

doi: 10.3969/j.issn.1000-8349.2015.04.04

# 宇宙大尺度结构和原初扰动非高斯性 研究进展

杨 圣

(北京师范大学 天文系, 北京 100875)

**摘要:** 宇宙微波背景辐射 (CMB) 探测和大尺度结构 (LSS) 巡天是两种研究宇宙演化的重要手段。近年来高精度的宇宙微波背景辐射实验和大尺度结构巡天可以对今天的宇宙学模型给出精确测量, 比如宇宙曲率、暴胀模型参数和原初非高斯性等。宇宙大尺度结构的原初非高斯性揭示了在甚早期宇宙中引发大尺度结构的原初扰动的物理诱发机制, 有效地对不同暴胀模型进行区分, 是研究甚早期宇宙的有效工具和手段。但是最近的研究结果表明, 很多大尺度结构巡天的数据被系统误差严重污染, 如大气视宁度、天光背景等, 这些误差导致由大尺度结构巡天数据给出的原初非高斯的研究结果与实际情况相差甚远。大尺度结构巡天数据的系统误差消除和它对限制宇宙原初非高斯性的改善, 对更好地了解甚早期宇宙具有重要意义。

**关 键 词:** 原初非高斯性; 功率谱; 大尺度结构; 系统误差

**中图分类号:** P159

**文献标识码:** A

## 1 引 言

宇宙学成为一门学科始于爱因斯坦在 1917 年建立的广义相对论, 到了 20 世纪 80 年代, 宇宙演化的标准模型——热大爆炸模型和原初核合成理论基本建立。可这一模型存在一些明显的缺陷, 如: 奇点问题、磁单极、宇宙空间平坦性和视界问题等。1980 年 Guth 和 So-Young<sup>[1, 2]</sup> 提出暴胀模型, 很好地解决了其中的大部分问题。基于暴胀的宇宙学扰动理论预言了一个原初扰动, 它是产生微波背景和大尺度涨落的原初“种子”, 大量天文观测数据表明, 这个扰动在大尺度上是近标度不变、绝热和近高斯的。而现存的暴胀模型有很多种, 可以利用今天的天文观测数据对其预言的暴胀模型参数和原初非高斯参数进行研究。

暴胀模型主要是用来解决标准宇宙模型中存在的平坦性问题、熵问题、视界问题和原初

---

收稿日期: 2015-03-23; 修回日期: 2015-09-24

资助项目: 国家自然科学基金 (11422323)

通讯作者: 杨圣, yangs@mail.bnu.edu.cn

扰动问题等难题而提出的一种唯像模型,标准宇宙学的这些疑难其实是如何精细调节宇宙初始条件的问题:对于平坦性问题和熵问题,要求宇宙必须经历过一个非绝热的演化过程;而对于视界问题和原初扰动起源的问题,可以通过假设宇宙经历了一个快速膨胀的相变过程,这个过程使得宇宙物理尺度的演化比视界更快,这一时期结束时的视界如果比大爆炸模型对宇宙的期望视界还要大,就可以解决标准宇宙遇到的诸多疑难。暴胀模型假设暴胀过程中引起宇宙加速膨胀的负压成分是一种零自旋的标量场,即暴胀子场,它与引力场的耦合会传递到弗里德曼方程中,从而影响宇宙的膨胀。不是所有的标量场都可以使共动粒子视界快速膨胀,充分条件是标量场处于慢滚区,即定义的慢滚参数  $\eta < 1$ ,宇宙暴胀直到慢滚条件不再成立为止,中间经历的时间定义为一个叫做 e 折时间的量 (e-folding time)<sup>[3]</sup>。大部分暴胀模型都可以归结为单场暴胀的模型,它的根本特点在于仅由一个动力学参量来描述。依据慢滚参数单场暴胀模型又可以分为大场模型 (Large-field models)、小场模型 (Small-field models) 和混合模型 (Hybrid models) 三大类<sup>[4]</sup>。在暴胀宇宙模型中,原初扰动源于暴胀场的真空量子扰动,真空量子扰动是因为引入的暴胀标量场本身的扰动所产生的,在暴胀期间它会因为出视界而被“冻结”,之后慢慢重新进入视界范围,变成经典扰动,最后演化为如今的宇宙结构。 $\zeta = \delta(x) - \frac{3}{5}f_{\text{NL}}(\delta^2(x) - \langle \delta^2(x) \rangle)$ <sup>[5]</sup> 是定义在等能动超曲面上的一个空间曲率扰动参量<sup>[6]</sup>,不同的暴胀模型预言不同幅度、形状和尺度依赖的物质非高斯分布,这个特性对于用原初非高斯来甄别暴胀模型非常重要。

对原初非高斯的研究,可以近似地将原初引力势  $\Phi$  对偏离高斯情况做泰勒展开<sup>[7]</sup> (高阶效应微弱,可以忽略):  $\Phi = \phi + f_{\text{NL}}(\phi^2 - \langle \phi^2 \rangle)$ ,其中  $\phi$  是高斯随机场。这样就可以用非高斯系数  $f_{\text{NL}}$  来描述非高斯性了。对非高斯性研究的方法也有很多种,比较常用的是 CMB 的三点关联函数或其傅里叶空间上的双功率谱 (Bi-spectrum) 来对非高斯参数进行限制。目前用 CMB 来限制原初非高斯系数  $f_{\text{NL}}$  的限制结果有 ( $f_{\text{NL}}$  信号有三种,详见第 2 章,其中 local 类型比较明显): WMAP7<sup>[8]</sup> 给出在 95% 的置信区间上  $f_{\text{NL}} < +74$ , WMAP9<sup>[9]</sup> 给出在 95% 的置信区间上  $f_{\text{NL}}^{\text{local}} = 37.2 \pm 19.9$ , Planck<sup>[10]</sup> 给出在 68% 的置信区间上  $f_{\text{NL}}^{\text{local}} = 2.7 \pm 5.8$ 。除了利用 CMB 的三点函数和双谱研究以外,还可以利用非高斯性在 LSS 上的局域成团性,用 LSS 的二阶统计效应来进行研究。这种用 LSS 统计效应的方法对  $f_{\text{NL}}$  的限制结果包括: Slosar 等人<sup>[11]</sup> 用 SDSS<sup>[12]</sup> DR5<sup>[13]</sup> 的类星体数据限制得到在 95% 的置信区间上  $-82 < f_{\text{NL}} < +74$ ,在此基础上 Xia 等人<sup>[14]</sup> 用 NVSS 巡天的自相关函数数据得到在 1 个  $\sigma$  误差范围内  $f_{\text{NL}} = 62 \pm 27$ , 2 个  $\sigma$  置信区间上  $25 < f_{\text{NL}} < 117$ 。Xia 等人<sup>[15]</sup> 用 SDSS DR6<sup>[16]</sup> 的类星体分类数据<sup>[17]</sup> 和 NVSS、MegaZ 巡天的 LRG DR7 数据、SDSS II 亮红星系 (LRG) 光度红移巡天的数据联合限制得到  $f_{\text{NL}} = 48 \pm 20$ 。同样,射电波段的巡天项目也可以运用到这方面的研究中来,文献 [18] 用 skA 来限制原初非高斯。skA 测量亿级别的中性氢星系,覆盖 30 000 平方度的天区,最深达到红移  $z \geq 2$ ,如此大体积的巡天可以很好地用大尺度结构功率谱来限制原初非高斯系数,限制结果是  $\delta(f_{\text{NL}}) = 1.54$ 。

但是近期的研究<sup>[19-21]</sup>表明,这种用 LSS 局域物质密度功率谱或相关函数来限制原初非高斯系数的方法受到巡天系统误差的影响较大,比如:视宁度、大气质量、星际消光等,这些

系统误差产生的功率谱往往会混入到结果中来, 很难分辨, 并且误差的量级与预期信号的量级相比拟或者更大, 因而不能忽略。所以, 研究结果表明, 之前用 LSS 方法对原初非高斯系数的限制结果可能因为这些系统误差的效应而失真, 因而消除系统误差给 LSS 统计特性带来的影响对于限制原初非高斯非常重要。

本文从 CMB 限制非高斯引入, 着重介绍用 LSS 处理宇宙原初非高斯的方法和这个过程中系统误差处理的重要性。第 2 章着重介绍用 CMB 限制原初非高斯的相关内容; 第 3 章对用 LSS 来限制原初非高斯的过程进行讨论; 第 4 章介绍可能涉及到的系统误差以及如何消除系统误差对结果的影响; 第 5 章进行简单的总结与展望。

## 2 CMB 对非高斯的限制

对于 CMB, 要研究它的光子涨落对非高斯性的限制, 常常利用一些统计工具来对它进行处理分析。对于 CMB 而言, 能够分辨高斯性和非高斯性的最低阶统计工具是三点相关函数或在傅里叶空间中的双功率谱 (Bi-spectrum), 它们在傅里叶变化下的形式为  $\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3) \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\Phi(k_1, k_2, k_3)$ , 其中  $\Phi$  是规范不变的巴丁引力势<sup>[7]</sup>, 双谱  $B_\Phi(k_1, k_2, k_3) = f_{\text{NL}} F(k_1, k_2, k_3)$ <sup>[10]</sup>,  $f_{\text{NL}}$  是定义的非高斯系数,  $F(k_1, k_2, k_3)$  是双谱的 3 个波矢  $k_1, k_2, k_3$  构成的“动量位型”<sup>[10]</sup>。

