Spatial Heterodyne Observations of Atmospheric Water Vapour from a High Altitude Aircraft

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Physics and Engineering Physics
University of Saskatchewan
Saskatoon

By
Daniel Letros

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Physics and Engineering Physics
University of Saskatchewan
Room 163, 116 Science Place
Saskatoon, Saskatchewan
S7N 5E2
Canada

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan
S7N 5C9
Canada

Abstract

The atmosphere of the Earth is divided into layers, and the boundaries of these layers are positioned at altitudes where inversions in the atmospheric temperature lapse rate occur. The content of the present work is concerned with water vapour within the upper troposphere - lower stratosphere (UTLS) which is a region of the atmosphere within five kilometres of the tropopause - the boundary between the troposphere and stratosphere. The behaviour of this region is very important in the formation of both the troposphere and stratosphere with water vapour being a principle factor. While broad characteristics of the UTLS, and the role of water vapour within, are well understood questions do remain. To get a better understanding of the answers to these questions there is a need for a continuous high resolution scientific data set of UTLS constituents like water vapour.

The Spatial Heterodyne Observation of Water (SHOW) is a scientific instrument designed to measure atmospheric water vapour. The premise of the SHOW measurement technique is Spatial Heterodyne Spectroscopy (SHS) which performs a frequency decomposition on input light to create an interferogram. SHOW measures the atmosphere by observing the UTLS region in the limb. The measurements of SHOW can be combined with a radiative transfer model, a SHS instrument model, and an inversion technique to infer the water content of the observed atmosphere.

Due to the nature of the SHS technique, SHOW has the ability to obtain measurements at a high spatial and altitude resolution. This gives SHOW, and SHS technology, the potential to obtain the desired high resolution data sets of the UTLS. However, SHS based instruments, including SHOW, are largely unproven for the application of atmospheric research like measuring water vapour. As a demonstration to validate SHOW, and SHS technology, as applicable to atmospheric science, SHOW was deployed on NASA's ER-2 high altitude science aircraft in July of 2017 for a scientific campaign. The goal of this campaign was to determine the abilities of SHOW with the desired results being to measure atmospheric water vapour to within ± 1 ppm, with a vertical resolution of less than five hundred meters, and with high spatial sampling. To provide a comparison for the assessment of SHOW, a Vaisala RS41 radiosonde was launched from the Jet Propulsion Laboratory facility located

close to Table Mountain. This radiosonde was launched in-situ with some SHOW measurements taken during the engineering flight on July 18th, 2017 and measured the atmospheric water vapour within the region above the facility.

The SHOW data which corresponded to the radiosonde measurements was analyzed and found that the measurements between the two instruments agreed largely within the goal of ± 1 ppm. Furthermore, SHOW was able to do so at a vertical resolution of two hundred and fifty meters and achieved a spatial resolution of 0.005° to 0.01° in latitude (roughly 500 m to 1000 m) along a north-south flight track when deployed on the ER-2 platform. These results lend strong supporting evidence that SHOW is capable of providing the desired high resolution UTLS water vapour data set and should be considered for further development and deployment in the future. Furthermore, SHS based instruments should be considered viable atmospheric instruments, and should potentially be used to measure other atmospheric constituents within the UTLS.

Acknowledgements

We all stand on the shoulders of giants. Here are the ones I stand on:

- **My family** for raising me and providing me support whenever I needed it.
- My supervisor, **Adam Bourassa**, for providing me with academic guidance and countless opportunities that have vastly enriched my education. The knowledge and experience I have gained at the University of Saskatchewan would only be a fraction of what it is now without his intervention.
- My closest coworker on the SHOW project, **Jeff Langille**, who taught me a lot about optics, instrumentation, drinking, and gremlins as we worked well together on the content in this thesis.
- **Paul Loewen**, who has been a very close and constant coworker with me on multiple projects during my time in ISAS. Vitally important to much of the work our group does. I can't think of a campaign we could participate in without his work.
- **Nick Lloyd**, another close coworker of mine. He taught me a lot on the proper thought processes and procedures a scientist or engineer should take when doing their job. He's all ducky in my book.
- Last but not least are **the teachers and staff members of ISAS including all the students in room 312**, past and present. My opportunities at ISAS would not be possible without our capabilities as a team. I would like to extend special thanks to **Doug Degenstein**, **Daniel Zawada**, **Matt Kozun**, **Landon Rieger**, and **Seth Dueck**. These individuals have all had potent impacts on me academically in both my current graduate studies and past undergraduate studies. If they didn't help out directly with the content in this thesis then they definitely did indirectly.

Contents

List of Tables

List of Figures

List of Abbreviations

AFRC	Armstrong Flight Research Center
ECCC	Environment and Climate Change Canada
ER-2	Earth Resource 2
FFT	Fast Fourier Transform
GloSSAC	Global Space-based Stratospheric Aerosol Climatology
MJD	Modified Julian Date
SHOW	Spatial Heterodyne Observation of Water
SHS	Spatial Heterodyne Spectroscopy
SNR	Signal to noise ratio
STE	Stratosphere-troposphere Exchange
TIL	Tropopause Inversion Layer
UTC	Coordinated Universal Time
UTLS	Upper Troposphere - Lower Stratosphere

$\Delta\sigma$	Observable spectral range
γ	Angle of a diffraction grating's exiting wavefront with respect to the optical axis
λ	Wavelength
σ	Wavenumber
λ_L	Littrow wavelength
σ_L	Littrow wavenumber
θ_L	Littrow angle

Chapter 1

Introduction

Water vapour makes up less than 1% of the constituents in the Earth's atmosphere but plays an extremely important role in both everyday weather and long term climate trends [Caballero, 2014]. The influence of water on the atmosphere is largely driven by its role in the atmospheric processes that take place in the upper troposphere - lower stratosphere (UTLS), a region where the troposphere and stratosphere interact. The processes that take place here drive temperatures in both the troposphere and stratosphere, influence concentrations of other species like ozone, and dictate the radiative balance of the region [Sherwood et al., 2010; Gettelman et al., 2011]. Due to its importance, the UTLS is of considerable interest to the field of atmospheric science. To this end there is cause to obtain continuous, and higher vertical and spatial resolution measurements than what is currently available [Gettelman et al., 2011; Randel and Jensen, 2013; SPARC et al., 2017].

The Spatial Heterodyne Observation of Water (SHOW) is a scientific instrument intended to pursue this need of an UTLS data set. It was originally prototyped at York University [Lin, 2010] as part of a Canadian Space Agency Advanced Studies Project. A balloon and aircraft worthy flight version of SHOW [Langille et al., 2017] followed this development which was done by industry partner ABB, and in cooperation with the Canadian Space Agency, as well as the University of Saskatchewan. The University of Saskatchewan is currently leading the SHOW project.

SHOW is designed to measure atmospheric water vapour within the UTLS by employing the Spatial Heterodyne Spectroscopy (SHS) technique [Harlander, 1991] to perform a frequency decomposition of atmospheric light observed in the limb and record the resulting interferograms. Specifically, SHOW observes a vibrational band of water between 1363 nm and 1366 nm to a resolution of ≈ 0.02 nm.

SHS technology is unproven for the application of measuring atmospheric water vapour and needs validation. As such, the fundamental mission of SHOW, and the objective of the present work, is to validate the technique as applicable to the field. To this end a campaign was carried out in mid July of 2017 involving the University of Saskatchewan, the Canadian Space Agency, NASA, and industry partner ABB, in which SHOW was deployed on NASA's Earth Resource 2 (ER-2) high altitude science aircraft. This campaign was done to provide an opportunity where SHOW could take measurements of UTLS water vapour concentration.

The requirements for meeting the objective of validation was to demonstrate that SHOW measurements could result in measured water vapour profiles to within ± 1 ppm, with a vertical resolution of less than five hundred meters, and with high spatial sampling capability [Langille et al., 2018]. This validation is done through comparison of SHOW measurements to data from a Vaisala RS41 radiosonde which was launched from the Jet Propulsion Laboratory facility located close to Table Mountain. The radiosonde was launched in-situ with some measurements taken during the campaign. If SHOW proves itself feasible the goal is to one day deploy in low Earth orbit for long term measurements.

As part of this endeavour, a significant amount of work was needed to build the necessary infrastructure in support of the mission objective. Most of this infrastructure involved the development of a model that is capable of turning the measured interferograms into water vapour concentration. This is done through a processes of accurately simulating atmospheric radiance for a particular limb geometry, accurately simulating the SHOW instrument, and combining these with a retrieval algorithm that determines how the simulated radiances needs to change in order to match the observed SHOW measurements. In addition, calibration work needed to be done on SHOW itself so that its measurements could be correctly interpreted and the developed model needed to be evaluated for its own capabilities.

The content of the present work concerns itself with documenting the work done in the effort and scope of validating SHOW and SHS technology as being viable for atmospheric measurement of water within the desired accuracy and resolution goals. Following this intent background information on water vapour in the UTLS, different atmospheric measurement techniques and their trades, radiative transfer theory, current atmospheric water observing instruments, SHS theory, and details of SHOW as an instrument are provided. In addition,

the details of the theory, construction, and resulting abilities of the aforementioned model are presented as well. Finally, information on the SHOW instrument calibration, the July 2017 campaign, and the results of analyzing the campaign data with respect to water vapour are given which culminates in a reported conclusion of the abilities of SHOW.

Chapter 2

Background

2.1 Water Vapour in the Upper Troposphere - Lower Stratosphere

Earth's atmosphere can be sectioned into five different layers. From lowest to highest these layers are the troposphere, stratosphere, mesosphere, thermosphere, and exosphere. The boundaries of these layers are defined by the changes in the lapse rates of the atmospheric temperature profile [Salby, 2012]. The present work only concerns itself with the troposphere and stratosphere, specifically the region around the boundary between the two. The troposphere is warmest closest to the surface of the Earth and cools with increasing altitude. This temperature gradient promotes vertical mixing of the gases in this region that drives much of the everyday weather experienced on the surface [Salby, 2012]. In contrast, the layer above the troposphere, the stratosphere, has a temperature profile that increases with altitude promoting atmospheric stability. The tropopause marks the boundary between these two layers in which the positive lapse rate of the troposphere turns into the negative lapse rate of the stratosphere. The most common definition, the thermal definition, for the location of the tropopause comes from the World Meteorological Organization in which it is defined as *"the lowest level at which the lapse rate decreases to 2 °C/km or less, provided also the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C/km."* [Organization, 1957]. There are a number of factors which determine at what altitude this lapse rate occurs but typically the tropopause is located at an altitude of approximately nine kilometres in the polar regions and approximately sixteen kilometres around the equator. The UTLS is generally classified as the region of the atmosphere which is within 5 km

of the tropopause [Gettelman et al., 2011]. Figure 2.1 below shows the UTLS region for a tropopause located at sixteen kilometres along with example temperatures and pressure with respect to altitude.

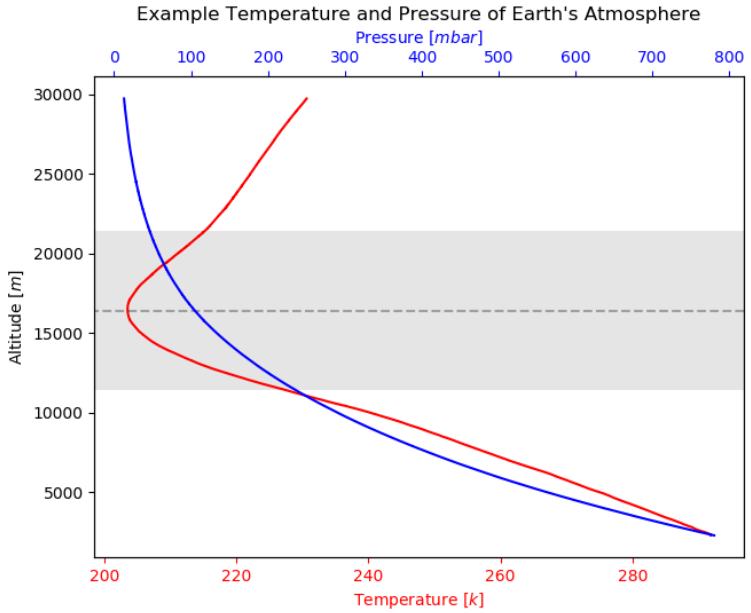

Figure 2.1: Example pressure and temperature of the troposphere and stratosphere. UTLS region shaded in gray.

In general, the atmosphere is made up of 78% nitrogen and 21% oxygen. The remaining 1% of constituents in the atmosphere is made up of argon, carbon dioxide, helium, hydrogen, krypton, methane, neon, water, and other trace species such as aerosols and ozone [Caballero, 2014]. As Figure 2.1 above shows, the pressure exponentially decrease with altitude. The rate of this decrease is characterized by the scale height which is defined as the altitude in which the pressure decreases by a factor of $1/e$. The temperature profile seen in this same figure is principally controlled by the solar flux incident on the Earth's surface and the concentration of ozone. As sunlight strikes the surface of the Earth it heats up the ground which then re-radiates this energy back into the atmosphere in the infrared region of the spectrum acting as a heat source. As altitude increases the pressure surrounding any parcel of air deceases and causes it to expand. The expansion reduces the temperature until the tropopause. Above the tropopause ozone is absorbing energy from the ultra-violate radiation of the Sun

and acts as another heat source to warm the atmosphere. This causes the increase in the temperature profile above the tropopause as a function of altitude.

The atmospheric processes that occur within the UTLS play a pivotal role in the characteristics of both the troposphere and stratosphere. Principal among these processes are the effects associated with the presence of water vapour. Water vapour accounts for the majority of upward heat transfer and drives convection in the troposphere [Sherwood et al., 2010], a primary force behind weather in this region. In addition, convection and heat transfer in the troposphere is heavily linked to the structure and composition of the stratosphere through stratosphere-troposphere exchange (STE) within the UTLS. This exchange influences concentrations of radiative and chemically important species like ozone [Gettelman et al., 2011]. For example, ozone is extremely important to temperatures in the stratosphere as well as habitability of the troposphere through absorption of harmful ultra-violet radiation. Influenced by STE, water vapour in the UTLS is thought to play a major role in the radiative balance of the region and forming of the tropopause inversion layer (TIL) which is a region of static stability above the tropopause. The region acts as a convective barrier hindering vertical transport of constituents and is linked to the temperature inversion of the tropopause giving it its name [Gettelman et al., 2011].

While a lot of the behaviour and composition of the UTLS is understood, there are some uncertainties to the role of water vapour in the formation of the TIL as well as questions related to the radiative feedback mechanisms associated with water vapour in the UTLS, such as, how the processes which control the chemical composition of the UTLS vary temporally and spatially, the relative importance in TIL formation between dynamical and radiative processes and its relation to STE [Gettelman et al., 2011]. There are broader questions concerning the UTLS role in climate change through mechanisms of STE [Randel and Jensen, 2013]. In the pursuit of answers to these questions, a continuous high resolution data set of UTLS constituents, especially of water vapour, would play a significant and vital role in the study of climate processes and improvement of modelling capabilities [Gettelman et al., 2011; Randel and Jensen, 2013; SPARC et al., 2017; Hegglin et al., 2013]. For example, it is known that there are steep chemical gradients and large dynamic variability within the UTLS which is a likely reason for the observed disagreement between some instruments which

have made UTLS measurements in the past [Hegglin et al., 2013]. Capturing this variability continuously will allow for significant reduction on the uncertainty in UTLS observations which provide insight to the processes that take place there [Gettelman et al., 2011].

2.2 Atmospheric Measurement Techniques

Scientific instruments which measure the atmosphere can do so directly or indirectly. Direct methods involve an instrument or sensor measuring the atmosphere through some controlled method of evaluation that directly measures a property. As an example, a capacitive humidity sensor uses a hygroscopic dielectric material between two electrodes to form a capacitor. As the relative humidity of the atmosphere surrounding the dielectric material changes so does its dielectric constant which changes the capacitance. This allows for a *direct* relationship between the relative humidity of the atmosphere in contact with the sensor and the measured capacitance.

In contrast to direct measurements, indirect measurements work by measuring one property and using that to infer the state of another property. In atmospheric science, indirect measurements *often* imply measuring the spectrum of the atmosphere and then using knowledge of spectroscopy and atmospheric physics to determine some other property. Using the example of humidity again, indirect measurements differ from direct measurements in that where a humidity sensor's output directly resembles the relative humidity, the output of an indirect measurement only directly resembles the spectrum and additional knowledge or information is needed to get the humidity information from the spectrum.

While direct measurements have the advantage of being simpler to process, they suffer from the amount of data they can acquire. Simply put, a humidity sensor can report about the water content immediately surrounding it but the spectrum taken from some detector can potentially tell you about the water content in its entire optical field of view. It is for this reason that indirect measurements, commonly refereed to as remote sensing in atmospheric science, is a large focus of work within the field.

Instruments which indirectly measure the atmosphere are deployed from the ground, high-altitude balloons, high-altitude aircraft, or space faring satellites. Ground based deployment

is obviously the cheapest and most accessible platform since it often requires little infrastructure aside from the instrument itself. However, the solar geometry and spectral properties involved in ground based measurements can often prove constricting on the amount of useful information that can be recorded, depending on the experiment. As an example, the instrument of concern in the present work measures spectral water absorption features and, if it were deployed on the ground, the measured features would be so broad and saturated that its effectiveness would be highly reduced. This is in addition to the ground platform's inherent lack of spatial or global coverage as well. Both high-altitude balloons and aircraft offer a platform that operates at an altitude ranging from twenty kilometres to about forty kilometres. This altitude range allows for ideal observation of the troposphere and stratosphere. The advantages of a balloon platform over the aircraft platform is the lower cost and longer flight time yielding more observations for the cost. However, the aircraft typically provides a more controlled flight path which can potentially cover a much larger area than a balloon yielding higher quality measurements. Finally, satellites which are typically launched in low earth orbit provide an expensive but very long term observation platform with missions lasting years or decades. In addition, the viewing geometry provided by satellites allows for global coverage in comparison to sub-orbital methods but can come at the trade off of much lower spatial resolution.

Regardless of deployment method, indirect scientific measurements employ at least one of the following viewing methods: nadir, occultation, and limb. The nadir method works by pointing the instrument down at the ground to image a column of the atmosphere between the surface and the instrument [SPARC et al., 2017]. The nadir method clearly does not work for ground based observations but is employed by the other platforms. The occultation and limb methods are very similar in that they are both employ a limb viewing geometry [SPARC et al., 2017]. They work by imaging the atmosphere above and tangentially to the surface. In this geometry, the geographical coordinates of the point along the line of sight which is closest to the Earth's surface is called the tangent point. The altitude of this tangent point is called the tangent altitude and it defines the lowest layer of the atmosphere the line of sight is observing. In addition, a reference point and altitude refers to some representative point that can inform what multiple lines of sight are observing. Figure 2.2 below conceptually

depicts this geometry. The fundamental difference between the limb and occultation methods is that the occultation method observes the Sun (solar occultation) or another star (stellar occultation) at the end of the line of sight where as limb has no requirement for this.

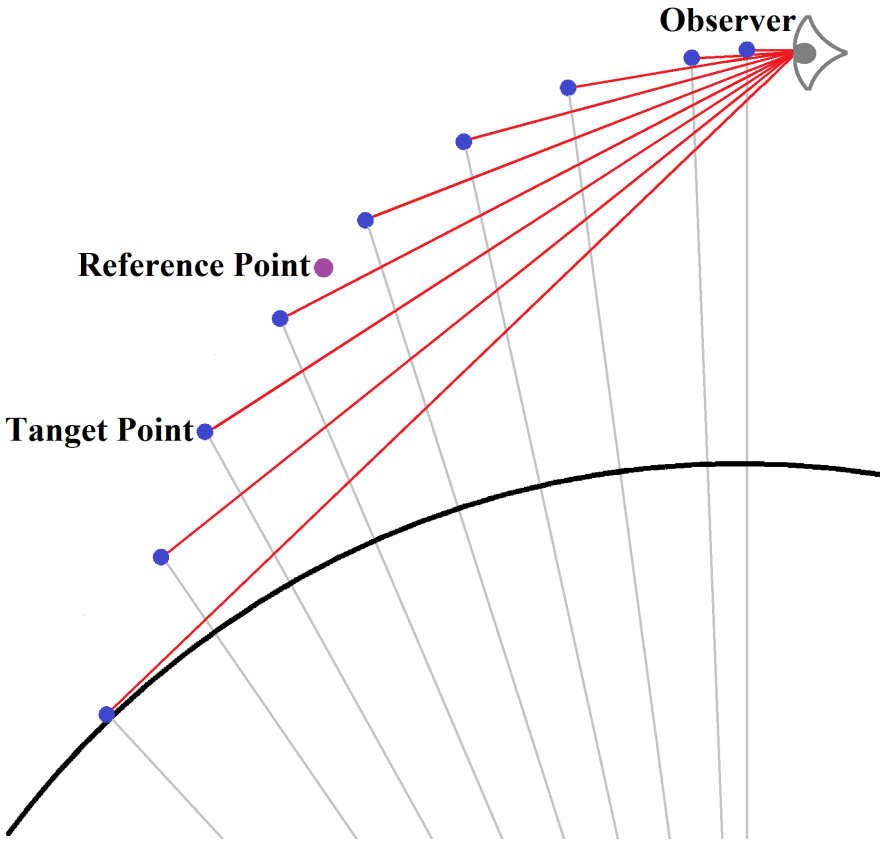

Figure 2.2: Conceptual depiction of limb viewing geometry. The tangent point is defined as the point along the line of sight closet to the the Earth. The reference point informs what multiple lines of sight are observing.

The nadir method provides high spatial resolution in comparison to the two limb geometry methods [SPARC et al., 2017] and can deal with weather effects such as clouds easier; however, it has poorer altitude resolution which can only be resolved through observation of the pressure broadening of spectral lines or through optical depth calculations [SPARC et al., 2017]. Limb geometry measurements have increased altitude resolution due to the vertically resolved lines of sight. They are also better at observing trace species of the atmosphere since the total line of sight of these geometries is longer than nadir and provides a higher

signal from the increased path length. The increased path length and the ability to avoid the abundance of signal from the lower troposphere makes the limb and occultation methods a superior choice over nadir when observing the stratosphere or upper troposphere. As for the trade between the limb and occultation methods, recall that occultation requires observation of a light source. This requirement imposes a restriction on the potential viewing geometry of the measurements which the limb method does not suffer from. Not only is pointing direction limited but the measurements can only be made while the source is right above the horizon. This limits the times of measurement to, in the case of solar occultation for example, sunrise and sunset which reduces the quantity and coverage of data. However, using a bright source like the Sun in solar occultation can provide an extremely high Signal to noise ratio (SNR) and using stellar occultation can provide measurements during nighttime.

2.3 Radiative Transfer Theory

Measurements taken by remote sensing instruments often need some extra tools to get at the desired information. In the field of atmospheric science, one of these tools *usually* takes the form of a radiative transfer model. Radiative transfer models are used in atmospheric science to interpret optical measurements of the atmosphere where the spectral signals are being used to infer some property of the atmosphere. The goal of these models is to accurately predict the photons seen by an observer given some viewing geometry and atmospheric state. Generally speaking, a user can define the chemical species which make up the atmosphere within the model along with other parameters such as the Sun's position and ground albedo to define the atmospheric state. Further supplied with lines of sight for an observer, such as an optical instrument, these models then calculate a simulated spectrum that would be seen by the observer. Accomplishing this provides a tool which can be used to infer the atmospheric state at the time of an observed measurement.

The backbone of these models is the equation of radiative transfer initially put forth by [Chandrasekhar, 1960]. This equation requires a set of five coordinates to be properly defined. The first three are the spatial position vector $\vec{r_o}$ and the last two comprise an angular vector $\hat{\Omega}$ which specifies the direction of radiation propagation at the point of $\vec{r_o}$.

This setup constructs the definition of the radiation path length, s, which is a scalar value defined as zero at the observer position ($\vec{r_o}$) and increasing along the direction of propagation specified by $\vec{r_o}$ and $\hat{\Omega}$. Essentially, the radiation path length describes a position along the observer's line of sight. Under this framework the radiative transfer equation can be written as

$$I(\vec{r_o}, \hat{\Omega}) = \bar{I}(s_1, \hat{\Omega}) e^{-\tau(s_1,0)} + \int_{s_1}^{0} J(s, \hat{\Omega}) e^{-\tau(s,0)} k(s) ds \tag{2.1}$$

where $I(\vec{r_o}, \hat{\Omega})$ is the observed radiance, $\bar{I}(s_1, \hat{\Omega})$ is the radiance at the end of the radiation path length defined by path position s_1, $\tau(s, 0)$ is the optical depth for some segment of the radiation path with respect to the observer's position, $J(s, \hat{\Omega})$ is a source term describing additional sources of radiance along the radiation path, and $k(s)$ is the extinction at a particular point along the radiation path. What this equation is saying in simple terms is that the measured radiance for an observer looking in some direction from some position is the radiance of a possible source at the very end of the line of sight attenuated through the observation medium (in this case the atmosphere) to the observer. In addition, any more radiance sources which exist along the line of sight (such as a scattering event, thermal emission, or photochemical sources) are also attenuated back to the observer and added to the measured signal. It should be noted that if the observer is looking through the atmosphere into space, as is often the case with limb measurements, then the $\bar{I}$ is zero.

Equation 2.1 mentioned two quantities, extinction and optical depth. Along the radiation path length (s) the radiance can get attenuated through absorption or scattering by the atmospheric species present. The extinction quantifies this attenuation for any single point on the radiation path and is written as

$$k(s) = \sum_i \sigma_i(s) n_i(s) \tag{2.2}$$

where the summation index i specifies an atmospheric species, $\sigma_i(s)$ is the cross section for a species, and $n_i(s)$ is the volume number density of the species. Essentially, the cross section defines the probability of scattering or absorption with the species per unit area and when multiplied with the volume number density results in amount of attenuation per unit length

caused by that species. Summing the attenuation effects for all the species present leads to the extinction at point s along the radiation path. The optical depth is simply the integral of extinction written as

$$\tau(s_a, s_b) = \int_{s_a}^{s_b} k(s)ds \tag{2.3}$$

to represent attenuation over some segment of the radiation path rather than at a point. It is through these two quantities that Equation 2.1 models the scattering and absorption for some given atmospheric state.

Multiple radiative transfer models exist such as SASKTRAN (discussed in Section 3.2), VLIDORT [Spurr, 2006], SCIATRAN [Rozanov et al., 2014], and GSLS [Loughman et al., 2015] for example. Each of these models uses different methodologies and techniques surrounding the radiative transfer equation to calculate atmospheric radiance. Due to the differences, one model may lend itself better to an application than another and it is important that a model be selected with this in mind.

2.4 The Inverse Problem

As discussed in Section 2.2 and Section 2.3, remote sensing in the context of atmospheric science needs tools to infer the state of the atmospheric property one actually wants to measure. One of these tools is *typically* the radiative transfer model, but these models only simulate atmospheric radiance and need another tool to interpret measurements taken by an optical instrument into an atmospheric property. This tool is called a retrieval algorithm and these algorithms solve what is known as the inverse problem. In the context of atmospheric science, the inverse problem is taking some atmospheric radiance and determining the atmospheric state that produced it. Note how this is inverse to what a radiative transfer model dose.

This inverse process works by making an initial guess of the atmospheric property one wants to determine. Then a radiative transfer model (also known as a forward model in this context) calculates the observed radiance using the guessed profile which is input into an instrument model (discussed more in Section 3.3) that imprints instrument effects onto the simulated data. The simulated data are then compared with the real data and an improved

estimate of the atmospheric state is produced. The new state is input back into the forward model as an updated guess profile where the whole process repeats until the newest estimate no longer changes significantly from the previous one. In addition to this, an *a priori* profile can be supplied to the process which helps regulate and solve the problem. The *a priori* profile can be the same as, or different from, the initial profile guess but should have characteristics of the expected trend.

The retrieval algorithm used to solve the inverse problem in the present work follows the non-linear optimal estimation approach put forth in *Inverse Methods for Atmospheric Sounding: Theory and Practice* [Rodgers, 2000] which is widely accepted in the field as standard. The algorithm takes the form of

$$\mathbf{x_{i+1}} = \mathbf{x_i} + (\mathbf{S_a^{-1}} + \mathbf{R}^T\mathbf{R} + \mathbf{K}^T\mathbf{S}_\epsilon\mathbf{K} + \lambda\, diag(\mathbf{K}^T\mathbf{S}_\epsilon\mathbf{K}))^{-1}[\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x_i})) - ({\mathbf{S}_\mathbf{a}}^{-1} + \mathbf{R}^T\mathbf{R})(\mathbf{x_i} - \mathbf{x_a})] \tag{2.4}$$

where the index i represents the current iteration of the inverse problem, $\mathbf{x}$ is the state vector (the desired atmospheric property) of length n, $\mathbf{x_a}$ is the *a priori* state vector of length n, $\mathbf{y}$ is the measurement vector (vectorized form of the measurement data) of length m, $\mathbf{S_a}$ is the *a priori* covariance matrix of n by n size, $\mathbf{R}$ is a regularization matrix of dimensions n by n, $\mathbf{K}$ is the Jacobian matrix of size m by n, $\mathbf{S}_\epsilon$ is the measurement covariance of size m by m, λ is a Levenberg-Marquardt type damping parameter [Marquardt, 1963], and $\mathbf{F}(\mathbf{x_i})$ is a vector the results from the forward model and instrument model with a length of m. Note that lengths n and m are arbitrary and defined by the parameters of the inverse problem.

Equation 2.4 basically describes the application of the Gauss Newton Method to the minimization of a cost function

$$\chi^2 = (\mathbf{y} - \mathbf{F}(\mathbf{x_i}))^T{\mathbf{S}_\epsilon}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x_i})) + (\mathbf{x} - \mathbf{x_a})^T{\mathbf{S}_\mathbf{a}}^{-1}(\mathbf{x} - \mathbf{x_a}) \tag{2.5}$$

which maximizes the likelihood of $\mathbf{x}$ producing $\mathbf{y}$ [Rodgers, 2000] with the addition of damping and regularization terms to stabilize the inverse problem.

The covariance matrices, $\mathbf{S_a}$ and $\mathbf{S}_\epsilon$, are constructed from the uncertainty in their respective quantities. Essentially, these allow one to weigh particular measurements or points in

the state vector more than others based on the confidence of their accuracy. λ prevents the retrieval algorithm from updating its guess of the state vector by too much in one iteration. Complementary to this, $\mathbf{R}$ provides regularization to the inverse problem which reduces oscillation in the updated state vectors. The retrieval algorithm used in the present work employs a second derivative Tikhonov type constraint such that $\mathbf{R} = \lambda {\mathbf{L_2}}^T \mathbf{L_2}$, where $\mathbf{L_2}$ is a second order difference operator of size n by n. The Jacobian $\mathbf{K}$ is defined as the sensitivity of a particular state vector point to a particular measurement vector such that $\mathbf{K} = \delta\mathbf{F}(\mathbf{x})/\delta\mathbf{x}$.

As [Rodgers, 2000] also describes, this formulation of the inverse problem allows the calculation of the gain matrix ($\mathbf{G}$) with

$$\mathbf{G} = (\mathbf{S_a^{-1}} + \mathbf{R}^T\mathbf{R} + \mathbf{K}^T\mathbf{S}_\epsilon\mathbf{K})^{-1}\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1} \tag{2.6}$$

which is a n by m matrix. The gain matrix can be used to find the solution covariance matrix ($\mathbf{S_s}$), which describes the error in the final state vector due to the measurement covariance, by

$$\mathbf{S_s} = \mathbf{G}\mathbf{S}_\epsilon\mathbf{G}^T \tag{2.7}$$

which is a n by n matrix. Finally the gain matrix can also be used in finding the averaging kernel ($\mathbf{A}$), which provides indication of state vector resolution and the influences of the retrieval algorithm on the final answer, with

$$\mathbf{A} = \mathbf{G}\mathbf{K} \tag{2.8}$$

which is a n by n matrix. These tools are extremely helpful in quantifying the precision on the reported state vector and interpreting the actual quality of the retrieval.

2.5 Water Observing Instruments

There are a number of scientific instruments that provide data sets for the study of atmospheric water vapour including, but not limited to, ACE-FTS, MIPAS, SCIAMACHY and Aura-MLS. These instruments take indirect measurements of the atmosphere which are used

in conjunction with radiative transfer models to retrieve water concentration. However, these instruments do not solely focus on water concentration in the UTLS and include observations of other species at other atmospheric regions as well. While the tactic of obtaining a larger picture of the atmosphere has its advantages, generally speaking, doing so often leads to sacrifices in the ability to observe a single species as well as possible since it is not the sole purpose of the instrument. It has been shown that current observations lack the spatial and temporal resolution within the UTLS to observe water vapour to the necessary degree [Hegglin et al., 2013].

The instrument which is the focus of the present work (SHOW) is intended to fill this empty niche by one day measuring water vapour in the UTLS to a high spatial and temporal resolution from low earth orbit. Until then, instruments like these are what is available. While these instruments have no influence on the present work, it is worthwhile to know and understand their respective missions and abilities.

