

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

ALESSANDRO RIBEIRO MARINS

**TÉCNICAS DE SEPARAÇÃO DE COMPONENTES
APLICADAS AO BINGO TELESCOPE**

SÃO PAULO
JANEIRO/2022

Alessandro Ribeiro Marins

**Técnicas de Separação de Componentes aplicadas ao
BINGO Telescope**

Monografia apresentada ao Instituto de Física da Universidade de São Paulo como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

Orientador: Élcio Abdalla

Coorientador: Filipe Abdalla

Universidade de São Paulo – USP

Instituto de Física

São Paulo

Janeiro/2022

*Este trabalho é dedicado às minhas filhas, Laura e Valentina.
Àquelas que dedico não somente esse e outros trabalhos, mas todo o meu amor.*

Agradecimentos

Agradeço aos meus companheiros de grupo, meu orientador Prof. Dr. Élcio e colaboradores externos à Universidade de São Paulo, que têm me auxiliado imensamente nesse trilhar científico. Agradeço, em destaque, à CNPq pelo apoio e suporte financeiro que viabiliza tal projeto.

Resumo

A natureza da Energia Escura é um dos principais enigmas científicos nos tempos atuais e um dos principais métodos para estudá-la é por emissões de sinais de hidrogênio em períodos recentes do Universo. Em específico, o uso da linha de 21-cm do hidrogênio neutro através do método de mapeamento de intensidade é uma das técnicas mais promissoras para estudos de tal natureza. O BINGO Telescope é um projeto que objetiva fazer mapeamento de tal tipo de sinal em baixos redshifts - no regime de domínio da Energia Escura. Contudo, uma das principais dificuldades é que na mesma faixa de frequência que será observada pelo projeto há forte contaminação por diferentes outras fontes (astrofísicas, cosmológica, terrestres, de equipamentos e estatísticas) que sobrepuja-se em muitas ordens de grandeza o sinal desejado. Através de técnicas de Separação de Componentes do tipo *blind* (sem o uso de templates das diferentes contribuições do sinal), o atual trabalho de doutorado tem se dedicado a reconstruir os sinais de 21-cm através de selecionadas técnicas - já aplicadas à Radiação Cósmica de Fundo - e compará-las, apresentando aquela que reconstrói o sinal o mais acurado possível, dentro das especificações do BINGO.

Palavras-chaves: Cosmologia; 21-cm; Separação de Componentes; Mapeamento de Intensidade; BINGO Telescope.

Abstract

The nature of Dark Energy is one of the main scientific puzzles actually and one of the main methods to study about it is through emissions of hydrogen signals in recent epochs of the Universe. Specifically, the use of the 21-cm line of neutral hydrogen through intensity mapping method is one of the most promising techniques for studying of that one. The BINGO Telescope project aims to map 21-cm lines at low redshifts when Dark Energy comes to dominate. However, one of the main difficulties is that in same frequency band that will be observed by the project there is strong contamination by different other sources (astrophysical, cosmological, terrestrial, instrumental and statistical) that overlap by many orders of magnitude the desired signal. Using blind Component Separation techniques the current doctoral work has been dedicated to reconstructing the 21-cm signals by selected techniques - already applied to Cosmic Microwave Background Radiation - and compare them, presenting the one that reconstructs the signal as accurately as possible, within the BINGO specifications.

Key-words: Cosmology; 21-cm; Component Separation; Intensity Mapping; Bingo Telescope.

Sumário

	Introdução	11
I	HIDROGÊNIO E O ESTUDO DO UNIVERSO	13
1	UMA BREVE CONTEXTUALIZAÇÃO.	15
2	SINAIS ASTROFÍSICOS E COSMOLÓGICOS	17
2.1	Sinais de 21-cm	18
2.2	Sinais de foregrounds	19
2.2.1	Synchrotron	21
2.2.2	Free-Free	22
2.2.3	Anomalous Microwave Emission	23
2.2.4	Radio Point Sources	24
2.2.5	Cosmic Microwave Background	25
2.3	Ruído Térmico	26
II	SEPARAÇÃO DE COMPONENTES	27
3	MÉTODOS DE SEPARAÇÃO DE COMPONENTES APLICADOS AOS DADOS DO BINGO TELESCOPE	29
3.1	Modelando a emissão do céu	31
3.2	Métodos de Separação de Componentes	34
3.2.1	ICA	34
3.2.2	GMCA	35
3.2.3	GNILC	38
3.3	Espectro de Potência Angular dos Mapas	40
3.4	Noise Debias Process	41
III	RESULTADOS PRELIMINARES	43
3.5	Resultados preliminares	45
3.6	Conclusões	54
IV	OBSERVAÇÕES E PRÓXIMOS PASSOS	57
3.7	Próximos passos	59

3.8	Observações	59
------------	--------------------	-----------

REFERÊNCIAS	61
--------------------	-----------

Introdução

Os sinais de 21-cm pós-reionização provêm de regiões de densidade de coluna de HI¹ alta ($N_{\text{HI}} > 2 \cdot 10^{20} \text{ cm}^{-2}$), que blindam uma pequena fração do HI disponível no período ($x_{\text{HI}} \sim 2\%$), protegendo-os de radiação ionizante de fundo (Wolfe; Gawiser; Prochaska, 2005; Bird et al., 2014; Ho; Bird; Garnett, 2021). Tais regiões são conhecidas como *Damped Ly α* (DLA) e, por conterem uma grande fração do HI após a reionização, são uma prova direta da distribuição de gás neutro. O HI que permanece em tais regiões possui temperatura de spin (medida que mensura a razão entre a densidade numérica de triplete e singuleto) maior do que a temperatura da CMB (Kanekar; Chengalur, 2003), o que possibilita detectar emissões de 21-cm.

A temperatura de brilho de 21-cm pode ser obtida por duas maneiras distintas: (1) temperatura de brilho média sobre todos os ângulos do céu (sinal global) (Shaver et al., 1999) e (2) por flutuações na temperatura de brilho (Zaldarriaga; Furlanetto; Hernquist, 2004). As flutuações da temperatura de brilho do sinal de 21-cm pós-reionização, objeto de análise do BINGO, são da ordem de 0.1 mK. Contudo, na mesma faixa de frequência observada, que contempla todo sinal de 21-cm do período pós-reionização, diversas outras fontes de emissão em radiofrequência, principalmente as provenientes da Galáxia, não somente são observadas juntas ao sinal de 21-cm como são de temperaturas muito maiores do que a mesma. Sinais como *synchrotron* e *free-free* chegam a ser ~ 5 ordens maior do que o de 21-cm pós-reionização. Apesar dos sinais não desejados contaminarem o sinal alvo tais sinais astrofísicos e cosmológico possuem comportamento espectral suave (Dickinson; Davies; Davis, 2003; Santos; Cooray; Knox, 2005; McQuinn et al., 2006), diferente do sinal de 21-cm, em que flutuações na distribuição de HI ao longo da linha de visada levam a pequenas flutuações espectrais do sinal de 21-cm (Liu; Shaw, 2020). Portanto, diferente do que ocorre em CMB, pode-se explorar as características espectrais dos sinais por análise tomográfica em levantamentos do tipo mapeamento de intensidade para estimar os sinais alvos através de métodos não-paramétricos de *Separação de Componentes*.

De modo geral, denomina-se qualquer fonte astrofísica e cosmológica de sinal, na (faixa de) frequência a ser observada e que não contemple o sinal alvo, como *foreground*. Fontes de contaminação instrumental são denominados genericamente de *noise*; e, de modo geral, qualquer fonte de sinal de *componente*. Separação de Componentes é a área que objetiva reconstruir, ou melhor dizendo, estimar, as diferentes componentes observadas. No caso de métodos não-paramétricos, como os atualmente usados no BINGO e muito comum em dados de CMB, explora-se as características espectrais e morfológicas dos sinais através dos diferentes canais de frequência para realizar a estimação, e o que

¹ $N_{\text{HI}} = \int_{s_0}^{s_1} ds n_{\text{HI}}(s)$, com n_{HI} sendo a densidade numérica de HI.

diferencia cada método de estimação são as características espetrais e as informações estatísticas exploradas para estimar tais comportamentos espetrais.

O método GNILC (Olivari; Remazeilles; Dickinson, 2016a; Olivari et al., 2017; Fornazier et al., 2021) tem demonstrado conseguir resultados satisfatórios dentro das características do BINGO. Técnicas como FastICA e GMCA (Carucci; Irfan; Bobin, 2020; Cunningham et al., 2021a) já foram aplicadas em outros futuros levantamentos de 21-cm pós-reionização por mapeamento de intensidade com bons resultados, e estão sendo aplicados ao BINGO no trabalho em curso pelo doutorando. Contudo, tais métodos somente separam os foregrounds e fornecem como resíduo 21-cm + noise, de forma a precisarmos posteriormente estimar o sinal de 21-cm a partir de diferentes resíduos (*noise debias process*), ou seja, a partir de diferentes realizações de mapas de 21-cm. Tal processo busca corrigir tanto o efeito médio de estimação em diferentes realizações através de um fator de supressão quanto a contribuição média do ruído por diferentes templates (Fornazier et al., 2021; Marins et al., 2022). Essa etapa de processamento dos sinais é de grande importância para a identificação das BAOs e, por conseguinte, para as restrições de parâmetros.

Objetiva-se neste trabalho apresentar o status atual do processo de estimativa dos sinais de 21-cm por três métodos diferentes: GNILC, GMCA e FastICA.

A monografia está organizada como segue: primeiro capítulo inicia com uma breve contextualização do por que se usar sinais de 21-cm como medida para obter informações do Universo, seguida da descrições de como foi construído e o que compõe o conjunto de dados utilizados nas análises. O segundo capítulo é dedicado a descrever os métodos de separação de componentes utilizados para remoção dos foregrounds e reconstrução (estatística) do sinal de 21-cm. O terceiro capítulo contém os resultados a serem apresentados nessa monografia. Por fim, o quarto e último capítulo é dedicado aos próximos passos até a defesa da tese.

Parte I

Hidrogênio e o estudo do Universo

1 Uma breve contextualização.

Ao menos desde 1929, quando Edwin Hubble publica um trabalho analisando medidas de redshift e distância de 18 galáxias e conclui que o Universo está em expansão homogênea, sabe-se que o Universo é dinâmico. Tal descoberta pode ser vista como demarcando (no tempo) o início da cosmologia experimental. Não somente, causa uma revolução inclusive nos trabalhos teóricos sendo desenvolvidos naquele tempo, como os de Einstein que resistia até então à ideia de um Universo dinâmico, ao ponto de ter adicionado um termo constante às suas equações com objetivo de contrabalancear a atração gravitacional da matéria, hoje chamado de *constante cosmológica* Λ . Tal termo cai em desuso até final da década de 80 (do mesmo século), e na década de 90, quando dois grupos independentes liderados por A. Riess (*Supernova Cosmology Project*, 1998) e S. Perlmutter (*High-z Supernova Search Team*, 1999) ao analisarem dados de supernovas do tipo Ia (SNIa) convergem, de maneira independente, ao fato do Universo não ser inteiramente preenchido de matéria ordinária e com fortes evidências para estar se expandindo aceleradamente segundo uma constante cosmológica, que contém a maior parte da densidade de energia do Universo em períodos recentes (Weinberg et al., 2013). O problema estava posto: O que de fato causa tal efeito antigravitacional, seria uma constante cosmológica, ou seria algum outro tipo de fluido? ou ainda, será que as equações da relatividade geral não funcionam em escalas cosmológicas? Ao atacar a primeira pergunta, estamos lidando com a questão da expansão acelerada dentro do escopo da relatividade geral e assumindo que alguma fonte de energia, a qual denomina-se *Energia Escura*, está causando tal efeito. Já no caso da segunda, busca-se explicações fora da relatividade geral, principalmente com extensões da mesma a qual os modelos denominam-se genericamente como modelos de *Gravidade Modificada*.

No mesmo período dos resultados das SNIa, estava em operação o satélite *Cosmic Background Explorer* (COBE), que operou de 1989 à 1996, e fez a medida mais precisa, até o período, dos sinais mais antigos possíveis de se obter do Universo, quando o mesmo tinha ainda algumas centenas de milhares de anos, e com espectro de corpo negro, compatível com um Universo primordial denso e quente. Os resultados científicos do COBE apontavam para um Universo inflacionário plano com boa concordância com uma constante cosmológica, e a matéria sendo composta por duas componentes: uma bariônica e uma exótica denominada de *escura fria* (*cold dark matter*, CDM) (Bennett et al., 1993). Tal modelo cosmológico é denominado de modelo flat- Λ CDM e é considerado o modelo padrão cosmológico. Sem dúvida a(s) natureza(s) do que venham a ser matéria escura e a energia escura, formando o que se denomina hoje de *setor escuro*, se põe como um dos maiores enigmas modernos da ciência. Atualmente, há diversas evidências independentes

do setor escuro para além das duas citadas acima, e, por exemplo, medidas precisas feitas pelo PLANCK (Aghanim et al., 2020), que reforçam o modelo Λ CDM. Assim sendo, há diversos modelos que tentam descrever o setor escuro, assim como diversos projetos que objetivam coletar informações que possibilitem restringir tais modelos (para mais informações sobre modelos e métodos observacionais cosmológicas, veja (Abdalla; Marins, 2020)). No que consente aos métodos, as próximas décadas tendem a ter um avanço significativo nas que exploram emissões de hidrogênio neutro (HI), em especial a de 21-cm, como fonte de informação. No período após a época da Reionização¹, que se denomina justamente por *pós-reionização*, da ordem de 1% dos bárions estão na forma de HI dentro de regiões extremamente densas denominadas de *Damped Ly α system* (DLAs), $\Omega_{\text{HI}} \sim 0.01$, e esses dentro de halos de matéria com massas entre $10^{10} - 10^{13} M_{\odot}$ (ícone para unidades de massas solares) (Liu; Shaw, 2020), o que justifica a afirmação de HI pós-reionização ser um bom traçador da distribuição total de matéria. Ao mesmo tempo, para compensar a baixa probabilidade de emissão do sinal de 21-cm a partir de um átomo, a abundância de hidrogênio no Universo possibilita o seu uso através da emissão integrada das contribuições de fôtons das diferentes linhas de visada cobertas por um levantamento, técnica conhecida como Mapeamento de Intensidade (*Intensity Mapping*, IM), que tem o benefício de, não somente cobrir um grande volume do Universo com medida precisa da distância radial, como não sofrer com a limitação de um limiar mínimo de luminosidade, uma vez que necessita justamente da contribuição integrada das fontes emissoras, e portanto, mesmo fontes fracas contribuem.

Segundo o modelo Λ CDM, a partir de um redshift de $z \sim 0.3$ ocorre o início do domínio da Energia Escura na dinâmica do Universo, ou seja, no período pós-reionização. Assim sendo, o projeto BINGO está sendo desenvolvido para operar através da técnica de IM com sinais de 21-cm pós-reionização na faixa de frequência que contempla a transição para o domínio da Energia Escura, 0.13-0.48. E para tal, objetiva obter a medida estatística das *Oscilações Acústicas de Bárions* (*Baryon Acoustic Oscillations*, BAOs)² como método de restrição de parâmetros cosmológicos, e em especial, de modelos de Energia Escura (Abdalla et al., 2021).

Na próxima seção, estarei interessado em modelar os sinais que se espera observar com o BINGO, o que constitui parte da pipeline do projeto.

¹ Após o desacoplamento da radiação, que dá origem a CMB, prótons e elétrons se recombinam, principalmente, em hidrogênio neutro e o Universo se torna transparente e inicia um período denominado de *Dark Ages*. Após centenas de milhões de anos, com a formação das primeiras fontes luminosas, tais como estrelas de população III e II e quasares, que emitem radiação ionizante, o Universo passa a um novo período denominado de *Reionização*, em que quase todo HI disponível fora destruído. Esse período, basicamente, termina quando o hidrogênio ionizado (HII) preenche muito do volume entre galáxias (Liu; Shaw, 2020).

² As BAOs são informações que podem ser obtidas da distribuição de matéria do Universo e remetem ao período do acoplamento entre fôtons e elétrons. Para mais detalhes, ver (Bassett; Hlozek, 2010).

2 Sinais astrofísicos e cosmológicos

A pipeline do BINGO está sendo desenvolvida concomitantemente à sua construção instrumental. Os dados de emissão de 21-cm que estão sendo simulados se dão por duas maneiras: simulação do tipo (distribuição) lognormal, gerada por uma extensão do código FLASK (Xavier; Abdalla; Joachimi, 2016) para 21-cm; e de simulação de N-corpos (Zhang et al., 2021). Para o atual trabalho os dados usados foram os de distribuição lognormal, que usa como input espectros de potência angulares teóricos provenientes do código UCLC ℓ (McLeod; Balan; Abdalla, 2017). O módulo da pipeline que simula as observações do radiotelescópio (que consiste nas coletas dos *dados brutos*, registros em TODs¹ e o conversão desses em mapas, ou seja, o *map-making*) ainda está em desenvolvimento. Assim, os dados usados nesse trabalho ainda não estão sendo pré-processados de forma a incluir efeitos observacionais do mapeamento. O que fora até então feito é restringir a região analisada à região de cobertura e com a resolução angular esperada pelo projeto.

A região do céu a ser coberta pelo projeto corresponderá ao intervalo de declinação de -22.5° à -7.5° (em coordenadas celestiais), que será realizada através de um *drift scan* por 28 cornetas receptoras com mobilidade longitudinal (sobre o plano focal). Como descrito em (Abdalla et al., 2021), durante a fase 1, cada corneta terá um posicionamento fixo no plano focal durante cerca de 1 ano, e sendo a fase 1 estimada em 5 anos, haverá 5 regiões distintas a serem mapeadas por cada corneta. Os receptores possuem temperatura de sistema de aproximadamente 70 K e deverão medir os parâmetros de Stokes I e V (Wuensche et al., 2021), uma vez que o design óptico é do tipo crossed-Dragonian (Dragone, 1978) que tem como uma das características minimizar as polarizações cruzadas. A faixa de frequência a ser observada será 980-1260 MHz, escolhida dentre os outros motivos por mitigar RFI (*radio frequency interference*), composta por 30 canais de resolução de 9.33 MHz. Isso faz com que usemos uma análise tomográfica com 30 mapas observacionais (compostos por sinais de 21-cm + foregrounds). Cada um dos mapas nesse trabalho é mascarado (como será melhor descrito no Parte 3, capítulo 4) com uma máscara binária, restringindo a declinação à região de cobertura. Além disso, os mapas observados são convoluídos com feixe gaussiano com *full width half maximum* (FWHM) de 40 arcmin, que corresponde a resolução angular do feixe do BINGO, e posteriormente adicionados aos mesmos ruídos instrumentais gaussianos. Ou seja, nesse trabalho não há inclusão de efeitos de *polarization leakage* e ruído 1/f, que esperamos realizar em breve.

No que segue, procurarei fazer uma breve descrição dos mapas criados e utilizados nas análises da Parte 3, capítulo 4. E naquele capítulo, descreverei como eles são usados

¹ *time ordered datasets*

para montar o dataset. O objetivo aqui não será desenvolver todos os cálculos para chegar até a(s) expressão(ões) do sinal - o que está feito para a tese - mas sim, fornecer um mínimo necessário para a compreensão do trabalho apresentado e do que constitue o conjunto de dados utilizados.

