UiO : University of Oslo

Håvard Tveit Ihle

Bayesian Data Analysis for Intensity Mapping and CMB Experiments

Thesis submitted for the degree of Philosophiae Doctor

Institute of Theoretical Astrophysics
Faculty of Mathematics and Natural Sciences

2021

Series of dissertations submitted to the

Faculty of Mathematics and Natural Sciences, University of Oslo

No. 2390

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

To Ludvig, for being a good boy

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* at the University of Oslo. The research presented here was conducted at the University of Oslo and, under the supervision of Ingunn K. Wehus and Hans Kristian K. Eriksen. This work was supported by the Norwegian Research Council through grant 251328.

Most of my work has been on the COMAP data analysis pipeline, including calibration, low level data processing, systematics, simulation, signal validation, data selection and power spectrum methods. Much of this work is presented in Paper I (Foss et al., in prep) which describes the data analysis pipeline all the way from raw data to final maps.

For the last year or so I have also worked on developing power spectrum methods that have a high sensitivity, yet are robust to instrumental systematics. This work is presented in Paper II (Ihle et al., in prep).

Papers I and II are both meant to be part of a set of papers to be published together making up the first release of COMAP results. As they are both in a fairly complete state we present them here as drafts, but we note that both may receive significant modifications before final submission.

I have also worked more abstractly on inference from line intensity maps: How do you get as much information possible out of an observed map? Paper III (Ihle et al., 2019) demonstrates how, for a COMAP-like experiment, using a combination of the power spectrum and the voxel intensity distribution gives significantly stronger constraints on the CO luminosity function than using either observable separately. Lately my focus has been more on how to bridge the gap between forecasts like this, with a bunch of simplifying assumptions, and real world data. As I discuss in the thesis, this gap needs to be closed for the work we do on inference methods to be useful.

For the last two years or so I have also spent some of my time working on the BeyondPlanck project (Paper IV, BeyondPlanck Collaboration, 2020). My main focus starting out was on implementing the methods for sampling the correlated noise and the corresponding noise parameters. As the work evolved we encountered some interesting systematics which seem to involve both the noise estimation and calibration, which makes sense since they are so closely related. This work involved using housekeeping data from the Planck satellite to interpret and understand the results. Our main findings and more discussion can be found in Papers V and VI (Ihle et al., 2020, Gjerløw et al., 2020).

Acknowledgements

I want to thank my supervisors **Ingunn** and **Hans Kristian** for believing in me, for giving me the chance to develop as a researcher, and for all the help you have provided. Thanks for suggesting these projects for me, I have learned so much during these years.

Thank you to **Marie**, it's been a pleasure working with you on COMAP for the last four and a half years.

Thanks to my master's students, **Maren**, **Jonas**, **Jowita** and **Nils**, you are all brilliant, and it has been really fun supervising and working with you!

Thanks to everyone I have worked with in COMAP, **Kieran**, **Tim**, **David**, **James**, **Rick**, **Stuart**, **Bade**, **Junhan**, **Liju**, **Dongwoo**, **Marco**, **George**, **Patrick**, **Marta** and **Hamsa**, it has been so interesting to meet and work with great people like you from all over the world!

Thanks to all the people I have met here at ITA over these many years, both in the basement, at the first floor and at the institute in general. I really enjoy it at ITA and you people are what makes it what it is.

Thanks to **Frode**, for for giving me the chance to lecture the Cosmology II course. That was a really cool and inspiring experience, and I learned so much!

To **Caroline**, my love, you are kind, beautiful and wise! Thank you for putting up with me, and for all your love and support.

Thank you also to my family, **Mor**, **Far**, **Aslak**, **Maria** and **Bjarte**, for the unconditional love that you provide!

Håvard Tveit Ihle
Oslo, March 2021

List of Papers

Paper I

M. Foss, H. T. Ihle and the COMAP collaboration. “First Season COMAP Results: CO Data Processing” (draft).

Paper II

H. T. Ihle and the COMAP collaboration. “First Season COMAP Results: Power spectrum methodology and preliminary data quality assessment” (draft).

Paper III

H. T. Ihle, D. Chung, G. Stein and the COMAP collaboration. “Joint Power Spectrum and Voxel Intensity Distribution Forecast on the CO Luminosity Function with COMAP”. In: ApJ, Jan 2019, volume 871, no. 1 p. 75. DOI: 10.3847/1538-4357/aaf4bc.

Paper IV

The BeyondPlanck collaboration. “BeyondPlanck I. Global Bayesian analysis of the Planck Low Frequency Instrument data”. In: A&A, to be submitted, arXiv: 2011.05609.

Paper V

H. T. Ihle, M. Bersanelli, C. Franceschet, E. Gjerløw and the BeyondPlanck collaboration. “BeyondPlanck VI. Noise characterization and modelling”. In: A&A, to be submitted, arXiv: 2011.06650.

Paper VI

E. Gjerløw, H. T. Ihle, S. Galeotta and the BeyondPlanck collaboration. “BeyondPlanck VII. Bayesian estimation of gain and absolute calibration for CMB experiments”. In: A&A, to be submitted, arXiv: 2011.08082.

Contents

Chapter 1

Cosmology

Cosmology is the study of the universe as a whole. What are the physical properties of the universe? What is the history and future of the universe? What does the universe consist of?

Although we have had a fairly good qualitative understanding of some parts of these questions for about a century, it is really first in the last few decades that we have been able answer any part of these questions with any kind of certainty or precision.

This development has mostly been led by experimental advances in the study of the cosmic microwave background (CMB) and the large scale structure of the universe.

1.1 Standard Cosmology

Here we will discuss some of the main theoretical components underlying modern cosmology, including general relativity (GR), dark matter (DM), dark energy (DE) and inflation, culminating in the ΛCDM model.

1.1.1 General Relativity

GR is an incredibly successful theory of gravity describing everything from gravitational phenomena measured in laboratories here on earth to the expansion and evolution of the universe as a whole. The fundamental equation of GR is the *Einstein equation*

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where $G_{\mu\nu}$ is the *Einstein tensor* and $T_{\mu\nu}$ is the *energy-momentum tensor*. The Einstein equation relates the curvature of the universe, represented by $G_{\mu\nu}$, to the matter and energy content of the universe, represented by $T_{\mu\nu}$. This means that any distribution of matter and energy will create curvature in the spacetime around it, much like any distribution of electric charge will create an electromagnetic field.

The ΛCDM model assumes that the spatial part of the universe is homogeneous, isotropic and flat, and that the energy content of the universe is given by radiation, baryonic matter, cold dark matter (CDM) and DE. In this case the Einstein equation can be reduced to

$$H^2(t) = H_0^2\left(\frac{\Omega_\mathrm{r}}{a^4} + \frac{\Omega_\mathrm{b} + \Omega_\mathrm{c}}{a^3} + \Omega_\Lambda\right), \tag{1.2}$$

where $a(t)$ is the scale factor, which keeps track of the size of the universe at any given time, $H(t) \equiv \frac{\dot{a}(t)}{a(t)}$ is the Hubble parameter denoting the expansion rate of the universe, $\dot{a}(t)$ denotes the time derivative of $a(t)$, H_0 is the current value of the Hubble rate, often called the *Hubble constant* and $\Omega_{\mathrm{r,b,c},\Lambda}$ denotes the fractional energy density today in radiation, baryonic matter, CDM and DE respectively.

1.1.2 Dark Matter

Dark matter (DM) refers to a matter component additional to the visible matter we can see and interact with. The term was first used by Fritz Zwicky (Zwicky, 1933) in 1933 when he found larger velocity dispersion in the Coma cluster than could be explained by the visible matter, indicating another "dark" matter component.

Today DM is a cornerstone of cosmology with multiple independent lines of evidence in its favor, including rotation curves of spinning galaxies (Borriello and Salucci, 2001), velocity dispersion in galaxy clusters (Saro et al., 2013), weak (van Uitert et al., 2012) and strong (Moustakas and Metcalf, 2003) gravitational lensing, the CMB and non-linear structure formation (Planck Collaboration VI, 2020).

These lines of evidence, as well as null results in direct- and indirect detection experiments, put strong constraints on the properties of dark matter:

- It must be *dark* (i.e. non-luminous): In practice this means no coupling (or extremely weak) to the photon. In general DM can not have any large coupling to any light standard model particle.
- It must be non-interacting: The self-interaction of dark matter is severely constrained, especially on large scales.
- It must be cold: DM has to be non-relativistic during structure formation.
- It must be stable: It must have a lifetime significantly longer than the age of the universe.

It is remarkable that simply positing a new heavy non-SM particle with small or no interactions with the visible sector can explain such a breadth of independent observations.

It is interesting to note, however, that in all the stated cases, DM is inferred from its gravitational effects. This has led some to suggest a modification to gravity to explain these phenomena (see e.g. Milgrom, 2010). It has, however, not been possible to devise a modification of gravity that can explain more than a few of the above lines of evidence for DM at a time.

1.1.3 Dark Energy

In 1998 supernova measurements suggested, for the first time, that the expansion of the universe was accelerating (Riess et al., 1998). This acceleration cannot be achieved by ordinary matter, but requires a cosmological component with positive energy and negative pressure. This new component is what is called *Dark Energy*. Since then multiple other lines of evidence have confirmed this accelerated expansion (Huterer and Shafer, 2018) and we now know that DE is actually the largest contribution to the energy budget of the universe at around 70%.

We do not know the nature of DE, but the simplest explanation is that empty space has a certain energy associated with it, a *vacuum energy*. This constant energy density has exactly the properties we need to explain the accelerated expansion we observe. However, the big open question is: Why does the vacuum energy take this

specific value, and not another? Or, more specifically: Why is the energy density of the vacuum of comparable magnitude as the energy density of matter? This is called the *Coincidence Problem*, and we do not have any satisfactory answers to it. Perhaps even more disconcertingly, if we try to estimate the contributions to the vacuum energy from quantum field theory we naively expect contributions of the order of the Planck scale, which is about 120 orders of magnitude off from the value we measure in cosmology. Clearly there is a significant piece missing in our understanding of DE, and it remains one of the major challenges for the future.

1.1.4 Inflation

Inflation is the theory that the universe went through a period of accelerated expansion in the very early universe. During inflation, the universe expanded by at least a factor of 10^{26}. This accelerated expansion solves two problems (and other related problems) with the traditional big bang model, *the flatness problem* and *the horizon problem* (Guth, 1981).

If the universe is dominated by matter or radiation, which it was during most of its history, any initial spatial curvature would tend to increase as the universe expands. This means that since we observe the universe to be very close to flat today, it must have been extremely close to flat in the very early universe, which would require a large degree of fine-tuning, unless we have some mechanism for making space flat to begin with. This is called the flatness problem, and inflation solves it since during inflation, the universe becomes more and more flat as the expansion goes on.

The horizon problem can be understood by observing that the CMB photons reaching us from opposite points of the sky have the same temperature, even though these points in space have never, in the traditional big bang model, been in causal contact and so have no reason to be at the same temperature. This is called the horizon problem, and inflation solves this by ensuring that regions that were in causal contact before inflation got spread out to super-horizon scales. The effect of this is that even regions which (if we did not know about inflation) we would think never had been in causal contact, actually could have had time to reach thermal equilibrium.

The simplest model of inflation posits a single scalar field, $\phi(\boldsymbol{x}, t)$, rolling slowly down towards the minimum of its potential. In such a *slow-roll* scenario, the energy of the scalar field is dominated by the potential term that acts almost like a vacuum energy and leads to exponential expansion.

Another crucial effect of the accelerated expansion is that small scale quantum fluctuations in the scalar field get blown up to large scales and then frozen in as these scales move outside the cosmological horizon. After inflation, when these modes eventually move into the horizon again, these initial perturbations act as seeds for the subsequent growth of structure in the universe.

Single field slow-roll inflation makes three predictions for the generated perturbations, they will be:

- Gaussian: Since the quantum state of the perturbation field is basically in the vacuum state of an harmonic oscillator.

- (Nearly) Scale invariant: Since the conditions under which the fluctuations were generated changes only very slowly, the properties of the generated perturbations will only depend very weakly on scale.
- Adiabatic: Since the perturbation in all the different cosmological components are generated from the same scalar perturbations, they will be in phase with each other.

Observations of the CMB confirms all these predictions, which is why inflation is generally accepted, even though we do not have very direct evidence for it. On the other hand, one can argue that in each of these three cases, if you had asked the question "What is the simplest thing you would expect?" the answer would be adiabatic Gaussian fields with a scale invariant spectrum. In this sense we can say that these are very generic predictions, and thus less impressive. Nevertheless inflation is still the simplest theory on the market to explain these observations (as well as solving the horizon problem etc.).

Another important prediction of inflation is the presence of tensor perturbations in addition to the scalar perturbations we have been discussing. Tensor perturbations are primordial gravitational waves induced by the inflation field. Tensor modes are harder to detect than scalar modes, but they can be observed because they induce so-called B-modes in the polarization spectrum of the CMB.

The scalar and tensor power spectra are conventionally defined as follows (Baumann, 2018)

$$P_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} A_{\mathrm{s}} \left(\frac{k}{k_*}\right)^{n_{\mathrm{s}}-1}, \tag{1.3}$$

$$P_h(k) = \frac{k^3}{2\pi^2} A_{\mathrm{t}} \left(\frac{k}{k_*}\right)^{n_{\mathrm{t}}}, \tag{1.4}$$

where $\mathcal{R}$ is the scalar curvature perturbation, h is the tensor perturbation, A_{s} and A_{t} are the scalar and tensor perturbation amplitudes, respectively. n_{s} and n_{t} are the scalar and tensor spectral indices, respectively. The amplitudes are conventionally defined at $k_* = 0.05\ \mathrm{Mpc}^{-1}$ (Planck Collaboration VI, 2020).

We then define the tensor-to-scalar ratio, r, as

$$r \equiv \frac{A_{\mathrm{t}}}{A_{\mathrm{s}}}. \tag{1.5}$$

Inflation does not predict the overall level of the tensor perturbations, but it predicts the following relation $n_{\mathrm{t}} = -r/8$. We have not observed any cosmological tensor modes, and the current observational constraints on the tensor-to-scalar ratio is given by $r < 0.044$ (95 % CL) (Tristram, M. et al., 2021). Therefore, since we already know that r is fairly small, this means that in practice, since it will be very hard to measure, inflation predicts $n_{\mathrm{t}} \approx 0$.

Measuring the tensor-to-scalar ratio, r, is one of the main goals of future CMB experiments. This would not only be more evidence for inflation (especially if we could confirm that the tensor spectral index is consistent with inflation), but it would also tell us (at least in the simplest inflation models) about the energy scale of inflation, and about the shape of the inflationary potential.

1.1.5 The ΛCDM Model

The ΛCDM model combines the theories of GR and inflation[1] with DE and (Cold) Dark Matter (CDM), as well as known particle physics and thermodynamics, into a single 6 parameter model for explaining the expansion and evolution of the universe, in addition to the origin and growth of structures in the universe.

The model assumes that the universe is spatially flat, that the energy density of dark energy is constant over time and that there are no primordial tensor perturbations. Table 1.1 shows the constraints on the six parameter model from the latest Planck release. We see that all these parameters, except for the optical depth to reionization, τ, are constrained to more than one percent accuracy, which is quite remarkable.

Table 1.1: Constraints on the 6 ΛCDM parameters based on Planck data, in addition to lensing and baryon acoustic oscillation data (Planck Collaboration VI, 2020). Here $h \equiv H_0 \,/\, 100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, θ_* is the angular acoustic scale and τ is the optical depth to reionization.

Parameter	68% limits
$\Omega_\mathrm{b}h^2$	0.02242 ± 0.00014
$\Omega_\mathrm{c}h^2$	0.11933 ± 0.00091
$100\theta_*$	1.04119 ± 0.00029
τ	0.0561 ± 0.0071
$\ln\left(10^{10}A_\mathrm{s}\right)$	3.047 ± 0.014
n_s	0.9665 ± 0.0038

1.2 Observational Cosmology

Observational cosmology is a wide ranging field with many different sources of data, including galaxy surveys, observations of supernovae, gravitational lensing, Lyman-α forest, CMB, line intensity mapping (LIM) and gravitational waves. Here we will focus on CMB and LIM, as those are the topics of this thesis.

1.2.1 CMB observations

The cosmic microwave background (CMB) is the radiation left over from the hot big bang. It essentially captures a snapshot of the distribution of structure in the very early universe, only about 400 thousand years after the big bang. This radiation has been a treasure trove of information about our universe and has brought us into the era of precision cosmology.

Planck was the fourth satellite mission to study the CMB, following the *RELIKT* (Klypin et al., 1987), *COBE* (Mather et al., 1994) and *WMAP* (Bennett et al., 2013) satellites. Planck produced the ultimate map of the temperature anisotropies of the CMB,

[1] The ΛCDM does not explicitly include inflation, but assumes that the initial power spectrum is of the form given in Eq. 1.3 and that $r = 0$.

limited only by cosmic variance and fundamental physical processes. This has made Planck one of the most important datasets for constraining cosmological parameters, further cementing the ΛCDM model as the standard model for cosmology. With the temperature map all but settled by Planck, the future of CMB lies in the polarization signal.

The success of the CMB field comes from the combination of precise observations and precise theoretical predictions. The early universe was incredibly homogeneous, allowing the use of linear perturbation theory to study the growth of perturbations, which means that the coupled Einstein-Boltzmann equations can be solved very precisely (see Dodelson, 2003 for more details). Figure 1.1 shows the close agreement between theoretical predictions and CMB observations.

1.2.2 Line intensity mapping

Line intensity mapping (Madau, Meiksin, and Rees, 1997) (LIM) aims to map out the 3D structure of the early universe in large volumes using various different spectral lines, like 21 cm, Lyα, [CII], or CO. Compared to galaxy surveys, these surveys do not rely on resolving and detecting individual sources of radiation, but instead collect the aggregate emission from the full population of sources from each cross-sectional area of the sky (see Fig. 1.2). By extracting the information present in the 3D maps LIM aims to help us understand both galaxy evolution, reionization and fundamental cosmology.

Intensity mapping is now a growing field with multiple efforts both experimental and theoretical. For a nice review of the motivation for LIM see (E. D. Kovetz et al., 2017; E. Kovetz et al., 2019). The number of running and planned LIM experiments is growing, including SKA (Koopmans et al., 2015; Square Kilometre Array Cosmology Science Working Group et al., 2020), GBT (Masui et al., 2013), BINGO (Battye et al., 2013), HIRAX (Newburgh et al., 2016), LOFAR (van Haarlem et al., 2013), CHIME (Bandura et al., 2014), HERA (DeBoer et al., 2017), GMRT (Pen et al., 2009), MWA (Tingay et al., 2013), PAPER (Parsons et al., 2014) and MeerKAT (Santos et al., 2017) targeting 21 cm, COMAP (Cleary et al., 2016), COPSS (Keating, Bower, et al., 2015; Keating, Marrone, Bower, Leitch, et al., 2016), mmIME (Keating, Marrone, Bower, and Keenan, 2020) and AIM-CO (Bower et al., 2016) targeting CO, CCART-prime (Stacey et al., 2018), TIME (Crites et al., 2014) and CONCERTO (Lagache, Cousin, and Chatzikos, 2018) targeting [CII] and CDIM Cooray et al., 2016, SPHEREx (Doré et al., 2014), Origins (The OST mission concept study team, 2018) and HETDEX (Hill et al., 2008) targeting other lines.

A simple analytic model for the signal power spectrum, $P(k)$, of a line intensity map is given by

$$P(k) = \langle T_{\text{line}} \rangle^2 b_{\text{line}}^2 P_m(k) + P_{\text{shot}}, \tag{1.6}$$

where $\langle T_{\text{line}} \rangle$ is the mean brightness temperature of the line, b_{line} is the luminosity weighted bias of the line, $P_m(k)$ is the matter power spectrum and P_{shot} is the scale invariant shot noise power spectrum. Together the three factors of the first term are referred to as the *clustering* power spectrum as it follows the underlying clustering of the matter distribution. The clustering term dominates at large scales, while the shot noise term dominates at small scales (see Fig. 1.3).

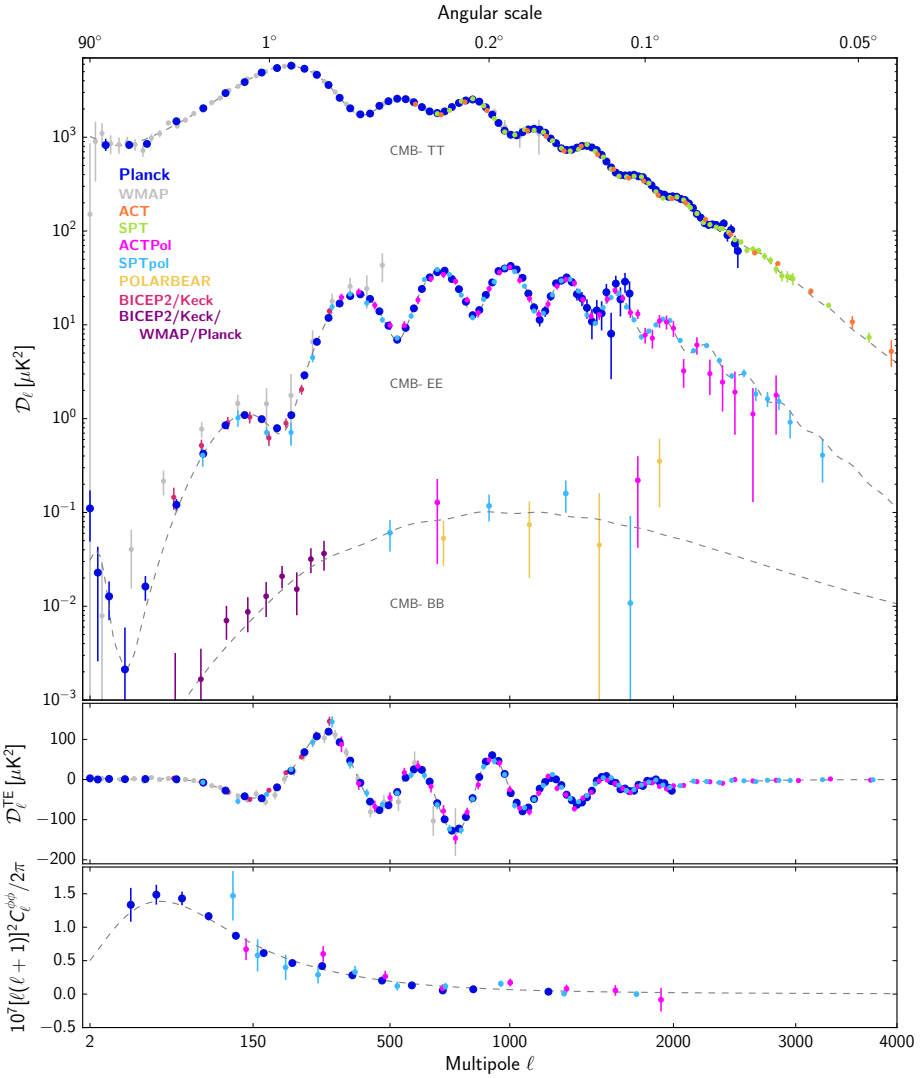

Figure 1.1: Comparison of CMB power spectrum measurements and theoretical predictions from the ΛCDM model. Top panel: Temperature (TT), E-mode (EE) and B-mode (BB) polarization auto spectra measured by several different CMB experiments. Middle panel: Cross-spectrum between temperature and E-mode polarization maps. Bottom panel: Lensing deflection power spectrum. Note that the B-mode spectrum measured here is consistent with that induced by lensing, and shows no sign of cosmological tensor modes. For more details see Planck Collaboration et al., 2020, from where this figure was taken.

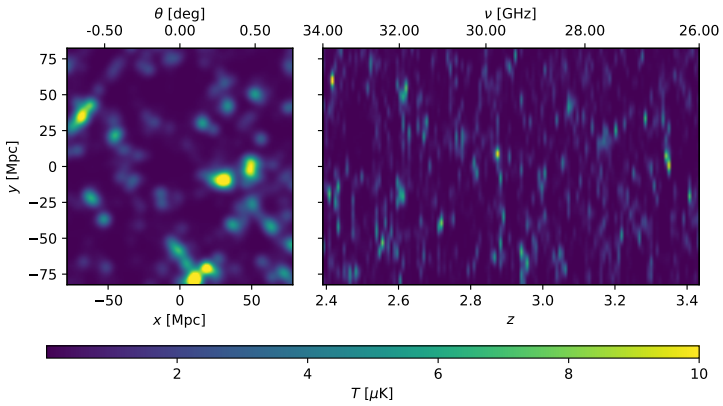

Figure 1.2: Simulated 3D cube of CO line emission brightness temperature. This corresponds roughly to the survey volume of one of the COMAP fields.

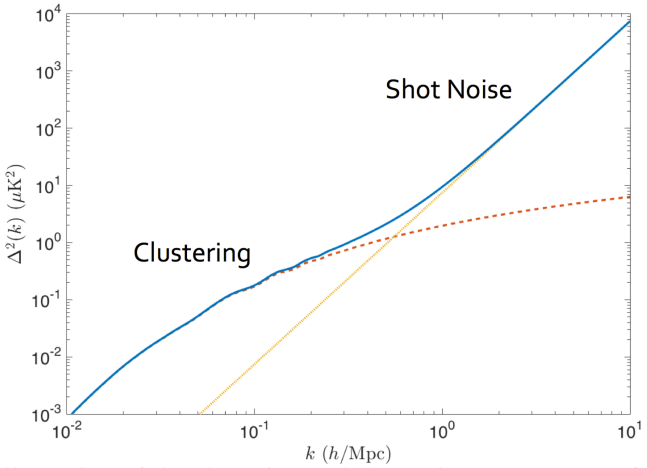

Figure 1.3: Illustration of the clustering and shot noise components of the intensity mapping power spectrum. Figure taken from E. D. Kovetz et al. (2017).

By tracing spectral lines originating in star forming regions, the interstellar medium and the intergalactic medium, LIM can map out galaxy formation and evolution over cosmic time, and, crucially, follow the process of reionization by tracing both the sources of ionizing radiation, and the neutral hydrogen gas being gradually ionized. These data will, among other things, help constrain the optical depth to reionization, τ, which is the least constrained parameter in the ΛCDM model.

In addition to these unique insights into astrophysics, LIM also provides a great opportunity to test fundamental cosmology. As we can see from Eq. 1.6, LIM traces the underlying matter distribution, and as such, it gives us access to new cosmological volumes largely inaccessible to other probes. Where CMB observations are limited to the number of 2D angular modes from the surface of last scattering, LIM can give us access to the full 3D modes up to high redshifts.

This could allow us, as an example, to follow the baryon acoustic oscillation (BAO) scale through cosmic time, which would give us a robust, model independent, measurement of the cosmic expansion history (Bernal, P. C. Breysse, and E. D. Kovetz, 2019). This would be a powerful tool to test things like modified gravity and dark energy properties. We could also get much tighter constraints on primordial non-Gaussianity (e.g. Furlanetto et al., 2019), which could help us distinguish different models of inflation. Even dark matter decay or annihilation could be detected using LIM (Creque-Sarbinowski and Kamionkowski, 2018). These are just some of the use cases of LIM in cosmology currently being discussed. As the field develops, many more will surely emerge.

1.3 Unresolved Questions and Future Prospects

The standard model of particle physics has been incredibly successful, explaining all particle interactions that we have observed, with the detection of the Higgs boson at the large hadron collider (LHC) in 2012 putting the last puzzle piece into place (Chatrchyan et al., 2012; Aad et al., 2012).

Motivated by the WIMP miracle and the hierarchy problem (Jungman, Kamionkowski, and Griest, 1996; Susskind, 1979), there was a significant expectation that the LHC would detect supersymmetry (SUSY), or other new beyond standard model (BSM) physics, at the weak scale. No such signs have yet been seen. While low energy SUSY is still not ruled out, the fact that we have not seen any signs of it at least lets us consider the possibility that we will not make much progress towards BSM physics using particle colliders in the near future.

From cosmology, however, we know that 95 % of the energy content of the universe (DE and DM) comes from sources that are not part of the standard model of particle physics. Inflation is also an important process in cosmology that cannot be explained within the standard model.

This suggests that at this point cosmology might be the most likely way to move fundamental physics forwards. What is the nature and properties of DM and DE? Are either of these explained by modifications of gravity? Did inflation happen? If so, what is the source of inflation? If not, then what explains the flatness and horizon

problem? And what gives us the spectrum of initial perturbations? How does any of this fit together in a coherent picture with the standard model?

With new probes like line intensity mapping coming online, in addition to well established ones like CMB analysis and galaxy surveys, observational cosmology and astrophysics is now in a position to shed light on these questions. Detecting primordial gravitational waves, non-Gaussianities or large scale anisotropies would shed light on inflation or any other source of primordial perturbations. Mapping out the whole expansion history of the universe, and the growth rate of structure, could help us nail down the properties of DE, possibly modified gravity, or even new early universe physics. Detecting the small scale structure of galaxies can tell us about the properties and interactions of DM. All this makes observational cosmology an exciting area to work in.

Chapter 2

COMAP

2.1 COMAP Experiment

The CO molecule is the second most abundant molecule in the interstellar medium, after H_2, and is a great tracer of cold molecular clouds, and thus star formation. CO has a ladder of equally spaced rotational transitions at integer multiples 115.27 GHz, which opens the possibility of studying different transitions of the same CO sources by observing at different frequencies.

The CO Mapping Array Pathfinder (COMAP, Cleary et al., 2016) is a line intensity mapping experiment targeting the rotational transitions of the CO molecule. It is an international collaboration between California Institute of Technology, Canadian Institute of Theoretical Astrophysics, Jet Propulsion Laboratory (NASA), University of Manchester, Maryland University, Miami University, Stanford University, University of California Berkeley, Princeton University and University of Oslo. The Oslo team is responsible for the end-to-end data analysis of the main science observations from COMAP.

Phase One of COMAP consists of a single telescope observing in the K_a-band at 26-34 GHz, which for CO $1 \rightarrow 0$, emitted at 115 GHz, corresponds roughly to the redshift range 2.4 – 3.4, the Epoch of Galaxy Assembly. The K_a-band will also pick up a, presumably weaker, signal from the CO $2 \rightarrow 1$ transition, emitted at 230 GHz, or the redshift range 6 – 8, during the Epoch of Reionization. See Fig. 2.1 for a group photo of the COMAP collaboration in front of the COMAP telescope.

COMAP Phase One is a pathfinder experiment meant to demonstrate the feasibility of large scale line intensity mapping using CO. Future planned phases of COMAP will involve more telescopes observing at K_a, as well as the addition of one or several telescopes observing the K_u-band. The K_u-band, at 12-20 GHz, would pick up the $1 \rightarrow 0$ emission from the same redshifts as the $2 \rightarrow 1$ emission from the K_a-band. This will allow us to separate out the $2 \rightarrow 1$ from the $1 \rightarrow 0$ signal in the K_a data. The long term goal is to do CO line intensity mapping from space (Bowman et al., 2012; M. B. Silva et al., 2019).

We started the main science observations for Phase One in June of 2019, and are currently analyzing the data taken during the first year of observations.

2.2 COMAP Instrument

For COMAP Phase One we utilize a 10.4 m telescope located at the Owens Valley Radio Observatory (OVRO) in California. The telescope was part of an array of dishes previously used for the Combined Array for Research in Millimeter-wave Astronomy (CARMA) experiment, but has now been moved back to OVRO to be used for other experiments. The angular resolution obtained is roughly 4.5 arcmin FWHM at 30 GHz.

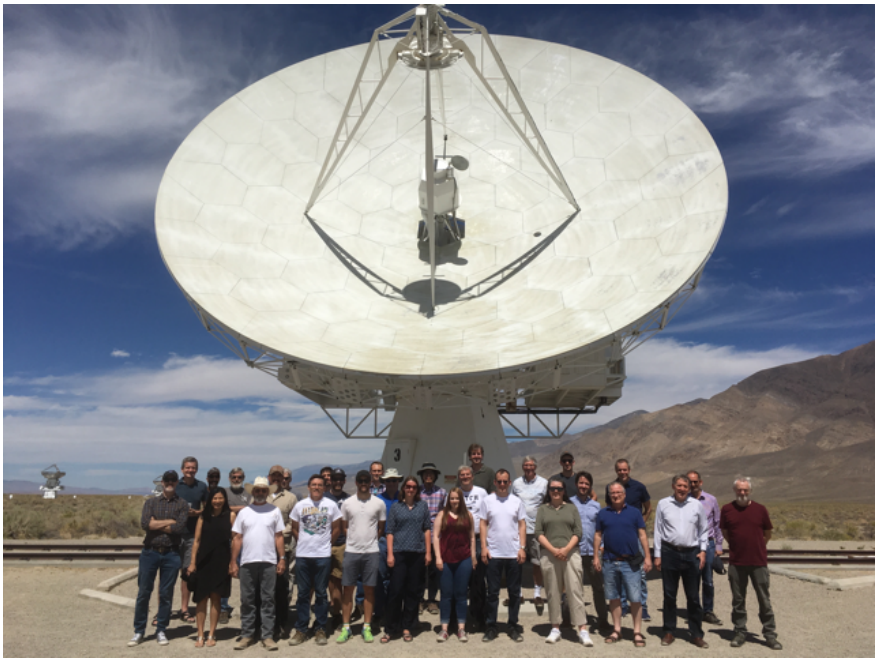

Figure 2.1: The COMAP collaboration standing in front of the COMAP Phase One telescope in August 2018.

