SUBMITTED TO APJ AUGUST 31 1998; ACCEPTED FEBRUARY 8
Preprint typeset using LATEX style emulateapj

COMPARING AND COMBINING CMB DATASETS

MAX TEGMARK[a,b]

[a]Institute for Advanced Study, Princeton, NJ 08540; max@ias.edu
[b]Hubble Fellow

Submitted to ApJ August 31 1998; accepted February 8

ABSTRACT

One of the best ways of spotting previously undetected systematic errors in CMB experiments is to compare two independent observations of the same region. We derive a set of tools for comparing and combining CMB data sets, applicable also in the common case where the two have different resolution or beam shape and therefore do not measure the same signal. We present a consistency test that is better than a χ^2-test at detecting systematic errors. We show how two maps of different angular resolution can be combined without smoothing the higher resolution down to the lower one, and generalize this to arbitrary beam configurations. We also show how lossless foreground removal can be performed even for foreground models involving scale dependence, latitude dependence and spectral index variations in combination.

Subject headings: cosmic microwave background — methods: data analysis

1. INTRODUCTION

Undetected systematic errors are one of the main obstacles to using cosmic microwave background (CMB) measurements to constrain cosmological models. One of the best ways to address this problem is to compare experiments whose data sets overlap both in sky coverage and angular scale, to see whether they are consistent.[1] If they are consistent, a useful second step is to combine them into a single sky map retaining all their cosmological information.

Both of these steps are simple for maps with identical resolution and beam shape. For *e.g.* the 53 and 90 GHz COBE DMR maps (Bennett *et al.* 1996), comparing (1) was done by subtracting the maps and checking whether the difference was consistent with pure noise, whereas combining (2) was done by simply averaging the two maps, weighting pixels by their inverse variance. Unfortunately, both steps are usually more complicated. MAP, Planck and most current experiments have different angular resolution in different channels. Many current experiments probe the sky in an even more complicated way, with *e.g.* double beams, triple beams, interferometric beams or complicated elongated software-modulated beams. Correlated noise further complicates the problem.

Despite these difficulties, precision comparisons between different experiments are crucial. Some of the best evidence so far for detection of CMB fluctuations comes from the success of such comparisons in the past — between FIRS and DMR (Ganga *et al.* 1993), Tenerife and DMR (Lineweaver *et al.* 1995), MSAM and Saskatoon (Knox *et al.* 1998; hereafter K98), two years of Python data (Ruhl *et al.* 1995), three years of Saskatoon data (Tegmark *et al.* 1996a) and two flights of MSAM (Inman *et al.* 1997). The fact that many of these non-COBE data sets were contaminated by systematic errors made the success of these cross-checks even more encouraging.

Data sets are currently growing rapidly in number, size and quality, often overlapping. It is therefore quite timely to develop methods that generalize both steps (1) and (2) to arbitrary experiments. This is the purpose of the present *Letter*. In the larger context of CMB data analysis, this is important between the steps of mapmaking (Wright *et al.* 1996; Wright 1996; Tegmark 1997a) and power spectrum estimation (Tegmark 1997b; Bond *et al.* 1998) in the pipeline.

2. NOTATION

Let us first establish some notation that will be used throughout this paper. Consider a pixelized CMB sky map at some resolution consisting of m numbers $x_1, ..., x_m$, where x_i is the temperature in the i^{th} pixel. Suppose two experiments $i = 1, 2$ have measured n_i numbers $y_1, ..., y_{n_i}$, each probing some linear combination of the sky temperatures x_i. Grouping these numbers into vectors $\mathbf{x}$, $\mathbf{y}_1$ and $\mathbf{y}_2$ of length m, n_1 and n_2, we can generally write[2]

$$\mathbf{y}_1 = \mathbf{A}_1\mathbf{x} + \mathbf{n}_1, \quad \mathbf{y}_2 = \mathbf{A}_2\mathbf{x} + \mathbf{n}_2 \qquad (1)$$

for some known matrices $\mathbf{A}_i$ incorporating the beam shapes and some random noise vectors $\mathbf{n}_i$ with zero mean ($\langle \mathbf{n}_i \rangle = \mathbf{0}$). We will refer to $\mathbf{x}$ as the "true sky". $\mathbf{y}_1$ and $\mathbf{y}_2$

[1]The best way to address the problem is clearly to design CMB experiments to be more immune to systematic errors in the first place. The next best thing to do is carefully examine the raw time-ordered data from an experiment for specific forms of systematic errors that may be expected, and for general signs of systematic errors for data removal or correction. The cross-check between experiments that are discussed in this paper are by no means a substitute for this, but rather a way of catching additional systematic errors that have slipped through the cracks and not been detected by the team that reduced the raw data.

