes. This issue was mitigated via the introduction of an activity parameter. The choice of free parameters provides a challenge to the method since some parameters have a strong impact on the results.

Application to time series obtained via measurements of the TRISTAR device brings some difficulties. The study results indicate that similar experiments are necessary to fully analyze sets of other atmospheric variables. The set of measurable species does not contain the variables analyzed in the study. Dynamical models containing HCHO feature a lot of reaction processes and will increase the complexity of the system. Application to airborne measurements is most likely not possible since fast exchange of the sampled air masses break the Markov condition that each data point is connected to the previous one. Hence, this method needs further testing before it can be useful to increase the data quality of the TRISTAR device.

In the fourth study (Chapter 6) the proposed idea to use neural networks for noise reduction in absorption spectra was tested in depth. Novel neural network architectures, e.g., convolutional neural networks and transformers, were considered that have shown promising results in image recognition and natural language processing tasks. Using neural networks as an alternative to linear interpolation between background measurements and as an alternative to linear fitting of the mixing ratio was tested. For both tasks the input data was interpreted in various ways to obtain image or sequence-like data. Up to four different neural network architectures were trained to increase the choice of well performing models. The outputs of the fitter models contained large biases which rendered the model approach not applicable despite showing a great increase in MSE. A similar denoising behavior was also observed in Chapter 4. Subsequent tests with a classifier objective rejected the possibility of a poorly chosen loss function. As mentioned in Chapter 1, it cannot be excluded that a different model architecture or optimizing scheme might achieve a better outcome due to the vast size of the search space. Considering the chosen architectures that outperform humans

in image recognition tasks, it may be assumed that the target objective cannot be solved. The interpolator model showed more appropriate denoising behavior and was shown to be transferable to a different absorption spectrometer. Application of the interpolator to the test set lowered the detection limit by a factor of 2.

In summary, all advanced analysis methods analyzed show potential for application in various tasks in atmospheric measurements and infrared absorption spectroscopy. However, all of them are limited to specific properties of the input data and target objectives. In depth analysis of these limitations are mandatory to prevent biases and false conclusions emerging in the data.

As an outlook, many future studies based on the research done are possible. These include detailed modeling of the noise structure of the TRISTAR instrument, application of the Allan-Werle-plot to other airborne measurement instruments, further testing and application of the SMC method, transfer of the neural network architectures to other spectrometers.

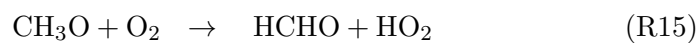
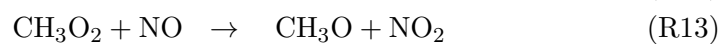
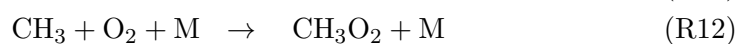
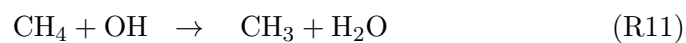
Many open problems and challenges presented in the conclusions are connected to the interference fringe patterns that limit the minimum detectable concentration of absorption spectrometers. Building a physical model to reproduce the observed patterns can lead to new insights in the field and implications to the optical design and possible data processing schemes.

Additionally, some new approaches to data quality analysis and improvements are presented that can have a great impact on the analysis, description and optimization of atmospheric measurement instruments. Spreading the utility of Allan-Werle-plots from the field of infrared absorption spectroscopy to other experiment designs, e.g., mass spectrometers, laser induced fluorescence or luminescence detection, can help investigating the temporal characteristics and measurement uncertainties in field experiments. Establishing the SMC method as a typical tool in atmospheric chemistry data analysis may open new doors to the quantitative determination of unknown parameters and improvements of overall data quality. Improvements in the data quality of atmospheric measurements will lead to a better understanding of chemical processes in the atmosphere and increase the accuracy of global circulation models to predict weather, climate and air pollution.

Appendix

1 Chemical reactions and formulas

The reactions given in Sect. 2.1.1 consist of these individual reactions [21]:



2 Supplement to Chapter 3

Parameter	First Plateau	Second Plateau	High Altitude	Landing	Ground
σ_W [ppbv]	0.421 ± 0.007	0.456 ± 0.005	–	–	0.4293 ± 0.0022
σ_B [ppbv]	0.0230 ± 0.0013	–	–	–	0.0309 ± 0.0004
σ_F [ppbv]	0.189 ± 0.006	0.1918 ± 0.0025	–	0.552 ± 0.008	–
A_{sin} [ppbv]	0.385 ± 0.014	0.730 ± 0.022	3.31 ± 0.04	7.03 ± 0.37	–
T_{sin} [s]	144 ± 5	897 ± 26	12.82 ± 0.06	251 ± 17	–
A_{MH} [ppbv]	–	–	1.68 ± 0.04	–	–
T_{MH} [s]	–	–	3.07 ± 0.17	–	–

Table A1: Optimal fit parameters of each Allan-Werle-plot shown in Figure 3.4 when fitting against (3.8). Parameters marked – are neglected since they do not contribute major features of the Allan variance. For comparison, the fit results from ground operation is also appended to the table.

3 Supplement to Chapter 4

Electronic supplementary material to Chapter 4 is available in the online version of [100].

4 Supplement to Chapter 5

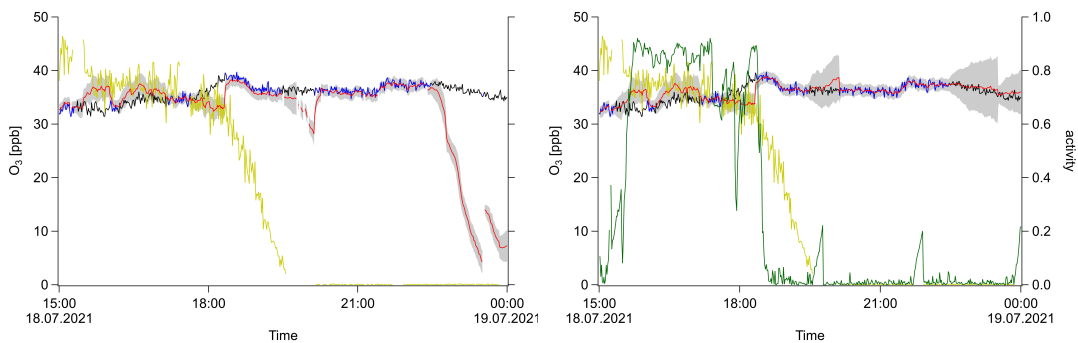


Figure A1: Visualization of the effect of the activity parameter η . The SMC shows undesired behavior when data is missing and the provided transition function does not correspond to the dominant system dynamics. This problem is shown on an example of the SMC used for interpolation of ozone. O_3 mixing ratio as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial gaps (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded grey region). Top: The basic algorithm described in section 5.3.1 without activity parameter η is used. When ozone data is missing during the night, the algorithm quickly decreases the resulting mixing ratio to values close to the PSS. Bottom: The extended algorithm 1 with activity parameter η is used. Its average value is plotted in green. At sunset the system becomes passive and the algorithm does not use the provided chemistry anymore. The stochastic part of the prior dominates and ozone no longer vanishes where no data is provided.