非高斯的“动量位型” $F(k_1, k_2, k_3)$  是 CMB 双谱对应的 3 个波矢  $k_1, k_2, k_3$  组成的一个三角形, 它描述了这个双谱的形状<sup>[22]</sup>。天文观测可以得到不同位型的非高斯信号, 用这个三角形的形状将非高斯信号分类: (1) local 的非高斯位型<sup>[23-26]</sup> 对应了一个看上去很“瘦”的三角形, 这个非高斯信号在  $k_1 \ll k_2 \approx k_3$  时达到极大, 其产生机制是原初非高斯在超视界尺度上产生的; (2)  $k_1 \approx k_2 \approx k_3$  时极大的非高斯信号称作 equilateral 非高斯<sup>[27]</sup>, 它对应一个近等边三角形位型, 是由于三种扰动模式在同时穿视界时波长相当, 从而相互关联导致的; (3) 最后一种叫 folded 或 flattened 的动量位型<sup>[28]</sup>, 它对应的非高斯信号是当  $k_1 + k_2 \approx k_3$  时取极大值, 这个动量位型看上去很“胖”, 这个非高斯信号也是由于三种扰动模式之间的相互关联导致的。这三种动量位型的示意图见图 1。

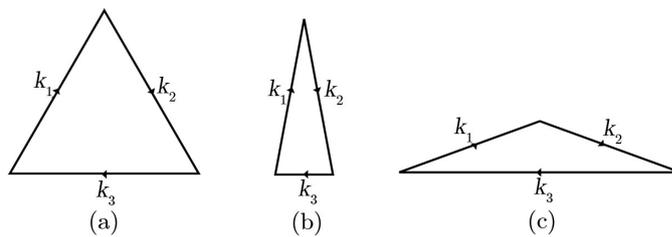


图 1 三种动量位型的示意图<sup>[85]</sup>

(a) “瘦”三角形非高斯信号; (b) 等边三角形非高斯信号; (c) “胖”三角形非高斯信号。

可以通过对 CMB 巡天的数据进行高阶统计, 得到它的双谱或者三点相关函数, 从而得

到对非高斯系数  $f_{NL}$  的限制。下面将具体介绍如何利用 CMB 对非高斯进行限制。

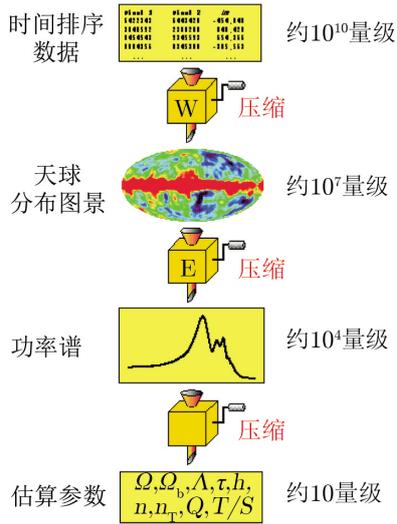


图2 CMB 观测的基本流程示意图<sup>[35]</sup>  
第3章讲到的 LSS 观测流程也是类似的。

从观测数据出发,图2给出了CMB从观测数据到最终限制结果的一个简单示意图:通过CMB探测仪器得到的数据是一组依赖于时间、坐标的温度强度分布数据,这组数据包含极其丰富的信息(约 $10^{10}$ 数量级的信息量),这些信息并不能够直接利用,需要采用统计技术对其进行“压缩”,通过压缩把这组三维数据降到二维,数据量也下降3个数量级,这就是得到的CMB关于天球的分布图景(map)。得到的这组数据仍然需要进行处理,于是通过统计工具——相关函数或功率谱——继续对这个map进行降维,得到数据量大约在 $10^4$ 数量级的角功率谱,角功率谱已经能够比较直接地反映温度在不同尺度上的幅值或与某些对应map在尺度上的相关性。当然,如果需要更直接地限制宇宙学,会通过蒙特卡洛方法或者费舍尔矩阵等工具来利用这些功率谱,最终得到需要的关于宇宙模型的相关参数,当然这其中也包括了非高斯系数 $f_{NL}$ 。

## 2.1 从CMB观测数据到角功率谱的过程

文献[29, 30]非常详细地讲述了CMB从卫星上观测的原始数据到功率谱的过程:CMB探测器在时间 $t$ 时接收到的CMB信号强度为 $d_i = s_i + n_i$ ,其中 $s_i, n_i$ 是已经被Healpix格点化的CMB温度的信号和噪声, $i$ 是Healpix格点化过程给出的每个格点对应的标示号。于是可以给出信号和噪声各自的自相关函数(此处以两点相关函数为例,更高阶的类似)为: $C_{Tii'} = \langle s_i s_{i'} \rangle$ ,  $C_{nii'} = \langle n_i n_{i'} \rangle$ 。而在 $\mathbf{x}_i$ (坐标 $\mathbf{x}$ 是 $i$ 的一维数组)处的探测器信号中的CMB真实信号可以反映该处的CMB强度 $T(\mathbf{x})$ :  $s_i = \int d\Omega H(\mathbf{x}, \mathbf{x}_i) T(\mathbf{x})$ ,其中 $H$ 又被称为天线模式,即天线对接收到的CMB光子的一个响应函数。又因为CMB信号是在 $360^\circ$ 全天球分布的,所以对其做球谐函数展开: $T(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi)$ ,则可以通过展开系数 $a_{lm}$ 的关联函数 $\langle a_{lm} a_{l'm'}^* \rangle = C_l \delta_{ll'} \delta_{mm'}$ 定义得到角功率谱 $C_l$ ,它与信号关联函数之间的关系为 $C_{Tii'} = \sum_l \frac{2l+1}{4\pi} C_l W_{ii'}(l)$ ,其中窗口函数 $W_{ii'}(l) = \sum_{nn'} H_{in} H_{i'n'} P_l(\cos \theta_{nn'})$ , $\theta_{nn'}$ 表征对应CMB上的这两个点 $\mathbf{x}(n)$ 和 $\mathbf{x}(n')$ 的夹角。

## 2.2 伪功率谱的处理——PCL判据和QML判据

实际上,受制于仪器、技术和某些系统误差的影响,不可能测量得到CMB对全天的map,因而只能通过已有天区的数据来估量实际的情况,这就好像给全天打上了一个“马赛克”(mask,见图3),我们只能得到mask以后CMB的伪功率谱(Pseudo Power Spectrum),并以此来估计实际的功率谱,而这个用伪功率谱来估量实际功率谱的方法叫做伪功率谱判据(PCL<sup>[30-33]</sup>)。

对应CMB温度观测量的格点化矢量为 $\mathbf{s} = (s_0, \dots, s_{N_{pix}-1})$ ,其中 $N_{pix}$ 为对全天



图 3 银道坐标系下, 用于处理 CMB 和 LSS 的 mask<sup>[20]</sup>  
(a) CMB mask; (b) LSS mask。它们是分别依据 WMAP KQ85 和 SDSS DR6 做出来的。

格点化以后的格点总量, 可以取一组对应的 mask 量—— $\mathbf{m} = (m_0, \dots, m_{N_{pix}-1})$ , 其中  $m_i = 0$  表示这个格点的数据观测不到或者污染严重, 于是被打上了 mask, 而  $m_i = 1$  表示这个格点是可视的, 数据点有效。对于 CMB 光子, 全天的二维密度场  $\delta(\mathbf{n})$  作球谐展开为:  $\delta(\mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} Y_{lm}(\mathbf{n})$ , 其中展开系数  $a_{lm} = \int d\Omega_n \delta(\mathbf{n}) Y_{lm}^*(\mathbf{n})$ 。而对应的观测得到的展开系数  $\tilde{a}_{lm} = \int d\Omega_n \delta(\mathbf{n}) W(\mathbf{n}) Y_{lm}^*(\mathbf{n})$ , 其中  $W(\mathbf{n})$  是巡天 mask 带来的权重函数, 波浪线代表观测结果。所以观测得到的伪功率谱 (PCL) 的期望值为  $\tilde{C}_l = \frac{1}{2l+1} \sum_{m=-l}^l |\tilde{a}_{lm}|^2$ 。伪功率谱  $\tilde{C}_l$  与真实功率谱  $C_l$  之间的关系为:  $\langle \tilde{C}_l \rangle = \sum_{l'} C_{l'} G_{ll'}$ , 其中相关系数  $G_{l_1 l_2} = (2l_2 + 1) \sum_{l_3=|l_1-l_2|}^{l_1+l_2} \frac{2l_3+1}{4\pi} \tilde{W}_{l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2$ ,  $\tilde{W}_l = \frac{1}{2l+1} \sum_{m=-l}^l |\tilde{w}_{lm}|^2$  是权重函数  $W(\mathbf{n})$  的角功率谱, 而  $\tilde{w}_{lm} = \int d\Omega_n W(\mathbf{n}) Y_{lm}^*(\mathbf{n})$  是对应的权重系数。同时, PCL 方法给出的误差与真实误差之间的关系为:  $F_{ll'}^{\text{full}} = \frac{(2l+1)f_{\text{sky}}}{2(C_l + N_l)(C_{l'} + N_{l'})} \delta_{ll'}$ , 其中  $C_l$  是理论功率谱,  $N_l = \Delta\Omega/N$  被称为 shot noise 的噪声项<sup>[14]</sup>。

PCL 方法比较简单, 但是它并没有将因为 mask 存在而导致的信息丢失降到最低, 同时 PCL 比较适合处理平的功率谱<sup>[32, 34]</sup>。当然除了这种 PCL 伪功率谱判据以外, 还有一种比较主流的 QML 判据 (Quadratic Maximum Likelihood, 二次最大似然估计), Tegmark<sup>[35]</sup> 在 1998 年首次结合二次判据法 (Quadratic Methods, QM) 和最大似然法 (Maximum-Likelihood, ML) 提出 QML 方法, 并论证了 QML 是一种没有因为 mask 存在而导致信息丢失的方法。

