Testing Gaussianity With the Smooth Tests of Goodness of Fit

Andrés Curto

IFCA, Santander

Outline

- Introduction
- The goodness-of-fit Gaussianity test
  1. Smooth tests of goodness of fit
  2. Signal-to-noise decomposition
- Archeops experiments
- Archeops data sets
- Gaussianity analysis on Archeops data
- Gaussianity analysis on WMAP data

Introduction

- The detection of non-Gaussianity in cosmic microwave background (CMB) anisotropies can question the validity of several standard inflationary theories
- The detection of non-Gaussianity is also a tool to find “contamination” in CMB maps

Smooth Tests of Goodness of Fit (I)

$\{x_i\}$ set of n random numbers distributed with a p.d.f. $f(x,\theta)$

$\theta=(\theta_1,\ldots,\theta_m)\in\Theta$ m parameters to define the p.d.f. space

Null Hypothesis H_0: $\theta=\theta_0$

vs.

Alternative Hypothesis K: $\theta\neq\theta_0$

Smooth Tests of Goodness of Fit (II)

- The “order k” alternative p.d.f. (Rayner & Best, 1989, 1990)

$$g_k(y,\theta)=C(\theta)\exp\left[\sum_{i=1}^{k}\theta_i h_i(y)\right]f(y)$$

- To evaluate the null hypothesis H_0 we use the “score statistic”

$$S_k=\sum_{i=1}^{k}U_i^2$$

Smooth Tests of Goodness of Fit (III)

- The U_i^2 statistics are computed from the set of random numbers through

$$U_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} h_i(x_j)$$

- The quantities $h_i(x)$ depends on the p.d.f. that we want to test. For the Gaussian case, these are the "normalized Hermite-Chebyshev" polynomials

Smooth Tests of Goodness of Fit (IV)

The U_i statistics in terms of the x_i random numbers

$$U_1^2 = n(\hat{\mu}_1)^2 \qquad U_2^2 = \frac{n}{2}(\hat{\mu}_2 - 1)^2$$

$$U_3^2 = \frac{n}{3}(\hat{\mu}_3 - 3\hat{\mu}_1)^2 \qquad U_4^2 = \frac{n}{24}(\hat{\mu}_4 - 6\hat{\mu}_2 + 3)^2$$

...

$$\hat{\mu}_\alpha = \sum_{j=1}^{n} x_i^\alpha$$

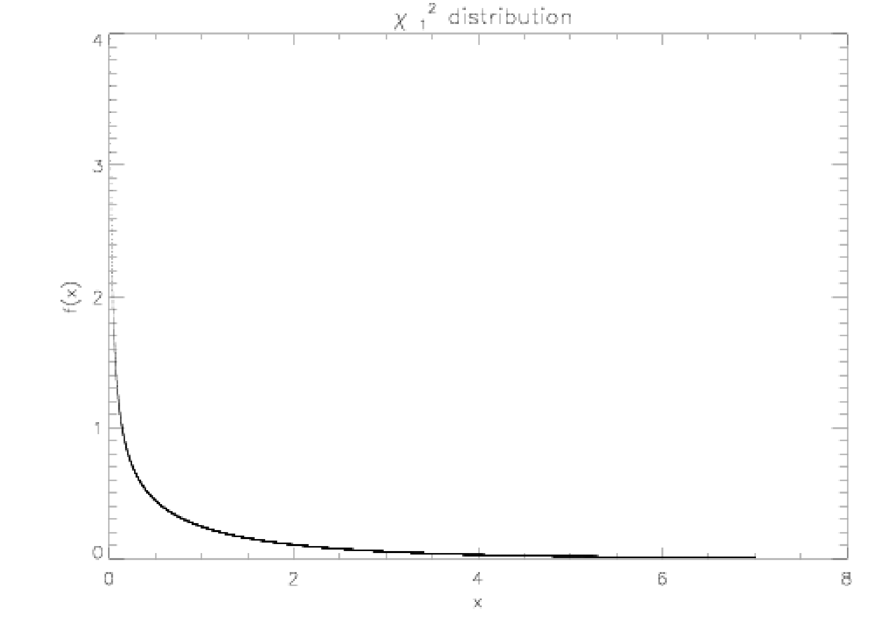
If the null hypothesis H_0 is satisfied then for n>>1 it can be proved that

$$U_i^2 \sim \chi_1^2 \qquad S_k \sim \chi_k^2$$

The χ_1^2 Distribution

$$f(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}$$

x_0	3.5	4.0	5.0	6.0	7.0
$P(x_0 \leq x)$	0.939	0.955	0.975	0.986	0.992



Signal-to-noise Eigenmode Decomposition (I)

The set of *$\{x_i\}$* numbers that we analyze comes from a CMB anisotropies map $\Delta T(\vec{n}_i)$

$\vec{n}_i$ is the coordinates vector of a pixel *i* from a set of n_{pix} pixels