Appendix

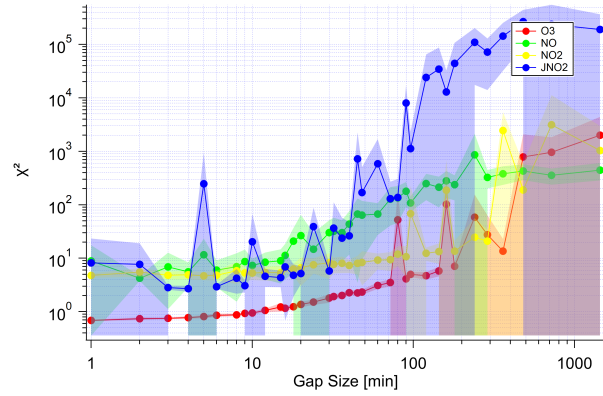


Figure A2: χ^2 of the SMC estimation as a function of artificial data gap size to study the interpolation capabilities. Mean χ^2 as lines and markers and standard deviation as shaded region for the ensemble of repetitions. The plot shows the results of all variables ozone (red), NO (green), NO₂ (yellow) and j_{NO_2} (blue).

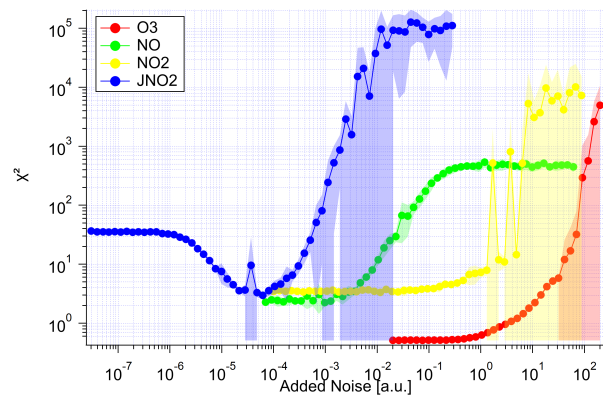


Figure A3: χ^2 of the SMC estimation as a function of artificial noise to study the precision enhancement abilities. Mean χ^2 as lines and markers and standard deviation as shaded region for the ensemble of repetitions. The plot shows the results of all variables ozone (red), NO (green), NO₂ (yellow) and j_{NO_2} (blue).

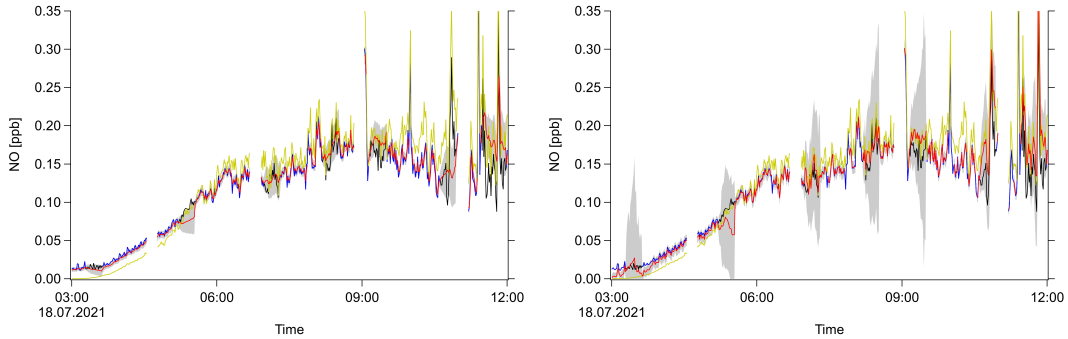


Figure A4: Variation of the free parameter σ_0 and its exemplary effect on the interpolation of NO. $\sigma_{0,\text{const}}$ is decreased (top) and increased (bottom) by a factor of 10. NO mixing ratio as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial gaps (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded grey region).

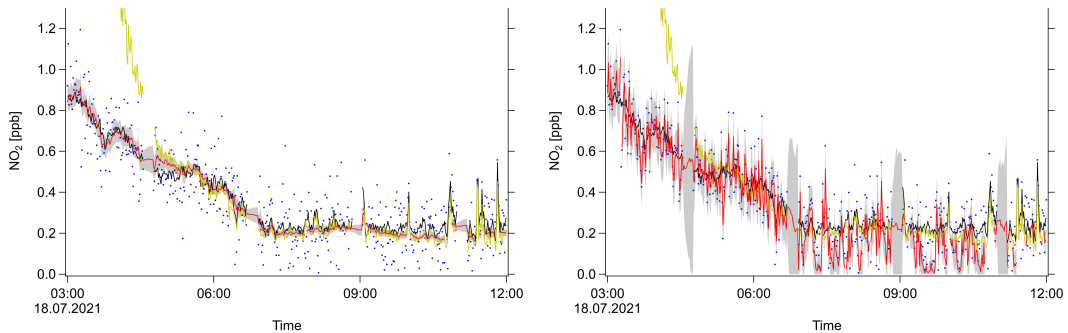


Figure A5: Variation of the free parameter σ_0 and its exemplary effect on the denoising of NO₂. $\sigma_{0,\text{const}}$ is decreased (top) and increased (bottom) by a factor of 10. NO₂ mixing ratio as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial noise (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded grey region).

Appendix

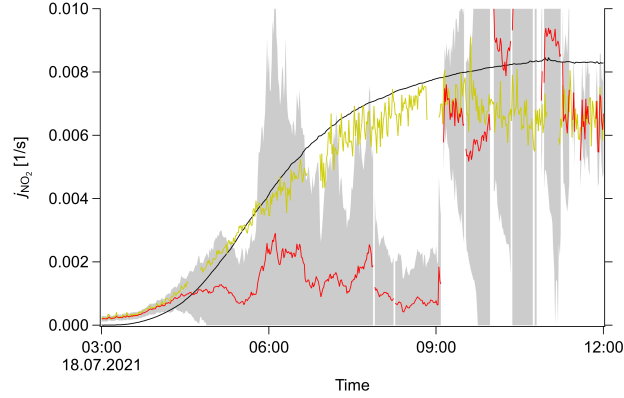


Figure A6: Variation of the free parameter σ_0 and its exemplary effect on the inference of j_{NO_2} . $\sigma_{0,\text{rel}}$ is increased by a factor of 3. Photolysis frequency j as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded grey region).

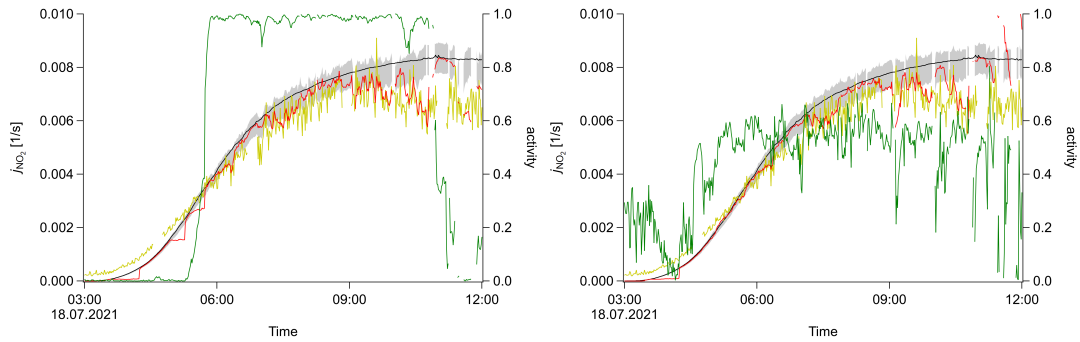


Figure A7: Variation of the free parameter p_η and its exemplary effect on the interpolation of j_{NO_2} . p_η is decreased (top) and increased (bottom) by a factor of 10. Photolysis frequency j as a function of time (UTC) for an arbitrary day of the field campaign. Original measurement (black), measurement with artificial gaps (blue), PSS calculation (yellow), SMC ensemble mean (red) and $\pm 1\sigma$ -interval (shaded grey region), average activity (green).

