

Feasibility and performances of compressed sensing and sparse map-making with *Herschel*/PACS data

N. Barbey¹, M. Sauvage¹, J.-L. Starck¹, R. Ottensamer², and P. Chanial¹

¹ SAp/Irfu/DSM/CEA, Centre d'Études de Saclay, Orme des Merisiers, Bât. 709, 91191 Gif sur Yvette, France
e-mail: nicolas.barbey@cea.fr

² University of Vienna, Department of Astronomy, Türkenschanzstr. 17, 1180 Wien, Austria

Received 17 September 2010 / Accepted 2 December 2010

ABSTRACT

The *Herschel* Space Observatory of ESA was launched in May 2009 and has been in operation ever since. From its distant orbit around L2, it needs to transmit a huge quantity of information through a very limited bandwidth. This is especially true for the PACS imaging camera, which needs to compress its data far more than what can be achieved with lossless compression. This is currently solved by including lossy averaging and rounding steps onboard. Recently, a new theory called compressed sensing has emerged from the statistics community. This theory makes use of the sparsity of natural (or astrophysical) images to optimize the acquisition scheme of the data needed to estimate those images. Thus, it can lead to high compression factors.

A previous article by Bobin et al. (2008, IEEE J. Selected Topics Signal Process., 2, 718) has shown how the new theory could be applied to simulated *Herschel*/PACS data to solve the compression requirement of the instrument. In this article, we show that compressed sensing theory can indeed be successfully applied to actual *Herschel*/PACS data and significantly improves over the standard pipeline. To fully use the redundancy present in the data, we perform a full sky-map estimation and decompression at the same time, which cannot be done in most other compression methods. We also demonstrate that the various artifacts affecting the data (pink noise and glitches, whose behavior is a priori not very compatible with compressed sensing) can also be handled in this new framework. Finally, we compare the methods from the compressed sensing scheme and data acquired with the standard compression scheme. We discuss improvements that can be made on Earth for the creation of sky maps from the data.

Key words. instrumentation: photometers – methods: numerical – infrared: general – methods: observational

1. Introduction

The *Herschel* Space Observatory (Pilbratt et al. 2010) is a three and a half year mission carrying instruments to observe in the far infrared. It is dedicated to investigating galaxy formation, evolution mechanisms, star formation, interaction with the interstellar medium, molecular chemistry in the universe, and finally chemical analysis of the atmospheres of bodies in the Solar System. With its 3.5 m telescope, *Herschel* is the largest space observatory to date. Its three instruments are the Photodetector Array Camera and Spectrometer (PACS) (Poglitsch et al. 2010), the Spectral and Photometric Imaging Receiver (SPIRE), (Griffin et al. 2010), and the Heterodyne Instrument for the Far Infrared (HIFI) (de Graauw et al. 2010).

Power and dissipation constraints, allow routine operations on *Herschel* to see only one instrument powered per observing day (OD). A special mode exists, called the SPIRE+PACS parallel mode (see below), where both SPIRE and PACS can be operated. The duty cycle of *Herschel* rests on the concept of the OD, which typically consists in a period of autonomous observations of 20–22 h, and a daily telecommunication period (DTCP) of 2–4 h during which the data recorded onboard is downlinked and the program of the next ODs is uplinked. (*Herschel* always stores the program of its next two ODs onboard, in case of problems during the DTCP.)

Limits to the amount of data that can be generated by the instruments are imposed by (1) the speed of the internal communication links between the instruments and the service module;

and (2) the speed and duration of the communication link between the satellite and its ground stations. These limits typically translate into an acceptable average data rate for the instrument during the OD of 130 kbit/s, while the PACS photometer, featuring the largest number of pixels (~2500) and a high sampling frequency (40 Hz), provides a native data rate of 1600 kbit/s, to which 40 kbit/s have to be added for housekeeping information. The situation is even more complex for the PACS spectrometer at 4000 kbit/s (but the integration ramps can be compressed onboard in a natural fashion), while SPIRE and HIFI have much lower data rates around or below 100 kbit/s, requiring no compression onboard.

We focus on the PACS imaging camera, which is made of two arrays of bolometers operating in the 55 to 210 μm range. The above restrictions mean that PACS data need to be compressed by a factor of 16 to 32. For this purpose, onboard data reduction and compression are carried out by dedicated subunits (Ottensamer & Kerschbaum 2008).

In this paper, we investigate alternative compression modes to attain this high compression factor without sacrificing noise or resolution. The study in this paper is based on the theory of compressed sensing and represents a real-world implementation of the ideas presented in Bobin et al. (2008). Simply put, compressed sensing is a promising theory that leverages the sparsity properties of data to allow measurements beyond the Shannon-Nyquist limit (Donoho 2006; Candès 2008). In the context of PACS, it offers the potential of optimizing onboard compression. It has the very strong advantage that it is not

computationally expensive on the acquisition side, a very important point for a space mission. An additional constraint for PACS is that CS had to be implemented after the conception of the instrument, while the instrument hardware cannot be changed as well as most of the onboard software.

We first present the PACS instrument, describe the properties of its data, and explain how we model the instrument. Then, we explain how we perform mapmaking using PACS data, which is a central stage in data analysis. Finally, we introduce our compressed sensing implementation and show how the map-making process can be modified in the compressed sensing framework to estimate state-of-the-art sky maps or better.

2. PACS data

2.1. The PACS imaging camera

The PACS imaging camera is made of two arrays of bolometers. One array is made of two 16×16 matrices forming a 32×16 array. It is associated with the “red” filter ranging from 130 to 210 μm . The other array is made of eight 16×16 matrices and thus has a shape of 64×32 pixels. It is associated with the filters “green” and “blue” ranging from 85 to 130 μm and from 60 to 85 μm , respectively. Figure 3 shows a single frame of the PACS blue array.

The PACS camera can be used in two modes: the chop-nod mode and the scan mode. The chop-nod mode consists in using an internal chopper to alternate between two sky positions (ON and OFF) and the satellite to exchange the ON and OFF beam positions. It was originally dedicated to the observation of point sources, with the chop-nod modulation used to get rid of the low-frequency noise originating in the detector. However, it turned out that even for point sources, scan-mode observations are more efficient in terms of achieved sensitivity for a given observing time. This mode is therefore not considered here. The scan mode scans the sky in order to observe larger areas than the detector field-of-view. When PACS is the only instrument powered, the prime mode, there are three possible scan speeds: 10, 20 and 60 arcsec/s. The PACS camera can also work in scan mode in conjunction with SPIRE, the parallel mode, in which case observations are driven by SPIRE requirements, and the scan speed is either 30 or 60 arcsec/s. Most of the PACS data are therefore obtained in scan mode, either in prime or parallel mode, and in either case, the bolometer arrays are read at 40 Hz.

