Characterization of CHIME′s Complex Gain Using New Transits of CygA and CasA

by

Emilie Marie Hertig

A HONOURS THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Bachelor of Science

in

THE FACULTY OF SCIENCE
(Physics and Astronomy)

The University of British Columbia
(Vancouver)

April 2019

Abstract

This work aims to improve the calibration of the Canadian Hydrogen Intensity Mapping Experiment (CHIME), a new radiotelescope built in Penticton (BC) and inaugurated in 2017. CHIME's main goal is to observe baryon acoustic oscillations in order to probe the evolution of the universe at redshifts between 0.8 and 2.5, in the period where the standard model of cosmology predicts that dark energy started to dominate over matter and radiation. Accurate expansion measurements during this period would provide tighter constraints on the Hubble parameter and the dark energy equation of state. This would lead to invalidation or further verification of our current cosmological theories, therefore improving our understanding of the nature of dark energy.

It has been observed that the the steel structures of the telescope, as well as the cables and antennae have temperature-dependent behaviour, which affects the quality of the data. Previous attempts of characterizing this dependency weren't successful; therefore, this works aims to explore an innovative method based on new direct sky observations and allowing to determine the influence of external temperature on CHIME's complex gain. Thermal susceptibilities are obtained from analyzing observations of bright radiosources, mainly CygA and CasA, and calculating linear fits of the gain fractional variation as a function of external temperature. A correction for the nonlinear behaviour of antennae, based on laboratory experiments, is included in later stages of the analysis, as well as a detailed investigation and treatment of outliers.

This project is a step towards the making of a full thermal model that could be included in CHIME's calibration algorithm in order to significantly improve the quality of cosmological data.

Table of Contents

List of Figures

Glossary

CHIME Canadian Hydrogen Intensity Mapping Experiment

DRAO Dominion Radio Astrophysical Observatory

CMB Cosmic Microwave Background

FRB Fast Radio Bursts

BAO Baryon Acoustic Oscillations

SDSS Sloan Digital Sky Survey

WMAP Wilkinson Microwave Anisotropy Probe

LNA Low Noise Amplifier

RFI Radio-Frequency Interference

SVD Singular Value Decomposition

FLA Filter Amplifier

RMS Root Mean Square

Acknowledgments

I would like to extend my heartfelt thanks to my supervisor Dr. Gary Hinshaw for giving me the opportunity to be a part of his team and contribute to this fascinating project, as well as for his advice and guidance throughout the process of research. A special thanks goes to Mateus Fandino for his precious help, his clear explanations, his kindness and patience. Many thanks also to Dr. Seth Siegel for storing and pre-processing the gain data. Finally, I am very grateful to the whole CHIME Collaboration for the enjoyable research experience and the friendly atmosphere in the experimental cosmology lab.

Chapter 1

Introduction

1.1 Motivation

Since its first observation in 1998, the accelerated expansion of the universe has remained an unsolved mystery and a flourishing domain of investigation for modern cosmologists. While an intuitive view of the Big Bang model suggests that the initial expansion should be slowing down under the effect of gravity, observations yielded by a type Ia supernovae survey conducted in 1998 [15] and numerous subsequent experiments show the exact opposite: an unknown force seems to be pushing galaxies away from each other at an increasing rate. The mysterious phenomenon, due to its puzzling nature and its apparent lack of directly observable properties, has been given the name of "dark energy"; understanding its physical origin has become one of the greatest challenges of today's cosmology.

The most widely accepted cosmological theory at present corresponds to the ΛCDM model, where Λ stands for a cosmological constant representing dark energy, and CDM refers to "cold dark matter". The adjective "cold", in this context, means that dark matter is seen as non-baryonic particules moving at sub-relativistic velocities. In this model, the initial stage of the universe was characterized by extremely high density and temperature; the emergence of expanding space and time, commonly known as "Big Bang", was triggered 13.79 billion years ago [14] by quantum fluctuations of yet unknown nature. Within the first few fractions of a second, a very brief stage of extreme exponential expansion known as inflation is

believed to have taken place. Such a sudden and violent event (supposedly increasing the size of the universe by a factor of 10^{26} in the span of 10^{32} second [19]) can seem difficult to conceive, but this model actually solves some fine-tuning issues arising from Big Bang cosmology and provides an explanation for Cosmic Microwave Background (CMB) anisotropies [16]. Inflation was followed by a long period of decelerating expansion during which the universe was successively dominated by radiation and matter. The present acceleration phase suggests that the dominating component is now dark energy, the proportion of which has been estimated to 68.89% by the latest Planck results [14]. Representing it by a cosmological constant Λ indicates that the dark energy density is assumed to remain constant in time. This means that dark energy doesn't "dilute" with expansion like matter and radiation; while this would justify why it only becomes dominant in later stages of the universe's evolution, it still doesn't explain what dark energy actually is. Many hypotheses have been made, none of them has been confirmed yet. One of the candidates is the vacuum energy predicted by quantum field theory [9], which would be consistent with a cosmological constant; however, the predicted energy density exceeds the measured cosmological constant by no less than 120 orders of magnitude. Several modern cosmologists are also questioning the fundamental assumptions of ΛCDM and proposing alternative theories such as a time-dependant dark energy density [20], modified gravity [17] or scale invariance of vacuum [11].

Most of the observational evidence for the ΛCDM model is contained in the CMB, which constitutes the relic of a much hotter and denser state of the universe. The near-perfect correlation between the CMB spectrum and the predicted blackbody at 2.7255 K definitely tipped opinions in favor of Big Bang theory; since then, ΛCDM has been generally accepted as today's leading cosmological model. In its simplest form, ΛCDM only requires the knowledge of 6 cosmological parameters in order to reconstitute the past evolution of the universe and predict its future. These parameters have to be constrained by experiments, most of them consisting in measuring the relation between distance and redshift of various remote celestial objects. Such measurements are the main goal of the Canadian Hydrogen Intensity Mapping Experiment (CHIME) telescope. Numerous surveys with a similar purpose have already been conducted, mostly focusing on the very early universe (e.g. Wilkinson Microwave Anisotropy Probe (WMAP) [10] and Planck [14]) or the

closer neighbourhood of our galaxy (e.g. Sloan Digital Sky Survey (SDSS) [18]). CHIME's wide frequency range and use of intensity mapping will allow for observations at redshifts $0.8 < z < 2.5$, an intermediate range of major importance as it is believed to enclose the moment at which dark energy became dominant. Measurements of the Hubble parameter (see Chapter 2) and of the dark energy equation of state deduced from CHIME's 3D map of this zone could be the key to a better understanding of the nature of dark energy, by determining if it really is a cosmological constant. These results might allow to invalidate or further verify ΛCDM, and maybe even to discover a relation between the current accelerating expansion and the initial inflation stage.

Additionally, CHIME is already used for Fast Radio Bursts (FRB) detection [6] (for which its wide field of view is a significant advantage) and pulsar monitoring [12]. Two FRBs have already been observed [4], and many more are expected to be detected in the near future. While this is a very encouraging result, my work will be focused on the cosmological purpose of the instrument, and I will therefore not go into more detail on the topic of FRBs.

With foreground signals 10^3 to 10^5 times brighter than the actual cosmological sources [5], CHIME requires a very precise calibration. Previously conducted laboratory experiments as well as studies of the CHIME Pathfinder have indicated that the telescope's components present a temperature-dependent response which affects the quality of the data. This paper will focus on the effect of external temperature on the amplitude of the complex gain, corresponding to the direction-independant part of the total gain as generated by amplifiers and cables. Thermal expansion of cables and steel structures produces variations of the complex gain which need to be analyzed and taken into account in the calibration algorithm. Previous attempts of characterizing CHIME's thermal susceptibility consisted in generating a controlled artificial signal and correlating the telescope's response to the temperature at the time of measurement. This Radio-Frequency Interference (RFI) producing method had the disadvantage of perturbing other instruments at the Dominion Radio Astrophysical Observatory (DRAO), and had to stop being used before it could yield satisfactory results. This work therefore aims to explore an alternative method in which the artificial signal is replaced by direct sky data. The new archive of stable observations of CygA, CasA and more recently also

TauA will constitute the core of the present analysis; these bright and well-known radiosources will be used to deduce CHIME's complex gain for each frequency bin and channel over a few months' period. Variations of this complex gain will then be analyzed in the hopes of determining CHIME's thermal susceptibility and improving its calibration algorithm.

1.2 Overview

This paper will begin with a theoretical section summarizing the ΛCDM model's fundamental concepts and equations. Techniques used to constrain cosmological parameters will be introduced, with a focus on Baryon Acoustic Oscillations (BAO) measurements. CHIME's design and working principle will then be described.

The chapter on experimental methods will cover the various steps of gain extraction and data pre-processing, before describing the different stages of the development of the thermal model. Details will be provided on the fitting techniques and statistical concepts used to obtain the results.

All the important results will then be presented and commented, in a logical order following the various stages of investigation. The significance of these results will be evaluated and opportunities for further investigation will be discussed.

Chapter 2

Theoretical concepts

2.1 Notions of cosmology

This section will provide some conceptual and mathematical details about the ΛCDM model and CHIME's scientific purpose.

2.1.1 ΛCDM model

The fundamental framework for modern cosmology is given by Einstein's general relativity; the theory describes a four-dimensional spacetime whose curvature is influenced by matter-energy distribution, and where the trajectories of all particles (including photons) correspond to geodesics. The separation between two events in spacetime is determined by a metric; at cosmological scales, where the universe can be assumed to be isotropic and homogeneous, the Robertson-Friedmann-Walker metric is commonly used. Its mathematical expression is as follows [16]:

$$ds^2 = -c^2dt^2 + a(t)^2[dr^2 + S_\kappa(r)^2 d\Omega^2] \tag{2.1}$$

In this equation, expressed in spherical coordinates, $S_\kappa(r)$ is a curvature-dependent factor defined in Equation 2.2 and $d\Omega^2 = d\theta^2 + \sin(\theta)^2 d\phi^2$. R_0 represents the radius of curvature, and the curvature scalar κ equals 1 for positive curvature, 0 for flat space and -1 for negative curvature.

$$S_\kappa(r) = \begin{cases} R_0 \sin(r/R_0) & (\kappa = 1) \\ r & (\kappa = 0) \\ R_0 \sin(r/R_0) & (\kappa = -1) \end{cases} \tag{2.2}$$

The $a(t)$ factor appearing in Equation 2.1 is called "scale factor", and is of fundamental importance as it characterizes the expansion of the universe. It is defined to be equal to 1 at present time. The coordinates (r, θ, ϕ) in Equation 2.1 are referred to as "comoving coordinates"; two points having constant comoving coordinates implies the absence of proper motion relative to each other, even though the separation between the two objects increases due to the expansion of spacetime. At large scales, expansion can be assumed to be isotropic and the proper distance between two objects can be written $d_p(t) = a(t)r$ with r the comoving distance. In this case, Equation 2.3 holds.

$$\dot{d}_p(t) = \dot{a}(t)r = \frac{\dot{a}(t)}{a(t)} d_p(t) \tag{2.3}$$

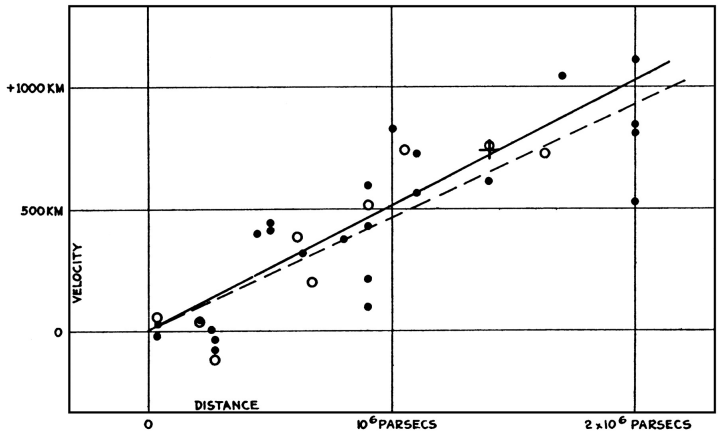

Figure 2.1: Edwin Hubble's original diagram showing the linear relation between distance and recession speed of neighboring galaxies. The distances were underestimated at the time, leading to an excessive value for H_0. Source: https://astro.unl.edu/naap/distance/graphics/hubble_orig.png

The proportionality constant $H(t) = \frac{\dot{a}(t)}{a(t)}$ is called "Hubble parameter", referring to Edwin Hubble's first observation of the expansion of the universe in 1929 (Figure 2.1). Equation 2.3, evaluated at present time $t = t_0$, is known as Hubble's law.

While Hubble's law expresses the fact that the universe is expanding (i.e. $\dot{a}(t) > 0$), it doesn't tell us anything about the sign of $\ddot{a}(t)$. Evidence of accelerating expansion was only discovered decades later, in 1998 [15], when a survey conducted by the High-Z Supernova Search Team yielded larger luminosity distances to various Type Ia Supernovae than would be observed in the absence of a positive acceleration. Since then, this result has been confirmed by multiple experiments and it is now generally accepted that $\ddot{a}(t) > 0$, despite the lack of satisfactory physical explanation for this phenomenon.

The evolution of $a(t)$ as a function of cosmic time can be computed using Friedmann's equation. Derived from Einstein's field equations, this equality relates the Hubble parameter to the energy distribution and the curvature of the universe. Its mathematical form is as follows, with G Newton's gravitational constant, ε the total energy density (summed over all components of the universe), κ the curvature scalar and Λ the cosmological constant:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\varepsilon - \frac{\kappa c^2}{R_0^2 a^2} + \frac{\Lambda}{3} \tag{2.4}$$

The other equation involving a derivative of $a(t)$ is the acceleration equation, where P represents the pressure:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\varepsilon + 3P) + \frac{\Lambda}{3} \tag{2.5}$$

Two important observations can be drawn from Equation 2.4 and Equation 2.5. First, the only way to produce accelerating expansion is to include a component with negative pressure in Equation 2.5. More precisely, we can define the equation of state as the relation between energy density and pressure for a given substance:

$$P = w\varepsilon \tag{2.6}$$

Then a component producing positive acceleration must have $w < -\frac{1}{3}$. The

second observation is that, neglecting Λ in Equation 2.4, there exists a critical energy density for which spacetime is flat ($\kappa = 0$). This value is determined by Equation 2.7.

$$\varepsilon_{crit}(t) = \frac{3c^2}{8\pi G} H(t)^2 \tag{2.7}$$

The time evolution of the energy density for a component with equation of state w is ruled by the fluid equation (Equation 2.8), derived from the first law of thermodynamics.

$$\dot{\varepsilon} + 3\frac{\dot{a}}{a}(\varepsilon + P) = 0 \tag{2.8}$$

Substituting Equation 2.6 into Equation 2.8 and solving the resulting differential equation for ε, the following general expression can be obtained:

$$\varepsilon(a) = \varepsilon_0 a^{-3(1+w)} \tag{2.9}$$

From Equation 2.7, we can define a density parameter $\Omega = \frac{\varepsilon}{\varepsilon_{crit}}$ for each component of the universe; written with a subscript 0, Ω_0 refers to the density parameter at present time. The latest Planck results [14] yield a value very close to 0 for the curvature of our universe; therefore, ΛCDM assumes spacetime to be flat, which means the total energy density is equal to the critical density ($1 - \Omega_m - \Omega_r - \Omega_\Lambda = 0$). The Friedmann equation can now be rewritten as a function of the density parameters at present time and equations of state of baryonic matter ($w = 0$), radiation ($w = 1/3$) and dark energy (according to ΛCDM, $w = -1$):

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{r,0} a^{-4} + \Omega_{m,0} a^{-3} + \Omega_\Lambda \tag{2.10}$$

Equation 2.10 can be expressed as a function of an observable quantity, redshift (z), thanks to the following relation:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{1}{a} - 1 \tag{2.11}$$

Therefore, the scale factor at the time of emission is related to the redshift of the observed object by the simple formula $a^{-1} = 1 + z$. Friedmann's equation now

takes the form:

$$\left(\frac{H(z)}{H_0}\right)^2 = \Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_\Lambda \tag{2.12}$$

From Equation 2.12, it appears that measurements of $H(z)$ can be used to yield constraints on cosmological parameters. Using the values obtained by previous experiments such as WMAP [10] or Planck [14] for H_0 and the density parameters, the theoretical evolution of $H(z)$ can be computed and compared to measurements; any observed deviations could for example invalidate the cosmological constant assumption and yield new constraints on the dark energy equation of state. This is the main objective of the CHIME mission: by measuring the distance-redshift relation for standard rulers at $0.8 < z < 2.5$, the radiotelescope will be able to retrace the expansion history of the universe during this crucial period and distinguish possible disagreements with the predictions of ΛCDM. This new study of the dark energy equation of state might enable us to confirm or rule out some of our current hypotheses, and could therefore be a step towards a better understanding of the physical nature of one of the universe's biggest mysteries.

