Mars 2008.

Statistical Image Deconvolution with uncertainties of observed maps of spectra

G. Rydbeck[1]

Onsala Space Observatory, Onsala 43992 SWEDEN

ABSTRACT

An imminent problem in mapping an astronomical object is that the true source image is convolved with the sensitivity pattern of the observing telescope , and, in addition, degraded by noise and other various sources of error. To find the image which is as close as possible to the true source image, observed images are analyzed from a probabilistic point of view. It is demonstrated that one's knowledge of the true image can be described by conditional probability distributions of possibly true pixel intensities. The distributions are derived using Bayes' equation. Based on the idea that a "perfect observer" must be totally impartial in the analysis of his data, a solution to the well-known problem of defining prior probabilities is put forth. Approximate equations for solving out the expected image intensities are then derived. The expected image is defined as the mean of all possibly true images that are allowed by the observational constraints. The expected true image is in effect a deconvolution of the observed image. The effectiveness of the deconvolution is greatly enhanced if also velocity data (i.e., the whole position-velocity data cube) is included in the analysis. The approach allows derivation of uncertainties of the deconvolved intensities.

Subject headings: methods:data analysis techniques:image-processing techniques:statistical techniques:numerical ISM: individual objects: Orion KL- ISM: molecules

1. Notation conventions

When not separated by a mathematical symbol, repeated indices are summed i.e. $\alpha_\nu\beta_\nu = \sum_{\nu=1}^{N} \alpha_\nu\beta_\nu$, while $\alpha_\nu \cdot \beta_\nu$ is not summed.

$\{\alpha_i\}_{i\in N}$ is a compact notation for the set $\{\alpha_1, \ldots, \alpha_N\}$. Boldfaced symbols represent a vector e.g. $\mathbf{a} = (a_1, a_2, \ldots, a_N)$, and $da = da_1 da_2 \cdots da_N$.

$\int_{R+} = \int_0^\infty \cdots \int_0^\infty$. Further we let $\langle\rangle$ represent expectation values so that $\langle a_\nu \rangle = \int_{R+} a_\nu P(\mathbf{a}) da$, where $P(\mathbf{a})$ is a probability distribution of $\mathbf{a}$.

A pixel intensity a_ν is the intensity a(x,y) averaged over pixel ν i.e.,

$a_\nu = \int_{pix_\nu} a(x,y) dxdy$

Unless explicitly stated otherwise, beam-width, BW or $HPBW$ always refers to $FWHM$, i.e. full width at half maximum. For a symmetric Gaussian beam

$$B(x,y) = \frac{1}{\sigma_{beam}^2 2\pi} \exp -\frac{x^2+y^2}{2\sigma_{beam}^2} \tag{1}$$

$HPBW = 2\sqrt{\ln(4)} \times \sigma_{beam} \approx 2.35 \times \sigma_{beam}$, where σ_{beam} is the Gaussian width.

2. Introduction

2.1. Background

In mapping astronomical objects with an instrument, the true source intensities over the map are corrupted by the limited resolution of the telescope, the limited sampling rate, noise addition to the signal and receiver gain fluctuations etc. A great number of schemes have been developed to recover true source intensities (See eg. Starck et al. (2002)). Most of these have been based on variations of the maximum entropy method (MEM) others on Bayes' theorem or a mixture of these. It has been suggested (cf e.g. Bryan and Skilling (1984)) that one may define an entropy in terms of map-pixel intensities I_j so that

$$S = -\sum_{\nu=1}^{q} I_\nu \ln I_\nu \tag{2}$$

where q is the number of map pixels. Sometimes one has proposed

$$S = -\sum_{\nu=1}^{q} \ln I_\nu \tag{3}$$

It is shown in appendix E and F that these forms of entropy are of the kind where the intensities I_j are actually expected intensities of the typical exponential distributions obtained for quantized thermal heat baths in equilibrium. The assumptions underlying heat bath physics or thermal physics is that the system under consideration consists of granules or atoms of some kind and that these can assume quantized states of energy (or in this case emission intensity). The introduction of these scales, 1) the spatial scale of the granularity

and two the 2) the scale of the emission intensity levels are totally artificial in the context of image analysis of objects which are not associated with any such scales. The resulting set of images, constrained by artifial scales, would simply not be sufficiently rich. It is certainly not rich enough to allow the dispersion to be determined by observational accuracy since it is already determined by the fixed exponential distribution. The conclusion is that the above forms of entropy are not relevant for the type of problem considered here. Details of these problems are further discussed in appendix F.

The idea of maximizing the entropy to find a probability distribution is still attractive. To be precise we let an image be defined as a map of source intensities. The entropy of a set of N images is simply $\log(N)$. If N is the number of possibly true images that could give rise to an observed image, maximizing the entropy is then synonymous to saying that N should include all images that are not excluded by the observational constraints. One may express this entropy in terms of probabilities P_i that "i" is the true image. The entropy "S" is then

$$S = -\sum_{i=1}^{N} P_i \ln P_i \tag{4}$$

The equation is somewhat diffuse however since the image density is left unspecified, or in other words, how the i'th image is selected. One may write eq. (4) in terms probabilities of intensities $(a_i)_\nu$ of the individual map positions or pixels.

$$S = \sum_{\nu=1}^{q} \sum_{i=1}^{\infty} P_\nu(a_i) \ln P_\nu(a_i) \tag{5}$$

Where i denotes discreet intensity levels and ν map positions. The equation is more problematic than eq. (4) since both pixel-size (granularity) and level increment is unspecified. Introducing a level density $\rho(a)_j$, eq. (5) can be written

$$S = \sum_{\nu=1}^{q} \int_0^\infty da\ P_\nu(a) \ln P_\nu(a)/\rho_\nu(a) \tag{6}$$

This equation is equivalent with eqn. (1) in Caticha and Preuss ((2004)), identifying $\rho_j(a)$ with a prior density. For this equation to be valid, it is clear that the granularity problem must be overcome.

The statistical and Bayesian approach to deconvolution taken by Ariel Caticha and Roland Preuss ((2004)) is similar to ours. Their approach does not seem to include intensity scale invariance for the prior however. Including this in our work leads to a reduced approximate prior which is compact and easy to apply. Actual deconvolution codes are derived producing expected intensities and dispersions.

As pointed out by Caticha and Preuss the selection of a prior is a controversial and difficult issue which has generated a vast literature (see e.g. Kass and L. Wasserman (1988) and references therein). The general approach in MEM is to maximize the entropy expression subject to observational constraints thereby obtaining the $P_j(a)$'s. Another route is to use MEM with some basic constraints to select a prior and then use Baye's equation to apply observational constraints.

It seems however, that still after all these years, a satisfactory understanding of the problem of recovering true source intensities has not been reached.

In the present approach it is suggested that a particular kind of scalefree prior, removing all traces of granularity, will provide a satisfactory solution. Preliminary notes and applications of the approach, developed in the early 1990's have been published earlier (e.g. Rydbeck et al. (1993) and Rydbeck (2000)).

One well known maximum entropy related method is the Pixon image restoration, (R.C. Puetter and R.K. Pina (1994)). If I understand the Pixon method correctly it avoids the problem of granularity and the size and form of local regions by postulating that these can be described by Pixons, functions of a certain form and size, which may vary or be constant over the map, depending on need or approach. These Pixons may then be considered as a non-orthogonal set of base functions in which the deconvolved image can be expanded. The image is found by minimizing the number of Pixons needed and maximizing the fit to the data. Minimizing the number of Pixons (the external no. of degrees of freedom) maximizes the Pixon size and the individual support of a Pixon amplitude by the data. All this makes reasonable intuitive sense but the logic behind the approach seems rather complicated especially concerning the techniques to derive the Pixon size and form from spatial correlations. The connection of all these steps to Bayes' relation appears somewhat unclear. On the other hand the approach makes intuitive sense and it seems experimentally clear that the approach can produce very good deconvolved maps.

The approach to deconvolution presented in this work is very different in basic philosophy to the Pixon approach. The methods may in the end be approximately equivalent in the same sense as macroscopic thermodynamics is equivalent to microscopic thermodynamics.

2.2. A summary of the central ideas of this work

The aim of this work to define correct probability distributions $P_j(a)$ of the pixel intensities $\{(a)_j\}$ and to use these to derive expected map intensities and dispersions. The most basic and important proposal guiding us in this pursuit is that the approach must be totally

impartial or objective. It means that we must not introduce the artificial scale constraints inherent in many MEM methods. It means in fact that we (the investigator) must be totally impartial with respect to spatial (cf Axiom 2 Caticha and Preuss (2004)) and map intensity scales and let these be determined by observational results and only by observational results. More precisely we must be able to form a prior distribution of maps which is scale-free in both intensity and space.

Assume then that observations are made for example with a single-dish telescope with a certain beam towards a certain source. Assume further that the observed intensities are corrupted by noise fluctuations with known properties, and that the observations result in a map of intensities.

The work described in this paper is based on the following four propositions:

1. *It is possible to form, in a general sense and independent of any observation, a totally scale-free ensemble* [1] *of* <u>*all possible*</u> *maps of intensity distributions.*

2. *The above ensemble constitutes the most impartial choice for a prior ensemble of maps.*

3. *The only complete way to represent the result of the above observations is to form, from the above ensemble, the (sub-)ensemble of all possibly true map intensities (EPTM) which, when observed, may lead to those results.*

4. *In those cases where the possibly true map intensities are gathered around their mean values, the single map which best represents the set of possibly true maps is the map of mean (mean of EPTM) intensities.*

The mean of the EPTM is called the *expected map.* The variance of the EPTM, or the uncertainty map, may also be formed.

The above recipe for forming the EPTM is simply a special case of Bayes' relation (see e.g., Stuart and Ord (1987) or Caticha and Preuss (2004)). The most difficult part of the approach is that of forming the prior ensemble of maps. A solution is proposed in section 3 .

---

[1] it appears that set and ensemble are equivalent words, statisticians seem to prefer the former and physicist the latter .

2.3. Bayes' relation

It has long been recognized that a fully satisfactory analysis of observed data must be done in a probabilistic framework. One available road to accomplish this is to use Bayes' equation (cf. Stuart and Ord (1987), especially paragraph 8.6 p 283, vol 1). With it one may relate the probability distributions of observed variables to observational results. Bayes' equation and related quantities, are briefly introduced below. Some further discussions on probabilities are given in appendix E. For a full introduction to this somewhat controversial subject, consult e.g. Stuart & Ord (1987). When applied to probability densities Bayes' equation takes on the form,

$$P(\mathbf{a}|\mathcal{A},\mathcal{O}) = C_N P(\mathbf{a}|\mathcal{A}) P(\mathcal{O}|\mathbf{a}) \tag{7}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_q)$ is a set of variables,which are estimated in an observation. $\mathcal{A}$ and $\mathcal{O}$ represent the prior information and the outcome of the observation respectively. C_N is a normalization constant. Thus the first probability density is that of $\mathbf{a}$, given that $\mathcal{A}$ and $\mathcal{O}$ are true. The second, called the prior probability density or just the prior,is that of $\mathbf{a}$, if $\mathcal{A}$ alone is known to be true. The third, which should be noted in particular, is the probability of the outcome $\mathcal{O}$ if $\mathbf{a}$ is given. In this context, the latter quantity is, up to a constant, called the *likelihood function*. One should also note here, that the most probable value of $\mathbf{a}$ is that which maximizes $P(\mathbf{a}|\mathcal{A},\mathcal{O})$, while the most likely value of $\mathbf{a}$ is that which maximizes the likelihood function $P(\mathcal{O}|\mathbf{a})$. If the prior is constant, the most probable and the most likely value are identical.

3. The prior model

From our previous discussion it is clear that the prior distribution should be scalefree in intensity and also in space. Power law distributions are generally scalefree in intensity. A number of investigators e.g. Jeffreys (1939) and Stuart & Ord (1987) (p 288, 8.15) and references therein) have suggested that a power law with index minus one, i.e.

$$P(\mathbf{a}|\mathcal{A}) \propto \prod_{\nu=1}^{q} 1/a_\nu \tag{8}$$

is the preferable prior distribution in problems of the kind considered in this work. In appendix A it is shown that demanding also a spatially scalefree behavior in fact limits the power index to minus one. There is an obvious and serious problem with the above distribution in that it is not normalizable. And it is not clear how to alter it (for example by cutoffs) so it can be sensibly normalized by constraints from observed data. The problem

becomes crucial when intensities approach the noise level, as the precise form of a lower cutoff will have a major influence on averages calculated by eqn. 7. There is a second more subtle problem which has to do with continuity or spatial nearness. I.e. if two points in space are nearby it is more likely that they have similar matter densities than if they are far apart. So our prior distribution should allow a "nearness" property, which again must have no preferred scale, properties which distribution (8) lacks. It is basically a one dimensional distribution which lives in one "pixel", and generates a correlation function which singles out the pixel scale.

To construct a normalizable scalefree distribution of intensities over a limited map $\mathcal{M}$ *we suggest the following procedure:*

1. *start with a finite positive intensity* **a**
2. *use a scalefree (in both space and intensity) random process to partition* **a** *into a map of size n pixels.*
3. *make* **a** *and n sufficiently large so that the resulting map is much larger than* $\mathcal{M}$ *and the intensity over a sub-region of size* $\mathcal{M}$ *can take on any conceivable value*

By construction then, the distribution obtained over such a subregion is normalizable and "sufficiently" scalefree.

It is important to note that these properties have nothing to do with the source or observational constraints such as the pixel-size or the nature of the "messengers", in this case photons. They express the initial state of the observer: (1) he assumes that physical quantities are finite and (2) he is initially devoid of any preferred scale of any kind. He makes his statistical calculations as if the object he is observing was picked at random from an ensemble having the above properties. Any scale, average or whatever can only be imposed by the likelihood ratio which via Bayes' equation brings observational constraints into the resulting probability distribution. To obtain the above scale free distribution, consider a finite amount a_0 of a positive quantity such as energy. Distribute this energy in N maximally random "identical decay" (MRD) operations to the maximum possible n pixels along one axis. An initial pixel in which the initial energy is placed is divided into two new pixels along one dimension, into which the energy is randomly separated by a maximally random partitioning, i.e., with even probability, between zero and a_0. Note that the even distribution has maximum entropy. If this procedure is repeated N times one has $n = 2^N$ pixels with a random distribution of energies. This is a recipe for a fractal distribution, and it is not

difficult to see (Appendix B) that the single-pixel distribution of a must obey

$$P_{N+1}(a) \propto \int_a^{a_0} 1/a' P_N(a') da' \tag{9}$$

with $P_0(a) = \delta(a - a_0)$, one has

$$P_N(a) = \begin{cases} \frac{1}{a_0(N-1)!} \left(-\ln \frac{a}{a_0}\right)^{(N-1)} & 0 < a < a_0 \\ 0 & 0 > a \text{ or } a > a_0 \end{cases} \tag{10}$$

which when a is not close to the limits 0 or a_0 for large N's can be approximated to

$$P_N(a) \propto a^{-1/\ln 2} \sim a^{-1.44}. \tag{11}$$

This limiting form is interesting because it demonstrates the scale-free behavior of the MRD process but, as will be clear, it is the form (eq. (10)) with a limited N, relating nearby pixel intensities, which is useful in statistical image deconvolution. A crucial property of the MRD distribution is that in contrast to 11, the probability for the intensity to have a value between 0 and ϵ is finite as long as the pixel area is greater than zero. As the pixel area approaches zero the distribution approaches 11. An equally important property is that in contrast to thermal (exponential) distributions where the atomic scale can be obtained by studying intensity fluctuations over different scales this is not possible for the MRD distribution. The fluctuations over a given scale is by construction totally independent on which smaller scale the MRD process was halted. I.e. the "granularity problem" is solved. Properties of the MRD distribution are discussed further in Section 4.

We have thus managed to construct a normalizable distribution which lacks any prior scales. That is, the scales a_0 and n are arbitrarily large but will, as we shall see, be reduced by constraints imposed strictly by observational information. The distribution has maximum entropy on condition that it is scale-free.

To generalize to two or more dimensions one may let the partitioning alter along different axis or let it decay into several parts at each step. This would alter the relation between n and N, but would otherwise be unimportant. Thus, with N and a_0 sufficiently large, the probability distribution (10) is adopted as the prior distribution function for pixel intensities. The infinite set of maps generated by this distribution define the initial state of knowledge of the observer. The object to be observed is, of course, itself in a very precise state (cf appendix E).

In summary of this section, a unique prior distribution was obtained by the requirement that it is totally scale-free in both intensity and space. Because of its fundamental nature,

the MRD distribution, is in all likelihood used previously in Statistical Physics. In spite of occasional searches for many years however, I have not been able to find it in the literature. The much used gamma priors (cf e.g. Hsiao et al. (2002),Chang et al.(2005)) and Moussaoui et al. (2006) , allow independent mean and variance, but are neither scalefree in space nor in intensity. It does not relate in a scalefree manner 'area' integrated intensities and variances to pixel variances and intensities. It would therefore seem, it is not comparable to the prior in this work, and it cannot lead to correct 'resolved areas' and uncertainties. It is now time to specify basic details of the observations.