[2]This assumes both that the experimental data have perfect linearity, and that there are no non-zero experimental offsets. Ideally, experimentalists should model and remove both nonlinearities and offsets as part of their data reduction, thereby making equation (1) applicable to their final data product. If offsets of an unknown amplitude remain nonetheless, multiplication of a data set $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} + offsets$ by a matrix $\mathbf{P}$ that projects out these offsets will produce a new data set $\mathbf{y}_1 = \mathbf{P}\mathbf{y}$ that satisfies equation (1), merely with the slightly more complicated noise vector $\mathbf{n}_1 = \mathbf{P}\mathbf{n}$ and with $\mathbf{A}_1 = \mathbf{P}\mathbf{A}$. A detailed example of this procedure can be found in de Oliveira *et al.* (1998).

can be either time-ordered data or some linear combination thereof, for instance pixelized maps. It is sometimes useful to define the larger matrices and vectors

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}, \quad \mathbf{y} \equiv \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \mathbf{n} \equiv \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{pmatrix}, \tag{2}$$

which allows us to rewrite equation (1) as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \tag{3}$$

Let us write the noise covariance matrix as

$$\mathbf{N} \equiv \langle \mathbf{n}\mathbf{n}^t \rangle = \begin{pmatrix} \mathbf{N}_1 & \mathbf{N}_{12} \\ \mathbf{N}_{12}^t & \mathbf{N}_2 \end{pmatrix}, \tag{4}$$

where $\mathbf{N}_1 \equiv \langle \mathbf{n}_1\mathbf{n}_1^t \rangle$, $\mathbf{N}_2 \equiv \langle \mathbf{n}_2\mathbf{n}_2^t \rangle$ and $\mathbf{N}_{12} \equiv \langle \mathbf{n}_1\mathbf{n}_2^t \rangle$.

We derive useful consistency tests in §2 both for the special case of identical observations ($\mathbf{A}_1 = \mathbf{A}_2$) and for the general case $\mathbf{A}_1 \neq \mathbf{A}_2$, then show how to combine data sets without destroying information in §3.

3. COMPARING DATA SETS: ARE THEY CONSISTENT?

3.1. *The "null-buster" test*

Let us first consider the simple case where the two data sets measure the same thing, *i.e.*, $\mathbf{A}_1 = \mathbf{A}_2$. This often applies for two different channels of the same experiment at the same frequency. We can then form a difference map $\mathbf{z} \equiv \mathbf{x}_1 - \mathbf{x}_2$, which in the absence of systematic errors should consist of pure noise.

Consider the null hypothesis H_0 that such a data set $\mathbf{z}$ consists of pure noise, *i.e.*, $\langle \mathbf{z} \rangle = \mathbf{0}$, $\langle \mathbf{z}\mathbf{z}^t \rangle = \mathbf{N}$ for some noise covariance matrix $\mathbf{N}$. Suppose we have reason to suspect that the alternative hypothesis H_1 is true, where $\langle \mathbf{z} \rangle = \mathbf{0}$, $\langle \mathbf{z}\mathbf{z}^t \rangle = \mathbf{N} + \mathbf{S}$ for some signal covariance matrix $\mathbf{S}$, and want to try to rule out H_0 by using a test statistic q that is a quadratic function of the data:

$$q \equiv \mathbf{z}^t\mathbf{E}\mathbf{z}^t = \mathrm{tr}\,\{\mathbf{E}\mathbf{z}\mathbf{z}^t\}. \tag{5}$$

Depending on whether H_0 or H_1 is true, the mean of q will be $\langle q \rangle_0 = \mathrm{tr}\,\mathbf{EN}$ or $\langle q \rangle_1 = \mathrm{tr}\,\mathbf{EN} + \mathrm{tr}\,\mathbf{ES}$, respectively. If $\mathbf{z}$ has a multivariate Gaussian probability distribution[3], then q will have a variance $(\Delta q)^2 = 2\,\mathrm{tr}\,\mathbf{ENEN}$ if H_0 is true. Therefore the quantity $\nu \equiv (q - \langle q \rangle_0)/\Delta q$ gives the number of standard deviations ("sigmas") by which the observed q-value exceeds the mean expected under the null hypothesis. If we observe $\nu \gg 1$, we can thus conclude that H_0 is ruled out at high significance. Which choice of $\mathbf{E}$ has the greatest statistical power to reject H_0 if H_1 is true, *i.e.*, which $\mathbf{E}$ maximizes the expectation value

$$\langle \nu \rangle \equiv \frac{\langle q \rangle_1 - \langle q \rangle_0}{\Delta q} = \frac{\mathrm{tr}\,\mathbf{ES}}{[2\,\mathrm{tr}\,\mathbf{ENEN}]^{1/2}}? \tag{6}$$

