Astronomy & Astrophysics manuscript no. aanda
January 8, 2016

On the reliability of observational measurements of column density probability distribution functions

V. Ossenkopf[1], T. Csengeri[2], N. Schneider[1], C. Federrath[3], and R.S. Klessen[4]

[1] I. Physikalisches Institut, Universität zu Köln, Zülpicher Straße 77, 50937 Köln, Germany
e-mail: `ossk@ph1.uni-koeln.de`
[2] Max-Planck Institut für Radioastronomie, Auf dem Hügel 69, Bonn, Germany
[3] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[4] Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, 69120 Heidelberg, Germany

Received ; accepted

ABSTRACT

Context. Probability distribution functions (PDFs) of column densities are an established tool to characterize the evolutionary state of interstellar clouds.

Aims. Using simulations, we show to what degree their determination is affected by noise, line-of-sight contamination, field selection, and the incomplete sampling in interferometric measurements.

Methods. We solve the integrals that describe the convolution of a cloud PDF with contaminating sources such as noise and line-of-sight emission and study the impact of missing information on the measured column density PDF. In this way we can quantify the effect of the different processes and propose ways to correct for their impact in order to recover the intrinsic PDF of the observed cloud.

Results. The effect of observational noise can be easily estimated and corrected for if the root mean square (rms) of the noise is known. For values below 40 % of the typical cloud column density this involves almost no degradation of the accuracy of the PDF parameters. For higher noise levels and narrow cloud PDFs the width of the PDF becomes increasingly uncertain. A contamination by turbulent foreground or background clouds can be removed as a constant shield if the PDF of the contamination peaks at a lower column or is narrower than that of the observed cloud. Uncertainties in the definition of the cloud boundaries mainly affect the low-column density part of the PDF and the mean density. As long as more than 50 % of a cloud are covered, the impact on the PDF parameters is negligible. In contrast, the incomplete sampling of the uv plane in interferometric observations leads to uncorrectable distortions of the PDF of the produced maps. An extension of ALMA's capabilities would allow us to recover the high-column density tail of the PDF but we found no way to measure the intermediate and low column density part of the underlying cloud PDF in interferometric observations.

Key words. ISM: clouds – ISM: structure – ISM: dust, extinction – Methods: data analysis Methods: statistical – Instrumentation: interferometers

1. Introduction

Probability distribution functions (PDFs) of column density maps of interstellar clouds have been proven to be a valuable tool to characterize the properties and evolutionary state of the clouds (see e.g. Berkhuijsen & Fletcher (2008); Kainulainen et al. (2009); Froebrich & Rowles (2010); Lombardi et al. (2011); Kainulainen et al. (2011); Schneider et al. (2012, 2013); Russeil et al. (2013); Kainulainen & Tan (2013); Hughes et al. (2013); Alves et al. (2014); Druard et al. (2014); Schneider et al. (2015a,b,c); Berkhuijsen & Fletcher (2015) for the observational side and Klessen (2000); Federrath et al. (2008, 2010); Kritsuk et al. (2011); Molina et al. (2012); Konstandin et al. (2012); Federrath & Klessen (2013); Girichidis et al. (2014) for the theoretical side).

Column density maps of dense clouds typically show a log-normal PDF at low column densities, produced by interstellar turbulence (Vazquez-Semadeni 1994; Padoan et al. 1997; Passot & Vázquez-Semadeni 1998; Federrath et al. 2008; Price et al. 2011; Federrath & Banerjee 2015; Nolan et al. 2015), and a power-law tail in the PDF at high densities due to gravitational collapse (Klessen 2000; Kritsuk et al. 2011; Federrath & Klessen 2013; Girichidis et al. 2014). The PDFs seen in 2-D projection allow to estimate the underlying 3-D density PDFs when assuming global isotropy (Brunt et al. 2010a,b; Ginsburg et al. 2013; Kainulainen et al. 2014). The width of the log-normal part allows to quantify the turbulent driving of the interstellar medium (Nordlund & Padoan 1999; Federrath et al. 2008, 2010). If the driving is known, PDF width and Mach number can be used to constrain the magnetic pressure (Padoan & Nordlund 2011; Molina et al. 2012). The slope of the PDF power-law tail can be translated into radial profiles of collapsing clouds when assuming a spherical or cylindrical density distribution (Kritsuk et al. 2011; Federrath & Klessen 2013).