4. The observation model and pixel intensity probability distributions

The object observed is a piece of area which is covered by a given intensity distribution. The mapping is accomplished by means of point-wise noise-corrupted observations of this area with a telescope with a given beam pattern. Let the observation towards k positions be defined as follows

$$o_i = B_{i\nu}a_\nu + n_i \qquad i = 1, \ldots, k \tag{12}$$

where o_i is the observed intensity in position i, $B_{i\nu}$ is the value of the beam function at pixel ν when centered at position i, and a_ν the true intensity of (averaged over) pixel ν. n_i is the observational noise or uncertainty in measuring $B_{i\nu}a_\nu$. To distinguish between observed positions, which is any continuous x, y position, and the pixel positions of the estimated intensities, we denote the former by Roman and the latter by Greek indices. In line with the discussions above one should now use (10) as the prior.

The single-pixel prior distribution is

$$P(a_\nu|\mathcal{C}) = \begin{cases} \frac{1}{a_0(N-1)}\left(-\ln\frac{a_\nu}{a_0}\right)^{N-1} & a_\nu < a_0 \\ 0 & a_\nu > a_0 \\ & N = \dfrac{\ln n}{\ln 2} \end{cases} \tag{13}$$

and one has the total prior distribution

$$P(\mathbf{a}|\mathcal{C}) = \prod_{\nu=1}^{q} P(a_\nu|\mathcal{C}) \tag{14}$$

where $\mathcal{C}$ represent MRD conditions and q is the total number of pixels. *Note now that, n needs to be sufficiently large* ($n >> q$) *such that the prior distribution is scale-free in intensity and space over the map.* This means that the total intensity over the the pixels in question have totally lost any memory of a_0 and can take on any conceivable value.

The likelihood function is simply the probability distribution for observed intensities, given the beam response function, true pixel intensities, and noise. If n_i is Gaussian random noise with variance σ_i, one has that the probability distribution for observed intensities o_i are

$$P(\{o_i\}_{i\in N}|\{B_{i\nu}\},\mathbf{a},\{\sigma_i\}) \propto \exp\left(-\sum_{i=1}^{k}\frac{(o_i - B_{i\nu}a_\nu)^2}{2\sigma_i^2}\right) \tag{15}$$

According to Bayes (cf. eq. (7)), with the prior eq. (14) and the the likelihood function eq. (15) the conditional probability distributions of the pixel intensities are,

$$P(\mathbf{a}|\mathcal{C},\mathcal{O}) \propto P(\mathbf{a}|\mathcal{C})\cdot\exp\left(-\sum_{i=1}^{k}\frac{(o_i - B_{i\nu}a_\nu)^2}{2\sigma_i^2}\right) \tag{16}$$

Let the average or expected intensity in a single pixel be denoted $\alpha_\nu = \langle a_\nu\rangle$. One may formally write the expected map (the mean of the ensemble of possibly true maps).

$$\alpha_\nu = \langle a_\nu\rangle = \int_{R+} a_\nu P(\mathbf{a}|\mathcal{C},\mathcal{O})\,da \tag{17}$$
$$\nu = 1,\ldots,q$$

Since this map is a deconvolution of the observed intensities obtained by statistical means it may be called a Statistical Image Deconvolution or SID. One may also write the pixel intensity dispersions,

$$\varsigma_\nu = (\langle a_\nu^2\rangle - \langle a_\nu\rangle^2)^{1/2} \tag{18}$$

Note that both these averages are finite and well defined, so this 'inversion' is never ill-defined in the way which may occur when probabilities or likelihoods are maximized.

Equation (16) describes how probabilities of intensities integrated over small (compared to the beam-width) spatial scales are described by the prior distribution function, allowing large (scale-free) fluctuations while intensities averaged over larger spatial scales are more and more with increasing precision determined by the observed ones. As the spatial scale approaches the beam width, intensity distributions are, if the signal to noise ratio is good, almost totally determined by the likelihood function to a narrow Gaussian around the observed intensities. This fact allows a simplification of eq. (16), since observations forces the MRD process on scales larger than a particular region, approximately of beam size, to be irrelevant. I.e. one may treat the MRD process as if it started from such "transition" regions. Let C_μ be both name and area of the transition region, and $\bar{a}_\mu$ the intensity averaged over the region. Since the fluctuations of the $\bar{a}_\mu$'s are very small, one has

$$\bar{a}_\nu \approx <\bar{a}_\nu> = \bar{\alpha}_\nu \tag{19}$$

It should be emphasized, as will be seen in the following, that the above steps lead to an enormous simplification of eq. (16). The single-pixel prior distribution is simplified to the "reduced prior",

$$P(a_\nu|\mathcal{C}) = \begin{cases} \frac{1}{n\cdot\bar{\alpha}_\nu(N-1)!}\left(-\ln\frac{a_\nu}{n\cdot\bar{\alpha}_\nu}\right)^{(N-1)} & a_\nu < n\cdot\bar{\alpha}_\nu \\ 0 & a_\nu > n\cdot\bar{\alpha}_\nu \end{cases} \tag{20}$$

$$\nu = 1,\ldots,q \qquad N = \frac{\ln n}{\ln 2}.$$

The choice of n is determined by

$$n \gtrsim \frac{C_\nu}{S_\nu}; \qquad \text{all } \nu \tag{21}$$

where S_ν is the pixel area.

Remember that the constraint on C_ν is that it should be sufficiently large that the corresponding signal to noise ratio is large. A further practical requirement is that the pixel area is constant. Thus if one wants to keep C_ν constant (which is practical but not necessary) the size of C_ν should be set such that the signal to noise ratio is good over the whole source (see however the discussion in sect. 9.4). This size is usually on the order of the beam size. Note that n or N are now much smaller than before reducing eq. (16) to transition regions. While n in eq. (16) was much greater than k, the number of pixels in the map, it has now been reduced to approximately the number of pixels in the telescope beam which is certainly much smaller than k. Earlier decays started from a very large a_0 covering a large area and decayed down to the pixel-size or smaller. Now pixel intensity scales are set by observationally well determined total intensities over beam sized areas. These intensities are not affected by the prior, which by construction is fluxconserving. The intensities then decay in relatively fewer steps to pixel scales. The pixel intensity fluctuations, constrained by the prior in combination with the likelyhood function, depend on how many steps are needed to get to the pixel level. It is easy to see (by direct integration of eq. (20)) that the prior pixel intensity average and dispersions are

$$\langle a_\nu\rangle = \bar{\alpha}_\nu \tag{22}$$

and

$$\varsigma_\nu^2 = \langle\Delta a_\nu^2\rangle = \langle a_\nu^2\rangle - \langle a_\nu\rangle^2 = \{(4/3)^N - 1\}\langle a_\nu\rangle^2 \tag{23}$$

Analysis of simulated observations of models show that the results are insensitive to the exact size of the C_μ regions. If the observed intensities become smaller than the noise, the C_μ regions should be increased allowing the possibly true pixel intensities, now limited only

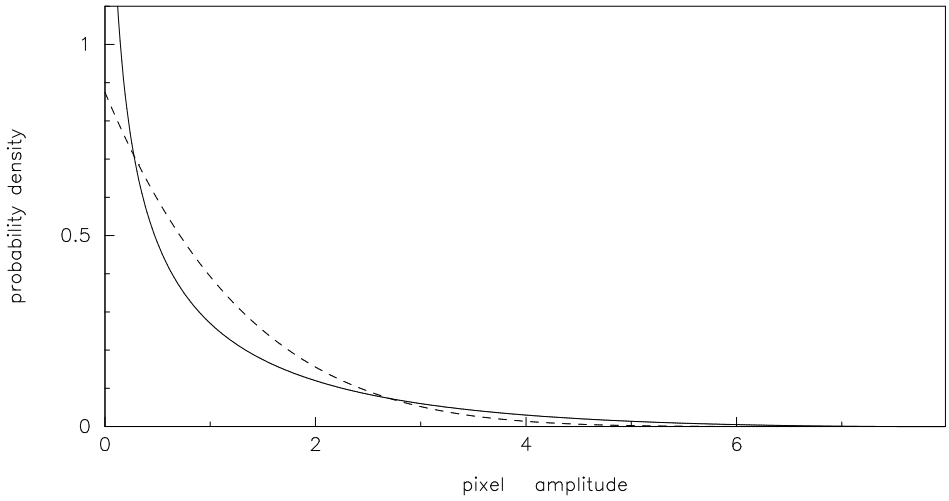

Fig. 1.— Maximum random decay (MRD, solid line)(cf. eq. (20)) and maximum entropy (exponential, dashed line) distribution of a pixel intensity which is constrained so that when added to the intensities of seven associated pixels the sum is eight or equivalently the average pixel intensity is one. Note that the ME distribution is much more condensed around its average strongly discriminating against high intensities and perhaps more importantly against intensities approaching zero. Note further that the ME distribution has preferred scales, in distance a pixel is the preferred scale, and in intensity, the preferred scale is the average intensity.

by an upper limit, to fluctuate on larger spatial scales. Increasing the size of these regions could be done but would introduce unnecessary complexity to our procedures. Neglecting the above effect may introduce small (on the order of the noise fluctuations of beam averaged observed intensities) positive regional offsets of estimated intensities (i.e. small violations of fluxconservation) as the true pixel intensities approach zero in a larger region. Since the resulting errors are very small (not directly noticeable in residuals between observed maps and beam-convolved expected maps (cf. fig 7) the effect is not further considered. See however the discussions following eq. (47) and in section 9.2.

In Fig. (1) a single pixel MRD distribution (eq. (20)) is compared with a single pixel exponential (Thermal ME) distribution. The MRD distribution is constrained such that it's sum with other MRD intensities over n=8 pixels is limited to $\sum_{C_\nu} a_\mu/n = \bar{\alpha}_\nu$, where $\bar{\alpha}_\nu$ is set to one. The exponential distribution also has the average intensity one. Note that while a local average of a subset of the MRD intensities may have a value very different from the total average, a subset of the ME intensities will have an average which is very close to the total average. This may be exemplified by the insignificant thermal fluctuations of a small but macroscopic region of a heat bath. I.e., in contrast to ME intensities, MRD intensities fluctuate on all spatial scales different from that for which they are determined.

Note further that distribution (20) in contrast to (8) produces finite and unique results when integrated to zero. This property gives well defined expectation values (cf. eq. (17)) even for small intensities while in the latter case the expectation value becomes dependant on an artificial cutoff of distribution (8).

5. The parameters of the reduced prior and calculation of the expected map

It would be nice if the integrals determining the expected map intensities ,eq. (17), could be evaluated directly in some way e.g. by a (speeded up) Monte Carlo method. Presently, a different route has been explored. In order to derive a system of equations for the expected intensities, consider the following. Since $a_\nu P(\mathbf{a}|\mathcal{C},\mathcal{O}) \to 0$, if $a_\nu \to 0$ and $a_\nu \to \infty$, one has

$$0 = \int_{R+} \frac{d}{da_\nu} \left(a_\nu P(\mathbf{a}|\mathcal{C},\mathcal{O})\right) da \quad \nu = 1,\ldots,q \tag{24}$$

, where q is the number of pixels in the map.

Inserting eq. (20) in eq. (16) and taking the derivative of latter, eq. (24) becomes,

$$0 = \int_{R_+} \left(1 - \frac{\frac{\ln n}{\ln 2} - 1}{\ln \frac{n\bar{\alpha}_\nu}{a_\nu}} - \sum_{j=1}^{k} \frac{a_\nu B_{\nu j}(B_{\mu j} a_\mu - o_j)}{\sigma_j^2}\right) P(\mathbf{a}|\mathcal{C},\mathcal{O})\, da \tag{25}$$

$$\nu = 1,\ldots,q$$

Which more compactly may be written ,

$$\langle \mathcal{G}'(a_\nu, \bar{\alpha}_\nu, n)\rangle - \quad A_{\mu\nu}\alpha_\mu + \omega_\nu = 0 \qquad \nu = 1, \ldots, q \tag{26}$$

where

$$\begin{aligned} A_{\mu\nu} &= \sum_{j=1}^{k} \frac{B_{\nu j} B_{\mu j}}{\sigma_j^2} \\ \omega_\nu &= \sum_{j=1}^{k} \frac{B_{\nu j} o_j}{\sigma_j^2} \end{aligned} \tag{27}$$

$$\mathcal{G}'(a_\nu, \bar{\alpha}_\nu, n) = \frac{1}{\alpha_\nu}\left(1 - \frac{\frac{\ln n}{\ln 2} - 1}{\ln \frac{n\bar{\alpha}_\nu}{a_\nu}}\right) - \frac{\mathcal{D}_\nu}{\alpha_\nu} \tag{28}$$

$$\mathcal{D}_\nu =< a_\nu \cdot A_{\mu\nu} a_\mu > -\alpha_\nu \cdot A_{\mu\nu}\alpha_\mu \tag{29}$$

Note that the matrix $A_{\mu\nu}$ is sparse, i.e. an element is different from zero only if the positions μ and ν are close enough for the beams to overlap. Thus if the map is much larger than the beam, the matrix will be very sparse. $\mathcal{D}_\nu$ is a dispersion term and its expectation value is normally negligible since $A_{\mu\nu}a_\mu$ is an average of a_μ over an area exceeding that of the beam, and may therefore be replaced by $A_{\mu\nu}\alpha_\mu$. The expectation value of the other part of $\mathcal{G}'$ is more tricky to evaluate. However, it turns out that if the C_μ regions are limited to about 7 to 10 pixels, the derivative of the $\mathcal{G}'$ is always a strongly negative and slowly varying function of the $\{a_\nu\}'s$ (see Appendix C). In the case that $((a_\nu - \alpha_\nu)/\alpha_\nu)^2$ is relatively small one may then with good approximation set

$$\langle \mathcal{G}'(a_\nu, \bar{\alpha}_\nu, n)\rangle = [\mathcal{G}'(a_\nu, \bar{\alpha}_\nu, n)]_{a_\nu = \alpha_\nu} \qquad n \approx 9 \pm 1 \tag{30}$$

This approximation becomes less accurate when $((a_\nu - \alpha_\nu)/\alpha_\nu)^2$ is not small. The effect of the $\mathcal{G}'$ function in eq. (26) is to drive the expected pixel intensity towards the local average $\bar{\alpha}_\nu$(see discussion below). Agreeing with intuition, it is from Fig. 16 clear that when the uncertainty becomes larger the drive towards the local average should become stronger (i.e. $\mathcal{G}(\alpha, \bar{\alpha}, n) = \int_{\alpha-\Delta\alpha}^{\alpha+\Delta\alpha} \mathcal{G}'(a, \bar{\alpha}, n)da$ is steeper than $\mathcal{G}'(a, \bar{\alpha}, n)$, the index is dropped for brevity). Thus one expects that for low signal to noise ratios the right side of eq. (31) should be multiplied by a function increasing with the inverse signal to noise ratio. The "brake function" introduced section 7 approximates this effect. Note that for normal signal to noise ratios the C_μ region has about the same size as the beam. i.e. the pixelsize should should

be about a third of the beamwidth. For extreme signal to noise ratios, the pixel-size can be smaller.

Let

$$\mathcal{G}(\alpha_\nu, \bar{\alpha}_\nu, n) = [\mathcal{G}'(a_\nu, \bar{\alpha}_\nu, n)]_{a_\nu=\alpha_\nu} = \frac{1}{\alpha_\nu}\left(1 - \frac{\frac{\ln n}{\ln 2} - 1}{\ln \frac{n\bar{\alpha}_\nu}{\alpha_\nu}}\right) \tag{31}$$

where

$$\bar{\alpha}_\nu = \sum C_{\mu,\nu}\alpha_\nu \ , \quad n = \left(\sum_\nu C_{\mu,\nu}^2\right)^{-1} \tag{32}$$

and $C_{\mu,\nu}$ is a unitary convolution over the area C_μ covering n pixels. One then has

$$\mathcal{G}(\alpha_\nu, \bar{\alpha}_\nu, n) - \quad A_{\mu\nu}\alpha_\mu + \omega_\nu = 0 \qquad \nu = 1, \ldots, q \tag{33}$$

One may now note that eq. (33) consists of a least-squares part and the $\mathcal{G}$ part. The latter part will play a role only if the least-squares part lacks sufficient signal to noise ratio to allow an exact determination of expected intensities. Considering the case of a spatially constant signal it becomes obvious that $\mathcal{G}$ must be zero for $\alpha_\nu = \bar{\alpha}_\nu$, i.e. that $\mathcal{G}$ will push α_ν towards the local average if there is no more exact information from the least squares part. Thus, in order that the approximation above does not violates fluxconservation, one has according to eq. (31)

$$\frac{\ln n}{\ln 2} - 1 = \ln n \tag{34}$$

$$n = \exp\left(\frac{\ln 2}{1 - \ln 2}\right) \approx 9.57 \tag{35}$$

a value consistent with the constraints required in eq. (31) above. It might sometimes be handy to have a little bit of freedom in the choice of n. This can be accomplished by replacing $\frac{\ln n}{\ln 2} - 1$ in eq. (33) by $\ln(n)$, requiring that $8 \lesssim n \lesssim 10$. Trials indicate that this approximation has little impact on results. Since $\mathcal{G}$ acts as guide when other information is lacking it is henceforth called the guide function.