Since rescaling $\mathbf{E}$ by a constant leaves $\langle \nu \rangle$ invariant, let us for simplicity normalize $\mathbf{E}$ so that the denominator equals unity. We thus want to maximize $\mathrm{tr}\,\mathbf{ES}$ subject to the constraint that $\mathrm{tr}\,\mathbf{ENEN} = 1/2$. Using the method of Lagrange multipliers with $L = \mathrm{tr}\,\mathbf{ES} - \lambda \mathrm{tr}\,\mathbf{ENEN}/2$ and differentiating L with respect to the components of $\mathbf{E}$, this gives the solution $\mathbf{E} \propto \mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}$. This leaves us with our optimal "null-buster" statistic[4]

$$\nu \equiv \frac{\mathbf{z}^t\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}\mathbf{z} - \mathrm{tr}\,\mathbf{N}^{-1}\mathbf{S}}{[2\,\mathrm{tr}\,\{\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}\mathbf{S}\}]^{1/2}}, \tag{7}$$

which will rule out the null hypothesis H_0 with the largest average significance if H_1 is true. For the special case $\mathbf{S} = \mathbf{N}$, we see that this reduces to a standard χ^2-test with $\nu = (\chi^2 - n)/\sqrt{2n}$, $\chi^2 \equiv \mathbf{z}^t\mathbf{N}^{-1}\mathbf{z}$. Whenever we have reason to suspect systematic errors of a certain form (producing a signal $\propto \mathbf{S}$), the null-buster test will thus be more sensitive to these systematic errors than the χ^2-test, which is a general-purpose tool. This issue is elaborated in K98, which also provides a useful general discussion of consistency tests.

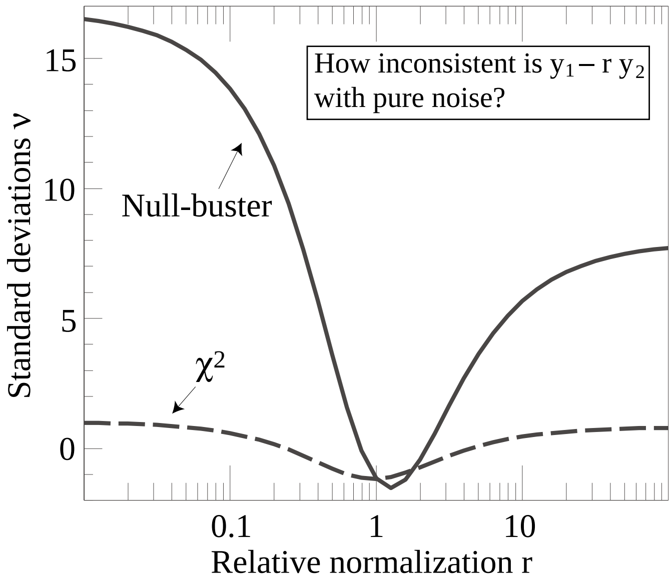

FIG. 1.— The null-buster test against calibration errors was applied to the QMAP experiment (see de Oliveira-Costa et al. 1998). The figure (from Devlin et al. 1998) shows the number of σ at which signal is detected in the weighted flight 1 Ka-band difference maps $\mathbf{y}_1 - r\mathbf{y}_2$. The χ^2-test is seen to be weaker.