For us a CMB map is a vector of n_{pix} real numbers d_i

Signal to noise decomposition $\vec{d} = \vec{s} + \vec{n}$

$\vec{s}$ Signal Contribution: CMB

$\vec{n}$ Noise contribution: the instrumental noise

Signal-to-noise Eigenmode Decomposition (II)

Consider the 2-point correlation function

$$\xi(\vec{x}_1, \vec{x}_2) = <g(\vec{x}_1) \cdot g(\vec{x}_2)>¿$$
¿

$$g(\vec{x}) = \Delta T(\vec{x})$$

We can compute it for signal and noise

$$S(\vec{x}_1, \vec{x}_2) \equiv \xi_S(\vec{x}_1, \vec{x}_2) = <\Delta T_S(\vec{x}_1) \cdot \Delta T_S(\vec{x}_2)>¿$$
¿

$$N(\vec{x}_1, \vec{x}_2) \equiv \xi_N(\vec{x}_1, \vec{x}_2) = <\Delta T_N(\vec{x}_1) \cdot \Delta T_N(\vec{x}_2)>¿$$
¿

Signal-to-noise Matrix (III)

Equivalent to divide signal by noise

$$A = L_N^{-1} S (L_N^t)^{-1}$$

where L_N is the Cholesky matrix of N $N = L_N L_N^t$

$$L_N = R_N D_N^{1/2}$$

- R_N eigenvector matrix of N
- D_N eigenvalues matrix of N

$$R_N^T N R_N = D_N$$

The Signal-to-noise Eigenmodes (IV)

First we need the A matrix eigenvectors and eigenvalues

$$R_A^T A R_A = D_A$$

- R_A eigenvectors matrix
- D_A eigenvalues matrix

The set of signal-to-noise eigenmodes $\{y_i\}$ for a given map

$$\xi = R_A^T L_N^{-1} d$$

$$y_i = \frac{\xi_i}{\sqrt{1+(D_A)_i}}$$

Signal-to-noise Eigenmodes (V)

The properties of the *{y_i}* eigenmodes when the map is Gaussian:

- *Independent numbers* $\langle y_i y_j \rangle = \delta_{ij}$
- *Each* y_i *behaves statistically as N(0,1)*
- $(D_A)_i$ *is a measure of the signal-to-noise^2 ratio*

The goal is to analyze the *{y_i}* eigenmodes with the previous statistical method

The Archeops Experiment

- A balloon borne experiment (1999-2002)
- First link in C_l determination between large angular scales (COBE) and first acoustic peak (BOOMERanG, MAXIMA)
- Archeops shares the same technological design with the very same Planck satellite
- 21 bolometers at frequencies 143, 217, 353, 545 GHz
- 20% of sky with a beam of 6’ to 8’

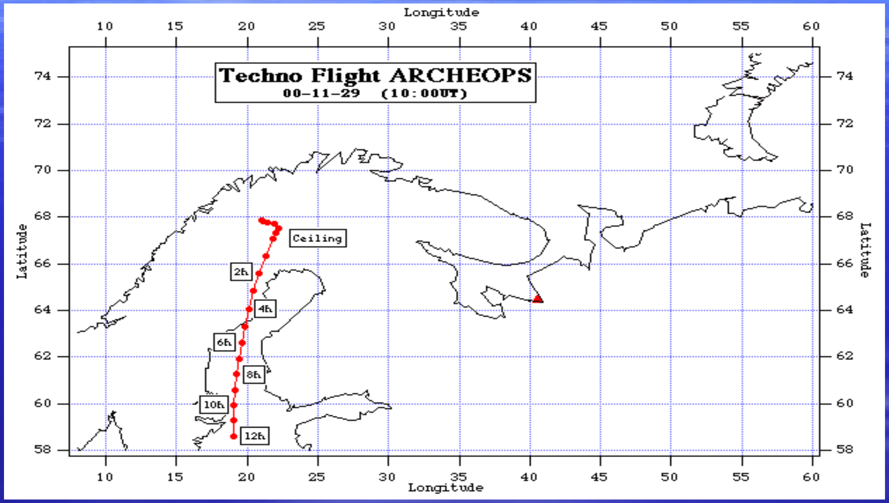
The Archeops "Face"