2.1.2 Constraining cosmological parameters

Measuring the distance-redshift relation mentioned in Section 2.1.1 is not as easy as it seems; indeed, it is important to realize that the proper distance, defined as the length of the spatial geodesic between two events in spacetime, is not an observable quantity. The mathematical expression of the proper distance at present time is given by Equation 2.13, where t_e is the emission time:

$$d_p(t_0) = c\int_{t_e}^{t_0} \frac{dt}{a(t)} \tag{2.13}$$

From this definition, it is clear that the time evolution of the proper distance to an object directly depends on the time evolution of the scale factor $a(t)$, which is determined by the parameters of the model used to describe the universe. This explains why distance measurements are so widely employed to constrain cosmological parameters and confirm or invalidate possible cosmological models.

The two observable quantities related to proper distance are luminosity distance

and angular diameter distance. To determine luminosity distance, one needs to observe the energy flux F received from a source of known luminosity L. Luminosity distance is then given by Equation 2.14.

$$d_L = \sqrt{\frac{L}{4\pi F}} \tag{2.14}$$

Luminosity distance is related to proper distance by a simple equation involving another observable quantity, reshift (z):

$$d_L = (1+z)d_p \tag{2.15}$$

Experiments aiming to measure luminosity distances in order to constrain cosmological parameters are conducted by observing bright sources of known luminosity, called "standard candles". The first discovered standard candles were Cepheid variables: a relation was established between the period of variation and the luminosity of these very bright stars, which allows astronomers to deduce the intrinsec luminosity by analyzing the variation pattern. The obtained value is then compared to the received flux and Equation 2.14 is used to deduce the luminosity distance. Due to their supergiant nature and intense brightness, Cepheids can yield measurements of intergalactic distances; however, their range is limited to about 30 Mpc [16]. Other methods are required in order to explore the universe at high redshifts.

Another famous category of standard candles includes type Ia supernovae. These stellar explosions occur when a white dwarf orbiting around a companion star attracts enough matter to exceed the Chandrasekhar mass and collapse under the effect of its self-gravity. It has been discovered that the light curve of such a supernova is related to its peak brightness. Therefore, by observing the time taken by the supernova to reach its peak and to fade away, it is possible to deduce the maximal luminosity and to use Equation 2.14 again in order to determine the luminosity distance. As supernovae are extremely violent events, their brightness can exceed that of an entire galaxy, which makes them visible from very far away; SN Ia observations have been conducted up to $z \approx 1$. At such redshifts, the observed distances present a significant deviation when compared to those computed for a decelerating or steady expansion. Therefore, type Ia supernovae measurements re-

sulted in the first experimental evidence for dark energy and are also used to place constraints on the Hubble parameter and on the energy densities of matter and of the cosmological constant.

While SN Ia surveys have been very successful over the years, our growing understanding of the early universe and of the mechanisms of gravitational instability and structure formation have led us to explore a new class of objects, known as "standard rulers". Standard rulers correspond to great structures of known, constant comoving size. Measuring the apparent diameter of such objects allows to determine the angular diameter distance d_A, related to proper distance by Equation 2.16:

$$d_A = \frac{d_p}{1+z} \tag{2.16}$$

In Section 2.1.3, we will see how CHIME uses a specific type of standard rulers, Baryon Acoustic Oscillations (BAO), in order to retrace the expansion of the universe at relatively high redshifts and constrain the dark energy equation of state.

Finally, it is important to note that constraints on cosmological parameters can also be deduced from observations of the very early universe, especially the CMB. The state of the universe at the time of last scattering contains a mine of information that can be used to analyze the underlying physics and test our current models. The most accurate constraints are obtained by combining CMB and SN Ia or BAO measurements.

2.1.3 BAO as standard rulers

During the first few fractions of a second of the universe's history, when spacetime had just been formed and matter hadn't been created yet, quantum fluctuations occured and were exponentially amplified by inflation. These perturbations remained after inflation, as potential wells to which matter (both dark and baryonic) was gravitationnally attracted, producing slight density anisotropies in the very early cosmos; these initially infinitesimal anomalies would later become the source of the great structures (galaxy clusters and superclusters) observed in the present-day universe. Before recombination, photons and hot, ionized baryonic matter constantly interacted through Thomson scattering. As this photon-baryon

plasma was attracted to the aforementioned potential wells, pressure would build up until a state of maximum compression was reached. The plasma would then "bounce back" from this contracted state, and the spherical density perturbations would propagate outwards as acoustic waves. Once the temperature of the universe had decreased enough for hydrogen atoms to form and matter to become electrically neutral, photons decoupled from baryons; the absence of radiation pressure caused these waves to stop propagating and leave their mark as concentric hot (overdense) and cold (underdense) rings on the CMB (Figure 2.2). Since then, these rings (which are actually the projection of three-dimensional spherical shells on the two-dimensional last scattering surface) have kept a constant comoving size, and their maximal radius, called "sound horizon", has been calibrated from CMB measurements, yielding a value of 146.8 ± 1.6 Mpc [2].

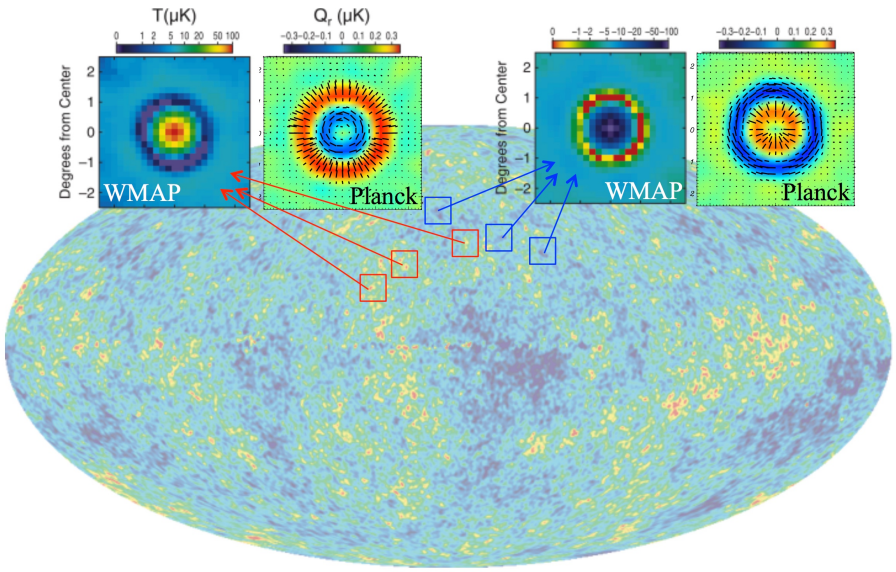

Figure 2.2: Zoom on hot and cold spots and rings observed in the CMB. The WMAP data shows temperature anisotropies, while the Planck results focus on polarization. Inverse temperature and polarization patterns are observed in both cases. These anisotropies constitute evidence of the existence of BAO and have allowed to calibrate the comoving size of the sound horizon. Image credit: CHIME Collaboration.

As more and more dark and baryonic matter was attracted to the primordial

overdense shells, gravitational instability progressively led to the formation of stars, galaxies, clusters and superclusters. Billions of years after the emission of the CMB photons, the imprint of BAO is still detectable in the great structures of the universe, as a preferred scale for galaxy clustering resulting in a peak in the correlation function of galaxy distributions. This feature was first observed by Eisenstein *et al.* in 2005 [8] and is shown in Figure 2.3.

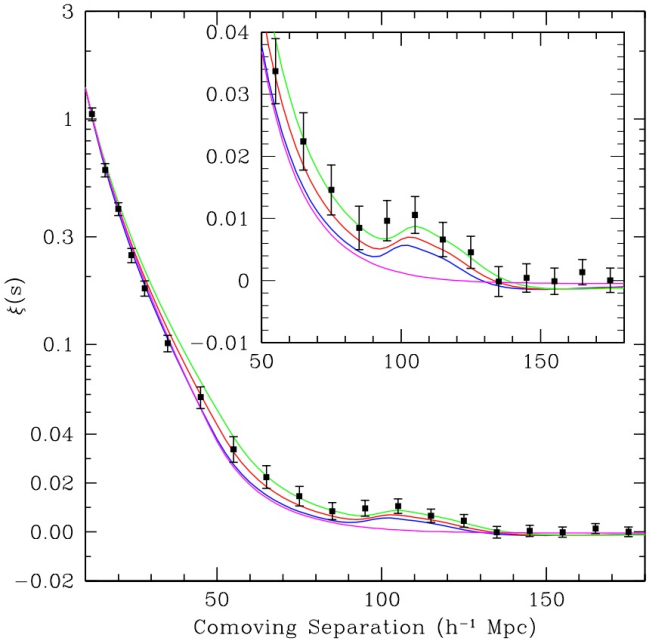

Figure 2.3: Correlation function of the galaxy distribution obtained from the Sloan Digital Sky Survey (SDSS). The peak at $d \approx 100h^{-1}$ Mpc illustrates the higher correlation between galaxies separated by a comoving distance d. With $h = H_0/100$ and H_0 the Hubble constant (67.66 km s^{-1} Mpc^{-1}), the observed preferred scale is in good agreement with the sound horizon determined from CMB measurements. Image credit: Eisenstein et al. [8].

BAO structures having a constant comoving radius means that they can be used as a reference distance. As they can only be recovered from sky maps using correlation functions, they belong to the category of statistical standard rulers.

Once a BAO structure has been detected at a certain redshift z, observations

across and along the line of sight allow the computation of the angular diameter distance $d_A(z)$ and the Hubble parameter $H(z)$ respectively. With s the diameter of the sound horizon, $\Delta\theta$ the measured angular diameter across the sky and Δz the redshift difference between both extremities of the structure, the relevant formulae [2] are given by Equation 2.17 and Equation 2.18.

$$d_A(z) = \frac{s}{\Delta\theta(1+z)} \tag{2.17}$$

$$H(z) = \frac{c\Delta z}{s} \tag{2.18}$$

Therefore, using BAO as standard rulers has the significant advantage of providing measurements of two important quantities at once, and offers the possibility of cross-checking the obtained results using Equation 2.19.

$$d_A(z) \propto \int_0^z \frac{dz'}{H(z')} \tag{2.19}$$

Consequently, for the same quality of data, BAO measurements can provide tighter constraints on the dark energy equation of state than type Ia supernovae observations, for which part of the physical mechanism is very complex and still not fully understood [13], and which only provide measurements of $d_L(z)$. Determining $d_A(z)$ and $H(z)$ simultaneously is more powerful than only knowing one of these quantities. The simulation in Figure 2.4 illustrates how combining BAO data with CMB observations by Planck can significantly improve constraints on the dark energy equation of state, and yield similar results to much more expensive surveys proposed for the next few years (EUCLID and BigBOSS).

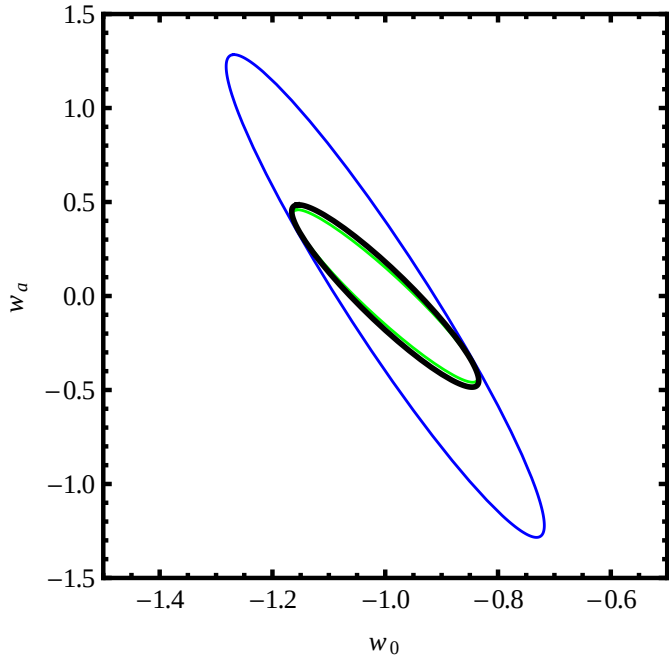

Figure 2.4: Comparison of simulated constraints from Planck + DEFT Stage II (blue), PLanck + DEFT Stage II + EUCLID/BigBOSS (green) and PLanck + DEFT Stage II + CHIME (black). Combining CMB and BAO results in the tightest constraints on the dark energy parameters w_0 and w_a, with the model assuming $w = w_0 + w_a \frac{z}{1+z}$. ΛCDM corresponds to $w_0 = -1$ and $w_a = 0$. Image credit: CHIME Collaboration.

2.2 The CHIME telescope

2.2.1 Interferometry

CHIME is a ground-based transit interferometer, which has no moving parts but uses the Earth rotation in order to map half of the sky every day. One of the main principles behind the obtention of these maps is interferometry, which consists in combining the signals received by separate antennae in a way that preserves the phase data, and analyzing the interference pattern in order to reconstitute the full sky information. Increasing the number of antennae results in a more accurate reconstruction of the signal.

In the case of CHIME, four adjacent half-cylinders oriented North-South each reflect the incoming light towards their respective focal line. Along each focal line are placed 256 dual-polarization feed antennae. The resolution requirements for intensity mapping (see Section 2.2.2) are met in the East-West direction by the focusing effect of the cylindrical dishes, and in the North-South direction by the instantenous correlation of the signals received by each pair of feeds.

2.2.2 Intensity mapping

The technique of intensity mapping consists of detecting the 21 cm spin-flip emission line of neutral hydrogen. The transition from triplet to singlet state in the ground state of the hydrogen atom is extremely rare, but the resulting emission remains detectable at cosmological scales due to the abundance of neutral hydrogen in regions such as damped Lyman alpha systems, mostly situated inside galaxies (often quasars). The intensity of the emission is related to the density of matter; detecting this signal can therefore allow to reconstitute the power spectrum of matter distribution in the zone of interest.

The significant advantage of intensity mapping is that it does not require resolving individual galaxies, as denser regions simply appear as brightness temperature fluctuations on the map. Consequently, the use of this technique allows CHIME to have a much lower resolution than instruments dedicated to galactic surveys such as SDSS, from which the first evidence of BAO was obtained. An comparative example is shown in Figure 2.5. This greatly increases the mapping speed, and is the reason why CHIME's maps will be able to encompass a wide redshift range and a much larger volume of sky than previous surveys.