The receiver consists of 19 independent detectors, each with a separate signal path from the 19 individual feed horns, through polarizers and low noise amplifiers to downconversion and eventually frequency separation and digitization which happens in two CASPER "Roach2" FPGA-based spectrometers for each signal chain (Cleary, 2015). We usually use the term "feed" to refer to each of these 19 signal chains. For more details on the COMAP instrument see J. Lamb et al. (in prep).

The observed signal in the frequency range 26-34 GHz gets, during the downconversion, split into two different bands, band A coming from 26-30 GHz and band B from 30-34 GHz. One Roach2 module processes the signal from one band, meaning that a total of 38 Roach2 modules are used to collect the full science data from the 19 feeds. In the Roach2 modules the data gets further split into a lower and an upper sideband, each with a bandwidth of 2 GHz.

For the raw science data, the Roach2 modules separate the 2 GHz bandwidth of each sideband into 1024 separate frequency channels, leading to a native frequency resolution of about 2 MHz. The data is also integrated over a period of about 0.02 s, leading to a sample rate of roughly 50 Hz in time.

2.3 Observations

During observations we employ two different scanning strategies, Constant Elevation Scans (CES) and Lissajous scans.

For the CES observations we keep the telescope at a constant elevation while slewing it back and forth in azimuth. During a Lissajous scan the telescope performs independent sinusoidal motion in both elevation and azimuth at the same time as follows

$$\text{Azimuth}(t) = A\sin(at + \phi) \tag{2.1}$$

$$\text{Elevation}(t) = B\sin(bt). \tag{2.2}$$

A Lissajous scan can thus create many different scanning patterns depending on which values used for A, B, a, b and ϕ. We typically use values of $B \approx 0.6$ degrees, while A is scaled with elevation to correspond to roughly the same on-sky distance. The other parameters are randomized in order to not repeat the same scanning pattern each time. Typical values for a and b are in the range 0.3–0.5 Hz, leading to typical periods of 12–25 seconds.

Circular scans, which are sometimes used, are just a special case of the Lissajous scan. In fact, even for the CES observations, we use a sinusoidal motion in the azimutal direction in order to minimize the strain on the telescope, so the CES can also be thought of as a special case of the Lissajous scan (with $a = 0$).

Lissajous scans typically give a more even coverage of the observed field, with better cross-linking, but they have the disadvantage of getting a large signal from the different amount of atmosphere at the different elevations, which needs to be removed. Ground signal is also harder to remove from Lissajous scans, since the scanning pattern is not repeating, in the same way as it is for a CES. A Lissajous scan typically covers a larger area than a CES, so it can recover more of the large scale cosmological signal, at the cost of a higher noise level on small scales.

For the first year of observations, we split the observation time roughly evenly between Lissajous and CES. However, due to the problems we have found in the Lissajous data (see Paper II, Ihle et al., in prep), the observations taken afterwards have been exclusively CES, and unless we can figure out and solve the issues with the Lissajous data, we will continue to rely exclusively on CES observations for the future.

The general observation strategy is to start by pointing the telescope a bit ahead of the on-sky field that you want to observe (i.e. point to where the centre of the field will be in a few minutes), and then performing CES or Lissajous scans centered on this azimuth and elevation position while the field drifts through. Once the field has drifted past the azimuth and elevation you are observing at, you repeat the sequence by moving slightly ahead of the field again etc. Each such period where you stay centered at at a given azimuth and elevation performing CES or Lissajous motion while the field drifts through is called a "scan". When you move the telescope ahead of the field again and repeat the sequence you start a new scan. Figure 2.2 shows the telescope pointing for 27 subsequent scans. This illustrates the general observation strategy, and how the telescope follows the field across the sky. We can also see the randomization of the Lissajous parameters clearly.

A scan typically lasts for 3-10 minutes, and multiple scans of the same field are combined together into an observation. An observation usually contains around an hour

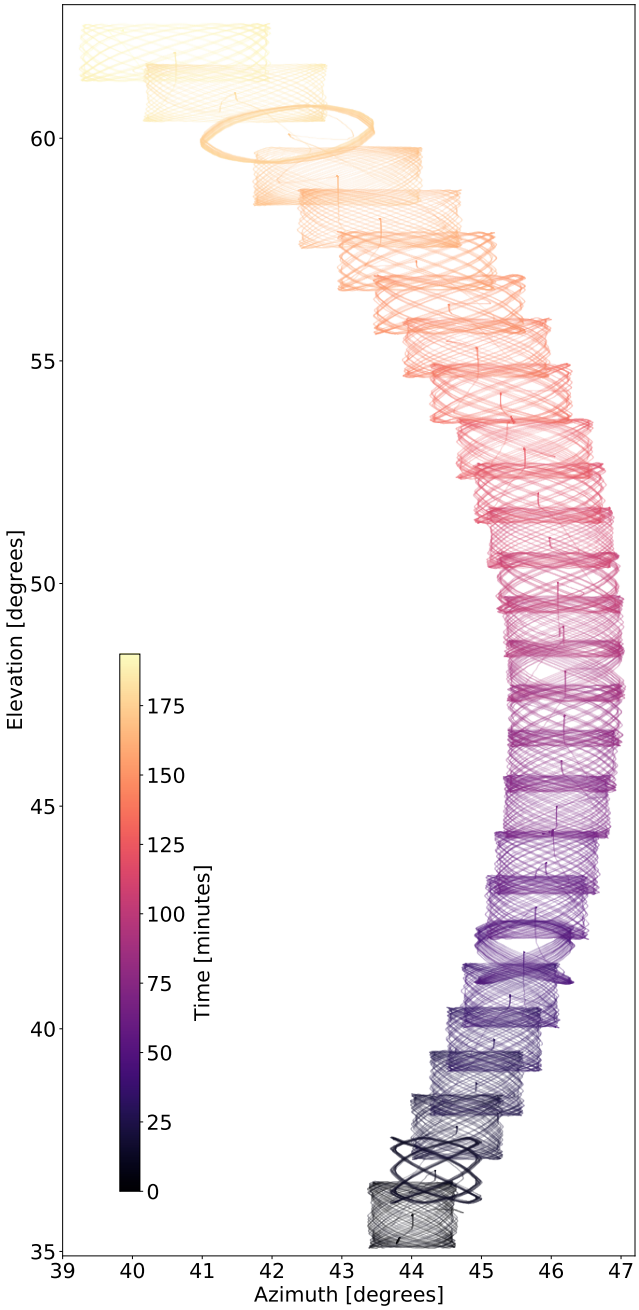

Figure 2.2: Telescope pointing during 27 subsequent Lissajous scans of the same field. Figure courtesy of Jonas Lunde.

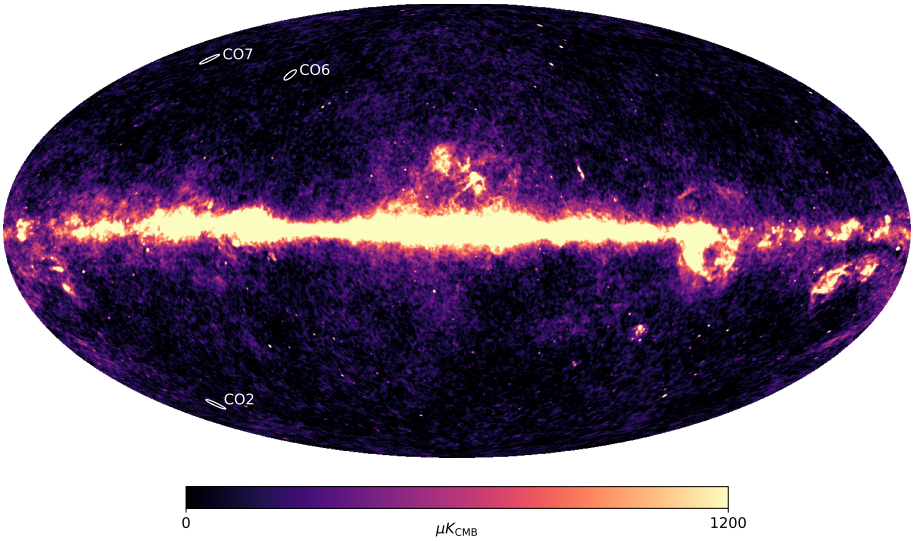

Figure 2.3: The three main CO fields, overplotted on the Planck LFI 30 GHz map. Map downloaded from the PLA, https://pla.esac.esa.int (Planck Collaboration I, 2020). Figure courtesy of Nils O. Stutzer.

Field Name	RA (J2000)	Dec (J2000)
CO2	01:41:44.4	+00:00:00.0
CO6	15:04:00.0	+55:00:00.0
CO7	11:20:00.0	+52:30:00.0

Table 2.1: COMAP main science fields

of data, and is labeled by a, monotonically increasing, integer number called the "obsid" (we refer to the observations as e.g. obsid 7456 etc.). The raw data is collected together into what we call "level 1" files, which contain all the data, from all 19 feeds, for a single obsid, together with pointing information and a large set of housekeeping data.

We observe three main science fields, labeled "co2", "co6" and "co7". These are chosen to be in low foreground regions, and to overlap with other surveys. The position of these fields can be seen in Fig. 2.3 and in Table 2.1. The sizes of the fields are roughly 2 x 2 deg^2.

2.4 Data and Systematics

The basic properties of the raw data are discussed in Paper I (Foss et al., in prep). Here we give a short summary and focus on some of the main systematics present in the data and what we know about them.

2.4.1 Data model

The power picked up by any of the detectors is given by

$$P_{\text{out}} = k_{\text{B}} G \Delta\nu T_{\text{sys}}, \tag{2.3}$$

where G is the instrumental gain, k_{B} is the Boltzmann constant, $\Delta\nu$ is the frequency bandwidth, and T_{sys} is the total system temperature of the instrument.

The system temperature is basically a measure of the overall noise level, and it has many different sources,

$$\begin{aligned} T_{\text{sys}} = {} & T_{\text{receiver}} + T_{\text{atmosphere}} + T_{\text{ground}} \\ & + T_{\text{CMB}} + T_{\text{foregrounds}} + T_{\text{CO}}, \end{aligned} \tag{2.4}$$

where T_{receiver} is the effective noise temperature of the receiver, $T_{\text{atmosphere}}$ is the temperature contribution from the atmosphere, T_{ground} is the ground pickup from far sidelobes, T_{CMB} is the signal from the CMB, $T_{\text{foregrounds}}$ are all continuum foregrounds (mostly from the galaxy), and T_{CO} is the line emission signal from extragalactic CO, which is what we are ultimately trying to measure.

Making some simple assumptions we can make the following model for the system temperature as a function of feed, frequency and time,

$$T^i_{\text{sys},\nu}(t) = \langle T^i_{\text{sys},\nu} \rangle \left[1 + P^i_{\text{cel}} (\Delta s_{\text{cont}} + \Delta s^\nu_{\text{CO}}) + P^i_{\text{tel}} \Delta s_{\text{ground}} + n_{\text{corr}} + n^{\nu i}_{\text{w}} \right]. \tag{2.5}$$

Here $\langle\rangle$ denotes average in time, P^i_{cel} and P^i_{tel} are the pointing matrices in celestial and telescope coordinate systems, respectively; Δs_{cont} denotes the mean subtracted celestial continuum sources, like the CMB or galactic foregrounds; Δs^ν_{CO} is the mean subtracted CO line emission; Δs_{ground} is the mean subtracted ground signal picked up by the far sidelobes; and n_{corr} is the (zero mean) correlated noise component, consisting mostly of atmosphere fluctuations and standing waves. The superscript i is a feed index. Terms with no feed index are assumed to be similar (or at least strongly correlated) between different feeds, while terms with a ν label indicate parts of the model that are assumed to have non-smooth frequency dependence.

Likewise we can make the following simple model for the gain

$$G^i_\nu(t) = \langle G^i_\nu \rangle \left[1 + \delta^i_G(t) \right], \tag{2.6}$$

where $\delta^i_G(t)$ is a correlated noise term coming from gain fluctuations in the low noise amplifiers. These gain fluctuations are assumed to be the same for all frequencies of a single feed, but are independent between different feeds.

We can then combine these into a single model for the power measured by the detectors

$$d^i_\nu(t) \approx \langle d^i_\nu \rangle \left[1 + P^i_{\text{cel}} (\Delta s_{\text{cont}} + \Delta s^\nu_{\text{CO}}) + P^i_{\text{tel}} \Delta s_{\text{ground}} + \delta^i_G + n_{\text{corr}} + n^{\nu i}_{\text{w}} \right], \tag{2.7}$$

where $d^i_\nu(t)$ is the raw time ordered data and we have assumed that the deviations from the mean are small, so that we can neglect second order terms.

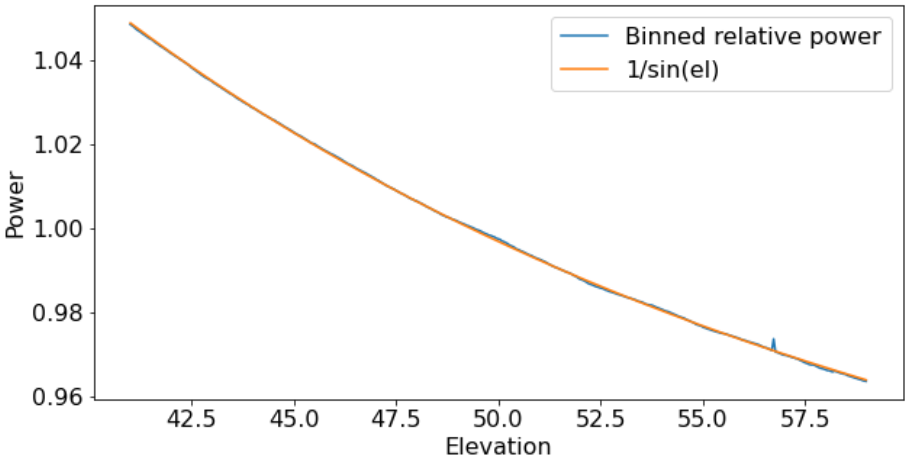

Figure 2.4: Normalized power measured during sky-dip observations compared to a fitted 1/ sin(el) model. Figure courtesy of Jonas Lunde.

2.4.2 Atmosphere

The atmosphere is one of the major contributors to the system temperature as well as to the correlated noise. The brightness temperature of the atmosphere can be roughly modeled as that of an optically thin medium at a certain temperature. The observed brightness temperature from the atmosphere, $T_{\mathrm{atm}}^{\mathrm{b}}$, is then given by

$$T_{\mathrm{atm}}^{\mathrm{b}} = T_{\mathrm{atm}}(1 - e^{-\tau}) \approx T_{\mathrm{atm}}\tau, \tag{2.8}$$

where T_{atm} is the physical temperature of the atmosphere and τ is the optical thickness of the atmosphere.

The optical thickness of the atmosphere depends on the elevation at which we are observing. We can take this elevation dependence into account using a simple model

$$\tau(\mathrm{el}) = \frac{\tau_0}{\sin(\mathrm{el})}, \tag{2.9}$$

where "el" denotes the elevation of the telescope pointing, and τ_0 is the optical thickness at zenith (el = 90 degrees). Figure 2.4 shows data from sky-dip observations (where the telescope quickly sweeps down in elevation and up again) taken with the COMAP telescope and a fitted optical depth model. This is a good illustration that the model works well.

Typical measured values of $T_{\mathrm{atm}}\tau_0$ are around $\sim$ 12 – 15 K. Since we typically observe at elevations between 35 and 65 degrees, the atmosphere typically contributes around 15 – 25 K to the system temperature.

The atmosphere also has a significant contribution to the correlated noise. To a large extent, the atmospheric column observed by the different feeds should be the same, so the correlated noise from the atmosphere should be highly correlated in the different

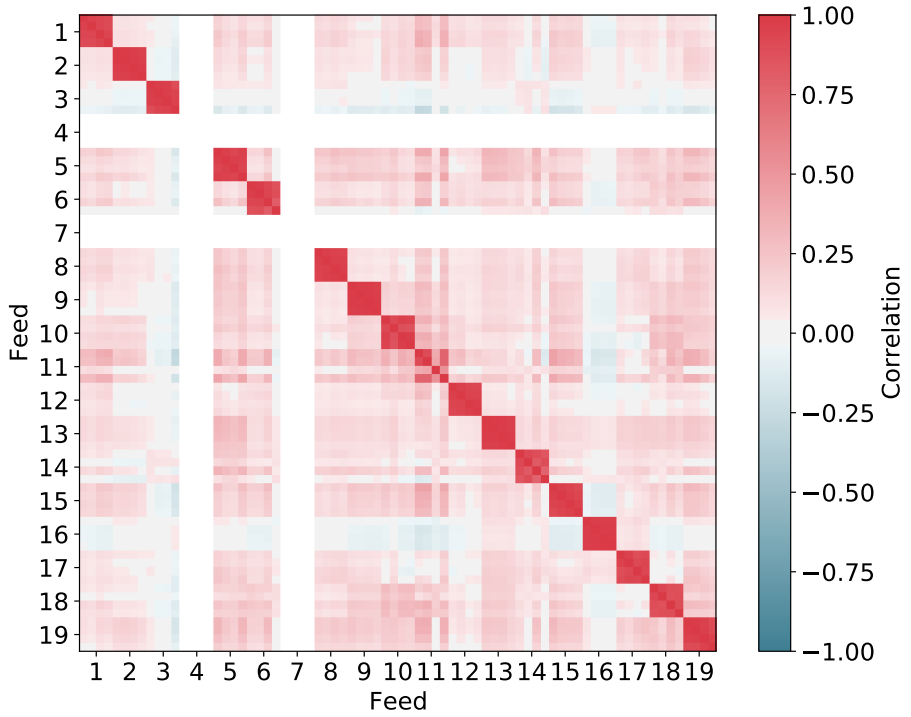

Figure 2.5: Correlation matrix of sideband averaged data from a single CES scan.

feeds, it should also be very smooth in frequency. If this is true, and there are no other major continuum contributions to the correlated noise that are also correlated between the different feeds, then the amount of correlation between the data of different feeds can be used to estimate the the rough magnitude of the atmospheric contribution to the correlated noise.

Figure 2.5 shows the correlation between the raw data averaged over all the frequencies of each sideband. We see that the four sidebands of the same feed are essentially completely correlated, this is because both the gain fluctuations and the atmosphere fluctuations are common to all four sidebands. For the different feeds, however, the gain fluctuations are independent, so the main contribution to the correlation here is from the atmosphere. This suggests that the correlations due to the atmospheric contribution are typically around 10 – 40 %.

2.4.3 Ground Contamination

Figure 2.6 shows a far sidelobe beam model for a single feed of the COMAP instrument. As we can see, we have significant beam sidelobes all the way out to around 70 degrees. This means that, as long as the telescope is pointing at an elevation lower than about 70 degrees, some of our main sidelobes will be hitting the ground. As long as the

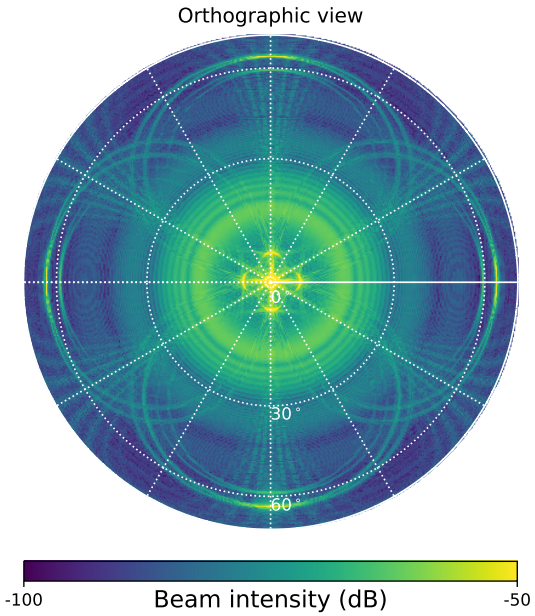

Figure 2.6: Simulated far sidelobe beam model for a single feed of the COMAP instrument. Beam model courtesy of James Lamb.

contribution from the sidelobes is fairly stable in time, this small ground pickup only has the effect of increasing our system temperature by up to a few Kelvin. However, if one of the more sharply defined sidelobes is right at the edge of the horizon, then the contribution from the ground can change quite abruptly, and we get a very strong signal in our time streams.

One thing that makes ground contamination more worrisome than other systematics is that the ground signal is correlated with where the telescope is pointing. This means that it will not necessarily integrate down properly when we add in more data. In general if some systematic is present in the data, this may not be ideal, but as long as the systematic is random and independent between each time we take data, then the contribution of the systematic will generally integrate down roughly as $\sim 1/\sqrt{N_{\text{obs}}}$, where N_{obs} is the number of independent observations.

The ground contribution, however, will be roughly the same each time we are observing at the same point in the sky, which means that it is not so simple. Of course, the ground is roughly constant in AZ/EL coordinates, while the cosmological signal is mapped in RA/DEC, so the signals are not completely degenerate, however, we do observe the same cosmological fields along the same path in AZ/EL every day, so the ground contamination will tend to affect the same modes in RA/DEC again and again.

While the ground itself is fairly close to a blackbody spectrum, which is something we would not be much worried about, the important thing for the ground contamination

is the frequency dependence of the sidelobes themselves. This is because if the sidelobes are very different at different frequencies, the ground contamination seen at different frequencies will also be very different, even though the ground is almost a blackbody. Fortunately, the positions of the sidelobe peaks is something that is not expected to change with frequency, but the sidelobe structure will in general be more sharp at high frequencies, and more diffuse at lower frequencies. The resulting ground contamination is expected to be fairly smooth in frequency, and thus something that can be removed quite effectively by filtering. Note that since the beam model (Fig. 2.6) is fairly expensive to run, we do not have one at another frequency, which means that these are mostly qualitative educated guesses about the beam properties.

While the smooth frequency structure should make it possible to remove the ground contamination quite effectively, we should note that the ground signal is still perhaps five orders of magnitude stronger than the cosmological signal we are trying to measure, so we need to remove it very precisely for it not to affect our measurements. This is what makes ground contamination one of the major challenges for COMAP.

The current version of the COMAP data analysis pipeline uses simple pointing template fitting in time domain in order to remove as much as possible of the ground signal. We use the simple atmosphere model from Sect. 2.4.2 together with using the azimuth position itself as a template

$$d_{\text{after}} = d_{\text{before}} - \frac{g}{\sin(\text{el})} - a\,\text{az} - \left\langle \frac{g}{\sin(\text{el})} + a\,\text{az} \right\rangle, \tag{2.10}$$

here g and a, as well as an overall constant, has been fit to the data. The elevation template is mostly there to remove the effect of the atmosphere, but it will also pick up some ground signal. The azimuth template is basically removing the linear term in the Taylor expansion of the ground signal about the centre of the field. These templates are subtracted from each frequency channel independently. Figure 2.7 shows the average values for these fitted template amplitudes along the paths followed by our three cosmological fields. We see large amplitudes both at high and low elevations. At low elevations (around el $\sim$ 30) the main sidelobe is hitting the edge of the horizon, while at around el $\sim$ 70 the lower of the four far sidelobe peaks (see Fig. 2.6) is just at the edge of the horizon.

The most extreme azimutal amplitudes observed are found in the region az, el $\sim$ $315, 30$ degrees. Here we believe what is happening is that one of the 60-70 degree sidelobes are just at the edge of Black Mountain (see bottom of Fig. 2.7 at az $\sim$ 60). Figure 2.8 shows the feed and frequency averaged raw time stream from a co6 observation of this area, together with the azimuth template. We here see a very strong effect. Even though we can remove most of this effect by fitting and subtracting templates, there is still contamination left, and this will be hard to remove completely at a later stage. This is why we currently do not use any data from below 35 degrees elevation.

In the next iteration of the analysis pipeline, the plan is to model the ground explicitly. This work is led by master student Jonas Lunde, under my supervision. We are using all the raw data to make continuum maps of the ground using the destriping mapmaker (Delabrouille, 1998; Keihänen et al., 2010). These ground maps can then be used to

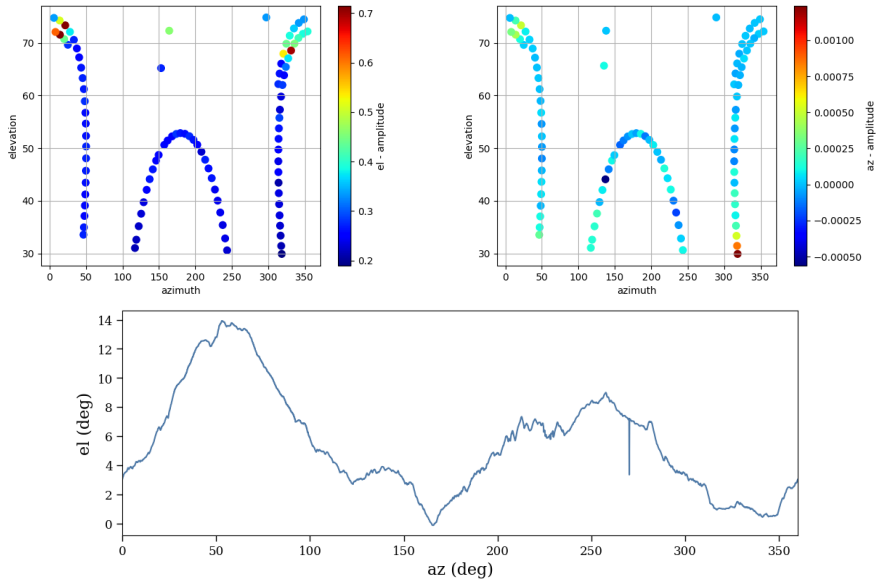

Figure 2.7: Average pointing template amplitudes fitted to the data in the data analysis pipeline (top). Results are binned along the path of the three cosmological fields. Elevation profile of the horizon as seen from the telescope location at OVRO (bottom). Elevation profile courtesy of Duncan Watts and Kieran Cleary.

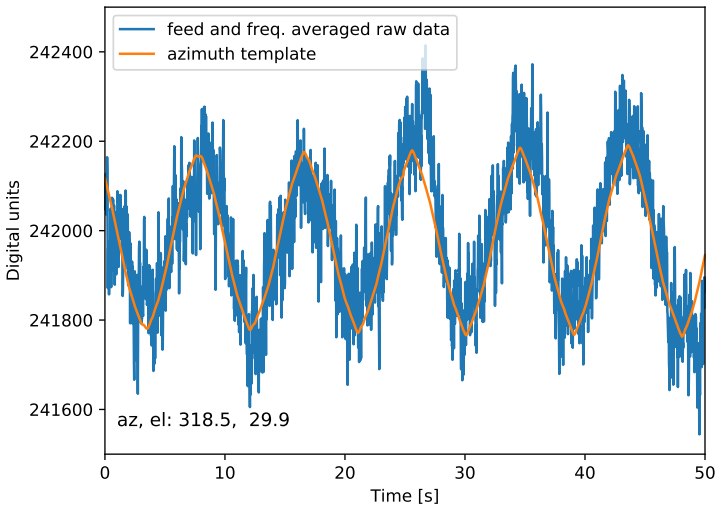

Figure 2.8: Raw time ordered data, averaged over feeds and frequencies, plotted together with the azimuth template.

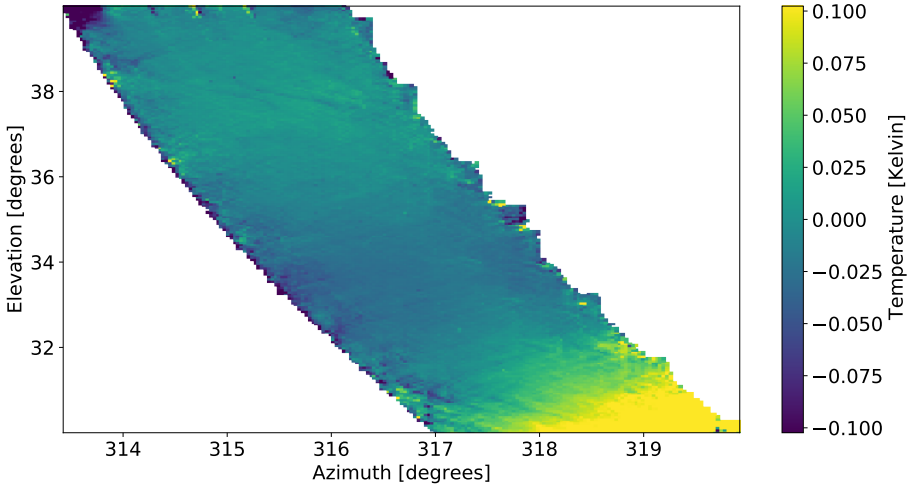

Figure 2.9: Continuum map of ground contamination. Made using a destriper and all the available Lissajous data taken in this region. Figure courtesy of Jonas Lunde.

make more precise templates to remove from the time streams during the regular data analysis. Figure 2.9 shows the previously discussed region with very strong ground contamination. In the lower right corner we see a strong gradient in both azimuth and elevation. By using maps like these, we can produce more precise templates that can remove the ground better than the simple templates we are currently using.

In addition to subtracting the ground contamination from the time streams, we can also design power spectrum methods that are as robust as possible to ground contamination, see Paper II (Ihle et al., in prep) for more details.

2.4.4 Standing Waves

Standing waves are another one of the major systematics in the raw COMAP data. They can form at various stages of the signal chain, from the sky to the digital backend, in the various potential electromagnetic cavities.

A given cavity length D will resonate with frequencies separated by

$$\Delta\nu_{\mathrm{SW}} = \frac{c}{2D}, \tag{2.11}$$

where $\Delta\nu_{\mathrm{SW}}$ is the distance, in frequency between peak resonances and c is the speed of light. A simple model for the standing wave structure in frequency is then given by (D. Chung, 2020)

$$T_{\mathrm{SW}} = T_{\mathrm{sys}} r_{\mathrm{SW}} \sin\left(2\pi \frac{\nu}{\Delta\nu_{\mathrm{SW}}} + \phi\right), \tag{2.12}$$

where T_{SW} is the contribution to the system temperature coming from the standing wave, r_{SW} is the (assumed to be frequency independent) standing wave amplitude and ϕ is some overall phase.

As we will see, a constant standing wave is not an issue for us since it will be removed in the bandpass normalization. It is only if a standing wave changes over time that it will show up in our data. These changes will arise whenever the cavity length changes, due to temperature fluctuations, vibrations etc. In the data the standing waves will then contribute to the correlated noise.

Standing waves can be distinguished from the other sources of correlated noise by their particular structure in frequency. Gain or atmospheric fluctuations are both (in slightly different senses) continuum sources in frequency, while standing waves will contribute with different sign to different frequencies as it goes in and out of resonance, as can bee seen from Eq. 2.12.

If the standing wave cavity is common to all the feeds, like the one between the receiver and the secondary, then the standing wave will change as a single function in time, common to all feeds, that contributes with different amplitudes to different frequencies. As we will see in Sect. 2.6.7, this is a perfect signal to pick up using Principal Component Analysis (PCA). Standing waves that only show up in individual feeds are much harder to measure and remove, but they will also affect a smaller part of the data, so they will integrate down significantly as we combine the different feeds.

Figure 2.10 shows the leading PCA mode from a CES scan of co6. We see that as a function of time, the mode looks like some sort of correlated noise. The amplitudes as a function of frequency show signs of the sinusoidal behavior expected from Eq. 2.12. Largest PCA amplitudes are found in feeds 3 and 5, this is because these feeds have a single stage polarizer, which makes them more susceptible to the standing waves than the other feeds which have two stage polarizers. If we try to estimate the frequency gap between the resonances, we see that we get a bit different results depending on which feeds we use, but the results are in the range $\Delta\nu_{\mathrm{SW}} \sim 0.3$ GHz, which corresponds to a cavity length of about 0.5 m. This fits fairly well with the cavity between the polypropylene weather shield that covers the receiver and the receiver itself (J. W. Lamb, 2020).

This reciever-weather shield standing wave is the one we see most clearly in our time ordered data, and the one we have studied the most. We find it in all the data, but the stability of the standing wave changes significantly, including a clear correlation with the windspeed. In June 2020, we augmented the the weather shield with a polystyrene backing support in order to make the weather shield more stable. This change was successful at suppressing this standing wave, although it is still there at a lower level. The other standing waves we have observed are typically more unstable and sporadic. For an example of another standing wave, with a frequency gap of $\Delta\nu_{\mathrm{SW}} \sim 0.07$GHz, suggestive of the ~ 2 m distance between the receiver and the secondary, see Fig. 2.11. We are still in the early phase of understanding and classifying these standing waves, but we are making progress in this area, especially in understanding which standing waves affect our measured cosmological power spectrum.

