SPLITTER

A data model and algorithm for detecting spectral lines and continuum emission of high-redshift galaxies using DESHIMA 2.0

S.A.Brackenhoff

TU Delft

SPLITTER

A data model and algorithm for detecting spectral lines and continuum emission of high-redshift galaxies using DESHIMA 2.0

A thesis by

S.A.Brackenhoff

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday June 17th, 2021 at 9:00 AM.

Student number: 4619005
Project duration: August 31, 2020 – June 17, 2021
Thesis committee:

Prof. dr. ir. A.J. v.d. Veen,	TU Delft, supervisor
Dr. A. Endo,	TU Delft, supervisor
Dr. A. Taniguchi,	Nagoya University, external research expert

TU Delft Delft University of Technology

Abstract

We propose the Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR (SPLITTER) for removing wideband atmospheric noise from observations of high-redshift galaxies. This algorithm has specifically been developed for the DEep Spectroscopic HIgh-redshift MApper (DESHIMA) 2.0, a spectrometer that is designed to observe the waveband from 220 GHz to 440 GHz in 347 spectral channels. This octave bandwidth poses a challenge, due to the spectrotemporal changes in the atmosphere column between the instrument and the target source. Removing the time-varying nonlinear interference and distortion caused by the atmosphere is a difficult task, as the atmospheric emission is much stronger than a typical galaxy signal.

The goal of this thesis is to develop a method that can estimate both narrow spectral lines and the broad continuum emission with a higher sensitivity than the currently used method of directly subtracting noisy on- and off-source spectra. We develop a logarithmic data model for separating atmospheric noise from the galaxy signal in position switching-observations. Because the atmospheric transmittance appears as a multiplicative term in both the atmospheric interference and the signal modulation, the logarithmic model allows for an additive decomposition of the data. The atmospheric transmittance behaves as a low-rank component in this model.

Using the model, we develop an optimization algorithm (SPLITTER) to perform the separation of the signal and the low-rank atmospheric transmittance. Several implementations are discussed. The final algorithm uses a Singular Value Decomposition (SVD) to estimate the atmosphere component and the Alternating Directions Method of Multipliers (ADMM) for estimating the source signal. Instead of subtracting the noisy estimate of the source from the data directly, a denoised model is used in this step, such that we can trade some spectral resolution for a higher sensitivity.

SPLITTER is tested on simulated data using the Time-dependent End-to-end Model for Post-process Optimization of the DESHIMA spectrometer (TiEMPO), a dedicated software package for simulating DESHIMA observations. We show that SPLITTER is able to estimate the spectrum with a higher sensitivity than the conventional method. The improvement factor in our weighted root mean squared error is up to ~ 1.7 for the full spectrum and up to ~ 1.3 for the spectral lines only compared to the conventional method. The larger improvement for the full spectrum is achieved by trading spectral resolution for a higher sensitivity in the smooth continuum. With these results, we have an indication that a statistically driven method for DESHIMA observations can provide better estimates than the current method with the same amount of observing time.

More work is needed to create a robust version of the algorithm, because although the sensitivity benefit of SPLITTER is larger in the continuum regions, there are also situations where the continuum is overestimated. The conditions for this to occur are not yet clear. A more robust version could make SPLITTER a reliable new method that can replace current data reduction methods for wideband atmospheric noise removal. In this way, it can be used to make background-limited direct detection spectrometers on both existing and future telescopes observe more efficiently.

Acknowledgements

This project would not have been possible without the many people who have helped me along the way. First of all, I would like to thank my daily supervisor Akira Endo for this tireless enthusiasm and support during this project. Your optimism and support have been invaluable to me during the work, especially with the current corona crisis. It was always a pleasure to discuss and look for answers together on Slack, over Zoom and later even in person. In the limited options available due to COVID-19, you have done everything in your power to be available and make me feel like a part of the academic community and I want to thank you for this. It was extremely interesting to learn about your perspective on astronomy as you often use completely different techniques from the ones I have been exposed to in my bachelor.

I also want to thank my my official supervisor Alle-Jan for his guidance and ideas. Your thought-provoking comments have helped me shape the thesis in a much more clear way. Our meetings were an important quality check for me. I have learned a lot about simplifying my models and being more precise in my statistical descriptions and look forward to putting this to use in my PhD project in collaboration with you.

As a third point, I would like to thank Akio Taniguchi for the great discussions we have had on the project, even at times that were often late and inconvenient in Japan. I really appreciate the effort you have put into helping me in my thesis, despite me not being your own student. I hope you will enjoy reading the result.

Additionally, I want to thank both the Terahertz Sensing group and the Circuits and Systems group for including me and showing me the, often very different, perspectives on science. I have really enjoyed being welcomed into both groups and learning from both. Aside from the groups at Delft University of Technology, I would also like to thank the DESHIMA team and Astronomical Instrumentation subgroup as a whole. Through interactive discussions with the entire group you have given me, as a student, the opportunity to really be a part of the project and work as a scientist and I have enjoyed this immensely.

Aside from the people at University, I want to thank my family. First of all, I want to thank Alex for his amazing support, patience and love over the course of this (sometimes difficult) year. I would not have been able to do it without you and I cannot thank you enough. I also want to thank my parents for always encouraging me in my studies and supporting me, even when I had crazy ideas like starting a second bachelor in my third year. My brother, Joeri, has been an inspiration to me during this year and I want to thank you for your amazing stories and advice. I really enjoyed our weekly online stand-ups during the second peak of the pandemic.

Last but not least I want to thank my friends from both Delft University and Technology and Leiden University for your company over the years. ‘Studeren doe je samen’, even if you come from completely different countries or study very different subjects. Without you guys, studying definitely would not have been as much fun!

Stefanie Brackenhoff
Leiden, June 11, 2021

Contents

Nomenclature

Acronyms

ADC	Analog-to-Digital Converter
ADMM	Alternating Directions Method of Multipliers
ARIS	Astronomical Radio Interferometer Simulator
ASTE	Atacama Submillimeter Telescope Experiment
CW	(Room temperature blackbody) Chopper Wheel
DESHIMA	DEep Spectroscopic HIgh-redshift MApper
MKID	Microwave Kinetic Inductance Detector
PSD	Power spectral density
PSW	Position SWitch
PWV	Precipitable Water Vapor
SPLITTER	Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR
SVD	Singular Value Decomposition
TiEMPO	Time-dependent End-to-end Model for Post-process Optimization of DESHIMA
TLS	Two Level System

Mathematical Notation

$\circ$	Hadamard product
$\hat{\cdot}$	Estimate of $\cdot$
∇	Gradient operator
$\mathbf{A}$	(Upper case) matrix
$\mathbf{a}$	(Lower case) vector

Physical Qunatities

η_{atm}	Atmospheric transmittance
η_{fwd}	Forward efficiency
λ	Wavelength
ν	Center frequency of Microwave Kinetic Inductance Detector
τ	Optical depth
B_ν	Spectral brightness
c	Speed of light in a vacuum
D	Telescope dish diameter

f	Operating frequency in the noise spectrum
G	(Variable) gain
h	Planck constant
k_{B}	Boltzmann constant
$L_{\odot}$	Solar luminosity
P	Power
T_{amb}	Ambient temperature at the telescope
T_{ast}	Astronomical temperature (galaxy brightness temperature or zero)
T_{measured}	Brightness temperature measured by DESHIMA
T_{ph}	(Zero-mean) photon noise (around sky temperature)
T_{sky}	Noiseless sky brightness temperature
T_{TLS}	TLS noise
T_{μ}	Mean temperature of TLS noise plus photon noise
T_{σ}	Variation around mean temperature of TLS noise plus photon noise
T_A	Astronomical source brightness temperature with atmospheric absorption
T_A^*	Astronomical source brightness temperature without atmospheric absorption
$T_{\mathrm{blackbody}}$	(Room temperature) blackbody temperature
T_{noise}	Noise temperature of a heterodyne receiver
$T_{B,\mathrm{atm}}$	Brightness temperature of the atmosphere
$T_{p,\mathrm{atm}}$	Physical temperature of the atmosphere
z	Redshift

SPLITTER Parameters and variables

$\mathbf{L}$	Low-rank atmosphere matrix
$\mathbf{N}$	Residual noise matrix
$\mathbf{S}$	Source matrix
$\mathbf{X}$	Data matrix
$\boldsymbol{\theta}$	Line locations
$\boldsymbol{\nu}$	Vector of MKID center frequencies
$\boldsymbol{\xi}$	Continuum vector
$\mathbf{a}$	Spectral vector of $\mathbf{L}$ (Convex approximation method)
$\mathbf{b}$	Temporal vector of $\mathbf{L}$ (Convex approximation method)
$\mathbf{m}$	Temporal source indicator vector
$\mathbf{s}$	Source vector
$\mathbf{y}$	Dual variable in the Alternating Directions Method of Multipliers

$\mathbf{z}$	Central collector in the Alternating Directions Method of Multipliers
γ	Slack variable for the convex approximation method
ρ	Slack variable for the Alternating Directions Method of Multipliers
Θ	Number of spectral lines
l	Number of data blocks for the low-rank atmosphere matrix and residual noise matrix
M	Number of spectral channels
m	Indicator function for on-source measurement
N	Total number of (full-spectrum) measurements
n	Number of vectors in the Alternating Directions Method of Multipliers
t_l	Time support of of $\mathbf{L}_i$
k	Iteration number

1 Introduction

Astronomy and the technology used for astronomical observations have always developed together. Where astronomers used to simply measure using a telescope and the naked eye. Now, increasingly complex instruments are available to them. In imaging astronomy, photographic plates have been replaced by CCDs with a much higher sensitivity and more reliable linearity [1]. This has opened up many new science cases and simultaneously given rise to the demand for even better instruments. A similar development has occurred in spectroscopy, where the demand for more wideband and higher resolution spectrometers is ever present.

One of the frequency bands where such spectrometers are not yet widely available is the 100 GHz to 10 THz band [2],[3]. This is a waveband that contains much information about starburst galaxies [4], which are galaxies with a high amount of star-formation that are important astronomical targets. Although instruments that can observe in this waveband have been developed, they are often limited in bandwidth. Therefore, more technological developments are needed to truly close the gap. One of such efforts is the development of the DEep Spectroscopic HIgh-redshift MApper (DESHIMA). After succesful tests with a first generation DESHIMA instrument [5], DESHIMA 2.0 is now in development. This is a wideband spectrometer that is able to measure in 347 channels in the 220 to 440 GHz band [6]. This is an unprecedented bandwidth for a spectrometer in this frequency band.

1.1. Motivation

With the development of new instruments, new challenges emerge. The octave bandwidth of the DESHIMA spectrometer does not only collect data from the source it is observing, but also integrates radiation from the terrestrial atmosphere. Traditionally, atmospheric influences have been modelled using a linear model that could be normalized out [5],[7], but the spectrally nonlinear atmospheric absorption in the frequency band of DESHIMA 2.0 cannot be adequately modelled in this way. This results in a distorted galaxy spectrum, as shown in Fig. 1.1. To make matters worse, this distortion is temporally changing.

In order to remove the atmospheric components from the data, measurements of both the target area and empty sky are performed. This is done using a chopper-wheel called a Position SWitching wheel (PSW) that has a rotation frequency of a few hertz and can switch between two fields of view of the telescope. The telescope is positioned such that one field of view contains a galaxy and the other only empty sky. In the convential method, the empty sky measurements are then subtracted from the measurements that contain the source. This results in a spectrum similar to the orange line in Fig. 1.1. In order to estimate the blue line, a second Chopper Wheel (CW) is used, that switched between on-source measurements and calibration measurements. The multiplicative factor between the blue and the orange line can be estimated by the difference between the two positions of the second chopper wheel [8].

The conventional method introduces a factor $\sqrt{2}$ in the variance of the final estimate, due to the direct subtraction of two noisy spectra. A novel approach based on the statistical properties of the noise might improve upon this method, by limiting the transfer of the noise of one spectrum to the other. This can be done by exploiting what is known about the shape of the spectrum and trading some resolution for a

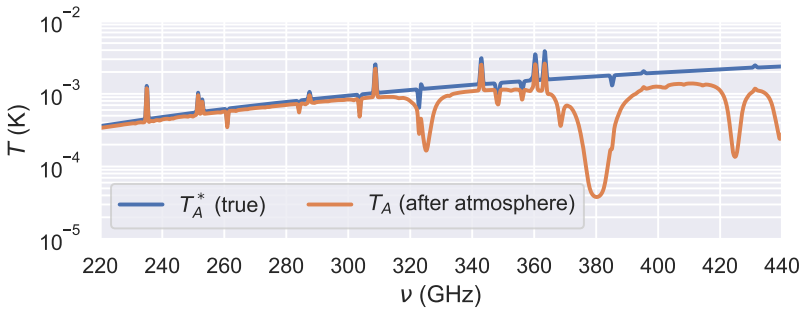


Figure 1.1. A spectrum of galaxy with and without atmospheric absorption. The blue curve is the spectrum of the galaxy itself, the orange curve shows the remainder of the signal after absorption in the atmosphere.

higher sensitivity in areas where this is possible.

A method based on statistical information is plausible due to the amount of measurements. A typical observation will last several hours at a sampling frequency of 160 samples per second [9].

1.2. Research Goals

The goal of this project is to formulate a new method for atmospheric noise removal from DESHIMA 2.0 spectroscopic data. This method includes a statistical data model and a noise removal method based on this model. Furthermore, the implications of this new method of noise removal for the optimal observing strategy are investigated. In line with this goal, the research question to be answered is:

> "Is there a way to efficiently remove the wideband atmospheric noise from DESHIMA 2.0 data in order to recover both the narrow line spectra and the broad continuum emission with a higher sensitivity than the current direct subtraction method?"

In order to answer this question, a data model for the DESHIMA 2.0 observations is formulated and an algorithm named Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR (SPLITTER) is designed. In this design, we take inspiration from Taniguchi et al. [10], who have created a similar method for a different type of spectrometer.

We assume the existing observing strategy, but an optimal observing strategy is also briefly touched upon in this thesis. This is because the optimal observing strategy is directly dependent upon the developed noise removal method. In the traditional method of on-target and empty sky subtraction for example, the switching between on- and off-source needs to happen sufficiently fast that the atmosphere has remained approximately constant between the two positions. This means that a fast mechanical chopper wheel is needed for position switching. A method that is able to predict the behaviour of the atmosphere during the on-source measurements based on the off-source measurements might be able to observe without the added complexity of this mechanical component.

1.3. Contributions

The contributions made in this thesis, in chronological order of execution, are:

- Contributions to a software package for simulating high redshift galaxies, named `GalSpec` [11].
- Contributions to a software package for an end-to-end simulation of the DESHIMA 2.0 spectrometer named `tiempo_deshima` and a conference contribution about this package [12]. The simulations from this package start with the simulated galaxy from `GalSpec` and include effects from the atmosphere, telescope and instrument. The output is the DESHIMA output power and calibrated sky temperature as measured by the frequency dependent detectors in the spectrometer.
- A data model tailored toward removing atmospheric noise from DESHIMA 2.0 measurements without requiring direct subtraction of two equally noisy spectra. This model describes the DESHIMA

output signal, including several of its noise properties in matrix form. The goal of this data model is to function as a starting point to formulate more efficient data reduction methods.

- An algorithm that processes DESHIMA 2.0 output data based on the data model and computes an atmosphere corrected galaxy spectrum. This algorithm is called Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR (SPLITTER).
- An analysis of the performance of SPLITTER on `TiEMPO` simulations in various situations. It is shown that SPLITTER is able to reduce the error in the galaxy estimate with a factor of up to ~ 1.65 for the entire spectrum. The noise removal is less efficient in spectral line areas, where the improvement factor is only up to ~ 1.3.
- A brief summary of the implications this noise removal method could have on the observing strategy used for future DESHIMA 2.0 measurements.

1.4. Outline

This thesis consists of 8 chapters. In Chapter 2, the relevant background information regarding the thesis topic is discussed, as well as current state-of-the-art wideband atmospheric noise removal methods. Using the knowledge from Chapter 2, a data model is developed in Chapter 3. This model describes how the statistical properties of the signal and various noise sources in the data relate to each other, such that these properties can be exploited later on. We explore the mean values and variances of these matrices numerically in Chapter 4. In Chapter 5, the data model is reformulated into various optimization problems with an accompanying solution strategy in order to formulate the algorithm. The properties of the matrices shown in Chapter 4 are compared to those same properties for the matrices that are estimated by running the various versions of the algorithm in Chapter 6. From these analyses, the best performing algorithm is selected. The performance of this algorithm in various situations is then further analyzed in Chapter 7. Finally, the most important findings of the thesis are summarized in Chapter 8, accompanied by recommendations for future work.

2

Background

This chapter explains the physical background of the optimization problem. First, the astronomical science cases and difficulties are explained to motivate the need for instruments such as DESHIMA. Secondly, the properties and working of DESHIMA itself are discussed. This section also contains a subsection on the data produced by DESHIMA as well as the calibration that is performed to remove most noise sources. Subsequently, the noise sources that cannot (easily) be removed with calibration and therefore need to be removed in post-processing are listed. Finally, current state-of-the-art methods for post-processing the data are explained. These methods set a standard for the sensitivity of the galaxy estimate, and are therefore the baseline that a new method has to compete with in order to be considered as a choice for post-processing method.

2.1. Radio Astronomy

To begin, we briefly discuss the reason why the waveband from 220 GHz to 440 GHz is of scientific interest and why there are not yet many instruments that observe over wide bandwidths in this area.

2.1.1. Science Targets

The most exuberant star forming regions within a galaxy are also naturally the parts of galaxies that contain many thick gas clouds and interstellar dust. As galaxies evolve, they accrete more matter, either by merging with other galaxies or due to the gravitational attraction of gas from their own viralized gas cloud [5]. As the gas and dust clouds within the galaxies become more dense, they collapse under their own gravitational pull and create stars [13]. The parts of the dust cloud that did not collapse into stars remain in the galaxy. These dust clouds obscure the optical and ultraviolet light that is emitted from heavily stellar populated areas [4]. This light is then re-emitted at lower far-infrared wavelengths.

Many of these galaxy forming processes occurred in an earlier stage of the Universe, in the order of 10 billion years ago [5],[14]. In the time that the light has been travelling from these distant galaxies to the Earth, the Universe has expanded. This expansion is described by Hubble's law, which states that distant objects move away from an observer, at a speed proportional to their distance. As the distant galaxies move away from the Earth in the expanding Universe, a Doppler shift towards longer wavelengths is introduced to their radiation, which is known as the cosmological redshift z:

$$z = \frac{\lambda_o - \lambda_s}{\lambda_s}, \tag{2.1}$$

where z denotes the dimensionless cosmological redshift, λ is the wavelength of the light in arbitrary units and the subscripts o and s represent the observed wavelength and the wavelength when the light was emitted by the source, respectively.

Specific atoms and molecules in the interstellar matter can absorb light through quantummechanical effects, resulting in absorption lines in the galaxy spectrum. Furthermore, emission also happens at quantized frequencies, resulting in emission lines where the spectrum is brighter.

Many spectral lines of interest for studying distant galaxies are found in the (sub)millimeter regime, due to the emission and absorption in the interstellar medium and the cosmological redshift. As such,

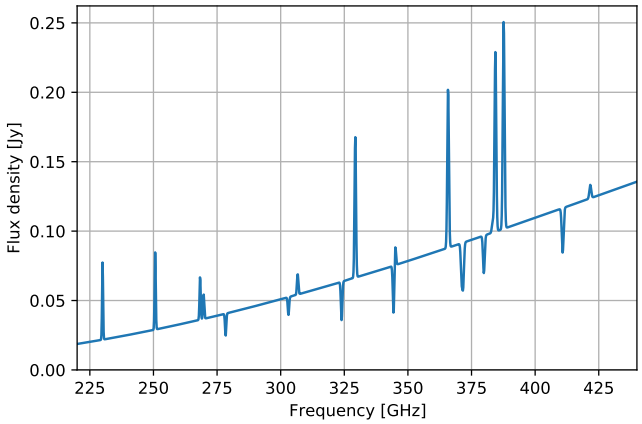


Figure 2.1. Example of a galaxy spectrum with blackbody radiation and spectral lines. The galaxy has a redshift of 2 and a luminosty of $L = 10^{13.7} L_{\odot}$. Figure generated using `GalSpec` [11].

spectral measurements, especially at $z \geq 4$, are needed to properly create a model of the amount of dust-obscured star formation in the early Universe [15]. This will shed more light on the history of cosmic star formation.

2.1.2. Source Description

The source signal from a galaxy before it reaches Earth consists of two components: a broad continuum and narrower spectral lines. This results in a spectrum such as shown in Fig. 2.1. In the scope of this thesis, it is mainly important to know that the spectrum consists of these two shapes, where the continuum extends across the entire spectrum and is smooth, but the spectral lines are only visible in a few neighbouring spectral channels. The data model and SPLITTER algorithm are designed to be applicable to generic galaxies. As such, there are no firm constraints on the type of radiation that can be in a galaxy. For the interested reader, a short description of the emission mechanisms inside high-redshift galaxies is given below, however.

The broad continuum consists mainly of blackbody radiation[1]. This radiation is caused by the interstellar dust in thermal equilibrium with the incoming radiation emitted by stars. The dust then radiates following Planck's law,

$$B_\nu(T) = \frac{2h\nu^3/c^2}{\exp(h\nu/k_\mathrm{B}T) - 1}. \tag{2.2}$$

In this equation, B_ν is the spectral brightness in units of power per unit area per unit frequency per unit solid angle. The number h is the Planck constant, c the speed of light and k the Boltzmann constant. The frequency of the light is denoted by ν. The dust in the galaxies of interest typically has temperatures in the order of 10 K. In the submillimeter wavelength regime, this means that the radiation is in the Rayleigh-Jeans limit of $h\nu \ll kT$. Expanding the exponential then results in the Rayleigh-Jeans law,

$$B_\nu(T) = \frac{2\nu^2}{c^2} k_\mathrm{B} T. \tag{2.3}$$

Aside from the continuum, there are the sharp spectral lines. As a part of this thesis, the author contributed to the open source `GalSpec`[2] package which can be used to model the blackbody continuum and spectral lines of a high-redshift galaxy. The model contains spectral lines from CO, SIII, SiII, OIII, OI, NIII, NII, CII and CI with various ratios with respect to the continuum. The line amplitudes of these lines are based on [16]–[19].

[1] In reality, there is also a power law contribution of synchrotron radiation and a nearly flat contribution of free-free emission at these frequencies. However, these contributions are smaller than that of blackbody radiation, and not included in the simulations used in this thesis.

[2] `10.5281/zenodo.4279061`

2.1.3. Terahertz Gap

Unfortunately, the submillimeter wavelength regime of interest is located inside the terahertz gap, which ranges from 100 GHz to 10 THz [2],[3], corresponding to wavelengths from 30 μm to 3 mm. This 'gap' is a name for a region in the electromagnetic spectrum, for which it is very difficult to create good detectors.

Coherent receivers are able to perform high resolution spectroscopy in the terahertz gap, but are not scalable to large bandwidths or are not able to create a larger image by multiplexing to many pixels at once. Due to this, many instruments that can cover a larger range of wavelengths in the submillimeter regime need to be tuned between observations to be able to access different wavebands, resulting in a very time-costly observation. An instrument that is currently operational and suited for this type of observation is the Atacama Large Millimeter Array (ALMA), but observing time on this instrument is very scarce. Therefore, obtaining many wideband spectra with ALMA is not a feasible endeavour [20]. This means that there are currently no extensive maps of the sky available across broad frequency ranges in the terahertz gap, limiting our knowledge of the Universe in the redshift range of dust-obscured star formation. Because of this, a lower resolution solution without employing coherent reception, but rather direct detection needs to be investigated. Such an instrument requires a filter or spectrometer between the collector and detector in order to separate the different wavebands.

Possible existing solutions include a grating, a Fabry-Pérot interferometer and a Fourier transform spectrometer [21]. Scaling a grating up to larger multipixel arrays is difficult as multiple gratings need to be installed which increases the size of the instrument. Grating components need to be cooled down to a few kelvins in order to function with a high enough sensitivity, but building cryostats at these sizes is not (yet) possible. Therefore, the grating cannot be scaled to a larger pixel array whilst also being wideband. The Fabry-Pérot interferometer and a Fourier transform spectrometer can image more naturally, but also need to be large for the frequencies of interest. Furthermore, although they can observe with multiple pixels at once, they cannot instantaneously obtain a full spectrum. A Fabry-Pérot interferometer observes one frequency band at a time and then needs to be adjusted to move on to the next. This is similar for the Fourier transform spectrometer, except this measures one Fourier component at a time. Because of this, they achieve a lower sensitivity than a filterbank in the same amount of observation time when obtaining a full spectrum for a point source. Therefore, traditional instruments are either very narrow band, are not scalable to larger pixel arrays, or have sensitivity penalties.

A new instrument that can measure wideband spectra with multiple pixels simulateneously whilst maintaining a higher sensitivity is therefore desired. A possible solution is a filterbank that has a smaller volume than a grating, as filterbanks naturally lend themselves to observing full spectra. If a filterbank can be made sufficiently small, it is possible to put multiple filterbanks in the focal plane. In this way, each filterbank can function as one pixel that observes a full spectrum. This solution allows for mapping that is fast, has high sensitivity and can be used for measuring broadband spectra in the terahertz regime.

2.2. The Instrument

A step on the road to such an instrument is the integrated filterbank of the DEep Spectroscopic HIgh-redshift MApper (DESHIMA). It uses the novel technology of microwave kinetic inductance detectors. This technology is easily integrated on a chip, as only one readout line is needed to read out many different channels. Therefore, in later versions the instrument will be scalable to multiple pixels.

In this section, the signal pipeline of DESHIMA is described, including the telescope in which it is installed. The section follows the same pipeline as the signal, starting at the point where it first reaches the telescope and ending in the readout signal.

2.2.1. The Atacama Submillimeter Telescope Experiment

The Atacama Submillimeter Telescope Experiment (ASTE) is a single dish Cassegrain focus telescope located in the Atacama desert in northern Chile [22]. The ASTE site is favorable, as it is at a high altitude (4,860 m above sea level), which means the atmosphere above it is thin compared to the atmospheric layer above a telescope at sea level. Additionally, the air above the Atacama desert is extremely dry. Both these factors reduce the amount of water vapor above the telescope and therefore make it easier to observe distant galaxies.

Galaxies that are potential science targets for DESHIMA can be treated as point sources, which means that only one telescope pointing is needed to capture their light. This is because the diameter of the telescope dish is 10 m. For light with a frequency of 300 GHz, corresponding to a wavelength of

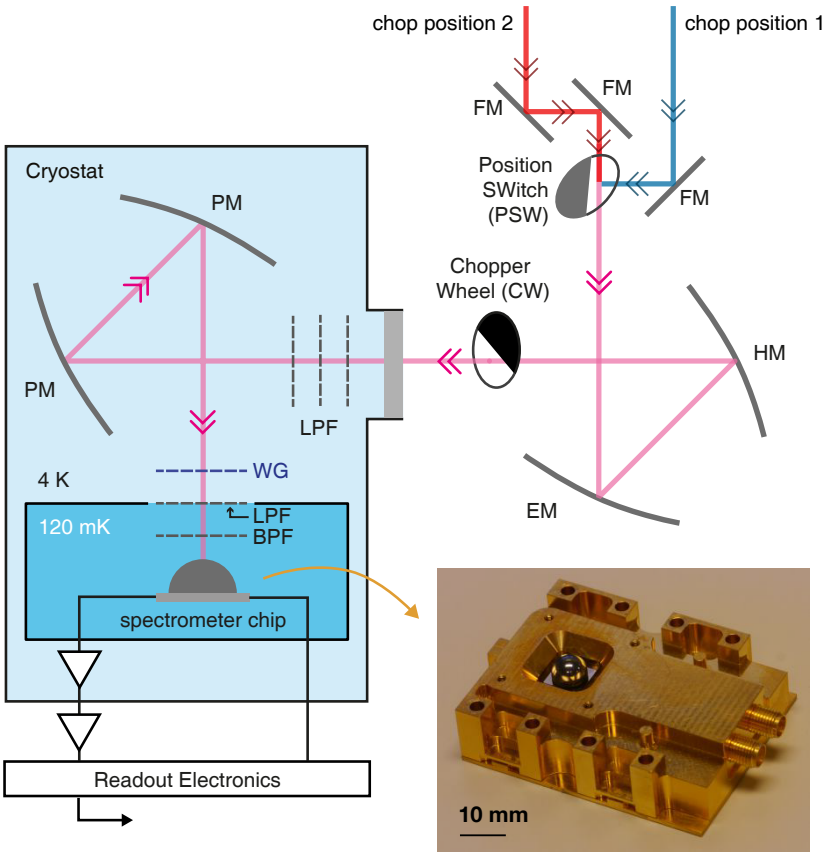


Figure 2.2. The signal pipeline inside the ASTE telescope. Light comes in from the sky, where it is directed to the Position SWitching chopper wheel (PSW) using Flat Mirrors (FM). There are two different paths to the PSW wheel, corresponding to two possible on-sky chop positions. 50% of the wheel is reflective, if the reflective part is in the signal path, the light from chop position 1 is reflected into the telescope (blue path). The other half is transparent, if this is in the signal path, the light from chop position 2 is transmitted into the telecope. After passing the PSW wheel, the light is first focused further into the telescope using the two mirrors in the Cassegrain focus. Subsequently, it is reflected by an ellipsoidal mirror (EM) and a hyperbolic mirror (HM) in order to focus it into the 4 K cryostat. Before reaching the cryostat, the light must pass through the room temperature blackbody Chopper Wheel (CW). Inside the cryostat, the light passes through a low-pass filter (LPF), is reflected by two parabolic mirrors (PM) and passes through a wiregrid (WG) before reaching the colder 120 mK cryostat when it reaches the second low-pass filter. Finally, it is filtered by a band-pass filter (BPF) and guided into the DESHIMA instrument. The enlarged picture is a photograph of the DESHIMA 1.0 instrument. Figure adapted from [5].

~ 1 mm, this means that the angular resolution of the telescope is $\sim \lambda/D = 10^{-4}$ rad or $\sim 20''$ [23]. As this number is significantly higher than the angular size of most high-redshift galaxies, light emitted from every part of the source that is collected by the dish couples to a single pixel.

2.2.2. Chopper Wheels

The path of the light inside the telescope is shown in Fig. 2.2. A very important part of this setup are the two chopper wheels, namely the Position SWitching wheel (PSW) and the blackbody Chopper Wheel (CW).

The PSW wheel is used to rapidly change the on-sky position of the telescope. The chopper wheel alternates the line of sight of the telescope between the on-source position and an off-source position nearby. The positions are chosen such that the primary telescope beams of the pointings overlap mostly within the terrestrial atmosphere, but the beams are spatially separated in the far field. This is illustrated in Fig. 2.3. As the atmosphere is mostly shared between the two beams, the empty sky around the galaxy can be estimated using the off-source measurement. The wheel rotates such that the beam is pointed off-source and then back on-source again 10 times per second[3]. Combined with the readout

[3]The wheel is divided into four quarters and rotates at 5 rotations per second. This achieves the same chopping frequency as a

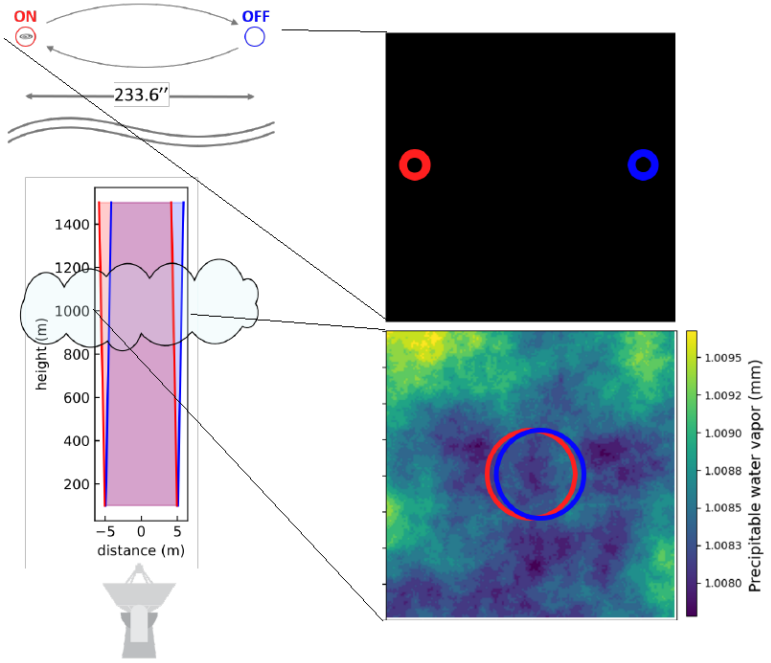


Figure 2.3. Illustration of how the chopper wheel is used to estimate the empty sky brightness. **Left:** The telescope pointings for the on- (red) and off (blue) positions. **Top right:** Primary beam outline in the far field. The beams, denoted here as colored circles, are spatially separated in this regime. **Bottom right:** The same beams, at 1 km above the telescope. The primary beams largely overlap in this regime, which means that the atmospheric brightness in one beam can be estimated using the brightness in the other. Figure not to scale. Figure adapted from [24].

frequency of 160 samples per second, this means that during each rotation of the chopper wheel, 8 samples on-source are obtained subsequently, followed by 8 samples off-source[4]. In Section 2.3.1, we describe how the off-source measurements are used to estimate the galaxy brightness temperature.

Due to the position switching, the light follows slightly different paths through the telescope for the on- and off-source measurements. As an effect, the main beam efficiency can differ between the chopping positions. In order to compensate for this, the entire telescope is nodded at a set time interval that is larger than the chopping interval [25]. The nodding is done such that the on-source position is moved to the other 'side' of the chopper wheel after nodding. In this way, the path that the light follows for the off-source chopping position at one nod position now becomes the path of the light for the on-source position at the other nod position (see Fig. 2.7 for an illustration). This chopnodding scheme is referred to as ABBA nodding[5].

The CW chopper wheel at the entrance to the cryostat is used for calibration of the temperature that is measured by DESHIMA. Half of this wheel lets the light pass through unchanged, the other half blocks it. This blocking half wheel behaves like a blackbody at room temperature, such that the light emitted from this half of the wheel can be described using Eq. (2.2). Since this temperature is known, measured brightness temperatures can be calibrated using the measurements of the opaque half of the wheel.

2.2.3. DESHIMA

After reaching the cryostat, the light is guided onto the DESHIMA superconducting integrated filterbank. A description of how this filterbank operates can be read in Appendix A. The most important part to note here is that DESHIMA is a direct detector. It uses Microwave Kinetic Inductance Detectors (MKIDs) to

wheel that is divided into two halves and rotates 10 times per second.

