Elements of Data Analysis for COSMOSOMAS

Benoît Revenu[1], Jacques Delabrouille[1] and Julio Gallegos[2]
[1] : PCC-Collège de France, France — [2] : IAC, Ténérife, Spain

May 3, 2000

1 A first analysis of the noise for COSMOSOMAS

1.1 Introduction

We give here some results of a preliminary analysis of COSMOSOMAS noise properties as seen on data from 02 january 2000 and 01 february 2000. We discuss methods of data reduction to minimize the effect of this noise.

In this note, the data sample k on circle i is denoted as $m_{i,k}$. We have n circles of $n_{\text{ech}} = 212$ samples each.

1.2 Noise spectra

We compute the noise spectrum for the data in three ways. First, we compute the FFT of the signal for each circle, $\widetilde{m}_{i,\ell}$, where ℓ indexes fourier modes and i indexes circles. Thus $\widetilde{m}_{i,0}$ is the average of circle i, and $\widetilde{m}_{i,\ell}$ for $\ell \geq 1$ are higher order modes of the signal on circle i.

1. we compute the average $\widetilde{m}_{\text{avg}}(\ell) = \sum_i \widetilde{m}_{i,\ell}/n$
2. we compute the median $\widetilde{m}_{\text{med}}(\ell) = \text{MEDIAN}_{(i)}\widetilde{m}_{i,\ell}$
3. we compute the maximum value $\widetilde{m}_{\max}(\ell) = \text{MAX}_{(i)}\widetilde{m}_{i,\ell}$

Results for the two days and for the three channels of the 15GHz instrument are plotted in Figures 1 and 2. The difference between the average and the median illustrates the skewness of the distribution. The max, very high as compared to the other two, clearly indicates the presence of a few very bad scans (note we have not done any cutting yet, not even for Sun and Moon effects). The most significant spectrum for the bulk of our data is probably the median (i.e. the best of the three).

1.2.1 02 january

The spectrum of the noise on january 2 exhibits clear pickup lines at about 23 Hz, 52 Hz, 74 Hz, 98 Hz and 102 Hz. The 52 Hz is probably 50 Hz pickup from power lines. The 74 Hz is almost as strong as the 52 Hz, the other lines are much weaker. The lines are quite broad (a few Hz), which may be due to a small modulation of the 50 Hz source signal during the day, or a small modulation of the main mirror rotation speed, or a combination of both.

The spectrum exhibits a rise at frequencies below 40 Hz, compatible with the presence of a $1/f$ spectrum with knee frequency f_{knee} of about 20 Hz (the plots are in linear-log axes). The first two modes, very strong, correspond to the effect of atmosphere pickup, sidelobe effects, and very slow drifts.

The spectrum is significantly *not* white over the whole range of useful frequencies.

1.2.2 01 february

The spectrum of the noise for february 1 is significantly lower than that of january 2. Only three lines are clearly visible, at 52 Hz, 74 Hz, and (much weaker) 97 Hz.

It exhibits also a $1/f$ rise at low freqencies with f_{knee} of about 20 Hz, but is much cleaner than that of january 2 otherwise.