ACE-FTS

ACE-FTS (Atmospheric Chemistry Experiment - Fourier transform spectrometer) is a Michelson interferometer that is part of a Canadian satellite mission which indirectly measures the Earth's atmosphere from low earth orbit. The mission overview [Bernath et al., 2005] gives the following information on the instrument along with the four goals of the endeavour:

- Understanding the chemical and dynamical processes that control the distribution of ozone in the stratosphere and upper troposphere, particularly in the Arctic.
- Exploring the relationship between atmospheric chemistry and climate change
- Studying the effects of biomass burning on the free troposphere.
- Measuring aerosols and clouds to reduce the uncertainties in their effects on the global energy balance.

To accomplish these goals ACE-FTS was given an orbit that provides coverage in the tropical, mid-latitude, and polar regions with a designed observational spectral window of 750 cm^{-1} to 4400 cm^{-1} at a resolution of 0.02 cm^{-1}. This spectral window allows for

the retrievals of a significant number of species, only one of which is water vapour. ACE-FTS takes these measurements using solar occultation providing data at sunrise and sunset at a number of tangent altitudes. After the inversion of these measurements the resulting atmospheric profiles have a vertical resolution of about four kilometres beginning at the cloud top altitude up to about one hundred and fifty kilometres.

MIPAS

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) [Fischer et al., 2008] is another Michelson interferometer designed to measure multiple atmospheric species for the study of the chemistry and physics of the upper troposphere, stratosphere, mesosphere, and lower thermosphere. This includes a focus on stratospheric ozone and STE. MIPAS flew aboard ENVISAT which is a satellite run by the European Space Agency until the mission ended in April 2012. The interferometer itself is designed to measure mid-infrared spectral emission in the limb within the range of 4.15 μm to 14.6 μm. The nature of the infrared spectrum allows measurements to be taken during both the day and the night.

The data set that MIPAS provides is used to retrieve atmospheric pressure, temperature, aerosol/cloud parameters, and roughly twenty five atmospheric species including O_3, H_2O, CH_4, N_2O, NO_2, and HNO_3. The altitude resolution of these measurements range from one to eight kilometres over an observable range of five to one hundred and sixty kilometres. Within the UTLS specifically, MIPAS is able to resolve data on a roughly three kilometre altitude grid.

SCIAMACHY

The Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) [Bovensmann et al., 1999] is a spectrometer that measures light from 240 nm to 2380 nm. Unlike both ACE-FTS and MIPAS which are designed around one viewing geometry, SCIAMACHY uses a combination of nadir, limb, solar occultation, and lunar occultation to make its measurements. Also aboard ENVISAT, the primary mission of SCIAMACHY is to study the chemistry and physics of the atmosphere on a global scale with the more particular goals of studying:

- Tropospheric pollution.
- STE
- Polar and mid-latitude stratospheric chemistry.
- Changes in atmospheric composition as a result of from natural phenomena (volcanic eruptions, solar cycle, or solar proton events).

SCIAMACHY observed a number of atmospheric species, including water vapour, from surface level up to ninety kilometres with varying altitude resolution. Within the UTLS region, the vertical resolution depends on the measurement mode. Occultation measurements of the UTLS provided a roughly 2.5 km vertical resolution while limb geometry yielded roughly three kilometres of resolution.

Aura-MLS

The Aura Microwave Limb Sounder (Aura-MLS) [Waters et al., 2006] is a microwave limb sounder on-board the Aura satellite that launched July 15th, 2004. It is part of the Earth Observing System (EOS) on Aura and is the successor to another microwave limb sounder that was part of the Upper Atmosphere Research Satellite [Barath et al., 1993]. Aura-MLS uses heterodyne radiometers to measure thermal emission in the limb. It specifically measures regions centred near 118 GHz, 190 GHz, 240 GHz, 640 GHz, and 2.5 THz. Observing these thermal emissions provides Aura-MLS information on: OH, HO_2, H_2O, O_3, HCl, ClO, HOCl, BrO, HNO_3, N_2O, CO, HCN, CH_3CN, and volcanic SO_2. By observing these species Aura-MLS pursues its scientific missions of:

- Determining if stratospheric ozone chemistry is recovering.
- Quantifying aspects of how composition affects climate.
- Studying aspects of pollution in the upper troposphere.

As for as the performance of the water vapour data product, the products made by Aura-MLS have a vertical resolution of roughly three kilometres in the UTLS. The full range of the

product goes up to an altitude of about one hundred and thirty kilometres but the resolution decreases with height.

2.6 Spatial Heterodyne Spectroscopy

SHS is an optical technique developed to measure a spectrum of light within a small passband to a high resolution without any moving parts. The basic principles of a SHS operate in a very similar way to a Michelson Interferometer but, unlike the Michelson, the SHS has fixed diffraction gratings instead of scanning mirrors. The use of diffraction gratings facilitates the frequency decomposition of light with respect to physical space by creating Fizeau fringes with a frequency dependence on λ. The fringes are measured by an imaging detector to create an interferogram with the spectrum encoded in the interference pattern. The spectrum can be recovered from the frequency patterns in the interferogram through a Fourier transform. This section will construct and explain the scientific operating principles of the SHS technique.

2.6.1 Fundamental Theory

For the purposes of this thesis, the SHS technique is throughly explained by [Harlander, 1991]. The following presents a re-derivation of fundamental SHS theory, constructed from core optical and signal processing principles, in agreement with Harlander's work.

Basic Formulation

The most basic configuration of a SHS system, shown in Figure 2.3, uses a beam splitter to send collimated light down two arms of equal length. At the end of each of these arms is an identical diffraction grating which diffracts the light. As these two sets of diffracted wavefronts pass through the beam splitter a second time and exit towards the detector they will now be crossed with an angle of 2γ between them, where γ is the angle of the grating's outgoing wavefront with respect to the optical axis. These exiting wavefronts interfere with each other to create the Fizeau fringes measured by the imaging detector, creating the interferogram. The entrance optics of the SHS are responsible for collimating the light while the exit optics are needed to focus the fringes and localize the interference onto the detector for imaging.

It is important to note that the interference is localized onto the imaging detector when the detector is imaging the diffraction gratings.

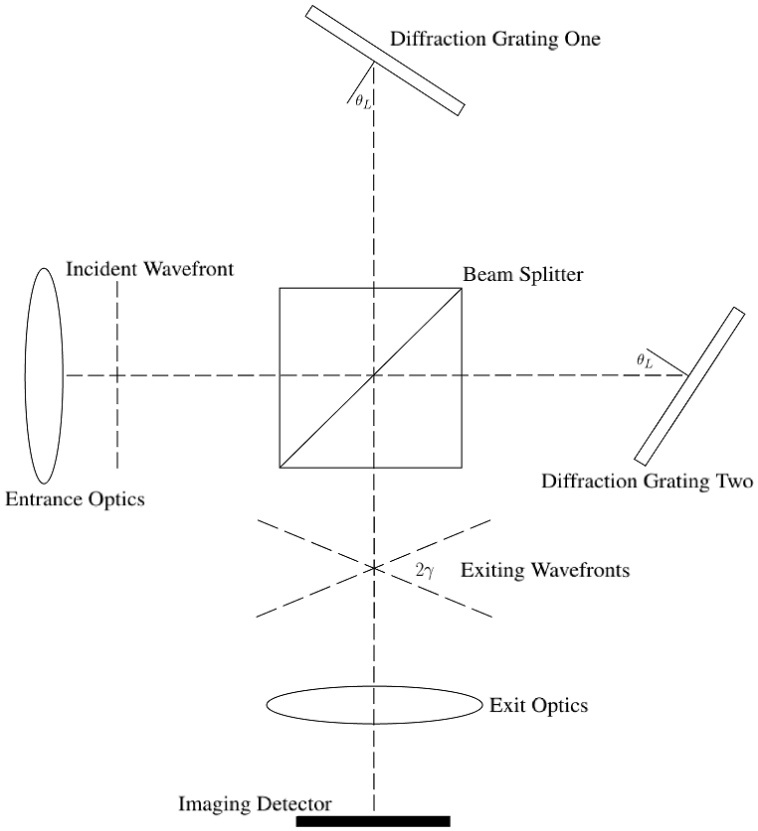

Figure 2.3: Basic SHS configuration. Input light is split by a beam splitter that travels down two different arms. At the end of each arm the light is diffracted by a diffraction grating and returns to the beam splitter resulting in crossed wavefronts. As the light reaches the imaging detector the crossed wavefronts form Fizeau fringes which are measured by the detector.

To understand how the interferogram contains information about spectra, begin with the definition of the Poynting vector ($\mathbf{S}$) expressed as

$$\mathbf{S} = \frac{1}{\mu_o}(\mathbf{E} \times \mathbf{B}) \tag{2.9}$$

where μ_o is absolute permeability equal to $4\pi \times 10^{-7} Hm^{-1}$. Assuming plane waves, $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields of a monochromatic EM wave given by

$$\begin{aligned} \mathbf{E} &= E_o \cos(kz - \omega t)\hat{\mathbf{i}} \\ \mathbf{B} &= B_o \cos(kz - \omega t)\hat{\mathbf{j}} \end{aligned} \tag{2.10}$$

in which E_o and B_o are the magnitudes of the electric and magnetic fields respectively, k is the propagation constant equal to $2\pi/\lambda$ where λ is wavelength, z is the position, t is time, and ω is angular frequency. Equation 2.9 describes the energy flow per unit time per unit area of an EM wave but the property of light measured by the imaging detector is the intensity (I). The intensity is calculated by taking the time average of the Poynting vector. Substituting Equation 2.10 into Equation 2.9 gives $\mathbf{S} = (E_i)(B_j)/\mu_o\hat{\mathbf{k}}$ which results in the time average becoming

$$I =< |\mathbf{S}| >= \frac{1}{\mu_o t}\int_0^t (E_o \cos(kz - \omega t'))(B_o \cos(kz - \omega t'))dt' \ . \tag{2.11}$$

Using $B_o = E_o/c$ derived from Faraday's law of induction, where c is the speed of light in vacuum, to equate the magnitudes of the electric and magnetic fields, in addition to, applying $c = (\mu_o\varepsilon_o)^{-1/2}$ where ε_o is absolute permittivity, reforms Equation 2.11 into

$$I = \frac{\varepsilon_o c}{t}\int_0^t E^2 dt \ . \tag{2.12}$$

Note that Equation 2.12 was derived using only one EM wavefront and that the solution for a single wavefront is $I = E_o^2\varepsilon_o c/2$ using the electric field given in Equation 2.10. However, in the case of the SHS there are two wavefronts, one from each arm, incident on the detector at the same time. Incorporating the second wavefront into Equation 2.12 yields

$$I = \frac{\varepsilon_o c}{t}\int_0^t (E_1 + E_2)^2 dt = \frac{\varepsilon_o c}{t}\int_0^t E_1^2 + E_2^2 + (2)E_1E_2 dt \ . \tag{2.13}$$

The E_1^2 and E_2^2 terms in Equation 2.13 both contribute a $E_o^2\varepsilon_o c/2$ respective to their electric fields with the remaining term describing the interaction between the two wavefronts. Within the context of an ideal SHS system both wavefronts would have the same angular frequency, same initial intensity, and same initial phase because they both originated from the same initial input source. Under these conditions the interaction between the wavefronts

will depend entirely upon the different optical paths taken between them. Applying the electric field from Equation 2.10 to carry out the integration Equation 2.13 becomes

$$I = 2I_o(1 + \cos(kz_1 - kz_2)) \tag{2.14}$$

where I_o is equal to $E_o^2 \varepsilon_o c/2$. Note that Equation 2.14 assumes polarization has no effect on the intensity (linear or unpolarized light) and that the optical components within the interferometer are ideal, as in, the transmittance and reflectance of the components are such that there is equal intensity distribution between the two arms. Furthermore, this equation does not take into account that half of the light from the initial input source will exit back out the front via the return pass through the beam splitter.

Equation 2.14 shows that the intensity measured by the imaging detector for a monochromatic wave will depend on a constant dictated by the initial source intensity as well as the phase difference induced by the optical path. To better reflect this relation within a SHS this equation can be reformed into

$$I = \int_0^\infty B(\sigma)(1 + \cos(2\pi\sigma(u)))d\sigma \tag{2.15}$$

where σ is the wavenumber equal to $1/\lambda$, $B(\sigma)$ is the spectral density of a particular σ given by the initial source taking into account the 1/2 loss of the SHS, and u is the optical path difference induced by $z_1 - z_2$. The integral over σ adds together the contribution of each wavenumber being radiated from the input source.

Diffraction Gratings and Optical Path Difference

To find the expression for the optical path difference the SHS will impart upon the wavefronts as a result of travelling down the different arms begin by looking at the general diffraction grating equation

$$\sigma[\sin(\theta_i) + \sin(\theta_m)] = \frac{m}{d} \tag{2.16}$$

where m is the order of diffracted light, $1/d$ is the grating groove density, θ_i is the angle the incoming wavefront makes with the grating's normal axis, and θ_m is the angle the outgoing

wavefront of order m makes with the normal. An important note to make is that this form of the diffraction grating equation assumes that angles can be negative or positive depending on what side of the grating's normal axis the angle is on. It does not particularly matter which side of the normal is defined as positive or negative as long consistency is maintained. A diagram of the grating geometry within a SHS is shown in Figure 2.4.

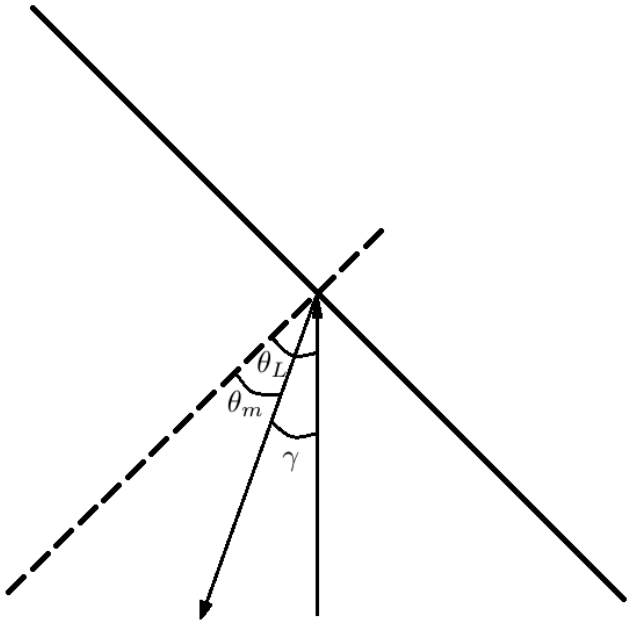

Figure 2.4: Diffraction grating geometry in a SHS. The optical axis of the diffraction grating is shown by the dashed line. The optical axis of the instrument is shown by the incoming ray.

Each diffraction grating in a SHS is fixed at an equal distances from the beam splitter and is rotated by an angle called the Littrow angle (θ_L). The Littrow angle, which is chosen by design, is defined by the Littrow wavenumber (σ_L) which is the wavenumber of light that will have its first order diffracted wavefront return along the path of incidence. Using these definitions the grating equation from Equation 2.16 can be rewritten to a more useful form for a SHS

$$\sigma[\sin(\theta_L) + \sin(\theta_L - \gamma)] = \frac{1}{d} \tag{2.17}$$

where γ is the angle the outgoing wavefront makes to the optical axis. Note that this expression of the diffraction grating equation assumes that m is equal to one. This is because a basic (non-echelle) SHS system, like SHOW (Section 2.7), is designed to only observe first order diffracted light.

From Equation 2.17 two expressions can be determined. The first of these expressions can be found by setting σ to be σ_L. In this case the angle γ becomes zero by definition and the expression becomes

$$2\sigma_L \sin(\theta_L) = \frac{1}{d} \ . \tag{2.18}$$

The second expression is found by substituting the right hand side of equation 2.17 with the left hand side of Equation 2.18 and solving for γ by applying the small angle approximation after using the sum and difference identity. Note that when solving for γ, one should not forget that the value can be positive or negative depending on the sign convention with respect to the diffraction grating's normal axis. This yields

$$\gamma = \pm 2 \tan(\theta_L)(\frac{\sigma - \sigma_L}{\sigma}) \ . \tag{2.19}$$

Equation 2.18 shows the relationship between the grating groove density, σ_L, and the θ_L. Knowing or designing any two of these quantities allows for the third to be calculated. Equation 2.19 provides a means to calculate γ for any σ given a particular Littrow configuration. Notice that γ is dependent on the difference between σ and σ_L. This observation is the first piece of evidence that the SHS heterodynes the light entered into the system with respect to σ_L. That is to say, observations made by the SHS for some σ will be with respect to σ_L. This will become more clear when looking at Equation 2.24 in which σ_L was input into the system will produce zero frequency and the frequency produced by some σ will increase in frequency the further it is away from σ_L.

The expression for γ facilitates finding the optical path difference generated between the exiting wavefronts in Figure 2.3. These wavefronts, shown to more detail in Figure 2.5, create a scenario similar to that of an optical cavity in which constructive and destructive interference takes place when localized. Complete constructive interference of these wavefronts occurs when the distance transversed in the "cavity" is an integer multiple of the λ. The

dashed line in this figure represents the projection of the detector onto which the fringe pattern caused by the constructive interference is recorded; distance along this dimension is quantified by x.

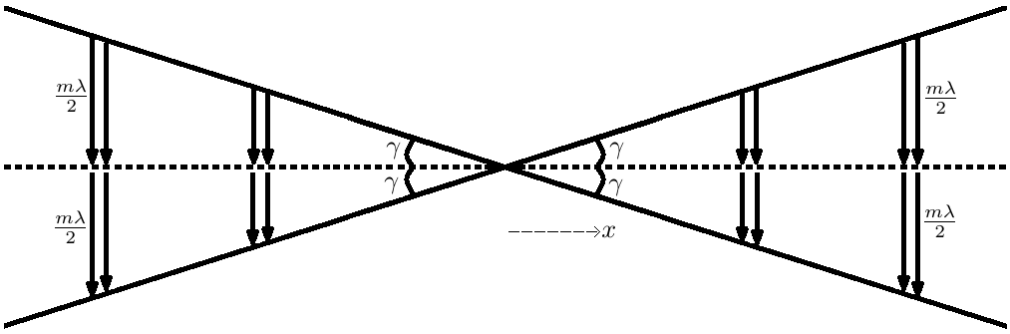

Figure 2.5: Fizeau fringe producing wavefronts. Condition for constructive interfere between the crossed wavefronts occurs only at specific spatial points.

In the case of an input σ_L, the wavefronts from each arm travel along the optical axis of the interferometer and simply add with only destructive interference due to γ being zero. As the input σ moves away from σ_L by $\delta\sigma$ such that $\sigma = \sigma_L \pm \delta\sigma$ just begins to produce fringes then γ is no longer zero and the interfering wavefronts have travelled a different path than σ_L. Using this argument and the expression for γ in Equation 2.19 it can be said that the optical path travelled through the instrument by a wavefront of wavenumber σ depends on the difference of σ from σ_L. The optical path difference (u) can be determined by calculating the extra distance travelled by a wavefront compared to the Littrow wavefront. The calculation can be done using one quarter of Figure 2.5 in which half of the optical path difference will be equal to half of the length of the cavity in which the first constructive interference occurs. This yields

$$\frac{\frac{u}{2}}{x} = \tan(\gamma) \tag{2.20}$$

and after applying the small angle approximation the substitution of Equation 2.19 gives the σ dependent optical path difference of

$$u = \pm 4x \tan(\theta_L)(\frac{\sigma - \sigma_L}{\sigma}) \ . \tag{2.21}$$

Although it can be useful to view optical path difference within a SHS as being wavenumber dependent it is not conventional to do so. Conventionally, the optical path of the SHS is defined only as $4x\tan(\theta_L)$ and the wavenumber dependence is imposed on the propagation constant such that $k = 2\pi(\sigma - \sigma_L/\sigma)$. In either framework of the optical path difference, the equation for the intensity as detected by the imaging detector can be found by substituting these findings into Equation 2.15 yielding

$$I(x) = \int_0^\infty B(\sigma)(1 + \cos(8\pi x \tan(\theta_L)(\sigma - \sigma_L)))d\sigma \ . \tag{2.22}$$

Note that the $\pm$ of Equation 2.21 can be dropped since the cosine is an even function. This equation describes the intensity measured by the imaging detector as a function of physical position (x) on the projected imaging detector. Note that any magnification caused by the exit optics needs to be taken into account when determining values for x in the equation.

Resolving Power

The ideal resolving power of a SHS system depends on the interplay between the abilities of the diffraction gratings and the ability of the imaging detector. Inherently, the source spectrum cannot be resolved any better than the resolving power of the gratings because, as far as the SHS system is concerned, it is the first limiting element in the optical chain. However, since the light undergoes a frequency decomposition (see discussions around Figure 2.5 above and Equation 2.24 below in which input light is turned into fringes of a particular frequency recorded by the detector) the detector must be able to meet the Nyquist criteria for the observed spectral window. In order to incorporate the limits of the imaging detector it is useful to have an expression which describes the frequency (f) at which the fringes will occur for a particular σ. Rearranging Equation 2.20 to solve for x instead of optical path difference yields

$$\frac{\frac{\lambda}{2}}{\sin\gamma} = \frac{x}{\cos(\gamma)} \rightarrow x = \frac{\cos(\gamma)}{2\sigma\sin(\gamma)} \tag{2.23}$$

where using the small angle approximation on γ and substituting Equation 2.19 results in an expression for x, the reciprocal of which gives the frequency of the fringes

$$f = 4\tan(\theta_L)(\sigma - \sigma_L) \ . \tag{2.24}$$

To meet the Nyquist criterion the detector must be able to spatially sample at least twice as often as the frequency caused by the furthest observable wavenumber from σ_L (σ_{max}). This defines an observable spectral range ($\Delta\sigma$) and a maximum detectable fringe frequency (f_{max}) of

$$\begin{aligned} \Delta\sigma &= |\sigma_{max} - \sigma_L| \\ f_{max} &= 4\tan(\theta_L)\Delta\sigma \end{aligned} \tag{2.25}$$

which results in the physical width of the detector (W_{det}) becoming

$$W_{det} = \frac{NM}{(2)4\tan(\theta_L)\Delta\sigma} \tag{2.26}$$

where N is the desired number spatial samples (number of pixels) contained in the imaging detector along x, and M is the magnification factor provided by the exit optics in Figure 2.3. Note that this definition of M is to project the actual size of the diffraction gratings into the perceived size as seen by the imaging detector. Knowing that each diffraction grating is rotated by θ_L from the optical axis, the required width of the illuminated diffraction gratings (W_{grat}) to capture f_{max} can be found using Equation 2.26 giving

$$W_{grat} = \frac{W_{det}}{M\cos(\theta_L)} = \frac{N}{(2)4\sin(\theta_L)\Delta\sigma} \ . \tag{2.27}$$

To relate W_{det} and W_{grat} to the resolution of the SHS start with the definition of resolving power in wavenumber space

$$R = \frac{\sigma}{\delta\sigma} \tag{2.28}$$

where $\delta\sigma$ is the system's minimum wavenumber separation that can be measured. For the SHS, $\delta\sigma$ can be found by determining how much of the full spectral range each element of the detector can resolve. To do this divide the spectral range ($\Delta\sigma$) by half the number of detector samples, recalling that the half comes from the Nyquist criterion. This results in the minimum observable wavenumber separation becoming

$$\delta\sigma = \frac{2\Delta\sigma}{N} \ . \tag{2.29}$$

Solving Equation 2.27 for $2/N$ and substituting that into Equation 2.29, then substituting this into Equation 2.28 yields an expression for the ideal limiting resolving power of a SHS system

$$R_{ideal} = 4W_{grat}\sigma\sin(\theta_L) \ . \tag{2.30}$$

It is critically important to be aware that the imaging detector images the illuminated portions of the diffraction gratings, as such, the effective width of W_{det} is limited by the width of the illuminated gratings. Likewise, if the size of the illuminated gratings is larger than that of the detector the effective width of the illuminated gratings is limited by the detector. It is this trade that marks the interplay between the detector and the diffraction grating in the resolving power of the SHS such that

$$R_{ideal\ limited} = minimum(4W_{grat}\sigma\sin(\theta_L), \frac{4W_{det}\sigma\tan(\theta_L)}{M}) \ . \tag{2.31}$$

It is worth pointing out that another trade exists between $R_{ideal\ limited}$ and $\Delta\sigma$ depending on the value of M. Say for instance that θ_L, N, and W_{det} have all been specified based on the selection of the diffraction gratings and imaging detector, then from Equation 2.26 it is clear to see that the ratio of $M/\Delta\sigma$ has to be fixed. As a result increasing M in the optical design naturally increases $\Delta\sigma$ as well which in turn lowers R as seen in Equations 2.28 and 2.29. Likewise, $\Delta\sigma$ can be sacrificed in exchange for a better R.

Fourier Transform

Since a SHS measures the spatial frequency decomposition of the source spectrum, a Fourier transform can be applied to the interferogram to recover the spectrum. Recall the definition of the Fourier transform

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} F(k)e^{2\pi ikx}dk = \int_{-\infty}^{\infty} F(k)(\cos(2\pi kx) + i\sin(2\pi kx))dk \\ F(k) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi ikx}dx = \int_{-\infty}^{\infty} f(x)(\cos(2\pi kx) - i\sin(2\pi kx))dx \end{aligned} \tag{2.32}$$

where $F(k)$ is the Fourier transform of the function $f(x)$ and $e^{2\pi ikx}$ is the kernel. It may not be immediately clear how Equation 2.22 can take the appropriate form of a Fourier transform but recall Equation 2.13. In this equation there is a constant term as a result of the two wavefronts from each arm adding together; this term needs to be removed from Equation 2.22. In practise, this constant term is removed from an interferogram by subtracting off the mean signal but mathematically it can just be dropped. This turns Equation 2.22 into

$$I(x) = \int_0^\infty B(\sigma)(\cos(8\pi x \tan(\theta_L)(\sigma - \sigma_L)))d\sigma \tag{2.33}$$

which now only depends on the modulated part of the interferogram - the part which contains the useful information about the input spectrum.

To show that Equation 2.33 is of the form required by equation 2.32 start by working back from equation equation 2.32 and break apart the integral

$$f(x) = \int_{-\infty}^\infty F(k)e^{2\pi ikx}dk = \int_{-\infty}^0 F(k)e^{2\pi ikx}dk + \int_0^\infty F(k)e^{2\pi ikx}dk \tag{2.34}$$

then apply a simple proof presented in *Introductory Fourier Transform Spectroscopy* by Bell [Bell, 1972] which shows

$$\int_{-\infty}^0 F(k)e^{2\pi ikx}dk = \int_0^\infty F^*(k)(e^{2\pi ikx})^*dk \tag{2.35}$$

when $f(x)$ is purely real. Since $I(x)$ is a purely real measurement this proof can be applied in the context of equation 2.33. As such, Equation 2.35 becomes

$$f(x) = \int_{-\infty}^\infty F(k)e^{2\pi ikx}dk = \int_0^\infty (F(k)e^{2\pi ikx} + [F(k)e^{2\pi ikx}]^*)dk \tag{2.36}$$

and as long as $F(k)$ is also purely real then it follows that

$$f(x) = \int_0^\infty F(k)cos(2\pi kx)dk = \frac{1}{2}\int_{-\infty}^\infty F(k)e^{2\pi ikx}dk\ . \tag{2.37}$$

Since $B(\sigma)$ is a real quantity Equation 2.33 can be directly applied to Equation 2.37 which, after some substitution of variables, yields

$$I(x) = \int_0^\infty B(\sigma')cos(2\pi\sigma' x)d\sigma' = \int_{-\infty}^\infty B_e(\sigma')e^{i(2\pi x\sigma')}d\sigma' \tag{2.38}$$

where $\sigma' = 4\tan(\theta_L)(\sigma - \sigma_L)$, $B(\sigma')$ is $B(\sigma)$ formulated to σ', and $B_e(\sigma') = \frac{1}{2}[B(\sigma') + B(-\sigma')]$ so that the spectral density can extend to negative wavenumbers and still have the amplitude conserved with the factor of one half. This mathematically demonstrates that a Fourier transform of the interferogram will provide a path to the input spectrum. Note that since $B(\sigma)$ (the desired quantity) and $B(\sigma')$ (the recovered quantity after a transform of Equation 2.38) are not of the same form $B(\sigma')$ needs to be converted back into $B(\sigma)$.

2.6.2 Aliasing

Aliasing is the misidentification of a signal from one or more other signals. In the specific context of a SHS, aliasing refers to the inability to determine which σ created an observed fringe pattern in the interferogram. This occurs under two circumstances, when violating the Nyquist criteria by observing a σ outside of $\Delta\sigma$ and when two σ of equal absolute difference from σ_L are input into the system. The circumstance of the Nyquist criteria is inherent to all signal sampling but the aliasing caused by the absolute wavenumber difference of σ from σ_L, hence forth called σ_L aliasing, is unique to heterodyne systems. Mathematically, the σ_L aliasing is immediately apparent in equation 2.22 due to the even nature of the cosine function acting upon the $(\sigma - \sigma_L)$ term. When present, aliasing creates major challenges to correctly reconstructing the input spectrum as the magnitude of a particular fringe pattern is created from at least two spectral sources without any knowledge of how much intensity each source contributed. The following will explain aliasing within the SHS and methods to identify and control it.

Aliasing Within the Spectral Passband

The behaviour of aliasing in a SHS depends on the arrangement of the instrument's spectral passband with respect to the spectral window (acceptable input created from $\Delta\sigma$), σ_L, and the chosen wavenumber domain (observed side of σ_L after the Fourier transform). The *direct* affect of the σ_L aliasing is that the spectra contained within the spectral window and not within the wavenumber domain gets reflected about σ_L into the wavenumber domain. However, the σ_L aliasing also has an *indirect* effect on aliasing when the Nyquist criteria is not satisfied.

The aliasing attributed to the Nyquist limit ($\Delta\sigma$ aliasing) has two sides, a "high" side which occurs when $\sigma > \sigma_L + \Delta\sigma$ and a "low" side when $\sigma < \sigma_L - \Delta\sigma$. Each of these sides behave a little differently depending on the wavenumber domain. As both the high and low sides of the $\Delta\sigma$ aliasing fold into the observed spectrum they are first shifted to the inner walls of the observed spectral window. This shifting is such that the high side will be observed on the right side of the $\sigma_L - \Delta\sigma$ window limit and the low side will be seen to the left of the $\sigma_L + \Delta\sigma$ window limit. It is now that the wavenumber domain plays a role as that choice determines which region of the spectral window gets reflected by the σ_L aliasing. As a result, the *indirect* affect of the σ_L aliasing is the reflection of whatever $\Delta\sigma$ aliasing happens to reside in the non-observed part of the spectral window. This reflection may heavily favour one side of the $\Delta\sigma$ aliasing over the other depending on the shape of the passband and the relative position of the spectral window.

Figure 2.6 shows an example of how all the possible aliasing will skew the observed spectrum for both wavenumber domains given a particular passband, like that of an optical filter, within a SHS. Notice that in this particular case, the indirect effect of the σ_L aliasing is much more prominent in the right wavenumber domain convention then it is in the left. Naturally, this is because with a right domain convention a much larger amount aliasing is reflected about σ_L.

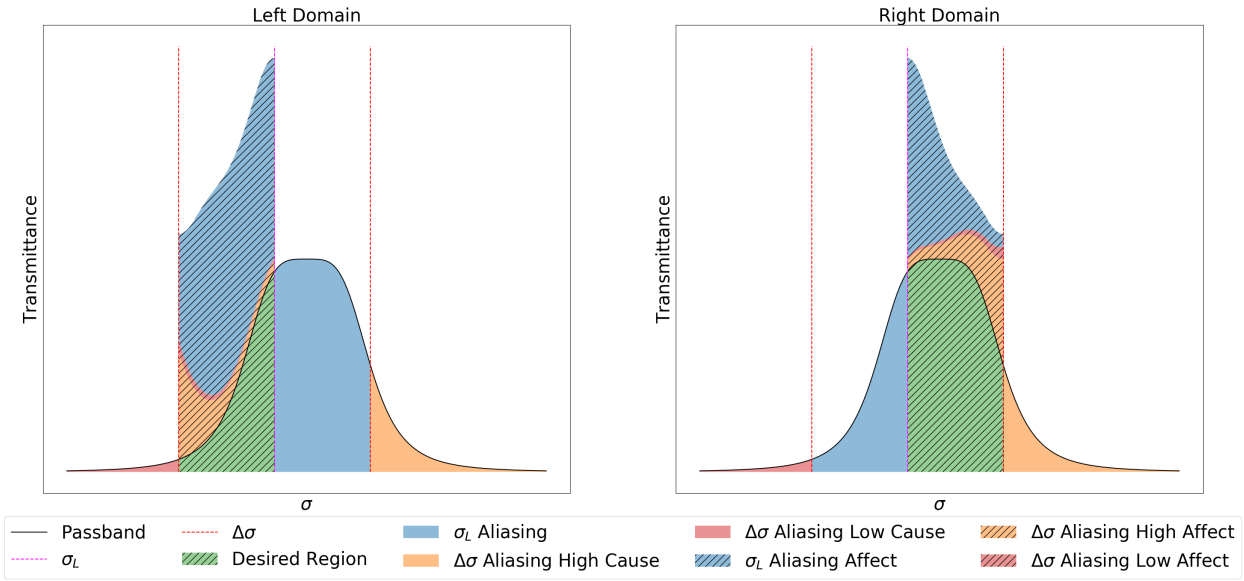

Figure 2.6: Effects of aliasing with respect to the passband and wavenumber domain.