[4] It should be noted that in reality, some samples near the moment the chop is performed will be unusable due to diffraction if the edge of the chopper wheel is still too close to the aperture at the moment the sample is taken. This effect is not considered in the scope of this thesis, however.

[5] This is not an acronym. The letters refer to nod positions A and B. When the telescope has moved from nod A to nod B it stays there for several chopping cycles before returning back to nod A. For this reason, the two Bs follow each other in the name of the method.

measure the spectrum. These MKIDs change resonance frequency when they are optically loaded [26]. This shift in resonance frequency can be read out using frequency division multiplexing on a readout line. In operation, DESHIMA performs a sweep of the full filterbank every 6.25 ms. This is a sampling frequency of 160 Hz[6]. The phase shift in the resonance frequencies can directly be converted to a sky brightness temperature. This is fundamentally different from coherent receivers that often measure a voltage, which needs to be squared to compute a power.

In order to ensure the phase shift translates to the correct sky brightness temperature, DESHIMA needs to be calibrated. This calibration is done using the measured temperature of the CW blackbody (CW in Fig. 2.2) and a skydip line of sight measurement. This is an iterative procedure, where the MKID readout signal is compared to the expected sky brightness at an interval of elevations, from 32° to 88° above the horizon. Using this procedure, a responsivity model is created toconvert the MKID readout signal to sky brightness temperature [27]. This calibration is performed every 1 to 2 hours, in order to ensure the telescope remains calibrated. The aperture efficiency is also measured, using a planet scan [9].

2.3. Simulations of Data and Noise Sources

Despite the calibrations performed on DESHIMA, some systematic errors remain, as well as some noise sources. In the scope of this thesis, we focus on three main noise sources, namely atmospheric noise, photon noise and two level system noise. In total the measured brightness temperature can be written as:

$$T_{\text{measured}}(\nu, t) = T_{\text{sky}}(\nu, t) + T_{\text{ph}}(\nu, t) + T_{\text{TLS}}(\nu, t) \tag{2.4}$$

Where $T_{\text{measured}}(\nu, t)$ is the total measured brightness temperature, $T_{\text{sky}}(\nu, t)$ is the deterministic brightness temperature of the sky. This is composed of the brightness of the astronomical source that we want to measure and noise introduced by the atmosphere. $T_{\text{ph}}(\nu, t)$ is the photon noise, which is the random variation of the sky brightness temperature around the mean $T_{\text{sky}}(\nu, t)$. $T_{\text{TLS}}(\nu, t)$ is called Two Level System (TLS) noise. Below, these three noise sources are explained in more detail, as they are important for the development of the data model. After these three main noise sources, a short list of errors and effects that are not in the scope of this thesis is given.

As there is no DESHIMA 2.0 data available yet, it is not possible to apply the methods discussed in this thesis on real on-sky data. Instead, simulations were made using `TiEMPO`[7] [12]. This is an open source Python package for simulating the full signal pipeline in DESHIMA. The author has contributed to this package as a part of this thesis. A `TiEMPO` simulation starts with a `GalSpec`[8] simulation of a galaxy, and includes a simulation of the atmosphere (including different telescope pointings), the optical chain and the detectors. In this way a realistic simulation of DESHIMA observations can be obtained. `TiEMPO` does not include two level system noise, but a model for the TLS noise based on work by Kaushal Marthi [28] is used to create a noise source with the required statistical properties. This noise is generated and added in the power domain and then converted to a brightness temperature using a simulation of the skydip calibration.

2.3.1. Atmospheric Noise

The atmospheric 'noise' is the dominant noise source in the signal. It is very difficult to remove, as it is temporally correlated and nonlinear between the channels of the detector. Furthermore, it appears in both an additive and a multiplicative term, which renders noise removal even more difficult.

The atmosphere noise is caused by the frequency dependent absorption of incident light in the atmosphere. This absorption is the multiplicative term applied to the signal, as it dims the galaxy at frequencies with much absorption. On the other hand, the atmosphere is also an emitter itself, causing additive noise. This can be summarized by the radiation transfer equation [29]:

$$T_{\text{sky}}(\nu, t) = T_A^*(\nu) e^{-\tau(\nu,t)} + T_{p,\text{atm}}(t)\left(1 - e^{-\tau(\nu,t)}\right). \tag{2.5}$$

[6] When mentioning a 'sample' in this thesis, we are referring to a full spectrum. This means that a sample contains measurements in all 347 channels of the DESHIMA spectrometer.

[7] 10.5281/zenodo.4279085

[8] 10.5281/zenodo.4279061

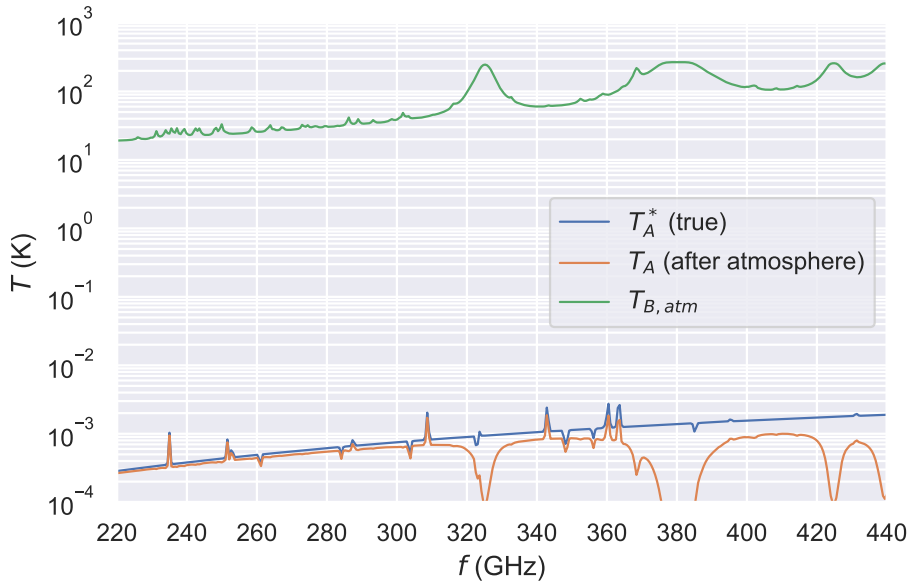


Figure 2.4. Contrast between the atmospheric brightness temperature $T_{B,\mathrm{atm}}$ (green) and the source signal after transmission through the atmosphere (orange). The original signal arriving at the atmosphere is also shown in blue. waveband is chosen to correspond with that of DESHIMA 2.0. In this figure, a galaxy at redshift 2.2 with a luminosity of $10^{13.7} L_\odot$ is used. The atmosphere has a Precipitable Water Vapor (PWV) level of 1.3 mm.

In this equation, T_{sky} is the measured brightness temperature in K, T_A^* is the brightness temperature of the astronomical object (a galaxy), where the $*$ denotes that atmospheric effects are removed[9], also in K. The optical depth of the atmosphere is denoted by the dimensionless τ. If the optical depth is very large, the mean free path length of a photon travelling through the atmosphere is very small compared to the thickness of the atmosphere layer and much of the incident light will be absorbed. Finally, $T_{p,\mathrm{atm}}$ denotes the *physical* temperature of the atmosphere in K. Note the difference with the spectral *brightness* temperature of the atmosphere, $T_{B,\mathrm{atm}}(\nu, t) = T_{p,\mathrm{atm}}(t)\left(1 - e^{-\tau(\nu,t)}\right)$.

Instead of the optical depth, the atmospheric transmittance $\eta_{\mathrm{atm}}(\nu, t)$is often used to write Eq. (2.5) more compactly. The atmospheric transmittance is a number between 0 and 1 that describes what fraction of incident light is able to travel through the atmosphere to the telescope dish and is computed with $\eta_{\mathrm{atm}}(\nu, t) = e^{-\tau(\nu,t)}$. Using this transmittance, the radiation transfer function can be written as

$$T_{\mathrm{sky}}(\nu, t) = T_A^*(\nu)\eta_{\mathrm{atm}}(\nu, t) + T_{p,\mathrm{atm}}(t)(1 - \eta_{\mathrm{atm}}(\nu, t)). \tag{2.6}$$

This second form of the radiative transfer equation is the form we will use throughout this thesis. There are some frequency bands where the transmittance is close to 1 and most of the light from the source can pass through the atmosphere. These bands are called atmospheric windows. Outside the windows, the signal is much harder to detect. This is because of two reasons: first of all, more of the signal emitted by the galaxy is absorbed, so the source signal at the telescope is weaker. Secondly, the atmosphere itself is radiating more energy in these low transmission regions. Therefore, the areas where the signal is at its weakest are also the areas where the noise is the strongest. This is illustrated in Fig. 2.4, where the green line is the brightness temperature of the atmosphere, and the orange line shows what is left of the galaxy signal after transmission through the atmosphere.

Now that we can express the brightness temperature of the light arriving at the telescope, we can look at its variation in frequency and time. The brightness temperature of the astronomical source can be viewed as static in time, as the timescales in which galaxies evolve are much larger than relevant for an observation, or even several observations over several years. There is a frequency dependence, however, as described in Section 2.1.2. The physical temperature of the atmosphere is frequency inde-

[9]A convention in astronomy, where the observed signal usually still contains atmospheric effects, so the removal of these effects must be explicitly stated.

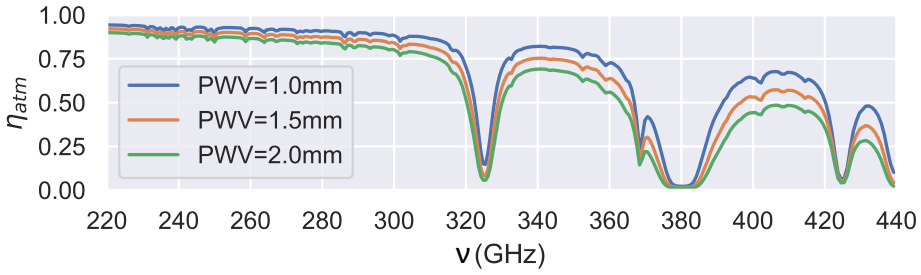


Figure 2.5. Atmospheric transmittance as a function of frequency within the DESHIMA 2.0 bandwidth for three different levels of Precipitable Water Vapor (PWV).

pendent as it is a physical temperature rather than a brightness temperature. It is slightly time dependent, as the outside temperature can vary by a few kelvins over multiple hours [30].

As the variation of the physical temperature of the atmosphere is slow, both the temporal and spectral variation of the atmospheric noise are dominated by the atmospheric transmittance. This transmittance varies mostly based on the variation in water vapor content in the line of sight of the telescope [31]. Even at a dry and elevated site such as that of ASTE, water vapor influences the atmospheric transmission of the atmosphere drastically [32]. In this thesis, the metric used for water vapor is the Precipitable Water Vapor (PWV), a measure for how many millimeters of rainfall would be observed if all the water in the atmosphere column above the telescope would be precipitated as rain. A higher PWV indicates more water in the atmosphere.

Although the change in atmospheric transmittance is dependent on the change in PWV, the conversion between PWV and transmittance is not straightforward and there is no single equation to describe it in closed form for long timescales. The frequency dependence of the atmospheric transmittance for several levels of water vapor is shown in Fig. 2.5.

Although there is no direct conversion between the PWV and the atmospheric transmittance for the full spectrum, the optical depth in a single channel can be written as a linear function of the PWV [10],[33],[34]:

$$\tau(\nu, t) = -\ln \eta_{\mathrm{atm}(\nu,t)} \approx a(\nu) PWV(t) + b(\nu) \tag{2.7}$$

The parameters of this function, $a(\nu)$ and $b(\nu)$ depend on the weather conditions and channel frequency and are not known prior to the observation. Using this equation, the change in optical depth for the full spectrum can be described with the change in PWV, on short timescales [31]. For longer timescales, the linearization is no longer valid. In `TiEMPO`, the atmospheric transmittance is simulated used the model created by Pardo et al. [35]. More information on how the temporally changing PWV is created is given in Appendix B. The first order and second order moments of the PWV are dependent on weather conditions, the time of day, the time of the year and even variations between years. At the telescope site, the mean PWV is expected to typically be around 1.2 mm [36]. A typical standard deviation for the PWV used in this thesis is $\sim 1.5 \cdot 10^{-2}$ mm [24], which corresponds to a variance of $\sim 2.3 \cdot 10^{-4}$ mm. This number is small, but leads to variances up to a Kelvin in the atmospheric brightness temperature.

2.3.2. Photon Noise

Aside from the atmospheric noise, there are also the two additive noise sources, the first of which is the photon noise $T_{\mathrm{ph}}(\nu, t)$. This photon noise is a fundamental noise, that is not caused by the detector but the incoming light itself. In reality, the photon noise is a part of the signal, such that the signal has a mean value of the sky temperature, with a variance characterized by the photon noise. Here, we treat photon noise as a zero mean noise source added to the sky temperature instead.

Photon noise is created by shot noise and photon bunching. Shot noise is caused by the fact that photons are only measured at the time of their arrival at the detector. This arrival time is discretized in time and the arrivals occur at random time intervals. It is thus subject to a Poisson distribution. Photon bunching occurs because of the wave nature of the incoming light. Photons can interfere with each other

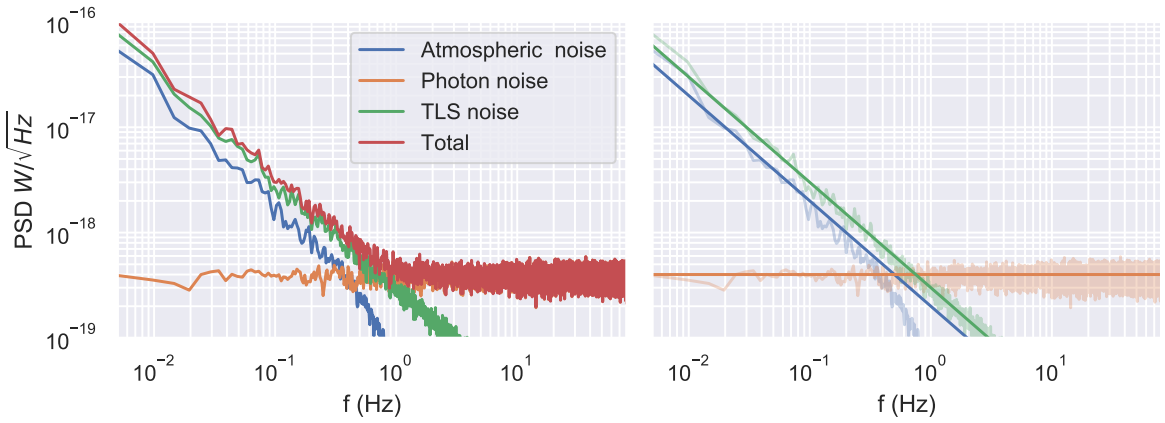


Figure 2.6. Frequency noise spectrum of the different noise sources in a DESHIMA channel near 300 GHz. Note that the frequency axis shown here is an operating frequency axis (denoted by f), rather than the frequency of the light incident on the detector (denoted by ν). **Left**: PSDs computed from `TiEMPO` data using the Welch method. **Right**: The PSDs of the different noise sources with an overlay of their true spectrum.

at the detector and arrive in 'bunches' because of this, which also gives rise to a Poisson distribution.

The number of photons arriving at the detector is very large and the sampling frequency of 160 Hz that DESHIMA uses is much smaller than the effective bandwidth of the filters. Therefore, the central limit theorem can be used to approximate the Poisson distribution of the photon noise with a Gaussian distribution [24]. As such, all photon noise in `TiEMPO` is modeled as a Gaussian. The standard deviation of this Gaussian depends on the power incident upon an MKID. As the shot noise power increases with the square root of the incident power and the photon bunching noise increases with the incident power itself, a higher incident power in a channel also results in more photon noise at that detector[10]. As discussed above, the Gaussian is separated, such that the variance caused by shot noise and photon bunching is modeled as a zero mean additive noise source. The mean of this Gaussian is put into the sky temperature as the signal we want to detect.

Naturally, an observation is much longer than a single sample. When we estimate the galaxy temperature, we average over many measurements in time. If $\sigma_{T(\nu)}$ denotes the standard deviation of the detected temperature in a single sample and we integrate over N measurements, the standard deviation on the time averaged temperature $\overline{T(\nu)}$ is $\sigma_{\overline{T(\nu)}} = \sigma_{T(\nu)}/\sqrt{N}$ due to the Gaussian properties of the noise. This is only the case if the underlying temperature $T(\nu)$ stays constant, however. Therefore, time-varying effects need to be removed prior to averaging.

2.3.3. Two Level System Noise

Another noise source in the spectrum is the Two Level System (TLS) noise $T_{\mathrm{TLS}}(\nu, t)$. This is noise contribution is caused by small electric defects [37]. The TLS noise is spectrally uncorrelated, since it is created in each channel independently. It is temporally correlated, however. Its power spectral density drops off reciprocally with frequency. Therefore, the noise drops below the photon noise level if the sampling frequency is sufficiently high.

2.3.4. Noise Comparison

In order to understand how the different noise sources behave in time and what noise sources dominate at what operating frequencies, it is helpful to look at the power spectral density (PSD) of a `TiEMPO` simulation. This is plotted in Fig. 2.6. The PSD has been computed using the Welch method [38] on the detected power in a single MKID with a center frequency near 300 GHz. The spectra are noisy as only a finite amount of data (30 mins) was used. The operating frequency is the effective frequency in the noise spectrum. When using the direct subtraction method, the operating frequency is the reciprocal of the time between two samples that are subtracted from each other. Using the PSW wheel, the operating frequency is 10 Hz for DESHIMA 2.0.

[10] The interested reader is referred to [12] for a more thorough treatment of the Poisson and bunching noise.

At different operating frequencies in the spectrum, different noise sources dominate. The photon noise is white and therefore has a constant noise power across all frequencies. The atmospheric noise and TLS noise, on the other hand, drop off in a $1/f$ shape. The atmosphere noise drops off faster past an operating frequency of approximately 0.5 Hz in this plot. This is not a physical effect, but due to discretization effects in `TiEMPO` and `ARIS` (see Appendix C for a more detailed explanation).

Because of the difference in the shape of the power of the photon noise on the one hand, and the TLS and atmospheric noise on the other, the total noise has a knee around 1 Hz. In practice, the exact location of the knee for both the TLS and the atmospheric noise can be at a slightly different operating frequency, but the overall shape of the spectrum remains. This means that different noise sources dominate at different timescales. At time differences of longer than $\sim T_{\mathrm{knee}} = 1/f_{\mathrm{knee}} = 1$ s, the temporal correlation of the TLS noise and atmospheric noise causes an error in the measurement that is larger than the noise introduced by the photon noise.

In order to exploit the photon noise limited properties of the MKID detectors in DESHIMA and obtain the best estimate possible, samples that are directly compared should therefore not be further apart than T_{knee}. In the current method of direct subtraction, this means that the on- and off-chopping of the telescope must occur with a frequency of ≥ 1 sample per second. This is also the reason why the chopper wheel rotates at 10 cycles per second, as this is safely above the knee frequency of both $1/f$ noise sources.

2.3.5. Remaining Errors

Aside from the noise sources listed above, some systematic errors remain ,as well as noise sources that can be neglected compared to the noise sources mentioned above. These errors and noise sources are not analyzed in the scope of this thesis, but can be relevant to the interested reader. For completeness, they are briefly listed below.

Responsivity shape: Inside `TiEMPO`, the responsivity of all MKIDs is assumed to be of Lorentzian shape. In reality, the responsivity has a more complex shape [5]. This shape influences the waveband over which the MKID integrates, as well as its peak frequency. A different responsivity therefore needs to be carefully calibrated for, or it will introduce systematic errors in the spectral estimates.

Different telescope paths: The different telescope paths for the two different positions of the chopper wheel are not included in `TiEMPO` and therefore the systematic error caused by the difference in optical efficiency and main beam efficiency is not included in the simulations. The chopping and nodding behaviour of the telescope can be included in terms of sky positions, however.

Readout Noise: The readout line of DESHIMA is connected to an Analog to Digital Converter (ADC), which introduces ADC noise. Additionally, there is noise introduced by the amplifier. This noise is zero mean and white and expected to be at a lower level than the photon noise [28]. Therefore, the noise is dominated by other noise sources than the ADC noise for all sampling frequencies and the ADC noise can safely be omitted from the simulations.

2.4. Existing Data Reduction Methods

Generally, astronomers reduce photon noise by using longer integration times and approximate atmospheric effects with a baseline that is linear in frequency and can be subtracted to remove atmospheric effects. In the wide bandwidth of DESHIMA 2.0, however, the atmosphere cannot accurately be approximated using a model that is linear in frequency (see Fig. 2.5). Therefore, more sophisticated data reduction models are being developed. Below I give an overview of a currently existing post processing methods, along with a method similar to ours that is still in development.

2.4.1. Direct Subtraction

In direct subtraction chopping and nodding, the noisy empty sky signal is directly subtracted from the source signal. See Fig. 2.7 for an illustration of the sky positions at the different telescope setups. Note that the telescope can be in two different positions called nodding positions (denoted by A and B). These two nodding positions both have two different pointings, called chopping positions (denoted by a and b). In this way, both nod A chop b and nod B chop a are on-source positions. Nod A chop a and nod B chop b on the other hand are off-source. Below, we use subscripts to denote the different chopping and nodding positions of several variables. The uppercase subscript denotes the telescope nod, and the lowercase subscript denotes the PSW chopping position. At a single point in time, the sky temperatures

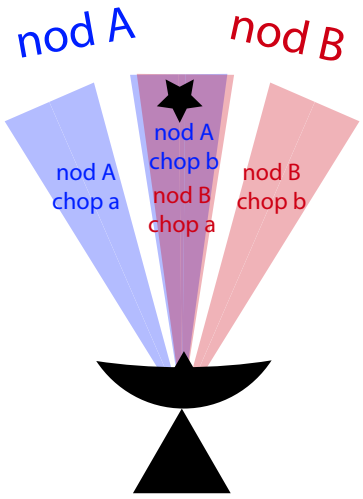


Figure 2.7. Chopping and nodding. The blue beam indicates the possible fields of view for nodding position A. If the chopper wheel is in position a, the left side (blue) of this beam is observed. If the chopper wheel is in position b, the right side (purple) is observed. When the dish is put in nodding position B, the red beam represents the possible pointings. Nodding position A with the chopper wheel in position b and nodding position B with the chopper wheel in position a both point towards the target galaxy (shown as a star here).

measured at the different positions can be described as:

$$\begin{aligned}
T_{\mathrm{Ab}}(\nu,t) &= (1-\eta_{\mathrm{atm,Ab}}(\nu,t))T_{p,\mathrm{atm}}(t) + \eta_{\mathrm{atm,Ab}}(\nu,t)T_A^*(\nu) + T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t),\\
T_{\mathrm{Aa}}(\nu,t) &= (1-\eta_{\mathrm{atm,Aa}}(\nu,t))T_{p,\mathrm{atm}}(t) + T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t),\\
T_{\mathrm{Ba}}(\nu,t) &= (1-\eta_{\mathrm{atm,Ba}}(\nu,t))T_{p,\mathrm{atm}}(t) + \eta_{\mathrm{atm,Ba}}(\nu,t)T_A^*(\nu) + T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t),\\
T_{\mathrm{Bb}}(\nu,t) &= (1-\eta_{\mathrm{atm,Bb}}(\nu,t))T_{p,\mathrm{atm}}(t) + T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t). \qquad (2.8)
\end{aligned}$$

The telescope cannot observe instantaneously, so the dependence on t will be replaced by a subscript that denotes the sample, i. So if sample i is collected at time t_i, it is denoted $T_{\mathrm{Ab},i}(\nu)$, rather than $T_{\mathrm{Ab}}(\nu,t_i)$. Since both the telescope and the PSW wheel can only be in one position at a given time, the measured sky temperature will always be only one of the temperatures in Eq. (2.8). In order to denote this, an indicator function for the position is introduced. This function is denoted as $m_{\mathrm{Ab},i}$ for sample i in position Ab. The function is equal to 1 when the telescope is in position Ab and zero otherwise.

Subsequently, the off-positions Aa and Bb are subtracted from the on-positions Ab and Ba and a sum is taken over all samples. Since the galaxy is present in half of the N samples, the sum is multiplied by $2/N$. Effectively, this is the mean of the on-source positions minus the off-source positions.

$$\hat{T}_A(\nu) = \frac{2}{N}\sum_i T_{\mathrm{Ab},i}(\nu)m_{\mathrm{Ab},i} + T_{\mathrm{Ba},i}(\nu)m_{\mathrm{Ba},i} - T_{\mathrm{Aa},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{Bb},i}(\nu)m_{\mathrm{Bb},i}$$

Plugging in the equations listed above gives:

$$\begin{aligned}
\hat{T}_A(\nu) = \frac{2}{N}\sum_{i=1}^{N} &\left[(1-\eta_{\mathrm{atm,Ab},i}(\nu))T_{p,\mathrm{atm},i} + \eta_{\mathrm{atm,Ab},i}(\nu)T_A^*(\nu) + T_{\mathrm{ph},i}(\nu) + T_{\mathrm{TLS},i}(\nu)\right]m_{\mathrm{Ab},i}\\
&+\left[(1-\eta_{\mathrm{atm,Ba},i}(\nu))T_{p,\mathrm{atm},i} + \eta_{\mathrm{atm,Ba},i}(\nu)T_A^*(\nu) + T_{\mathrm{ph},i}(\nu) + T_{\mathrm{TLS},i}(\nu)\right]m_{\mathrm{Ba},i}\\
&-\left[(1-\eta_{\mathrm{atm,Aa},i}(\nu))T_{p,\mathrm{atm},i} + T_{\mathrm{ph},i}(\nu) + T_{\mathrm{TLS},i}(\nu)\right]m_{\mathrm{Aa},i}\\
&-\left[(1-\eta_{\mathrm{atm,Bb},i}(\nu))T_{p,\mathrm{atm},i} + T_{\mathrm{ph},i}(\nu) + T_{\mathrm{TLS},i}(\nu)\right]m_{\mathrm{Bb},i}.
\end{aligned}$$

Note that the TLS noise and the photon noise are independent of the sky-position and therefore have not been given a subscript. Reordering using $m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i} + m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i} = 1 \forall i$ gives

$$\begin{aligned}\hat{T}_A(\nu) = \frac{2}{N}\sum_i^N &\left[\eta_{\mathrm{atm,Ab},i}(\nu)T_A^*(\nu)m_{\mathrm{Ab},i} + \eta_{\mathrm{atm,Ba},i}(\nu)T_A^*(\nu)m_{\mathrm{Ba},i}\right] \\ &+ \left[-\eta_{\mathrm{atm,Ab},i}(\nu)m_{\mathrm{Ab},i} - \eta_{\mathrm{atm,Ba},i}(\nu)m_{\mathrm{Ba},i} + \eta_{\mathrm{atm,Aa},i}(\nu)m_{\mathrm{Aa},i} + \eta_{\mathrm{atm,Bb},i}(\nu)m_{\mathrm{Bb},i}\right]T_{p,\mathrm{atm},i} \\ &+ \left[T_{\mathrm{ph},i}(\nu)m_{\mathrm{Ab},i} + T_{\mathrm{ph},i}(\nu)m_{\mathrm{Ba},i} - T_{\mathrm{ph},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{ph},i}(\nu)m_{\mathrm{Bb},i}\right] \\ &+ \left[T_{\mathrm{TLS},i}(\nu)m_{\mathrm{Ab},i} + T_{\mathrm{TLS},i}(\nu)m_{\mathrm{Ba},i} - T_{\mathrm{TLS},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{TLS},i}(\nu)m_{\mathrm{Bb},i}\right].\end{aligned} \tag{2.9}$$

Since the on-off chopping occurs above the knee-frequency of the TLS noise, $T_{TLS,i}(\nu)$ in the on-source position is almost equal to its subtracted off-source counterpart half a chopping period later. Therefore, the TLS terms that are being subtracted from each other become negligible, and can be omitted. Furthermore, if the positions are spaced together closely enough, the atmosphere at the two different chopping positions for one nod position can be considered equal. If the chopping occurs quickly enough (above the knee frequency), we can assume that the atmosphere has stayed constant in the half chopping period that is between the two positions. If half a chopping period is j samples, $\eta_{\mathrm{atm,Ab},i}(\nu) \approx \eta_{\mathrm{atm,Aa},i+j}(\nu)$ and $\eta_{\mathrm{atm,Ba},i}(\nu) \approx \eta_{\mathrm{atm,Bb},i+j}(\nu)$. Eq. (2.9) then becomes more attractive:

$$\hat{T}_A(\nu) = \frac{2}{N}\sum_i^N \eta_{\mathrm{atm},i}(\nu)T_A^*(\nu)(m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i}) + T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i} - m_{\mathrm{Aa},i} - m_{\mathrm{Bb},i}) \tag{2.10}$$

The photon noise still has an effect on the final estimate, but since it is modeled as zero mean[11], $\mathbb{E}\{T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i} - m_{\mathrm{Aa},i} - m_{\mathrm{Bb},i})\} = 0$, so the larger N becomes, the smaller this noise term will be. Unfortunately, we do not yet have an estimator for the galaxy brightness temperature $T_A^*(\nu)$ itself, but rather for $\eta_{\mathrm{atm}}(\nu, t)T_A^*(\nu)$ at the on-source positions. This is what is left of the brightness temperature of the galaxy after passing through the atmosphere.

In order to remove the $\eta_{\mathrm{atm}}(\nu, t)$ term, the blackbody chopper wheel (CW) in the telescope is used. This chopper wheel can alternate between the sky position and a room temperature calibration blackbody to estimate the galaxy brightness before atmospheric transmission. Once every 1 or 2 hours, the blackbody is observed to ensure $T_{\mathrm{Blackbody}}$ is still well-estimated[12]. The PSW chopper wheel that switches between the on- and off-position is then paused. Using the CW chopper wheel, the source brightness temperature can be estimated. Instead of using Eq. (2.10), we use Eq. (2.11):

$$\hat{T}_A^* = \frac{2}{N}\sum_i^N \frac{T_{\mathrm{Ab},i}(\nu)m_{\mathrm{Ab},i} + T_{\mathrm{Ba},i}(\nu)m_{\mathrm{Ba},i} - T_{\mathrm{Aa},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{Bb},i}(\nu)m_{\mathrm{Bb},i}}{T_{\mathrm{Blackbody}} - T_{\mathrm{Aa},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{Bb},i}(\nu)m_{\mathrm{Bb},i}} T_{\mathrm{Blackbody}}. \tag{2.11}$$

where $T_{\mathrm{Blackbody}}$ represents the blackbody temperature. If the temperature of the blackbody is sufficiently close to the physical temperature of the atmosphere, this can be used to divide out the atmospheric transmission. So, we assume that $T_{\mathrm{Blackbody}} \approx T_{p,\mathrm{atm},i}$. Using the same approximations as above $T_{\mathrm{Aa},i}(\nu)m_{\mathrm{Aa},i} + T_{\mathrm{Bb},i}(\nu)m_{\mathrm{Bb},i} \approx (1 - \eta_{\mathrm{atm},i}(\nu))T_{p,\mathrm{atm},i}(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i}) + T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i})$, such that:

$$T_{\mathrm{Blackbody}} - T_{\mathrm{Aa},i}(\nu)m_{\mathrm{Aa},i} - T_{\mathrm{Bb},i}(\nu)m_{\mathrm{Bb},i} \approx \eta_{\mathrm{atm},i}(\nu)T_{p,\mathrm{atm},i}(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i}) - T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i}).$$

This means that

$$\hat{T}_A^*(\nu) \approx \frac{2}{N}\sum_i^N \frac{\eta_{\mathrm{atm},i}(\nu)T_A^*(\nu)(m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i}) + T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Ab},i} + m_{\mathrm{Ba},i} - m_{\mathrm{Aa},i} - m_{\mathrm{Bb},i})}{\eta_{\mathrm{atm},i}(\nu)T_{p,\mathrm{atm},i}(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i}) - T_{\mathrm{ph},i}(\nu)(m_{\mathrm{Aa},i} + m_{\mathrm{Bb},i})} T_{p,\mathrm{atm},i}. \tag{2.12}$$

[11] In reality, the mean value of the photon noise is the true brightness temperature, but this is separated from the photon noise in our model (see Section 2.3.2).

[12] This is measured during the skydip calibration, described in Section 2.2.3

Therefore, since the effects of the photon noise tend to zero for longer integration times,

$$\hat{T}_A^*(\nu) \approx \frac{2}{N}\sum_i^N \frac{\eta_{\mathrm{atm},i}(\nu)T_A^*(\nu)(m_{\mathrm{Ab},i}+m_{\mathrm{Ba},i})}{\eta_{\mathrm{atm},i}(\nu)T_{p,\mathrm{atm},i}(m_{\mathrm{Aa},i}+m_{\mathrm{Bb},i})}T_{p,\mathrm{atm},i} = \frac{2}{N}\sum_i^N \frac{\eta_{\mathrm{atm},i}(\nu)(m_{\mathrm{Ab},i}+m_{\mathrm{Ba},i})}{\eta_{\mathrm{atm},i}(\nu)(m_{\mathrm{Aa},i}+m_{\mathrm{Bb},i})}T_A^*(\nu)$$

Such that $\hat{T}_A^* \approx T_A^*$. A systematic error remains, however, due to the difference between $T_{\mathrm{Blackbody}}$ and $T_{p,\mathrm{atm}}$ [8]. Typically, this error is in the order of $0.1T_A^*$.

Although the direct subtraction method has been shown to be able to detect spectral lines with DESHIMA 1.0 [5], it suffers from two disadvantages. First of all, the room temperature blackbody needs to be held as close as possible to the physical temperature of the atmosphere, as any deviations cause a systematic error. Secondly, the computation of the numerator, where the on- and off-source measurements are subtracted from each other, introduces a factor $\sqrt{2}$ to the noise of the final estimate, as both spectra contain noise.

2.4.2. Noise Removal Using Common Modes for Heterodyne Receivers

Another method that has recently been developed uses a Singular Value Decomposition (SVD) to remove the atmospheric noise and other common noise modes. This method is by Taniguchi et al. and is described in [10]. This method has the advantage that it does not require the direct subtraction of two full noisy spectra, and is therefore used as a starting point for the work presented in this thesis.

There are two fundamental differences between the work by Taniguchi et al., and the work presented in this thesis. First of all, Taniguchi et al. focus on spectrally sparse sources, rather than a full spectrum. This means that only line emission is estimated. As such, the on-source measurements that do not contain a spectral line from the source can also be used to estimate the atmospheric transmission, which improves the estimate. As a downside, however, continuum emission cannot be recovered using this method.