With the system matrix

$$D_{\mu\nu} = A_{\mu\nu} - \frac{d\mathcal{G}_\mu}{d\alpha_\nu} \tag{36}$$

the iterative equations for solving eq. (33) are

$$D_{\mu\nu}(l)\Delta\alpha_\mu(l) = \mathcal{G}(\alpha_\nu(l), \bar{\alpha}_\nu(l)) + \omega_\nu - A_{\mu\nu}\alpha_\mu(l) \qquad \alpha_\nu(l+1) = \Delta\alpha_\nu(n) + \alpha_\nu(n) \tag{37}$$

where l is the order of the iteration. If, as asserted above, $\frac{d\mathcal{G}_\mu}{d\alpha_\nu}$ is zero if $\mu \neq \nu$ and strongly negative if $\mu = \nu$, then the iterative equations 37 are soluble since the diagonal elements of the system matrix are then strongly positive.

Formally the solution to eq. (37) is

$$\Delta\alpha_\mu(l) = D^{-1}_{\mu\nu}(l)\left(\mathcal{G}(\alpha_\nu(l), \bar{\alpha}_\nu(l)) + \omega_\nu - A_{\mu\nu}\alpha_\mu(l)\right) \qquad \mu = 1, \ldots, q \tag{38}$$

Note again, that for large maps, $D^{-1}_{\mu\nu}$ is the inversion of a very sparse matrix.

6. Dispersion

There are two different types of intensity dispersions which should be considered. The first one, which we have already noted in eq. (18), is the variance in possible pixel intensity obtained from its probability distribution. The second one is the variance in the expected pixel intensity due to noise fluctuations. The latter should always be smaller than the first one. Consider first the variance of the second case. From eq. (37) one has that

$$\delta\alpha_\nu D_{\mu\nu} = \delta\omega_\mu \;\text{ or }\; \delta\alpha_\nu = D^{-1}_{\mu\nu}\delta\omega_\nu \tag{39}$$

where δ refers to noise fluctuations related to observations. Squaring eq. (39) and taking the average one has

$$\langle\delta\alpha^2_\nu\rangle = D^{-1}_{\mu\nu}D^{-1}_{\gamma\nu}\langle\delta\omega_\mu\delta\omega_\gamma\rangle = D^{-1}_{\mu\nu}D^{-1}_{\gamma\nu}A_{\mu\gamma} \tag{40}$$

It is shown in Appendix D that eq. (40) can be approximated to the form

$$\sigma^2_{\alpha_\nu} = \sigma^2_n \frac{S_{pix}\ln(4)}{S_{beam}}\left(\exp\left(-\frac{S_{beam}}{S_{pix}\ln(4)}\right) + \frac{\sigma^2_n}{2\bar{\alpha}^2_\nu}\right)^{-1} \tag{41}$$

where $\sigma^2_{\alpha_\nu} = \langle\delta\alpha^2_\nu\rangle$ and S_{beam} and S_{pix} are the beam area and pixel area respectively. With a particular value of n, eg. 9, $S_{beam}/S_{pix} = 9$, eq. (41) is simplified to,

$$\sigma_{\alpha_\nu} = \frac{10}{\left(1 + \left(\frac{18\sigma_n}{\bar{\alpha}_\nu}\right)^2\right)^{1/2}}\;\sigma_n \tag{42}$$

Defining the resolved area S_{res} as that particular pixel area for which the two terms in the denominator of eq. (41) are equal one has, (see Appendix D)

$$S_{res} = S_{beam}\left(\ln(4)\ln\left(\frac{2\bar{\alpha}^2_\nu}{\sigma^2_n}\right)\right)^{-1} \tag{43}$$

Denoting the beam averaged intensity with α_{beam} and the beam area noise with σ_{beam}, it is also shown that one may write eq. (43) in the form

$$\frac{S_{res}}{S_{beam}} = \left(\ln(4) \ln \left(\frac{\alpha_{beam}^2 S_{beam}}{32 \sigma_{beam}^2 S_{res}} \right) \right)^{-1} \tag{44}$$

where an iterative procedure can be used to find S_{res}/S_{beam}. The uncertainty in the determination of the true intensity can be expressed (cf. Appendix D) as

$$\sigma_a = \begin{cases} \left[\{(4/3)^N - 1\} \alpha_{res}^2 + \sigma_{\alpha_\nu}^2 \right]^{1/2} & N > 0 \\ \sigma_{\alpha_\nu} & N < 0 \end{cases} \tag{45}$$

where

$$N = \ln(S_{res}/S_{pix}) / \ln(2) \tag{46}$$

These results agree with what has been found for deconvolved intensities of simulated observations of model sources described later in the paper.

Intensity errors may in special cases arise because the iterative procedure fails to reach the exact solution of equation eq. (33). This failure has so far appeared only in trials with extremely high signal to noise ratios in combination with extreme dynamics (cf. Fig. 6). These trials were performed to test the limits of mathematical precision, and the errors probably result from rounding-off errors in the equation-solving codes. In the code using the "peaked Mexican hat" like operator, (cf. section 7) imperfections in the "hat" may result in these type of errors.

7. Solving the iterative equations

The code which solves the iterative equations (37) has been dubbed Imagine. Note that the 'IMAGINED' map is an approximation of the SID map. The outputs (intensities and uncertainties) produced by the actual computer codes, depend on the accuracies of several steps. The first step lies in the derivation of the probability distributions, where certain approximations are done. The second step lies in the derivation of expected amplitudes where again certain approximations are introduced. The third step lies in the ability of the iterative computer code to converge towards the correct solution. As is discussed later (see section 9.1), the convergence may in difficult cases not be perfect. When deconvolved maps or spectra are convolved with the beam, they are called SYNTHESIZED. The convergence of the iterative procedure can be judged by comparing OBSERVED quantities with SYNTHESIZED. Ideally the difference should only contain random noise. The goodness of the fit of SYNTHESIZED to OBSERVED intensities can be measured in terms of the error

normalized root mean squared deviation, or equivalently the mean Chi-square value, defined below (cf. eq. (16)).

$$MChi^2 = k^{-1} \left(\sum_{i=1}^{k} \frac{(o_i - B_{i\nu} a_\nu)^2}{2\sigma_i^2} \right) \tag{47}$$

If every position is sufficiently sampled the $MChi^2$ should approach one if the fit is good. Under-sampled observations may lead to a $MChi^2$ which is less than one. To avoid 'over-fitting', a 'brake' (increasing the strength of the guide function) becomes active if the local beam averaged $MChi^2$ decreases below one. An $MChi^2$ larger than one indicates either that noise contributions are not properly estimated or that the computer code has poor convergence power.

As the number of possibly true maps may be large, a good $MChi^2$ does not guarantee that the solution is a good approximation of expected intensities. In case not otherwise is stated, deconvolved means imagined. In simulated test observations of models discussed later (cf. section 9) the deconvolved image is compared to the model. Also here the quality of the deconvolution is measured by the root mean squared deviation. In order to avoid confusion, notation "RMSD" is reserved as the goodness of fit for this latter case.

Standard methods for solving systems of equations are in most cases not well-fitted for the above types of equations, since they become extremely slow for large maps. Therefore, a special fast IMAGINE code for solving eq. (37) has been developed. The Beam-deconvolution is based on the Beam-deconvolving effect of a "peaked Mexican hat" type of convolution, fully utilizing the sparsity and convolving nature of the "system matrix" in eq. (38), approximating the effect of $D_{\mu\nu}^{-1}$. Since eq. (38) is an iterative equation, a good approximation of $D_{\mu\nu}^{-1}$ will make the iterative process reach a good approximation of the exact solution, i.e. making the right hand side of eq. (37) zero. The 'Mexican hat' function is the Laplacian of a Gaussian function. It is utilized in various problems dealing with image analysis and especially in wavelet applications (e.g., Freeman et al. (2002)). The deconvolving function used here is better described as a central one-pixel peak minus a Gaussian with a width of a few pixels. The total sum over the function must be 1. The fast code is very efficient and the time for solving eq. (37) grows only linearly with the size of the map. It may in difficult cases (large dynamics) not fully reach the correct solution,i.e. the $MChi^2$ value may not reach that obtained with the more accurate code. Normally the errors are small enough not to cause a problem however, but see the discussions in sections 9.1 and 9.4.

It is very important to have correctly estimated parameters for noise statistics in the equations. If statistics of noise, bad baselines or pointing errors are not correctly included in the calculation of σ, large errors might appear in the resulting pixel-intensities. In the presently used code, system noise is estimated from input spectra. In addition one may set

parameters to handle the influence of other types of statistical sources of error. Pointing errors can in this version of the code only be handled in the linear regime, i.e., to about 10% of the beam-width. The limit for effective resolution is set by the actual signal errors and noise fluctuations. Typical single-dish observations will usually lead to an effective resolution of one half to one third of the beam width.

It is important to note that the above approach has limited the pixel area to about one tenth of an area over which the integrated intensity is well determined. Such an area normally corresponds to about one beam-width, which if the pixel size is one third of the beam-width, consists of 9 pixels. In case a pixel size of one third of the beam-width is used, the iteration of eq. (37) works well and converges rapidly and safely in all cases for which it has been tried.

In cases for which the signal-to-noise ratio is extremely high, the area for which the signal is well determined, might correspond to one half beam-width. The pixel size could then be as small as one sixth of the beam-width.The code has also been tested for this case where it, as predicted works, well. One might wish to have a small pixel-size also for for cases with intermediate signal to noise ratios. An approach to this case has been worked out which generally seems to work well, but results are less solid and the intensities obtained, even though they solve equation eq. (33), might not always be good approximations of expected intensities. The details of this approach are outside the scope of this paper and will not be further discussed, suffice it to say that it would be nice if one could more directly evaluate the integral eq. (17). A further step is however considered where the integral eq. (17) would be evaluated around the iterated solution. More accurate true expected intensities could then possibly be obtained.

8. Calculating the expected position-velocity cube

Assuming a scale-free prior distribution of emission intensities also in velocity space one may average as well over spectral channels as spatial pixels in the construction of local averages (cf. eq. (20)). This approach provide unique equations for the cube which guarantees that averages on all spatial as well as velocity (or frequency) scales come out right and that the total flux is conserved.

Instead of directly solving a gigantic system of equations for the whole cube, the cube equations may be solved in two steps. First for an average of about 10 channels (sufficient to keep average uncertainty low), and secondly, using this first solution to construct local averages, one may solve for individual channels. The approach to the highest channel resolution

may be done in more steps, but in most cases two steps appear sufficient.

9. Testing Statistical Image Deconvoluion on simulated observations of Model distributions

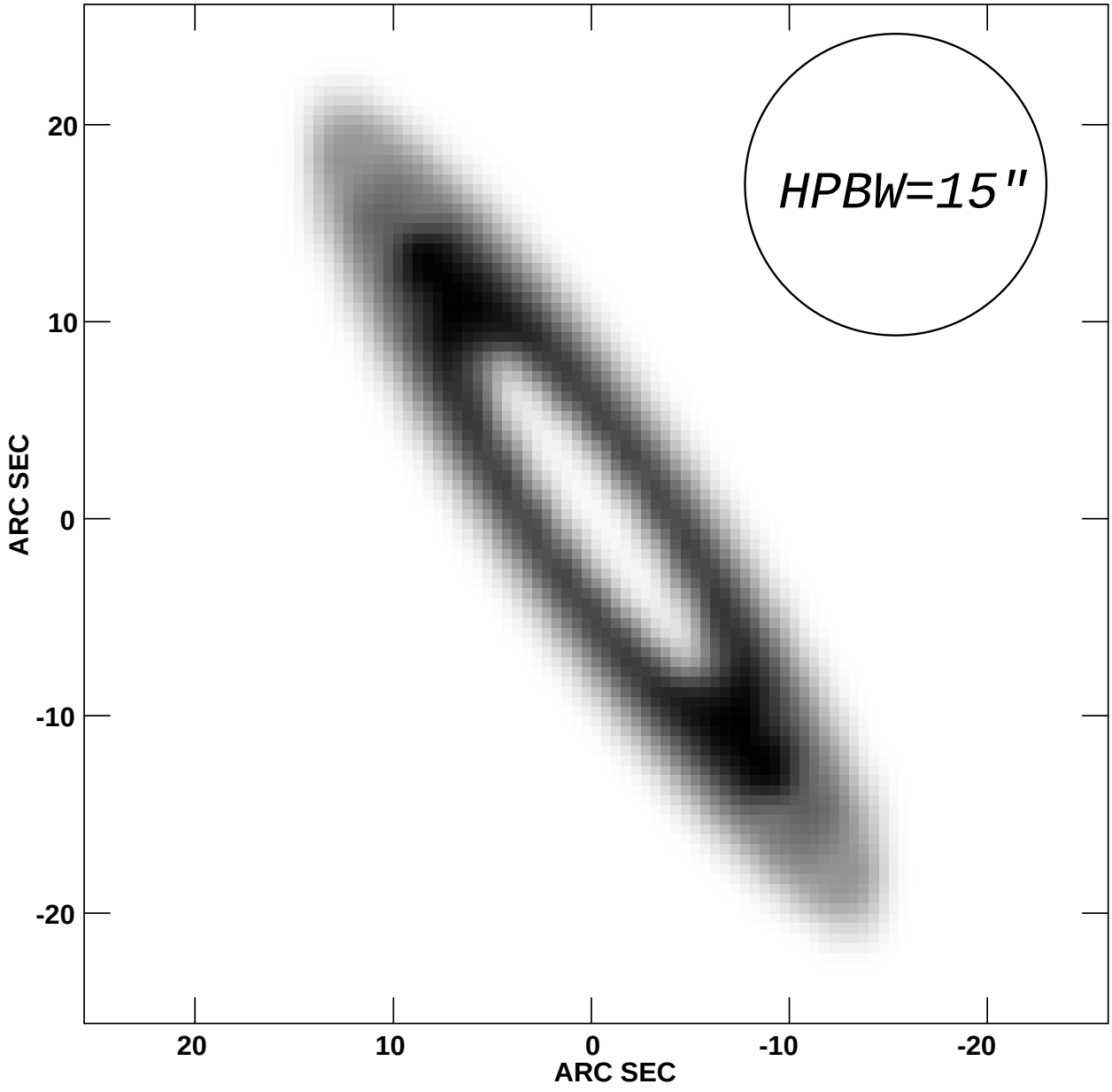

Fig. 2.— The model. The full half power width (15") of the observing beam is shown by the circle. The maximum intensity is about 850 Kkm/s.

Note first that the expected map by no means equals the true map. It is hard to know how close the "imagined" observed model maps are to expected model maps which can only be known by evaluating eq. (17). It is still useful to compare deconvolved observed model maps with the true model map. In order to judge the performance of IMAGINE or some other code, we use the following expression for the root-mean-square deviation (RMSD) of the deconvolved map to the TRUE map.

$$RMSD = \left(\frac{\sum (0.5I_i + 0.5T_i) \frac{(I_i - T_i)^2}{(I_i + T_i)^2}}{\sum (0.5I_i + 0.5T_i)} \right)^{1/2} \tag{48}$$

If the compared maps are identical, RMSD is zero, and if they are 'maximally unequal' RMSD is one. Imagine routinely calculates $Mchi^2$ and the total flux-difference. In all cases where convergence is achieved, $Mchi^2 \approx 1.0$ and the flux-difference is within noise limits.

In the simulated observations, multichannel spectra are constructed. The spectra are a function of the source structure in space and velocity, the observed position, channel noise , baseline fluctuations, random calibration errors, and pointing errors. All these conditions, plus the size of the sampling interval, have been tested separately and in combination, to study the the effect on the expected image. The studies have mostly been limited to Gaussian beams. The effects of approximating a distorted beam including side-lobes with a Gaussian beam have not been studied in detail. As suggested by some simple tests, however, the resulting relative errors in intensities are expected to be smaller than relative beam errors. A correct inclusion of an accurate beam-map is in principle trivial, but leads to very complex data handling if the effect of side-lobes or other time changing beam properties are included.

In this paper tests are limited to the analysis of simulated observations, under some different conditions, of a model consisting of a rotating disk (flat toroidal) distribution of gas similar in size and shape to the ones observed in CenA (Rydbeck et al. (1993)) and in Circinus (Curran et al. (1999; 1998)), shown in Fig. 2. The analysis was done on the whole data cube. The velocity field is in this particular case of relatively little importance to the resolution since different velocities are relatively well separated in space.

9.1. Simulated almost noise free observations of the model disk

The limit of the ability of IMAGINE to recover the original disk distribution when observational interferences are insignificant is here investigated. At some point, as noise interference approaches zero, mathematical rounding-off errors or other code imperfections will limit the resolution actually obtained. Fig. 3 shows the integrated intensities of the

model observed (almost noise free) with a 15" beam. Fig. 7 shows the observed intensities minus the model first observed with a 15" beam, second deconvolved, and third convolved back with a 15" beam. A perfect analysis should here give pure noise, but as is evident, other errors dominate the map. Fig. 4 shows the model convolved with a 5" beam, and Fig. 5 the model first observed with a 15" beam, second deconvolved, and third convolved with a 5" beam. Fig. 6 shows the difference between 4 and 5, i.e., residuals of the same origin as above but now 'deconvolved'. The residual shows an interesting regular pattern. The origin is unknown at present, but is probably due to computational rounding off errors. Applying eq. (48) to compare convolved versions of the true map with the deconvolved results it appears that the best fitting width of the convolution is about $HPBW/6$. This resolution limit is probably set by the double precision Fortran codes. These limits occur at signal-to-noise ratios which are far higher and thus resolutions which are far better than what could be reached in most actual observations. Extremely bright sources, such as 22 GHz emitting water vapor masers, could reach sufficient signal to noise ratios, however.