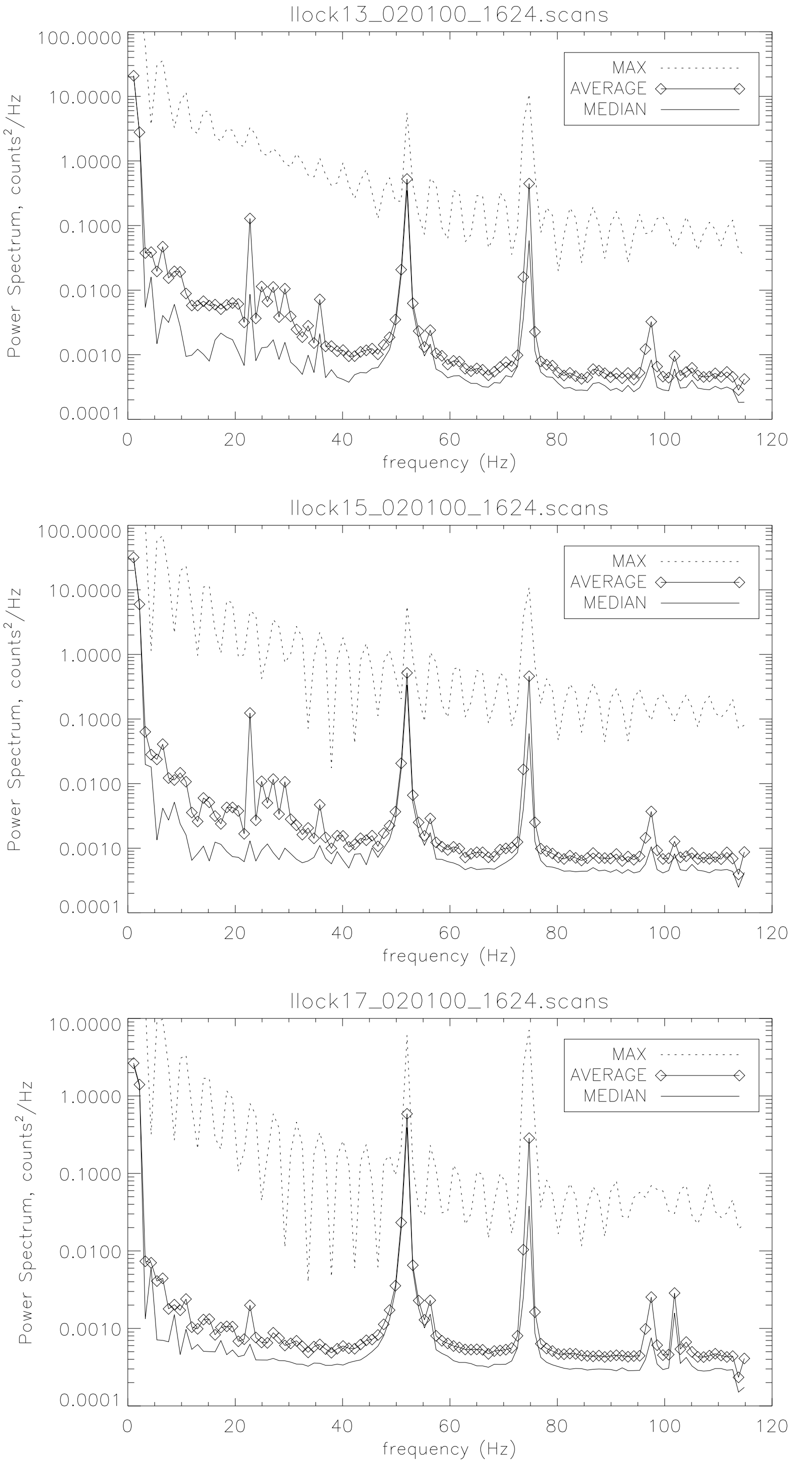


Figure 1: Power spectra for january 2nd, 2000

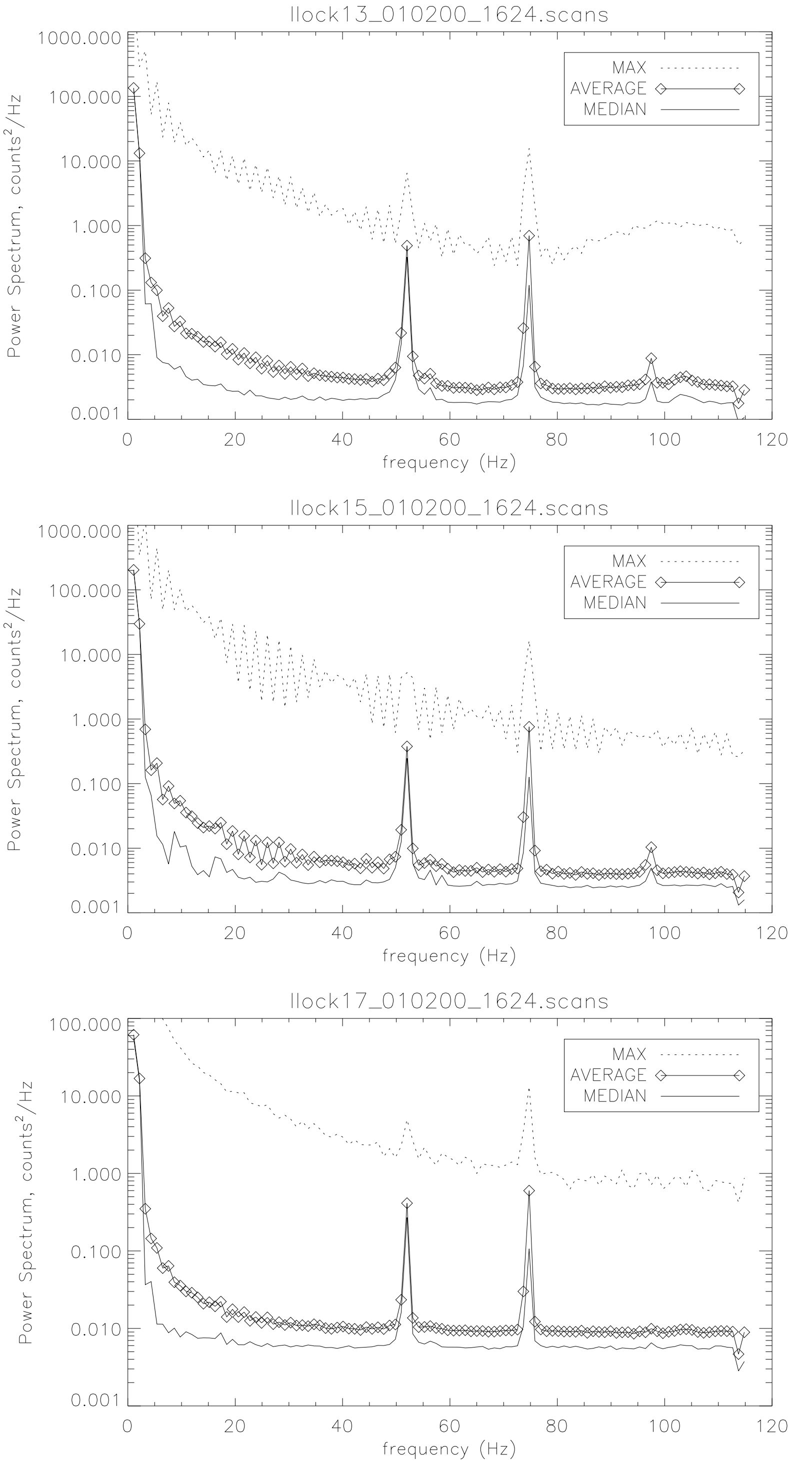


Figure 2: Power spectra for february 1st, 2000

1.3 Correcting for lines in the spectrum

Ideally, the pollution of isolated Fourier modes should be fixed by hardware (wire shielding, isolation). The origin of the 50 Hz is clear, that of the other lines may be its harmonics, cycles of engines related to the coolers, modulation of other frequencies by the acquisition frequency ...

1.3.1 Current methodology

Currently, the acquisition computes the average of the FFT of about 30 consecutive scans. For each of the computed FFT, the mode nearest to each line is corrupted. If the source lines are broad, if the frequency changes within 30 seconds, or if the mirror spinning speed changes, several modes are corrupted (something of this order happens for the spectrum estimated over the full day, as for each line three modes at least are significantly polluted).

The simplest to deal with present data files is to do what has been done at IAC for the 50 Hz line: cut corrupted modes by putting them to 0. The method should be extended to other lines as well. Unfortunately, lines are broad enough that at present at least a few modes are to be cut for each line. Of course, if the amplitude of monochromatic lines is present only for some circles, not all, it should be possible to refine on the cutting by performing it only on circles significantly contaminated. This has to be done with extreme caution however, as the monochromatic signal may be sub-dominant for each individual scan, but become clearly detectable when averages are performed (with clear consequences on the final errors on C_ℓ.

With a scanning speed of 65 turns per minute, and 62.51 degrees on the sky for each turn (for a circle diameter of 20 degrees), we can compute the scales on the sky corresponding to the main lines in the spectrum of the two days of data analysed here. The result is summarised in table 1.

frequency (Hz)	23	52	74	97
scale (degrees)	2.94	1.30	0.915	0.684

Table 1: Angular scales corresponding to the main lines in the spectrum.

1.3.2 Correction before the coaddition of scans

If the frequency lines of the contaminant are originally very pure, a major improvement of the rejection of such lines (if impossible by hardware) may be obtained by software, either in the acquisition with the present data storage, or in offline data processing if the full raw data is available.

This possible improvement relies on the fact that the resolution in frequency scales as the inverse of the total acquisition time. In the current data acquisition process, about 30 consecutive scans are averaged in Fourier space to give an average Fourier transform (which by linearity is the Fourier transform of the average). However, the frequency resolution for each of the FFTs is the spinning frequency of about 1.08 Hz. If the FFT had been computed for a signal of 30 consecutive scans instead, the resolution in frequency would have been 30 times better, i.e. 0.0361 Hz.

Because of this, in the ideal case that the contaminant has a very stable frequency (stability better than the resolution over 30 scans, i.e. about 30 seconds), and that the spinning period is also very stable, in the FFT computed over 30 scans all the useful astrophysical signal is concentrated within 0.0361 Hz of the harmonics of the spinning frequency, and the contaminant is concentrated within 0.0361 Hz of the pure true contaminant frequency.

This is illustrated with Figure 3 where we show the monolateral spectrum of a 212-sample white random signal repeated 30 times, to which we add a sine wave at a pure 50.1234 Hz frequency, simulating a 27.6923 second long signal, similar to that of COSMOSOMAS. As can be seen, the monochromatic signal can be separated easily from the useful periodic signal (diamonds) by filtering the FFT.

If the 50 Hz (and other) is not very monochromatic, wings may overlap with frequencies where the useful signal is, and there is little that can be done except shielding cables. If on the other hand the signal is not very periodic itself due to fluctuations of the rotation speed, then the monochromatic

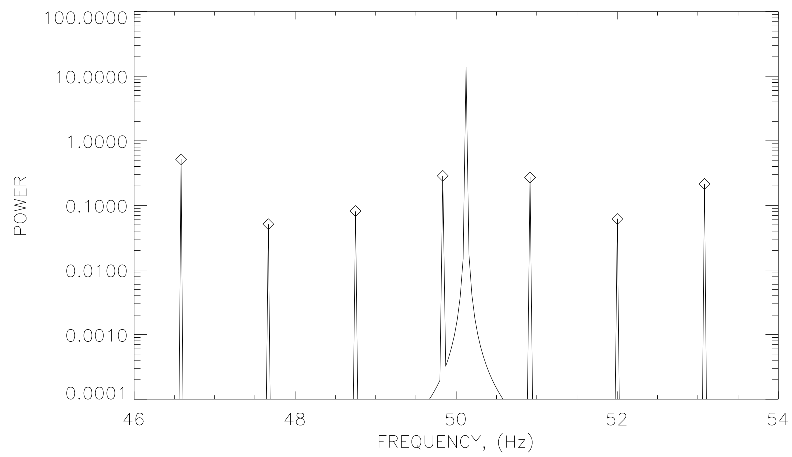


Figure 3: Power spectrum of a periodic signal contaminated by a non-synchronous monochromatic signal.

contaminants can still be separated from the useful signal. The better the exact pointing is known (i.e. the rotation speed as a function of time), the better this can be done. The method to separate the contaminants in this case is somewhat technical and we will not describe it here.

1.3.3 Subtraction using redundancies

For data as they are stored for the two days discussed here (no access to the raw data, only to co-added circles), an other method for subtracting monochromatic signals of fluctuating amplitude consists in extending the method we developped for subtracting low-frequency drifts on Planck data. Again, the method is somewhat intricate and we will not discuss it here.

1.4 Time-frequency analysis of COSMOSOMAS noise

In previous section, we introduced a discussion on the importance of the stability of contaminants for their subtraction by software, either in the acquisition, or offline. We will rediscuss this a bit in the next section on sampling.

we now turn to a time-frequency analysis of the signal and noise, in order to check how does the spectrum depend on the data. On Figure 4 we display the power spectral density as a function of frequency and circle number for the 13 GHz channel of february 1st. This permits to check whether the spectrum depends much on the circle.