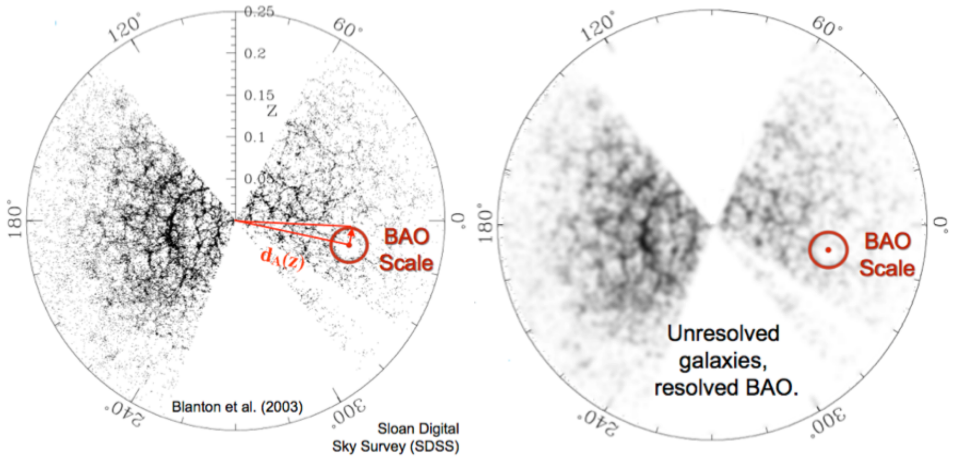

Figure 2.5: Left panel: map of a slice of sky up to $z = 0.25$ obtained by the Sloan Digital Sky Survey. Each dot is an individual galaxy. Observing every single galaxy for a significant amount of time results in a low mapping speed. Right panel: the same map with a resolution similar to that of CHIME. No individual galaxies are visible, but density fluctuations appear as intensity variations and the BAO scale is resolved. Image credit: [3] and CHIME Collaboration.

2.2.3 CHIME's working principle

As mentioned in Section 2.2.1, CHIME consists of four adjacent parabolic cylinders made of steel, whose role is to reflect and focus incoming light towards the North-South oriented focal lines. Figure 2.6 shows a photograph of the external structure of the telescope. After being reflected by the cylinders, light is captured by the 1024 dual-polarization antennae distributed between the four focal lines. These cloverleaf shaped antennae [7] are sensitive to both linear polarizations, East-West and North-South, between 400 and 800 MHz. This range was chosen in order to match the frequency of the redshifted 21 cm emission line between $z = 0.8$ and $z = 2.5$. One of CHIME's feeds is pictured in Figure 2.7. Directly after being captured by an antenna, the signal goes through a Low Noise Amplifier (LNA), and is transported by a 50 m coaxial cable to the Filter Amplifier (FLA). These different steps before digitization correspond to the analog signal chain, which is schematized in Figure 2.8.

Figure 2.6: Photograph of the CHIME instrument taken at the DRAO. The metallic structures of the four cylinders and focal lines are visible, as well as the containers enclosing the X-engine. The smaller dish antennae in the background belong to a different instrument. Image credit: CHIME Collaboration [1].

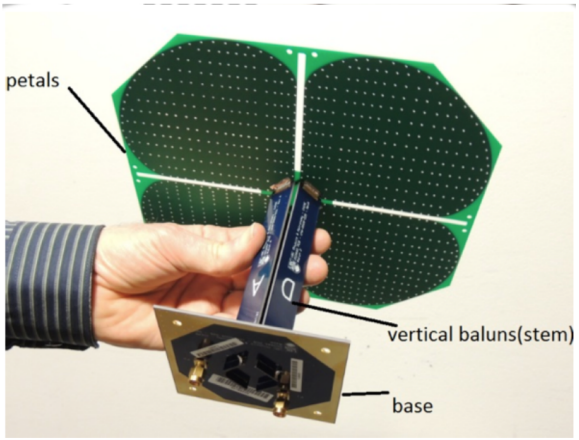

Figure 2.7: Photograph of one of CHIME's cloverleaf shaped dual-polarization antennae. Image credit: CHIME Collaboration.

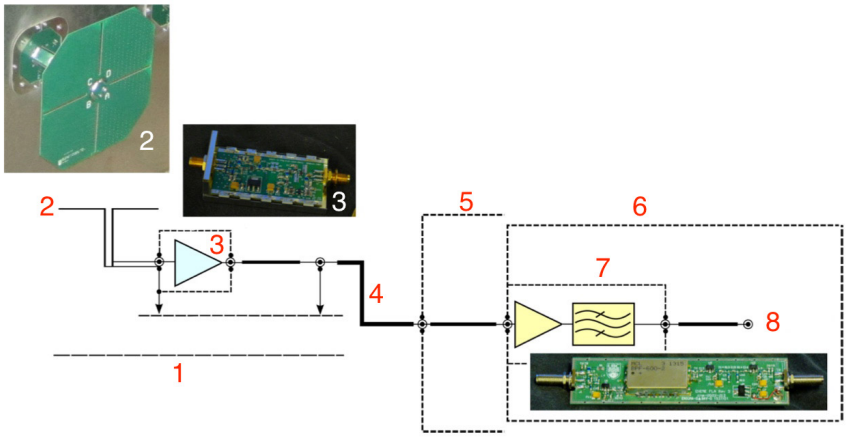

Figure 2.8: Diagram of CHIME's analog signal chain. (1) Reflector. (2) Antenna. (3) LNA. (4) 50 m coaxial cable. (5) Outer bulkhead. (6) Shielded room. (7) FLA. (8) A/D converter input. Image credit: CHIME Collaboration.

The analog chain is the source of CHIME's thermal susceptibility. Thermal expansion of the steel structures of the reflectors and the 50 m cable are the most dominant factors, but the temperature-dependant response of the LNAs has also been confirmed and analyzed in the lab.

After being digitized, the signal arrives to the F-engine which performs a fast Fourier transform in real time, with an input data rate of 13 terabits/s, in order to convert data from position space to frequency space, divided in 1024 frequency bins. Finally, the signal is sent to the X-engine, which computes the visibility matrix (see Chapter 3). From the X-engine, data is transferred to three different backends, each characterized by different sampling rates and analysis methods: the FRB search engine, the pulsar timing monitor and the cosmology backend.

Table 2.1: Summary of CHIME's main characteristics [1].

No. of cylinders	4
Dimensions of a cylinder	20 m $\times$ 100 m
No. of feeds	256/cylinder
No. of channels	2048 (4 cyl., 256 feeds each, 2 pol.)
Redshift range	0.8-2.5
Frequency range	400-800 MHz
No. of frequency bins	1024
Frequency resolution	0.39 MHz
Angular resolution	15'-25'
Spatial resolution	15 Mpc
Instantaneous FOV	200 sq. deg.

Table 2.1 summarizes the main characteristics of the instrument. The angular resolution has been adjusted to correspond to 1/10 of the BAO scale at the redshifts of interest. The large instantaneous field of view combined with the Earth rotation allows the telescope to map half of the celestial sphere each day.

Chapter 3

Methods

3.1 Gain calibration

The signals detected by CHIME arrive in the form of electromagnetic waves, composed of oscillating electric and magnetic fields. A general way of mathematically expressing such an electric field is by using a complex number:

$$E = E_0 e^{i\phi} \tag{3.1}$$

If one feed, identified by the subscript i, receives a signal described by Equation 3.1, another feed j separated from the first one by a vector $\vec{b}_{ij}$ will receive the signal $E_0 e^{i\left(\phi + 2\pi \frac{\hat{n}\cdot\vec{b}_{ij}}{\lambda}\right)}$. $\hat{n}$ corresponds to the unit vector in the direction of the incoming radiation, and λ is the wavelength of the signal.

Cross-correlating these two inputs, each multiplied by the corresponding complex gain g_i introduced by the system, yields the following expression:

$$E_i E_j^* = g_i E_0 e^{i\phi} \left(g_j E_0 e^{i\left(\phi + 2\pi \frac{\hat{n}\cdot\vec{b}_{ij}}{\lambda}\right)} \right)^* = g_i g_j^* |E_0|^2 e^{-2\pi i \frac{\hat{n}\cdot\vec{b}_{ij}}{\lambda}} \tag{3.2}$$

Replacing $\vec{b}_{ij}$ by $\vec{r}_j - \vec{r}_i$, Equation 3.2 becomes:

$$E_i E_j^* = |E_0|^2 g_i e^{i\frac{\hat{n}\cdot\vec{r}_i}{\lambda}2\pi} \left(g_j e^{i\frac{\hat{n}\cdot\vec{r}_j}{\lambda}2\pi} \right)^* \tag{3.3}$$

Equation 3.3 shows that a matrix containing elements $V_{ij} = E_i E_j^*$ for all possible pairs of feeds can be written as the outer product $|E_0|^2 \vec{v}\vec{v}^\dagger$, with $\vec{v}_i = g_i e^{i\frac{\hat{n}\cdot\vec{r}_i}{\lambda}2\pi}$. This implies that this matrix can be expressed as $V = BDB^\dagger$, with B a unitary matrix and D a diagonal matrix containing the eigenvalues of V. In the ideal case described above, only one eigenvalue is nonzero ($|E_0|^2$) and the corresponding eigenvector is equal to $\vec{v}$. Therefore, if the complex gain of the system is known, the power of the received signal can be calculated by performing an eigendecomposition of the matrix V_{ij}. Inversely, if the intensity of the observed source is known, the complex gains can be deduced from the eigenvector corresponding to the nonzero eigenvalue. This property is very useful for the calibration procedure as it allows to retrieve and analyze the system gains based on the observation of bright radiosources.

In the case of CHIME, the situation is slightly more complicated than the simple and ideal example derived here. The output of the correlator, called "visibility matrix", is obtained by cross-correlating signals for each pair of feeds and integrating over every possible direction $\hat{n}$; it also incorporates a noise factor n_{ij}. Mathematically, the visibility matrix elements are given by Equation 3.4.

$$V_{ij} = \int |E_0(\hat{n})|^2 g_i(\lambda) g_j(\lambda)^* e^{2\pi i \frac{\hat{n}\cdot\vec{r}_i}{\lambda}} e^{-2\pi i \frac{\hat{n}\cdot\vec{r}_j}{\lambda}} d^2\hat{n} + n_{ij} \tag{3.4}$$

The previously mentioned eigendecomposition procedure can be generalized to Equation 3.4, therefore CHIME's complex gains are obtained from the eigenvalues and eigenvectors of the visibility matrix.

3.2 Data pre-processing

Once the complex gains have been deduced from the correlator outputs, a few steps of pre-processing have to be performed before the data is ready to be analyzed.

The first of these steps is the removal of digital gains. Indeed, the system gains calculated from the eigendecomposition of Equation 3.4 include the complex gain (the component of interest here, produced by the steel structures and cables), and an additional component generated by the digitization of the data. These digital gains obviously do not depend on external conditions such as temperature, therefore their influence on the analyzed data must be removed. The digital gain values are stored

separately from the system gains in the CHIME database, and a function built in the CHIME-specific Python environment allows for their removal.

The second pre-processing step is the filtering of the data. Transits of the observed radiosources can be contamined by external perturbations such as a simultaneous solar transit or Radio-Frequency Interference (RFI); contamined datasets have to be excluded from the analysis. This step is performed by a "flagging" algorithm, which takes into account a complex combination of external parameters such as weather conditions, solar transits and interference sources, and subsequently evaluates the validity of a given dataset. The full development of the flagging algorithm is not the object of this paper, therefore it will not be presented in more detail; it is however important to note that flagged data is systematically filtered out before any analysis is performed. This is, once again, made possible by the CHIME-specific Python environment used throughout this project.

3.3 Data analysis

3.3.1 Linear fits

The main objective of this work is to determine CHIME's thermal susceptibility; in order to do this, a relation has to be found between the telescope's complex gain and the external temperature at the time of observation. While the complex gain is, as its name indicates, a complex number, this paper will focus on its amplitude. Analyzing the gain phase constitutes a perspective for future investigation.

Expressing the complex gain for channel i and frequency f as $g(i,f) = g_0(i,f)e^{i\phi(i,f)}$, the amplitude then corresponds to $g_0(i,f)$. The absolute value of the gain itself is of little interest in this context; however, relative fluctuations between different transits are of crucial importance. Therefore, we can define the fractional gain variation for a transit at time t:

$$g_{frac}(i,f,t) = \frac{g_0(i,f,t) - \bar{g}_0(i,f)}{\bar{g}_0(i,f)} \tag{3.5}$$

In this equation, $\bar{g}_0(i,f)$ represents the gain amplitude averaged over time for channel i and frequency f.

The fundamental hypothesis of this work is that g_{frac} should follow a linear trend as a function of external temperature. This assumption is supported by the fact that the expansion of the 50 m cable, which is considered to be the main source of CHIME's thermal susceptibility, obeys a linear law. Previous experiments performed on the CHIME Pathfinder instrument also indicated a satisfactory linear model.

The frequency-dependent thermal susceptibility $b_p(f)$ for one polarization p will therefore be given by Equation 3.6:

$$\bar{g}_{frac}(f,t) = a + b_p(f)T(t) \tag{3.6}$$

$\bar{g}_{frac}(f,t)$ corresponds to the median or mean (both will be investigated) of the fractional gain variation over all channels coupled to the same polarization. It would also be possible to fit individually for each channel, and find the median or mean of the susceptibilities afterwards; however, results detailed in Chapter 4 show that averaging before fitting is the best choice as it suppresses some isolated outliers which would significantly impact individual channel susceptibilities.

3.3.2 LNA correction

Lab experiments performed by the CHIME Collaboration prior to this project have shown that the LNAs do not have a linear response to temperature variations. Therefore, the simple model presented in Section 3.3.1 does not apply to the component of the complex gain produced by the antennae. In order to determine the significance of the LNA contribution to the total gain, it is possible to predict the expected LNA response to a given temperature. This operation is performed by a Python code developed as a result of the aforementioned laboratory experiments.

Temperature sensors have been placed directly at 59 of CHIME's LNAs; the data collected by these thermometers allows to calculate the median LNA temperature at the time of a CygA or CasA transit, and deduce the expected LNA gain response. The total complex gain is then divided by the obtained value, and fitted to the external temperature measured by the on-site meteorological station as explained in Section 3.3.1.

In order to evaluate the quality of the fits, the Root Mean Square (RMS) of

residuals given by Equation 3.7 will be calculated. N is the number of transits considered, $f(T)$ represents the linear fit (Equation 3.6) and $\bar{g}_{frac}(T)$ is the fractional gain variation, averaged over channels, measured at temperature T.

$$rms = \left(\frac{\sum_0^N (f(T) - \bar{g}_{frac}(T))^2}{N} \right)^{\frac{1}{2}} \tag{3.7}$$

Equation 3.7 shows that a lower RMS indicates a more accurate fit; therefore, the LNA contribution can be considered significant if the RMS of residuals is lower after the correction than before. The RMS of the data itself will also be calculated over time, frequencies or channels depending on the context (see Chapter 4).

3.3.3 Singular value decomposition

In order to verify that temperature really is the main explanatory factor for the complex gain variations, a linear algebra technique known as Singular Value Decomposition (SVD) can be employed. This method consists of building a matrix in which each column contains complex gain values for all channels, for a given date and frequency. Therefore, there will be 1024 matrices (one per frequency bin) with dimensions $m \times n$ with m the number of channels (2048) and n the size of the dataset.