2.4.5 Correlated noise

There are three main sources of correlated noise. Atmospheric fluctuations contributes to the correlated noise at timescales longer than one or a few seconds, and these fluctuations are common to all frequencies and feeds. Gain fluctuations are typically the main source

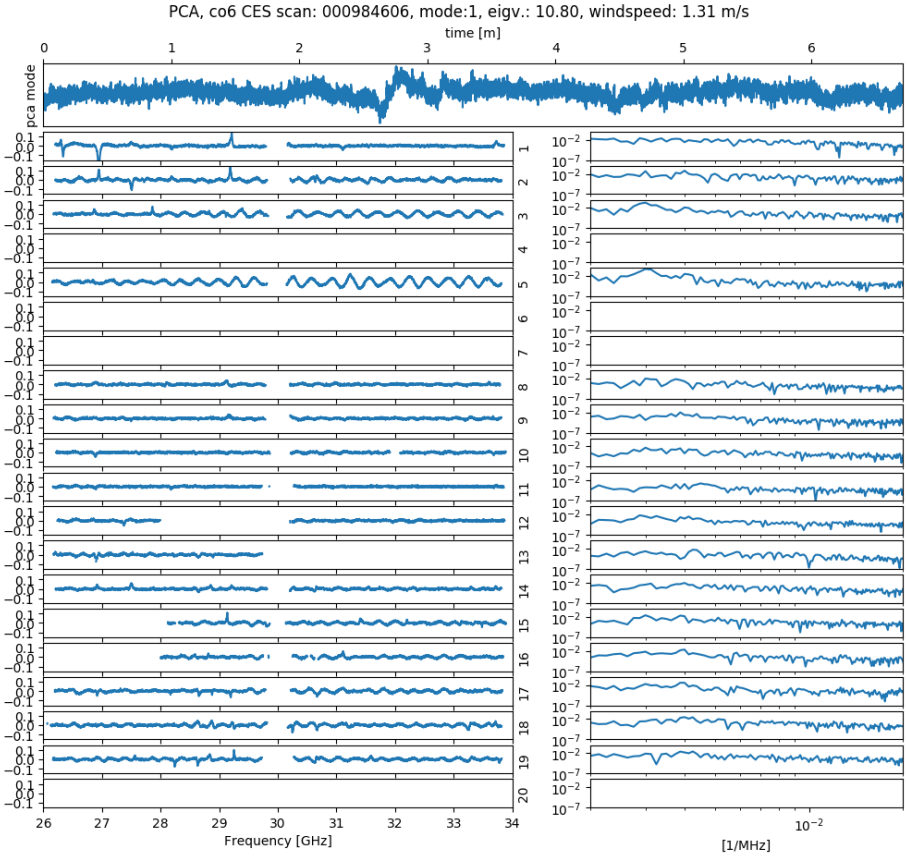

Figure 2.10: Leading PCA mode as a function of time (top panel). The rest of the panels show the corresponding PCA amplitudes as a function of frequency for each of the 19 feeds (left) and power spectral density, in frequency, of these amplitudes (right). Data is from a CES scan of co6.

of correlated noise, but they are also common to all the frequencies within a single feed. Some standing waves contribute significantly to the correlated noise, however, the same frequency structure is typically preserved over time, so these can be removed using a PCA filter.

Since all these sources of correlated noise are also highly correlated between frequencies, they can usually be removed very precisely in the filtering steps of the data analysis pipeline, as can be seen in Fig. 2.12. For this reason, correlated noise is not a systematic that we are very worried about at this stage in COMAP.

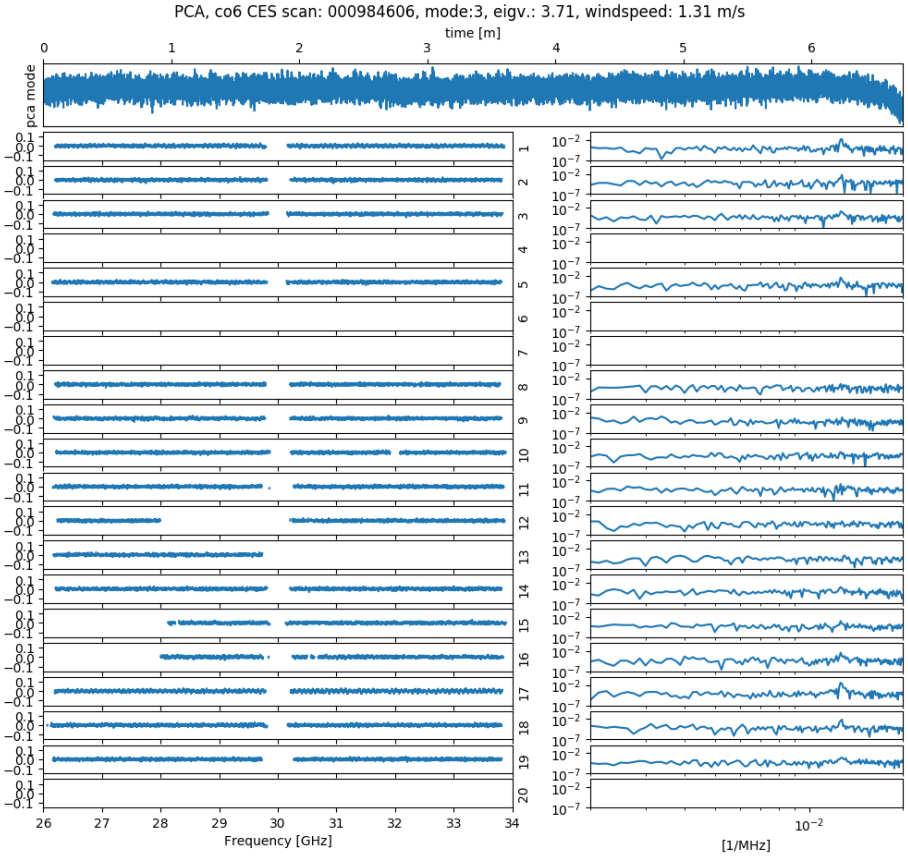

Figure 2.11: Third PCA mode as a function of time (top panel). The rest of the panels show the corresponding PCA amplitudes as a function of frequency for each of the 19 feeds (left) and power spectral density, in frequency, of these amplitudes (right). Data is from a CES scan of co6.

2.4.6 Continuum Foregrounds

Our cosmological fields are selected to be in regions with low galactic foregrounds, for this reason we have not been too worried about these up until now, as they will be fairly effectively removed in the frequency filter. However, the foregrounds are still up to two orders of magnitude stronger than our CO signal, so we do need to suppress them significantly, if we want to measure our signal.

This will be a more and more significant issue the more sensitivity we get, so it is something that we will probably have to focus more on in the future.

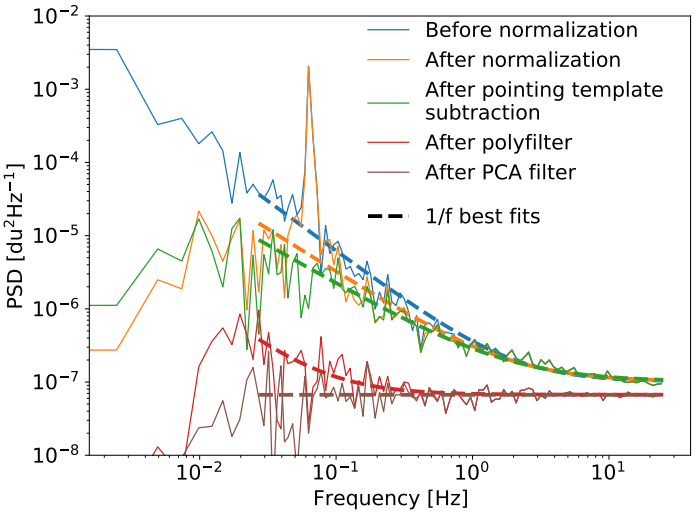

Figure 2.12: Power spectral density of the COMAP time ordered data at various different stages of the analysis pipeline. Figure courtesy of Jonas Lunde.

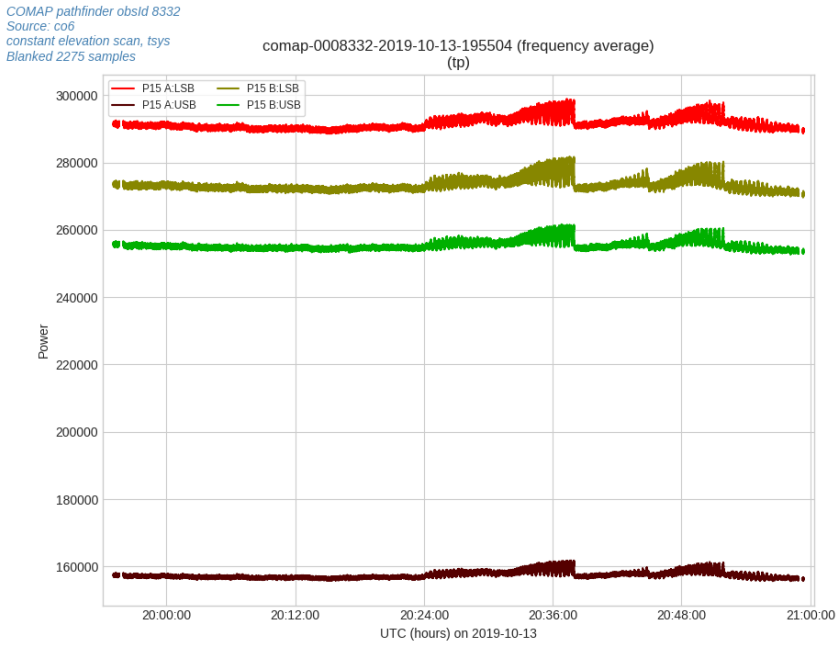

Figure 2.13: Example of time ordered data taken with the sun in the far sidelobe. Here we see data from the four sidebands of feed 15.

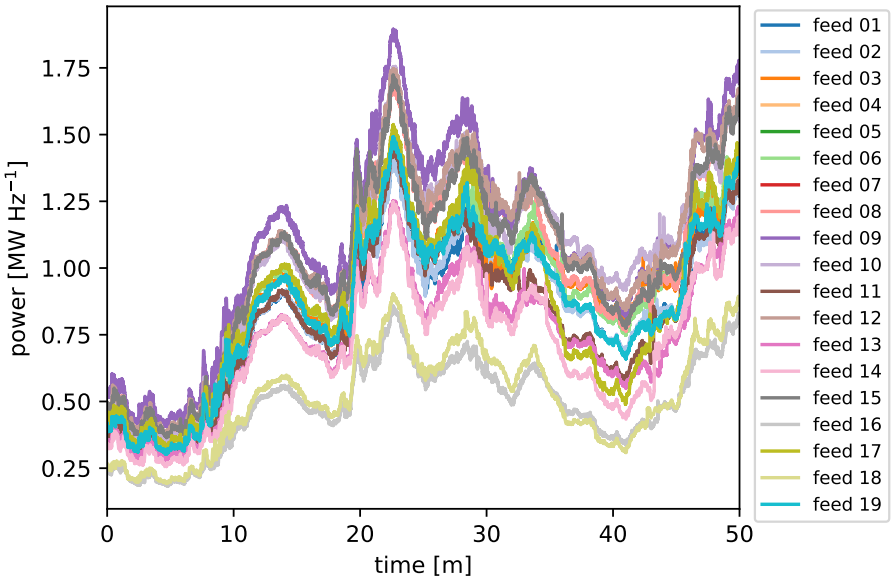

Figure 2.14: Example of data taken during bad weather.

2.4.7 Sun/moon in sidelobes

Another systematic that we did not anticipate the significance of in advance were the times when the sun, or even the moon, is exactly at one of the four peaks in the far sidelobes (see Fig. 2.6 at about 65 degrees). If this happens, we get a very large signal in the time streams, as can be seen in Fig. 2.13 after about 30 minutes. When this effect was first observed, we did not know what was the cause, but after looking at the data in the right way we could see the effect very clearly (see Fig. 2.23). We now track the relative position of the sun carefully and this allows us to remove the affected data during data selection (see Sect.2.7.1).

2.4.8 Weather

If there is significant cloud coverage, or other bad weather, then the data taken is typically not usable. Figure 2.14 shows an extreme example of data taken during bad weather. As we will discuss in Sect. 2.7.1 bad weather actually affects a significant part of the data, and is dealt with during data selection.

2.4.9 Spikes

We occasionally see spikes in the raw data, especially during the summer. We do not know for sure what the cause of the spikes are. Some of them seem to come from insects flying in front of the receiver, but we do not know if all of them are caused by

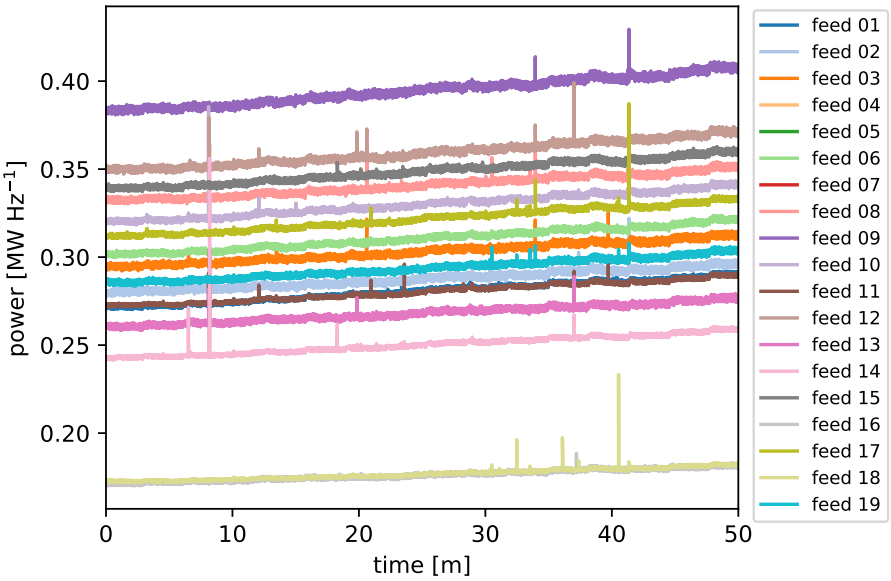

Figure 2.15: Example of an observation with a very large number of spikes.

this. Sometimes the spikes show up sporadically, one or two in an entire obsid, and sometimes the data is flooded by spikes (see Fig. 2.15).

The spikes are continuous in frequency, so we expect most of the signal to be removed in the frequency filters. We have not seen any clear signs that the spikes make a significant impact on the final maps or power spectra, but we are tracking the spikes to look for any such affect.

2.5 Calibration

Calibration is one of the major challenges we need to solve in order to make use of the data. We need to translate the raw power readings of the detectors (given in arbitrary digital units) into brightness temperatures. COMAP uses a main calibration strategy based on a calibration vane at ambient temperature that is periodically moved in front of the receiver. This acts as a hot load reading that is compared to the cold load of the sky measurement.

We can verify the calibration by looking at a source with a known temperature, like Jupiter. Although this will not give a very precise calibration, it is still a useful test to perform, if nothing else than as a sanity check that we are not doing something completely wrong. It will also show if the calibration is consistent over time.

2.5.1 Ambient load calibration

The main method of calibration used for COMAP involves comparing an observation of the cold sky with observations of a calibration vane at ambient temperature. This ambient load is automatically moved in front of the receiver and back during calibration.

The power measured by the telescope at any given time is given by

$$P = GT_{\rm sys}, \tag{2.13}$$

where P is the total power, G is the gain[1] and $T_{\rm sys}$ is the system temperature combining the contribution from the receiver with the external signal coming into the feed horns.

The idea of the calibration method is simple. We observe two different signals at known temperatures, and use these two measurements to determine $T_{\rm sys}$ assuming the gain is constant.

The complication with this is that we want to calibrate to the cosmological signal, and some of this signal gets absorbed by the atmosphere, meaning that only part of the signal makes it down to our receiver. So for a given signal contribution the system temperature changes by

$$\Delta T_{\rm sys} = e^{-\tau}\Delta T_{\rm signal}, \tag{2.14}$$

where $\Delta T_{\rm sys}$ is the contribution to the physical system temperature of the signal, τ is the optical depth of the atmosphere and $\Delta T_{\rm signal}$ is the temperature of the cosmological signal. To account for this effect we define an effective system temperature (Penzias and Burrus, 1973)

$$T'_{\rm sys} \equiv e^{\tau}T_{\rm sys}, \tag{2.15}$$

where $T'_{\rm sys}$ is the effective system temperature and $T_{\rm sys}$, as before, is the physical system temperature measured by the instrument. This new definition ensures that

$$\Delta T'_{\rm sys} = \Delta T_{\rm signal}, \tag{2.16}$$

making the interpretation of our measurements much easier. Another way to think about this is to note that losing some of the cosmological signal in the atmosphere is equivalent to just having a higher noise level in the first place, and the effective temperature, $T'_{\rm sys}$, is just this higher noise level. In the same way we define the effective gain

$$G' \equiv e^{-\tau}G, \tag{2.17}$$

where G' is the effective gain and G is the physical gain. We can now rewrite Eq. 2.13 using our new definitions

$$P = GT_{\rm sys} = G'T'_{\rm sys}. \tag{2.18}$$

When the instrument is looking at the cold sky, the power is given by

$$P_{\rm c} = G\left(T_{\rm rcv} + T^b_{\rm atm} + e^{-\tau}T_{\rm CMB}\right) = GT_{\rm sys}, \tag{2.19}$$

where $P_{\rm c}$ is measured power when looking at the cold sky (the *cold* load), $T_{\rm rcv}$ is the receiver temperature contribution to the physical system temperature, $T^b_{\rm atm}$ =

[1] Note that, for convenience, we have absorbed the conventional constant factors $k_{\rm B}\Delta\nu$ into G for this derivation.

$(1 - e^{-\tau})T_{\rm atm}$ is the brightness temperature of the atmosphere, $T_{\rm atm}$ is the physical temperature of the atmosphere, which we assume to be in thermal equilibrium at a constant temperature, and $T_{\rm CMB}$ is the CMB monopole, which we assume dominates any other cosmological signal. Rewriting this in the effective coordinates we get

$$P_{\rm c} = G' \left(T'_{\rm rcv} + T'^{b}_{\rm atm} + T_{\rm CMB} \right) = G' T'_{\rm sys}, \tag{2.20}$$

where we have defined the effective receiver temperature $T'_{\rm rcv} \equiv e^{\tau} T_{\rm rcv}$ and atmosphere brightness temperature $T'^{b}_{\rm atm} \equiv e^{\tau} T^{b}_{\rm atm}$.

If the instrument is looking at the calibration vane (the *hot* load) the measured power is given by

$$P_{\rm h} = G \left(T_{\rm rcv} + T_{\rm h} \right), \tag{2.21}$$

where $T_{\rm h}$ is the physical temperature of the calibration vane. If we assume that the physical temperature of the calibration vane is the same as the physical temperature of the atmosphere, we can rewrite $T_{\rm h}$ in the following convenient way

$$T_{\rm h} = e^{-\tau} T_{\rm h} + (1 - e^{-\tau}) T_{\rm h} = e^{-\tau} T_{\rm h} + T^{b}_{\rm atm}. \tag{2.22}$$

If we now write Eq. 2.21 in the effective units we get

$$P_{\rm h} = G' \left(T'_{\rm rcv} + T_{\rm h} + T'^{b}_{\rm atm} \right) = G' \left(T_{\rm h} + T'_{\rm sys} - T_{\rm CMB} \right). \tag{2.23}$$

Using the measurements of the cold and the hot load, assuming the gain is constant and that we know the temperature of the calibration vane, we can solve for the effective system temperature

$$T'_{\rm sys} = \frac{T_{\rm h} - T_{\rm CMB}}{P_{\rm h}/P_{\rm c} - 1}. \tag{2.24}$$

During this derivation we have neglected the effect of ground spillover. However, a very similar argument can be made to take this into account as well, and if we also assume that the ground has the same physical temperature as the calibration vane and the atmosphere, the result is exactly the same as in Eq. 2.24 with the effective system temperature, $T'_{\rm sys}$, now taking into account both the effect of the atmosphere and the ground spill. Note that whenever the temperature of the system, atmosphere etc. is mentioned elsewhere in this thesis, we are referring to the effective temperatures, not the physical ones.

2.5.2 Jupiter calibration

We do periodic Jupiter observations for calibration and pointing correction. By fitting a Gaussian beam to these observations we can measure the antenna temperature of Jupiter. These measurements can be compared to model expectations depending on the current distance to Jupiter. This work is led by Stuart Harper in the Manchester group.

Figure 2.16 shows antenna temperature measurements of Jupiter from October 2019 to August 2020, for a single feed, compared to the model. We see that apart from an overall constant factor offset, the measurements fit the model very well. The overall offset factor, which we can see is given roughly by 0.7-0.8, is the effect of the aperture

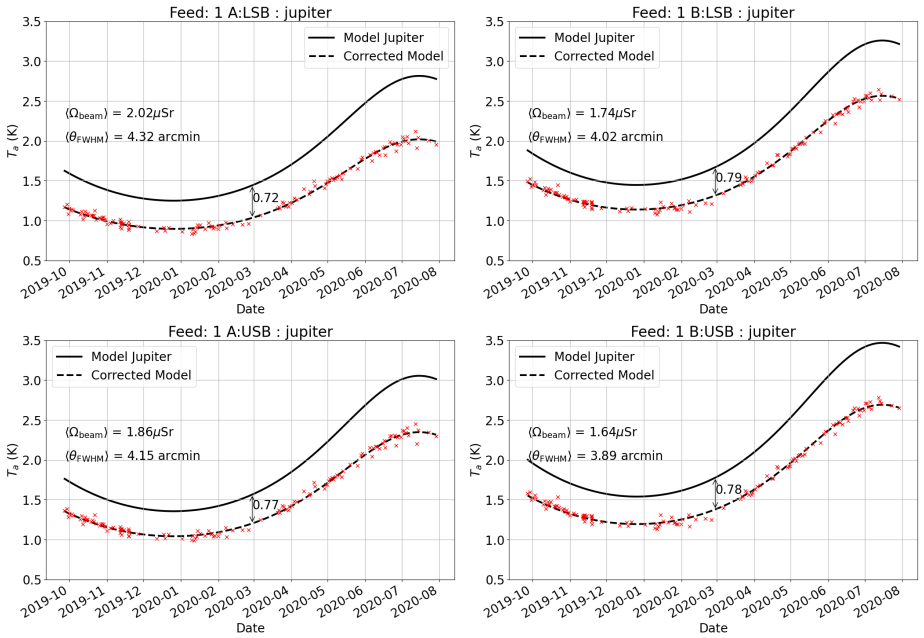

Figure 2.16: Antenna temperatures measured by feed 1, based on calibration observations of Jupiter, compared to model expectations. Figure courtesy of Stuart Harper.

efficiency[2], which is expected. These results show that the calibration is consistent over time, and at a reasonable value.

2.6 Data Analysis Pipeline

The COMAP analysis pipeline takes the raw data, from what we call *level 1* files, does calibration, filtering, performing data selection, making maps and calculating power spectra.

Figure 2.17 shows an overview of the various steps and modules that are part of the COMAP analysis pipeline. The first modules are `scan_detect`, which identifies, classifies and divides our data, and `l2gen`, which performs the low-level calibration, filtering and frequency masking of the data and producing *level 2* files. Level 2 files contain calibrated and filtered time ordered data, as well as a set of diagnostic data. Before the level 2 files are combined into maps, there is a second level of data selection, performed by the module `accept_mod`, which produces an `accept_list` denoting which observations to accept or reject when making the maps. The mapmaker, `tod2comap`, converts the time ordered data from the level 2 files into 3D maps of brightness temperature. The preceding steps are described in Paper I (Foss et al., in

[2]It is slightly different from the aperture efficiency since Jupiter is not a point source.

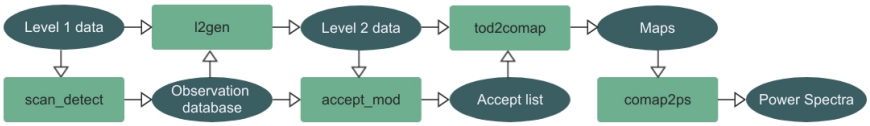

Figure 2.17: Overview of the COMAP CO analysis pipeline. Figure courtesy of Jowita Borowska.

prep). Here we will give a summary overview, while going into more details in some areas.

The last part of the current data analysis pipeline is `comap2ps`. This module takes in the 3D maps in brightness temperatures and calculates various power spectra. These procedures are described in Paper II (Ihle et al., in prep). We will not repeat much of the methods presented there, but will, in later sections, discuss more abstractly the problem of inference from line intensity maps. We will also go into some details on using the voxel intensity distribution (VID) that have not been published.

2.6.1 Scan Detect

The main goal of scan detect is to go through all the raw data files to classify and divide the data into individual scans. This information is then provided in terms of a runlist, which is used when running `l2gen` or `tod2comap`.

A *runlist* is a list of obsids that is divided according to the field or source observed during each obsid. For each obsid, we list all the scans within that obsid. For each scan the modified Julian date (MJD) of the start and end of the scan is provided as well as the scanning strategy (e.g. Lissajous or constant elevation) and some pointing information.

2.6.2 Level 2 File Generator: l2gen

The main goal of l2gen is to take the raw data files, called level 1 files, and turn them into masked and filtered data files, called level 2 files, that are ready for mapmaking. An additional function of l2gen is to keep track of various diagnostics for tracking data quality and for data selection.

The main steps in l2gen are (in chronological order):

- Bandpass normalization
- Removal of pointing templates
- Masking
- Polynomial filter in frequency

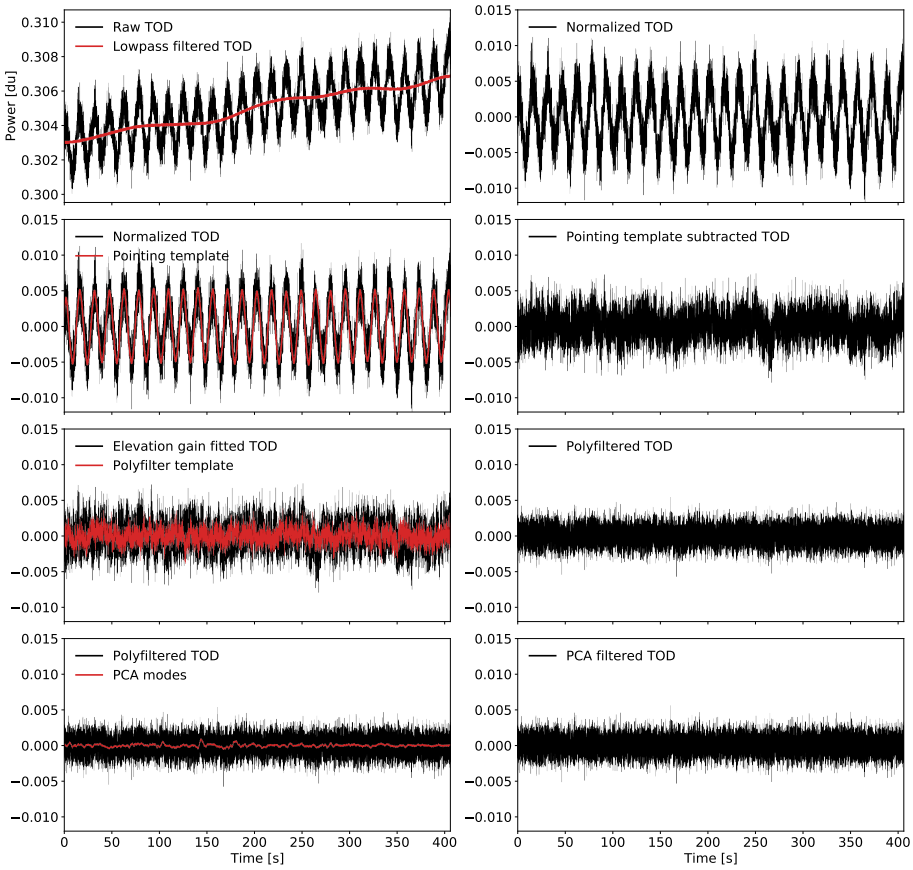

Figure 2.18: Time ordered data plotted along the various steps of the data analysis pipeline. Each row shows the data before (left column) and after (right column) applying the indicated filter. From top to bottom, the filters shown are 1) normalization; 2) pointing template subtraction; 3) polynomial filter in frequency; and 4) PCA filter. Data used is from scan 1445603, feed 5, in a 31.25 MHz band around 27.673 GHz. Figure courtesy of Jonas Lunde.

- PCA-filter
- Calibration
- Decimation

2.6.3 Bandpass normalization

Our model for the total power seen in the detectors is given by

$$P_{\text{out}} = k_{\text{B}} G \Delta\nu T_{\text{sys}}, \tag{2.25}$$

where k_B is Boltzmanns constant, G is the gain and $\Delta\nu$ is the bandwidth of the frequency channel.

The first of the actions that is performed on the data is bandpass normalization. The normalization is performed on each frequency channel separately and is done by dividing the time ordered data (TOD), d, by its *running mean*, $\bar{d}$, and subtracting one. As we see from Eq. 2.25, this normalization will cancel out both the average gain G and the average system temperature $T_{\rm sys}$ leaving us with a time stream with small-scale fluctuations with a variance of $1/\Delta\nu\tau$ as dictated by the radiometer equation, where τ is the duration in time of each sample.

The running mean is calculated by passing the TOD through a lowpass filter. See Fig. 2.18 (top row) for an example of this normalization step. The lowpass filter is implemented in Fourier space by multiplying with the following weight function:

$$W = \left[1 + \left(\frac{f}{f_{\rm knee}}\right)^{\alpha}\right]^{-1}, \tag{2.26}$$

where f is the temporal frequency, $f_{\rm knee} = 0.02$ Hz (corresponding roughly to a 50 s timescale) and $\alpha = 4$. So we have

$$\bar{d} = \mathcal{F}^{-1}\left\{\mathcal{F}\{d\}W\right\}, \tag{2.27}$$

where $\mathcal{F}$ denotes the Fourier transform. The normalized data then becomes

$$d_{\rm norm} = d/\bar{d} - 1. \tag{2.28}$$

2.6.4 Removal of Pointing Templates

As the telescope changes elevation we are looking through different amounts of atmosphere, leading to changes in the power received by the detectors. As we have discussed, this effect can be modeled by a simple expression for the optical depth of the atmosphere, τ

$$\tau(\mathrm{el}) = \tau_0/\sin(\mathrm{el}), \tag{2.29}$$

where τ_0 is the optical depth of the atmosphere at zenith, and el is the elevation.

We also know that sometimes there is significant ground contamination, and we therefore want to subtract an azimuth template, as well as the elevation template. Specifically we assume that

$$d = g/\sin(\mathrm{el}(t)) + a\,\mathrm{az}(t) + c + n, \tag{2.30}$$

where g, a and c are constants and n is some Gaussian noise with constant variance. We then find the best fit values (i.e. least squares) of g, a and c, and use the values of g and a to remove the pointing templates from the time stream:

$$d_{\rm after} = d_{\rm before} - g/\sin(\mathrm{el}(t)) - a\,\mathrm{az}(t) - \langle g/\sin(\mathrm{el}(t)) + a\,\mathrm{az}(t)\rangle, \tag{2.31}$$

where $\langle\ \rangle$ denotes the mean value. See Fig. 2.18, second row.

We fit these templates independently for each frequency in order to try to remove any frequency structure of the ground and atmosphere. For longer scans we divide the data into different segments of roughly 4 minutes each, and perform this template fit and removal separately on each of the segments of data.

2.6.5 Masking

We want to mask out frequency channels that are *bad* in the sense that they are not performing as expected. We do this mostly by using the correlation matrix between different frequency channels.

In order for the data to be as clean as possible before identifying bad channels, we make a copy of the original (level 2) dataset and perform the poly- and PCA filters, described below, on this copied data, before calculating the correlation matrices within each band. We use two main approaches to identifying individual or groups of channels to be masked. The first approach uses the fact that the expected correlation between two independent Gaussian variables (for large n_{samp}) is given by $1/\sqrt{n_{\text{samp}}}$, where n_{samp} is the number of samples the correlation is calculated on.

This means that, after subtracting the expected correlation induced by the polynomial frequency filter, we know the statistics describing *good* data, and can identify bad data as deviations from these statistics. We look at entries within squares of different sizes as well as sets of columns within the correlation matrix and see if the average of the absolute value of the correlations within this region deviates from the values expected, and mask out the corresponding channels if the deviations are outside the acceptable limits.