The data are compressed using a lossless entropic compression scheme that yields up to a factor of four, requiring another factor of four (eight in parallel mode) to be achieved with lossy steps. For this purpose, four (or eight) consecutive frames are averaged before lossless compression. Additionally, a bit-rounding step is performed. It consists in rounding the averaging that is performed onboard so that the number of bits required to encode the data is reduced by one or two bits. This rounding is randomly up and down so as not to bias the results. In prime mode, one bit is removed while two bits are removed in parallel mode. We do not consider the bit-rounding operation in this article. In parallel mode an averaging of eight frames is required because the satellite bandwidth is shared with SPIRE. In fast parallel mode, this averaging results in a blurring of the frames and thus a loss of resolution in the sky maps, which is highly undesirable. Indeed, in this mode the scan speed can be 60 arcsec/s, or 18.75 pixels/s. Frames are taken at 40 Hz, making the displacement between two frames approximately 0.5 pixel. In parallel mode eight frames are averaged with a total motion of around four pixels. This is very significant compared to the point-spread

function (PSF) in the blue filter, which has a full-width at half maximum (*FWHM*) of 1.6 pixels.

This blurring is a very strong incentive to investigate other compression schemes. Presently, it can only be avoided by configuring the so-called transparent compression mode, where only one-eighth of the frame area is transmitted, but at full temporal resolution. With this mode we can experiment with all the stages of the acquisition while incorporating all features and artifacts of the actual instrument.

Indeed, when using a new acquisition or compression mode care must be taken that it will not decrease our ability to separate the signal of interest from instrumental artifacts. PACS imaging data have some properties that have to be taken into consideration. The most salient one is a large signal offset that varies for each pixel. The dispersion of the offset among the detectors occupies half of the range of values allowed by the 16 bits digitisation, whereas the typical signal range is around 15 bits with a noise entropy of 4–6 bits. This noise measurement is not a theoretical value, but the entropy of decorrelated measured data. It has been determined in two ways: computing the Shannon entropy of the difference signal minus 0.5 bit (Differentiation increases the sigma by a factor $\sqrt{2}$ or half a bit in entropy, which needs to be subtracted to get the original value) and computing the Shannon entropy of the high-frequency coefficients of a wavelet (for example Haar) transform. Both methods give almost the same value. This is essentially the readout noise, so it does not consider the pink component of the noise which can be seen at longer time scales. So, the quantization noise is not negligible with respect to the signal and adds to the other sources of noise. As can be seen in Fig. 3a, the signal of interest is not visible when the offset is not removed, and we cannot subtract it onboard. If the offset is removed (Fig. 3b), it is still hard to detect physical structure because of the noise. Some glitches can be identified however that could not be seen in Fig. 3a. It corresponds to the isolated pixels of high value.

Furthermore, there is a drift in this offset (Fig. 1), which can be seen as pink noise ($1/f$ noise). More precisely, it is a $1/f^{1/2}$ noise, as its power spectrum decreases with the inverse of the square root of the frequency. One of the main issues in PACS data processing is the separation of the signal from this drift. Its amplitude is the same order of magnitude as the signal of interest. In addition, this pink noise is affecting all the temporal frequencies in the PACS bandwidth. In other words, the transition between pink noise and white noise is happening at higher frequencies than the sampling frequency of PACS.

Once the offset is removed, other features become visible, as can be seen in Fig. 1. For instance, the bad behavior of a whole line of pixels results in frequent discontinuities of the intensity as a function of time (pixels 176 to 192 in the figure). A cosmic ray hit in the electronics of the last line around the time index 45 000 results in an additional drift of these pixels. Other cosmic rays can hit individual detectors and be the cause of wrong intensity values (outliers). However, those cosmic ray hits are affecting less than 1 percent of the data owing to the spatial structure of the detector elements (because they are hollow). Outliers due to cosmic rays are usually filtered out by comparing the set of data values corresponding to one map pixel and masking values beyond a threshold (see Sect. 5.2 for details).

The faint vertical structures in Fig. 1 are in fact the nucleus of the galaxy that passes in front of the detectors multiple times. Only the nucleus is bright enough to be visible in the offset removed data. Projection of the data onto the sphere of the sky, as well as filtering of the $1/f^{1/2}$ noise, is required in order to see the fainter parts. After the projection, the contributions of each

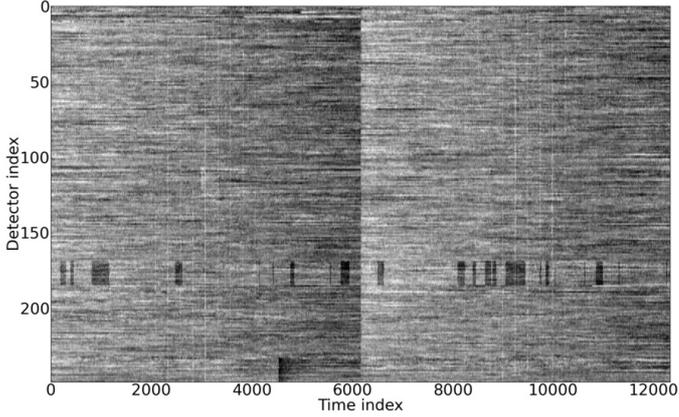


Fig. 1. Data set of the PACS imaging camera in the blue band in the transparent mode. This set consists of two cross-scans. The offset has been removed. All pixels have been reordered into a vector shape and displayed as a function of time (so that each pixel is shown as a horizontal line). Horizontal structures are the drift of the offset. The discontinuity in the offsets at the center of the figure comes from the temporal gap between the two cross-scans. An unstable line (pixels 176 to 192) results in frequent intensity discontinuities along both scans. At time index 45 000 the additional drift is due to a cosmic ray hit in the multiplexing electronics affecting a whole line. Finally, the faint vertical structures are due to the bright galaxy nucleus passing in front of the detectors.

pixel at each time index add up in order to increase the visibility of the fainter objects a lot.

2.2. Instrument model

Our optimal reconstruction scheme requires an understanding of the acquisition process, done here with an instrument model. For the rest of this article, the data sets and the sky maps are vectorized (i.e. re-indexed). Vectors are represented as bold lower-case letters. Linear operators are represented as matrices and written in bold upper-case letters.

The PACS data are a set of samples of the bolometers taken at different times and thus at different positions in the sky (in scan mode). Each sample is proportional to the sky surface brightness and to the area of the detector projected on the sky sphere (see Fig. 2). Thus, PACS data and the flux of the sky are linearly related so we can write Eq. (1) where \mathbf{y} stands for the data, \mathbf{x} denotes the sky pixels values and \mathbf{P} is the projector. The noise is considered additive and is denoted \mathbf{n} :

$$\mathbf{y} = \mathbf{P}\mathbf{x} + \mathbf{n}. \quad (1)$$

We consider the distortion of the field of view (FOV) by the optics of the instrument in the projector. The gains of the bolometers can be taken into account too. In our formalism, we would simply multiply \mathbf{P} by a diagonal matrix with the gain map on the diagonal (assuming that the gain does not vary with time, so the gain map would simply be the flatfield repeated n_f times, where n_f is the number of frames). The compression is taken into account separately.

In this paper, we used an implementation of the PACS instrument model provided by Chaniel (in prep.).