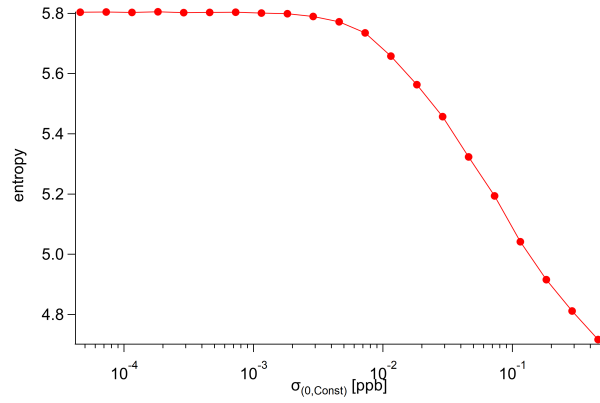


Figure A8: Variation of the free parameter $\sigma_{0,\text{const}}$ and its exemplary effect on the interpolation of NO. The median entropy is plotted for several runs as a function of each free parameter setting.

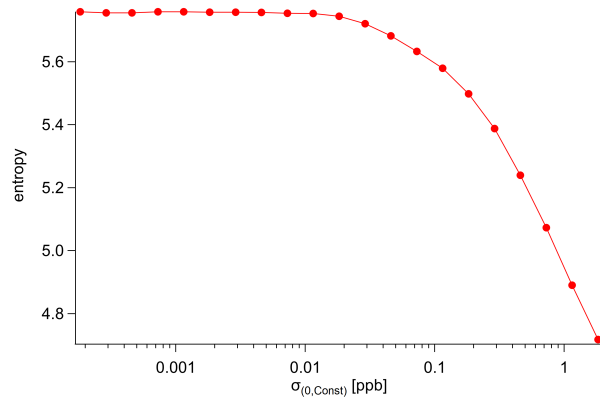


Figure A9: Variation of the free parameter $\sigma_{0,\text{const}}$ and its exemplary effect on the denoising of NO₂. The median entropy is plotted for several runs as a function of each free parameter setting.

Appendix

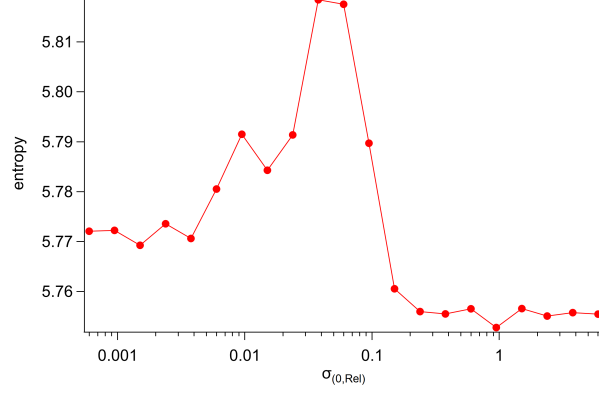


Figure A10: Variation of the free parameter $\sigma_{0,\text{rel}}$ and its exemplary effect on the inference of j_{NO_2} . The median entropy is plotted for several runs as a function of each free parameter setting. In this particular case the entropy shows a maximum and decreases again with lower σ_0 .

Variable	Detection Limit	Precision [%]	$\sigma_{0,\text{const}}$	$\sigma_{0,\text{rel}}$ [%]
O ₃	2 ppbv	2	0.68 ppbv	0.3
NO	7 pptv	4	4.6 pptv	6.6
NO ₂	10 pptv	12	18 pptv	2.7
j_{NO_2}	$3 \cdot 10^{-6} \text{ s}^{-1}$	10	$1.5 \cdot 10^{-5} \text{ s}^{-1}$	6

Table A2: System parameters used throughout the study. Detection Limit and Precision are taken from the measurement datasets. Note that the photolysis frequency is only calculated from solar radiation. The contribution from diffuse radiation is not considered. The σ_0 parameters are retrieved from linear fits of the difference between consecutive measurements vs the measurement value. Additional parameters are the switching probability $p_\eta = 2.5\%$, the ensemble size $K = 1000$ and the auxiliary ensemble size $R = 10000$.

5 Supplement to Chapter 6

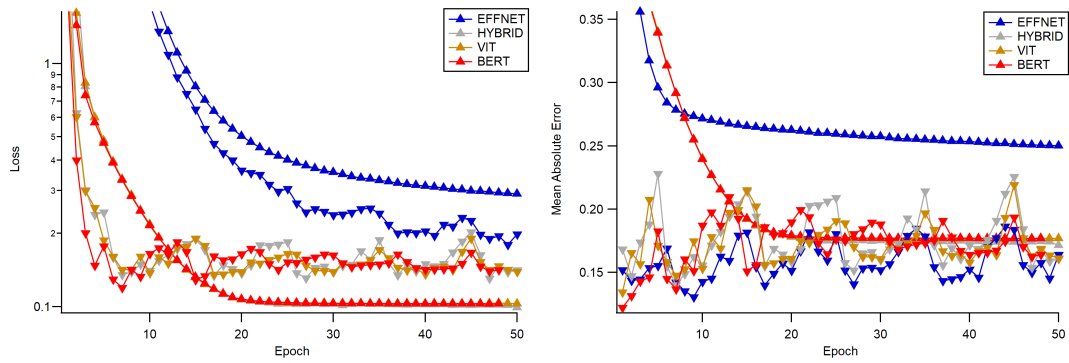


Figure A11: Training metric (upward triangle) and validation metric (downward triangle) of high-SNR fitter models based on BERT (red), VIT (yellow), Hybrid (grey) and EffV2 (blue) architecture: Left: Loss (mean squared error); Right: MAE (mean absolute error)

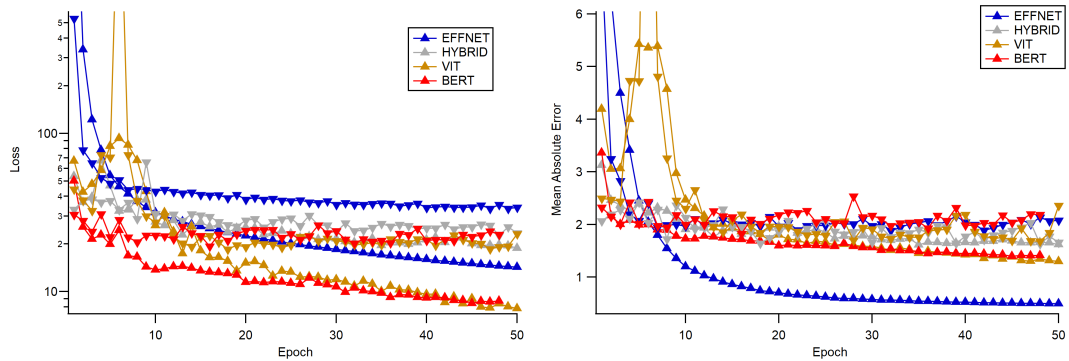


Figure A12: Training metric (upward triangle) and validation metric (downward triangle) of low-SNR fitter models based on BERT (red), VIT (yellow), Hybrid (grey) and EffV2 (blue) architecture: Left: Loss (mean squared error); Right: MAE (mean absolute error)

Appendix

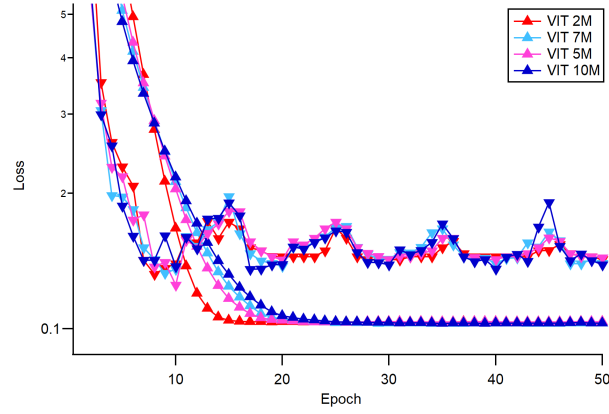


Figure A13: Ablation study. Training loss (upward triangle) and validation loss (downward triangle) of high-SNR fitter models based on ViT. Original model size performance in blue and reduced parameter versions in light blue, pink, red in the order of smaller model sizes.