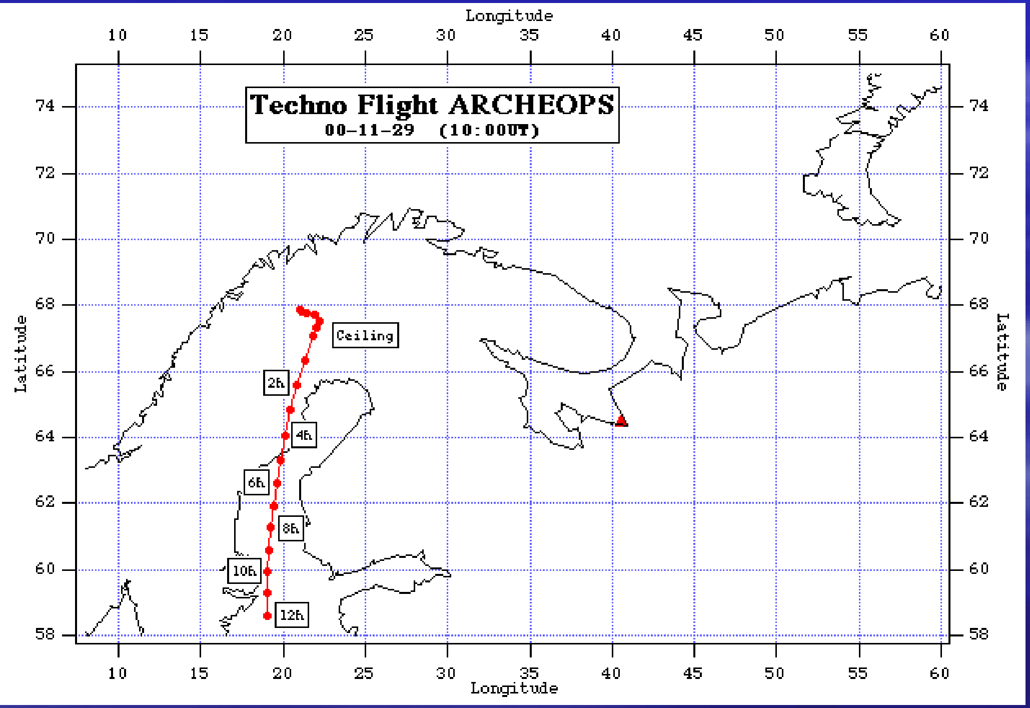

The Archeops Flights

- Trapani test flight (July 1999)
- Kiruna test flight (March/April 2000)
- Kiruna scientific flight (29 January 2001)
- Kiruna scientific flight (17 January 2002)
- Kiruna scientific flight (7 February 2002)



Kiruna First Flight

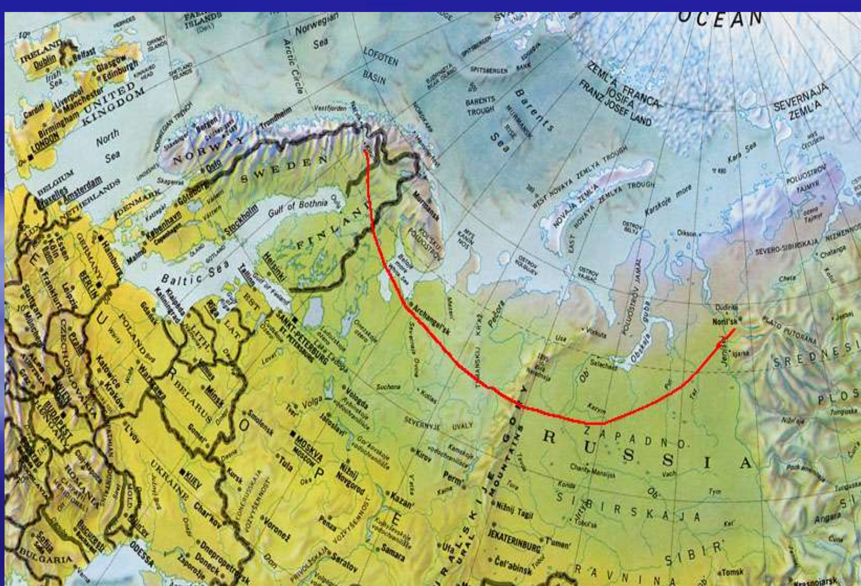
- The 21st January 2001 (Artic winter)
- In Kiruna, Sweden (not traveled to Urals Mountains)
- Almost lost in the sea