2.1 Sinais de 21-cm

Na quase totalidade das situações astrofísicas se tem que a temperatura de spin (medida que mensura a razão entre a densidade numérica de tripleto e singuleto) de um gás de Hidrogênio neutro satisfaz a relação $1 - \exp(-T_\star/T_S) \approx T_\star/T_S$, pois $T_\star \ll T_S$ e com $T_\star \doteq h\nu_{10}/k_B$, no qual $\nu_{10} = 1420.4$ MHz é a frequência de transição hiperfina em seu referencial de repouso, k_B e h constantes de Boltzmann e Planck, respectivamente. Isso nos possibilita descrever o coeficiente de emissividade do gás como sendo:

$$\alpha_\nu = \left(\frac{3hc^2 A_{10}}{32\pi k_B \nu_{10}} \right) \frac{n_{\text{HI}}}{T_S} \phi^{(10)}(\nu),$$

com $A_{10} = 2.85 \times 10^{-15} s^{-1}$ coeficiente de emissão espontânea da transição de 21-cm, n_{HI} densidade numérica de átomos HI e $\phi^{(10)}$ perfil de linha normalizado, que será assumido uma distribuição delta de Dirac centrada em ν_{10} . Integrando a contribuição do coeficiente ao longo da linha de visada através da nuvem de gás obtemos a profundidade ótica, que pode ser escrita como (Bharadwaj; Nath; Sethi, 2001; Bharadwaj; Ali, 2005; Battye et al., 2013):

$$\tau_{\nu_{10}} = \int_{\text{nuvem}} ds \alpha_\nu(s) \approx \left(\frac{9hc^3 A_{10}}{256\pi^2 G k_B \nu_{10}^2 m_{\text{HI}}} \right) \frac{\Omega_{\text{HI}}(z)}{(1+z)} \frac{H_0^2}{\|dv_{\parallel}/d\chi\|} \frac{1}{T_S(z)}, \quad (2.1)$$

em que m_{HI} é a massa do átomo de HI, G constante gravitacional, H_0 constante de Hubble e $\Omega_{\text{HI}}(z) = 8\pi G m_{\text{HI}} n_{\text{HI}}(z)/3H_0^2$ parâmetro de densidade de HI no redshift z. $\|dv_{\parallel}/d\chi\|$ é o gradiente do campo de velocidade peculiar ao longo da linha de visada, em coordenadas comóveis $\chi = a^{-1}s$, com o fator de escala $a = a(z)$. Como estamos lidando com o regime de frequência de rádio, o campo de radiação pode ser descrito em unidades de temperatura através da aproximação de Rayleigh-Jeans, que nesse caso leva à definição da temperatura de brilho (Field, 1958). Além disso, o caso aqui tratado é $\tau_{\nu_{10}} \ll 1$, o que, pela equação integral da transferência radiativa e levando em consideração que a emissão é afetada pela expansão do Universo durante o seu trajeto sobre a linha de visada até a observação, $\nu_{10} = \nu(1+z)$, escrevemos a temperatura de brilho de 21-cm observada (média) como sendo (Furlanetto; Oh; Briggs, 2006; Pritchard; Loeb, 2012)

$$T_{\text{HI}}(z) = \left(\frac{9hc^3 A_{10}}{256\pi^2 G k_B \nu_{10}^2 m_{\text{HI}}} \right) \frac{\Omega_{\text{HI}}(z)}{(1+z)^2} \frac{H_0^2}{\|dv_{\parallel}/d\chi\|} \left(1 - \frac{T_{\text{CMB}}(z)}{T_S(z)} \right)$$

em que $T_{\text{CMB}}(z) = 2.725(1+z)$ K. Assumiremos aqui que o processo de reionização aqueceu o campo de densidade do gás acima da temperatura de CMB tal que temos

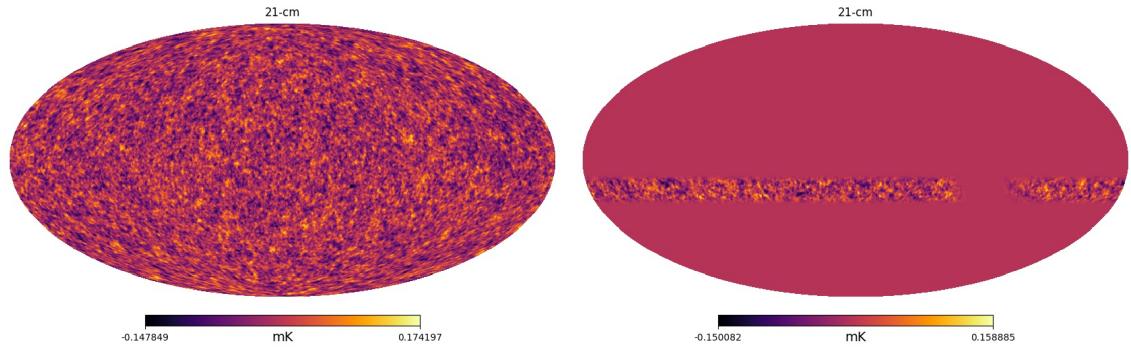


Figura 1 – Mapas de flutuações de temperatura de brilho de 21-cm (em mK) na faixa de frequência 1059-1070 MHz. Na esquerda, o mapa é de céu inteiro, e na direita mapa restrito à região de observação do BINGO Telescope.

$T_S(z) \gg T_{\text{CMB}}(z)$, para $z < 9$ ([Santos; Cooray; Knox, 2005](#); [Morales; Wyithe, 2010](#)). Isso também é válido para DLAs à $z < 2$. ([Kanekar; Chengalur, 2003](#)). Ou seja,

$$T_{\text{HI}}(z) = \left(\frac{9hc^3 A_{10}}{256\pi^2 Gk_B \nu_{10}^2 m_{\text{HI}}} \right) \frac{\Omega_{\text{HI}}(z)}{(1+z)^2} \frac{H_0^2}{\|dv_{\parallel}/d\chi\|}$$

expressão equivalente as contidas nos trabalhos ([Battye et al., 2013](#); [Costa et al., 2021](#)) quando se assume somente o efeito do fluxo de Hubble no gradiente (comóvel) do campo de velocidade peculiar. Contudo, diferentes mecanismos ao longo da linha de visada da emissão causam flutuações na distribuição de HI, que, em primeira ordem, $n_{\text{HI}}(z, \hat{n}) = \bar{n}_{\text{HI}}(z)(1 + \delta_{\text{HI}}(z, \hat{n}))$, na direção \hat{n} . O que, consequentemente, leva a uma perturbação, em primeira ordem, de $T_{\text{HI}}(z, \hat{n}) = \bar{T}_{\text{HI}}(z)(1 + \delta T_{\text{HI}}(z, \hat{n}))$. O termo perturbativo da temperatura é a medida objetivada pelo BINGO e tem em sua expressão (relativística) efeitos diversos tais como: *Redshift Space Distortion* (RSD), efeito *Sachs-Wolfe* integrado, perturbação da própria distribuição de matéria, dentre outros. A expressão completa pode ser vista em ([Costa et al., 2021](#)).

O código UCLC ℓ gera os mapas teórico δT_{HI} e constrói espectros de potência angulares HI-HI, ou seja, entre dois campos de 21-cm, em cada correspondente intervalo de redshift dos 30 canais de frequência. Tais resultados são usados como entrada do código FLASK, que por sua vez constrói mapas lognormais de tal forma que os mapas ainda obedeçam a função de correlação inicial. Tais mapas são usados como mapas de (flutuações de) 21-cm no Parte 3, capítulo 4.

2.2 Sinais de foregrounds

Os sinais de 21-cm são extremamente fracos quando comparados às demais emissões astrofísicas e cosmológicas que são detectadas pelos receptores. No caso da faixa de frequência de interesse do BINGO, 980-1260 MHz, que é uma faixa de baixa frequência quando comparada, por exemplo, ao PLANCK, que operou com nove canais entre

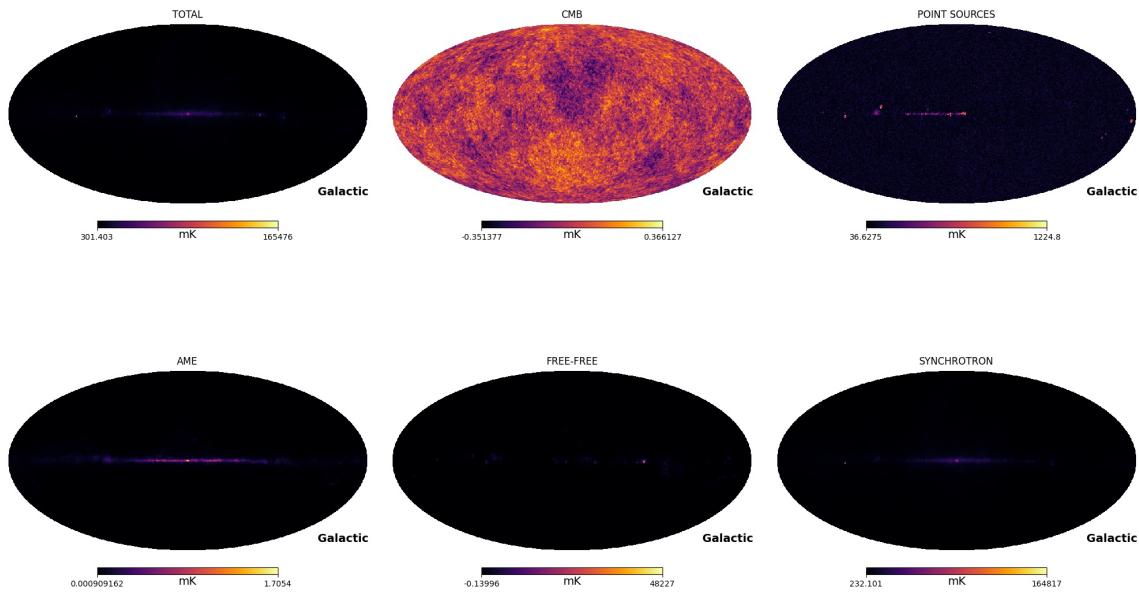


Figura 2 – Todos os mapas de temperatura estão em mK e em coordenadas Galácticas no intervalo 1059-1070 MHz.

30-857 GHz, as principais contribuições de fontes que não são 21-cm, e que chamaremos simplesmente de *foregrounds*, podem ser divididas em dois grupos: Emissões Galácticas difusas e emissões extragalácticas. De modo geral, as emissões Galácticas são provenientes do meio interestelar (ISM) da Galáxia (Via Láctea), que por sua vez é constituída por nuvens moleculares e atômicas frias, um meio entre nuvens que é parcialmente ionizado e regiões ionizadas presumivelmente formadas por supernovas. Tais meios estão fortemente concentrados no plano Galáctico. Em tal classificação usamos as três componentes com maiores contribuições: *synchrotron*, *free-free* e *anomalous microwave emission*. Já para emissões extragalácticas, no caso aqui posto, é assumido a componente chamada de *radio point sources*. Estarei incluindo no caso de emissão extragaláctica também as emissões de CMB.

De modo geral, os mapas de foregrounds são construídos a partir de templates usando o fato da temperatura de brilho das mesmas terem comportamento espectral suave, ou seja, cada uma podendo ser aproximada por alguma potência da frequência. Os mapas de foregrounds foram gerados através do código *Planck Sky Model* (PSM) (Delabrouille et al., 2013) nas 30 bandas entre 960-1260 MHz². A Figura 2 apresenta os foregrounds e a sua soma em unidades de mK nas coordenadas Galácticas, em que a Galáxia fica na região central da imagem de cada mapa. Já na Figura 3 a representação se dá em coordenadas celestiais, como será o caso do BINGO.

² A montagem do dataset a ser utilizada foi criada antes da mudança para frequência mínima de 980 MHz, devido a restrição instrumental de interferência.

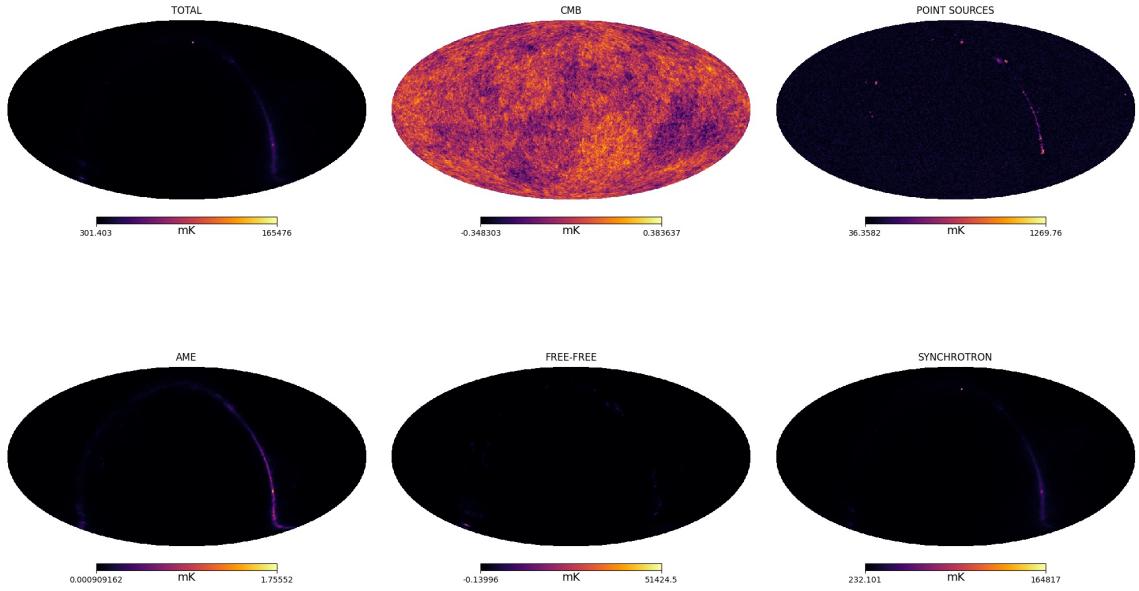


Figura 3 – Todos os mapas de temperatura estão em mK e em coordenadas celestiais no intervalo 1059-1070 MHz.

2.2.1 Synchrotron

Emissão synchrotron é uma emissão não-térmica emitida por partículas carregadas espiralando em um campo magnético. Na nossa galáxia, raios cósmicos relativísticos (basicamente, elétrons ou, por vezes, pósitrons (Chapman; Jelić, 2019)) são acelerados pelo campo magnético Galáctico interestelar. Como na nossa galáxia tais campos se estendem para fora do plano Galático, as emissões synchrotron são menos concentradas em tal plano do que as demais emissões Galácticas. Na Figura 4 temos a representação do mapa synchrotron (à esquerda) em unidades de mK e em coordenadas Galácticas para o intervalo de frequência de 1059-1070 MHz. Nesse fica nítido as emissões fora do plano Galático. Também podemos ver pelo mapa dos foregrounds sem a contribuição synchrotron (à direita) que a emissão synchrotron corresponde a maior intensidade de emissão entre os foregrounds e que as outras emissões são mais concentradas no plano Galático.

A densidade numérica de tais elétrons de raios cósmicos segue a relação de distribuição de energia $n_e(E)dE \sim E^\gamma dE$ (Chapman; Jelić, 2019), para algum índice de raios cósmico γ ; e, então, a intensidade específica segue uma lei de potência da forma $I_\nu \sim n_e \nu^{-\alpha}$, com $\alpha = (\gamma - 1)/2$. Usando o fato de estarmos interessado no regime de frequência de rádio e podermos usar a aproximação de Rayleigh-Jeans, a temperatura de brilho da emissão synchrotron é do tipo $T_{\text{syn}} \sim \nu^\beta$, com o índice espectral $\beta = 2 + \alpha$. Os templates numa dada frequência ν são descritos, em cada pixel (ou em uma dada direção \hat{n}), com relação a algum template observational a uma frequência ν_0 , segundo

$$T_{\text{syn}}(\nu, p) = T_{\text{syn}}(\nu_0, p) \left(\frac{\nu}{\nu_0} \right)^{\beta(p) + \kappa(\nu)},$$

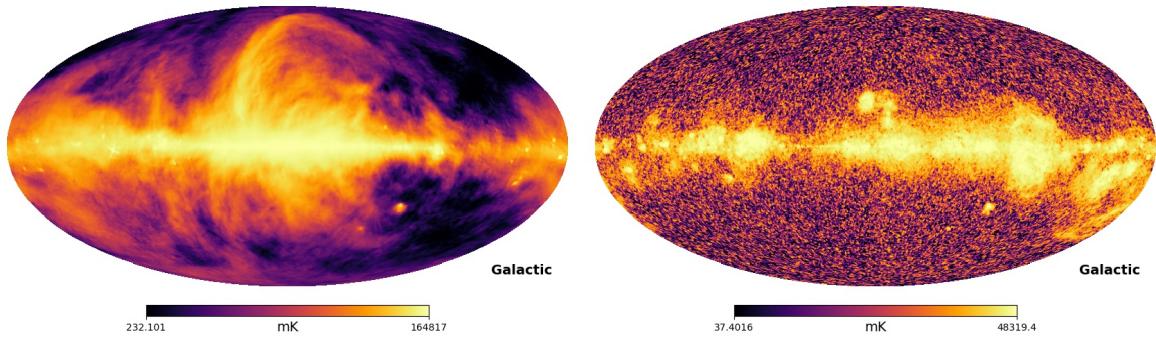


Figura 4 – O mapa à esquerda é de temperatura synchrotron e o à direita, o mapa das emissões de foregrounds menos synchrotron. Ambos em unidade de mK e em coordenadas Galácticas no intervalo 1059-1070 MHz. Os mapas foram gerados com normalização da cor para obter contrastes mais nítidos. Com tais contrastes fica nítido a localização do plano Galáctico e as emissões nos polos, longe do plano Galáctico, assim como as demais emissões que mais contribuem são provenientes do plano Galáctico.

com $\kappa(\nu)$ alguma possível correção devido a frequência (Kogut et al., 2007), como por exemplo algum outro mecanismo que quando o elétron cósmico se propaga através do ISM o faça perder energia por interação. O índice espectral β depende da posição, uma vez que podem ocorrer tanto variações na densidade numérica de raios cósmicos, quanto da intensidade do campo magnético perpendicular à linha de visada, a depender da posição. Assim, o índice espectral da lei de potência da temperatura de brilho synchrotron varia tipicamente entre 2.5 a 3.2. O template usado é o fornecido por (Remazeilles et al., 2015) a $\nu_0 = 408$ MHz, e do céu inteiro. Já o índice espectral, uma distribuição não uniforme no céu fornecido pelo modelo 4 em (Miville-Deschénes et al., 2008).

2.2.2 Free-Free

Também conhecida como *Bremsstrahlung térmico*, são emissões devido ao espalhamento de um elétron com um íon na parte ionizada do ISM Galáctico. Seu índice espectral varia muito pouco por todo céu. Diferente das emissões synchrotron, suas variações estão fortemente concentradas no plano focal. Tem um comportamento como $T_{\text{ff}}(\nu) \sim \nu^{-2.1}$, isso é, um índice espectral de $\beta \approx 2.1$. O template usado para emissão se baseia no template gerado por (Dickinson; Davies; Davis, 2003), que possui uma expressão fechada

$$T_b(T_e, \nu_{\text{GHz}}) = 8.396 \times 10^3 a(T_e, \nu_{\text{GHz}}) \left(\frac{T_e}{10^4} \right)^{0.667} \times 10^{0.029/T_e} \times 1.08 \nu_{\text{GHz}}^{-2.1},$$

em que o fator de correção Gaunt é dado por

$$a(T_e, \nu_{\text{GHz}}) = 0.366 \nu_{\text{GHz}}^{0.1} T_e^{-0.15} \times \left[\ln \left(4.995 \times 10^{-2} \nu_{\text{GHz}}^{-1} \right) + 1.5 \ln (T_e) \right],$$

e assumindo a temperatura eletrônica constante de $T_e = 7 \times 10^3$ K. Vide que ν_{GHz} significa que a frequência é assumida em unidades de GHz nessa expressão.