3. Mapmaking

The mapmaking stage is the inverse problem associated with Eq. (1), which is the estimation of \mathbf{x} knowing \mathbf{y} . It can be computed using classical estimation methods. We now review some

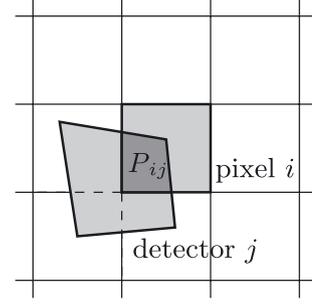


Fig. 2. Sketch of the intersection of the sky map pixel i and detector j . The matrix element (i, j) of the projection matrix \mathbf{P} is equal to the intersecting area of pixel i with detector j . The sky map pixels are square but the detectors are quadrilaterals because of the distortion of the optics. The distortion is exaggerated in this sketch.

of the methods that have been implemented for the mapmaking stage.

3.1. PACS pipeline

The *Herschel* interactive processing environment (HIPE)¹ offers backprojection of the data onto the sky map for the mapmaking pipeline step. To guarantee the conservation of the flux, the map must be divided by the coverage map (that is the number of times a pixel of the sky map has been seen by the detectors). The sky map estimation using the pipeline is thus described by:

$$\hat{\mathbf{x}}_{\text{pipeline}} = \frac{\mathbf{P}^T \mathbf{y}}{\mathbf{P}^T \mathbf{1}}. \quad (2)$$

Estimated sky maps are noted with a $\hat{\cdot}$ and \cdot^T is the transpose of a matrix, $\mathbf{1}$ a vector of ones. The division of two vectors here is the term-by-term division of their coefficients.

This method is very fast and straightforward to implement. But it does not provide a very accurate estimate of the sky. To illustrate this, one could compare this estimate to the least-squares estimate of the sky map. This latter estimate minimizes the criterion $\|\mathbf{y} - \mathbf{P}\mathbf{x}\|^2$ and can be written in closed form as in Eq. (3). The least-squares solution can be seen as the maximum likelihood solution under the assumption of independent and identically distributed Gaussian noise:

$$\hat{\mathbf{x}}_{\text{lsq}} = \left(\mathbf{P}^T \mathbf{P}\right)^{-1} \mathbf{P}^T \mathbf{y}. \quad (3)$$

The pipeline solution can now be interpreted as a very crude approximation of the least-squares solution by replacing the inversion of $\mathbf{P}^T \mathbf{P}$ by the inversion of the coverage map. This approximation ignores the correlations introduced by the projection matrix. In a more physical interpretation, one can say that the pipeline solution does not take the possibility of making use of superresolution from the PACS data into account. Reducing the pixel size in the pipeline sky map will not help in this matter. The projection matrix inversion is required. In fact, the PSF is not always sampled at the Nyquist-Shannon frequency and it is thus possible to obtain a better resolution in the sky map than in the raw data. As can be seen in Table 1, this is mainly beneficial for the blue filter (70 μm) in which aliasing occurs the most.

¹ “HIPE” is a joint development by the *Herschel* Science Ground Segment Consortium, consisting of ESA, the NASA *Herschel* Science Center, and the HIFI, PACS and SPIRE consortia. (See <http://herchel.esac.esa.int/DpHipeContributors.shtml>)

Table 1. PACS photometer overall characteristics/performances (taken from the PACS Observer’s Manual).

Filter	70 μm	100 μm	160 μm
wavelength range (μm)	60–85	85–130	130–210
resolution ($\lambda/\Delta\lambda$)	~3		~2
pixel size (arcsec)	3.2		6.4
FOV (arcmin)	3.5 \times 1.75		
<i>FWHM</i>	5.2	7.7	12

Notes. The resolution is the spectral resolution of the bandwidth.

3.2. MADmap

There is another approach called MADmap (Cantalupo et al. 2010), which tries to better estimate the sky map. It makes use of a preconditioned conjugate gradient in order to efficiently estimate the least-squares solution. Furthermore, MADmap adds a more accurate model of the pink noise. The noise is modeled as a multivariate normal distribution with a covariance matrix D diagonal in Fourier space (Eq. (4)). Denoting F as the matrix of the Fourier transform along the time axis, F^\dagger its conjugate transpose (and also its inverse), “diag” the operator that transforms vectors into diagonal matrices and $\tilde{\cdot}$ the Fourier transform of a vector or a matrix, we have

$$\mathbf{n} = \mathcal{N}(\mathbf{0}, D), \quad D = F^\dagger \tilde{D} F = F^\dagger \text{diag}(\tilde{d}) F. \quad (4)$$

This allows storing the full covariance matrix in an array of the size of the data, by storing the diagonal elements of the Fourier transform of D , which we denote here as \tilde{d} .

The map estimation in this case can be rewritten as a simple least-squares estimation following Eq. (5).

$$\begin{aligned} \mathbf{y}' &\leftarrow D^{\frac{1}{2}} \mathbf{y} \\ \mathbf{P}' &\leftarrow D^{\frac{1}{2}} \mathbf{P} \\ \hat{\mathbf{x}}_{\text{MADmap}} &= (\mathbf{P}'^T \mathbf{D} \mathbf{P}')^{-1} \mathbf{P}'^T \mathbf{D}^{\frac{1}{2}} \mathbf{y}' \\ &= (\mathbf{P}'^T \mathbf{P}')^{-1} \mathbf{P}'^T \mathbf{y}', \end{aligned} \quad (5)$$

where $D^{\frac{1}{2}}$ is such that $D^{\frac{1}{2}} D^{\frac{1}{2}} = D$.

The difficulty resides in finding a proper estimation for the noise covariance matrix. One has to assume that the noise covariance matrix is time invariant at very long time scales in order to have enough data samples for its estimation. In addition, one has to assume that there is only noise in the data that is taken to estimate the noise covariance matrix.

For *Herschel*/PACS data, we have data sets close enough to pure noise data that estimating the noise statistics is not a problem. In principle, the noise covariance matrix should be the size of the data set since the $1/f$ noise extends to the largest scales. In practice we can perform a median filtering on very large scales and use the noise covariance matrix estimation for intermediate scales. The median filter window can be made large enough to avoid filtering any scale corresponding to the physical sources of interest. This is easier for unresolved point sources, but even for extended emission the scales of interest do not exceed the size of the map.

Furthermore, the current MADmap implementation does not make use of an accurate projection matrix but instead assumes a one-to-one pixel association between sky map pixel

and data pixel. This limits the capacity of MADmap to take advantage of the super-resolution possibilities of PACS data since the projection matrix is not accurate. Another limitation of MADmap comes from the temporal-only Fourier transform. Because of it, the covariance matrix can only model temporal correlation, neglecting detector to detector correlations, which have to be removed by some pre-processing. However, most are gone after median filtering.