We first notice that the 50 Hz extends over the same samples for all circles (about 2-3 samples) with quite constant power. This indicates that the line is not strictly monochromatic, but is stable. This is somewhat less true with the 74 Hz, which seems to shift occasionally to 73 Hz. The 97 Hz has a fluctuating intensity.

We can clearly identify by comparing with Figure 5 one or a few circles around 370 in which a strong point source (or a glitch!) puts additionnal power at all frequencies. The source is seen clearly on Figure 5. There is a strong 74 Hz pickup around that circle. The source at circle 1785 or so induces only excess low frequencies (4) and thus is somewhat more extended than the previous one.

A similar time-frequency analysis can be done on the data of january 2. On Figure 6 we display the power spectral density as a function of frequency and circle number for the 13 GHz channel that day. We see that all scans between 2300 and 3000 exhibit significantly larger low frequency power, part of which is probably noise (especially for scans between 2300 and 2500). This is confirmed by a look at Figure 7. Around scans 850, 1000, 1900, 2080 and 2100 we can identify compact sources of varying extent in the co-scan and cross-scan direction. A few obviously bad scans are easily identified.

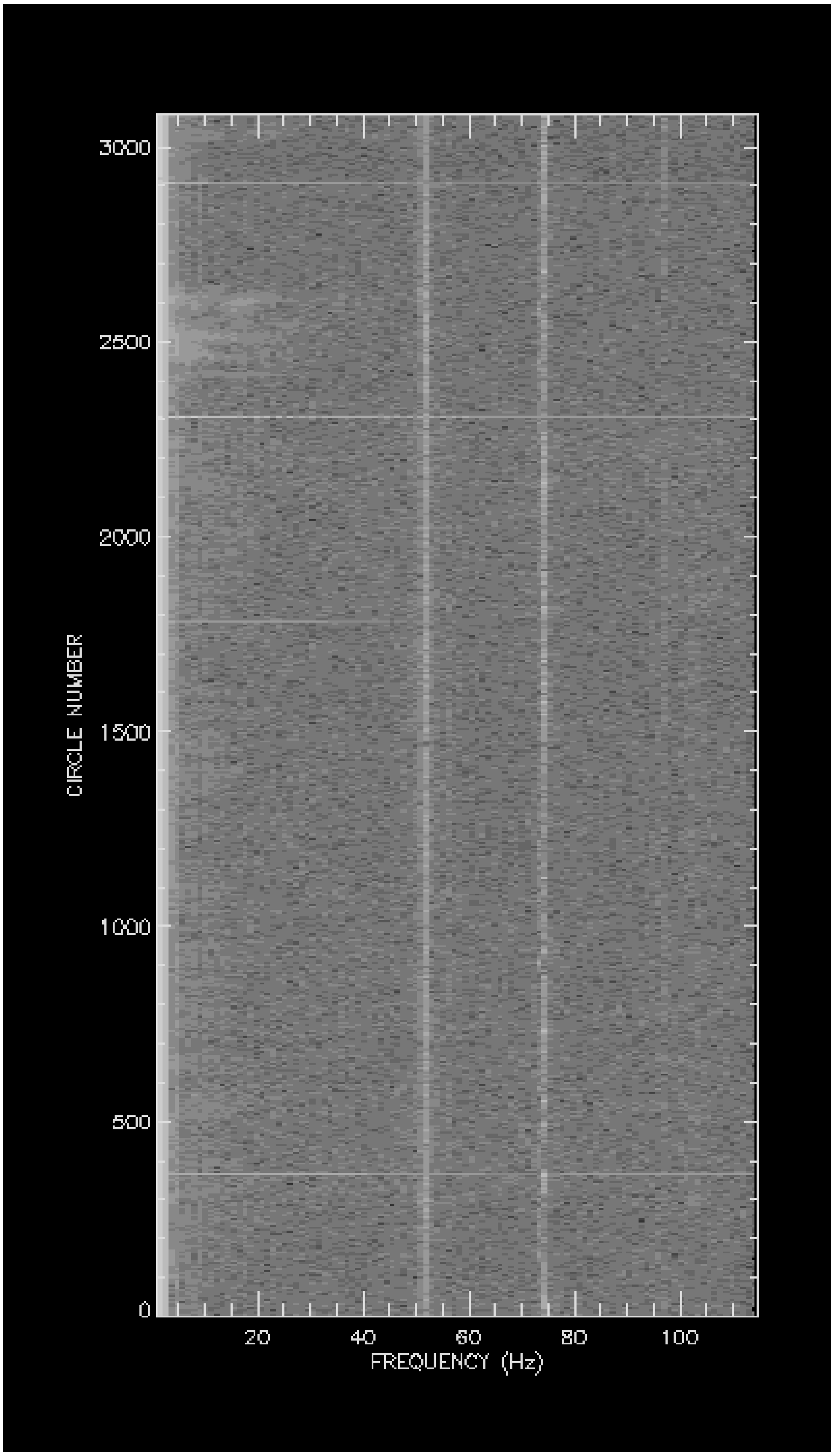


Figure 4: Time-frequency display of the signal of february 1 for the 13 GHz channel