The Two Dimensional Method

From the previous section it should be clear that the most obvious, straightforward, and generally ideal solution to deal with both kinds of aliasing is to design a passband such that it is eliminated. While this is the only method of combating $\Delta\sigma$ aliasing there is an additional method, the two dimensional method, which can deal with σ_L aliasing for some applications. Recall equation 2.22 and how the occurrence of the σ_L aliasing was the end result of constructing this equation from an one dimensional optical path difference within the cosine term beginning in equation 2.14. To understand the two dimensional method and how it changes this dynamic a more generalized approach to finding the optical path difference needs to be taken.

Begin by re-examining the wavefronts at the diffraction gratings but this time within the context of vectors. Let $\mathbf{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$ define a three dimensional propagation constant and $\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z}$ define the path of $\mathbf{k}$ such that the optical path difference is now $\mathbf{k_1} \cdot \mathbf{r} - \mathbf{k_2} \cdot \mathbf{r}$ instead of $kz_1 - kz_2$ as it was in equation 2.14. In addition to this allow for an arbitrary rotation of each diffraction grating (ϕ and ψ) about the x-axis. Figure 2.7 shows a diagram of this setup with **G1** and **G2** labelling diffraction gratings one and two respectively as well as $\mathbf{k_i}$ labelling the incoming wavefront to each grating. The reflection caused by the **G2** exiting wavefront passing through the beam splitter has been imposed in this diagram. Note that for sign conventions to be consistent, **G2** must have a counter-clockwise rotation of ψ be considered negative due to the beam splitter's reflection.

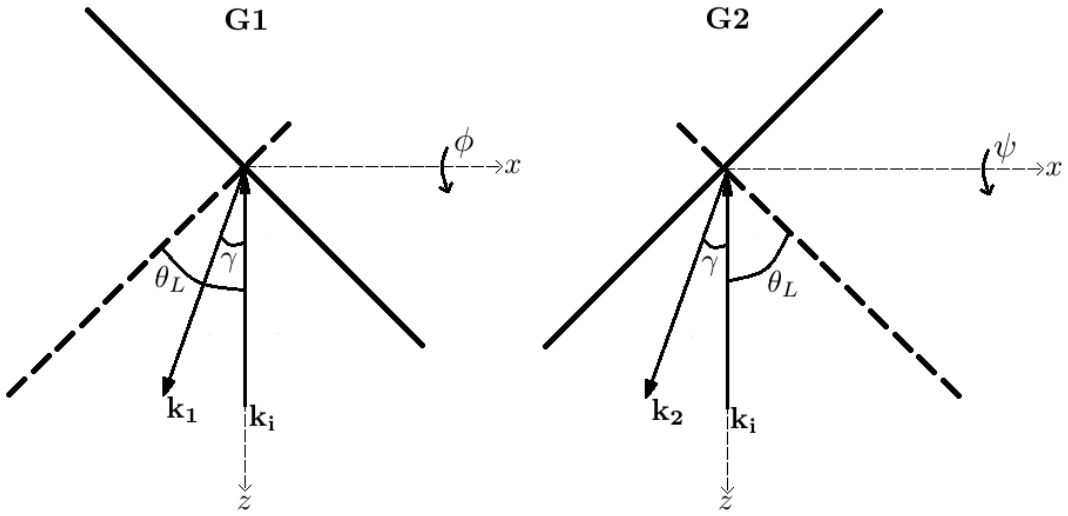

Figure 2.7: Diagram of the diffraction grating wavefronts in a vector framework.

Writing $\mathbf{k_1}$ and $\mathbf{k_2}$ in terms of their components, recalling that the x-axis rotations will result in these vectors making angles twice that of ϕ and ψ above or below the plane shown in Figure 2.7 as a result of the diffraction grating reflection, yields

$$\begin{aligned} \mathbf{k_1} &= k_1[\cos(2\phi)\sin(\gamma)\hat{x} + \sin(2\phi)\hat{y} + \cos(2\phi)\cos(\gamma)\hat{z}] \\ \mathbf{k_2} &= k_2[\cos(2\psi)\sin(-\gamma)\hat{x} + \sin(2\psi)\hat{y} + \cos(2\psi)\cos(-\gamma)\hat{z}] \end{aligned} \tag{2.39}$$

and applying small angle approximations to ϕ, ψ, and γ along with even/odd identities refines the above into

$$\begin{aligned} \mathbf{k_1} &= k_1[\gamma\hat{x} + 2\phi\hat{y} + 1\hat{z}] \\ \mathbf{k_2} &= k_2[-\gamma\hat{x} + 2\psi\hat{y} + 1\hat{z}] \ . \end{aligned} \tag{2.40}$$

Knowing that the initial light came from the same source the substitution $k_1 = k_2 = 2\pi\sigma$ can be made. In addition, define a new angle, α, which represents the relative difference between the x-axis rotations of the diffraction gratings such that $\alpha = 2(\phi - \psi)$. The optical path difference now becomes

$$\mathbf{k_1} \cdot \mathbf{r} - \mathbf{k_2} \cdot \mathbf{r} = 2\pi\sigma(2\gamma x + \alpha y) \ . \tag{2.41}$$

Entering this form of the optical path difference into the analysis done to reach equation 2.22 changes it to give

$$I(x,y) = \int_0^\infty B(\sigma)(1 + \cos(2\pi(4x\tan(\theta_L)(\sigma - \sigma_L) + \sigma y\alpha)))d\sigma \ . \tag{2.42}$$

There are a number of things to note about this equation. First, is that when substituting γ the $\pm$ can be dropped due to the sign convention being handled in the **G2** reflection. Second, if α is set to zero, as was the assumption in the previous analysis, then equation 2.22 is recovered again. Third, when the relative x-axis rotation between the diffraction gratings is nonzero then the interferogram depends on both x and y instead of just x. The y dependence will produce fringes in its dimension of a frequency (f_y) equal to

$$f_y = \sigma\alpha \tag{2.43}$$

which can be found in the exact same way equation 2.24 was except changing the perspective from the x-z plane to the y-z plane.

The fringes produced in the y overlay with the fringes in the x and result in the observed fringes being rotated by an angle ν, where

$$\tan(\nu) = \frac{f_x}{f_y} \tag{2.44}$$

and f_x is the frequency from equation 2.24. The direction of this rotation changes depending on the sign of f_x which is the result of where σ is with respect to σ_L. The end result of all of this is when α is nonzero then the σ_L aliasing can be resolved via the information stored in the direction of fringe rotation; however, a *two dimensional* Fourier transform is required to do this.

2.7 SHOW

SHOW is a SHS instrument originally prototyped at York University [Lin, 2010] as part of a Canadian Space Agency Advanced Studies Project. Later a balloon and aircraft worthy flight version of SHOW was developed by industry partner ABB also in cooperation with the

Canadian Space Agency and the University of Saskatchewan. The University of Saskatchewan now leads the SHOW project. Figure 2.8 below depicts the SHS system of SHOW which is designed to measure limb scattered sunlight to determine the atmospheric water vapour concentration in the UTLS by recording spectra between 1363 nm and 1366 nm to a resolution of ≈ 0.02 nm. This spectral region was chosen since it centres around a strong vibrational absorption band of water vapour and limb viewing geometry is used since the long term goal of the SHOW design is for eventual deployment in low earth orbit. This geometry provides high vertical resolution and along track spatial sampling which is ideal for meeting the scientific needs of the atmospheric community.

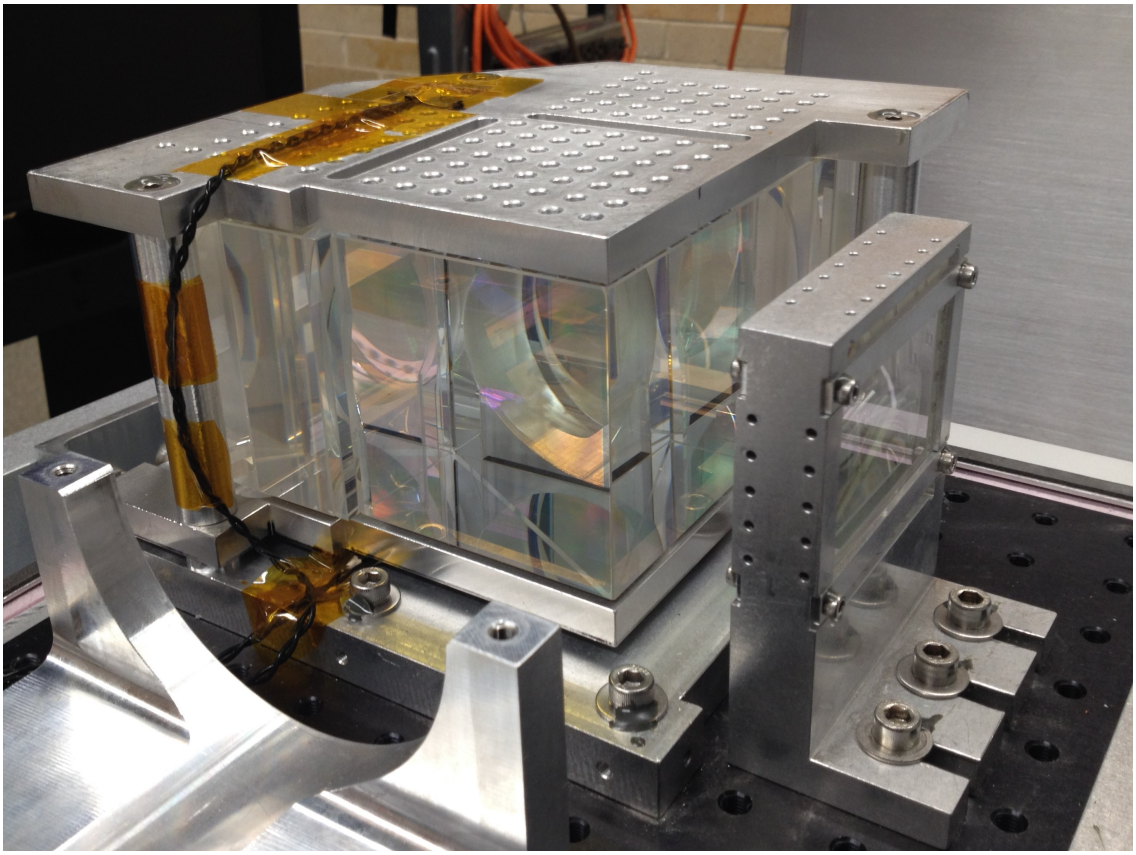

Figure 2.8: SHS system of SHOW. The large chunk of glass depicted is the monolithic design of the beam splitter and two diffraction gratings.

SHOW observes the limb through anamorphic entrance optics which collimate the light in the horizontal plane but focus the light in the vertical plane. This configuration images

the limb on the diffraction gratings and allows for the input light to be collimated in the interference plane allowing for the formation of interferograms while allowing for altitude information to still be independently resolved. In addition, this averaging collimates any structure in the field of view such as clouds. The diffraction gratings are imaged by the Raptor Photonics Owl 640 InGaAs detector inside of SHOW though a set of exit optics which provide appropriate magnification. This arrangement allows for a direct mapping of each row of the detector to a different observed tangent altitude. As such, the instrument has as many lines of sight as detector rows which image the illuminated diffraction gratings.

The Owl 640 detector in SHOW has a 640 x 512 detector array rotated so the 640 dimension is along the vertical (altitude) dimension of the instrument. The rotation was done so that each row interferogram would be read out along the same scheme as the readout electronics. This is necessary to minimize the inter-pixel variation along a row due to having different biases from the electronics that would make the interferograms much harder to interpret. In this configuration the bore sight of the instrument corresponds to the three hundred and twenty second row of the detector with the zeroth row being at the bottom. The top and bottom limits of the useful rows, or rows that imaged the illuminated gratings, are on the one hundred and ninety seventh row and the four hundred and ninety first row of the detector inclusively and respectably. This results in two hundred and ninety five usable lines of sight from the instrument and through verification done at ABB each row maps to a 0.0125° vertical field of view of the atmosphere. This results in the total vertical field of view of SHOW being 3.69°.

Table 2.1 below shows the remaining relevant specifications of SHOW. These specification come from [Langille et al., 2017] but due to practical effects like temperature variation some of these properties can vary and these values should be recognized as the baseline specifications. Note that the hight and width of the Owl 640 pixel has been modified slightly from the data sheet which states both the hight and width as being 1.46×10^{-3} cm . This was done to compensate for observed detector pixel response.

SHOW Specifications	
Littrow Wavelength	1363.715 nm
Littrow Angle	0.4975798 rad
Diffraction Grating Groove Density	7000 cm^{-1}
Relative Difference Between X-Axis Grating Rotation (α)	1.0158×10^{-4} rad
Horizontal Grating Width	4.4 cm
Vertical Grating Width	3.0 cm
Individual Transmission of Entrance & Exit Optics	0.8
Transmission of SHS	0.46
Individual Horizontal & Vertical Magnification	0.223
Individual Field Widen Half Angle Horizontal & Vertical	6.9813170×10^{-2} rad
Resolving Power	51527.84
Number of Usable Detector Pixels in the Horizontal	494
Number of Usable Detector Pixels in the Vertical	295
Individual Hight & Width of Detector Pixel	1.46×10^{-3} cm
Detector Quantum Efficiency	0.85 e^- photon^{-1}
Low Gain Detector ADU Conversion	39.67 e^- ADU^{-1}
High Gain Detector ADU Conversion	0.61 e^- ADU^{-1}
Low Gain Detector Well	6.5×10^5 e^-
High Gain Detector Well	1.0×10^5 e^-
Low Gain Detector Readout Noise	195 e^-
High Gain Detector Readout Noise	50 e^-
Detector Dark Current	5.3×10^3 e^- second^{-1}

Table 2.1: Table of typical SHS specifications for SHOW. Note that some properties change with effects like temperature.

2.8 Thesis Motivation

SHOW was installed on board the ER-2 Science aircraft for a demonstration campaign in July of 2017. SHOW as an instrument, and more broadly SHS technology, is largely unproven for the application of atmospheric measurements like water vapour. To this end, the main mission of this campaign was not to obtain scientifically useful results but to verify SHOW as being capable of measuring water vapour concentration in the UTLS within ± 1 ppm, with a vertical resolution of less than five hundred meters, and with high spatial sampling [Langille et al., 2018] as proof of the instrument concept. The long term goal of this exercise is to further develop the SHOW concept and one day provide another tool to the atmospheric community that can meet the call for high resolution and continuous observation of the UTLS.

The purpose of this thesis is to provide insight into the work required to support and fulfil the mission of the campaign. This support primarily required the development of a model to facilitate extracting the water vapour concentration from the observed spectra. To do this, a forward model capable of simulating the atmospheric radiance seen by SHOW is used in combination with an instrument model that accurately simulates the interferograms and spectra that SHOW produces. This simulation is then input to a retrieval algorithm that determines the needed state of the forward model to reproduce the observed measurements of SHOW. Through this process the water vapour concentration is determined.

In addition to the model development, there was a need to simulate the ER-2 campaign before it took place. This served two purposes, one being a test of the software in processing SHOW measurements, and two was to determine expectations and strategies for the campaign itself. Further support for the SHOW mission also involved doing instrument calibrations for SHOW, performing the analysis of the data, and reporting on the final results of the capability of SHOW. All of which is work that needed to be done in the eventual accomplishment of the end goal of the SHOW mission and to move a step closer to providing a better data set of UTLS water vapour for scientific study.

Much of the work done in this thesis has contributed to [Langille et al., 2018] and [Langille et al., 2019] but these publication are comprised of efforts from other sources as well. For

clarity, the original efforts of the present work which contributed to these two publications are:

- The overall development and construction of the model.
- Development of required supporting algorithms for the model.
- Development of the instrument model.
- Setting up, executing, and analyzing the ER-2 campaign simulation.
- The collection of the raw data for the SHOW calibrations.
- Processing and analyzing the ER-2 campaign data.

It is also worth to further clarify that the efforts listed above had the scientific support of other individuals as well. These individuals are those listed in the acknowledgements section with relation to the University of Saskatchewan and the Institute of Space and Atmospheric studies. However, particular recognition should be given to the head author of the two publications, Jeff Langille. Jeff provided support to the efforts of the present work as a co-worker and formulated the work into a cohesive publication.

Chapter 3

Modelling and Simulation

To preform simulation and analysis of SHOW data a robust model was developed. The purpose of the model is to not only accurately simulate the workings of a SHS, but to extend that into atmospheric water vapour retrievals using SHOW data. Figure 3.1 below conceptually depicts the general process by which the model produces water vapour concentration from observed spectra.

To accomplish this process atmospheric radiance must be reliably simulated and used as input to a SHS instrument model (Figure 3.1 top left). The instrument model must use the simulated radiance to produce an interferogram faithful to an interferogram of SHOW (Figure 3.1 top right). Applying a Fast Fourier Transform (FFT) to the interferograms of both SHOW and the simulation yields spectra. When the spectra is applied to a retrieval algorithm (Section 2.4) the spectra made from an initially guessed profile is updated by comparison to SHOW spectra until the modelled spectra converges (Figure 3.1 bottom left). This results in an estimation of the water vapour present at the time of the measurements (Figure 3.1 bottom right).

This chapter will go into detail the architecture used to construct the model, the forward model used to simulate atmospheric radiances, the SHS instrument model, the retrieval algorithm, and the other important tools required by the model.

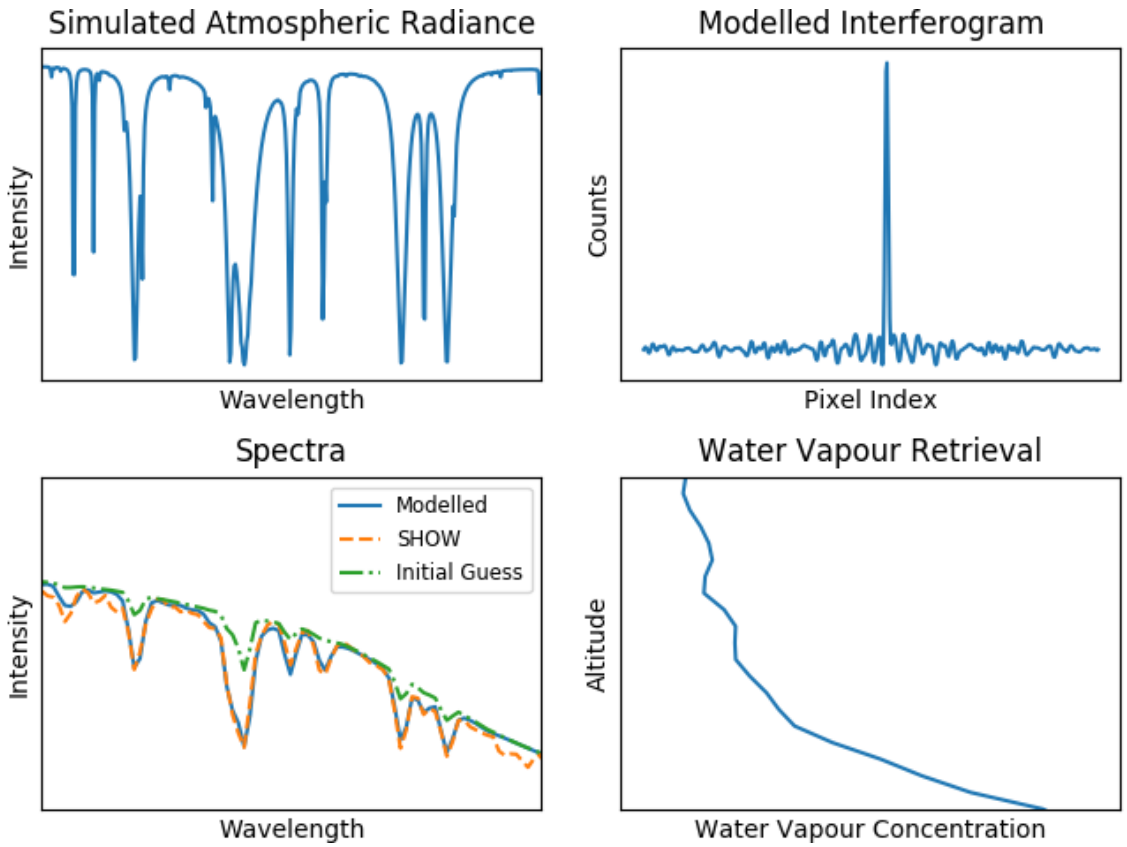

Figure 3.1: Conceptual depiction of the model process. Simulated radiance from an estimated water vapour profile is given to a instrument model to produce simulated measurements. Real measurements are compared with simulation to produce a better estimation of a water vapour profile to use in the model. This is iteratively repeated until simulation matches real, resulting in a water vapour profile which would produce the real measurement.

3.1 Modelling Approach

Here a succinct overview of the approach to the model's architecture will be presented to provide some context on the flow of the SHOW data analysis. The model is written entirely in Python 3.x and follows an object orientated design. The only exception to this is the base code of SASKTRAN which is written in C/C++ but still makes use of object orientated design principles [Bourassa, 2007]. The overall structure of the model, information flow, and

brief descriptions of the functionality of individual parts are shown in Figure 3.2 below. The portrayal in this figure is a simplified picture of the actual code but is true to the essence of the code's object layout and purpose.

As Figure 3.2 shows, the model is broken down into eight objects each with its own function. The backbone of the model is comprised of SASKTRAN, the SHS Model, and the Retrieval Algorithm. SASKTRAN is a radiative transfer model and is used to simulate atmospheric radiance following the geometry present in some SHOW measurement. This radiance is given to the SHS instrument model which has been configured to be as true to the state of SHOW at the time of the measurement as possible. The purpose of the instrument model is to artificially imprint the instrumental effects of SHOW onto the simulated radiance from SASKTRAN. Following a FFT of the modelled interferograms, the spectra is then input into the retrieval algorithm to provide an estimation of the water vapour in the atmosphere.

The Water Species object creates the measurement vectors used by the retrieval algorithm by applying normalization and formatting to the radiances. The Optimization object determines what parameters are desired by the SHS instrument model to best match the conditions SHOW experienced during a measurement. Finally, the Systematic Correction uses a sample set of real SHOW data to provide a correction which combats behavioural affects of within the detector of SHOW that are difficult to calibrate out.

Five of the six previously mentioned sections will be discussed in further detail later in this chapter with the Systematic Correction being discussed in Section 4.4.1; however, the Model Launcher and Wrapper Classes will not be discussed at all. These two sections are important to the functionality of the model but are entirely coding infrastructure to aggregate and distribute data through out the model and contain nothing more worth mentioning beyond that.

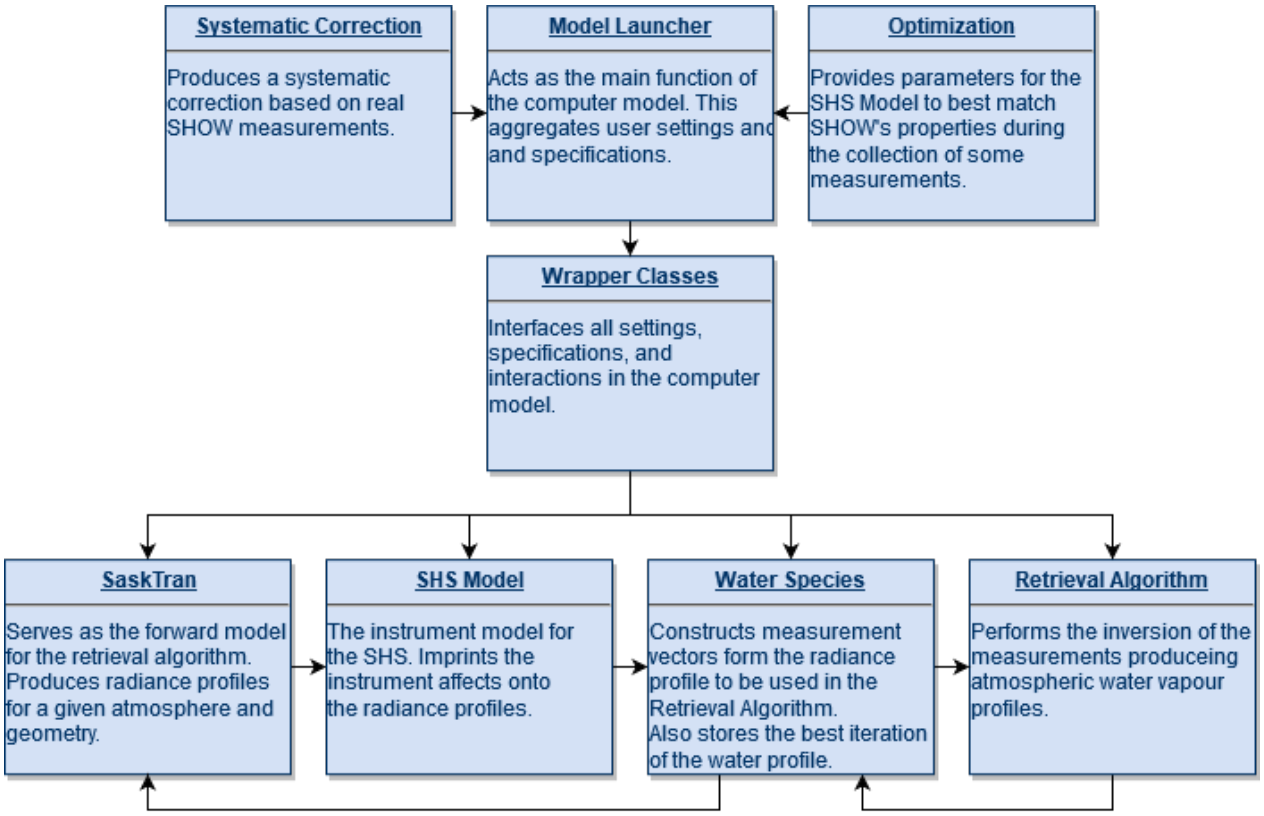

Figure 3.2: Simplified architecture of the modelling approach. The bottom four boxes display the iterative loop which results in water vapour profiles.

3.2 SASKTRAN

SASKTRAN is a radiative transfer model that was developed by A. E. Bourassa [Bourassa, 2007], D. A. Degenstein, and E. J. Llewellyn for use in processing Osiris measurements [Bourassa et al., 2008]. It employs the same radiative transfer theory described in Section 2.3 but for the purposes of this thesis the more detailed workings specific to SASKTRAN will be covered here in brief to contextualize its role as a forward model for the simulation and analysis of SHOW data.

Following the needs of Osiris, an instrument which measures light in the limb, SASKTRAN was developed specifically with limb measurements in mind making it an ideal choice for SHOW. It breaks up the atmosphere into a number of homogeneous spherical layers of

variable thickness above the earth. As stated before, the specific chemical composition and properties of these layers, such as amount of water or particle size of aerosols, can be user defined but sources such as MSIS90 and ECMWF are readily available to provide background climatology. This atmosphere is then set in a geocentric coordinate system with the $\hat{z}$ axis pointing from the Earth's centre to the North Pole and the $\hat{x}$ axis leading to the prime meridian; the $\hat{y}$ axis is set according to the conventional right hand rule. The position of the Sun can be set manually as a unit vector within the geocentric system or it can be set automatically by providing a Modified Julian Date (MJD). An observer is then placed within the model with its position and lines of sight defined within the geocentric system. Over a wavelength region of interest, SASKTRAN then calculates the radiance observed by the instrument following the scattering and emission properties of the defined atmospheric state, including the user defined albedo of the Earth's surface. Figure 3.3 below conceptually displays the scenario that SASKTRAN models for an observer with a single line of sight. If more then one light of sight is desired, as is the case with SHOW, the scenario is repeated using the different geometry. What makes SASKTRAN ideal for limb observations is its ability to accurately calculate different cases of photon scattering in the atmosphere including single scattering, multiple scattering, and albedo scattering.

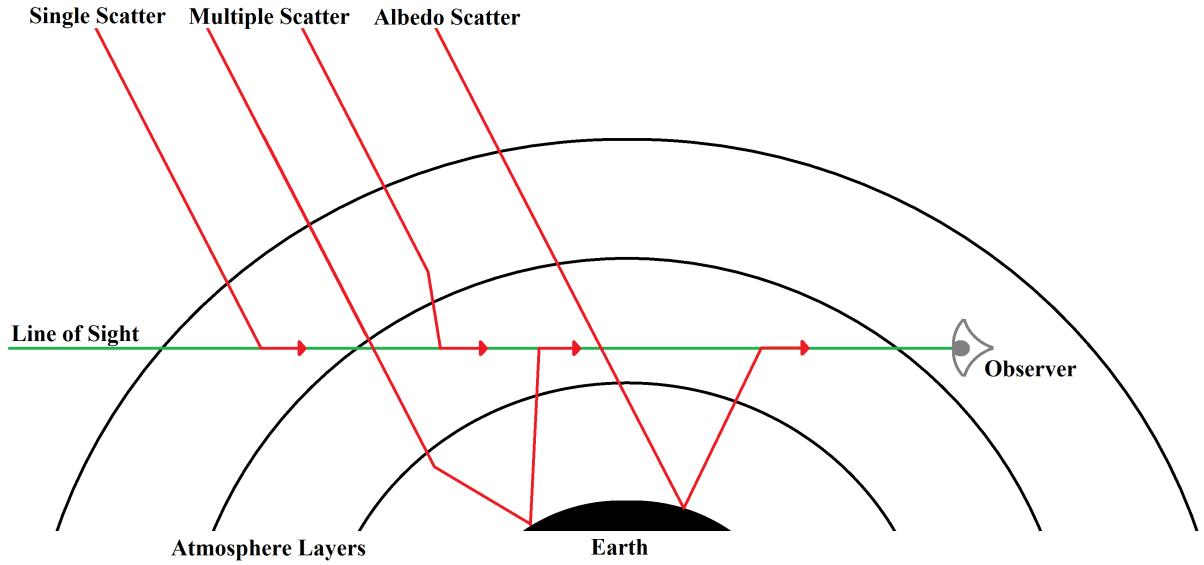

Figure 3.3: Conceptual picture of SASKTRAN modelling. Photons from different scattering cases are seen by an observer (limb observing instrument) along one of its lines of sight through different homogeneous layers of the atmosphere.

To implement the calculation of these different scattering events, SASKTRAN uses the method of successive orders. Essentially the radiative transfer equation (Equation 2.1) is recursively repeated to model a photon path of multiple scattering points before entering the observer's line of sight. Modelled light from the Sun becomes incident on the atmospheric state specified by the user and the calculation of the scattered light from this event creates a first scattering term. The term from the first incident of scattered light is then proliferated through the atmosphere again to make a second term and so on. This processes can be repeated as many times as the user desires at the cost of increased computational time.

3.3 SHS Instrument Model

The ultimate purpose of the instrument model within the context of the simulation is to accurately simulate the instrumental affects of SHOW and imprint them onto the radiances produce by SASKTRAN for the purposes of preforming the water retrieval. Examples of these instrumental affects include the resolution of the measurements, the aliasing affects imposed by the SHS method, intensity losses due to transmission coefficients of the optical components, noise caused by the electronics of the detector, affects of non-perfect optical alignment, the imprint of the filter bandpass, and various temperature dependant changes.

The SHS instrument model was designed and constructed to be as abstract as possible and as such is not inherently specific to SHOW. It allows for each design property of a SHS instrument, including detector properties, to be specified or automatically calculated to meet the user's needs. As such, the model could be used in any python modelling project involving a SHS. However, in the context of the current work it will be discussed from the perspective of modelling SHOW specifically where applicable. The following will explore capabilities of the model and how the results were achieved.

3.3.1 Interferogram Modelling

Fundamentally, the instrument model needs to accurately calculate interferogram signals (Equation 2.42, shown again below as Equation 3.1 for easy reference) to produce the interferogram measurement results of SHOW (or any other SHS instrument) in the *ideal* case.