4. Compressed sensing

4.1. Theory

It is now well known that natural images and astronomical images can be sparsely represented in wavelet bases. This finding has been successfully applied to a number of problems and to compression in particular (JPEG2000 Skodras et al. 2001; or Dirac video compression Borer & Davies 2005). But, until recently, it has not been realized that it is possible to make use of this knowledge to derive more efficient data acquisition schemes. This is precisely the idea behind the compressed sensing (CS) theory. Compressed sensing states that if a signal is sparse in a given basis, it can be recovered using fewer samples than would be required by the Shannon-Nyquist theorem, hence the term compressed sensing. See Candes & Wakin (2008) for a tutorial introducing compressed sensing theory or Candès (2006) and Donoho (2006) for seminal articles. You can also look at Starck et al. (2010), which has a lot of example applications in astrophysics.

More formally, let us assume that our signal of interest \mathbf{x} is sparse in the vector basis B , and we are taking samples \mathbf{y} through the measuring matrix A (i.e. $\mathbf{y} = A\mathbf{x}$). In this context sparse means that $\mathbf{x} = B\mathbf{a}$ have few non-zero elements, and \mathbf{x} is said to be approximately sparse if it only has few “significant” elements (far from zero). The theory proposes two categories of measurement matrices that allow precise estimation of \mathbf{x} : incoherent matrices and random matrices. Random measurements are likely to give incoherent measurements. Incoherence is the fundamental requirement for A in the compressed sensing framework. Quantitatively, it means that the mutual coherence $\mu = \max_{j,k} \|\mathbf{m}_j^T \mathbf{b}_k\|$ is low where \mathbf{m}_j and \mathbf{b}_k are the columns of A and B respectively. Under the assumptions of exact sparsity and incoherence, it can be shown that \mathbf{x} can be estimated exactly using few measurements. Precisely, if m is the number of measurements, s the number of non-zero elements, and n the number of unknowns, we need $m \geq C \cdot \mu^2 \cdot s \cdot \log(n)$, where C is a constant (Candès & Wakin 2008). Under the approximate sparsity hypothesis, good performances can be demonstrated too (Stojnic et al. 2008).

In order to get the benefits of CS, it is assumed that it is possible to solve the inverse problem under the assumption of sparsity (i.e. to find a sparse solution to the inverse problem). In other words, we need to find the solution to problems of the type of Eq. (6):

$$\hat{\mathbf{x}}_{\text{CS}} = \arg \min_{\mathbf{x}} \|\mathbf{B}\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 < \epsilon, \quad (6)$$

where $\|\cdot\|_0$ is the L_0 norm, which is the number of non-zero elements of the vector, and ϵ is a low value linked to the variance of the signal. This estimation problem is of combinatorial complexity (all solutions need to be tested) and is thus intractable for our applications.

Since Eq. (6) is of combinatorial complexity, it is often replaced by Eq. (7) (see Bruckstein et al. 2009), where the L_0 norm

has been replaced by the L_1 norm ($\|\cdot\|_1$ is the L_1 norm) or the sum of the absolute values of the coefficients.

$$\hat{x}_{\text{CS}} = \arg \min_x \|\mathbf{B}\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 < \epsilon, \quad (7)$$

we can choose to write the estimation problem as in Eq. (8). It can be seen as a reformulation of Eq. (7) using Lagrange multipliers. Statistically, it can also be seen as the co-log-posterior (the opposite of the logarithm of the posterior distribution) of an inverse problem assuming a normal likelihood and a Laplace prior:

$$\hat{x}_{\text{CS}} = \arg \min_x \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{B}\mathbf{x}\|_1. \quad (8)$$

This modeling of the problem ignores some artifacts of PACS data such as glitches and pink noise. We will see how we can handle it in preprocessing.

4.2. PACS implementation of compressed sensing

We now describe how we implement the CS theory to PACS data in practice. Let us stress that we cannot modify the way *Herschel* observes and that we are severely limited by computer resources on board. We have chosen to apply a linear compression matrix \mathbf{C} to each frame. \mathbf{C} is the compressed sensing part of the acquisition model. The new inverse problem can now be described as in Eq. (9).

$$\mathbf{z} = \mathbf{C}\mathbf{P}\mathbf{x} + \mathbf{n}. \quad (9)$$

The severe onboard computational constraints mean we can use neither random measurement matrices nor wavelet transforms. We choose the Hadamard transform, which can be implemented in linear complexity using the fast Hadamard transform algorithm (Pratt et al. 1969). The compression is then made using pseudo-random decimation of the Hadamard transform of the frame. The Hadamard transform has been chosen since it mixes all the pixels uniformly. Thus, in conjunction with the decimation, it is a good approximation of a random matrix. To illustrate this, one can look at Fig. 3, which displays the result of the Hadamard transform of a single frame of PACS (without removing the offset).

Our method is not CS in a strict meaning. Because we have more data than unknowns, we cannot tell from the CS theory which compression matrix is optimal. But we think CS-friendly compression matrices are worth being investigated as there is no other theory that points to other optimal compression matrices under sparsity assumptions. We plan to test a wider range of compression matrices in future works.

We also point out that the standard compression scheme of PACS scans (averaging) can also be modeled as in Eq. (9) when replacing \mathbf{C} by an averaging matrix. The exact same inversion scheme can be applied to both the CS inversion problem and the standard inversion problem, allowing a clear comparison by separating effects of the acquisition matrix from those due to the sparsity constraint.

Equation (10) shows the averaging compression that is usually performed on PACS data. In this equation \mathbf{I} is the identity matrix of size the number of detectors:

$$\mathbf{C}_{\text{AV}} = \begin{pmatrix} \mathbf{I} & \dots & \mathbf{I} & & & \\ & & & \ddots & & \\ & & & & \mathbf{I} & \dots & \mathbf{I} \end{pmatrix}. \quad (10)$$