3.2. *An example: calibration errors*

The null-buster is useful for comparing two CMB maps $\mathbf{y}_1$ and $\mathbf{y}_2$ that have the same shape, beam size and pixelization. Let $\mathbf{S}$ denote the expected contribution to $\mathbf{y}_i{\mathbf{y}_i}^t$ from CMB fluctuations, *i.e.*, $\mathbf{S} = \mathbf{A}_1\mathbf{C}\mathbf{A}_1^t = \mathbf{A}_2\mathbf{C}\mathbf{A}_2^t$, where $\mathbf{C}$ is the map covariance matrix

$$\mathbf{C}_{ij} = \sum_{l=2}^{\infty} \left(\frac{2l+1}{4\pi}\right) P_l(\widehat{\mathbf{r}}_i \cdot \widehat{\mathbf{r}}_j) C_l; \tag{8}$$

[3]We will only make the assumption of Gaussianity for the CMB and the detector noise, not for the systematic errors. In fact, systematics, such as foreground signals, data spikes and atmospherics seldom have a Gaussian probability distribution, but they vanish under the null hypothesis that we are trying to rule out.

[4]K98 discuss a test using the likelihood ratio, which is the best solution to a slightly different problem. Translated into our notation, it corresponds to the choice $\mathbf{E} = \mathbf{N}^{-1} - [\mathbf{N}+\mathbf{S}]^{-1} = \mathbf{N}^{-1}\mathbf{S}[\mathbf{N}+\mathbf{S}]^{-1}$. Note that whereas equation (7) is independent of the normalization of $\mathbf{S}$ (the "shape" of the signal matters, but not its amplitude), the likelihood ratio test requires an assumed amplitude.

the unit vector $\widehat{\mathbf{r}}_i$ gives the direction towards the i^{th} pixel and C_l is the expected or observed CMB power spectrum (the normalization of C_l is irrelevant here — only the shape matters). Now suppose one or both of the data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ have a linear calibration error, *i.e.*, are off by some constant multiplicative factors. Consider a difference map of the form $\mathbf{z} \equiv \mathbf{y}_1 - r\mathbf{y}_2$ for some factor r, and plot ν as a function of r using equation (7) with the null hypothesis being that $\mathbf{z}$ is pure noise, *i.e.*, that $\mathbf{N} \equiv \mathbf{N}_1 + r[\mathbf{N}_{12} + \mathbf{N}_{12}^t] + r^2\mathbf{N}_2$. An example is shown in Figure 1. If $\nu \gg 1$ for $r = 1$, then we have a significant detection of signal not common to the two maps. If $\nu(1) \gg 1$ but $\nu(r) \lesssim 1$ for some other r-value, this would show that there is a relative error of r in the normalization between the two maps. If the maps are at different frequencies, this could also indicate that they are dominated by foreground contamination with a frequency dependence different from the CMB.

3.3. *If the beam shape differs*

Above we found the null-buster to be a useful consistency test when applied to $\mathbf{z} \equiv \mathbf{y}_1 - r\mathbf{y}_2$, since $\mathbf{A}_1 = \mathbf{A}_2$ implied that $\mathbf{y}_1 - \mathbf{y}_2$ should consist of mere noise and be independent of the (a priori unknown) signal. If $\mathbf{A}_1 \neq \mathbf{A}_2$, this is no longer true, since the two data sets are not measuring the same thing. We therefore perform our null-buster test with the difference map redefined to be

$$\mathbf{z} \equiv \mathbf{A}_3\mathbf{y}_1 - r\mathbf{A}_4\mathbf{y}_2 \tag{9}$$

for some matrices $\mathbf{A}_3$ and $\mathbf{A}_4$, and choose these matrices so that the new data vectors $\mathbf{A}_3\mathbf{y}_1$ and $\mathbf{A}_4\mathbf{y}_2$ measure at least approximately the same sky signal. Substituting equation (1) into equation (9) shows that this corresponds to the requirement

$$\mathbf{A}_3\mathbf{A}_1 \approx \mathbf{A}_4\mathbf{A}_2, \tag{10}$$

which would make $\mathbf{z}$ approximately signal-independent for $r = 1$. In addition to equation (10), we clearly want $\mathbf{A}_3\mathbf{A}_1$ and $\mathbf{A}_4\mathbf{A}_2$ to have as large rank as possible, to avoid destroying more information than necessary (otherwise say $\mathbf{A}_3 = \mathbf{A}_4 = \mathbf{0}$ would do the trick).

If one data set can be written as a linear combination of the other (say $\mathbf{A}_2 = \mathbf{M}\mathbf{A}_1$ for some matrix $\mathbf{M}$), then the best choice is clearly $\mathbf{A}_3 = \mathbf{M}$, $\mathbf{A}_4 = \mathbf{I}$. This is the case for identically sampled maps where $\mathbf{y}_1$ has higher resolution than $\mathbf{y}_2$, as well as for cases where a map (say DMR or QMAP) is compared with more complicated weighted averages (say Saskatoon) of the same sky region at the same or lower resolution. This approach was adopted by, *e.g.*, Lineweaver *et al.* (1995), for comparing the (smoothed) Tenerife data to DMR.