Such a matrix can be written as a product of two unitary matrices (A and B) and a rectangular diagonal matrix (D).

$$M = ADB^{\dagger} \tag{3.8}$$

The values on the diagonal of D are known as singular values. In an ideal case where one single factor dominates the behaviour of the complex gain, the matrix M can be written as an outer product of two vectors multiplied by the only nonzero singular value. In reality, we are not expecting all the other singular values to be zero; however, a first singular value significantly bigger than the other ones is a good indication that the one-parameter model considered so far is a decent approximation. This method will therefore be used to test the assumption that temperature variations are the main reason behind the observed gain fluctuations.

Chapter 4

Results

4.1 Preliminary observations

Before performing any fitting or advanced analysis, some general, qualitative observations about the raw gain data can be mentioned. Figure 4.1 and Figure 4.2 show the gain amplitude values obtained as a function of frequency for 4 different channels, for the same CygA transit (08/10/18). Three main characteristics directly appear.

First, a global increasing pattern is visible on all 4 panels, meaning that the complex gain amplitude tends to be greater for higher frequencies.

Second, an oscillatory pattern with a characteristic frequency of 30 MHz appears, and is more defined in Figure 4.2 than in Figure 4.1. This oscillation can be explained by an interference effect produced by multiple reflections between the cylindrical structure and the antennae. At certain frequencies, this interference is constructive and results in an amplified received intensity, thus increasing the measured gain; this occurs when the distance between the cylinder and the focal line is an integer multiple of the wavelength of the signal. In between these peaks, lower gain values are consequences of destructive interference. Calculations carried out prior to this project predicted a period of about 30 MHz for this interference phenomenon, which is in agreement with the observed pattern. Thermal expansion of the cylinder will modify the distance to the focal line; therefore, we can expect the whole interference figure to be shifted as a consequence of temperature variations.

The third interesting feature is the presence of some isolated outliers which are significantly higher or lower than the rest of the data. Despite being clearly off the trend, these points were not filtered out by the flagging algorithm. As the next few sections will show, such events can have a non-negligible impact on the measured susceptibilities and will therefore have to be understood and possibly even predicted.

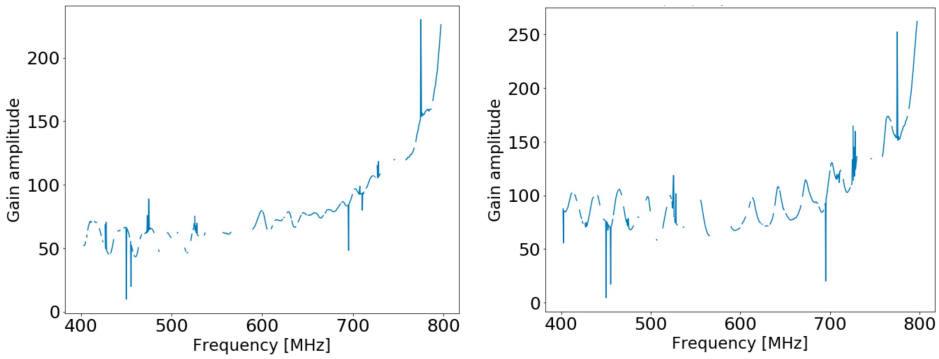

Figure 4.1: Gain amplitude for the CygA transit on 08/10/18. The left panel shows the data obtained for channel 1, and the right panel for channel 500. Solar transits and RFI-contaminated frequencies have been removed, which explains the blank spots. A global rising pattern combined with a 30 MHz oscillation can be observed.

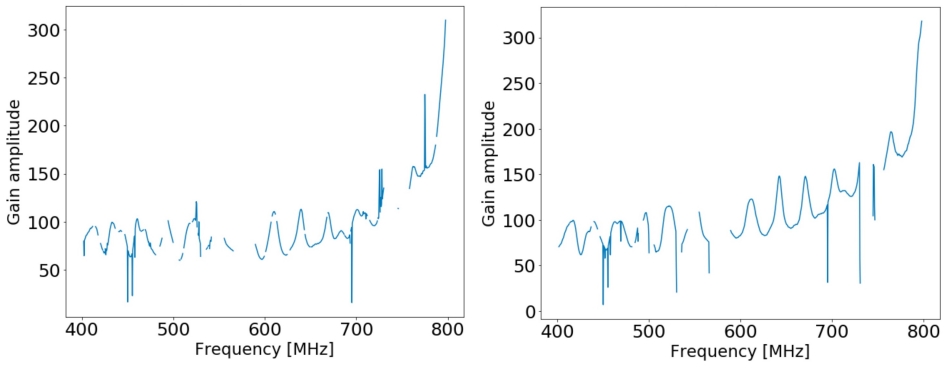

Figure 4.2: Gain amplitude for the CygA transit on 08/10/18. The left panel shows the data obtained for channel 1000, and the right panel for channel 1500. Similar trends as in Figure 4.1 are visible.

The waterfall plot shown in Figure 4.3 displays the gain amplitude obtained from the same CygA transit, as a function of frequency and channel. The white stripes represent missing data (frequencies that have been filtered out by the flagging algorithm). While channels seem to behave rather uniformly, it is difficult to distinguish any details; this is due to the presence of a few isolated extremal gain values, about one order of magnitude greater than the average. Being so rare, these values are not properly visible on the plot; however, their existence is indicated by the extended range of the colorbar. Understanding what produces such extreme values and why they are not filtered out will be an important step in the later stages of analysis.

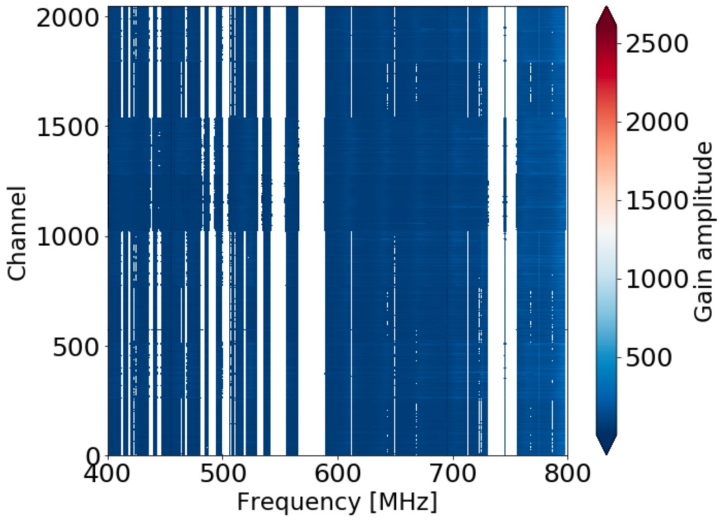

Figure 4.3: Waterfall plot of the gain amplitude obtained for the CygA transit on 08/10/18, over all frequencies and all channels. The white lines correspond to filtered out data. The wide global amplitude range is due to the presence of a few extreme outliers, but most of the amplitudes are contained in a range of 0 to 400.

Figure 4.4 shows two slices of Figure 4.3, for two different sets of 200 channels each. The behaviour of different channels appears to be sufficiently unifom over this range to justify the averaging process explained in Chapter 3. Indeed, it only makes sense to average the fractional gain variation over all channels if they all follow a similar trend, which appears to be the case in both of these plots. Similar

results have been observed for most other channels (except the very few containing extremal gain amplitudes).

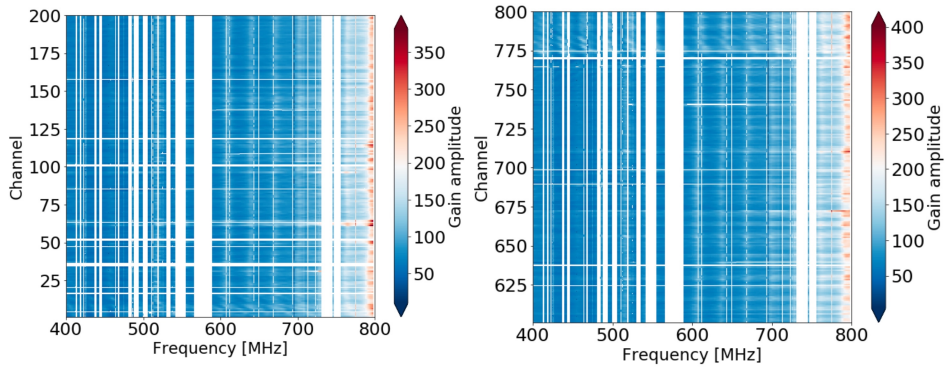

Figure 4.4: Gain amplitude for the CygA transit on 08/10/18. The left panel shows the data obtained for channels 1 to 200, and the right panel for channels 601 to 800. These two ranges don't contain any extreme outliers, therefore more detailed patterns than in Figure 4.3 can be distinguished. Comparing the two panels shows that the gain amplitude tends to behave rather uniformly over channels.

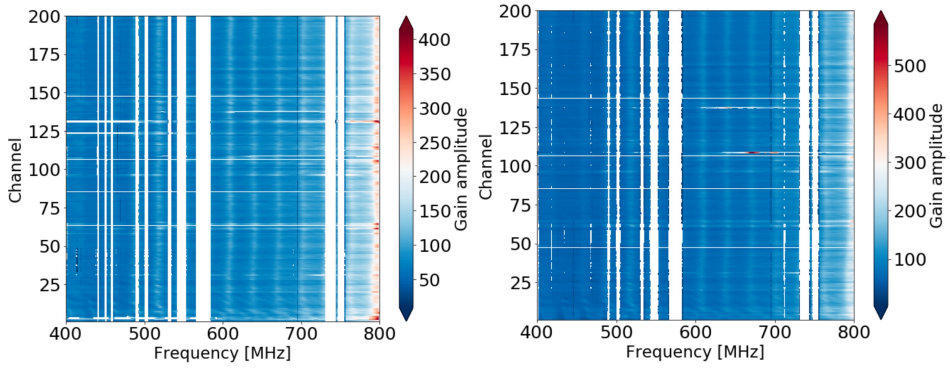

Figure 4.5: Gain amplitude for CygA transits displayed over channels 1 to 200. The left panel shows the data obtained on 09/24/18, and the right panel on 10/24/18. Here again, the uniform behaviour of channels is noticeable. Slight differences can be observed between the two dates, but the general pattern is similar and the 30 MHz oscillation is visible.

Analogous comments can be made about Figure 4.5. This time, the group of

200 channels was kept the same, but two different CygA transits are represented. Both plots still display a uniform tendency for all considered channels. The visible differences between the two graphs are an illustration of the fact that the gain amplitude varies depending on the considered transit, a phenomenon that is primarily explained by the instrument's sensitivity to external temperature and will be studied in detail in the next few sections.

4.2 First susceptibility results

In order to qualitatively test the hypothesis that the gain amplitude varies linearly with temperature, Figure 4.6 displays the correlation between median gain fractional variation and external temperature, for 107 transits of CygA and 4 different frequencies. It appears that the data follows the expected linear trend, which confirms the fundamental hypothesis of this project (at least in first approximation). A large majority of the points are moderately scattered around an obvious linear pattern. A few isolated values are significantly off the expected trend; they are greatly outnumbered by the other points and are, in this case, not sufficient to visibly perturb the fits.

Performing similar fits for individual channels can result in a much less distinct linear behaviour, as shown in Figure 4.7. In the left panel, the data is way too scattered to follow any clear pattern, and the linear fit is completely off. A slight tendency to align can be seen in the left half of the graph, but the group of points in the bottom right corner prevents us from drawing any conclusion from this figure. The right panel, however, shows that while a channel sometimes does not display the expected pattern for a given frequency, it can follow a linear trend (similar to that observed in Figure 4.6 for the median of all channels) at another frequency. In general, this comparison indicates that it is preferable to average the fractional gain variation over all channels before performing the fits, as the fits for individual channels can be more easily perturbed by the scattering of data or the presence of extremal values.

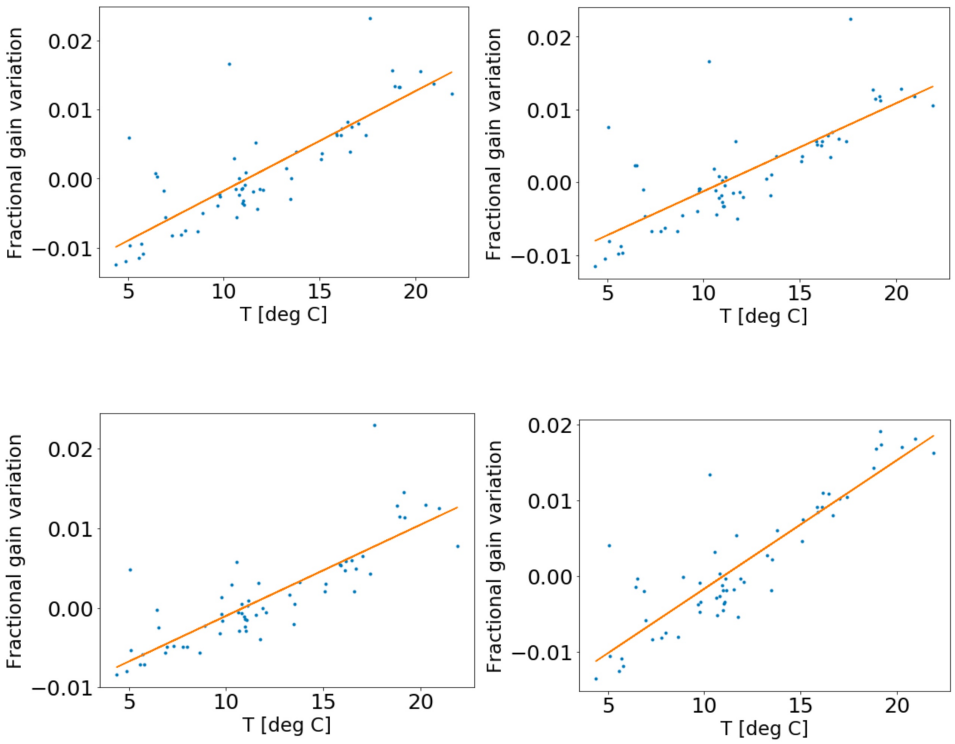

Figure 4.6: Linear fits of the median gain fractional variation as a function of temperature for different frequencies (pol. E-W). Top left: 721.875 MHz. Top right: 682.815 MHz. Bottom left: 643.75 MHz. Bottom right: 604.6875 MHz. The represented dataset includes 107 CygA transits between 05/31/18 and 10/30/18. The data appears to follow a clear linear trend despite moderate scattering and some isolated outliers.

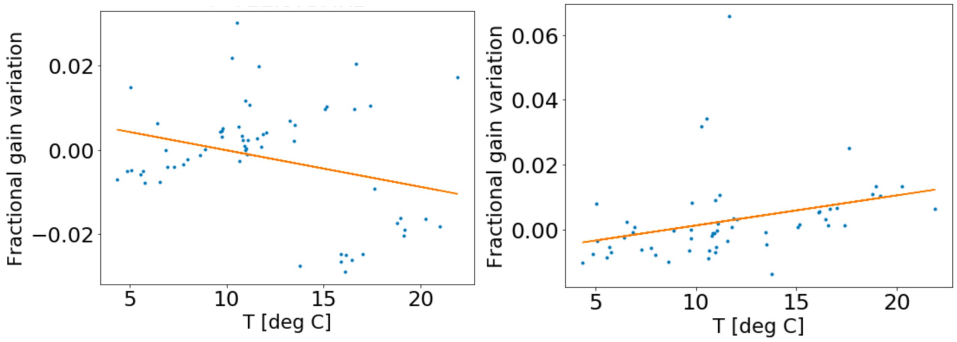

Figure 4.7: Linear fits of the gain fractional variation as a function of temperature for channel 1001 and two different frequencies. Left panel: 721.875 MHz. Right panel: 643.75 MHz. The fit in the left panel is clearly off as the points do not follow a linear trend. However, the same channel exhibits a mostly linear behaviour at a different frequency.