文献 [20] 主要通过计算程序的复杂度和估算得到的功率谱效果这两个方面比较 QML 和 PCL 两种方法的优劣: (1) PCL 和 QML 都依赖于 mask, 但是 QML 还依赖于格点协方差矩阵的先验模型, 即用 QML 方法运行程序的时候, 必须输入一个先验的  $C_l$ 。(2) 通过如图 3 的 LSS 和 CMB 的 mask 比较 PCL 和 QML 的性能, 发现在  $l$  比较小, 即大尺度范围时, QML 性能更好; 而  $l$  比较大, 即小尺度范围内用 PCL 会更好一点。(3) PCL 算法速度更快, 需要的内存也更小, 比较适合作星系巡天的判据, 但是 PCL 在小尺度范围时很难避免这个范围内的 shot noise; 相对来说, QML 更复杂, 计算量更大, 但是它可以与 PCL 互补, 进行大尺度的计算。

### 2.3 程序实现

事实上, 这个从 CMB 观测数据到角功率谱的实际处理过程在程序上会更加复杂, 幸运的是, 有很多开源的宇宙学软件可以帮助处理这个过程: Healpix 能够对全天进行格点化, 从而给出数据的 map; 而 Polspice 可以计算这个 map 的角功率谱。

#### 2.3.1 得到全天的 map——Healpix

Healpix<sup>[36-38]</sup> 是 Hierarchical Equal Area iso-Latitude Pixelization 的缩写, 直译为分层等面积等纬度格点化, 顾名思义, 它是用来做格点化的工具。简单的理解, Healpix 是一个集成在多程序语言环境下对球面做格点化并进行统计分析的子程序集。图 4 和图 5 是非常直观的 Healpix 的格点层次示意图。如图 4, Healpix 把一个球分为 12 个四边形 (上中下各 4 个), 每个四边形里的点数取作  $N_{\text{side}}^2$ , 则球面共有  $N_{\text{pix}} = 12N_{\text{side}}^2$  个格点。而图 5 是把球表示在二维上的一个示意图, Healpix 逆时针把每个四边形里的格点编号。于是 Healpix 就完成了对球面进行格点化的任务, 并对每个格点作一个标识  $i$  ( $i = 0, \dots, 12N_{\text{side}}^2$ )。对于巡天数据, Healpix 可以根据每个点的坐标信息算出该点对应的  $i$ , 这样把二维坐标降到一个数组, 比较方便处理坐标信息; 然后通过 Healpix 算出坐标为  $i$  时的巡天强度量, 把数据根据坐标辐射到全天图景, 得到数据的 map。而 Healpix 的标识号  $i$  的排列方式有两种: 线排列 (ring) 和网排列 (nest)。同时选取不同坐标系, 例如赤道坐标系和银道坐标系, 也会有不同的 map, 做关联函数或求功率谱的时候必须保证坐标系和排列方式一致。

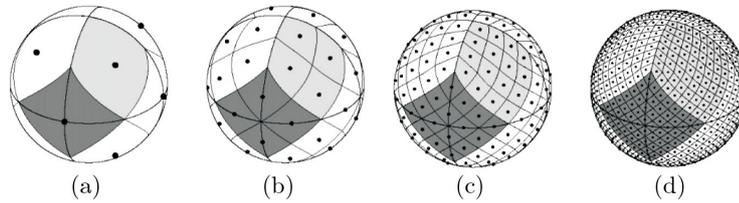


图 4 Healpix 软件的三维分层示意图<sup>[36]</sup>  
(a), (b), (c), (d) 分别对应  $N_{\text{side}} = 1, 2, 4, 8$  的情形。

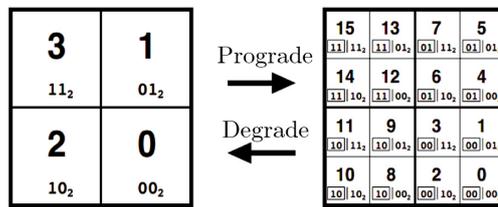


图 5 Healpix 软件的二维格点示意图<sup>[36]</sup>

它具体给了每个点一个相对应的标识  $i$ , 以及在  $N_{\text{side}}$  加倍时,  $i$  是如何确定的。其中每个格点的上方是十进制自然数、下方是二进制机器数。

### 2.3.2 计算功率谱——Polspice

虽然 Healpix 的 anafast 子程序就可以用来直接计算功率谱, 但是 Polspice 是一个更方便更智能的计算功率谱的程序<sup>[39]</sup>。它可以读入信号和噪声的 map、mask 向量等, 通过计算输出相关函数和功率谱, 例如温度的两点相关函数:  $\xi_{12} = \langle T(q_1)T(q_2) \rangle = \sum_l \frac{l+1/2}{l(l+1)C_l P_l(\cos\theta)}$ , 其中  $P$  是球谐展开过程中的勒让德函数。同样地, Polspice 也可以估算功率谱的误差:  $\Delta C_l = \left(\frac{2}{(2l+1)f_{\text{sky}}}\right)^{\frac{1}{2}} [C_l + w^{-1}l(l+1)W_l^{-2}/2\pi]$ 。

### 2.4 通过对功率谱的似然分析限制非高斯性

至此, 从 CMB 的观测数据得到了功率谱, 接下来要利用功率谱来得到相关宇宙学信息。通常用到的统计分析工具有费舍尔矩阵 (Fisher Matrix) 和蒙特卡洛—马尔科夫链 (MCMC) 这两种方法。文献 [40] 对比了这两种方法, 费舍尔矩阵是用得比较多的一种方法, 因为它相对来说会更快, 但是它依赖于参数是高斯分布。而事实上, 参数往往由于简并或者范围是有界的而导致不能满足高斯分布。这个时候 MCMC 就非常重要了。在宇宙学上, 有一款非常著名的 MCMC 开源软件 cosmoMC<sup>[41]</sup> (<http://cosmologist.info/cosmomc/>), 它的出现使得宇宙学上的统计分析工作变得相对简单。cosmoMc 是 Cosmological MonteCarlo 的简称, 它是在宇宙学模型的基础上引入蒙特卡洛—马尔科夫链这种随机撒点取样的方法, 基于 Fortran 程序语言环境而开发出来的一款开源软件。

在统计学上有一个非常有名的“贝叶斯公式”, 也被称作“逆概公式”, 人们可以通过参数的验前分布和贝叶斯公式, 推导出该参数的验后分布。但是有时候参数的验前分布不能直接得到, 于是产生了蒙特卡洛取样方法, 它通过对参数进行随机取样求平均的方法, 并结合统计中的“大数定理”: 当取样数足够大, 并且每次取样服从独立原则的时候, 这个平均值趋向于参数期望的真实值; 如果一条采样链某时刻的状态只与前一刻的状态相关就称为马尔科夫链, 马尔科夫链只有一个唯一静态点, 当马尔科夫链足够长以至于达到静态点以后, 该链就会收敛于一个恒定值。有了 MCMC, 就可以通过随机取样的方法估算参数的分布, 从而借助贝叶斯公式完成参数估计。

cosmoMc 由两个主要的部分组成: camb 主要功能是做理论功率谱, 另一部分主要做 MCMC 主程序。camb 本身是由一款叫 cmbfast 的软件演化过来的, 它是一个计算理论模型的功率谱的程序, 它的作用是给 MCMC 主程序提供算好的理论功率谱。而 MCMC 主程序的工作流程是: (1) 调用 camb 给的理论功率谱; (2) 从一个叫做 data 的文件夹里调用实际巡天项目的数据, 并按照本文第三章和第四章的方法, 求得 CMB 和 LSS (也包括别的类型的数据, 例如 BAO 的数据, 超新星的数据等) 的实际功率谱; (3) 通过理论功率谱和实际功率谱求得  $\chi^2$  项; (4) 最后是 MCMC 过程: 确定收敛判据, 计算机在给定的初始值位置开始, 按照给定步长随机撒点, 得到马尔科夫链, 当链跑到收敛判据成立以后, 分析得到宇宙学限制结果。当然为了提高工作效率, 中间也可能会涉及到一些并行计算。

### 3 LSS 对非高斯的限制

第 2 章讲了 CMB 对非高斯的限制, LSS 做宇宙学限制基本步骤是类似的, 只不过是把 CMB 温度强度的全天分布变成了 LSS 示踪体数密度的全天分布。但是 LSS 对非高斯的限制原理却与 CMB 不大一样, CMB 是利用温度分布的三点相关函数或双谱来研究温度 map 上的偏离高斯的起伏, 而 LSS 是利用大尺度结构巡天示踪体数密度分布的两点相关函数或功率谱, 通过一个在物质成团结构上预言存在的、偏离高斯的偏移量  $\Delta b$  来限制非高斯系数。

#### 3.1 LSS 观测数据与实际 LSS 物质分布之间的偏移量 $b$

LSS 示踪体给出的物质密度分布并不是实际的物质密度分布, 它只是发光物质对整个物质的一个示踪而已, 他们之间必定会有一个偏移量  $b$ 。

以类星体为例<sup>[42]</sup>, 类星体的物质分布与宇宙 LSS 实际密度分布之间有:  $\delta_q(\mathbf{n}) = \int_0^\infty \delta[\mathbf{n}, \chi(z)] f_i(z) dz$ ,  $\chi(z)$  是共动距离。根据文献 [42],  $f_i(z) = b(z) \Pi(z) + \int_z^\infty W(z, z') [\alpha(z') - 1] \Pi_i(z') dz'$ ,  $\Pi(z) \equiv \frac{dN_i}{dz(z)}$  是类星体的红移分布, 则第一项  $b(z) \Pi(z)$  是线性偏移量 (linear bias), 第二项是放大偏移量 (magnification bias), 其中  $W(z, z') = \frac{3}{2} \Omega_m H_0^2 \frac{1+z}{cH(z)} \chi^2(z) \left[ \frac{1}{\chi(z)} - \frac{1}{\chi(z')} \right]$  是透镜窗函数, 第二项只在大尺度范围内作用比较明显, 在局域宇宙中一般可以不考虑。又对类星体的密度扰动求相关函数, 有  $C^{g_i g_j} = \langle \delta g_i(\mathbf{n}) \delta g_j(\mathbf{n}) \rangle$ , 将 QSO 的自相关函数在球谐空间中展开有  $C^{g_i g_j} = \sum_{l=2}^{\infty} \frac{2l+1}{4\pi} C_l^{g_i g_j} P_l[\cos(\vartheta)]$ , 其中,  $C_l^{g_i g_j}$  是 LSS 的角功率谱,  $P_l$  是勒让德函数,  $\vartheta$  是  $i$  和  $j$  点之间的角分离距离。于是有:  $C^{g_i g_j} = \frac{2}{\pi} \int k^2 dk P(k) [I_l^q(k)]^2$ , 其中,  $I_l^q(k) = \int b_q(z) \frac{dN}{dz}(z) \delta_m(k, z) j_l[k\chi(z)] dz$ <sup>[14]</sup>。