We will now tackle the problem for the general case. Our solution (there may be others) involves performing a signal-to-noise eigenmode analysis (Bond 1994; Bunn & Sugiyama 1995; Tegmark *et al.* 1996b) three times, in an unusual way. We will first present this procedure with no derivation, then show that it solves our problem. We start by solving the generalized eigenvalue problems

$$\begin{aligned} \left[\mathbf{A}_1\mathbf{C}_1\mathbf{A}_1^t\right]\mathbf{B}_1 &= \mathbf{N}_1\mathbf{B}_1\mathbf{\Lambda}_1, & (11)\\ \left[\mathbf{A}_2\mathbf{C}_2\mathbf{A}_2^t\right]\mathbf{B}_2 &= \mathbf{N}_2\mathbf{B}_2\mathbf{\Lambda}_2, & (12)\end{aligned}$$

where the eigenvectors are the columns of the matrices $\mathbf{B}_1$ and $\mathbf{B}_2$, normalized so that $\mathbf{B}_i^t\mathbf{N}_i\mathbf{B}_i = \mathbf{I}$, and the corresponding eigenvectors are elements of the diagonal matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$, sorted in decreasing order. We then reduce the width of $\mathbf{B}_1$ and $\mathbf{B}_2$ by throwing away all eigenvectors with eigenvalues below some cutoff λ_{min}, and define a new smaller data set

$$\tilde{\mathbf{y}} = \begin{pmatrix} \mathbf{B}_1^t\mathbf{y}_1 \\ \mathbf{B}_2^t\mathbf{y}_2 \end{pmatrix}. \tag{13}$$

This will have the covariance matrix $\langle\tilde{\mathbf{y}}\tilde{\mathbf{y}}^t\rangle = \tilde{\mathbf{S}} + \tilde{\mathbf{N}}$, where

$$\begin{aligned} \tilde{\mathbf{S}} &= \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{B}_1^t\mathbf{A}_1\mathbf{C}\mathbf{A}_2^t\mathbf{B}_2 \\ \mathbf{B}_2^t\mathbf{A}_2\mathbf{C}\mathbf{A}_1^t\mathbf{B}_1 & \mathbf{\Lambda}_2 \end{pmatrix}, & (14)\\ \tilde{\mathbf{N}} &= \begin{pmatrix} \mathbf{I} & \mathbf{B}_1^t\mathbf{N}_{12}\mathbf{B}_2 \\ \mathbf{B}_2^t\mathbf{N}_{12}^t\mathbf{B}_1 & \mathbf{I} \end{pmatrix}. & (15)\end{aligned}$$

We then solve the generalized eigenvalue problem

$$\tilde{\mathbf{S}}\mathbf{B} = \tilde{\mathbf{N}}\mathbf{B}\mathbf{\Lambda}, \tag{16}$$

where the eigenvectors are the columns of the matrix $\mathbf{B}$, normalized so that $\mathbf{B}^t\tilde{\mathbf{N}}\mathbf{B} = \mathbf{I}$, and the corresponding eigenvectors are elements of the diagonal matrix $\mathbf{\Lambda}$, sorted in *increasing* order. Finally, we reduce the width of $\mathbf{B}$ by throwing away all eigenvectors with eigenvalues *above* some cutoff λ_{max}, leaving us with a matrix of the form

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_3 \\ \mathbf{B}_4 \end{pmatrix}. \tag{17}$$

By choosing a tiny cutoff (say $\lambda_{max} = 10^{-2}$), we ensure that the transformed data vector $\mathbf{B}^t\tilde{\mathbf{y}}$ is completely dominated by detector noise, with a for all practical purposes negligible CMB signal. This means that

$$\mathbf{B}^t\tilde{\mathbf{y}} = [\mathbf{B}_3^t\mathbf{B}_1^t\mathbf{A}_1 + \mathbf{B}_4^t\mathbf{B}_2^t\mathbf{A}_2]\mathbf{x} \approx \mathbf{0}, \tag{18}$$

so we can solve our original problem by defining

$$\begin{aligned} \mathbf{A}_3 &\equiv \mathbf{B}_3^t\mathbf{B}_1^t, & (19)\\ \mathbf{A}_4 &\equiv -\mathbf{B}_4^t\mathbf{B}_2^t. & (20)\end{aligned}$$