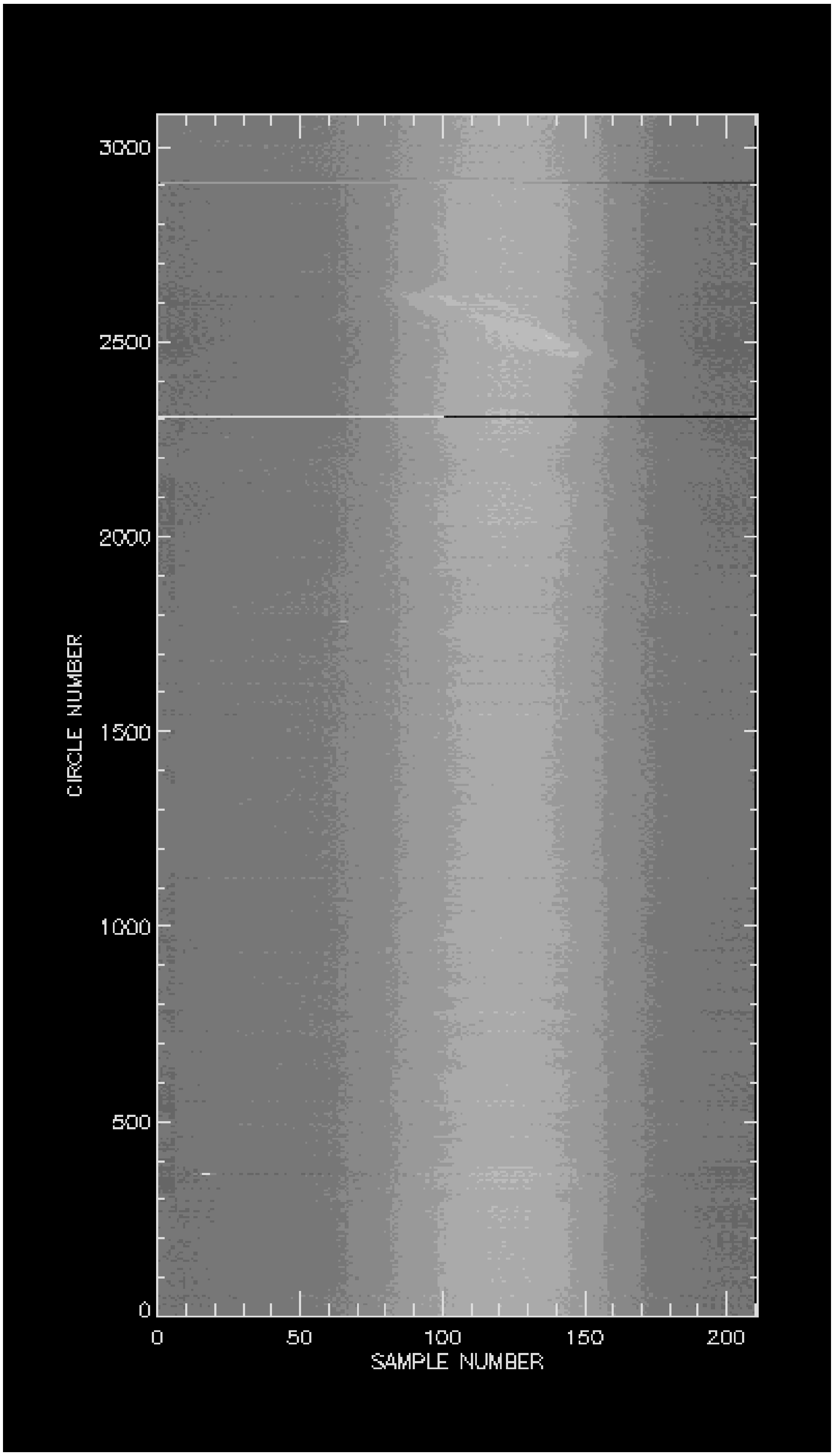


Figure 5: Image display of the signal of february 1 for the 13 GHz channel

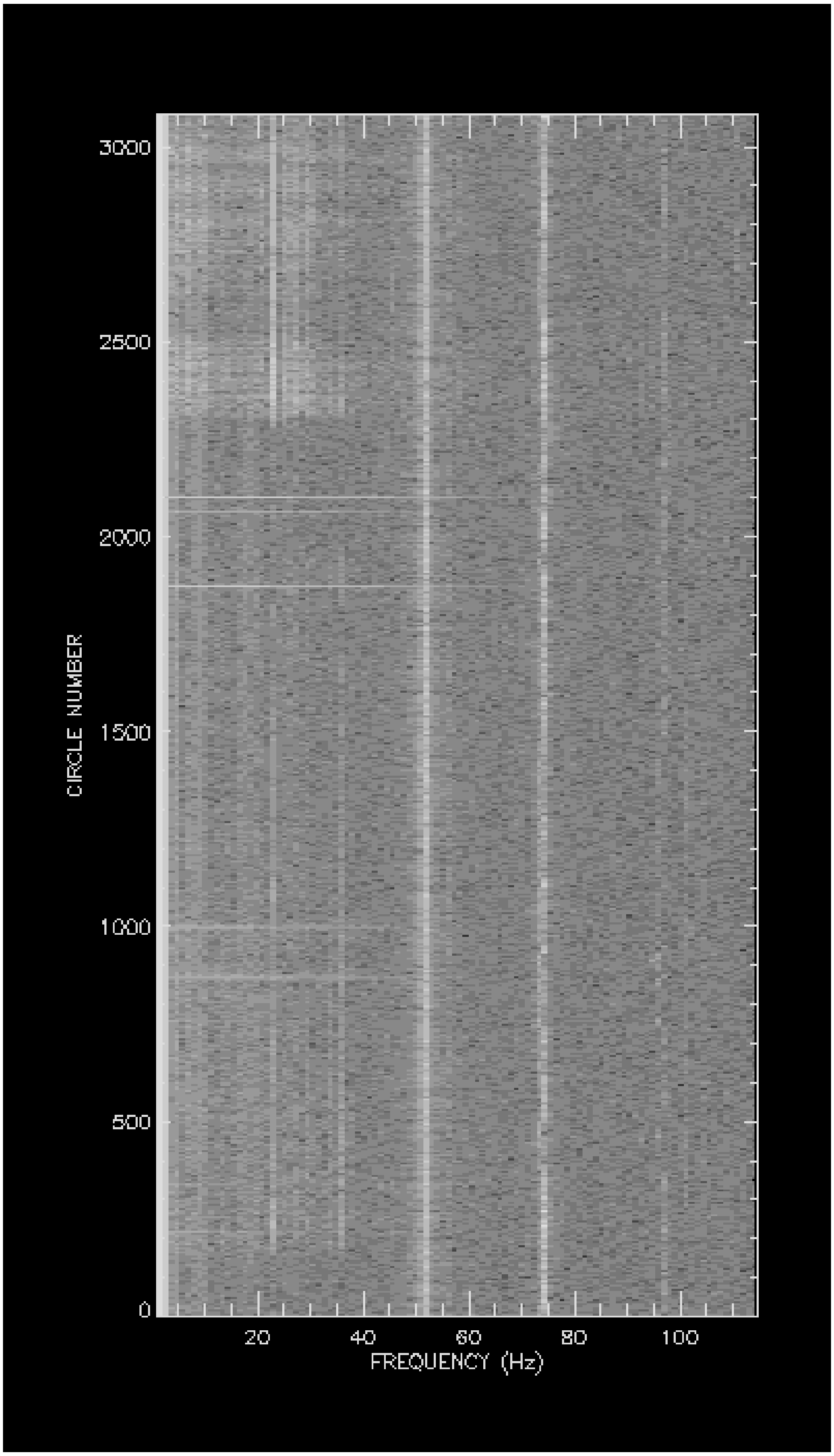


Figure 6: Time-frequency display of the signal of january 2 for the 13 GHz channel

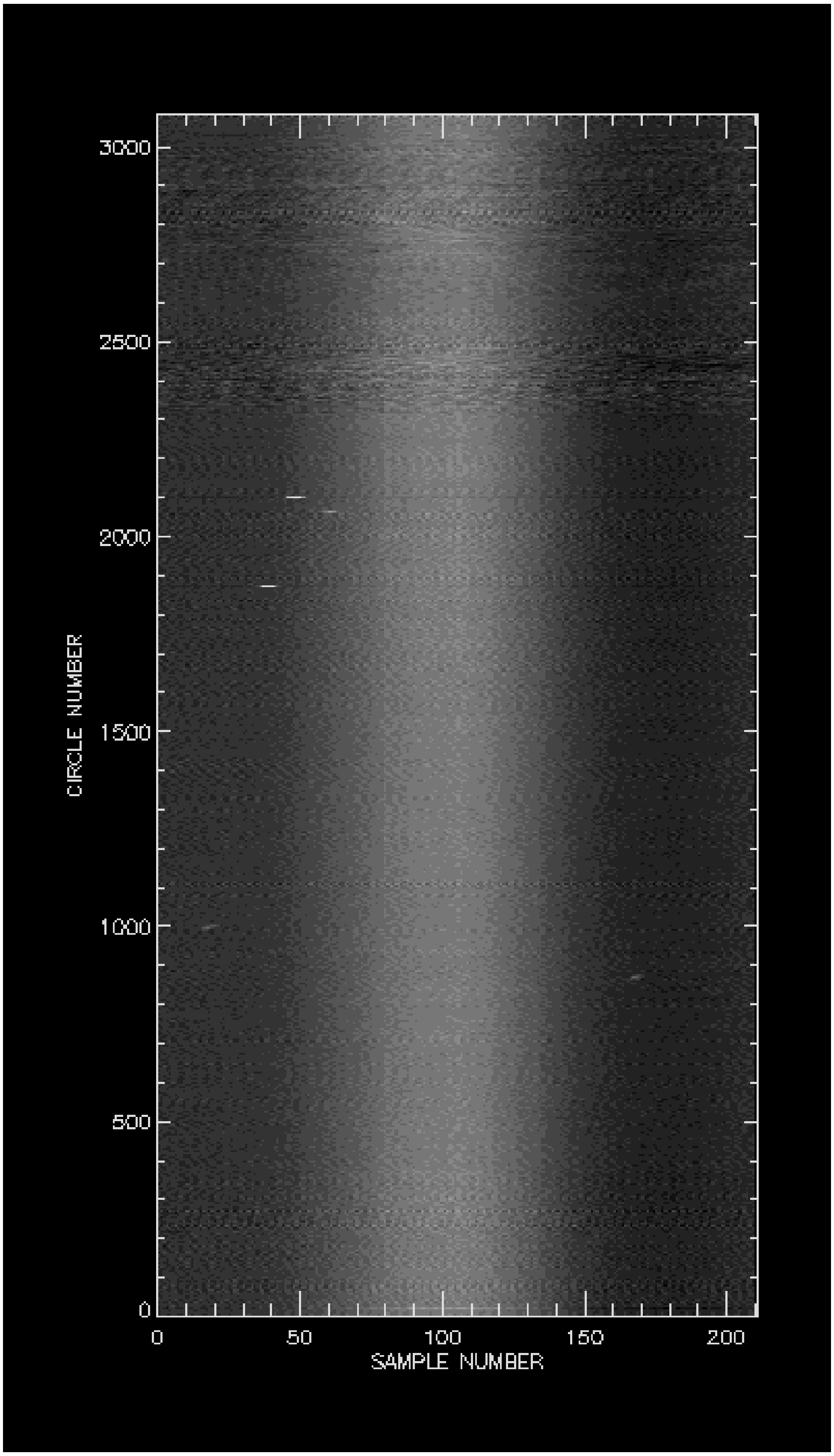


Figure 7: Image display of the signal of january 2 for the 13 GHz channel

We also notice in Figure 6 that the 23 Hz line fluctuates in amplitude over the whole day. A 36 Hz line is significantly present in the data as well as the 97 Hz discussed before.