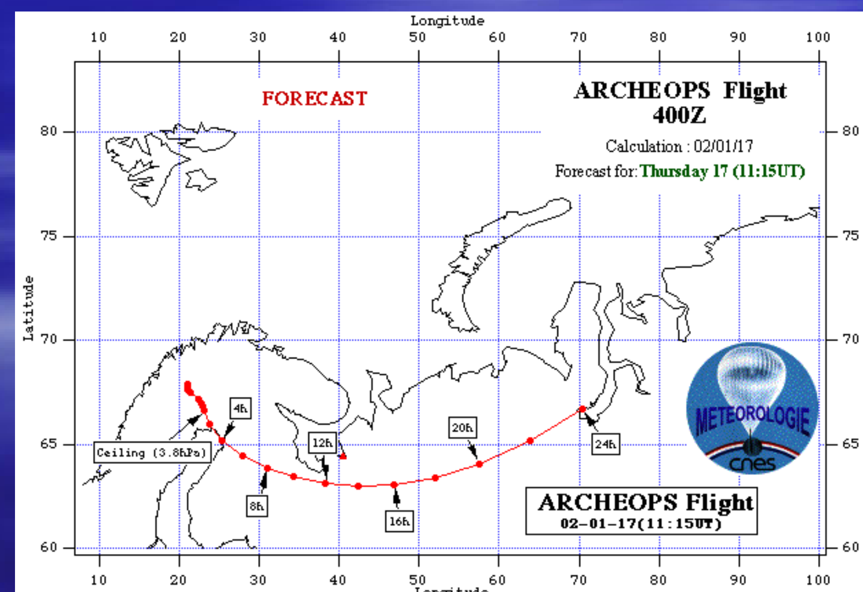


Andrés Curto, BICOS 2007

Kiruna Second Flight

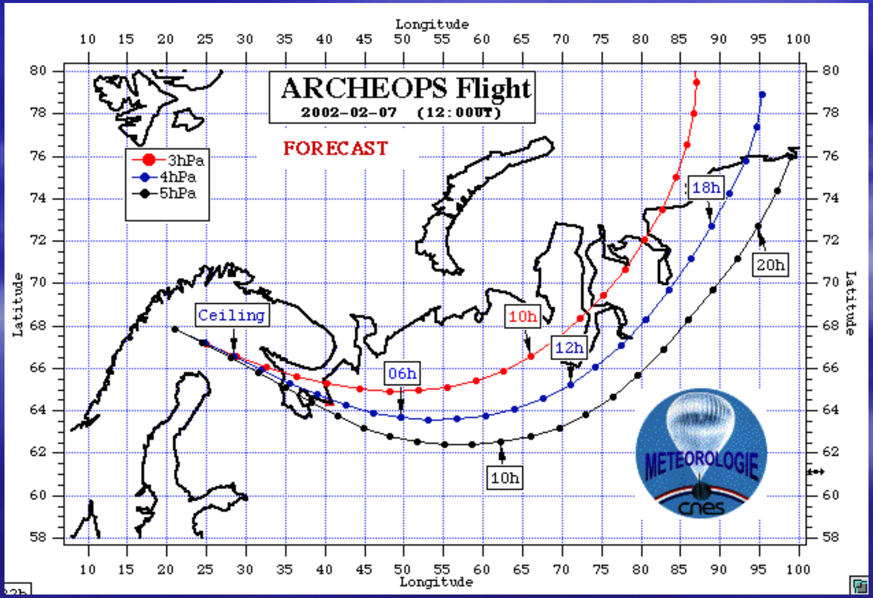
- The 17th January 2002 (Artic winter)
- From Sweden to Russia (Siberia)





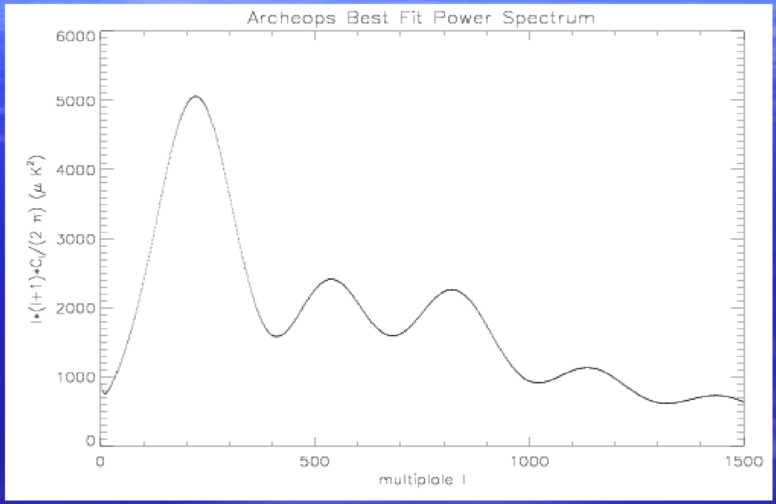
Kiruna Third Flight

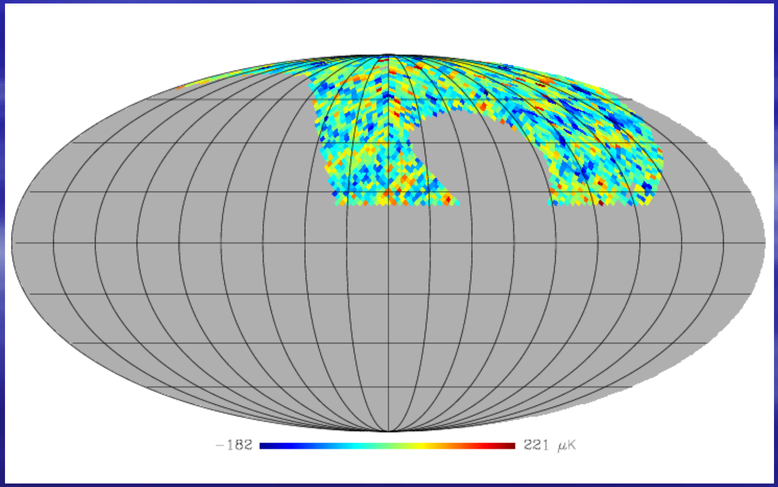
- The 7th february 2002 (Artic winter)
- From Sweden to Russia (Siberia)
- Onboard computer problems