However, to model non-ideal cases such as temperature effects within a simulation, adjustments to this equation need to be made which will be outlined here.

$$I(x,y) = \int_0^\infty B(\sigma)(1+\cos(2\pi(4x\tan(\theta_L)(\sigma-\sigma_L)+\sigma y\alpha)))d\sigma \ . \tag{3.1}$$

Recall that Equation 3.1 describes the measured intensity as a function of a position on the diffraction gratings as defined by x and y within the coordinate system of the SHS. Since x and y are functionally positions on the imaging detector (after the projection of the exit optics), Equation 3.1 has to be evaluated at every pixel on the detector. For large detector arrays this can add up to a non-trivial amount of computational time, especially when coupled with retrieval algorithms which will repeat the whole calculation many times over. Therefore, Equation 3.1 is re-written in terms of matrix operations for computational speed starting with the definitions of x and y. Based on the pixel dimensions of the detector *projected through the exit optics onto the diffraction gratings*, the arrays of x and y can be defined as

$$\begin{aligned}
\mathbf{x} &= \begin{bmatrix} -x_{\frac{w}{2}} + x_{offset} & -x_{\frac{w}{2}} + 1 + x_{offset} & \cdots & 0 & \cdots & x_{\frac{w}{2}} - 2 + x_{offset} & x_{\frac{w}{2}} - 1 + x_{offset} \end{bmatrix} \\
\mathbf{y} &= \begin{bmatrix} -y_{\frac{h}{2}} + y_{offset} & -y_{\frac{h}{2}} + 1 + y_{offset} & \cdots & 0 & \cdots & y_{\frac{h}{2}} - 2 + y_{offset} & y_{\frac{h}{2}} - 1 + y_{offset} \end{bmatrix}
\end{aligned} \tag{3.2}$$

where $x_{\frac{w}{2}}$ and $y_{\frac{h}{2}}$ are respectively the number of pixels which make up the width and height of the detector divided by two. This imposes a Cartesian grid system on the projected detector pixels in which the centre pixel is the origin. In a perfectly aligned SHS the centre pixel would correspond to the zero path difference in the horizontal ($\hat{x}$) dimension and the principle horizontal fringe in the vertical ($\hat{y}$) dimension (should $\alpha \neq 0$) which occurs at the optical axis of the instrument. The values x_{offset} and y_{offset} are shifts which move the origin of the imposed detector grid from the geometric centre pixel of the detector to whatever pixel these fringe conditions occur on thus modelling any misalignment of the imaging detector with respect to the optical axis.

It is worth noting further that the y_{offset} shift can also incorporate a pseudo-correction to the angle α. Since α is determined by the physical geometry of the diffraction gratings an

error in its calculation will be constant. Due to this an *effective* correction can be applied just by changing the value that y_{offset} is adding; however, the nature of the correction is such that it would have to be determined numerically in comparison with real interferograms to be properly applied.

The coordinate system of the projected detector in Equation 3.2 simply labels each projected pixel with an index but the required units need to be the inverse of the wavenumber units for Equation 3.1 to make sense. Therefore, the units of $\mathbf{x}$ and $\mathbf{y}$ need to be converted into a length. Additionally, as stated before, the detector images the diffraction gratings through the exit optics so the magnification of the exit optics needs to be factored into this unit conversion as well to compensate for the aforementioned projection. This is a simple matter of mapping the physical length of each pixel in both dimensions to the physical length on the diffraction gratings which it images. This can be written as

$$\begin{aligned} \mathbf{x_p} &= \mathbf{x}(\frac{W_{det}}{M_x N_x}) \\ \mathbf{y_p} &= \mathbf{y}(\frac{H_{det}}{M_y N_y}) \end{aligned} \tag{3.3}$$

where $\mathbf{x_p}$ and $\mathbf{y_p}$ are the projected pixel arrays in the $\hat{x}$ and $\hat{y}$ directions respectively, W_{det} is the physical width of the detector array in the $\hat{x}$ direction, H_{det} is the physical width of the detector array in the $\hat{y}$ direction, M_x and M_y are the magnification factors of the exit optics acting on the $\hat{x}$ and $\hat{y}$ directions respectively, and N_x and N_y are the number of pixels that makes up the detector in the $\hat{x}$ and $\hat{y}$ directions respectively.

Recall that the entrance optics of SHOW are anamorphic such that light in the horizontal is collimated but light in the vertical is focused to resolve altitude. This allows for a mapping of atmospheric tangent altitudes to detector rows via the mapping of the detector to the gratings done above. So the implementation of the affects of the entrance optics begins with determining the spectral density (previously noted as $B(\sigma)$) for each tangent altitude observed by each row using SASKTRAN. This produces a radiance matrix ($\mathbf{R}(\mathbf{y}, \boldsymbol{\sigma})$) in units of photons sr^{-1} nm^{-1} cm^{-2} s^{-1} and with dimensions equal to the number of detector rows (N_y) and the number of wavenumbers used in the SASKTRAN calculation (N_σ). This is then converted and scaled into an appropriate spectral density ($\mathbf{B}(\mathbf{y}, \boldsymbol{\sigma})$) of units ADU nm^{-1}

observed by the detector with

$$\mathbf{B}(\mathbf{y},\boldsymbol{\sigma}) = \mathbf{R}(\mathbf{y},\boldsymbol{\sigma})diag(\mathbf{F}(\boldsymbol{\sigma}))(\frac{(QE)(\tau_f\tau_b\tau_{SHS})\Omega M_x W_{pixel} M_y H_{pixel} T}{g}) \tag{3.4}$$

where QE is the quantum efficacy of the detector, Ω is the solid angle of a projected pixel, W_{pixel} and H_{pixel} are the physical widths and heights respectively of an individual pixel, τ_f and τ_b are the transmission coefficients of the front and back end optics respectively, τ_{SHS} is the transmission coefficient inherent to the SHS itself which is nominally one half due to half the input light escaping back out the front, T is the integration time of the detector's exposure, $\boldsymbol{\sigma}$ is an array comprising the wavenumbers used in the radiative transfer calculation, $\mathbf{F}(\boldsymbol{\sigma})$ is a vector created from the wavenumber dependant transmission function of the SHS's modelled filter, and g is the detector's gain or a conversion from measured electrons to analog-digital units.

Equations 3.2, 3.3, and 3.4 have formulated the quantities within Equation 3.1 to incorporate various sources of imperfections and practical parameters. All that remains is finishing the description of the matrix implementation. The term within the cosine of Equation 3.1 can easily be broken into two parts, each describing the horizontal and vertical behaviour of the interferogram (Θ_x, Θ_y respectively), and rewritten in matrix form giving,

$$\begin{aligned} \boldsymbol{\Theta}_\mathbf{x} &= 8\pi \tan(\theta_L)\mathbf{x_p}(\boldsymbol{\sigma} - \sigma_L\mathbf{1}) \\ \boldsymbol{\Theta}_\mathbf{y} &= 2\pi\alpha\mathbf{y_p}\boldsymbol{\sigma} \end{aligned} \tag{3.5}$$

where $\mathbf{1}$ denotes the identity matrix. Thus, we can define

$$\boldsymbol{\Theta}_\mathbf{xy} = \boldsymbol{\Theta}_\mathbf{x} + \boldsymbol{\Theta}_\mathbf{y} \tag{3.6}$$

but note that $\boldsymbol{\Theta}_\mathbf{x}$ has dimensions of N_x and N_σ and that $\boldsymbol{\Theta}_\mathbf{y}$ has dimensions of N_y and N_σ. Therefore, in order for Equation 3.6 to be valid and produce the correct result both $\boldsymbol{\Theta}_\mathbf{x}$ and $\boldsymbol{\Theta}_\mathbf{y}$ need to be repeated and reshaped with respect to the other so that they are both of dimensions N_x, N_y, N_σ. With this in mind Equation 3.1 can be rewritten as

$$\mathbf{I}(\mathbf{x},\mathbf{y}) = \int_0^\infty \mathbf{B}(\mathbf{y},\boldsymbol{\sigma})(1 + \cos(\boldsymbol{\Theta}_\mathbf{xy}))d\sigma \tag{3.7}$$

where the matrix inside the integral is of dimensions N_x, N_y, N_σ. The integration is then performed by the composite trapezoidal rule resulting in the interferogram matrix ($\mathbf{I}(\mathbf{x}, \mathbf{y})$) of dimensions N_x and N_y.

By inputting parameters into the model which accurately describe the state of the real SHS instrument at the time of its measurement Equation 3.7 does accurately simulate an interferogram with the same instrument affects in the actual measurement. It is worth mentioning that there are still improvements that can be made to the model that incorporate even more non-ideal effects such as a detector that is rotated with respect to the optical axis, off-axis rays, poor focus in the optical system, and other various optical aberrations or imperfections. However, as will be demonstrated later in the present work, further modelling of such affects was not necessary in the scope of this work. Such imprints were combated in the real data with various calibrations so that the present model agreed to SHOW measurements within acceptable tolerances. With that said, implementing these affects into the model remains a source of future work.

3.3.2 FFT Procedure

Included in the modelling package are the algorithms which turn the SHS interferograms into usable spectra. The algorithms have the ability to do both one dimensional and two dimensional FFTs on the interferograms. Recall that the advantage of the two dimensional FFT is its ability to resolve σ_L aliasing where an one dimensional FFT cannot. However, for the purposes of SHOW the two dimensional FFT is not useful since the altitude information resolved in the vertical is lost. To maintain the altitude resolved spectral information the fringe patterns on each individual row must be put through an one dimensional FFT.

From Equation 2.42 it is clear that there are two additive components which make up the fringe pattern on the interferogram. There is a constant term which depends only on the spectral density and a cosine term which results from the interaction between the wavefronts of the two interferometer arms. The constant term is of no interest since it contains no information about the interfering wavefronts where as the cosine term, hereby called the phase dependent term, is of interest since it contains the spectral information which can be extracted with the FFT. In a modelling case, the constant term can be removed from an

interferogram easily by calculating and subtracting off $\int_0^\infty \mathbf{B}(\mathbf{y}, \boldsymbol{\sigma})d\sigma$ but in a realistic case this cannot be done so a row by row subtraction of the mean of the full interferogram row is done instead. In either case, one ends up with an interferogram of the form

$$I(x, y) = \int_0^\infty B(\sigma)(\cos(2\pi(4x\tan(\theta_L)(\sigma - \sigma_L) + \sigma y\alpha)))d\sigma \tag{3.8}$$

which is appropriate for a two dimensional FFT but does break the ideal form of an one dimensional FFT unless $\alpha = 0$. However, as will be shown later, treating Equation 3.8 as one dimensional for the intents and purposes of the present work produces acceptable and interpretable results.

The next step is to perform an apodization on the phase dependent interferogram. This is done to smooth the discontinuities at the ends of the interferogram improving the Fourier Transform. A hanning window was chosen to do this so as to smooth the interferogram down to zero on the end points ensuring continuity in the spatial FFT. For ideally modelled data this has little affect since the fringe patterns are very clean to begin with but for noisy data this can significantly improve the quality of the spectra. Figure 3.4 below shows examples of both the one dimensional and two dimensional versions of the hanning window used in the apodization. In the case of the one dimensional window, the function is multiplied with each row of the interferogram individually. In the two dimensional case, the window is applied to the whole interferogram at once.

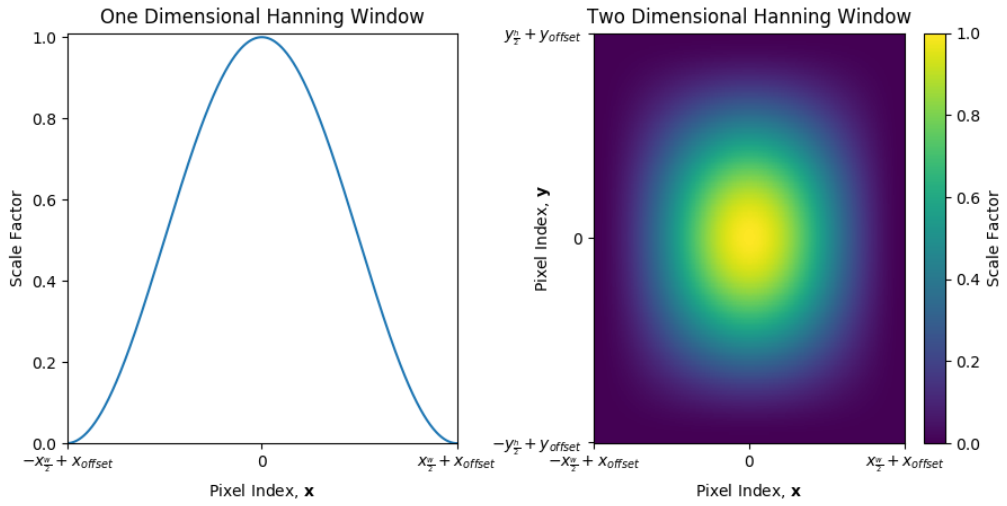

Figure 3.4: One and two dimensional Hanning windows used for apodization.

From this point extracting the spectra from the apodized interferogram is a simple matter of performing an one dimensional or two dimensional FFT as appropriate. The absolute value of this needs to be taken to incorporate any imaginary component. Then there is still work somewhat unique to a SHS remaining on identifying the wavenumbers/wavelengths of the observed spectra thus converting $B_e(\sigma')$ back into $B_e(\sigma)$ which is different from $B(\sigma)$ only by an amplitude scale factor. If desired, the exact scale factor can be found by inputting flat uniform radiance into the model, taking the FFT of the interferogram, and finding the scale factor to turn the amplitude of the resulting spectra back into one.

Recall, the expression for the max observable spatial frequency in Equation 2.25. In order to capture this frequency correctly and meet the Nyquist criteria the sample frequency has to be twice as high as this value such that

$$f_{sample} = (2)4\tan(\theta_L)\Delta\sigma \tag{3.9}$$

where f_{sample} is the sampling frequency along the horizontal ($\hat{x}$) direction. Knowing that the signal length is simply N_x, the observed spatial frequencies can then be found with

$$\mathbf{f_o} = (f_{sample}/N_x)\mathbf{x_{idx}} \tag{3.10}$$

where $\mathbf{f_o}$ is the observed frequencies along the plane measuring the Fizeau fringes, and $\mathbf{x_{idx}}$ is $\mathbf{x}$ as described in Equation 3.2 but with a x_{offset} equal to zero since now the data is being viewed in a purely signal processing context. The observed wavenumbers can be found using

$$\boldsymbol{\sigma_o} = \frac{\mathbf{f_o}}{4\tan(\theta_L)} + \sigma_L\mathbf{1} = (\frac{2\Delta\sigma}{N_x})\mathbf{x_{idx}} + \sigma_L\mathbf{1} \tag{3.11}$$

where $\boldsymbol{\sigma_o}$ is the index of wavenumbers corresponding to the spectra. It should be noted that one of the limits of the returned wavenumbers will fall short of the SHS's spectral range by one increment of the resolution due to Equation 3.2. Since there is required to be one and only one zero pixel in this index both ends of the $\mathbf{x}$ array cannot be at $x_{\frac{w}{2}}$ and will not mathematically meet the spectral range. However, in practise this is only noticeable when taking the two dimensional transform. When the one dimensional transform is taken half the returned spectrum will be a reflection of the other half due to the aliasing nature of the

SHS which cannot be resolved using α in this case.

Mathematically speaking, the FFT is being taken on purely real input so the output is Hermitian-symmetric. The negative frequency terms are redundant complex conjugates of the positive terms. When taking the one dimensional FFT it makes more sense to only look at half of the full FFT and then take care to assign the new $\mathbf{x}$ array based on the domain begin observed as shown in Figure 2.6. Depending on the domain, $\mathbf{x}$ in Equation 3.11 can be recast into a truncated form

$$\begin{aligned} \mathbf{x_R} &= \begin{bmatrix} -x_{\frac{w}{2}} & -x_{\frac{w}{2}}+1 & \cdots & 0 \end{bmatrix} \\ \mathbf{x_L} &= \begin{bmatrix} 0 & \cdots & x_{\frac{w}{2}}-1 & x_{\frac{w}{2}} \end{bmatrix} \end{aligned} \tag{3.12}$$

where $\mathbf{x_R}$ is used in a right domain case and $\mathbf{x_L}$ is used in a left domain case. The first, more two dimensionally purposed, description yielded N_x points; this description is effectively a real-valued FFT which returns $(N_x/2)+1$ points and will meet the spectral range in the domain being examined.

3.3.3 Noise

For the purposes of simulating the operation of a real SHS instrument such as SHOW with accuracy it is important to have the ability to model and quantify noise. In the context of the present work the modelling of noise is critical in simulation of SHOW campaigns to get a sense of desired SNR for the purpose of meeting retrieval precision goals. Furthermore, it provides a mechanism to check if one can accurately quantify the amount of noise in a single interferogram. This section outlines the cause of noise on the interferograms, how it can be modelled, and how it can be estimated in real measurements.

As explained in *Electronic Imaging In Astronomy - Detectors and Instrumentation* [McLean, 2008] if instrument calibrations have been applied such that relevant values for electronic read out noise and the gain can be established, then the generic noise on the measurement will have only two sources: Poisson noise (p_n, also called photon or shot noise) which is caused by the uncertainty associated to the distribution of photons over time, and the electronic readout noise (r_n) associated to the effects of operating the electronics. Both of these noise

components physically present themselves as electrons in the measurement electronics and add in quadrature giving the expression

$$\left(\frac{n}{g}\right)^2 = \left(\frac{p_n}{g}\right)^2 + \left(\frac{r_n}{g}\right)^2 \tag{3.13}$$

where n is the noise in electrons, and g is gain which converts each term into the desired units of of ADU. The left hand side of this expression is simply the variance in the measurement and, in accordance with Poisson statistics, p_n can be expressed as the square root of the mean signal. Accounting for the unit conversion between electrons and ADU this takes the form of $p_n = \sqrt{(g)s_m}$, where s_m is the signal in ADU averaged over time. Making this substitution into Equation 3.13 yields

$$v_m = \frac{s_m}{g} + \left(\frac{r_n}{g}\right)^2 \tag{3.14}$$

where v_m is the variance of the measurement in ADU.

It is worth highlighting here that the ability to model or quantify noise in a SHS interferogram depends significantly on the quality of the calibrations such as the flat field. The higher the quality of calibrations, more applicable the data sheet values for for r_n and g get, and the more physical Equation 3.14 becomes. When simulating noise on a modelled interferogram s_m is the signal mean after the dark current has been added into $\mathbf{I}(\mathbf{x}, \mathbf{y})$. The dark current is calculated to match the dark calibrations applied to any real interfeorgram that the model is simulating. Finally, the square root of Equation 3.14 is taken to find the standard deviation of the noise which is used to construct a normal Gaussian distribution that adds random noise to each pixel in $\mathbf{I}(\mathbf{x}, \mathbf{y})$. However, it should be noted that when using the instrument model with a retrieval algorithm injecting artificial noise into the model only serves to make the retrieval less precise. Instead the model should be run without adding any noise to the simulated interferograms and only use an uncertainty on the real measurement data.

To quantify the uncertainty in an interferogram, without using the quantities to construct it in the case of a simulated interferogram or if they are simply unknown as the case for a real interferogram, a SNR is calculated. Finding an estimation of the SNR in interferogram space is a simple processes shown in *Fourier Transform Spectrometry* [Davis et al., 2001].

This processes assumes that, in a two arm interferometer like a SHS, the signal at the zero path is dependent only on the intensity of the input beam and that its error is independent of path position such that

$$I_0 = I_a(0) + I_b(0) \pm \sqrt{\epsilon_a^2 + \epsilon_b^2} \tag{3.15}$$

where I_0 is the intensity at the zero path position, $I_a(0)$ and $I_b(0)$ are the intensity contributions of each arm at the zero path, and ϵ_a and ϵ_b are the errors associated with each arm added in quadrature. In this scenario the SNR simply becomes the ratio of the observed signal and the error in the measurement which is expressed as

$$SNR_x = \frac{I_0}{\sqrt{\epsilon_a^2 + \epsilon_b^2}} \tag{3.16}$$

where SNR_x is the signal to noise ratio in interferogram space.

When looking at a real or simulated interferogram in the context of SHOW the SNR is taken at a row by row basis since the interferogram is vertically resolved in altitude and each row will have independent intensity levels. As such the I_0 term for a row can simply be taken as the intensity at the zero path position for that row. The corresponding error term is made with the assumption that the error is totally dependent on Poisson statistics which is only true if corrections such as dark and readout noise have been applied. However, in this case the error can simply be taken as the square root of the signal I_0 at that row such that

$$SNR_{xi} = \frac{I_{0i}}{\sqrt{\frac{I_{0i}}{g}}} \tag{3.17}$$

where the subscript i now denotes a particular row on the interferogram. The gain factor has to be included in the square root to make the units of SNR dimensionless since the quantities are in units of ADU and not raw counts.

Equation 3.16 and 3.17 describe the SNR in interferogram space but, as said already and discussed more later, the retrieval algorithm is designed to work with SASKTRAN in spectral space. As such, calculating the SNR in spectral space is required to assign uncertainty in the measurement for the purposes of retrieving water vapour. *Fourier Transform Spectrometry*

[Davis et al., 2001] also lays out this transformation by following the power of the noise from the interferogram domain into the spectral which results in

$$SNR_\sigma = \frac{B_e(\sigma)}{\epsilon_\sigma} = \sqrt{\frac{2}{N_x}}\frac{B_e(\sigma)}{B_{e\ avg}}(SNR_x) \tag{3.18}$$

where SNR_σ is the signal to noise ratio in spectral space, ϵ_σ is the spectral error, and $B_{e\ avg}$ is the average spectral intensity. However, in the case of the SHS modifications to this formula need to be made. First a factor of square root two needs to be removed since in the initial formulation of Equation 3.18 this factor comes about by assuming only the noise in the real domain is applicable. In the case of the FFT procedure used in the present work, the absolute value of the resulting complex spectral density is taken. This incorporates the imaginary parts so the imaginary noise also needs to be accounted for by getting rid of this root two factor. Next, the spectral SNR is taken on a row by row basis for the same reason, and in the same way, as was done with the interferogram SNR. Finally, since the interferogram undergoes apodization a scale factor needs to be included to account for its effects. Applying these adjustments leads to Equation 3.18 becoming

$$SNR_{\sigma i} = A\sqrt{\frac{1}{N_x}}\frac{B_e(\sigma)_i}{B_{e\ avg\ i}}(SNR_{xi}) \tag{3.19}$$

where A is the scale factor for the apodization which a Monte Carlo style simulation found numerically to be approximately 1.6 for the hanning windows used. It was confirmed that the apodization factor was actually a result of the hanning apodizaton when the simulation was run again without the hanning window and the factor was no longer needed. Section 3.3.4 has more detail on the execution of the Monte Carlo style simulation.

3.3.4 Results

With the theory behind the SHS technique explained in Section 2.6 and its implementation to the simulation described in the other parts of this section all that remains is the evaluation of the SHS model's performance. As mentioned earlier the model is capable of doing both one and two dimensional transforms of the interferograms; however, the two dimensional transform trivially returns a full spectrum which is an effective average of all the spectra in all vertical rows. Since no vertical information can be kept it is of no use to the application of the present work and as such only the one dimensional transform will be discussed here. To begin, the model was configured with the SHOW parameters from Table 2.1 with different filter functions used to accommodate the evaluation and various forward model inputs given to examine the performance of the model.

First, the most basic scenario involving quasi-monochromatic light. In this case there is no tilt in the diffraction gratings such that $\alpha = 0$ and the quasi-monochromatic input used is at 1364.0 nm with a width of 0.02 nm. This wavelength was picked since it is very close to λ_L. From the theory, this case should produce uniform Fizeau fringes along the entire vertical dimension of the detector with a frequency predicted by Equation 2.24. This equation can be slightly modified to find the frequency in arbitrary pixels ($f_{x\ pixel}$) such that it becomes

$$f_{x\ pixel} = 4\tan(\theta_L)(\sigma - \sigma_L)\frac{W_{pixel}}{M} \tag{3.20}$$

where W_{pixel} is the physical width of a pixel on the detector. Doing this calculation yields an expected, theoretical, fringe frequency of 0.0218 pixel^{-1} which, when considering the number of pixels in the detector of SHOW, should produce eleven observed fringes. The resulting interferogram of the SHS model under these conditions is shown in Figure 3.5 which indeed shows eleven uniform fringes across the entire detector as expected.

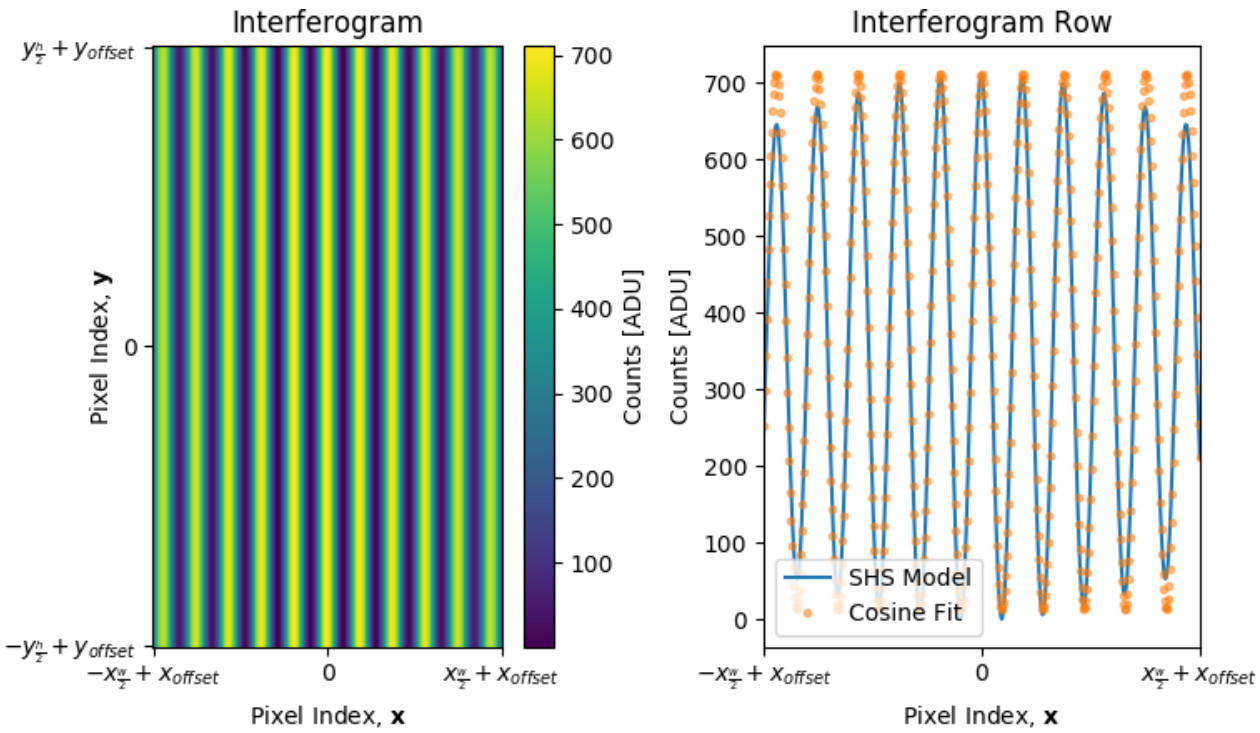

Figure 3.5: Simulated interferogram for a basic quasi-monochromatic case.

To confirm the spatial frequency of the fringes more accurately, a cosine fit was preformed on the interferogram row shown in Figure 3.5 which resulted in a fit frequency of 0.0215 pixel^{-1}. The fit frequency of 0.0215 pixel^{-1} and the theoretical frequency of 0.0218 pixel^{-1} differ by only 3.0×10^{-4} pixel^{-1} or 1.39% which shows high agreement between theory and model. In addition, notice the shape of the envelope which differs the interferogram row from the cosine fit. This envelope is caused by the quasi-monochromatic light not being of infinitely finite width. Increasing or decreasing the width of the quasi-monochromatic input will increase and decrease the envelop imprint respectively. Another consequence of using the trapezoidal integration is an under calculation of spectral magnitude for sharp gradients in the spectrum as can be seen in Figure 3.6 below. This effect is less of a concern for analysis of a continuous spectrum than it is for quasi-monochromatic light but it is still a notable limitation of the model especially around absorption features. A continuous spectrum example will be shown later to better show the limitation in a SHOW application.

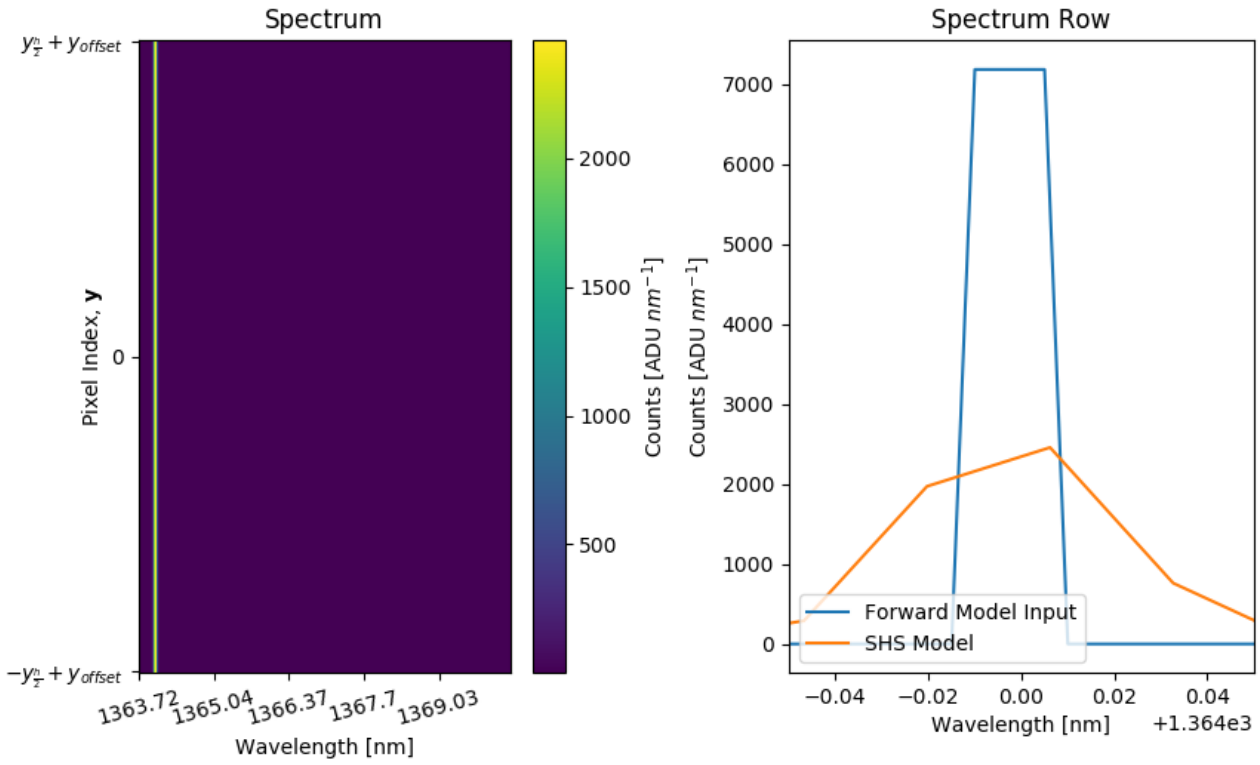

Figure 3.6: Simulated spectrum for a basic quasi-monochromatic case.

Regardless of this inaccuracy the model still identifies the input source with correct wavelength. Furthermore, you can observe the line broadening of the quasi-monochromatic light due to the resolving power of the SHS. Given SHOW specifications the expected resolution is about 0.0265 nm at this wavelength. With the input spectrum being made of light in the range of 1363.99 nm to 1364.01 nm we should expect signal to be observed in the range of 1363.9635 nm to 1364.0365 nm which is seen in the above figure.

Limitations of the numerical integration aside, the above example show the model correctly simulating the interferogram, fringe frequency, spectrum, and resolution expected from the SHS for quasi-monochromatic light. Repeating the example above but adding in a relative rotation between the diffraction gratings, such that $\alpha \neq 0$, should produce the same interferogram pattern in Figure 3.5 expect with a row dependent phase shift. The phase shift should present itself by tilting the fringes and a phase offset be added to the observed cosine curve predicted by $2\pi\sigma y\alpha$. Running this scenario through the model produced the interferogram seen in Figure 3.7 below which again shows the expected eleven fringes but

tilted as expected.