The second approach is to calculate a set of diagnostics for individual frequency channels (like the average correlation of the channel to all the others in that band or the average of the absolute value of the same). We can then compare the values of these diagnostics for the different channels and look for large outliers. The disadvantage of the first approach is that it is much harder to detect these deviations in a short scan than it is in a long scan, so it is hard to choose the parameters for the acceptable deviations in a way that will mask out enough of a short scan without masking too much of a long scan. The disadvantage of the second approach is that if the overall data quality is bad, then a bad channel will not necessarily be an outlier, and will not be masked out.

In addition to these approaches we look specifically for *edge correlations*, that is, correlations between individual frequencies at the edge of each sideband with the corresponding frequencies at the edge of another sideband. This is to look for and mask out frequencies affected by a known aliasing effect. We also have some overall hard cuts on large individual correlations and on large individual variances (compared to what is expected from the radiometer equation). After this mask has been found we perform all the analysis on the original dataset, but now only using the frequency channels that are not masked.

2.6.6 Polynomial filter in frequency

The goal of the polynomial filter in frequency is to remove the $1/f$-noise specific to each feed, as well as to remove any continuous foregrounds from the data. This filter greatly reduces the correlated noise, and often leaves the TOD close to white noise.

The poly filter is performed separately on each time step, by fitting and removing a polynomial (usually linear) in frequency across each sideband. Specifically, we assume

$$d_\nu = a + b\nu + c\nu^2 .. + n, \tag{2.32}$$

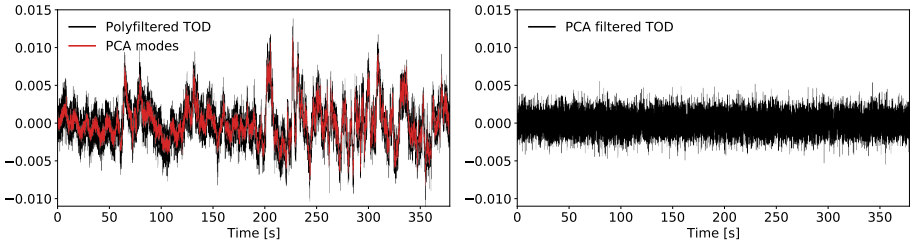

Figure 2.19: Example of time ordered data before and after PCA filter. This is a very extreme case with a very strong standing wave. Figure courtesy of Jonas Lunde.

where d_ν is the data across one sideband at a specific time step, a, b and c etc. are constants and n is some Gaussian noise with constant variance. We then find the best fit values (i.e. least squares) for the constants and subtract the polynomial (see Fig. 2.18, third row)

$$d_\nu^{\text{after}} = d_\nu^{\text{before}} - \left(a + b\nu + c\nu^2 ...\right). \tag{2.33}$$

Our current version of the pipeline uses a linear polynomial for the polyfilter, which appears to be sufficient for our purposes.

One possible issue with the current version of the polyfilter is that, depending on what the origin of the effects we want to remove with this filter, we may want to divide the polynomial in Eq. 2.32 by the T_{sys} corresponding to the frequency channel, since a channel with high T_{sys} is expected to have a relatively lower effect of any signal compared to the noise. This is certainly what we would expect from a continuous foreground.

We are working on improved ways to do this frequency filtering in a way to remove both the gain fluctuations as well as continuum foregrounds in a consistent manner.

2.6.7 PCA-filter

Let us consider all the data corresponding to a single scan as a data matrix, D_{ij}, with the TOD (time stream) corresponding to a single frequency as a row of the matrix. This will be a matrix with dimensions $n_{\text{freq}} \times n_{\text{samp}}$, where $n_{\text{freq}} = 19 \cdot 4 \cdot 1024$, i.e. the total number of frequencies added up over all the sidebands and feeds, and where n_{samp} is the number of samples in time (typically of the order of ~ 20000 for a single scan).

$$D = \begin{bmatrix} D_{11} & \dots & D_{1n_{\text{samp}}} \\ \vdots & \ddots & \vdots \\ D_{n_{\text{freq}}1} & \dots & D_{n_{\text{freq}}n_{\text{samp}}} \end{bmatrix}. \tag{2.34}$$

If we treat the columns (all the different time steps) of the data matrix D as representing a set of n_{samp} random variables, then each frequency of each feed is a new sample of each of these random variables. The covariance matrix for these random

variables is then given by

$$C^{\text{samp}} \propto D^{\text{T}}D = \begin{bmatrix} C_{11} & \dots & C_{1n_{\text{samp}}} \\ \vdots & \ddots & \vdots \\ C_{n_{\text{samp}}1} & \dots & C_{n_{\text{samp}}n_{\text{samp}}} \end{bmatrix} \tag{2.35}$$

The PCA components are the eigenvectors of this matrix with the highest eigenvalues. These eigenvectors are basically just functions of time (one value for each time step). Specifically, the leading PCA-components are the functions of time that explain most of the variance between the different frequencies.

After finding the properly normalized eigenvectors, v_k, corresponding to the four largest eigenvalues of C^{samp} we calculate the PCA amplitude corresponding to each of these eigenvectors for each frequency channel, and subtract it from the data:

$$a_k = d \cdot v_k = \sum_{i=1}^{n_{\text{samp}}} d_i v_k^i$$

$$d_{\text{after}} = d_{\text{before}} - \sum_{k=1}^{4} a_k v_k,$$

where d is the time stream of a single frequency channel, a_k is the PCA amplitude of that same channel corresponding to the eigenvector v_k.

The PCA filter is perfect for removing standing waves that are common to all (or at least many) feeds. This is because, as we discussed in Sect. 2.4.4 the standing waves typically change as a single function of time, but affect different frequencies differently depending on if they are in or out of resonance. The resonance structure of the standing waves will then show up in the PCA amplitudes of the different frequencies, as we see in Figs. 2.10 and 2.11.

The PCA filter also picks up other systematics that do not have the features one would expect from standing waves. These effects are mostly of unknown origin at this point, but the PCA filter can still usually remove them quite effectively.

2.6.8 Calibration

For each frequency we interpolate the measured power and temperature of the hot load (calibration vane), P_{hot} and T_{hot} to the time, t_{scan}, of our current scan, i.e.:

$$P_{\text{hot}}(t_{\text{scan}}) = \frac{P^1_{\text{hot}}(t_2 - t_{\text{scan}}) + P^2_{\text{hot}}(t_{\text{scan}} - t_1)}{t_2 - t_1},$$

$$T_{\text{hot}}(t_{\text{scan}}) = \frac{T^1_{\text{hot}}(t_2 - t_{\text{scan}}) + T^2_{\text{hot}}(t_{\text{scan}} - t_1)}{t_2 - t_1},$$

where 1 and 2 denote the hot load measurement at the start and the end of the current obsid, respectively. We can then calculate T_{sys} for our current frequency

$$T_{\text{sys}} = \frac{T_{\text{hot}}(t_{\text{scan}}) - T_{\text{CMB}}}{P_{\text{hot}}(t_{\text{scan}})/\langle P_{\text{cold}}\rangle - 1}, \tag{2.36}$$

where $\langle P_{\text{cold}}\rangle$ is the mean power of the scan in this frequency channel.

2.6.9 Decimation

All the analysis up until now have been performed on the full frequency grid (1024 channels per sideband). For mapmaking purposes, however, we typically don't need this kind of resolution, so we want to co-add several frequency channels together to a single low resolution channel.

$$d_i^{\text{lowres}} = \frac{1}{\sum_m w_m} \sum_{m=(i-1)n_{\text{dec}}+1}^{in_{\text{dec}}} w_m d_m^{\text{highres}}, \tag{2.37}$$

where d_i^{lowres} is the time stream of frequency channel i in the low resolution frequency grid, d_m^{highres} is the time stream of frequency channel m in the high resolution frequency grid, n_{dec} is the number (usually 16) of high resolution frequencies to be combined in each low resolution frequency channel and $w_m = 1/\sigma_m^2$ is the inverse variance of the time stream in frequency channel m of the high resolution data (w_m is zero for masked frequencies).

2.6.10 Mapmaker, `tod2comap`

The mapmaker takes a set of level 2 files with calibrated and filtered time ordered data, and makes 3D maps of brightness temperature.

The mapmaker we use is very simple. It basically assumes that the input data is dominated by white noise, so that every sample is independent. The value in each pixel of the map is then just the noise weighted sum of all the samples that hit this pixel

$$m_p = \frac{\sum_{d\in p} \sigma_i^{-2} d_i}{\sum_{d\in p} \sigma_i^{-2}}, \tag{2.38}$$

where d_i is sample number i of the time ordered data, σ_i is the white noise level of sample i and where m_p is the value of the map in pixel p. This is typically referred to as a *binned* mapmaker. Figure 2.20 shows single frequency maps based on all the (accepted) data for each of the three main science fields. The maps look white noise dominated and smooth.

A large advantage with using this simple mapmaker is that we do not need to load in all the time-ordered data into memory at once, but can add more data incrementally. The disadvantage is that we lose a lot of the information that is in the TOD. This results in lower sensitivity to the signal on large scales (see Paper I, Foss et al., in prep for more details).

We are working on an improved mapmaker, using a destriper, and this has shown some promising results. This work is lead by master student Nils-Ole Stutzer, under my supervision. Figure 2.21 compares the pipeline transfer function[3] for the binned map and a map created by the destriper. We see that the destriper recovers a lot more of the large scale structure in the angular direction, it is however, much more costly process to run. Note that this destriper still treats each frequency channel independently, so it is still far from an optimal mapmaker.

[3] See Paper I (Foss et al., in prep) for details on how the transfer function is defined and estimated.

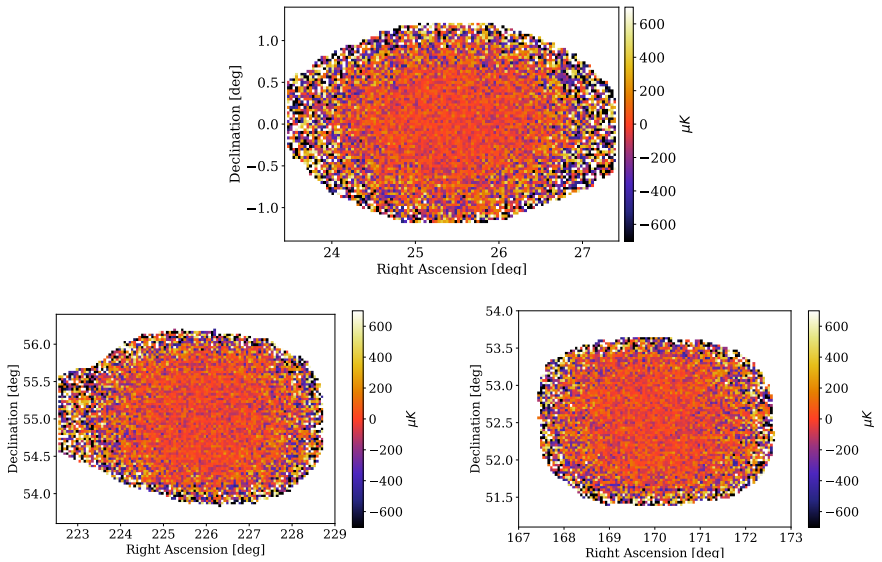

Figure 2.20: Feed averaged maps from of the co2 (top), co6 (bottom left) and co7 (bottom right) fields. These are single frequency maps at 32 GHz with a bandwidth of 31.25 MHz, including all the accepted data. Figure courtesy of Nils O. Stutzer.

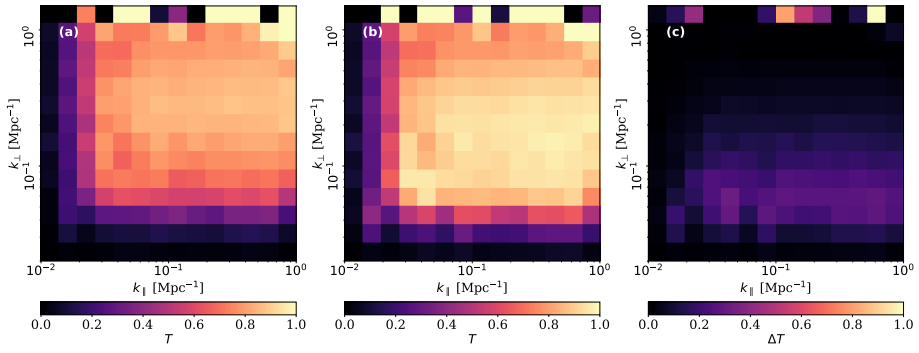

Figure 2.21: Comparison of the transfer function using the current (binned) mapmaker (a), the destriper (b), difference between the two (c). Figure courtesy of Nils O. Stutzer.

2.7 Data selection

There are two main stages of data selection in the COMAP pipeline. First, as discussed in the previous section, we mask out badly performing or outlier frequencies as part of the l2gen process. The second stage happens after we have created level 2 files, but before we run them through the mapmaker. This selection step is performed in the `accept_mod` module, and is based on gathering a large number of statistics for each chunk of data, and then deciding based on these statistics if we should accept or reject that data.

2.7.1 Statistics database

The first and most extensive part of the `accept_mod` module is the generation of the statistics database. This is a database where a number of statistics, or diagnostic variables, are calculated or gathered for each sideband of each feed for each scan. So the entire scan database is an array of dimensions $(n_{\text{scans}}, n_{\text{feeds}}, n_{\text{sidebands}}, n_{\text{stats}})$, where n_{stats} is the number of different diagnostic variables gathered for each chunk of data. Each statistic is represented by a floating point number.

Some of the statistics are calculated from the data itself, like various χ^2-values or the average kurtosis of the data, while others, like air temperature or windspeed are collected from the housekeeping data. In Tab. 2.2 we list all the different statistics currently used for the database, together with a short description of each one.

As the statistics database is a fairly small dataset (in total about 1 GB), it is very useful to look for patterns in this dataset, in order to understand the raw data, and the effect of the data analysis pipeline, better. Perhaps the most basic analysis we can do on a dataset like this is to do a basic correlation analysis. Fig. 2.22 shows the correlation between the different statistics from all the co2 data. Note that we do an automatic removal of extreme outliers, to avoid spurious correlations. Here we see a number of interesting correlations. While the mere correlation between two variables is often not in itself actionable information, the full correlation plot is still very useful to see which variables are related and which distributions to look closer at.

A particularly useful set of statistics are the `ps_chi2`-type statistics. These are measures of the excess power spectrum of the data as compared to what is expected from white noise. These are calculated by making one 3D map for each of the chunks of data, taking the power spectrum of this map, and comparing that to power spectra of simulated white noise maps with the same noise levels, using a normalized χ^2 statistic. The different types of `ps_chi2` statistics corresponds to using chunks of data of different dimensions for this analysis.

The `ps_chi2` statistics measure something that is fairly close to what we are actually interested in measuring, so it serves as a great diagnostic for identifying bad data. If another statistics is correlated with `ps_chi2`, this can help us identify what is the source of the excess power.

An excellent example of this is shown in Fig. 2.23. Here we see the far sidelobe model for the instrumental beam compared to the `ps_chi2` values projected onto a map of the sun position relative to the pointing when that data was taken. We see that if the sun is in one of the 4 main far sidelobe peaks there is a large excess in the average

Table 2.2: List of statistics in the statistics database and a short description of each.

Statistic label	Description	Statistic label	Description
mjd	mean MJD of scan	night	distance in h from 2 AM (UTC - 7)
sidereal	sidereal time in degrees (up to a phase)	az	mean azimuth of scan
el	mean elevation of scan	chi2	χ^2 statistic for all timestreams of whole sb
acceptrate	acceptance rate of sideband	az_chi2	χ^2 of azimuth binned timestreams
max_az_chi2	max χ^2 of az binned single freq TODs	med_az_chi2	median χ^2 of az binned single freq TODs
fbit	feature-bit of scan (indicates scanning mode)	az_amp	average amplitude of fitted az-template
el_amp	average amplitude of fitted el-template	n_spikes	number of spikes
n_jumps	number of jumps	n_anomalies	number of anomalies
n_nan	number of nan-samples	tsys	average T_{sys} value of scan
pca1	average variance of removed PCA mode 1	pca2	average variance of removed PCA mode 2
pca3	average variance of removed PCA mode 3	pca4	average variance of removed PCA mode 4
weather	estimated probability of bad weather	kurtosis	kurtosis of timestreams
skewness	skewness of timestreams	scan_length	length of scan in minutes
saddlebag	saddlebag number (1-4)	sigma_poly0	σ of mean in poly filter
fknee_poly0	f_{knee} of mean in poly filter	alpha_poly0	α of mean in poly filter
sigma_poly1	σ of slope in poly filter	fknee_poly1	f_{knee} of slope in poly filter
alpha_poly1	α of slope in poly filter	power_mean	mean power of sideband mean
sigma_mean	σ of sideband mean	fknee_mean	f_{knee} of sideband mean
alpha_mean	α of sideband mean	airtemp	hk: air temp, C
dewtemp	hk: dewpoint temp, C	humidity	hk: relative humidity, (0-1)
pressure	hk: pressure, millibars	rain	hk: rain today, mm
winddir	hk: az from where wind is blowing, deg	windspeed	hk: windspeed m/s
moon_dist	distance to moon in deg	moon_angle	az of moon relative to pointing
moon_cent_sl	moon close to central sidelobe	moon_outer_sl	moon close to outer (feedleg) sidelobes
sun_dist	distance to sun in deg	sun_angle	az of sun relative to pointing
sun_cent_sl	sun close to central sidelobe	sun_outer_sl	sun close to outer (feedleg) sidelobes
sun_el	elevation of sun	ps_chi2	χ^2 excess in power spectrum (old)
ps_s_sb_chi2	ps_chi2 for single sb single scan	ps_s_feed_chi2	ps_chi2 for single feed single scan
ps_s_chi2	ps_chi2 for all feeds for single scan	ps_o_sb_chi2	ps_chi2 for single sb full obsid
ps_o_feed_chi2	ps_chi2 for single feed full obsid	ps_o_chi2	ps_chi2 for all feeds for full obsid
ps_z_s_sb_chi2	ps_chi2 for avg of z-direction 1D ps	ps_xy_s_sb_chi2	ps_chi2 for avg of xy-direction 2D ps
sw_01	standing wave magnitude $k \approx 0.012\,\mathrm{Mpc}^{-1}$	sw_02	standing wave magnitude $k \approx 0.017\,\mathrm{Mpc}^{-1}$
sw_03	standing wave magnitude $k \approx 0.025\,\mathrm{Mpc}^{-1}$	sw_04	standing wave magnitude $k \approx 0.036\,\mathrm{Mpc}^{-1}$
sw_05	standing wave magnitude $k \approx 0.051\,\mathrm{Mpc}^{-1}$	sw_06	standing wave magnitude $k \approx 0.073\,\mathrm{Mpc}^{-1}$
sw_07	standing wave magnitude $k \approx 0.10\,\mathrm{Mpc}^{-1}$	sw_08	standing wave magnitude $k \approx 0.15\,\mathrm{Mpc}^{-1}$
sw_09	standing wave magnitude $k \approx 0.21\,\mathrm{Mpc}^{-1}$	sw_10	standing wave magnitude $k \approx 0.30\,\mathrm{Mpc}^{-1}$
sw_11	standing wave magnitude $k \approx 0.44\,\mathrm{Mpc}^{-1}$	sw_12	standing wave magnitude $k \approx 0.62\,\mathrm{Mpc}^{-1}$
sw_13	standing wave magnitude $k \approx 0.89\,\mathrm{Mpc}^{-1}$	sw_14	standing wave magnitude $k \approx 1.3\,\mathrm{Mpc}^{-1}$

ps_chi2, indicating clear excess power during these observations. Although we had not anticipated the severity of this effect in advance, the statistics database lets us see this effect very clearly, and it also makes it much easier to reject the relevant part of the data once it has been noticed. The bottom part of the figure shows the sun-sidelobe mask used to reject data that is close to these far sidelobes, as well as the azimutally symmetric main sidelobes.

Another important effect that we have worked a lot on dealing with is weather. As we discussed in Sect. 2.4.8 weather can affect the data significantly. We have therefore implemented a weather classifier using a deep neural network, to detect weather contamination. This work was lead by master student Maren Rasmussen, under my supervision. Her results classified about 38% of the data as affected by bad weather, and this data is thus rejected. For more details and results regarding the weather classifier see Rasmussen, 2020.

2.7.2 Acceptance rates

With the help of the statistics database we set various thresholds for the different statistics, and decide what data to accept or reject based on the values of these statistics.

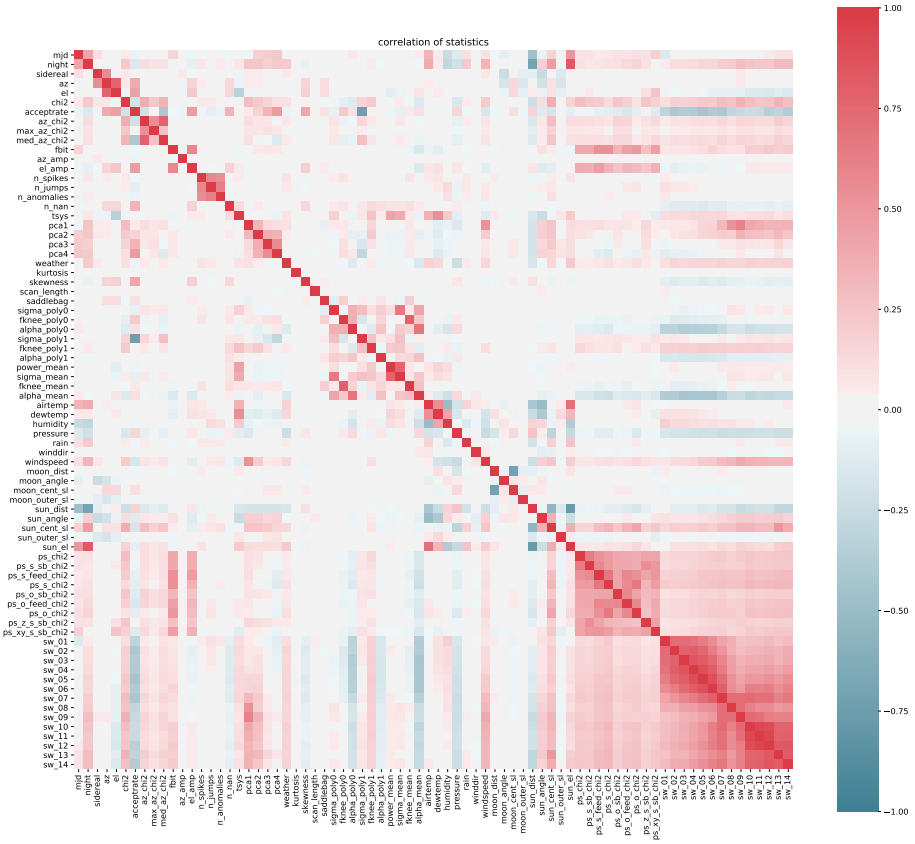

Figure 2.22: Correlation between the different diagnostic variables in the statistic database.

Figure 2.24 shows the fraction of data left over after each cut is applied for the co7 field. For the first year, we lost all the data from feeds 4, 6 and 7, due to very strong systematics, so we lose 3 out of 19 parts of the data right out of the box. After the frequency cuts in `l2gen` we are down to a bit more than 60 % of the data. After observing for close to a year, we understood that data taken at higher than 65 degrees or lower than 35 degrees in elevation was contaminated by ground signal, so we lost a significant fraction of our data from the co6 and co7 fields. After the elevation cut, we see that we are down to a bit more than 40 % of the original data volume. After this, a bunch of smaller cuts due to e.g. weather, large T_{sys} values or the Sun in the sidelobe, we are down to a bit more than 20 %. Now we do the cuts on the `ps_chi2` which gets us down to the roughly 14 % of data. This is the fraction of data that goes into the maps. The corresponding acceptance rates for co2 and co6 is given by about 22 % and 14 %, respectively.

As we see, with the current cuts, only a fraction of the original data volume is retained. We have been deliberately conservative in these cuts, in order to make sure

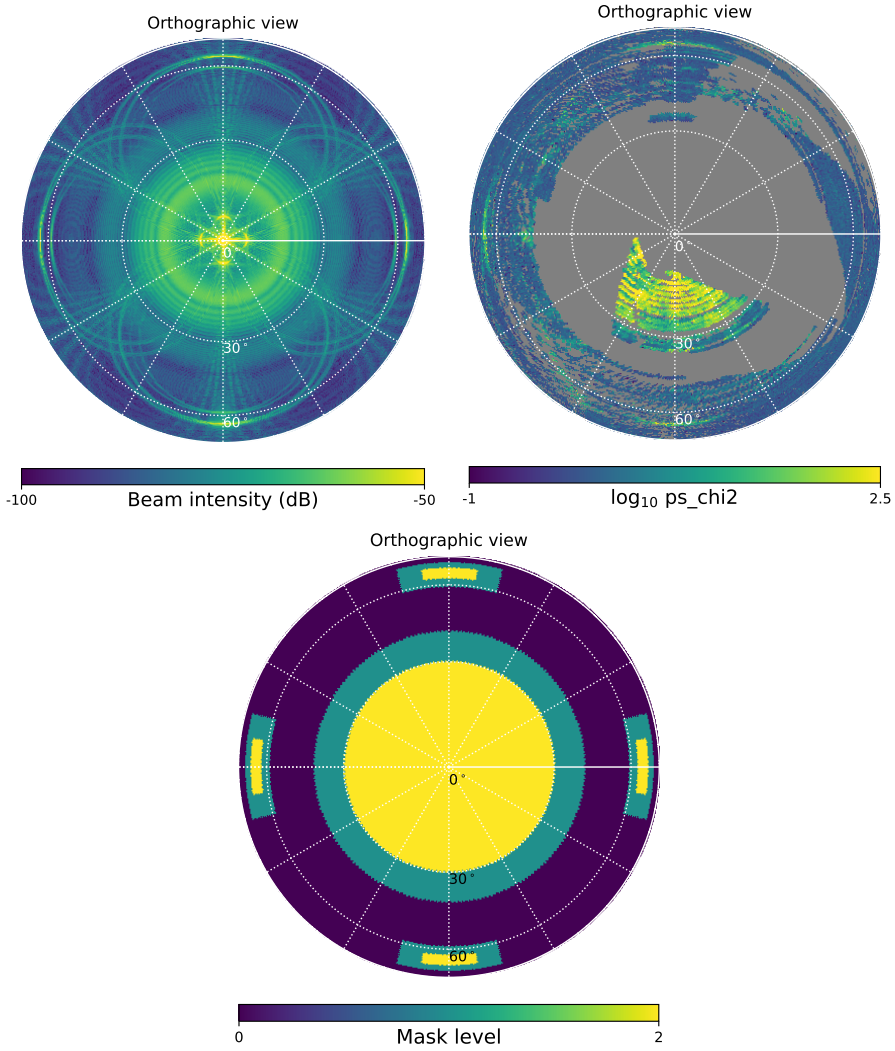

Figure 2.23: Simulated beam model for a single feed (top left), the derived statistic `ps_chi2` binned according to the position of the sun relative to the pointing when the data was taken (top right) and far sidelobe sun mask used for data selection (bottom).

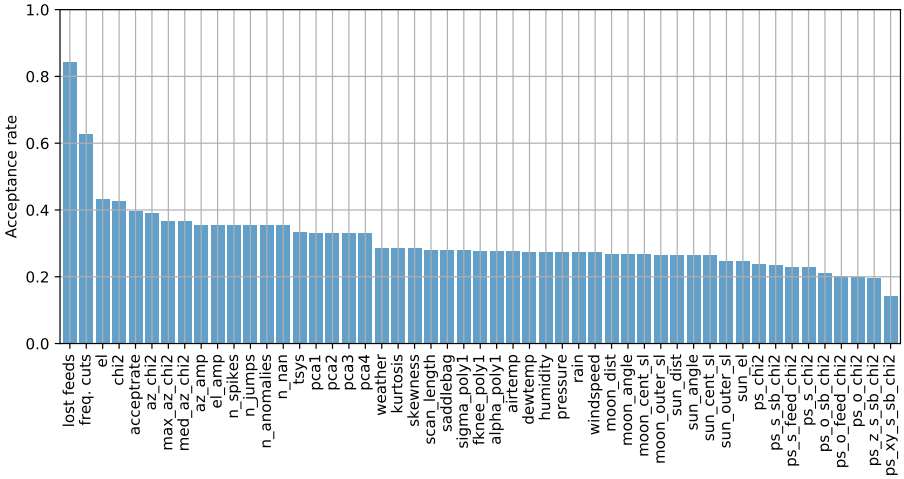

Figure 2.24: Acceptance rate of the data after each individual cut is applied for all the co7 data.

that only the best data makes it through. Once we understand the data better, and have experimented more with different cuts, we will hopefully be able to use more of the data. We now only observe between 35 and 65 degrees in elevation, so we will not lose more data to the elevation cuts in the future, and the lost feeds from the first year have been fixed, so we will be in a significantly better situation for our second year of data. With new improvements in the pipeline, we also hope to clean the data better during `l2gen`, allowing us to use more of the data still. With all this in mind we hope that we can at least increase the fraction of the data used by a factor of 2 or 3 from our current level.

2.7.3 Accept list

All the different data cuts result, in the end, in the accept list, which tells the mapmaker which parts of the data to include or not include when making the maps. The accept list is a boolean array of dimension $(n_{\text{scan}}, n_{\text{feed}}, n_{\text{sideband}})$, which stores the value "True" if the data from a given sideband on a given feed of a given scan is to be included in the map. This allows us to play around with more or less strict data cuts when we make the maps, without having to run the very expensive low level time ordered data filtering.

`accept_mod` can also split the data in any way you specify. This essentially runs the mapmaker independently for the two (or more) different sets of data and produces two (or more) different maps. This is very useful for the cross-spectrum methods we are using (see Paper II (Ihle et al., in prep) for details), and when doing null-tests, jack-knifes etc.

2.8 Inference from line intensity maps

The ultimate goal of observational astronomy and cosmology is to learn about the universe, so the final question we need to ask is how can we infer anything about astrophysics and cosmology from a line intensity map.

This problem is in general very hard. The physics involved in producing the CO signal we are looking for involves many different processes spanning so many different length scales, from large-scale clustering of dark matter, to the physics, chemistry and radiative transfer of cold molecular clouds. On the other hand, this means that there is potentially a lot of information, about all these different processes, in the map, if we can only figure out how to extract it.

2.8.1 Modeling CO line emission

Cosmological simulations of dark matter (DM) only are vastly cheaper to run than simulations including baryonic physics as well. This is because dark matter is only affected by gravity, while baryons have much more complicated interactions. Baryons form structure all the way down to stellar scales, and things like stellar feedback means that what happens at small scales affects what happens at every other scale, especially during reionization. This coupling between scales means that cosmological scale simulations including baryons cannot be run at the resolution required to include all relevant physics.

To include baryonic effects in a cosmological simulations you therefore typically need to employ a model to take into account the physics that happens at smaller scales than you can resolve. Such sub-grid models can then be calibrated using smaller scale simulations and observations, so that it captures as much as possible of the features of the small scale physics that are relevant for the large scale physics.

2.8.1.1 Dark matter halo based models

A simple approach to sub-grid modeling for CO line intensity, is to model the connection between each dark matter halo and the CO line luminosity of the galaxy (or multiple galaxies) within that halo. This way you can run DM only simulations and then replace each DM halo in the simulation with a point source of a given luminosity, which can then be turned into a 3D intensity map, by slicing the DM simulation along a lightcone.

Perhaps the simplest possible such DM halo-$L_{\rm CO}$ model would be to assume a linear relation between the DM halo mass, $M_{\rm halo}$, and the CO luminosity, $L_{\rm CO}$. This would be a one-parameter model with the amplitude of the linear relation being the only free parameter.

A generalization of this model would be a general parameterized relation between $M_{\rm halo}$ and $L_{\rm CO}$

$$L_{\rm CO}(M_{\rm halo}, \theta_i), \tag{2.39}$$

where θ_i is a set of free variables parametrizing the relation between $M_{\rm halo}$ and $L_{\rm CO}$. A more advanced model could also take into account other properties of the DM halos, like the virial velocity, the formation time or the merger history of the halo. It could

also predict not only the CO luminosity, but also for example the line width of the CO emission from each halo.

2.8.1.2 The Li et al., 2016 model

One of the most used models for CO line emission at intermediate redshifts ($z \sim 2-3$) the last years is the model developed in Li et al. (2016). This is a parametric model relating DM halo masses, $M_{\rm halo}$, star formation rates (SFR), infrared (IR) luminosities, $L_{\rm IR}$, and CO luminosities, $L_{\rm CO}$, using five free parameters. The model utilizes the relations between average SFR from DM halo masses derived in Behroozi, Wechsler, and Conroy (2013a) and Behroozi, Wechsler, and Conroy (2013b), and adds an additional log-normal scatter to the resulting SFR, determined by the parameter, $\sigma_{\rm SFR}$.