#### 3.2 LSS 如何限制非高斯——局域偏移量 $\Delta b$

根据文献 [44], 非高斯性会导致这个可视物质与实际物质的偏差  $b$  上再产生一个  $\Delta b \propto \frac{f_{\text{NL}}}{k^2}$  的尺度依赖的偏移量 (bias)。这个偏移效应主要体现在暗物质晕和星系等物质成分上, 导致暗物质晕和星系团等物质集团的密度与标准模型  $\Lambda$ CDM 宇宙预言的  $b$  不符, 相关分析文章见文献 [11, 45–51]。

Slosar 等人<sup>[11]</sup> 比较详细地介绍了这个局域偏差产生的原理。Cole<sup>[52]</sup> 提到了一种叫作峰值背景分离 (peak-background split) 的分类方法, 可以把宇宙中物质成团形成的晕 (如暗物质晕, 星系团等类似结构) 的密度场分为长波模式  $\delta_l$  和短波模式  $\delta_s$  两个类别, 则晕的密度可以表示成  $\rho(\mathbf{x}) = \bar{\rho}(1 + \delta_l + \delta_s)$ 。对长波模式, 晕的数密度  $n[\delta_l(\mathbf{x}), P_s(k_s); M]$  对应有  $n(\mathbf{x}) = \bar{n}(1 + b_L \delta_l)$ , 其中  $b_L = \bar{n}^{-1} \frac{\partial n}{\partial \delta_l}$  是拉格朗日偏移量, 则实际偏差  $b = \frac{n(\mathbf{x})}{\bar{n}} = b_L + 1$ 。

由于引力势是决定宇宙物质分布的动力, 所以对应的高斯引力势也可以相应地分为长波部分和短波部分:  $\phi = \phi_l + \phi_s$ , 将其代入到  $\Phi = \phi + f_{\text{NL}} \phi^2$  中, 有:  $\Phi = \phi_l + f_{\text{NL}} \phi_l^2 + (1 + 2f_{\text{NL}} \phi_l) \phi_s + f_{\text{NL}} \phi_s^2 + \text{const}$ 。假设密度和引力势之间存在  $\delta(k) = \alpha(k) \phi(k)$  这

样的关系, 将引力势和密度场联系起来, 通过物理推导可以知道这个关联的系数:  $\alpha(k) = \frac{2c^2 k^2 T(k) D(z)}{3\Omega_m H_0^2}$ ,  $T(k)$  是密度的传播函数,  $D(z)$  是宇宙的线性增长因子,  $c$  是光速,  $\Omega$  是宇宙的物质密度组分,  $H_0$  是今天的哈勃常数; 则对应的大尺度结构上的长波部分的密度扰动  $\delta_l(k) = \alpha(k)\phi_l(k)$ , 所以相应短波部分的密度扰动为  $\delta_s = \alpha[(1 + f_{NL}\phi_l)\phi_s + f_{NL}\phi_s^2]$ 。而  $n$  与  $M$  之间的关系涉及到晕质量函数 (halo mass function) 的研究, 具体见文献 [?]。于是可以求出偏移量  $b_L$ , 而  $b_L = b_L^{\text{GAUSS}} + \Delta b$ ,  $b_L^{\text{GAUSS}}$  是理论算出来的, 因而可得  $\Delta b$  的大小。

Slosar<sup>[11]</sup> 最后通过计算给出了偏移量  $\Delta b(M, k) = 3f_{NL}(b - p)\delta_c \frac{\Omega_m}{k^2 T(k) D(z)} \left(\frac{H_0}{c}\right)^{[11]}$ , 所以可以通过限制偏移量  $b$  来限制原初非高斯系数  $f_{NL}$ 。

### 3.3 LSS 的红移分布

LSS 示踪体随红移的分布关系  $\Pi(z)$  在选源的时候就可以得到。图 6 是 SDSS 类星体在不同红移段的一个分布情况, 其中图 6 (a) 是根据 ross 2011 的类星体数据画出的各个红移处的类星体数目图<sup>[54]</sup>; 图 6 (b) 是通过样本的光度红移和光谱红移进行分类给出的类星体数目随红移的分段图<sup>[55]</sup>。

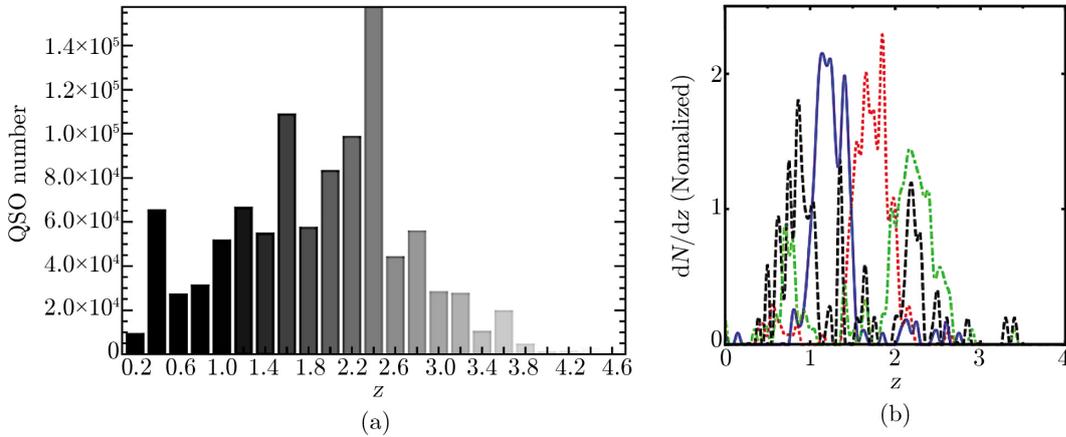


图 6 (a) 根据文献 [54] 的类星体数据画出的各个红移处的类星体数目图; (b) 文献 [55] 给出的类星体数目随红移的分 bin 图

### 3.4 程序实现

#### 3.4.1 从 LSS 数据到功率谱

现阶段关于 LSS 的星系红移巡天有很多, 比较著名的有: 低红移的斯隆数字化巡天 (SDSS)、2 度视场星系红移巡天 2df、高红移的深度 2 红移巡天 (DEEP2) 和 VIMOS-VLT 深度巡天 (VVDS) 等。以斯隆数字巡天为例, 它的巡天数据包括星系的坐标、红移、各个波段的信息 (斯隆分为 5 个波段巡天 u、g、r、i、z)、误差信息等, 而星系又分为很多种, 不能将全部星系样本都拿来限制宇宙模型, 这样得到的功率谱往往很“粗糙”, 给不出有效的限制结果。可以通过一些判据将星系进行细分, 文献 [56] 介绍了如何用似然函数法得到类星体。ross 2011 给出了用这种方法筛选出来的斯隆巡天类星体样本<sup>[54]</sup>

(<http://portal.nersc.gov/project/boss/galaxy/photoz/>), 用 Healpix 格点化后得到的 map 如图 7 所示。然后用 Polspice 就可以通过 map 算出功率谱。

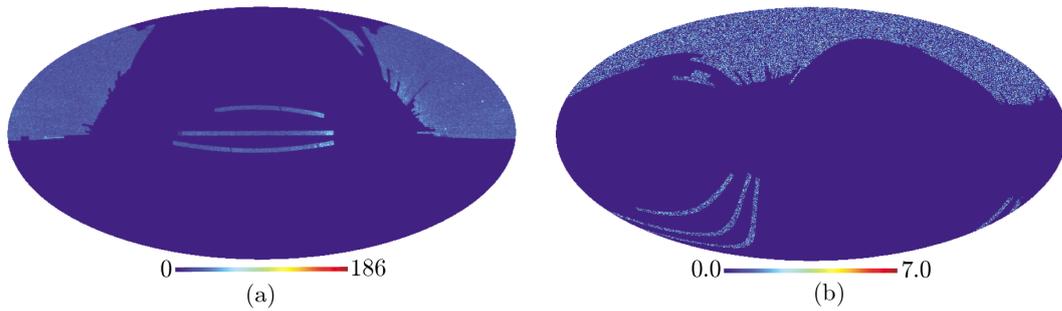


图 7 SDSS 类星体的 map<sup>[54]</sup>  
(a)  $N_{\text{side}} = 128$ , 赤道坐标系; (b)  $N_{\text{side}} = 512$ , 银道坐标系。