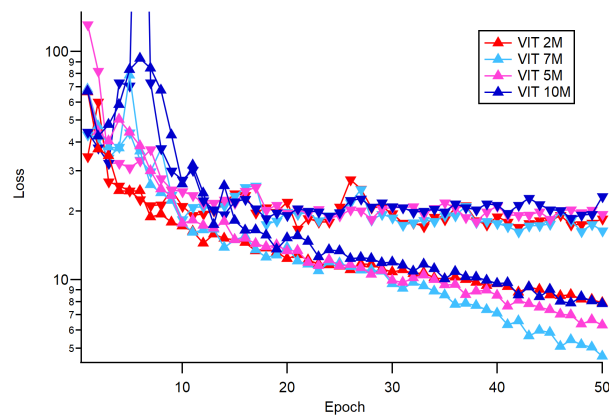


Figure A14: Ablation study. Training loss (upward triangle) and validation loss (downward triangle) of low-SNR fitter models based on ViT. Original model size performance in blue and reduced parameter versions in light blue, pink and red in the order of smaller model sizes.

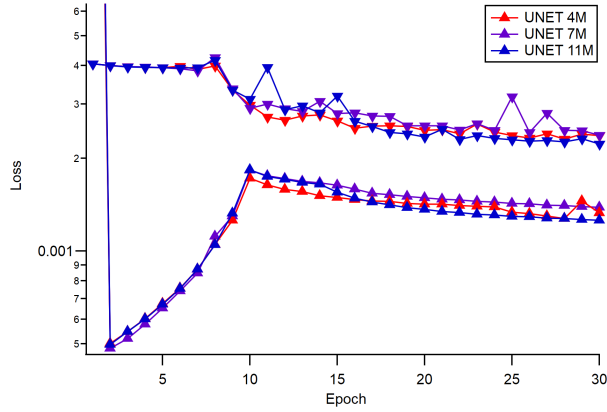


Figure A15: Ablation study. Training loss (upward triangle) and validation loss (downward triangle) of interpolator models based on UNET. Original model size performance in blue and reduced parameter versions in purple and red in the order of smaller model sizes.

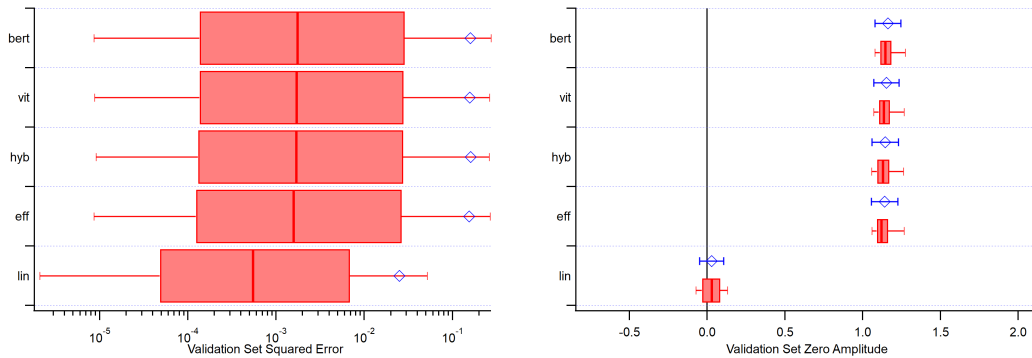


Figure A16: Evaluation of fitter models trained on high SNR range and comparison to linear fit. Red boxes indicate 25%-quantile, median and 75%-quantile; red whiskers indicate 10% and 90% quantile. Blue diamonds show the mean, blue whiskers (if provided) show one standard deviation. Left: absolute squared error between prediction and true value. Maxima are in the order 10^1 . Right: predicted amplitude of pure background spectra.

Appendix

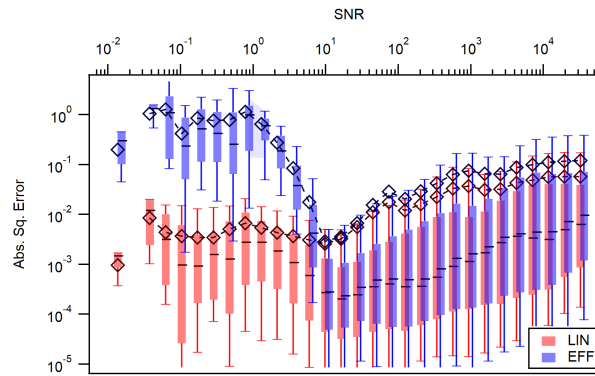


Figure A17: Fitter squared error as a function of input SNR. Linear fit in red, high SNR trained EffV2 variant prediction in blue. Boxes indicate 25%-quantile, median and 75%-quantile; whiskers indicate 10% and 90% quantile. Diamonds and lines show the mean. For better visualization the EffV2 plot is slightly offset along the x -axis.

Bibliography

- [1] P. Werle, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **54**(2), 197 (1998)
- [2] J.U. White. Optical system providing a long optical path (1957)
- [3] D. Herriott, H. Kogelnik, R. Kompfner, *Applied Optics* **3**(4), 523 (1964)
- [4] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, *Science* **264**(5158), 553 (1994). DOI 10.1126/science.264.5158.553
- [5] A. Fried, D. Richter, *Infrared Absorption Spectroscopy* (Blackwell Publishing Ltd, 2006), chap. 2, pp. 72–146. DOI 10.1002/9780470988510.ch2. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470988510.ch2>
- [6] J. Li, B. Yu, W. Zhao, W. Chen, *Applied Spectroscopy Reviews* **49**(8), 666 (2014)
- [7] Z. Wang, P. Fu, X. Chao, *Applied Sciences* **9**(13), 2723 (2019)
- [8] J. Nicely, T. Hanisco, H. Riris, *Journal of Quantitative Spectroscopy and Radiative Transfer* **211**, 115 (2018)
- [9] D. Richter, P. Weibring, J.G. Walega, A. Fried, S.M. Spuler, M.S. Taubman, *Applied Physics B* **119**(1), 119 (2015)
- [10] C. Torrence, G.P. Compo, *Bulletin of the American Meteorological Society* **79**(1), 61 (1998). DOI 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2
- [11] F. Wienhold, H. Fischer, P. Hoor, V. Wagner, R. Königstedt, G. Harris, J. Anders, R. Grisar, M. Knothe, W. Riedel, et al., *Applied Physics B* **67**(4), 411 (1998)
- [12] J. Rudolph, S. Tan, S. Tan, *Journal of Applied Learning and Teaching* **6**(1) (2023)
- [13] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., *Information fusion* **58**, 82 (2020)
- [14] D. Kaur, S. Uslu, K.J. Rittichier, A. Durresti, *ACM Computing Surveys (CSUR)* **55**(2), 1 (2022)