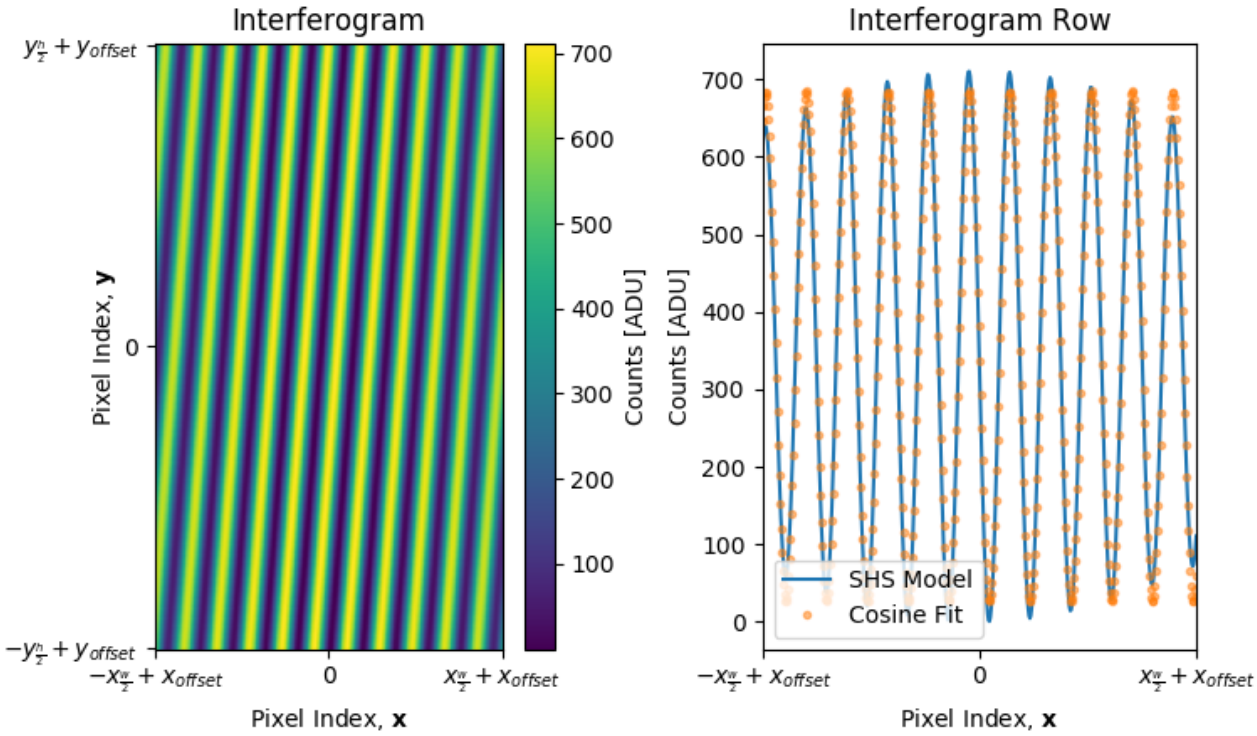

Figure 3.7: Simulated interferogram for a basic quasi-monochromatic case with alpha rotation.

The non-arbitrary row plotted in this figure is row fifty which results in a predicted phase shift of $2\pi\sigma y\alpha\frac{W_{pixel}}{M}$ equal to -1.55. The cosine fit to this interferogram has the same (and expected) fringe frequency as the $\alpha = 0$ but now included a phase shift of -1.65 which differs from the theoretical value by 0.1 or 6.25% showing reasonable agreement between theory and model.

To demonstrate that, for the purposes of SHOW, treating the two dimensional interferogram as one dimensional produces acceptable results Figure 3.8 below shows the spectrum given by the one dimensional FFT applied to each row of the two dimensional interferogram above. Notice that despite the phase shift given by the relative rotation of the gratings and the tilted interferogram that the resulting spectrum is identical to what was seen in the $\alpha = 0$ case. Note that additional complications of treating the two dimensional case as one dimensional do arise when aliasing is involved but to better quantify that aliasing in the

one dimensional case should be evaluated first.

Recall that within the SHS system there are two potential sources of aliasing: Nyquist and σ_L aliasing. The σ_L aliasing is caused by the even nature of the cosine term in equation in Equation 2.42 and, in the case where $\alpha = 0$, should present itself simply by adding the input intensities of one wavenumber domain into the other. In Figure 3.9 below a completely flat spectrum was given to the SHS instrument model which had intensities on both sides of λ_L and the one dimensional transform was taken with respect to the right domain.

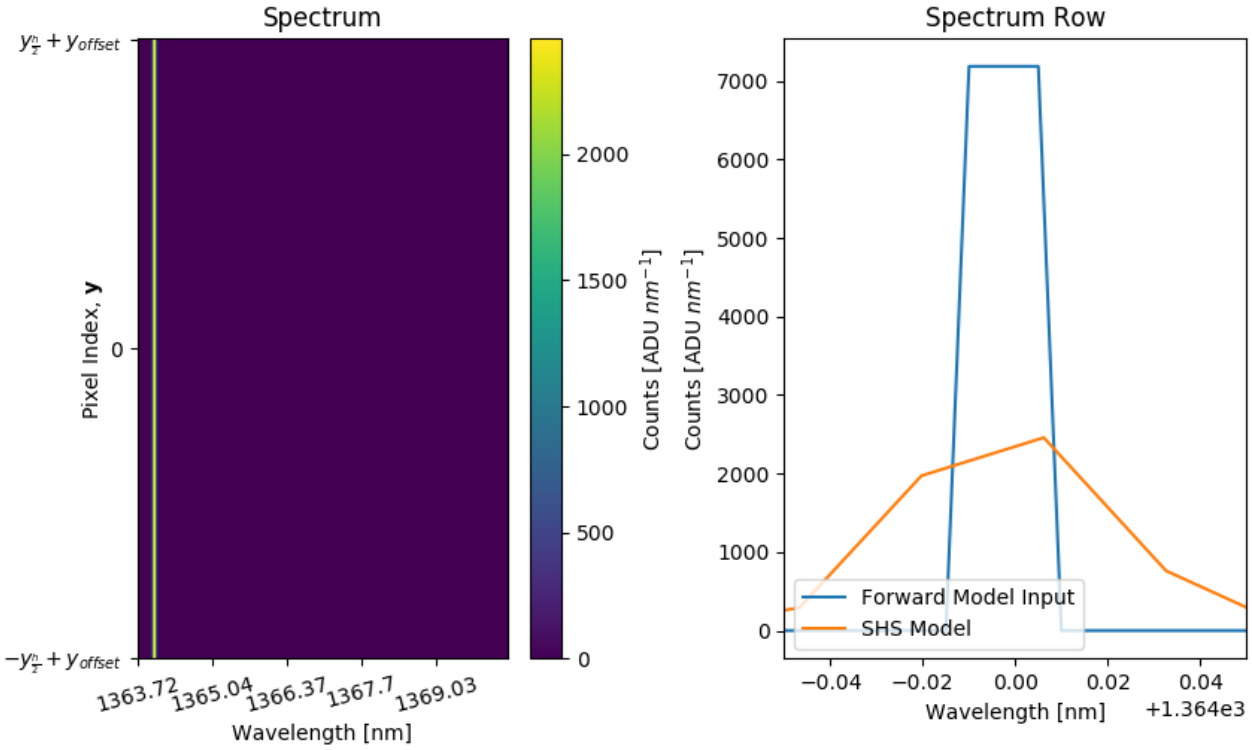

Figure 3.8: Simulated spectrum for a basic quasi-monochromatic case with alpha rotation.

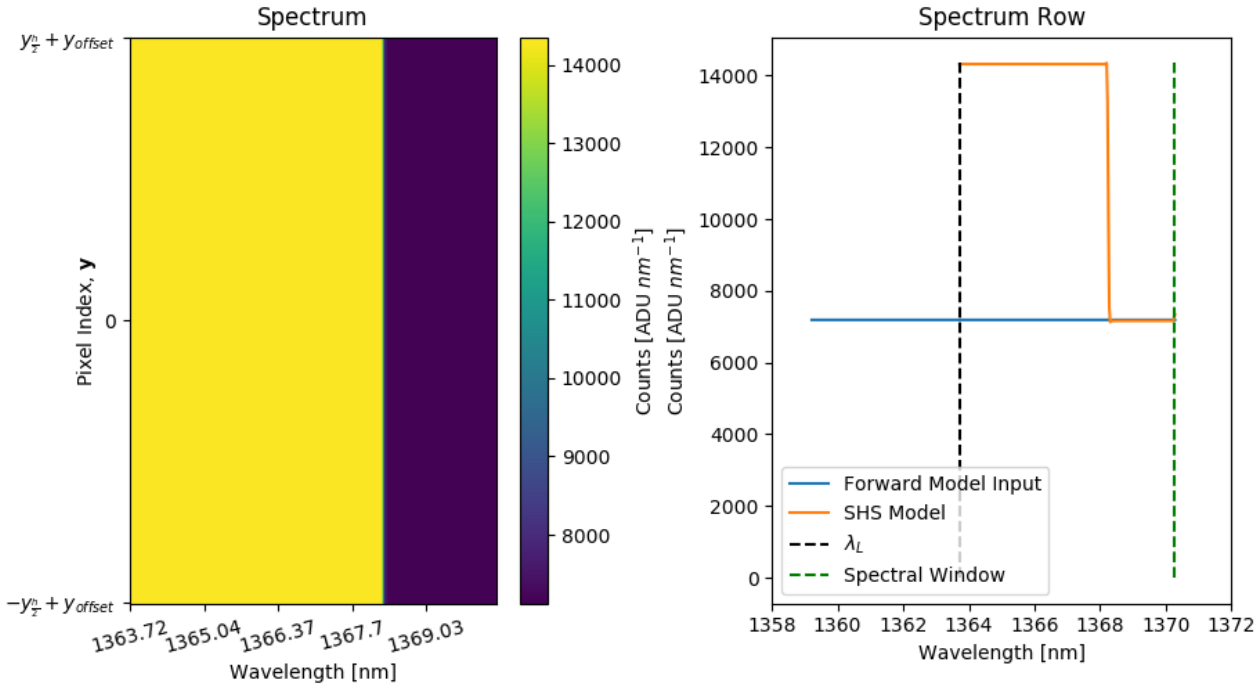

Figure 3.9: Simulated flat spectrum showing aliasing about the littrow wavelength.

As predicted by theory, the intensity contribution from the left domain is directly added onto the corresponding wavelengths in the right domain in this simple case. Likewise, the aliasing caused by violating the Nyquist criterion can also be tested in the exact same way. Figure 3.10 below again show the simple case of flat uniform radiance input into the model. This time however the input went past the Nyquist sampling limit and the one dimensional transform was taken in the right domain. From theory and basic knowledge of aliasing the expected result of violating this limit should shift the aliased signal to the right of the left Nyquist limit; however, given the σ_l aliasing of the SHS system, this signal is then reflected about σ_L such that it appears as it was reflected about the violated limit.

As Figure 3.10 shows the model again produces the result predicted by theory. Coupled with Figure 3.9 the conclusion is reached that aliasing in the SHS model is handled correctly in the one dimensional, $\alpha = 0$, case as demonstrated by the input of simple flat uniform radiance. As said before, the introduction of a non-zero α can significantly complicate the behaviour of the aliasing. The addition of the phase shift in interferogram space translates into spectral space in a sophisticated way for aliased signal. This presents itself as the aliased signal oscillating between adding and subtracting itself to the non-aliased signal based on detector row. Figure 3.11 below demonstrates this by repeating the exact same exercises of the two aliasing cases shown in Figures 3.9 and 3.10 but now with $\alpha \neq 0$. Section 3.6 will show that the oscillation of the aliasing is physical and observed in real SHOW measurements in addition to the SHS instrument model's ability to correctly simulate the phenomenon.

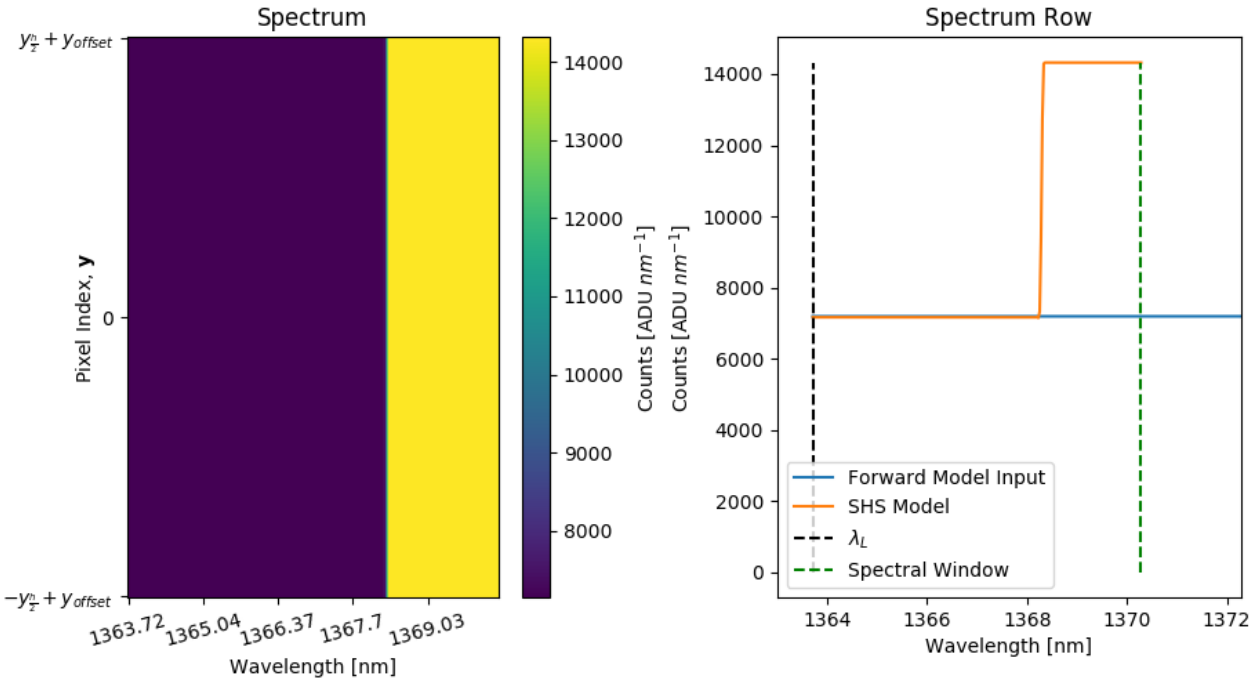

Figure 3.10: Simulated flat spectrum showing aliasing caused by the nyquist limit.

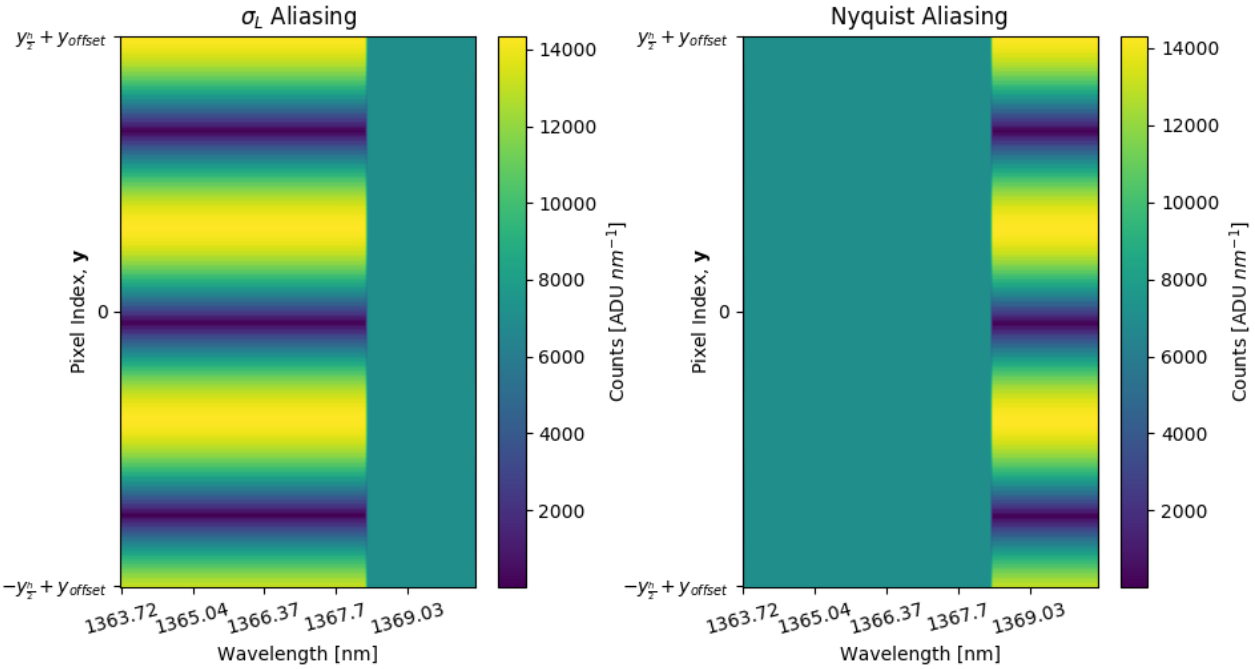

Figure 3.11: Aliasing behaviour for simulated flat spectrum when $\alpha \neq 0$. Note the vertical modulation as a result of the aliased signal.

Now that the fundamental behaviour of a SHS system has been shown to be correctly simulated by the instrument model, the model can now be used with more complicated spectra produced by SASKTRAN to observe the performance on realistic atmospheric spectra that SHOW will observe. Figure 3.12 below shows the model's ability to accurately capture the spectral features of atmospheric radiance using 0.005 nm resolution input from SASKTRAN. Notice that overall shape and intensity of the spectrum is captured very well; however, the ≈ 0.02 nm resolution of SHOW cannot capture a lot of the fine structure provided by the input. The consequence of this is the spectrum returned by the model is a version of the input signal which has been convolved down to the resolution of the instrument model.

Finally, to ensure that the theory behind the SHS instrument noise is correct, iterative testing was done using the instrument model. Flat uniform radiance was given to the SHS instrument model in a SHOW like configuration over its full spectral range and the one dimensional transform of the resulting interferogram was taken. The resulting spectra was taken as a true no noise SHS spectrum of the input. Next, at each of fifty iterations the same input was given again but the spectra produced by the SHS instrument model had random noise added in accordance with Equation 3.14. The standard deviation of the noisy (but

otherwise flat) spectra was taken and was divided out of the true flat signal on a row by row basis to calculate a SNR. This calculated SNR was divided by the theoretical SNR given by Equation 3.19 to produce a row by row ratio between the calculated and theoretical cases. The mean of this row by row ratio was taken to be reprsentative of the SNR ratio of that iteration. The results of running the full fifty iterations are shown in Figure 3.13 below where the ratio is approximantly one showing that the theory linking Equations 3.14 and 3.19 are correct for application to a SHS system.

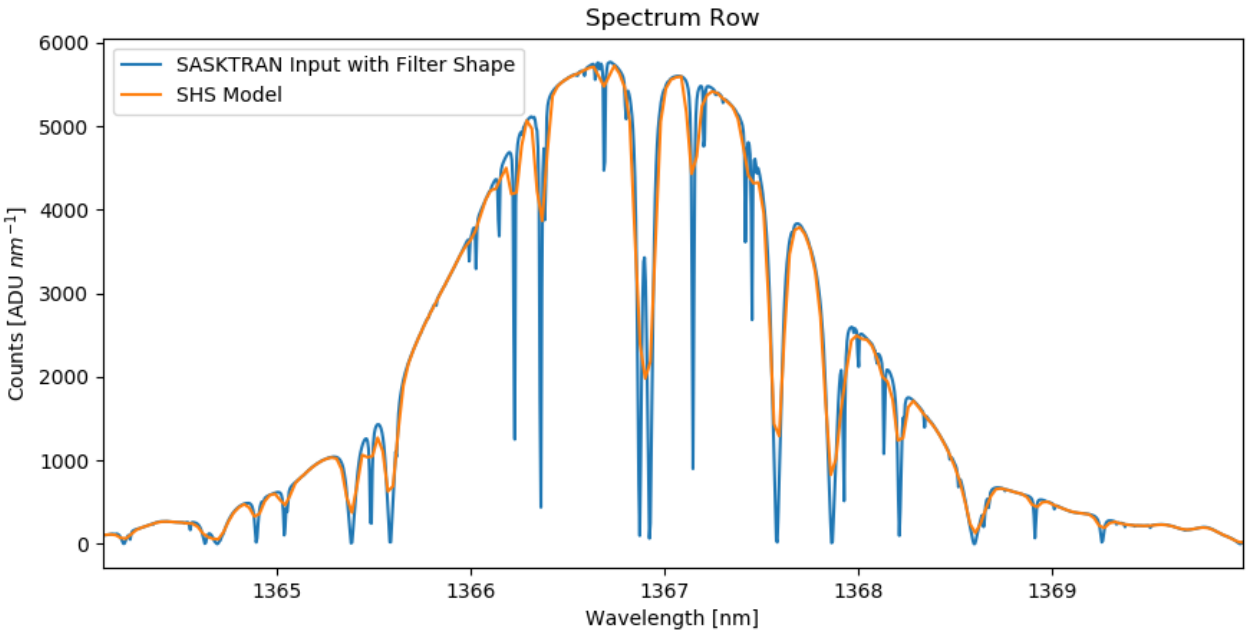

Figure 3.12: SHS model performance on atmospheric radiance.

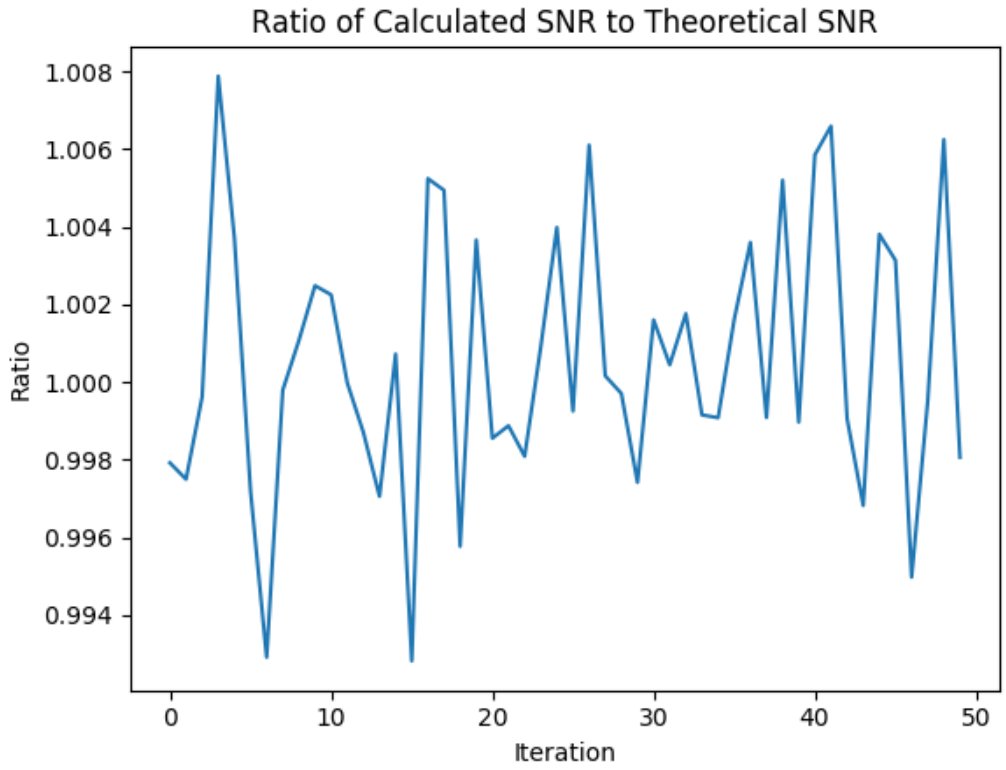

Figure 3.13: Iterative testing confirming the accuracy of modelled noise.

3.4 Retrieval Algorithm

The ultimate goal of the model is to retrieve the vertical profile of atmospheric water vapour from the spectral radiance measurements made by SHOW. As alluded to before, the basic idea behind doing this is to have a forward model simulate the atmosphere with variable water vapour. The radiance produced by the forward model is then compared against SHOW measurements and the water vapour within the forward model is updated based on this comparison. The algorithm which does this comparison is called a retrieval algorithm and the one used by the present work is the non-linear optimal estimation approach put forth in *Inverse Methods for Atmospheric Sounding: Theory and Practice* [Rodgers, 2000]. This algorithm has already been discussed in detail in Section 2.4 but it is worth making clear that the state vector, as it is in the present work, is water vapour and that the Jacobian is calculated by the used forward model - SASKTRAN.

3.5 Retrieval Measurement Vector

As discussed in Section 2.4, the retrieval algorithm requires a measurement vector $\mathbf{y}$, and a forward model output $\mathbf{F}(\mathbf{x_i})$ to solve the inverse problem. However, constructing these quantities requires some thought and processing in order to optimize the end results of the retrieval. This section will cover how these quantities are made and why they are made the way they are for the purposes of SHOW water retrievals.

The vectors $\mathbf{F}(\mathbf{x_i})$ and $\mathbf{y}$ are, for the intents and purposes of construction and formating, the same vector with their only difference being the source of the data used to make them. Vector $\mathbf{y}$ is constructed from a set of experimental spectra which measured the true state of the atmosphere and vector $\mathbf{F}(\mathbf{x_i})$ is made from the iterative output of the forward model being put through the SHS instrument model. Ideally, the iterations of the retrieval algorithm will update $\mathbf{F}(\mathbf{x_i})$ such that $\mathbf{F}(\mathbf{x_i})$ and $\mathbf{y}$ become equal within noise so the two vectors must be constructed with the intent of comparing the two. Furthermore, the data used to make these vectors is initially in the matrix format defined by the observed wavelengths and image height of size $(N_x/2) + 1$ by N_y as explained in Section 3.3.2. This adds the additional requirement that the matrix of instrument data needs to be reshaped into a vectorized format for the retrieval algorithm.

To begin ensuring that the two vectors are comparable to each other, the wavelengths used in the forward model need to make sense for the application. In a real application a SHS like SHOW will measure a continuous spectrum of light within the passband of the filter. However, a forward model like SASKTRAN will only calculate radiance for discrete wavelengths. For the best results, the forward model should calculate a high resolution discrete spectrum which is much finer than the resolution of the SHS. This allows the forward model to approximate a continuous spectrum. However, this can be computationally expensive so it is possible that only select wavelengths will be calculated by the forward model. In either case, the setup of the retrieval defines a set of wavelengths, called the state wavelengths, which will be used in the construction of both $\mathbf{y}$ and $\mathbf{F}(\mathbf{x_i})$ to ensure that they are spectrally comparable to each other. The state wavelengths should be selected to accurately capture the desired spectral features and instrument effects for the retrieval, furthermore, these wavelengths should be

directly calculated by the forward model for highest accuracy.

After the state wavelengths have been defined, the two sets of spectral data can be further refined for the retrieval vectors by selecting only particular altitudes (N_y rows) to perform a retrieval on. Doing this prevents altitudes which are too optically thin or thick to provide a good water vapour signature from being used in the inverse problem since the injection of insensitive altitudes will worsen the overall quality of the retrieval. It is difficult to determine which altitudes should be used and which should be cut since it depends on the amount of water vapour in the atmosphere, which is the quantity the retrieval is solving for. The most robust way to determine the altitudes to use is to look at the sensitivity of the Jacobian matrix during the retrieval. A second, more direct way, is to simply employ the trial and error method over multiple retrievals and observe which altitudes produce the most stable and reasonable solutions. Either way, once common wavelengths and altitudes have been established for $\mathbf{F}(\mathbf{x_i})$ and $\mathbf{y}$ the corresponding data can be extracted from the modelled and measurement spectral data sets respectively for further processing.

Once the spectral data to use has been isolated for either $\mathbf{F}(\mathbf{x_i})$ or $\mathbf{y}$ the next step is to divide out the mean of the spectrum on a row by row (altitude) basis. The reason this is done is to remove the dependence of intensity from the data and to compare only on the shape and characteristics of the spectral lines in the retrieval. The line shapes and their relative depths are directly affected by the amount of water absorption and provide a much better indicator of water vapour concentration than raw intensity values. After this the final step in making the vectors is to take the logarithm of the normalized spectral data set and then flatten the two dimensional matrix into an one dimensional vector by stacking one altitude's spectrum after the other. The logarithm is taken for two reasons. The first reason is to help linearize the inverse problem. Since the intensity of limb scattered radiance will follow the exponential trend of atmospheric density with altitude, taking the logarithm of the spectral data will turn the exponential trend into a linear trend which is easier for the retrieval algorithm to fit. The second reason is that doing the retrieval in log space will help prevent unphysical negative water densities from being a valid solution.

As one final note on the construction of $\mathbf{F}(\mathbf{x_i})$ and $\mathbf{y}$ it is that typical that one would perform a high altitude normalization on the spectral data instead of simply dividing by

the mean intensity. By dividing all the spectral data by the spectrum of the highest used altitude two things are accomplished. The first is that the relative intensity of the spectral lines are highlighted better than what is done in a mean normalization. And the second is this method can easily combat systematic errors in the data set or systematic inconsistencies between a real instrument and the instrument model simply by dividing them out. The reason this was not done for the retrievals in the present work is because of the aliasing affects demonstrated in Figure 3.11. SHOW does exhibit this aliasing behaviour and when a high altitude normalization is applied to these row dependant spectral affects it emphasizes the instrument model's ability to perfectly simulate the working state of SHOW as well as to normalize the data by aliased spectral features which are ultimately unknown. Due to this it was found during testing that just doing the mean normalization provided better results than high altitude normalization.

3.6 Instrument Optimization

As Sections 3.3.4 and 3.5 allude to, having the SHS instrument model configured to be as close to SHOW as possible is important to correctly match the instrument affects and preform the retrieval. Furthermore, as Section 2.7 indicates, the properties of SHOW are not invariant and change with temperature. While the changes are subtle they result in large impacts on the measured interferograms, particularly with respect to aliasing, which need to be accounted for in accurately modelling the instrument. To do this a method needed to be developed which minimizes the difference between the state of SHOW during a window of measurements and the SHS instrument model, thus optimizing the model to the true state of the real instrument. This section will discuss this method with an example of its application with lab measurements of SHOW.

The idea behind this optimization method is to model a high resolution spectrum which is representative of the spectrum SHOW saw during some window of measurements and give it as input to the model. The SHS model is then iteratively run while adjusting one parameter such as the position of the filter or the α angle. On each iteration the resulting spectra produced by the model in its current configuration is tested for its similarity to the

SHOW spectra and given a score. After all desired permutations have been tested the model configuration which produced the most realistic spectra is reported and used in simulating SHOW measurements for data taken around the window of interest.

To demonstrate this in detail, laboratory measurements were taken with SHOW using a large integrating sphere as the source input. A two dimensional FFT was taken of the recorded interferogram to provide an example of the input spectrum to SHOW. Then a high resolution simulation of the recorded spectrum was done in an attempt to reverse engineer an appropriate input to the SHS model. Figure 3.14 below shows the modelled and measured spectra plotted on top of one another for comparison. It is clear that despite the instrumental effects in the recorded spectrum that the two spectra do not match perfectly; however, for the purpose of minimizing the difference between SHOW and the model a perfect match is not required.

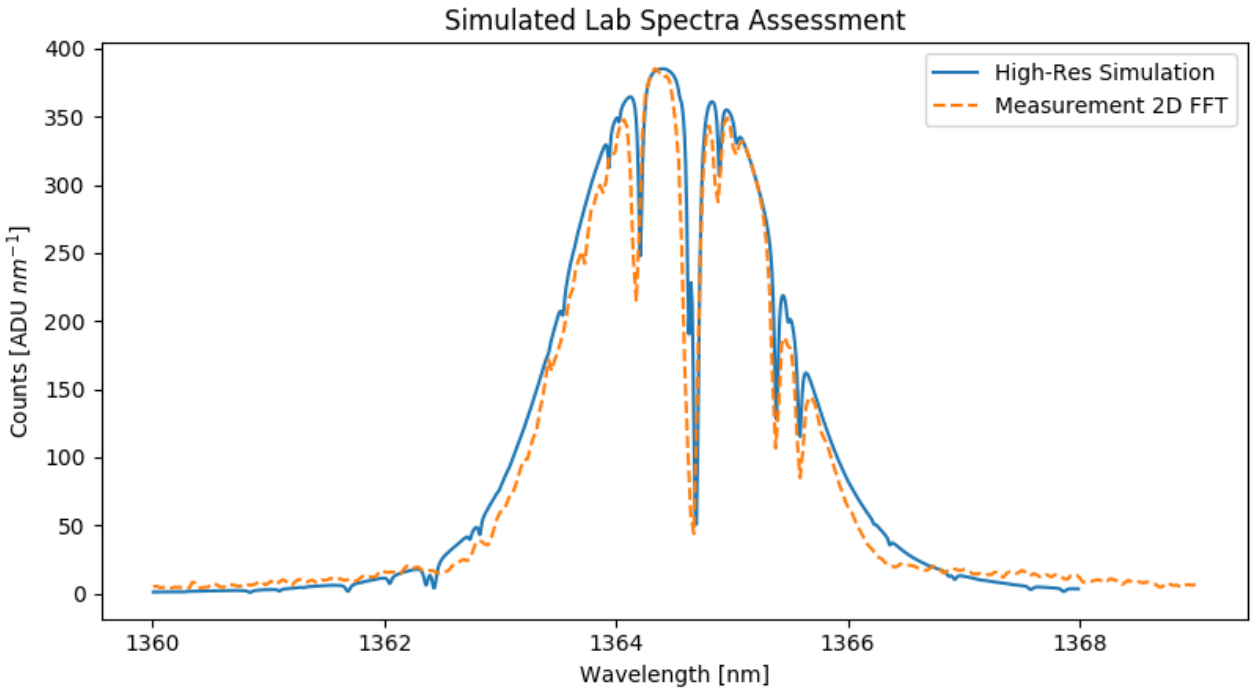

Figure 3.14: Comparison of recorded and simulated lab spectra.