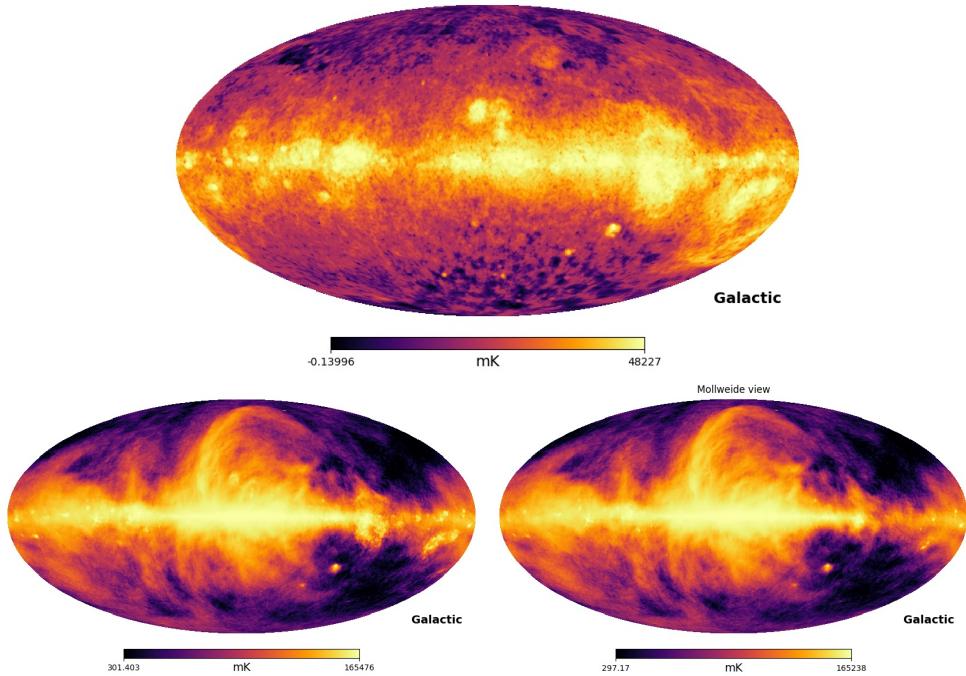


Figura 5 – O mapa superior é o de temperatura free-free, o inferior à esquerda o mapa de emissões totais de foregrounds e o inferior à direita o mapa total de foregrounds retirada a contribuição free-free. Todos estão em unidade de mK e em coordenadas Galácticas no intervalo 1059-1070 MHz. O mapa foi gerado com normalização da cor para obter contrastes mais nítidos.

Na Figura 5 podemos ver o mapa de free-free para o intervalo de 1059-1070 MHz, em coordenadas Galácticas. Nela podemos ver que no mapa de foregrounds sem a contribuição free-free, quando comparado com o mapa com free-free, tem um diminuição nítida de emissão na parte central direita (leste). A mesma estrutura é visível na imagem das emissões free-free (mapa superior).

2.2.3 Anomalous Microwave Emission

Anomalous Microwave Emission (AME) é uma componente que foi observada pela primeira vez em meados da década de 1990 em experimentos de CMB. Dados do COBE (Kogut et al., 1996) apresentaram um sinal significativo nos resultados de correlação cruzada entre dados da CMB (do radiômetro DMR, *Differential Microwave Radiometer*, a 53 GHz) com mapas de poeira no *far-infrared* (FIR) (mapas do radiômetro multibandas DIRBE, *Diffuse Infrared Background Experiment*). Na ocasião, a componente foi vista como sendo uma componente da emissão free-free correlacionada com emissões de poeira no FIR, e como compõe ao menos um terço da emissão total free-free à 53 GHz. Já observações feitas pelo OVRO (telescópio *Owens Valley Radio Observatory*) (Leitch et al., 1997), para CMB em 14.5 e 32 GHz, apresentavam um forte sinal em ambas as bandas com índice espectral inclinado da ordem de 2 e amplitude de temperatura de brilho de

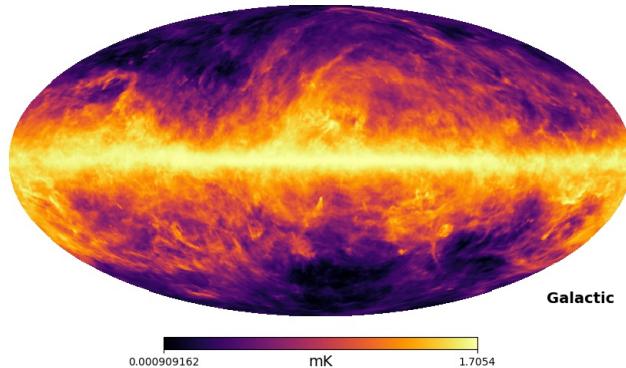


Figura 6 – Mapa de temperatura AME em unidade de mK e em coordenadas Galácticas no intervalo 1059-1070 MHz. O mapa foi gerado com normalização da cor para obter contrastes mais nítidos.

$\sim 300 \mu\text{K}$ à 14.5 GHz. Em seguida, uma possibilidade de explicação que toma força é a da emissão "anômala" ser devido às emissões de grãos de poeira muito pequenos, o que leva ao modelo Draine-Lazarian para emissões de poeiras muito pequenas em rotação no meio interestelar (Draine; Lazarian, 1998), chamado de *Spinning Dust Emission* (SDE), e que em (de Oliveira-Costa et al., 1999) é usado num ajuste de combinação linear entre emissões por spinning dust e emissões por vibrações de grãos maiores (*Thermal Dust Emission*) em análise de correlação entre dados de CMB, do projeto TENERIFE (Gutiérrez et al., 2000), com mapas de poeira do DIRBE e com mapa synchrotron HASLAM 408GHz (Haslam et al., 1981). O trabalho demonstra que a emissão não é consistente com emissão free-free e que o ajuste fornece boas evidências para tal componente ser um spinning dust. As atuais detecções da AME se dão entre ~ 10 -100 GHz e a mais plausível explicação até o momento é de que tal emissão se deva a radiação proveniente da interação do dipolo elétrico, ou magnético, de pequenos grãos de poeira ($\sim 1 \text{ nm}$) girantes no ISM, tal que a potência total emitida segue a fórmula de Larmor (Dickinson et al., 2018). Portanto, supõe-se que a emissão AME seja uma SDE segundo uma simples lei de potência baseada na combinação do modelo Draine-Lazarian extrapolado para frequências menores do que as com medições; e, para frequências maiores, segundo um template de thermal dust do PLANCK a 353 GHz e escalonado para spinning dust a 22.8 GHz, como descrito em (Fornazier et al., 2021). O mapa gerado para intervalo de frequências de 1059-1070 MHz pode ser visto na Figura 6.

2.2.4 Radio Point Sources

Devido a baixa resolução da observação por IM, diferentes mecanismos físicos contribuem para tal componente. As fontes de emissões extragalácticas de rádio são provenientes, principalmente, de núcleos ativos de galáxias (AGNs, *active galactic nuclei*) e por galáxias com formação estelar (SFGs, *star-forming galaxies*). As AGNs emitem radiação

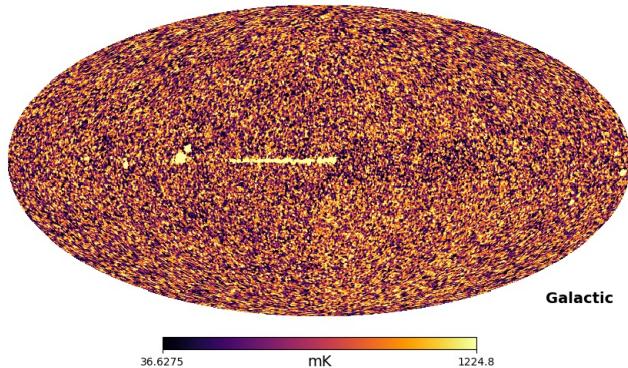


Figura 7 – Mapa de temperatura radio point sources em unidade de mK e em coordenadas Galácticas no intervalo 1059-1070 MHz. O mapa foi gerado com normalização da cor para obter contrastes mais nítidos. Podemos ver que a fontes fortes não resolvidas próximo ao plano focal, e um fundo difuso

synchrotron a partir da acreção de matéria pelo buraco negro supermassivo central, que em seguida as ejeta num feixe perpendicular ao plano de acreção. Já SFGs produzem emissões synchrotron similares as da nossa galáxia e emissões free-free a partir de regiões de hidrogênio ionizado (HII) (Chapman; Jelić, 2019). No código PSM³, *point sources* são classificadas em três categorias: fontes de rádio (que contém também as fontes provenientes do WMAP), fontes de infravermelho e regiões ultra-compactas de HII. O modelo usado neste trabalho usa a composição entre fontes pontuais muito brilhantes (fontes fortes) catalogadas com medidas entre 0.85-4.85 GHz (mais especificamente, em três frequências 0.843, 1.4 e 4.85 GHz. Catálogo disponível na tabela 2 em (Delabrouille et al., 2013)) e um fundo difuso composto pela soma de fontes não resolvidas em um simples pixel (fontes fracas), como pode ser visto na Figura 7. Este último mapa é convoluído com um feixe gaussiano no espaço harmônico. Cada radio point source é modelada com uma amplitude e um índice espectral, ou determinada diretamente através de dados multifrequência, ou extraídas aleatoriamente a partir de distribuição observada de índices espetrais (Delabrouille et al., 2013). Fontes com fluxos de densidade acima do limite 5σ do PLANCK nas bandas 30, 70, 353 e 857 GHz são fontes que deverão ser removidas dos dados do BINGO e, por isso, não são usadas neste trabalho (Fornazier et al., 2021).

2.2.5 Cosmic Microwave Background

A Cosmic Microwave Background (CMB) é uma radiação proveniente do desacoplamento entre elétrons e prótons no Universo primordial, que ocorre na época em que a taxa de expansão do Universo se torna maior do que a taxa de interação, em $z = 1089$ (Aghanim et al., 2020). Sua temperatura média hoje é de $T_{\text{CMB}} = 2.725 \text{ K}$, com pequenas flutuações anisotrópicas, e seu espectro é um espectro de corpo negro. Suas anisotropias,

³ O link para o manual do usuário do PSM (vol 1.9.0) se encontra [aqui](#).

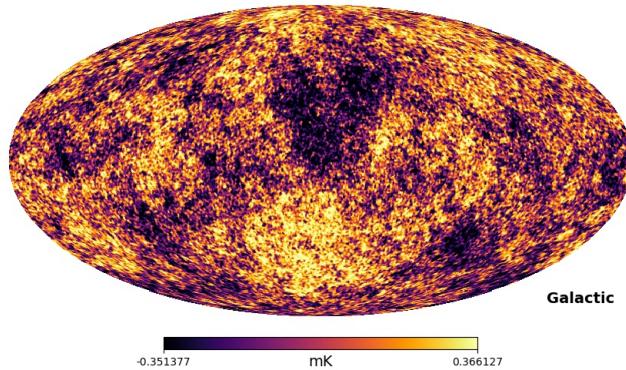


Figura 8 – Mapa de temperatura de CMB em unidade de mK e em coordenadas Galácticas, 1059-1070 MHz. O mapa foi gerado com normalização da cor para obter contrastes mais nítidos. O mapa de flutuações da CMB contém o efeito de dipolo.

tanto em medidas de intensidade quanto de polarizações, são fontes riquíssimas para compreensão do Universo primordial, assim como efeitos físicos que afetam os fótons de CMB na linha de visada, tais como: lentes gravitacionais, reionização, dentre outros. Os mapas de anisotropias da intensidade da CMB que foram usadas são baseados no best-fit do PLANCK ([Aghanim et al., 2020](#)). Na Figura 8 temos a representação das emissões da CMB para uma representação em coordenadas Galácticas entre 1059 e 1070 MHz. Vide que a distribuição de emissões não se apresenta homogênea e há um efeito de dipolo bem nítido.

2.3 Ruído Térmico

O ruído térmico é do tipo *white noise* (segue uma distribuição gaussiana no céu), e assume o fato do levantamento ser composto por 28 cornetas. Para obter o mapa de ruído térmico, usei o fato do feixe possuir FWHM de 40 arcmin, com 15% de cobertura do céu feita através das 28 cornetas com tempo de observação de um ano. Também usei uma pixelização na esfera do tipo HEALPIX ([Górski et al., 2005](#)) com parâmetro $N_{\text{side}} = 512$ (sendo o número de pixels, a partir desse parâmetro, dado por $12 \times N_{\text{side}}^2$), fornecendo um nível de ruído térmico por feixe, em cada canal, da ordem de 5 mK.

Parte II

Separação de Componentes

3 Métodos de Separação de Componentes aplicados aos dados do BINGO Telescope

Separação de Componentes é um termo genérico que codifica a busca por conseguir reconstruir sinais de fontes físicas (e instrumentais) diferentes que são obtidas conjuntamente numa observação ou simulação. Para ser preciso, não ocorre reconstrução a menos que se saiba o que realmente compõe os sinais observados e todas as suas informações. Logo, o que ocorre é uma estimativa dos sinais¹. Como infelizmente experimentos tais como de CMB e de Intensity Mapping não identificam sinal(is) alvo separadamente, os sinais que são obtidos são misturas, em cada canal, de sinais provenientes de fontes físicas vindas do céu que convoluem com o feixe e sofrem a adição de ruídos instrumentais. Ou seja, há, em princípio, dois processos que dificultam a obtenção do sinal alvo. É comum chamarmos todas as diferentes fontes de sinais - sejam astrofísicas ou cosmológicas -, pré-processados até a reconstrução dos mapas por map-making, de *componentes* e por isso do nome Separação de Componentes. Vide que com isso tanto cada fonte de foreground e quanto 21-cm é uma componente. A área de Separação de Componentes em IM incorpora os conhecimentos, técnicas e algoritmos que foram desenvolvidos para o caso de CMB, adaptando-os às suas necessidades.

O sinal de 21-cm, ou melhor dizendo, as flutuações da temperatura de brilho de 21-cm², são extremamente fracas frente às outras contaminações do céu, inclusive mais fracas que o ruído térmico e às flutuações de temperatura da CMB, como apresentado nas Figuras 2 e 3. No caso, por exemplo, de emissões synchrotron Galácticas, estas chegam a ser de 5 a 6 ordens de magnitude maiores do que sinais de 21-cm no plano Galáctico. Contudo, diferentes informações podem ser extraídas diretamente dos mapas - e aqui ter uma análise multifreqüência melhora e muito os resultados - que podem auxiliar na estimativa dos sinais. A começar pelo comportamento espectral das diferentes fontes serem distintos, e em especial, ao contrário dos foregrounds que possuem comportamento espectral suave³, os sinais de 21-cm e de white noise não. Isso significa que em uma análise multifreqüência os sinais de foregrounds são fortemente correlacionados entre as frequências, enquanto 21-

¹ Por toda essa monografia, estarei assumindo que Separação de Componentes (*Component Separation*) e Remoção de Foregrounds (*Foreground Removing*) são sinônimos, mesmo que em um sentido preciso isso não seja verdade. Para ser detalhista, o que será feito é um *Foreground Removing*, pois não há interesse nesse trabalho em outra informação de sinais que não as de 21-cm. Assim, só haverá remoção de foregrounds e não reconstrução das diferentes componentes que compõe o sinal.

² Daqui para frente, quando eu me referir à sinal de 21-cm estarei me referindo justamente as flutuações de temperatura de brilho de 21-cm.

³ O caso da CMB é um pouco mais sutil. Quando expressa em unidades termodinâmicas, seu comportamento espectral é unitario para todas as bandas. Contudo, o mesmo não é verdade em unidades de temperatura de brilho.

cm e ruído não são. Os sinais de 21-cm, devido as flutuações primordiais, descritas pelos modelos inflacionários mais simples como, p.ex., modelo *slow-roll inflation*, são (quase-)Gaussianos⁴. Outras características são as suas morfologias distintas, como pudermos ver nos mapas da Parte 1, capítulo 2. Sinais de 21-cm são homogêneos, free-free são altamente concentrados no plano Galáctico, synchrotron também tem forte concentração no plano Galáctico mas devido as linhas de campos magnéticos da Galáxia sairem do seu plano as emissões também surgem de regiões fora do mesmo, emissões AME também se concentram no plano Galáctico e fontes pontuais de emissão de rádio podem ser identificadas por técnicas morfológicas e filtros wavelets isotrópicos (Starck; Murtagh; Bertero, 2011; Carrón Duque et al., 2019). Sinais da CMB, após remoção do efeito de dipolo, são homogêneos, mas com comportamento espectral de corpo negro⁵. Fazer análises dos sinais no espaço real (ou, equivalentemente, dos pixels) tem os benefícios de pudermos fazer análises em partes separadas do céu. Mas há ainda a possibilidade de fazer análises em outros espaços matemáticos tais como no espaço harmônico, que fornece melhor controle e modelagem do feixe; e no espaço wavelet, que assim como o espaço real consegue ser feito por partes do céu e explorar melhor características morfológicas e informativas dos pixels enquanto media contribuições de ruídos, uma vez que wavelets são muito bem localizadas tanto no espaço real (seja em relação ao espaço ou do tempo) quanto no espaço de frequências (Starck; Fadili; Murtagh, 2007).

Os resultados fornecidos pela Separação de Componentes, que podemos pensar como a última etapa do pré-processamento das observações, são de suma importância para a detectabilidade das BAOs e para a qualidade das restrições de parâmetros cosmológicos, e consequentemente, das restrições de modelos de Energia Escura. Uma reconstrução de mapas com alto nível de resíduos (remanescentes) de foregrounds pode inviabilizar o objetivo científico final, de tal forma que a compreensão dos algoritmos e as suas efetivas aplicações verificando quais são os que melhor estimam os mapas de 21-cm é central na construção da pipeline do BINGO.

Nas próximas seções objetivo apresentar como métodos os do tipo *blind*, ou seja, os métodos que não assumem conhecimento prévio dos mecanismos de emissão, e modelam o problema a ser atacado de forma a estimar os mapas. Não estarei nessa monografia me dedicando a descrição minuciosa dos algoritmos - fato a ser feito na tese - mas sim na compreensão dos seus funcionamentos e nas informações relevantes para as análises finais. Vale também ressaltar que os modelos de separação de componentes não se restringem aos

⁴ Termo *quase* deve ser tomado com cuidado, pois, dependendo da análise estatística que se faça tais características não gaussianas do sinal não são identificadas, necessitando de análises estatísticas de ordem maiores (p.ex., Bispectrum). Também que os efeitos não-Gaussianos que surgem não se devem (necessariamente) às flutuações primordiais, mas sim, são efeitos introduzidos nos dados por efeitos físicos secundários tais como gradientes de velocidade e densidades das emissões (Fornazier et al., 2021), por resíduos de foregrounds e por sistemáticas instrumentais.