Secondly, the work by Taniguchi assumes heterodyne receivers. These receivers have a different noise characteristic than DESHIMA. They are not background limited and therefore have their own noise temperature and frequency dependent gain, which is already calibrated for in the case of DESHIMA. On the other hand, they do not have the $1/f$ TLS noise that is present in the direct detectors used in DESHIMA. Aside from noise behaviour, the second major difference is that heterodyne receivers retain both phase and amplitude information when downconverting the signal. As such, the spectrum can be digitally computed at a very high spectral resolution for a narrow bandwidth. This means that spectral lines are wide compared to the width of a channel. In the case of DESHIMA, the spectral resolution is determined by the instrument itself and as such, spectral lines often only span one or a few of the much wider spectral channels. This mainly has implications for the shape of a line in the measured spectrum, rather than its sensitivity, as the sensitivities of DESHIMA and a heterodyne receiver are the same (unless the DESHIMA channel is wider than the line itself).

In the method by Taniguchi et al., the on-sky measurements are modeled as:

$$P_{\mathrm{sky}}(f,t) = G(\nu,f)k_{\mathrm{B}}[\eta_{\mathrm{fwd}}(\nu)T_{\mathrm{sky}}(\nu,t) + (1-\eta_{\mathrm{fwd}}(\nu))T_{\mathrm{amb}} + T_{\mathrm{noise}}(\nu)]. \tag{2.13}$$

In this equation, P_{sky} is the power measured in an on-sky measurement. f is the intermediate or downconverted frequency, rather than the observed frequency ν. The two are converted using $\nu = f + m$, where $|m|$ is the local oscillator frequency of the heterodyne receiver. $G(\nu,f)$ is the channel and observing frequency dependent gain of the receiver. k_{B} is the Boltzmann constant. $\eta_{\mathrm{fwd}}(\nu)$ is the forward efficiency of the telescope feed, T_{amb} is the ambient temperature near the telescope and $T_{\mathrm{noise}}(\nu)$ is the noise temperature. Subsequently, T_{sky}, the sky brightness temperature in the absence of photon noise, is expressed as:

$$T_{\mathrm{sky}}(\nu,t) = \begin{cases} T_A^*(\nu)\eta_{\mathrm{atm}}(\nu,t) + T_{p,\mathrm{atm}}(t)(1-\eta_{\mathrm{atm}}(\nu,t)) & \text{, on-source} \\ T_{p,\mathrm{atm}}(t)(1-\eta_{\mathrm{atm}}(\nu,t)) & \text{, off-source.} \end{cases} \tag{2.14}$$

These two equations are then unified using a new variable $T_{\mathrm{ast}}(\nu,t)$, which is equal to $T_A^*(\nu)$ when the telescope is pointed on-source and equal to 0 when it is pointed off source, such that $T_{\mathrm{sky}}(\nu,t) = T_{\mathrm{ast}}(\nu,t)(\nu)\eta_{\mathrm{atm}}(\nu,t) + T_{p,\mathrm{atm}}(t)(1-\eta_{\mathrm{atm}}(\nu,t))$ for all measurements.

Taniguchi et al. also define the power that is measured when the room temperature blackbody of the CW is measured:

$$P_{\mathrm{cal}}(f,t) = G(\nu,f)k_{\mathrm{B}}[T_{\mathrm{blackbody}} + T_{\mathrm{noise}}(\nu)]. \tag{2.15}$$

The difference between the two is then defined as $\mathrm{d}P_{\mathrm{sky}}(f,t) = T_{\mathrm{sky}}(\nu,t) - P_{\mathrm{cal}}(f,t)$. The expression is simplified by assuming that the atmosphere, the ambient air around the telescope and the calibrator blackbody all have the same, constant, temperature. This leads to the following expression:

$$\mathrm{d}P_{\mathrm{sky}}(f,t) = T_{\mathrm{sky}}(\nu,t) - P_{\mathrm{cal}}(f,t) = G(\nu,f)k_{\mathrm{B}}\eta_{\mathrm{fwd}}(\nu)\eta_{\mathrm{atm}}(\nu,t)(T_{\mathrm{ast}}(\nu,t) - T_{p,\mathrm{atm}}). \tag{2.16}$$

Subsequently, a new dimensionless variable is defined:

$$X(f,t) \equiv \ln\left(-\frac{\mathrm{d}P_{\mathrm{sky}}(f,t)}{k_{\mathrm{B}}T_{p,\mathrm{atm}}}\right) = \ln\{G(\nu,f)\eta_{\mathrm{fwd}}(\nu)\eta_{\mathrm{atm}}(\nu,t)\} + \ln\left(1 - \frac{T_{\mathrm{ast}}(\nu,t)}{T_{p,\mathrm{atm}}}\right). \tag{2.17}$$

This expression is then put into matrix form,

$$\mathbf{X} = \left[\ln\{G(\nu,f)\eta_{\mathrm{fwd}}(\nu)\eta_{\mathrm{atm}}(\nu,t)\}\right] + \left[\ln\left(1 - \frac{T_{\mathrm{ast}}(\nu,t)}{T_{p,\mathrm{atm}}}\right)\right], \tag{2.18}$$

where the measured frequency is put on one axis and time on the other. The two components of $\mathbf{X}$, where the bold typesetting represents a matrix, are defined as:

$$\mathbf{L} \equiv \left[\ln\{G(\nu,f)\eta_{\mathrm{fwd}}(\nu)\eta_{\mathrm{atm}}(\nu,t)\}\right] \tag{2.19}$$

and

$$\mathbf{S} \equiv \left[\ln\left(1 - \frac{T_{\mathrm{ast}}(\nu,t)}{T_{p,\mathrm{atm}}}\right)\right]. \tag{2.20}$$

Matrix $\mathbf{L}$ is only dependent on time in the dependence on $\eta_{\mathrm{atm}}(\nu,t)$. As explained in Section 2.3.1, the temporal changes in $\eta_{\mathrm{atm}}(\nu,t)$ are dominated by the changes in PWV. For small PWV changes, this change in PWV can be linearized as in Eq. (2.7). The time dependence can therefore be modelled as $L(\nu,t) = a(\nu) + b(\nu)PWV(t)$ on timescales in the order of an hour[13]. This suggests that $\mathbf{L}$ can be expressed as a low-rank matrix.

As a line spectrum is expected, $T_{\mathrm{ast}}(\nu,t)$ is zero in most channels, since most channels do not contain a spectral line. Taniguchi et al. place the limit for the fraction of lines at around 15% of the spectrum, although an exact limit is not known. Additionally, $T_{\mathrm{ast}}(\nu,t)$ is also equal to zero in all off-source measurements. Note that the entries of $\mathbf{S}$ are equal to zero if $T_{\mathrm{ast}}(\nu,t) = 0$. It is therefore a sparse matrix with blocks of nonzero data.

In order to separate the two terms of $X(f,t)$, an iterative low-rank plus sparse approximation is used. In this approach, a truncated Singular Value Decomposition (SVD) is used to approximate the low-rank matrix with a matrix of r components. The SVD is an operation where an $M \times N$ matrix $\mathbf{A}$ of rank r is decomposed as[14]: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. $\mathbf{U}$ is an $M \times r$ matrix that contains the left singular vectors of $\mathbf{A}$. These vectors are the eigenvectors of $\mathbf{A}\mathbf{A}^T$. Similarly, the columns of $\mathbf{V}$ contain the right singular vectors of $\mathbf{A}$, that are the eigenvectors of $\mathbf{A}^T\mathbf{A}$. $\mathbf{\Sigma}$ is an $r \times r$ diagonal matrix that has the singular values of $\mathbf{A}$ on its main diagonal. These values are the square roots of the eigenvalues of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$. By convention, $\mathbf{\Sigma}$ is usually ordered in descending order of singular values. In order to find the optimal low rank approximation of $\mathbf{X} - \mathbf{S}$, the SVD can be truncated to the appropriate rank. In the case of the algorithm under consideration, $\mathbf{X} - \mathbf{S}$ is full rank due to the role of various noise sources. In order to reduce its rank to r, the first r columns of $\mathbf{U}$ and $\mathbf{V}$ are taken, and only the first r singular values are used.

[13] Note that the signs of $a(\nu)$ and $b(\nu)$ are flipped here, which does not matter as we use no prior information on them and they are arbitrary constants.

[14] as only real values are used in the context of this thesis, we treat the real case here and therefore use the transpose instead of the Hermitian operator.

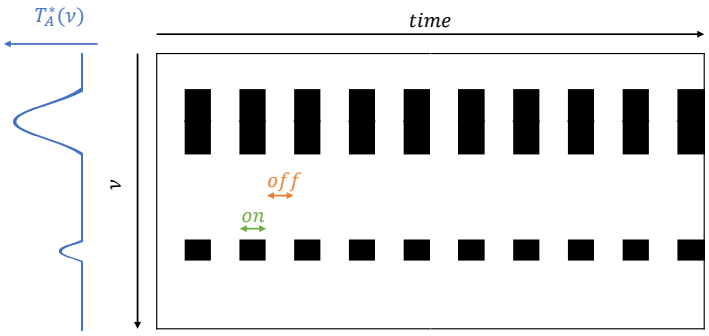


Figure 2.8. Illustration of a typical $\mathbf{\Omega}^{k+1}$. The black blocks are entries that are equal to 1, the white blocks are equal to 0. Where the spectrum (shown in blue) has a spectral line, $\mathbf{\Omega}^{k+1}$ is set to 1 for the on-source positions. This creates blocks where the matrix is equal to one, surrounded by zeros.

Subsequently, this estimate of $\mathbf{L}$, denoted $\mathbf{L}^k$ for the k^{th} iteration, is subtracted from $\mathbf{X}$ and the n largest components in the absolute value of the temporally integrated spectrum are identified. $\mathbf{S}^k$ is then estimated as the value of $\mathbf{X} - \mathbf{L}^k$ at a selected set of frequencies in on-source measurements. All other components are set to zero. This estimate is then subtracted from X and a new iteration begins. The full algorithm as proposed by Taniguchi et al. is summarized by Alg. 1.

Algorithm 1: The algorithm proposed by Taniguchi et al. [10].

Result: $\mathbf{L}^k, \mathbf{S}^k$
Input: $\mathbf{X}, n, \epsilon$
$\mathbf{L}^0 \coloneqq \mathbf{0}$;
$\mathbf{S}^0 \coloneqq \mathbf{0}$;
$k \coloneqq 0$;
while $|\mathbf{X} - \mathbf{L}^k - \mathbf{S}^k|_F^2 / |\mathbf{X}|_F^2 \geq \epsilon$ **do**
  $\mathbf{L}^{k+1} \coloneqq \mathrm{SVD}_r(\mathbf{X} - \mathbf{S}^k)$;
  $\mathbf{\Omega}^{k+1} \coloneqq \mathrm{SparseID@PSW}(|\mathbf{X} - \mathbf{L}^{k+1}|; n)$;
  $\mathbf{S}^{k+1} \coloneqq \mathbf{\Omega}^{k+1} \circ (\mathbf{X} - \mathbf{L}^{k+1})$;
  $k \coloneqq k + 1$
end

In this algorithm SVD_r denotes an SVD truncated to rank r. $\circ$ denotes the Hadamard product (elementwise multiplication). $\mathrm{SparseID@PSW}(\mathbf{A}; n)$ returns a matrix of the same dimensions as $\mathbf{A}$ that is equal to 1 at the on-source positions that correspond to n frequencies where a line is detected. Typically, $\mathbf{\Omega}^k$ is therefore a matrix that is filled with zeros, but has 'blocks' of ones at the frequencies where spectral lines are detected. These blocks periodically appear and disappear with the chopping of the telescope, as illustrated in Fig. 2.8. The matrix with the spectrum is then found as:

$$\hat{\mathbf{T}}_{\text{ast}}(\nu, t) = T_{p,\text{atm}}[1 - \exp(\mathbf{X} - \mathbf{L}^k)], \tag{2.21}$$

And the galaxy temperature is equal to the time average of all on-source samples. This algorithm has shown to improve the noise characteristic with a factor 1.65 compared to conventional direct subtraction methods. Due to this promising result, this method is taken as a starting point for the analysis in this thesis, but adapted to the situation of DESHIMA.

2.5. Research questions

Using the toolbox of terms and techniques described in this chapter, the research question posed in Section 1.2 can be refined and divided into several subquestions. The subquestions that are answered in this thesis are listed here, and we explain in which parts of this thesis they are answered. The main research question is:

> “Is there a way to efficiently remove the wideband atmospheric noise from DESHIMA 2.0 data in order to recover both the narrow line spectra and the broad continuum emission with a higher sensitivity than the current direct subtraction method?”

As mentioned above, the starting point for the research done in the context of this thesis is the method proposed by Taniguchi et al. Since the data model posed in [10] is not directly applicable to data obtained by the DESHIMA 2.0 instrument, the first question that is addressed is:

- “How can we formulate a variation of the model presented by Taniguchi et al. in [10] that is applicable to DESHIMA 2.0?”

This is done in Sections 3.1 and 3.3. Since the noise behaviour of the DESHIMA spectrometer is very different from that of a heterodyne receiver, a natural extension of this subquestion is:

- “What is the behaviour of the various noise sources (i.e. atmosphere, photon and TLS noise) in this model?”

Which is discussed mathematically in Section 3.3. In Chapter 4, the statistical properties of the model and the noise are further explored though simulations.

After the data model has been created and understood, it can be used to remove the wideband atmospheric noise. Therefore, the next subquestion in this thesis is:

- “How can this data model be used to formulate an algorithm that estimates both the continuum emission and the spectral lines of high-redshift galaxies?”

Several methods of formulating the problem with accompanying solution strategies are given in Chapter 5. The results of running these algorithms are compared to the desired results from Chapter 4 in Chapter 6. In Chapter 6, we also argue what set of solutions is best to use in SPLITTER.

In order to determine if the algorithm removes the atmospheric noise *with a higher sensitivity*, as the main research question states, the results produced by the algorithm must be compared to a standard. The standard used in this thesis is the direct subtraction method outlined in Section 2.4.1. The question to answer then becomes:

- “How do the estimates by SPLITTER compare to the estimates created using direct subtraction?”

Some preliminary results for this are shown in the end of Chapter 6, but Chapter 7 focuses on this question specifically.

Another main concern for Chapter 7 is the observing strategy for SPLITTER. Since SPLITTER is a fundamentally different technique from direct subtraction, it is possible that an observing strategy that is (nearly) optimal for direct subtraction, is not optimal at all for SPLITTER. If this is the case, it can potentially occur that the results found using SPLITTER are worse than using direct subtraction, whereas SPLITTER could provide better results for a different observing strategy. Additionally, it is possible that an observing technique with more relaxed requirements can be used for SPLITTER, such as a PSW chopper wheel that rotates more slowly. Therefore, the final subquestion touched upon in this thesis is:

- “Should the observing strategy of DESHIMA 2.0 be changed for this algorithm?”

Together, these five subquestions are used to answer the main research question.

3

Data Model

A suitable data model is required for developing an algorithm for noise removal. This is so important, since the steps that the algorithm needs to take to remove noise are determined by the way the signal and the noise behave in the model. Here, the model is described starting at the physical model from the previous chapter. We briefly review the statistical properties of all terms in this physical model. Subsequently, the assumptions we need for deriving the model are listed and motivated. Using the groundwork laid by the statistical properties and the assumptions, the data model is then derived in a way such that it can easily be used in the algorithm explained in the next chapter.

3.1. Physical Model

Recall that the equation for the brightness temperature measured by DESHIMA 2.0 is

$$T_{\text{measured}}(\nu,t)=\begin{cases}T_A^*(\nu)\eta_{\text{atm}}(\nu,t)+T_{p,\text{atm}}(t)(1-\eta_{\text{atm}}(\nu,t))+T_{\text{ph}}(\nu,t)+T_{\text{TLS}}(\nu,t), & \text{on source}\\ T_{p,\text{atm}}(t)(1-\eta_{\text{atm}}(\nu,t))+T_{\text{ph}}(\nu,t)+T_{\text{TLS}}(\nu,t), & \text{off source.}\end{cases} \tag{3.1}$$

For developing the algorithm, it is helpful to have both positions in a single more general equation. To this end, we now define a deterministic and known variable that discerns whether the telescope is pointed on-source or off-source:

$$m(t)=\begin{cases}1, & \text{on source}\\ 0, & \text{off source.}\end{cases} \tag{3.2}$$

We call this the indicator function. This equation can be used to combine the on-source and off-source positions in Eq. (3.1) into one equation, namely

$$T_{\text{measured}}(\nu,t)=\eta_{\text{atm}}(\nu,t)m(t)T_A^*(\nu)+(1-\eta_{\text{atm}}(\nu,t))T_{p,\text{atm}}(t)+T_{\text{ph}}(\nu,t)+T_{\text{TLS}}(\nu,t). \tag{3.3}$$

Before deriving the model, it is important to have a clear overview of which terms in this equation are deterministic, and which ones are random variables. They are listed here from left to right:

- $T_{\text{measured}}(\nu,t)$**:** This is the measurement from the telescope. A random variable of which the value is known.
- $\eta_{\text{atm}}(\nu,t)$**:** The atmospheric transmittance is a random variable that cannot be directly measured. It is correlated in time and between channels.
- $m(t)$**:** The indicator function is a known deterministic variable.
- $T_A^*(\nu)$**:** This is the unknown but deterministic temperature of the galaxy that is under observation.

- $T_{p,\mathrm{atm}}(t)$: This is the physical temperature of the atmosphere. It is a very slowly changing deterministic variable [30], that is assumed to be known due to sensors at the telescope or the room temperature blackbody.
- $T_{\mathrm{ph}}(\nu, t)$: This is a zero mean unknown Gaussian distributed random variable.
- $T_{\mathrm{TLS}}(\nu, t)$: This is an unknown random variable that is strongly correlated in time, but uncorrelated between channels.

3.2. Assumptions

Several assumptions are needed for the data model. First of all, we assume that the physical temperature of the atmosphere $T_{p,\mathrm{atm}}$ is known and can be approximated as constant within the observation. This assumption is made, since the temperature of the atmosphere varies slowly compared to all other variables and its variations are small compared to its value. This means that the time dependence of the physical temperature is dropped from here on in the notation. Note that this assumption is comparable to the assumptions in the existing data reduction method and the assumptions by Taniguchi et al. in [10], where $T_{p,\mathrm{atm}}$ is assumed to be equal to the room temperature blackbody between skydip calibrations.

Secondly, we assume that $T_{p,\mathrm{atm}}$ is much larger than T_A^*, $T_{\mathrm{ph}}(\nu, t)$ and $T_{\mathrm{TLS}}(\nu, t)$. To place this in perspective, the physical temperature of the atmosphere is expected to be near the freezing point of water at $T_{p,\mathrm{atm}} \sim 270$ K [30]. A typical galaxy brightness temperature T_A^* is a few millikelvins. The photon noise, $T_{\mathrm{ph}}(\nu, t)$ is zero mean and channel dependent. In the channels where the photon noise is strongest, the standard deviation is ~ 0.1 K. Finally, the TLS noise is also maximally on the subkelvin level. These three conditions also ensure that $T_{p,\mathrm{atm}} > T_{\mathrm{measured}}(\nu, t)$ as long as $\eta_{\mathrm{atm}} > 0$. We assume this is the case in all channels.

Due to the temporally correlated nature of the TLS noise, it stays nearly at a constant level for short timescales in the order of a few seconds. Therefore, the TLS noise and zero mean photon noise can be modeled together by a single Gaussian with a nonzero mean. This is illustrated in Fig. 3.1 for 10 seconds of data. For long timescales (illustrated in Fig. 3.2 for 30 minutes), the TLS noise can vary, leading to a random variable that is difficult to model. This means that we assume that the TLS noise and photon noise form this Gaussian variable together. The mean value is constant for these short timescales and denoted as $T_\mu(\nu)$. We also define a new zero mean Gaussian random variable $T_\sigma(\nu, t)$, such that $T_{\mathrm{TLS}}(\nu, t) + T_{\mathrm{ph}}(\nu, t) = T_\mu(\nu) + T_\sigma(\nu, t)$.

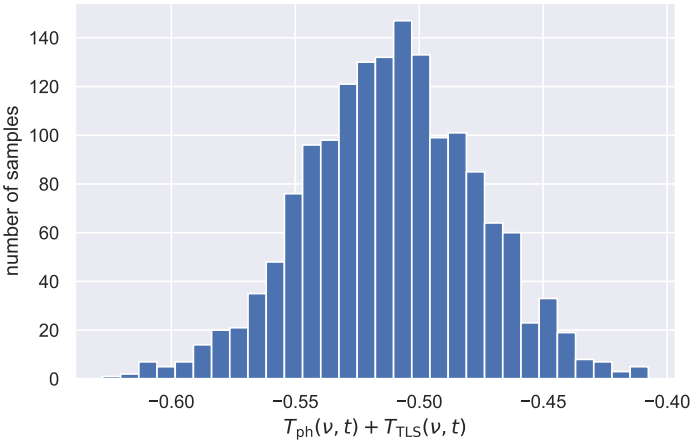


Figure 3.1. Histogram of the photon noise and TLS noise for 10 seconds of data in channel 28.

Finally, there is an assumption on the variation of the atmospheric transmittance. Recall from Section 2.3.1 that the optical depth is $\tau = -\ln \eta_{\mathrm{atm}}(\nu, t)$ and that small changes to this can be linearized by the PWV. When this optical depth is placed in a matrix with the channels on one axis and the time on the other, it can be described with a low rank matrix [10]. For short timescales of a few second, variations in the transmittance can be described with a single component [31]. Therefore, in the regime where the TLS noise can be modeled as a constant, the optical depth can be modeled using a rank 1

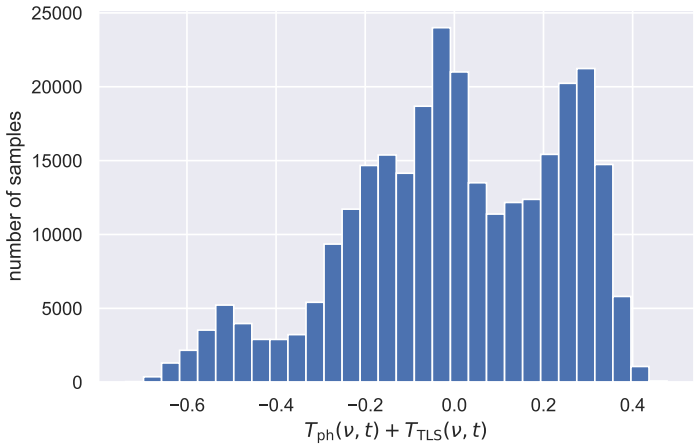


Figure 3.2. Histogram of the photon noise and TLS noise for 30 minutes of data in channel 28.

matrix. We assume this to still be true with a small temporally constant overhead per channel. That is, we assume that the matrix with entries $-\ln\{\eta_{\mathrm{atm}}(\nu,t)-T_{\mu}(\nu)/T_{p,\mathrm{atm}}\}$ is rank 1 if its width is such that it only describes a few seconds of data. This is because $T_{\mu}(\nu)/T_{p,\mathrm{atm}}$ is such a small number that small changes in the atmosphere can be linearized without significant perturbations.

3.3. Matrix Formulation

In order to estimate the source, $T_A^*(\nu)$ needs to be isolated from the other variables. In order to do this, the nonlinearly varying multiplication by $\eta_{\mathrm{atm}}(\nu,t)$ must be removed. Since the optical depth $\tau(\nu,t)=-\ln\eta_{\mathrm{atm}}(\nu,t)$ can be written as a linear function of the PWV, transforming the data to the logarithmic domain enables us to use methods that exploit low-rank properties of the optical depth to decompose the data. As such, we follow the example of Taniguchi et al. in [10] and define a new dimensionless random variable:

$$X(\nu,t)=-\ln\left(-\frac{T_{\mathrm{measured}}(\nu,t)-T_{p,\mathrm{atm}}}{T_{p,\mathrm{atm}}}\right). \tag{3.4}$$

This variable can be calculated for each channel at each sample. This results in an $M\times N$ matrix, where $M=347$ for the number of channels in DESHIMA and N is the number of measurements. N Can be a very large number, as the sampling frequency is 160 Hz and an observation is typically in the order of hours for high redshift galaxies. The matrix is therefore very wide. Using bold capitals to denote matrices, the matrix is then written as:

$$\mathbf{X}=\left[-\ln\left(-\frac{T_{\mathrm{measured}}(\nu,t)-T_{p,\mathrm{atm}}}{T_{p,\mathrm{atm}}}\right)\right]. \tag{3.5}$$

We call this matrix the ‘data matrix’. The goal of this chapter is to separate the different parts of the data matrix, such that the galaxy can ultimately be isolated. This is done by rewriting $\mathbf{X}$ into the form

$$\mathbf{X}=\mathbf{S}+\mathbf{L}+\mathbf{N}. \tag{3.6}$$

Where $\mathbf{S}$ denotes a matrix that represents the source, $\mathbf{L}$ represents a low-rank matrix that contains the effects of the atmospheric transmittance and $\mathbf{N}$ represents a matrix that contains all residual noise. To get to this form, the equation for $T_{\mathrm{measured}}(\nu,t)$ (Eq. (3.3)) is plugged in and the $T_{p,\mathrm{atm}}$ in the numerator of the fraction is rewritten:

$$\mathbf{X}=\left[-\ln\left(-\frac{\eta_{\mathrm{atm}}(\nu,t)(m(t)T_A^*(\nu)-T_{p,\mathrm{atm}})+T_{\mathrm{ph}}(\nu,t)+T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}}}\right)\right]. \tag{3.7}$$

Subsequently, the fraction is split, resulting in:

$$\mathbf{X} = \left[-\ln\left\{ \eta_{\mathrm{atm}}(\nu,t)\left(1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right) - \frac{T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}}} \right\} \right]. \tag{3.8}$$

Note that the part of the equation between round brackets only depends on the desired variable $T_A^*(\nu)$ and the known quantities $m(t)$ and $T_{p,\mathrm{atm}}$. We therefore isolate this part into a factor within the logarithm:

$$\mathbf{X} = \left[-\ln\left\{ \left(1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right)\left(\eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{ph}}(\nu,t) + T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)}\right) \right\} \right]. \tag{3.9}$$

If the time support of $\mathbf{X}$ is sufficiently short, $T_{\mathrm{TLS}}(\nu,t) + T_{\mathrm{ph}}(\nu,t) = T_\mu(\nu) + T_\sigma(\nu,t)$. Note that the TLS noise $T_{\mathrm{TLS}}(\nu,t)$ is correlated for the full observation, whereas the photon noise is an uncorrelated random variable. Therefore, the fraction that contains them is split:

$$\mathbf{X} = \left[-\ln\left\{ \left(1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right)\left(\eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)} - \frac{T_{\mathrm{ph}}(\nu,t)}{T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)}\right) \right\} \right], \tag{3.10}$$

and subsequently, the correlated variables $\eta_{\mathrm{atm}}(\nu,t)$ and $T_{\mathrm{TLS}}(\nu,t)$ are put into a separate factor as well:

$$\mathbf{X} = \left[-\ln\left\{ \left(1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right)\left(\eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)}\right) \left(1 - \frac{T_{\mathrm{ph}}(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)[T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)] - T_{\mathrm{TLS}}(\nu,t)}\right) \right\} \right]. \tag{3.11}$$

Now, since $\ln(ab) = \ln a + \ln b$, the logarithm can be split into a sum:

$$\mathbf{X} = \left[-\ln\left\{ 1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}} \right\} \right] + \left[-\ln\left\{ \eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)} \right\} \right] + \left[-\ln\left\{ 1 - \frac{T_{\mathrm{ph}}(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)[T_{p,\mathrm{atm}} - m(t)T_A^*(\nu)] - T_{\mathrm{TLS}}(\nu,t)} \right\} \right]. \tag{3.12}$$

Since we assume that $T_A^*(\nu) \ll T_{p,\mathrm{atm}}$, we can simplify the denominators in the second and third matrix:

$$\mathbf{X} \approx \left[-\ln\left\{ 1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}} \right\} \right] + \left[-\ln\left\{ \eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}}} \right\} \right] + \left[-\ln\left\{ 1 - \frac{T_{\mathrm{ph}}(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_{\mathrm{TLS}}(\nu,t)} \right\} \right] \tag{3.13}$$

This final equation has a form where the brightness temperature is the only unknown in the first matrix. The second matrix contains the temporally correlated random variables $\eta_{\mathrm{atm}}(\nu,t)$ and $T_{\mathrm{TLS}}(\nu,t)$ and the last matrix contains the photon noise. We can now define:

$$\mathbf{S} \equiv \left[-\ln\left\{ 1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}} \right\} \right] \tag{3.14}$$

$$\mathbf{L} \equiv \left[-\ln\left\{ \eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}}} \right\} \right] \tag{3.15}$$

$$\mathbf{N} \equiv \left[-\ln\left\{ 1 - \frac{T_{\mathrm{ph}}(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_{\mathrm{TLS}}(\nu,t)} \right\} \right]. \tag{3.16}$$

Such that:

$$\mathbf{X} = \mathbf{S} + \mathbf{L} + \mathbf{N}. \tag{3.17}$$

In the remainder of this chapter, the properties of these matrices are discussed.

3.3.1. Source Matrix

The source matrix is given by:

$$\mathbf{S} \equiv \left[-\ln\left\{1 - \frac{m(t)T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right\}\right]. \tag{3.18}$$

As described above, the purpose of the source matrix is to have a matrix from which $T_A^*(\nu)$ can be estimated. Since the galaxy does not change, the entire time dependence is contained in $m(t)$. In the columns where $m(t) = 0$, the fraction disappears. This means that when $m(t) = 0$, the entire column of $\mathbf{S}$ is equal to zero. When $m(t) = 1$, the columns contain positive numbers, as $1 - T_A^*(\nu)/T_{p,\mathrm{atm}} < 1$. Additionally, all of these columns are equal to each other, since the time dependence is only present in $m(t)$. We now define an $M \times 1$ vector that describes the nonzero columns of $\mathbf{S}$:

$$\mathbf{s} \equiv \left[-\ln\left\{1 - \frac{T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right\}\right]. \tag{3.19}$$

The lowercase bold typesetting denotes a vector. Therefore, we call $\mathbf{s}$ the source vector. We then also define an $N \times 1$ vector $\mathbf{m}$. The components of this vector are 0 and 1, denoting whether the telescope points on-source or off-source. This means that $\mathbf{S}$ van be written as:

$$\mathbf{S} = \mathbf{s}\mathbf{m}^T. \tag{3.20}$$

This is a completely deterministic matrix with columns either equal to $\mathbf{0}$ or $\mathbf{s}$.

3.3.2. Atmosphere Matrix

For convenience, the low rank matrix with the atmospheric effects is repeated here:

$$\mathbf{L} \equiv \left[-\ln\left\{\eta_{\mathrm{atm}}(\nu,t) - \frac{T_{\mathrm{TLS}}(\nu,t)}{T_{p,\mathrm{atm}}}\right\}\right]. \tag{3.21}$$

Since both $\eta_{\mathrm{atm}}(\nu,t)$ and $T_{\mathrm{TLS}}(\nu,t)$ are correlated in time, the time support of this matrix is very important to its statistical behaviour. If the time support of the matrix is sufficiently small, the TLS noise becomes a constant per channel and the effects of the atmospheric transmittance can be described by a rank 1 matrix. In order to achieve this, the matrix is split into l narrower matrices:

$$\mathbf{L} = \left[\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \ldots, \mathbf{L}_l\right]. \tag{3.22}$$

That all have a time support t_l equal to the total length of the observation divided by l. We can now operate on the smaller matrices $\mathbf{L}_i$. This split also has the benefit that discontinuities in the atmosphere can be incorporated naturally when the observation needs to be paused and started again (for example for calibrations). This is possible, since each block of data can be processed independently. If the matrices $\mathbf{L_i}$ have a sufficiently small time support, the time dependence of the TLS noise disappears and we can replace it with $T_\mu(\nu)$:

$$\mathbf{L}_i = \left[-\ln\left\{\eta_{\mathrm{atm}}(\nu,t) - \frac{T_\mu(\nu)}{T_{p,\mathrm{atm}}}\right\}\right], \quad (i-1)t_l < t \leq it_l. \tag{3.23}$$

Each matrix $\mathbf{L}_i$ is a rank one matrix. In this context, only $\eta_{\mathrm{atm}}(\nu,t)$ is a random variable. Furthermore, since $T_\mu(\nu)/T_{p,\mathrm{atm}}$ is a very small number and $\eta_{\mathrm{atm}}(\nu,t)$ is always between 0 and 1, the matrix $\mathbf{L}_i$ has only nonnegative values.

3.3.3. Residual Noise Matrix

The final matrix in the model is the residual noise matrix $\mathbf{N}$. This matrix can be split up in the same way as the low rank atmosphere matrix, where we replace $T_{\mathrm{TLS}}(\nu,t)$ and $T_{\mathrm{ph}}(\nu,t)$ by $T_\mu(\nu)$ and $T_\sigma(\nu,t)$:

$$\mathbf{N}_i = \left[-\ln\left\{1 - \frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right\}\right], \quad (i-1)t_l < t \leq it_l. \tag{3.24}$$

Although there is a time dependence in $\eta_{\mathrm{atm}}(\nu,t)$, the time dependence of $T_\sigma(\nu,t)$ is much more variable. Therefore, this matrix is dominated by the zero mean Gaussian random variable $T_\sigma(\nu,t)$. The

realizations of this variable are scaled due to the temporal changes in the atmospheric transmittance, but these changes are relatively small for channels that are either well within or well outside of an atmospheric window. Since $T_{p,\mathrm{atm}} \gg T_\sigma(\nu,t)$, the fraction inside the expression for $\mathbf{N}_i$ is small as long as $\eta_{\mathrm{atm}}(\nu,t)$ is not too small. The matrix can thus be linearized around $T_\sigma(\nu,t)/[\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)] = 0$ using the Taylor expansion.

$$\begin{aligned}\mathbf{N}_i \approx &\left[-\ln 1+ \quad \frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right] + \left[\mathcal{O}\left\{\left(\frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right)^2\right\}\right] \\ = &\left[\frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right] + \left[\mathcal{O}\left\{\left(\frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right)^2\right\}\right], \quad (i-1)t_l < t \le it_l.\end{aligned} \tag{3.25}$$

Where $\mathcal{O}\{\cdot\}$ denotes higher order terms of the expression inside the brackets. When $T_{\mathrm{ph}}(\nu,t)/[\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_{\mathrm{TLS}}(\nu)]$ is very small, the dependency on the higher order terms is also very small and they can be neglected. In this case, $\mathbf{N}_i$ behaves as a zero mean Gaussian in the first order. If $T_{\mathrm{ph}}(\nu,t)/[\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_{\mathrm{TLS}}(\nu)]$ is larger, the quadratic terms will play a role, and that row of the matrix will have a positive mean value. To circumvent this issue, the channels with an atmospheric transmittance below 10% are removed from the matrices, reducing the number of rows. Under typical weather conditions at the ASTE site, this is about 5% of the channels, where the signal to noise ratio is lowest. This means that the loss of information by removing these channels is small, whereas separating the signal and noise matrices becomes easier. Due to the photon noise in the numerator, $\mathbf{N}_i$ behaves as a Gaussian in the temporal direction. Therefore, it is expected to be a full-rank matrix with an (almost) diagonal covariance matrix. The denominator is correlated across channels due to the $\eta_{\mathrm{atm}}(\nu,t)$ term, and can introduce off-diagonal terms in the covariance matrix.