Bibliography

- [15] D.E. Heard, *Field Measurements of Atmospheric Composition* (Blackwell Publishing Ltd, 2006), chap. 1, pp. 1–71. DOI 10.1002/9780470988510.ch1. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470988510.ch1>
- [16] M. Martinez, H. Harder, D. Kubistin, M. Rudolf, H. Bozem, G. Eerdeken, H. Fischer, T. Klüpfel, C. Gurk, R. Königstedt, U. Parchatka, C.L. Schiller, A. Stickler, J. Williams, J. Lelieveld, *Atmospheric Chemistry and Physics* **10**(8), 3759 (2010). DOI 10.5194/acp-10-3759-2010. URL <https://acp.copernicus.org/articles/10/3759/2010/>
- [17] R. Dörich, P. Eger, J. Lelieveld, J.N. Crowley, *Atmospheric Measurement Techniques* **14**(8), 5319 (2021). DOI 10.5194/amt-14-5319-2021. URL <https://amt.copernicus.org/articles/14/5319/2021/>
- [18] C.M. Nussbaumer, U. Parchatka, I. Tadic, B. Bohn, D. Marno, M. Martinez, R. Rohloff, H. Harder, F. Kluge, K. Pfeilsticker, et al., *Atmospheric measurement techniques* **14**(10), 6759 (2021)
- [19] Z. Hamryszczak, A. Hartmann, D. Dienhart, S. Hafermann, B. Brendel, R. Königstedt, U. Parchatka, J. Lelieveld, H. Fischer, *Atmospheric Measurement Techniques Discussions* **2023**, 1 (2023)
- [20] D.E. Heard (ed.), *Analytical Techniques for Atmospheric Measurement* (Blackwell Publishing Ltd, 2006). DOI 10.1002/9780470988510
- [21] J.H. Seinfeld, S.N. Pandis (eds.), *Atmospheric chemistry and physics* (John Wiley & Sons, Inc., 1998). DOI 10.1002/9780470988510
- [22] C.E. Junge, *Tellus* **26**(4), 477 (1974). DOI 10.3402/tellusa.v26i4.9853. URL <https://doi.org/10.3402/tellusa.v26i4.9853>
- [23] J.N. Crowley, N. Pouvesle, G.J. Phillips, R. Axinte, H. Fischer, T. Petäjä, A. Nölscher, J. Williams, K. Hens, H. Harder, et al., *Atmospheric chemistry and physics* **18**(18), 13457 (2018)
- [24] C.M. Nussbaumer, J.N. Crowley, J. Schuladen, J. Williams, S. Hafermann, A. Reiffs, R. Axinte, H. Harder, C. Ernest, A. Novelli, K. Sala, M. Martinez, C. Mallik, L. Tomsche, C. Plass-Dülmer, B. Bohn, J. Lelieveld, H. Fischer, *Atmospheric Chemistry and Physics* **21**(24), 18413 (2021). DOI 10.5194/acp-21-18413-2021. URL <https://acp.copernicus.org/articles/21/18413/2021/>
- [25] J. Lelieveld, S. Gromov, A. Pozzer, D. Taraborrelli, *Atmospheric Chemistry and Physics* **16**(19), 12477 (2016). DOI 10.5194/acp-16-12477-2016. URL <https://acp.copernicus.org/articles/16/12477/2016/>
- [26] J. McManus, M. Zahniser, D. Nelson, J. Shorter, S. Herndon, D. Jervis, M. Agnese, R. McGovern, T. Yacovitch, J. Roscioli, *Applied Physics B* **119**, 203 (2015)

- [27] J. Hodgkinson, R.P. Tatam, *Measurement Science and Technology* **24**(1), 012004 (2012)
- [28] J.B. McManus, P.L. Kebabian, M. Zahniser, *Applied Optics* **34**(18), 3336 (1995)
- [29] B. Tuzson, M. Mangold, H. Looser, A. Manninen, L. Emmenegger, *Optics letters* **38**(3), 257 (2013)
- [30] J.B. McManus, M.S. Zahniser, D.D. Nelson Jr, J.H. Shorter, S. Herndon, E. Wood, R. Wehr, *Optical Engineering* **49**(11), 111124 (2010)
- [31] C.S. Goldenstein, R.M. Spearrin, J.B. Jeffries, R.K. Hanson, *Progress in Energy and Combustion Science* **60**, 132 (2017)
- [32] P. Werle, *Applied Physics B* **102**, 313 (2011)
- [33] J.B. McManus, C. Dyroff. Spectroscopic measurement response to interference fringes: Fundamental and aliased fringes (2022). FLAIR Conference
- [34] J.A. Silver, A.C. Stanton, *Applied optics* **27**(10), 1914 (1988)
- [35] P.W. Werle, B. Scheumann, J. Schandl, *Optical Engineering* **33**(9), 3093 (1994)
- [36] J. Li, U. Parchatka, H. Fischer, *Applied Physics B* **108**(4), 951 (2012)
- [37] I. Mapped-Fogaing, L. Joly, G. Durry, N. Dumelié, T. Decarpenterie, J. Cousin, *Applied Physics B* **108**(4), 933 (2012)
- [38] C. Schiller, H. Bozem, C. Gurk, U. Parchatka, R. Königstedt, G. Harris, J. Lelieveld, H. Fischer, *Applied Physics B* **92**(3), 419 (2008)
- [39] P. Werle, R. Mücke, F. Slemr, *Applied Physics B* **57**(2), 131 (1993)
- [40] A. LeNail, *J. Open Source Softw.* **4**(33), 747 (2019)
- [41] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning* (MIT press, 2016)
- [42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *The journal of machine learning research* **15**(1), 1929 (2014)
- [43] S. Ioffe, C. Szegedy, arXiv preprint arXiv:1502.03167 (2015)
- [44] O. Ronneberger, P. Fischer, T. Brox, in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241
- [45] K. He, X. Zhang, S. Ren, J. Sun, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778
- [46] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu, arXiv preprint arXiv:1706.06083 (2017)

Bibliography

- [47] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Communications of the ACM* **63**(11), 139 (2020)
- [48] M. Tan, Q. Le, in *International Conference on Machine Learning* (PMLR, 2021), pp. 10,096–10,106
- [49] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Proceedings of the IEEE* **86**(11), 2278 (1998)
- [50] A. Krizhevsky, I. Sutskever, G.E. Hinton, in *Advances in Neural Information Processing Systems*, vol. 25, ed. by F. Pereira, C. Burges, L. Bottou, K. Weinberger (Curran Associates, Inc., 2012), vol. 25. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [51] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, C. Bregler, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 648–656
- [52] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**(7553), 436 (2015)
- [53] S. Hochreiter, J. Schmidhuber, *Neural computation* **9**(8), 1735 (1997)
- [54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, *Advances in neural information processing systems* **30** (2017)
- [55] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, arXiv preprint arXiv:1810.04805 (2018)
- [56] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, D. Tran, in *International conference on machine learning* (PMLR, 2018), pp. 4055–4064
- [57] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., arXiv preprint arXiv:2010.11929 (2020)
- [58] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., *Advances in neural information processing systems* **33**, 1877 (2020)
- [59] OpenAI. ChatGPT (2023). URL <https://chat.openai.com>. [Large language model]
- [60] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., *nature* **550**(7676), 354 (2017)