⁵ O espectro das flutuações da intensidade da CMB são da derivada do espectro de corpo negro.

do tipo blind. Há outros métodos, como por exemplo, o método paramétrico baseado em inferência bayesiana COMMANDER (Eriksen et al., 2008); e métodos do tipo *semi-blind*, como por exemplo, sGMCA (Bobin; Acero; Picquenot, 2019; Cunningham et al., 2021b) que utiliza informações da lei de emissão de algumas componentes, como por exemplo, para CMB e Sunyaev-Zeldovich.

3.1 Modelando a emissão do céu

Esqueçamos por um momento o efeito observacional do feixe sobre o sinal observado e imaginemos também, por ora, que não há sinal de 21-cm nas observações - o motivo disso ficará claro mais a frente -, nos restringindo a um problema muito comum em CMB que é de estimar as leis de emissão. Consideremos, então, que o i -ésimo canal registra o sinal de uma dada localização p^6 do céu como sendo uma mistura linear e instantânea (ruídososa) de componentes estatisticamente independentes. Se chamarmos a observação, no i -ésimo canal, de $x_i(p)$, e de $y_j(\nu_i, p)$ o sinal dá j -ésima fonte no dado canal e na dada posição, temos que para N_s fontes

$$x_i(p) = \sum_{j=1}^{N_s} y_j(\nu_i, p) + n_i(p),$$

com n_i sendo o ruído instrumental. A maioria dos modelos de Separação de Componentes são do tipo *mixing model*, que são modelos que assumem que o sinal de cada fonte pode ser representado por uma parte estritamente espacial e outra estritamente espectral; ou seja, uma parte que contém o padrão de emissão no céu de uma certa componente e outra parte que representa a lei de emissão (a informação espectral). Seja $s_j(p)$ o padrão de emissão da j -ésima fonte na posição p do céu, e a_{ij} a lei de emissão da j -ésima componente no i -ésimo canal de frequência, temos a relação

$$x_i(p) = \sum_{j=1}^{N_s} a_{ij} s_j(p) + n_i(p). \quad (3.1)$$

A passagem dessa expressão para uma forma vetorial é direta ao assumirmos que cada componente do vetor é um localização, $\mathbf{x}_i = [x_i(p_1), \dots, x_i(p_k), \dots]$ compreendendo todas as posições do céu⁷ e cada vetor representará um mapa do céu. Em uma representação matricial, útil para uma análise multifrequência, reescrevemos a relação acima como sendo

$$\mathbb{X} = \mathbb{A}\mathbb{S} + \mathbb{N}, \quad (3.2)$$

⁶ Aqui p pode indicar um direção no céu descrita por um certo par de ângulos (θ_p, ϕ_p) , ou, como estou interessado, o pixel de uma certa pixelização que é centrado nesses ângulos.

⁷ Vide que são valores discretos, pois, estamos assumindo uma certa pixelização do céu, ou seja, haverá um número finito N_{pix} de valores a serem usados.

que para as matrizes \mathbb{X} e \mathbb{N} cada linha representa um mapa de um dado canal⁸. Assim, tais matrizes têm dimensões de número de canais por número de posições no céu (pixels) $N_{\text{ch}} \times N_{\text{pix}}$. A matriz \mathbb{A} é chamada de *mixing matrix* e cada coluna sua representa a lei espectral de uma fonte, e por isso sua dimensão é $N_{\text{ch}} \times N_s$. Por fim, \mathbb{S} tem em cada linha o padrão de emissão no céu de uma dada fonte, com dimensão $N_s \times N_{\text{pix}}$. Logo, o problema de estimar as diferentes componentes contidas nos mapas observados é um problema de estimar as matrizes \mathbb{A} e \mathbb{S} , que é um típico problema de inversão linear, e estimar fontes por esse método por vezes é referido como *Blind Source Separation* (BSS).

No caso de $N_{\text{ch}} = N_s$, $\det(\mathbb{A}) \neq 0$ e de não haver ruído, o problema possui solução única⁹ e essa é dada pela inversão da mixing matrix, \mathbb{A}^{-1} . Contudo, em problemas de interesse astrofísico, isso quase nunca é verdade. Além disso, por ser um método blind sequer devemos saber *a priori* o número de fontes que compõe as emissões do céu. É justamente a maneira de estimar ambas as matrizes que caracteriza um método (um algoritmo) em Separação de Componentes do tipo BSS.

Uma forma de resolver esse problema é buscar algum tipo de operador que ao ser aplicado sobre a observação as "desmistura". Algo como um \mathbb{W} , que leva à estimativa $\hat{\mathbb{S}} = \mathbb{W}\mathbb{X}$ ¹⁰. Ou ainda, a estimativa dos mapas,

$$\hat{\mathbb{Y}} = \hat{\mathbb{A}}\hat{\mathbb{S}} = \hat{\mathbb{A}}(\mathbb{W}\mathbb{X}) = \mathbb{W}_{\text{FG}}\mathbb{X}, \quad (3.3)$$

em que defino o operador

$$\mathbb{W}_{\text{FG}} \doteq \hat{\mathbb{A}}\mathbb{W}. \quad (3.4)$$

Vide que o operador \mathbb{W} quando aplicado às observações leva a relação

$$\hat{\mathbb{S}} = \mathbb{W}(\mathbb{A}\mathbb{S} + \mathbb{N}) = (\mathbb{W}\mathbb{A})\mathbb{S} + \mathbb{W}\mathbb{N}$$

que para um caso sem ruído térmico e na qual a inversa da mixing matrix existe, esse operador é exatamente a inversa da mixing matrix, $\mathbb{W} = \mathbb{A}^{-1}$, e a estimativa é perfeita. No caso de não haver ruído térmico e a mixing matrix não possuir inversa, \mathbb{W} deve ser uma pseudo-inversa. O problema pode ser encarado com $\mathbb{W}\mathbb{A}$ o mais próximo possível da matriz identidade e $\mathbb{W}\mathbb{S}$ o mais perto possível da matriz nula. Visto de uma maneira mais lúdica, podemos imaginar \mathbb{W} como um filtro, que se aplicado às observações filtra as fontes astrofísicas e cosmológicas e retira o ruído instrumental. Uma forma natural e a mais simples de resolver esse problema é pelo critério de mínimos quadrados (*least*

⁸ Meu interesse é em falar em termos voltados para o caso de observações de Intensity Mapping de tal forma que as observações são feitas em canais, em bandas de frequências. Mas a mesma ideia se aplica a uma única frequência, como é o ocorre em alguns casos de observações e análises em experimentos da CMB.

⁹ No caso sem ruído, se $N_{\text{ch}} < N_s$, o problema não possui solução. Já se $N_{\text{ch}} > N_s$, há infinitas soluções.

¹⁰ Usarei o chapéu para representar a estimativa de algo.

squares, LS), tentando minimizar o ruído em relação da matriz fonte \mathbb{S} ,

$$\begin{aligned}\nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{S}}} \varepsilon_{\text{LS}} &= \nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{S}}} \left(\frac{1}{2} \|\mathbb{N}\|^2 \right) = 0 \\ &= \nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{S}}} \left(\frac{1}{2} (\mathbb{X} - \mathbb{A}\mathbb{S})^T (\mathbb{X} - \mathbb{A}\mathbb{S}) \right) = 0 \\ &= (\hat{\mathbb{A}}^T \hat{\mathbb{A}}) \hat{\mathbb{S}} - \hat{\mathbb{A}}^T \mathbb{X} \\ \hat{\mathbb{S}} &= \left[(\hat{\mathbb{A}}^T \hat{\mathbb{A}})^{-1} \hat{\mathbb{A}}^T \right] \mathbb{X},\end{aligned}$$

com $\hat{\mathbb{A}}$ e $\hat{\mathbb{S}}$ valores estimados das respectivas mixing matrix e matriz fonte. O termo entre colchetes é conhecido como *pseudo inversa de Moore-Penrose* e é simbolizado por \mathbb{A}^+ . Assim, esse tipo de solução leva ao filtro ser uma pseudo-inversa de Moore-Penrose,

$$\therefore \mathbb{W} = \mathbb{A}^+ = (\hat{\mathbb{A}}^T \hat{\mathbb{A}})^{-1} \hat{\mathbb{A}}^T \quad (3.5)$$

Note-se que usamos na minimização uma métrica canônica, usando a norma em ℓ_2 , $\|\cdot\|_{\ell_2}$, mas podemos também usar outra norma como a $\|\cdot\|_{\mathbb{Q}}$, segundo alguma matriz \mathbb{Q} . Nesse caso, a minimização é uma generalização do caso anterior e chamada de mínimos quadrados generalizados (*Generalized Least Squares*, GLS)

$$\begin{aligned}\nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{S}}} \varepsilon_{\text{GLS}} &= \nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{S}}} \left(\frac{1}{2} \|\mathbb{N}\|_{\mathbb{Q}}^2 \right) = 0 \\ &= \nabla_{\mathbb{S}} \Big|_{\hat{\mathbb{A}}, \hat{\mathbb{A}}} \left(\frac{1}{2} (\mathbb{X} - \mathbb{A}\mathbb{S})^T \mathbb{Q} (\mathbb{X} - \mathbb{A}\mathbb{S}) \right) = 0 \\ &= (\hat{\mathbb{A}}^T \mathbb{Q} \hat{\mathbb{A}}) \hat{\mathbb{S}} - (\hat{\mathbb{A}}^T \mathbb{Q}) \mathbb{X} \\ \hat{\mathbb{S}} &= \left[(\hat{\mathbb{A}}^T \mathbb{Q} \hat{\mathbb{A}})^{-1} (\hat{\mathbb{A}}^T \mathbb{Q}) \right] \mathbb{X}.\end{aligned}$$

Logo, o filtro desse caso é dado por

$$\therefore \mathbb{W} = (\hat{\mathbb{A}}^T \mathbb{Q} \hat{\mathbb{A}})^{-1} (\hat{\mathbb{A}}^T \mathbb{Q}), \quad (3.6)$$

que também é conhecido em Teoria da Informação como *estimador de Gauss-Markov*. Comumente, toma-se como matriz \mathbb{Q} a inversa da matriz de correlação do ruído, $\mathbb{C}_{\mathbb{N}}^{-1}$, ou mesmo a matriz de correlação das observações, $\mathbb{C}_{\mathbb{X}}^{-1}$. Há ainda outros estimadores como, por exemplo, o filtro de Wiener (Paykari; Starck, 2012). Outro exemplo é, como veremos a frente, quando é assumido uma norma de Frobenius na restrição mais a condição dos mapas decompostos em dicionário sobrecompleto serem os mais esparsos possíveis.

Os cálculos feitos acima que resultaram em 3.5 e 3.6 foram feitos seguindo os raciocínios iniciais no espaço de pixels (domínio real), mas o mesmo vale para outros espaço de interesse como o harmônico e wavelet. Uma Transformada Wavelet Redundante (Starck; Fadili; Murtagh, 2007) é antes de mais nada um operador linear. Veremos como é tratado o caso harmônico no método GNILC.

Agora, a observação a ser feita é que em nenhum momento eu disse estar lidando com 21-cm e o que de certa forma os métodos BSS fazem é estimar a matriz das respostas espetrais das fontes, que são suaves. No caso do problema de inversão linear do tipo BSS, os métodos não estimam o sinal de 21-cm. Esse, por ter uma resposta espectral (quase-)gaussiana, é lido como um ruído. Portanto, o que fazemos é tentar estimar a reconstrução dos foregrounds como um todo o mais preciso possível e então extraímos o seu resíduo, que será composto de 21-cm + ruído térmico. Disso fica claro que haverá uma etapa posterior ao método de Separação de Componentes que deverá estimar os mapas de 21-cm a partir dos mapas de resíduos. Expressando de forma mais clara,

$$\mathbb{R} = \mathbb{X} - \hat{\mathbb{Y}} = (1 - \mathbb{W}_{\text{FG}}) \mathbb{X},$$

em que \mathbb{R} é uma matriz $N_{\text{ch}} \times N_{\text{pix}}$ que contém a estimação de 21-cm + ruído térmico. O processo de estimação dos N_{ch} mapas de 21-cm será chamado de *Noise Debias Process*, como será descrito em detalhes mais a frente.

Ainda em tempo, o problema de inversão linear se torna mais complicado ao termos de considerar o efeito instrumental, que impõe uma convolução do feixe com o céu observado e que pode ser diferente para cada canal,

$$x_i(p) = b_j(p) * \left(\sum_{j=1}^{N_s} a_{ij} s_j(p) \right) + n_i(p), \quad (3.7)$$

sendo $b_j(p)$ o feixe instrumental, que aqui eu assumi ser gaussiano.

3.2 Métodos de Separação de Componentes

Neste trabalho de monografia estarei me restringindo a três métodos de separação de componentes: GNILC, GMCA e ICA (através do algoritmo FastICA). Todos esses são métodos do tipo *blind*, não assumem nenhum conhecimento dos processos físicos das fontes emissoras. Contudo, cada um busca fazer a separação entre foregrounds e 21-cm + ruído explorando técnicas diferentes, como serão descritas a seguir.

3.2.1 ICA

O ICA (*Independent Component Analysis*) é uma extensão do PCA (*Principal Component Analysis*) desenvolvido na década de 1980 na área de redes neurais e que se tornou muito comum em resoluções de problemas de inversão linear. Neste método não é preciso fazer hipóteses sobre a escala de frequência ou a mixing matrix, a hipótese que é feita é a de que as fontes emitem sinais estatisticamente independentes e que tais sinais são não gaussianos. Os sinais serem estatisticamente independentes entre si significa que

um sinal não carrega informações dos demais sinais. Em termos matemáticos, dois sinais são ditos serem estatisticamente independentes se a função de distribuição conjunta dos sinais é igual ao produto das funções de distribuição individuais (marginalizadas). A mixing matrix relaciona as componentes independentes dos sinais com os sinais observados e a sua estimativa é feita maximizando sua não gaussianidade. Seu algoritmo FastICA foi introduzido para problemas de Separação de Componentes de interesse astrofísico em (Maino et al., 2002) objetivando estimar as matrizes de padrão espacial \mathbb{S} e espectral \mathbb{A} assumindo que diferentes fontes astrofísicas e cosmológicas, que compõe um sinal observado, são estatisticamente independentes, e que ao menos uma das componentes é não gaussiana, numa análise multifrequência. FastICA não estima o sinal de 21-cm, nem o ruído térmico. Ele estima a informação proveniente de fontes não-gaussianas, de forma que não há distinção entre 21-cm e ruído; ambos sendo tratados como um único resíduo.

Seja \mathbf{s}_i a i -ésima linha da matriz \mathbb{S} , e \mathbf{w}_i a i -ésima linha da matriz \mathbb{W} , estimada pelo ICA, então

$$\mathbf{s}_i = \sum_{j=1}^{N_{\text{ch}}} W_{ij} \mathbf{x}_j = \mathbf{w}_i \mathbb{X}$$

é uma componente independente (*Independent Component*, IC), obtido pela minimização da gaussianidade do termo $\mathbf{w}_i \mathbb{X}$. Essa minimização é feita segundo alguma medida que mensura não gaussianidade, que no caso do FastICA é a *negentropy* (Hyvärinen, 1999), que mede quão maior a entropia de um valor dista da entropia de um valor gaussiano.

O ICA foi utilizado em dados de 21-cm em (Chapman et al., 2012) no contexto de estudo da época da reionização, onde foi usada a maximização da não gaussianidade do sinal mistura das componentes independentes dos foregrounds, para modelar e extrair os foregrounds do sinal de 21-cm e, neste caso, mostrou-se um método eficiente e promissor na reconstrução dos sinais dos foregrounds.

3.2.2 GMCA

Um sinal y é dito ser esparso em um dicionário¹¹ Φ se ele pode ser representado por poucos elementos do dicionário. Por exemplo, o sinal y pode ser representado por

$$y = \alpha \Phi = \sum_{\gamma \in \Gamma} \alpha_\gamma \phi_\gamma$$

com α_γ sendo os coeficientes da representação. Se y for esparso em Φ , a quantidade de coeficientes não nulos será muito pequena. De modo mais preciso, y é dito ser k -esparso em Φ se o suporte dos seus coeficientes tenha cardinalidade k , ou seja, se houverem somente

¹¹ Dicionário é uma generalização de uma base ortonormal formada por uma coleção de formas de onda $\{\phi_\gamma\}_{\gamma \in \Gamma}$, com Γ sendo algum conjunto de índices. Um dicionário não é necessariamente completo, ele também pode ser subcompleto ou mesmo sobrecompleto, como por exemplo, sendo formado por uniões de bases ortogonais, o que leva a representações não únicas de um sinal (Chen; Donoho, 1994; Chen; Donoho; Saunders, 2001).

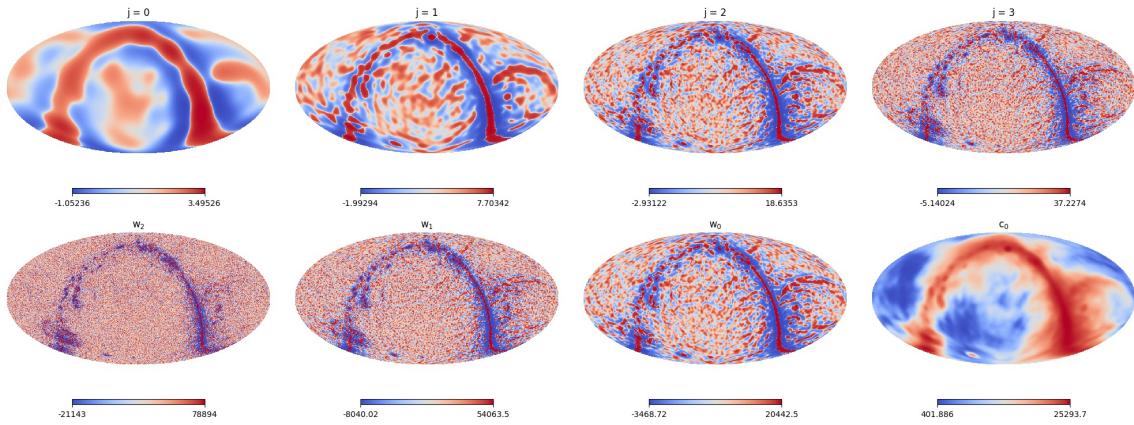


Figura 9 – Os mapas da linha superior são os mapas de coeficientes da decomposição needlet (do tipo *mexican*) usando o código MTNeedlet (Pietrobon et al., 2010). Nos mapas de baixo são os mesmos mapas de foregrounds na mesma faixa de frequência só que decomposto em três camadas starlets. c_0 sendo a camada com a descrição das informações suaves, e as w os detalhes.

k elementos não nulos. Contudo, sinais de interesse prático não são em geral estritamente esparsos (com k muito menor que a quantidade de elementos α_j). Mas, ao invés disso, podem ser compressíveis ou fracamente esparsos, no sentido de que as magnitudes dos coeficientes decaem muito rapidamente (Starck; Murtagh; Fadili, 2010). Logo, seria interessante encontrar representações dos sinais em que eles sejam esparsos. Nesse sentido, o uso de wavelets para compressão de informações é vasta, como, por exemplo, no uso em armazenamento de informações de digitais de dedos e nos algoritmos de imagens do tipo JPEG2000 (Starck; Siebenmorgen; Gredel, 1997; Starck; Fadili; Murtagh, 2007), por serem muito bem localizadas tanto no domínio do espaço (ou tempo) quanto da frequência. Diferente do caso de Fourier, que é adequado para sinais estacionários, ou seja, só é bem localizada no domínio das frequências. Por essa característica de localidade (suporte localizado), wavelets são úteis para representação esparsa de um dado sinal. Wavelets na esfera são bem localizadas tanto no espaço harmônico quanto no espaço real e se tornaram muito comuns em análises de mapas em astrofísica. Dois exemplos são as needlets (Mariucci et al., 2008) e as starlets (Starck; Murtagh; Bertero, 2011). Na Figura 9 apresento a decomposição de um mapa observacional com a soma das componentes de foregrounds entre 1059-1070 MHz. As primeiras wavelets esféricas, needlets, são usadas no algoritmo GNILC, que veremos a frente, para fazer uma análise localizada no espaço harmônico. As segundas, starlets, são comuns para identificação de imagens astronômicas de fontes pontuais que são isotrópicas, ou aproximadamente isotrópicas.