Significant patterns in the signal spectrum can be seen below 20 Hz in both time-frequency panels (for both days), appearing as vertical stripes in the time-frequency panel. These patterns are not of astrophysical origin (they are significantly different in the two files) and are probably due to residual sidelobe pickup. They are much weaker on february 2, but still visible.

2 Raw data analysis

We use in this note the data taken on 1st february, 2000 at 13 GHz. We noticed the presence of 50 Hz signal and we removed this noise contribution. The bad circles have been removed. We see in Figure 8

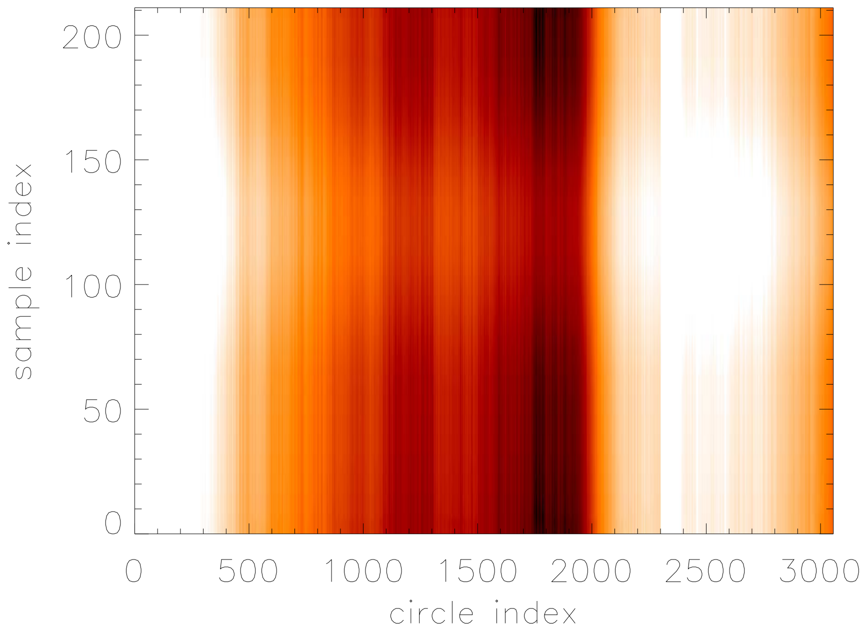


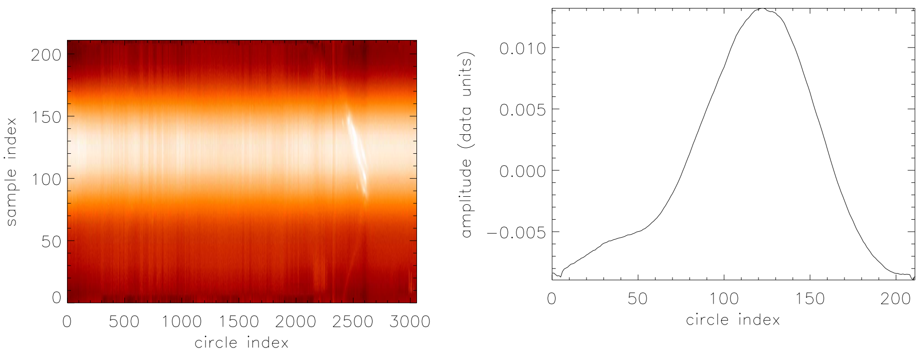
Figure 8: **Raw data**. The 50 Hz has been filtered out and the bad circles has been removed.

the presence of very low frequency drifts inducing an offset on each circle.

3 Destriping

A simple glance at the averaged circles (`lock_DDMMYY_HHmm.scans` files) show that the main contribution to the data is that of a systematic effect, probably mainly due to atmosphere and side lobes. This effect has been reduced by a factor of 5 (as compared with the data of may, 1999) since the shield on the floor.

The key point in the destriping achieved here is that the *shape* of this systematic effect is very stable, it only changes in amplitude. The first step is to have a good estimation of this shape. To do this, we temporarilly get rid of the baselines of all the circles removing the lowest frequency drifts visible as offsets on each circle, setting their averages to zero (see Figure 9(a)). We then first compute the simple arithmetic average of the data at a given angular position along the circles, for each position k (k ranges from 0 to 211). We obtain a crude approximation of a "template" (see Figure 9(b)), a common function depending only on k for all the circles. We model the measurements $m_{i,k}$ at sample



(a) Raw data after setting the average of each circle to zero.

(b) Shape of the template.

Figure 9: **Determination of the shape of the template**. In figure (a), the average of each circle has been set to zero. We see that the template is quite constant on time. The sources are now visible. A first crude estimation of the shape of the template is shown in figure (b); for each sample k, we compute the average over all the circles.

k on circle i as:

$$m_{i,k} = \underbrace{O_i}_{\substack{\text{baseline}\\ \text{for each circle}}} + \underbrace{a_i\, y_k}_{\substack{\text{amplitude}\\ \times\text{template}}} + \underbrace{s_{i,k}}_{\text{sky}}$$

and we want to estimate the amplitudes a_i and the offsets O_i given the measurements $m_{i,k}$ and the template y_k in order to construct the sky $s_{i,k}$.
The computation of this template is done iteratively using weights: in the first step, all the samples of all the circles ($\sim 3000 \times 212$) contributes with the same amount; we are going to use 3 maps: the raw data $M_{i,k}$, the raw data with zero averaged circles $z_{i,k}$ and the residual map $s_{i,k}$ obtained by subtracting to the raw data map the offsets O_i and the template. At the first iteration, all the measurements $m_{i,k}$ contribute with the same weight since we do not know at the moment what is in the map. Then we iterate:

1. using raw data, we fit the offsets O_i and the amplitudes a_i *given* the template y_k and using the weights;

2. we compute the residual map $r_{i,k}$. The sources appear in this map so that we can compute the new map of weights (see Figure 10). This map is obtained in this way: for each angular position k we make an histogram of the values $r_{i,k}$. We fit on this histogram a gaussian (representing the sky background) and a polynom (representing the sources). We affect a zero weight to points distant more than few sigmas (typically 2 or 3) from the average of the mean of the gaussian.

3. with this new map of weights, we make an histogram of the values $z_{i,k}$ *with a non zero weight.* We fit a gaussian on these values and the mean of this gaussian if the value y_k of the tenplate at angular position k.

4. go to first step of the iteration.