The reason the two spectra do not need to be identical is because the optimization algorithm does not care about the reproduction of the initial spectrum but rather reports on the configuration that has the smallest difference from the initial spectrum. The goal being

to match aliasing affects seen in Figure 3.11, the detector offset discussed in Equation 3.2, and other such properties. This is done by testing the spectra along both dimensions. Along the spectral dimension (horizontal $\hat{x}$) the absolute difference between model's spectra and SHOW spectra is taken on a row by row basis. The mean of this difference is taken as a measure of the each row's similarity, with zero being ideal. Along the vertical ($\hat{y}$) dimension the non-absolute difference is taken on a column by column basis. The standard deviation of each column difference is taken as the measure of similarity, with zero again being ideal. The configuration which produced scores closest to zero along each dimension is recorded. If the optimal configuration for the horizontal test does not match the optimal configuration for the vertical test then the solution is not used and the whole test is run again with modified parameters to find a different configuration in which they do agree.

Figures 3.15 and 3.16 below show both the interferogram and spectra of the measured SHOW lab data taken from the large integrating sphere along side the results of the SHS instrument model after the optimization was applied. In the applied optimization one single lab measurement was reproduced using the same exposure time for the model as the actual measurement. The optimization found that the modelled filter should be shifted by -0.08 nm, α was equal to -1.0216×10^{-4} rad, and the detector needed to be shifted by 15.77 pixels horizontally and by -15.52 pixels vertically. It is important to note that the interferogram and spectral characteristics such as the zero path position, zero path angle, and aliasing modulation are virtually identical between the two. This is evidence that the optimization algorithm ensured the SHS model was in an accurate configuration compared to SHOW.

While the two figures below may show that both the interferograms and spectra look a like, it is important that a quantitative measure on the optimization's effectiveness is supplied. A direct comparison of the spectra shown in Figure 3.16 yields that the mean absolute difference between the SHOW measurements and the simulation is 13.754 ADU nm^{-1}. Since both spectra should have the same noise this absolute mean difference should be attributed entirely to the difference in the input spectra if the optimization worked. Comparing the spectra shown in Figure 3.14 shows that the mean absolute difference between them is 13.367 ADU nm^{-1} which is only 0.387 ADU nm^{-1} or 2.85% lower than the difference found between the final spectra. This suggests that the spectral difference of the final result is almost entirely due

to the input spectra and that the optimization was very successful in reproducing the state of SHOW, furthermore, that the SHS instrument model is also very capable of being an accurate virtual substitute for the real instrument.

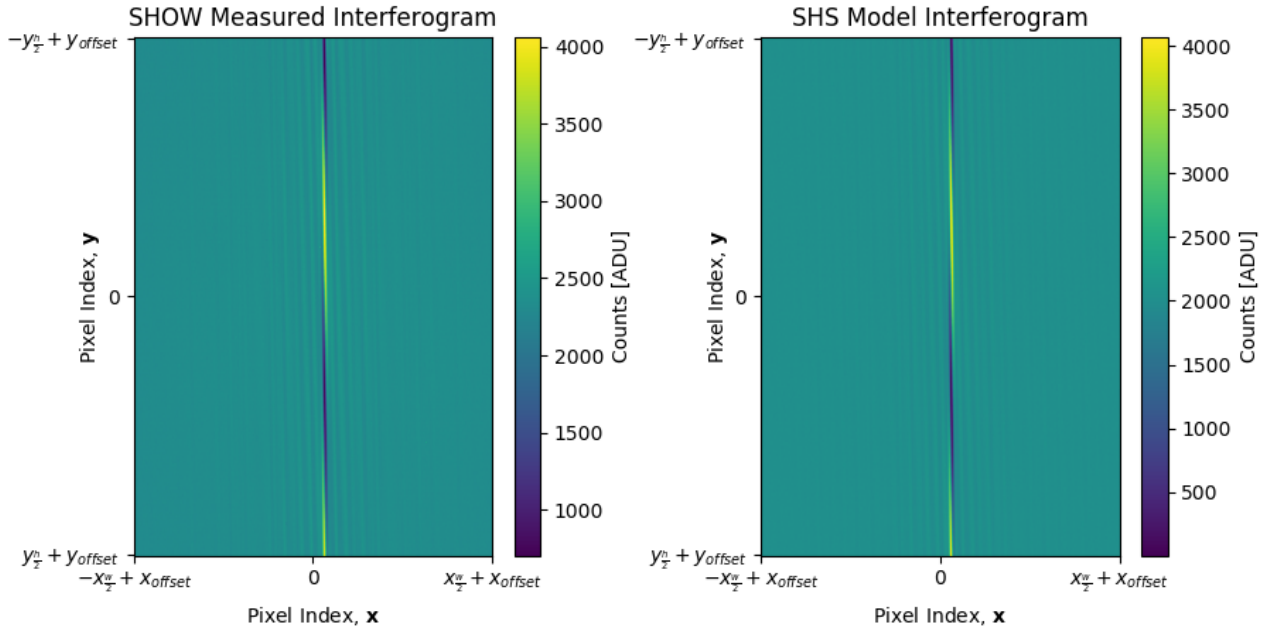

Figure 3.15: Comparison of SHOW lab measurement and SHS model interferograms.

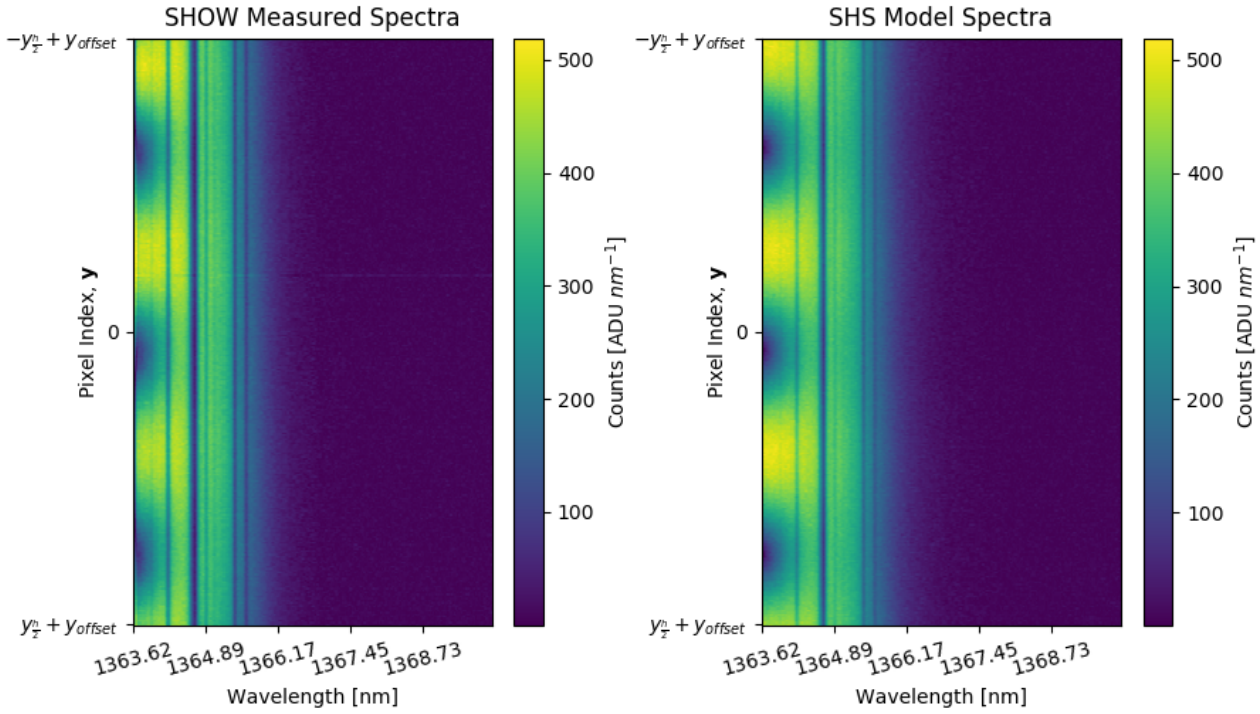

Figure 3.16: Comparison of SHOW lab measurement and SHS model spectra.

Chapter 4

SHOW ER2 2017 Campaign

It was the intention to fly SHOW on NASA's ER-2 science aircraft since the present work was started. The goal of this flight was to asses capability of SHOW to measure water vapour spectra of high enough quality to facilitate atmospheric water vapour retrievals within ± 1 ppm accuracy and a vertical (altitude) resolution less than five hundred meters [Langille et al., 2018]. Of course, any scientific information which could be gained by studying the UTLS water content observed in this campaign would be desirable but the primary mission was to validate SHOW, and more widely SHS technology, as applicable to atmospheric remote sensing. To this end, only the validation of SHOW is within the scope of the present work.

The work prior to the flight taking place involved two tasks. The first was to simulate the campaign to asses the performance of the modelling software and to make campaign expectations. The second was to develop a database of instrument calibrations to correct the collected data for non-ideal effects. The campaign itself took place in mid July of 2017 out of the Armstrong Flight Research Center. During the campaign three different flights took place. The first was a qualification flight which was to test the integration of SHOW into the ER-2 aircraft. The second flight was an Engineering flight purposed with experimentally determining integration times for SHOW, establish flight procedures for the science flight and, most notably, collect some in-situ radiosonde measurements to assist in later verification of SHOW. The last flight was the science flight which collected measurements all along the west coast of the United States of America. The final phase of work was to process the collected data from the campaign, particularly, to demonstrate that SHOW was capable of meeting its scientific goals through comparison to the radiosonde data. This chapter will describe the work done during these phases of the project.

4.1 Campaign Simulation

Prior to the flights of the ER-2 campaign there was a desire to run a simulation of the expected science flight. This simulation was done to form expectations of the altitude sensitivity of water, SNR of the measurements needed, and the performance of the retrieval algorithm along with its modelling capabilities. The methods and results of this work have already been published in the *Journal of Quantitative Spectroscopy & Radiative Transfer* [Langille et al., 2018] in which the present work contributed by constructing, executing, and analyzing the results of the simulation. However, the topic will be covered here in a more succinct manner for inclusion.

The need for better defining the SNR and sensitivity expectations is because limb scattered spectral measurements fail to contain a quality imprint of the water features when the observing line of sight is too optically thick or too optically thin as described by Equation 2.3. In the case where there is very little water content, as there is in the higher instrument lines of sight, the spectral imprint of the water is too weak against the noise of SHOW for the retrieval algorithm and forward model to develop an accurate guess. These lines of sight are too optically thin to be of use. Likewise, when the water concentration is very high, as they are in the lower instrument lines of sight, there is so much water that the depth of spectral features bottom out, resulting in loss of information about the concentration. Furthermore, the pressure broadening of the features becomes so significant in these optically thick lines of sight that the overall signal level gets lowered closer to the noise floor, again resulting in a worse SNR. Inclusion of these poor quality measurements in the inversion problem will worsen the overall quality of the retrieval since it will attempt to fit to the poor data at the expense of higher quality data.

The simulation to understand these limits better was done using the model described in Chapter 3 but without the inclusion of the systematic corrections or optimization. The simulation and its retrievals were run in an entirely artificial capacity in which the instrument model was used as the stand in for real SHOW data. SASKTRAN, which substituted Earth's atmosphere, provided the input to the instrument model to construct the measurement set. The water vapour content in the measurement state of the atmosphere within SASKTRAN

was set to an Environment and Climate Change Canada (ECCC) Global Environmental Multi-scale - Modelling Air quality and CHemistry forecast model [Talbot et al., 2008] for relative realism. Note that sharp discontinuous were observed in the forecast model which were smoothed over in the simulation. Integration times of the exposures was selected to be three hundred milliseconds to provide observed SNR between 250 and 600.

The flight path taken by the simulated SHOW instrument in constructing the measurement set was broken into two legs: north and south. Both legs were simulated with the instrument at a constant twenty two kilometres above sea level. Figure 4.1 below shows the two flight paths, both of which start at 34° N, 118° W - the approximate location of the Armstrong Flight Research Center (AFRC). The north and south legs take linear paths to 48° N, 125° W and 14° N, 125° W respectively. At the time of the simulation this was the best knowledge of the planned flight path. In order to manage processing time and to match the spatial resolution of the ECCC forecast model, five rows of the interferogram were binned together to provide an interferogram which is smaller than the actual interferogram expected from SHOW and measurements were only conducted every twenty nine kilometres along the flight track.

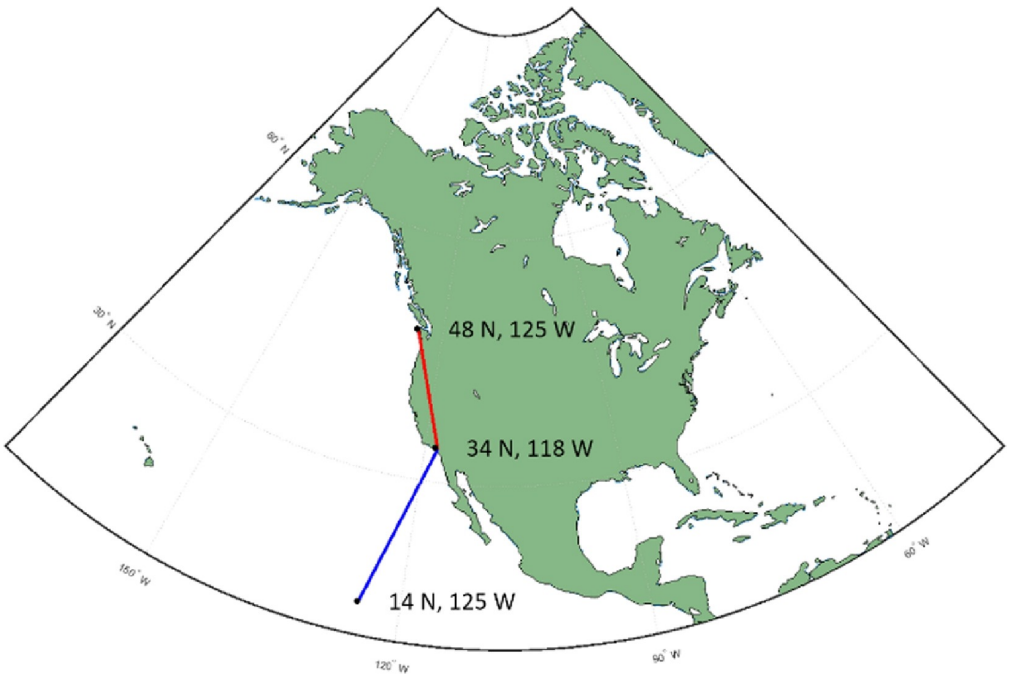

Figure 4.1: SHOW ER-2 simulation flight path. Altitude was 22 km.

Noise as described in Section 3.3.3 was included in the simulated data set but there are still some realistic effects which are not present. As already noted, there is the systematic correction which is required for real data that is not included in the simulation. This step is simply not applicable under the artificial conditions. There is also pitch variation of the aircraft in a real scenario which was not simulated in the constructed data set. In reality the pitch variation would cause each $\mathbf{y}$ row of the interferogram to likely view a different altitude from exposure to exposure or even cause smearing of the observed altitudes within a single exposure. In practise this can be corrected to a large extent with accurate knowledge of the aircraft position and orientation during flight but is not included in the present simulation. Finally, there was no inclusion of clouds in the simulated data set but this is not particularly relevant since interferograms of real data which show a localized increase of intensity due to clouds are neglected in processing.

The results of the simulation are shown in Figure 4.2 below. This figure shows the true state of the atmosphere as was in SASKTRAN along the two flight tracks, the retrieved water concentration along the flight track, and the absolute difference between the two to show the true error of the retrieval. The first take away from the results is that the scientific goals of SHOW (retrievals within ± 1 ppm accuracy and an altitude resolution less than five hundred meters) are achievable, at least in theory, with the methods described in Chapter 3. The error is largely less than ± 1 ppm; however, the vertical resolution in the particular retrievals shown above is almost double that of the target five hundred meter goal. Reducing the vertical resolution down to an acceptable level is not a significant challenge since, as demonstrated more robustly in the published work [Langille et al., 2018], simply increasing the SNR by averaging more exposures or increasing exposure time will reduce the vertical resolution. The second result of the simulation was a better understanding of the water sensitivity and retrieval limits. By virtue of the water content being so low at high altitudes, the upper altitude limit of useful information is determined by the path length of the observation. Through experimentation the upper limit was constantly found to be about eighteen kilometres. The lower limit, which is now dominated by the amount of water vapour instead of path length, is determined by the examination of the Jacobian matrices $\mathbf{K}$ produced by the retrieval algorithm. Since this matrix expresses the effect on the forward model caused

by changing water concentrations at different altitudes then the largest values in this matrix should be at their corresponding altitudes; i.e. if a change in water concentration at altitude X is not observed most prominently in the line of sight observing altitude X then this line of sight is too optically thick for use. The white space in the retrieval shown in Figure 4.2 below is where a dynamic algorithm determined the lines of sight to be too optically thick. While the fundamental idea behind this method is sound, and in general the lower limit of useful information is at about twelve kilometres, the algorithm used at the time failed to reliably produce the correct results all the time. The resulting Jacobian matrices proved difficult to interpret with an algorithm and in all retrieval exercises done after this point in time of the present work the lower limit was found in a manual fashion.

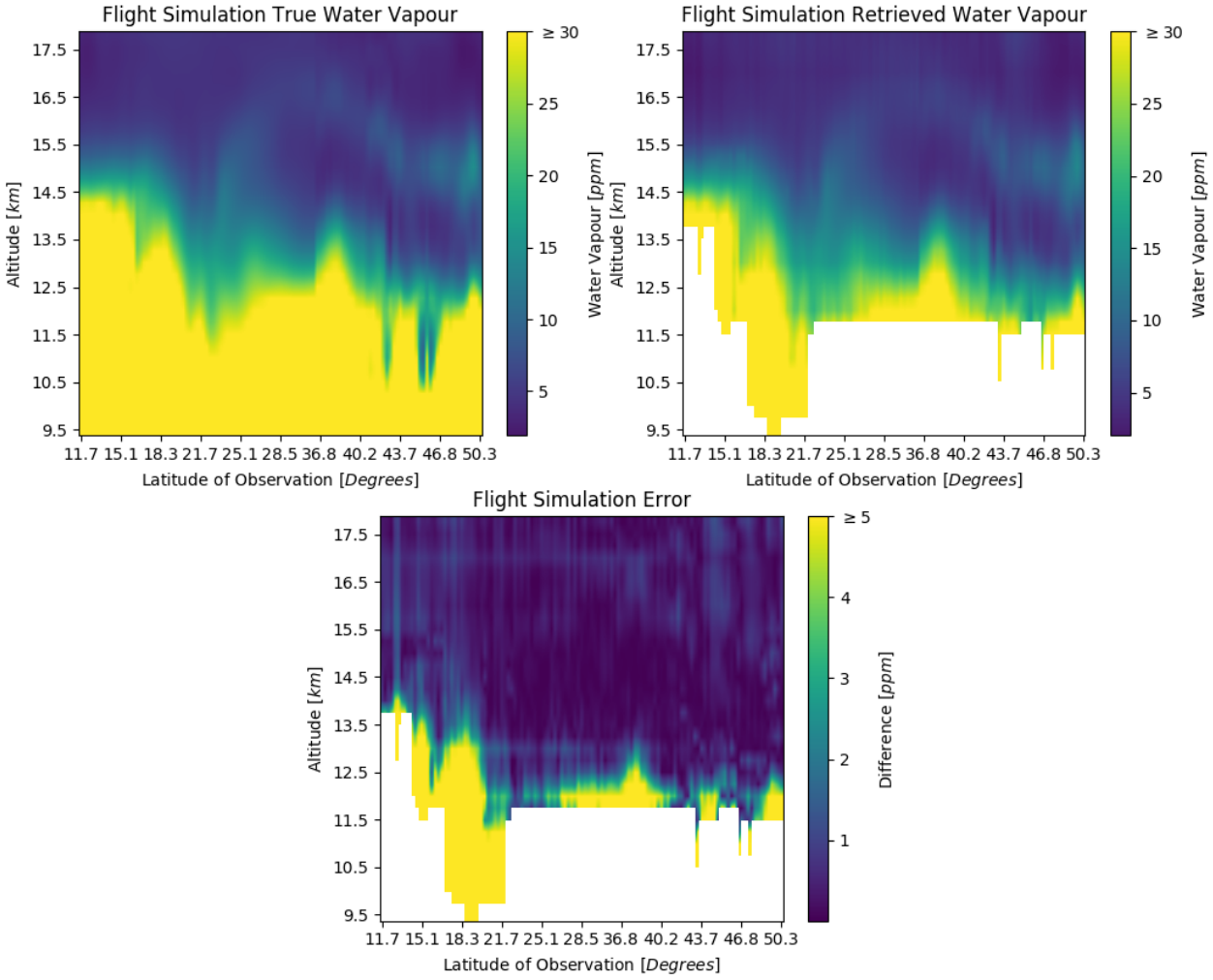

Figure 4.2: Water vapour in flight simulation.

4.2 SHOW Calibration and Characterization

As with any scientific instrument, there are imperfections and idiosyncrasies in SHOW which need to be understood and calibrated out so that the collected data can be interpreted to the truest physical extent possible. These properties can be broadly categorized into the characteristic of the imaging detector and characteristic of the SHS. This section outlines all of the properties that these two categories encompass, the procedures taken to isolate and quantify them, and the data pipeline structure which corrects them. Much of the information presented here can also be found in [Langille et al., 2019] which, as stated before, is a publication the present work contributed to. Furthermore, while this section will be discussed within the context of SHOW, the general procedures can be applied to other scientific instruments of a similar nature.

4.2.1 Detector Characterization

The imaging detector within SHOW is a commercial Raptor Photonics Owl 640 InGaAs camera. It should be noted that, somewhat unexpectedly, the frame rate of this detector significantly changes its behaviour as far as these characterizations are concerned. It is not well understood why the frame rate matters but the dependent behaviour is constantly repeatable. While this may not be the true cause, a possible explanation as to why is within the manual for the detector. The manual describes the timings of the detector operations within a single frame. Within this frame there is a delay between when the camera is told to begin an exposure and when the actual integration takes place. During this delay the detector behaves in a non-linear way. Since the frequency of this non-linear behaviour exhibiting itself is tied to the frame rate, it is *possible* that this is the reason. However, since the behaviour at a particular frame rate is consistent and repeatable, the issue was not investigated further and rather it was decided that the camera would only be operated in two frame rates: one hertz and one half hertz. This was to narrow the scope of the detector's behaviour into one which could be practically calibrated.

In addition to, and possible related with, the frame rate dependency was some observed non-linearity in the behaviour of the detector. This non-linearity was also observed to have

a behaviour dependent on the intensity of light incident on the detector which made some of the calibrations discussed here difficult and ultimately lead to the need for the systematic correction mentioned in Section 4.4.1.

Regardless, any InGaAs detector will have the non-ideal characteristics of dark current, non-uniform photo response, and faulty pixels. Without quantifying and correcting these errors it will be impossible to correctly interpret the spectral information contained in the interferograms this detector is responsible for collecting. The following are the particulars of each effect along with the procedure used to quantify them.

Dark Current

Dark current presents itself as a detected signal level in the collected image even when no light is incident. This results in an intensity offset of the collected interferograms. The magnitude of the dark current is dependant on exposure time, temperature, and the frame rate of the camera. The limitation of the frame rate discussed above limits the range of exposure times which can be used. It was decided that for images requiring less than one second exposure would used the one hertz rate and images which needed more than one second would use the half hertz rate. As such, collecting dark current calibrations in the range of ten milliseconds to the nine hundred and ninety milliseconds at ten millisecond intervals would provide enough calibration points for the one hertz data. Likewise, collecting data from one thousand one hundred milliseconds to eighteen hundred milliseconds at 10 millisecond intervals would be adequate for the half hertz data. As for the temperature, it is highly controllable due to the thermal electric cooler built into the detector. However, as the cooler works to drive the temperature of the detector lower it produces more and more heat which is dispersed into the rest of the SHS instrument. To balance the performance of the detector with the thermal needs of the rest of the instrument, it was decided that the detector would always be operated at zero degrees Celsius.

The campaign simulation done in Section 4.1 showed that an exposure time of three hundred milliseconds was adequate so the constraints set above should cover the needed operational range of the detector with plenty of margin for adjustment. Under these constraints the procedure to collect the dark current calibration is as follows:

- Ensure that the entrance cover of SHOW is on so that no light can enter the instrument.

- Turn on the detector's cooler and allow the temperature to reach 0° C.

- Set the frame rate of the detector to either 1 Hz or 0.5 Hz.

- If the frame rate is 1 Hz then the exposure range will be 10 ms to 990 ms at intervals of every 10 ms. If the frame rate is 0.5 Hz then the range will be 1100 ms to 1800 ms at intervals of every 10 ms.

- At this frame rate collect between one hundred and one thousand images at each exposure time of the frame rate's range and average these images together.

- Repeat for the other frame rate.

Figure 4.3 below shows an example dark image correction and the mean signal of all averaged dark images with respect to exposure time. In general, the dark images for both frame rates at all exposure times were comparable to the example figure below; falling within the 2000 ADU to 2200 ADU range with longer exposure times tending toward higher values. The noise in each individual dark image was about 4 ADU which, after averaging, was reduced to approximately 0.4 ADU in each final calibration image. Note, the trend of the dark images is clearly not linear and there is a discontinuous jump between the two sets of frame rate data. The discontinuous jump is evidence of the unexpected frame rate dependence discussed before; however, the non-linear trend that is evident here is typical of this type of detector and was expected. The issue with the non-linearity which lead to the need for the systematic correction is a result of the observed intensity dependence and not the inherent trend of the dark current. This will be discussed more in the photo response non-uniformity correction below.

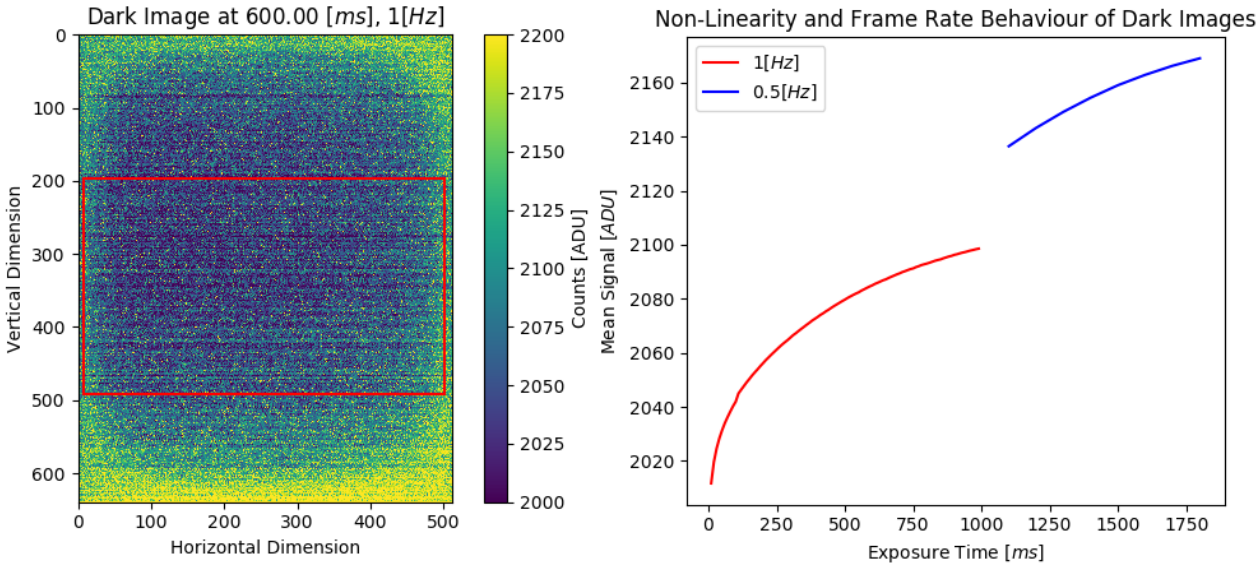

Figure 4.3: (left) Dark calibration image (600 ms, 1 Hz) after averaging. Red outline shows usable region of image. (right) The mean signal of the dark images vs exposure time to display non-linearity and frame rate effect.

Photo Response Non-Uniformity

Each pixel in an InGaAs detector has an unique response to illumination due to manufacturing limitations. This means that every pixel of the detector needs to be thought of as an individual sensor with an unique offset for calibration purposes. The non-uniform response of the pixels leads to systematic offsets in the interferograms which distributes error into the resulting spectra. For this reason it is important that this non-uniformity response be quantified and corrected. Normally, this correction would be done on a bare detector; however, at this point in time the detector was already installed into the optics of SHOW and could not be removed. It is because of this constraint that this correction was intertwined with the flat field correction (Section 4.2.2).

The idea behind quantifying this response, under the constraints presented, is to do a pixel by pixel linear fit to the detector with data collected at two different intensities of uniform light; one intensity obviously higher than the other. To act as a source of uniform light a large, high-grade, and calibrated integration sphere was used. By setting the current of the sphere's lamp to two different settings the two different uniform intensities could be

achieved. The particular settings used for the SHOW calibration were 2.67 A and 3.63 A. These two settings allowed for good coverage over the range of the detector well. Collecting the data to produce this pixel by pixel linear fit to the response was done in the following steps:

- Allow the integrating sphere to stabilize at either the 2.67 A or the 3.63 A current setting.
- Let the detector cool to 0° C for consistency with the dark current considerations.
- Select either the 1 Hz or 0.5 Hz frame rate.
- While the sphere is illuminating the entrance aperture of SHOW, take 100 to 1000 images at each exposure time of: 100 ms, 150 ms, 200 ms, 250 ms, 300 ms, 350 ms, and 400 ms.
- Repeat the above step two more times. Each time with a different arm of the interferometer blocked.
- Repeat the whole thing again for the all combinations of current and frame rate.
- Construct the photo response non-uniformity correction following the procedure below.

The reason to take the arm blocked measurements is to incorporate the flat field correction which will be discussed more in Section 4.2.2. The fundamental idea is that the observed signal for a particular pixel takes the form of

$$y_{[i,j]} = g_{[i,j]}x_{[i,j]} + b_{[i,j]} \tag{4.1}$$

where i and j are pixel indices, y is the observed signal, g is the slope or gain, x is the incident radiance, and b is the offset. This means that the corrected image can be obtained pixel by pixel using

$$
\begin{aligned}
x_{[i,j]} &= g_{[i,j]}(y_{[i,j]} - b_{[i,j]}) \\
g_{[i,j]} &= \frac{I_1 - I_2}{I_{1[i,j]} - I_{2[i,j]}} \\
b_{[i,j]} &= I_{1[i,j]}
\end{aligned}
\tag{4.2}
$$

where I_1 and I_2 are the average signal over the full image for the high current and low current settings respectively, and $I_{1[i,j]}$ and $I_{2[i,j]}$ are the measured signals at a particular pixel in the same images.

The idea is to do the flat field correction for two lamp currents, apply the lower current flat field correction to both the high and low current images such that the lower current image was basically constant and the higher current image highlighted the gain variation. The non-uniformity response fit described above was applied to these images to find a guess of the response correction. This response correction was applied to the lower current image of the flat field correction but not the higher current image and then process repeated to further isolate the response correction.

After the procedure was done it was found that the photo response non-uniformity could not be fully separated from the flat field due primarily to non-linearity in the pixel response. To compensate for this the flat field corrections applied have the photo response non-uniformity included. This means that flat field corrections have to be obtained for each signal level. For the purposes of the SHOW campaign, this method worked but left residual variations which required an additional systematic correction discussed in Section 4.4.1.

Faulty Pixels

In a similar vein to the photo response non-uniformity correction, manufacturing limitations prevent every pixel in the detector from operating correctly. Some pixels in the detector array will always be "hot" and have a high signal regardless of incident light. Likewise, some pixels will be "dead" and respond with a very low (if any) signal when the array is otherwise saturated.

The process of identify these pixels is very straight forward. Looking at the dark images it is expected that under no light conditions each pixel should give no more than about

two thousand and two hundred counts. Pixels which repeatably surpassed this value in well averaged (over one hundred images) dark images by a clear margin were deemed hot. To identify dead pixels the detector was saturated to ninety five percent of its full well depth of about sixteen thousand counts. Pixels which repeatably showed a weak response under these saturation conditions in well averaged images were deem dead. The correction to the identified pixels is a simple interpolation of their value using the values of the surrounding pixels and it is applied in the level 0 to level 1A conversion discussed in Section 4.2.3. Figure 4.4 below shows the resulting bad pixel map of the detector of SHOW for the usable region of the detector.