The above demonstrates that it can be possible to retrieve map Fourier components at frequencies higher than $1/HPBW$. This possibility is partly due to tight sampling (G) and partly due to properties inherited from the set of prior maps which only contains maps with positive intensities. Thus, the set of allowed intensities containing higher-frequency components is greatly reduced, so that their expectation values are well determined.

9.2. Simulated noisy observation of the model disk

Analysis of noisy observations of the disc is here discussed. Frequently it is preferable to convolve deconvolved data so that one has a more even effective resolution width (d_{eff}) corresponding to an area S_{eff}. The Observations are done with a spacing of 4.5". The channel width is 10km/s and the channel noise 0.1K(sigma). The quality of the observations is demonstrated by the spectral map in Fig. 10 and the quality of the deconvolved in Fig.11. These spectra are comparable to the ones obtainable from CO observations of strong CO sources like the galaxies NGC4945 or Circinus. The map of deconvolved map intensities are shown in Fig. 8. The number of integrated channels is 41 and the channel width 10 km/s, so the pixel-noise is 6.4 Kkm/s. Residuals, i.e., observed intensities minus synthesized (beam-convolved deconvolved intensities) are shown in Fig. 9. Note that the residual fluctuations are symmetric around zero demonstrating that IMAGINE approximately conserves flux. As noted in section 4 the approximations involved in reducing the prior to transition regions in combination with the positivity constraint, induces small negative offset in the expected total map residual. The total offset is then proportional to the size of the map while the

integrated noise only grows as the square root of the map-size. At some map size the total "transition region" offset will be greater than the total noise fluctuations. The total average intensity of the map in Fig. 9 is -1.0 Kkm/s, i.e. a sigma of about 3.3 since the noise sigma is about 0.3 Kkm/s. The total no of iterations for this particular map is 25.

It should be clear that residuals, although random, might well differ in fluctuating intensity on and off the source. Off the source the fluctuations should exactly correspond to the observational noise. On the source, however, the fluctuations will depend on the sampling density. This means that if one uses a sampling rate of $HPBW/2$ there is no observational redundancy and synthesised intensities will closely follow the observed ones, except in the case this is hindered by requiring negative source intensities. If one uses a dense sampling rate, in this case slightly better than a third of the $HPBW$, the redundancy will cause a more even distribution in residual fluctuations. Again using eq. (48) to find the convolution of the true map which best correspond to the deconvolved map we obtain a convolution with the width 4.5", or 30% of the $HPBW = 15$". I.e., if CO(2-1) is observed towards NGC4945 with the SEST telescope ($HPBW = 23$") one may reach an effective resolution of about 7".

It is interesting to compare these numbers with those obtained using eq. (41) and eq. (43), where the resolution in fact varies depending on emission intensity, and also to check weather in fact, the theoretical dispersion formuli eq. (41) and eq. (45) do agree with the dispersions obtained in simulated observations. To this end, a number of observations were performed, all with different noise realizations, to directly see how the noise influences the expected intensities. The results are shown as cerise spectra in Fig. 11. True model spectra are drawn in black.

Consider as an example the high intensity channels of the spectrum in pos. $(4."5, 0")$ in Fig. 11 where one finds one of the largest deviations between true and expected model spectrum. Let S_{eff} be the A ratio $S_{beam}/S_{eff} = 9$, and $\sigma_n = 0.1K$, gives (cf. eq. (41)) $\sigma_\alpha = 0.49$K. Eye inspection of the max. width of the (cerise) fluctuations give about 1.5K$\approx 3\sigma_\alpha$

The neighborhood average $\bar{\alpha}_{(4.5,0)}$ is approximately 1.0 K (cf. the same pos. in Fig. 10). From eq. (43) the resolved area (eq. (43)) is $S_{res} = S_{beam}/7.35$, or $d_{res} = 5.53$" and $\alpha_{res} \approx 3$, a value obtained by convolving α_{pix} over S_{res} . The degeneracy is $n = S_{res}/S_{eff} \approx 5.53^2/5^2 \approx 1.223$ and $N \approx \ln(1.223)/\ln(2) \approx 0.29$ so (cf. eq. (45)), $\sigma_a \approx (0.087 * 9 + 0.24)^{1/2} \approx 1$K. The offset between true and expected intensity in the spectrum considered is about 3 K, i.e. about 3 times the calculated σ_a. Note that offsets are usually not this large, i.e. expected spectra are usually a good guess of the true spectra. On the other hand possible offsets grow extremely rapidly (cf. eq. (45)) as the pixel area becomes smaller than the resolved area.

9.3. Simulated observations with pointing errors of the model disc

If pointing errors are Gaussian and small with respect to the beam-size, the effect can approximatively be described as noise fluctuations which are proportional to the slope of the observed distribution multiplied by Gaussian fluctuations. Small random pointing errors may thus be correctly included in our analysis. Simulated observations were done on the flat toroidal distribution (Fig. 2) with a spacing of 2.25". The width ($FWHM$) of the Gaussian random pointing errors were 1.5". The deconvolved map was done on a 4.5" spacing. The obtained quality is similar to the deconvolved noisy map (Fig. 5). Pointing errors could be included as separate pixel parameters which would then be solved out in an extended system of equations (cf. eq. (37)). The system of equations would become considerably more complex however. Normally one would not expect the gain in information to be worth the effort.

9.4. Comparison with other deconvolution schemes

A fair comparison of our method with others is difficult because of the vast number of methods in existence. We have nevertheless decided to compare with two different methods, of which one is a maximum entropy method (MEM, see (Olofsson et al. (1996))) and the other a Pixon method. I'm most grateful to Vincent Eke [2] who on my request kindly did the Pixon deconvolution. To compare the methods we have applied the different methods to the simulated observations of the model shown in figure (2). The observational spacing as well as the deconvolved map spacing was 4.5" while the beam was 15". Since a deconvolution results in a (corrupted) version with limited resolution of the true map, different convolutions of the true map is compared to the deconvolved result using eq. 48. Note that the "true" map of the model is constructed with very small pixel spacing. This map is first convolved with various beam-widths (HPBW) and then pixelised to 4.5". The pixelization is accomplished by averaging all emission within the pixel. For comparison, the RMSD of the unconvolved true map relative to the convolved true map is included. The result is shown in the table 1. Imagine and Pixon are very similar on the highest resolution, while MEM becomes somewhat better than Pixon on lower resolution but not as good as Imagine.

Computing times is of considerable practical importance. Comparing computing times may in this case be misleading since the codes I compare with may not have been optimized for speed. The computing speed may depend more on the algorithm used for solving the

[2]at the Physics dept. at Durham University

system of (deconvolving)equations than the the type of deconvolution. In this work a 2.8GHz computer with an internal memory of 2 G-byte is used. It takes about 0.22 sec. with the fast Imagine code (column four). We note again that the fast imagine code is linear in time consumption so that the deconvolution time is proportional to the number of pixels in the map. Consider for example the M51 CO(1-0) map (see section 10, made with the Onsala 20m telescope with a HPBW of 33". The map has 1603 input positions and 1910 output positions on a grid with a size 36×66 pixels and a spacing of 11". Deconvolving this map with the slow Imagine code using a standard algorithm for solving a system of equations takes about 10 minutes in all. It iterates about 10 times. The fast IMAGINE code iterates about 40 times and takes about one second in total, about four times the time for the model map which is about four times smaller. If the dynamics is extreme however, the fast code may take longer time, perhaps needing one 1000 iterations. In addition it may not reach the same Chi^2 as the slow code (cf. 9.4)

10. Statistical Image Deconvolution of real observed data

Maps of molecular emission from a number of galaxies have been deconvolved using the IMAGINE code. Most recently M83 (see Lundgren et al. (2004a; 2004b; 2000)), where the careful and extensive observations led to extremely high signal to noise ratios, providing resolutions on the order of 0.22 times the $HPBW = 22.5$" i.e. 5". For more details see Appendix D.

Earlier Circinus (Curran et al. (1999; 1998)) and Centaurus A (Rydbeck et al. (1993)) were analyzed. Analysis of the galaxy M51 has been presented at several meetings e.g. (Rydbeck et. al (2003)), and a detailed paper should be ready for publication soon.

Recently sub-millimeter CO, ^{13}CO and H_2O emission across the Orion-KL molecular cloud core, mapped by the Odin satellite, has been analyzed. The beam-width is about 120" in all three cases and the imagine pixel-size 40". The basic map grid spacing was 60" but frequent changes in the map center created as many inter-grid positions as on grid positions.

The observed Orion center CO$(5-4)$ spectrum is shown in Fig. 12. The deconvolved center spectrum from the same mapping observations is shown in Fig. 13. The difference is dramatic. The narrow spectrum emitted from a large region is almost gone while a new medium broad component ($\sim$ 30km/s), known as the low velocity outflow is present together with a magnified high velocity outflow (broad) component, both originating from a small spatial region. The presence of the medium broad component could not have been guessed by eye inspection of original observed spectra. The observed spectral map of ^{13}CO is shown

in Fig. 14 and the deconvolved in Fig. 15. Note the pronounced ^{13}CO low velocity outflow (the broadest) component ($\sim$ 30 km/s) in the convolved center spectrum, also shown in Fig. 13 for comparison withe same CO$(5-4)$ component. The deconvolved Odin results also are briefly discussed in a recent Odin progress report by Hjalmarson et al. (2005), where the low velocity outflow also is seen in the deconvolved H_2O spectra.

There is a striking similarity between the deconvolved map and the CN($N = 1 \rightarrow 0, 2 \rightarrow 1$ and $3 \rightarrow 2$) map in Rodriguez-Franco et al., (2001; 1998)). Compare eg with fig. 1 in the latter reference. This similarity may indeed be expected since the ^{13}CO(5-4) emission should be a good measure of the H^2 column density of the interface layer between the Orion KL molecular cloud and the M42 HII region.

There are hints of some extra noise or error contributions to the ^{12}CO data. These are certainly not of the magnitude that could have created the medium broad component, but calculation of uncertainties becomes difficult. The ^{13}CO data appears very reliable. With a signal(observed) and noise temperature of $\sim$ 15K and $\sim$ 0.4K respectively the resolved area is better than a pixel-area (cf. Appendix D). The uncertainty in the determination of deconvolved temperatures are then given by eq. (42), so $\sigma_{a_\nu} \sim 4K$.

11. Conclusions

An new approach to the problem of deconvolution and noise reduction of astronomical data observed with a single dish radio telescope has been developed. The approach is statistical in nature and based on the assumption that the analysis should be maximally objective. Specifically this means that the prior should be totally scale free i.e. it should have no scale dependence either in space or intensity. Another way to express the scale independence is to say that the deconvolution procedure must not exclude or suppress any possible map of intensities prior to observations.

With the existence of this Prior, observational results will then via Bayes' equation define an ensemble of possibly true source maps(or cubes), i.e. maps which when observed may produce the observed result. This ensemble define an expected or mean possibly true (i.e. deconvolved) intensity map and uncertainties around this expected map. More specifically:

1. Based on fundamental statistical concepts a new iterative deconvolution equation has been derived.

2. Equations expressing uncertainties and variances of the map intensities have been derived.

3. A very fast computer code solving the iterative equations and calculating uncertainties has been developed.

4. The performance of the code has been compared with two other codes (MEM and Pixon) and it was found to be better. Our code is very fast especially for large maps since the computing time is linearly proportional to the map size.

5. The code has been applied to real data demonstrating that new and essential information can be brought to light.

I am particularly grateful to Åke Hjalmarson for a long time support and interest in this work. In addition I thank him, Hans Olofsson and Per Friberg for carefully reading the manuscript and suggesting valuable improvements to the text. I also thank Per Bergman for suggesting a solution to the problem posed in Appendix C

A. The power index of a scale-free distribution

Consider a set of maps where intensities a_i of pixel (i) have a power law distribution. Require that the distribution must be free of any spatial scale dependence and that it is spatially uncorrelated. This means that the intensity distribution resulting from the sum of the intensities from two pixels must have the same form as that over one pixel. Let the distribution in pixel (i) be

$$P(a_i|\mathcal{A}) \propto a_i^{-\gamma} \qquad 0 < a < \infty \tag{A1}$$

where $\mathcal{A}$ represent the applied conditions and γ is the power index.

In other words, given the two-pixel distribution of uncorrelated intensities a_1 and a_2, one has

$$P(a_1, a_2|\mathcal{A}) \propto a_1^{-\gamma} a_2^{-\gamma} \tag{A2}$$

For $\bar{a} = k(a_1 + a_2)$, where k is an arbitrary constant, the following must hold

$$P(\bar{a}|\mathcal{A}) \propto \bar{a}^{-\gamma} \tag{A3}$$

As is shown below this condition limits γ to minus one.

Let the distribution of x_1 and x_2 be given by

$$p(x_1, x_2)dx_1dx_2 = x_1^{\gamma} x_2^{\gamma} dx_1 dx_2 = (x_1 x_2)^{\gamma} dx_1 dx_2 \qquad x_1 \geq 0, x_2 \geq 0 \tag{A4}$$

Make the coordinate transformation (a rotation by 45°)

$$\begin{aligned} y_1 &= \frac{1}{\sqrt{2}}(x_1 + x_2) \\ y_2 &= \frac{1}{\sqrt{2}}(x_1 - x_2) \end{aligned} \tag{A5}$$

Equation A5 then transforms as

$$p(y_1, y_2)dy_1dy_2 = \left(\frac{1}{2}(y_1^2 - y_2^2)\right)^{\gamma} dy_1 dy_2 \tag{A6}$$

Now substitute $\frac{y_2}{y_1}$ with z. One then has

$$p(y_1, z)dy_1dz = 2^{-\gamma} y_1^{2\gamma+1}(1 - z^2)^{\gamma} dy_1 dz \tag{A7}$$

z can now be integrated away ($-1 \leq z \leq 1$) and the y_1 distribution remains. With the condition that the power distribution for y_1 has the same power index γ as the distribution for x_1

$$y_1^{2\gamma+1} = y_1^{\gamma} \tag{A8}$$

follows, and therefore $\gamma = -1$. Note that the above is valid only if the intensities are linear quantities or equivalently that they have a constant density of states.The relation would for example not hold if a_i's where substituted by $\sqrt{a_i}'s$. Thus assuming no pixel to pixel correlation, the most impartial prior distribution for positive (linear) pixel intensities $\mathbf{a} = (a_1, \ldots, a_n)$ is

$$P(\mathbf{a}|\mathcal{A}) \propto \prod_{i=1}^{n} 1/a_i \tag{A9}$$

See also Stuart & Ord (1987) (p 288, 8.15), Jeffreys (1939), and Gelman et al. (2003).