Why was the first eigenmode step necessary? We went through the extra trouble of solving equations (11) and (12) and applying the cutoff λ_{min} because otherwise, a number $(\mathbf{B}^t\tilde{\mathbf{y}})_i$ in our final data vector could be noise-dominated for two different reasons:

1. Because the signal contribution from $\mathbf{y}_1$ approximately cancels that from $\mathbf{y}_2$.
2. Because it is a noise-dominated mode from $\mathbf{y}_1$ or $\mathbf{y}_2$ (or some combination thereof).

It is clearly only the first case that interests us when comparing data sets. In the latter case, applying the null-buster only tests for systematic errors internally, within each data set, and this is best done before comparing it with other data sets. We therefore throw away all noise-dominated modes from the individual data sets, choosing say $\lambda_{min} = 1$, before proceeding to the final eigenvalue problem (16). A lower threshold may be appropriate as well — as long as we choose $\lambda_{min} \gg \lambda_{max}$, we know that the lack of signal in $\mathbf{B}^t\tilde{\mathbf{y}}$ will be due mainly to subtracting the data sets.

4. COMBINING DATA SETS

4.1. *Combining maps of different beam shape*

Suppose that we have performed all the tests described above and conclude that the data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ are consistent. We then wish to simplify future calculations by combining the two data sets into a single map $\tilde{\mathbf{x}}$, inverting the (usually over-determined) system of linear equations (3). A physically different but mathematically identical problem was solved in Tegmark (1997a), showing that the minimum-variance choice

$$\tilde{\mathbf{x}} = [\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{y} \tag{21}$$

retains all the cosmological information of the original data sets. Substituting this into equation (3) shows that the combined map is unbiased ($\langle\tilde{\mathbf{x}}\rangle = \mathbf{x}$) and that the pixel noise $\boldsymbol{\varepsilon} \equiv \tilde{\mathbf{x}} - \mathbf{x}$ has the covariance matrix

$$\boldsymbol{\Sigma} \equiv \langle\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t\rangle = [\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}. \tag{22}$$

A common special case is that where the two data sets have uncorrelated noise ($\mathbf{N}_{12} = \mathbf{0}$), simplifying the solution to

$$\begin{aligned} \tilde{\mathbf{x}} &= \boldsymbol{\Sigma}\left[\mathbf{A}_1^t\mathbf{N}_1^{-1}\mathbf{y}_1 + \mathbf{A}_2^t\mathbf{N}_2^{-1}\mathbf{y}_2\right], & (23)\\ \boldsymbol{\Sigma} &= [\mathbf{A}_1^t\mathbf{N}_1^{-1}\mathbf{A}_1 + \mathbf{A}_2^t\mathbf{N}_2^{-1}\mathbf{A}_2]^{-1}. & (24) \end{aligned}$$

An even simpler case occurs if the first data set is already a sky map (say the QMAP map), so that $\mathbf{A}_1 = \mathbf{I}$, and we wish to combine it with a more complicated data set covering the same sky region (say the Saskatoon observations). For this case, equation (23) can be rewritten as

$$\tilde{\mathbf{x}} = \mathbf{y}_1 + \boldsymbol{\Sigma}\mathbf{A}_2^t\mathbf{N}_2^{-1}(\mathbf{y}_2 - \mathbf{A}_2\mathbf{y}_1), \tag{25}$$

which has a simple interpretation. The vector $\mathbf{A}_2\mathbf{y}_1$ is just map 1 convolved with the observing strategy of experiment 2, so the factor $(\mathbf{y}_2 - \mathbf{A}_2\mathbf{y}_1)$ contains only noise. The map $\tilde{\mathbf{x}}$ is thus obtained by correcting $\mathbf{y}_1$ with a pure noise term that partially cancels some of its noisiest modes.