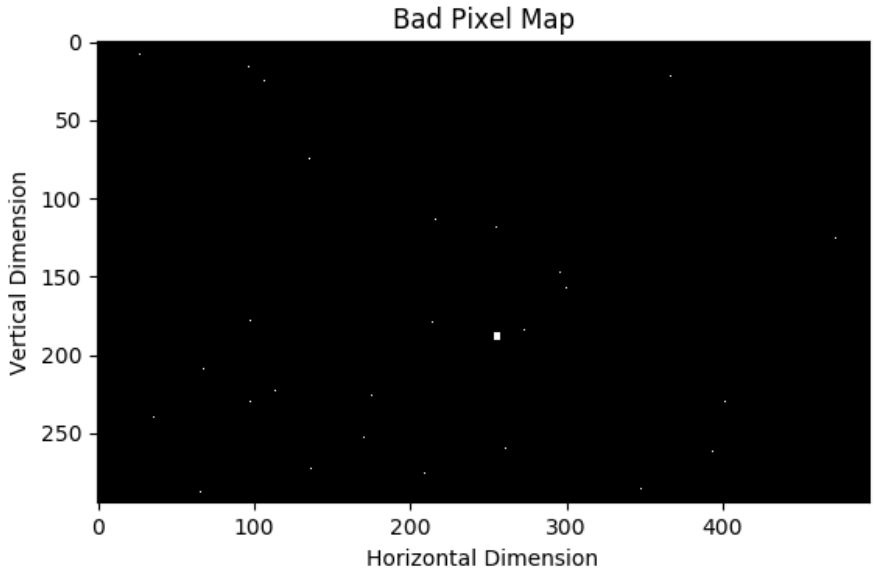

Figure 4.4: Dark pixel map of SHOW detector. Bad pixels are the white marks. Only bad pixels within usable region shown.

4.2.2 SHS Characterization

In contrast to the last section which covered the characterization of the SHOW imaging detector, this section concerns itself with the characterization of the optical components of the instrument. The properties of interest are the magnification of the exit optics, the passband of the optical filter, the spectral resolution, measuring λ_L, and determining a flat field calibration. Quantifying the magnification and the filter passband are not very important for

direct calibration but are instead very important for assessing the health of the instrument and interpreting the data. In addition, these also provide information required to correctly model SHOW as Chapter 3 has discussed. Determining the spectral resolution of the instrument, as well as measuring λ_L, are not very useful beyond an instrument checkout since these properties fluctuate with temperature and are also "calibrated" through the instrument optimization of Section 3.6 as well. What is very useful to calibration is finding the flat field of the instrument. The intent of the flat field is to combat the optical imperfections imprinted in the observed measurements. Accomplishing this greatly improves the quality of the instrument's scientific observations.

It should be noted that there are obviously additional optical properties which could be included in this section. Properties like the focus of the optics, transmission efficiencies, verification of the field of view and angular resolution of the instrument, grating groove density, and more. However, *direct* measurement of such things laid outside of the scope of the present work. As far as the present work was concerned SHOW was given as a complete and unmodifiable instrument with these kinds of properties confidently verified by previous parties. With that said, some of these properties, such as the grating groove density, were *indirectly* measured through the instrument optimization process discussed in Section 3.6 when possible. The following will detail the procedures used to measure all of the aforementioned properties that fell within the scope of the present work.

Exit Optic Magnification

The concept behind measuring the magnification of the exit optics is very simple. The height of the diffraction gratings is known to be three centimetres and the pixel pitch of the Raptor Owl 640 is fifteen micrometres. Since the exit optics of the SHS image the gratings onto the detector, then the magnification can be measured by imaging the gratings, measuring the observed height, and then calculating its scale from the known true height. The formal procedure is:

- Uniformly illuminate the entrance aperture of SHOW with white light. Ensure the signal is strong enough to produce clear intergerograms.

- Take roughly 100 images at a constant exposure time.
- Turn off the light and cover the entrance aperture of SHOW.
- Take roughly 100 dark images at the same exposure time as before.
- Average the illuminated images together.
- Average the dark images together.
- Subtract the average dark image from the average illuminated image.
- Determine the number of pixels that image the grating in the vertical and convert that to a height.
- Take the ratio of observed height and the actual height to find the magnification.

In the case of SHOW the fully illuminated grating was imaged by four hundred and forty six rows of pixels. Note that this range extends beyond the usable range of the detector. This means that the imaged height of the diffraction gratings is $0.669cm$. Dividing this number by the true three centimetre height of the diffraction gratings the magnification becomes 0.223 as listed in Table 2.1.

It is worth noting that measuring the magnification is an indirect way to measure the focus of the exit optics with respect to the imaging detector from the optical design as on paper. If one assumes that the optical system was constructed as designed and that design states there is a magnification factor of 0.223 then it follows that measuring that magnification also confirms the proper focus. This was the case with SHOW which had a designed magnification of 0.22.

Filter Passband

All light which enters the front aperture of SHOW goes through an optical filter that isolates a narrow wavelength range of the input spectrum. It is important to understand the transmission profile and position of the filter when looking at observational data since it directly impacts the observed intensities of the spectrum. The filter does have a temperature dependence which is fine tuned using the instrument optimization method during data

analysis. However, to get the first quantification of the filter passband requires the ability to produce near monochromatic light within the observable range of the SHS. The present work accomplished this with the use of a monochromator that takes a broad spectrum source as input and uses diffraction gratings to output desired wavelengths of light with a maximum passband of 0.18 nm. The procedure to map the filter passband of the SHS instrument is done as follows using the monochromator:

- Using coupling optics illuminate the entrance aperture of SHOW with the output of the monochromator.
- Determine a source intensity and exposure time combination which produces clear signal on the interferograms.
- Scan the output of the monochromator across the entire range of the SHS instrument. In the case of SHOW this was 1355 nm to 1372 nm in steps of 0.2 nm.
- At each monochromator step take a minimum of 25 images.
- Repeat all measurements with each arm of the interferometer blocked.
- Correct the measurements for dark current and then average them together to produce a singular intensity value for each monochromator step.
- Internally normalize the three different full measurement sets (SHS fully open, one arm blocked, other arm blocked) to the maximum observed intensity seen in each.

It should be made clear that this method does not allow one to exactly determine the transmission of the filter since the intensity of the initial source is not quantified. However, this method does still determine the relative shape and position of the filter. The idea behind doing the three different measurement sets is primarily to perform due diligence in taking more than one measurement of the filter.

Spectral Resolution and Littrow Wavelength

Measuring both λ_L and the spectral resolution of the instrument is a straight forward process, however, both are influenced by operating temperature making the measurement mostly

useful for instrument checkout purposes. To measure both of these things a monochromatic source is required. For SHOW a Krypton spectral lamp served as the monochromatic source since there is a spectral line in Krypton at 1363.422 nm (in air). This wavelength closely corresponds to λ_L of SHOW which makes the process more widely. Regardless of the source used the procedure is as follows:

- Stabilize the temperature of SHOW as much as possible.
- Using the monochromatic source antebellum the front aperture of SHOW.
- Take roughly 25 exposures.
- Take roughly the same number of dark images.
- Correct the interferograms for the dark current and average.
- Select some example rows of the interferogram.
- Do a least mean square cosine fit of the form $A\cos(2\pi fx + c)$ to the interferogram pattern.
- Determine the frequency of the cosine fit and use it to solve for σ_L in Equation 2.24, incorporating the magnification such that $f_x = 4M_x \tan(\theta_L)(\sigma - \sigma_L)$.
- Convert σ_L into λ_L to complete the wavelength measurement.
- Take the FFT of the interferogram to get an observed spectrum of the monochromatic source.
- Estimate the full width half max of the observed peak in the spectrum to obtain a resolution.

Flat Field

The theory behind flat fielding a SHS instrument is described in *Flatfielding in spatial heterodyne spectroscopy* [Englert and Harlander, 2006] and will not be discussed here. For the purposes of the present work all that is required to know is that there are two terms:

$$FF_1 = I_A(x) + I_B(x)$$
$$FF_2 = \frac{2[I_A(x)I_B(x)]^{\frac{1}{2}}}{I_A(x) + I_B(x)} \tag{4.3}$$

where $I_A(x)$ and $I_B(x)$ are interferograms where one arm is blocked and then the other. The FF_1 term corrects for intensity variations caused by the entrance and exit optics. The FF_2 term is used to measure the optical imbalance between the two interferometer arms. The procedure to collect the data is:

- Uniformly illuminate the entrance aperture of SHOW with white light.
- Stabilize the detector temperature at $0°C$.
- Select either the 1 Hz or 0.5 Hz frame rate.
- The exposure time and light intensity combo should fill roughly 60% of the detector well for ideal conditions.
- Take 100 exposures with one interferometer arm blocked.
- Take another 100 exposure with the other interferometer arm blocked.
- Take 100 dark images and correct both previous sets for dark current.
- Average the two corrected image sets and then normalize them by their average value.

For SHOW it was found that the FF_2 was equal to one within 0.2% displaying that the two interferometer arms are very well balanced with each other. As a result of this the FF_2 was deemed not needed in the calibration. An example of a FF_1 correction is shown in Figure 4.5 below. This term is significantly more dominate than the the FF_2 term and it is applied as a correction through division of the interferograms by this term during calibration.

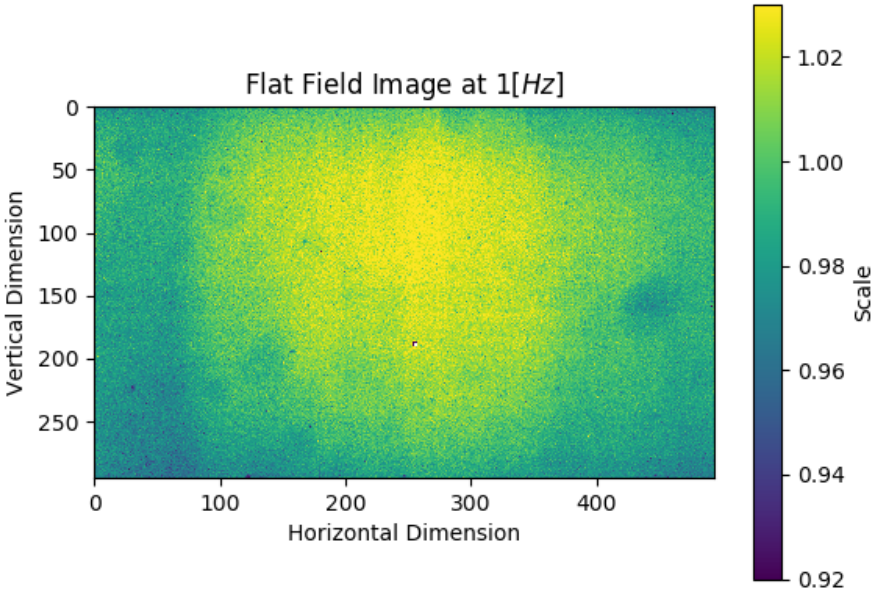

Figure 4.5: Example flat field correction for SHOW data. Only usable region shown.

4.2.3 Data Pipeline

The data pipeline is a Python 3.x code package which takes raw interferograms directly from the instrument, calibrates them, and turns them into usable spectra. The raw interferograms, referred to as Level 0 data, are completely uncorrected. A set of Level 0 data from a particular section of data, such as a set of calibration measurements, is then turned to Level 1A data. Level 1A data are a collection of Level 0 data bundled together into a netcdf file with calibrations and relevant house keeping data attached to each interferogram. For the purposes of the present work the Level 1A data are of limited interest since the retrieval algorithm works in the spectral space. As a result of this, the Level 1A data are transformed into another netcdf file referred to as Level 1B data. The Level 1B data are the Level 1A data transformed into spectral space with aircraft position and attitude information (if applicable) now being attached to the scientific data. Table 4.1 below shows the calibrations and associated data of the two data stages. Finally, there is a third Level 1C stage in which the systematic correction of Section 4.4.1 has been applied. Other than this one change there is no other difference between Level 1B and Level 1C.

To apply the Level 1A calibrations to an interferogram the dark current is first subtracted

Level 1A Corrections	Level 1A Data
Dark current	Corrected interferogram
Faulty pixels	Exposure time of interferogram
Flat field	UTC / MJD time stamp
Unusable region discarded	Interferogram error
	Various temperature data
Level 1B Corrections	**Level 1B Data**
Interpolation over faulty pixels	Spectrum
Apodization	Wavelength
	Exposure time of interferogram
	UTC / MJD time stamp
	Spectral error
	Various temperature data
	Raw aircraft information (if applicable)
	Processed geocentric aircraft information (if applicable)

Table 4.1: Table of level 1 A & B data stages, applied corrections, and associated data.

off and then the interferogram is divided by the flat field. The pixels which were identified as faulty (pixels which are too hot or too cold) are assigned a NAN value to identify them in further processing. Finally, to reduce file size the unused region of the interferogram (region outside of red box in Figure 4.3 (LEFT)) is discarded. The faulty pixels identified in the Level 1A data are interpolated over using the surrounding pixels in the Level 1B data before the being apodized with an one dimensional Hanning window (see Figure 3.4) and turned into a spectrum via a FFT. All data from rows 384 - 388 where the large chunk of bad pixels are is kept in the Level 1B data but discarded for the retrieval since the interpolation fails due to lack of data in the surrounding pixels.

In regards to the aircraft data attached to the Level 1B data, this is only done if there is aircraft data to be attached from one of the science flights discussed in Section 4.3 below. The raw information is provided from NASA as an ER-2 data product. This information contains time, GPS position, aircraft speed, and aircraft attitude. Of vital importance to the retrieval algorithm is the aircraft position and attitude. SHOW is mounted in the ER-2 wingpod during flight and what it observes is dictated by the attitude, heading, and location of the aircraft. This information is correlated to the timestamps of the interferograms and then converted into the geocentric coordinate frame used by SASKTRAN where individual lines of sight for each row of the instrument can be calculated. This step is extremely important for determining the geometry of the retrieval and while careful effort has been put into making this as accurate as possible it is not perfect. Ultimately, the accuracy of the geometry can only be as good as the measuring devices on board the ER-2 which, along with conventional instrument error, does not measure the steadiness of the wing which SHOW is mounted on. *Spatial Heterodyne Observations of Water (SHOW) from a high altitude airplane: Characterization, performance and first results* [Langille et al., 2019] delves into the analysis of this error but it was found that as a result the angular resolution of the instrument is roughly twice as large as it could be without unaccounted wing moment.

4.3 Campaign Flight Details

In mid July of 2017 the SHOW campaign took place which involved flying SHOW for a total of three flights on board NASA's ER-2 aircraft. The first flight was a qualification flight that took place on July 13th and was primarily done to test the integration of SHOW into the ER-2 aircraft. The second flight was the engineering flight which took place on July 18th and was focused on instrument performance. Finally, the last flight was the main science flight which took place on July 21st and collected a large set of scientific data for study. For all three of these flights SHOW was mounted inside of a wing pod of the ER-2 facing forward of the aircraft. SHOW was also angled down from the aircraft frame. The measured angle between the SHOW optical axis and the forward vector of the aircraft as measured by its instruments was $2.40° \pm 0.1°$. This section will document the details of all three of these flights and give perspective on the data they collected.

Qualification Flight

For the purposes of the present work the qualification flight is of little interest because its mission was not to collect any meaningful data with SHOW. However, the flight is still covered here in brief for context. As said before, the purpose of this flight was to ensure integrating of SHOW into the ER-2 systems and to test control, thermal systems, and telemetry during flight.

The flight took off from AFRC at 14:30 coordinated universal time (UTC) and flew north to a latitude of of 40° N before turning around and following the same flight path home to land at 16:20 UTC. During this flight telemetry between the ground station and the ER-2 performed well. Images taken by SHOW could be downloaded in near real time and would allow changes to the aircraft attitude to be made for the engineering and science flights based on the observed interferograms if needed.

Engineering Flight

The goal of the engineering flight was to asses the performance of SHOW on board the ER-2 aircraft. To do this two main objectives were given to this flight: narrow in on the exposure

times to use for the science flight, and to collect coincident in-situ measurements with a radiosonde. The flight itself took off from AFRC at 15:02 UTC and flew the path shown in Figure 4.6 below. The flower looking pattern in the upper left was for the first objective. This pattern is called a compass rose and it allowed testing exposure times versus solar scattering angle. By downloading images in flight and looking at the number of counts with respect to the relative position of the sun a better sense of what exposure times to use during the science flight could be determined. At this time it was decided that the flight track for the science flight would be very different than what was originally thought in Figure 4.1. The science flight would consist of a north track following the west coast of the United States and a return south track along the same path. Ultimately, this compass rose exercise lead to the conclusion to use an eighteen hundred millisecond exposure at an one half Hertz frame rate for the north flight track given the foreseen sun position and a nine hundred millisecond exposure at one Hertz frame rate for the south track.

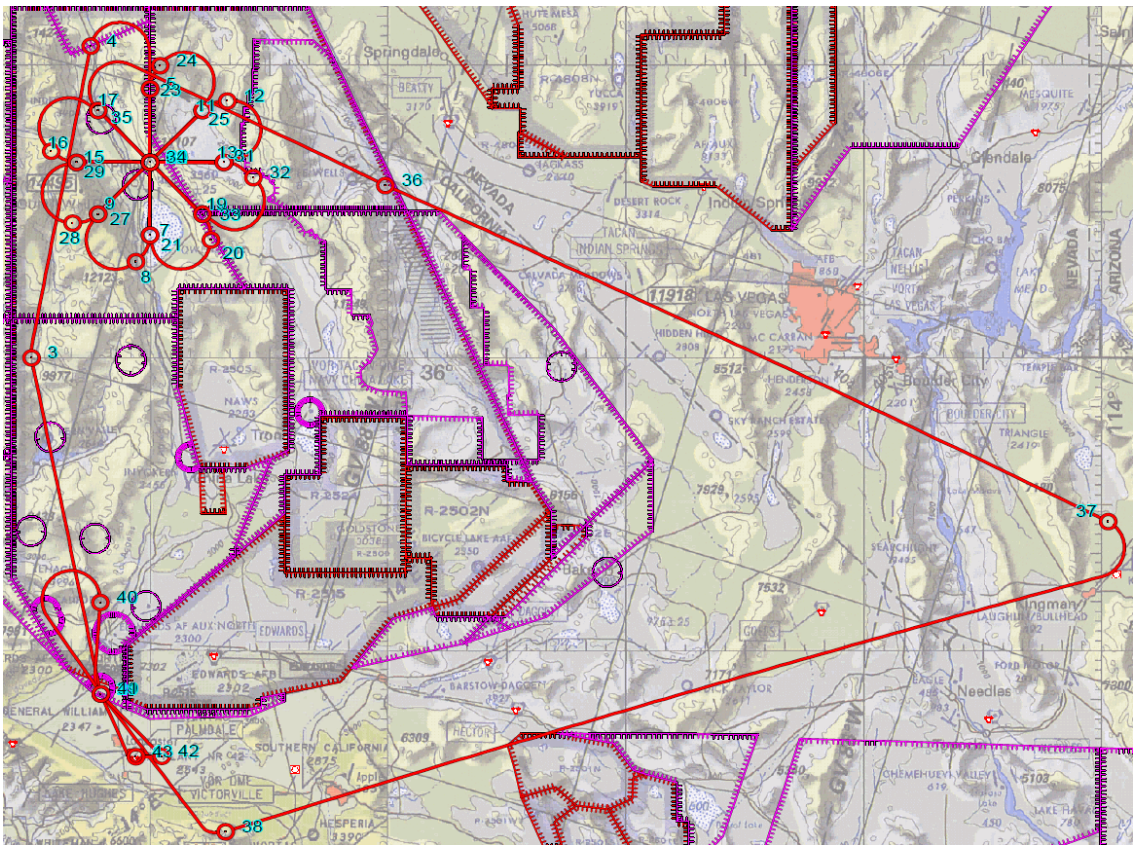

Figure 4.6: Planned flight path of the engineering flight, July 18th 2017.

The ideal behind the second goal of the flight, collecting in-situ measurements, was to provide a second data set to which the SHOW data could be compared to. After the turn marked by the number thirty seven in Figure 4.6 which occurred at 17:41 UTC to the point marked by thirty eight which was at 18:12 UTC, SHOW and the ER-2 were in straight and level flight at twenty one kilometres collecting a scientific data set. At this same time a Vaisala RS41 radiosonde was launched from the Jet Propulsion Laboratory facility located close to Table Mountain. The full exercise of performing the retrieval on the ER-2 data set of this time period and comparing it to the radiosonde data is discussed in Section 4.4.1.

Science Flight

The final flight of the SHOW ER-2 campaign followed a flight path along the west coast of the United Sates. This path was different than what was expected during the course of the campaign simulation in Section 4.1 simply because that flight path was a rough preliminary one. SHOW took off from AFRC at 15:02 UTC on board the ER-2 aircraft and headed west until it was over the ocean. At 16:00 UTC the aircraft began its turn north to start the first leg of the science mission. The ER-2 flew at an altitude of twenty one kilometres north until a latitude of 48°. At 17:38 UTC the ER-2 reached the end of the north leg and turned south following the same path back. At 19:22 UTC the ER-2 made its turn east to head back to AFRC where it landed at 20:30 UTC.

There was a large amount of interesting data taken during this science flight but it is not the focus of the present work. As already noted, the ultimate purpose of the present work is to validate SHOW as an atmospheric instrument capable of measuring water vapour in the UTLS to within ±1 ppm, with a vertical resolution of less than five hundred meters, and with high spatial sampling [Langille et al., 2018]. To this end the data collected from the radiosonde during the engineering flight is more useful and the full analysis of the data taken from the science flight remains a topic of future work. However, the performance over the science flight can yield results about the spatial sampling. It was found that along this north-south flight track SHOW, on-board the ER-2 platform, achieved a spatial resolution of 0.005° to 0.01° in latitude depending on the length of interferogram exposure. This translates into roughly a 500 m to 1000 m spatial resolution.

4.4 Data Analysis

The primary goal of the ER-2 campaign was to validate SHOW as an instrument able to measure water vapour in the UTLS. To this end, the data collected from the ER-2 campaign, specifically the engineering flight, was processed with the intent to meet the campaign's goal. The objective was to perform a retrieval using the Table Mountain data from the engineering flight in-situ with the radiosonde data as an example of the ability of SHOW to successfully retrieve water vapour.

In the processes of analyzing the campaign flight data for these retrievals it became quickly apparent that detector behaviour observed in the calibration (discussed in Section 4.2) would require a systematic correction to be applied which is constructed from the data itself. This correction is applied in addition to the calibrations done within the data pipeline but before the retrieval takes place. This section will explain the systematic correction and how the Table Mountain retrieval wes done and the results in comparison to the radiosonde.

4.4.1 Systematic Correction

There are effects associated with the measurements of SHOW which are very difficult to calibrate out. Specifically, the residual error left over from the flat field correction being intertwined with the photo response non-uniformity as discussed in Section 4.2.1. This introduces a fixed relative error in the interferograms in both the horizontal and vertical dimensions. When the error carriers over into spectral space after the FFT it results in a fixed variation in spectral space which requires correction for the retrieval algorithm to work correctly. To combat this in the retrievals a systematic correction was developed which normalizes this bias out of the data. This correction works by observing each row's response to an identical source, such as a particular tangent altitude, and isolates the behaviour relative to the other rows. In order to create and apply this correction a data set is required in which a number of different rows observe the same light source. Since the correction is required for the retrievals and will be different depending on the source(s) being looked at it makes sense to take advantage of aircraft pitch and create the correction for whatever individual data set is being looked at using the data set itself. This section will outline the creation of the

systematic correction using the data set of the Table Mountain Retrieval (discussed further in Section 4.4.2) for example purposes.

To begin constructing the correction for flight data, a reasonably sized temporal window of the data being looked at needs to be chosen in which there is adequate amount of vertical pitching that a single tangent altitude is observed by multiple rows. It is worth emphasizing that this correction assumes the source intensity at a particular tangent altitude is constant throughout the temporal window selected. Each observed wavelength of the spectral data in these measurements, along with their reported uncertainty, is binned onto a finely resolved tangent altitude gird. The binned information is then correlated to the detector row which observed it and for each detector row the spectral data, along with the uncertainty, is averaged to produce a representative intensity recorded by the detector row for observed tangent altitude. Note that if a detector row did not see enough data points to make a reasonable average then the row was discarded for the rest of the analysis. The result of doing this is presented in Figure 4.7 at an example tangent altitude of 13122 m and at a representative wavelength of 1364.81 nm.

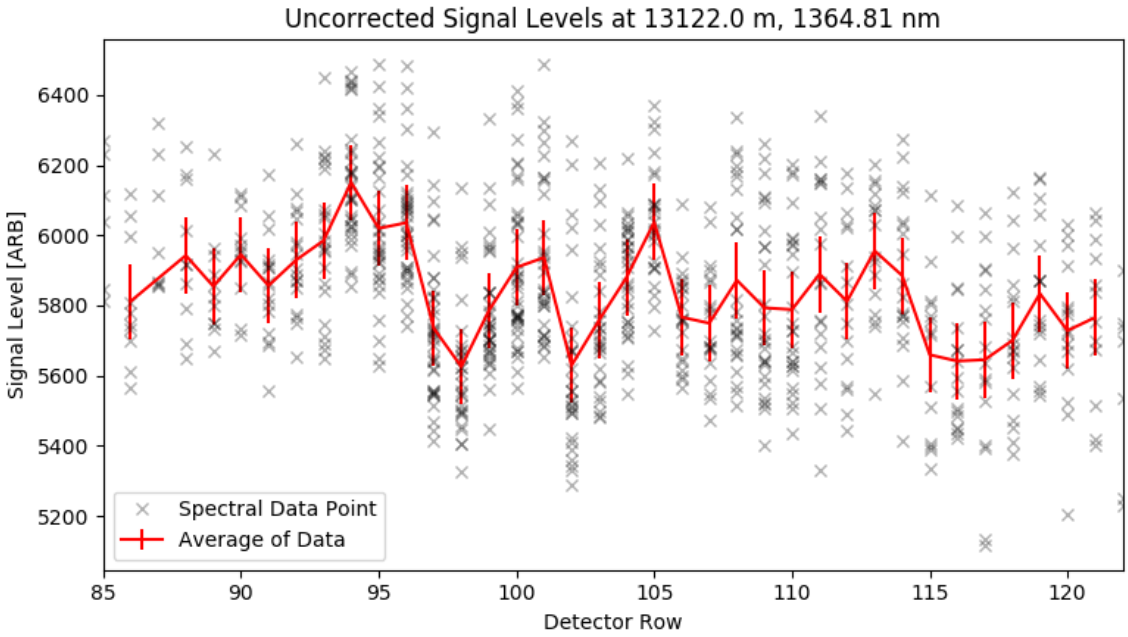

Figure 4.7: SHOW detector signal levels without correction. Note the non-uniformity of the red line depicting the systematic error.

If the detector was behaving in an ideal way than one would expect that each row's cluster of black markers be centred at some consistent signal level with a spread which can be correlated to the uncertainty of the measurement. Instead what is observed in the above figure is that the clusters of black markers jump around in average signal level such that a straight flat line cannot be drawn within the corresponding error bars. If the assumption that the intensity of the observed tangent altitude is constant with the selected temporal window is true than this row deviation from a consistent signal level is a systematic error which needs to be corrected.

To continue in creating the systematic correction, the row bias seen in Figure 4.7 above was overlaid with all other row biases seen at all other tangent altitudes at the same wavelength. This provided highly correlated bias behaviour for each individual detector row regardless of tangent altitude. All biases from all tangent altitudes pertaining to a particular row were then normalized by their mean and averaged together to form one coherent bias correction for all detector rows at the current wavelength. The whole process was repeated for all observed wavelengths which produce a bias correction value at each row and wavelength in the data set. As a final step to creating the systematic correction, recall from Figure 3.11 there is modulation imposed onto the spectra by the aliased signal. Since the SHS instrument model does a good job of simulating this modulation it is important that the systematic correction does not change it in the real data. To this end a high order polynomial is divided out of the systematic correction in the wavelength regions where aliased data is expected in an effort to eliminate any redundant modulation of the systematic correction itself.

The final result of this effort is a normalized matrix correction which when divided out of relevant spectral measurement made by SHOW will flatten the row biased response and give the data an uniformity expected by the forward model. Figure 4.8 below is the same data shown in Figure 4.7 above but now with the correction applied. Note that the consistent signal level expected is now achieved and the intensity dependant row bias which cannot easily be removed in lab calibrations is now compensated for. It was found through testing that the application of the systematic correction makes a significant improvement to the behaviour of the retrieval.

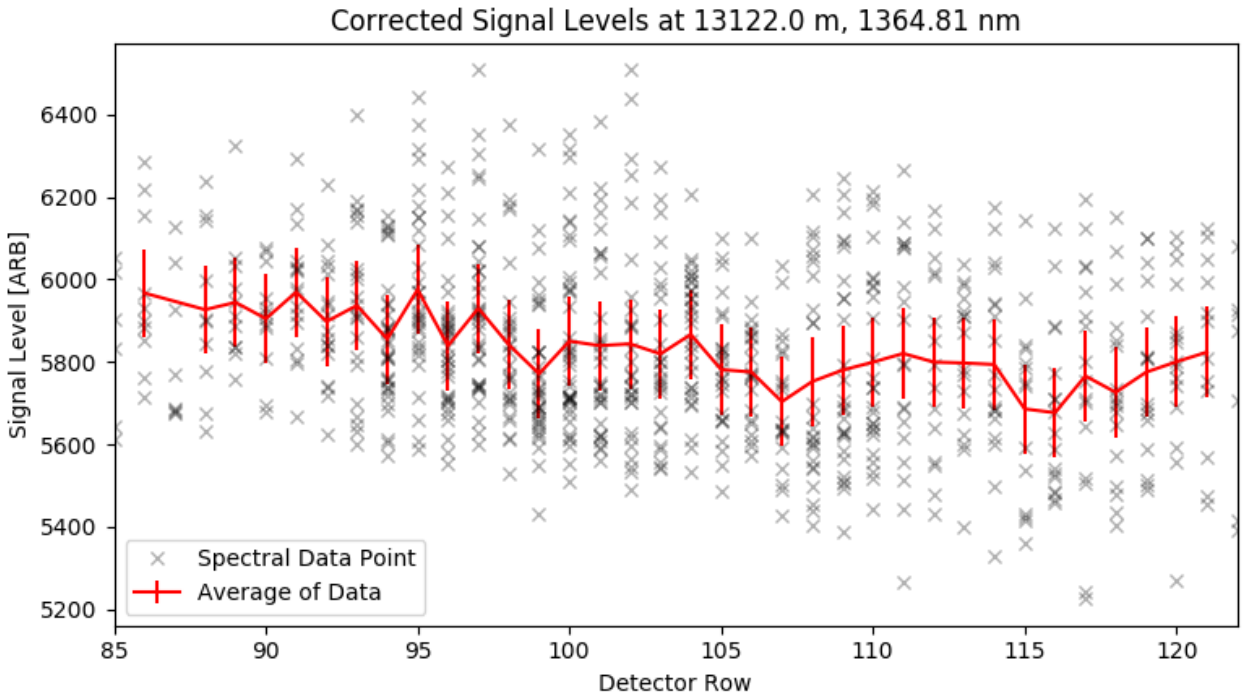

Figure 4.8: SHOW detector signal levels with correction. Note the uniformity of the red line suggesting significant reduction of the systematic error.

It is worth noting that this systematic error is actually an additive error and not a multiplicative error as the application of the correction would suggest, such that

$$S' = S + \delta s \tag{4.4}$$

where S' is the observed signal at a tangent altitude for a particular row, S is the actual signal, and δS is the systematic error of the row. When all the biases from all the tangent altitudes are normalized by their mean value this makes any difference from a value of one proportional to δS for a particular row if, and only if, the assumption is made that the average of the rows which observed the tangent altitude is equal to S. This assumption is valid if the inter-row variation appears random and that a reasonable number of rows observed the tangent altitude. So the effective correction being made and applied is

$$\frac{S'}{\bar{S'}} \tag{4.5}$$

where $\bar{S}'$ is the average of S' and since under the assumption above $\bar{S}'$ is equal to S this clearly produces S when divided out of Equation 4.4. The reason this method is applied as opposed to one which employs a subtractive application is that such a correction would also have to make the same assumption. Furthermore, through this method overlapping data points from other tangent altitudes for a single row can be averaged together to incorporate more information.

4.4.2 Table Mountain Retrieval

The Table Mountain retrieval serves as the primary piece of supporting evidence of the ability of SHOW to retrieve water vapour in the UTLS. A fair number of retrievals have been done using the Table Mountain data set employing different parameters; however, only one will be discussed here. The retrieval which will be discussed is an ideal example and displays the best ability of the retrieval algorithm applied to SHOW measurement data. This section will go into detail of how the SHOW retrieval was done in terms of the geometry and image averaging of the retrieval, the retrieval settings used, how SASKTRAN was setup, SHS insturment parameters, and finally, how the retrieval compares to the radiosonde data.

Retrieval Geometry and Image Averaging

The retrieval uses ten interferograms taken at 17:57:41 UTC to 17:57:59 UTC on the day of the engineering flight, July 18th 2017. The spectra obtained from these interferograms after all corrections were applied was averaged together to maximize the SNR while maintaining a respectable spatial resolution. As said before, each row of the imaging detector can be mapped to an observed tangent altitude. So the averaging of the spectral information from each interferogram is done via matching of the tangent altitude. A base image is selected to define the geometry of the retrieval and then all other lines of sight from all other images are averaged with their closest matching line of sight in the base image. Averaging the measurement images together in this fashion means that for optimal accuracy of the retrieval the same averaging needs to be done for each iteration of the retrieval on side of SASKTRAN and the SHS instrument model. Since this can be computationally time consuming the geometry of all averaged images are kept track of and then a single radiance profile is calculated at a higher

tangent altitude resolution than any of the measurements. Next, the radiance profiles used to make the simulated interferograms for averaging are consturcted through interpolation of this high resolution radiance profile. Note that this makes the assumption that the averaged images are temporally and spatially close enough that the observed atmosphere won't have changed significantly to invalidate this method. Attention should be paid to the images being selected for averaging with respect to this assumption.