Archeops Data Sets

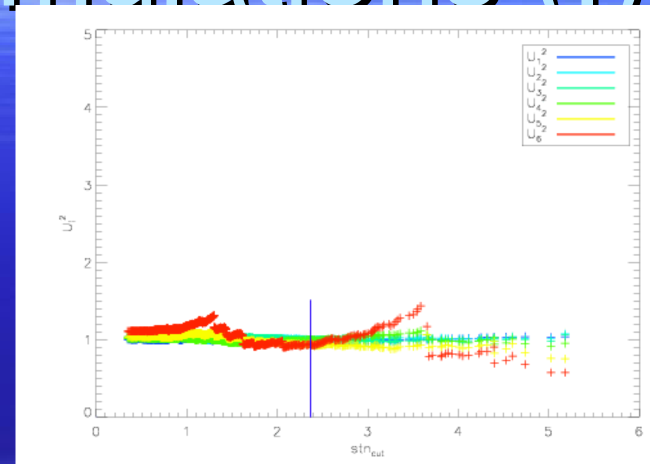
- Analyzed 143 GHz bolometer map
- 16 % north galactic hemisphere analyzed
- 1.8º pixel resolution
- 1995 available pixels
- ~$5 \cdot 10^5$ Gaussian signal simulations computed with Archeops best fit power spectrum (Tristram et al. 2005)
- ~$5 \cdot 10^5$ Gaussian noise simulations

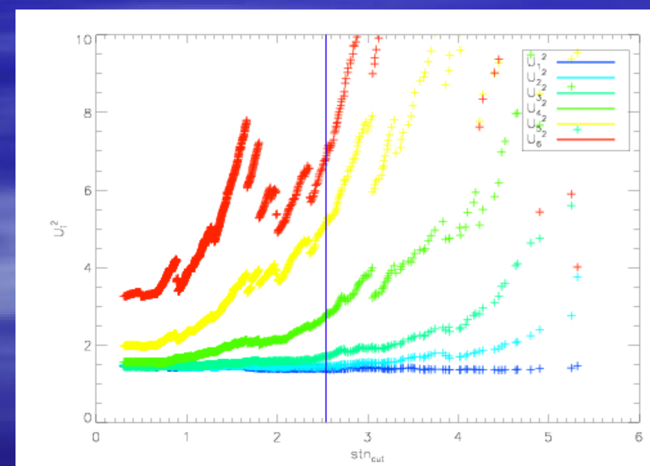




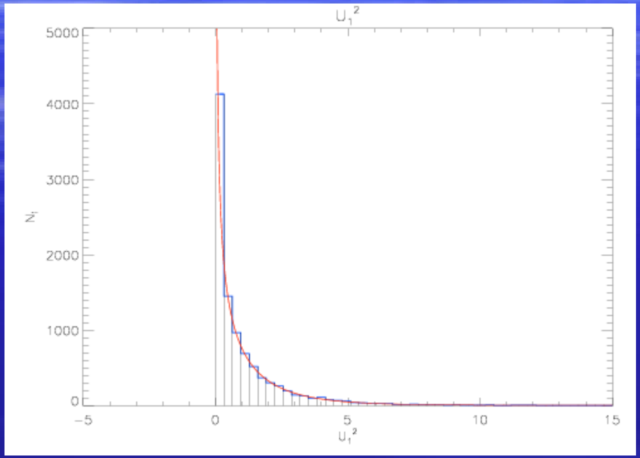
The Archeops Analysis: Calibration With Simulations (I)

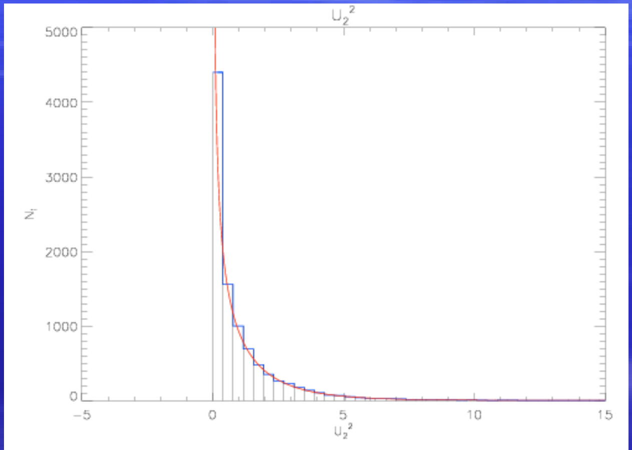
- First we calibrated the method with Gaussian simulations (Gaussian simulations should return the χ_1^2 distribution for U_i^2)
- This means $\langle U_i^4 \rangle - \langle U_i^2 \rangle^2 = 2$ $\quad \langle U_i^2 \rangle = 1$
- High order, numerical errors are increased