Given a SFR, the IR luminosity is then found using the linear relation

$$\mathrm{SFR} = \delta_{\rm MF} \times 10^{-10} L_{\rm IR}. \tag{2.40}$$

Further, CO-luminosities are approximated by the following relation

$$\log L_{\rm IR} = \alpha L'_{\rm CO} + \beta, \tag{2.41}$$

before a final step of log-normal scatter is added determined by the parameter, $\sigma_{L_{\rm CO}}$. We then get a total of five parameters $\theta = \{\log \delta_{\rm MF}, \alpha, \beta, \sigma_{\rm SFR}, \sigma_{L_{\rm CO}}\}$, determining the function $L_{\rm CO}(M_{\rm halo})$. See Li et al. (2016) for a detailed discussion of the physical and observational motivation for this model.

2.8.1.3 Simulations

Another approach to modeling CO line intensity is to use smaller hydrodynamical simulations (Hopkins et al., 2014; Schaye et al., 2014; Vogelsberger et al., 2014). These resolve more of the physics related to molecular clouds, star-formation, feedback etc, but they typically cover a much smaller volume than what is observed with COMAP. In order to do forecasts, or use these simulations to interpret the results from a CO intensity map, we would need to somehow extrapolate these results to the full volume, which is hard to do in a statistically rigorous way. For some work along these lines, see Lakhlani (2019) or M. Silva et al. (in prep).

2.8.2 Observables

When doing inference from a line intensity map, the general approach is to have some kind of astrophysical and cosmological model which predicts how we would expect the cosmological signal to look like, and then compare this to the observed map. Of course, we don't expect the model to predict the specific spatial distribution of the CO signal, since that depends on the random initial conditions. Rather, we expect the model to predict the statistical properties of the ensemble that any observed CO intensity map is a specific realization of. This means that instead of directly comparing the map itself to some model prediction, we typically calculate some summary statistics, or observables, from the map, which hopefully captures the statistical information in the map. We can

then try to derive predictions for these different observables from the model, and then compare these predictions to the values derived from the observed map.

In CMB analysis, and (to some degree) analysis of large scale structure, where the signal is extremely close to a Gaussian field, the power spectrum contains all the statistical information in the map. A line intensity map, however, is often significantly non-Gaussian, which means that new statistics need to be developed to extract all the information present in the map.

2.8.2.1 Power spectrum

The power spectrum, $P(\boldsymbol{k})$, sets the variance of the Fourier components of a map

$$P(\boldsymbol{k}) = \frac{V_{\rm vox}}{N_{\rm vox}} \langle |f_{\boldsymbol{k}}|^2 \rangle, \tag{2.42}$$

where k is the wave vector of a given Fourier component, $V_{\rm vox}$ is the volume of each voxel of the map and $N_{\rm vox}$ is the total number of voxels in the map. A voxel is the 3D equivalent of a pixel, it is the small volume element defined by the resolution used for the 3D map.

If the Fourier components have a Gaussian distribution, then the map is called a Gaussian map, in which case the power spectrum fully describes the statistical ensemble that a given map is a single realization from. In the more general case, where the map is not Gaussian, the power spectrum still carries a lot of important information about the map, even if it does not describe all the information.

The cross spectrum, $C(\boldsymbol{k})$, sets the covariance between the Fourier components of two different maps, m_1 and m_2

$$C(\boldsymbol{k}) = \frac{V_{\rm vox}}{N_{\rm vox}} \left\langle \mathrm{Re}\{f^*_{1\boldsymbol{k}} f_{2\boldsymbol{k}}\} \right\rangle \tag{2.43}$$

where f_1 and f_2 are the Fourier components of the two maps and where Re{} denotes the real part of a complex number or function.

For more details on different power spectrum methods and how to robustly estimate the signal power spectrum from an observed line intensity map see Paper II (Ihle et al., in prep). For other recent examples of the use power spectrum analysis on intensity mapping data see e.g. (Keating, Marrone, Bower, and Keenan, 2020; Mertens et al., 2020; Uzgil et al., 2019).

2.8.2.2 Voxel Intensity Distribution (VID)

A useful statistic that is complementary to the power spectrum is the Voxel Intensity distribution (VID) (P. Breysse et al., 2017). The VID, $\mathcal{P}(T)$, is the probability distribution of voxel brightness temperatures. This observable does not use any of the spatial information in the map, but is more sensitive to the bright end of the underlying CO luminosity function and small-scale clustering.

A natural observable related to the VID is the histogram of voxel temperatures, B_i. The expectation value of these are given by the VID

$$\langle B_i \rangle = N_{\rm vox} \int_{T_i}^{T_{i+1}} \mathcal{P}(T) dT, \tag{2.44}$$

where B_i is the number of voxels with a brightness temperature between T_i and T_{i+1} and $N_{\rm vox}$ is the total number of voxels.

2.8.2.3 Pseudo-VID

In the same way as for the auto-spectrum, when a line intensity map has an uneven distribution of noise, then the VID statistic will be dominated by the highest noise parts of the map, so unless you have a large signal to noise, this will affect the usefulness of the statistic significantly. We therefore want to think of a statistic that preserves the nice properties of the VID, but that is less susceptible to uneven noise distribution in the map.

In order to investigate this let us use the following simplified data model

$$d = n + s, \tag{2.45}$$

where d is a vector of the measured temperature in each voxel, n is a vector of the noise in each voxel and s is the signal in each voxel. For simplicity we assume that the noise level, $\sigma_{\rm vox}$, in each voxel is known and that the signal is independent between different voxels.

Let us then define the pseudo-VID as the VID of the following data

$$\tilde{d} = \frac{d}{\sigma_{\rm vox}^{\alpha}}, \tag{2.46}$$

where α is some positive power that we will try to find an optimal value for.

If the signal was just an overall constant number, then $\alpha = 2$ is gives us the maximum likelihood estimate of s, so this is perhaps not a bad initial guess for what value of α to choose. However, this is not a very realistic model of the signal. Another simple model for the signal, which is more realistic as a toy model, is to assume that the signal is an independent Gaussian with a constant standard deviation[4], $\sigma_{\rm signal}$. We can then try to find what is a good value for α to get the best signal to noise estimate of the signal in this model.

We can make a mock observation by generating noise and signal for 10^6 voxels and dividing the data by $\sigma_{\rm vox}^{\alpha}$ and then binning it to obtain the pseudo bin counts, $B_i(\tilde{d})$. The pseudo bin counts follow (under our assumptions) a binomial distribution, and we can thus map out the likelihood of the data, $P(\tilde{d}|\sigma_{\rm signal})$, by using many simulations to determine the expectation, $\langle B_i(\tilde{d})\rangle$, for each value of $\sigma_{\rm signal}$. We can then examine the width of this likelihood as a function of $\sigma_{\rm signal}$ and use this as an estimate of the sensitivity of the pseudo-VID statistic for a given value of α.

Figure 2.25 shows the sensitivity of the pseudo-VID statistics for different values of the exponent, α. These results are obtained in the noise dominated limit (with the signal component of the mock data set to zero), and we show results for different noise level distributions to investigate the effect of the distribution of the noise levels has on the optimal value for α. We find that values in the range $\alpha = 1 - 2$ give consistent good results. Figure 2.26 shows the results from the same analysis using noise level

[4]My work on this toy model is building on some previous work by Patrick Breysse on these issues.

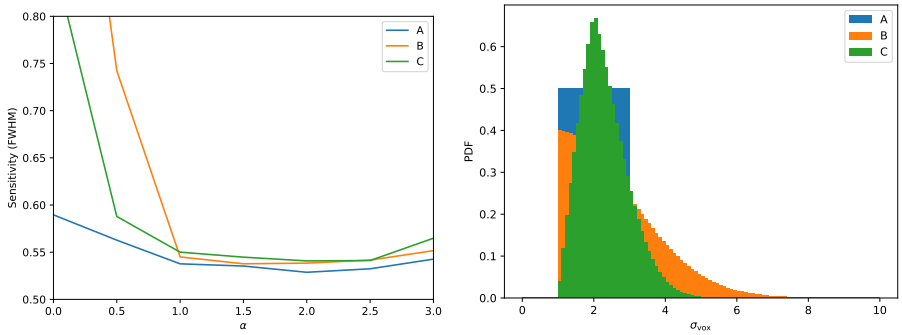

Figure 2.25: Sensitivity of the pseudo-VID statistic for different exponents, α, (left) and the three different noise level distributions used (right).

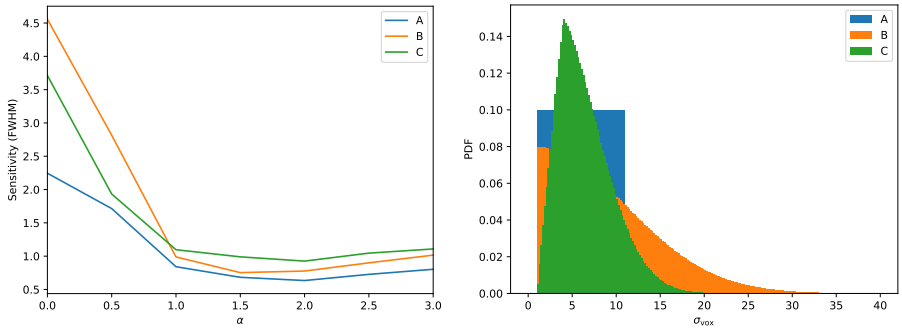

Figure 2.26: Sensitivity of the pseudo-VID statistic for different exponents, α, (left) and the three different noise level distributions used (right).

distributions with significantly more variance, as can be seen in the right panel. The optimal values for α found here are fairly consistent in all cases, suggesting that values between $\alpha = 1 - 2$ are good values for the pseudo-VID in a large range of different noise level distributions.

2.8.3 Likelihood

We can combine all observables into a data vector, d_i. For the case of using the PS and the VID we get

$$d_i = (P_{k_i}, B_i). \tag{2.47}$$

If all the components of d_i were independent, they would (under some simplifying assumptions) have the following variance, which we will denote as the *independent variance*

$$\mathrm{Var_{ind}(P_{k_i})} = \langle P_{k_i} \rangle^2 / N_{\mathrm{modes}}, \tag{2.48}$$

$$\mathrm{Var_{ind}(B_i)} = \langle B_i \rangle. \tag{2.49}$$

2.8.3.1 Gaussian likelihood

Because of spatial clustering, there typically are correlations between the different elements of the data vector. We can take this into account using a full covariance matrix

$$\xi_{ij} = \mathrm{Cov}(d_i, d_j). \tag{2.50}$$

We can then build a Gaussian likelihood of the form (up to a constant)

$$-2\ln P(d|\theta) = \frac{N_s}{N_s+1}\sum_{ij}[d_i - \langle d_i\rangle](\xi^{-1})_{ij}[d_j - \langle d_j\rangle] + \ln|\xi|, \tag{2.51}$$

where $\langle d\rangle(\theta)$ and $\xi(\theta)$ are the mean values and covariance matrix of the observables d_i for specific parameters θ. N_s is the number of simulations used to estimate $\langle d\rangle$, and the factor $N_s/N_s + 1$ takes into account the effect of the uncertainty in the estimate of $\langle d\rangle$.

2.8.3.2 Binomial VID likelihood

In the previous section, we implicitly assumed that all the bin counts, B_i, are Gaussian random variables, with some covariance matrix. However, as we already know, even if the samples are completely independent, the bin counts will follow a binomial distribution. When the expected bin counts are high ($B_i \gg 5$), the binomial distribution can be very well modeled by a Gaussian. When we only expect a few voxels in each bin, however, the Gaussian approximation breaks down.

If we know that the samples are independent, there is no problem, since the likelihood is just given by the joint distribution of n_{bins} independent binomial variables

$$P(B|\theta_l) = \prod_{i=1}^{n_{\mathrm{bins}}}\binom{N_{\mathrm{samples}}}{B_i}p_i^{B_i}(1-p_i)^{N_{\mathrm{samples}}-B_i}, \tag{2.52}$$

where $p_i \equiv p_i(\theta_l) = \langle B_i\rangle(\theta_l)/N_{\mathrm{samples}}$ is the probability of a voxel being placed in bin i. As we have argued, however, when we are not completely noise dominated, the samples will not be independent, and we have no general way to calculate the likelihood. On the other hand, we expect the clustering effect to be dominated by the shot noise when we get to bins with very few samples.

This suggests a possible approximation we can make. We can try to separate the bins according to the expected number of samples. For bins where we expect many samples, $B_i > 5$, we can assume that they are approximately Gaussian, and use simulations to estimate the covariance matrix between these bins. For the bins with few expected samples $B_i \leq 5$ we can assume that they follow independent binomial distributions, because the variance will presumably be dominated by shot noise, and clustering effects are (hopefully) negligible. The full likelihood, in this approximation, is then given by

$$P(B|\theta_l) = P_{\mathrm{Binomial}}(B|\theta_l)P_{\mathrm{Gaussian}}(B|\theta_l), \tag{2.53}$$

where

$$P_{\mathrm{Binomial}}(B|\theta_l) = \prod_{\{i|B_i\leq 5\}}^{n_{\mathrm{bins}}}\binom{N_{\mathrm{samples}}}{B_i}p_i^{B_i}(1-p_i)^{N_{\mathrm{samples}}-B_i}, \tag{2.54}$$

and

$$P_{\text{Gaussian}}(B|\theta_l) = \frac{1}{(2\pi)^{\frac{n_{\text{bins}}}{2}}|\xi_{>5}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\chi^2_{>5}\right), \tag{2.55}$$

and where

$$\chi^2_{>5} \equiv \sum_{\{i,j|B_i,B_j>5\}} [B_i - \langle B_i\rangle(\theta_l)](\xi^{-1})_{ij}[B_j - \langle B_j\rangle(\theta_l)] \tag{2.56}$$

and

$$\xi_{>5} \equiv \xi_{\{i,j|B_i,B_j>5\}}. \tag{2.57}$$

One issue associated with bins with very few expected samples is that if you are estimating p_i from some limited number of simulations, you will sometimes have no samples in the bin from either simulation. In that case it is not clear how to evaluate the likelihood. This problem can occur quite frequently when there is not that much signal to noise in the data, and you are exploring a large region in parameter space. We can deal with this problem by imposing some prior on the expectation value of the binomial distribution, and then update this prior with the values from the simulation. The details for how to do this are discussed in App. A.1.

2.8.4 Exploring the posterior distribution

Using the methods we have sketched in the previous sections we can now try to put them all together. To summarize what we need to do this, here is a list:

- A set of observables, d_i, that can be estimated from a line intensity map.
- A model with parameters θ, which can generate simulated intensity maps, or at least in some way calculate the mean observables, $\langle d_i\rangle(\theta)$, covariance matrix, $\xi_{ij}(\theta)$ (or any other quantity needed in the likelihood) for any given values of θ. Any prior on θ is also assumed to be part of the model.
- A likelihood $P(d_i|\theta)$, that can be evaluated using the model output.
- Some observed (or simulated) line intensity mapping data, summarized in the observables data, d_i^{observed}.

We can then use a sampling procedure, like a Markov-Chain Monte Carlo (MCMC) method, or other methods to explore the posterior distribution $P(\theta|d_i^{\text{observed}})$.

At this abstract level this looks fairly simple, however, the details matter a lot here, and can make things much more complicated. There are several important questions about the model and the parameter space θ. How degenerate are the parameters θ? And how easy are the parameters to interpret? A given set of observed data d_i^{observed} is often good at constraining one combination of parameters, but is completely insensitive to other combinations of parameters, so if you make bad choices of parameters, or simply just use too many parameters, you will not be able to use the data to constrain much. If you make a model that is fairly closely related to the physics, and where each parameter has a simple interpretation, then you typically get a complicated model, with lots of parameters, and also lots of degeneracy between the parameters. A simple

model, however, can avoid any unnecessary degeneracies, but it will in turn often be much harder to connect these parameters back to cosmology and astrophysics. A good model will be one that finds a good compromise between these considerations, for the particular dataset you are considering. For a detailed discussion of these issues and more see D. Chung et al., in prep.

Another set of issues relate to the observables, d_i, and the likelihood. In Ihle et al. (2019) we show that using a combination of the power spectrum and the voxel intensity distribution is better at constraining the CO luminosity function than using either observable individually, for a COMAP-like experiment. While this was a very useful finding, there are many simplifying assumptions made in these (and similar types of) forecasts. If we want to do a similar analysis on actual data from an experiment, we will have to address many issues first. How do you model or estimate the noise in the data? How do you deal with an uneven noise distribution in the map? How do you build a good likelihood function? How do you take into account systematic effects and propagate systematic uncertainties into the likelihood? Some of these issues are discussed in detail for the power spectrum in Paper II (Ihle et al., in prep), and we have made some preliminary investigation into some issues with the VID here, in Sects. 2.8.2.3 and 2.8.3.2. However, we still have quite a way yet to go in order to use the VID on real data.

2.8.5 Future Prospects

CO intensity mapping is in a somewhat special position in that there is a fairly low level of foregrounds, and no significant interloper lines (D. Chung et al., 2017). This makes the analysis and inference from these maps much simpler.

CII intensity mapping, for example, has several interloper CO-lines from lower redshifts. This makes the inference much more complicated (Cheng, Chang, and Bock, 2020), but significantly increases the amount of information in the maps.

As the field matures, more areas of the sky, and more ranges in redshift, will be covered by different experiments, meaning that we will have access to the same galaxy population and large scale structure from multiple tracers. This will allow us to get more robust results, break degeneracies and give unique insights into the astrophysics of galaxies and their chemistry and gas dynamics.

Cross-correlation (D. T. Chung et al., 2019) as well as new tools being developed (Bernal, P. C. Breysse, Gil-Marin, et al., 2019; P. C. Breysse, Anderson, and Berger, 2019; Cheng, Chang, and Bock, 2020; D. T. Chung, 2019; Gong, Chen, and Cooray, 2020; Yang et al., 2020), will be needed to deal with these combined datasets. As we get more different datasets that we want to analyze together the more important it becomes to use a consistent statistical framework for inference, like we use in Paper III (Ihle et al., 2019) and like we have discussed here.

The prospects for doing fundamental cosmology with LIM is very interesting. Getting access to the largely unmeasured large scale modes from the early universe will fill in many gaps in our knowledge of the expansion and evolution of the universe (Bull et al., 2015; Creque-Sarbinowski and Kamionkowski, 2018; Dinda, Sen, and Choudhury, 2018; Bernal, P. C. Breysse, and E. D. Kovetz, 2019; Furlanetto et al., 2019; Liu and

P. C. Breysse, 2020). The ultimate dream, of course, is going to space (M. B. Silva et al., 2019).

Chapter 3

BeyondPlanck

BeyondPlanck (BeyondPlanck Collaboration, 2020) is an ambitious project to develop an end-to-end Bayesian data analysis pipeline for CMB experiments, combining everything from low-level instrumental parameters like gains and bandpasses to foreground components and CMB power spectra into one consistent statistical model. As a demonstration, we apply this framework to the Planck Low Frequency Instrument (LFI) time ordered data. Here I will give a short summary of the general BeyondPlanck approach, and of the two papers that I have been most involved in, on noise modeling (Paper V, Ihle et al., 2020) and calibration (Paper VI, Gjerløw et al., 2020).

3.1 The BeyondPlanck approach

The first part of the BeyondPlanck approach is to define an explicit statistical model of the data, including all the components and effect that we want to take into account. In BeyondPlanck, for each of the three LFI bands, at frequencies of 30, 44, and 70 GHz, we model the raw time-domain data, $\boldsymbol{d}$, as follows,

$$d_{j,t} = g_{j,t}\mathsf{P}_{tp,j}\left[\mathsf{B}^{\mathrm{symm}}_{pp',j}\sum_c \mathsf{M}_{cj}(\beta_{p'},\Delta^j_{\mathrm{bp}})a^c_{p'} + \mathsf{B}^{\mathrm{asymm}}_{pp',j}\left(s^{\mathrm{orb}}_{j,t}+s^{\mathrm{fsl}}_{j,t}\right)\right] + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t}. \quad (3.1)$$

The subscript t is the index in time domain; j is the radiometer index; p is the pixel index; c is the signal component index; $\boldsymbol{g}$ is the gain; P is the pointing matrix; $\mathsf{B}^{\mathrm{symm}}$ and $\mathsf{B}^{\mathrm{asymm}}$ is the symmetric beam matrix and the asymmetric beam matrix, respectively; $\boldsymbol{a}$ are the astrophysical signal amplitudes; $\boldsymbol{\beta}$ are the corresponding spectral parameters; Δ_{bp} are the bandpass corrections; M_{cj} is the bandpass-dependent component mixing matrix; $\boldsymbol{s}^{\mathrm{orb}}$ is the orbital dipole; $\boldsymbol{s}^{\mathrm{fsl}}$ are the far sidelobe corrections; $\boldsymbol{n}^{\mathrm{corr}}$ is the correlated noise; and $\boldsymbol{n}^{\mathrm{w}}$ is the white noise. For more details and discussion of each of these parameters, and how they are modeled, see BeyondPlanck Collaboration (2020) and companion papers.

Given this model, the goal of the Bayesian approach is to explore the full posterior distribution of all the free parameters in the model, given the observed data, $\boldsymbol{d}$,

$$P(\boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta, C_\ell \mid \boldsymbol{d}). \quad (3.2)$$

This is a really complicated distribution, with a very large number of parameters that depend on each other in complicated ways. The way we deal with this distribution is to use Gibbs sampling. The idea of Gibbs sampling is simple. Instead of sampling directly from the full distribution, we sample iteratively from each of the conditional distributions. In this way we can divide the complex problem of sampling from the whole distribution into a set of smaller manageable steps. We can summarize this

process as follows,

$$\begin{aligned}
\boldsymbol{g} &\leftarrow P(\boldsymbol{g} \mid \boldsymbol{d}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (3.3)\\
\boldsymbol{n}_{\rm corr} &\leftarrow P(\boldsymbol{n}_{\rm corr} \mid \boldsymbol{d}, \boldsymbol{g}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (3.4)\\
\xi_n &\leftarrow P(\xi_n \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \Delta_{\rm bp}, \boldsymbol{a}, \beta, C_\ell) && (3.5)\\
\Delta_{\rm bp} &\leftarrow P(\Delta_{\rm bp} \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \boldsymbol{a}, \beta, C_\ell) && (3.6)\\
\beta &\leftarrow P(\beta \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, C_\ell) && (3.7)\\
\boldsymbol{a} &\leftarrow P(\boldsymbol{a} \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, \beta, C_\ell) && (3.8)\\
C_\ell &\leftarrow P(C_\ell \mid \boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\rm corr}, \xi_n, \Delta_{\rm bp}, \boldsymbol{a}, \beta). && (3.9)
\end{aligned}$$

Here, $\leftarrow$ indicates sampling from the distribution on the right-hand side. The challenge then is to implement methods to sample from each of the conditional distributions, and to put everything together into a computational framework.

Perhaps the simplest way to illustrate the appeal of the full end-to-end Bayesian approach of BeyondPlanck is shown in Fig. 3.1. Here we see constraints on the optical depth to reionization, τ, derived in four different cases. The blue curve, labeled "WN", corresponds to fixing all the other parameters in the model, and only taking into account the uncertainty coming from the white noise. The red curve, labeled "TOD + WN", corresponds to fixing the foreground model, but marginalizing over all the instrumental parameters and the correlated noise in addition to the white noise. The green curve, labeled "FG + WN", marginalizes over the foreground parameters and the white noise, but fixes the instrumental parameters and the correlated noise, while the black curve marginalizes over all the parameters, in addition to the white noise. This figure illustrates how important it is to propagate the uncertainties from all the instrumental parameters and the foregrounds simultaneously, and in a statistically rigorous way, which is exactly what the BeyondPlanck pipeline was designed for.

3.2 Noise Modelling and Characterization

Here we will discuss two of the sampling steps, 3.4 and 3.5, from the full BeyondPlanck Gibbs chain in more detail.

The starting point for both conditional distributions is the following parametric data model,

$$\boldsymbol{d}' \equiv \boldsymbol{d} - g\boldsymbol{s}^{\rm tot} = \boldsymbol{n}^{\rm corr} + \boldsymbol{n}^{\rm wn}. \tag{3.10}$$

where we have defined the signal subtracted data, $\boldsymbol{d}'$, and where $\boldsymbol{d}$ is the raw time ordered data (TOD); g is the gain; $\boldsymbol{s}^{\rm tot}$ describes the total sky signal, comprising both CMB and foregrounds, projected into time-domain; $\boldsymbol{n}^{\rm corr}$ represents the correlated noise in time domain; and $\boldsymbol{n}^{\rm wn}$ is white noise.

We assume that both the two noise terms are distributed as a Gaussian, with the following covariance matrices $\mathsf{N}_{\rm corr} \equiv \langle \boldsymbol{n}_{\rm corr}\boldsymbol{n}_{\rm corr}^T\rangle$ and $\mathsf{N}_{\rm wn} \equiv \langle \boldsymbol{n}_{\rm wn}\boldsymbol{n}_{\rm wn}^T\rangle$, respectively. The complete noise power spectral density (PSD) is then given by

$$P(f) = \mathsf{N}_{\rm wn} + \mathsf{N}_{\rm corr} = \sigma_0^2 + \sigma_0^2\left(\frac{f}{f_{\rm knee}}\right)^{\alpha}, \tag{3.11}$$

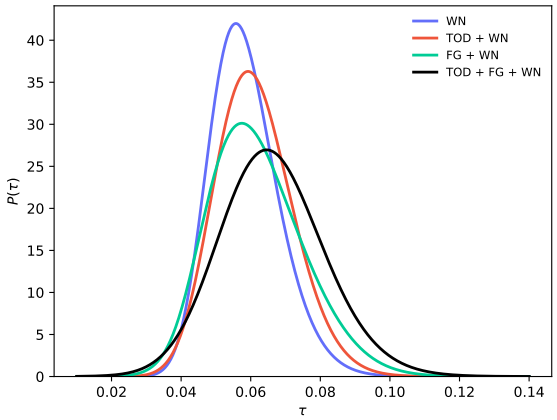

Figure 3.1: Posterior distribution derived for the optical depth to reionization, τ, marginalizing over different combinations of variables. Figure taken from (Paradiso et al., 2020).

where f denotes temporal frequency; σ_0 is the white noise level of the time-ordered data; α is the slope (typically negative) of the correlated noise PSD; and the knee frequency, $f_{\rm knee}$, denotes the (temporal) frequency at which the variance of the correlated noise is equal to the white noise variance. The three PSD parameters are collectively denoted $\xi_n = \{\sigma_0, f_{\rm knee}, \alpha\}$. This noise model is usually referred to as $1/f$-noise (Planck Collaboration II, 2014; Planck Collaboration II, 2020; Tauber et al., 2019).

We can sample the correlated noise, $\boldsymbol{n}^{\rm corr}$, by solving the following equation,

$$\left(\mathsf{N}_{\rm corr}^{-1} + \mathsf{N}_{\rm wn}^{-1}\right)\boldsymbol{n}^{\rm corr} = \mathsf{N}_{\rm wn}^{-1}\boldsymbol{d}' + \mathsf{N}_{\rm wn}^{-1/2}\boldsymbol{\eta}_1 + \mathsf{N}_{\rm corr}^{-1/2}\boldsymbol{\eta}_2, \tag{3.12}$$

where $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are two independent vectors of random variates drawn from a standard Gaussian distribution, $\boldsymbol{\eta}_{1,2} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$. Since we typically use a time-domain mask, this equation cannot be solved by simply going to Fourier space, but needs to be solved by a more general method. For a nice method to solve the problem in this case, see Keihänen et al. (2020). Figure 3.2 shows some examples of how the correlated noise solutions look like. We see that the $1/f$ model seems to be working quite well.

Once we have found the correlated noise, the noise PSD parameters, $f_{\rm knee}$ and α, are sampled from their exact conditional distributions

$$-\ln P(f_{\rm knee}, \alpha \mid \sigma_0, \boldsymbol{n}^{\rm corr}) = \sum_{f=f_{\rm min}}^{f_{\rm max}} \left[\frac{|n_f^{\rm corr}|^2}{\mathsf{N}_{\rm corr}(f)} + \ln \mathsf{N}_{\rm corr}(f)\right] - \ln P(f_{\rm knee}, \alpha), \tag{3.13}$$

where $\mathsf{N}_{\rm corr}(f) = \sigma_0^2\left(\frac{f}{f_{\rm knee}}\right)^{\alpha}$ and $P(f_{\rm knee}, \alpha)$ is an optional prior.

The LFI data are divided into roughly 45 000 pointing periods, denoted PIDs (pointing ID), each with a duration of 30–60 minutes. The official *Planck* LFI Data

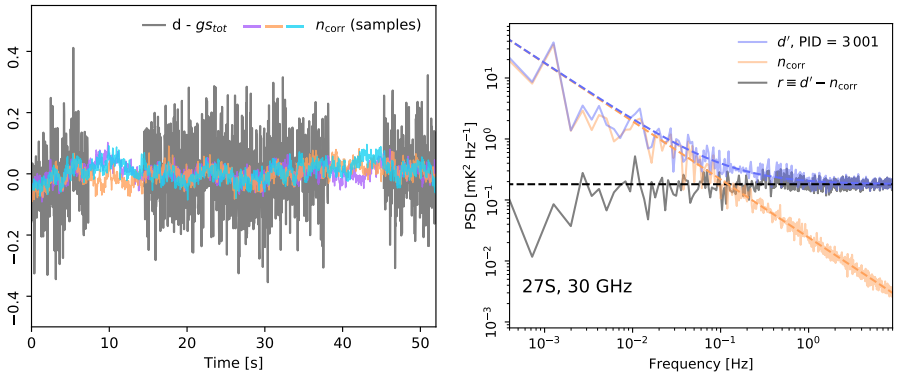

Figure 3.2: Example correlated noise samples in time domain (left) and as PSD (right).

Processing Center (DPC) (DPC, Planck Collaboration II, 2020) assumed that the noise parameters were constant throughout the entire mission. In our analysis, we sample the correlated noise, and the noise parameters independently for each PID, but we use the DPC results as priors for $f_{\rm knee}$ and α. In this way we use some information from the entire mission, but we also allow enough freedom to account for actual changes in noise properties over time.

The results we have found are very interesting. In addition to giving us a whole new insight into the noise properties of the LFI instrument, this detailed work on the noise modeling also highlights some important systematic effects. We demonstrate, for the first time, that the noise properties of most radiometers do indeed change significantly over time, mostly due to changes in the thermal environment of the instrument. Figure 3.3 shows very clearly the intimate relationship between the noise properties of the 70 GHz radiometers and the temperatures measured at the 20K cooling stage.

Although the noise properties of the 70 GHz radiometers change significantly over time, we are still able to fit the noise well using the $1/f$ model. At 30 and 44 GHz, however, some radiometers show signs of a significant power excess at intermediate timescales (0.1–10 Hz), which do not fit well with the $1/f$ model, as is seen in Fig 3.4. This is quite worrisome, since these timescales correspond roughly to 1–60 degrees on the sky, which is in the prime science range, where we don't want an unfitted systematic sloshing around in the model. This is something that will need more work, perhaps with a modified model for the noise PSD.

3.3 Calibration

The calibration step, Eq. 3.3, is a crucial step in any pipeline. We will now describe the approach we use for calibration in BeyondPlanck. We start with the data model,

$$\boldsymbol{d} = g\boldsymbol{s}^{\rm tot} + \boldsymbol{n}^{\rm corr} + \boldsymbol{n}^{\rm wn}. \tag{3.14}$$

The goal is to convert the raw time ordered data, $\boldsymbol{d}$, from voltages to Kelvin, effectively determining g. If we had a signal of known magnitude, then we could estimate g directly.

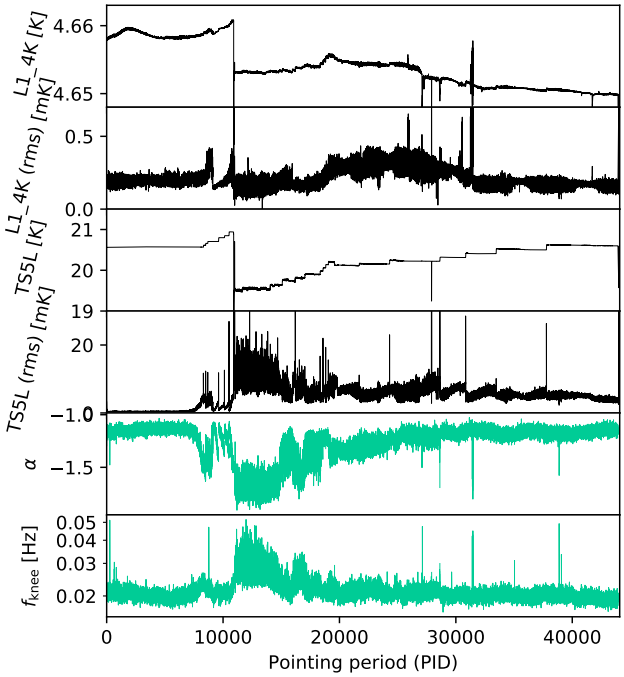

Figure 3.3: Average correlated noise properties of the 70 GHz radiometers (bottom two panels) compared with 4 K (top two panels) and 20 K (middle two panels) temperature sensor read-outs for the full mission.