### 3.4.2 LSS 功率谱的程序实现和统计分析

类似 CMB 的理论功率谱要用 cosmoMC 的 camb 软件实现一样, LSS 的理论功率谱在程序上用 CAMB.sources (<http://camb.info/sources/>) 软件包实现。而 LSS 的实际功率谱则依托于 Healpix 和 Polspice。同时 Max Tegmark 领导开发了一款名叫 Mangle 的开源软件 (<http://casa.colorado.edu/~ajsh/mangle/>), 这款软件事实上是对 Healpix 的一个很好的补充。如前所述, 给 CMB 和 LSS 等巡天找 mask 是利用 CMB、LSS 做宇宙学研究不可避免的一个步骤, 而随着技术的进步, LSS 巡天项目的覆盖天区更大, 也更加复杂, 制作这个 mask 变得更加困难了。文献 [57, 58] 提出用 Healpix 对每个数据点找标量  $i$ , 实际上是一个  $O(N^2)$  的操作, 随着巡天项目逐渐变大, 搜寻的对应体数目  $N$  越来越多, 计算机的计算量就会非常大, 而 Mangle 的算法使得  $O(N^2)$  的操作变成了  $O(N)$  的操作, 很好地对这个过程的计算进行了简化。同时 Mangle 也可以更好地处理 mask 上格点单元多边形的覆盖问题和边缘化问题, 这些覆盖问题是由 LSS 巡天区域的重叠所导致。图 8 对比做了用 Healpix 格点化和不做格点化的 Mangle 的工作原理, 显然 Mangle 的运行并不一定要先做格点化, 它可以直接把巡天的覆盖区域分离, 然后加权, 最后统一, 这样就得到比较优化的 LSS 的 map 和 mask。同时 Mangle 可以无障碍地和 Healpix 格式进行对接, 以辅助完成对天区数据的格点化。

## 4 系统误差的影响

近期的研究<sup>[19-21]</sup>表明, LSS 这种用局域物质密度功率谱或者相关函数来限制原初非高斯系数的方法受到相关系统误差的影响较大, 如大气视宁度、湍流、恒星点源的污染效应等等, 这些系统误差产生的功率谱往往会混入到结果中来, 并且误差的量级与预期数据的量级相比拟或者更大, 因而不能忽略。研究表明, 之前对原初非高斯系数的限制结果可能因为这

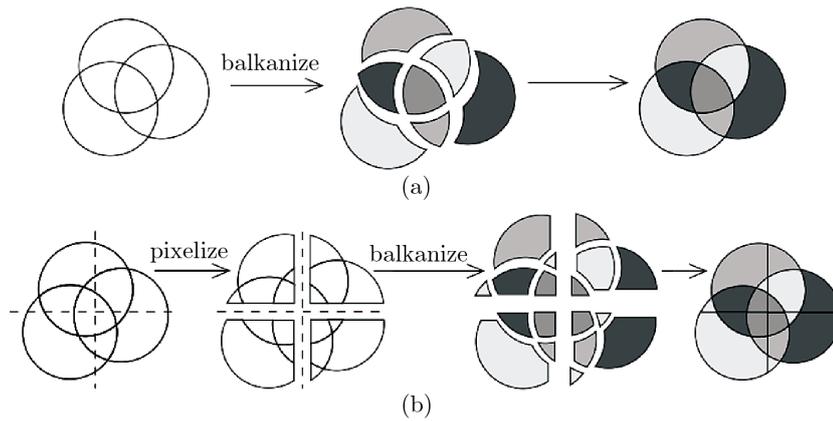


图 8 Mangle 软件的基本操作<sup>[58]</sup>

包括分裂、加权、统一,最后可以得到比较好的 mask,同样可以减小计算消耗。Mangle 是对 Healpix 的一个很好的补充。图 (a) 是 Mangle 单独制作 mask; 图 (b) 是配合 Healpix 一起工作的示意。很明显配合 Healpix 工作量会更大,但是 mask 细节更好,效果可能会更好。

些系统误差的效应而失真,因而消除系统误差给大尺度结构统计特性带来的影响,从而更好地限制原初非高斯就显得非常重要。

仅仅简单将每个系统误差减掉也不可取,如表 1 所示,给出主要系统误差模板之间的相关函数,很明显系统误差之间并不相互独立,即两种不同的系统误差可能含有共同部分,如果简单地将每个误差减掉,也可能导致最终的功率谱失真。所以必须采用一些特殊的方法消除误差:早在 20 世纪, Tegmark 等人<sup>[35, 59]</sup>, Bond 等人<sup>[29]</sup>, Halverson 等人<sup>[61]</sup> 就开始对 CMB 的数据进行系统误差消除,他们普遍使用一种叫做“模式抛射”(model projection)的方法,形象地把不同的成分比喻为不同的模式,将系统误差的模式抛射掉,从而得到相对“干净”的模式。同样的道理,可以利用这个模式抛射原理来处理 LSS 数据。文献 [21] 也提出一种新的方法来消除系统误差。

#### 4.1 可能的系统误差

可能影响结果的系统误差主要包括:消光 (extinction), 恒星点源污染 (stellar contamination), 红恒星污染 (red stellar contamination), 测光色偏差和流量偏差 (stellar color locus offset and stellar flux-error locus offset), 大气质量 (air mass), 视宁度 (seeing), 天空亮度 (sky brightness), 约化儒略日 (modified Julian data), 相机列 (camera column) 和大气反射 (atmospheric refraction)<sup>[19]</sup>。下面对这些误差模板作简单介绍如下。

##### (1) 消光

消光主要是尘埃对电磁波吸收和散射造成光度减弱的现象 (包括星际消光、大气消光等):尘埃颗粒对波长小于或与其尺度相当的电磁波消光显著,而电磁波波长远大于尘埃颗粒尺度时,由于衍射效应消光能力不强,因而电磁波的长波部分得以穿越尘埃,而短波部分被尘埃吸收和散射。这导致了观测得到的电磁波呈现“红化”现象,因而消光 and 红化之间有关系,通过研究给出  $R_v \equiv A_v/E(B - V)$ , 其中  $A_v$  表征消光强度,  $B - V$  是定义的

表 1 主要系统误差模板之间的相关函数<sup>[19]</sup>

	ebv	star	rstar	ugr	gri	riz	uerr	airu	seeu	seer	skyu	skyi	mjd	cam1	cam2	cam3	cam4	cam5	ref
ebv	+1.000	+0.327	+0.118	+0.016	-0.068	-0.252	-0.140	+0.004	-0.043	-0.087	-0.025	-0.069	+0.002	+0.017	-0.004	-0.010	+0.006	-0.008	-0.092
star	+0.327	+1.000	+0.364	+0.254	+0.082	-0.128	-0.182	-0.036	-0.164	-0.140	-0.011	-0.086	-0.006	-0.013	+0.024	+0.026	+0.009	-0.014	+0.011
rstar	+0.118	+0.364	+1.000	+0.082	+0.065	+0.035	+0.132	-0.056	-0.177	+0.050	+0.005	+0.030	+0.005	-0.026	+0.015	+0.005	+0.001	-0.022	+0.006
ugr	+0.016	+0.254	+0.082	+1.000	+0.006	+0.018	+0.020	-0.006	+0.013	+0.011	-0.002	-0.016	-0.039	-0.015	+0.065	+0.019	+0.021	-0.009	+0.052
gri	-0.068	+0.082	+0.065	+0.006	+1.000	+0.086	+0.121	-0.019	+0.024	+0.024	+0.079	+0.073	-0.012	-0.010	-0.040	-0.003	-0.003	+0.004	+0.035
riz	-0.252	-0.128	+0.035	+0.018	-0.086	+1.000	+0.133	-0.103	-0.110	-0.021	+0.137	+0.084	+0.013	+0.002	+0.067	-0.018	-0.050	-0.019	+0.039
uerr	-0.140	-0.182	+0.132	+0.020	+0.121	+0.133	+1.000	-0.037	+0.012	+0.031	+0.116	+0.102	+0.016	-0.018	-0.003	+0.016	+0.019	-0.020	+0.031
airu	+0.004	-0.036	-0.056	-0.006	-0.019	-0.103	-0.037	+1.000	+0.135	+0.114	-0.099	-0.041	-0.006	-0.016	+0.003	-0.001	-0.016	-0.009	-0.076
seeu	-0.043	-0.164	-0.177	+0.013	+0.024	-0.110	+0.012	+0.114	+1.000	+0.804	-0.119	-0.040	-0.123	+0.033	-0.065	-0.012	-0.020	+0.010	-0.027
seer	-0.087	-0.140	-0.127	+0.011	+0.024	-0.021	+0.031	+0.116	+0.804	+1.000	-0.127	-0.058	-0.149	+0.040	-0.022	-0.002	-0.019	-0.017	-0.013
skyu	-0.025	-0.011	+0.050	-0.002	+0.079	+0.137	+0.116	-0.099	-0.119	-0.127	+1.000	+0.492	+0.041	+0.008	-0.039	+0.000	+0.014	-0.026	+0.077
skyi	-0.069	-0.086	+0.005	-0.016	+0.073	+0.084	+0.102	-0.041	-0.040	-0.058	+0.492	+1.000	-0.151	-0.008	+0.010	+0.014	+0.031	-0.002	+0.069
mjd	+0.002	-0.006	+0.030	-0.039	-0.012	+0.013	+0.016	-0.006	-0.123	-0.149	+0.041	-0.151	+1.000	+0.019	-0.025	-0.041	-0.007	+0.042	-0.111
cam1	+0.017	-0.013	-0.026	-0.015	-0.010	+0.002	-0.018	-0.016	+0.033	+0.040	+0.008	-0.008	-0.008	+0.019	+1.000	+0.000	+0.000	+0.000	+0.026
cam2	-0.004	+0.024	+0.015	+0.065	-0.040	+0.067	-0.003	+0.003	-0.065	-0.022	-0.039	+0.010	-0.025	+0.000	+1.000	+0.000	+0.000	+0.000	-0.006
cam3	-0.010	+0.026	+0.005	+0.019	-0.003	-0.018	-0.016	-0.001	-0.012	-0.002	+0.014	-0.041	-0.000	+0.000	+1.000	+0.000	+0.000	+0.000	-0.015
cam4	+0.006	+0.009	+0.001	+0.021	-0.003	-0.050	+0.019	-0.016	-0.020	-0.019	+0.014	+0.031	-0.007	+0.000	+0.000	+0.000	+1.000	+0.000	+0.022
cam5	-0.008	-0.014	-0.022	-0.009	+0.004	-0.019	-0.020	-0.009	+0.010	-0.017	-0.026	-0.002	+0.042	+0.000	+0.000	+0.000	+0.000	+1.000	-0.001
ref	-0.092	+0.011	+0.006	+0.052	+0.035	+0.039	+0.031	-0.076	-0.027	-0.013	+0.077	+0.069	-0.111	+0.026	+0.006	+0.015	-0.022	+0.001	+1.000