3.3.4. Full Data Model

Using the derivations above, the full data model for each data block is:

$$\mathbf{X}_i = \mathbf{S}_i + \mathbf{L}_i + \mathbf{N}_i, \quad (i-1)t_l < t \le it_l. \tag{3.26}$$

Here, the new matrix $\mathbf{S}_i$ is introduced. This matrix is formed by splitting up $\mathbf{m}$, the vector that describes the telescope pointing, into subvectors, $\mathbf{m} = [\mathbf{m}_1^T, \mathbf{m}_2^T, \ldots \mathbf{m}_l^T]^T$. The deterministic matrix is simply

$$\mathbf{S}_i = \mathbf{s}\mathbf{m}_i^T, \quad (i-1)t_l < t \le it_l. \tag{3.27}$$

$$\text{with} \quad \mathbf{s} = \left[-\ln\left\{1 - \frac{T_A^*(\nu)}{T_{p,\mathrm{atm}}}\right\}\right] \tag{3.28}$$

Whereas the two stochastic matrices are given by:

$$\mathbf{L}_i = \left[-\ln\left\{\eta_{\mathrm{atm}}(\nu,t) - \frac{T_\mu(\nu)}{T_{p,\mathrm{atm}}}\right\}\right], \quad (i-1)t_l < t \le it_l \tag{3.29}$$

and

$$\mathbf{N}_i = \left[\frac{T_\sigma(\nu,t)}{\eta_{\mathrm{atm}}(\nu,t)T_{p,\mathrm{atm}} - T_\mu(\nu)}\right], \quad (i-1)t_l < t \le it_l. \tag{3.30}$$

Bear in mind that the time split of the data is done to make sure the TLS noise can be modeled as a constant and that the atmospheric transmittance can be described with a single component. There is no such boundary on the deterministic source matrix. Therefore, the vector $\mathbf{s}$ can be obtained from any row or combination of rows of the data, as long as there is an estimate for the rows of $\mathbf{X}$, $\mathbf{L}$ and $\mathbf{N}$ there. As such, $\mathbf{S}$ can be estimated using the complete dataset still, where $\mathbf{L}$ and $\mathbf{N}$ are simply a concatenation of the data blocks like in Eq. (3.22).

4

Analysis

In this chapter, the properties of the matrices in the data model are simulated. This is done by looking at a typical examples of the matrices, as well as the sampling mean of each channel and the sampling covariances of the matrices. The assumptions of Chapter 3 can be checked with these simulations.

4.1. Simulations

The simulations in this chapter are created using `TiEMPO`[1].Due to its modular setup, `TiEMPO` allows for calculating the different variables inside the matrices in the full data model.

First of all, $T_{p,\mathrm{atm}}$ is fixed at 273 K in `TiEMPO`. Since the photon noise can be turned off in the simulations, $\eta_{\mathrm{atm}}(\nu, t)$ can be calculated from the empty sky simulation. The difference between an empty sky simulation with and without photon noise is used to calculated $T_{\mathrm{ph}}(\nu, t)$. Additionally, since TLS noise is added to the output of `TiEMPO` after execution, both measurements with and without TLS noise are available, which allows for calculating $T_{\mathrm{TLS}}(\nu, t)$. In order to satisfy the model, the TLS noise and photon noise are added to create a Gaussian, the mean of which is constant per timescale t_l. The deviations around this constant level are put into the photon noise, as not to remove noise from the simulations. Finally, the galaxy brightness temperature $T_A^*(\nu, t)$ can be simulated using the `GalSpec`[2] package and the same responsivity model as in `TiEMPO`. The other variables used in the simulation are listed in Table I.

pwv_0	telescope elevation	redshift	observation time	luminosity	velocity width of line
1.0 mm	60°	3	30 mins	$10^{13.8} L_\odot$	600 km/s

TABLE I. Input parameters of `TiEMPO`

4.2. Typical Matrices

Fig. 4.1 shows a typical realization of the matrices in the data model. In line with expectation, low-rank atmosphere matrix $\mathbf{L}_i$ is very similar to data matrix $\mathbf{X}_i$. This confirms that the measured data is dominated by the atmosphere prior to noise removal (see Fig. 2.4 for the brightness of the atmosphere compared to the galaxy). Furthermore, $\mathbf{L}_i$ is nearly constant in the horizontal temporal dimension, confirming that the atmosphere does not change much over the 160 samples (one second) that this matrix is wide, especially compared to the much larger channel to channel variations. The fact that the temporal changes are small re-enforces the idea that they can be linearized. The bright frequencies in $\mathbf{L}_i$ coincide with the frequencies where the atmospheric transmittance is low (examples of the atmospheric transmittance can be found in Fig. 2.5). From this, we conclude that the value of $\mathbf{L}_i$ is dominated by the atmospheric transmittance rather than the TLS noise.

If all matrices are plotted on the same color scale as data matrix $\mathbf{X}_i$, the structure in $\mathbf{S}_i$ and $\mathbf{N}_i$ is invisible. This is because $\mathbf{X}_i$ and $\mathbf{L}_i$ have much larger numbers, even inside the atmospheric windows. Therefore, the matrices are plotted again in Fig. 4.2, where each matrix has its own color scale. Note

[1] 10.5281/zenodo.4279085
[2] 10.5281/zenodo.4279061

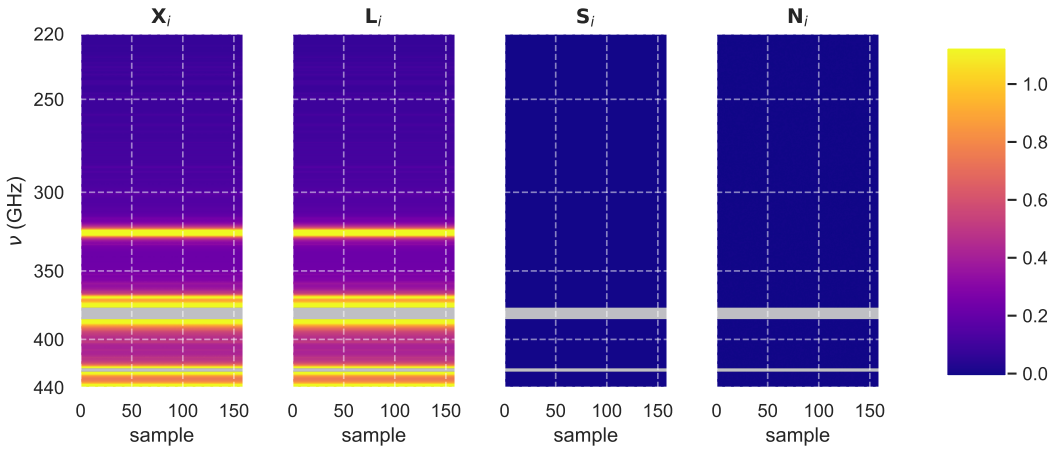


Figure 4.1. Typical example of the four matrices in the data model at the same color scale. The color denotes the dimensionless quantity in the matrices. The greyed out values represent ommitted samples where $\eta_{\text{atm}} < 10\%$. Matrix $\mathbf{L}_i$ dominates over $\mathbf{S}_i$ and $\mathbf{N}_i$.

that the colorscale of $\mathbf{N}_i$ is 10 times as large as that of $\mathbf{S}_i$. It becomes clear that the source matrix $\mathbf{S}_i$ is the weakest component inside a single block of data $\mathbf{X}_i$. Only when multiple blocks are combined to remove $\mathbf{L}_i$ and $\mathbf{N}_i$ from $\mathbf{X}_i$, the signal can be recovered. The matrix $\mathbf{S}_i$ has a a pattern of vertical stripes, caused by the on-off chopping of the telescope. In the off position, the matrix is equal to zero. In the other positions, the structure of the galaxy brightness temperature can be seen. There is a color gradient caused by the continuum, as well as a few sharper horizontal lines signifying spectral lines. Residual noise matrix $\mathbf{N}_i$ is randomly varying in the horizontal direction that represents time, as is expected from the Gaussian distributed photon noise. Additionally, the noise amplitude is larger in the channels where the atmospheric transmission is low, as is evident by the yellow and blue dots at the same frequencies where atmosphere matrix $\mathbf{L}_i$ is large.

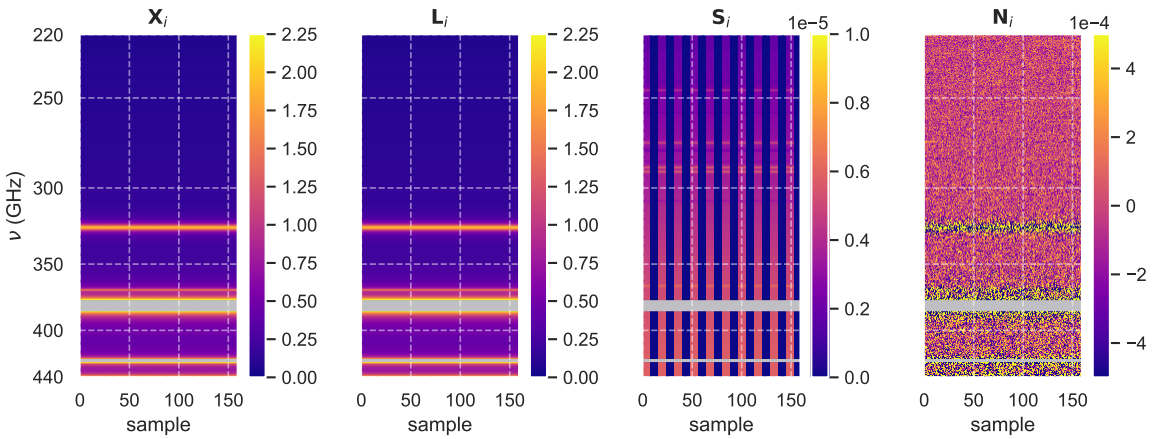


Figure 4.2. Typical example of the four matrices in the data model at different color scales to improve contrast. The color denotes the dimensionless quantity in the matrices. The greyed out values represent ommitted samples where $\eta_{\text{atm}} < 10\%$. The structure of data matrix $\mathbf{X}_i$ is mostly the same as that of $\mathbf{L}_i$. The source matrix $\mathbf{S}_i$ has vertical stripes, that are cause by the on-off chopping. The amplitude of $\mathbf{N}_i$ is larger at the frequencies where $\mathbf{L}_i$ is high.

An important assumption in the data model is that $\mathbf{L}_i$ is rank 1 and $\mathbf{N}_i$ is full rank. In order to investigate this, the singular values of the matrices have been plotted in Fig. 4.3. The first singular value of $\mathbf{L}_i$ is over a factor 100 larger than all other singular values of $\mathbf{L}_i$, $\mathbf{S}_i$ and $\mathbf{N}_i$. Therefore, it can safely be assumed that $\mathbf{L}_i$ can be approximated as a rank 1 matrix. The size of the first singular value of $\mathbf{X}_i$ is comparable to the first singular value of $\mathbf{L}_i$. All other singular values of $\mathbf{X}_i$ are comparable to those of $\mathbf{N}_i$. The fact that there are no large differences between the singular values shows that $\mathbf{N}_i$ is a full rank matrix. The

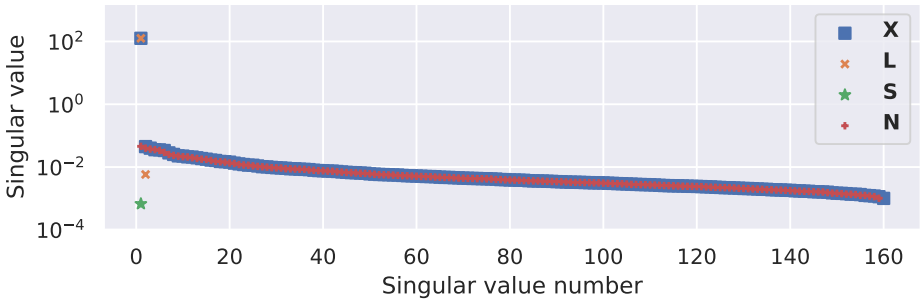


Figure 4.3. Singular values of the different matrices on a logarithmic scale. There is only one very large singular value that dominaties over all other singular values. This is the first singular value of the low-rank atmosphere matrix $\mathbf{L}_i$. On the other hand, $\mathbf{N}_i$ has a much more gradual decrease in singular values. Note that most of the singular values of $\mathbf{L}_i$ and $\mathbf{S}_i$ are so small that they are not shown in this plot.

singular values are not all equal, but this is statistically expected in a full rank noise matrix [39]. The singular values of both $\mathbf{L}_i$ and $\mathbf{N}_i$ behave as expected from the reasoning Chapter 3.

4.3. Statistical Properties

In order to fully understand the matrices, their statistical properties must also be understood. To this end, the sampling means of the full dataset and the sampling covariances of the data blocks have also been plotted.

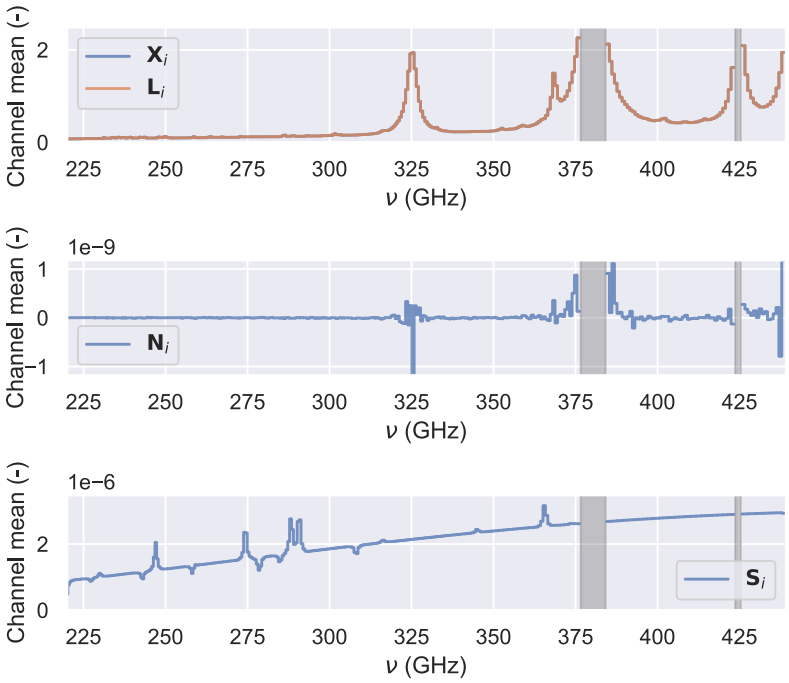


Figure 4.4. Sampling means of all matrices in the data model. Since the contrast between the matrices that compose $\mathbf{X}$ is so large, they are plotted on different scales.

The means were found by averaging over 30 minutes of data. They are shown in Fig. 4.4. The shape of the optical depth of the atmosphere is clearly present in atmosphere matrix $\mathbf{L}$ and data matrix $\mathbf{X}$. In source matrix $\mathbf{S}$, the shape of a spectrum with a broad continuum and more narrow spectral lines is shown. It is important to note that half of the $\mathbf{S}$ matrix is equal to zero due to the off-source measurements, so the mean of the timesamples that do contain the galaxy is actually twice as large as the mean of the entire matrix shown here.

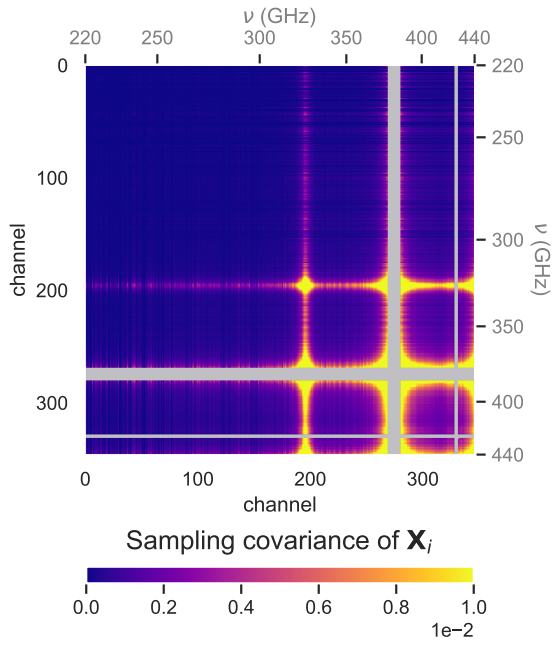


Figure 4.5. Sampling covariance matrix of data $\mathbf{X}_i$, the color shows the covariance. The grey areas represent the channels that have been omitted from the analysis, as the atmospheric transmittance is too low at these frequencies. Note that this matrix is very similar to that of the atmosphere matrix $\mathbf{L}_i$, which is shown in Fig. 4.6.

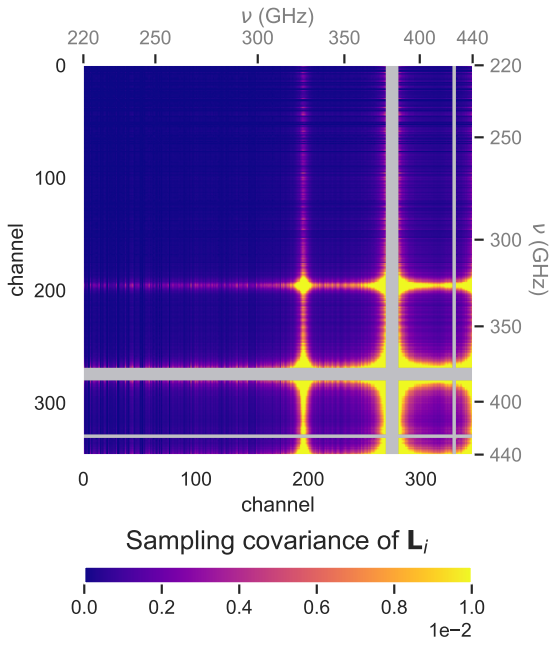


Figure 4.6. Sampling covariance matrix of $\mathbf{L}_i$, the color shows the covariance and the channels where the atmospheric transmissions is lowest omitted once more. This matrix has strong vertical and horizontal lines, which are also faintly visible in the darker areas of the matrix. This striped profile occurs because $\mathbf{L}_i$ is a rank 1 matrix.

The mean of residual noise matrix $\mathbf{N}$ is always nearly zero, but its mean deviates from 0 more in channels with low atmospheric transmittance. This is due to the the division by $\eta_{\text{atm}}(\nu,t)T_{p,\text{atm}} - T_\mu(\nu)$, which can be a very small number if the atmospheric transmittance is also small. This division by a small number van make the Gaussian noise is made relatively large. Since the values are still very small, this does not seem to be a problem. However, we must bear in mind that the current form of the residual noise matrix (Eq. (3.30) instead of Eq. (3.24)) is only valid if $T_\sigma(\nu,t) \ll \eta_{\text{atm}}(\nu,t)T_{p,\text{atm}} - T_\mu(\nu)$ in that channel. Since $T_\sigma(\nu,t)$ is largest where $\eta_{\text{atm}}(\nu,t)$ is smallest, channels with a low atmospheric transmittance are naturally not well modeled with a zero mean Gaussian timestream in $\mathbf{N}_i$. Therefore, higher order terms of the Taylor polynomal for $T_\sigma(\nu,t)/(\eta_{\text{atm}}(\nu,t)T_{p,\text{atm}} - T_\mu(\nu))$ come into play, and the expected value of $\mathbf{N}_i$ becomes larger than zero. Since this effect is present in both on and off samples and is dependent on the atmospheric transmittance, this offset could be added to the estimate of atmosphere $\mathbf{L}_i$ by the algorithm during running. This would not be a problem, however, as $\mathbf{L}_i$ and $\mathbf{N}_i$ do not strictly need to be separated from each other, only from $\mathbf{S}_i$.

The sampling covariance matrices are created using $\text{cov}(\mathbf{X}) = \frac{1}{l}\sum_{i=1}^{l}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T$, where $\mathbf{X}_i$ is the matrix of which the covariance is calculated. The notation $\overline{\mathbf{X}}$ denotes a matrix that has the same width as $\mathbf{X}_i$ and contains the sampling mean of the matrix in its columns.

The covariance matrices of $\mathbf{X}_i$ and $\mathbf{L}_i$ are once again very similar (see Fig. 4.5 and Fig. 4.6), further confirming that $\mathbf{L}_i$ is the dominant matrix in $\mathbf{X}_i$. The fact that $\mathbf{X}_i$ and $\mathbf{L}_i$ are so similar is what allows for estimating $\mathbf{L}_i$ by a low rank approximation of $\mathbf{X}_i$.

The sampling covariance matrix of $\mathbf{N}_i$ is shown in Fig. 4.7. This is a diagonal matrix. The main diagonal of the matrix varies between channels. This is caused by the photon noise, as it has a higher standard deviation in the channels where the atmospheric brightness is highest. Therefore, the main diagonal of the sampling covariance matrix of $\mathbf{N}_i$ also contains higher values at the channels where the atmosphere is brightest. The diagonal structure of this matrix means that the noise in $\mathbf{N}_i$ is not (strongly) correlated across different channels. In combination with the mean in Fig. 4.4, this shows that $\mathbf{N}_i$ is an uncorrelated noise matrix with (approximately) zero mean.

Finally, the sampling covariance of $\mathbf{S}_i$ is shown in Fig. 4.8. Essentially, this matrix is the mean of $\mathbf{S}_i$ correlated with itself. This occurs because the mean is equal to half of $\mathbf{s}$, so after subtraction of the

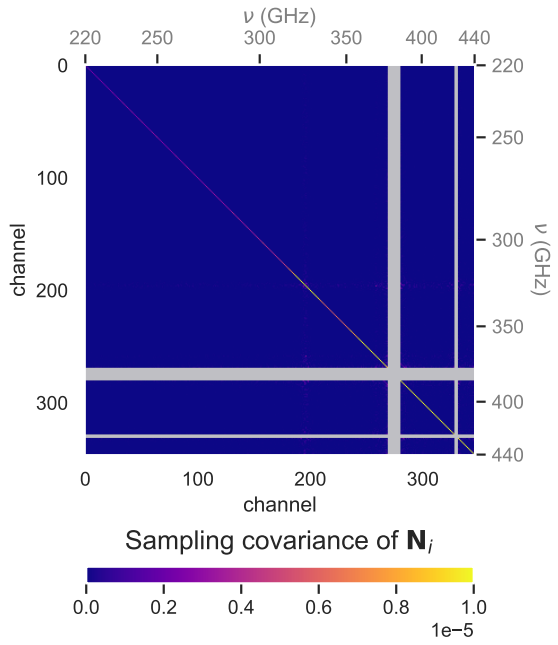


Figure 4.7. Sampling covariance matrix of residual noise matrix $\mathbf{N}_i$ where the color shows the covariance. The main diagonal dominates in this matrix. The channel dependent nature of this diagonal is visible from the changes in brightness.

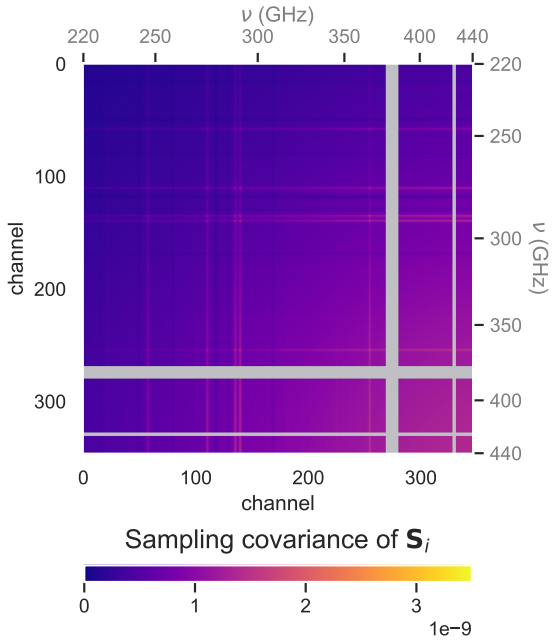


Figure 4.8. The covariance matrix of $\mathbf{S}_i$, the color shows the covariance. The profile of the galaxy (smooth continuum with spectral lines) is clearly visible in this matrix.

mean, there is power in all columns of the matrix. The on-source columns become equal to $\mathbf{s}/2$ and the off-source columns become equal to $-\mathbf{s}/2$. By subsequently multiplying this matrix with its transpose, the matrix shown in Fig. 4.8 is created.

5 Algorithm

In order to separate the noise matrices $\mathbf{L}$, $\mathbf{S}$ and $\mathbf{N}$, the Stationary source Plus Low-Rank Transmit-Tance EstimatoR (SPLITTER) is developed. In this chapter, a number of possible algorithm designs are discussed. These are presented as modular options for estimating the various matrices. Not every possibility presented in this chapter is fully pursued to a working version within the scope of this thesis. The different possible options are listed here, however, accompanied by an explanation of why they were not chosen.

First, the optimization problem needs to be formulated. Since $\mathbf{X}_i = \mathbf{S}_i + \mathbf{L}_i + \mathbf{N}_i$ for each chunk of data i and since the residual noise matrix $\mathbf{N}_i$ is Gaussian in time, $\mathbf{N}_i$ can be estimated by minimizing the Frobenius norm of $\mathbf{X}_i - \mathbf{L}_i - \mathbf{S}_i$. Adding the constraints formulated in Chapter 3 leads to the generic form of the optimization problem:

$$\begin{aligned} \underset{\mathbf{L}_i, \mathbf{s}}{\text{minimize}} \quad & \frac{1}{2} \sum_{i=1}^{l} \|\mathbf{X}_i - \mathbf{L}_i - \mathbf{S}_i\|_F^2 \\ \text{subject to} \quad & \text{rank}\{\mathbf{L}_i\} = 1 \\ & \mathbf{S}_i = \mathbf{s}\mathbf{m}_i^T \\ & \mathbf{L}_i \geq 0 \\ & \mathbf{s} \geq 0. \end{aligned} \tag{5.1}$$

An analytic closed form solution of this problem does not exist, nor can it easily be solved using convex optimization methods, as the low-rank constraint is a quasiconcave constraint [40]. Due to the large number of zero-entries in $\mathbf{S}$ (where the entries of $\mathbf{m}_i$ are zero, due to the telescope pointing off-source), the problem is related to low rank plus sparse estimation problems, such as those discussed by Slavakis et al. [41]. Slavakis et al. assume a much less structured type of matrix than used here, so methods from this source should be implemented with caution. We cannot apply methods that rely on randomly distributed nonzero samples. Another more closely related problem is the method described by Taniguchi et al. in [10], where the matrix is also much more structured. In Taniguchi et al, the nonzero entries appear in blocks however, whereas those in our data model appear in columns. In our model, the number of zero entries is smaller than in Taniguchi et al.

Despite the structure in $\mathbf{S}$, the general solution structure proposed by Slavakis et al. and Taniguchi et al. can be applied. Their approach is to split the optimization problem in a minimization problem where one of the optimization matrices is kept at a fixed value while optimizing the other, and then keeping the latter fixed at its optimal value and optimizing the first matrix. By iterating this process, both matrices are then expected to converge to the optimum. This method is widely used, although it has not been mathematically proven to converge [41].

If the problem is simply solved for $\mathbf{s}$ and $\mathbf{L}_i$ separately, the noise limit that can be reached has no improvement over a simple direct subtraction. This is because we subtract $\mathbf{S}^k$ from $\mathbf{X}$ in order to estimate $\mathbf{L}^{k+1}$. This subtraction means that there is now more noise in the on-source positions of $\mathbf{X} - \mathbf{S}^k$ than there was originally in those of $\mathbf{X}$. Taking a rank one approximation of $\mathbf{X} - \mathbf{S}^k$ cannot remove all of the

added noise and therefore there is additional noise in $\mathbf{L}_i^{k+1}$. As $\mathbf{S}_i^{k+1}$ is estimated from $\mathbf{X}_i - \mathbf{L}_i^{k+1}$, the photon noise is transferred back into $\mathbf{S}_i^{k+1}$ and there is no noise removal compared to direct subtraction. In order to avoid this, denoising strategies need to be investigated as well, such that the photon noise is left in $\mathbf{N}_i$ as much as possible.

The general form of the algorithm is shown in Alg. 2. Note that unlike in Taniguchi et al., we use an iteration number (denoted by k) as a stopping condition, rather than the Frobenius norm. This is because we want to leave the photon noise in $\mathbf{N}_i$, such that it cannot converge to zero. Furthermore, in this algorithm $\mathbf{s}$ is optimized, which is equivalent to optimizing $\mathbf{S}$ or $\mathbf{S}_i$, as the columns of these matrices are given by $\mathbf{s}$.

Algorithm 2: Basic form of the algorithm under development.

Result: $\mathbf{s}, \mathbf{L}_i$
Input: $\mathbf{X}, \mathbf{m}, k_{max}$
$\mathbf{s}^0 = \mathbf{0}$;
$k = 0$;
while $k < k_{max}$ **do**
  update $\mathbf{L}_i^{k+1} \forall i$ using $\mathbf{s}^k$;
  update $\mathbf{s}^{k+1}$ using $\mathbf{L}^{k+1}$;
  denoise $\mathbf{s}^{k+1}$;
  $k \coloneqq k + 1$;
end

The structure of this chapter is the same as that of the algorithm: first, the optimization of $\mathbf{L}_i$ is discussed, followed by that of $\mathbf{s}$. After that, we discuss the denoising of $\mathbf{s}$ and finally the full algorithm in its modular form.

5.1. Optimizing the Low Rank Atmosphere Matrix

When the signal matrix $\mathbf{S}_i$ is held constant and only $\mathbf{L}_i$ is optimized, the problem takes the form

$$\begin{aligned} \underset{\mathbf{L}_i}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{l} \|\mathbf{X}_i - \mathbf{S}_i - \mathbf{L}_i\|_F^2 \\ \text{subject to} \quad & \text{rank}\{\mathbf{L}_i\} = 1 \\ & \mathbf{L}_i \geq 0. \end{aligned} \tag{5.2}$$

This nonconvex optimization problem cannot be solved using local optimization solutions directly due to the quasiconcave rank constraint. Despite this constraint, several solutions are available. In this thesis, 3 options were considered:

- Transforming the problem into a convex problem using slack variables and solving it with descent algorithms.
- Truncating $\mathbf{X}_i - \mathbf{S}_i$ to a rank 1 matrix using a Singular Value Decomposition (SVD).
- Approximating $\mathbf{L}_i \approx -[\ln \eta_{\text{atm}}(\nu, t)]$ and fitting the time dependence of basis spectra of $\boldsymbol{\eta}_{\text{atm}}(\nu)$ for several PWV values.

All options can become very computationally expensive, due to the large data sizes involved for long observations. Fortunately, all smaller matrices $\mathbf{L}_i$ are decoupled from each other and can therefore be optimized independently. The optimization of $\mathbf{L}_i$ can thus be done in parallel, reducing the execution time of the algorithm and bringing it down to feasible lengths.

The first option is a commonly used approach and allows us to optimize the problem using relatively efficient convex optimization methods, but a downside to it is that it is only an approximation of the original problem and therefore the results may be less reliable (this is further discussed in Chapter 6). The second method is more precise, as an SVD truncated to rank 1 is the optimal approximation of a full rank matrix at rank 1. SVDs can be very computationally expensive, however.

The final option was not implemented in the scope of this thesis. The most important reason for this is that it is no longer possible to compensate for the TLS noise in the model if basis spectra are used.

In Section 3.3.2, we defined $\mathbf{L}_i = [-\ln\{\eta_{\text{atm}}(\nu,t) - T_\mu(\nu)/T_{p,\text{atm}}\}]$. The TLS noise term in this matrix cannot be modeled by a prior distribution because it is uncorrelated across channels. Since the TLS noise cannot be modeled in $\mathbf{L}_i$ if basis spectra are used, and it cannot be put into $\mathbf{N}_i$ either, as it is a constant value that will be pushed out of the Frobenius norm in optimization, it will be forced into the source matrix $\mathbf{S}_i$. This is not a desirable situation, that the top two methods do not suffer from.

5.1.1. Transforming the Problem into a Convex Problem

The first method to optimize the low rank atmosphere matrix is to transform the problem into a convex problem. The only part of the problem that is not convex yet is the rank constraint. The Frobenius norm is a convex function in $\mathbf{L}_i$ and the $\mathbf{L}_i \geq 0$ constraint describes a convex set. Using a slack variable of the nuclear norm, the problem can then be transformed into a convex problem [41]:

$$\begin{aligned} \underset{\mathbf{L}_i}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{l} \|\mathbf{X}_i - \mathbf{S}_i - \mathbf{L}_i\|_F^2 + \gamma \|\mathbf{L}_i\|_* \\ \text{subject to} \quad & \mathbf{L}_i \geq 0. \end{aligned} \tag{5.3}$$

The nuclear norm $||\cdot||_*$ denotes the sum of the singular values of the matrix inside the operator and is a convex envelope of the rank constraint [40]. The relative importance of the original minimization objective versus the slack variable that represents the rank constraint is controlled by the parameter γ.