Thankfully we have a very strong dipole signal which each detector measures, typically with a large signal-to-noise ratio, every minute, as the Planck scanning strategy orbits a great circle on the sky every sixty seconds. The problem, however, is that the overall magnitude of the dipole is not something we know, but one of the things we are trying to measure. On the other hand, there is one signal that we do know very precisely, and that is the component of the overall dipole coming from the movement of the Planck satellite relative to the sun, called the *orbital* dipole. The, much stronger, but unknown, dipole signal from the movement of the sun relative to the CMB rest frame, we denote as the *solar* dipole (see Fig. 3.5).

The idea then is to use the orbital dipole for the overall calibration of the entire system, but to use the solar dipole to measure the relative calibration of the different detectors and the change over time of single detectors. We therefore decompose the gain, g, into the following parts,

$$g_{t,i} = g_0 + \Delta g_i + \delta g_{t,i}, \tag{3.15}$$

where t is a time index and i is an index labeling the different detectors of a given frequency band. g_0 is then the absolute calibration, Δg_i, is the detector specific offset and $\delta g_{t,i}$ tracks the time evolution of the gain for each specific detector. The two last

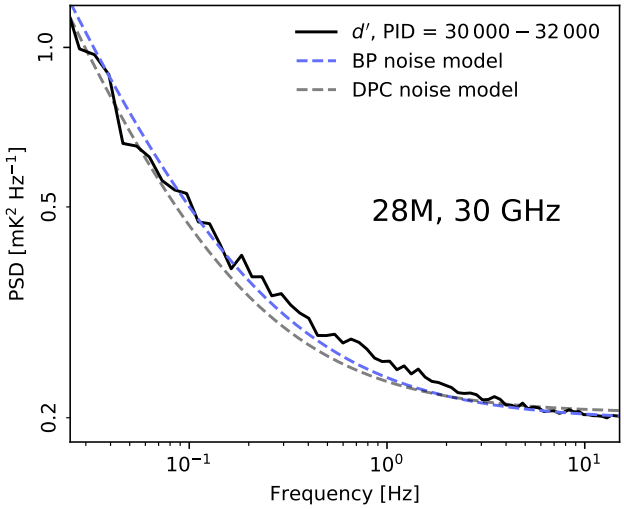

Figure 3.4: PSD of signal-subtracted data from radiometer 28M, averaged over 18 PIDs (black). Compared to the BeyondPlanck (dashed blue) and LFI DPC (dashed gray) noise models.

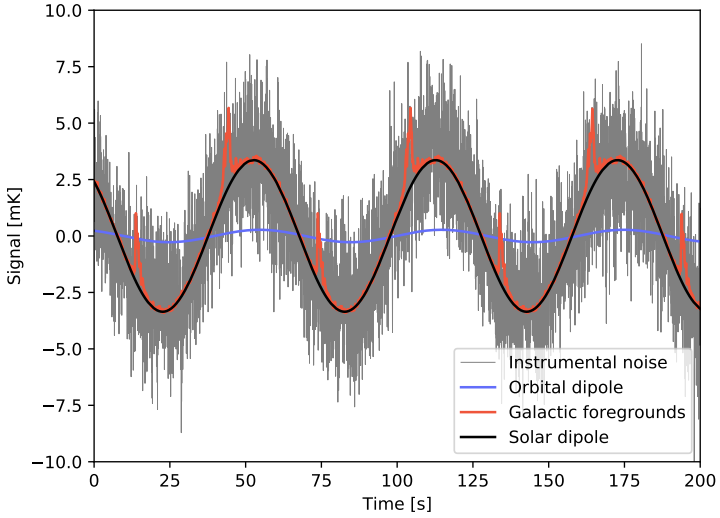

Figure 3.5: Different contributions to the 30 GHz time-ordered data. The orbital and the solar dipole are shown in blue and black, respectively, while the red line shows the modeled small-scale CMB fluctuations and Galactic foregrounds. The residual, given by the instrumental noise, is shown in gray. Figure courtesy of Eirik Gjerløw.

terms are subject to the constraints $\sum_i \Delta g_i = 0$ and $\sum_t \delta g_{t,i} = 0$. We can then use all the data, from all of the detectors, to determine g_0 using the known orbital dipole, and use the full signal (dominated by the solar dipole) to determine Δg_i and $\delta g_{t,i}$ subject to the constraints. See Paper VI (Gjerløw et al., 2020) for the detailed sampling steps, which we will not repeat here.

There are several challenges we need to deal with during calibration. First of all, there is typically not enough signal to noise in a single PID to measure the time dependent gain, $\delta g_{t,i}$, precisely enough. To solve this we apply a smoothing window, essentially using data from a few hundred consecutive PIDs to measure the time dependent gain. Another complication comes from the fact that for some periods the Planck scanning is almost completely orthogonal to the dipole, meaning that it is very hard to use the dipole for calibration. To deal with this we need to periodically employ very wide smoothing windows, making us less sensitive to changes in the gain for these periods. There are also abrupt jumps in the gain, causing the smoothing window approach to break down. If we know when these jumps happen, we can explicitly include them in the model, and fit the gain independently on each side of the jump. Although we know the origin of some of the gain jumps, we do not know the origin of all of them. This leaves us vulnerable to the presence of gain-jumps that we have not included in our model. Any such missing gain-jumps could lead to significant gain errors effecting the entire model (Eq. 3.1).

Figure 3.6 shows the correlated noise in map domain for the Stokes Q 44 GHz channel. We see some very clear stripes along an orbit from the middle of the map and up to the right. While we do not know the origin of these stripes, we suspect they are related to the gain, and quite possibly unmodeled gain-jumps in particular. This is an area we are still working to understand better using the large library of Planck housekeeping data. The stripes we see in the southern hemisphere of the 44 GHz band may be related to the problems we have had using the CMB polarization data from the southern hemisphere (Colombo et al., 2020; Paradiso et al., 2020), which means that this is one of the most important things we are working to figure out (see Paper VI (Gjerløw et al., 2020) for more discussion).

3.4 Future prospects

The future of the CMB is in polarization, particularly in B-mode polarization from primordial gravitational waves. These signals are very weak, which means that we will need to dig deep into both foregrounds and instrumental systematic effects in order to detect it. This is the challenge that BeyondPlanck has set out to tackle, to build a consistent statistical framework to deal with the complex interactions and degeneracies between instrumental effects, foregrounds and the cosmological signal.

The BeyondPlanck results presented here, analysing the Planck LFI data, is a proof of concept, demonstrating that end-to-end analysis from time ordered data to cosmological parameters is actually possible, and that this approach has several advantages over the more traditional, modular, approach. In the future, under the umbrella of *Cosmoglobe*[1], we want to combine as many datasets as possible into the

[1] http://cosmoglobe.uio.no

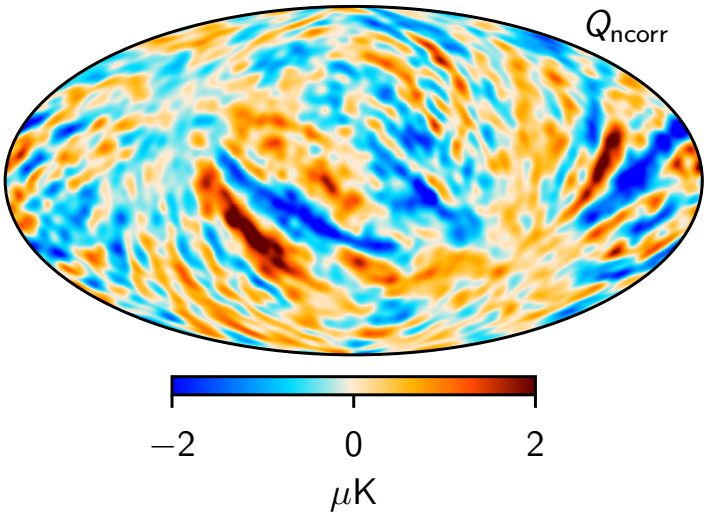

Figure 3.6: Stokes Q-map of the correlated noise, n_{corr}, 44 GHz band, smoothed at an angular resolution of 5 degrees FWHM. Figure courtesy of Eirik Gjerløw.

analysis, to get the most complete and robust sky model, and to be in the best position as possible to analyze data from future experiments like *LiteBIRD*, *PICO* and CMB-S4 (Abazajian et al., 2019; Hanany et al., 2019; Sugai et al., 2020; Suzuki et al., 2018).

Summary and Outlook

In this thesis I have given an introduction to the COMAP experiment and summarized some of mine and others work on the data analysis for COMAP. Papers I and II expand on this and go into more details on the various steps in the data analysis pipeline, and the power spectrum methods used. I also give a short summary of the BeyondPlanck project and the work I have done on noise estimation, systematics and calibration. In addition to discussing the experiments, the data analysis and the results, I have tried to connect the individual experiments with their significance to astrophysics and cosmology more generally.

The COMAP experiment is in a very exciting period right now, where we are working to finalize our analysis of the first year of data and publishing our first results, including an upper limit on the CO clustering power spectrum at around $z \sim 3$. On a longer time horizon we are already working on the next iterations of improvements in the COMAP pipeline, which involves better modeling and mitigation of ground contamination, better characterization of standing waves and their effects and an improvement in the mapmaking techniques, some of which have already been mentioned. The future plans for the COMAP experiments involve an expansion to multiple new telescopes, including instruments observing at lower frequencies, which would give us a whole new window into the epoch of reinonization ($z \sim 6$–8).

A bright future for intensity mapping in general lies in combining the data from multiple different experiments studying different spectral lines of the same galaxies giving us a more complete picture all the way from small scale chemistry and astrophysics to the large scale structure and expansion history of the universe as a whole. Combining together multiple complicated datasets increased the need for robust and statistically rigorous data analysis methods.

Our results in BeyondPlanck is a good demonstration of the unique advantages of the end-to-end approach. By including all the parts of the data analysis into a single model, not only does it help us deal with degeneracies and error propagation consistently, it also highlights problems in the data and where the model does not fit. With BeyondPlanck and *Cosmoglobe* we are laying the groundwork for the CMB community to face its greatest challenge, detecting B-modes from primordial gravitational waves. We believe that the best way to do this is to build an international community focused on global joint end-to-end analysis of cosmological datasets using open source, community built, software. By combining many datasets in this way we can not only get the most accurate sky model possible, but we can also break the degeneracies present in any one dataset, making us less susceptible to the systematic effects in either dataset.

Bibliography

Abazajian, K. et al. (July 2019). "CMB-S4 Science Case, Reference Design, and Project Plan". In: *arXiv e-prints*, arXiv:1907.04473, arXiv:1907.04473. arXiv: 1907.04473 [astro-ph.IM].

Bandura, K. et al. (July 2014). "Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder". In: *Ground-based and Airborne Telescopes V*. Ed. by Stepp, L. M., Gilmozzi, R., and Hall, H. J. Vol. 9145. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 914522. DOI: 10.1117/12.2054950. arXiv: 1406.2288 [astro-ph.IM].

Battye, R. A. et al. (Sept. 2013). "H I intensity mapping: a single dish approach". In: MNRAS vol. 434, no. 2, pp. 1239–1256. DOI: 10.1093/mnras/stt1082. arXiv: 1209.0343 [astro-ph.CO].

Baumann, D. (July 2018). "TASI Lectures on Primordial Cosmology". In: *arXiv e-prints*, arXiv:1807.03098, arXiv:1807.03098. arXiv: 1807.03098 [hep-th].

Behroozi, P. S., Wechsler, R. H., and Conroy, C. (Jan. 2013a). "On the Lack of Evolution in Galaxy Star Formation Efficiency". In: ApJ vol. 762, L31, p. L31. DOI: 10.1088/2041-8205/762/2/L31. arXiv: 1209.3013 [astro-ph.CO].

— (June 2013b). "The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8". In: ApJ vol. 770, 57, p. 57. DOI: 10.1088/0004-637X/770/1/57. arXiv: 1207.6105 [astro-ph.CO].

Bennett, C. L. et al. (Oct. 2013). "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results". In: ApJS vol. 208, no. 2, 20, p. 20. DOI: 10.1088/0067-0049/208/2/20. arXiv: 1212.5225 [astro-ph.CO].

Bernal, J. L., Breysse, P. C., Gil-Marin, H., et al. (Dec. 2019). "User's guide to extracting cosmological information from line-intensity maps". In: Phys. Rev. D vol. 100, no. 12, 123522, p. 123522. DOI: 10.1103/PhysRevD.100.123522. arXiv: 1907.10067 [astro-ph.CO].

Bernal, J. L., Breysse, P. C., and Kovetz, E. D. (Dec. 2019). "Cosmic Expansion History from Line-Intensity Mapping". In: Phys. Rev. Lett. Vol. 123, no. 25, 251301, p. 251301. DOI: 10.1103/PhysRevLett.123.251301. arXiv: 1907.10065 [astro-ph.CO].

BeyondPlanck Collaboration (2020). "**BeyondPlanck I. Global Bayesian analysis of the *Planck* Low Frequency Instrument data**". In: A&A*, in preparation*. arXiv: 2011.05609.

Borriello, A. and Salucci, P. (May 2001). "The dark matter distribution in disc galaxies". In: MNRAS vol. 323, pp. 285–292. eprint: astro-ph/0001082.

Bower, G. C. et al. (Jan. 2016). "Cosmic Structure and Galaxy Evolution through Intensity Mapping of Molecular Gas". In: *American Astronomical Society Meeting Abstracts #227*. Vol. 227. American Astronomical Society Meeting Abstracts, p. 426.04.

Bowman, J. et al. (2012). *The First Billion Years: report of a study program*. Workshop report.

Breysse, P. et al. (May 2017). "Insights from probability distribution functions of intensity maps". In: MNRAS vol. 467, no. 3, pp. 2996–3010. DOI: 10.1093/mnras/stx203. arXiv: 1609.01728 [astro-ph.CO].

Breysse, P. C., Anderson, C. J., and Berger, P. (Dec. 2019). "Canceling Out Intensity Mapping Foregrounds". In: Phys. Rev. Lett. Vol. 123, no. 23, 231105, p. 231105. DOI: 10.1103/PhysRevLett.123.231105. arXiv: 1907.04369 [astro-ph.CO].

Bull, P. et al. (Apr. 2015). "Late-time Cosmology with 21 cm Intensity Mapping Experiments". In: ApJ vol. 803, no. 1, 21, p. 21. DOI: 10.1088/0004-637X/803/1/21. arXiv: 1405.1452 [astro-ph.CO].

Chatrchyan, S. et al. (Sept. 2012). "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* vol. 716, no. 1, pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

Cheng, Y.-T., Chang, T.-C., and Bock, J. J. (Oct. 2020). "Phase-space Spectral Line Deconfusion in Intensity Mapping". In: ApJ vol. 901, no. 2, 142, p. 142. DOI: 10.3847/1538-4357/abb023. arXiv: 2005.05341 [astro-ph.CO].

Chung, D. (2020). "Line-intensity mapping with the CO Mapping Array Pathfinder and beyond". PhD thesis. Stanford University.

Chung, D. et al. (in prep).

Chung, D. et al. (Sept. 2017). "On Estimation of Contamination from Hydrogen Cyanide in Carbon Monoxide Line-intensity Mapping". In: ApJ vol. 846, 60, p. 60. DOI: 10.3847/1538-4357/aa8624. arXiv: 1706.03005.

Chung, D. T. (Aug. 2019). "A Partial Inventory of Observational Anisotropies in Single-dish Line-intensity Mapping". In: ApJ vol. 881, no. 2, 149, p. 149. DOI: 10.3847/1538-4357/ab3040. arXiv: 1905.00209 [astro-ph.CO].

Chung, D. T. et al. (Feb. 2019). "Cross-correlating Carbon Monoxide Line-intensity Maps with Spectroscopic and Photometric Galaxy Surveys". In: ApJ vol. 872, no. 2, 186, p. 186. DOI: 10.3847/1538-4357/ab0027. arXiv: 1809.04550 [astro-ph.GA].

Cleary, K. (2015). *COMAP Memo 5: System Overview*. Internal COMAP memo.

Cleary, K. et al. (Jan. 2016). "The CO Mapping Array Pathfinder (COMAP)". In: *American Astronomical Society Meeting Abstracts #227*. Vol. 227. American Astronomical Society Meeting Abstracts, p. 426.06.

Colombo et al. (2020). "**BeyondPlanck XI. CMB map and power spectrum posteriors**". In: A&A*, in preparation*. arXiv: 201x.xxxxx.

Cooray, A. et al. (Feb. 2016). "Cosmic Dawn Intensity Mapper". In: *arXiv e-prints*, arXiv:1602.05178, arXiv:1602.05178. arXiv: 1602.05178 [astro-ph.CO].

Creque-Sarbinowski, C. and Kamionkowski, M. (Sept. 2018). "Searching for decaying and annihilating dark matter with line intensity mapping". In: Phys. Rev. D vol. 98, no. 6, 063524, p. 063524. DOI: 10.1103/PhysRevD.98.063524. arXiv: 1806.11119 [astro-ph.CO].

Crites, A. T. et al. (2014). "The TIME-Pilot intensity mapping experiment". In: *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*. Ed. by Holland, W. S. and Zmuidzinas, J. Vol. 9153. International Society for Optics and Photonics. SPIE, pp. 613–621. DOI: 10.1117/12.2057207.

DeBoer, D. R. et al. (Apr. 2017). "Hydrogen Epoch of Reionization Array (HERA)". In: PASP vol. 129, no. 974, p. 045001. DOI: 10.1088/1538-3873/129/974/045001. arXiv: 1606.07473 [astro-ph.IM].

Delabrouille, J. (Feb. 1998). "Analysis of the accuracy of a destriping method for future cosmic microwave background mapping with the PLANCK SURVEYOR satellite". In: A&AS vol. 127, pp. 555–567. DOI: 10.1051/aas:1998119.

Dinda, B. R., Sen, A. A., and Choudhury, T. R. (Apr. 2018). "Dark energy constraints from the 21 cm intensity mapping surveys with SKA1". In: *arXiv e-prints*, arXiv:1804.11137, arXiv:1804.11137. arXiv: 1804.11137 [astro-ph.CO].

Dodelson, S. (2003). *Modern Cosmology*. ISBN: 9780122191411. Academic Press.

Doré, O. et al. (Dec. 2014). "Cosmology with the SPHEREX All-Sky Spectral Survey". In: *arXiv e-prints*, arXiv:1412.4872, arXiv:1412.4872. arXiv: 1412.4872 [astro-ph.CO].

Foss, M. K. et al. (in prep). *First Season COMAP Results: CO Data Processing*.

Furlanetto, S. et al. (May 2019). "Fundamental Cosmology in the Dark Ages with 21-cm Line Fluctuations". In: BAAS vol. 51, no. 3, 144, p. 144.

Gjerløw et al. (2020). "**BeyondPlanck VII. Gain and absolute calibration**". In: A&A, *in preparation*. arXiv: 2011.08082.

Gong, Y., Chen, X., and Cooray, A. (May 2020). "Cosmological Constraints from Line Intensity Mapping with Interlopers". In: ApJ vol. 894, no. 2, 152, p. 152. DOI: 10.3847/1538-4357/ab87a0. arXiv: 2001.10792 [astro-ph.CO].

Guth, A. H. (Jan. 1981). "Inflationary universe: A possible solution to the horizon and flatness problems". In: Phys. Rev. D vol. 23, no. 2, pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

Hanany, S. et al. (Feb. 2019). "PICO: Probe of Inflation and Cosmic Origins". In: *arXiv e-prints*, arXiv:1902.10541, arXiv:1902.10541. arXiv: 1902.10541 [astro-ph.IM].

Hill, G. J. et al. (Oct. 2008). "The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results". In: *Panoramic Views of Galaxy Formation and Evolution*. Ed. by Kodama, T., Yamada, T., and Aoki, K. Vol. 399. Astronomical Society of the Pacific Conference Series, p. 115. arXiv: 0806.0183 [astro-ph].

Hopkins, P. F. et al. (Nov. 2014). "Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation". In: MNRAS vol. 445, no. 1, pp. 581–603. DOI: 10.1093/mnras/stu1738. arXiv: 1311.2073 [astro-ph.CO].

Huterer, D. and Shafer, D. L. (Jan. 2018). "Dark energy two decades after: observables, probes, consistency tests". In: *Reports on Progress in Physics* vol. 81, no. 1, 016901, p. 016901. DOI: 10.1088/1361-6633/aa997e. arXiv: 1709.01091 [astro-ph.CO].

Ihle et al. (2020). "**BeyondPlanck VI. Noise characterization and modelling**". In: A&A, *in preparation*. arXiv: 2011.06650.

Ihle, H. T. et al. (in prep). *First Season COMAP Results: Power spectrum methodology and preliminary data quality assessment*.

Ihle, H. T. et al. (Jan. 2019). "Joint Power Spectrum and Voxel Intensity Distribution Forecast on the CO Luminosity Function with COMAP". In: ApJ vol. 871, no. 1, 75, p. 75. DOI: 10.3847/1538-4357/aaf4bc. arXiv: 1808.07487 [astro-ph.CO].

Jungman, G., Kamionkowski, M., and Griest, K. (Mar. 1996). "Supersymmetric dark matter". In: Phys. Rep. Vol. 267, pp. 195–373. DOI: 10.1016/0370-1573(95)00058-5. arXiv: hep-ph/9506380 [hep-ph].

Keating, G. K., Bower, G. C., et al. (Dec. 2015). "First Results from COPSS: The CO Power Spectrum Survey". In: ApJ vol. 814, no. 2, 140, p. 140. DOI: 10.1088/0004-637X/814/2/140. arXiv: 1510.06744 [astro-ph.GA].

Keating, G. K., Marrone, D. P., Bower, G. C., and Keenan, R. P. (Oct. 2020). "An Intensity Mapping Detection of Aggregate CO Line Emission at 3 mm". In: ApJ vol. 901, no. 2, 141, p. 141. DOI: 10.3847/1538-4357/abb08e. arXiv: 2008.08087 [astro-ph.GA].

Keating, G. K., Marrone, D. P., Bower, G. C., Leitch, E., et al. (Oct. 2016). "COPSS II: The Molecular Gas Content of Ten Million Cubic Megaparsecs at Redshift z ∼ 3". In: ApJ vol. 830, no. 1, 34, p. 34. DOI: 10.3847/0004-637X/830/1/34. arXiv: 1605.03971 [astro-ph.GA].

Keihänen et al. (2020). "**BeyondPlanck II. Optimal mapmaking by Gibbs sampling**". In: A&A*, in preparation*. arXiv: 2011.06024.

Keihänen, E. et al. (Feb. 2010). "Making cosmic microwave background temperature and polarization maps with MADAM". In: A&A vol. 510, A57, A57. DOI: 10.1051/0004-6361/200912813. arXiv: 0907.0367 [astro-ph.CO].

Klypin, A. A. et al. (Apr. 1987). "Limits on Microwave Background Anisotropies - the Relikt Experiment". In: *Soviet Astronomy Letters* vol. 13, p. 104.

Koopmans, L. et al. (Apr. 2015). "The Cosmic Dawn and Epoch of Reionisation with SKA". In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 1. arXiv: 1505.07568 [astro-ph.CO].

Kovetz, E. D. et al. (Sept. 2017). "Line-Intensity Mapping: 2017 Status Report". In: *arXiv e-prints*, arXiv:1709.09066, arXiv:1709.09066. arXiv: 1709.09066 [astro-ph.CO].

Kovetz, E. et al. (May 2019). "Astrophysics and Cosmology with Line-Intensity Mapping". In: BAAS vol. 51, no. 3, 101, p. 101. arXiv: 1903.04496 [astro-ph.CO].

Lagache, G., Cousin, M., and Chatzikos, M. (Jan. 2018). "The [CII] 158 μm line emission in high-redshift galaxies". In: A&A vol. 609, A130, A130. DOI: 10.1051/0004-6361/201732019. arXiv: 1711.00798 [astro-ph.GA].

Lakhlani, G. (Jan. 2019). "The Structure and Dynamics of the Interstellar Medium in the FIRE Simulations". PhD thesis. University of Toronto (Canada).

Lamb, J. W. (2020). *Standing waves due to optics and cables in the COMAP system*. Internal COMAP memo.

Lamb, J. et al. (in prep).

Li, T. Y. et al. (Feb. 2016). "Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly". In: ApJ vol. 817, 169, p. 169. DOI: 10.3847/0004-637X/817/2/169. arXiv: 1503.08833.

Liu, R. H. and Breysse, P. C. (Feb. 2020). "Coupling parsec and gigaparsec scales: primordial non-Gaussianity with multi-tracer intensity mapping". In: *arXiv e-prints*, arXiv:2002.10483, arXiv:2002.10483. arXiv: 2002.10483 [astro-ph.CO].

Madau, P., Meiksin, A., and Rees, M. J. (Feb. 1997). "21 Centimeter Tomography of the Intergalactic Medium at High Redshift". In: ApJ vol. 475, pp. 429–444. DOI: 10.1086/303549. eprint: astro-ph/9608010.

Masui, K. W. et al. (Jan. 2013). "Measurement of 21 cm Brightness Fluctuations at z ~0.8 in Cross-correlation". In: ApJ vol. 763, no. 1, L20, p. L20. DOI: 10.1088/2041-8205/763/1/L20. arXiv: 1208.0331 [astro-ph.CO].

Mather, J. C. et al. (Jan. 1994). "Measurement of the Cosmic Microwave Background Spectrum by the COBE FIRAS Instrument". In: ApJ vol. 420, p. 439. DOI: 10.1086/173574.

Mertens, F. G. et al. (Apr. 2020). "Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z ≈ 9.1 from LOFAR". In: MNRAS vol. 493, no. 2, pp. 1662–1685. DOI: 10.1093/mnras/staa327. arXiv: 2002.07196 [astro-ph.CO].

Milgrom, M. (June 2010). "New Physics at Low Accelerations (MOND): an Alternative to Dark Matter". In: *American Institute of Physics Conference Series*. Ed. by Alimi, J.-M. and Fuözfa, A. Vol. 1241. American Institute of Physics Conference Series, pp. 139–153. arXiv: 0912.2678 [astro-ph.CO].

Moustakas, L. A. and Metcalf, R. B. (Mar. 2003). "Detecting dark matter substructure spectroscopically in strong gravitational lenses". In: MNRAS vol. 339, pp. 607–615. eprint: astro-ph/0206176.

Newburgh, L. B. et al. (Aug. 2016). "HIRAX: a probe of dark energy and radio transients". In: *Ground-based and Airborne Telescopes VI*. Ed. by Hall, H. J., Gilmozzi, R., and Marshall, H. K. Vol. 9906. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 99065X. DOI: 10.1117/12.2234286. arXiv: 1607.02059 [astro-ph.IM].

Paradiso et al. (2020). "**BeyondPlanck XII. Cosmological parameter posteriors**". In: A&A, *in preparation*. arXiv: 201x.xxxxx.

Parsons, A. R. et al. (June 2014). "New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at z = 7.7". In: ApJ vol. 788, no. 2, 106, p. 106. DOI: 10.1088/0004-637X/788/2/106. arXiv: 1304.4991 [astro-ph.CO].

Pen, U.-L. et al. (Oct. 2009). "The GMRT EoR experiment: limits on polarized sky brightness at 150 MHz". In: MNRAS vol. 399, no. 1, pp. 181–194. DOI: 10.1111/j.1365-2966.2009.14980.x. arXiv: 0807.1056 [astro-ph].

Penzias, A. A. and Burrus, C. A. (Jan. 1973). "Millimeter-Wavelength Radio-Astronomy Techniques". In: ARA&A vol. 11, p. 51. DOI: 10.1146/annurev.aa.11.090173.000411.

Planck Collaboration II (2014). "*Planck* 2013 results. II. Low Frequency Instrument data processing". In: A&A vol. 571, A2. DOI: 10.1051/0004-6361/201321550. arXiv: 1303.5063.

Planck Collaboration I (2020). "*Planck* 2018 results. I. Overview, and the cosmological legacy of *Planck*". In: A&A vol. 641, A1. DOI: 10.1051/0004-6361/201833880. arXiv: 1807.06205.

Planck Collaboration II (2020). "*Planck* 2018 results. II. Low Frequency Instrument data processing". In: A&A vol. 641, A2. DOI: 10.1051/0004-6361/201833293. arXiv: 1807.06206.

Planck Collaboration VI (2020). "*Planck* 2018 results. VI. Cosmological parameters". In: A&A vol. 641, A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209.

Planck Collaboration et al. (Sept. 2020). "Planck 2018 results. I. Overview and the cosmological legacy of Planck". In: A&A vol. 641, A1, A1. DOI: 10.1051/0004-6361/201833880. arXiv: 1807.06205 [astro-ph.CO].

Rasmussen, M. (2020). "Data selection for intensity mapping experiments using machine learning". MA thesis. University of Oslo.

Riess, A. G. et al. (Sept. 1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: AJ vol. 116, no. 3, pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph].

Santos, M. G. et al. (Sept. 2017). "MeerKLASS: MeerKAT Large Area Synoptic Survey". In: *arXiv e-prints*, arXiv:1709.06099, arXiv:1709.06099. arXiv: 1709.06099 [astro-ph.CO].

Saro, A. et al. (July 2013). "Toward Unbiased Galaxy Cluster Masses from Line-of-sight Velocity Dispersions". In: ApJ vol. 772, 47, p. 47. arXiv: 1203.5708.

Schaye, J. et al. (Nov. 2014). "The EAGLE project: simulating the evolution and assembly of galaxies and their environments". In: *Monthly Notices of the Royal Astronomical Society* vol. 446, no. 1, pp. 521–554. DOI: 10.1093/mnras/stu2058. eprint: https://academic.oup.com/mnras/article-pdf/446/1/521/4139718/stu2058.pdf.

Silva, M. et al. (in prep).

Silva, M. B. et al. (Aug. 2019). "Mapping Large-Scale-Structure Evolution over Cosmic Times". In: *arXiv e-prints*, arXiv:1908.07533, arXiv:1908.07533. arXiv: 1908.07533 [astro-ph.CO].

Square Kilometre Array Cosmology Science Working Group et al. (Mar. 2020). "Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical specifications and performance forecasts". In: PASA vol. 37, e007, e007. DOI: 10.1017/pasa.2019.51. arXiv: 1811.02743 [astro-ph.CO].

Stacey, G. J. et al. (July 2018). "CCAT-Prime: science with an ultra-widefield submillimeter observatory on Cerro Chajnantor". In: *Ground-based and Airborne Telescopes VII*. Ed. by Marshall, H. K. and Spyromilio, J. Vol. 10700. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 107001M. DOI: 10.1117/12.2314031. arXiv: 1807.04354 [astro-ph.GA].

Sugai, H. et al. (Jan. 2020). "Updated Design of the CMB Polarization Experiment Satellite LiteBIRD". In: *Journal of Low Temperature Physics* vol. 199, no. 3-4, pp. 1107–1117. DOI: 10.1007/s10909-019-02329-w. arXiv: 2001.01724 [astro-ph.IM].

Susskind, L. (Nov. 1979). "Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory". In: Phys. Rev. D vol. 20, no. 10, pp. 2619–2625. DOI: 10.1103/PhysRevD.20.2619.

Suzuki, A. et al. (Dec. 2018). "The LiteBIRD Satellite Mission: Sub-Kelvin Instrument". In: *Journal of Low Temperature Physics* vol. 193, no. 5-6, pp. 1048–1056. DOI: 10.1007/s10909-018-1947-7. arXiv: 1801.06987 [astro-ph.IM].

Tauber, J. A. et al. (2019). "Characterization of the in-flight properties of the *Planck* telescope". In: A&A vol. 622, A55. DOI: 10.1051/0004-6361/201833150.

The OST mission concept study team (Sept. 2018). "The Origins Space Telescope (OST) Mission Concept Study Interim Report". In: *arXiv e-prints*, arXiv:1809.09702, arXiv:1809.09702. arXiv: 1809.09702 [astro-ph.IM].

Tingay, S. J. et al. (Jan. 2013). "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies". In: PASA vol. 30, e007, e007. DOI: 10.1017/pasa.2012.007. arXiv: 1206.6945 [astro-ph.IM].

Tristram, M. et al. (2021). "Planck constraints on the tensor-to-scalar ratio". In: *A&A* vol. 647, A128. DOI: 10.1051/0004-6361/202039585.

Uzgil, B. D. et al. (Dec. 2019). "The ALMA Spectroscopic Survey in the HUDF: Constraining Cumulative CO Emission at $1 \lesssim z \lesssim 4$ with Power Spectrum Analysis of ASPECS LP Data from 84 to 115 GHz". In: ApJ vol. 887, no. 1, 37, p. 37. DOI: 10.3847/1538-4357/ab517f. arXiv: 1911.00028 [astro-ph.GA].

van Haarlem, M. P. et al. (Aug. 2013). "LOFAR: The LOw-Frequency ARray". In: A&A vol. 556, A2, A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

van Uitert, E. et al. (Sept. 2012). "Constraints on the shapes of galaxy dark matter haloes from weak gravitational lensing". In: A&A vol. 545, A71, A71. arXiv: 1206.4304.