At the end of the iteration process we have a map of weights (it corresponds roughly to the clean data map in negative) and, of course, a clean data map. With this fit (amplitude times the template plus baselines), the sources are very well reconstructed (point sources as weel as extended sources) but there remain some residual in the background along the circles. It means that our hypothesis was not correct: the shape of the template is not so constant. So we add a perturbation to the template: a sine plus a cosine for instance and we fit, in addition to the offsets and the amplitudes of the template,

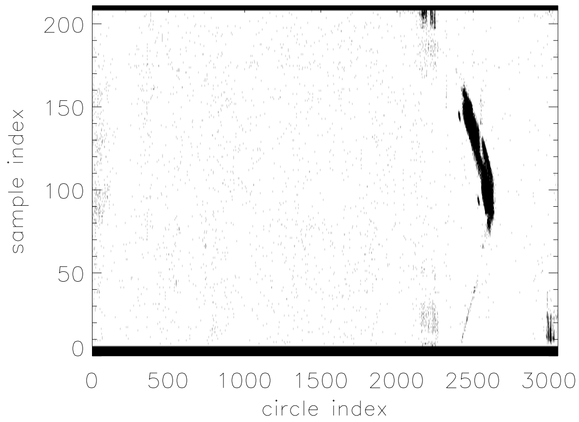


Figure 10: **Map of weights obtained after 10 iterations**. The black parts of this map has a weight set to zero, the white ones a weight set to one. This map permits us to localize the sources so that the determination of the template is not contaminated by the sources. This map is roughly the negative of the map of residuals (see Figure 11). In the file used here, the first and last samples of each circle are bad so that their weights are set to zero.

the amplitudes for the sine and cosine. With this new fit, the result is much better and the iteration converges after 5 or 6 steps and is shown in Figure 11.

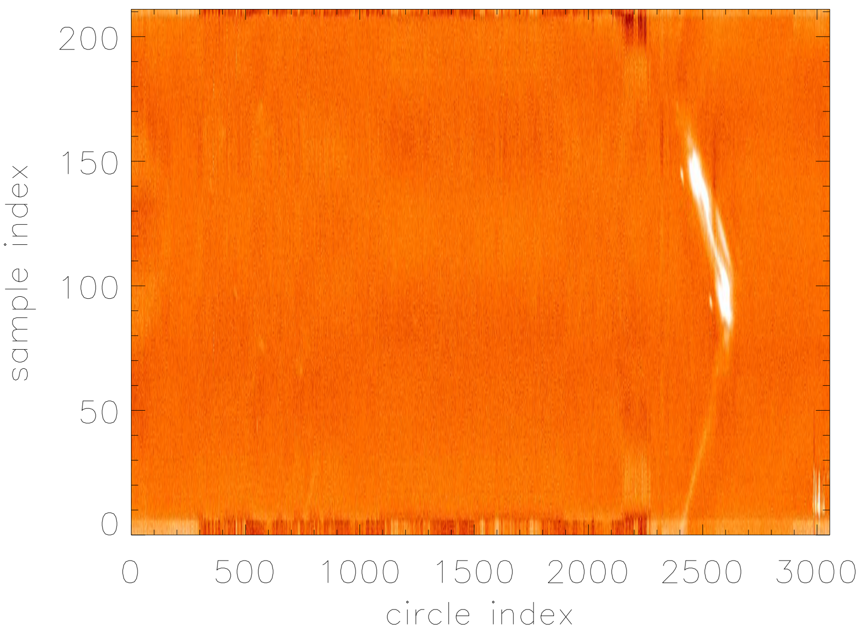


Figure 11: **Data map obtained after 10 iterations**. The template y_k has been iteratively estimated from the first shape given in Figure 9 using weights. We have fitted the offsets, the amplitudes of the template and amplitudes for the perturbations functions sin and cos. Do not pay attention to the first samples of all the circles (bad systematic data).

3.1 Phase determination

The determination of the phase using the galaxy has been explained in a previous note.
Here we use the presence of a point source in the data. This method is much more precise for determining the phase. During the scanning of the sky by the instrument, the declination of the circles is constant so that a point source is seen at two different moments t_S and $t_S + \Delta t$ corresponding to two angular positions along the circles such that their sum must be equal two 2π (see Figure 12). It

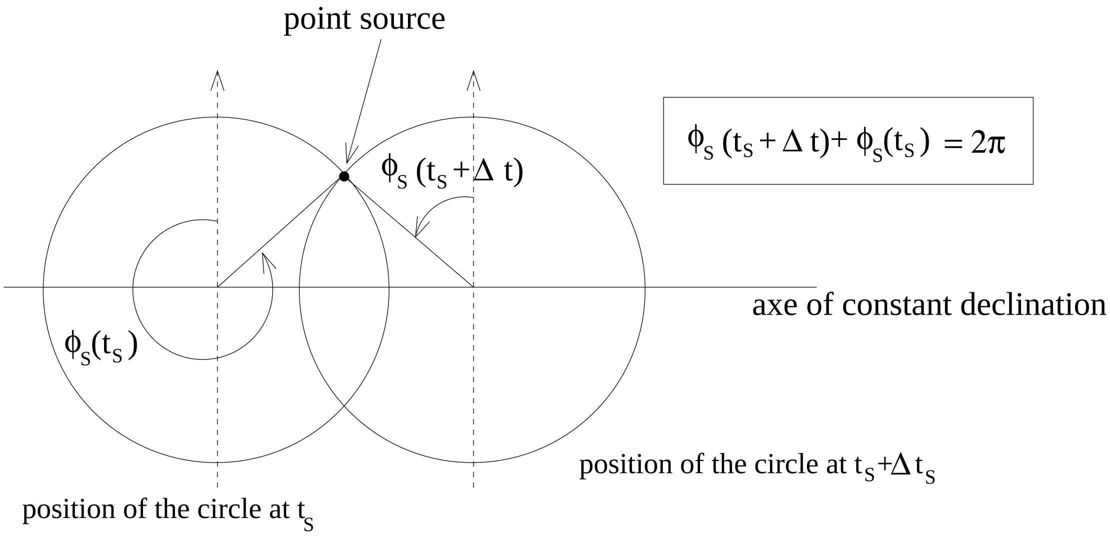


Figure 12: **Phase determination using a point source**. Given the scanning strategy of COSMO-SOMAS, a point source is seen twice and with angular positions such that the sum of the angles is 2π. These angles are symetric with respect to the axis represented in dashed line.

is easy to show that the shift in phase is given by:

$$\Delta\phi = 180^\circ \left(1 - \frac{\phi_S(t_S) + \phi_S(t_S + \Delta t)}{n_{\text{ech}}}\right)$$

where n_{ech} is the number of samples on the circles. The sample number 0 whose angular position was arbitrarily set to 0 is in fact at the angle:

$$\phi_0 = \Delta\phi \quad \text{or} \quad \phi_0 = 180^\circ + \Delta\phi.$$

The degeneracy between the two phases is lifted as soon as we know the declination of the source used*. This method is precise because we use a 2D gaussian fit to determinate the two angles $\phi_S(t_S)$ and $\phi_S(t_S + \Delta t)$.

4 Map making

Once we know the random phase, we can construct the sky map. The data used here cover the Cygnus region. The source at ra $\sim$ 20 h and dec $\sim 40^\circ$ is Cygnus A. The dense region close to this point source is the complex Cygnus X. Both are in the galactic plane. The projection of clean data is shown in Figure 13.

5 Beams

We determine the beams using the point sources seen in the data. We use here Cygnus A. We see this sources at two different moments so that we get two beams. These two beams correspond to two

*If the mirror rotates clockwise, the angle $\Delta\phi$ would be changed into its opposite.

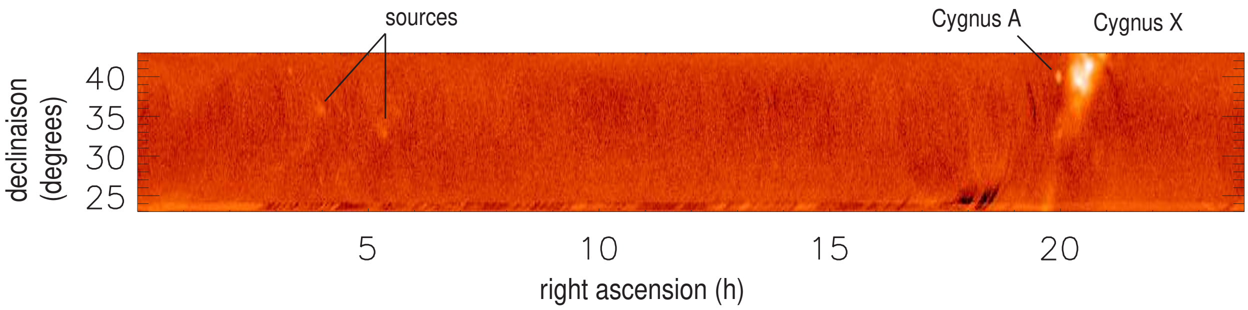
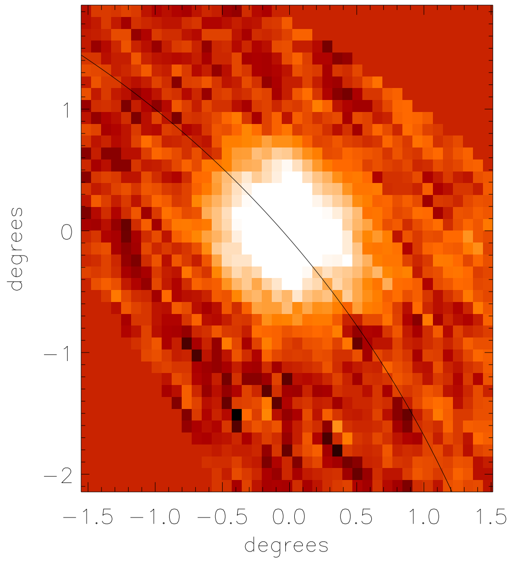