This important case applies also to combining two maps with different angular resolution. For instance, if $\mathbf{y}_1$ and $\mathbf{y}_2$ have narrow Gaussian beams of resolution θ_1 and θ_2, with $\theta_2 > \theta_1$, then we define $\mathbf{x}$ to be the true sky map smoothed by θ and have $(\mathbf{A}_2)_{ij} = \exp[-\theta_{ij}^2/2\theta^2]/2\pi\theta^2$ where $\theta_{ij} = \cos^{-1}(\widehat{\mathbf{r}}_i\cdot\widehat{\mathbf{r}}_j)$ is the angular separation between pixels i and j and $\theta = (\theta_2^2 - \theta_1^2)^{1/2}$ is the extra smoothing in map 2. Despite occasional claims to the contrary, this shows that two maps at different resolution can be combined without destroying any information, without first degrading the higher resolution down to the lower one by smoothing. Instead, equation (21) will use the lower resolution map $\mathbf{y}_2$ to improve the accuracy of the large-scale fluctuations in $\mathbf{y}_1$ (as was done by Schlegel *et al.* 1998 when combining the DIRBE and IRAS maps), retaining all the information from both maps.

Note that even the case of two identical maps ($\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{I}$) can be non-trivial. Equation (23) shows that the optimal combination is $\tilde{\mathbf{x}} = \boldsymbol{\Sigma}\left[\mathbf{N}_1^{-1}\mathbf{y}_1 + \mathbf{N}_2^{-1}\mathbf{y}_2\right]$. This was used in the combined QMAP analysis (de Oliveira-Costa *et al.* 1998), and reduces to separate averaging for each pixel only if the two maps have vanishing or identical noise correlations.

4.2. *Combining maps at different frequencies to remove foregrounds*

As described in Tegmark (1998) and further elaborated in White (1998), foregrounds can be treated as simply an additional source of noise that is correlated between channels ($\mathbf{N}_{12} \neq \mathbf{0}$). This means that if we have d data sets measured at different frequencies ν_α, $\alpha = 1, 2, ..., d$, each defined by their own matrix $\mathbf{A}_\alpha$, the best[5] way to combine them is still given by equation (21) — we simply need to include more physics in the noise covariance matrix $\mathbf{N}$. Let us be more explicit about this. The noise covariance matrix $\mathbf{N}$ will be of size $n \times n$, where $n = \sum_{\alpha=1}^d n_\alpha$, *i.e.*, the total number of numbers in all data sets combined. We will therefore write the elements of $\mathbf{N}$ as $\mathbf{N}_{\alpha i\beta j}$, where α and β determine the data sets and i and j the numbers therein. If there are f foreground components, this noise matrix becomes a sum $\mathbf{N} = \sum_{k=0}^f \mathbf{N}^{(k)}$, where $\mathbf{N}^{(0)}$ is the contribution from instrumental noise and the other terms are the contributions from foregrounds (synchrotron emission, bremsstrahlung, dust, point sources, *etc.*). Each of these foreground matrices will be of the form

$$\mathbf{N}^{(k)}_{\alpha i\beta j} = \mathbf{F}^{(k)}_{\alpha\beta}[\mathbf{A}_\alpha\mathbf{C}^{(k)}\mathbf{A}_\beta^t]_{ij}, \tag{26}$$

where the $d \times d$ matrix $\mathbf{F}^{(k)}$ gives the covariance of the i^{th} foreground between frequencies, and the $m \times m$ matrix $\mathbf{C}^{(k)}$ gives its spatial covariance between pixels. For example, if the data sets are identical maps ($\mathbf{A}_\alpha = \mathbf{I}$) at $d = 2$ different frequencies, we obtain the $(2m) \times (2m)$ block matrix

$$\mathbf{N}^{(k)}_{\alpha i\beta j} = \mathbf{F}^{(k)} \otimes \mathbf{C}^{(k)} = \begin{pmatrix} F_{11}^{(k)}\mathbf{C}^{(k)} & F_{12}^{(k)}\mathbf{C}^{(k)} \\ F_{21}^{(k)}\mathbf{C}^{(k)} & F_{22}^{(k)}\mathbf{C}^{(k)} \end{pmatrix}. \tag{27}$$

Explicit models specifying the foreground dependence on frequency and position (the matrices $\mathbf{F}$ and $\mathbf{C}$) can be found in Tegmark (1998). Although it has been common to characterize $\mathbf{C}$ by a power spectrum as in equation (8), the above formalism clearly works even if we break the isotropy by introducing an additional dependence on galactic latitude, say.

This shows how to best combine data sets at different frequencies. When combining different multifrequency experiments, it is desirable to first combine corresponding maps at the same frequency and apply the null-buster test for systematic errors. The different frequencies can then be merged afterwards, as a second step.