ebv: 消光, 给出消光项  $E(B - V)$  的值, map 来自文献 [63];

star: 恒星污染, 恒星同样作为污染源被误认为星系导致的污染;

rstar: 红恒星污染, 和 star 类似, 但是限制  $g - r > 1.4$  而排出了蓝恒星;

ugr, gri, riz: 测光校准, ugr 代表色差, 即  $g - r$  和  $u - g$  的差值, 当然也有很多文献是直接给出  $u - g$ ,  $g - r$ ,  $r - i$ ,  $i - z$  的 map 来代表色差 [65];

uerr: 星系巡天选源的时候存在的一个统计误差, 它会导致某些不在巡天内的源偏移而进入了考察区域, 从而污染了星系巡天的源表, 同理, 也可能导致在巡天范围内的源偏移出了考察区域,

这个误差的 map 同样有 u, g, r, i, z 五个波段的数据, 此处 uerr 代表 u 波段;

airu: u 波段的大气质量, 大气质量是光子通过大气层时大气的影响, map 来源于 SDSS CAS [12];

seeu, seer: u 波段和 i 波段的视宁度, 视宁度是每个格点的点扩散函数的半高全宽, 同样它的 map 可以来自 SDSS CAS;

skyu, skyi: u 波段和 i 波段的视亮度, 顾名思义, 天光亮度是观测时大气的亮度, 同样它的 map 来自 SDSS CASS 服务器;

mjd: 约化儒略日, 这是一个天文上常用的时间单位制, 考虑它是为了排除观测时间带来的影响, 它的 map 同样来自 SDSS CAS 服务器;

cam1, cam2, cam3, cam4, cam5: SDSS 的 CCD 照相机, 用 CCD 拍全天的时候可能会有机器故障导致星系巡天的误差, 它的 map 同样来自 SDSS CAS;

ref: 大气反射, 大气层对光子的反射, 可能导致源的位置偏离, 从而导致密度场扭曲, 它来自 SDSS DAS 服务器。

UBV 测光系统的色指数,  $U$ 、 $B$ 、 $V$ 、 $R$ 、 $I$  是 UBV 滤镜系统定义的五个不同波段的测光强度量(可以类比 SDSS 巡天的  $u$ 、 $g$ 、 $r$ 、 $i$ 、 $z$  测光系统), 色指数可以表征红化强度, 对于银河系,  $R_V$  的标准值是  $3.1^{[62]}$ 。Schlegel 等人<sup>[63]</sup> 根据 COBE/DIRBE 的 map 和 IRAS/ISSA 的 map, 结合银道消光, 并移除了当时已经证认的恒星点源, 给出了一个  $100 \mu\text{m}$  波长的全天尘埃谱。Peek 等人<sup>[64]</sup> 在此基础上, 针对 SDSS 巡天给出一个修正后的北半球 map。

#### (2) 恒星点源污染

恒星作为点源, 很难和星系区分, 从而被星系巡天误认为是星系, 导致功率谱误差, 所以在算 LSS 功率谱的时候必须移除恒星点源。Ashley 等人<sup>[54]</sup> 详细描述了如何排除这种误差的方法, 主要是对不同波段, 恒星和星系表现出不同的特性和强度, 从而可以近似地区分出恒星。而所谓红恒星污染是限制  $g - r > 1.4$  而排出蓝恒星的恒星点源 map。

#### (3) 测光偏差

因为我们证认 LSS 巡天的示踪体时必须使用一个判据, 比如从 SDSS 巡天项目中挑选出类星体的巡天样本, 会用到一个被称为 CMASS 的选择方法<sup>[54]</sup>:  $17.5 < i_{\text{mod}} < 19.9$ ,  $r_{\text{mod}} - i_{\text{mod}} < 2$ ,  $d_{\perp} > 0.55$ ,  $i_{\text{fiber2}} < 21.7$ ,  $i_{\text{cmod}} < 19.86 + 1.6(d_{\perp} - 0.8)$ 。此时, 测光不准就会带来一些误差, 导致我们选的类型体样本是有误差的: 某些不在巡天内的源因偏移而进入了考察区域, 从而污染了星系巡天的源表, 同理, 也可能导致在巡天范围内的源因偏移而出了考察区域。测光色偏差的校准主要是用恒星的双色图(或称为双色指数图, color-color diagrams) 来直接校正或测试光学和红外巡天光谱的颜色和星等。双色图是一个比较恒星在不同波长下视星等差异的图表, 在天文学上不同波段间的光度差异称为色指数。类似, 测光流量偏差的校准将双色图的一个轴从色指数换成某波段内的测光强度即可。

#### (4) 大气质量

大气质量指天顶距(天体与天顶之间的角距离) 方向上大气光学厚度和天顶方向大气光学厚度的比值。地球上观测得到的电磁信息不可能都是从天顶方向, 即垂直于地平的天空过来的, 所以不同方向过来的电磁信息穿越大气的厚度也就不一致, 大气的影晌也各不同。

#### (5) 视宁度

视宁度是由地球大气引起的星像的亮度闪烁(当然观测室本身也会产生视宁度), 它主要体现在星像的形状破坏和位置变化。视宁度使焦面上的星像位置不断变化, 通过测量其变化可以反过来推算出视宁度的大小。

#### (6) 天空亮度

天光亮度是一种来自非信号源的背景误差信号, 用 CCD 做天体测光时, 不仅包含了天体的位置和亮度分布信息, 也包含了天光背景的贡献。天光背景来源很多, 包含大气本身的辉光、太阳光的散射、人造光污染等等。

#### (7) 约化儒略日

儒略日是一种时间单位制, 它不使用年和月, 是一种长期的纪日法。天文上常采用约化儒略日来记时。考虑儒略日的误差, 主要是为了排除观测时间不同带来的影响。

#### (8) 相机列

这个误差是用 CCD 拍全天的时候,可能会发生机器故障从而导致星系巡天的误差。对于每个巡天项目,这个误差是不同的,由于我们的类星体样本来自 SDSS 巡天,所以相机故障误差也必须从 SDSS 网站上得到。

### (9) 大气反射

大气反射是电磁波穿过大气时,被大气中的云层和较大尘埃将其一部分反射回宇宙空间去,从而削弱到达地面的辐射信息。

很多系统误差的信息来源于 SDSS 网站,例如文献 [20] 中,处理视宁度、大气质量、天光亮度、恒星污染等就是直接从 SDSS 网站上得到,然后可以像对待 CMB 巡天数据或者 LSS 巡天数据一样,将这些误差也通过 Healpix 格点化后对应到全天得到 map,同样也可以得到这些误差的功率谱,具体 SDSS 网站可以提供的误差信息, map 后有图 9。当然,还有很多系统误差也可以从别的研究中直接引用,例如消光是 Peek 和 Graves<sup>[64]</sup> 在 2010 年对 Schlegel 等人<sup>[63]</sup> 1998 年提供的尘埃 map 的一个修正 map,见图 10。

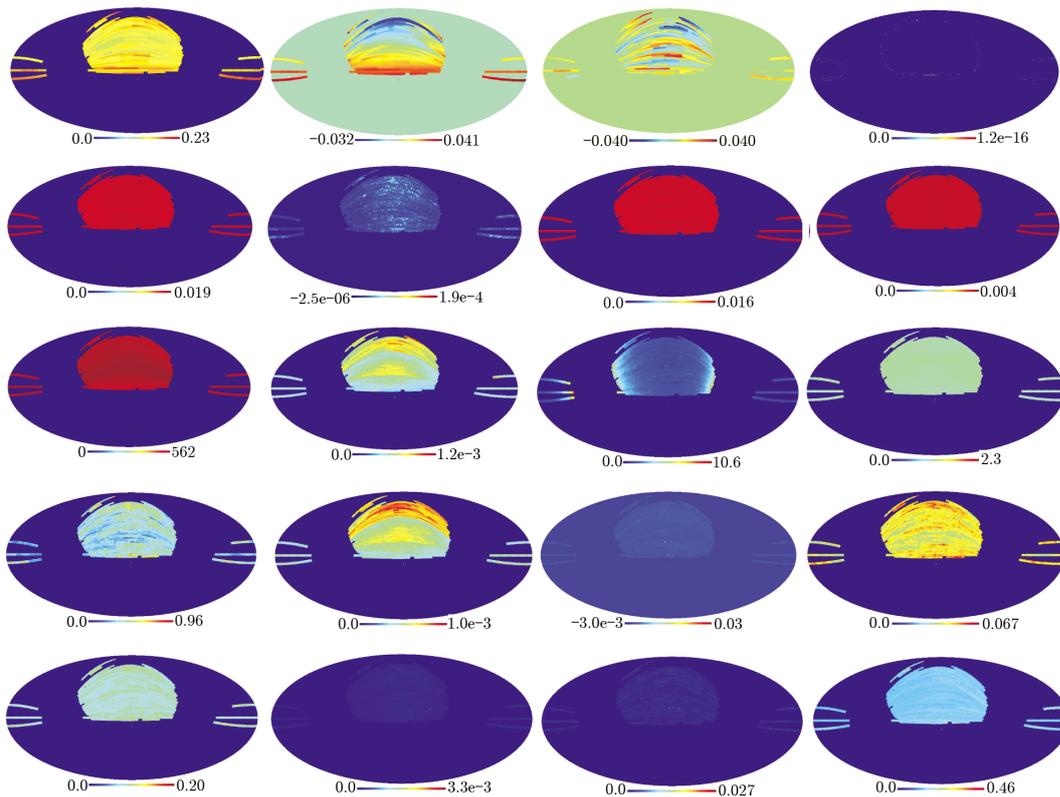


图 9 全天 map

从左上到右下依次对应 SDSS 的如下模板: airmass、ccCol、ccRow、ColOffset、expApCorrection、expApCorrectionErr、expModelApCorrection、gain、mjd、muErr、nBrightObj、nCR、nEffPsf、nuErr、psf2GB、psf2GSigma1、psf2GSigma2、psfApCorrectionErr、psfB、psfBeta。其中每个模板都有对应的 u、g、r、i、z 五个波段,此图只选取了 r 波段的情况