Figure 4.9 below shows the tangent altitudes of the base image used to define the geometry of the Table Mountain retrieval. Note that the indices shown are with respect to the usable region discussed in the Section 4.2. The particular reference latitude and longitude of the retrieval is 33.9° N and $-118.5°$ W with a reference altitude of $15.0km$.

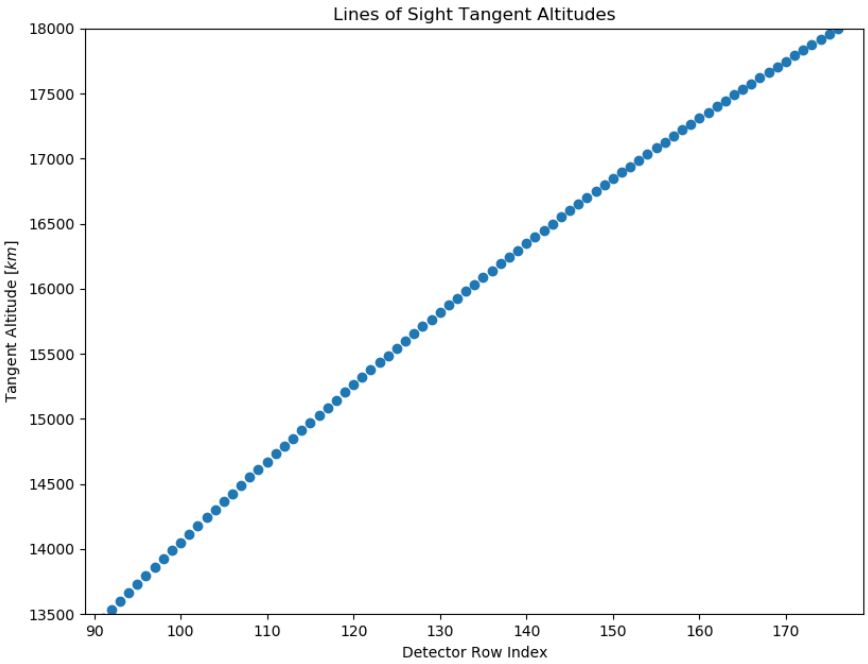

Figure 4.9: Tangent altitudes used in Table Mountain retrieval.

Retrieval Settings

The particular settings of the retrieval can make significant impacts on the results. These settings should be optimized so the retrieval is well behaved and converges. To this end it is worth mentioning the settings used in this retrieval. The *a priori* covariance ($\mathbf{S_a}$) for this

retrieval is zero such that no confidence was placed in the *a priori* profile. The idea behind this is to allow the retrieval to mold the initially guessed profile to the measurement data with as little external influence as possible. The retrieval was performed using fifteen iterations, a damping factor of five used to construct the Levenberg-Marquardt damping parameter, and a Tikhonov constraint damping factor of 0.1. The altitude limits of the retrieval were set with eighteen kilometres being the high altitude limit and thirteen and a half kilometres defining the low limit. These limits were set using the techniques developed during the campaign simulation of Section 4.1. Furthermore, the retrieval was performed on a two hundred and fifty meter altitude grid which works well for computation time and desired performance. The amount of scattering events within SASKTRAN was set to one for this retrieval which was well known from the simulation study. It was also found during testing that increasing the scattering events did not make any significant difference in the resulting retrieval for comparison to the radiosonde data. Because of these reasons the scattering was kept at one to reduce computational time.

The ideal end result of doing the inverse problem is that the spectrum produced by the last iteration of the retrieval will be identical to the measurement spectrum taken by SHOW. This is an impractical task for the SHOW retrievals of the present work since there are still imperfections in the simulation. For instance, not having any certain knowledge of the aerosol profile at the time of the measurements, not narrowing in on the temperature affects perfectly through the optimization process, and imperfections in the systematic correction are among the reasons why the spectrum won't be identical. After the retrieval is done, particular wavelengths at different altitudes will match better than others. It is possible to have the retrieval artificially weigh some wavelengths more than others with respect to their altitude but such weights would be determined somewhat arbitrarily based on human judgement. Therefore, the strategy to just simply discount some wavelengths, regardless of altitude, from being incorporated into the retrieval was adopted to reduce the amount of human judgement. Note that these are the wavelengths chosen to make the vectors $\mathbf{y}$ and $\mathbf{F}(\mathbf{x_i})$ and do not define what wavelengths SASKTRAN calculates to construct the instrument model interferograms and spectra. Furthermore, the mean averaging described in Section 3.5 is done based on these wavelengths and not the wavelengths of SASKTRAN. The Table Mountain retrieval has been

tried with many different combinations of these wavelengths but the retrieval presented here uses only the wavelengths which matched very well. Table 4.2 below lists these wavelengths.

1364.80	1364.96	1365.11	1365.26	1365.41	1365.57
1364.83	1364.98	1365.14	1365.29	1365.44	1365.60
1364.87	1365.01	1365.16	1365.32	1365.47	1365.62
1364.88	1365.03	1365.19	1365.34	1365.49	1365.65
1364.91	1365.06	1365.21	1365.37	1365.52	1365.67
1364.93	1365.09	1365.24	1365.39	1365.55	1365.70

Table 4.2: Table listing the wavelengths used to make vectors $\mathbf{y}$ and $\mathbf{F}(\mathbf{x_i})$ for the Table Mountain retrieval. Units of nm.

SASKTRAN Settings

Setting up SASKTRAN for the this, or any other, retrieval involves specifying the atmospheric climatology to begin the retrieval with. Recall Section 3.2 which described briefly how SASKTRAN works, the atmospheric species and their concentrations for each shell need to be defined for at least the first iteration. The atmosphere for this retrieval included a background climatology provided by MSIS90, aerosol, and of course water. In addition, the surface albedo was given a factor of 0.3 which is a common average value for the Earth's surface.

The aerosol profile used in the retrieval comes from Global Space-based Stratospheric Aerosol Climatology (GloSSAC) [Thomason et al., 2018]. While this is obviously not perfectly ideal it is the most realistic aerosol state that is available to the retrieval. Figure 4.10 below shows the particular aerosol profile used from the GloSSAC data which is the most recent profile corresponding to the reference latitude and longitude of the SHOW measurements. The work done in [Langille et al., 2018] explores the effect of using incorrect aerosol profiles in the retrieval. These results show that a poor choice of aerosol in the retrieval can lead to an inaccuracy largely on the order of one part per million of the retrieved profile. While undesirable and unknown to an exact degree, this error by no means defeats the purpose of validating SHOW.

The initially guessed water profile is the only atmospheric state in the retrieval which will change from iteration to iteration. The initial guess given can have an influence on the results of the retrieval. In testing with artificial data during the campaign simulation of Section 4.1 it was observed that the choice of a starting guess had no effect on the results of the retrieval; however, experimenting with real SHOW measurements showed a clear dependence of the results on the initial input. Due to this the initially guessed profile was selected to be as benign as possible. The profile as a whole follows the general trend of atmospheric water but is very dry, furthermore, the region where the retrieval is being done is forced to be flat at a value of 0.25 ppm. Outside the retrieval range the estimated profile profile is simply shifted to remain continuous with where the retrieval puts the limits. Figure 4.11 below shows the initial water profile used in the retrieval.

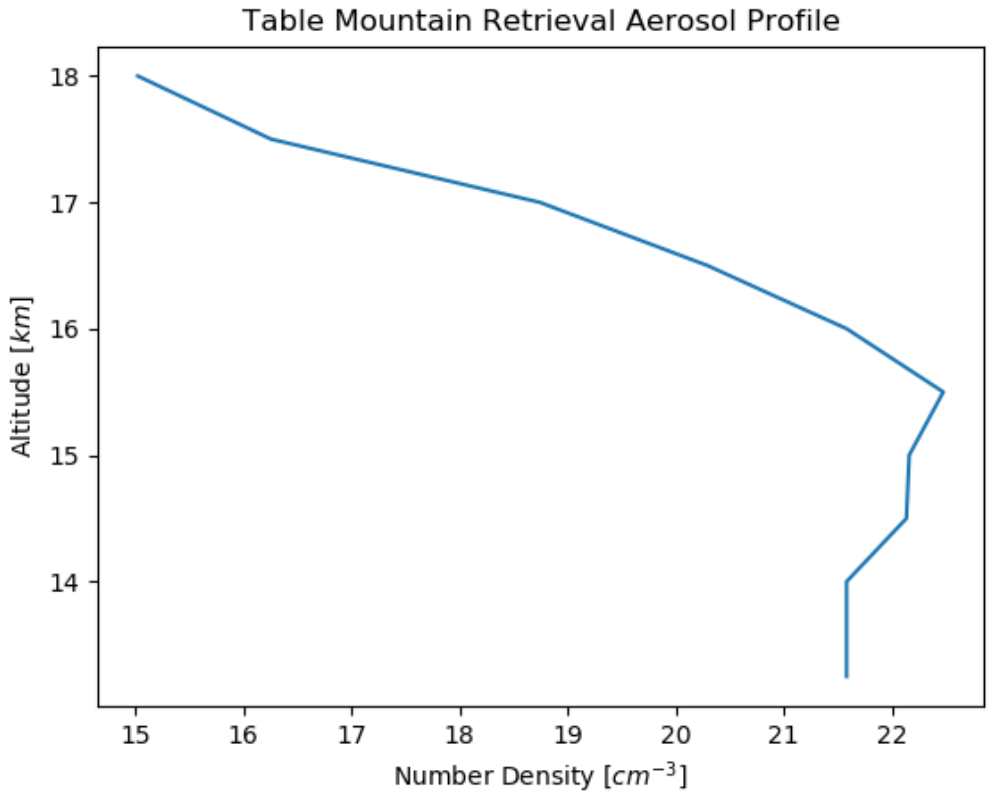

Figure 4.10: GloSSAC aerosol profile used in Table Mountain Retrieval.

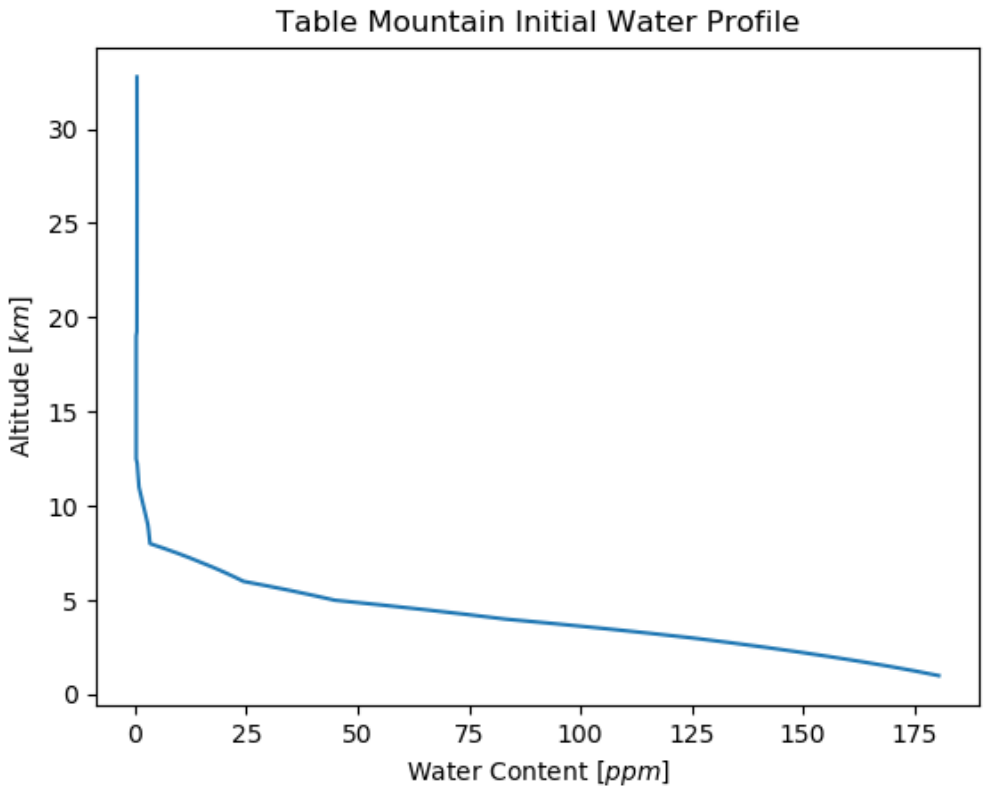

Figure 4.11: Initial water profile used in Table Mountain retrieval.

SHS instrument parameters

Like with both the SASKTRAN settings and the retrieval settings, the settings given to the instrument model are also important. From the optimization procedure discussed in Section 3.6 a x and y offset of 17.27 and -18.581 pixels were used respectively. The filter determined by the same procedure is shown in Figure 4.12 below which is the filter shape reported by the manufacturer shiftedby -0.048 nm. Finally, a λ_L value of 1363.76 nm was determined.

The wavelengths calculated by SASKTRAN and input to the instrument model was a high resolution array between the wavelengths of 1360 nm to 1370 nm at a resolution of 0.005 nm. It is important that the spectrum calculated by SASKTRAN be of a high enough resolution to capture the absorption features. This is because if the high resolution detail is either not there or simulated through interpolation of a lower resolution spectrum, the shape of the instrument spectrum made as the input spectrum is convolved down to the resolution of the instrument will be incorrect.

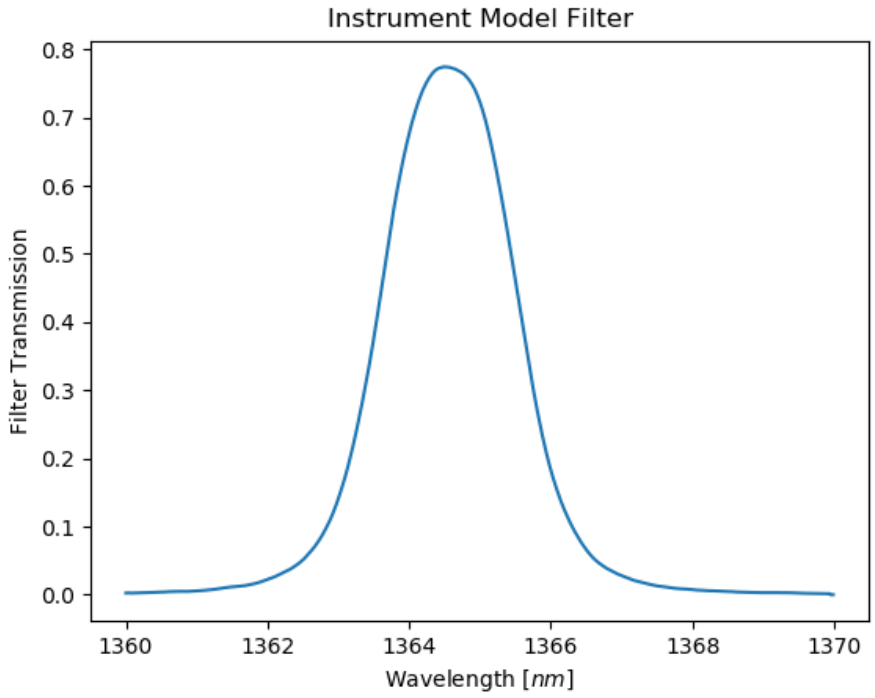

Figure 4.12: Instrument model filter used in Table Mountain retrieval.

Radiosonde Comparison

To properly compare the results of the retrieval to the data of the radiosonde, its data must be put into context. The radiosonde used was a Vaisala RS41 which reports relative humidity, pressure, temperature, and altitude. These quantities had to be converted into parts per million of water vapour for comparison. The method to do this was done using the Hyland and Wexler formulation [Hyland and Wexler, 1983]. Of course, the radiosonde itself is an instrument with its own uncertainty. To calculate the error in parts per million from the data recorded the maximum uncertainty as reported by the data sheet was added onto the raw data. The erroneous data were then processed with the same procedure as before. The difference in parts per million between this maximum error profile and the unaltered measurements made the error in parts per million.

The measurements provided by the radiosonde also need to be matched to the retrieval grid resolution for a proper comparison. The raw data of the radiosonde has an altitude resolution of roughly four meters. So to produce a profile comparable to the retrieval profile,

the data points which best matched the altitudes on the retrieval grid and their surrounding data were convolved with a gauss constructed with a full width half max of the averaging kernel at that same altitude. This lower resolution profile resembles the profile that one could expect to see from the retrieval if both are accurately measuring the same atmospheric water content. The original radiosonde data, the convolved radiosonde data, and the results of the retrieval algorithm are showed in Figure 4.13 below. This figure also shows the subtraction of the convolved radiosonde data and the retrieval results as well as example spectra to display the level of agreement between the two.

As Figure 4.13 shows, the Table Mountain retrieval performed from SHOW data and the radiosonde data are in very good agreement. The strong majority of points agree within error and the rest agree to within, or very close to within, one part per million which resembles the accuracy goal of the SHOW instrument. However, there are a number of important things to note for proper context of these results. First and foremost these measurements are not perfectly coincidental in either time or space. The radiosonde measurements take place between 17:59:16 UTC and 18:04:14 UTC where as the full set of SHOW data used in this retrieval takes place before in the last twenty seconds of 17:57 UTC. Spatially, the reference coordinates of the SHOW data are over the Marina Del Rey in Los Angeles, CA. The radiosonde did not report its position but presumably it measured a column of atmosphere above the facility from which it was launched. Needless to say there is spatial uncertainty between what atmospheric water is being compared. Furthermore, as discussed before, there is an error of uncertain magnitude in the pointing of SHOW which affects the geometry of the retrieval as well as uncertainty in the atmospheric state of SASKTRAN compare to the true atmospheric state. Lastly, while the uncertainties of both instruments were carried with the analysis there is still a question of both instruments can truly measure the atmospheric water. SHOW remains in its infancy in this regard but studies of this and other radiosondes showed that accuracy changes as a function of humidity and temperature differently for different radiosondes [Miloshevich et al., 2009] and that errors between sensors to be on the order of 2 - 5% [Jensen et al., 2016]. However, with all this in mind, the agreement between the two measurements is strong supporting evidence to the measurement abilities of SHOW.

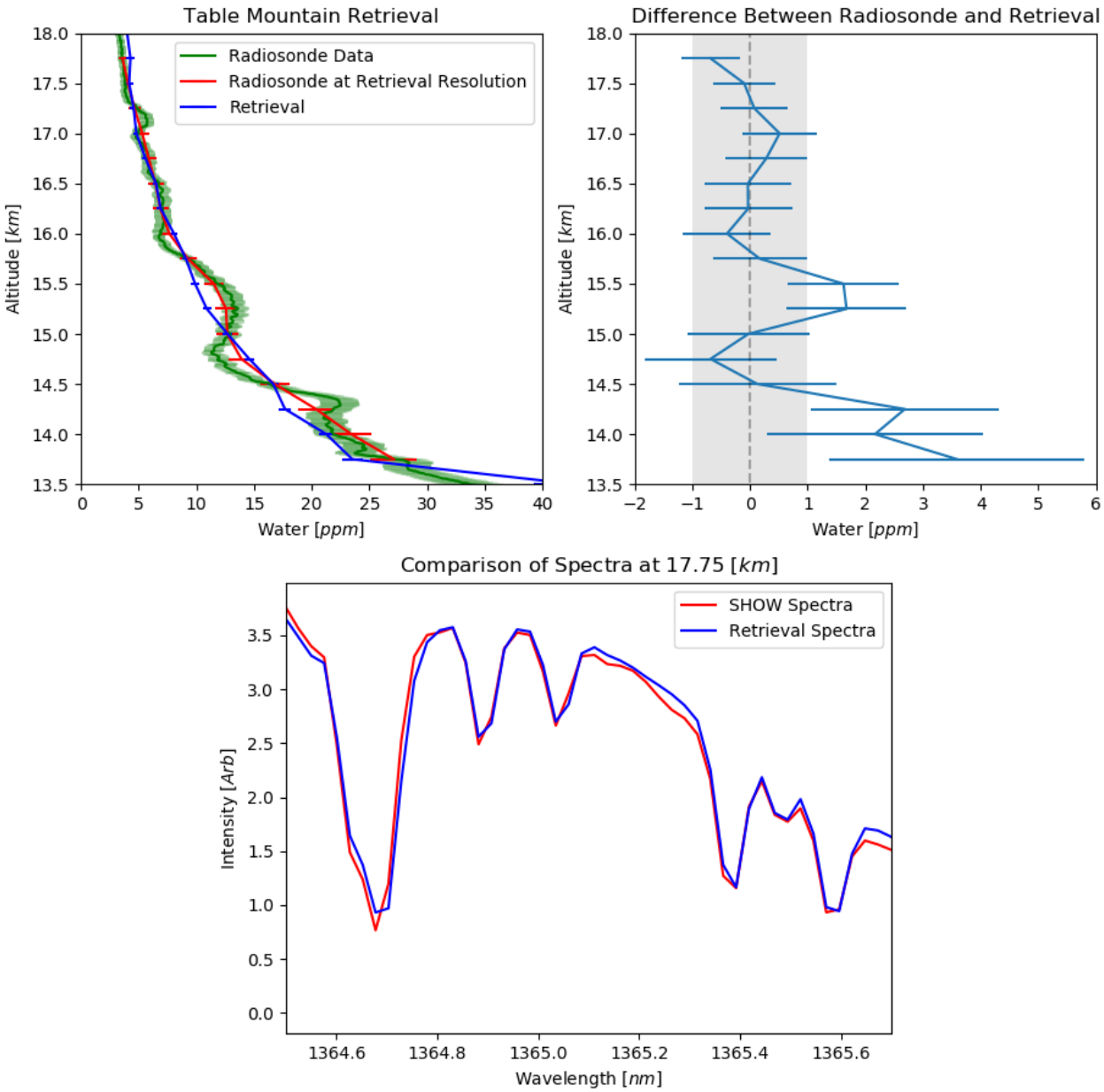

Figure 4.13: Results of the Table Mountain retrieval. (top left) The SHOW retrieval at an altitude resolution of 250 m in comparison to the radiosonde data. (top right) The SHOW results and the radiosonde data largely agree within one part per million. This lends strong supporting evidence that SHOW is capable of measuring water to the desired level. (bottom) An example of the spectral fit the retrieval algorithm was able to accomplish.

Chapter 5

Conclusion

The present work detailed the efforts of validating SHOW, and SHS technology, as applicable to collecting atmospheric observations. In presenting these efforts, background information on water vapour in the UTLS, different atmospheric measurement techniques and their trades, radiative transfer theory, current atmospheric water observing instruments, SHS theory, and details of SHOW as an instrument have been provided. In addition, the details of the theory, construction, and resulting abilities of the model have been presented as well. Finally, information on the SHOW instrument calibration, the July 2017 campaign, and the results of analyzing the campaign data with respect to water vapour are given which culminates in a reported conclusion of SHOW's abilities.

As Section 2.1 has mentioned, there is a need for high spatial, and vertical, resolution measurement of water vapour within the UTLS. To validate SHOW as an atmospheric measurement tool it needed to succeed in providing high resolution measurements within an accuracy of ± 1 ppm of water vapour. During the July 2017 campaign a radiosonde was launched to collect in-situ measurements for comparison against SHOW. An analysis of this data showed that SHOW agreed to the radiosonde very well, largely within the goal of one part per million of water vapour. This result provides very strong supporting evidence that SHOW is able to measure water within the desired accuracy goal. Furthermore, the measurements taken by SHOW yield water vapour profiles with a vertical resolution of two hundred and fifty meters using the analysis methods and tools described in the present work. This resolution is very high relative to the other atmospheric water observing instruments mentioned in Section 2.5 and within the desired goal of five hundred meters. The spatial resolution of SHOW is largely controlled by the speed and viewing geometry of the craft that it is on. However, in the slowest operating mode used in the present work a useful measurement was

taken every two seconds. Given the speed and geometry of the ER-2 platform this resulted in a latitude resolution of 0.005° to 0.01° (roughly 500 m to 1000 m). This conclusion makes the SHOW SHS design a valid candidate to fulfil the missing water vapour observations of the UTLS. Furthermore, SHOW should be considered for future development, and potential deployment, in low earth orbit and SHS based instruments should be considered as potential solutions to measuring other UTLS constituents.

As mentioned in Section 4.3 there was a science flight over the west coast of the United States. A very large amount of data was collected in this flight and to process all of this data in a similar fashion to the Table Mountain retrieval will take roughly three hundred and seventy core years of computation time. Processing this data is an ongoing task using the Cedar Supercomputer and at the time of writing the present work there is intent to make another future publication presenting the results of that analysis.

References

Barath, F. T., Chavez, M. C., Cofield, R. E., Flower, D. A., Frerking, M. A., Gram, M. B., Harris, W. M., Holden, J. R., Jarnot, R. F., Kloezeman, W. G., Klose, G. J., Lau, G. K., Loo, M. S., Maddison, B. J., Mattauch, R. J., McKinney, R. P., Peckham, G. E., Pickett, H. M., Siebes, G., Soltis, F. S., Suttie, R. A., Tarsala, J. A., Waters, J. W., and Wilson, W. J. (1993). The upper atmosphere research satellite microwave limb sounder instrument. *Journal of Geophyscial Research*, **98**:10751–10762.

Bell, R. J. (1972). *Introductory Fourier Transform Spectroscopy*. Academic Press.

Bernath, P., McElroy, C., Abrams, M., Boone, C., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMaziere, M., Drummond, J., Dufour, D., Evans, W., Fast, H., Fussen, D., Gilbert, K., Jennings, D., Llewellyn, E., Lowe, R., Mahieu, E., McConnell, J., McHugh, M., McLeod, S., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C., Rochon, Y., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K., Walkty, I., Wardle, D., Wehrle, V., Zander, R., and Zou, J. (2005). Atmospheric chemistry experiment (ace): Mission overview. *Geophysical Research Letters*, **32**. L15S01, doi:10.1029/2005GL022386.

Bourassa, A. E. (2007). *Stratospheric Aerosol Retrieval From OSIRIS Limb Scattered Sunlight Spectra*. PhD thesis, University of Saskatchewan.

Bourassa, A. E., Degenstein, D. A., and Llewellyn, E. J. (2008). Sasktran: A spherical geometry radiative transfer code for efficient estimation of limb scattered sunlight. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **109**:52–73.

Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V., Chance, K., and Goede, A. (1999). Sciamachy: Mission objectives and measurement modes. *Journal of the Atmospheric Sciences*, **56**:127–150.

Caballero, R. (2014). *Physics of the Atmosphere*. IOP Publishing.

Chandrasekhar, S. (1960). *Radiative Transfer*. Dover Publications, Inc.

Davis, S. P., Abrams, M. C., and Brault, J. W. (2001). *Fourier Transform Spectrometry*. Academic Press.

Englert, C. and Harlander, J. (2006). Flatfielding in spatial heterodyne spectroscopy. *Applied Optics*, **45**:4583–4590.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., Lopez-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R. (2008). Mipas: an instrument for atmospheric and climate research. *Atmospheric Chemistry and Physics*, **8**:2151–2188.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., and Birner, T. (2011). The extratropical upper troposphere and lower stratosphere. *Rev. Geophys.*, **49**. RG3003, doi:10.1029/2011RG000355.

Harlander, J. M. (1991). *Spatial Heterodyne Spectroscopy: Interferometric Performance at any Wavelength Without Scanning*. PhD thesis, University of Wisconsin-Madison.

Hegglin, M., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., , and Weigel, K. (2013). Sparc data initiative: Comparison of water vapor climatologies from international satellite limb sounders. *Journal of Geophysical Research: Atmospheres*, **118**:11824–11846.

Hyland, R. W. and Wexler, A. (1983). Formulations for the thermodynamic properties of the saturated phases of h2o from 173.15k to 473.15k. *ASHRAE Trans*, **89(2A)**:500–519.

Jensen, M., Holdridge, D., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. (2016). Comparison of vaisala radiosondes rs41 and rs92 at the arm southern great plains site. *Atmospheric Measurement Techniques*, **9**:3115–3129.

Langille, J., Letros, D., Bourassa, A., Solheim, B., Degenstein, D., Dupont, F., Zawada, D., and Lloyd, N. D. (2019). Spatial heterodyne observations of water (show) from a high altitude airplane: Characterization, performance and first results. *Atmospheric Measurement Techniques Discussions*, **12**:431–455.

Langille, J., Letros, D., Zawada, D., Bourassa, A., Degenstein, D., and Solheim, B. (2018). Spatial heterodyne observations of water (show) vapour in the upper troposphere and lower stratosphere from a high altitude aircraft: Modelling and sensitivity analysis. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **209**:137–149.

Langille, J., Solheim, B., Bourassa, A., Degenstein, D., Brown, S., and Shepherd, G. (2017). Measurement of water vapor using an imaging field-widened spatial heterodyne spectrometer. *Applied Optics*, **56**:4297–4308.

Lin, Y. (2010). *Instrument development for spatial heterodyne observations of water*. PhD thesis, York University.

Loughman, R., Flittner, D., Nyaku, E., and Bhartia, P. K. (2015). Gauss-seidel limb scattering (gsls) radiative transfer model development in support of the ozone mapping and profiler suite (omps) limb profiler mission. *Atmospheric Chemistry and Physics*, **15**:3007–3020.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11(2)**:431–441.

McLean, I. S. (2008). *Electronic Imaging In Astronomy: Detectors and Instrumentation*. Springer.

Miloshevich, L., Vomel, H., Whiteman, D., and Leblanc, T. (2009). Accuracy assessment and correction of vaisala rs92 radiosonde water vapor measurements. *J. Geophys. Res.*, **114**. D11305, doi:10.1029/2008JD011565.

Organization, W. M. (1957). A three-dimensional science: Second session of the commission for aerology. *WMO Bulletin*, **4**:134–138.

Randel, W. and Jensen, E. J. (2013). Physical processes in the tropical tropopause layer and their roles in a changing climate. *Nature Geoscience*, **6**:169–176.

Rodgers, C. D. (2000). *Inverse Methods For Atmospheric Sounding: Theory and Practice*. World Scientific Publishing.

Rozanov, V., A.V.Rozanov, A.A.Kokhanovsky, and J.P.Burrows (2014). Radiative transfer through terrestrial atmosphere and ocean: Software package sciatran. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **133**:13–71.

Salby, M. L. (2012). *Physics of the Atmosphere and Climate*. Cambridge University Press.

Sherwood, S. C., Roca, R., Weckwerth, T. M., and Andronova, N. (2010). Tropospheric water vapor, convection, and climate. *Rev. Geophys.*, **48**. RG2001, doi:10.1029/2009RG000301.

SPARC, Hegglin, M. I., and Tegtmeier, S. (2017). The sparc data initiative: Assessment of stratospheric trace gas and aerosol climatologies from satellite limb sounders. *SPARC Report No. 8, WCRP-05/2017*. doi:10.3929/ethz-a-010863911, available at www.sparc-climate.org/publications/sparc-reports/.

Spurr, R. (2006). Vlidort: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **102**:316–342.

Talbot, D., Moran, M. D., Bouchet, V., Crevier, L. P., Ménard, S., Kallaur, A., and the GEM-MACH Team (2008). *Development of a New Canadian Operational Air Quality Forecast Model. In: Borrego C., Miranda A.I. (eds) Air Pollution Modeling and Its Application XIX. NATO Science for Peace and Security Series Series C: Environmental Security*. Springer, Dordrecht.

Thomason, L. W., Ernest, N., Millán, L., Rieger, L., Bourassa, A., Vernier, J.-P., Manney, G., Luo, B., Arfeuille, F., and Peter, T. (2018). A global space-based stratospheric aerosol climatology: 1979–2016. *Earth Syst. Sci. Data*, **10**:469–492.

Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., Holden, J. R., Lau, G. K., Livesey, N. J., Manney, G. L., Pumphrey, H. C., Santee, M. L., Wu, D. L., Cuddy, D. T., Lay, R. R., Loo, M. S., Perun, V. S., Schwartz, M. J., Stek, P. C., Thurstans, R. P., Boyles, M. A., Chandra, K. M., Chavez, M. C., Chen, G., Chudasama, B. V., Dodge, R., Fuller, R. A., Girard, M. A., Jiang, J. H., Jiang, Y., Knosp, B. W., LaBelle, R. C., Lam, J. C., Lee, K. A., Miller, D., Oswald, J. E., Patel, N. C., Pukala, D. M., Quintero, O., Scaff, D. M., Snyder, W. V., Tope, M. C., Wagner, P. A., and Walch, M. J. (2006). The earth observing system microwave limb sounder (eos mls) on the aura satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **44**:1075–1092.