B. The MRD distribution

Assume that the positive variable a_N has the distribution $p_N(a_N)$, and that $p_N(a_N) = 0$ for $a_N > a_0$. Let a_N be partitioned into two parts $a_{1,N+1}$ and $a_{2,N+1}$ in a 'maximum random decay' i.e. in such a way that in this decay, every value of $a_{1,N+1}$ in its range $0 \leq a_{1,N+1} \leq a_N$ is equally likely (Every value of $a_{2,N+1}$ is also equally likely in the same range). Since the range is $0 - a_N$, the decay probability distribution of $a_{1,N+1}$ is

$$P_D(a_{1,N+1})da_{1,N+1} = \frac{da_{1,N+1}}{a_N} \tag{B1}$$

Multiplying with the distribution for a_N one has,

$$P_{N+1}(a_{N+1}, a_N)da_{1,N+1}da_N = \frac{da_{1,N+1}}{a_N} p_N(a_N)da_N \tag{B2}$$

Integrating a_N from $a_{1,N+1}$ to a_0 and simplifying the notation by letting $a_{N+1} = a_{1,N+1}$ one has finally

$$P_{N+1}(a_{N+1}) = \int_{a_{N+1}}^{a_0} \frac{p_N(a_N)}{a_N} da_N \tag{B3}$$

C. the guide function

The guide function eq. (31) is

$$\mathcal{G}(\alpha_\nu, \bar{\alpha}_\nu, n) = \frac{1}{\alpha_\nu}\left(1 - \frac{\frac{\ln n}{\ln 2} - 1}{\ln \frac{n \cdot \bar{\alpha}_\nu}{\alpha_\nu}}\right) \tag{C1}$$

The derivative $\frac{d}{d\alpha_\nu}\mathcal{G}$ will have a maximum for a particular $\alpha_\nu = \alpha_{\nu,0}$. We want to find the n that makes this maximum as small as possible. To this end let

$$x = \frac{\alpha_\nu}{n \cdot \bar{\alpha}_\nu} \quad \text{, and}$$
$$R = \left(\frac{\ln(n)}{\ln(2)} - 1\right) \tag{C2}$$

Now eq. (C1) assumes the form

$$\mathcal{G}(\alpha_\nu, \bar{\alpha}_\nu, C_{\nu\nu}) = (n\bar{\alpha}_\nu)^{-1} \cdot \frac{1}{x}\left(1 + \frac{R}{\ln(x)}\right) \tag{C3}$$

Consider

$$f(x) = \quad \tfrac{1}{x}\left(1 + \tfrac{R}{\ln(x)}\right) \tag{C4}$$
$$f'(x) = \quad -\tfrac{1}{x^2} - \tfrac{R}{(x^2\ln(x))} - \tfrac{R}{(x\ln(x))^2} \tag{C5}$$
$$f''(x) = \quad \tfrac{2}{x^3} + R\left\{\tfrac{2(1+\ln(x))}{(x\ln(x))^3} + \tfrac{2\ln(x)+1}{x(x\ln(x))^2}\right\} \tag{C6}$$

Now $0 < x < 1$, and thus $\ln(x) \neq 0$. The slope has a maximum if $f''(x) = 0$ i.e. if,

$$2 + 3u + 2u^2 + 2u^3/R = 0 \tag{C7}$$

where $u = ln(x)$ This equation has at least one real solution,

$$u_0 = g(R) \quad or \tag{C8}$$
$$x_0 = e^{g(R)} \tag{C9}$$

Insert this solution in eq. (C5) which then takes the form,

$$h(R) = f'(x_0) = -x_0^{-2}\left(1 + \frac{R}{\ln(x_0)} + \frac{R}{\ln^2(x_0)}\right) = \tag{C10}$$
$$-e^{2g}(1 + R/g + R^2/g^2)$$

To find the minimum with respect to R, take the derivative with respect to R

$$\begin{aligned} h'(R) \quad &= 2e^{-2g}g'(1 + R/g + R^2/g^2) - \\ &\quad e^{-2g}(1/g - Rg'/g^2 + 1/g^2 - 2Rg'/g^3) = \\ e^{-2g}(\quad &1/g + 1/g^2 - g'(2 + 2R/g + 3Rg^2 + 2R/g^3)) \end{aligned} \tag{C11}$$

or

$$-e^g h'(R) = g^2 + g - 2g'g^3 - Rg'(2g^2 + 3g + 2) = \tag{C12}$$

$$g^2 + g \tag{C13}$$

since according to eq. (C7), $2g^2 + 3g + 2 = -2/Rg^3$. Thus, if h(R) has a minimum then $g(1+g) = 0$. Since $g \neq 0$, one has that $g = -1$. From eq. (C5), one has $2 - 3 + 2 - 2/R = 0$ or $R = 2$. From eq. (C7), then

$$n = 8 \tag{C14}$$

That is, $C_{\mu\nu}$ should optimally be a convolution over 8 pixels corresponding to a convolution diameter of

$$d_C = 3.19 \text{ pixels.} \tag{C15}$$

We have in Fig. 16 plotted the guide function as a function of α_ν for different widths ω_c of the neighborhood C_μ . It is easy to see from C5 that as long as $R < 4$, $\frac{d}{d\alpha_\nu}\mathcal{G}$ is negative. I.e. the system of equations should be stable as long as R is closer to 2 than to 4 or equivalently n closer to 8 than to 36. That is $n \approx 10$ should work well. Trials have indicated however that one may soon run into problems if n is made much bigger than this. For the slope $\frac{d\mathcal{G}_\mu}{d\alpha_\nu}$ at $\alpha_\nu = \bar{\alpha}_\nu$ and $n = 8$, one has

$$\left.\frac{d\mathcal{G}_{\mu\nu}}{d\alpha_\nu}\right|_{\alpha_\nu=\bar{\alpha}_\nu} = \begin{cases} \lesssim -0.5/\bar{\alpha}_\nu^2 & \mu = \nu \\ = 0 & \mu \neq \nu \end{cases} \tag{C16}$$

in terms of the Kronecker delta,

$$\left.\frac{d\mathcal{G}_{\mu\nu}}{d\alpha_\nu}\right|_{\alpha_\nu} \lesssim -\frac{\delta_{\mu\nu}}{2\alpha_\nu^2} \tag{C17}$$

note that in terms of continuous coordinates $\delta_{\mu\nu}$ becomes a 'top hat' function with the area S_{pix} and the height $1/S_{pix}$ so that it integrates to 1 over the pixel surface, i.e.

$\delta(r, r_{pix})/S_{pix} = 1$ if $r < r_{pix}$ and zero elsewhere.

D. dispersion

Consider a map intensity α_ν centered on the ν'th pixel extending out to a radius r_ν, the rest having zero intensity. Taking the Fourier transform of eq. (39) and noting that $\Delta\tilde{\omega}(w) = \Delta\tilde{o}(w)\tilde{B}(w)/\tilde{\sigma}(w)^2$ one has

$$\Delta\tilde{a}(w)_\nu \left(\tilde{A}(w) + \frac{d\tilde{\mathcal{G}}_{\mu\nu}}{d\alpha_\nu}(w) \right) = \Delta\tilde{o}(w)\tilde{B}(w)/\tilde{\sigma}(w)^2 \tag{D1}$$

where $\Delta\tilde{o}(w)$, $\tilde{B}(w)$ and $\tilde{\sigma}(w)^2$ are the (spatial) fourier transform the of the observed map of noise fluctuations, beam and noise power respectively. Note further that $\Delta o_i = n_i$ of eq. (12) so that $< \Delta o_i^2 >= \sigma_i^2$. Here $'\tilde{}\ '$ denotes the Fourier transform. To keep expressions compact we let w represent the vector (w_x, w_y) where x and y are the orthogonal spatial coordinates of the map. Using eq. (C16) this equation may be approximated to

$$\Delta\tilde{a}(w)_\nu \approx \Delta\tilde{o}(w)\tilde{B}(w)/\tilde{\sigma}(w)^2 \left(\tilde{A}(w) + \frac{\tilde{\delta}(w, w_r)}{2S_{pix}\bar{\alpha}_\nu^2}\right)^{-1} \tag{D2}$$

where $\tilde{\delta}(w, w_r)$ integrates to unity over $\tilde{S}_{pix}(w) = S_{pix}^{-1}$, S_{pix} being the pixel area. I.e. $\tilde{\delta}(w) = \tilde{S}_{pix}^{-1}(w)$ if $|w| \leq w_r = \sqrt{\tilde{S}_{pix}/\pi}$ and zero if $w \geq w_r$. Note that in effect all frequency components with a frequency $w \geq w_r$ are zero since they describe structures which are smaller than the pixel scale. Thus w_r is the upper limit of fourier integrals. Square eq. (D2) and take expectation value to obtain,

$$< \Delta\tilde{a}(w)_\nu^2 > \approx \tilde{A}(w) \left(\tilde{A}(w) + 0.5s(w_r, w)/\bar{\alpha}_\nu^2\right)^{-2} \tag{D3}$$

where $s(w_r, w)$ is the step function, and $\Delta\tilde{a}(w)$ are fluctuations in the expected (deconvolved) intensities due to noise fluctuations in the observed intensities. To further simplify, assume that the noise (noise fluctuations in the observed intensities) is constant with position. Then $\tilde{\sigma}(w)^2 = \sigma_n^2/\pi w_r^2$, where σ_n^2 is the total observed noise power. We may then write $\tilde{A}(w) = \sigma_n^{-2}\exp(-w^2\sigma_B^2)$, where $\sigma_B = (8ln(2))^{-1/2}HPBW$. Integrating eq. (D3) to the pixel resolution frequency to get the noise total power of the deconvolved pixel intensity a_ν we get

$$\sigma_{\alpha_\nu}^2 = \frac{\sigma_n^2}{w_r^2}\int_0^{w_r} \frac{\exp(-w^2\sigma_B^2)\ w\ dw}{\left(\exp(-w^2\sigma_B^2) + 0.5\sigma_n^2/\bar{\alpha}_\nu^2\right)^2} = \frac{\sigma_n^2}{\sigma_B^2 w_r^2}\left[\left(\exp(-w_r^2\sigma_B^2) + \frac{\sigma_n^2}{2\bar{\alpha}_\nu^2}\right)^{-1} - \left(1 + \frac{\sigma_n^2}{2\bar{\alpha}_\nu^2}\right)^{-1}\right] \tag{D4}$$

Where $\sigma_{\alpha_\nu,n}^2 =< \Delta\tilde{a}(w)_\nu{}^2 >$. The last term is only important if $\bar{\alpha} \lesssim \sigma_n$ and may in most cases be ignored.

Dropping the last term and writing eq. (D4) in terms of the beam and pixel areas S_{beam} and S_{pix} respectively one has,

$$\sigma_{\alpha_\nu}^2 = \sigma_n^2 \frac{S_{pix}\ln(4)}{S_{beam}}\left(\exp\left(-\frac{S_{beam}}{S_{pix}\ln(4)}\right) + \frac{\sigma_n^2}{2\bar{\alpha}_\nu^2}\right)^{-1} \tag{D5}$$

Note that if the exponential term in eq. (D5) is dominant then the source is resolved to a pixel. If the second term is dominant then the diagonal terms in the inversion matrix are

increased, which means that the inversion is 'smeared' i.e. full resolution to a pixel is not obtained. One may thus write the resolved area

$$S_{res} = S_{beam}\left(\ln(4)\ln\left(\frac{2\bar{\alpha}_\nu^2}{\sigma_n^2}\right)\right)^{-1} \tag{D6}$$

The resolved area S_{res} is here defined as the area for which uncertainties due to observational noise or other error sources are overcome by uncertainties due to the intensity degeneracy in areas smaller than the resolved area. The origin of the degeneracy being that intensity fluctuation within S_{res} cannot be detected as long as α_{res}, the mean intensity over S_{res}, is constant.

Equation eq. (D6) can be put in a more user-friendly form. Position $\{\mu\}$ is not essential in this discussion and the reference to it is therefore on occasion dropped. It was found (cf. eq. (C14)) that the optimal size of the area C_μ is 8 pixels. Choosing the pixel area S_{pix} to equal the resolution area S_{res} one has that $C_\mu = 8S_{res|\mu}$. For observational noise fluctuations one has

$$\frac{\sigma_{res}^2}{\sigma_{beam}^2} = \frac{S_{beam}}{S_{res}} \tag{D7}$$

Where $\sigma_{beam}^2 = (\sum_{S_{beam}} \sigma_{pix}^{-2})^{-1}$ and σ_{pix} is the pixel observational noise. Assume now that the source is contained in the resolution area. Then $\alpha_{res|\mu} = 8\bar{\alpha}_\mu$. Remember that $\bar{\alpha}_\mu$ is the intensity averaged over the area C_μ. Since $\bar{\alpha} = \alpha_{beam}\frac{S_{beam}}{C}$ one has

$$\bar{\alpha} = \alpha_{beam}\frac{S_{beam}}{8S_{res}} \tag{D8}$$

Equation D6 can now be rewritten

$$\frac{S_{res}}{S_{beam}} = \left(\ln(4)\ln\left(\frac{\alpha_{beam}^2 S_{beam}}{32\sigma_{beam}^2 S_{res}}\right)\right)^{-1} \tag{D9}$$

The signal to noise ratio of the resolved area may as well be calculated. Remembering that for the resolved area

$$\exp\left(-\frac{S_{beam}}{S_{pix}\ln(4)}\right) = \frac{\sigma_n^2}{2\bar{\alpha}_\nu^2} \tag{D10}$$

one has from eq. (D5) that

$$\sigma_{\alpha_{res}}^2 = \sigma_{res}^2\frac{S_{res}}{S_{beam}}\left(\frac{\sigma_{res}}{\bar{\alpha}}\right)^{-2} = \bar{\alpha}^2\frac{S_{res}\ln(4)}{S_{beam}} \tag{D11}$$

Inserting $\bar{\alpha} = \alpha_{res}/8$ one has,

$$\frac{\alpha_{res}}{\sigma_{\alpha_{res}}} = 6.8\sqrt{\frac{S_{beam}}{S_{res}}} \tag{D12}$$

The resolved area over the beam area has been tabulated (2) as a function of observed signal to noise ratio. It is apparent that, for typical signal to noise ratios, the resolved area is about one tenth of the beam area or about a third of the beam-width in terms of length scale. It is of course important that the sampling interval is less than the resolution diameter (cf. appendix G).

For comparison, the CO(2-1) data for M83 has an intensity of about .6K towards the center of the map, and a channel noise sigma of about 3 mK and a channel width of .52 km/s. The mapping was done with a spacing of 7" which is about a third of the beam width at this frequency. Thus the σ_{beam} is about 1 mK, so the signal to noise ratio is about 600. Over eg four channels it is 1200. According to the table, the resolution in terms of one dim. width is .22 times the $HPBW$ which is 22.5", i.e. 5" This data is probably the best that has ever been analyzed with Imagine. It is striking that the signal to noise ratio actually increases as the resolution is increased. This happens because even if the noise fluctuations increases as the resolution is improved, the intensity increases faster, since it was assumed that the source of emission is within the resolved area.

Expressions for the dispersion in the determination of true pixel intensities is a bit more complex than that of expected intensities. The number of pixels in the resolved area S_{res} is $n = S_{res}/S_{pix}$. For areas smaller than the S_{res} area the intensity statistics is of MRD type. One may then use eq. (23) to find the variance of pixel intensities. Thus,

$$\sigma_a^2 = \{(4/3)^N - 1\}\alpha_{res}^2 + \sigma_{\alpha_\nu^2} \tag{D13}$$

where

$$N = \ln(S_{res}/S_{pix})/\ln(2) = \ln\left(\frac{S_{beam}}{S_{pix}}\ln(4)\ln\left(2\bar{\alpha}_\nu^2/\sigma_n^2\right)\right)/\ln(2) \tag{D14}$$

Note that the MRD part of this uncertainty becomes zero for pixel-sizes greater than S_{res}. Since there is no reason to expect otherwise it was assumed, excluding a deeper discussion on the subject, that eq. (D14) is valid also for non-integer n and N.

With $\bar{\alpha} \approx 1$K which it is in the central parts of the disc model and the noise 0.1 K, $S_{res} = S_{beam}/7.35$ and the number of pixels in the resolved area (degeneracy), n=1.23. From eq. (D13), $\sigma_a = 0.61\alpha_{res}$. To get α_{res} the cube is convolved to the resolution area.In e.g. position (0,4.5) (in map 11) $\alpha_{res} \approx 3.5$ K. Thus $\sigma_a \approx 2.1$ K. As in this position the deviation of true pixel intensities compared to expected in all positions are within bounds estimated by eq. (D13) and D14. Although calculations are somewhat involved a computer code can trivially produce a map of uncertainties for the determination of true pixel intensities.

Conv. HPBW	RMSD MEM	RMSD PIXON	RMSD IMG	RMSD IMG_{cube}	RMSD TRUE
6.00	0.225	0.274	0.165	0.196	0.385
5.25	0.203	0.257	0.143	0.165	0.354
4.5	0.190	0.247	0.133	0.137	0.315
3.75	0.193	0.244	0.146	0.128	0.274
3.0	0.219	0.251	0.180	0.148	0.213
2.25	0.284	0.273	0.241	0.209	0.114
1.50	0.346	0.297	0.291	0.262	0.004

Table 1: Deconcolved intensities of simulated observations of model intensities, compared to different convolutions of true model intensities. Column one displayes convolving beam-widths, and the following three columns the RMSD's of MEM, PIXON and IMAGINE(IMG) applied to observed integrated intensities. The fourth column is the result of the standard imagine code. The fifth column results from the standard imagine code applied to the whole observed data cube. The "fast" imagine code was used at all times. The sixth column results from the RMSD's of the true map compared to different convolutions of itself.

$\alpha_{beam}/\sigma_{beam}$	S_{res}/S_{beam}	C/S_{beam}	$\frac{\alpha_{res}}{\sigma_{\alpha_{res}}}$
10	0.32	2.56	12
30	0.136	1.09	18.4
90	0.091	0.73	22.5
180	0.076	0.61	24.7
1200	0.052	0.42	29.8

Table 2:

E. probabilities

The definition of probabilities is a subject that has been hotly debated for many years. One side in this discussion, "the frequentists", holds that a proper definition requires that probabilities are based on the frequency of occurrence, while the other side holds that probabilities may be used as a measure of knowledge or information. While everybody agrees that Bayes' equation is correct, the frequentists are vehemently opposed to the Bayes' postulate which states that in the absence of any information to the contrary, every possibility or state should be considered equally likely. Bayes' postulate is similar to the "maximum entropy principle" in physics, which asserts that in thermal equilibrium a physical system is equally likely to be found in any of its accessible states. The difference is that one has well understood reasons for assuming that certain physical systems are in equilibrium. Maximizing the entropy may in such situations be an elegant way to derive probability distributions. For "non trivial systems", unlike trivial systems like heat baths in equilibrium, notions like "every accessible state" or Jaynes' (1957) "most non committal" and "maximum entropy" become very complex issues. At least in the type of inversion problem treated in this paper it is wise to avoid the term "maximum entropy", since depending on the implied constraints, it can mean almost anything (see appendix F).

Even if one accepts prior probabilities and Bayes' postulate, it has been considered very difficult to translate a sometimes vague prior knowledge into probabilities. However, in a simple well-defined situation as below it is easy to apply Bayes' equation.