Following the methods described in [41], this problem can be further simplified by replacing $\mathbf{L}_i$ by the vectors that span the matrix: $\mathbf{L}_i = \mathbf{a}_i\mathbf{b}_i^T$. Here, $\mathbf{a}_i$ and $\mathbf{b}_i$ represent $M \times 1$ and $N \times 1$ vectors respectively, that through multiplication create the $M \times N$ matrix $\mathbf{L}_i$. To ensure that $\mathbf{L}_i$ is nonnegative in all positions, $\mathbf{a}_i$ and $\mathbf{b}_i$ need to both be either be nonnegative or nonpositive in all entries. We opt for the former here. Furthermore, the slack variable $\|\mathbf{L}_i\|_* = \|\mathbf{a}_i\mathbf{b}_i^T\|_*$ can be approximated [1] by $(\|\mathbf{a}_i\|_2^2 + \|\mathbf{b}_i\|_2^2)/2$, such that the new formulation becomes

$$\begin{aligned} \underset{\mathbf{a}_i,\mathbf{b}_i}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{l} \|\mathbf{X}_i - \mathbf{S}_i - \mathbf{a}_i\mathbf{b}_i^T\|_F^2 + \frac{\gamma}{2}(\|\mathbf{a}_i\|_2^2 + \|\mathbf{b}_i\|_2^2) \\ \text{subject to} \quad & \mathbf{a}_i \geq \mathbf{0}. \\ & \mathbf{b}_i \geq \mathbf{0}. \end{aligned} \tag{5.4}$$

A reasonable assumption at the timescales considered in this problem is that the atmosphere remains nearly constant. In such cases, neglecting the mean value of the TLS noise, $\mathbf{a}_i$ represents the optical depth of the atmosphere. This means that $\mathbf{a}_i \geq 0$. Since $\mathbf{a}_i\mathbf{b}_i^T \geq 0$, is must be true that $\mathbf{b}_i \geq 0$ as well. This means that Eq. (5.4) can be rewritten to form a nonnegative matrix factorization problem,

$$\begin{aligned} \underset{\mathbf{a}_i,\mathbf{b}_i}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{l} \|\mathbf{X}_i - \mathbf{S}_i - \mathbf{a}_i\mathbf{b}_i^T\|_F^2 + \frac{\gamma}{2}(\|\mathbf{a}_i\|_2^2 + \|\mathbf{b}_i\|_2^2) \\ \text{subject to} \quad & \mathbf{a}_i \geq 0 \\ & \mathbf{b}_i \geq 0. \end{aligned} \tag{5.5}$$

Unfortunately, this problem is once again not convex, as $\|\mathbf{X}_i - \mathbf{S}_i - \mathbf{a}_i\mathbf{b}_i^T\|_F^2$ is not convex when both $\mathbf{a}_i$ and $\mathbf{b}_i$ are optimized simultaneously. However, a low-complexity approach that circumvents this problem is available: alternating nonnegative least squares [42]. In this approach, once again, one

[1] Note that this approximation is not absolutely necessary, as $\|\mathbf{a}_i\mathbf{b}_i^T\|_* = \|a\|_2\,\|b\|_2$. Not using the approximation makes the derivations of an optimal step such as that listed in Eq. (5.7) difficult, as $\nabla_{\mathbf{a}_i}\|a\|_2\,\|b\|_2 = \left(\|b\|_2 / \|a\|_2\right)\mathbf{a}$. A term of $\|a\|_2$ will then appear in the derivative of the objective function, such that $\mathbf{a}$ cannot easily be separated anymore, and we can no longer take an optimal step by setting the derivative to 0.

optimization variable is kept constant, while the other is optimized. This is effective because for a fixed $\mathbf{a}_i$, the problem is convex in $\mathbf{b}_i$ and vice versa.

Since the problem is now convex, a local optimization method can be used to obtain the global minimum. Therefore, in order to optimize the problem, a step must be found for $\mathbf{a}_i$ and $\mathbf{b}_i$ that comes as close as possible to the minimum of the problem in each iteration. For a convex problem, all extrema of the objective function are a global minimum. Therefore, the optimization can be achieved by setting the derivatives of the objective to zero for each iteration. If we define the objective as $f_{ab}(\mathbf{a}_i, \mathbf{b}_i) = \frac{1}{2}\sum_{i=1}^{l} \left\|\mathbf{X}_i - \mathbf{S}_i - \mathbf{a}_i\mathbf{b}_i^T\right\|_F^2 + \frac{\gamma}{2}(\|\mathbf{a}_i\|_2^2 + \|\mathbf{b}_i\|_2^2)$, the derivatives are given by

$$\begin{aligned}\nabla_{\mathbf{a}_i} f_{ab}(\mathbf{a}_i, \mathbf{b}_i) &= (\mathbf{a}_i\mathbf{b}_i^T - (\mathbf{X}_i - \mathbf{S}_i))\mathbf{b}_i + \gamma\mathbf{a}_i \\ \nabla_{\mathbf{b}_i} f_{ab}(\mathbf{a}_i, \mathbf{b}_i) &= (\mathbf{a}_i\mathbf{b}_i^T - (\mathbf{X}_i - \mathbf{S}_i))^T\mathbf{a}_i + \gamma\mathbf{b}_i.\end{aligned} \tag{5.6}$$

In order to find an optimal $\mathbf{a}_i$ for a given $\mathbf{b}_i$, we use the $\mathbf{a_i}$ that results in $\nabla_{\mathbf{a}_i} f_{ab}(\mathbf{a}_i, \mathbf{b}_i) = \mathbf{0}$ and vice versa for $\mathbf{b}_i$. The update equations for this are

$$\begin{aligned}\mathbf{a}_i^{k+1} &= \frac{(\mathbf{X}_i - \mathbf{S}_i)\mathbf{b}_i^k}{\left(\mathbf{b}_i^k\right)^T \mathbf{b}_i^k + \gamma} \\ \mathbf{b}_i^{k+1} &= \frac{(\mathbf{X}_i - \mathbf{S}_i)^T\mathbf{a}_i^{k+1}}{\left(\mathbf{a}_i^{k+1}\right)^T \mathbf{a}_i^{k+1} + \gamma}.\end{aligned} \tag{5.7}$$

Since an estimate of $\mathbf{a_i}$ is needed to estimate $\mathbf{b}_i$ and vice versa, a proper initialization is necessary. This can be done by initializing all $\mathbf{a}_i$ with the mean of the data in the temporal direction. Since the low rank atmosphere matrix is much larger than all other components in $\mathbf{X}$, this is a good first order estimate of the atmospheric transmittance. For long datasets, the dataset can be cut up into smaller sets of a few minutes to initialize the different $\mathbf{a}_i$, to avoid problems with atmospheric drift. Regardless of separating the dataset, $\mathbf{b}_i$ is initialized by executing the update equation in Eq. (5.7).

5.1.2. Truncating $\mathbf{X}_i - \mathbf{S}_i$ to a Rank 1 Matrix Using a Singular Value Decomposition

Instead of transforming the problem, an SVD can also be used of $\mathbf{X}_i - \mathbf{S}_i$ restrict its rank. Since $\mathbf{X}_i = \mathbf{S}_i + \mathbf{L}_i + \mathbf{N}_i$, subtracting $\mathbf{S}_i$ from $\mathbf{X}_i$ results in $\mathbf{L}_i + \mathbf{N}_i$. As discussed in Chapter 4, the first singular value of $\mathbf{L}_i$ dominates over that of $\mathbf{N}_i$. This means that if only the first singular vectors of $\mathbf{L}_i + \mathbf{N}_i$ are used, most of the photon noise is removed and $\mathbf{L}_i$ is estimated. Effectively, this is a projection that separates the noise subspace with $\mathbf{N}_i$ from the atmosphere plus noise subspace. By only using the first singular vectors, most of the photon noise is removed from $\mathbf{L}_i$.

Computing a full SVD is computationally expensive. Therefore, computationally it is wiser to directly compute the truncated SVD rather than the full SVD to be truncated. The implementation used in this thesis uses `sklearn` [43] to find an approximate truncated SVD and has a complexity $\mathcal{O}(MP\log(r) + (M+P)r^2)$ operations for an $M \times P$ matrix that is truncated to rank r [44]. As the width of each matrix is reasonable due to the dataset being divided into smaller matrices in order to remove the TLS noise, this SVD operation becomes feasible to use in the algorithm.

5.2. Optimizing the Source Matrix

After each iteration of the optimization of the low rank matrix $\mathbf{L}_i$, the source matrix is updated as well. Since $\mathbf{S}$ does not contain the atmospheric transmittance nor the TLS noise (see Eq. (3.19)), there is no reason stemming from the data model to limit the time support of the source matrix. Therefore, we can work with the full matrix $\mathbf{X}$ again, rather than the blocks $\mathbf{X}_i$. In order to do this, the blocks of $\mathbf{L}_i$ and $\mathbf{N}_i$ are concatenated back into $M \times N$ matrices $\mathbf{L}$ and $\mathbf{N}$. Only one vector, $\mathbf{s}$, needs to be found to optimize $\mathbf{S}$, as $\mathbf{m}$ is perfectly known prior to processing.Eq. (5.1) can then be rewritten to

$$\begin{aligned}\underset{\mathbf{s}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2 \\ \text{subject to} \quad & \mathbf{S} = \mathbf{s}\mathbf{m}^T \\ & \mathbf{s} \geq \mathbf{0}.\end{aligned} \tag{5.8}$$

The matrices in Eq. (5.8) can become millions of samples wide. This makes many optimization algorithms too inefficient to realistically use. Therefore, a way to process the data in parallel is desirable. A difficulty in parallel processing the data is that, unlike the case of the atmosphere matrix, the source vector is equal across all on-source samples. Therefore, the problem is coupled across all data blocks and it is not straightforward to parallel process the problem. In order to still be able to employ parallel methods, two solutions are considered: the Alternating Direction Method of Multipliers and a solutions where pieces of data are randomly paired every iteration.

In the parallel processed methods, the dataset is divided into n parts denoted by the number j. The number n is generally different from the number of blocks of data used to compute $\mathbf{L}$ which is denoted by l. As such the split in the data set happens in a different way. This is because the split in $\mathbf{L}$ is limited by time, whereas the split in $\mathbf{s}$ is done for parallel processing (and later denoising). The widths of the data blocks used for estimating the source matrix are thus generally much larger. Additionally, since $\mathbf{s}$ is a constant, the data blocks do not need to be a constant time stream. Only the on-source samples are needed, and they do not have to be in consecutive order.

5.2.1. Alternating Direction Method of Multipliers

Alternating Direction Method of Multipliers (ADMM) [45] is a form of consensus optimization. In this method a slack variable is used to formulate a problem equivalent to Eq. (5.8). Every block of data j has a vector $\mathbf{s}_j$ that is optimized. In other words, for each block j, a function $f_j(\mathbf{s}_j)$ is defined as a separate objective that is independent on the other data blocks. In order to ensure that one final solution will be found, a vector $\mathbf{z}$ is introduced as a central collector. All $\mathbf{s}_j$ are forced to be equal to the central collector through the constraints:

$$\begin{aligned} \underset{\mathbf{s}_j}{\text{minimize}} \quad & \frac{1}{2}\sum_{j=1}^{n} f_j(\mathbf{s}_j) \\ \text{with} \quad & f_j(\mathbf{s}_j) = \left\| \mathbf{X}_j - \mathbf{L}_j - \mathbf{s}_j \mathbf{m}_j^T \right\|_F^2 \\ \text{subject to} \quad & \mathbf{s}_j = \mathbf{z} \\ & \mathbf{z} \geq \mathbf{0}. \end{aligned} \tag{5.9}$$

Due to the constraint that forces all $\mathbf{s}_j$ to be equal to $\mathbf{z}$, the problem is not fully decoupled yet and therefore the data blocks cannot be processed independently. In order to decouple it, $\mathbf{z}$ must be put into the cost function as a slack variable, rather than as a constraint. This is done using an augmented Lagrangian,

$$\begin{aligned} \underset{\mathbf{s}_j}{\text{minimize}} \quad & L_\rho(\mathbf{s}_j, \mathbf{z}, \mathbf{y}_j) \\ \text{with} \quad & L_\rho(\mathbf{s}_j, \mathbf{z}, \mathbf{y}_j) = \sum_{j=1}^{n} f_j(\mathbf{s}_j) + \mathbf{y}_j^T(\mathbf{s}_j - \mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{s}_j - \mathbf{z}\right\|_2^2 \\ & f_j(\mathbf{s}_j) = \frac{1}{2}\left\| \mathbf{X}_j - \mathbf{L}_j - \mathbf{s}_j \mathbf{m}_j^T \right\|_F^2 \\ \text{subject to} \quad & \mathbf{s}_j \geq \mathbf{0}, \end{aligned} \tag{5.10}$$

where the dual variables $\mathbf{y}_j$ are introduced. Each full summand of L_ρ can now be optimized separately, after which $\mathbf{y}_j$ and $\mathbf{z}$ can be updated. This is done with the algorithm shown in Alg. 3.

In this algorithm, first $\mathbf{s}_j$ is updated for each block of data. Subsequently, the central collector and $\mathbf{y}_j$ are updated to the new average of the different $\mathbf{s}_j$. This ensures that $\mathbf{z}$ and $\mathbf{y}_j$ always increase the cost if the different $\mathbf{s}_j$ are more different. If all $\mathbf{s}_j$ are equal to each other, $\mathbf{z}$ is equal to $\mathbf{s}_j$, and $\mathbf{y}_j = \mathbf{0}$.

Updating $\mathbf{z}$ and $\mathbf{y}_j$ is straightforward, but updating $\mathbf{s}_j$ is more complicated as a minimization is required. The equation that is to be minimized is convex. Therefore, similarly to the first solution for updating the low rank atmosphere matrix, this can be done by finding the $\mathbf{s}_j^{k+1}$ for which $\nabla_{\mathbf{s}_j} L_\rho(\mathbf{s}_j, \mathbf{z}, \mathbf{y}_j) = \mathbf{0}$. The gradient is

$$\nabla_{\mathbf{s}_j} L_\rho(\mathbf{s}_j, \mathbf{z}, \mathbf{y}_j) = \left[\mathbf{s}_j \mathbf{m}_j^T - \left(\mathbf{X}_j - \mathbf{L}_j\right)\right]\mathbf{m}_j + \mathbf{y}_j + \rho(\mathbf{s}_j - \mathbf{z}). \tag{5.11}$$

Algorithm 3: ADMM

Result: $\mathbf{s}_j^{k+1}, \mathbf{z}^{k+1}$

Input: $\mathbf{X}, \mathbf{L}^k, \mathbf{m}, \mathbf{s}_j^k, \mathbf{z}^k, \mathbf{y}_j^k$

$\mathbf{s}_j^{k+1} \coloneqq \underset{\mathbf{s}_j}{\operatorname{argmin}} \left\| \mathbf{X}_j - \mathbf{L}_j - \mathbf{s}_j \mathbf{m}_j^{T\ T} \right\|_F^2 + (\mathbf{y}_j^k)^T (\mathbf{s}_j - \mathbf{z}^k) + \frac{\rho}{2} \left\| \mathbf{s}_j - \mathbf{z}^k \right\|_2^2;$

$\mathbf{z}^{k+1} \coloneqq \frac{1}{n} \sum_{j=1}^{n} (\mathbf{s}_j^{k+1} + \frac{1}{\rho} \mathbf{y}_j^k);$

$\mathbf{y}_j^{k+1} \coloneqq \mathbf{y}_j^k + \rho\ (\mathbf{s}_j^{k+1} - \mathbf{z}^{k+1})\ k \coloneqq k + 1$

Subsequently, we use $\left[\mathbf{s}_j \mathbf{m}_j^T - \left(\mathbf{X}_j - \mathbf{L}_j\right)\right] \mathbf{m}_j = \mathbf{s}_j \mathbf{m}_j^T \mathbf{m}_j - \left[\mathbf{X}_j - \mathbf{L}_j\right] \mathbf{m}_j$. The number $\mathbf{m}_j^T \mathbf{m}_j$ is a scalar value. As all entries of $\mathbf{m}_j$ are either 0 or 1, it is equal to $\sum_{p=1}^{N/n} (m_j)_p$, the number of on-source samples in block j. This scalar can be moved in front $\mathbf{s}_j$, such that $\mathbf{s}_j \mathbf{m}_j^T \mathbf{m}_j = \mathbf{m}_j^T \mathbf{m}_j \mathbf{s}_j$. The exact update for $\mathbf{s}_j$ that sets the gradient to zero is then

$$\mathbf{s}_j = \frac{\rho \mathbf{z} + \left(\mathbf{X}_j - \mathbf{L}_j\right) \mathbf{m}_j - \mathbf{y}_j}{\mathbf{m}_j^T \mathbf{m}_j + \rho}. \tag{5.12}$$

5.2.2. Data Pairing and a Mean Value

If $\mathbf{L}$ is estimated correctly, all noise that is left in $\mathbf{s}$ is temporally white Gaussian noise. Therefore, the sampling mean in time of $(\mathbf{X} - \mathbf{L})|_{\mathbf{m}=1}$ is the maximum likelihood estimator of the underlying source signal. Here $\cdot|_{\mathbf{m}=1}$ denotes the samples where $(m_j)_p = 1$ and the telescope is thus pointing towards the source. The sum of the on-source samples is given by $(\mathbf{X} - \mathbf{L})\, \mathbf{m}$. We can therefore use:

$$\mathbf{s} = \frac{(\mathbf{X} - \mathbf{L})\, \mathbf{m}}{\mathbf{m}^T \mathbf{m}}. \tag{5.13}$$

This method is a very straight forward way of obtaining an estimate of $\mathbf{s}$. For reasons outlined in Section 5.3, it is beneficial to have several different estimates of $\mathbf{s}$ that can be combined. In ADMM, this was automatically the case due to the way the data was split into n blocks. A similar split in the mean value approach is not directly possible, since there is no global collector forcing the estimates together. As such, the different $\mathbf{s}_j$ would not necessarily converge if the data set is split in consecutive blocks.

In order to reach convergence with the mean value approach, the matrices and vector can instead be shuffled in time and randomly split up into n equally sized chunks prior to taking the mean every time $\mathbf{s}_j$ is estimated. In this way, n separate estimate $\mathbf{s}_j$ are created. Since the brightness temperature of the galaxy is constant in time, the expectation value of the mean is the same for all data pairings. By randomly dividing the data every iteration, we avoid systematically treating timesamples differently from others, and the estimates of $\mathbf{s}_j$ will not diverge. If we define $\mathbf{A}$, the matrix with the shuffled columns of $\mathbf{X}|_{\mathbf{m}=1} - \mathbf{L}|_{\mathbf{m}=1}$, we can define the mean value method as Alg.4.

Algorithm 4: Mean value method

Result: $\mathbf{s}_j^{k+1}$

Input: $\mathbf{X}|_{\mathbf{m}=1}, \mathbf{L}^k|_{\mathbf{m}=1}, \mathbf{m}_j$

$\mathbf{A} \coloneqq$ randomly shuffle columns of $\mathbf{X}|_{\mathbf{m}=1} - \mathbf{L}|_{\mathbf{m}=1}$;

$[\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n] \coloneqq \mathbf{A}$;

$\mathbf{s}_j^{k+1} \coloneqq \mathbf{A}_j \mathbf{m}_j / (\mathbf{m}_j \mathbf{m}_j^T)$;

5.3. Denoising

If $\mathbf{L}_i^{k+1}$ is directly estimated from $\mathbf{X}_i - \mathbf{S}_i^k$, any noise that remains in the estimate of $\mathbf{s}^k$ will also introduce noise in the estimate of $\mathbf{L}_i^{k+1}$. If $\mathbf{s}^{k+1}$ is then reestimated from $\mathbf{L}_i^{k+1}$, the noise that was present in $\mathbf{s}^k$ will still be present in $\mathbf{s}^{k+1}$. Therefore, no improvement in sensitivity is expected without denoising, as the addition of noise through direct subtraction is still present. The estimate of $\mathbf{S}^k$ is noisy, because it is not

possible to completely remove $\mathbf{N}^k$ from $\mathbf{S}^k$. Therefore, it is better to create a denoised version of $\mathbf{s}^k$ to subtract from $\mathbf{X}$ prior to estimating $\mathbf{L}^{k+1}$. This does mean that $\mathbf{s}^k$ cannot be used as a final estimate of the spectrum anymore, since it has been artificially denoised and therefore does not give an accurate representation of the noise level in the spectrum. The final estimate of the galaxy brightness temperature is therefore best estimated from $\mathbf{X}^k - \mathbf{L}^k$, as this still contains the noise:

$$\hat{\mathbf{t}}_A^* = T_{p,\mathrm{atm}}\left(1 - \exp\left\{\frac{(\mathbf{X}-\mathbf{L})\,\mathbf{m}}{\mathbf{m}^T\mathbf{m}}\right\}\right). \tag{5.14}$$

In order to denoise $\mathbf{s}$, information about the shape of the spectrum can be used. Since $\mathbf{s} = [\ln(1 - T_A^*(\nu)/T_{p,\mathrm{atm}})]$, the only part that varies with channel in $\mathbf{s}$ is $T_A^*(\nu)$. This means that the broad continuum shape with much sharper spectral lines in $T_A^*(\nu)$, is also reflected in the spectrum of $\mathbf{S}$. Since both the broad and narrow features must remain in the estimate, simple filtering methods for denoising cannot be applied. Low pass filtering or other smoothing methods reduce the resolution in the spectral line channels and thereby change the shape of the lines to a wider line with a lower amplitude. Because of this, smoothing with a model spectrum was decided upon[2].

In this method, we start with a number of noisy estimates $\mathbf{s}_j^k$, obtained by either using ADMM or the mean value method with random data order. Based on this noisy estimate, a model of the continuum is created. Subsequently, spectral lines are estimated from the noisy spectra and added to the modeled continuum to create a full model of the galaxy. This full model is then used as $\mathbf{S}_i^k$.

In this way, the noise in the continuum is reduced, such that no noise peaks are subtracted from $\mathbf{X}_i$. This means that the new estimate $\mathbf{L}_i^{k+1}$ is less noisy. This noise reduction of the continuum is reduced because information from multiple channels is used to create a model for the continuum. This, in turn, means that the spectral resolution in the continuum is lower. This is not a problem, as there are no sharp features in the continuum itself.

The used model for the continuum is a third order polynomial[3]. This shape was chosen, as it has a reasonable shape for approximating the spectral shapes expected from the continuum whilst having only four degrees of freedom. The fit is weighted by $\hat{\boldsymbol{\eta}}_{\mathrm{atm}}^2$. This weighting is chosen, since the lower the atmospheric transmittance is, the more of the galaxy signal is absorbed, so the weaker the signal is. Additionally, the lower the atmospheric transmission, the more light from the atmosphere is in that channel. Therefore, the noise pollution is higher for a weaker signal if the atmospheric transmission decreases. This means that the signal to noise ratio is scaled approximately by $1/(\eta_{\mathrm{atm}}(\nu))^2$. To counter this effect, the weighting by $(\eta_{\mathrm{atm}}(\nu))^2$ is used. The resulting continuum that is fitted to $\mathbf{s}_j^k$ is denoted by $\boldsymbol{\xi}_j^k$.

After the continua $\boldsymbol{\xi}_j^k$ are estimated, they are subtracted from the noisy spectra they were fitted to. This results in residuals $\mathbf{s}_j^k - \boldsymbol{\xi}_j^k$ containing only the spectral lines and noise. In order to locate the lines, the residual is weighted by an estimate of $\hat{\boldsymbol{\eta}}_{\mathrm{atm}}^2$, such that the highest noise peaks are reduced and false line detections at the edges of the spectral windows are avoided. The estimate of $\hat{\boldsymbol{\eta}}_{\mathrm{atm}}$ is created using $\hat{\boldsymbol{\eta}}_{\mathrm{atm}} = \exp\{-\mathbf{L1}\}$. Then, the Θ highest peaks in $\hat{\boldsymbol{\eta}}_{\mathrm{atm}}^2 \circ |\mathbf{s}_j^k - \boldsymbol{\xi}_j^k|$ are selected as line locations. The absolute value is used, since this also allows for detecting absorption lines, which become negative when the continuum is subtracted. The locations of the Θ highest peaks are denoted by $\boldsymbol{\theta}_j^k$. This vector is equal to 1 when a line is present, but 0 when there is no line.

In some channels, there could be a noise peak that is larger than some of the spectral lines in the spectrum, especially in earlier iterations. If this noise peak is added back onto the continuum, it will remain in the spectrum in later iterations. In order to avoid this, a heuristic is used. A random other block of data denoted by $q \leq n$ is assigned to block j at every iteration. The line positions from block q are used as the line positions for block j. The amplitudes from block j are used. The spectral lines added back to block j are then $\boldsymbol{\theta}_q^k \circ (\mathbf{s}_j^k - \boldsymbol{\xi}_j^k)$, such that the denoised version of $\mathbf{s}_j^k$ is equal to $\boldsymbol{\xi}_j^k + \boldsymbol{\theta}_q^k \circ (\mathbf{s}_j^k - \boldsymbol{\xi}_j^k)$. As the noise lines are stochastic, it is less likely to have a strong noise peak in both blocks of data. Using the heuristic, if a false line location is selected in block q, this is not necessarily problem as the amplitude noise amplitude in the selected channel in block j is most likely small. Therefore, only a small noise peak

[2] By separating the continuum and the spectral lines as we do below, smoothing through a regularizer also becomes an option. In the scope of this thesis, this was not fully worked out, but an alternative optimization problem with a regularizer is described in Appendix D

[3] Low pass filtering and other smoothing methods were also considered for estimating the spectrum, but a polynomial fit proved to work better experimentally.

is added and it is more likely to disappear in the next pass. Since the blocks used for selecting the line locations are randomly assigned every iteration, lines must have a large amplitude in each block of data to survive the procedure. Whilst it is possible for noise lines to survive this procedure, this happens much less often than when the same residual is used for estimating both the line position and amplitude.

The hyperparameter Θ that determines the number of lines needs to be chosen with care. If the value is too low, some lines are missed. On the other hand, choosing a value that is too large can result in noise lines being misidentified as spectral lines. Even if Θ is chosen well, there is a probability that unresolved or barely resolved spectral lines that are too buried in noise are missed.

Algorithm 5: Algorithm for creating a model of the spectra estimated for $\mathbf{S}$

Result: $\mathbf{model}_j^k$
Input: $\mathbf{s}_j^k, \hat{\boldsymbol{\eta}}_{\mathrm{atm}}, n$
$j \coloneqq 1;$
while $j < n$ **do**
  $\mathbf{a} \coloneqq \underset{a_i}{\operatorname{argmin}} \left\| \hat{\boldsymbol{\eta}}_{\mathrm{atm}}^2 \circ \left(\mathbf{s}_j^k - \sum_{i=0}^{3} a_i \boldsymbol{\nu}^{(i)} \right) \right\|_2^2;$
  $\boldsymbol{\xi}_j^k \coloneqq \sum_{i=0}^{3} a_i \boldsymbol{\nu}^{(i)};$
  $j \coloneqq j + 1;$
end
$j \coloneqq 1;$
while $j < n$ **do**
  $q \coloneqq \text{random number } \{1, 2, \ldots, n\};$
  $\boldsymbol{\theta}_q^k \coloneqq \text{largest positions } \left\{ \hat{\boldsymbol{\eta}}_{\mathrm{atm}}^2 \circ \left| \mathbf{s}_q^k - \boldsymbol{\xi}_q^k \right| \right\};$
  $\mathbf{model}_j^k \coloneqq \boldsymbol{\xi}_j^k + \boldsymbol{\theta}_q^k \circ \left(\mathbf{s}_j^k - \boldsymbol{\xi}_j^k \right);$
  $j \coloneqq j + 1;$
end

The full algorithm for creating a model of the spectrum is summarized in Alg. 5. In this algorithm, $\boldsymbol{\nu}^{(i)}$ denotes a vector with the center frequencies of the channels in DESHIMA 2.0 to the power i. This is used for the polynomial fit.

5.4. Full Algorithm

Finally, all steps are combined in the algorithm shown in Fig. 5.1. The modular structure of the algorithm allows for easily changing the update steps in $\mathbf{L}$ and $\mathbf{s}$. In this way, changing the algorithm to a more optimal solution in future implementations is easily done.

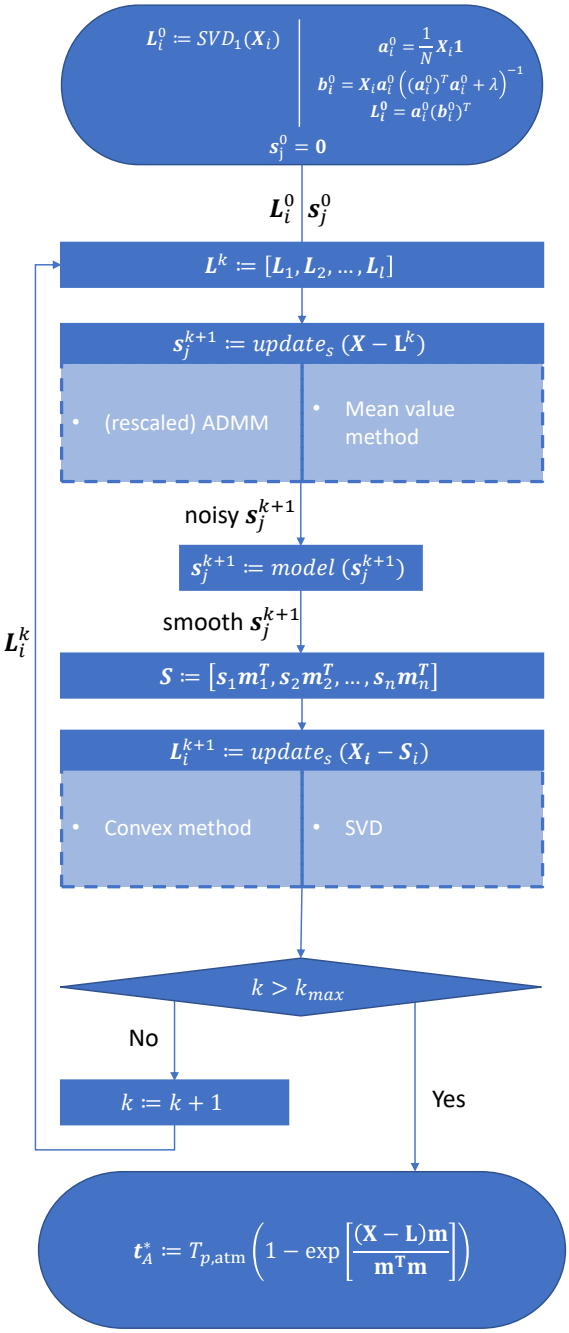


Figure 5.1. The full algorithm is shown in this figure. The rounded rectangles indicate initializations and the final estimate. The diamond shape shows a conditional statement and the rectangles are operations. In light blue, the options for performing an operation are indicated. When the algorithm is run, only one of these options is executed. This is a more detailed version of the algorithm shown in Alg. 2.

6

Simulation and Method Selection

If the algorithm is able to separate the signal from the noise and interference matrices, the properties of the estimated matrices will be comparable to those of the matrices in Chapter 4. In order to gauge how well each method performs, simulations using `TiEMPO`[1] data are presented below. The dataset used has the same properties as in Chapter 4, to make a comparison easier. First, the estimate of the low-rank matrix is discussed, followed by that of the signal matrix.

6.1. Estimating the Low-Rank Atmosphere Matrix

Since the low rank atmosphere matrix dominates the data, it is best to first find a suitable method to remove this. In order to test the methods presented in Section 5.1, a dataset of the telescope observing empty sky without chopping is used. This is the simplest problem that can be presented to the algorithm. In this section, the results of the convex method and the SVD method are presented side-by-side. The convex method requires one full iteration (one update for $\mathbf{b}_i$ and one for $\mathbf{a}_i$) after initialization to obtain a reasonable estimate of $\mathbf{L}_i$. Therefore, results after one iteration are shown here.

The means of $\mathbf{X}$, $\mathbf{L}^k$ and $\mathbf{N}^k$ are shown in Fig. 6.1 for the convex method and Fig. 6.2 for the method that uses an SVD. The mean of $\mathbf{L}^k$ fully overlaps with that of $\mathbf{X}$ for both methods, as is also the case in the model presented in Chapter 4. The mean of $\mathbf{N}^k$ looks different from the expectation, however. In the channels with a high atmospheric transmittance (where the mean of $\mathbf{L}^k$ is low), $\mathbf{N}^k$ is close to zero, as expected. In the low atmospheric transmittance channels (where the mean of $\mathbf{L}^k$ is high), however, $\mathbf{N}^k$ is much smoother than before. This means that the separation between zero mean noise and temporally correlated noise has not been entirely successful in the low transmittance channels.

[1] `10.5281/zenodo.4279085`

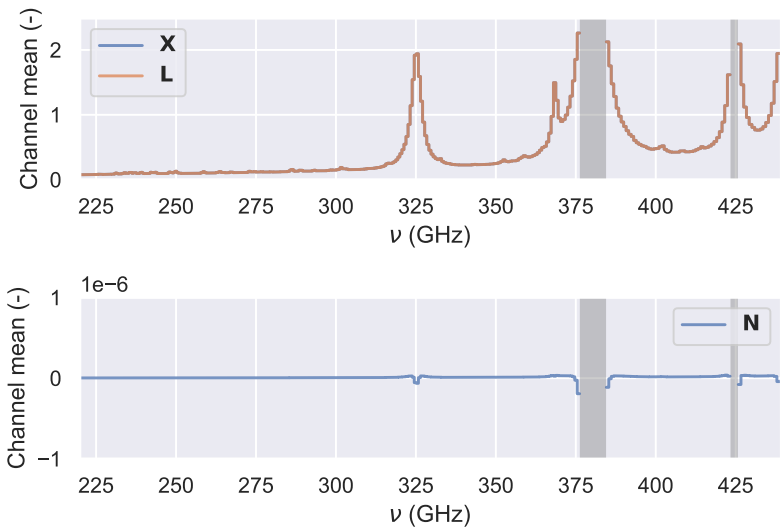


Figure 6.1. Means of $\mathbf{X}$, $\mathbf{L}^k$ and $\mathbf{N}^k$ after applying one iteration of the convex atmosphere matrix estimator. The grey areas represent omitted channels.

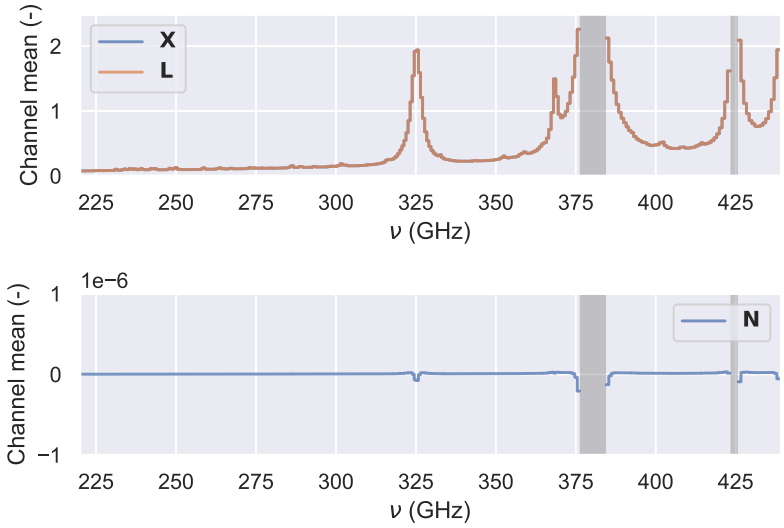


Figure 6.2. Means of $\mathbf{X}$, $\mathbf{L}^k$ and $\mathbf{N}^k$, after $\mathbf{L}^k$ has been estimated using an SVD. The grey areas represent omitted channels.