Vogelsberger, M. et al. (Oct. 2014). "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe". In: MNRAS vol. 444, no. 2, pp. 1518–1547. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921 [astro-ph.CO].

Yang, S. et al. (Sept. 2020). "Multitracer cosmological line intensity mapping mock lightcone simulation". In: *arXiv e-prints*, arXiv:2009.11933, arXiv:2009.11933. arXiv: 2009.11933 [astro-ph.GA].

Zwicky, F. (Jan. 1933). "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helvetica Physica Acta* vol. 6, pp. 110–127.

Aad, G. et al. (Sept. 2012). "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* vol. 716, no. 1, pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

Appendix A

Appendix

A.1 Binomial likelihood and mean estimation

Say you have some model for a statistic with a binomial distribution. The binomial mean, $\mu(\theta) = Np$, is then a function of the model parameters θ. Say you don't know the function $\mu(\theta)$, but you can simulate samples from the distribution at fixed θ. After one sample x_1, the posterior distribution of $\mu(\theta)$ is given by

$$P_1(\mu) \equiv P(\mu|x_1) = \frac{P(x_1|\mu)P_0(\mu)}{\int_{\mu'} P(x_1|\mu')P_0(\mu')d\mu'}, \tag{A.1}$$

where

$$P_0(\mu) = \int_\theta P(\mu|\theta)P(\theta)d\theta, \tag{A.2}$$

where $P(\theta)$ is the prior on the model parameters. In general, after n, samples we get

$$P_n(\mu) = P(\mu|x_n, \cdots, x_1) = \frac{P(x_n|\mu)P_{n-1}(\mu)}{\int_{\mu'} P(x_n|\mu')P_{n-1}(\mu')d\mu'}. \tag{A.3}$$

The estimated likelihood of some data d given by the model at parameters θ is then given by

$$P(\theta|d) = \frac{P(\theta)}{P(d)} \int_{\mu(\theta)} P(d|\mu(\theta))P_n(\mu(\theta))d\mu. \tag{A.4}$$

In practice $P_0(\mu)$ could perhaps be set to some weak prior, although the details will matter. Choosing an exponential prior we get

$$P_0(\mu) = \frac{\exp\left(-\frac{\mu}{\mu_0}\right)}{\int_0^\infty d\mu' \exp\left(-\frac{\mu'}{\mu_0}\right)} = \frac{\exp\left(-\frac{\mu}{\mu_0}\right)}{\mu_0}, \tag{A.5}$$

where μ_0 is the scale parameter of the prior, if needed, this parameter could be tuned to approximate Eq. A.2, or simply chosen at a value we think is reasonable. We can now

perform the integral in equation A.1

$$I_1 = \frac{1}{\mu_0}\int P(x_1|\mu')\exp\left(-\frac{\mu'}{\mu_0}\right)d\mu', \tag{A.6}$$

$$= \frac{1}{\mu_0}\int \binom{x_1}{N}\left(\frac{\mu'}{N}\right)^{x_1}\left(1-\frac{\mu'}{N}\right)^{N-x_1}\exp\left(-\frac{\mu'}{\mu_0}\right)d\mu', \tag{A.7}$$

$$\approx \int \frac{\binom{x_1}{N}}{N}\left(\frac{\mu'}{N}\right)^{x_1} e^{-\mu'}\exp\left(-\frac{\mu'}{\mu_0}\right)d\mu', \tag{A.8}$$

$$= \frac{N}{\mu_0}\binom{x_1}{N}\left(\frac{\mu_0}{N(\mu_0+1)}\right)^{x_1+1}\int u^{x_1}e^{u}\,du, \tag{A.9}$$

$$= \frac{N}{\mu_0}\binom{x_1}{N}\left(\frac{\mu_0}{N(\mu_0+1)}\right)^{x_1+1}\Gamma(x_1+1), \tag{A.10}$$

where we have assumed $N \gg \mu, d$, and used the substitution $u = \mu'\left(\frac{\mu_0}{\mu_0+1}\right)$.

Inserting this we can now find the analytic expression for Eq. A.1,

$$P_1(\mu) \approx \frac{\frac{1}{\mu_0}\binom{x_1}{N}\left(\frac{\mu'}{N}\right)^{x_1}e^{-\mu}\exp\left(-\frac{\mu'}{\mu_0}\right)}{\frac{N}{\mu_0}\binom{x_1}{N}\left(\frac{\mu_0}{N(\mu_0+1)}\right)^{x_1+1}\Gamma(x_1+1)} \tag{A.11}$$

$$= \frac{\mu^{x_1}e^{-\mu}\exp\left(-\frac{\mu'}{\mu_0}\right)}{\left(\frac{\mu_0}{(\mu_0+1)}\right)^{x_1+1}\Gamma(x_1+1)}. \tag{A.12}$$

In general, we get

$$P_n(\mu) = \frac{\mu^{x_1+x_2+\cdots+x_n}e^{-n\mu}\exp\left(-\frac{\mu'}{\mu_0}\right)}{\left(\frac{\mu_0}{(n\mu_0+1)}\right)^{x_1+x_2+\cdots+x_n+1}\Gamma(x_1+x_2+\cdots+x_n+1)}. \tag{A.13}$$

The posterior distribution then becomes

$$P(\theta|d) = \frac{P(\theta)}{P(d)}\int P(d|\mu)P_n(\mu)d\mu, \tag{A.14}$$

$$= \frac{P(\theta)}{P(d)}\binom{d}{N}\frac{1}{N^d}\frac{1}{\left(\frac{\mu_0}{(n\mu_0+1)}\right)^{x+1}\Gamma(x+1)}\int \mu^{x+d}e^{-(n+1)\mu}\exp\left(-\frac{\mu'}{\mu_0}\right), \tag{A.15}$$

$$= \frac{P(\theta)}{P(d)}\binom{d}{N}\left(\frac{\mu_0}{N(n+1)\mu_0+1}\right)^{d}\left(\frac{n\mu_0+1}{(n+1)\mu_0+1}\right)^{x}\frac{\Gamma(x+d+1)}{\Gamma(x+1)}, \tag{A.16}$$

where we have defined $x \equiv x_1 + x_2 + \cdots + x_n$.

Papers

Paper III

Joint Power Spectrum and Voxel Intensity Distribution Forecast on the CO Luminosity Function with COMAP

H. T. Ihle, D. Chung, G. Stein and the COMAP collaboration

Published in ApJ, Jan 2019, volume 871, no. 1 p. 75. DOI: 10.3847/1538-4357/aaf4bc.

III

The Astrophysical Journal, 871:75 (11pp), 2019 January 20 https://doi.org/10.3847/1538-4357/aaf4bc

Joint Power Spectrum and Voxel Intensity Distribution Forecast on the CO Luminosity Function with COMAP

H. T. Ihle[1], D. Chung[2], G. Stein[3,4], M. Alvarez[5], J. R. Bond[4], P. C. Breysse[4], K. A. Cleary[6], H. K. Eriksen[1], M. K. Foss[1], J. O. Gundersen[7], S. Harper[8], N. Murray[4], H. Padmanabhan[9], M. P. Viero[2], and I. K. Wehus[1]
(COMAP collaboration)
[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway; h.t.ihle@astro.uio.no
[2] Kavli Institute for Particle Astrophysics and Cosmology and Physics Department, Stanford University, Stanford, CA 94305, USA
[3] Department of Astronomy and Astrophysics, University of Toronto, 50 St. George St., Toronto, ON, M5S 3H4, Canada
[4] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George St., Toronto, ON M5S 3H8, Canada
[5] Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA
[6] California Institute of Technology, Pasadena, CA 91125, USA
[7] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
[8] Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[9] Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang-Pauli-Strasse 27, CH 8093 Zurich, Switzerland
Received 2018 August 22; revised 2018 November 8; accepted 2018 November 27; published 2019 January 23

Abstract

We develop a framework for joint constraints on the CO luminosity function based on power spectra (PS) and voxel intensity distributions (VID) and apply this to simulations of CO Mapping Array Pathfinder (COMAP), a CO intensity mapping experiment. This Bayesian framework is based on a Markov chain Monte Carlo (MCMC) sampler coupled to a Gaussian likelihood with a joint PS + VID covariance matrix computed from a large number of fiducial simulations and re-calibrated with a small number of simulations per MCMC step. The simulations are based on dark matter halos from fast peak patch simulations combined with the $L_{\rm CO}(M_{\rm halo})$ model of Li et al. We find that the relative power to constrain the CO luminosity function depends on the luminosity range of interest. In particular, the VID is more sensitive at large luminosities, while the PS and the VID are both competitive at small and intermediate luminosities. The joint analysis is superior to using either observable separately. When averaging over CO luminosities ranging between $L_{\rm CO} = 10^4$–$10^7\,L_\odot$, and over 10 cosmological realizations of COMAP Phase 2, the uncertainties (in dex) are larger by 58% and 30% for the PS and VID, respectively, when compared to the joint analysis (PS + VID). This method is generally applicable to any other random field, with a complicated likelihood, as long a fast simulation procedure is available.

Key words: diffuse radiation – galaxies: high-redshift – large-scale structure of universe

1. Introduction

Intensity mapping (Madau et al. 1997; Battye et al. 2004; Peterson et al. 2006; Loeb & Wyithe 2008) appears promising for mapping large 3D volumes cheaply in a relatively short period of time, using specific bright emission lines as matter tracers. This is an interesting avenue for advancing precision cosmology, with a multitude of ongoing efforts (Kovetz et al. 2017), following on the successes of the CMB field in the last few decades. One such line intensity mapping experiment currently under construction is called the CO Mapping Array Pathfinder (COMAP; Cleary et al. 2016; Li et al. 2016), which aims to observe frequencies between 26 and 34 GHz, corresponding to redshifted CO line emission from the epoch of galaxy assembly (redshifts between $z = 2.4$ and 3.4) for the CO $J = 1 \rightarrow 0$ line at 115 GHz rest frequency, and CO emission from the epoch of reionization ($z = 5.8$–6.7) for the CO $J = 2 \rightarrow 1$ line at 230 GHz rest frequency.

One important scientific target for studying and understanding the epoch of galaxy assembly, the main goal of the first COMAP phase, is the so-called CO luminosity function, which measures the number density of CO emitters as a function of luminosity. Several methods for extracting this function from real data have already been suggested in the literature, the most prominent being the power spectrum (PS) approach, for instance as implemented by Li et al. (2016). A second complementary method is the one-point function, or voxel intensity distribution (VID), $\mathcal{P}(T)$, as suggested by Breysse et al. (2016, 2017).

In this paper, we consider the prospect of combining the VID and PS approaches when constraining the CO luminosity function, and we study this approach within the context of the COMAP experiment. To do so, we first define a joint likelihood that includes both the VID and the PS and construct a joint covariance matrix for both observables. This covariance matrix is constructed from a large set of dark matter (DM) light-cone halo catalogs from so-called "peak patch" cosmological simulations (Bond & Myers 1996, Stein et al. 2019), coupled to an empirical $L_{\rm CO}(M_{\rm halo})$ model (Li et al. 2016) that infers CO luminosities, $L_{\rm CO}$, from DM halo masses, $M_{\rm halo}$. We then investigate the posterior distribution of the resulting model parameters for each of the first two anticipated phases of the COMAP experiment (see Table 1). Finally, we compare the constraints on the CO luminosity function derived from joint PS and VID measurements to those obtained from the PS or VID separately.

2. Idealized Simulations of the COMAP Experiment

We start our discussion by reviewing some central properties of the COMAP experiment, focusing in particular on those required for generating representative yet computationally

Table 1
Experiment Setup for the Two COMAP Phases

Parameter	COMAP1	COMAP2
System temperature, $T_{\rm sys}$ [K]	40	40
Number of feeds	19	95
Beam FWHM (arcmin)	4	4
Frequency band [GHz]	26–34	26–34
Channel width, $\delta\nu$ (MHz)	15.6	15.6
Observing time [hr]	6000	9000
Noise per voxel [μK]	11.0	8.0
Field size [deg^2]	2.25	2.25
Number of fields	1	4

affordable simulations. For convenience, these properties are summarized in Table 1.

In Phase 1, COMAP will employ a single telescope equipped with 19 single-polarization detectors, each with 512 frequency channels with width $\delta\nu \approx 15.6$ MHz[10] covering frequencies between 26 and 34 GHz. The system temperature is expected to be around $T_{\rm sys} \approx 40$ K and the angular resolution corresponds to a Gaussian beam with $4'$ full width at half maximum (FWHM). We anticipate two years of observation time targeting a single field of $1.^\circ5 \times 1.^\circ5$ close to the north celestial pole, and we assume a conservative observing efficiency of 35% for a total of 6000 hr of total integration time on the field.

In Phase 2, the experiment will be expanded to five telescopes with the same setup as in Phase 1 and will observe for three additional years. In this phase, we assume that the observation time will be split between four fields of the same size as in Phase 1. The two COMAP phases will be referred to as COMAP1 and COMAP2 in the following.

2.1. Noise

The simulations used in this work consist of two components only, namely the target CO signal and random white noise with properties corresponding to the parameters described above. Explicitly, the noise per voxel is given by

$$\sigma_T = \frac{T_{\rm sys}}{\sqrt{\tau\ \delta\nu}} = \frac{T_{\rm sys}\sqrt{N_{\rm pixels}}}{\sqrt{\tau_{\rm tot} e_{\rm obs} N_{\rm feeds}\delta\nu}}, \quad (1)$$

where $T_{\rm sys}$ is the system temperature, τ is the observation time per pixel, $\tau_{\rm tot}$ is the total observation time, $e_{\rm obs}$ is the observation efficiency, $N_{\rm feeds}$ is the number of feeds, $N_{\rm pixels}$ is the number of pixels, and $\delta\nu$ is the frequency resolution. This gives us $\sigma_T \approx 11\,\mu$K and $8\,\mu$K for the COMAP1 and COMAP2 phases, respectively. For simplicity we assume that the noise is evenly distributed over all voxels.

A voxel is the 3D equivalent of a pixel. Two of the dimensions correspond to a regular pixel on the sky, while the third dimension corresponds to a small range of redshifts from where line emission would redshift into a given frequency bin of our instrument.

Both instrumental systematics and astrophysical foreground contamination are neglected in the following. However, since our estimator is inherently simulation based, these effects can be added at a later stage when a sufficiently realistic instrument model is available. For discussion of foreground contamination in similar line intensity surveys see, e.g., da Cunha et al. (2013), Breysse et al. (2015, 2017), and Chung et al. (2017).

[10] Higher spectral resolutions are available, but these are most likely useful only for systematics mitigation rather than science due to limited signal-to-noise per voxel.

2.2. DM Simulations

The signal component is based on the peak patch DM halo approach described by Bond & Myers (1996) and Stein et al. (2019), coupled to the $L_{\rm CO}(M_{\rm halo})$ model presented by Li et al. (2016). Additionally, we adopt the same cosmological parameters as the Li et al. (2016) analysis for the DM simulations: $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.96$.

The DM simulations in this paper were created using the peak patch method of Bond & Myers (1996) and Stein et al. (2019). To cover the full redshift range of the COMAP experiment we simulated a volume of $(1140\,{\rm Mpc})^3$ using a particle-mesh resolution of $N_{\rm cells} = 4096^3$. Projecting this onto the sky results in a $9.^\circ6 \times 9.^\circ6$ field of view covering the redshift range $2.4 < z < 3.4$, with a minimum DM halo mass of $2.5 \times 10^{10}\,M_\odot$.

The resulting halo catalog contains roughly 54 million halos, each with a position, velocity, and mass. The peak patch method can simulate continuous light cones on-the-fly, so stitching snapshots together was not required to create the light cone. Although peak patch simulations result in quite accurate halo masses, the DM halo catalogs were additionally mass corrected by abundance matching along the light cone to Tinker et al. (2008) to ensure statistically the same mass function as the simulations used in the Li et al. (2016) analysis. For a detailed study of the clustering properties of peak patch simulations and other approximate methods, see Lippich et al. (2019), Blot et al. (2018), and Colavincenzo et al. (2019).

A single run required 900 s of computation time on 2048 Intel Xeon EE540 2.53 GHz CPU cores of the Scinet-GPC cluster, with a memory footprint of $\simeq$2.4 TB. This efficiency of the peak patch method allowed for 161 independent realizations of the full 1140 Mpc, $N_{\rm cells} = 4096^3$ volume, taking a total computation time of only $\sim$82,000 CPU hours, over three orders of magnitude faster when compared to an N-body method of equivalent size.

2.3. Converting to CO Brightness Temperature

There are many approaches in the literature for estimating the expected CO signal based on DM halos (e.g., Righi et al. 2008; Obreschkow et al. 2009; Visbal & Loeb 2010; Carilli 2011; Gong et al. 2011; Lidz et al. 2011; Fu et al. 2012; Carilli & Walter 2013; Pullen et al. 2013; Breysse et al. 2014; Greve et al. 2014; Mashian et al. 2015; Li et al. 2016; Padmanabhan 2018), with resulting estimates of the CO luminosities spanning roughly an order of magnitude.

Here we adopt the model described by Li et al. (2016) to convert from simulated light cones populated with DM halos to observed CO brightness temperature. This model is defined by a set of parametric relations between DM halo masses, star formation rates (SFR), infrared (IR) luminosities, $L_{\rm IR}$, and CO luminosities, $L_{\rm CO}$.

The model uses the results from Behroozi et al. (2013a, 2013b) to obtain average SFR from DM halo masses and adds an additional log-normal scatter on top of the average, determined by $\sigma_{\rm SFR}$. IR luminosities are then obtained through

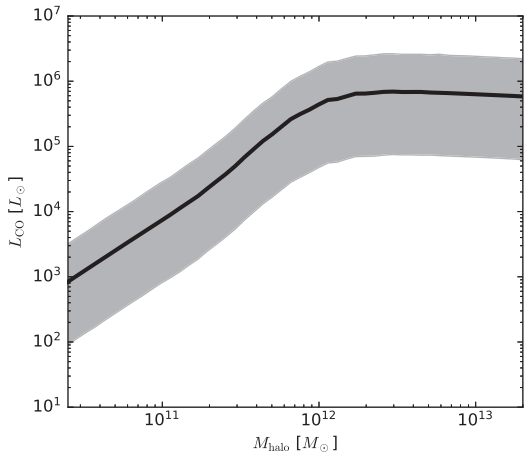

Figure 1. Plot of CO luminosity, $L_{\rm CO}$, as a function of dark matter halo mass, $M_{\rm halo}$, in the Li et al. (2016) model. Here, ($\sigma_{\rm SFR}$, $\log\delta_{\rm MF}$, α, β, $\sigma_{L_{\rm CO}}$) = (0.3, 0.0, 1.17, 0.21, 0.3) (our fiducial model), and we have evaluated the function at redshift 2.9. The solid line corresponds to the mean relation with no scatter added, while the shaded region corresponds to the 95% confidence intervals after adding log-normal scatter at the two appropriate steps.

the relation

$$\mathrm{SFR} = \delta_{\rm MF} \times 10^{-10}\, L_{\rm IR}. \tag{2}$$

Further, to obtain CO luminosities, the relation

$$\log L_{\rm IR} = \alpha \log L'_{\rm CO} + \beta \tag{3}$$

is used before a second round of log-normal scatter is added, determined by the parameter $\sigma_{L_{\rm CO}}$.

This gives us a $L_{\rm CO}(M_{\rm halo})$ model with five free parameters, $\theta = \{\sigma_{\rm SFR}, \log\delta_{\rm MF}, \alpha, \beta, \sigma_{L_{\rm CO}}\}$. The relation between $L_{\rm CO}$ and $M_{\rm halo}$ for our fiducial model parameters is shown in Figure 1. For more discussion of the physical and observational motivation for this model, see the original paper, Li et al. (2016).

This model is applied to each DM halo separately and we create high-resolution maps from the resulting CO luminosities by converting the total luminosity in a given voxel into brightness temperature. These maps were created using the publicly available `limlam_mocker` code.[11] Next, we convolve these maps with the (Gaussian) instrumental beam profile, degrade to the low-resolution voxel size used in the analysis, and, finally, we add Gaussian uncorrelated noise with standard deviation σ_T, as specified above.

3. Algorithms

The ultimate goal of this work is to constrain cosmological and astrophysical parameters from CO line intensity observations. The computational engine for this work is a standard Metropolis Markov chain Monte Carlo (MCMC) sampler (see, e.g., Gilks et al. 1995), coupled to a posterior distribution with a corresponding likelihood and prior. For this task to be computationally tractable, though, the full CO line intensity data set must first be compressed to a smaller set of observables that may be modeled in terms of the desired astrophysical parameters, fully analogous to how CMB sky maps are compressed to a CMB power spectrum from which cosmological parameters are derived (e.g., Bond et al. 2000). As described above, we adopt the power spectrum and the VID as representative observables, each of which may be approximated in terms of multivariate Gaussian random variables. However, in order to perform a joint analysis of these two observables, we need to construct their joint covariance matrix, and that is the primary goal of this section. Before doing that, however, we review for completeness each observable individually, and our posterior sampler of choice, referring to relevant literature for full details.

3.1. The Power Spectrum

The estimated power spectrum, $P(k_i)$, is calculated simply by taking the 3D Fourier transform of the temperature cube, binning the absolute squared values of the Fourier coefficients according to the magnitude of corresponding wave number k, and averaging over all the contributions within each bin. For a Gaussian map, the Fourier components within each bin follow a perfect normal distribution with mean zero and variance given by the value of the power spectrum. For a non-Gaussian field, the distribution of the Fourier components is more complicated, and thus the power spectrum does not contain all the statistical information in the map. We expect the CO signal to form a highly non-Gaussian map, therefore, in this paper we simply consider the power spectrum as a useful observable that carries some, but far from all, of the statistical information in the map.

As an observable, the power spectrum needs to be accompanied by a covariance matrix $\xi^P_{ij} \equiv \mathrm{Cov}(P(k_i), P(k_j))$ in the analysis, since there are correlations between the power spectrum at different k values.

3.2. The Voxel Intensity Distribution

We consider the VID as another observable, complementary to the PS and more closely related to the luminosity function.

Unlike in many other works on $\mathcal{P}(D)$ analysis (e.g., Lee et al. 2009; Glenn et al. 2010; Vernstrom et al. 2014; Breysse et al. 2017; Leicht et al. 2019), we do not try to estimate the VID analytically, rather we estimate it based on simulations. This allows us to fully take into account the effects of the beam, clustering, and covariance between temperature bins in a very straightforward manner.

We consider two contributions to the VID, namely the CO signal itself and the instrumental noise. Together they result in the the full probability distribution of voxel temperatures, $\mathcal{P}(T)$, where T is the observed brightness temperature from a voxel. Since we assume the noise to be uniformly distributed over all voxels in the observed field and assume that the CO signal itself is statistically homogeneous and isotropic, the total probability distribution, $\mathcal{P}(T)$, is the same across all voxels.

The basic observable related to the VID are the temperature bin counts (i.e., the histogram of voxel temperatures), B_i. The expectation value of these are given by the VID itself,

$$\langle B_i \rangle = N_{\rm vox} \int_{T_i}^{T_{i+1}} \mathcal{P}(T) dT, \tag{4}$$

where $N_{\rm vox}$ is the number of voxels observed and B_i is the number of voxels with a temperature between T_i and T_{i+1}.

[11] https://github.com/georgestein/limlam_mocker

If the temperatures of all the voxels that we sample were completely independent, then each of the voxel bins would be approximately independent and would follow a binomial distribution with variance $\mathrm{Var}_{\mathrm{ind}}(B_i) = \langle B_i \rangle (1 - \langle B_i \rangle / N_{\mathrm{vox}})$. However, even in this ideal case they would not be perfectly independent. We only have a finite number of voxels, and, therefore, if one bin contains a number of voxels above average, then the other bins must have a number lower than average.

In general, the samples will not be independent for many reasons, including correlated sky or noise structures and processing effects, and we therefore need the full covariance matrix between bins, $\xi_{ij}^{B} \equiv \mathrm{Cov}(B_i, B_j)$. This covariance matrix will depend on the DM density field, the CO bias, and the luminosity function, and we will estimate it using simulations.

3.3. The Joint PS+VID Covariance Matrix

The main missing component in the above method is definition of a joint power spectrum and VID covariance matrix. By having access to the computationally cheap yet realistic Monte Carlo simulations described above, we can approximate this matrix with simulations. In addition to giving us covariance matrices to do our analysis, this also allows us to check under what conditions the full covariance matrix is necessary and when we can get away with assuming that individual samples are independent.

In this paper, we start with 161 independent simulated light-cone cubes of DM halos, each covering about $9.^{\circ}6 \times 9.^{\circ}6$ on the sky and a frequency range between 26 and 34 GHz, corresponding to redshifts between 2.4 and 3.4. The frequency dimension is divided equally into 512 frequency bins, each spanning $\delta\nu \approx 15.6$ MHz, corresponding to a redshift resolution of $\delta z \approx 0.002$. Since the COMAP field only spans $1.^{\circ}5 \times 1.^{\circ}5$ on the sky, we sub-divide each of the $9.^{\circ}6 \times 9.^{\circ}6$ light-cone cubes, after beam convolution, into 36 square fields, each covering $1.^{\circ}5 \times 1.^{\circ}5$, resulting in a total of 5796 semi-independent sky realizations. The final pixelization of these maps is a 22×22 grid of square pixels, resulting in a pixel size of $\delta\theta \approx 4.'1$. To these maps, we add uniformly distributed white noise at the appropriate levels for the COMAP1 and COMAP2 experiment setups described above.

When choosing the pixel size to use for the analysis, we follow Vernstrom et al. (2014). They show that, for $\mathcal{P}(D)$ analysis, choosing a pixel size to be equal to the FWHM of the beam is a good tradeoff between picking a small pixel size to include the maximal information, and choosing a larger pixel size to reduce the pixel to pixel correlations induced by the beam.

We combine our two observables into a joint one-dimensional vector of the form

$$d_i = (P_{k_i}, B_i), \tag{5}$$

where P_{k_i} is the binned power spectrum and B_i are the temperature bin counts. Let us first consider the ideal case in which all elements in this vector are independent and the Fourier components are approximately Gaussian. In that case we can compute the expected variance, which we will simply call the *independent variance*, analytically,

$$\mathrm{Var}_{\mathrm{ind}}(P_{k_i}) = \langle P_{k_i} \rangle^2 / N_{\mathrm{modes}}(k_i), \tag{6}$$

$$\mathrm{Var}_{\mathrm{ind}}(B_i) = \langle B_i \rangle (1 - \langle B_i \rangle / N_{\mathrm{vox}}) \approx \langle B_i \rangle, \tag{7}$$

where $N_{\mathrm{modes}}(k_i)$ is the number of modes in the ith k bin and where we have introduced the notation $\mathrm{Var}_{\mathrm{ind}}(d_i)$ for this conditionally independent variance.

With this notation in hand, we define a "pseudo-correlation matrix" as

$$c_{ij} \equiv \frac{\xi_{ij}}{\sqrt{\mathrm{Var}_{\mathrm{ind}}(d_i)\,\mathrm{Var}_{\mathrm{ind}}(d_j)}}, \tag{8}$$

where, as in Section 3.4, ξ_{ij} is the full covariance matrix. Note that c_{ij} is the exact correlation matrix in the limit that $\mathrm{Var}_{\mathrm{ind}}(d_i)$ is the true full variance. An important advantage of the pseudo-correlation matrix, however, is the fact that $\mathrm{Var}_{\mathrm{ind}}(d_i)$ may be estimated directly from the average data itself, and this is required for our MCMC procedure to be sufficiently fast.

The full covariance matrix ξ is estimated for the model described by Li et al. (2016), adopting the fiducial parameters θ_0, using the set of 5796 simulations described above. However, for the MCMC sampler described in Section 3.4, we actually need the full covariance matrix, corresponding to different model parameters θ, at each step in the Markov chain. Generating the full covariance matrix with the above procedure at each MC step is clearly not computationally feasible and we therefore need to approximate this somehow.

With regard to this last point, we introduce the following proposal: we assume that the full covariance matrix scales, under a change of model parameters from θ_0 to θ, the same way as the independent variance, $\mathrm{Var}_{\mathrm{ind}}(d_i)$,

$$\hat{\xi}_{ij}(\theta) \approx \xi_{ij}(\theta_0) \frac{\sqrt{\mathrm{Var}_{\mathrm{ind}}^{\theta}(d_i)\,\mathrm{Var}_{\mathrm{ind}}^{\theta}(d_j)}}{\sqrt{\mathrm{Var}_{\mathrm{ind}}^{\theta_0}(d_i)\,\mathrm{Var}_{\mathrm{ind}}^{\theta_0}(d_j)}}, \tag{9}$$

where $\mathrm{Var}_{\mathrm{ind}}^{\theta_0}(d_i)$ is the independent variance for the fiducial model and $\mathrm{Var}_{\mathrm{ind}}^{\theta}(d_i)$ is the independent variance for arbitrary parameters θ. Since this latter function only depends on the average quantities $\langle d_i \rangle$, it is computationally straightforward to compute $\hat{\xi}_{ij}(\theta)$ at any position in an MCMC sampler. Note also that $\hat{\xi}_{ij}(\theta)$ is, by construction, positive definite, as required for a proper covariance matrix.

For a noise-dominated experiment, where all samples are approximately independent, the independent variance, $\mathrm{Var}_{\mathrm{ind}}(d_i)$, is the correct variance and Equation (9) is the correct scaling of the covariance matrix. However, we use this scaling as a first approximation even in cases where there is some covariance in the data.

Intuitively, Equation (9) is equivalent to postulating that the pseudo-correlation matrix, c_{ij}, is approximately constant (i.e., independent of the specific parameters in question). For real-world applications, we recommend testing this assumption explicitly by computing the covariance matrix by brute force simulation for a few extreme parameter combinations drawn from the Markov chains produced during the analysis.

The above prescription applies straightforwardly to single-field observations as, for instance, planned for COMAP1. In contrast, COMAP2 will, under our assumptions, span $N = 4$ independent but statistically identical fields. Since the mean vector of observables evaluated across those four fields equals the average of the four corresponding independent observable vectors, the full covariance matrix is simply given by the

single-field covariance matrix divided by the number of fields:

$$\xi_{ij}^{N\ \mathrm{field}} = \frac{\xi_{ij}^{1\ \mathrm{field}}}{N}. \quad (10)$$

Note that c_{ij} then, assuming the fields are of the same size, only depends on the noise level per field, so for a given noise level per field, c_{ij} is independent of the number of fields.

Finally, we note that the total number of degrees of freedom in our joint PS+VID statistic is in this paper equal to 45, corresponding to 20 power spectrum bins and 25 VID bins. For this number of degrees of freedom, a set of 5796 (semi-independent) simulations provides a very good estimate of all numerically dominant components of the covariance matrix, including both the diagonal and the leading off-diagonal modes, and ξ_{ij} is well conditioned.

3.4. Posterior Mapping by MCMC

As previously mentioned, we use an MCMC algorithm to sample from the posterior distribution of the $L_{\mathrm{CO}}(M_{\mathrm{halo}})$ model parameters, $\theta = \{\sigma_{\mathrm{SFR}}, \log \delta_{\mathrm{MF}}, \alpha, \beta, \sigma_{L_{\mathrm{CO}}}\}$. This posterior distribution is, as usual, given by Bayes' theorem,

$$P(\theta|d) \propto P(d_i|\theta)P(\theta), \quad (11)$$

where d represents our compressed data set, $P(d|\theta)$ is the likelihood defined below, and $P(\theta)$ is some set of priors. We use the *emcee* package (Foreman-Mackey et al. 2013) and its implementation of an affine-invariant ensemble MCMC algorithm, with 142 walkers.

We use a burn-in period of 1000 steps, and use the next 1000 steps for the posterior estimation.

We assume a Gaussian likelihood for our observables d_i of the form (up to an additive constant)

$$-2\ln P(d|\theta) = \sum_{ij}[d_i - \langle d_i\rangle](\xi^{-1})_{ij}[d_j - \langle d_j\rangle] + \ln|\xi|, \quad (12)$$

where the means $\langle d_i\rangle$ depend on the model parameters θ, and the covariance matrix ξ_{ij} is approximated by the expression given in Equation (9). (Note that we do not need to assume that the low-level data are Gaussian, but only that the compressed observables may be well approximated by a multivariate Gaussian distribution. Due to the central limit theorem, this is in practice very often an excellent approximation.)

For both the power spectrum and the low and intermediate temperature VID bins, for which there is a large number of voxel counts per bin, this Gaussian approximation holds to a high degree. However, for the highest VID temperature bins, where there are only a few voxel counts per bin, the discrete nature of the bin count may become relevant and the full binomial distribution should, in principle, be taken into account. However, this effect can also be easily remedied by increasing the bin width, albeit at the cost of a slight loss of information, as is suggested in Vernstrom et al. (2014), and we therefore neglect it in the following, since our primary focus is the dominant Gaussian component of the likelihood. A more thorough analysis may take this issue into account either analytically or by simulations.