O primeiro trabalho a usar esparsividade como condição de distinção de fontes, problemas do tipo BSS, foi (Zibulevsky; Pearlmutter, 2000). Já (Starck; Donoho; Elad, 2004) apresenta o método (algoritmo) MCA (*Morphological Component Separation*), que usa a ideia de esparsividade no caso de sinais formados por mistura linear instantânea de

sinais oriundos de diferentes fontes, usando dicionários sobrecompletos para representar os mesmos. O método aproveita a representação esparsa dos dados em tais dicionários para separar características baseado em suas morfologias. Ou seja, o método explora esparsividade e diversidade morfológica. Sua extensão multifrequência dá origem ao MMCA (*Multichannel MCA*), que assume que cada fonte de sinal s_j pode ser bem representada, de uma maneira esparsa, em um dicionário Φ_j . Por fim, (Bobin et al., 2007) apresenta o GMCA (*Generalized MCA*), que expande o MMCA ao assumir que s_j pode ser representado não por um dicionário, mas por uma composição de diferentes dicionários (que aqui são assumidos serem bases ortogonais), ou seja, um superdicionário $\mathcal{D} = [\Phi_1^T, \dots, \Phi_D^T]^T$. Portanto, cada fonte s_j tem a representação

$$\begin{aligned} s_j &= \sum_{k=1}^D \phi_{jk} = \sum_{k=1}^D \alpha_{jk} \Phi_k, \\ \mathbb{S} &= \alpha \mathcal{D} \end{aligned} \quad (3.8)$$

Dizemos assim que s_j é modelada como sendo a combinação linear de *D-componentes morfológicas*, ϕ_j , sendo cada componente morfológica esparsa em uma certa base ortogonal. A questão é que a decomposição sobrecompleta leva a uma infinidade de representações esparsas. Por isso, o GMCA impõe a restrição de que o algoritmo deve buscar, dentre as possíveis representações, a mais esparsa. Ou seja, aquela que contém o menor número de coeficientes não nulos. O problema de optimização pode ser escrito como

$$\{\hat{\mathbb{A}}, \hat{\alpha}\} = \underset{\mathbb{A}, \alpha}{\operatorname{argmin}} \left\{ \|\mathbb{X}\mathcal{D}^T - \mathbb{A}\alpha\|_{F, \mathbb{C}_N^{-1}}^2 + 2\lambda\|\alpha\|_{\ell_1} \right\} \quad (3.9)$$

com a norma do primeiro termo sendo a norma de Frobenius em relação a inversa da matriz de covariância do ruído¹². Vide que eu usei como restrição a norma ℓ_1 e não a ℓ_0 , que representaria justamente se o coeficiente é ou não não-nulo. Isso porque o caso de regularização- ℓ_0 é um problema não convexo e combinatório, que demanda muito processamento computacional. Ao mesmo tempo, diversos trabalhos apresentam o caso de regularização- ℓ_1 como equivalente, além de ser um problema convexo.

Sem entrar em detalhes do algoritmo, que será melhor detalhado na tese e pode ser encontrado em (Bobin et al., 2007), o algoritmo estima a matriz mistura no espaço formado por \mathcal{D} . O seu uso comum é através das starlets.

GMCA foi usado em 21-cm para dados da reionização em (Chapman et al., 2013), em que utilizou-se do fato de que fosse o sinal de 21-cm forte o suficiente para ser detectado pelo uso direto do método, esperaria que ele fosse esparsa em certas escalas, dando o tamanho característico das bolhas (bolsões) ionizadas. Como o tamanho dessas bolhas

¹² Seja \mathbb{Z} matriz. A norma de Frobenius podenrada por \mathbb{Q} será definida como sendo $\|\mathbb{Z}\|_{F, \mathbb{Q}} = \operatorname{Tr}(\mathbb{Z}\mathbb{Q}^{-1}\mathbb{Z}^T)$. Logo, o caso em que \mathbb{Q} é a matriz unitária, a norma é o traço da matriz e a denotarei simplesmente por $\|\cdot\|_F$

muda com o redshift então o sinal esparsa a partir destes elementos surgeria como um padrão como uma função do comprimento de onda. Contudo, uma vez que os foregrounds possuem uma forma espectral suave, algumas descrições de esparsidade delas na mesma escala do sinal de 21-cm devem permanecer imutáveis com a frequência. Explorando esse fato o trabalho buscou obter 21-cm e ruídos como resíduos, ou seja, modelar os foregrounds. Foram recuperados os espectros de potência dos foregrounds com alto grau de precisão. Já para dados de 21-cm de IM, ([Carucci; Irfan; Bobin, 2020](#); [Cunnington et al., 2021c](#)) apresentam resultados que apontam não haver diferença dos obtidos pelo GMCA e pelo FastICA. Esse mesmo fato já fora reportado em outro trabalho, ([Bobin et al., 2006](#)), que então sugere o uso de uma análise mais útil para verificar a qualidade da estimativa analisando especificamente as mixing matrix, que então apresentou GMCA melhor do que algoritmos de ICA.

3.2.3 GNILC

No contexto astrofísico/cosmológico, o método ILC (*Internal Linear Combination*) foi o primeiro a ser proposto para remoção de foregrounds dos dados do COBE-DMR ([Basak; Delabrouille, 2013](#)). Isso pois é uma técnica simples e efetiva para extrair os mapas da CMB, enquanto rejeita os contaminantes. Também fora usado nos dados do WMAP. Ambas aplicações se deram no espaço de pixels e usando o filtro GLS, equação [3.6](#), com \mathbb{Q} sendo a matriz de covariância dos dados observacionais, $\mathbb{R}_{\mathbb{X}} = \langle \mathbb{X} \mathbb{X}^T \rangle$. Tanto que é comum encontrar artigos se referindo à equação [3.6](#) com a matriz de covariância observacional como sendo *filtro ILC*. Um pouco depois, estendou-se sua aplicação nos dados do WMAP para o espaço harmônico, e anos depois ao espaço needlets (que são wavelets esféricas com simetria axial) ([Delabrouille et al., 2009](#)), passando a se chamar NILC (*Needlet Internal Linear Combination*). Nesse último, não somente faz a estimativa de \mathbb{W} no espaço harmônico como filtra os coeficientes harmônicos dos mapas por needlets de diferentes escalas. Ou seja, sendo $x_{\ell m}$ o coeficiente do mapa observacional (a projeção do mapa numa casca esférica), e $h_{\ell}^{(j)}$ o filtro que define a needlet numa dada escala j , tem-se

$$\chi^{(j)}(p) = \sum_{\ell, m} h_{\ell}^{(j)} x_{\ell m} Y_{\ell m}(p),$$

com p a posição do céu, e com $\chi^{(j)}(p)$ sendo o mapa de coeficientes needlets dos coeficientes harmônicos do mapa observacional na escala j . Ou seja, a estimativa se faz em diferentes bandas de multipolos e a matriz \mathbb{Q} da equação [3.6](#) se faz em cada escala j . Observemos que sua análise por diferentes canais de frequência pode ser vista ao tomar que o mapa de um canal ν_1 , com $(x_{\nu_1})_{\ell m}$, fornece $\chi_{\nu_1}^{(j)}(p)$.

GNILC é a extensão natural do NILC para uma análise multifrequência ([Remazeilles; Delabrouille; Cardoso, 2011](#)), no contexto da CMB. No caso, com o objetivo de estimar a emissão difusa de uma componente (complexa) originada a partir de multiplas

fontes de emissão correlacionadas, tais como, emissões de foregrounds Galácticos ou radiação CIB (*cosmic infrared background*) e foi utilizado nos dados do Planck Collaboration (Planck Collaboration et al., 2016; Planck Collaboration et al., 2020b; Planck Collaboration et al., 2020). O GNILC foi adaptado ao caso de 21-cm IM recentemente (Olivari; Remazeilles; Dickinson, 2016b), objetivando modelar não os foregrounds ou a CMB, mas os mapas de 21-cm a baixos redshifts.

O objetivo geral para obtenção do filtro ILC, e consequentemente o seu caso em needlets, é estimar a CMB enquanto minimiza o ruído. O problema fora proposto similar ao feito no início dessa seção. No caso, a ideia era a de \mathbb{W} ter resposta unitária ao ser aplicado à mixing matrix. O problema proposto é do tipo de minimizar $\|\mathbb{W}\|_{F,\mathbb{R}_X^{-1}}^2$, sendo a norma a de Frobenius ponderada pela matriz de covariância obsevacional, enquanto impõe a condição $\mathbb{W}\mathbb{A} = \mathbb{A}$. O que pode ser formulado como

$$\{\hat{\mathbb{A}}, \hat{\mathbb{S}}\} = \underset{\mathbb{A}, \mathbb{S}}{\operatorname{argmin}} \left\{ \|\mathbb{W}\|_{F,\mathbb{R}_X^{-1}}^2 + \|\Lambda^T(\mathbb{A} - \mathbb{W}\mathbb{A})\|_F \right\}, \quad (3.10)$$

com Λ sendo uma matriz com parâmetros de regularização. Esse é um tipo de problema de optimização, que em (Remazeilles; Delabrouille; Cardoso, 2011) é resolvido por multiplicadores de Legendre em que Λ é a matriz de multiplicadores. Assumindo a lagrangiana

$$\mathcal{L}(\mathbb{W}, \Lambda) = \operatorname{Tr}(\mathbb{W}\mathbb{R}_X^{-1}\mathbb{W}^T) + \operatorname{Tr}(\Lambda^T(\mathbb{A} - \mathbb{W}\mathbb{A})), \quad (3.11)$$

que ao diferenciarmos em relação à matriz \mathbb{W} ,

$$\begin{aligned} \nabla_{\mathbb{W}} \Big|_{\hat{\mathbb{A}}} \mathcal{L}(\mathbb{W}, \Lambda) &= 0 \\ &= 2\mathbb{W}\mathbb{R}_X - \Lambda\hat{\mathbb{A}}^T, \end{aligned}$$

e pela restrição inicial, fornece a matriz $\Lambda = 2\hat{\mathbb{A}}(\hat{\mathbb{A}}^T\mathbb{R}_X^{-1}\hat{\mathbb{A}})^{-1}$. E portanto, ao filtro, também chamada de matriz ponderada ILC de foregrounds, como

$$\mathbb{W} = \hat{\mathbb{A}} \left(\hat{\mathbb{A}}^T\mathbb{R}_X^{-1}\hat{\mathbb{A}} \right)^{-1} \left(\hat{\mathbb{A}}^T\mathbb{R}_X^{-1} \right). \quad (3.12)$$

Vide que como adiantado acima, o filtro ILC é justamente o caso GLS com $\mathbb{Q} = \mathbb{R}_X^{-1}$. Mas o GNILC não é simplesmente o modelo do filtro \mathbb{W} , mas todo processo de como se estima a mixing matriz e se obtêm o sinal alvo; portanto, é um algoritmo. Na sua adaptação para análises de mapas e 21-cm em IM, em que os dados observacionais são limitadas pelo número de canais do experimento, N_{ch} , e as componentes dos ruídos são correlacionadas sobre as frequências, sendo representadas por uma combinação linear de um número finito N_S de templates independentes, há $N_{ch} - N_S$ graus de liberdade para descrever o sinal de HI através da mixing matrix. Fixando a dimensão da mixing matrix como sendo $N_{ch} \times N_{ch}$, que é assumida como sendo parcialmente correlacionado em canais adjacentes - justamente pelos efeitos de não-gaussianidades introduzidos por diferentes

mecanismo físicos e instrumentais, como já fora dito no início deste capítulo -, através de templates independentes que forma uma submatriz de \mathbb{S} . No caso, o GNILC busca estimar a dimensão das componentes que são altamente correlacionadas, os foregrounds, e faz isso usando o modelo de seleção *Akaike Information Criterion* (AIC) (Bozdogan, 1987; Liddle, 2007) para que assim consiga identificar a submatriz da mixing matrix que representa 21-cm + ruído. Com essa informação ele consegue obter a submatriz da mixing matrix que corresponde ao sinal alvo. Para cada escala needlet j , a submatriz estimada $\hat{\mathbb{A}}_{\text{HI+N}}^{(j)}$, é usada para construir a matriz filtro na dada escala needlet e segunda a matriz de correlação de dois canais ν_1 e ν_2 , $\mathbb{W}_{(\nu_1 \nu_2)}^{(j)}$, de tal forma que a estimativa do sinal de 21-cm + ruído é feito em cada escala j ,

$$\hat{y}_{21\text{cm+N}(\nu)}^{(j)}(p) = \sum_{\nu'} \mathbb{W}_{(\nu, \nu')}^{(j)}(p) x_{\nu'}^{(j)}(p).$$

Após a operação estimar os mapas no espaço needlet, retorna-se ao espaço real.

Os trabalhos (Olivari; Remazeilles; Dickinson, 2016b; Fornazier et al., 2021) mostram que, em condições muito próximas ao que se acredita que o BINGO operará, o GNILC é um método eficiente para a recuperação do sinal de HI, com a estimação dos espectros de potência angular de HI com boa precisão, principalmente ao ser comparado com o método PCA (principal component analysis). Isto é devido ao fato dos graus de liberdade serem fixos em todo domínio de ℓ no PCA e serem ajustados em intervalos de ℓ diferentes no GNILC, o que permite corrigir maiores dependências com as escalas.

3.3 Espectro de Potência Angular dos Mapas

Campos gaussianos condensam todas as suas informações no 1-momento e no 2-momento (estatístico), ou seja, na média e na variância. Todos outros momentos são nulos. Assim, uma forma de avaliar se um campo é ou não gaussiano é avaliando sua medidas estatísticas de alta ordem. Seja $a_{\ell m}^{(X)}$ o coeficiente angular da projeção de um mapa de temperatura X no espaço harmônico. No caso, se o mapa segue uma distribuição gaussiana (com média zero),

$$\langle |a_{\ell m}^{(X)}| \rangle = 0, \quad (3.13)$$

$$\langle |a_{\ell m}^{(X)}|^2 \rangle = \langle a_{\ell m}^{(X)} a_{\ell' m'}^{(X)*} \rangle = \delta_{\ell \ell'} \delta_{mm'} C_{\ell}^{(X)}, \quad (3.14)$$

em que $C_{\ell}^{(X)}$ é o espectro de potência angular do mapa X , e os "brackets" simbolizam o valor esperado da quantidade interna a eles. Os espectros de potência tanto da CMB (Bunn; White, 1997) quanto de 21-cm (Zaldarriaga; Furlanetto; Hernquist, 2004; Battye et al., 2013) dependem dos parâmetros cosmológicos, ou seja, da cosmologia. Isso faz com que o espectro de potência angular seja muito usado para análises de restrição de parâmetros cosmológico (Costa et al., 2021; Planck Collaboration et al., 2020a).

Aqui estarei interessado em conseguir estimar os espectros de potência dos mapas de 21-cm o mais preciso possível. Isso porque são esses que serão usados para identificação (estatística) das BAOs e para restrição de modelos de Energia Escura.

3.4 Noise Debiais Process

Como foi visto na seção acima, sobre os métodos de separação de componentes, todos os métodos aqui utilizados são do tipo *blind* em que estima os mapas de foregrounds, devido principalmente a sua alta correlação espectral e pelas suas características morfológicas. Assim, os métodos fornecem 21-cm como um termo, junto ao ruído térmico, residual. Se usarmos \mathbb{W} , e, consequentemente, a matriz \mathbb{W}_{FG} , da equação 3.4, construído a partir de algum método de Separação de Componentes dos três descritos acima, a estimativa dos foregrounds, através dos mapas observacionais em \mathbb{X} , se dá como¹³

$$\hat{\mathbb{X}}_{\text{FG}} \doteq \hat{\mathbb{Y}} = \mathbb{W}_{\text{FG}} \mathbb{X}.$$

Logo,

$$\hat{\mathbb{X}}_{\text{21cm+N}} = \mathbb{X} - \mathbb{W}_{\text{FG}} \mathbb{X}.$$

É possível estimar as informações de 21-cm, como explicado na subseção anterior, estimando diretamente seu espectro de potência angular. Aqui eu usei o processo *Noise Debiais*, inspirado pelo modelo utilizado em (Fornazier et al., 2021). Para isso, é necessário a estimativa de diversos espectros de potências, provenientes de diferentes realizações (simulações). A ideia é simular M processos de Separação de Componentes, usando um dado método de Separação de Componentes, provenientes de realizações diferentes, para 21-cm e para o White Noise. Seja (cs) , símbolo para representar que a estimativa foi feita por um certo método, e $\{L_i\}_{i=1}^M$, conjunto de índices que identifica a origem do espectro de potência angular. Temos, por exemplo, a seguinte representação a ser utilizada: $C_\ell^{(\text{cs}), L_k}$, significando que é o espectro de potência angular de um dado mapa (canal), gerado por uma realização identificada como L_k (k -ésima realização), estimada pelo método (cs) e sendo avaliada no multipolo ℓ . De maneira similar se dará para mapas que não foram estimados por nenhum método, mas são simplesmente obtidos seu espectro de potência, somente trocando (cs) pela respectiva identificação. Assim sendo, o espectro de potência angular, num dado multipolo ℓ , estimado por um método (cs) , é dado por

$$\hat{C}_\ell^{(\text{cs}), L_0} = \frac{C_\ell^{(\text{cs}), L_0}}{S} - \langle C_\ell^{(\text{N}), L_i} \rangle, \quad (3.15)$$

em que o valor é analisado tomando uma realização como referência, como se fosse o "nosso" universo, que aqui eu identifico como L_0 . Os valores esperados são assumidos

¹³ A partir de agora, procurarei codificar as matrizes com mapas de uma subclasse pela matriz \mathbb{X} com o *subscript* que identifica a subclasse. Isso simplificará a nomeclatura.

como médias sobre as realizações sem usar a realização assumida como referência. S é definida como

$$S = \left\langle \frac{C_{\ell}^{(\text{cs}), L_i}}{C_{\ell}^{(\text{prior}), L_i}} \right\rangle, \quad (3.16)$$

com *prior* sendo o nome dado para valores provenientes de mapas (não estimados) de 21-cm + white noise, e novamente, o valor esperado sendo a média na quantidade sobre as realizações, exceto a de referência. O termo S serve como um termo de supressão que busca corrigir a amplitude da estimação devido aos resultados imperfeitos dos métodos de Separação de Componentes. O termo para o ruído busca, por sua vez, estimar a partir de diferentes realizações que não seria a do Universo, a influência do ruído.