In the sampling covariances, the same type of effect is visible, where the matrices behave similar to the expectations, except for the residual noise matrix in low atmospheric transmittance channels. The estimated covariance matrix of $\mathbf{L}_i^k$ resembles that of the model in all channels (see Fig. 6.3, Fig. 6.4 and Fig. 6.5). Additionally, there is a clear diagonal visible in for $\mathbf{N}_i^k$, as can be seen in Fig. 6.7 and Fig. 6.8. In the low transmittance channels, however, $\mathbf{N}_i^k$ has a stronger correlation between different channels when the estimate is computed than according to the model from Chapter 4. This means that some of the noise from $\mathbf{L}_i^k$ remains in $\mathbf{N}_i^k$. This potentially influences the estimate of $\mathbf{S}^k$ when a signal is present, as there is residual correlated noise in $\mathbf{X}-\mathbf{L}^k$. Since this effect is bound to the low transmittance channels, an option is to raise the threshold on the atmospheric transmittance that is used to exclude low transmittance channels. This ensures that more correlated noise is removed, but also removes spectral information.

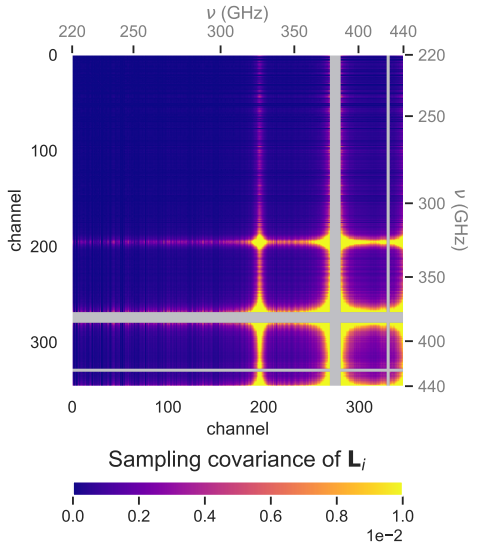


Figure 6.3. Sampling covariance matrix of $\mathbf{L}$, if the matrix is adhering to the model. The color shows the covariance.

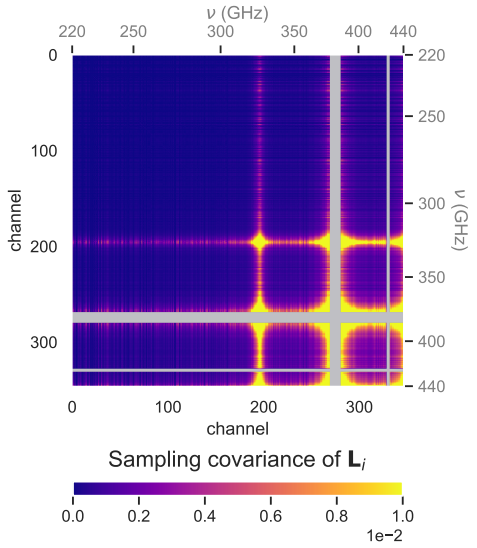


Figure 6.4. Sampling covariance matrix $\mathbf{L}_i^k$ after one iteration of the convex method with an empty sky dataset.

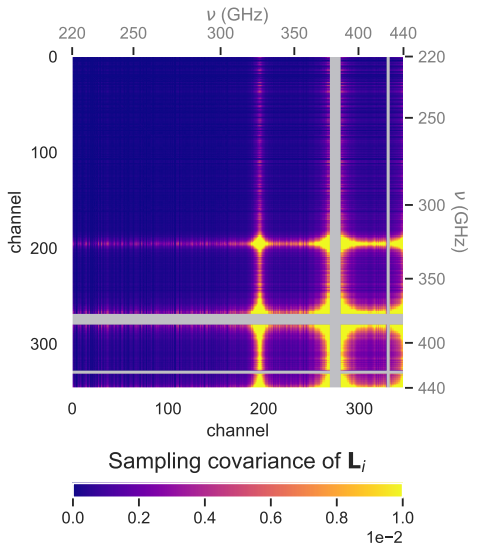


Figure 6.5. Sampling covariance matrix $\mathbf{L}_i^k$ obtained using an SVD with an empty sky dataset.

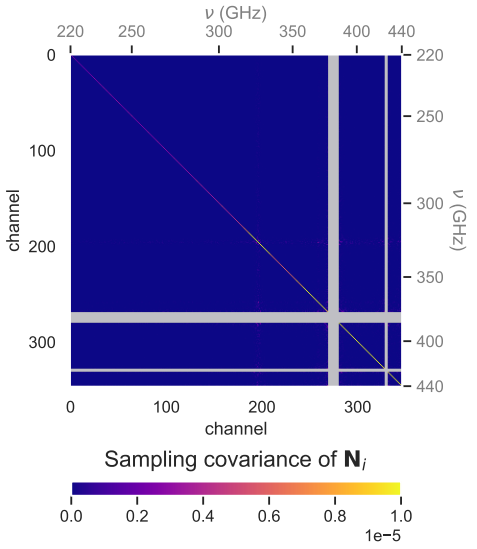


Figure 6.6. Sampling covariance matrix of $\mathbf{N}$, if the matrix is adhering to the model. The color shows the covariance.

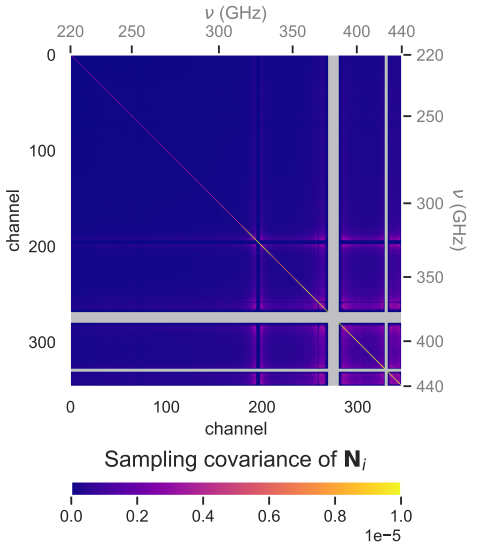


Figure 6.7. Sampling covariance matrix $\mathbf{N}_i^k$ after one iteration of the convex method with an empty sky dataset.

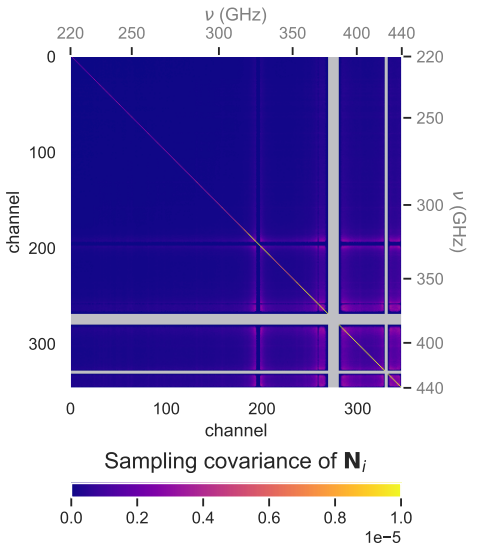


Figure 6.8. Sampling covariance matrix $\mathbf{N}_i^k$ obtained using an SVD to calculate $\mathbf{L}_i^k$ with an empty sky dataset.

The SVD is theoretically a better method to compute $\mathbf{L}_i$, as it is the optimal low-rank approximation of $\mathbf{X}-\mathbf{S}$. However, in all figures, there is no clear difference between the convex method and the method that uses an SVD. As such, we compare the results of estimating the source instead.

6.2. Estimating the Source Matrix

As both methods for estimating the low rank atmosphere matrix need to be explored further and there are two methods for estimating the source vector, four methods are tested here. All four methods are performed with the parameters in Table I, where those parameters apply.

Threshold on η_{atm}	ρ	γ	n	Θ	Iterations	Number of parallel processes
20%	10^5	10^{-6}	5	10	50	10

TABLE I. Parameters of the simulation. ρ controls how quickly the different estimates $\mathbf{s}_j$ are pushed together in ADMM, γ controls the slack variable for the nuclear norm in the convex method. The amount of different estimates $\mathbf{s}_j$ in both ADMM and the mean method is set by n. Θ controls the number of lines that the algorithm attempts to identify.

In order to keep the problem as simple as possible, a simplified chopping scheme is used. In this scheme, the telescope observes a single position in the sky, where a galaxy appears and disappears to simulate the on-off chopping. In real observations the telescope would need to move to a different field of view to observe the off-source position. The threshold on η_{atm}, was raised to 20% in this simulation, to exclude the areas where he residual noise matrix $\mathbf{N}$ cannot accurately be estimated. In the case of the estimates shown here, this means that a total of 24 channels are omitted (7% of the total amount of channels).

6.2.1. Estimating the Spectral Lines

The results of running the different versions of the algorithm are listed in Table II and Table III. The difference between these tables is the time support t_l of each matrix $\mathbf{L}_i$. In Table II, the matrices have a length of 160 samples and in Table III they have a length of 1600 samples. A shorter length means that the atmosphere and TLS noise level have less time to change significantly, so they are easier to model. On the other hand, since there are fewer samples in total per matrix the photon noise is more difficult to remove from the low-rank atmosphere matrices.

The simulations are run using `Python` on a laptop with Windows 10, an Intel i7-9750H processor with clockspeed 2.6 GHz and 16 GB RAM. In the following section, a time estimate of the computation time is also given. It should be noted that the code run here is not completely optimized for efficiency, but the execution time can give an indication of the relative complexity of the different versions of the algorithm.

The other discussed metrics are more interesting, as they give an indication of how well the algorithm converges to the desired result. The first metric is the final loss in the Frobenius norm: $||\mathbf{X}-\mathbf{L}-\mathbf{S}||_F^2$. This is the most basic version of the cost function on which all versions of the algorithm are based. Minimizing this is not the ultimate goal of the algorithm, as we ultimately only care about the estimate of $T_A^*(\nu)$. It is a good metric to check if the algorithm has converged, however.

A reasonable metric to gauge how well the algorithm recovers $T_A^*(\nu)$ would be the mean square error between the true value of $T_A^*(\nu)$ and its estimate $\hat{T}_A^*(\nu)$. The problematic part of this error is that the signal to noise ratio of the channels varies wildly. The lower the atmospheric transmittance is, the weaker the signal and the stronger the interference of the atmosphere. This means that the error estimate would be dominated by the channels in which the SNR is very low, and a good estimate is therefore often impossible. A more reasonable estimate of the error gives a larger significance to the channels inside atmospheric windows. Therefore, the difference should be weighted by $\eta_{\text{atm}}(\nu,t)$. In order to do this, the same estimator for $\eta_{\text{atm}}(\nu,t)$ as in the algorithm is used, namely $\hat{\boldsymbol{\eta}}_{\text{atm}} = \exp\{-\mathbf{L}^k\mathbf{1}\}$. The metric for how well $T_A^*(\nu)$ is computed is then $\sqrt{\frac{1}{M}\sum_\nu[\hat{\eta}_{\text{atm}}(\nu)(T_A^*(\nu)-\hat{T}_A^*(\nu))]^2}$, where M is the number of channels used to estimate the spectrum. This is listed as WRMSE (Weighted Root Mean Square Error) in Table II and Table III.

Since the fraction of lines is relatively small compared to the fraction of continuum in the spectrum, the WRMSE error is mainly useful in finding out how well the method fits the continuum. In order to judge the peak estimation, the continuum areas should be excluded. Therefore the WRMSE of only the channels

that contain spectral lines is also listed. The channels that are still included in the error calculation are highlighted in green in Fig. 6.9. The WRMSE of the spectral lines only is listed as 'WRMSE lines' in Table II and Table III.

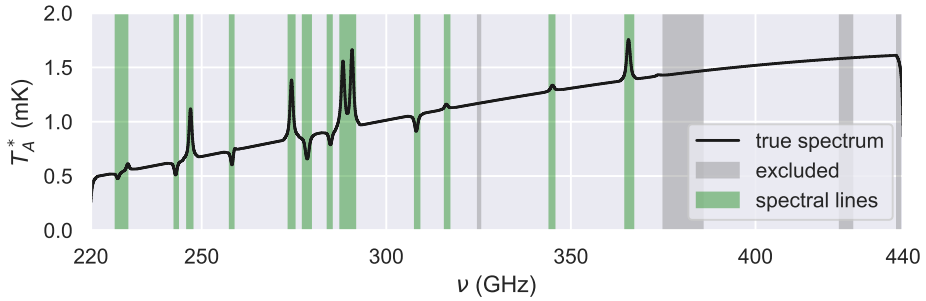


Figure 6.9. Spectral line locations selected for the computation of the error of the spectral lines.

Method	$\|\|\mathbf{X}-\mathbf{L}-\mathbf{S}\|\|_F^2$	WRMSE	WRMSE lines	Execution time of iterative part
Convex with ADMM	13.72	$7.44 \cdot 10^{-5}\mathrm{K}^2$	$9.02 \cdot 10^{-5}\mathrm{K}^2$	9m35s
Convex with mean	13.72	$7.17 \cdot 10^{-5}\mathrm{K}^2$	$9.66 \cdot 10^{-5}\mathrm{K}^2$	8m53s
SVD with ADMM	13.72	$7.17 \cdot 10^{-5}\mathrm{K}^2$	$8.65 \cdot 10^{-5}\mathrm{K}^2$	19m47s
SVD with mean	13.72	$7.10 \cdot 10^{-5}\mathrm{K}^2$	$9.80 \cdot 10^{-5}\mathrm{K}^2$	18m07s

TABLE II. Results for a width of 160 samples for $\mathbf{L}_i$

Method	$\|\|\mathbf{X}-\mathbf{L}-\mathbf{S}\|\|_F^2$	WRMSE	WRMSE lines	Execution time of iterative part
Convex with ADMM	14.72	$7.35 \cdot 10^{-5}\mathrm{K}^2$	$8.54 \cdot 10^{-5}\mathrm{K}^2$	8m18s
Convex with mean	14.72	$7.20 \cdot 10^{-5}\mathrm{K}^2$	$9.94 \cdot 10^{-5}\mathrm{K}^2$	7m06s
SVD with ADMM	14.72	$7.08 \cdot 10^{-5}\mathrm{K}^2$	$8.87 \cdot 10^{-5}\mathrm{K}^2$	13m59s
SVD with mean	14.72	$7.11 \cdot 10^{-5}\mathrm{K}^2$	$9.73 \cdot 10^{-5}\mathrm{K}^2$	11m14s

TABLE III. Results for a width of 1600 samples for $\mathbf{L}_i$

In both tables, the squared Frobenius norm does not change significantly between the different methods. Since the Frobenius norm is expected to be mostly determined by the fit of $\mathbf{L}_i^k$, this indicates once again that both the convex method and the SVD are able to fit the atmosphere matrix reasonably well, also in the presence of the signal. Furthermore, the Frobenius norm does not drop to zero, indicating that there is residual noise inside it. This is desired and in line with expectations, as it means that the photon noise is not completely contained in $\mathbf{L}^k$ or $\mathbf{S}^k$. Additionally, the Frobenius norm for the time support of $\mathbf{L}_i$ being 1600 is larger than that of $\mathbf{L}_i$ with time support 160. This is unsurprising, as there are more samples that need to be fit per free parameter in $\mathbf{L}_i^k$.

In terms of the WRMSE in the spectral line channels only, the methods that use ADMM are performing better than their counterparts that use the mean. As the detection of spectral lines is one of the main goals of this research, the lower line sensitivity for the mean method is not deemed acceptable. Therefore, for the remainder of this thesis, only the ADMM method will be considered.

The total WRMSE level is comparable with all methods on both timescales, but generally a bit larger for the convex method. Based on this, we conclude that the SVD based method is slightly better at estimating the continuum than the convex method. The results are very close, however. Since the convex method is much faster and can therefore more easily be applied to large datasets, we investigate this trade-off a bit further. Both methods converge at a similar level at the same iteration number, as is shown in Fig. 6.10. However, since the SVD method takes longer per iteration, its convergence is about twice as slow in terms of computing power.

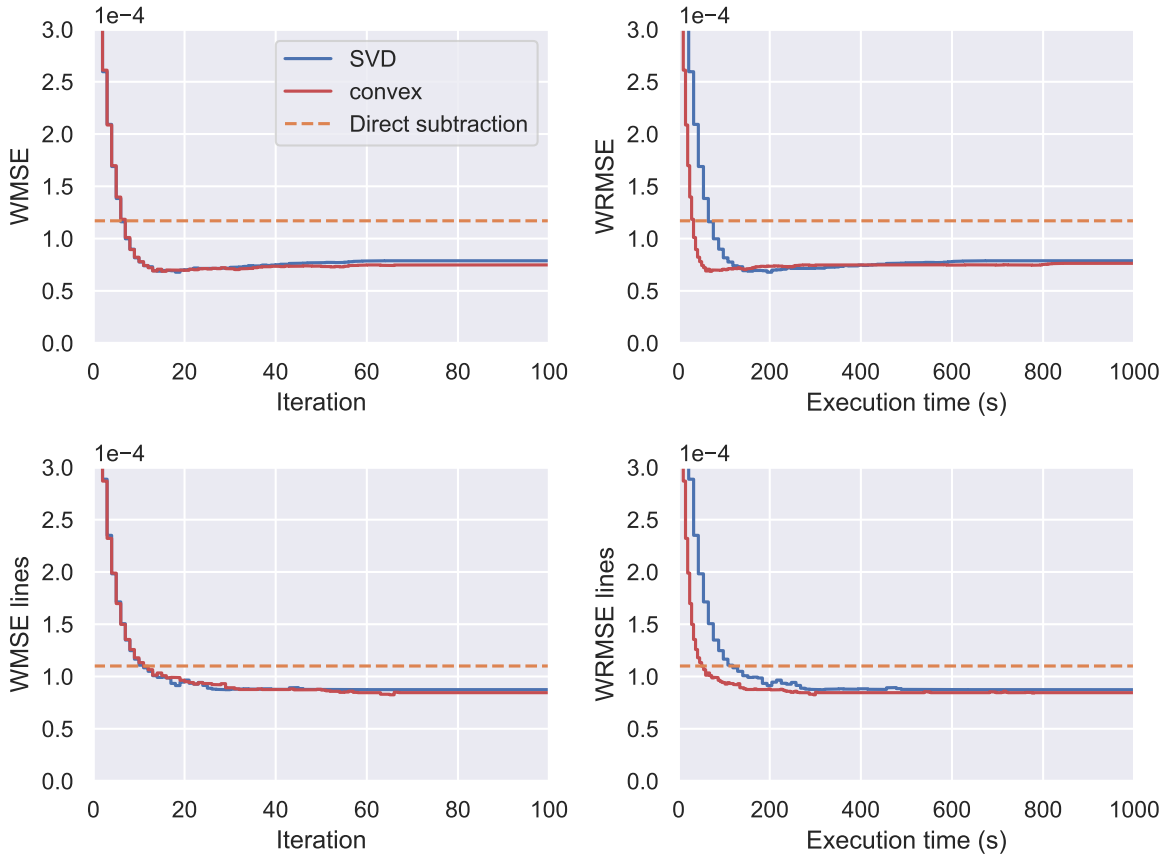


Figure 6.10. Convergence of the SVD and convex method with ADMM as a function of both iterations and execution time for $t_l = 1600$.

It should be noted that after a short period of 10-20 iteration in which both the TLS and TLS lines decrease very rapidly, the overall WRMSE increases again, whilst the WRMSE lines does not. This is most likely due to the tendency of the algorithm to slightly overestimate the signal, as is shown in Fig. 6.13. As the algorithm converges, the estimate moves from the initialization at zero to the slightly overestimated level, causing the dip. Since WRMSE lines is still decreasing at this iteration number, the algorithm cannot be stopped yet. The latter still changes because the spectral lines can only be found once the continuum is near its true position.

Fig. 6.10 also shows the error of the direct subtraction method, which is the conventional method for these estimates. SPLITTER does perform better than the direct subtraction method, but the two are much closer together for the spectral line error than for the full estimate. This is because SPLITTER cannot use the polynomial model to make a trade-off between resolution and sensitivity on the spectral lines, but it can do so on the continuum.

As can be concluded from Table II and Table III, the choice of t_l has an impact on the final estimate. Therefore, the metrics of these tables are investigated as a function of t_l for both the convex and the SVD method with ADMM. The results are shown in Fig. 6.11. A number of things stand out in this plot.

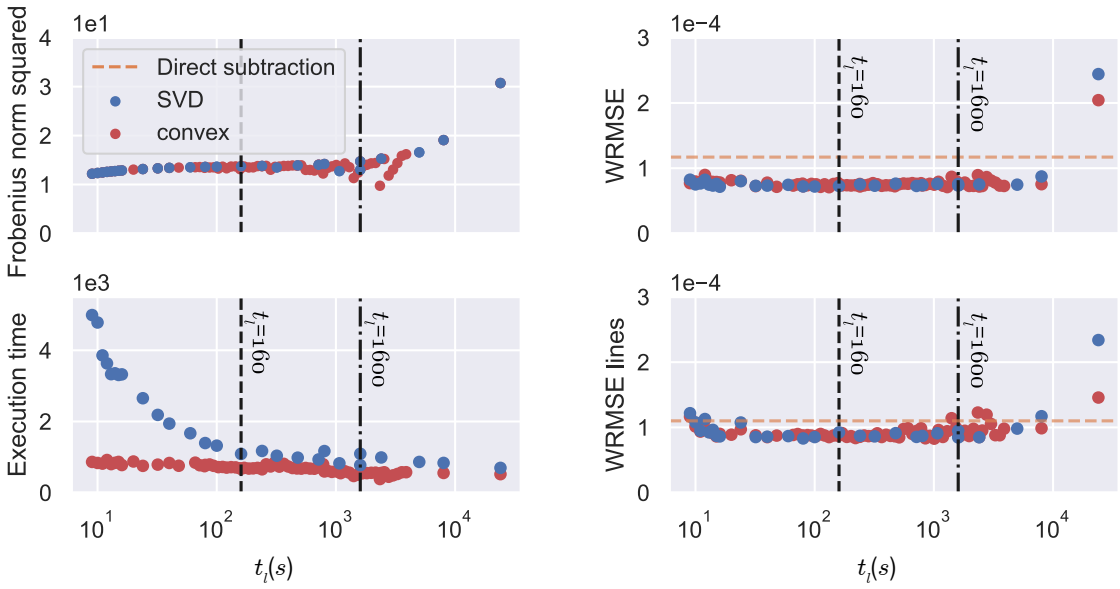


Figure 6.11. Squared Frobenius norm, WRMSE, WRMSE lines and the total execution time for the convex and SVD based methods using ADMM. The t_l values shown in Table II and Table III are shown in black.

In the top left panel, a slight increase in the trend of the squared Frobenius norm is visible as t_l increases for both methods. This is expected, but the Frobenius norm starts increasing at a faster rate past $t \sim 5 \cdot 10^3$. This indicates that a good fit can no longer be found. This is also visible in the WRMSE and WRMSE lines plots, where the estimate deteriorates for these higher values of t_l.

For very low values of t_l, in the order of $t \sim 10^1$, the estimates of the spectrum also deteriorate, especially for the spectral line regions. This could be due to two reasons. First of all, for low values of t_l, there can be a strong imbalance between the number of on- and off-source measurements. For example, if $t_l = 24$, each chunk of $\mathbf{L}_i$ contains 1.5 on-off cycles. This means that it is possible for one chunk to contain 16 on-source samples and only 8 off-source samples. This could lead to problems in estimating the level of the continuum. An argument against this is that the WRMSE does not increase a lot, whereas the WRMSE in the spectral line regions does. This leads to a second hypothesis, which is that the time support of the $\mathbf{L}_i$ matrices is so short that photon noise is fitted more strongly compared to atmospheric and TLS noise than before. This means $\mathbf{L}_i$ is estimated less well, such that the estimate of the spectral lines also becomes worse.

A third item of interest in the plots is the behaviour of the convex method and higher values of t_l. As t_l increases, the estimates of the convex method stop lying on a single line and start spreading out. This is especially visible for the Frobenius norm and the WRMSE in the spectral line areas. Interestingly, the Frobenius norm drops, whereas the WRMSE in the line regions increases. As such, the estimates obtained using the convex method become less reliable as t_l increases.

Due to the spread in results for the higher values of t_l and due to the slightly higher WRMSE in Table II and Table III, the convex method is deemed less reliable than the SVD method. Therefore, the SVD is adopted in the next chapter. It is important to note that this does not mean that the convex approximation method is not useful, however. At values near $t_l = 160$ it behaves very comparably to the SVD based methods, whilst having a shorter execution time. For large volumes of data, it can therefore be a better method.

6.3. Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR

The final version of the algorithm in the scope of this thesis is shown in Fig. 6.12. For the remainder of this thesis, this is the algorithm that is referred to as Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR (SPLITTER).

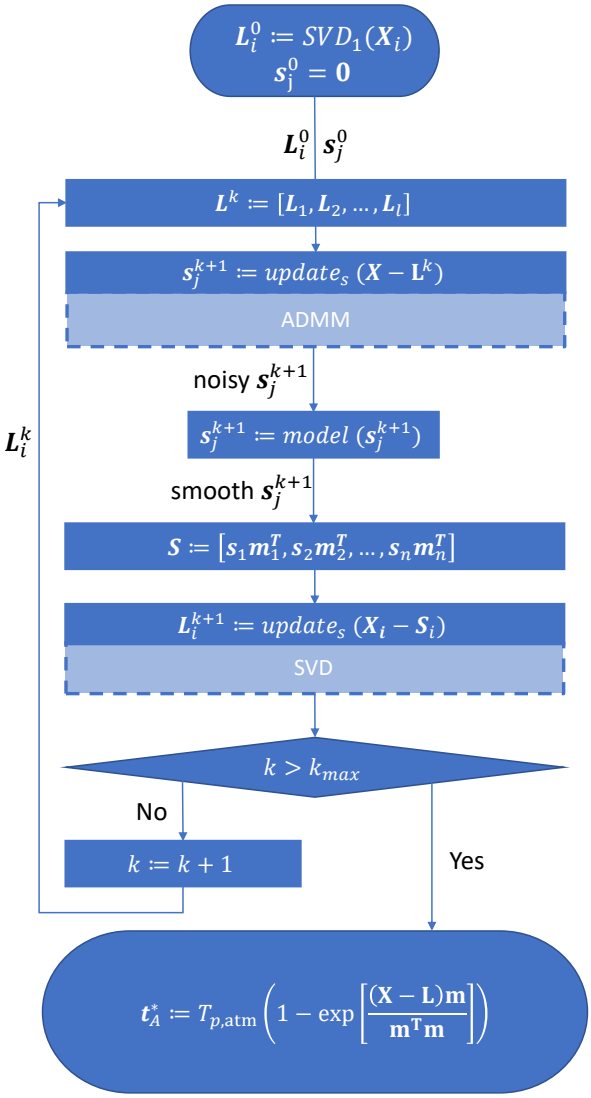


Figure 6.12. The final version of the SPLITTER algorithm used in this thesis.

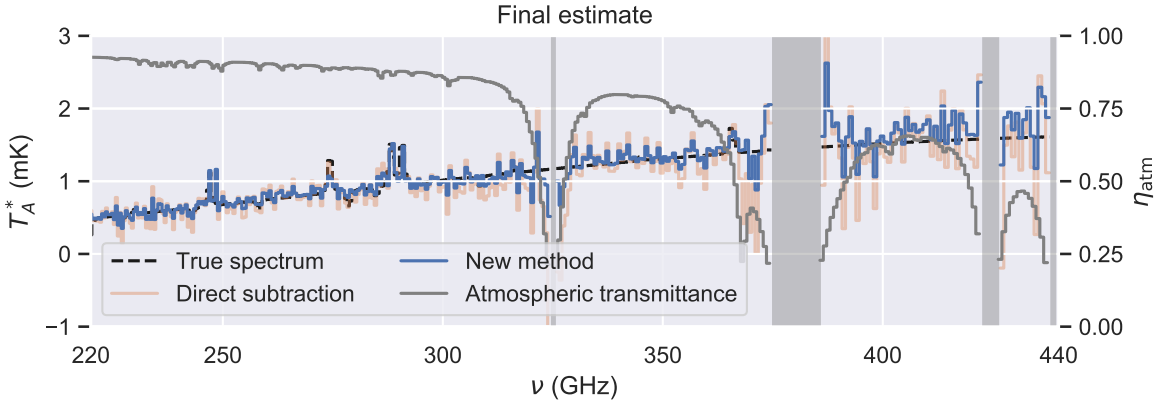


Figure 6.13. Final result of the SVD with ADMM algorithm running for $t = 1600$. The greyed out areas are omitted due to low atmospheric transmittance. There is a slight overestimate of the spectrum at higher frequencies.

Fig. 6.13 showns a final estimate produced by SPLITTER. The improvement factor in total WRMSE for this estimate is 1.65, where the improvement factor is defined as the WRMSE of the direct subtraction estimate, divided by the WRMSE of the SPLITTER estimate. In the spectral line areas, the improvement factor is only 1.26. The larger improvement in the full spectrum is due to the fact that the continuum model is used.

It is interesting to note that the 1.65 improvement is in perfect agreement with the improvement of a factor 1.65 that Taniguchi et al. report [10] in their line-free channels for a heterodyne receiver. The fact that the values are exactly equal is likely to be a coincidence, as the other versions of the algorithm result in slightly different values, and the improvement factor for SPLITTER varies from estimate to estimate. Note that the error metric used in Taniguchi et al. is different in 3 ways, namely that only the line-free channels are used, that the standard deviation of the estimate is used rather than the root mean squared error and that the result is not weighted for atmospheric transmittance. In the case of this comparison, this is not a problem however. This is because the spectral lines are only significant in a relatively small fraction of the total spectrum for SPLITTER, which means their effect on the total WRMSE is negligible. Furthermore, since Taniguchi et al. also remove their continuum using the low-rank matrix, the emission is equal to zero outside of the emission lines. For an unbiased estimator, the sampling standard deviation and root mean squared error are equivalent. Finally, since the bandwidth of the spectrum by Taniguchi et al. is much narrower than that of DESHIMA 2.0, the weighting of all channels would be nearly equal if the WRMSE would be computed. Therefore, the WRMSE would simply be the RMSE scaled by a scalar weighting factor that is common for all channels. Since this weighting factor is the same for both direct subtraction and the new method, it will drop out when the two are divided. This means that weighting has (nearly) no effect on the improvement factor for Taniguchi et al.

A final comment that is important to make before moving on to the experiments is that SPLITTER can be prone to overestimating the continuum of a galaxy for some parameter choices and noise realizations. The exact conditions for this to occur are unclear. When this occurs, the spectral lines are often still correctly identified, however. An example of this is shown in Fig. 6.14. In these cases, the estimate can be of use for applications that solely focus on the spectral lines, and remove the continuum.

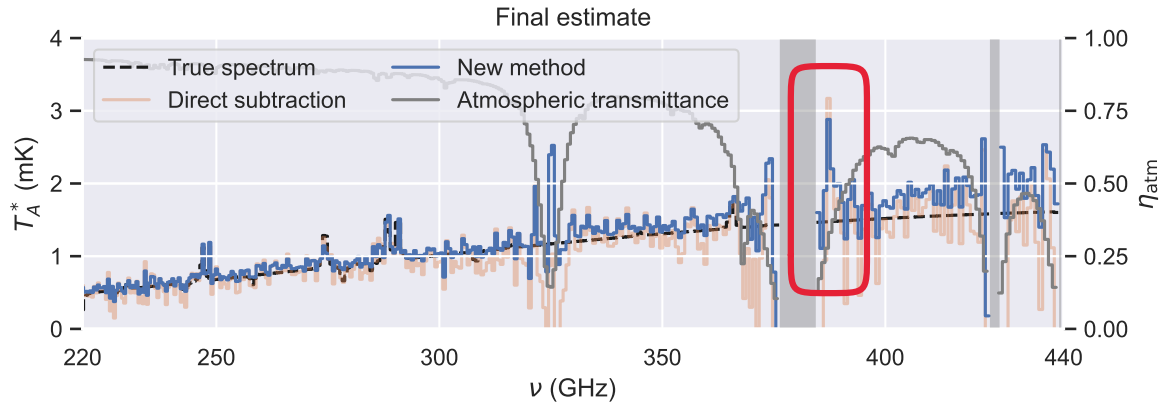


Figure 6.14. Example of an estimate where the continuum is overestimated, but the spectral lines are still detected. A false spectral line is circled in red.

Another potential issue is that SPLITTER is able to 'create' false spectral lines in the lower transmission areas, such as the one below 400 GHz in Fig. 6.14. This happens when a higher number of spectral lines Θ is given as an input than the number of lines SPLITTER can identify. In this case, the next highest peaks in the source estimates $\mathbf{s}_j$ are selected. If a noise peak is particularly strong in several $\mathbf{s}_j$, a positive feedback loop can accidentally be created that amplifies such a noise peak. For this reason, the parameter Θ that determines the amount of lines that are selected must be chosen carefully.

7

Tuning and Experiments

After determining what version of the algorithm to use as SPLITTER in the previous chapter, the performance of SPLITTER in various situations can be evaluated. In this part of the thesis, we present the performance of SPLITTER in several small experiments. These are experiments with regard to the tuning of SPLITTER's parameters and implications on observing strategies and the requirements on DESHIMA. Due to the many nonlinearities involved, this chapter is not intended as a comprehensive guide to how the parameters should be set or how the telescope should be used. It is simply intended as a first exploration of the capabilities and limitations of the algorithm.

7.1. Rebinning the Final Spectrum

Since SPLITTER relies on smoothing the estimate of the source inside the algorithm to denoise the on-off subtraction, it is important to be careful that SPLITTER is not simply smoothing the final estimate of the galaxy, but instead truly avoids adding additional noise by avoiding direct subtraction. If SPLITTER is only smoothing the final estimate, averaging over several bins in the final estimate of direct subtraction would have the same sensitivity as SPLITTER, but be much less computationally expensive. One way to tentatively check for this is to check the effect of post-rebinning both the direct subtraction estimate and the SPLITTER estimate. This is a method of post-processing where the widths of the frequency frequency channels in the final estimate are changed, such that several measurements fit inside a bin. This increases the number of measurements in a bin and reduces the variance on the estimate. Here, we simulate this by taking the average over two channels at a time. If SPLITTER is simply an artificially smoothed version of direct subtraction that contains no new information, there will be little to no gain in averaging over bins.