The slopes of all the linear fits of the fractional gain variation as a function of temperature can be represented in a single plot, as frequency-dependent thermal susceptibility values. Such a graph is displayed in Figure 4.8, containing the susceptibility values obtained from the aforementioned 107 CygA transits and from another dataset including 74 CasA transits. It is immediately clear that the 30 MHz oscillation that affects the gain amplitudes, also has an influence on the thermal susceptibility. The oscillatory pattern is mostly visible between 600 and 750 MHz for both polarizations. As expected, the susceptibilities obtained from CygA and CasA data are very close; the complex gain being defined as direction-independent, noticing a strong dependency on the observed source would be abnormal.

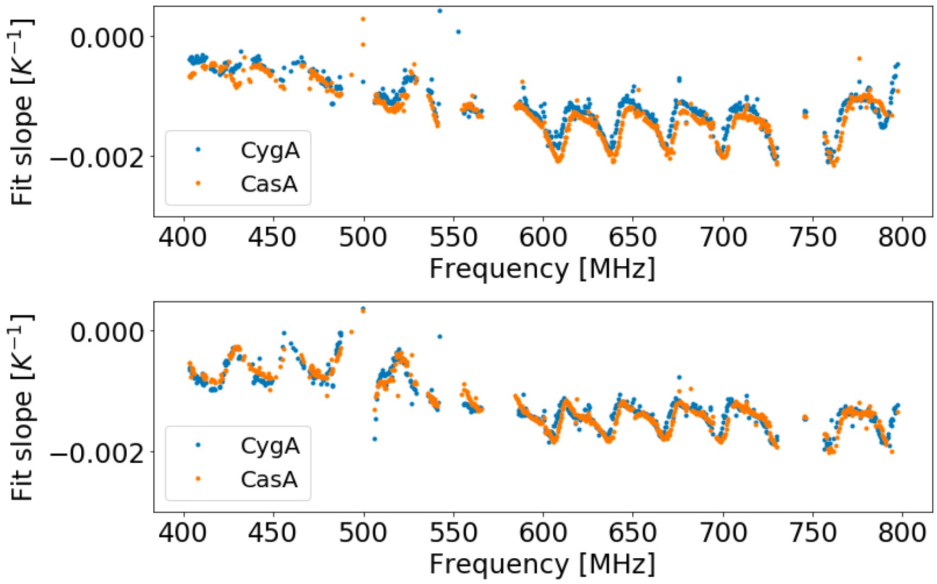

Figure 4.8: Thermal susceptibility obtained as a function of frequency, from the 107 CygA transits and from 74 CasA transits between 07/25/18 and 11/01/18. Top panel: pol. E-W. Bottom panel: pol. N-S. A clear oscillatory pattern is visible, especially between 600 and 750 MHz, with a characteristic frequency of 30 MHz. The susceptibility values exhibit little to no source dependency.

A plot very similar to Figure 4.8 was obtained with two additional datasets: the first one containing 34 CygA transits between 09/21/18 and 11/05/18, and the second one containing transits of TauA, another bright radiosource that could potentially be used for calibration. While the new CygA results follow the same trend as the two first datasets and are only slightly shifted upwards, the TauA data has a very strange behaviour. This dataset was introduced in order to confirm the absence of dependency on the observed source; however, the results actually indicate the exact opposite. The TauA susceptibilities are much lower than the rest and do not exhibit a distinct pattern. Furthermore, a significant fraction of the data is missing, meaning that it has been filtered out during pre-processing. While it would be interesting to understand the reason behind this odd behaviour, the scarceness of data and the encouraging results obtained for the two other sources pushed us to momentarily discard TauA. The TauA datasets will be further analyzed in Section 4.5.

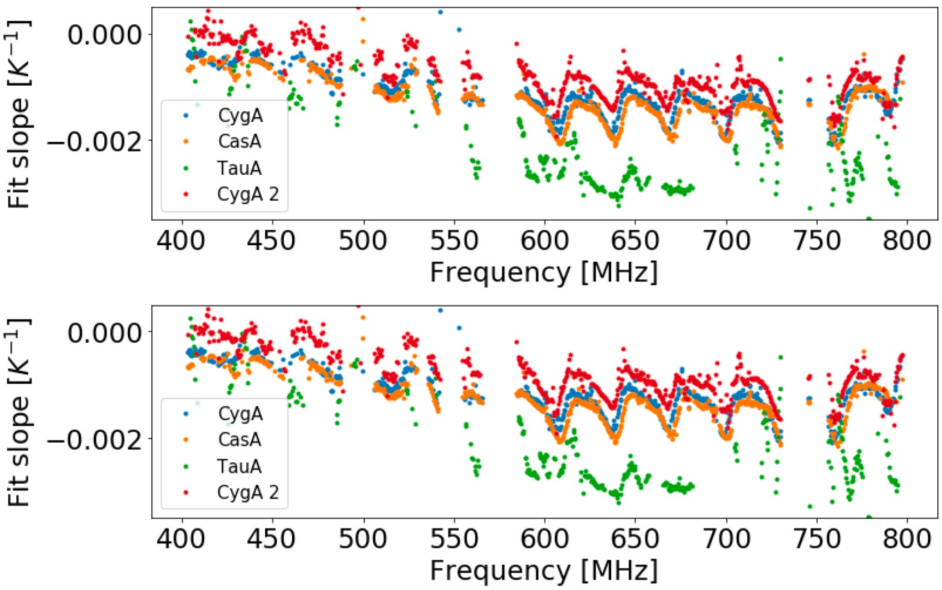

Figure 4.9: Thermal susceptibility obtained as a function of frequency. Red points: 34 CygA transits between 09/21/18 and 11/05/18. Green points: 46 TauA transits between 09/20/18 and 11/24/18. While the two CygA and the CasA datasets are in good agreement, apart for a slight shift upwards, the TauA susceptibilities are completely off the expected trend.

4.3 LNA correction

As explained in Chapter 3, we have attempted to account for the nonlinear behaviour of LNAs in the hopes of reducing scattering and improving the quality of the fits. The LNA temperatures measured by the 59 on-site sensors are displayed in Figure 4.10.

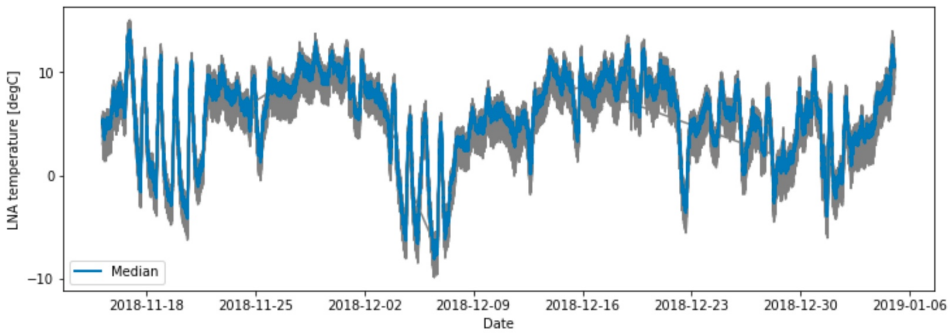

Figure 4.10: Temperature measurements as a function of time for the sensors placed at 59 of the LNAs. The grey curves represent the individual measurements, the blue curve corresponds to the median of the 59 values.

Here again, the first step of the analysis was to check the coherence of the linear model and to observe the influence of the correction on the slopes of the fits. Figure 4.11 shows the fits obtained before and after correction for a CasA dataset including 36 transits. The plots show a linear trend accompanied by quite significant scattering, both before and after correction; this scattering has a more important effect on these fits than on the ones obtained in the previous section due to the lower number of data points available. The correction influences the slopes, which means that it will produce a shift in the susceptibilities. However, it does not seem to significantly improve the scattering problem.

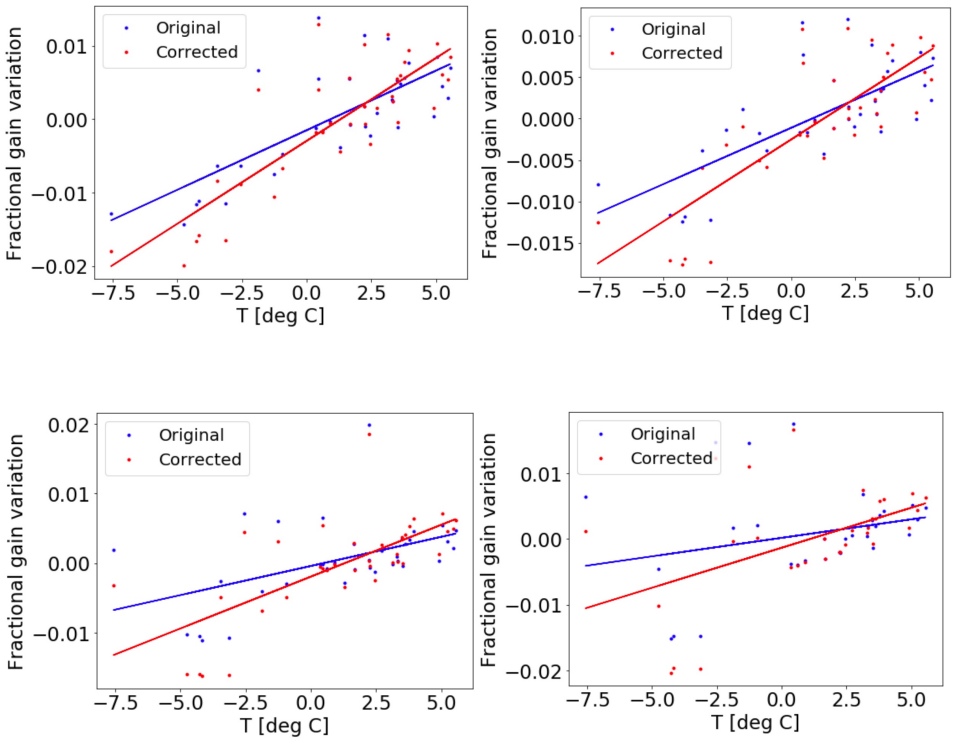

Figure 4.11: Linear fits of the median gain fractional variation as a function of temperature for different frequencies (pol. E-W), before and after removing the LNA gain component. Top left: 721.875 MHz. Top right: 682.815 MHz. Bottom left: 643.75 MHz. Bottom right: 604.6875 MHz. The represented dataset includes 36 CasA transits between 11/17/18 (date of first LNA temperature measurement) and 01/03/19. Linear trends are visible both before and after correction, and the correction visibly impacts the slope. Important scattering is present in both cases.

As suspected from the qualitative observation of Figure 4.11, the values of the RMS of residuals before and after the LNA correction are almost identical. This indicates that the correction does not significantly impact the quality of the fits and the accuracy of the model. As expected, most of the temperature-dependent response comes from the cables and steel structures; the antennae themselves do not play a major role.

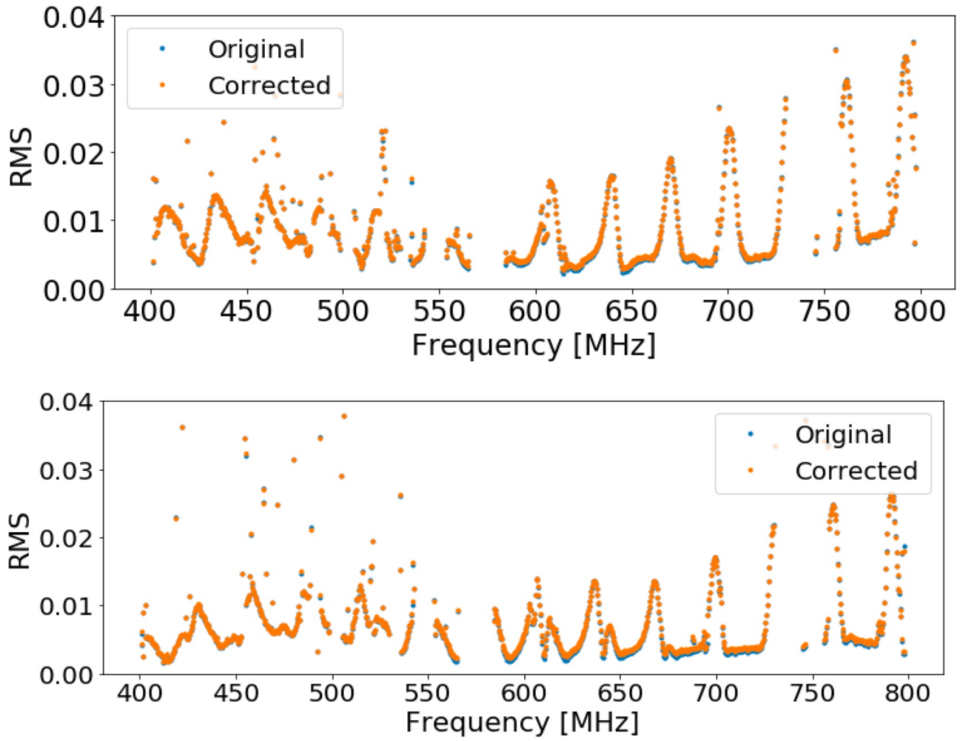

Figure 4.12: RMS of residuals before and after LNA correction, for pol. E-W (top panel) and pol. N-S (bottom panel). An oscillatory pattern appears with a characteristic frequency of about 30 MHz. The correction does not significantly lower the RMS values.

Figure 4.13 represents the susceptibility values deduced from the 36 CasA transits, before and after the correction. While the corrected values are shifted downwards (removing the LNA component slightly lowers the global susceptibility), they follow the same oscillatory trend as the original data points.

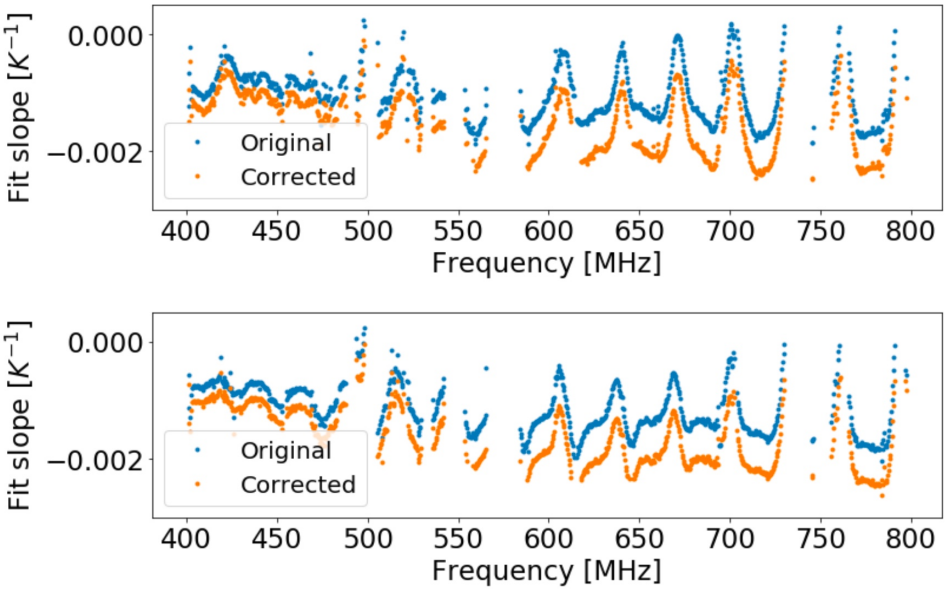

Figure 4.13: Thermal susceptibility obtained as a function of frequency from the 36 CasA transits, before and after LNA correction. Top panel: pol. E-W. Bottom panel: pol. N-S. The typical 30 MHz oscillation is still visible. The original and corrected data follow a very similar pattern, however the correction introduces a downwards shift in the susceptibility values.