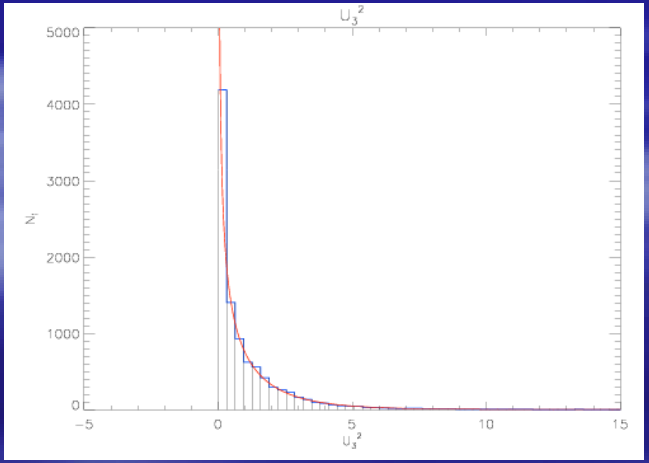


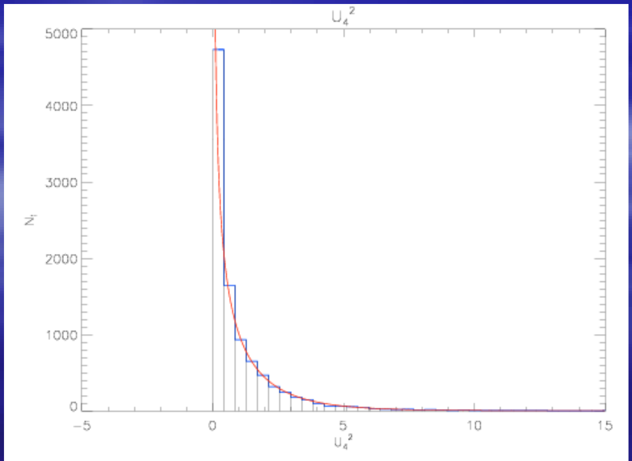


The Archeops Analysis: Calibration With Simulations (II)



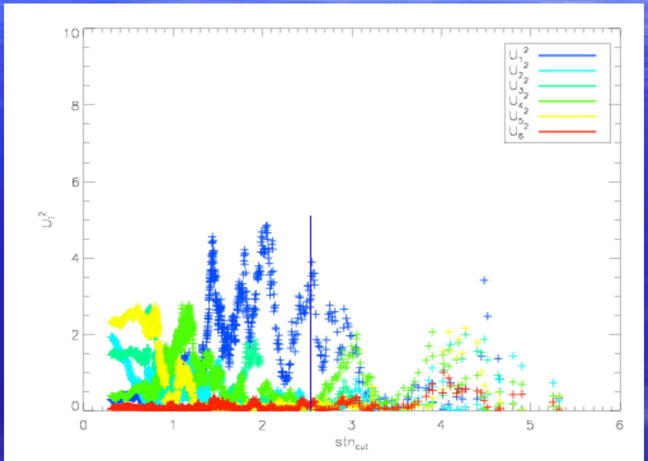


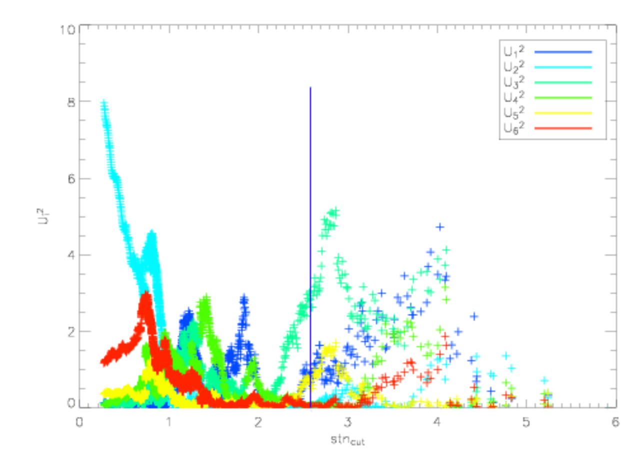




The Archeops Analysis: Analyzing The Data

- We obtained the U_i^2 statistics at different signal-to-noise cuts for two map making: MIRAGE and coaddition
- Some U_2^2 values are over 6.7, (99% cumulative probability)