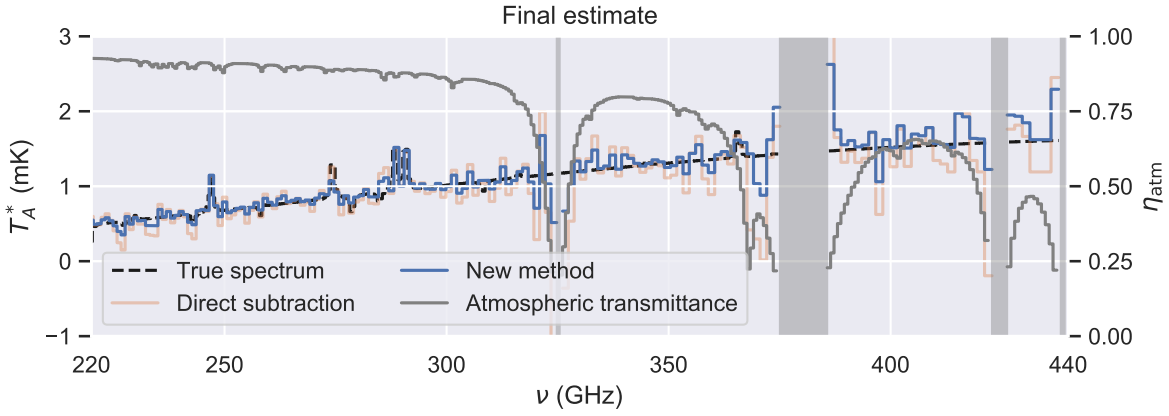


Figure 7.1. Rebinned version of Fig. 6.13. Every bin in this plot contains two of the channels in the original plot. The center frequencies correspond to the average of the two center frequencies in the bin.

Fig. 7.1 shows the result of rebinning the spectra from Fig. 6.13, where each bin contains two of the original channels. In this estimate, SPLITTER is still visually smoother than direct subtraction. The improvement factor in the WRMSE is 1.73. This is slightly higher than the original 1.65. For rebinning with an average of 3 bins instead of 2, the improvement factor is 1.58. With this test, we show that the improvement in sensitivity is not likely to be equivalent to sacrificing resolution after estimating the spectrum. This supports the claim that SPLITTER avoids increasing the noise value by avoiding the direct subtraction of noisy spectra.

7.2. Chopping Time and TLS Knee Frequency

In order to determine the effect of the knee in the noise spectrum (recall Fig. 2.6), two experiments are possible. First of all, the chop length of the observation can be changed using a set `TiEMPO` simulation. The second option is to change the knee frequency of the TLS noise.

In the first experiment, the noise spectrum is kept the same, but the frequency at which the observation changes from on-source to off-source is varied. For the direct subtraction method, this means changing the operating point on the noise spectrum to a different frequency, as it directly compares two samples. The reciprocal of the time interval between these samples is then the operating frequency. For SPLITTER, this is not as straightforward, since multiple time samples are combined to estimate the low-rank matrix. Therefore, there is no direct translation from the chopping time to the noise spectrum.

In the second option where the knee frequency of the TLS noise is changed, the TLS noise level is altered across the entire noise spectrum. This experiment is similar to the first experiment in the sense that the ratio between the knee frequency of the TLS noise and the chopping frequency is varied. On the other hand, it is fundamentally different from changing the chopping time, as it means that the knee frequency of the atmospheric interference and the TLS noise are no longer in the same range. This can be detrimental to the estimate for high TLS knee frequencies, as it is more difficult to estimate the atmospheric interference with only one vector if the TLS noise is more dominant.

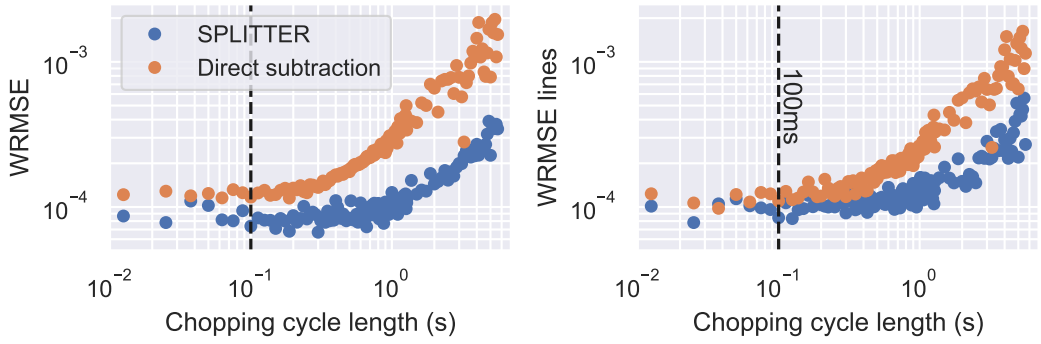


Figure 7.2. Errors as a function of chop length. The black line indicates the 100 ms chop length that corresponds to a the PSW wheel rotating at 10 Hz.

The first experiment is shown in Fig. 7.2. Here, it is clear that SPLITTER is more robust against a slower chopping speed than the direct subtraction method. The direct subtraction method starts suffering from additional noise at a chopping length of about 0.2 s, and as a result has a much higher error than SPLITTER at higher chopping cycle lengths. Direct subtraction already has a large penalty in both the continuum and the spectral line areas at a chopping time of 1 s, which corresponds to the knee frequency of 1 Hz. SPLITTER only starts to suffer around this chopping cycle length. This means that a slower PSW wheel can potentially be used when observing with SPLITTER. Unfortunately, SPLITTER does have a serious noise penalty at multisecond chopping cycle lengths. This means that a chopper wheel is still necessary to switch positions and we cannot resort to simply noddig the telescope. If a later version of SPLITTER can remain penalty free for larger chopping cycle lengths, the mechanical chopper wheel can potentially be omitted. This is therefore a potential area for future work.

The second experiment, where the knee frequency of the TLS noise is moved, is shown in Fig. 7.3. Here, the pattern looks very different. Both SPLITTER and the direct subtraction method start showing an increased WRMSE near a knee frequency of 10 Hz, which is the chop cycle frequency. For clarity, the

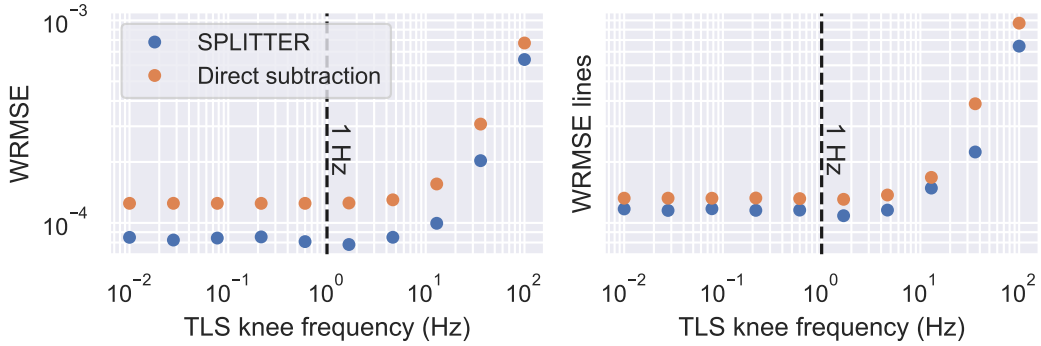


Figure 7.3. Errors as a function of TLS knee frequency. The black line indicates the 1 Hz where the knee frequency is usually located.

power spectra of the datasets used to create Fig. 7.3 are shown in Fig. 7.4. When the knee frequency of the TLS noise is increased, the $1/f$ part of the spectrum shifts up. When the knee frequency of the TLS noise becomes lower than that of the atmosphere noise, the spectrum does not change anymore, as the atmosphere noise dominates over the TLS noise at lower frequencies.

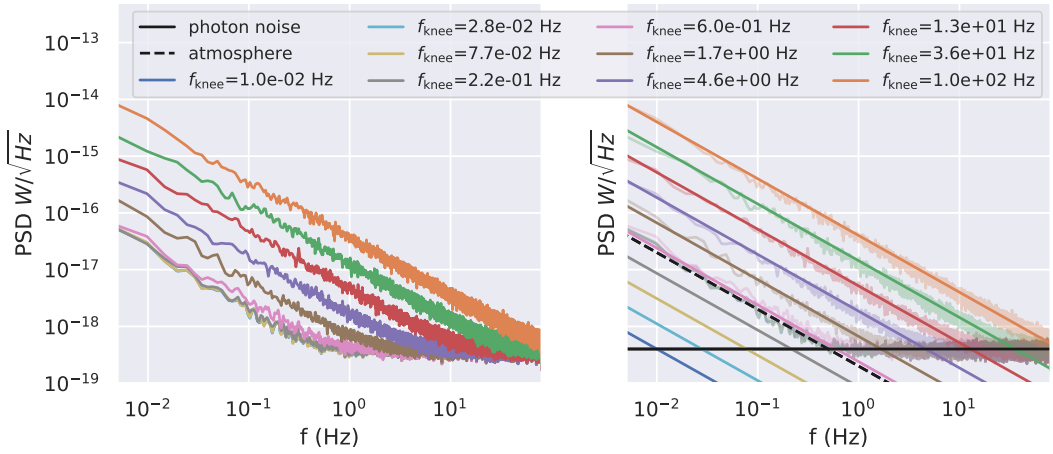


Figure 7.4. Noise power in a channel near 255 GHz. **Left**: The noise PSDs of the datasets used to create Fig. 7.3 determined using the Welch method. **Right**: The expected shapes of the components of the spectrum. The TLS $1/f$ spectra are shown in colors, in black the atmosphere spectrum and the photon noise are shown. For low knee frequencies, the TLS noise is not visible in the full PSD as the atmosphere noise and photon noise dominate over it across the entire spectrum.

From the TLS noise perspective, increasing the TLS knee frequency is equivalent to reducing the chopping cycle frequency. Therefore, the fact that SPLITTER responds differently to the change in knee frequency than it does to the change in chopping cycle length means that increase in error cannot be explained by TLS noise alone. A probable explanation is that the issue arises because increasing the TLS knee frequency means increasing the TLS noise across the entire noise spectrum (see Fig. 7.4). Therefore, the TLS noise PSD becomes larger than that of the atmosphere. This in turn results in more difficulty of fitting the atmospheric transmittance within the low rank matrix. From this, it can be concluded that the issue is most likely in the TLS noise dominating over the atmospheric noise, rather than the TLS noise dominating over the photon noise. If the TLS noise becomes too large, the approximation that the low-rank source matrix $\mathbf{L}_i$ is only dependent on time in the atmospheric transmittance is not valid anymore.

In order to check this idea, the experiment with the increased chop length is repeated without added TLS noise. The results are shown in Fig. 7.5. Here, we indeed see that a much slower PSW frequency is possible. SPLITTER is able to correctly estimate the galaxy up to a chop cycle length of almost 10 s,

or 0.1 Hz. This is a factor 10 under the knee frequency. The higher errors at chopping cycle lengths over 10 seconds long can be due to either the chopping cycle being too long, or due to the number of chopping cycles within t_l (the time support of one block of atmosphere data). Since $t_l = 1600$, chopping cycles that are just below that can have an unbalanced number of on- and off-source measurements, leading to an increased error, as in Fig. 6.11.

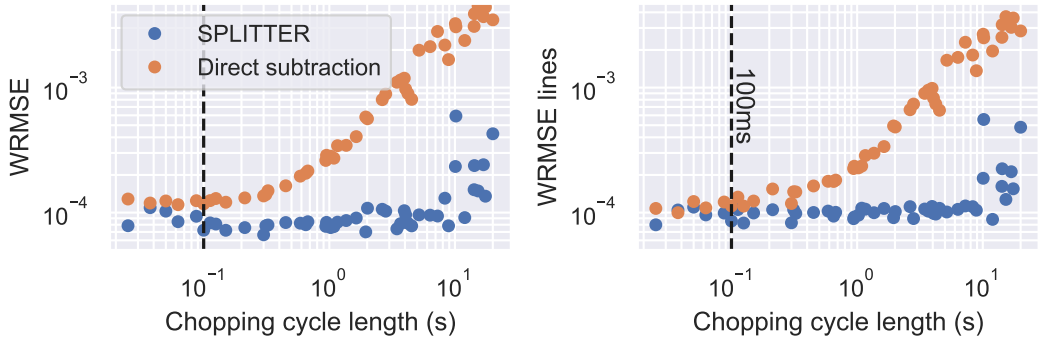


Figure 7.5. Errors as a function of chop length when TLS noise is omitted. The black line indicates the 100 ms chop length that corresponds to a the PSW wheel rotating at 10Hz. Note the different scale compared to Fig. 7.2.

7.3. Fraction of Samples On-Source

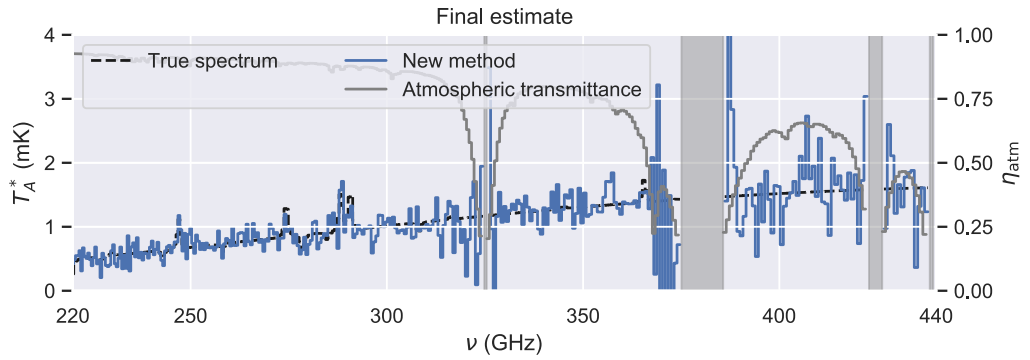
The current direct subtraction method requires the time spent observing the on-source sky to be equal to the time spent observing the on-source sky as is spent observing the empty sky. This is because the noise on the final estimate is the root of the square sum of the noise on the on-source and off-source spectra. Since this is not directly the case for SPLITTER, it is potentially possible to have more samples on-source than off-source. This leads to a higher effective observation time of the galaxy itself.

Three examples of a varied on-source fraction are shown in Fig. 7.6. The middle figure, Fig. 7.6b shows the current situation with 50% of samples on-source. The top figure has 3 out of 16 samples on-source ($\sim 20\%$) and the bottom figure has 13 out of 16 samples on-source ($\sim 80\%$).

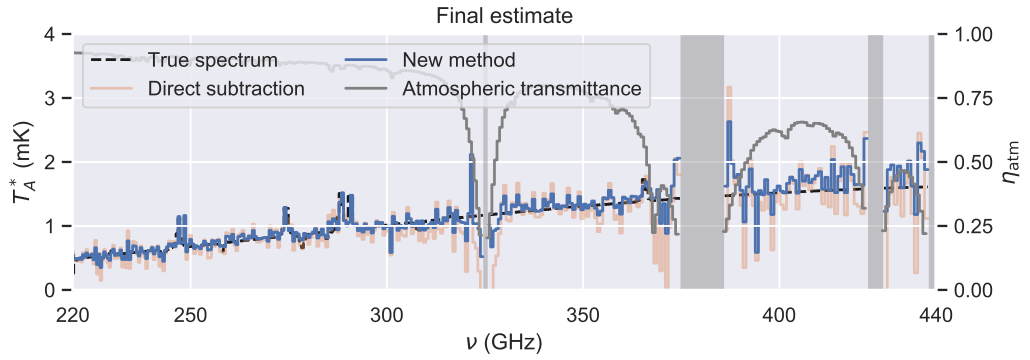
In the figure where the on-source fraction is reduced, there is more noise in the final estimate. This is comparable to the effect of reducing the total integration time. Although the noise is increased, the continuum level is estimated at the right brightness temperature, and the spectral lines near 250 GHz and 280 GHz are visible in the spectrum. In the second spectral window from the high end of the spectrum, between 380 GHz and and 420 GHz the spectrum is not very well estimated.

In the bottom plot where more time is spent on-source, the continuum is very smooth, but overestimated. Additionally, most of the spectral lines have not been found. This is most likely due to the estimate still suffering from the reduced off-source time. This makes the atmosphere more difficult to characterize. The estimate is therefore not an improvement over the traditional observing strategy.

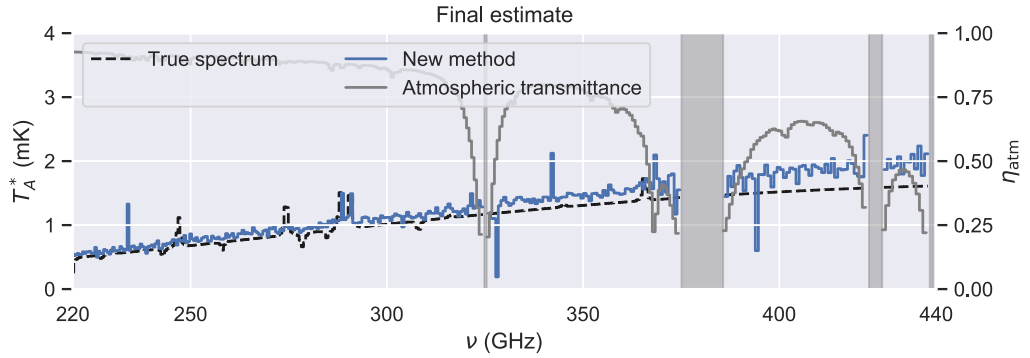
All in all, the hypothesis that SPLITTER could allow for a higher fraction on-source samples is not supported by these results. Even though the connection between the percentage of samples spent on-source and the quality of the final estimate is not as clear for SPLITTER as for direct subtraction, the estimate of the empty sky is deteriorated too much by spending less time off-source. Therefore, the observing strategy that is currently used is also a good observing strategy for the current version of SPLITTER.



(a) 3 samples on-source, 13 samples off-source for each chop cycle



(b) 8 samples on-source, 8 samples off-source for each chop cycle



(c) 13 samples on-source, 3 samples off-source for each chop cycle

Figure 7.6. Estimated spectra for different fractions of time spent on-source during the chopping cycle.

7.4. The number of Different Estimates of the Source

When using ADMM, the parameter n determines how many instances of the source vector $\mathbf{s}_j$ are created and therefore how large the data blocks of the source matrix are. Having too many can lead to a low accuracy per $\mathbf{s}_j$, as the observation time per estimate is limited. Having too few means that the spectral line detection will not work well anymore. This is because we randomly assign a block of data for the line detection to another block of data for the line amplitude. If there are too few blocks, a blocks will often be assigned to the same block for peak detection, which can result in noise lines being amplified

Fig. 7.7 shows the results for n ranging from 1 to 75 on the same dataset as before, using SVD. From this figure, it is clear that the dependence on n is much weaker than that on t_l, which determines the amount of blocks the atmosphere matrix is divided into. Most metrics stay (nearly) flat throughout the tested range. The most significant effect is in the line detection, as expected. For $n = 1$ and $n = 2$, the

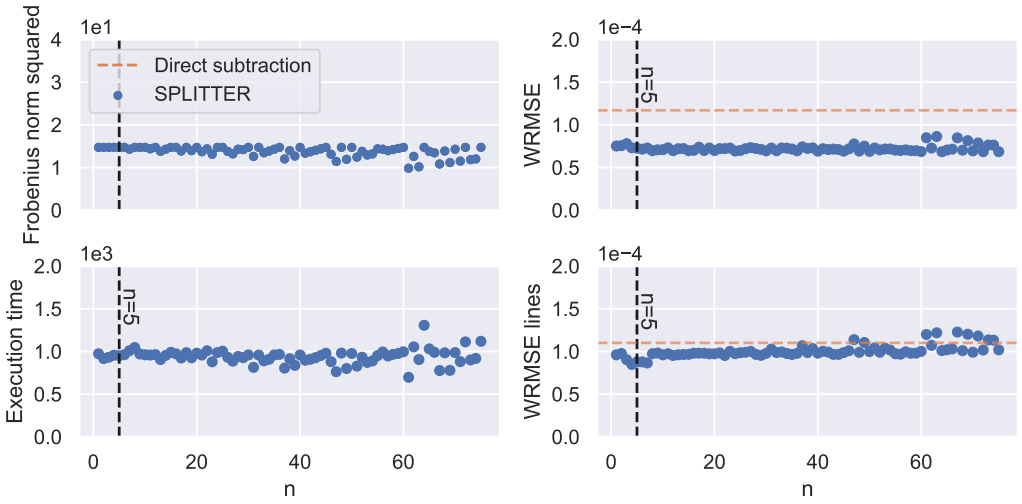


Figure 7.7. Plot of the dependence of the final estimate on n, the number of ADMM source vectors. In black, the value of $n = 5$ used in the previous chapter is shown. In the plots of the WLS and the WLS lines, the value for direct subtraction is shown in orange.

error is relatively high. For $n = 3$, there is a drop, and the error stays lower until $n = 8$, which confirms the idea that finding a suitable n is trade-off. For this dataset, a suitable n lies in the range $4 - 7$. For $n = 5$, the improvement in WRMSE compared to direct subtraction is a factor of 1.61, in the line areas the improvement is smaller at a factor of 1.25.

7.5. Threshold on the Atmospheric Transmittance

In the Chapter 6, the threshold on the atmospheric transmittance was increased from 10% to 20%, to remove the channels that induced strong non-diagonal elements to the residual noise matrix $\mathbf{N}$. The effects of this threshold can also be simulated. To this end, 30 different simulations with thresholds ranging from 0% to 29% transmittance are run. It should the noted that the threshold is a threshold on the transmittance of the atmosphere, not the amount of channels used directly. The results are shown in Fig. 7.8.

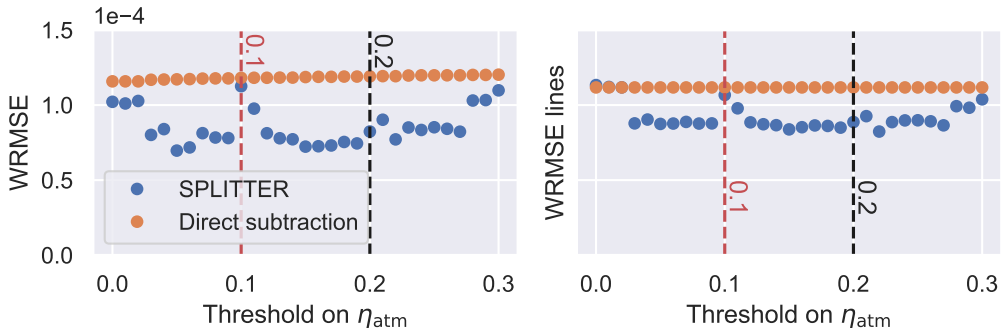


Figure 7.8. Error in the estimate as a function of the threshold on the atmospheric transmittance.

The dependence of the quality of the estimate on the threshold is not clear. There is an almost periodic looking effect, that the author is unable to explain at the time of writing. The initial choice of a threshold at 10% is very unfortunate, as this happens to be the worst outlier of the estimate shown in this plot. The increased threshold at 20% is indeed better, but not necessarily the best choice. A lower threshold of about 5% would lead to a better estimate. For the threshold at 5%, the improvement would be a factor of 1.28 in the line areas and a factor of 1.68 for the continuum.

Note that an estimate with the same data set and the exact same parameter set, aside from the

threshold on the atmospheric transmittance can lead to a very different result. This instability is unfortunate. However, the main error that SPLITTER often makes is to overestimate the continuum. Furtheremore, there is not yet a reason to use a very different observing strategy for SPLITTER and for the current direct subtraction method. This means that SPLITTER can be run on a dataset obtained for the direct subtraction method with various thresholds in use. The agreement of the mean level of the spectra obtained using SPLITTER and the direct subtraction method can then be used to find the optimal estimate. This method is far from ideal, but does offer a way of working around the instability that SPLITTER can present in its current form.

7.6. Chopping and Nodding

Thus far, only results using a perfectly aligned on-source and off-source position have been presented. In these simulations, the galaxy was 'turned on and off', whilst the telescope pointing remained stationary. Although this is possible in simulations, in the real world the telescope must have a different pointing to observe on-source and off-source positions. Therefore, a simulation using ABBA chopnodding (see Section 2.2.2) is also performed. For this simulation, a nodding cycle of 60 s was used. The result is shown in Fig. 7.9.

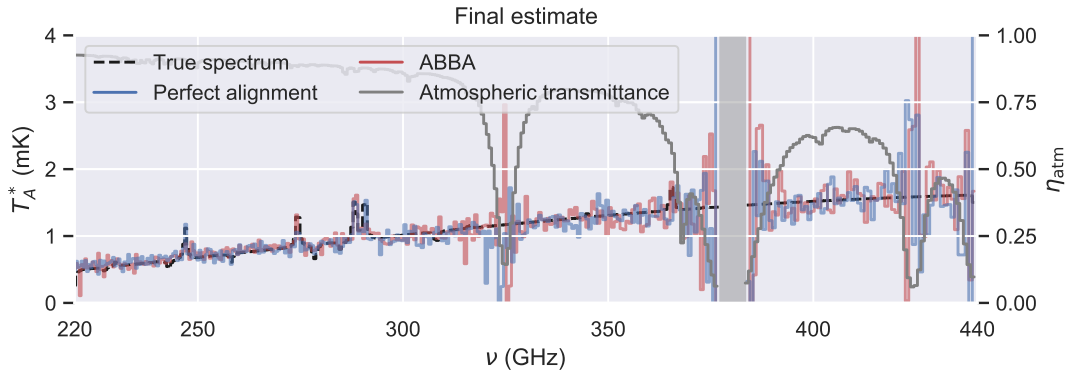


Figure 7.9. ABBA simulation vs perfect alignment of the on- and off-source beams for a galaxy with $z = 3$ with an atmospheric transmittance threshold of 5%.

The general shape of the ABBA and perfect alignment estimates are comparable, but SPLITTER has more trouble identifying the spectral lines in the ABBA estimate. One of the lines near 290 GHz is no longer found. Therefore, the estimate is negatively influenced by using ABBA chopnodding rather than the idealized situation where the beams are perfectly aligned.

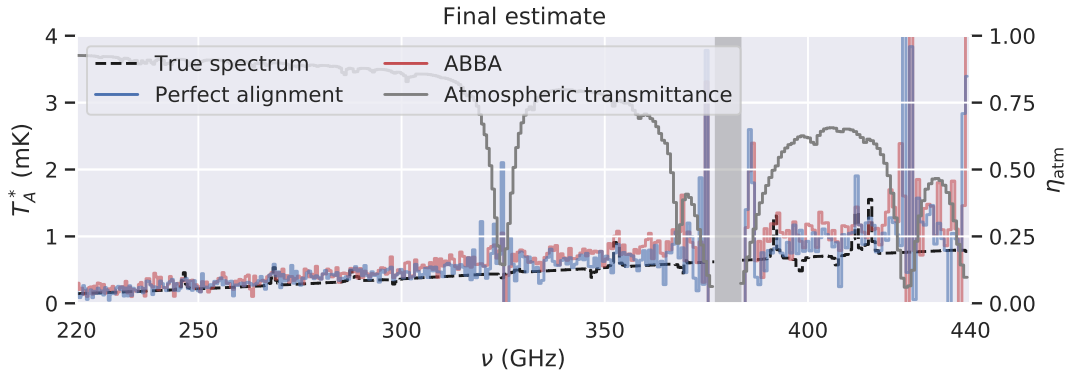


Figure 7.10. ABBA simulation vs perfect alignment of the on- and off-source beams for a galaxy with $z = 1.8$ with an atmospheric transmittance threshold of 5%.

A similar estimate, but for a galaxy with a redshift of 1.8 and a luminosity of $10^{13.4} L_\odot$ is shown in

Fig. 7.10. This emphasizes the instability that SPLITTER can display, as the galaxy is clearly overestimated in this plot, both for perfect alignment and for ABBA. This can (partially) be improved by increasing the threshold on the atmospheric transmittance, as shown in Fig. 7.11. Here, the estimate of the continuum has significantly improved, but SPLITTER is unable to correctly identify the spectral lines. For the same parameter set, a galaxy with $z = 4.43$ and a luminosity of $10^{13.8} L_{\odot}$ is also estimated, the result is shown in Fig. 7.12. Here, the ABBA estimates the continuum level better, but the large central line in the middle of the spectrum is underestimated. An overview of the improvement factors is shown in Table I.

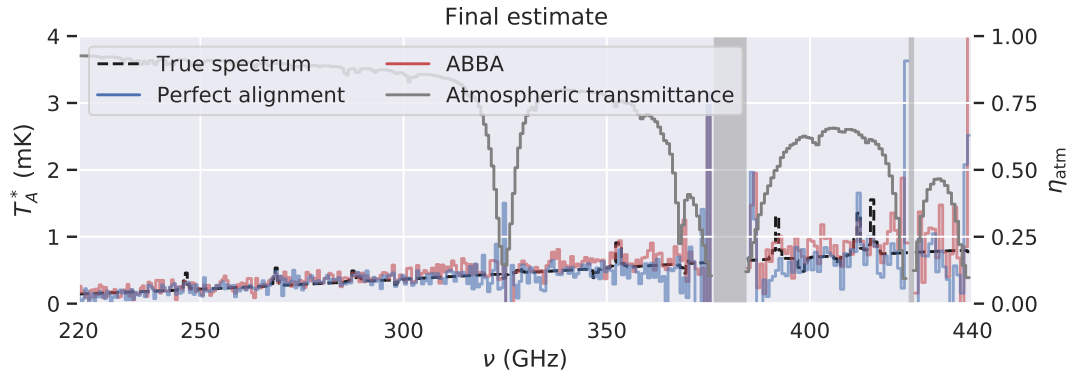


Figure 7.11. ABBA simulation vs perfect alignment of the on- and off-source beams for a galaxy with $z = 1.8$ with an atmospheric transmittance threshold of 9%.

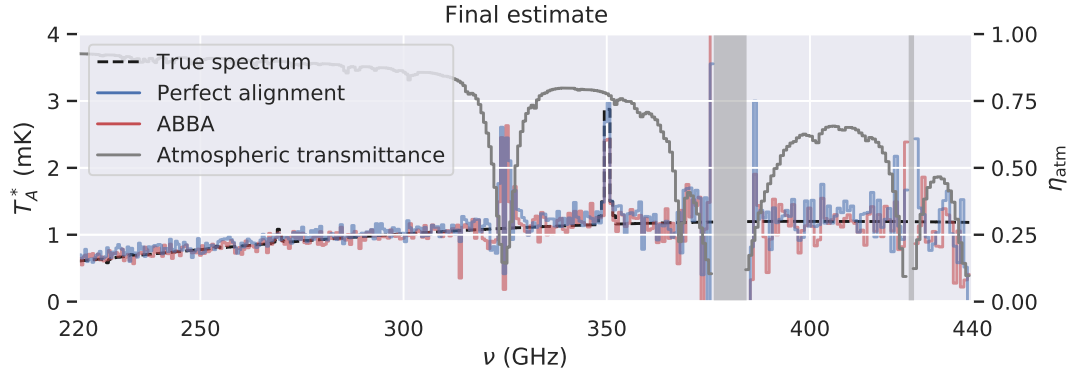


Figure 7.12. ABBA simulation vs perfect alignment of the on- and off-source beams for a galaxy with $z = 4.43$ with an atmospheric transmittance threshold of 9%.

Simulation		Shown in	Improvement factor		Line improvement factor	
z	Threshold on η_{atm}		ABBA	Perfect alignment	ABBA	Perfect alignment
3	5%	Fig. 7.9	1.68	1.57	1.27	1.16
1.8	5%	Fig. 7.10	1.02	1.43	0.88	1.23
1.8	9%	Fig. 7.11	1.56	1.74	1.29	1.32
4.43	9%	Fig. 7.12	1.74	1.67	0.98	1.41

TABLE I. Improvement factors for ABBA and perfect alignment. Line detection can be more difficult with ABBA, as can be seen by the two deteriorated errors in red. This is not the case for all ABBA experiments, however.

In general, SPLITTER is often able to recover a galaxy spectrum with a better sensitivity than the direct subtraction method. There are, however, still many cases where the estimate by SPLITTER fails to find the spectrum at the right level. From this, it can be concluded that a SPLITTER based method can be used to estimate spectra with a higher sensitivity using the same dataset, but that work is still needed in order to make the method robust enough to be reliably usable in practice.

8

Conclusions and Future Work

In this thesis, the removal of wideband atmospheric noise from spectra of high redshift galaxies obtained by the DEep Spectroscopic HIgh-redshift MApper (DESHIMA) 2.0 is investigated. This removal is done using the newly developed Stationary spectrum Plus Low-rank Iterative TransmiTtance EstimatoR (SPLITTER). This chapter describes the main findings in the thesis, the limitations of SPLITTER and provides suggestions for future work.

8.1. Discussion and Conclusions of the Research Questions

The presented work is structured around five research questions listed in Section 2.5. Together, these questions provide an answer to the main research question of how to remove wideband atmospheric noise from DESHIMA 2.0 data such that both the spectral lines and the continuum are well estimated. These subquestions are listed here with a short discussion for each question.

"How can we formulate a variation of the model presented by Taniguchi et al. in [10] that is applicable to DESHIMA 2.0?" A data model that is comparable to that of Taniguchi et al. is presented in Sections 3.1 and 3.3. Obvious differences between the data model presented in this thesis and in Taniguchi et al. are that there is no forward efficiency or variable gain term necessary for our DESHIMA 2.0 observations, as we assume that these terms are already calibrated in for in the sky temperature data we use as an input for SPLITTER. For our instrument, however, temporally correlated TLS noise must be incorporated. This means that the low-rank atmosphere matrix must be split up into shorter pieces of data. Splitting is not needed for the heterodyne receivers Taniguchi et al. work with. Additionally, the model by Taniguchi et al. focuses only on spectral lines and not the continuum, whereas we create a model that contains both. Furthermore, the spectral lines in Taniguchi et al. are much broader compared to the channel width than is the case for DESHIMA 2.0. Finally, the photon noise term is more explicitly considered in our model than in that of Taniguchi et al.

"What is the behaviour of the various noise sources (i.e. atmosphere, photon and Two Level System (TLS) noise) in this model?" The TLS noise and photon noise can be modelled together as a Gaussian. The atmospheric transmittance and the mean value of this Gaussian are put into a low-rank matrix. This allows us to remove them using low-rank estimation methods. The remainder of the Gaussian noise is put into a residual noise matrix. Chapter 4 shows the properties of these matrices. We see here that the low rank matrix is approximately a rank one matrix that has much larger values than the source and residual noise matrices. The residual noise matrix is almost a diagonal matrix, but it does contain off-diagonal terms in channels where the atmospheric transmittance is very low. If the time support of the low rank matrix is very short, this problem is circumvented, but the photon noise can distort the estimate of the low-rank matrix. Therefore, a balance must be stricken in the time support of the low-rank matrix.