Assume that an intensity a is picked at random in the interval 0 to 1 (i.e., Bayes' postulate is implicit), and then measured. The prior probability density of a is then

$$P(a|\mathcal{A}) = \begin{cases} 1.0 & 0 < a < 1 \\ 0.0 & \text{otherwise} \end{cases} \tag{E1}$$

Let the result of the measurement be described by

$$o = a \pm \sigma \tag{E2}$$

where o is the measured value, and σ the uncertainty in the measurement. If the uncertainty is due to Gaussian random noise with the variance σ, then the probability density of the outcome o given a is

$$P(o|a,\sigma) \propto \exp(-\frac{(a-o)^2}{2\sigma^2}) \tag{E3}$$

Since the crucial prior distribution is given, one has according to Bayes' equation,

$$P(a|\mathcal{A},o,\sigma) = \begin{cases} \propto \exp(-\frac{(a-o)^2}{2\sigma^2}) & 0 < a < 1 \\ 0 & \text{otherwise} \end{cases} \tag{E4}$$

From this distribution one may form the *expectation value* of a,

$$\langle a \rangle = \int_0^1 a\, P(a|\mathcal{A}, o, \sigma)\, da \tag{E5}$$

Note that if a new measurement is done, the probability distribution given by eq. (E4) serves as a the new prior distribution. In this way the probability distribution can be continually upgraded by new measurements.

Real situations rarely provide us with such clear cut prior probabilities however. Part of the difficulty is to understand the relation between frequency of ocurrance, information and conditional probability. The simple example below suggests that there is a well defined link between frequency of ocurrance and the subjective state of information of an observer.

Consider a "true" random machine that produces random numbers say from 1 to 6 like a die. If the die is honest, $P_i = 1/6,\ i = 1, \ldots, 6$. This probability is based on frequencies of occurrence of the numbers 1 to 6. Now consider a die which is cast but hidden under a cup. There are now no random machines or frequencies involved anymore, only the hidden die. This die is in a *precise state*, but the observer does not know which. The observer may however express his level of information (or confusion) about the state of the die in terms of the P_i given above. Note that, although associated, *there is an important difference in the meaning of these probabilities. The first is describing the indefinite state of a system and the second the indefinite state of the observer.* One may then conclude that to any set of acceptable prior "informative" probabilities there is a corresponding ensemble of "virtual" events whose frequencies correspond to these probabilities. Moreover there may be different observers. One on each side of the table with the die hidden by the upside down cup. Assume this cup has a small hole on one side, unnoticed by the observer on the other side of the cup. The observer on the side of the hole however gets a glimpse of one face of the die which happens e.g. to have one dot on it. The state of the two observers must now be described by two different sets of probabilities or different ensembles of possible outcomes. The state of one observer is described by $P_i = 1/6\ i = 1, \ldots, 6$ and that of the other by $P_1 = 0, P_6 = 0$ and $P_i = 1/4\ i = 2, \ldots, 5$, in spite of the fact that they are observing the same die. One might think of these ensembles of possible outcomes as models of the different observers state of mind.

The above discussion suggests that an analysis of an observation is defined by the observers prior state of information and the observational results. Moreover the observers state is not unique, different observers may have a different set of prior probabilities and different observational information.

F. on maximum entropy

The quantity which is needed for a correct analysis of an observed map are the probabilities $P(a_i)$ of possibly true pixel intensities a_i in pixel (i) as constrained by observations. Such a distribution would provide both an average intensity and a mean squared deviation from that intensity, i.e. an uncertainty.

It has been suggested (cf e.g. Bryan and Skilling (1984)) that one may define an entropy in terms of map-pixel intensities I_j so that

$$S = -\sum_{j=1}^{M} I_j \ln I_j \tag{F1}$$

Sometimes one has proposed

$$S = -\sum_{j=1}^{M} \ln I_j \tag{F2}$$

This corresponds to distributing a quantum of intensity to a pixel (i) according to the probability P_i. After many such distributions $I_i \approx n \times P_i$. It should be obvious that this distribution is not the one we search. Meaningful uncertainties cannot be defined. The probabilities concern a situation which is not at hand.

Consider now the "heat bath" case where the intensities in a given pixel $[i]$ have an exponential distribution

$$P_i(a_n) = (1 - \exp[-\epsilon/C]) \exp[-a_n/C] \tag{F3}$$

of quantized intensities $a_n = n \times \epsilon$. Such a distribution would follow from MEM (Maximum Entropy Method) if the ensemble intensities of a given pixel are constrained to an average $< a_i >$. The entropy of this distribution is

$$S = -\frac{\langle a \rangle}{\epsilon} \ln \frac{\langle a \rangle}{\epsilon} + (1 + \frac{\langle a \rangle}{\epsilon}) \ln(1 + \frac{\langle a \rangle}{\epsilon})$$

In the limit $\frac{\langle a \rangle}{\epsilon} \to \infty$ one has that

$$S \to \ln \frac{\langle a \rangle}{\epsilon} \tag{F4}$$

and in the limit $\frac{\langle a \rangle}{\epsilon} \to 0$, one has that

$$S \to -\frac{\langle a \rangle}{\epsilon} \ln \frac{\langle a \rangle}{\epsilon} \tag{F5}$$

Here the forms F1 and F2 of intensity related entropy are obtained. That the last form (eq. (F5))is identical to (eq. (F1)) is expected since in effect the (eq. (F5)) $\frac{\langle a \rangle}{\epsilon} \rightarrow 0$ constraint is identical to the constraint leading to (eq. (F1)).

It is clear then that there are "hidden" thermal type constraints in the entropy forms F1 and F2. Forms of entropy valid only if intensities are quantized and if their distribution is assumed to be exponential (thermal), both of which are highly artificial constraints. Also the size of the pixel is an artificial constraint. I.e. the distribution of area averaged intensities in a given area depends on the pixel size i.e. the number n of pixels enclosed by the area. The dispersion of the distribution of intensities in the given area depends on n (as $\sqrt{n}$).) i.e. it has nothing to do with observational uncertainties due to noise etc.

Applied to deconvolution problems, only eq. (F5) allows sufficient fluctuations to work reasonably well. The reason it allows fluctuations is the low occupation number. So it works fairly well, but for the wrong reason. And the dispersion has of course nothing to to with actual uncertainties in pixel intensities.

One might attempt to apply the maximum entropy method to find the actual probability distributions using observational constraints. The problem here is that if the entropy is maximized relative to a particular pixel-size, the result becomes pixel dependent. More precisely, in the limit of small pixel-size the result should be invariant in the sense that when averaged to a particular larger pixel-size the result should be independent with respect to the original pixel-size (cf Axiom 2 Caticha and Preuss (2004)). One should note here the important difference to thermodynamic systems where grain-sizes and quantizations are fundamental properties of the system. So, if one wants to use a maximum entropy approach one must maximize the entropy equally with respect to all scales. So we are back to MRD type constraints. With such constraints it appears that the route to obtain proper intensity distributions via Bayes' relation is simpler than the that via the maximum entropy principle. Properly done however the two methods should give the same result. In short, there may be several different routes to derive a correct distribution, one is the maximum entropy method another the Bayes' relation or a mixture of the two. The maximum entropy method introduces constraints via Lagrange multipliers while Bayes uses a prior combined with a likelihood ratio. Depending on the nature of the constraints, one or the other method may be more tractable than the other.

G. sampling rates

Writing eq. (43) in terms of diameters one has,

$$d_{res} = HPBW \left(\ln(4) \ln \left(\frac{2\bar{\alpha}_\nu^2}{\sigma_n^2} \right) \right)^{-1/2} \tag{G1}$$

To avoid loss of information the sampling interval should be less than the resolution diameter. Putting in numbers for example for the Orion^{13}CO observation (cf. Figures 14 and 15) where noise fluctuations are about 0.4K and observed signal $\sim$ 15 K one finds that $d_{res} \sim HPBW/3.3$. For such high signal to noise ratios, some information is lost if the sampling rate is less than $HPBW/3.3$. It should be emphasized that nothing is lost by a tight sampling, except possibly for technical reasons such as moving the telescope or reading and restarting a correlator. The noise considered here is derived only from noise fluctuations in the spectra. There are some indications of the presence of small other sources of error such as pointing errors or baseline fluctuations. A sampling rate of $HPBW/3$ is probably therefore quite sufficient in this case.

REFERENCES

Chang, J.H., Shin, J.W., Kim,N.S.and Mitra, K.S. IEEE SIGNAL PROCESSING LETTERS, VOL. 12, NO. 4, APRIL 2005

H. Jeffreys, Theory of Probability (Clarendon Press, Oxford, 1939).

Narayan, Ramesh; Nityananda, Rajaram, A&A Rev., 1986.24:127-70

Kass, R. E. and Wasserman, L. 1996, J. Am. Stat. Assoc. 91, 1343.

Bryan, R.K., Skilling, J., 1984, MNRAS, 191, 69

Caticha, A., Preuss, R. 2004, Phys. Rev. E, 70.046127

Cornwell, T. J.; Evans, K. F., 1985, A&A, ...143...77C

Curran, S. J.; Rydbeck, G.; Johansson, L. E. B.; Booth, R. S., A molecular outflow in the Circinus galaxy,1999 , A&A, v.344, p.767-778

Curran, S. J.; Johansson, L. E. B.; Rydbeck, G.; Booth, R. S., A molecular ring in the Circinus galaxy,1998 , A&A, v.338, p.863-873

Freeman, P. E.; Kashyap, V.; Rosner, R.; Lamb, D. Q.,2002 ,ApJS, ..138..185F

Gelman, A.,Carlin, B.C., Stern, H.S., and Rubin, D.B. 2003, Bayesian Data Analysis, Chapman Hall.

Hjalmarson, Å., Bergman, P., Bive'r et al.(2005) on behalf of the Odin team, submitted to the Advance in Space Research.

Hsiao, I-T., Rangarajan, A. , and Gene, G., IEEE Trans Image Processing, 11, 1466-1477

Jaynes, E.T., 1957, Phys. Rev. 106, 620.

Lundgren, A. A.; Wiklind, T.; Olofsson, H.; Rydbeck, G., 2004 ,A&A, v.413, p.505-523

Lundgren, A. A.; Olofsson, H.; Wiklind, T.; Rydbeck, G., 2004 ,A&A, v.422, p.865-881

Molecular Gas in the Barred Spiral Galaxy M83 A. A. Lundgren, Wiklind, T.; Olofsson, H.; Rydbeck, G. , Galaxy Disks and Disk Galaxies, Rome June 12-16 2000. ASP Conference Series, Vol. 230. pp. 377-378

Moussaoui, S., Brie, D., Mohammad-Djafari, A., and Carteret, C., IEEE Trans. on Signal Processing, Vol. 54, No. 11, NOV. 2006

Olofsson, H. Bergman, P., Eriksson, K., Gustafsson, B. 1996,A&A, v. 311, p.587-615

Puetter, R.C.; Pia, R.K. , 1994, The Restoration of HST Images and Spectra II.
Space Telescope Science Institute, 1994
R.J. Hanish and R.L. White, eds.

Rodrigues-Franco,A,Wilson.T.L.,Martin-Pintado,J., and Fuente, A.,2001 ,ApJ,559,985-992

Starck,J,L., Pantin,E. and Murtagh F., 2002, PASP, 114,1051-1069

Rodrigues-Franco,A,Martin-Pintado,J., and Fuente, A. 1998, A&A, 329,1097-1110

Rydbeck, G. in The neutral ISM in star-burst galaxies", Marstrand, Sweden, June 24-27, 2003.

Rydbeck, G.; Wiklind, T.; Cameron, M.; Wild, W.; Eckart, A.; Genzel, R.; Rothermel, H.,1993 ,A&A , vol. 270, no. 1-2, p. L13-L16

Rydbeck, G. in Imaging at Radio through Submillimeter Wavelengths, ASP Conference Proceedings, Vol. 217, edited by Jeffrey G. Mangum and Simon J. E. Radford. Astronomical Society of the Pacific, ISBN 1-58381-049-8, 2000., p.224

Stuart A. and Ord J.K., 1987, Kendall's Advanced theory of Statistics Vol 1, Charles Griffin, London, p. 283.

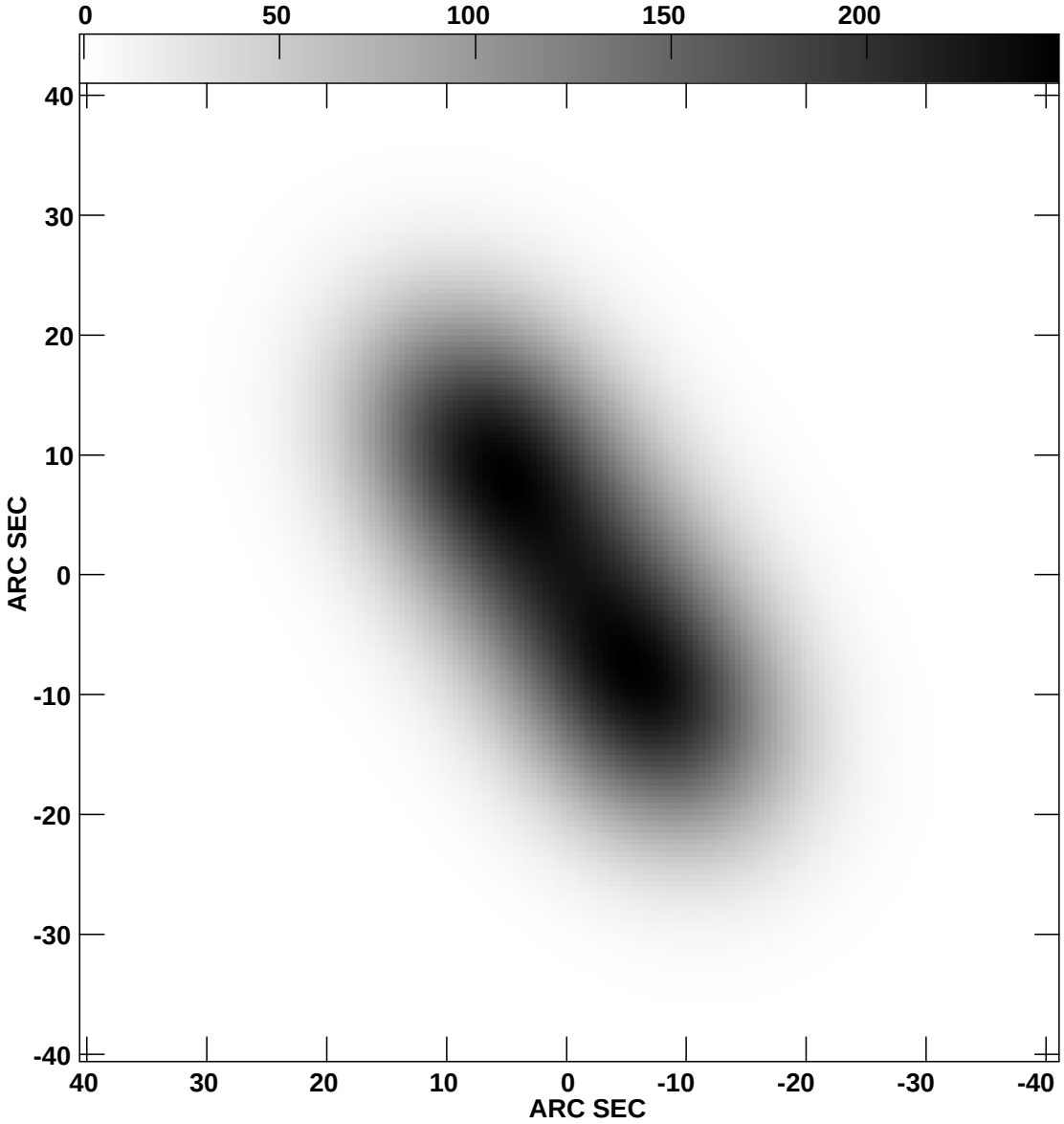

Fig. 3.— The model observed with a 15" beam on a 2.25" spacing and insignificant signal corruption from noise and other sources of error. The map has been interpolated. The intensity scale is Kkm/s