Finally, Figure 4.14 illustrates the strong dependency of the susceptibility on the considered dataset. The blue points are obtained from 70 CasA transits; the orange points are deduced from a 36 transits sample of this same dataset, which corresponds to the time period over which LNA temperature measurements have been available. Both datasets are represented without LNA correction. Such a dependency on the considered time period is a surprising feature; just like the complex gain is supposed to be source-independent, it is also assumed to be time-independent at constant temperature (if two different transits of the same source were observed with the same external temperature, they should yield the same complex gain). Nevertheless, Figure 4.14 seems to prove the opposite. This observation has led us to wonder if temperature really is the only dominating factor explaining gain variations; Section 4.4 and Section 4.5 seek to answer this question.

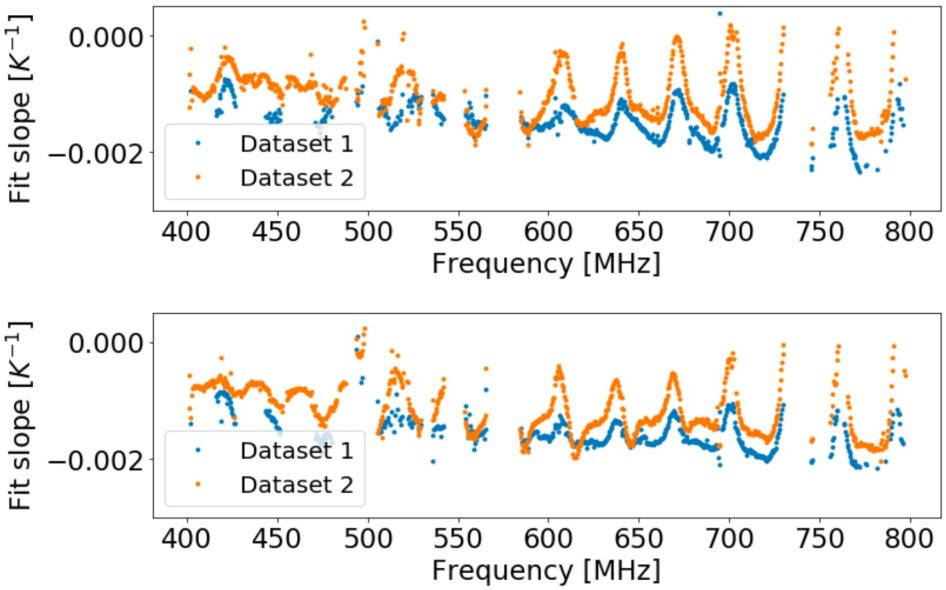

Figure 4.14: Thermal susceptibility obtained as a function of frequency from the full 70 CasA transits dataset (Dataset 1, from 09/21/18 to 01/03/19), compared to the ones deduced from the 36 transits sample (Dataset 2), without LNA correction. Top panel: pol. E-W. Bottom panel: pol. N-S. A strong dependency on the considered time period is visible, as the susceptibility values are significantly different despite the fact that the two time ranges overlap.

4.4 Singular value decomposition

As seen in the previous section, the RMS of residuals has an oscillatory behaviour, with peaks corresponding to the most imprecise fits and flat regions corresponding to zones in which the linear model accurately represents the data. The same observations can be made about the RMS of the data and the RMS of residuals for the 70 CasA transits without LNA correction.

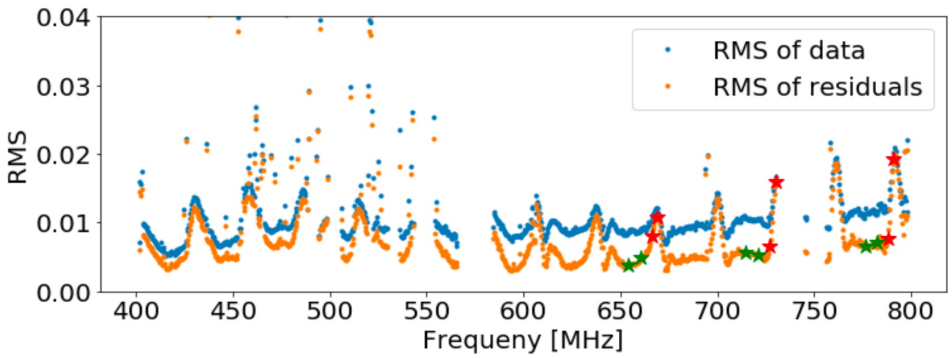

Figure 4.15: RMS of data and RMS of residuals for the 70 CasA transits (pol. E-W) without LNA correction. The values highlighted with red stars correspond to frequencies for which the RMS of residuals is high, indicating scattered data or an imprecise fit. The green stars indicate frequencies with lower RMS, where the linear fits are satisfactory. The RMS of residuals is globally lower than that of the data, which indicates that using the linear model already improves the quality of the data.

In order to better understand what causes some fits to be more accurate than others, we plotted the median fractional gain variation as a function of external temperature, with the corresponding linear fits, for 6 ”good” frequencies (green stars) and 6 ”bad” frequencies (red stars). Figure 4.16 and Figure 4.17 show that the peak RMS frequencies (on the left) systematically contain between 4 and 6 outliers that are way off the linear trend, while the rest of the points tend to be fairly well aligned. The good frequencies do not contain such clear outliers.

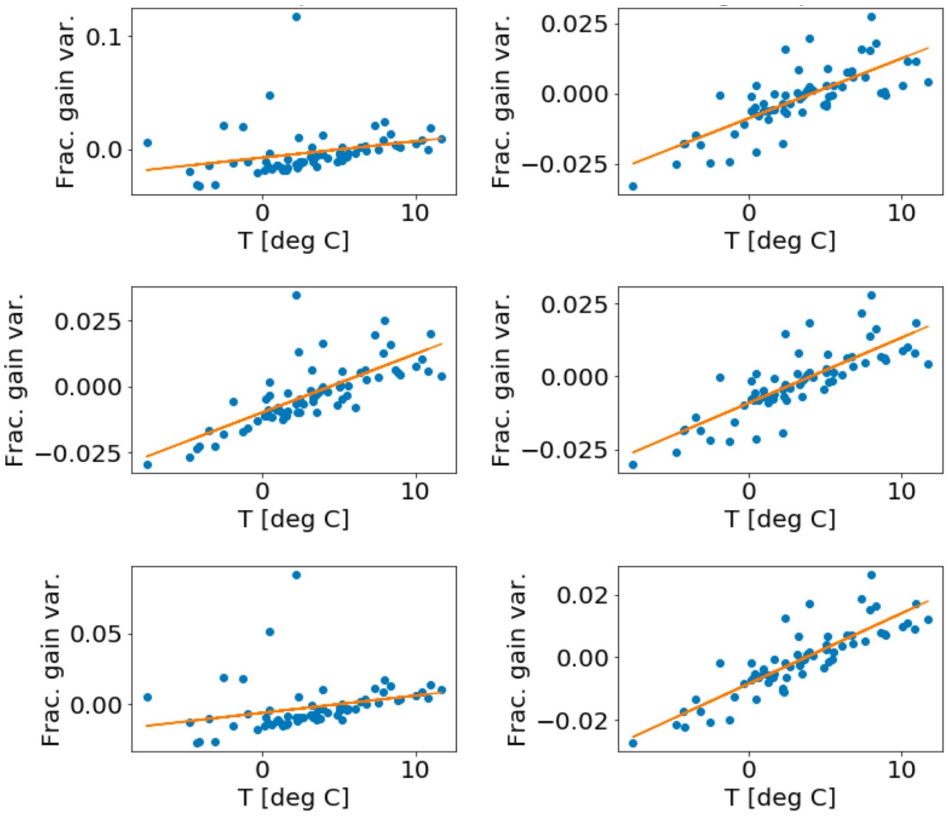

Figure 4.16: Linear fits of the median fractional gain variation (pol. E-W) obtained from the 70 CasA datasets, for peak RMS (left side) and low RMS frequencies (right side). From left to right and top to bottom: 791.016 MHz, 782.031 MHz, 787.891 MHz, 776.172 MHz, 730.078 MHz, 721.094 MHz. Frequencies with high RMS tend to contain isolated outliers that do not appear in the plots for the "good" frequencies.

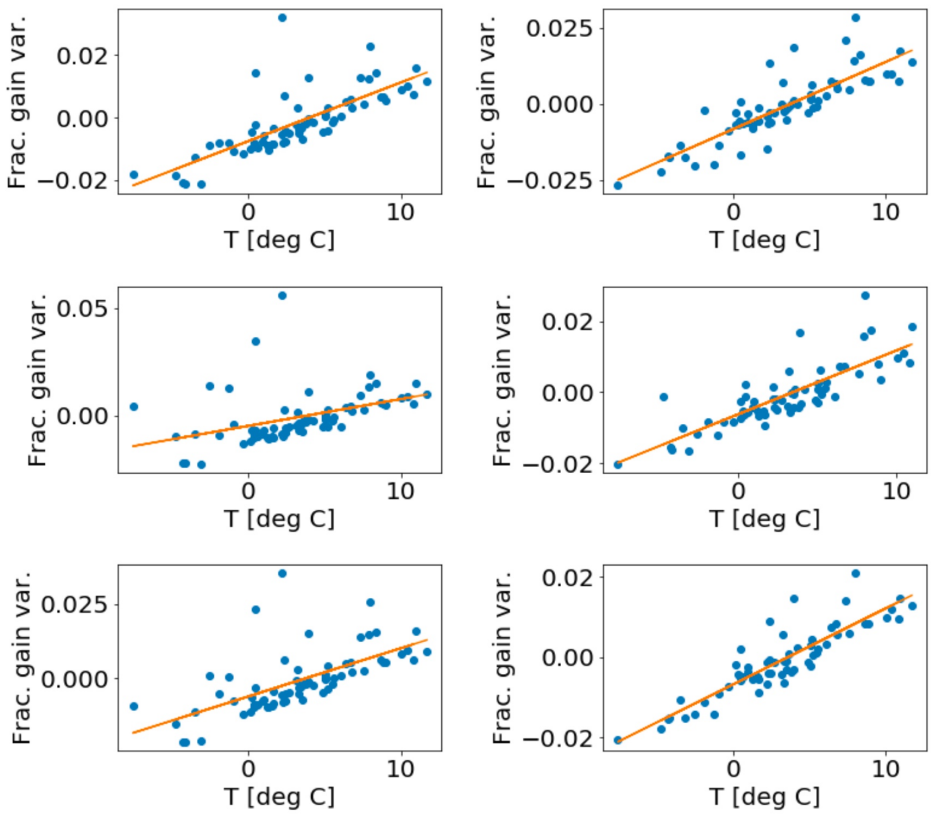

Figure 4.17: Same as Figure 4.16, for different frequencies. From left to right and top to bottom: 726.953 MHz, 714.844 MHz, 669.141 MHz, 660.156 MHz, 666.016 MHz, 653.906 MHz.

While the results of the LNA correction and the apparent dependency on the considered dataset seemed to indicate that temperature was not the only factor of importance, the results displayed in the two previous figures show that the linear thermal model is actually rather correct, except for a few frequencies for which a small number of isolated points are sufficient to perturb the fits. This might mean that there are a few particular days on which the transit data was contaminated by an external source, which could explain the recurring 4-6 clear outliers. The frequency dependent nature of these outliers, with a 30 MHz characteristic frequency, seems to indicate a relation between these points and the reflectors.

To make sure that temperature really is the main parameter, a singular value

decomposition was performed on the CasA data. Figure 4.18 shows the ratio between the first and second singular values, for the 6 peak RMS frequencies (orange) and the 6 low RMS frequencies (blue). In both cases, this ratio varies between 30 and 65; this result demonstrates the fact that the first singular value is largely dominant compared to the others, meaning that a one-parameter model is a good approximation of the situation.

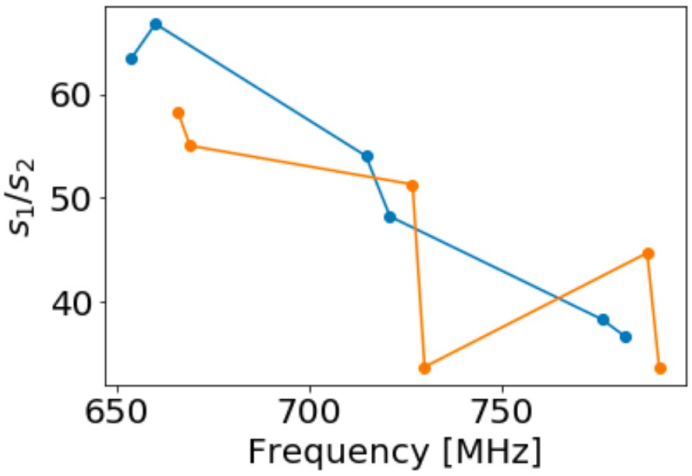

Figure 4.18: Ratio of first to second singular values for the 6 good frequencies and the 6 bad frequencies from Figure 4.16 and Figure 4.17. The first singular value clearly dominates in all cases.

If the dominating parameter is indeed temperature, the values contained in the right singular vector $\vec{v}$ should be proportional to the considered temperature values. This hypothesis was tested in Figure 4.19, where the right singular vector multiplied by the first singular value s_1 was plotted against temperature (right side). A linear trend is distinguishable, however the fit is thrown off by some clear outliers in the bottom of the plot. In the left panel, the values contained in $s_1\vec{v}$ are multiplied by the inverse slope a of the fit on the right and compared to the measured external temperatures. The two curves exhibit some similar tendencies but do not coincide.

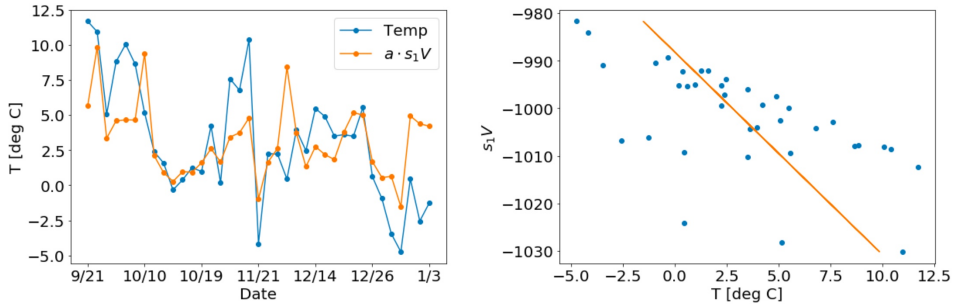

Figure 4.19: Right panel: linear fit of right singular vector values VS external temperature. The slope of the obtained fit is used in the left panel, in order to compare the temperature evolution to the singular vector values multiplied by a constant. Due to 7 points visible on the right side, the calculated fit does not follow the actual linear trend of the data and the agreement between the two curves in the left panel is questionable.

However, if the 7 clear outliers are removed, the linear trend in the right side of Figure 4.20 is evident and the agreement between $as_1\vec{v}$ and the measured temperatures is greatly improved. This result illustrates the fact that the thermal model is only perturbed by a restricted number of points.