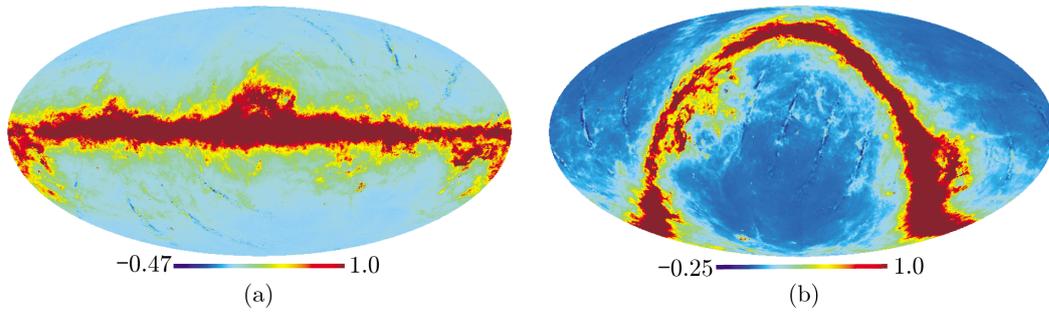


图 10 星际消光的 map  
(a)  $N_{\text{side}} = 128$ , 赤道坐标系; (b)  $N_{\text{side}} = 1024$ , 银道坐标系。

## 4.2 mask

有了系统误差的 map, 就可以通过 map 算出 mask 量。mask 本身实际上也是一个 map, 只不过这个特殊 map 每个格点的取值只有 0 和 1 两个数而已。对于考虑到系统误差模板的 map, 可以对不同的系统误差设定一个阈值, 当某个格点的对应误差值高于阈值时, 说明该成分在该格点的污染严重, 于是把它打上 mask, 不再考虑它对功率谱的贡献, 对应该格点的 mask 量为 0; 反之, mask 量为 1。当然 mask 阈值的选取是非常重要的: 阈值太小, 几乎没有考虑误差, 误差的影响仍然很大; 阈值太大, 就几乎把整个天区抹掉了, 计算功率谱的天区过小, 也会适得其反。如何选择需要考虑的误差, 并给出对应误差的阈值是非常重要的。Boris<sup>[20]</sup> 针对他们认为比较重要的五种误差模板——视宁度、消光、恒星、大气质量和天光亮度展开讨论, 给出了针对这五种误差模板的三种 mask 阈值选取方案及其对应的 mask 图景, 随后比较了这三种不同的阈值取值方法产生的 mask 对限制结果的影响, 并指出第三种经过改良后的 mask(如图 11) 的限制结果会更好。

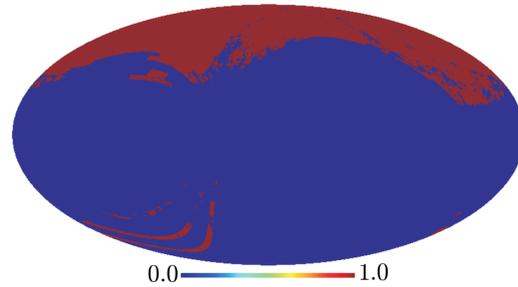


图 11 根据 Boris<sup>[20]</sup> 的第三种 mask 阈值选取方法, 扣除视宁度、消光、恒星、大气质量和天光亮度五种误差模板后得到的 mask 图

## 4.3 如何消除 LSS 系统误差

### 4.3.1 model projection

model projection 方法最早是 Tegmark 等人<sup>[35, 59]</sup>, Bond 等人<sup>[29]</sup>, Halverson 等人<sup>[61]</sup> 对 CMB 的数据进行系统误差消除时而提出来的, 把它借鉴到 LSS 数据上, 例如类星体的 map 为  $x$ , 它是由真实的信号和噪声组成的, 所以  $x^{\text{obs}} = x^{\text{true}} + \sum_{i=\text{sys}} \lambda_i \Phi_i$ , 于是对 map 做相关统计, 有  $C = \langle x^{\text{obs}} x^{\text{obs}} \rangle = C^{\text{true}} + \sum_{i=\text{sys}} \eta_i \Phi_i \Phi_i^T$ , 其中信号和噪声无关联, 所以他们的交叉相关函数为零。

不论 LSS 还是 CMB 的功率谱都依赖于很多宇宙学参数  $a_p$ , 这也是可以用 LSS 和 CMB 的数据来限制非高斯性系数的原因。利用功率谱对任意参数  $a_p$  的似然函数可以表示为<sup>[29]</sup>:  $\mathcal{L}_\Delta(a_p) = P(\Delta|a_p) = \frac{1}{(2\pi)^{N/2}|C_T(a_p) + C_N|^{1/2}} \times \exp\left\{-\frac{1}{2}\Delta^T[C_T(a_p) + C_N]^{-1}\Delta\right\}$ , 其中  $C_T(a_p)$  是信号的谱  $C^{\text{true}}$ ,  $C_N$  是噪声的谱  $\sum_{i=\text{sys}} \eta_i \Phi_i \Phi_i^T$ 。将似然函数做泰勒展开有  $\ln\mathcal{L}(a + \delta a) = \ln\mathcal{L}(a) + \sum_p \frac{\partial \ln\mathcal{L}(a)}{\partial a_p} \delta a_p + \frac{1}{2} \sum_{pp'} \frac{\partial^2 \ln\mathcal{L}(a)}{\partial a_p \partial a_{p'}} \delta a_p \delta a_{p'}$ , 高阶效应忽略, 后两项便是常用的宇宙学统计工具, 其中第三项, 即似然函数对  $a_p$  展开的二阶效应便是熟知的费舍尔矩阵。

model projection 提出, 若使  $\eta \rightarrow \infty$ , 此时噪声的功率谱  $C_N \rightarrow \infty$ , 于是似然函数的泰勒展开式中的后两项为零, 所以有  $\ln\mathcal{L}(a + \delta a) = \ln\mathcal{L}(a)$ , 因此无论该系统误差的模板怎么变化, 都影响不了对宇宙学参数  $a_p$  的限制, 于是就消除了误差。

#### 4.3.2 新方法

Shirley 等人<sup>[21]</sup> 提出了一种新的方法来消除系统误差。

由 4.3.1 节可知:  $x^{\text{obs}} = x^{\text{true}} + \sum_{i=\text{sys}} \lambda_i \Phi_i$ ,  $i$  代表需要考虑的误差数目, 假设  $i = 2$ , 则对  $x^{\text{obs}}$  做自相关函数有:  $\langle x^{\text{obs}} x^{\text{obs}} \rangle = \langle x^{\text{true}} x^{\text{true}} \rangle + \lambda_1^2 \langle \Phi_1 \Phi_1 \rangle + \lambda_2^2 \langle \Phi_2 \Phi_2 \rangle + 2\lambda_1 \lambda_2 \langle \Phi_1 \Phi_2 \rangle$ , 对  $x^{\text{obs}}$  分别与两种误差做交叉相关, 有:

$$\begin{cases} \langle x^{\text{obs}} \Phi_1 \rangle = \lambda_1 \langle \Phi_1 \Phi_1 \rangle + \lambda_2 \langle \Phi_1 \Phi_2 \rangle, \\ \langle x^{\text{obs}} \Phi_2 \rangle = \lambda_2 \langle \Phi_2 \Phi_2 \rangle + \lambda_1 \langle \Phi_1 \Phi_2 \rangle. \end{cases}$$

其中真实的信号假设是绝对“干净”的, 已经不再包含误差, 所以  $\langle x^{\text{true}} \Phi_i \rangle$  项为零。  $\langle x^{\text{obs}} \Phi_i \rangle$  和  $\langle \Phi_i \Phi_i \rangle$  是巡天项目可以给出的, 所以解这个二元二次方程组, 可以得到误差的系数  $\lambda_i$ , 代入到自功率谱  $\langle x^{\text{obs}} x^{\text{obs}} \rangle$  的公式中, 即可得到真实功率谱  $\langle x^{\text{true}} x^{\text{true}} \rangle$ 。

## 5 总结与展望

CMB 的各向异性和宇宙物质的成团性表明宇宙存在非高斯现象, 这个非高斯现象是由暴胀引起的, 所以对 CMB 和 LSS 的巡天观测以及对他们的统计研究可以帮助我们更好地认识非高斯性和暴胀模型。现阶段对非高斯性的限制实验种类很多, 本文比较详细地介绍了如何利用 CMB 的三点相关函数或双谱来限制非高斯系数, 以及如何利用大尺度结构的两点相关特性, 通过限制实际宇宙偏离高斯宇宙的偏移量来间接限制非高斯系数。但是在这个过程中往往不可避免地伴随着系统误差的影响, 最近的研究通过模式辐射方法以及一种新方法消除误差带来的影响, 从而得到比较真实的功率谱。