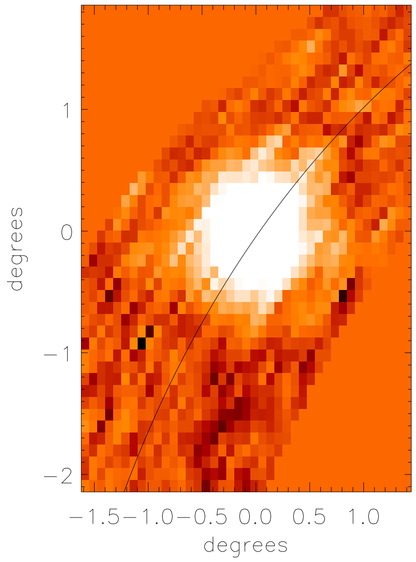
Figure 13: **Sky map obtained from the residuals shown in Figure 11**. After the determination of the phase, we can construct a sky map. The size of each pixel is $0.25° \times 0.25°$, and we have here 1440×81 pixels. The map covers 360° in right ascension and declinations 22° to 42°. The data has been oversampled in order that is pixel is seen at least once.

different orientations (in principle symetric) of the rotating mirror with respect to the detector.

We first make a zoom on the source, see Figures 14. We fit on each beam a 2D non circular



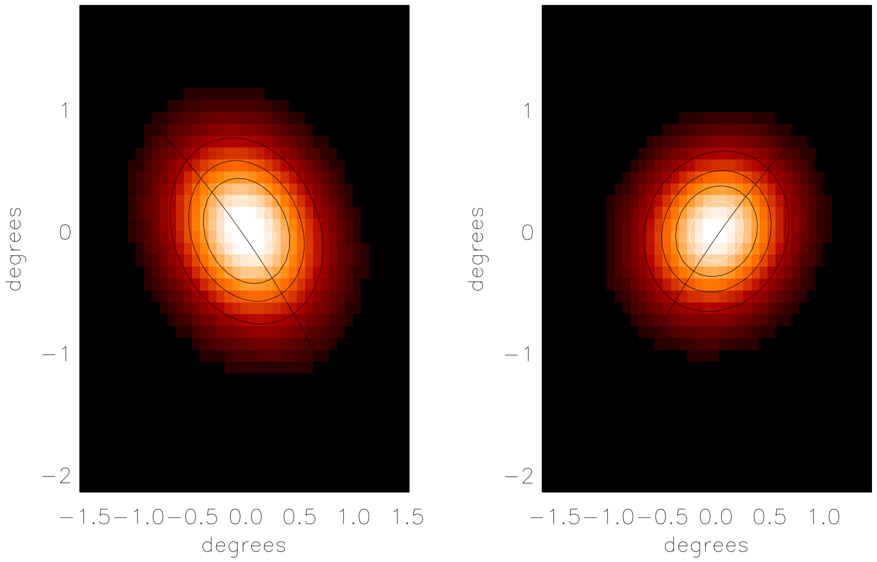
(a) First crossing on Cygnus A.



(b) Second crossing on Cygnus A.

Figure 14: **Cygnus A observations**. The pictures obtained give us two pictures of the beam. The superimposed line indicate the scanning direction. The data have been oversampled.

gaussian. The fit obtained are shown in Figures 15. s 15. The fit is very good: our beams are



(a) Fit of a 2D non circular gaussian on the first beam.

(b) Fit of a 2D non circular gaussian on the second beam.

Figure 15: **Fit of 2D non circular gaussians on the two beams**. In both cases, the fit is excellent. We see that the shape and size of the beam depend on the scanning direction and on the position of the rotating mirror with respect to the detector. The fit has been done over the oversampled beams.

gaussians. The caracteritics of these gaussians are:

$$B_i(x',y') = \frac{1}{2\pi\sigma^i_{x'}\sigma^i_{y'}} e^{-\left(\frac{x'^2}{{\sigma^i_{x'}}^2}+\frac{y'^2}{{\sigma^i_{y'}}^2}\right)}, \quad i = 1,2$$

$$\text{with} \quad \sigma^1_{x'} = 0.36°, \ \sigma^1_{y'} = 0.45°,$$

$$\text{and} \quad \sigma^2_{x'} = 0.39°, \ \sigma^2_{y'} = 0.35°$$

corresponding to

$$\boxed{\mathrm{FWHM}^1_{x'} = 0.86°, \ \mathrm{FWHM}^1_{y'} = 1.05°} \tag{1}$$

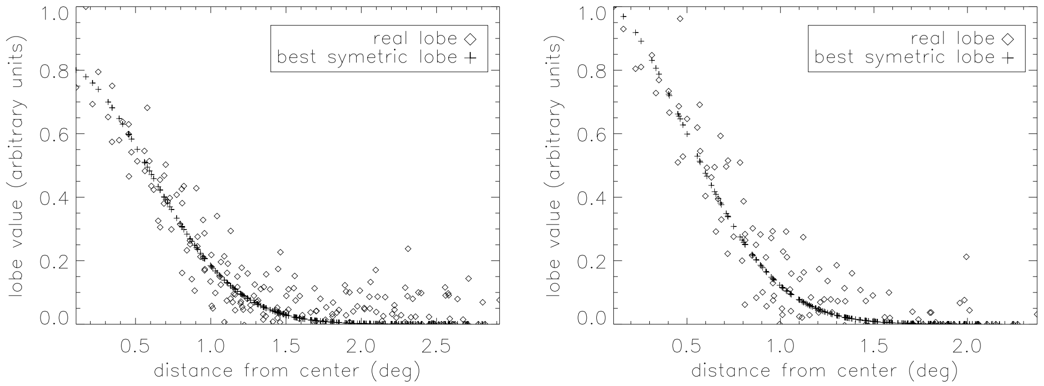
and

$$\boxed{\mathrm{FWHM}^2_{x'} = 0.92°, \ \mathrm{FWHM}^2_{y'} = 0.82°} \tag{2}$$

in a *rotated* reference system defined as:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} \cos\alpha_i & \sin\alpha_i \\ -\sin\alpha_i & \cos\alpha_i \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{with} \quad \alpha_1 \sim 34° \quad \text{and} \quad \alpha_2 \sim 38°.$$

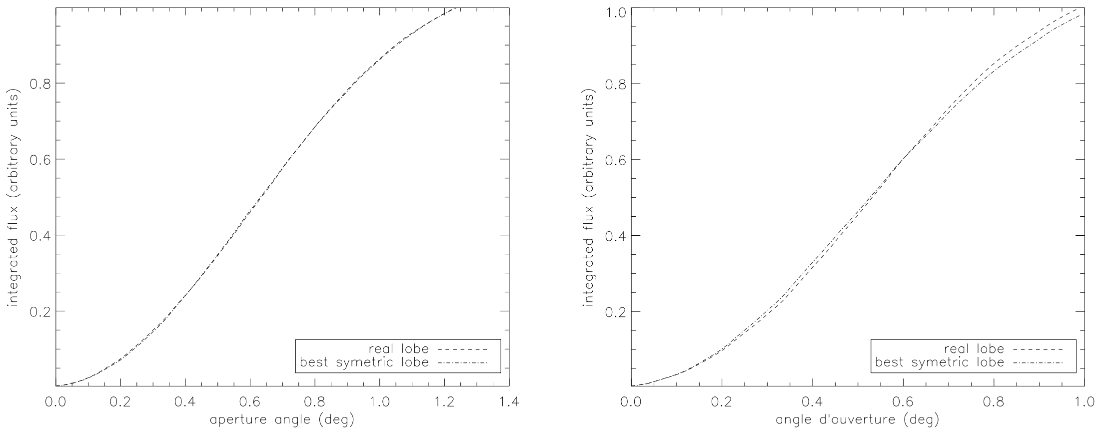
Later in the analysis, we will be interested in computing the power spectrum of the CMB temperature fluctuations. This work is much more easy when we use circular gaussian beam. We then compare our beams to the best circular beams obtained after a fit. These fits are shown in Figure 16. In order to estimate the error done when considering our beams as circular beams, we study the evolution of the integrated flux with respect to the aperture of the beam. The results are shown in Figures 17 and the computation show that the error is of the order of 5 pourcent.