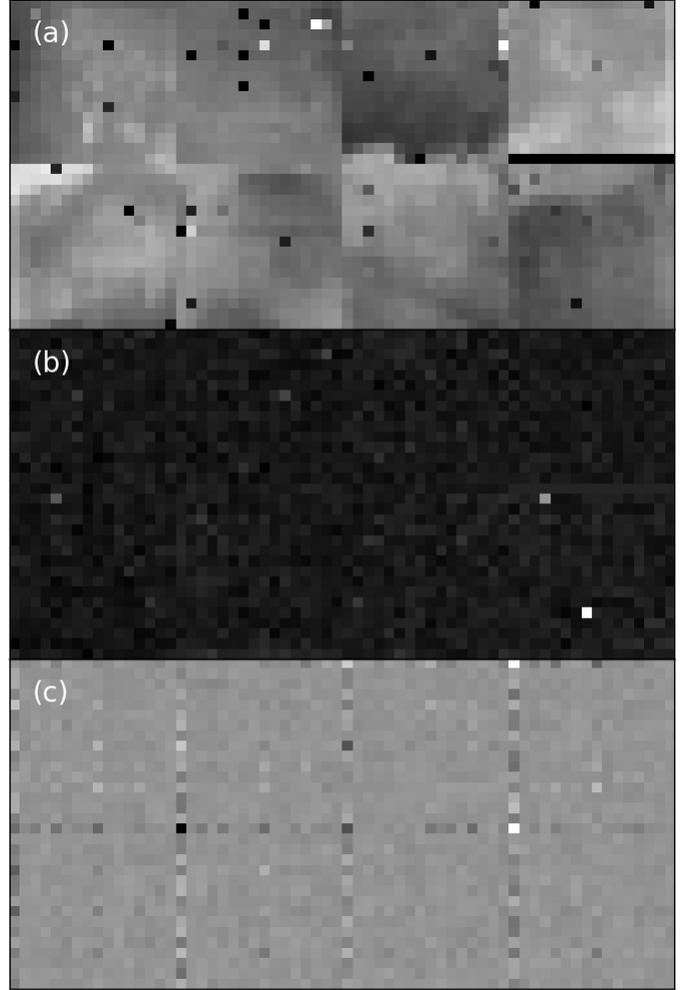


Fig. 3. a) Frame of the PACS imaging camera in the blue band. The darker and brighter isolated pixels are bad pixels. There is one complete bad line (at the bottom of the last matrix, ordered from the left to the right and from the bottom to the top). The observable structure is mainly the offset. Discontinuities in the offset correspond to the borders of the matrices of bolometers. There are physical gaps between matrices, as well as rotation and distortion. The matrix used in the transparent mode is the sixth one. b) The same frame with the offset removed. The brighter isolated pixels are glitches. c) The same frame after Hadamard transform. The offsets have not been removed.

Equation (11) shows the compressed sensing compression that we implemented, \mathbf{H} is the Hadamard transform of one frame, and \mathbf{M}_i a decimation matrix. A decimation matrix is a rectangular matrix with only one and zero values and exactly one non-zero element per row. It selects the elements of the Hadamard transform to be transmitted. For example, in the parallel mode, only one-eighth of the values are transmitted. The decimation matrices vary from one frame to the next to minimize the redundancy of the transmitted information, F is the compression factor:

$$\mathbf{C}_{\text{CS}} = \begin{pmatrix} \mathbf{M}_1 \mathbf{H} & & & & & \\ & \ddots & & & & \\ & & \mathbf{M}_F \mathbf{H} & & & \\ & & & \mathbf{M}_1 \mathbf{H} & & \\ & & & & \ddots & \\ & & & & & \mathbf{M}_F \mathbf{H} \end{pmatrix}. \quad (11)$$

All the matrices in this paper have not been implemented as matrices, but rather as linear functions, because it would not be feasible to store such large matrices and prohibitive to perform matrix multiplications.

We define sparse priors in both wavelet space and finite difference. The criterion can then be rewritten as in Eq. (12), where \mathbf{W} is a wavelet transform and \mathbf{D}_x and \mathbf{D}_y are finite difference operators along both axes of the sky map. The finite difference operators simply compute the difference between two adjacent pixels along one axis or the other. Thus, they favor smooth maps as a prior. Since we minimize a L_1 norm, it allows for some bigger differences, so the map can be piecewise smooth:

$$\hat{\mathbf{x}}_{\text{CSH}} = \arg \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{C}\mathbf{P}\mathbf{x}\|^2 + \lambda_W \|\mathbf{W}\mathbf{x}\|_1 + \lambda_x \|\mathbf{D}_x \mathbf{x}\|_1 + \lambda_y \|\mathbf{D}_y \mathbf{x}\|_1. \quad (12)$$

We estimate $\hat{\mathbf{x}}_{\text{CSH}}$ (CSH meaning Compressed Sensing for *Herschel*) using a conjugate gradient (CG) with Huber norm. The Huber norm is defined according to Eq. (13):

$$\begin{cases} \mathcal{H}(x) = x^2, & \forall x \in [-\delta, \delta] \\ \mathcal{H}(x) = 2|\delta x| - \delta^2, & \forall x \notin [-\delta, \delta]. \end{cases} \quad (13)$$

This norm is a way to approximate an L_1 norm ($\|\cdot\|_1$) with a function differentiable everywhere, which is mandatory for using CG methods. For this purpose, the Huber parameter δ can be chosen very small.

Finally, we minimize Eq. (14) as

$$\hat{\mathbf{x}}_{\text{CSH}} = \arg \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{C}\mathbf{P}\mathbf{x}\|^2 + \lambda_W \mathcal{H}(\mathbf{W}\mathbf{x}) + \lambda_x \mathcal{H}(\mathbf{D}_x \mathbf{x}) + \lambda_y \mathcal{H}(\mathbf{D}_y \mathbf{x}). \quad (14)$$

This optimization scheme is very similar to [Lustig et al. \(2007\)](#) in the idea of applying CG iterations and replacing the L_1 norm by a close function differentiable everywhere.

5. Preprocessing

It is important to note that, despite the significant alteration of the aspect of the data after CS compression (or any kind of linear compression), it is still possible to apply the same preprocessing steps as performed in the standard PACS pipeline. Let us demonstrate this in detail.

The following is usually applied to the data before the inversion step (backprojection or MADmap):

- removal of the offset by removing the temporal mean along each timeline;
- deglitching;
- pink noise filtering using median filtering (removes only the slowly drifting part in the case of MADmap).

5.1. Offset and pink noise removal

The removal of the offset is simple. Since the compression is linear, and applied independently on each frame, one can remove the offset by removing the temporal mean as in the averaging mode. The same applies for the pink noise removal.

To mitigate the effect of pink noise in the estimated maps, a filtering is performed during the preprocessing. Following the method of the pipeline, we use an high-pass median filtering. This can have the side effect of removing signal of interest on larger spatial scales in the map. This is of little importance in fields containing mostly point sources but can be dramatic in

fields with structures on every scale, such as fields taken in the galactic plane. One of the interests of MADmap is to be able to remove the pink noise artifacts in the maps without affecting the large-scale structures. But even with MADmap-like methods, one still have to remove the very low-frequency component of the pink noise not taken into account by the estimated noise covariance matrix. That is why it is important to ensure that one can perform median filtering in a CS framework.

We perform high-pass median filtering on the CS data exactly in the same way as done with the averaged data. Thanks to the linearity of the CS compression model, the pink noise in the raw data remains pink noise in the CS data. Thus, the median filtering affects the noise in the CS data the same way it does in the averaged data.

The only important parameter in the median filtering is the length of the sliding window used to compute the median in the neighborhood of each point. If the window is too large, there will be too much pink noise remaining in the filtered data. If it is too small, large-scale structures in the map will start to get affected. As a compromise, we choose the window to have the size of a scan leg, which is approximately the largest scale on the map. In the NGC 6946 maps this corresponds very roughly to a window of 1000 samples at a sample rate of 40 Hz at a speed of 60 arcsec/s. For the compressed data, we divide this length by the compression factor.

5.2. Deglitching

The deglitching is more complicated since it makes use of spatial information. To apply deglitching in the CS case, we first decompress the data without projecting onto the sky map (Eq. (15)).

$$\hat{\mathbf{y}} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{z}. \quad (15)$$

We can then deglitch on $\hat{\mathbf{y}}$ as we would have done on averaged data. We perform deglitching on $\hat{\mathbf{y}}$ using a method called *second-level deglitching*. It consists of regrouping all data samples that correspond to a given pixel map and masking out the ones that are beyond a given threshold using the median absolute deviation (MAD) method.