关于用 CMB 和 LSS 来限制非高斯的相关研究, 从 20 世纪起至今, 在继续并会深入进行下去。造成这方面研究困难的重要原因主要如下: (1) LSS 巡天项目的覆盖天区和深度不够, 这直接导致了对非高斯的认识比较局限; (2) 系统误差对 CMB 和 LSS 实验的严重污染,

这和第一条类似, 只能将系统误差污染严重的部分打上 mask 而不再讨论, 事实上这也就导致了比较局限地去认识非高斯性, 并且系统误差也是依赖巡天项目的; (3) 用来消除误差的方法, 或多或少用到一些近似条件, 这同样会使得对非高斯的限制产生偏差。

未来随着技术的进步, 各个波段大巡天项目的运行, 由巡天项目数据得到的功率谱以及噪声的功率谱会更加准确可靠; 同时, 相信更好消除误差的方法也会产生。对非高斯的限制会越来越精确, 宇宙的演化史随之也会越来越精确。

#### 参考文献:

- [1] Guth A H, So-Young Pi. PRL, 1982, 49: 1110
- [2] Guth A H. PRD, 1981, 23: 347
- [3] 汪浩. 博士论文. 北京: 北京师范大学, 2012: 101
- [4] Antonio R. <http://arxiv.org/pdf/hep-ph/0210162v1.pdf>, hep-ph/0210162, 2002
- [5] Lotfi B, et al. PRD, 2006, 73: 021301
- [6] David H L, et al. JCAP, 2005, 0510: 005
- [7] James M B. PRD, 1980, 22: 1882
- [8] Eiichiro K, et al. ApJS, 2011, 192: 18
- [9] Bennett C L, Larson D, et al. ApJS, 2013, 208: 2
- [10] Ade P A R, et al. A&A, 2014, 571: A24
- [11] Anze S, Christopher H, et al. JCAP, 2008, 08: 031
- [12] York D G, et al. AJ, 2000, 120: 1579
- [13] Jennifer K A, et al. ApJS, 2007, 172: 634
- [14] Xia J Q, et al. JCAP, 2011, 8: 33
- [15] Xia J Q, et al. ApJL, 2010, 717: L17
- [16] Jennifer K A, et al. ApJS, 2008, 175: 297
- [17] Gordon T R, Adam D M, et al. ApJS, 2009, 180: 67
- [18] Stefano C, Mário G S. MNRAS, 2015, 448: 1035
- [19] Pullen A R, Hirata C M. PASP, 2013, 125: 705
- [20] Boris L, et al. MNRAS, 2013, 435: 1857
- [21] Shirley H, Antonio C, et al. ApJ, 2012, 761: 1
- [22] Daniel B, Paolo C, et al. JCAP, 2004, 0408: 009
- [23] Gangui, et al. ApJ, 1994, 430: 447
- [24] Verde, et al. MNRAS, 2000, 313: 141
- [25] Wang B, et al. PRD, 2000, 61: 063504
- [26] Ashley J R, et al. MNRAS, 2012, 482: 1116
- [27] Maldacena J. Journal of High Energy Physics, 2003, 5: 13
- [28] Chen X, et al. JCAP, 2007b, 1: 2
- [29] Bond J, Jaffe A, et al. PRD, 1998, 57: 2117
- [30] Eric H, Krzysztof M G, et al. ApJ, 2002, 567: 2
- [31] Wandelt B D, et al. PRD, 2001, 64: 083003
- [32] Efstathiou G. MNRAS, 2004b, 349: 603
- [33] Brown M L, et al. MNRAS, 2005, 360: 1262
- [34] Pontzen A, Peiris H V. PRD, 2010, 81: 103008
- [35] Tegmark M. PRD, 1997, 55: 10
- [36] Krzysztof M G, Benjamin D W, et al. <http://healpix.sourceforge.net/>, 2015

- [37] Mark R C. <http://arxiv.org/pdf/astro-ph/0412607v1.pdf>, astro-ph/0412607, 2004
- [38] Gorski K M, Hivon E, et al. ApJ, 2005, 622: 759
- [39] Istvan S, Simon P, et al. ApJ, 2001, 548: L115
- [40] Laurence P, Julien L, et al. JCAP, 2006, 10: 013
- [41] Antony L, Sarah B, PRD. 2002, 66:103511
- [42] Shirley H, Christopher M H, et al. PRD, 2008, 78: 043519
- [43] Tommaso G, Robert C, et al. MNRAS, 2012, 426: 2581
- [44] Dalal N, et al. PRD, 2008, 77: 123514
- [45] Valageas P, A&A, 2014, 514: 46
- [46] Giannantonio T, Porciani C. PRD, 2010, 81: 063530
- [47] Matarrese S, Verde L. ApJL, 2008, 677: 77
- [48] Afshordi N, Tolley A J. PRD, 2008, 78: 123507
- [49] Schmidt F, Kamionkowski M. PRD, 2010, 82: 103002
- [50] Desjacques V, Jeong D, et al. PRD, 2011, 84: 061301
- [51] Smith K M, Ferraro S, et al. JCAP, 2012, 3: 032
- [52] Cole S, Kaiser N. MNRAS, 1989, 18: 1127
- [53] Tinker J, Kravtsov, et al. ApJ, 2008, 688: 709
- [54] Ashley J R, Shirley H, et al. MNRAS, 2011, 417: 1350
- [55] Shirley H, et al. JCAP, 2013, 5: 40
- [56] Gordon T R, et al. ApJS, 2004, 155: 257
- [57] Andrew J S H, Tegmark M. MNRAS, 2004, 349:115
- [58] Swanson M E C, Tegmark M, et al. MNRAS, 2008, 387: 1391
- [59] Tegmark M, et al. ApJ, 1998, 499: 555
- [60] Scott D, eds. *Morden Cosmology*, Singapore: Elsevier, 2003
- [61] Halverson N, et al. ApJ, 2002, 568: 38
- [62] Schultz D, Wiemer W, A&A, 1975, 43: 133
- [63] Schlegel D J, Finkbeiner D P, et al. ApJ, 1998, 500: 525
- [64] Peek J E G, Graves G J. ApJ, 2010, 719: 415
- [65] Schlafly E F, Finkbeiner D P, et al. ApJ, 2010, 725: 1175
- [66] Nishant A, Shirley H, et al. JCAP, 2014, 04: 007
- [67] Tommaso G, Ashley J R, et al. PRD, 2014, 89: 023511
- [68] Xia J Q, Matteo V, et al. JCAP, 2009, 0909: 003
- [69] Cullan H, Ashley J R. MNRAS, 2014, 449: 848
- [70] Marc M, Lado S. MNRAS, 2015, 447: 437
- [71] Mariana V M, Shirley H, et al. <http://arxiv.org/pdf/1312.4996v1.pdf>, arXiv: 1312.4996
- [72] Gianmassimo T, Matteo T. JCAP, 2014, 03: 032
- [73] Fernando d S, Flavia S, et al. MNRAS, 2013, 435: 4017
- [74] Carlos H M, Ashley J R, et al. MNRAS, 2014, 438: 1724
- [75] Donough R, Dipak M. MNRAS, 2015, 448: 2232
- [76] Komatsu E, Spergel D. PRD, 2011, 63: 063002
- [77] Komatsu E, et al. ApJS, 2009, 180:330
- [78] Steve W, ed. *Cosmology*, England: Oxford University Press, 2008
- [79] Richards G T, et al. MNRAS, 2005, 360: 839
- [80] Afshordi N, Loh Y S, et al. PRD, 2004, 69: 083524
- [81] Peiris H V, Spergel D N. ApJ, 2000, 540: 605
- [82] Cooray A. PRD, 2002, 65: 103510
- [83] Ade P A R, et al. A&A, 2014, 571: A1
- [84] Jaiseung K, et al. JCAP, 2013, 04: 021
- [85] Holman R, Andrew J T. JCAP, 2008, 0805: 001

[86] Liddle A R, Lyth D H. *Cosmological Inflation and Large-Scale Structure*. Cambridge: Cambridge University Press, 2000: 414

## Systematic Effects in Large Scale Angular Power Spectra and Implications for Constraining Primordial Non-gaussianity

YANG Sheng

*(Department of Astronomy, Beijing Normal University, Beijing 100875, China)*

**Abstract:** The cosmic microwave background radiation (CMBR) detection and the large scale structure (LSS) survey are two important tools to study on the evolution of the universe. The CMB experiment and the LSS detector could calculate and restrict on the cosmological parameters with very high precision, such as the curvature of the universe, the parameter of the inflation models and so on. The COBE satellite first showed the Non-Gaussianity among the observed CMB photons that in Gaussian distribution. The Primordial Non-Gaussianity (PNG) of the cosmic large scale structure explains the physical mechanism why there should exists a primordial perturbation and how it forms the large scale structure today in the very early universe. Also, it can help researcher to distinguish between different inflation models, and it is an effective tool for us to study very early universe. But recent research papers have shown that the large scale structure survey's data may be polluted badly by some systematic errors, such as seeing, airmass, sky brightness and so on, which would cause a big difference between the derived cosmic parameters and the actual counterparts. Therefore removing the systematic errors' effect for the LSS tracers on constraining primordial non-gaussianity is very meaningful so that we could get more real cosmic truth and know the early cosmology more clearly.

In this paper, we will review the way how the CMB and LSS tracer's maps restrict on the cosmological parameters firstly and then we will discuss about the systematic errors' effect on the observed angular power spectrum and also we introduce two important methods to remove the systematics. Finally, we will compare the modified parameter and the original one to discuss the goodness.

**Key words:** Primordial Non-Gaussianity; power spectrum; inflation; large scale structure; systematic errors