Parte III

Resultados Preliminares

3.5 Resultados preliminares

Nos capítulos e seções anteriores busquei, com o mínimo de fórmulas quanto fosse possível, fornecer o entendimento necessário para a compreensão deste capítulo. Este capítulo contém parte dos resultados obtidos até o momento e que está sendo preparado para ser apresentado de forma completa em um artigo que, espero, seja publicado em breve. A ideia do trabalho foi, dentro de um cenário que se espera ver o BINGO Telescope operando, avaliar três métodos de Separação de Componentes, do tipo *blind*, para conseguir estimar os mapas, e consequentemente as informações, das emissões de 21-cm emitidas de uma época pós-reionização no Universo entre redshifts 0.13 e 0.48, que enquadram o período de transição para o domínio da Energia Escura na dinâmica do Universo. Como vimos, a dificuldade em obter os sinais de 21-cm se dá pela instrumentação não ser capaz de diferenciar 21-cm de outras emissões que correspondem a mesma faixa de frequência observada, que para o dado intervalo de redshift corresponde à 980-1260 MHz.

A análise que fiz baseou-se em um conjunto de dados que foi gerado quando a faixa de frequência ainda era considerada entre 960-1260 MHz, com 30 canais. Modelei a observação das emissões (provenientes) do céu e a fiz como somas de diferentes mecanismos de emissão Galácticas e extragaláticas. Isso correspondeu, nas simulações aqui utilizadas, a somar os diferentes mapas, para cada banda de frequência. O BINGO operará numa faixa do céu que comprehende à declinação entre, aproximadamente, -22.5° e -7.5° , em coordenadas celestiais, com resolução angular de 40 arcmin. Assim, cada mapa do céu (foregrounds + 21-cm), de cada banda de frequência, foi convoluto com um feixe gaussiano de FWHM de 40 arcmin. Em todos os mapas (foregrounds, 21-cm e white noise) são usados com pixelização HEALPix (Górski et al., 2005) correspondente a um $N_{\text{side}}=256$, e possuem as regiões celestiais que não serão cobertas mascaradas com um máscara binária que simplesmente anula emissões fora da região de interesse. Também foram mascarados os 10% pixels com as maiores emissões. O ruido térmico assumido foi do tipo white noise correspondente à uma configuração com temperatura de sistema de 70 K, 28 cornetas coletoras e 1 mês de observação. O mapa de ruído foi gerado a $N_{\text{side}}=512$ e feito um *downgrade* para 256.

Exceto o GNILC, os métodos ICA e GMCA não estimam a dimensão que corresponde a influência dos foregrounds nos mapas observados. Dito de outra forma, esses métodos não estimam o que tenho chamado de número de fontes, N_S . Nesses casos o que fiz foi analisar diferentes valores. O GNILC, como anteriormente dito, faz essa estimativa pelo método AIC (Bozdogan, 1987). Os resultados aqui apresentados utilizam análises com o ICA no espaço de pixels, e com o GMCA no espaço wavelet esférico através de starlets (Starck; Murtagh; Bertero, 2011). O GNILC faz naturalmente a sua análise no espaço needlet. Também busquei analisar a influência do número de realizações assumidas na estimativa dos espectros de potência angular de 21-cm na qualidade do resultado. Essa,

por ora, só fora feita para o caso ICA.

Para estimar as variâncias dos resultados usei os método de *resampling Jackknife* e *Bootstrap* (Shao; Tu, 2012; Konishi, 2014), que por apresentarem resultados muito similares, vou apenas manter as análises com Jackknife que é muito mais rápido no quesito computacional. A idéia por trás desse método de estimação é que dado um conjunto de dados, todas as informações da distribuição que gerou os dados estão contidas neles. Então, o método gera novos conjuntos de dados a partir do conjunto original e estima a distribuição do novo conjunto para, por fim, estimar medidas estatísticas relacionadas a distribuição. No caso, estimei as variâncias.

Na Figura 10 estão os resultados (em azul) da estimação por FastICA do mapa de 21-cm para 400 simulações no processo de Noise Debiais com a variância de C_ℓ , em relação a cada multipolo, feita com o método de resampling Jackknife entre as diferentes realizações. Os resultados apresentados estão entre multipolos 3 e 200. O fato das barras verticais na parte inferior do gráfico parecerem muito grandes se deve ao uso da escala logarítmica no eixo vertical. O gráfico na primeira linha representa o resultado para o primeiro canal de frequência, a segunda linha com resultados para o décimo canal, a terceira linha para o vigésimo e o última para o último canal. As análises desses gráficos foram feitas assumindo que o número de fontes é igual a $N_S = 3$. A recuperação do primeiro canal tem uma região considerável entre 20-60 em que os resultados não são compatíveis com os valores fiduciais representados pela linha vermelha, representando os C_ℓ do mapa de 21-cm puro. Já para os demais canais a recuperação se apresenta compatível. Caso similar ocorre para as reconstruções pelo GMCA apresentadas na Figura 11, também para 400 simulações e a dimensão do número de fontes sendo 3, com os resultados no primeiro canal com uma ampla faixa sendo incompatível com o valor fiducial e os demais canais com aparente recuperação compativel. No caso do GNILC para 100 realizações, apresentados na Figura 12, o primeiro canal tem um recuperação incompatível em um intervalo de multipolos menores, o que corresponde a maiores escalas.

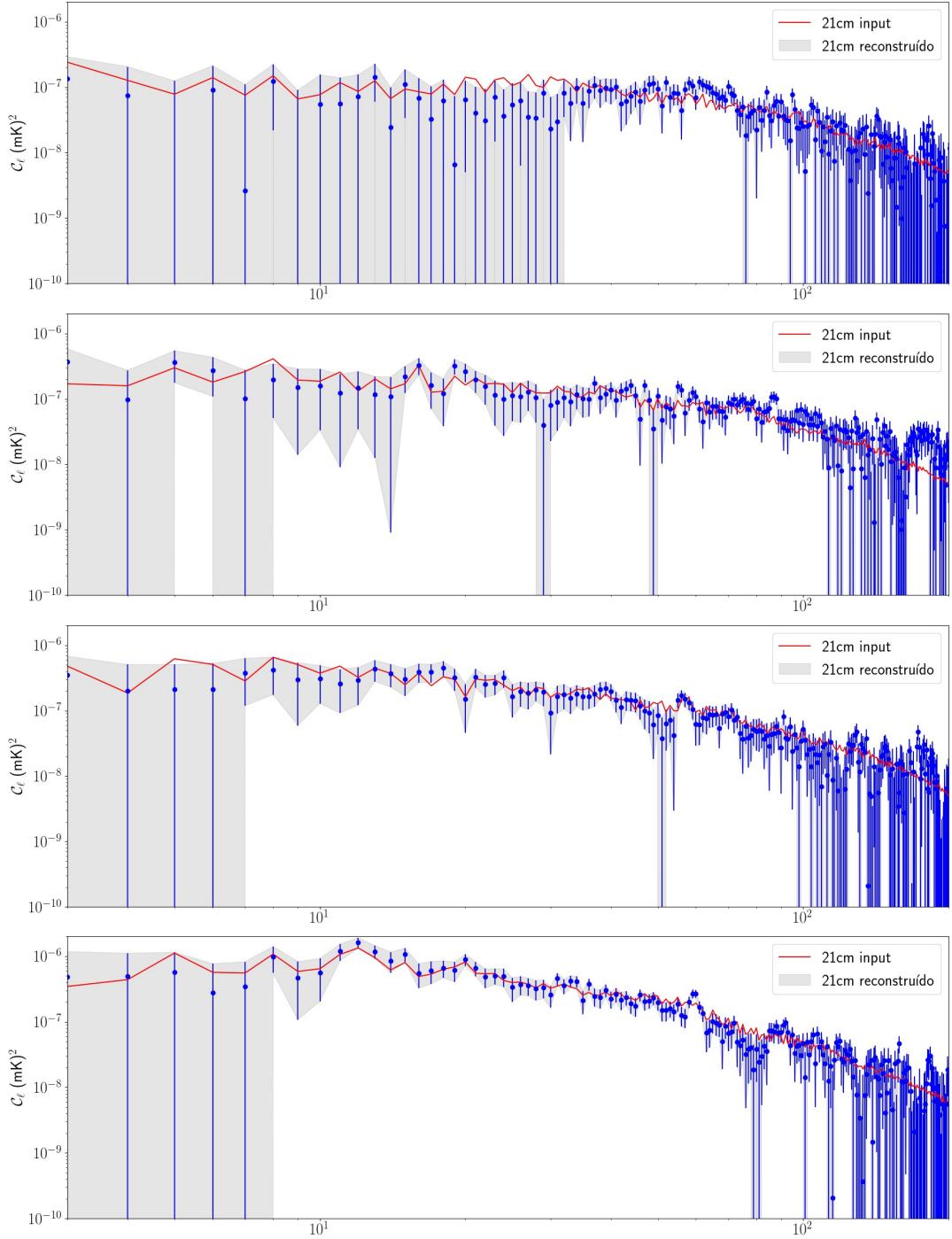


Figura 10 – Todos os quatro gráficos são referentes ao caso ICA com a curva em vermelho sendo o valor do espectro de potência do mapa de 21-cm puro, sem contaminações. Em azul estão os valores reconstruídos para cada multipolo de 3 à 200, com as variâncias estimadas pelo método Jacknife. No primeiro gráfico está o resultado para o primeiro canal; no segundo, para o décimo; para o terceiro, para o vigésimo canal; e, por fim, o resultado para o último canal.

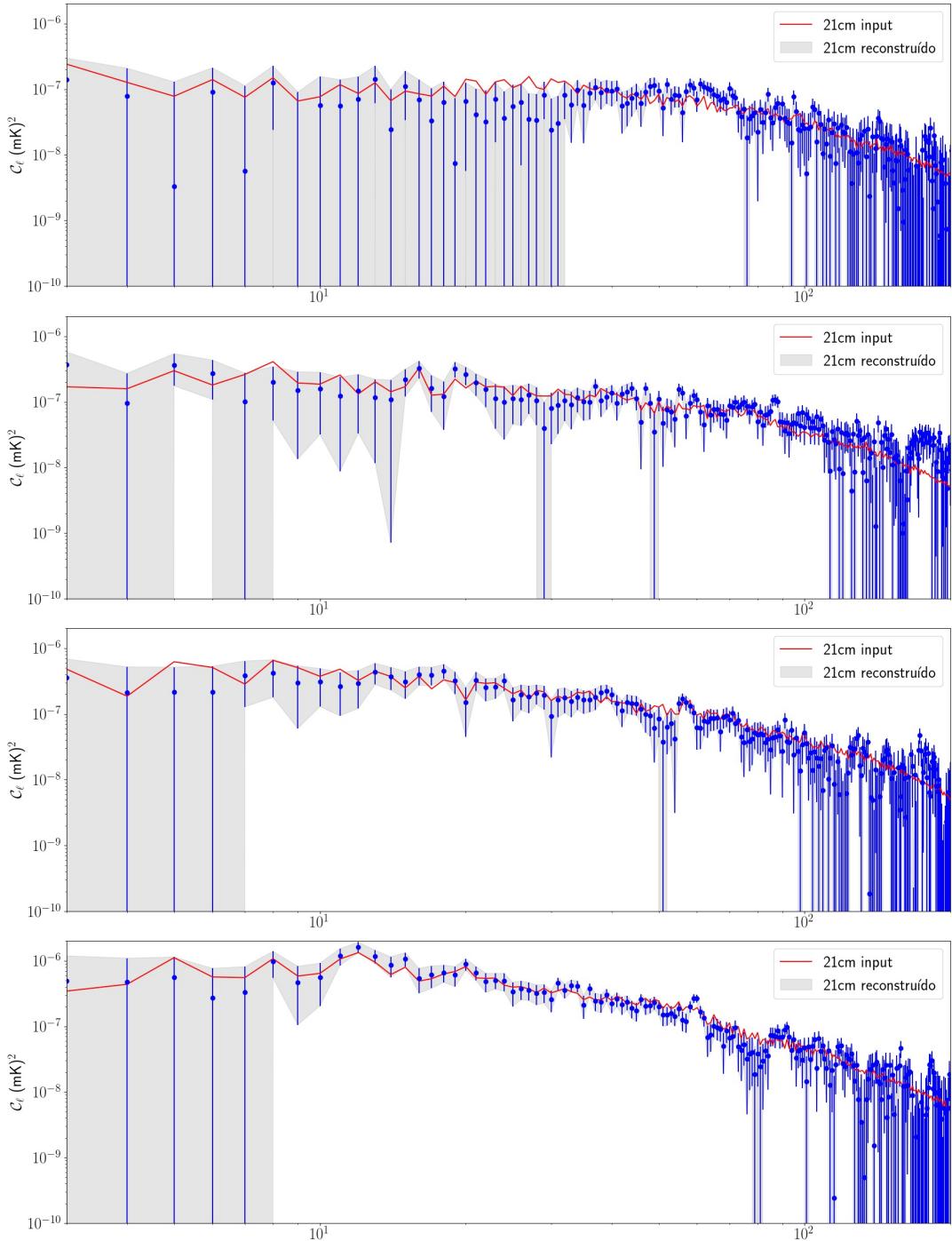


Figura 11 – Todos os quatro gráficos são referentes ao caso GMCA com a curva em vermelho sendo o valor do espectro de potência do mapa de 21-cm puro, sem contaminações. Em azul estão os valores restruídos para cada multipolo de 3 à 200, com as variâncias estimadas pelo método Jackknife. No primeiro gráfico está o resultado para o primeiro canal; no segundo, para o décimo; para o terceiro, para o vigésimo canal; e, por fim, o resultado para o último canal.

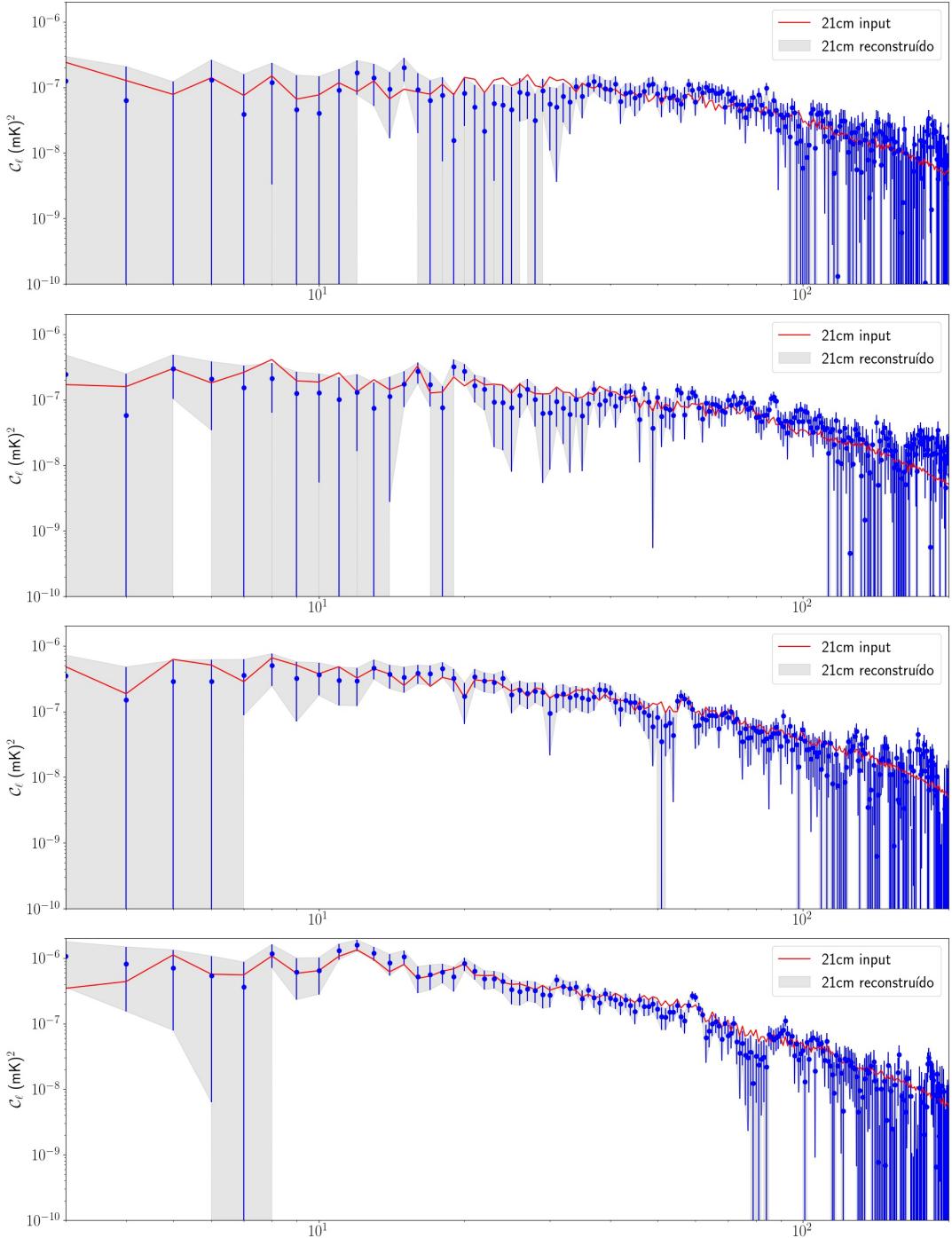


Figura 12 – Todos os quatro gráficos são referentes ao caso GNILC com a curva em vermelho sendo o valor do espectro de potência do mapa de 21-cm puro, sem contaminações. Em azul estão os valores restruídos para cada multipolo de 3 à 200, com as variâncias estimadas pelo método Jackknife. No primeiro gráfico está o resultado para o primeiro canal; no segundo, para o décimo; para o terceiro, para o vigésimo canal; e, por fim, o resultado para o último canal.

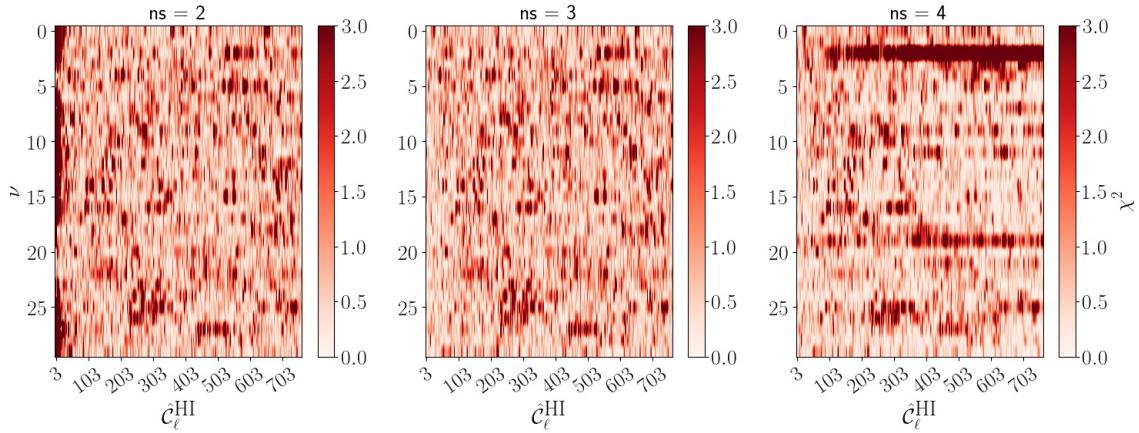


Figura 13 – Resultado dos χ^2 obtidos entre os espectros de potência angular estimados de 21-cm e os espectros fiduciais. O χ^2 contabilizado sobre 400 realizações diferentes. No eixo vertical estão representados os canais, que aqui constam como dispostos entre 0 à 29, e no eixo horizontal os multipolos. Os resultados são para as reconstruções do ICA com número de fontes (ns) 2, 3 e 4. Os valores estão restritos entre 0 e 3, para melhor identificação de regiões com estimação ruim.