- [61] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, in *Advances in Neural Information Processing Systems*, vol. 34, ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J.W. Vaughan (Curran Associates, Inc., 2021), vol. 34, pp. 12,077–12,090. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/64f1f27bf1b4ec22924fd0acb550c235-Paper.pdf
- [62] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 10,684–10,695
- [63] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, *Medical image analysis* **42**, 60 (2017)
- [64] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, K. Dietmayer, *IEEE Transactions on Intelligent Transportation Systems* **22**(3), 1341 (2020)
- [65] S. Youn, S.V. Jin, *Computers in Human Behavior* **119**, 106721 (2021)
- [66] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier, et al., *Learning and Individual Differences* **103**, 102274 (2023)
- [67] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al., arXiv preprint arXiv:1605.01156 (2016)
- [68] B.T. Pham, D. Tien Bui, H.R. Pourghasemi, P. Indra, M. Dholakia, *Theoretical and Applied Climatology* **128**, 255 (2017)
- [69] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T.N. Tran, L.M. Buydens, E. Marchiori, *Analytica chimica acta* **954**, 22 (2017)
- [70] J. Xia, Y. Huang, Q. Li, Y. Xiong, S. Min, *Environmental Chemistry Letters* **19**(5), 3547 (2021)
- [71] B.T. Ong, K. Sugiura, K. Zettsu, *Neural Computing and Applications* **27**, 1553 (2016)
- [72] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, *Environmental pollution* **231**, 997 (2017)
- [73] J. Gil, J. Kim, M. Lee, G. Lee, J. Ahn, D.S. Lee, J. Jung, S. Cho, A. Whitehill, J. Szykman, et al., *Atmospheric Environment* **247**, 118182 (2021)
- [74] J. Huang, H. Liu, J. Dai, W. Cai, *Journal of Quantitative Spectroscopy and Radiative Transfer* **218**, 187 (2018)

Bibliography

- [75] Y. Fu, R. Zhang, G. Enemali, A. Upadhyay, M. Lengden, C. Liu, in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (IEEE, 2022), pp. 1–5
- [76] C.D. Rankine, M.M. Madkhali, T.J. Penfold, *The Journal of Physical Chemistry A* **124**(21), 4263 (2020)
- [77] L. Tian, J. Sun, J. Chang, J. Xia, Z. Zhang, A.A. Kolomenskii, H.A. Schuessler, S. Zhang, *Measurement* **182**, 109739 (2021)
- [78] A. Nguyen, J. Yosinski, J. Clune, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015)
- [79] I.J. Goodfellow, J. Shlens, C. Szegedy, arXiv preprint arXiv:1412.6572 (2014)
- [80] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, D. Mané, arXiv preprint arXiv:1606.06565 (2016)
- [81] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, *ACM computing surveys (CSUR)* **54**(6), 1 (2021)
- [82] N. Akhtar, A. Mian, *Ieee Access* **6**, 14410 (2018)
- [83] L.L. Röder, L.M. Ort, J. Lelieveld, H. Fischer, (2023). DOI 10.21203/rs.3.rs-3619758/v1
- [84] P. Weibring, D. Richter, A. Fried, J. Walega, C. Dyroff, *Applied Physics B* **85**, 207 (2006)
- [85] P. Ehlers, A.C. Johansson, I. Silander, A. Foltynowicz, O. Axner, *J. Opt. Soc. Am. B* **31**(12), 2938 (2014). DOI 10.1364/JOSAB.31.002938. URL <https://opg.optica.org/josab/abstract.cfm?URI=josab-31-12-2938>
- [86] M. Giglio, P. Patimisco, A. Sampaolo, G. Scamarcio, F.K. Tittel, V. Spagnolo, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **63**(4), 555 (2016). DOI 10.1109/TUFFFC.2015.2495013
- [87] L. Dong, C. Li, N.P. Sanchez, A.K. Gluszek, R.J. Griffin, F.K. Tittel, *Applied Physics Letters* **108**(1), 011106 (2016). DOI 10.1063/1.4939452. URL <https://doi.org/10.1063/1.4939452>
- [88] D. Wilson, J.W. Phair, M. Lengden, *IEEE Sensors Journal* **19**(15), 6006 (2019). DOI 10.1109/JSEN.2019.2911737
- [89] J. Liao, G.M. Wolfe, R.A. Hannun, J.M. St. Clair, T.F. Hanisco, J.B. Gilman, A. Lamplugh, V. Selimovic, G.S. Diskin, J.B. Nowak, H.S. Halliday, J.P. Di-Gangi, S.R. Hall, K. Ullmann, C.D. Holmes, C.H. Fite, A. Agastra, T.B. Ryerson, J. Peischl, I. Bourgeois, C. Warneke, M.M. Coggon, G.I. Gkatzelis, K. Sekimoto, A. Fried, D. Richter, P. Weibring, E.C. Apel, R.S. Hornbrook,

- S.S. Brown, C.C. Womack, M.A. Robinson, R.A. Washenfelder, P.R. Veres, J.A. Neuman, *Atmospheric Chemistry and Physics* **21**(24), 18319 (2021). DOI 10.5194/acp-21-18319-2021. URL <https://acp.copernicus.org/articles/21/18319/2021/>
- [90] B. Yuan, L. Kaser, T. Karl, M. Graus, J. Peischl, T.L. Campos, S. Shertz, E.C. Apel, R.S. Hornbrook, A. Hills, J.B. Gilman, B.M. Lerner, C. Warneke, F.M. Flocke, T.B. Ryerson, A.B. Guenther, J.A. de Gouw, *Journal of Geophysical Research: Atmospheres* **120**(12), 6271 (2015). DOI <https://doi.org/10.1002/2015JD023242>
- [91] G.M. Wolfe, T.F. Hanisco, H.L. Arkinson, T.P. Bui, J.D. Crouse, J. Dean-Day, A. Goldstein, A. Guenther, S.R. Hall, G. Huey, D.J. Jacob, T. Karl, P.S. Kim, X. Liu, M.R. Marvin, T. Mikoviny, P.K. Misztal, T.B. Nguyen, J. Peischl, I. Pollack, T. Ryerson, J.M. St. Clair, A. Teng, K.R. Travis, K. Ullmann, P.O. Wennberg, A. Wisthaler, *Geophysical Research Letters* **42**(19), 8231 (2015). DOI <https://doi.org/10.1002/2015GL065839>
- [92] C.M. Nussbaumer, B.K. Place, Q. Zhu, E.Y. Pfannerstill, P. Wooldridge, B.C. Schulze, C. Arata, R. Ward, A. Bucholtz, J.H. Seinfeld, A.H. Goldstein, R.C. Cohen, *EGUsphere* **2023**, 1 (2023). DOI 10.5194/egusphere-2023-601. URL <https://egusphere.copernicus.org/preprints/2023/egusphere-2023-601/>
- [93] R. Kormann, H. Fischer, C. Gurk, F. Helleis, T. Klüpfel, K. Kowalski, R. Königstedt, U. Parchatka, V. Wagner, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **58**(11), 2489 (2002). DOI [https://doi.org/10.1016/S1386-1425\(02\)00066-5](https://doi.org/10.1016/S1386-1425(02)00066-5). URL <https://www.sciencedirect.com/science/article/pii/S1386142502000665>
- [94] I. Tadic, U. Parchatka, R. Königstedt, H. Fischer, *Applied Physics B* **123**, 1 (2017)
- [95] P. Hoor, H. Fischer, L. Lange, J. Lelieveld, D. Brunner, *Journal of Geophysical Research: Atmospheres* **107**(D5), ACL 1 (2002). DOI <https://doi.org/10.1029/2000JD000289>
- [96] A. Engel, H. Bönisch, D. Brunner, H. Fischer, H. Franke, G. Günther, C. Gurk, M. Hegglin, P. Hoor, R. Königstedt, M. Krebsbach, R. Maser, U. Parchatka, T. Peter, D. Schell, C. Schiller, U. Schmidt, N. Spelten, T. Szabo, U. Weers, H. Wernli, T. Wetter, V. Wirth, *Atmospheric Chemistry and Physics* **6**(2), 283 (2006). DOI 10.5194/acp-6-283-2006
- [97] L. Tomsche, A. Pozzer, N. Ojha, U. Parchatka, J. Lelieveld, H. Fischer, *Atmospheric Chemistry and Physics* **19**(3), 1915 (2019). DOI 10.5194/acp-19-1915-2019