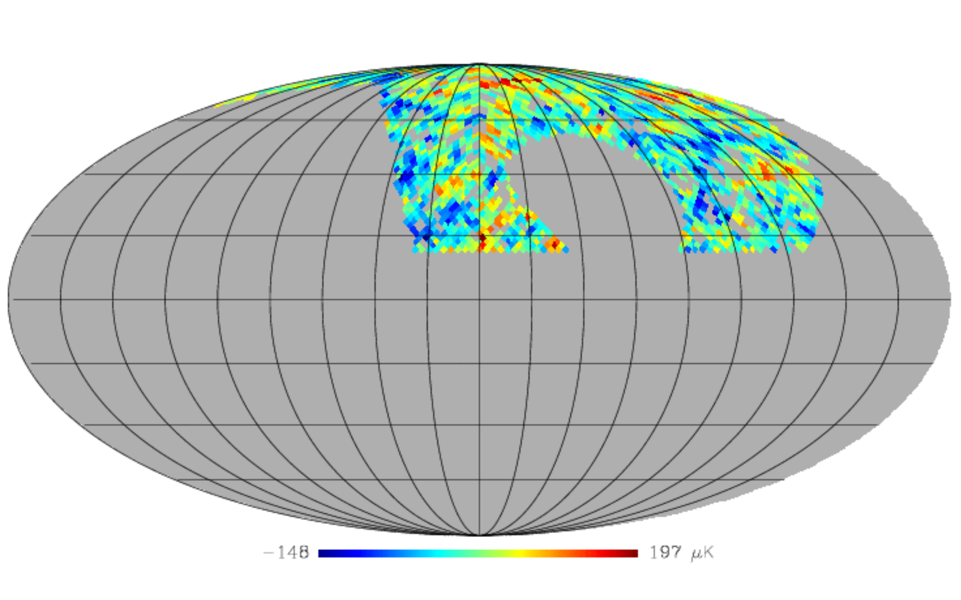


The Archeops Analysis: Significance and Explanation

- When p-value (the probability that the statistic takes a value at least as extreme as the observed) of U_2^2 is computed, p=15%
- There are many simulations (1500 from a set of 10000) which reach U_2^2 =7.97 in any s/n cut
- This detection is a “map-making issue”, due to it does not appear in the MIRAGE case

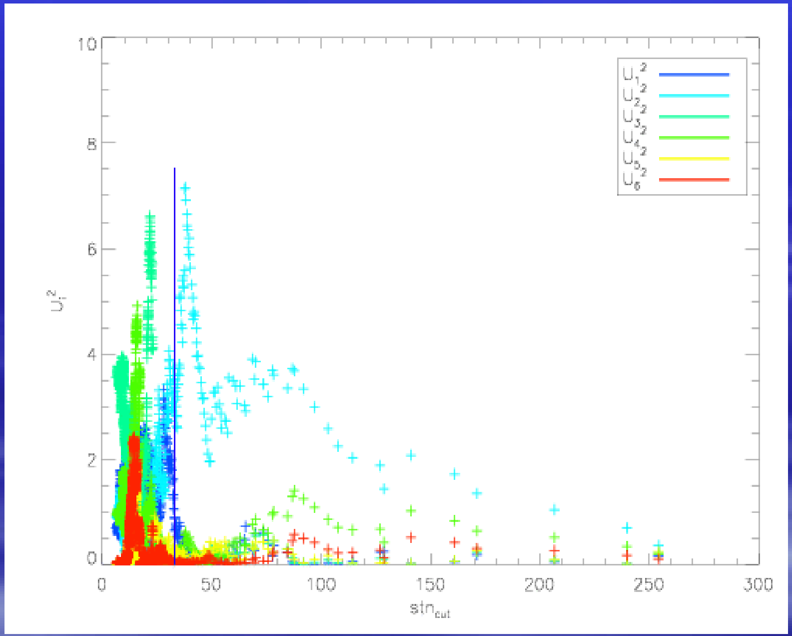
Comparing With WMAP (I)

- We analyzed the WMAP data to compare with Archeops
- Same pixel size: *1.8º*
- The map is a combination of maps at *41 GHz, 61 GHz* and *94 GHz*



Comparing With WMAP (II)

- WMAP data results comparable to those of Archeops
- WMAP has higher signal to noise range
- U_2^2 detection at signal to noise ratio *38.0*
- The p-value is *p=17.0%*