Any operation that is performed on standard data could be applied on CS data using Eq. (15). This is a very important advantage of using a linear compression method. Other compression methods, such as JPEG, do not satisfy this property, since a thresholding is performed. It is also the linearity of the compression that allows jointly performing the decompression and the mapmaking steps of the inversion, hence to obtain better maps.

When \mathbf{C} is an averaging matrix (the standard compression), the samples that are found to be glitches are simply masked. In our case, we define a masking matrix \mathbf{M} , which is simply an identity matrix with zeros on diagonal elements corresponding to where values are masked. We can finally perform the inversion replacing $\mathbf{C}\mathbf{P}$ by $\mathbf{C}\mathbf{M}\mathbf{P}$ in Eq. (14). This way, the inversion is done without altering the data set, but entries corresponding to glitches are not taken into account in the inversion.

As can be seen in the maps in Fig. 4, this is an effective method since no glitch remains visible. To ensure optimal detectability of the glitches, it is best to perform pink noise removal beforehand. One can use a shorter window for the median filtering to better remove the pink noise only for the deglitching step. Indeed, in this case, it does not matter if the map structures are affected since the filtered data are only used to compute a mask for the glitches. One can then apply the mask on the data filtered with a larger window to preserve large-scale structures.

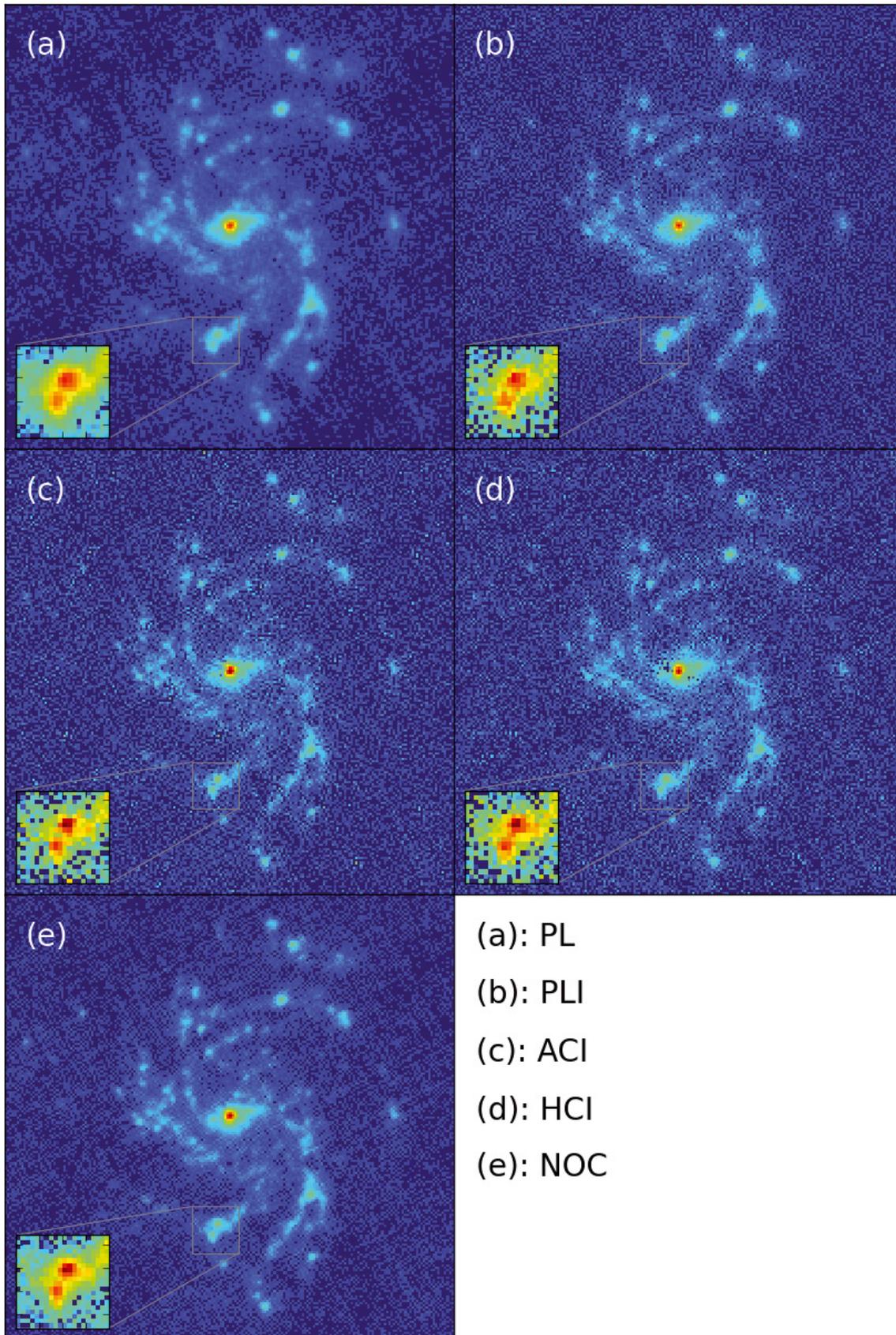


Fig. 4. Map estimations of NGC 6946 using **a)** the pipeline (PL) backprojection method, **b)** the pipeline model inversion (PLI) (without taking the compression into account), **c)** averaging compression inversion (ACI), **d)** Hadamard compression inversion (HCI), and **e)** reference map without any compression (NOC). Maps are presented on a fourth-root intensity scale. All maps have the same scale.

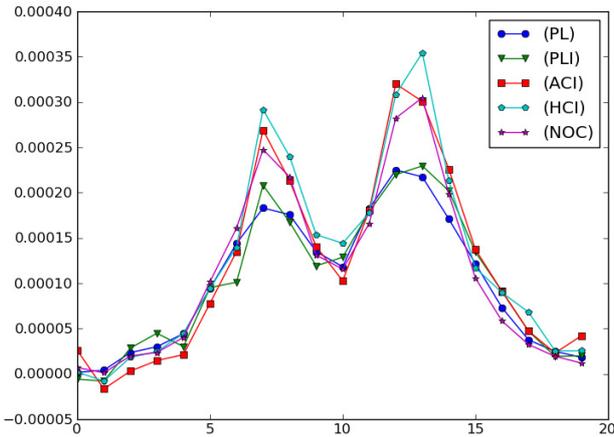


Fig. 5. Line cuts along the two sources displayed in the zoom of Fig. 4. Cuts are made by 2D linear interpolation of the maps. The x axis is the pixel index of the interpolation. The y axis is the intensity in Jy.

5.3. Conclusion

To summarize, we can perform the same processing as the pipeline in the CS mode. There is no obstacle remaining to prevent us from applying CS to PACS data. We now turn to the results we obtained with CS with real data.