"How can this data model be used to formulate an algorithm that estimates both the continuum emission and the spectral lines of high-redshift galaxies?" Several solutions were investigated

and implemented. The final version of SPLITTER uses an approximate Singular Value Decomposition (SVD) from `sklearn` [43] to estimate the low-rank atmosphere matrix. The source matrix is optimized using the Alternating Direction Method of Multipliers (ADMM).

This source matrix is denoised by creating a model spectrum for every iteration. In this model spectrum, some resolution in the continuum is traded for a higher sensitivity. Since the continuum is very broad and smooth, this loss in spectral resolution is permissible. Denoising the source matrix essentially means we do not transfer additional noise from the estimate of the source back into the low-rank atmosphere matrix. Therefore, the low rank atmosphere matrix has lower noise than is we would subtract the source from it without denoising. As there is less noise in the low-rank atmosphere matrix, estimating the source using the difference between the data matrix and the low-rank source matrix reaches a higher sensitivity than estimation using direct subtraction of noisy spectra.

"How do the estimates by SPLITTER compare to the estimates created using direct subtraction?" Especially in terms of sensitivity in the continuum, SPLITTER seems to be able to comfortably achieve a higher sensitivity than the direct subtraction method (see Section 6.2.1 and Chapter 7). However, there are also cases where the continuum is overestimated by SPLITTER, which does not occur for direct subtraction in the same situations. Therefore, SPLITTER seems to have a higher sensitivity than direct subtraction in principle, but can still arrive at a worse galaxy estimate due to robustness issues. It is not yet clear for what conditions this does and does not occur. Robustness is therefore one of the main issues that needs to be addressed.

A natural question to ask is whether the improved sensitivity thanks to the trade-off with resolution in the continuum is truly unique to SPLITTER. It is also possible to perform direct subtraction and then average over several channels to obtain a smoother final estimate. There are two ways in which SPLITTER challenges this idea. The first is that the averaging over channels is still possible with SPLITTER, as we do not actually rebin the spectrum. This means that the channels in which SPLITTER estimates the galaxy are the same as the channels in the input data set, aside from the removed channels outside the atmospheric windows. The second advantage that SPLITTER has over smoothing a direct subtraction estimate is that the telescope does not have to switch positions as fast as with direct subtraction. This means that the Position SWitching wheel (PSW) does not have to rotate as fast and other observing techniques can possibly be investigated in the future.

"Should the observing strategy of DESHIMA 2.0 be changed for this algorithm?" Aside from allowing for a lower PSW frequency (see Section 7.2), SPLITTER appears to require an observing strategy that is very similar to the optimal observing strategy for direct subtraction. Furthermore, using an observing scheme that can be used for direct subtraction allows for comparing the estimates by SPLITTER and direct subtraction. This enables the user to verify whether the continuum has been overestimated. Therefore, the current observing scheme cannot be replaced, unless a more robust version of SPLITTER is developed.

Main Research Question The results from these five questions combined can answer the main question of this thesis:

> "Is there a way to efficiently remove the wideband atmospheric noise from DESHIMA 2.0 data in order to recover both the narrow line spectra and the broad continuum emission with a higher sensitivity than the current direct subtraction method?"

The results indicate that SPLITTER is able to achieve a higher sensitivity than direct subtraction. This is shown up to a factor of ~ 1.7 over the full spectrum, with an improvement of a factor ~ 1.3 in the line areas. This means that SPLITTER is a proof of concept that such a more statistically driven method can improve the estimates of high redshift galaxies in both the line and continuum areas of the spectrum. For SPLITTER to truly have advantages over direct subtraction, more work is needed on robustness.

8.2. Contributions and Implications

Several contributions were made in this thesis. First of all, work was done on distributing existing simulation packages that are now open-source available. These tools can now be freely used by the scientific community to create wideband position switching simulations.

Secondly, a data model for DESHIMA 2.0 data was developed. This statistically oriented approach to DESHIMA data had not been worked out on this level before and can provide valuable insights for future work, even if another algorithm is adopted. Insights in the behaviour of noise sources other than the atmospheric noise, such as the photon noise, are also useful for the development of methods that only focus on spectral lines, such as the method by Taniguchi et al.

As a third point, the SPLITTER algorithm is able to reduce the error in the estimate of high-redshift galaxies in the same amount of observing time. This means that SPLITTER can potentially reach the same sensitivity in a shorter observing time. In DESHIMA 2.0, this means that more sources can be measured in the same amount of observing time, or the same amount of sources can be measured with a higher sensitivity.

A final contribution is the result that SPLITTER is able to function at a lower position switching frequency than the traditional direct subtraction method. In the current situation, SPLITTER seems to be limited by the position switching frequency compared to the TLS knee frequency, rather than the atmospheric noise frequency. Therefore, SPLITTER can potentially work with a much slower on-off frequency, if the TLS noise is sufficiently suppressed.

8.3. Limitations

Before the SPLITTER method can be used to replace other methods, however, robustness issues need to be resolved. In the current version of SPLITTER, parameters need to manually be tuned in order to remove the overestimation bias from the galaxy estimate. This means that SPLITTER is not reliable enough to replace existing methods at the current time.

Another important limitation is that SPLITTER has only been tested on simulated spectra that have a reasonably high signal-to-noise ratio. Because SPLITTER has only been tested on `TiEMPO` simulations, only a constant and known physical atmosphere temperature was used. In reality, the physical temperature can drift by a few Kelvins during a long observation. The effects of a change in atmosphere temperature are not yet investigated. Other effects not currently included are the difference between beam paths on both sides of the position switching chopper wheel. This difference will result in a different main beam efficiency for both on-sky positions, as mentioned in Section 2.2.2. The effect of these different main beam efficiencies should be investigated. Thirdly, due to the rotation of the PSW wheel in finite time and diffraction effects near the transition from on-source to off-source, a number of samples will be unusable for every chopping cycle. The impact of these missing samples should also be inspected. Finally, there is no readout noise in `TiEMPO`. This is most likely not an issue for SPLITTER due to how small the noise power is [28], but needs to be investigated nonetheless.

8.4. Future work

As described above, the most important step in developing a future version of SPLITTER is to solve the robustness issue where the continuum is overestimated. We would like to propose two potential methods to work on this issue. The first is to exploit the structure of the source matrix more. Since the positions where this matrix is equal to zero are known prior to the observation, it is known what columns of the data only contain the low-rank atmosphere matrix and the residual noise matrix. On these columns, a basis for the column space of the atmosphere matrix can be estimated before the iterative part of the algorithm starts. In this way, the estimate of the column space is not distorted by the estimate of the source. Subsequently, only the time behaviour of the atmosphere matrix needs to be fitted. A second option is to look at less restrictive smoothing methods, such as the one using a regularizer proposed in Appendix D.

Whilst the main robustness issue is in the continuum estimation, SPLITTER also occasionally misses weak spectral lines, that become smoothed like the continuum. On the other hand, SPLITTER can 'boost' false lines. Because of this, a more advanced line detection algorithm can also be very beneficial to SPLITTER.

Another interesting subject to investigate is whether a logarithmic data model is indeed the most

useful model. On the one hand, an advantage of the logarithmic model is that the logarithm of the atmospheric transmittance is well-approximated by a linear function of the precipitable water vapor. However, since the timescales of SPLITTER are very limited, it may be possible to do the same for the atmospheric transmittance directly. An advantage of this is that it simplifies the data model and fewer linearizations of the model itself are necessary.

Regardless of how robustness can finally be achieved, a robust version of SPLITTER can have big implications on the planning of future observation campaigns. Therefore, a robust version of SPLITTER also requires a more rigorous analysis of the observing strategy in order to find an optimal scheme for when direct subtraction is no longer needed to verify the results. Additionally, all simulated spectra in this thesis have a reasonably high signal-to-noise ratio. Therefore, a more thorough analysis with cases where there are only one or two barely resolved lines and a continuum below the noise level should be performed. On top of that, a better understanding of how the differences between real observations and `TiEMPO` simulations influence the results of SPLITTER is needed.

A

DESHIMA

In this appendix, a short overview of how the DESHIMA chip works is given

The DESHIMA chip is located inside a cryostat to reduce thermal noise and ensure superconductivity. A schematic of the DESHIMA chip is shown in Fig. A.1. The incoming light is guided to the fully on-chip filterbank using a leaky lens antenna. The remainder of the chip is shielded from radiation.

The detectors in DESHIMA are Microwave Kinetic Inductance Detectors (MKIDs). These detectors are suitable for weak-signal detection as they are photon noise limited. The filter bank (shown in blue) consists of 347 superconducting microstrip filters that couple the detectors to the sky signal line. This allows for extracting the energy from the line at a specific resonance frequency for each filter, without loading the line significantly at other frequencies [46]. In this way, each detector is only coupled to one specific center frequency between 220 GHz and 440 GHz. On the green side of the illustration, the shorted end of the MKID is coupled to the filter that feeds it, and the open end is coupled to the readout line [9].

MKIDS are direct detectors, which means they detect power directly. This works as follows. Inside the superconducting MKIDs, electrons condensate. This means that two fermionic electrons form a pair that can be regarded as a boson. These pairs are called Cooper pairs and exist just below the Fermi energy. Around the Fermi energy, an energy gap of width 2Δ exists. If an incident photon has an energy of $E = h\nu \geq 2\Delta$, it can break a Cooper pair up into quasiparticles that are a superposition of an electron above the gap and a hole below the gap.As Cooper pairs have a kinetic inductance, the breaking up of Cooper pairs changes the resonance frequency of the MKID (illustrated in Fig. A.2). The phase shift of this resonance frequency can then be detected in the readout line [26].

Due to the frequency dependent nature of the MKIDs, they naturally lend themselves to frequency division multiplexing. This means that only one pair of cables is needed to read out all MKIDs, which has two advantages. First of all, only two thermal loads are added, which makes the cryostat more

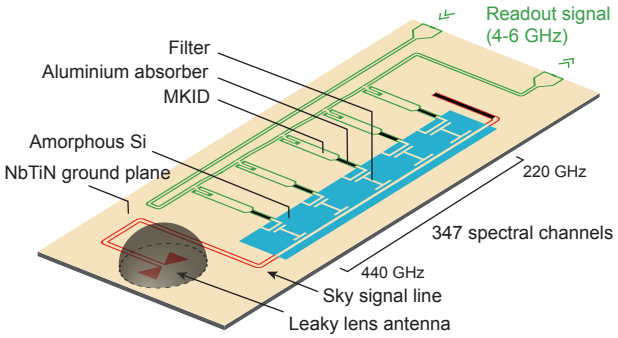


Figure A.1. Schematic of the DESHIMA 2.0 instrument. The signal arrives through the leaky lens antenna on the bottom left and then travels through the sky signal line. Subsequently, it is filtered between the 347 different filters and absorbed in the Microwave Kinetic Inductance Detectors (MKIDs). These 347 MKIDs are read out with a single readout line shown in green. Figure taken form [46].

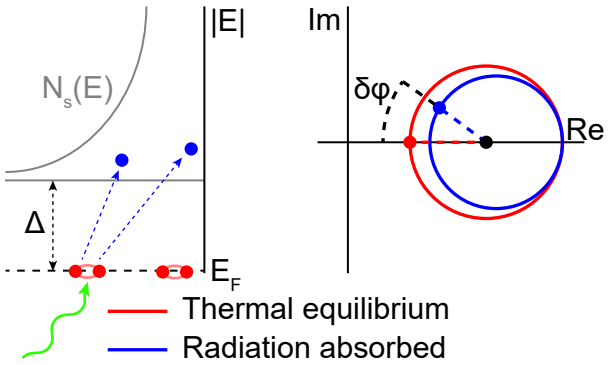


Figure A.2. Illustration of the operating principle of a Microwave Kinetic Inductance Detector (MKID). **Left:** Cooper pairs at the Fermi energy can be broken apart into quasiparticles in the upper conduction band. E_F denotes the Fermi energy in the normal state. **Right:** Transmission through a readout line with one MKID in the complex plane. As Cooper pairs are broken apart, the phase of the resonance frequency shifts. Figure based on [26].

efficient than a solution with more cables. Secondly, only one analog to digital converter is needed, which reduces the cost and complexity of the electronics. The readout signal ranges from 4 GHz to 6 GHz, but the line is read out at 160 samples per second. Which means that a full spectrum is obtained every 6.25 ms.

B

Structure function of the Precipitable Water Vapor

In TiEMPO, the change in PWV has been modeled using the Astronomical Radio Interferometer Simulator (ARIS), which in turn models the structure of the atmosphere using Kolmogorov turbulence [47],[48]. This is a thin-layer atmosphere model that is used to create a phase screen, which is later converted to a PWV screen. In this model, all water vapor in the atmosphere is modeled as being at a single layer at an altitude in the order of 1 km. The spatial variation of the atmosphere is modeled using a structure function for the spatial variation [49]. The spatial variation can subsequently be used to model the temporal behavior of the column of atmosphere above the telescope, by transforming the spatial changes into temporal changes at a stationary position using the windspeed.

The spatial structure function is the second order statistic of the difference in water vapor content,

$$D(s) = a^2 \mathbb{E}\left\{[PWV(x+s) - PWV(x)]^2\right\}. \tag{B.1}$$

In this equation, a is a constant, in this thesis $a = 6.587$ is used, based on work by [24]. The operator $\mathbb{E}\{\cdot\}$ denotes an expected value, x denotes a position in the layer of water vapor and s the displacement from this position. The structure function can be characterized using a minimum and maximum correlation length,

$$D(s) = \begin{cases} C^2 s^{5/3}, & s \leq L_1 \\ C^2 L_1 s^{2/3}, & L_1 < s \leq L_2 \\ C^2 L_1 L_2^{2/3}, & s > L_2. \end{cases} \tag{B.2}$$

Here, C denotes a constant structure coefficient and L_1 and L_2 denote the minimum and maximum correlation length, respectively. These can vary based on atmospheric conditions. In the simulations used in this thesis, C is set using $\sqrt{D(s = 100\mathrm{m})} = 50\mu\mathrm{m}$, which translates to $C \approx 1.077 \cdot 10^{-6}\mathrm{m}^{1/6}$. Furthermore, $L_1 = 1,638.4\mathrm{m}$ and $L_2 = 13,107.2\mathrm{m}$ are used.

As noted above, the structure function can also be written as a function of time for a stationary observer using a constant windspeed v_{wind}. Let τ' be a time difference. In that case, $s = v_{\mathrm{wind}}\tau'$ and the structure function becomes $D(\tau') = a^2 \mathbb{E}\left\{[PWV(t+\tau') - PWV(t)]^2\right\}$. Therefore, the equation can also be written as:

$$\mathbb{E}\left\{[PWV(t+\tau') - PWV(t)]^2\right\} = \begin{cases} \left(\frac{C}{a}\right)^2 v_{\mathrm{wind}}^{5/3} t^{5/3}, & \tau' \leq T_1 \\ \left(\frac{C}{a}\right)^2 v_{\mathrm{wind}}^{5/3} T_1 t^{2/3}, & T_1 < \tau' \leq T_2 \\ \left(\frac{C}{a}\right)^2 v_{\mathrm{wind}}^{5/3} T_1 T_2^{2/3}, & \tau' > T_2, \end{cases} \tag{B.3}$$

where $T_1 = L_1/v_{\text{wind}}$ and $T_2 = L_2/v_{\text{wind}}$. For a windspeed of 10 m/s, T_1 is approximately 2.5 minutes and T_2 is approximately 22 minutes. Inside `TiEMPO`, the PWV is only known at discrete positions, spaced at a distance of 20cm for this thesis.

C

Discretization in Power Spectral Density

In Fig. 2.6, the atmospheric noise shows a sharper drop-off near 1 Hz than the expected $1/f$ drop-off. In this appendix, we explain why this occurs, and why this is not problematic.

The PSD spectra were created using `TiEMPO`. In particular, the atmospheric noise spectrum is simply created using a timestream of an off-source beam with photon noise turned off. In order to simulate the temporal variations of the atmosphere, `TiEMPO` uses `ARIS` [48]. Without going into detail how the atmosphere is simulated, it suffices to know that `ARIS` simulates the atmosphere at discrete grid points. `TiEMPO` then interpolates the atmosphere between those points to create a screen of the atmosphere for the full observation.

A zoomed out version of Fig. 2.6 is shown in Fig. C.1. In this figure, the artifacts caused by the discrete grid points are much clearer. In the case of the simulation shown in Fig. C.1, the grid points are spaced 20 cm apart. Furthermore, a windspeed of 9.6 m/s is used. This means that a new grid point enters the beam every ~21 ms. This is at a frequency of 48 Hz, where the largest 'bump' in the atmospheric noise spectrum is shown. The other visible peaks are at 16 Hz, 32 Hz and 64 Hz. These are harmonics of the 48 Hz peak, with aliasing due to the 160 Hz sampling frequency of DESHIMA. This means that the steeper drop off and peaks are not a fundamental property of the atmosphere, but a modulation caused by the imperfect simulation.

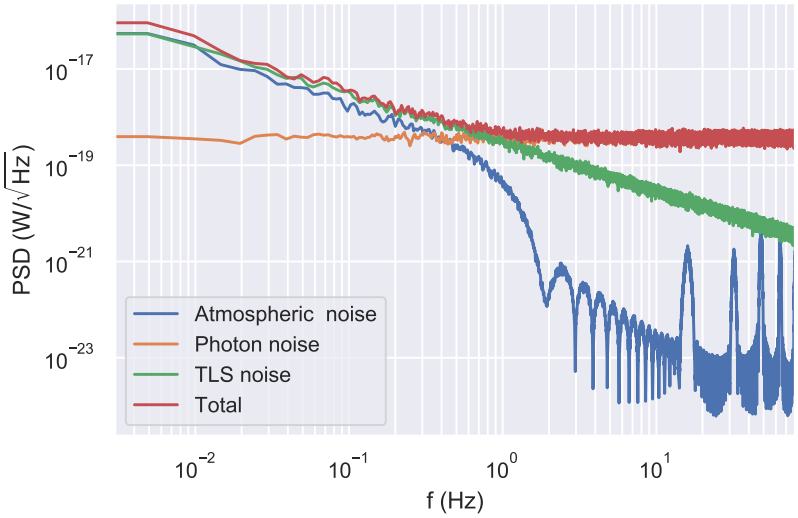


Figure C.1. Frequency noise spectrum of the different noise sources in a DESHIMA channel near 300 GHz. Note that the frequency axis shown here is a sampling frequency axis, rather than the frequency of the light incident on the detector. This zoomed out version of the figure clearly shows aliasing effects.

Although the modulation in the figure *looks* quite extreme, it is actually not expected to be prob-

lematic for the results of the simulation. This is because the photon noise (shown in orange) is much stronger than both the realized and the expected PSD of the atmospheric noise in the regions where the atmospheric noise does not follow a $1/f$ distribution. As such, the regions where the deviation is strong are dominated by photon noise and the atmospheric noise does not play a significant role at these frequencies.

D

Alternative Denoising Strategy

In the main method described in this thesis, a model spectrum is fitted to denoise the estimate of the source matrix. This is a restrictive way of denoising, since we force the continuum to adhere to the shape of the chosen model, in this case a third order polynomial. A method without a model spectrum means that the fit is more flexible. One such methods is to add a regularizing term to the spectrum.

The optimization of this new cost function can then be intialized by the noisy estimate $\mathbf{s}_j$. It takes the place where the continuum fit is in the current estimation method. This is because the regularizer will also smooth out the spectral lines, making them broader and lowering their amplitude.

A regularizer for the spectrum must punish large channel to channel variations in the continuum, as the shape of the spectrum is smooth in the continuum by definition. Small channel to channel variations should still be permissable, however. Therefore, an ℓ_2-norm regularizer is used, as this regularizer punishes large variations. The new cost function then becomes:

$$\begin{aligned} \underset{\mathbf{s_j}}{\text{minimize}} \quad & L_\rho(\mathbf{s_j}, \mathbf{z}, \mathbf{y_j}) \\ \text{with} \quad & L_\rho(\mathbf{s_j}, \mathbf{z}, \mathbf{y_j}) = \sum_{j=1}^{n} f_j(\mathbf{s_j}) + \mathbf{y_j}^T(\mathbf{s_j} - \mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{s_j} - \mathbf{z}\right\|_2^2 \\ & f_j(\mathbf{s_j}) = \frac{1}{2}\left\|\mathbf{X}_j - \tilde{\mathbf{L}}_j - \mathbf{s_j}\mathbf{m}_j^T\right\|_F^2 - \frac{\mu}{n}\mathbf{s_j}^T\mathbf{1} + \frac{\gamma}{n}\left\|\frac{\mathrm{d}\mathbf{s}_j}{\mathrm{d}\nu}\right\|_2^2 \\ \text{subject to} \quad & \mathbf{s_j} \geq \mathbf{0}. \end{aligned} \tag{D.1}$$

Where $\gamma \geq 0$ is the weight of the regularizer. The $(M+1) \times 1$ vector $\frac{\mathrm{d}\mathbf{s}_j}{\mathrm{d}\nu}$ represents the channel to channel variation and is given by $\frac{\mathrm{d}\mathbf{s}_j}{\mathrm{d}\nu} = ([\mathbf{s}_j^T, 0]^T - [0, \mathbf{s}_j^T]^T) \oslash ([\boldsymbol{\nu}^T, 0]^T - [0, \boldsymbol{\nu}^T]^T)$, where $\oslash$ denotes Hadamard or elementwise division. In this equation $\boldsymbol{\nu}$ is a vector that contains the center frequencies of the filters. This means that $\left(\frac{\mathrm{d}\mathbf{s}_j}{\mathrm{d}\nu}\right)_m = [(s_j)_m - (s_j)_{m-1}]/(\nu_m - \nu_{m-1})$. This cost function contains a new hyperparameter γ, which must be chosen with care.

Bibliography

[1] I. Furenlid, “A comparison of CCD’s and photographic plates.” *AAS Photo Bulletin*, vol. 36, pp. 5–8, Jan. 1984.

[2] S. S. Dhillon, M. S. Vitiello, E. H. Linfield, A. G. Davies, M. C. Hoffmann, J. Booske, C. Paoloni, M. Gensch, P. Weightman, G. P. Williams, E. Castro-Camus, D. R. S. Cumming, F. Simoens, I. Escorcia-Carranza, J. Grant *et al.*, “The 2017 terahertz science and technology roadmap,” *Journal of Physics D: Applied Physics*, vol. 50, no. 4, p. 043001, jan 2017. [Online]. Available: https://doi.org/10.1088/1361-6463/50/4/043001

[3] G. Davies and E. Linfield, “Bridging the terahertz gap,” *Physics World*, vol. 17, no. 4, pp. 37–41, apr 2004. [Online]. Available: https://doi.org/10.1088/2058-7058/17/4/34

[4] C. M. Casey, D. Narayanan, and A. Cooray, “Dusty star-forming galaxies at high redshift,” *Physics Reports*, vol. 541, no. 2, pp. 45–161, 2014.

[5] A. Endo, K. Karatsu, Y. Tamura, T. Oshima, A. Taniguchi, T. Takekoshi, S. Asayama, T. J. Bakx, S. Bosma, J. Bueno *et al.*, “First light demonstration of the integrated superconducting spectrometer,” *Nature Astronomy*, vol. 3, no. 11, pp. 989–996, 2019.

[6] A. Endo, K. Karatsu, Y. Tamura, T. Oshima, A. Taniguchi, T. Takekoshi, S. Asayama, T. J. L. C. Bakx, S. Bosma, J. Bueno, B. Buijtendorp, K. W. Chin, Y. Fujii, K. Fujita, E. Huijten, R. Huiting, S. Ikarashi, T. Ishida, S. Ishii, R. Kawabe *et al.*, “DESHIMA on ASTE: On-sky performance and upgrade to ultra wideband,” in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy X*, vol. 11453, International Society for Optics and Photonics. SPIE, 2020. [Online]. Available: https://doi.org/10.1117/12.2560669

[7] A. Taniguchi, Y. Tamura, T. Bakx, K. Suzuki, T. Takekoshi, K. Kohno, T. Tsukagoshi, T. Oshima, A. Endo, S. Ikarashi, and K. Karatsu, “DESHIMA on ASTE: Sky removal method for astronomical observations with an ultra-wideband submillimeter spectrometer,” in *18th International Workshop on Low Temperature Detectors*, 2018. [Online]. Available: https://agenda.infn.it/event/15448/contributions/95684/

[8] T. L. Wilson, K. Rohlfs, and S. Hüttemeister, *Tools of radio astronomy*. Springer, 2009, vol. 5.

[9] A. Endo, K. Karatsu, A. P. Laguna, B. Mirzaei, R. Huiting, D. J. Thoen, V. Murugesan, S. J. Yates, J. Bueno, N. van Marrewijk *et al.*, “Wideband on-chip terahertz spectrometer based on a superconducting filterbank,” *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 5, no. 3, p. 035004, 2019.

[10] A. Taniguchi, Y. Tamura, S. Ikeda, T. Takekoshi, and R. Kawabe, “A data-scientific noise-removal method for efficient submillimeter spectroscopy with single-dish telescopes,” Submitted for publication.

[11] T. Bakx and S. A. Brackenhoff, “deshima-dev/galspec: Update release,” Nov. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4279062

[12] E. Huijten, Y. Roelvink, S. A. Brackenhoff, A. Taniguchi, T. J. Bakx, K. B. Marthi, S. Zaalberg, J. J. Baselmans, K. W. Chin, R. Huiting *et al.*, “TiEMPO: Open-source time-dependent end-to-end model for simulating ground-based submillimeter astronomical observations,” in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy X*, vol. 11453. International Society for Optics and Photonics, 2020, p. 114533C.

[13] F. LeBlanc, *An introduction to stellar astrophysics*. John Wiley and Sons, 2010.

[14] E. L. Wright, “A Cosmology Calculator for the World Wide Web,” *Publications of the Astronomical Society of the Pacific*, vol. 118, no. 850, pp. 1711–1715, dec 2006.

[15] C. M. Casey, J. A. Zavala, J. Spilker, E. da Cunha, J. Hodge, C.-L. Hung, J. Staguhn, S. L. Finkelstein, and P. Drew, “The brightest galaxies in the dark ages: Galaxies’ dust continuum emission during the reionization era,” *The Astrophysical Journal*, vol. 862, no. 1, p. 77, 2018.

[16] J. Kamenetzky, N. Rangwala, J. Glenn, P. R. Maloney, and A. Conley, “L’CO/LFIR relations with CO rotational ladders of galaxies across the Herschel SPIRE archive,” *The Astrophysical Journal*, vol. 829, no. 2, p. 93, sep 2016. [Online]. Available: https://doi.org/10.3847/0004-637x/829/2/93

[17] M. J. F. Rosenberg, P. P. van der Werf, S. Aalto, L. Armus, V. Charmandaris, T. Díaz-Santos, A. S. Evans, J. Fischer, Y. Gao, E. González-Alfonso, T. R. Greve, A. I. Harris, C. Henkel, F. P. Israel, K. G. Isaak, C. Kramer, R. Meijerink, D. A. Naylor *et al.*, “The herschel comprehensive (u)LIRG emission survey (HERCULES): CO ladders, fine structure lines, and neutral gas cooling,” *The Astrophysical Journal*, vol. 801, no. 2, p. 72, mar 2015. [Online]. Available: https://doi.org/10.1088/0004-637x/801/2/72

[18] M. Bonato, M. Negrello, Z.-Y. Cai, G. De Zotti, A. Bressan, A. Lapi, C. Gruppioni, L. Spinoglio, and L. Danese, “Exploring the early dust-obscured phase of galaxy formation with blind mid-/far-infrared spectroscopic surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 438, no. 3, pp. 2547–2564, 01 2014. [Online]. Available: https://doi.org/10.1093/mnras/stt2375

[19] L. Spinoglio, K. M. Dasyra, A. Franceschini, C. Gruppioni, E. Valiante, and K. Isaak, “Far-IR/submillimeter spectroscopic cosmological surveys: Predictions of infrared line luminosity functions for z< 4 galaxies,” *The Astrophysical Journal*, vol. 745, no. 2, p. 171, jan 2012. [Online]. Available: https://doi.org/10.1088/0004-637x/745/2/171

[20] A. Endo, “Superconducting on-chip spectrometery for millimeter-submillimeter wave astronomy,” *IEICE Transactions on Electronics*, vol. 98, no. 3, pp. 219–226, 2015.

[21] B. Sibthorpe and W. Jellema, “Relative performance of dispersive and non-dispersive far-infrared spectrometer instrument architectures,” in *Modeling, Systems Engineering, and Project Management for Astronomy VI*, vol. 9150. International Society for Optics and Photonics, 2014, p. 91500J.

[22] H. Ezawa, R. Kawabe, K. Kohno, and S. Yamamoto, “The Atacama Submillimeter Telescope Experiment (ASTE),” in *Ground-based Telescopes*, vol. 5489, International Society for Optics and Photonics. SPIE, 2004, pp. 763 – 772. [Online]. Available: https://doi.org/10.1117/12.551391

[23] “ASTE instrumentation overview,” 2014, last visited February 22, 2021. [Online]. Available: https://alma.mtk.nao.ac.jp/aste/instruments/telescope.html

[24] E. Huijten, “TiEMPO:time-dependent end-to-end model for postprocess optimization of the deshima spectrometer,” B.Sc. Thesis, Delft University of Technology, 2020.

[25] Y. Roelvink, “Simulation of a high-redshift line-emitting galaxy detection with deshima using TiEMPO,” B.Sc. Thesis, Delft University of Technology, 2020.

[26] J. Baselmans, “Kinetic inductance detectors,” *Journal of Low Temperature Physics*, vol. 167, no. 3, pp. 292–304, 2012.

[27] T. Takekoshi, K. Karatsu, J. Suzuki, Y. Tamura, T. Oshima, A. Taniguchi, S. Asayama, T. J. Bakx, J. J. Baselmans, S. Bosma *et al.*, “DESHIMA on ASTE: On-sky responsivity calibration of the integrated superconducting spectrometer,” *Journal of Low Temperature Physics*, pp. 1–9, 2020.

[28] K. Marthi, “Modelling kinetic inductance detectors and associated noise sources,” Industrial internship project report, University of Groningen, 2020.

[29] G. B. Rybicki and A. P. Lightman, *Radiative processes in astrophysics*. John Wiley & Sons, 2008.

[30] A. Otarola, C. De Breuck, T. Travouillon, S. Matsushita, L.-Å. Nyman, A. Wootten, S. J. Radford, M. Sarazin, F. Kerber, and J. P. Pérez-Beaupuits, "Precipitable water vapor, temperature, and wind statistics at sites suitable for mm and submm wavelength astronomy in northern chile," *Publications of the Astronomical Society of the Pacific*, vol. 131, no. 998, p. 045001, 2019.

[31] Y. S. Sukul, "Principal component analysis on atmospheric noise measured with an integrated superconducting spectrometer," B.Sc. Thesis, Delft University of Technology, 2019.

[32] S. Matsushita, H. Matsuo, J. R. Pardo, and S. J. Radford, "FTS measurements of submillimeter-wave atmospheric opacity at Pampa la Bola II: supra-terahertz windows and model fitting," *Publications of the Astronomical Society of Japan*, vol. 51, no. 5, pp. 603–610, 1999.

[33] J. T. Dempsey, P. Friberg, T. Jenness, R. P. J. Tilanus, H. S. Thomas, W. S. Holland, D. Bintley, D. S. Berry, E. L. Chapin, A. Chrysostomou, G. R. Davis, A. G. Gibb, H. Parsons, and E. I. Robson, "SCUBA-2: on-sky calibration using submillimetre standard sources," *Monthly Notices of the Royal Astronomical Society*, vol. 430, no. 4, pp. 2534–2544, 03 2013. [Online]. Available: https://doi.org/10.1093/mnras/stt090

[34] F. Cortés, R. Reeves, and R. Bustos, "Analysis of the distribution of precipitable water vapor in the chajnantor area," *Radio Science*, vol. 51, no. 7, pp. 1166–1175, 2016. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RS005929

[35] J. R. Pardo, J. Cernicharo, and E. Serabyn, "Atmospheric transmission at microwaves (atm): an improved model for millimeter/submillimeter applications," *IEEE Transactions on Antennas and Propagation*, vol. 49, no. 12, pp. 1683–1694, 2001.

[36] R. Giovanelli, J. Darling, C. Henderson, W. Hoffman, D. Barry, J. Cordes, S. Eikenberry, G. Gull, L. Keller, J. Smith *et al.*, "The optical/infrared astronomical quality of high atacama sites. II. infrared characteristics," *Publications of the Astronomical Society of the Pacific*, vol. 113, no. 785, p. 803, 2001.

[37] O. Noroozian, J. Gao, J. Zmuidzinas, H. G. LeDuc, and B. A. Mazin, "Two-level system noise reduction for microwave kinetic inductance detectors," in *AIP Conference Proceedings*, vol. 1185, no. 1. American Institute of Physics, 2009, pp. 148–151.

[38] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[39] A. J. van der Veen, J. Romme, and Y. Cui, "Rank detection thresholds for hankel or toeplitz data matrices," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1911–1915.

[40] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[41] K. Slavakis, G. B. Giannakis, and G. Mateos, "Modeling and optimization for big data analytics: (statistical) learning tools for our era of data deluge," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 18–31, 2014.

[42] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Nonnegative matrix and tensor factorizations: An algorithmic perspective," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 54–65, 2014.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[44] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[45] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[46] A. P. Laguna, K. Karatsu, D. J. Thoen, V. Murugesan, A. Endo, and J. J. A. Baselmans, “Wideband superconducting integrated filter-bank for THz astronomy,” in *European Conference on Antennas and Propagation*, 2020.

[47] Y. Asaki, M. Saito, R. Kawabe, K.-i. Morita, Y. Tamura, and B. Vila-Vilaro, “Alma memo no. 535 simulation series of a phase calibration scheme with water vapor radiometers for the atacama compact array,” 2005.

[48] Y. Asaki, H. Sudou, Y. Kono, A. Doi, R. Dodson, N. Pradel, Y. Murata, N. Mochizuki, P. G. Edwards, T. Sasao *et al.*, “Verification of the effectiveness of VSOP-2 phase referencing with a newly developed simulation tool, aris,” *Publications of the Astronomical Society of Japan*, vol. 59, no. 2, pp. 397–418, 2007.

[49] M. Bester, W. C. Danchi, C. G. Degiacomi, L. J. Greenhill, and C. H. Townes, “Atmospheric fluctuations-empirical structure functions and projected performance of future instruments,” *The Astrophysical Journal*, vol. 392, pp. 357–374, 1992.