Bibliography

- [98] P. Dewald, C.M. Nussbaumer, J. Schuladen, A. Ringsdorf, A. Edtbauer, H. Fischer, J. Williams, J. Lelieveld, J.N. Crowley, *Atmospheric Chemistry and Physics* **22**(10), 7051 (2022). DOI 10.5194/acp-22-7051-2022. URL <https://acp.copernicus.org/articles/22/7051/2022/>
- [99] G. Li, I.E. Gordon, L.S. Rothman, Y. Tan, S.M. Hu, S. Kassi, A. Campargue, E.S. Medvedev, *The Astrophysical Journal Supplement Series* **216**(1), 15 (2015)
- [100] L.L. Röder, H. Fischer, *Applied Physics B* **128**(1), 1 (2022)
- [101] P. Werle, F. Slemr, K. Maurer, R. Kormann, R. Mücke, B. Jänker, *Optics and lasers in engineering* **37**(2-3), 101 (2002)
- [102] A. Hartmann, R. Strzoda, R. Schrobenauser, R. Weigel, *Applied Physics B* **115**(2), 263 (2014)
- [103] A.K. Leung, F. Chau, J. Gao, *Chemometrics and Intelligent Laboratory Systems* **43**(1-2), 165 (1998)
- [104] P.S. Addison, *Physiological measurement* **26**(5), R155 (2005)
- [105] A. Grossmann, J. Morlet, *SIAM journal on mathematical analysis* **15**(4), 723 (1984)
- [106] I. Daubechies, *Communications on Pure and Applied Mathematics* **41**(7), 909 (1988). DOI 10.1002/cpa.3160410705
- [107] S.G. Mallat, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**(12), 2091 (1989)
- [108] B.K. Alsberg, A.M. Woodward, M.K. Winson, J. Rowland, D.B. Kell, *Analyst* **122**(7), 645 (1997)
- [109] I. Mapped-Fogaing, L. Joly, G. Durry, N. Dumelié, T. Decarpenterie, J. Cousin, B. Parvitte, V. Zéninari, *Applied spectroscopy* **66**(6), 700 (2012)
- [110] D.L. Donoho, I.M. Johnstone, *Journal of the american statistical association* **90**(432), 1200 (1995)
- [111] H.Y. Gao, *Journal of Computational and Graphical Statistics* **7**(4), 469 (1998)
- [112] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* **454**(1971), 903 (1998)
- [113] Y. Meng, T. Liu, K. Liu, J. Jiang, R. Wang, T. Wang, H. Hu, *IEEE Photonics Journal* **6**(6), 1 (2014)

- [114] I. Gordon, L. Rothman, C. Hill, R. Kochanov, Y. Tan, P. Bernath, M. Birk, V. Boudon, A. Campargue, K. Chance, B. Drouin, J.M. Flaud, R. Gamache, J. Hodges, D. Jacquemart, V. Perevalov, A. Perrin, K. Shine, M.A. Smith, J. Tennyson, G. Toon, H. Tran, V. Tyuterev, A. Barbe, A. Császár, V. Devi, T. Furtenbacher, J. Harrison, J.M. Hartmann, A. Jolly, T. Johnson, T. Karman, I. Kleiner, A. Kyuberis, J. Loos, O. Lyulin, S. Massie, S. Mikhailenko, N. Moazzen-Ahmadi, H. Müller, O. Naumenko, A. Nikitin, O. Polyansky, M. Rey, M. Rotger, S. Sharpe, K. Sung, E. Starikova, S. Tashkun, J.V. Auwera, G. Wagner, J. Wilzewski, P. Wcisło, S. Yu, E. Zak, *Journal of Quantitative Spectroscopy and Radiative Transfer* **203**, 3 (2017). DOI <https://doi.org/10.1016/j.jqsrt.2017.06.038>. HITRAN2016 Special Issue
- [115] Q. Zou, P. Varanasi, *Journal of Quantitative Spectroscopy and Radiative Transfer* **75**(1), 63 (2002)
- [116] J. Reid, D. Labrie, *Applied Physics B* **26**(3), 203 (1981). DOI 10.1007/BF00692448
- [117] G. Vivó-Truyols, P.J. Schoenmakers, *Analytical chemistry* **78**(13), 4598 (2006)
- [118] M. Sadeghi, F. Behnia, R. Amiri, *IEEE Transactions on Instrumentation and Measurement* **69**(8), 5418 (2020)
- [119] S. Schilt, L. Thévenaz, *Infrared physics & technology* **48**(2), 154 (2006)
- [120] L.L. Röder, P. Dewald, C.M. Nussbaumer, J. Schuladen, J.N. Crowley, J. Lelieveld, H. Fischer, *EGU sphere* pp. 1–20 (2022)
- [121] H. Hidalgo, P. Crutzen, *Journal of Geophysical Research* **82**(37), 5833 (1977)
- [122] J. Lelieveld, E. Bourtsoukidis, C. Brühl, H. Fischer, H. Fuchs, H. Harder, A. Hofzumahaus, F. Holland, D. Marno, M. Neumaier, et al., *Science* **361**(6399), 270 (2018)
- [123] S. Wofsy, S. Afshar, H. Allen, E. Apel, E. Asher, B. Barletta, J. Bent, H. Bian, B. Biggs, D. Blake, N. Blake, I. Bourgeois, C. Brock, W. Brune, J. Budney, T. Bui, A. Butler, P. Campuzano-Jost, C. Chang, M. Chin, R. Commane, G. Correa, J. Crouse, P.D. Cullis, B. Daube, D. Day, J. Dean-Day, J. Dibb, J. DiGangi, G. Diskin, M. Dollner, J. Elkins, F. Erdesz, A. Fiore, C. Flynn, K. Froyd, D. Gesler, S. Hall, T. Hanisco, R. Hannun, A. Hills, E. Hintsa, A. Hoffman, R. Hornbrook, L. Huey, S. Hughes, J. Jimenez, B. Johnson, J. Katich, R. Keeling, M. Kim, A. Kupc, L. Lait, J.F. Lamarque, J. Liu, K. McKain, R. McLaughlin, S. Meinardi, D. Miller, S. Montzka, F. Moore, E. Morgan, D. Murphy, L. Murray, B. Nault, J. Neuman, P. Newman, J. Nicely, X. Pan, W. Paplawsky, J. Peischl, M. Prather, D. Price, E. Ray, J. Reeves, M. Richardson, A. Rollins, K. Rosenlof, T. Ryerson, E. Scheuer, G. Schill, J. Schroder,