6. Results

We present results of sky-map estimation using PACS data observations in transparent mode. The different modes of compression are simulated, where (PL) is the pipeline mode, i.e. a simple weighted backprojection ignoring that the data have been compressed (see Eq. (2)). Compression is done by averaging. (PLI) is a full inversion of the pipeline model. It means it does not take the correct compression model into account. Data for (PLI) have been compressed by averaging. In other words, (PLI) is the same as (PL) replacing the backprojection with a full inversion, (ACI) is the inversion of the model including the compression by averaging, (HCI) the inversion with Hadamard compression, and (NOC) is the inversion of the full data set (not compressed). All these methods but (PL) use the same sparse inversion algorithm as described in Sect. 4.2.

The compression factor on the acquisition side is always eight in the following results except for the reference map (Fig. 4e) for which no compression is applied. The data have been taken in fast scan mode (60 arcsec/s). Those are the conditions that give the biggest loss of resolution in standard (averaging) mode since the blurring of each frame is the biggest. The data set is made of two cross scans in the blue band of approximately 60 000 frames each. Each frame has 256 pixels (instead of 2048 because of the transparent mode). The estimated map is made of 192×192 pixels of 3 arcsec each, resulting in a 10 arcmin^2 map. The object observed is the galaxy NGC 6946 from the KINGFISH program (P.I. R. Kennicutt) but the observations are dedicated calibration observations. Results are presented in Fig. 4. Cuts along a line going through the two sources in the zoom are presented in Fig. 5.

To illustrate the impact on the spatial resolution of the different reconstruction methods, we list in Table 2 the mean of the $FWHM$, in pixels derived from Gaussian fits to four compact sources selected on the field. Although these sources are all not

Table 2. $FWHM$ in pixels of sources in Fig. 4 estimated using 2D Gaussian fits of 4 isolated sources randomly chosen from the map.

Maps	Maps statistics				
	(PL)	(PLI)	(ACI)	(HCI)	(NOC)
$FWHM$ (pixels)	4.30	4.01	2.75	3.27	3.15
std (10^{-6} Jy)	4.84	14.3	19.2	20.3	8.3
total flux (Jy)	0.65	0.62	0.64	0.67	0.63

Notes. Standard deviation for each map in Fig. 4, estimated from the corners of the maps ($\sim 50 \times 50$ pixels area at each corner for a total of $\sim 10\,000$ pixels in each map). Total flux of NGC 6946 in Jy as estimated from the maps in Fig. 4.

strictly point-like, comparing the achieved mean $FWHM$ is quite instructive.

To compare the global results quantitatively, we made radial profiles of NGC 6946 with maps from each method. Radial profiles were obtained using galaxy parameters from the NASA Extragalactic Database (NED): the position angle is 52.5° , the inclination 46° , the right ascension 308.71° and the declination 60.153° . Surface brightness was averaged over pixels contained in concentric ellipses. Background values were estimated using corners of the maps where no source can be seen. The standard deviation of the pixels in the background areas gives an estimate of the standard deviation of the noise in the whole map. Those values are reported in Table 2.

Results are presented in Fig. 6, and Fig. 6a shows radial profiles corresponding to each map from Fig. 4. Those profiles are supplemented with error bars estimated from the background measurements given in Table 2. We assumed normal noise meaning that the standard deviation of the surface brightness at each radius is given by $\sigma_{\text{profile}} = \frac{\sigma_{\text{bkg}}}{\sqrt{N(r)}}$, where $N(r)$ is the number of pixels in at radius r , and σ_{bkg} the standard deviation of the pixels in the background.

Figure 6b shows the ratio of those radial profiles over the radial profile of the reference map without compression. The reference map is not guaranteed to be perfectly correct. In particular, it is affected by the potential problems of the median filtering (used to remove pink noise) as much as the other maps. However, we assume it is closer to the ground truth since it does not suffer from a loss of information due to compression. We deduced the total flux of NGC 6946 from the maps in Fig. 4 and display the result in Table 2.

7. Discussion

There is a small gain in resolution between Figs. 4a and b. The $FWHM$ estimate in Table 2 illustrates this small gain, since the $FWHM$ drops from 4.3 pixels to 4.0 pixels. It comes from the inversion of the projection matrix. Physically, it means that we used the pointing information to get a super-resolved map. This result only holds for the blue band which is not sampled at the Nyquist frequency.

In Fig. 4, maps ACI and HCI are better resolved than maps PL and PLI. This is supported by the quantitative results presented in Table 2. Figure 5 shows that this gain in resolution allows close sources to be separated more. From Table 2, the gain in resolution is approximately 1 pixel, or 3 arcsec. This gain is due to the inversion of linear models, which take the compression step into account and to the inclusion of sparsity priors in the inversion algorithm. This is common to both ACI and HCI and not intrinsic to CS.

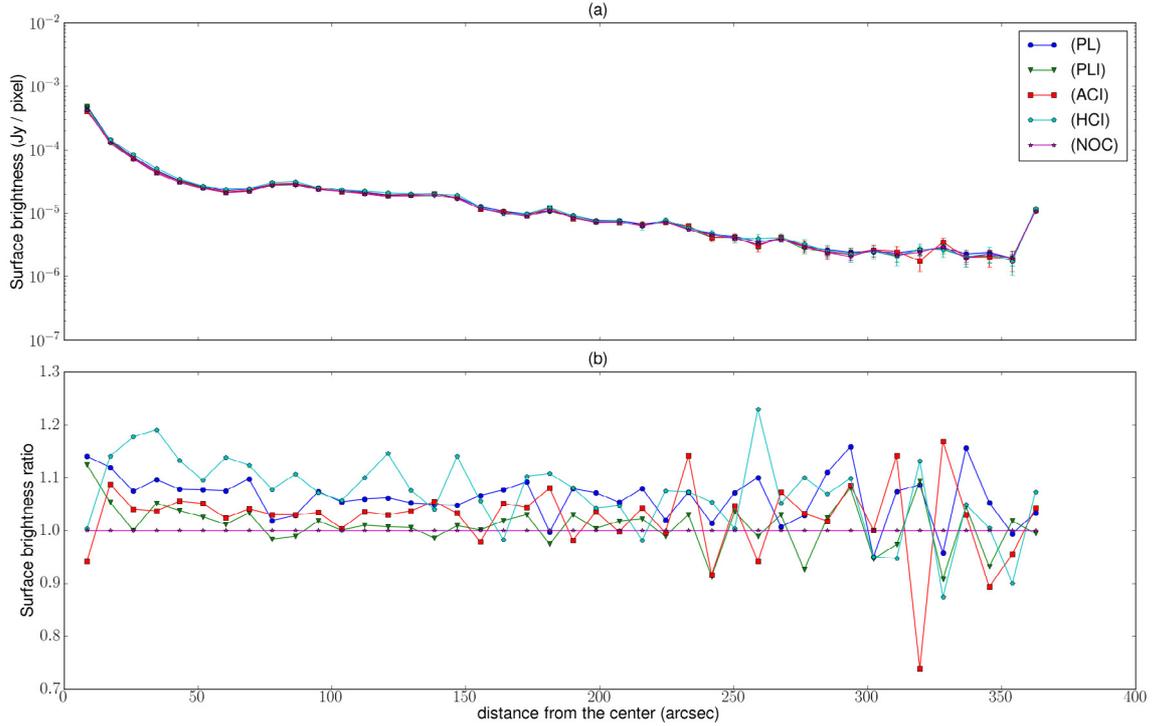


Fig. 6. a) Radial profiles of NGC 6946 in surface brightness (Jy) as a function of distance from the center of the galaxy (arcsec). Error bars are inferred from the background measurements reported in Table 2 and from the number of pixels at in the ellipses at each radius. **b)** Ratio of radial profiles of the different maps over the radial profile of the reference map obtained without compression. Radial profiles are obtained from the maps in Fig. 4. Acronyms refer to Fig. 4.