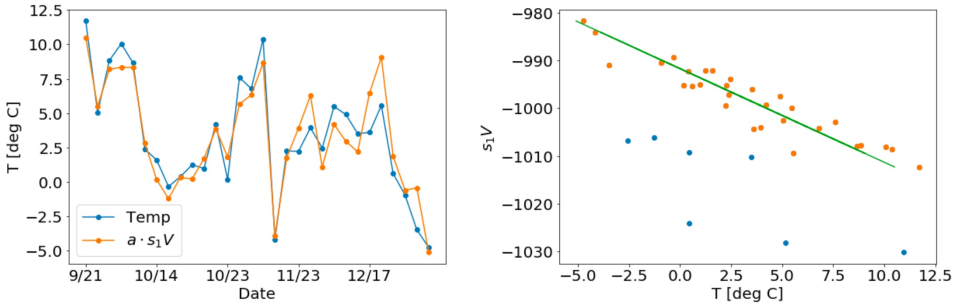

Figure 4.20: Same as Figure 4.19, without the 7 outliers (in blue). The linear fit now describes the remaining data (orange) in a satisfactory way, and the right singular vector is clearly related to temperature.

The next step of the analysis will be to determine to which transits the outliers in Figure 4.19 correspond. If they are the same as the transits producing the systematic outliers in Figure 4.16 and Figure 4.17, this would be a very encourag-

ing result: it would mean that, except for a restricted group of abnormal transits, CHIME's thermal response is well-described by a linear model. If this hypothesis is confirmed, the perturbed transits will have to be further investigated (rain and wind statistics, as well as potential interference problems, might have to be considered). Ideally, this would allow us not only to suppress these contaminated transits in order to obtain a more accurate model, but also to predict when such anomalies are susceptible of happening again.

4.5 Analysis of outliers

A more profound analysis allowed us to identify two distinct types of outliers. Studying the mean of the fractional gain variation instead of the median over channels exhibited another family of anomalous points, now referred to as "type 1 outliers", which did not appear in the previous investigations. In these particular transits, channels do not behave in a uniform way. It is therefore suspected that these outliers result from the influence of an external source on some isolated channels; the primary hypothesis, which will be strengthened by later results, is that rain causes water to enter some of the antennae and trigger an abnormal response. The phenomenon significantly perturbing a restricted number of feeds explains why type 1 outliers were undetectable when considering the median of channels, but are visible when working with the mean (see Figure 4.22 and Figure 4.23). As type 1 outliers are the only transits for which channels exhibit a non-homogeneous behaviour, they can be detected by taking the RMS over channels of the fractional gain variation. As shown in Figure 4.21 for CasA, the RMS values for type 1 outliers are clearly superior to those of normal days. The difference is especially significant between 400 and 500 MHz. This allowed to determine a criterion to detect type 1 outliers: the RMS over channels is superior to 0.1 in this frequency range. This method successfully identified outliers of this type in TauA datasets, as shown in Figure 4.22 and Figure 4.23.

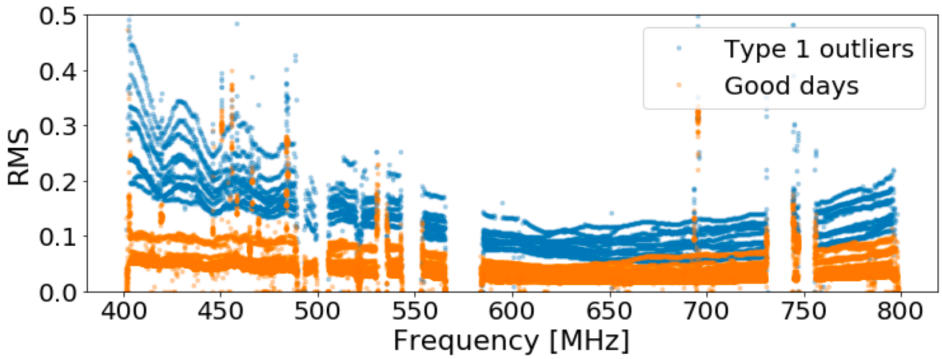

Figure 4.21: RMS of the fractional gain variation across channels, as a function of frequency, for the 34 CasA transits. Each curve corresponds to one transit. Type 1 outliers (blue) clearly have a higher RMS than good days (orange), especially between 400 and 500 MHz. This comes from the fact that only some of the antennae exhibit an anomalous response on these days, resulting in a higher variation across channels. Note that type 2 outliers are not detected by this method.

The second family of outliers is the one already observed in Section 4.4, which also appears when considering the median fractional gain variation and exhibits a strong frequency dependance. These anomalies, referred to as type 2 outliers, are not channel-dependent, so they cannot be detected by the previously mentioned method. However, these outliers being sensitive to frequency allows to define a new criterion: this time, RMS values are computed across frequencies instead of channels, around 0.1 of the median, and the threshold is fixed at 0.014. This criterion, defined by observing CasA data, effectively picked out TauA outliers as shown in pink on Figure 4.22 and Figure 4.23. These two figures clearly illustrate the two types of outliers, with type 1 points always appearing off and type 2 points only causing problems in ”peak” RMS frequencies. The linear fit of the mean fractional gain variation as a function of external temperature, when ignoring both types of outliers, accurately describes the remaining data.

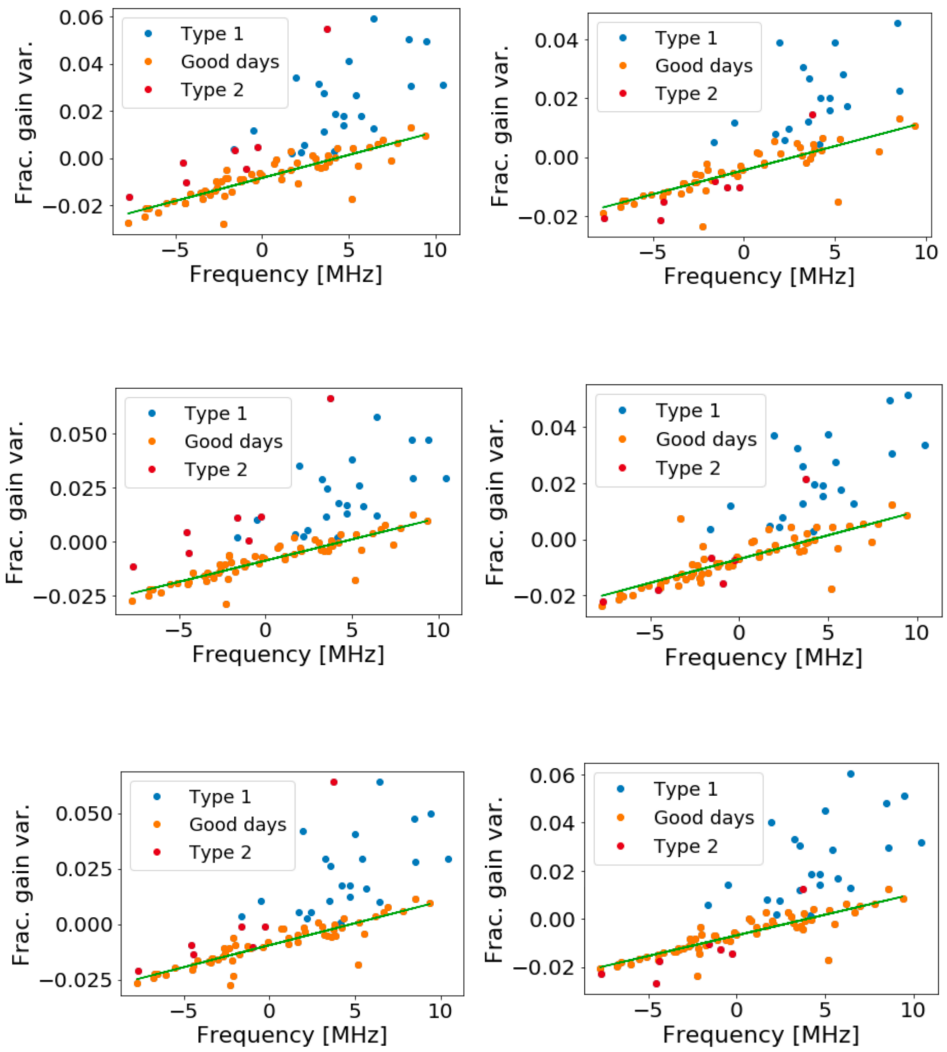

Figure 4.22: Linear fits of the mean fractional gain variation VS temperature, excluding type 1 (blue) and type 2 outliers (pink), for 97 TauA transits between 09/20/18 to 01/23/19 and different frequencies. The remaining data clearly follows the linear trend, and the frequency-dependent nature of type 2 outliers is visible. The left column shows peak RMS frequencies, the right one shows trough frequencies. From left to right and top to bottom: 669.1 MHz, 651.2 MHz, 671.1 MHz, 657.0 MHz, 703.5 MHz, 685.5 MHz.

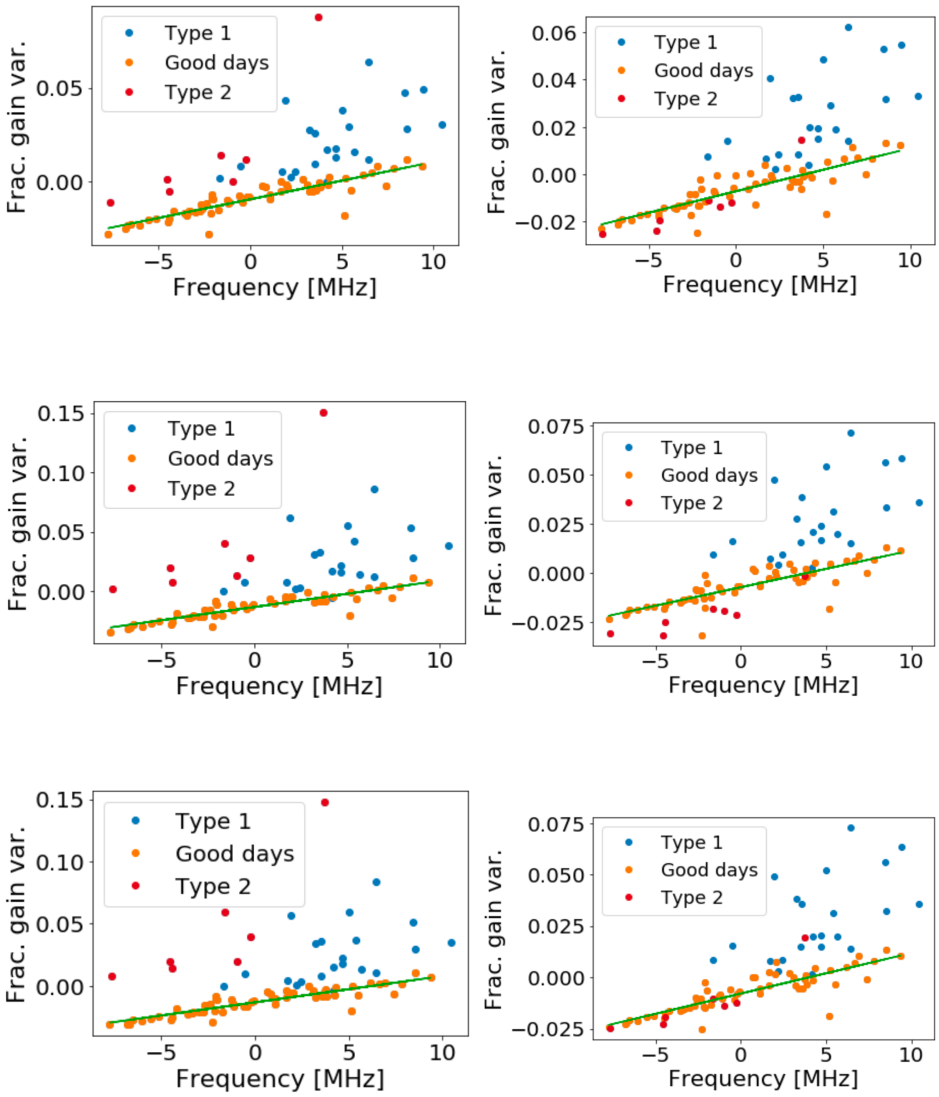

Figure 4.23: Same as Figure 4.22 for different frequencies. From left to right and top to bottom: 705.5 MHz, 691.4 MHz, 772.7 MHz, 719.9 MHz, 774.6 MHz, 726.2 MHz.

Having computed these new, improved fits after detecting and filtering out anomalous transits allows to calculate the new RMS of residuals and the RMS of the mean fractional gain variation across days. These two quantities are plotted in

Figure 4.24 for all three sources. The red and green stars correspond to the ”peak RMS” and ”low RMS” frequencies represented in the two previous figures; after filtering out outliers, peaks in the RMS of residuals have disappeared and no significant frequency dependance is observed. The target value of 0.003 is reached for part of the frequency range. This means that the linear model, combined with the appropriate treatment of outliers, improves the quality of the data sufficiently to meet the accuracy criteria of the instrument. This remark is also supported by the fact that the RMS of residuals is lower than that of the data without thermal model, especially in the 550-800 MHz range.

Now that the thermal model has been confirmed to be satisfactory when outliers are ignored, it is important to find the reasons behind these anomalous transits in order to be able to predict them. Figure 4.25 displays rain quantities cumulated for a six hour period at the DRAO, over the whole observation time. These statistics are superposed to the times of TauA transits, with type 1 and 2 outliers being represented in blue and pink respectively. It appears that type 1 outliers exhibit a correlation with the peaks in the graph: all type 1 points closely coincide with a rain event. This supports the primary hypothesis that these anomalies stem from water entering some of the antennae. A few of the minor peaks do not coincide with any type 1 outlier; this does not go against the hypothesis, as it is possible for rain to occur without causing significant amounts of water to perturb the response of some of the feeds. Type 2 outliers are more difficult to explain and no definite cause for their existence has been found yet. It is interesting to note that all the TauA type 2 points correspond to a single group of consecutive days, right after New Year. A similar constatation has been made with CasA type 2 outliers, which coincide with the TauA values. This series of dates is preceded by several significant peaks in the rain statistics, therefore it is possible that these outliers are also related to water in some way. However, this does not explain the inherently different characteristics of these transits compared to type 1 outliers, with the homogeneous channel behaviour and the frequency dependance seemingly indicating a relation with reflector effects.