Bibliography

- J. Schwarz, J. St.Clair, S. Steenrod, B. Stephens, S. Strode, C. Sweeney, D. Tanner, A. Teng, A. Thames, C. Thompson, K. Ullmann, P. Veres, N. Vieznor, N. Wagner, A. Watt, R. Weber, B. Weinzierl, P. Wennberg, C. Williamson, J. Wilson, G. Wolfe, C. Woods, L. Zeng. Atom: Merged atmospheric chemistry, trace gases, and aerosols [data set] (2018). DOI 10.3334/ORNLDAAAC/1581. URL https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1581
- [124] B. Ridley, S. Madronich, R. Chatfield, J. Walega, R. Shetter, M. Carroll, D. Montzka, *Journal of Geophysical Research: Atmospheres* **97**(D10), 10375 (1992)
- [125] G.K. Georgiou, T. Christoudias, Y. Proestos, J. Kushta, P. Hadjinicolaou, J. Lelieveld, *Atmospheric Chemistry and Physics* **18**(3), 1555 (2018)
- [126] G. Kitagawa, *Journal of computational and graphical statistics* **5**(1), 1 (1996)
- [127] M.K. Pitt, N. Shephard, *Journal of the American statistical association* **94**(446), 590 (1999)
- [128] P. Bauer, A. Thorpe, G. Brunet, *Nature* **525**(7567), 47 (2015)
- [129] P.J. Van Leeuwen, H.R. Künsch, L. Nerger, R. Potthast, S. Reich, *Quarterly Journal of the Royal Meteorological Society* **145**(723), 2335 (2019)
- [130] M. Krol, P.J. van Leeuwen, J. Lelieveld, *Journal of Geophysical Research: Atmospheres* **103**(D9), 10697 (1998)
- [131] T. Berkemeier, M. Ammann, U.K. Krieger, T. Peter, P. Spichtinger, U. Pöschl, M. Shiraiwa, A.J. Huisman, *Atmospheric Chemistry and Physics* **17**(12), 8021 (2017)
- [132] S. Guo, R. Yang, H. Zhang, W. Weng, W. Fan, *International Journal of Heat and Mass Transfer* **52**(17-18), 3955 (2009)
- [133] A. Wawrzynczak, P. Kopka, M. Borysiewicz, in *Parallel Processing and Applied Mathematics*, ed. by R. Wyrzykowski, J. Dongarra, K. Karczewski, J. Waśniewski (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014), pp. 407–417
- [134] N. de Freitas, C. Andrieu, P. Højten-Sørensen, M. Niranjana, A. Gee, *Sequential Monte Carlo Methods for Neural Networks* (Springer New York, New York, NY, 2001), pp. 359–379. DOI 10.1007/978-1-4757-3437-9_17. URL https://doi.org/10.1007/978-1-4757-3437-9_17
- [135] X. Ma, P. Karkus, D. Hsu, W.S. Lee, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34 (2020), vol. 34, pp. 5101–5108

- [136] A. Doucet, N. de Freitas, N. Gordon, *An Introduction to Sequential Monte Carlo Methods* (Springer New York, New York, NY, 2001), pp. 3–14. DOI 10.1007/978-1-4757-3437-9_1. URL https://doi.org/10.1007/978-1-4757-3437-9_1
- [137] N.J. Gordon, D.J. Salmond, A.F. Smith, in *IEE Proceedings F-radar and signal processing*, vol. 140 (IET, 1993), vol. 140, pp. 107–113
- [138] A. Doucet, S. Godsill, C. Andrieu, *Statistics and computing* **10**(3), 197 (2000)
- [139] R.E. Kalman, *Journal of Basic Engineering* (1960)
- [140] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, *Monthly Weather Review* **136**(12), 4629 (2008)
- [141] P. Fearnhead, H.R. Künsch, *Annual Review of Statistics and Its Application* **5**, 421 (2018)
- [142] M. Pulido, P.J. van Leeuwen, *Journal of Computational Physics* **396**, 400 (2019)
- [143] C.C. Hu, P.J. van Leeuwen, *Quarterly Journal of the Royal Meteorological Society* (2021)
- [144] P. Leighton, *Photochemistry of air pollution* (Academic Press, Inc., 1961)
- [145] M. Nicolet, *Journal of Geophysical Research* **70**(3), 679 (1965)
- [146] R. Atkinson, D. Baulch, R.A. Cox, J. Crowley, R. Hampson, R. Hynes, M. Jenkin, M. Rossi, J. Troe, *Atmospheric chemistry and physics* **4**(6), 1461 (2004)
- [147] P.J. Crutzen, *Annual review of earth and planetary sciences* **7**(1), 443 (1979)
- [148] D. Parrish, M. Trainer, E. Williams, D. Fahey, G. Hübler, C. Eubank, S. Liu, P. Murphy, D. Albritton, F. Fehsenfeld, *Journal of Geophysical Research: Atmospheres* **91**(D5), 5361 (1986)
- [149] J.N. Crowley, G. Schuster, N. Pouvesle, U. Parchatka, H. Fischer, B. Bonn, H. Bingemer, J. Lelieveld, *Atmospheric Chemistry and Physics* **10**(6), 2795 (2010). DOI 10.5194/acp-10-2795-2010. URL <https://acp.copernicus.org/articles/10/2795/2010/>
- [150] N. Sobanski, M. Tang, J. Thieser, G. Schuster, D. Pöhler, H. Fischer, W. Song, C. Sauvage, J. Williams, J. Fachinger, et al., *Atmospheric Chemistry and Physics* **16**(8), 4867 (2016)
- [151] I. Tadic, J.N. Crowley, D. Dienhart, P. Eger, H. Harder, B. Hottmann, M. Martinez, U. Parchatka, J.D. Paris, A. Pozzer, et al., *Atmospheric Chemistry and Physics* **20**(11), 6769 (2020)

Bibliography

- [152] B. Bohn, I. Lohse, Atmospheric Measurement Techniques **10**(9), 3151 (2017). DOI 10.5194/amt-10-3151-2017. URL <https://amt.copernicus.org/articles/10/3151/2017/>
- [153] E. Limpert, W.A. Stahel, M. Abbt, BioScience **51**(5), 341 (2001)
- [154] J. Wildt, D. Kley, A. Rockel, P. Rockel, H. Segschneider, Journal of Geophysical Research: Atmospheres **102**(D5), 5919 (1997)
- [155] K. Hens, A. Novelli, M. Martinez, J. Auld, R. Axinte, B. Bohn, H. Fischer, P. Keronen, D. Kubistin, A. Nölscher, et al., Atmospheric Chemistry and Physics **14**(16), 8723 (2014)
- [156] H. Riris, C.B. Carlisle, R.E. Warren, Applied optics **33**(24), 5506 (1994)
- [157] D. Leleux, R. Claps, W. Chen, F. Tittel, T. Harman, Applied Physics B **74**(1), 85 (2002)
- [158] L.L. Röder, (2023). DOI 10.21203/rs.3.rs-2832856/v1
- [159] I. Cortés-Ciriano, A. Bender, Journal of chemical information and modeling **59**(3), 1269 (2018)
- [160] J. Pyo, S.M. Hong, Y.S. Kwon, M.S. Kim, K.H. Cho, Science of the Total Environment **741**, 140162 (2020)
- [161] K.I. Funahashi, Neural networks **2**(3), 183 (1989)
- [162] K. Hornik, M. Stinchcombe, H. White, Neural networks **2**(5), 359 (1989)
- [163] G. Cybenko, Mathematics of control, signals and systems **2**(4), 303 (1989)
- [164] A. Krizhevsky, I. Sutskever, G.E. Hinton, Communications of the ACM **60**(6), 84 (2017)
- [165] X.J. Mao, C. Shen, Y.B. Yang, arXiv preprint arXiv:1606.08921 (2016)
- [166] A. Perrin, D. Jacquemart, F.K. Tchana, N. Lacome, Journal of Quantitative Spectroscopy and Radiative Transfer **110**(9-10), 700 (2009)
- [167] A. Sergeev, M.D. Balso, arXiv preprint arXiv:1802.05799 (2018)