Um passo necessário é quantificar a qualidade das estimativas, dado um método numa dada configuração. Para avaliar isso, até o momento tenho usado o χ^2 . Primeiro, é importante avaliar a dimensão do espaço dos foregrounds - que tenho chamado de número de fontes- para os métodos FastICA e GMCA, e então avaliei para os valores 2, 3 e 4, como apresentados na Figura 13, com eixo vertical identificando o respectivo canal ν e o eixo horizontal o multipolo ℓ , representa o χ^2 dos $C_{\ell}^{(\nu)}$ entre o respectivo valor estimado e o valor fiducial, somando sobre todas as realizações. Limitei ao intervalo de apresentação entre 0 e 3 para verificar a saturação em registos específicos e para melhorar comparação. Os multipolos foram usados estão entre 3 e 768. O resultado para o caso com $N_S = 4$ apresenta uma forte região de saturação, altos valores de χ^2 , à baixas frequências, e outra região por volta da banda 19 com forte saturação para pequenas escalas (altos multipolos). Já o caso para $N_S = 2$ apresenta forte saturação em altas escalas (baixos multipolos) em quase todos canais. As características de saturação do caso $N_S = 3$ são compartilhadas entre as outras, com forte saturação ao redor da banda 25 para multipolos entre 200-300, para 27 entre multipolos 400-500 e para as bandas 2 e 5 para multipolos entre 450-600. Caso similar ocorre para os resultados do GMCA, apresentado em 14. Também analisei um χ^2 efetivo, χ_{eff}^2 , como sendo a média dos χ^2 por canal de frequência. Os resultados estão apresentados em Figura 15. O gráfico superior é o resultado para o FastICA e o inferior para o GMCA. É possível ver que há muita similaridade entre os resultados, com diferenças estatísticas visíveis para $N_S = 2$, mas muito parecidos para valores maiores. Ambos apresentam melhor resultado para $N_S = 3$, com valores de χ_{eff}^2 entre, aproximadamente, 0.75 e 1.35.

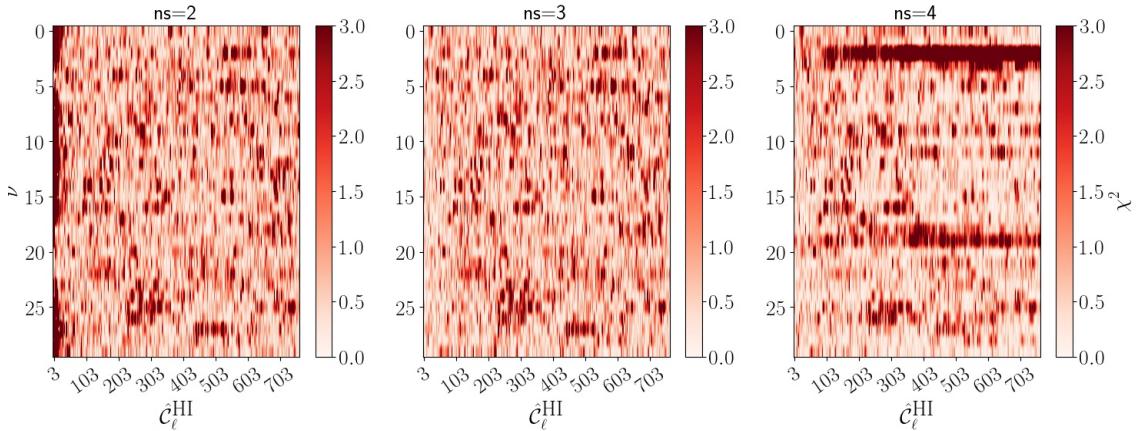


Figura 14 – Resultado dos χ^2 obtidos entre os espectros de potência angular estimados de 21-cm e os espectros fiduciais. O χ^2 contabilizado sobre 400 realizações diferentes. No eixo vertical estão representados os canais, que aqui constam como dispostos entre 0 à 29, e no eixo horizontal os multipolos. Os resultados são para as reconstruções do GMCA com número de fontes (ns) 2, 3 e 4. Os valores estão restritos entre 0 e 3.

Então, baseado nesse último resultado, considerei o valor de $N_S = 3$ para fins de comparação com o GNILC. O resultado da comparação do χ_{eff}^2 entre os três métodos está na Figura 16, com os dois primeiros resultados, GMCA e (Fast)ICA sendo os já apresentados anteriormente mas incluídos para fins de comparação. E o resultado para o GNILC. Os padrões de saturação são similares entre os três métodos com sutis diferenças como pode ser observado, por exemplo, na região entre o canal 7 e o canal 20, entre os multipolos 600-700, que é um padrão fraco nos dois primeiros métodos mas que é mais forte no GNILC. O GNILC também apresenta um maior padrão de saturação no primeiro canal.

Na Figura 17 apresento o resultado, até o momento, da comparação entre os três métodos. O padrão de similaridade nos resultados estatísticos entre (Fast)ICA e GMCA que foram elencados anteriormente se apresentam nessa imagem. Ambos apresentam o resultado ligeiramente mais próximo do valor de 1 para canais maiores, enquanto o GNILC para canais mais baixos, mas todos com resultados entre 0.7 e 1.4.

O fato de FastICA e GMCA, aparentemente, gerarem resultados equivalentes em 21-cm fora apontado em (Cunnington et al., 2021a), e para tratamentos de imagens em (Bobin et al., 2007). Neste último, é apresentado uma forma alternativa para avaliar a qualidade da reconstrução que se dá avaliando diretamente a mixing matrix. No mesmo trabalho, ao se utilizar essa medida GMCA se apresenta como melhor do que os demais algoritmos baseados em ICA. Uma outra possibilidade de avaliação é utilizar o índice de Gini para mensurar a esparsividade dos mapas e dos resultados, baseado na proposição feita em (Oppizzi et al., 2020), o que poderá dar uma direção do motivo dos resultados serem estatisticamente equivalentes.

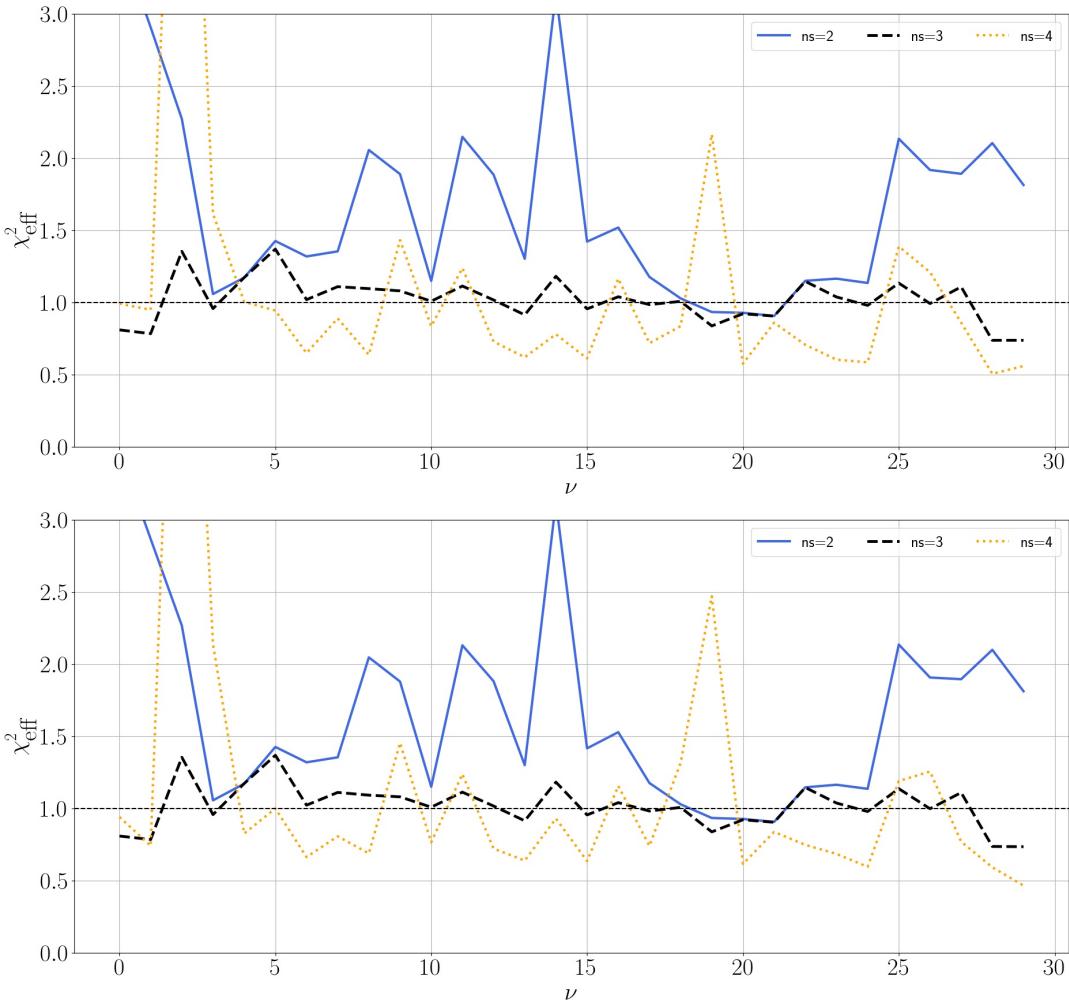


Figura 15 – Resultados dos χ^2_{eff} por banda para a estimação por ICA (superior) e por GMCA (inferior). Ambos os casos constam as estimações para o número de fontes de 2 (azul), 3 (preto) e 4 (amarelo).

Como uma análise complementar analisei a influência no número de simulações nas estimações. Por ora, só obtive os resultados para o caso ICA. Os resultados estão apresentados na Figura 18. Todos os gráficos correspondem a comparações entre 25, 50, 100 e 400 realizações. O primeiro gráfico é o caso para número de dimensões da matriz mistura para estimar as fontes sendo igual a 3 e o segundo para o caso de 4. Os resultados para o número de 4 (gráfico inferior), aparenta não haver forte dependencia com relação ao número de realizações. O valor com 25 realizações (azul) tem um valor de χ^2_{eff} mais distante do valor referência de 1 do que os demais em canais maiores do que 14. Resultado para 100 e para 400 são estatisticamente equivalentes, apresentando não haver grande diferença com o uso do número de realizações para mais do que 100. Já para o caso do número ser 3 (gráfico superior), a menor qualidade na estimativa com 25 realizações fica mais evidente, principalmente para canais 9 e 12, e para os canais 23 e 24¹⁴. Mas valores maiores do que

¹⁴ Observe que os gráficos estão numerados de 0 à 29. Assim, o valor de 0 corresponde ao 1 canal, o 10 ao canal 11, e assim por diante.

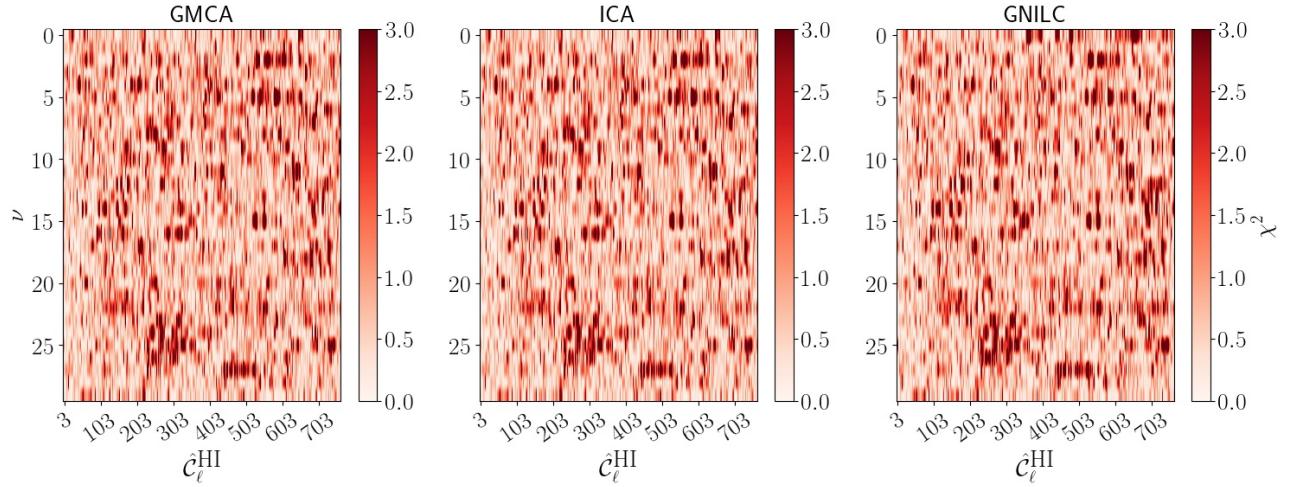


Figura 16 – Resultado dos χ^2 obtidos entre os espectros de potência angular estimados de 21-cm e os espectros fiduciais. Os resultados são para as reconstruções dos métodos: GMCA (à esquerda), ICA (central) e GNILC (à direita). Os valores estão restritos entre 0 e 3, para melhor identificação de regiões com estimativa ruim.

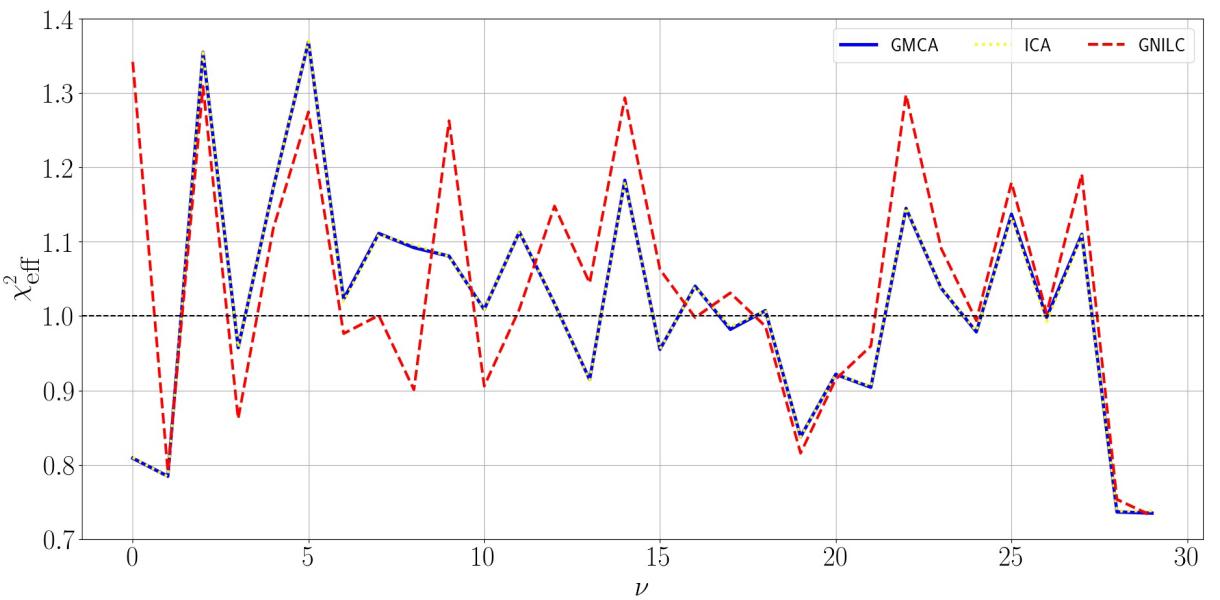


Figura 17 – Resultado dos χ_{eff}^2 por canal de frequência para os três métodos: GMCA (azul), ICA(amarelo) e GNILC (vermelho).

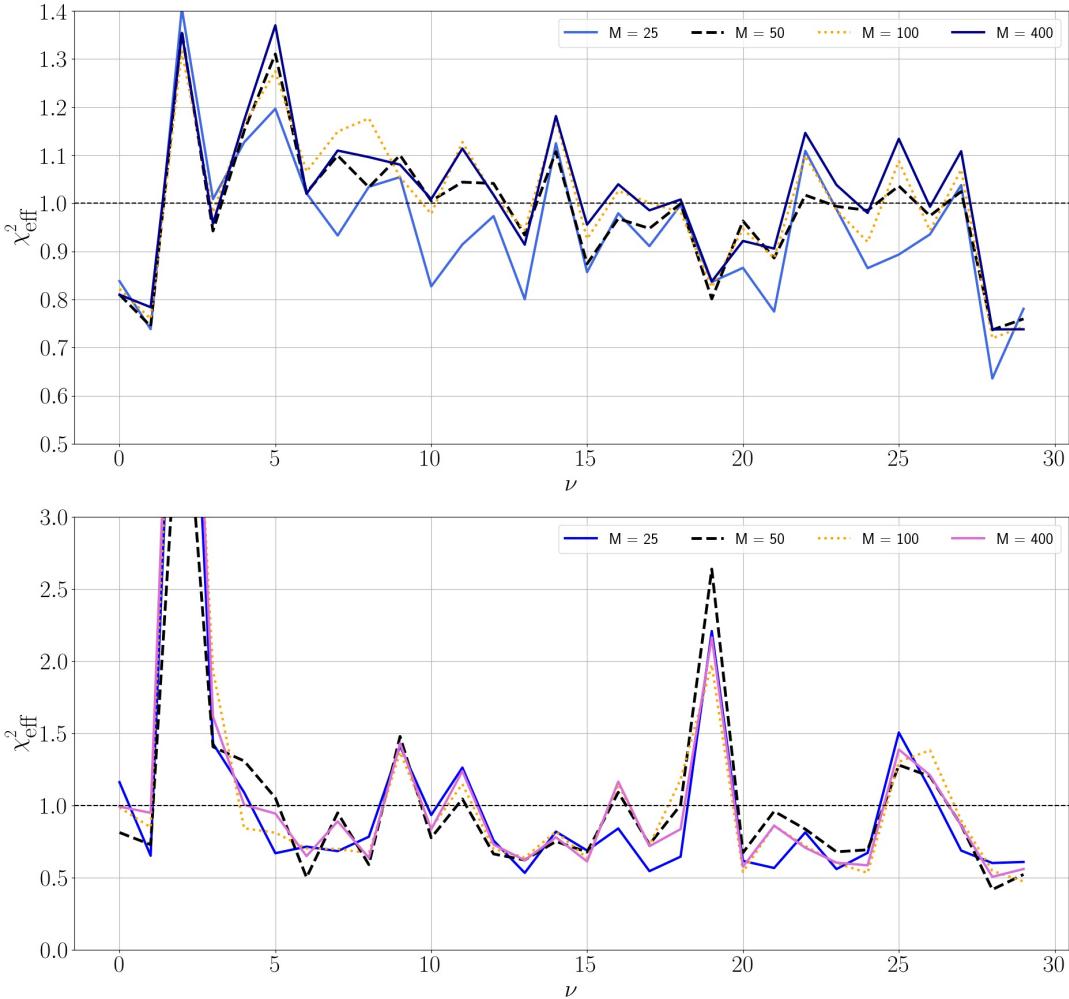


Figura 18 – Gráficos dos χ^2_{eff} para resultados usando FastICA com a dimensão de foregrounds na mistura de 3 (superior) e 4 (inferior). Os resultados são apresentados para 25, 50, 100 e 400 realizações.

100 realizações, novamente, parecem não resultar em melhorias expressivas nos resultados.