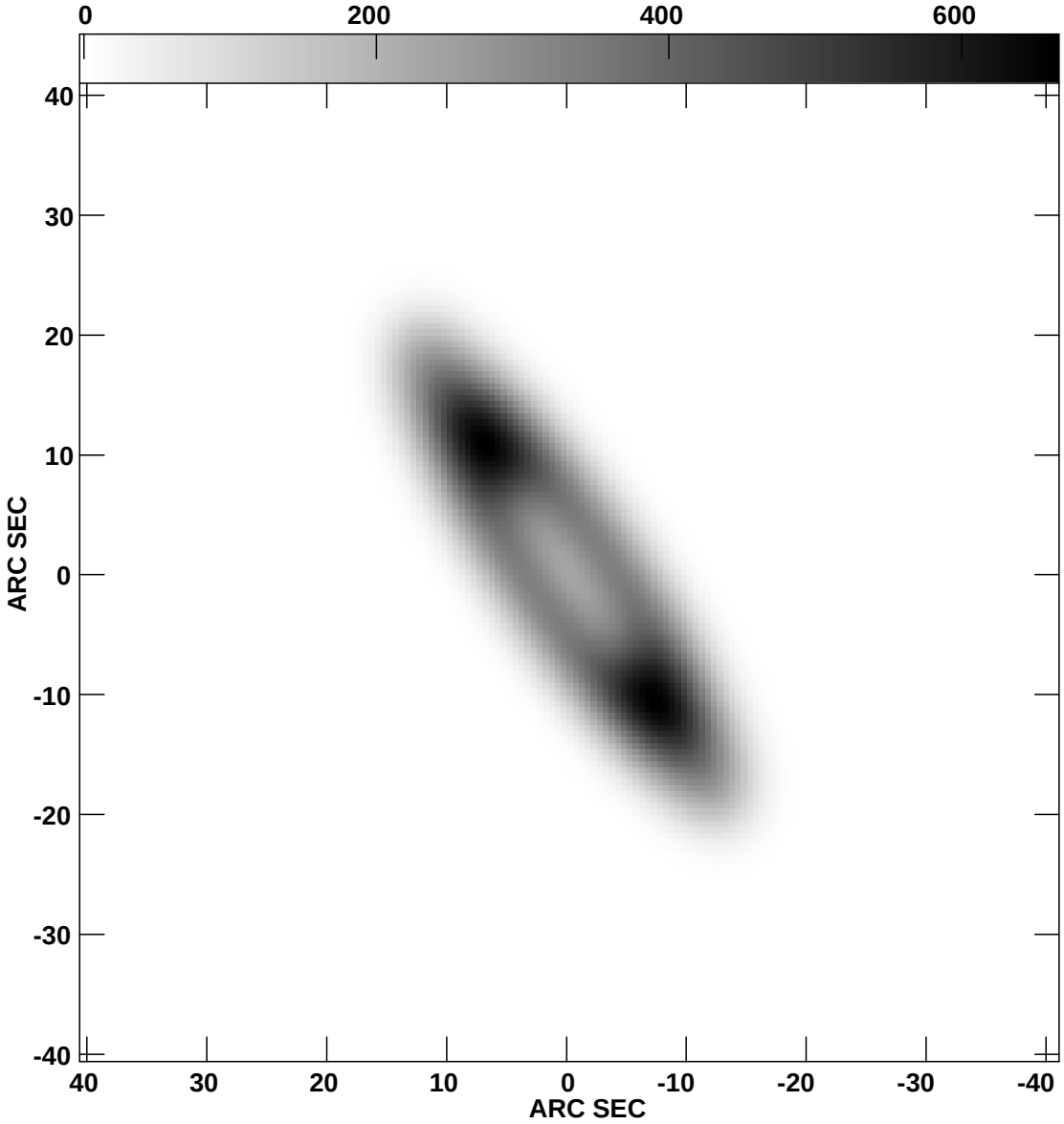

Fig. 4.— The model convolved with a 5" beam. The intensity scale is Kkm/s

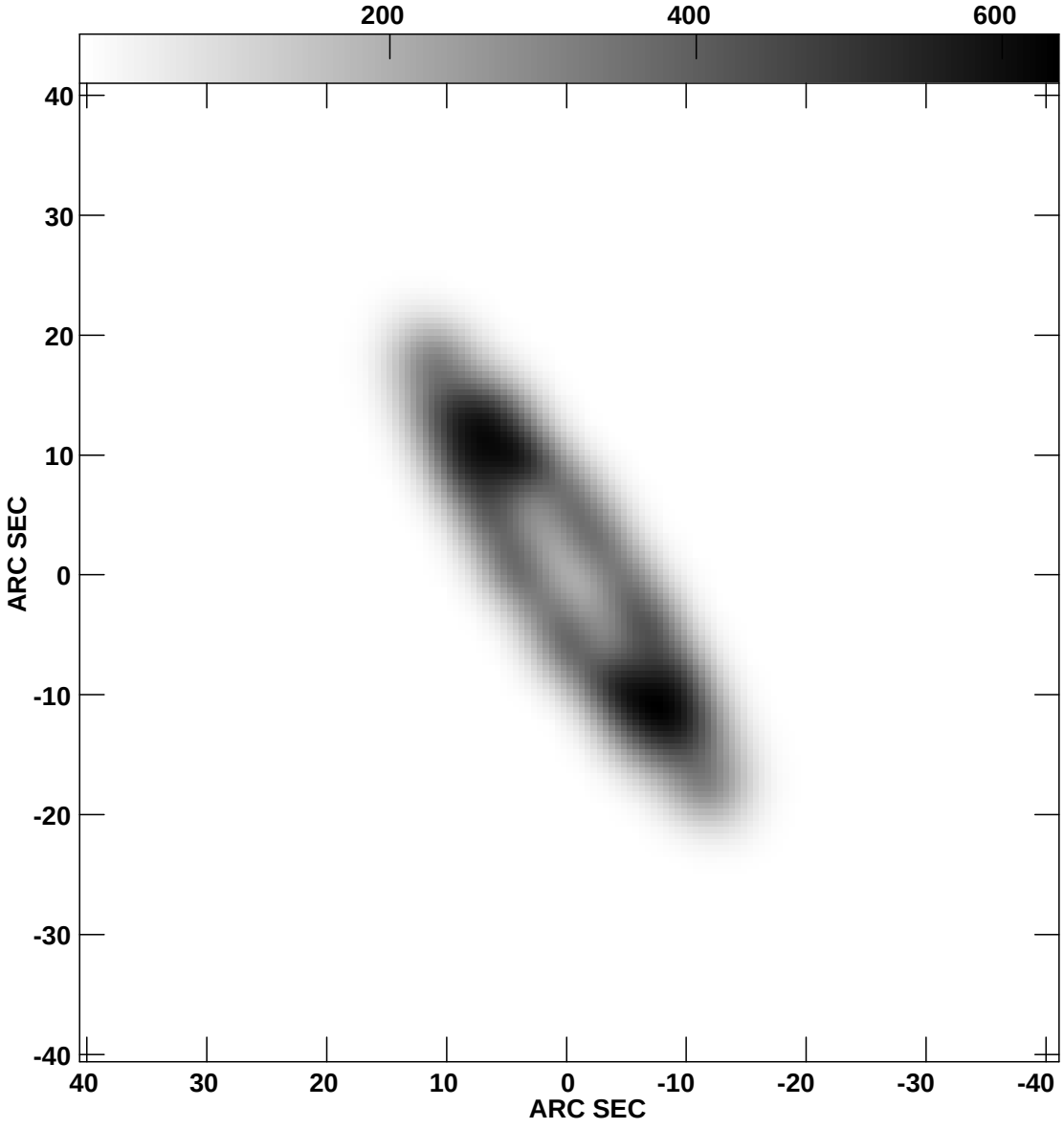

Fig. 5.— The model first observed (Fig. 3), second deconvolved, and third convolved with a 5" beam. The intensity scale is Kkm/s.

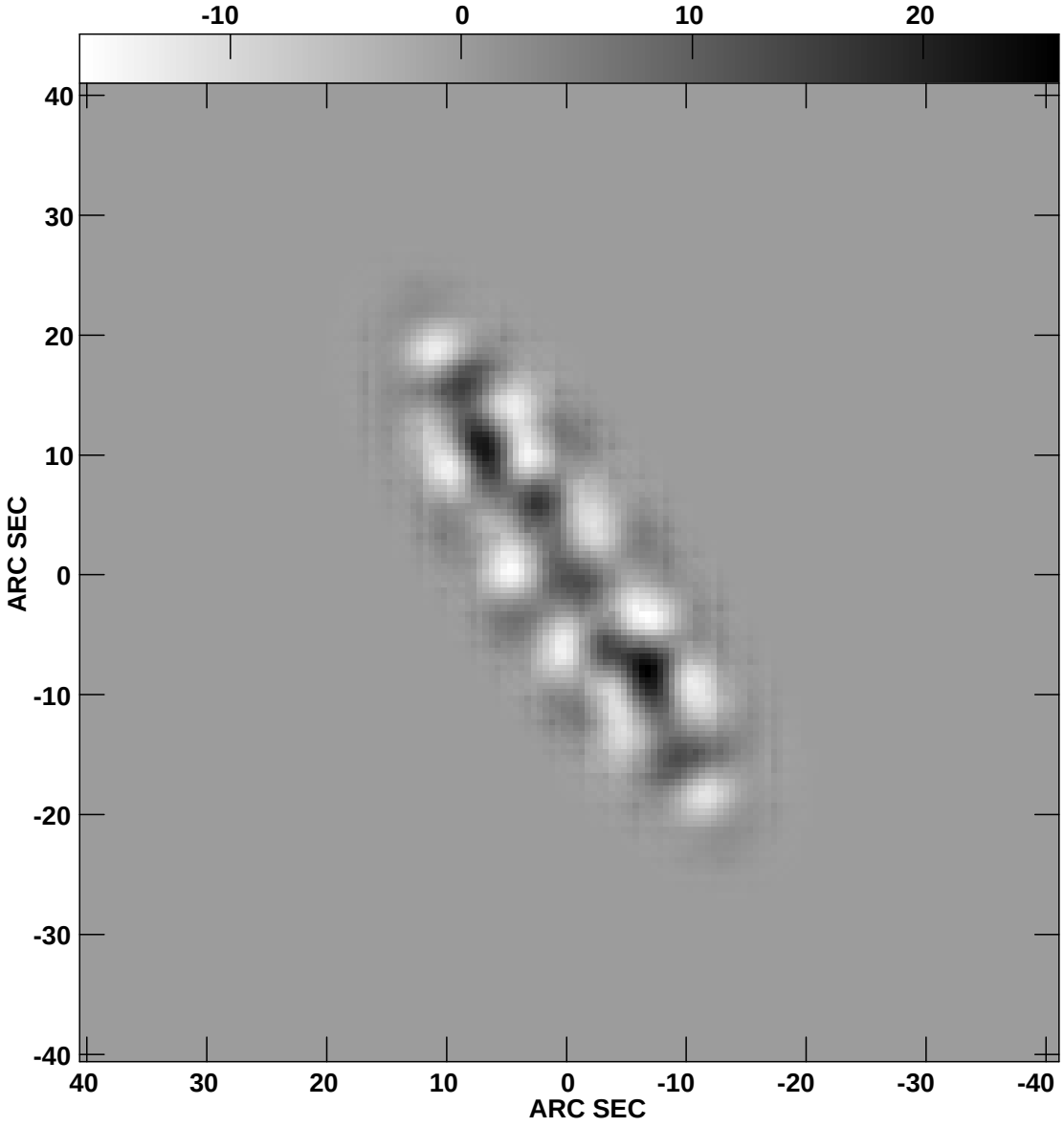

Fig. 6.— The difference between map 4 and 5, has a maximum which is less than 5% of that of map 4.The intensity scale is Kkm/s.

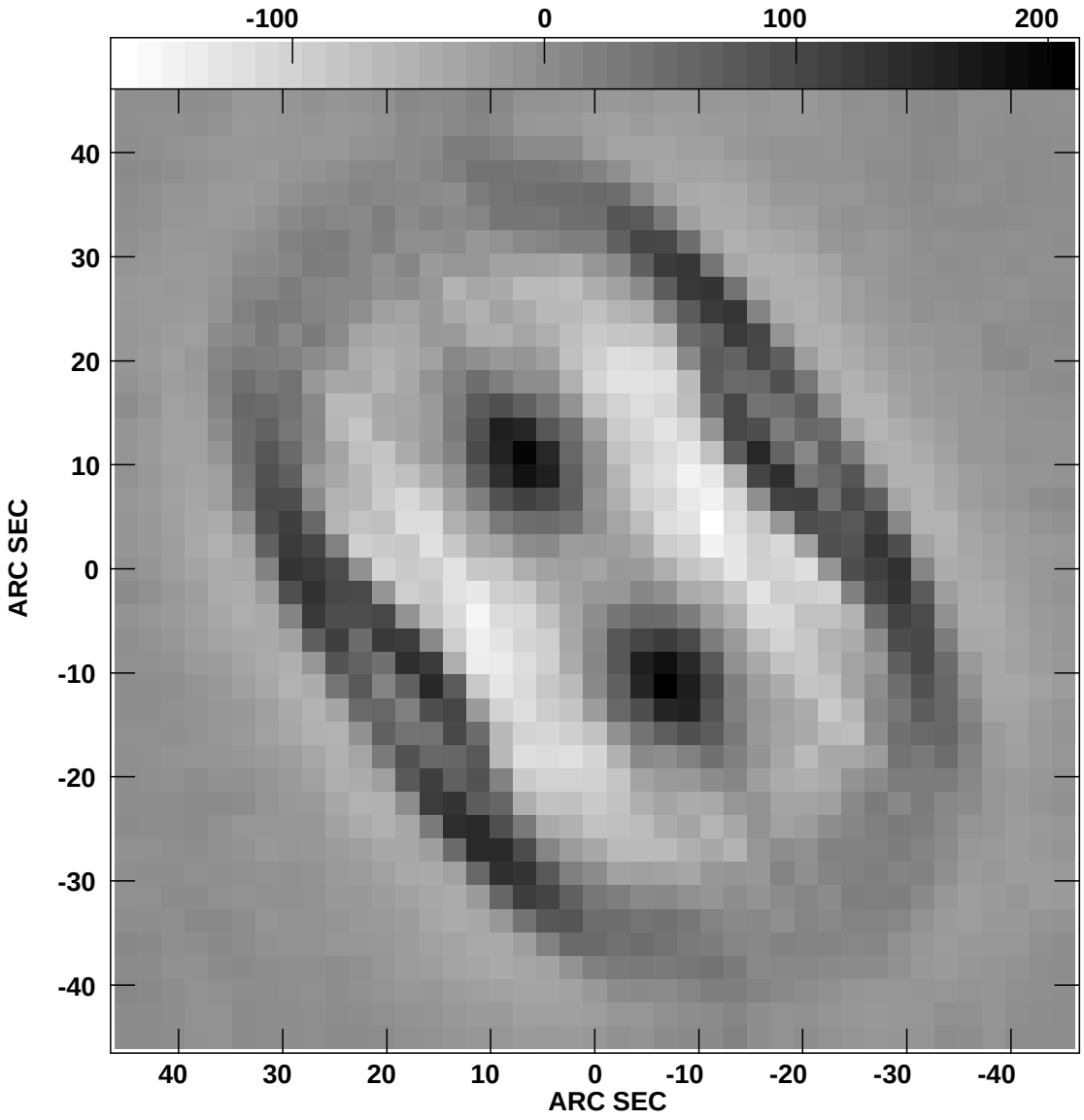

Fig. 7.— The difference between the almost noise free observed and the synthesized integrated intensities, should ideally show only noise cf. Fig. 9 . As the noise level approaches zero imperfections in the IMAGINE procedure will start to become obvious. The intensity scale is milli Kkm/s

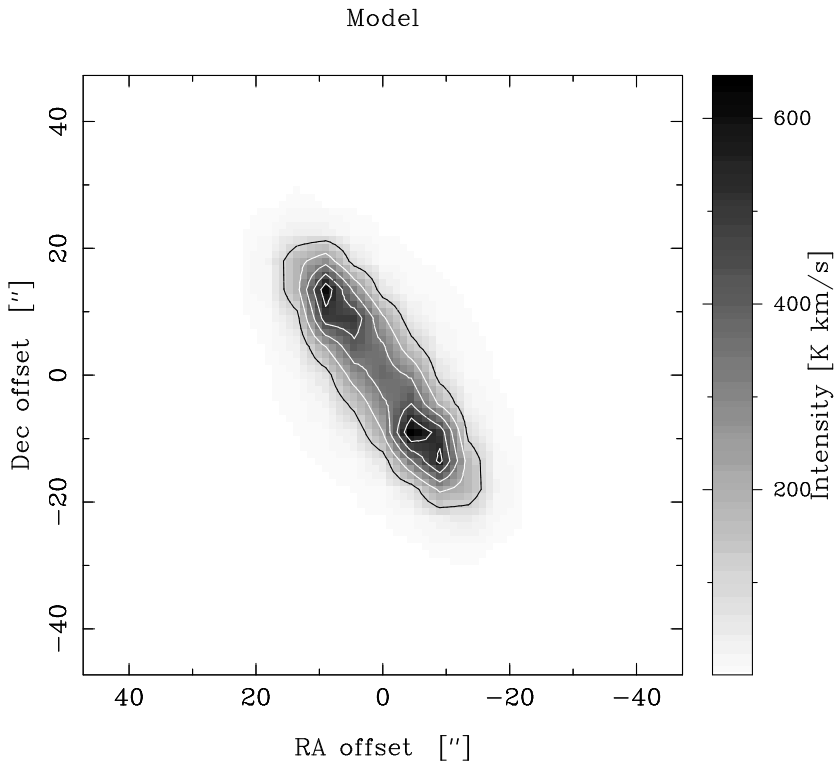

Fig. 8.— The model first observed with noise addition(cf. Fig. 10 and Fig. 9) and second deconvolved in space and frequency. The map has been interpolated.