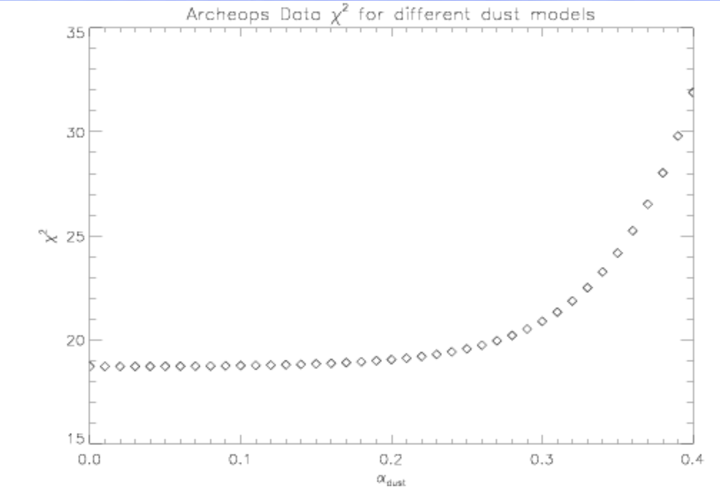


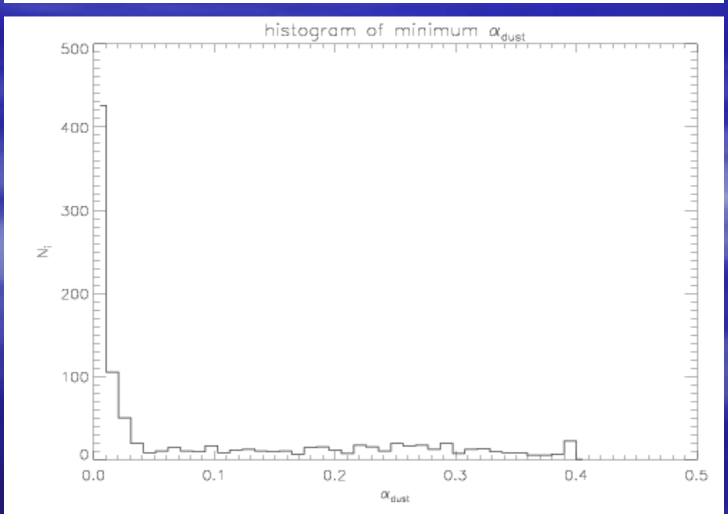
Archeops Levels of Dust Contamination

- Using an Archeops dust template
- We analyzed simulations

$$s + n + \alpha * dust$$

- With their U_2^2 values, we performed a χ^2 test
- The minimum χ^2 value gives us the level of dust
- The level corresponds to *3.2% (68%CL), 9.7% (95% CL)* when dispersions of Archeops map and dust template are compared





Archeops Levels of Dust Contamination (II)

$$U_2^2(\nu_1) - \text{¿}\, U_2^2(\nu_1, \alpha) > \text{¿}$$

$$U_2^2(\nu_2) - \text{¿}\, U_2^2(\nu_2, \alpha) > \text{¿}$$

$$\chi^2 = \sum_{\nu_1 \nu_2} \text{¿}$$

- Where $U_2^2(\nu)$ is the mean value of U_2^2 statistic over 50 adjacent signal-to-noise cuts
- There are 38 different cases: $\nu_1=1$ corresponds to the 50 lowest signal-to-noise cuts, $\nu_1=38$ corresponds to the 50 highest signal-to-noise cuts
- C is the correlation matrix among these 38 cases

$$C_{\nu_1 \nu_2} = < U_2^2(\nu_1, \alpha) U_2^2(\nu_2, \alpha) > - \text{¿}\, U_2^2(\nu_1, \alpha) >< U_2^2(\nu_2, \alpha) > \text{¿}$$

$$\text{¿}$$

- $<U_2^2(\nu, \alpha)>$ is the mean of $U_2^2(\nu)$ for simulations with dust, in a way

$$s + n + \alpha * dust$$

Conclusions

- Archeops data map at 143 GHz and 1.8º resolution are compatible with Gaussianity
- Comparable results with WMAP at same resolution and mask size
- 3.2% of dust contamination in ARCHEOPS (68% confidence level), and 9.7% (95% confidence level)

References

- Rayner, J. C. W. & Best, D. J. 1989, Smooth Tests of Goodness of Fit, Oxford University Press, New York
- Benoît, A, Ade, P., Amblard, A, et al. 2003, A&A, 399, L19
- Tristram, M., Patanchon, G., Macías-Pérez, J. F., et al. 2005, A&A, 436, 785
- Ponthieu, N., Macías-Pérez, J. M., Tristram, M., et al. 2005, A&A, 444, 327
- Cayón, L., Martínez-González, E., Argüeso, F., Banday, A. J., & Gorski, K. M. 2003a, MNRAS, 339, 1189
- Curto, A., Aumont, J., Macías-Pérez, J. F. , Martínez-González, E., Barreiro, R. B. et al. 2006, astro-ph/0612148

People

- Andrés Curto IFCA, Santander (Spain)
- Jonathan Aumont LPSC, Grenoble (France)
- Juan Francisco Macías-Pérez LPSC, Grenoble (France)
- Enrique Martínez-González IFCA, Santander (Spain)
- Belén Barreiro IFCA, Santander (Spain)
- Daniel Santos LPSC, Grenoble (France)
- Xavier Désert LAOG, Grenoble (France)