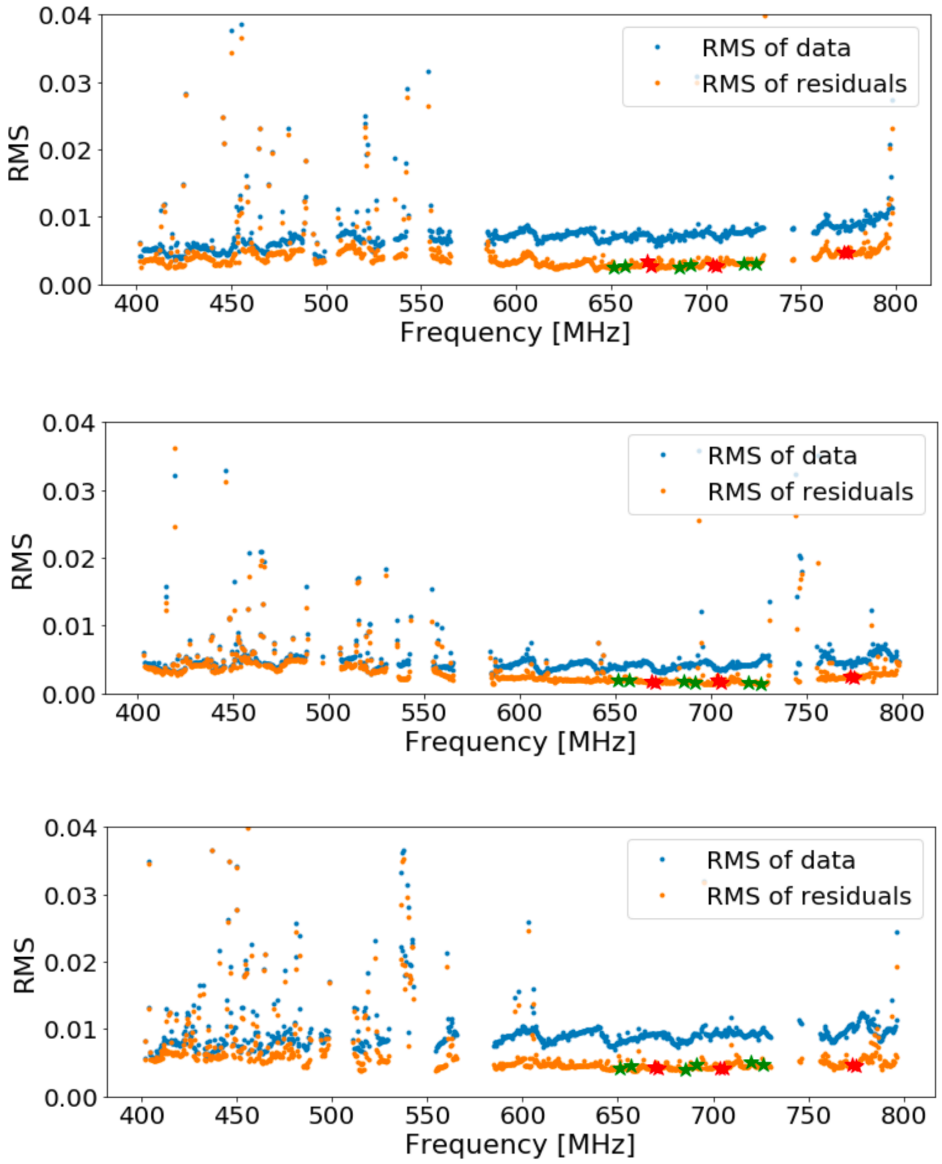

Figure 4.24: RMS of the data and RMS of residuals as a function of frequency, ignoring outliers, for the mean fractional gain variation of the 3 sources (E-W polarization). From top to bottom: CasA, CygA, TauA. The RMS of residuals is systematically lower than that of the data and does not exhibit a strong frequency dependance anymore. The target of 0.003 is attained between 600 and 700 MHz.

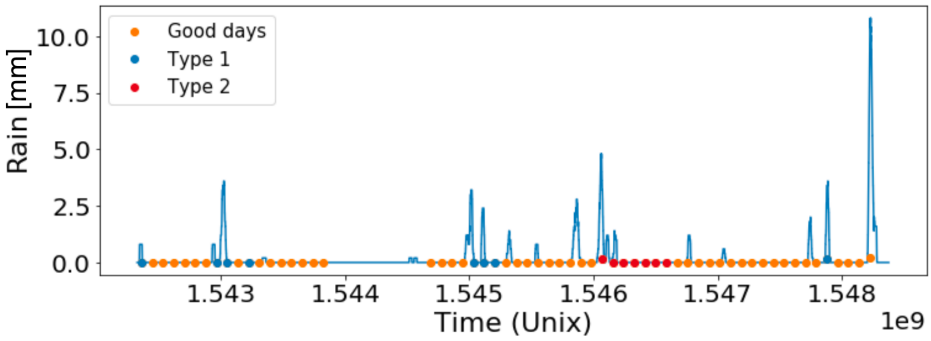

Figure 4.25: Comparison of TauA outlier days with rain accumulation over a 6 hour period. A correlation seems to exist between type 1 outliers and peak rain events. The series of type 2 outliers also follows an important rain accumulation, but the relation is not certain yet.

The improved linear fits resulting from the appropriate detection and treatment of outliers can be used to compute the final thermal susceptibility values for all three sources, over the whole frequency range. The results are displayed in Figure 4.26. Note that this figure shows a general increasing trend and positive values, while previous susceptibility results exhibited the opposite tendency: this is due to the fact that Figure 4.26 was computed using inverse gain values. To first order, the slopes of the fits are therefore opposite of those found for the gains themselves. This plots contains similar features as previous susceptibility plots: 30 MHz ripples and a slight dependency on the considered source. It appears that the quality of TauA datasets was greatly improved by the removal of outliers, as the TauA data now closely matches the CasA values. This is a significant progress compared to Figure 4.9.

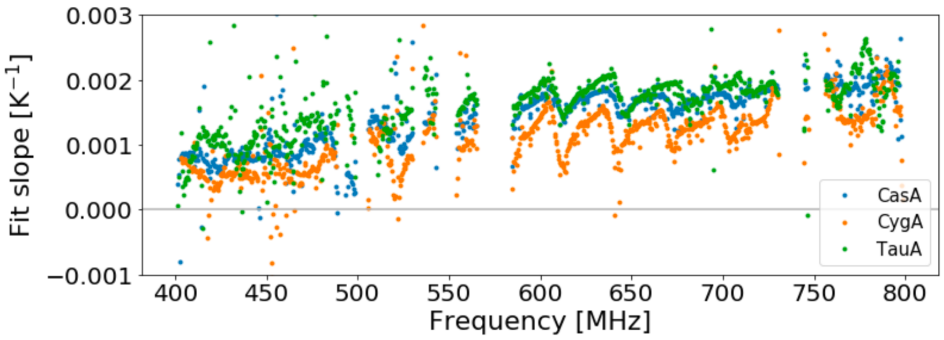

Figure 4.26: Susceptibilities obtained from linear fits of the mean fractional gain variation, filtering out outliers, for the 70 CasA transits, the 34 CygA transits and the 97 TauA transits. Despite being shifted with respect to each other, the three sources produce similar results and the behaviour of TauA is now coherent with the two other datasets. Scattering is significant and the clearest trend is visible between 600 and 700 MHz.

Finally, the dependency of susceptibilities on the considered portion of the dataset was tested by performing the linear fits using only half of the transits, in four different configurations. The first half and second half were used separately, then the even and odd days were also extracted and employed for the fits. The results shown in Figure 4.27, despite some differences, all exhibit the same general trend and the order of the susceptibility values is in agreement between the four panels. This is a progress compared to the results obtained before excluding outliers, which exhibited a very strong dependency on the considered time period. This improvement can be explained by the fact that removing outliers related to irregular events such as rain allowed to obtain a more reliable one-parameter model. An ideal one-parameter model would be completely independent on the chosen dates. In our case, the uncertainty introduced by considering different time periods is comparable to the difference between the three radiosources, and to the amplitude of the 30 MHz ripples. Within these limitations, the thermal model can be considered to be consistent with the direction-independent and dataset-independent nature of the complex gain.

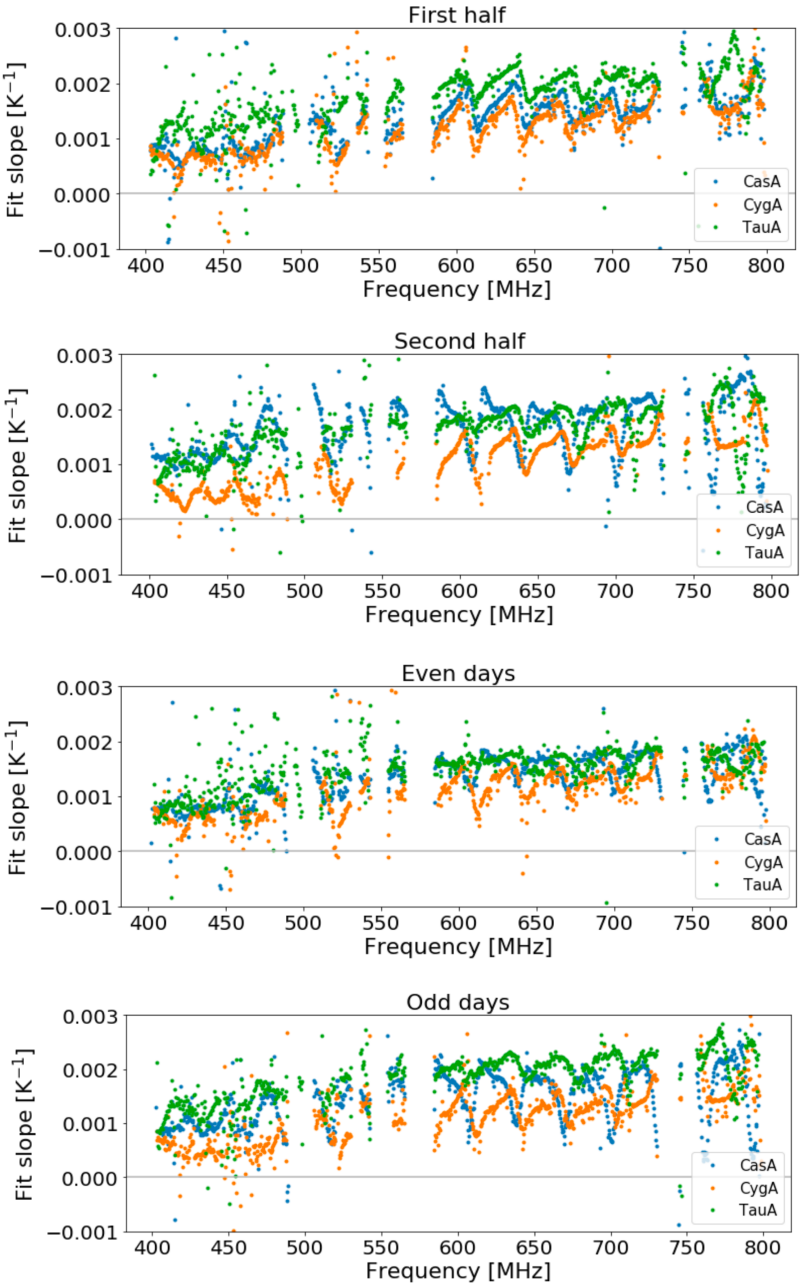

Figure 4.27: Susceptibilities obtained from only taking into account one half of each dataset, in different configurations. A dependency on the considered dataset is still present, but a lot less significant than before excluding outliers. The general increasing trend and the 30 MHz ripples are visible in each panel.

Chapter 5

Discussion

The results presented in this paper are illustrative of the validity of a linear thermal model, and also serve to understand the limitations of the employed methods. RMS of residuals values obtained after excluding outliers show how the thermal model can improve the data, and the target accuracy for the model being reached for a significant part of the frequency range is a very encouraging result. The final susceptibility plots demonstrate that the general increasing tendency and the order of the values are agreeing for all sources, independently of which part of the dataset is taken into account. It should therefore be possible to fit a smooth function to this general trend, and already significantly improve the calibration of the instrument. However, uncertainties are still present in all these measurements. Considering diverse sources or time periods results in differences that limitate the model. The 30 MHz oscillations linked to the expansion of the reflector introduce ripples which would also decrease the accuracy of a smooth fit through the susceptibility values. Therefore, taking into account these ripples by trying to model the expansion of the cylinders and the resulting shift in the interference pattern is an idea for further investigation. This could be facilitated by comparing current results with CHIME Pathfinder data in order to determine if a similar oscillation with lower amplitude was also occuring in the CHIME prototype. Other perspectives for future study include conducting an analysis similar to this work using the phase of the complex gain instead of its amplitude, and performing tests of the thermal model by applying it to simulated or actual sky maps and evaluating the resulting improvements.

Chapter 6

Conclusion

The objective of this study was to build a thermal model that would characterize the dependency of CHIME's complex gain on external temperature and be used to improve its gain calibration algorithm.

The results, obtained by analyzing transits of bright radiosources (CygA, CasA and later TauA), have shown that the complex gain amplitude variations can be satisfactorily described by a first order model combined with the appropriate data pre-processing.

After obtaining encouraging preliminary results despite significant scattering, different potential solutions were explored in order to reduce the impact of outliers as well as the dependency of the obtained susceptibilities on the considered source and dataset. The nonlinear response of LNAs was proposed as a first hypothesis and corrected for, but this modification only shifted the susceptibility values and did not improve the quality of the fits. External factors other than temperature were then considered; the SVD decomposition of the gain variation values demonstrated that a one-parameter model was a valid description of most of the data, but that an isolated group of outliers probably related to other factors such as rain needed to be explained, predicted and filtered out. Further investigation of outliers allowed to distinguish, detect and exclude two categories of anomalous transits, which greatly improved the quality of the linear fits and final susceptibility values and allowed to attain the target accuracy.

In the light of these results, we can consider the goals of this analysis to have

been met. Future work will be needed in order to include this model in the calibration algorithm and apply it to actual sky data, which will in time allow to perform innovative and possibly groundbreaking cosmological measurements, hopefully resulting in a better understanding of dark energy.

Bibliography

[1] CHIME. https://chime-experiment.ca/. Accessed on 2019-04-19.

[2] B. A. Bassett and R. Hlozek. Baryon acoustic oscillations. In P. Ruiz-Lapuente, editor, *Dark Energy*. Cambridge University Press, 2010. ISBN 9780521518888.

[3] M. R. Blanton et al. The galaxy luminosity function and luminosity density at redshift z=0.1. *The Astrophysical Journal*, 592:819–838, 2003.

[4] P. J. Boyle. First detection of fast radio bursts between 400 and 800 MHz by CHIME/FRB. http://www.astronomerstelegram.org/?read=11901, August 2018. Accessed on 2019-04-15.

[5] CHIME Collaboration. Calibrating CHIME, a new radio interferometer to probe dark energy. *The Astrophysical Journal*, submitted, 2018.

[6] CHIME/FRB Collaboration. The CHIME fast radio burst project: system overview. *The Astrophysical Journal*, submitted, 2018.

[7] M. Deng and D. Campbell-Wilson. The cloverleaf antenna: A compact wide-bandwidth dual-polarization feed for CHIME. *2014 16th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM), Victoria, BC*, pages 1–2, 2014.

[8] D. J. Eisenstein et al. Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *The Astrophysical Journal*, 633, 2005.

[9] A. Gomez-Valent. *Vacuum energy in quantum field theory and cosmology*. PhD thesis, University of Barcelona, 2017.

[10] G. Hinshaw et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208, 2013.

[11] A. Maeder. An alternative to the ΛCDM model: the case of scale invariance. *The Astrophysical Journal*, 834, 2016.

[12] C. Ng. Pulsar science with the CHIME telescope. *Proceedings of IAU Symposium No. 337*, submitted, 2017.

[13] S. Perlmutter, M. S. Turner, and M. White. Constraining dark energy with SNe Ia and large-scale structure. *Physical Review Letters*, 83:670–673, 1999.

[14] Planck Collaboration. Planck 2018 results. VI. cosmological parameters. *Astronomy & Astrophysics*, 2018.

[15] A. G. Riess. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astronomical Journal*, 116: 1009–1038, 1998.

[16] B. Ryden. *Introduction to Cosmology*. Cambridge University Press, 2 edition, 2017. ISBN 9781107154834.

[17] F. Sbisà. *Modified Theories of Gravity*. PhD thesis, University of Milan and University of Portsmouth, 2014.

[18] SDSS Collaboration. The fourteenth data release of the Sloan Digital Sky Survey. *The Astrophysical Journal Supplements*, forthcoming.

[19] J. A. Vazquez, L. E. Padilla, and T. Matos. Inflationary cosmology: from theory to observations. *Lecture notes in cosmology*, 2018.

[20] D. Wang and X. Meng. Observational constraints and diagnostics for time-dependent dark energy models. *Physics Letters B*, forthcoming.