(a) Fit over the first beam.

(b) Fit over the second beam.

Figure 16: **Comparison of our beams with circular beams**. We plot the repartition of the points of the beams with respect to their distance to the centers.



(a) Comparaison of integrated fluxes between circular beam and the first real beam.

(b) Comparaison of integrated fluxes between circular beam and the second real beam.

Figure 17: **Comparative aperture photometry between our real beams and the circular gaussian beams** .

6 Calibration using Cygnus A

6.1 Gain fluctuations correction

We first correct from gain variations using the "cal" signal; we just divide the raw data by the value of the cal signal for each circle.

6.2 Antenna temperature and emission of Cygnus A

We have the following relation between the temperature T_S of the source and its emission:

$$B_\nu(T_S) = \frac{2\,k\,T_S}{c^2}\,\nu^2 \quad \mathrm{W.m^{-2}.Hz^{-1}.sr^{-1}}\ ;$$

The total flux integrated over the solid angle of the source is simply:

$$B_\nu(T_S)\Omega_S = \frac{2\,k\,T_S}{c^2}\,\nu^2\,\Omega_S \quad \mathrm{W.m^{-2}.Hz^{-1}} \tag{3}$$

$$\widehat{=}\ S(\mathrm{Jy}) \times 10^{-26} \quad \mathrm{W.m^{-2}.Hz^{-1}}. \tag{4}$$

The flux of Cygnus A, in Jy is quite well known (see Figure 18). The analytic fit of the emission is

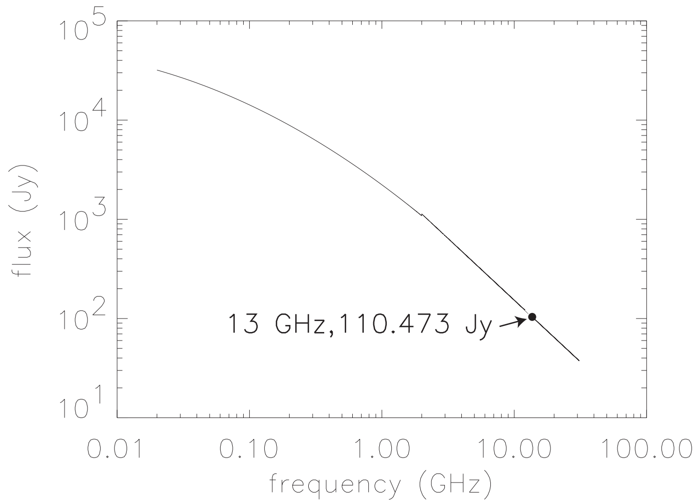


Figure 18: **Spectre de Cygnus A**. The emission of Cygnus A at 13 GHz is of the order of 100 Jy.

given by (IAC's formula):

$$\begin{aligned} \log S(\mathrm{Jy}) &= a + b \log \nu(\mathrm{MHz}) \\ \text{with} \quad a &= 7.161 \pm 0.053 \\ \text{and} \quad b &= -1.244 \pm 0.014. \end{aligned}$$

In our case, the flux is:

$$S = 110.47^{+32.03}_{-24.83}\ \mathrm{Jy} \quad \text{corresponding to an error of 30 \%.}$$

The source with temperature T_S induce a signal T_A in our antenna such that:

$$P = \frac{2\,k\,\nu^2}{c^2}\,T_A\,\Omega_A\ \widehat{=}\ \frac{2\,k\,\nu^2}{c^2}\,T_S\,\Omega_S$$

where Ω_A is the solid angle subtended by our beam. We can write:

$$T_A = \frac{\Omega_S}{\Omega_A}\,T_S \equiv \frac{c^2}{2\,k\,\nu^2\,\Omega_A}\,S \times 10^{-26}\ \mathrm{K}.$$

We have two measurements of the beam so we define an averaged beam $\widetilde{\Omega}$ defined by:

$$\frac{2}{\widetilde{\Omega}} = \frac{1}{\Omega_1} + \frac{1}{\Omega_2}$$

with $\Omega_i = 4\pi\sigma^i_{x'}\sigma^i_{y'}$, $i \in \{1,2\}$. The averaged expected antenna temperature is:

$$T_A = \frac{c^2}{2\,k\,\nu^2}\,S \times 10^{-26}\,\frac{1}{\widetilde{\Omega}}\ \mathrm{K} \tag{5}$$

and the error on this estimation is:

$$\frac{{\sigma_{T_A}}^2}{{T_A}^2} = \frac{{\sigma_S}^2}{S^2} + \frac{{\sigma_{\frac{1}{\Omega}}}^2}{\frac{1}{\Omega^2}}$$

With the two measurements of the beam, the value of $\sigma_{\frac{1}{\Omega}}$ is approximative; the best estimator we can make with the two measurements Ω_1 and Ω_2 is given by:

$$\sigma^2_{\frac{1}{\Omega}} \sim \frac{1}{4}\left(\frac{1}{\Omega_1} - \frac{1}{\Omega_2}\right)^2$$

and we obtain:

$$\frac{{\sigma_{\frac{1}{\Omega}}}^2}{\frac{1}{\Omega^2}} \sim 0.08.$$

The error due to the uncertainty over the flux of Cygnus A is:

$$\frac{{\sigma_S}^2}{S^2} \sim \frac{32^2}{110^2} \sim 0.08 \quad \text{so that} \quad \frac{\sigma_{T_A}}{T_A} \sim 0.4, \quad \text{or} \quad T_A = 75 \pm 30 \text{ mK}.$$

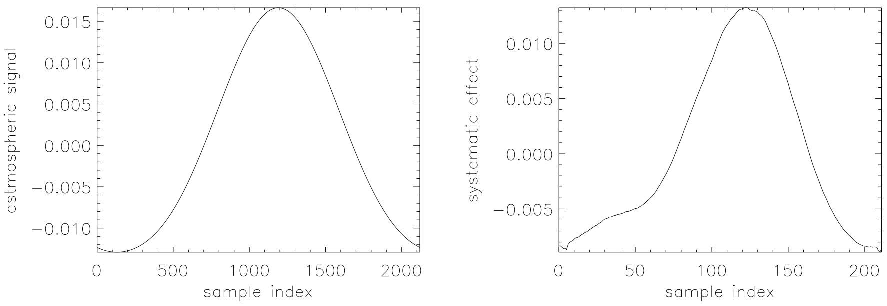
The corresponding value in data units is 0.0035 ± 0.0001 (0.0001 is the rms of the noise) and the calibration coefficient is:

$$\boxed{\alpha = \frac{75 \pm 30 \text{ mK}}{0.0035 \pm 0.0001} = 21 \pm 8 \text{ K.ud}^{-1}}.$$

The white noise has a rms of 2.1 ± 0.8 mK or 520 ± 20 μK per beam.

7 Systematic effect vs airmass effect

We didn't try in this analysis to include the contribution of airmass variations with respect to elevation. We should include this point in our analysis. The comparison of the shape of the systematic effect and of the airmass contribution is shown in Figures 19.



(a) Airmass effect.

(b) Systematic effect.

Figure 19: **Comparison of airmass contribution to the systematic effect contribution**.