Comparing ACI and HCI in Fig. 4, it appears that the structure of the remaining noise is different. HCI does not exhibit linear patterns in the noise, following the scan direction. This is due to the mixing of the pink noise component of each detector done by the Hadamard compression. Furthermore, Table 2 shows that ACI has a mean *FWHM* below NOC. This is unexpected since NOC has more data than ACI. But it can be explained by the median filtering performed to filter the pink noise. Median filtering can have a small effect on compact sources even when large sliding windows are used. In contrast, HCI seems more robust to this median filtering since the mean *FWHM* in the HCI map is just slightly above the NOC mean *FWHM* and not below it.

Of course, these lossy compression schemes still imply a loss of information but it mostly translates into an increase of the uncertainty of the map estimation (i.e. the noise). This is true for both compression matrices.

We note, however, that comparing results between CS and averaging compression matrices does not allow us to conclude anything on the absolute performance of compressed sensing. To end on this point, a study is needed to find the best CS compression matrix. In this paper we only try to highlight the applicability of CS methods to real data from a space observatory.

The amount of residual noise can seem surprising if one does not remember that the noise in PACS data is mostly pink, so is not taken into account well in our inversion model. We need to use the noise covariance matrix in our inversion scheme. In order to do so, we can update Eq. (14) by considering a noise covariance matrix. This can be done using methods inspired by MADmap. We will investigate this possibility in a forthcoming article; however, since the resolution is increased in our results, the noise increase cannot be directly translated into a sensitivity decrease.

In Fig. 6a, we can see that the radial profiles obtained using the maps displayed in Fig. 4 are very close to each other. It gives confidence in the capability of our methods to preserve the flux in the maps on larger scales. In particular, this shows that the high-pass filtering is not affecting the profiles differently for CS than for averaging compression. It validates the way we perform this preprocessing step in the CS pipeline.

Figure 6b shows the relative discrepancies between the profiles. They are higher from the center since the surface brightness are very low, but they are consistent with the uncertainties. Again the CS compression scheme and the averaging compression scheme are very similar.

Profiles from compressed maps seem to be systematically slightly overestimated between 0 and 150 arcsec from the center, especially for (HCI) and (PL) maps. However, as shown in Table 2 the total flux of the galaxy computed for each map is within 7% of the flux computed from the reference map. This result is in full accordance with the photometric accuracy quoted by the PACS team (Poglitsch et al. 2010).

We presented results only for a resolved galaxy but our method could benefit other objects. For example, deep survey observations are limited by the confusion limit at which the presence of numerous unresolved sources appear as noise. We plan to see whether the confusion limit could be improved with our approach in future work. The gain in resolution could also help for the source blending issue and could help improve the precision of catalogs of isolated unresolved sources (as estimated by SExtractor Bertin & Arnouts 1996, for instance). This could be tested on simulated field of sources with known positions. See also Pence et al. (2010) for a study on compression method using this idea.

8. Conclusions

In this paper we showed that methods inspired by the compressed sensing theory can be successfully applied to real *Herschel*/PACS data, taking all the instrumental effects into account. For this purpose, special PACS data were acquired in the “transparent mode” in which no compression is applied. We then performed a compressed sensing compression scheme, as well as the standard averaging compression scheme, on the very same data, which allows for an unbiased comparison. The data are affected by various artifacts such as pink noise and glitches. Handling of these artifacts is well known on standard averaged data, but we show that the same kind of techniques can be applied successfully on compressed sensing data.

Furthermore, we showed that it is possible to significantly improve the resolution in the sky maps over what can be obtained using the pipeline. This is done by considering the compression matrix in the acquisition model and by performing a proper inversion of the corresponding linear problem under sparsity constraints.

For now the resulting sky maps in CS and averaging compression modes have the same quality. However, it is important to note that we may not be applying the best CS method. In particular, the Hadamard transform is only one compression matrix among others that satisfy the requirements for a CS application. Thus, this article does not in any way conclude anything about the performance of CS over standard compression. We can only conclude that this particular approach inspired by the CS framework and this particular Hadamard transform give results that are as good as the averaging compression matrix. But choosing another compression matrix could further improve the CS results over averaging.

We plan to more systematically analyze other choices for the compression matrix in a later article. Other improvements can be

expected by deconvoluting the instrument PSF during the map-making stage and including the noise covariance matrix as in MADmap.

Acknowledgements. We thank the ASTRONET consortium for funding this work through the CSH and TAMASIS projects. Roland Ottensamer acknowledges funding by the Austrian Science Fund FWF under project number I163-N16.

References

- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bobin, J., Starck, J., & Ottensamer, R. 2008, *IEEE J. Selected Topics Signal Process.*, 2, 718
- Borer, T., & Davies, T. 2005, *EBU Tech. Rev.*, 303
- Bruckstein, A., Donoho, D., & Elad, M. 2009, *SIAM Rev.*, 51, 34
- Candès, E. 2006, in *Proc. the International Congress of Mathematicians*, Citeseer, 3, 14331452
- Candès, E. 2008, *Comp. R. Math.*, 346, 589
- Candès, E., & Wakin, M. 2008, *IEEE Signal Process. Mag.*, 25, 21
- Cantalupo, C., Borrill, J., Jaffe, A., Kisner, T., & Stompor, R. 2010, *Ap&SS*, 187, 212
- de Graauw, T., Helmich, F. P., Phillips, T. G., et al. 2010, *A&A*, 518, L6
- Donoho, D. 2006, *IEEE Transac. on Information Theory*, 52, 1289
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3
- Lustig, M., Donoho, D., & Pauly, J. 2007, *Magnetic Resonance in Medicine*, 58, 1182
- Ottensamer, R., & Kerschbaum, F. 2008, in *SPIE Conf. Ser.*, 7019
- Pence, W. D., White, R. L., & Seaman, R. 2010, *PASP*, 122, 1065
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, 518, L1
- Poglitich, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, 518, L2
- Pratt, W., Kane, J., & Andrews, H. 1969, *Proceedings of the IEEE*, 57, 58
- Skodras, A., Christopoulos, C., & Ebrahimi, T. 2001, *IEEE Signal Process. Mag.*, 18, 36
- Starck, J.-L., Murtagh, F., & Fadili, M. 2010, *Sparse Image and Signal Process.* (Cambridge University Press)
- Stojnic, M., Xu, W., & Hassibi, B. 2008, in *IEEE Int. Symp. on Information Theory*, 2182