[5] Specifically, under the assumption that foregrounds and detector noise have multivariate Gaussian probability distributions, one one can show (Tegmark 1997) that this foreground removal method retains all the cosmological information present in the multifrequency set of input maps, i.e., constitutes an information-theoretically lossless form of data compression reducing all data down to a single CMB map. One does not need to assume that the CMB itself is Gaussian. Systematics such as foreground signals, data spikes, and atmospherics seldom have a Gaussian probability distribution – in this more general case, the removal method is no longer strictly lossless, but retains the feature that it minimizes the total rms of foregrounds and noise assuming only their finite second moment.

5. CONCLUSIONS

We have derived a set of useful tools for comparing and combining CMB data sets, all based on simple matrix operations, and drawn the following conclusions:

1. The "null-buster" test is better at detecting systematic errors than a simple χ^2-test.
2. Such a consistency test can be performed even between two experiments with quite different beam shape and observing strategy.
3. When combining two maps of different angular resolution, one need not smooth the higher resolution down to the lower one.
4. When combining two identical maps, one should generally not do the averaging separately for each pixel.
5. Our map merging method also handles the case of *e.g.* Planck, where the beams are elliptical rather than round and the different detectors have different beam orientations.
6. The foreground removal method of Tegmark (1998) is a special case of the combination technique we derived, and can be carried out even for foreground models involving frequency dependence, scale dependence, latitude dependence and spectral index variations in combination.

As the available CMB data sets continue to increase in quantity and quality, it will be useful to perform such cross-checks against systematic errors and incrementally combine all consistent data sets into a single state-of-the art map containing our entire knowledge of the CMB sky.[6]

Both theories and new observations can then be tested against this combined map as it gradually grows in size, quality and resolution.

The author wishes to thank Angelica de Oliveira-Costa and Martin White for helpful comments. Support for this work was provided by NASA though grant NAG5-6034 and Hubble Fellowship HF-01084.01-96A from STScI, operated by AURA, Inc. under NASA contract NAS5-26555.

REFERENCES

Bennett, C. L. *et al.* 1996, *ApJ*, **464**, L1
Bond, J. R. 1995, *Phys. Rev. Lett.*, **74**, 4369
Bond, J. R., Jaffe, A. H., & Knox, L. 1998, *Phys. Rev. D*, **57**, 2117
Bunn, E. F., & Sugiyama, N. 1995, *ApJ*, **446**, L49
de Oliveira-Costa, A., Devlin, M., Herbig, T., Miller, A. D., Netterfield, C. B., Page, L. A., & Tegmark, M. 1998, *ApJL*, **509**, L77
Devlin, M., de Oliveira-Costa, A., Herbig, T., Miller, A. D., Netterfield, C. B., Page, L. A., & Tegmark, M. 1998, *ApJL*, **509**, L77
Ganga, K., Cheng, E., Meyer, S., & Page, L. 1993, *ApJL*, **410**, L57
Knox, L., Bond, J. R., Jaffe, A. H., Segal, M., & Charbonneau, D. 1998, astro-ph/9803272 (K98)
Inman, C. A. *et al.* 1997, *ApJL*, **478**, L1
Lineweaver, C. H. *et al.* 1995, *ApJ*, **448**, 482
Ruhl, J. E. *et al.* 1995, *ApJL*, **453**, L1
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
Tegmark, M. *et al.* 1996a, *ApJL*, **474**, L77
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1996b, *ApJ*, **480**, 22
Tegmark, M. 1997a, *ApJ*, **480**, L87
Tegmark, M. 1997b, *Phys. Rev. D*, **55**, 5895
Tegmark, M. 1998, *ApJ*, **502**, 1
White, M. 1997, *Phys. Rev. D*, **57**, 5273
Wright, E. L. 1996, astro-ph/9612006.
Wright, E. L., Hinshaw, G., & Bennett, C. L. 1996, *ApJL*, **458**, L53

This paper is available at *h t t p://www.sns.ias.edu/˜max/comparing.html*

[6]The signal and noise covariance matrices **S** and **N** of such a map could of course be rather complicated. However, the same complication is encountered if one maintains separate experiment-specific maps and only attempts to combine the power spectrum measurements. This is because the sample variance of the band power measurements from different experiments are generally not independent, and computation of the correlation requires knowledge of these matrices. A disadvantage of combining only at the power spectrum level is that this generally sacrifices useful information on relative phases, whereas the merged map retains all the cosmological information from both data sets. For instance, near degeneracies in the noise covariance matrices of two maps can often be broken by combining them.