3.6 Conclusões

Até o momento, foram feitas análises usando três métodos de separação de componentes do tipo blind: GMCA, FastICA e GNILC. Para os dois primeiros, foram feitas análises da dimensão dos foregrounds na mixing matrix na qualidade da reconstrução. O último realiza essa estimativa segundo o método AIC. Como resultado, a estimativa sugere que os foregrounds são bem representados por dimensão igual a 3. Posteriormente, usou-se esse resultado para comparar os três métodos. Os resultados para GMCA e ICA foram obtidos com número de dimensões da mixing matrix igual a 3 e 400 realizações, o GNILC para 100 realizações. Os resultados preliminares apresentam todos os métodos com χ^2_{eff} entre 0.75 e 1.4 para todos canais de frequência. Em especial, apresentam equivalência,

nessa configuração, entre GMCA e FastICA. Esse fato, que já fora apontado por outros trabalhos, pode ser melhor explorado e analisado por outras medidas estatísticas para melhor entendê-los. Por fim, fora feito uma análise na influência no número de realizações utilizadas para a estimação usando o FastICA. Os resultados apresentam que não há ganho significativo para número de realizações superior à 100, o que valida a análise entre os três métodos anteriores com o GNILC usando 100 realizações e os outros dois com 400 realizações.

Parte IV

Observações e Próximos Passos

3.7 Próximos passos

Para os próximos passos objetivo primeiro investigar o motivo dos resultados do FastICA e GMCA serem estatisticamente equivalentes, incluindo quantificando o grau de esparsividade dos mapas. Segundo, avaliar a influência do número de realizações sobre a qualidade das estimativas para o GNILC e o GMCA. Por fim, como proposto no projeto inicial, incluir análise com um método não blind.

3.8 Observações

Cabe observar que, como já reportado nos relatórios anuais de atividades para USP, o trabalho em Separação de Componentes não tem sido o único que venho realizando. Além desse, venho auxiliando na avaliação, estudos e testes do plano focal, com parte desses trabalho tendo levado ao artigo ([Abdalla et al., 2021](#)), em que conduzi toda a análise de optimização e avaliação de modelos para disposição das cornetas. Inclusive auxiliando no desenho da disposição das mesmas para o modelo de 28 cornetas. Também, ainda na parte óptica, venho desenvolvendo análises e modelos para construção dos espelhos do BINGO explorando simetrias dos mesmos e através de placas acopladas de $\sim 2\text{m}^2$.

Para além do BINGO, mas com futura intersecção com o mesmo em análise cruzada, trabalho com pesquisas em dados ópticos de diferentes levantamentos fotométricos e espectroscópicos para seleção de galáxias e quasares e inferência de seus redshifts utilizando modernas técnicas de aprendizado de máquina.

Referências

- Abdalla, E. et al. The BINGO Project I: Baryon Acoustic Oscillations from Integrated Neutral Gas Observations. *arXiv e-prints*, p. arXiv:2107.01633, jul. 2021. Citado na página 16.
- Abdalla, E.; Marins, A. The dark sector cosmology. *International Journal of Modern Physics D*, v. 29, n. 14, p. 2030014, jan. 2020. Citado na página 16.
- Abdalla, F. B. et al. The BINGO Project III: Optical design and optimisation of the focal plane. *arXiv e-prints*, p. arXiv:2107.01635, jul. 2021. Citado 2 vezes nas páginas 17 e 59.
- Aghanim, N. et al. Planck 2018 results. vi. cosmological parameters. EDP Sciences, 2020. Citado 3 vezes nas páginas 16, 25 e 26.
- Basak, S.; Delabrouille, J. A needlet ILC analysis of WMAP 9-year polarization data: CMB polarization power spectra. , v. 435, n. 1, p. 18–29, out. 2013. Citado na página 38.
- Bassett, B.; Hlozek, R. Baryon acoustic oscillations. *Dark energy: observational and theoretical approaches*, Cambridge University Press: Cambridge, UK, p. 246, 2010. Citado na página 16.
- Battye, R. A. et al. H I intensity mapping: a single dish approach. , v. 434, n. 2, p. 1239–1256, set. 2013. Citado 3 vezes nas páginas 18, 19 e 40.
- Bennett, C. L. et al. Scientific results from COBE. *Advances in Space Research*, v. 13, n. 12, p. 409–423, dez. 1993. Citado na página 15.
- Bharadwaj, S.; Ali, S. S. On using visibility correlations to probe the HI distribution from the dark ages to the present epoch - I. Formalism and the expected signal. , v. 356, n. 4, p. 1519–1528, fev. 2005. Citado na página 18.
- Bharadwaj, S.; Nath, B. B.; Sethi, S. K. Using HI to Probe Large Scale Structures at $z \sim 3$. *Journal of Astrophysics and Astronomy*, v. 22, p. 21, mar. 2001. Citado na página 18.
- Bird, S. et al. Damped lyman α absorbers as a probe of stellar feedback. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press, v. 445, n. 3, p. 2313–2324, 2014. Citado na página 11.
- Bobin, J.; Acero, F.; Picquenot, A. Metric learning for semi-supervised sparse source separation with spectral examples. In: IEEE. *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. [S.l.], 2019. p. 450–454. Citado na página 31.
- Bobin, J. et al. SZ and CMB reconstruction using Generalized Morphological Component Analysis. In: *Astronomical Data Analysis ADA'06*. Marseille, France: [s.n.], 2006. Disponível em: <<https://hal.archives-ouvertes.fr/hal-01160933>>. Citado na página 38.

Bobin, J. et al. Sparsity and Morphological Diversity in Blind Source Separation. *IEEE Transactions on Image Processing*, v. 16, n. 11, p. 2662–2674, nov. 2007. Citado 2 vezes nas páginas [37](#) e [51](#).

Bozdogan, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, Springer, v. 52, n. 3, p. 345–370, 1987. Citado 2 vezes nas páginas [40](#) e [45](#).

Bunn, E. F.; White, M. The 4 Year COBE Normalization and Large-Scale Structure. , v. 480, n. 1, p. 6–21, maio 1997. Citado na página [40](#).

Carrón Duque, J. et al. Point source detection and false discovery rate control on CMB maps. *Astronomy and Computing*, v. 28, p. 100310, jul. 2019. Citado na página [30](#).

Carucci, I. P.; Irfan, M. O.; Bobin, J. Recovery of 21-cm intensity maps with sparse component separation. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press, v. 499, n. 1, p. 304–319, 2020. Citado 2 vezes nas páginas [12](#) e [38](#).

Chapman, E. et al. The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis. , v. 429, n. 1, p. 165–176, fev. 2013. Citado na página [37](#).

Chapman, E. et al. Foreground removal using FASTICA: a showcase of LOFAR-EoR. , v. 423, n. 3, p. 2518–2532, jul. 2012. Citado na página [35](#).

Chapman, E.; Jelić, V. Foregrounds and their mitigation. *arXiv preprint arXiv:1909.12369*, 2019. Citado 2 vezes nas páginas [21](#) e [25](#).

Chen, S.; Donoho, D. Basis pursuit. In: IEEE. *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*. [S.l.], 1994. v. 1, p. 41–44. Citado na página [35](#).

Chen, S. S.; Donoho, D. L.; Saunders, M. A. Atomic Decomposition by Basis Pursuit. *SIAM Review*, v. 43, n. 1, p. 129–159, jan. 2001. Citado na página [35](#).

Costa, A. A. et al. The BINGO Project VII: Cosmological Forecasts from 21cm Intensity Mapping. *arXiv e-prints*, p. arXiv:2107.01639, jul. 2021. Citado 2 vezes nas páginas [19](#) e [40](#).

Cunnington, S. et al. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press, v. 504, n. 1, p. 208–227, 2021. Citado 2 vezes nas páginas [12](#) e [51](#).

Cunnington, S. et al. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. , v. 504, n. 1, p. 208–227, jun. 2021. Citado na página [31](#).

Cunnington, S. et al. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. , v. 504, n. 1, p. 208–227, jun. 2021. Citado na página [38](#).

de Oliveira-Costa, A. et al. Cross-Correlation of Tenerife Data with Galactic Templates—Evidence for Spinning Dust? , v. 527, n. 1, p. L9–L12, dez. 1999. Citado na página [24](#).

- Delabrouille, J. et al. The pre-launch planck sky model: a model of sky emission at submillimetre to centimetre wavelengths. *Astronomy & Astrophysics*, EDP Sciences, v. 553, p. A96, 2013. Citado 2 vezes nas páginas [20](#) e [25](#).
- Delabrouille, J. et al. A full sky, low foreground, high resolution CMB map from WMAP. , v. 493, n. 3, p. 835–857, jan. 2009. Citado na página [38](#).
- Dickinson, C. et al. The State-of-Play of Anomalous Microwave Emission (AME) research. , v. 80, p. 1–28, fev. 2018. Citado na página [24](#).
- Dickinson, C.; Davies, R.; Davis, R. Towards a free-free template for cmb foregrounds. *Monthly Notices of the Royal Astronomical Society*, Blackwell Science Ltd Oxford, UK, v. 341, n. 2, p. 369–384, 2003. Citado 2 vezes nas páginas [11](#) e [22](#).
- Dragone, C. Offset multireflector antennas with perfect pattern symmetry and polarization discrimination. *AT T Technical Journal*, v. 57, p. 2663–2684, set. 1978. Citado na página [17](#).
- Draine, B. T.; Lazarian, A. Diffuse Galactic Emission from Spinning Dust Grains. , v. 494, n. 1, p. L19–L22, fev. 1998. Citado na página [24](#).
- Eriksen, H. K. et al. Joint Bayesian Component Separation and CMB Power Spectrum Estimation. , v. 676, n. 1, p. 10–32, mar. 2008. Citado na página [31](#).
- Field, G. B. Excitation of the hydrogen 21-cm line. *Proceedings of the IRE*, IEEE, v. 46, n. 1, p. 240–250, 1958. Citado na página [18](#).
- Fornazier, K. S. F. et al. *The BINGO Project V: Further steps in Component Separation and Bispectrum Analysis*. 2021. Citado 6 vezes nas páginas [12](#), [24](#), [25](#), [30](#), [40](#) e [41](#).
- Furlanetto, S. R.; Oh, S. P.; Briggs, F. H. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. , v. 433, n. 4-6, p. 181–301, out. 2006. Citado na página [18](#).
- Górski, K. M. et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. , v. 622, n. 2, p. 759–771, abr. 2005. Citado 2 vezes nas páginas [26](#) e [45](#).
- Gutiérrez, C. M. et al. The Tenerife Cosmic Microwave Background Maps: Observations and First Analysis. , v. 529, n. 1, p. 47–55, jan. 2000. Citado na página [24](#).
- Haslam, C. G. T. et al. A 408 MHz all-sky continuum survey. I - Observations at southern declinations and for the North Polar region. , v. 100, p. 209–219, jul. 1981. Citado na página [24](#).
- Ho, M.-F.; Bird, S.; Garnett, R. Damped lyman-alpha absorbers from sloan digital sky survey dr16q with gaussian processes. *arXiv preprint arXiv:2103.10964*, 2021. Citado na página [11](#).
- Hyvärinen, A. Survey on independent component analysis. Citeseer, 1999. Citado na página [35](#).
- Kanekar, N.; Chengalur, J. N. A deep search for 21-cm absorption in high redshift damped lyman- α systems. *Astronomy & Astrophysics*, EDP Sciences, v. 399, n. 3, p. 857–868, 2003. Citado 2 vezes nas páginas [11](#) e [19](#).

Kogut, A. et al. Microwave Emission at High Galactic Latitudes in the Four-Year DMR Sky Maps. , v. 464, p. L5, jun. 1996. Citado na página [23](#).

Kogut, A. et al. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Foreground Polarization. , v. 665, n. 1, p. 355–362, ago. 2007. Citado na página [22](#).

Konishi, S. *Introduction to multivariate analysis: Linear and nonlinear modeling.* [S.l.]: CRC Press, 2014. Citado na página [46](#).

Leitch, E. M. et al. An Anomalous Component of Galactic Emission. , v. 486, n. 1, p. L23–L26, set. 1997. Citado na página [23](#).

Liddle, A. R. Information criteria for astrophysical model selection. , v. 377, n. 1, p. L74–L78, maio 2007. Citado na página [40](#).

Liu, A.; Shaw, J. R. Data analysis for precision 21 cm cosmology. *Publications of the Astronomical Society of the Pacific*, IOP Publishing, v. 132, n. 1012, p. 062001, 2020. Citado 2 vezes nas páginas [11](#) e [16](#).

Maino, D. et al. All-sky astrophysical component separation with Fast Independent Component Analysis (FASTICA). , v. 334, n. 1, p. 53–68, jul. 2002. Citado na página [35](#).

Marins, A. et al. Different methods of component separation for the bingo telescope. (em preparação). 2022. Citado na página [12](#).

Marinucci, D. et al. Spherical needlets for cosmic microwave background data analysis. , v. 383, n. 2, p. 539–545, jan. 2008. Citado na página [36](#).

McLeod, M.; Balan, S. T.; Abdalla, F. B. A joint analysis for cosmology and photometric redshift calibration using cross-correlations. , v. 466, n. 3, p. 3558–3568, abr. 2017. Citado na página [17](#).

McQuinn, M. et al. Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *The Astrophysical Journal*, IOP Publishing, v. 653, n. 2, p. 815, 2006. Citado na página [11](#).

Miville-Deschénes, M. A. et al. Separation of anomalous and synchrotron emissions using WMAP polarization data. , v. 490, n. 3, p. 1093–1102, nov. 2008. Citado na página [22](#).

Morales, M. F.; Wyithe, J. S. B. Reionization and Cosmology with 21-cm Fluctuations. , v. 48, p. 127–171, set. 2010. Citado na página [19](#).

Olivari, L. et al. Cosmological parameter forecasts for h i intensity mapping experiments using the angular power spectrum. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press, v. 473, n. 3, p. 4242–4256, 2017. Citado na página [12](#).

Olivari, L.; Remazeilles, M.; Dickinson, C. Extracting h i cosmological signal with generalized needlet internal linear combination. *Monthly Notices of the Royal Astronomical Society*, The Royal Astronomical Society, v. 456, n. 3, p. 2749–2765, 2016. Citado na página [12](#).

- Olivari, L. C.; Remazeilles, M.; Dickinson, C. Extracting H I cosmological signal with generalized needlet internal linear combination. , v. 456, n. 3, p. 2749–2765, mar. 2016. Citado 2 vezes nas páginas [39](#) e [40](#).
- Oppizzi, F. et al. Needlet thresholding methods in component separation. , v. 2020, n. 3, p. 054, mar. 2020. Citado na página [51](#).
- Paykari, P.; Starck, J.-L. S. Cosmic Microwave Background Data Analysis. In: _____. *Advances in Machine Learning and Data Mining for Astronomy*. [S.l.: s.n.], 2012. p. 55–87. Citado na página [33](#).
- Pietrobon, D. et al. NeedATool: A Needlet Analysis Tool for Cosmological Data Processing. , v. 723, n. 1, p. 1–9, nov. 2010. Citado na página [36](#).
- Planck Collaboration et al. Planck 2018 results. VI. Cosmological parameters. , v. 641, p. A6, set. 2020. Citado na página [40](#).
- Planck Collaboration et al. Planck 2018 results. XII. Galactic astrophysics using polarized dust emission. , v. 641, p. A12, set. 2020. Citado na página [39](#).
- Planck Collaboration et al. Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies. , v. 596, p. A109, dez. 2016. Citado na página [39](#).
- Planck Collaboration et al. Planck 2018 results. IV. Diffuse component separation. , v. 641, p. A4, set. 2020. Citado na página [39](#).
- Pritchard, J. R.; Loeb, A. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, v. 75, n. 8, p. 086901, ago. 2012. Citado na página [18](#).
- Remazeilles, M.; Delabrouille, J.; Cardoso, J.-F. Foreground component separation with generalized Internal Linear Combination. , v. 418, n. 1, p. 467–476, nov. 2011. Citado 2 vezes nas páginas [38](#) e [39](#).
- Remazeilles, M. et al. An improved source-subtracted and destriped 408-MHz all-sky map. , v. 451, n. 4, p. 4311–4327, ago. 2015. Citado na página [22](#).
- Santos, M. G.; Cooray, A.; Knox, L. Multifrequency Analysis of 21 Centimeter Fluctuations from the Era of Reionization. , v. 625, n. 2, p. 575–587, jun. 2005. Citado 2 vezes nas páginas [11](#) e [19](#).
- Shao, J.; Tu, D. *The jackknife and bootstrap*. [S.l.]: Springer Science & Business Media, 2012. Citado na página [46](#).
- Shaver, P. A. et al. Can the reionization epoch be detected as a global signature in the cosmic background? , v. 345, p. 380–390, maio 1999. Citado na página [11](#).
- Starck, J.; Murtagh, F.; Bertero, M. The starlet transform in astronomical data processing: Application to source detection and image deconvolution. *Handbook of Mathematical Methods in Imaging*, O. Scherzer ed., Springer, Citeseer, p. 1489–1531, 2011. Citado 3 vezes nas páginas [30](#), [36](#) e [45](#).
- Starck, J.-L.; Donoho, D.; Elad, M. Redundant multiscale transforms and their application for morphological component separation. *Adv. Imaging Electron Phys.*, CM-P00052061, n. DAPNIA-2004-88, 2004. Citado na página [36](#).

Starck, J.-L.; Fadili, J.; Murtagh, F. The Undecimated Wavelet Decomposition and its Reconstruction. *IEEE Transactions on Image Processing*, v. 16, n. 2, p. 297–309, fev. 2007. Citado 3 vezes nas páginas [30](#), [33](#) e [36](#).

Starck, J.-L.; Murtagh, F.; Fadili, J. M. *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. [S.l.]: Cambridge university press, 2010. Citado na página [36](#).

Starck, J.-L.; Siebenmorgen, R.; Gredel, R. Spectral Analysis Using the Wavelet Transform. , v. 482, n. 2, p. 1011–1020, jun. 1997. Citado na página [36](#).

Weinberg, D. H. et al. Observational probes of cosmic acceleration. , v. 530, n. 2, p. 87–255, set. 2013. Citado na página [15](#).

Wolfe, A. M.; Gawiser, E.; Prochaska, J. X. Damped lyman alpha systems. *Ann. Rev. Astron. Astrophys.*, v. 43, p. 861–918, 2005. Citado na página [11](#).

Wuensche, C. A. et al. The BINGO Project II: Instrument Description. *arXiv e-prints*, p. arXiv:2107.01634, jul. 2021. Citado na página [17](#).

Xavier, H. S.; Abdalla, F. B.; Joachimi, B. *FLASK: Full-sky Lognormal Astro-fields Simulation Kit*. 2016. ascl:1606.015 p. Citado na página [17](#).

Zaldarriaga, M.; Furlanetto, S. R.; Hernquist, L. 21 centimeter fluctuations from cosmic gas at high redshifts. *The Astrophysical Journal*, IOP Publishing, v. 608, n. 2, p. 622, 2004. Citado 2 vezes nas páginas [11](#) e [40](#).

Zhang, J. et al. The BINGO Project VI: HI Halo Occupation Distribution and Mock Building. *arXiv e-prints*, p. arXiv:2107.01638, jul. 2021. Citado na página [17](#).

Zibulevsky, M.; Pearlmutter, B. A. Blind source separation by sparse decomposition. In: Szu, H. H. et al. (Ed.). *Wavelet Applications VII*. [S.l.: s.n.], 2000. (Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, v. 4056), p. 165–174. Citado na página [36](#).