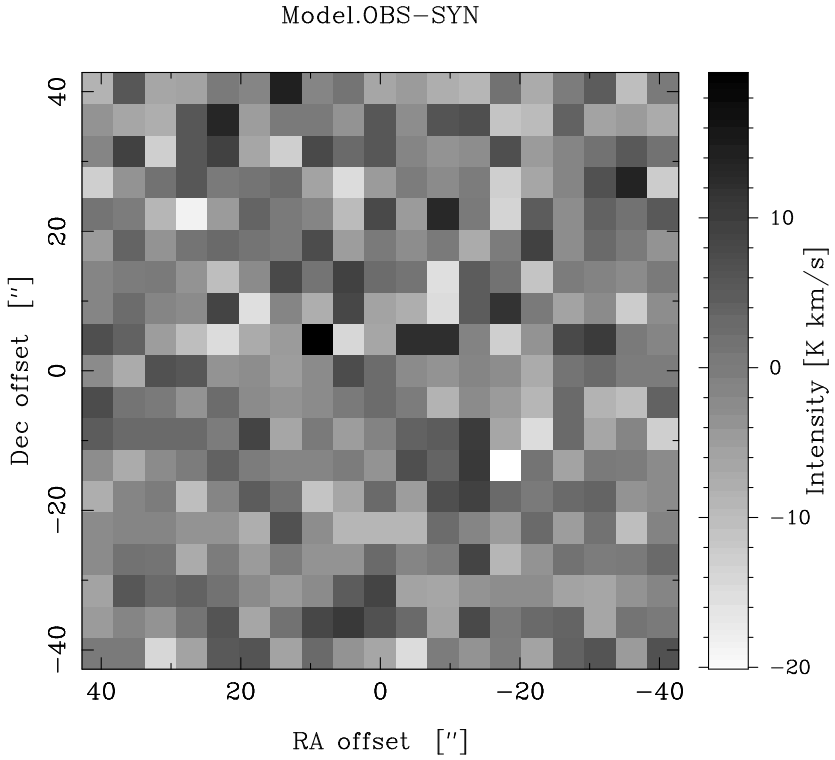

Fig. 9.— Residuals, i.e. the difference between integrated intensities of noisy observed model spectra and synthesized integrated intensities. Note that the residual fluctuations are symmetrically distributed around zero, demonstrating that the IMAGINED flux is conserved.

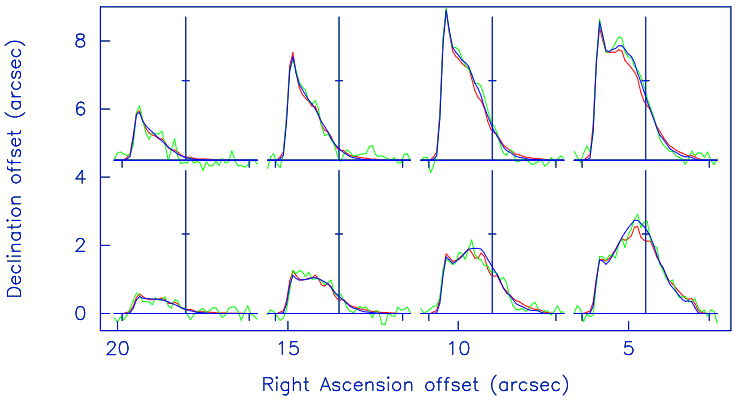

Fig. 10.— Examples of spectra from simulated noisy observations of the disc. The temperature increment is 1K and the velocity increment 200 km/s. The noise temperature is 0.1K at a channel width of 10 km/s. The $HPBW$ is 15" and the grid spacing 4.5". The blue spectra are noise free beam-convolved, the green are observed and the red deconvolved (using the green) and then beam-convolved. The red spectra are influenced by all spectra "within the beam" thereby reducing noise fluctuations.

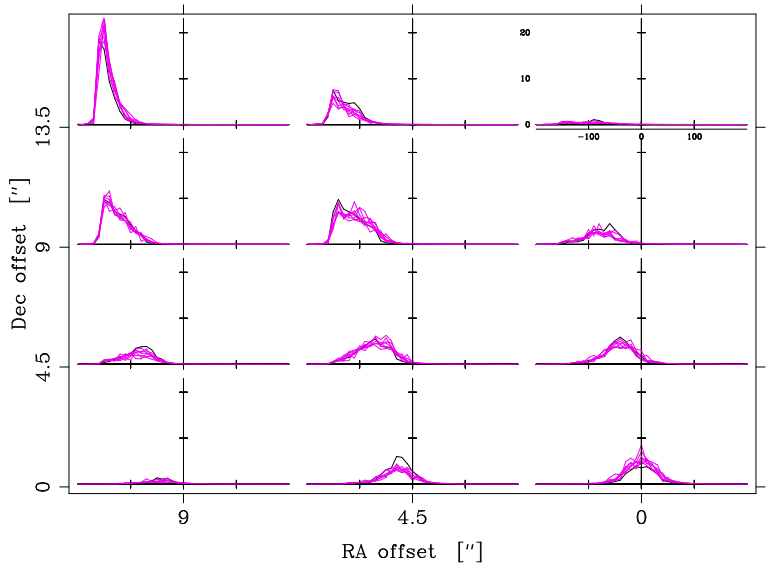

Fig. 11.— True model spectra(black) convolved to 5" resolution (FWHM=5") ,and deconvolved spectra (cerise), from 8 different noisy observations, also convolved to 5". The beam width was 15" and observations were done on a grid with 4.5" spacing. The 'noise temperature is constant but actual noise realizations are different to verify eq. (42). By eye measure the cerise spectra in pos.(0,4.5) have a width of about 1.5 K agreeing with a calculated σ of 0.49 K. The deviation from the true model spectrum is by eye about 3 K also agreeing with a calculated σ of 1 K

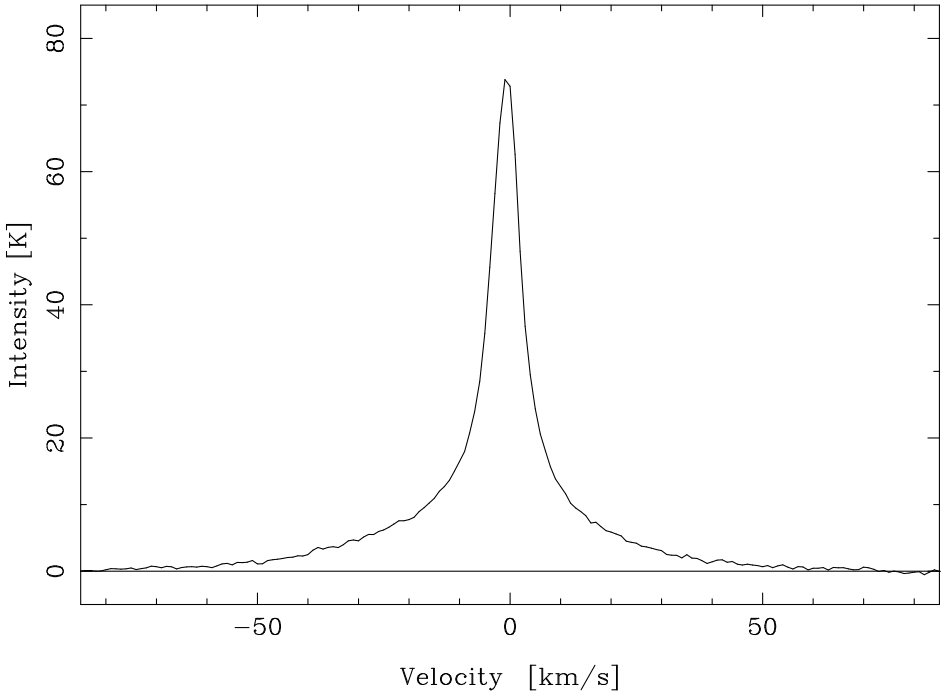

Fig. 12.— ^{12}CO(J=5-4) emission observed toward the center of Orion with the Odin satellite.

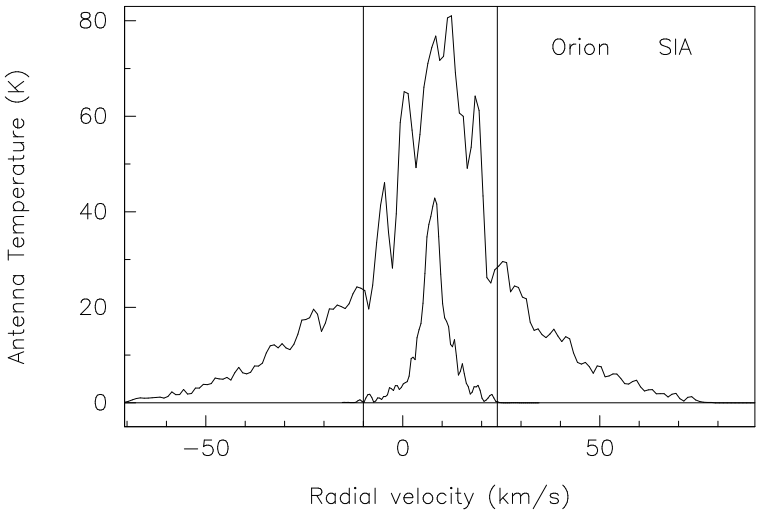

Fig. 13.— Deconvolved ^{12}CO(J=5-4) Orion center spectrum. Note the dramatic appearance of a 30 km/s broad component, marked by the vertical lines, not visible in the original spectra Fig. 12. For comparison the deconvolved ^{13}CO(J=5-4) center spectrum is shown with the actual temperature scale. Also in this spectrum the 30 km/s broad component is clearly seen. The intensity ratio of the ^{12}CO to ^{13}CO 30 km/s component varies between 7 and 10 (when the 150 km/s broad Gaussian ^{12}CO component has been subtracted. It is remarkable that the narrow large scale ^{12}CO component has vanished. It would suggest that that the 30 km/s component fills the beam, which the deconvolved map suggests, and that it has approximately the same temperature as the narrow component.

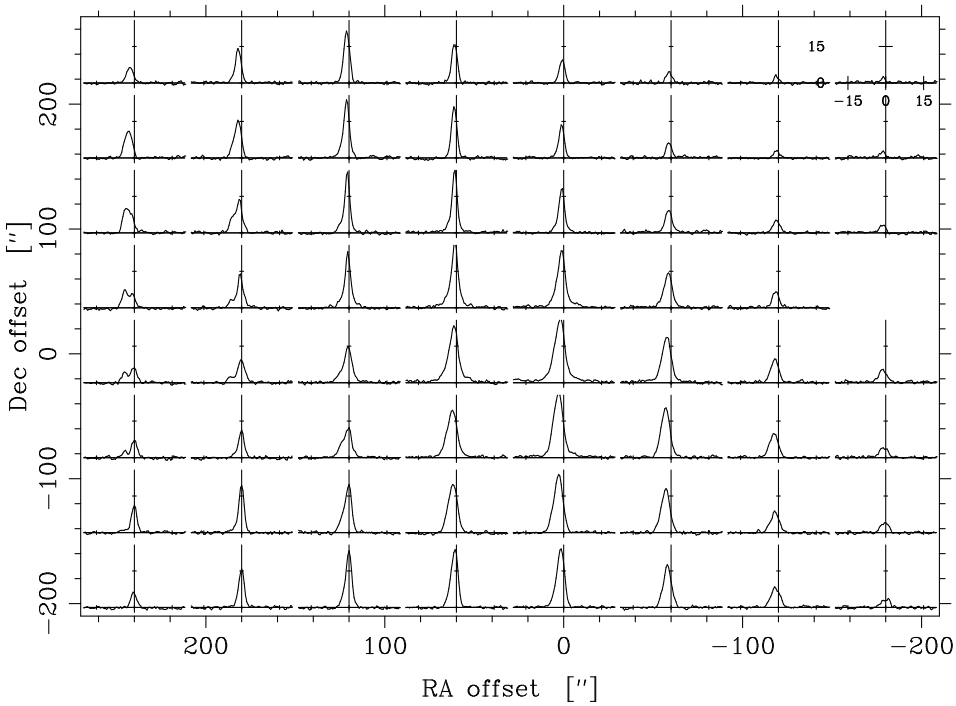

Fig. 14.— Spectral map obtained by observations of emission from the rotational ^{13}CO(J=5-4) transition towards Orion-KL with the Odin satellite telescope. The beam-width (FWHM) is 126".

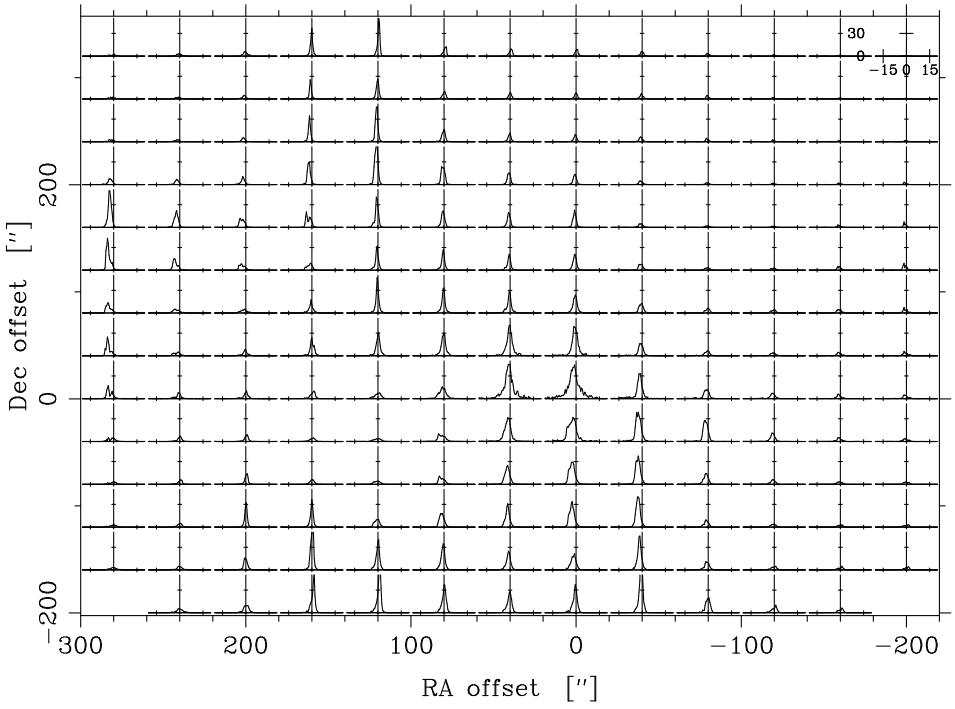

Fig. 15.— Deconvolved Orion ^{13}CO(J=5-4) map. Observational noise fluctuations is about 0.4 K per pixel (40”^{2}) (see text), implying deconvolved pixel channel uncertainties of about 4K, which should be compared to the central pixel intensities which are above 40K, i.e. on the order of 10%. The increased dynamics and appearance of condensed structures in comparison with the observed data in Fig. 14 is noteworthy.

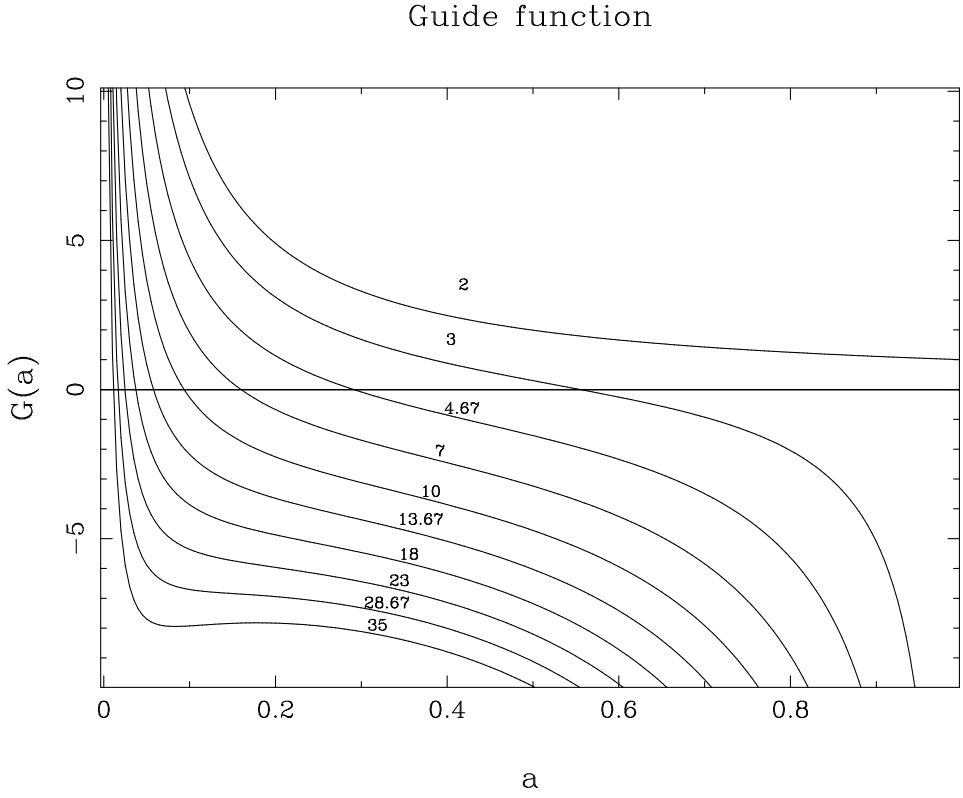

Fig. 16.— $\mathcal{G}$ functions plotted as a function of a, the pixel intensity, for different covering areas C_ν expressed in number of pixel areas n written above the curve. The total summed intensity in C_ν is assumed to be one, so that a can vary between zero and one. As shown in appendix C the function with n=8 has the greatest smallest negative slope of all the curves.