An advantage of using a Gaussian likelihood for the VID is that it gives us a straightforward way to take into account the correlations between temperature bins apparent in the covariance matrix, ξ_{ij} (e.g., in Figure 2). For another approach to building up a $\mathcal{P}(D)$ likelihood, see Glenn et al. (2010).

To estimate $\langle d_i\rangle$, we compute 10 maps of the survey volume at each step in the MCMC chain using the current model parameters θ with different DM halo realizations (randomly drawn from 252 independent catalogs corresponding to the survey volume). The specific number of realizations, 10 in our case, represents a compromise between minimizing the sample variance in the estimate of $\langle d_i\rangle$ and maintaining a reasonable computational cost per MC step. Finally, we bin all of the halos in the 10 realizations according to their luminosity and use this histogram to estimate the luminosity function at the current values of θ. This way the MCMC procedure gives us the luminosity function at different points in parameter space, sampled according to the posterior of the model parameters, which we can use to derive constraints on the luminosity function itself.

We adopt the same physically motivated priors as discussed by Li et al. (2016). Specifically, these read

$$P(\sigma_{\mathrm{SFR}}) = \mathcal{N}(0.3, 0.1) \quad (13)$$

$$P(\log \delta_{\mathrm{MF}}) = \mathcal{N}(0.0, 0.3) \quad (14)$$

$$P(\alpha) = \mathcal{N}(1.17, 0.37) \quad (15)$$

$$P(\beta) = \mathcal{N}(0.21, 3.74) \quad (16)$$

$$P(\sigma_{L_{\mathrm{CO}}}) = \mathcal{N}(0.3, 0.1), \quad (17)$$

where $\mathcal{N}(\mu, \sigma)$ corresponds to a Gaussian distribution with mean μ and standard deviation σ. Additionally, we require the two logarithmic scatter parameters, σ_{SFR} and $\sigma_{L_{\mathrm{CO}}}$, to be positive. We choose the mean of all these distributions as the fiducial model, θ_0.

To quantify the importance of joint PS+VID analysis, we perform the above analysis both with each observable separately and with the joint analysis. The main result in this paper may then be formulated in terms of the relative improvement on the CO luminosity function uncertainty derived from the joint analysis to those found in the independent analyses.

When calculating our observables (PS and VID), we assume that our survey volume can be treated as a rectangular grid of voxels with constant co-moving volume. We also neglect the evolution of our observables over redshifts between $z = 2.4$ and 3.4. That is, we assume that samples from different redshifts are drawn from the same distribution, whether they are power spectrum modes or voxel temperatures. We also assume that the instrument beam is achromatic and is equal to the value at the central frequency. This is of course just an approximation that we make in order for the analysis to be simple. If we were doing experiments with higher signal to noise, we might divide our data into two different redshift regions and do an independent analysis of each region. This could allow us to study the redshift evolution of the observables. For COMAP (1 and 2), however, we are probably best off combining all the data, like we do here, in order to increase the overall signal to noise.

Finally, since COMAP will not measure absolute zero levels, we subtract the mean from all maps. For the power spectrum, this has a negligible impact, as it simply removes one out of

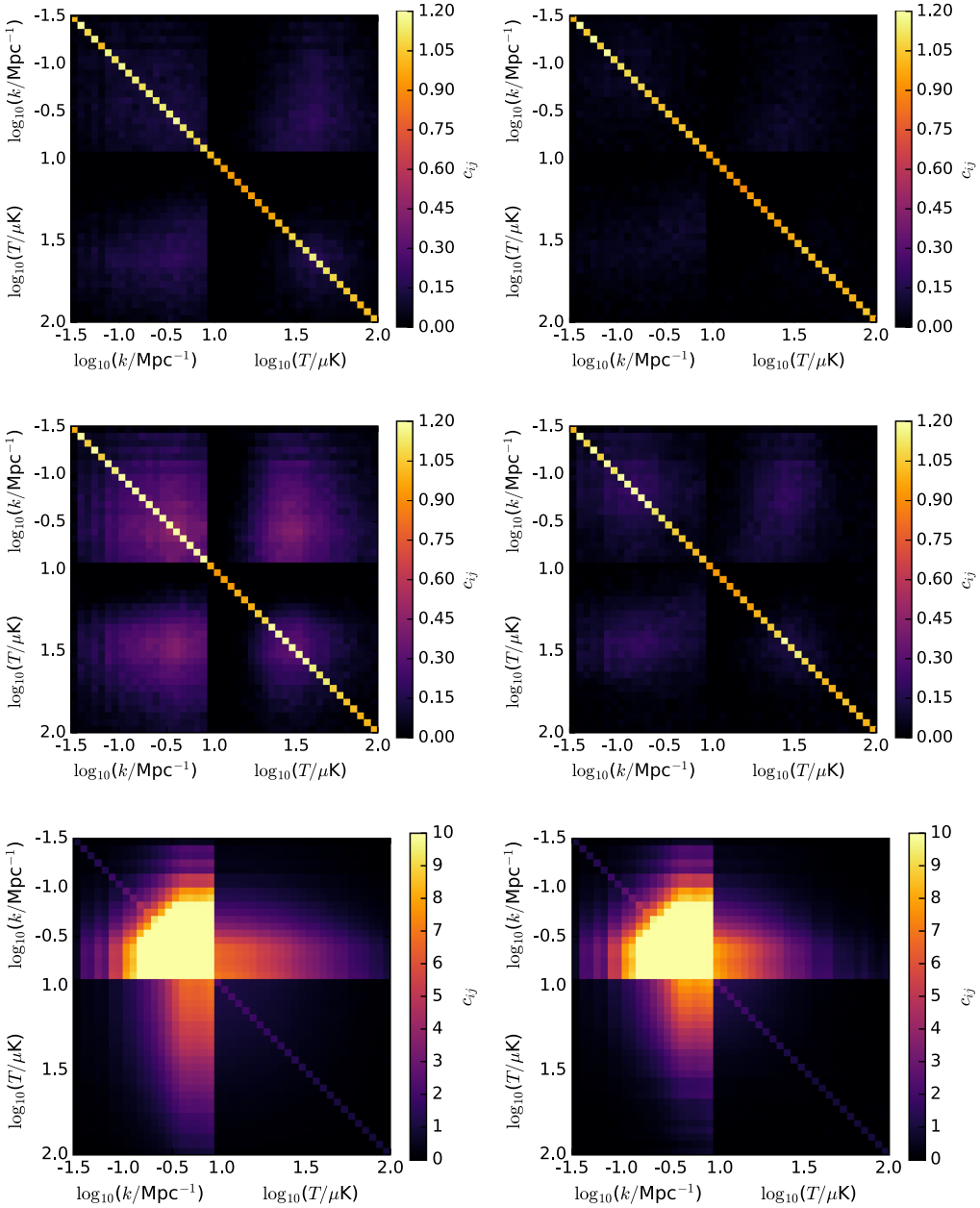

Figure 2. Estimated pseudo-correlation matrix of observables d_i, $c_{ij} = \mathrm{Cov}(d_i, d_j)/(\sqrt{\mathrm{Var}_{\mathrm{ind}}(d_i)\mathrm{Var}_{\mathrm{ind}}(d_j)})$, based on simulated maps with and without noise. The first block in each matrix corresponds to the power spectrum and the second block to the VID. Top: signal plus white noise corresponding to the COMAP1 experiment ($\sigma_{\mathrm{voxel}} \approx 11\ \mu$K). Middle: signal plus white noise corresponding to the COMAP2 experiment ($\sigma_{\mathrm{voxel}} \approx 8\ \mu$K). Bottom: signal alone. Note that here we have changed the color scale. Left: covariance matrices without beam smoothing. Right: covariance matrices with $\theta_{\mathrm{FWHM}} = 4'$ beam smoothing.

N_{vox} modes. However, it has a significantly higher impact for the VID. Specifically, it makes it much harder to distinguish a potential background of weak sources from noise. Indeed, as shown by Breysse et al. (2017), removing the monopole makes it much harder to detect a possible low luminosity cutoff in the CO luminosity function using the VID.

4. Results

We are now ready to present the main numerical results from our analysis, starting with an inspection of the joint PS+VID covariance matrix itself.

4.1. Visual Inspection of the PS+VID Covariance Matrix

Figure 2 shows the pseudo-correlation matrices, c_{ij}, for our two experimental setups, as well as for pure signal alone, for reference. In order to illustrate the effect of the beam, we show covariance matrices from maps both without and with beam smoothing in the left and right columns, respectively.

The first thing to notice is that instrumental noise significantly reduces the numerical values of the normalized covariance matrices, bringing it closer to the independent white noise case for which $c_{ij} = \delta_{ij}$. This agrees with intuition, since the noise itself is white and uncorrelated.

Beam smoothing also leads to weaker correlations. This is mainly due to the beam diluting the signal at small scales, where the correlation is otherwise strongest.

Next, we notice that the cross-correlations between the power spectrum and VID are of the same order of magnitude as the correlations internal to each observable itself. Thus, it is essential to account for all these correlations in any joint PS and VID analysis, as is done in the present paper.

Finally, we note that when designing an experiment like COMAP, one of the important trade-offs involves observation time per field. To obtain a fast signal detection it is in general advantageous to observe deep on the smallest possible field. However, this only holds true while the signal-to-noise per voxel is significantly less than unity. When the noise starts to become comparable to the signal, the signal-induced voxel–voxel correlations starts to become important, and the effective uncertainties no longer scale as $\mathcal{O}(1/\sqrt{\tau})$, where τ is the observation time per pixel. Generally, in such a tradeoff, any significant correlations between different power spectrum modes or voxel temperatures will tend to favor larger survey area or multiple fields, both effectively leading to more independent samples, and thereby higher overall integration efficiency.

4.2. Luminosity Function Constraints

We are now ready to present both individual and joint PS +VID constraints on the CO luminosity function, which are summarized in Figure 3 for COMAP1 (*left column*) and COMAP2 (*right column*). The top row shows the constraints obtained from the power spectrum alone; the middle row shows the constrains obtained from the VID alone; and the third row shows the constraints from the joint analysis. In each panel, the shaded colored region shows the 95% credibility region from the MCMC samples and the solid line with the same color shows the posterior median. The purple solid line shows the average luminosity function obtained from the mean of all available halo catalogs, and thus represents the ensemble average of our input model. Note that the colored regions correspond to one single realization and the uncertainties therefore contain contributions from instrumental noise, cosmic variance, and sample variance. The agreement between the estimated confidence regions and the ensemble mean is quite satisfactory in all cases, with uncertainties that appear neither too large nor too small.

Considering first the individual PS and VID estimates, shown in the top two rows, we see that the two observables are indeed complementary. In particular, the VID primarily constrains the high luminosity end of the luminosity function, while the power spectrum imposes relatively stronger constraints on the low luminosity end. This makes sense intuitively, since the VID is essentially optimized to look for strong outliers above the noise, whereas the power spectrum represents a weighted mean across the full field for each physical scale. It is interesting to note, however, that the VID provides, on average, stronger constraints on the luminosity function than the power spectrum does.

Due to this complementarity, the joint estimator provides the strongest constraints of all. To make this point more explicit, the fourth row compares the uncertainties of the independent power spectrum and VID analyses to the joint constraints. Of course, there is a significant amount of cosmic variance in each of these functions and the precise numerical value of the uncertainty ratio therefore varies significantly with luminosity; but the mean trend is clear: The individual analyses typically result in 20%–70% larger uncertainties than the joint analysis when averaged over luminosities between $L_{\rm CO} = 10^4$–$10^7\,L_\odot$. Over 10 cosmological realizations, the PS and VID resulted in, on average, 58% and 30% larger uncertainties (in dex) individually, than the joint analysis. This is the main novel result presented in this paper.

4.3. Posterior Distribution of Model Parameters

Lastly we present the constraints of the model parameters themselves. When doing the MCMC posterior mapping we explore the parameter space of the Li et al. (2016) $L_{\rm CO}(M_{\rm Halo})$ model. Figure 4 shows the posterior distribution for these parameters derived from one realization of the COMAP2 experiment (the same realization as the COMAP2 results in Figure 3).

Results for PS, VID, and joint PS+VID analysis are shown in blue, red, and black, respectively. Prior distributions are shown in green. The two curves of each color correspond 68% and 95% credibility regions.

We see that the two parameters that are mainly constrained are α and β, the two parameters from the average $L_{\rm CO}$–$L_{\rm IR}$ relation. These two parameters are fairly degenerate, and the direction in which they are degenerate is given roughly by the line $\alpha = -0.1\beta + 1.19$ (Li et al. 2016). In Figure 5, we show the luminosity function for different points on this line. For the figure, the values of $\sigma_{\rm SFR}$, $\sigma_{L_{\rm CO}}$, and $\log\delta_{\rm MF}$ are fixed at 0.3, 0.3, and 0.0, respectively. Although the overall signal strength, at least in terms of detectability, is fairly constant along this line, the shape of the luminosity function changes significantly. Lower values of alpha imply a more steep power-law relation between $L_{\rm CO}$ and $L_{\rm IR}$ leading to more sources with very high or very low luminosities. We see this as a flattening of the luminosity function. In such a case, a larger fraction of the overall signal will be given by high-luminosity sources.

The other parameter that is also slightly constrained is the log-normal scatter parameter from the $L_{\rm CO}$–$L_{\rm IR}$ relation, $\sigma_{L_{\rm CO}}$. This parameter is only slightly more constrained compared to the prior, with the highest values of $\sigma_{L_{\rm CO}}$ being disfavored. The posterior of the other scatter parameter, $\sigma_{\rm SFR}$, is basically given by the corresponding prior (i.e., this parameter is not very well constrained by the experiment), although, as expected from the

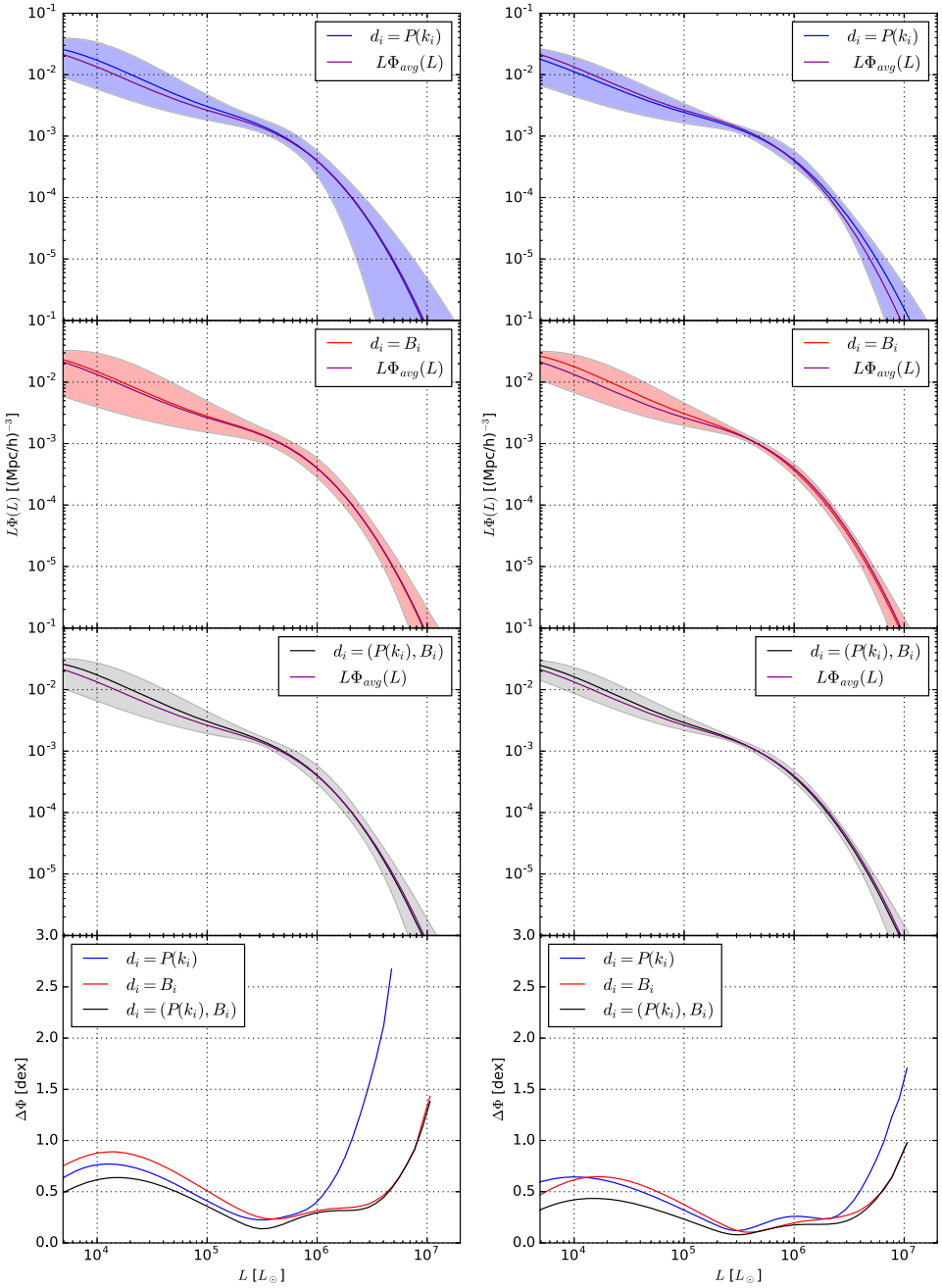

Figure 3. Constraints on the luminosity function from simulated experiments COMAP1 (left) and COMAP2 (right). The shaded area corresponds to 95% credibility intervals, solid lines correspond to the median, while the purple curve corresponds to the average luminosity function derived from all the available halo catalogs (i.e., the ensemble mean). Top row: constraints derived using only the power spectrum $P(k_i)$ as the observable. Middle row: constraints derived using only the temperature bin counts B_i as the observable. Bottom row: constraints derived by a joint analysis using both the power spectrum $P(k_i)$ and the temperature bin counts B_i as observables. Bottom: comparison of the uncertainty of the luminosity function constraints in dex, i.e., $\Delta\Phi \equiv \log_{10}\Phi_{97.5\%} - \log_{10}\Phi_{2.5\%}$.

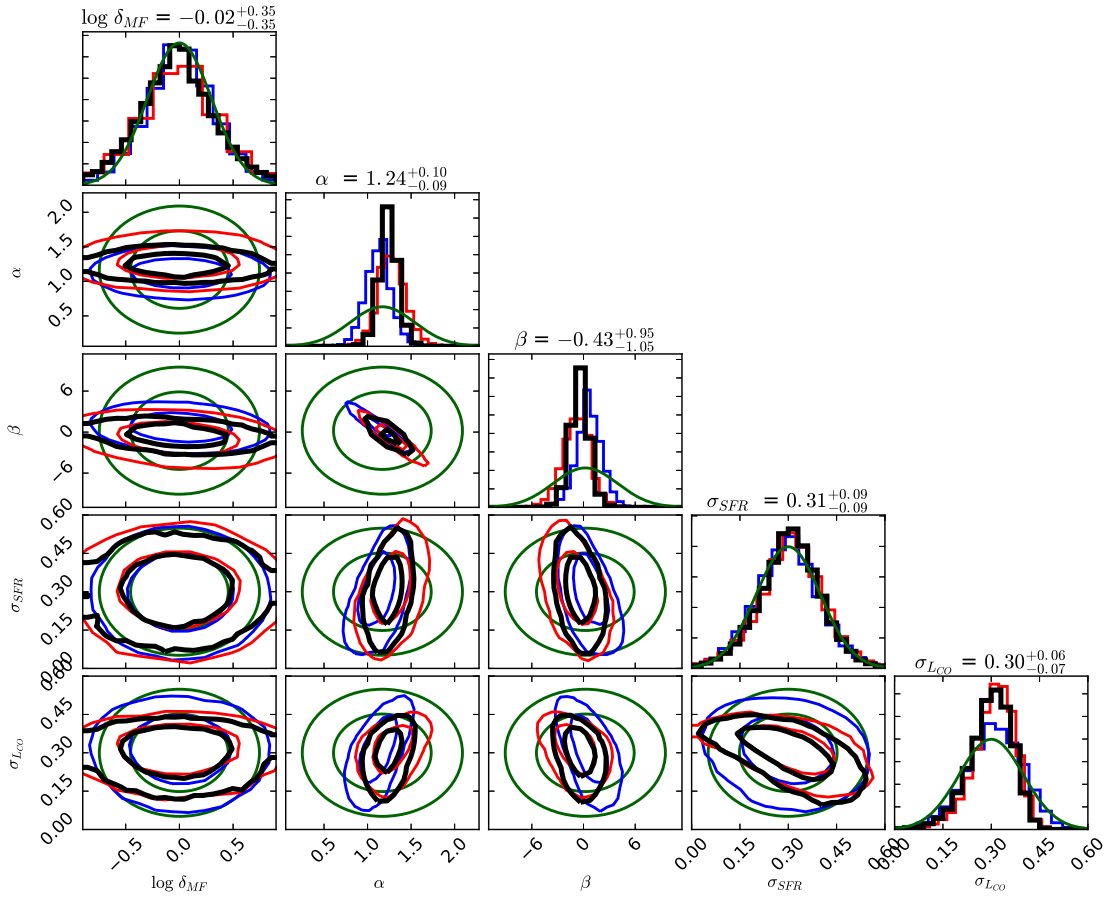

Figure 4. Posterior distributions for the Li et al. (2016) model parameters for a single realization of the COMAP2 experiment (the same realization as the COMAP2 results in Figure 3). Results for PS, VID, and joint PS+VID analysis are shown in blue, red, and (slightly bolder) black, respectively. Prior distributions are shown in green. The two curves of each color correspond to 68% and 95% credibility regions. The numbers on top of each column correspond to the 68% credibility interval for each parameter from the PS+VID analysis. We see that while the posterior of the two scatter parameters, $\sigma_{\rm SFR}$ and $\sigma_{L_{\rm CO}}$, is mostly set by the prior, the posterior on $\log \delta_{\rm MF}$, from the SFR–$L_{\rm IR}$ relation, is actually slightly wider than the prior, suggesting a significant intrinsic scatter in estimates of this parameter. These results are consistent with the corresponding results in Figure 7 in Li et al. (2016). The two parameters that are actually strongly constrained by the simulated experiment are α and β, the two parameters from the $L_{\rm CO}$–$L_{\rm IR}$ relation, and this figure shows that, at least for this realization, the constraints on these two get significantly improved in the combined analysis (PS+VID) as compared to analysis using the individual observables. This figure was created using the publicly available code (https://github.com/dfm/corner.py) `corner` (Foreman-Mackey 2016).

fact that the scatter parameters have basically the same effect, we see signs of the degeneracy between them in the posterior.

Interestingly, the normalization parameter in the SFR–$L_{\rm IR}$ relation, $\log \delta_{\rm MF}$, actually has a posterior that is wider than the prior. This may be because the best fit of this parameter from each of the different patches have an intrinsic scatter larger than the scatter in the prior. We note that we see the same effect in Li et al. (2016; their Figure 7).

From the mean relations in the Li et al. (2016) model, we have $\log L_{\rm CO} \sim -\beta - \log \delta_{\rm MF}$. Intuitively, we would then expect $\log \delta_{\rm MF}$ to be completely degenerate with β. However, since the SFR–$L_{\rm IR}$ is much better constrained by observations than the $L_{\rm CO}$–$L_{\rm IR}$ relation, the prior on $\log \delta_{\rm MF}$ is much tighter than the one on β. The degeneracy thus prevents us from constraining $\log \delta_{\rm MF}$ *until* β is constrained to a comparable level.

5. Discussion

We have developed a joint power spectrum and VID analysis for the CO luminosity function in the context of the COMAP CO intensity mapping experiment. We have implemented an efficient approach to estimating the joint covariance matrix for these two observables and have shown that accounting for both one- and two-point correlations leads to 20%–70% smaller uncertainties on the CO luminosity function for both COMAP1 and COMAP2.

The critical computational engine in our approach is the construction of fast yet semi-realistic simulations of the signal

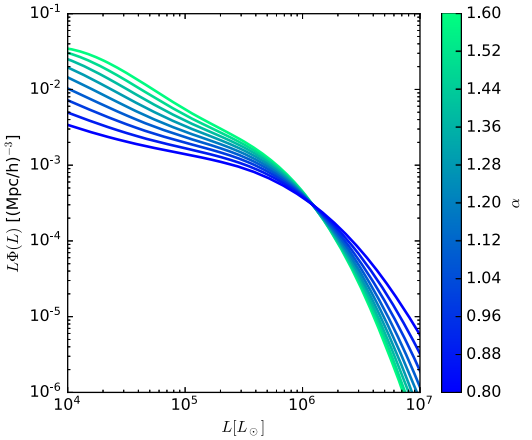

Figure 5. Plot of the CO luminosity function in the Li et al. (2016) model for different values of α and β. The colors of the lines indicate the values of α, the values of $\sigma_{\rm SFR}$, $\sigma_{L_{\rm CO}}$ and $\log \delta_{\rm MF}$ are fixed at 0.3, 0.3, and 0.0, respectively, while the value of β is determined from the relation $\alpha = -0.1\beta + 1.19$. This line corresponds roughly to the direction along which α and β are degenerate. Although the overall detectability of the signal remains roughly constant along this line, we see that the shape of the luminosity function changes significantly. We see that lower values of α correspond to less steep high-luminosity tails in the luminosity function, meaning that a larger proportion of the overall signal comes from high-luminosity sources.

in question. These simulations are based on the computationally cheap peak patch DM halo simulations produced by Bond & Myers (1996) and Stein et al. (2019), coupled to the semi-analytic CO luminosity model of Li et al. (2016). Of course, the results we derive are correspondingly limited by how well the model reproduces the true cosmological signal. If the true signal is significantly more complex than the model predicts, the constraints in Figure 3 will not be reliable.

The strength of the constraints on the CO luminosity function will depend on the overall level of the CO signal, which is highly uncertain. However, given the same rough level of signal, we expect the constraints on the luminosity function at the high luminosities to be less model dependent than the constraints on the $L_{\rm CO}$–$M_{\rm halo}$ relation or the luminosity function at lower luminosities. This is because the high-luminosity sources leave a fairly unique imprint on the maps that does not depend on the specific model used.

Additionally, we expect that the relative merits of using the PS or the VID as observables will change depending on the properties of the signal. In particular, anything that increases the shot noise of the signal, like a a strong galactic duty cycle, a large intrinsic scatter in luminosities or just a more top-heavy luminosity function, will make the resulting map more non-Gaussian, tending to favor observables like the VID more as compared to the PS. We can see this effect directly in Figure 4. The VID is better, compared to the PS, at ruling out low values of α and high values of $\sigma_{L_{\rm CO}}$, both of which correspond to cases where we would have a more top-heavy luminosity function and thus more shot noise.

We also expect the map to be more non-Gaussian on small scales than on large, so a wide survey with low resolution will tend to favor the PS, relative to the VID, more than a narrower high-resolution survey.

While the issues of model dependence are less relevant for low signal-to-noise measurements, where we are just trying to establish the rough level of the signal, they will become more important as the measurements improve.

Another potential issue with the simulations used in this paper is the minimum DM halo mass of $2.5 \times 10^{10}\,M_\odot$. While the model used here predicts that only a small fraction of the CO signal would come from halos lighter than this (see Li et al. 2016 and Chung et al. 2017), other models could disagree. If fact, searching for a low luminosity cutoff in the CO luminosity function is an interesting target for CO intensity mapping, and simulated halo catalogs with a smaller minimum DM halo mass would be useful both for forecasts and inference in such a scenario.

In general, it will be important to continuously improve the simulation pipeline as the experiment proceeds in order to account for more and more cosmological, astrophysical, and instrumental effects. However, the most important point in our approach is the fact that all such effects may be seamlessly accounted for, as long as the simulation procedure is sufficiently fast in order to be integrated into the MCMC procedure.

It should also be noted that our approach may be generalized in many different directions. For instance, the CO luminosity function does not play any unique role in our analysis, but is rather simply one specific worked example of a particularly interesting astrophysical function to be constrained. Many other functions may be constrained in a fully analogous manner, including, for instance, non-parametric $L_{\rm CO}(M_{\rm halo})$ models, or any of the parameters that are involved in converting the DM halo distributions to CO luminosities. The method is also not specific to CO intensity mapping, but should be equally well suited for other lines, or a combination of lines (Chung et al. 2018). Indeed, it should work for any type of random fields for which the covariance matrix must be estimated by simulations. Finally, we also note that there is nothing special about the power spectrum or VID as observables, but any other efficient data compression can be equally well included in the analysis, as long as the required compression step is sufficiently computationally efficient.

Support for the COMAP instrument and operation comes through the NSF cooperative agreement AST-1517598. Parts of this work were performed at the Jet Propulsion Laboratory (JPL) and California Institute of Technology, operating under a contract with the National Aeronautics and Space Administration. H.T.I., H.K.E., M.K.F., and I.K.W. acknowledge support from the Research Council of Norway through grant 251328. Research in Canada is supported by NSERC and CIFAR. These calculations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund—Research Excellence; and the University of Toronto. Work at Stanford University is supported by NSF AST-1517598 and by a seed grant from the Kavli Institute for Particle Astrophysics and Cosmology. J.O.G. acknowledges support from the Keck Institute for Space Studies, NSF AST-1517108, and the University of Miami. H.P.'s research is supported by the Tomalla Foundation.

We thank Sarah E. Church, Tim Pearson, and other members of the COMAP collaboration for useful discussion and comments on a draft of this paper.

ORCID iDs

D. Chung https://orcid.org/0000-0003-2618-6504

References

Battye, R. A., Davies, R. D., & Weller, J. 2004, MNRAS, 355, 1339
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJL, 762, L31
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013b, ApJ, 770, 57
Blot, L., Crocce, M., Sefusatti, E., et al. 2018, arXiv:1806.09497
Bond, J. R., Jaffe, A. H., & Knox, L. 2000, ApJ, 533, 19
Bond, J. R., & Myers, S. T. 1996, ApJS, 103, 1
Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, MNRAS, 467, 2996
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2014, MNRAS, 443, 3506
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2015, MNRAS, 452, 3408
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2016, MNRAS, 457, L127
Carilli, C. L. 2011, ApJL, 730, L30
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Chung, D. T., Li, T. Y., Viero, M. P., Church, S. E., & Wechsler, R. H. 2017, ApJ, 846, 60
Chung, D. T., Viero, M. P., Church, S. E., et al. 2018, ApJ, submitted (arXiv:1809.04550)
Cleary, K., Bigot-Sazy, M.-A., Chung, D., et al. 2016, AAS Meeting, 227, 426.06
Colavincenzo, M., Sefusatti, E., Monaco, P., et al. 2019, MNRAS, 482, 4883
da Cunha, E., Groves, B., Walter, F., et al. 2013, ApJ, 766, 13
Foreman-Mackey, D. 2016, JOSS, 1, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fu, J., Kauffmann, G., Li, C., & Guo, Q. 2012, MNRAS, 424, 2701
Gilks, W., Richardson, S., & Spiegelhalter, D. 1995, Markov Chain Monte Carlo in Practice, Chapman Hall/CRC Interdisciplinary Statistics (London: Taylor and Francis)
Glenn, J., Conley, A., Béthermin, M., et al. 2010, MNRAS, 409, 109
Gong, Y., Cooray, A., Silva, M. B., Santos, M. G., & Lubin, P. 2011, ApJL, 728, L46
Greve, T. R., Leonidaki, I., Xilouris, E. M., et al. 2014, ApJ, 794, 142
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv:1709.09066
Lee, S. K., Ando, S., & Kamionkowski, M. 2009, JCAP, 7, 007
Leicht, O., Uhlemann, C., Villaescusa-Navarro, F., et al. 2019, MNRAS, 484, 269
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, ApJ, 741, 70
Lippich, M., Sánchez, A. G., Colavincenzo, M., et al. 2019, MNRAS, 482, 1786
Loeb, A., & Wyithe, J. S. B. 2008, PhRvL, 100, 161301
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Mashian, N., Sternberg, A., & Loeb, A. 2015, JCAP, 11, 028
Obreschkow, D., Klöckner, H.-R., Heywood, I., Levrier, F., & Rawlings, S. 2009, ApJ, 703, 1890
Padmanabhan, H. 2018, MNRAS, 475, 1477
Peterson, J. B., Bandura, K., & Pen, U. L. 2006, arXiv:astro-ph/0606104
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Righi, M., Hernández-Monteagudo, C., & Sunyaev, R. A. 2008, A&A, 489, 489
Stein, G., Alvarez, M. A., & Bond, J. R. 2019, MNRAS, 483, 2236
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Vernstrom, T., Scott, D., Wall, J. V., et al. 2014, MNRAS, 440, 2791
Visbal, E., & Loeb, A. 2010, JCAP, 11, 016