PDFs can be determined from any mapped quantity, not only column densities. For instance, Schneider et al. (2015b) combined PDFs of dust temperatures with the column-density PDFs, Burkhart et al. (2013) compared PDFs of molecular lines with different optical depths, and Miesch & Scalo (1995); Miesch et al. (1999); Hily-Blant et al. (2008); Federrath et al. (2010); Tofflemire et al. (2011) studied PDFs of centroid velocities and velocity increments. Our analysis does not make any assumption on the quantity that is actually analyzed. As a pure statistical analysis, it can be applied to any kind of map. However, for the sake of convenience in naming the quantities and in the direct comparison to the frequent computation of PDFs from column density maps, we stick here to the column density as measured quantity, characterized either in terms of gas columns, visual extinction, or submm-emission from dust with constant temperature. Therefore, we use column density N and A_V synonymously.

Lombardi et al. (2015) discussed the impact of statistical noise, contamination with fore- and background material, and boundary biases on the PDFs of their near-infrared extinction maps and concluded that it is impossible to reliably determine the lower column-density tail of the observed molecular cloud PDFs, leaving only the high-column density power-law tail as significant structure. For column density maps obtained from Herschel PACS and SPIRE continuum observations, Schneider et al. (2015a) showed in contrast that 1) the foreground contamination can be corrected for by assuming a constant screen of material, 2) the selection of the map boundaries only affects column densities well below the column density peak, and 3) the observational noise in the Herschel maps was an order of magnitude lower allowing to fully resolve the log-normal part of the turbulent column-density PDF.

In order to provide a general framework to assess the reliability of column density PDFs obtained in observations, we perform here a parameter study where we simulate different observational biases and limitations with a variable strength to show under which conditions interstellar cloud PDFs can be reliably measured, when they can be recovered using parametric corrections, and when the observational limitations dominate the result.

In Sect. 2, we introduce the formalism and the simulated data used here. In Sect. 3, we study the impact of observational noise, Sect. 4 deals with line-of-sight contamination, Sect. 5 with the selection of map boundaries, and Sect. 6 with the limitations in interferometric observations. Sect. 7 summarizes our findings.

2. Mathematical description

2.1. PDFs

PDFs of column densities $p(N)$ are defined as the probability of the column density in a map to fall in the interval $N, N + dN$. As probabilities they are normalized to unity, i.e. $\int_0^\infty p(N)dN = 1$. If we discuss noisy quantities or other physical quantities that can become negative, the integral should start at $-\infty$. Due to the huge range of densities and resulting column densities in the interstellar medium, it is more appropriate to use logarithmic bins, i.e. we switch to a logarithmic scale $\eta = \ln(N/N_{\rm peak})$, where $N_{\rm peak}$ is the most probable column density in logarithmic bins, i.e. the peak of the distribution on a logarithmic column density scale. The translation between the PDFs on the linear and the logarithmic scale can be easily computed by

$$p_\eta(\eta)d\eta = p_N(N)\frac{\partial N}{\partial \eta}d\eta = N p_N(N)d\eta \; . \tag{1}$$

Most molecular cloud observations are characterized by a log-normal part for the high probability of low densities

$$p_\eta(\eta) = \frac{1}{\sqrt{2\pi}\sigma_{\eta,\rm cloud}} \exp\left(-\frac{\eta^2}{2\sigma^2_{\eta,\rm cloud}}\right) , \tag{2}$$

having typical widths on the logarithmic scale $\sigma_{\rm cloud}$ between 0.2 and 0.5 (see e.g. Berkhuijsen & Fletcher 2008; Kainulainen et al. 2009; Hughes et al. 2013; Schneider et al. 2015a), and a power-law tail at higher densities

$$p(\eta)d\eta = \eta^{-s}d\eta \; , \tag{3}$$

where the exponent s varies in observations between about $s = 1\dots4$ (see e.g. Schneider et al. 2015b; Stutz & Kainulainen 2015). For a free-fall gravitational collapse creating the tail, we expect values $s = 2\dots3$, depending on the geometry of the collapse (Kritsuk et al. 2011; Federrath & Klessen 2013; Girichidis et al. 2014).

All effects of noise, contamination, and edge selection mainly change the low-density regime, so that it is sufficient to concentrate on the log-normal part here. The measurement of the power-law tail can be affected by the impact of insufficient spatial resolution and binning. That was already studied in detail by Lombardi et al. (2010) and Schneider et al. (2015a). For our semi-analytic studies, focusing on low-density effects, we will therefore represent molecular clouds by a log-normal PDF, fully described by the two parameters $N_{\rm peak}$ and $\sigma_{\rm cloud}$, here. A power-law tail is only added for the full cloud simulations (see Sect. 2.3).

The normalization of the density to the peak of the distribution on a logarithmic scale $N_{\rm peak}(\eta)$, deviates from the normalization by $\langle N\rangle$ that is often employed in the literature and that we also used in Schneider et al. (2015a). In principle, the choice of the normalization constant does not affect any of our outcomes. Taking the logarithmic peak is simply more convenient for log-normal distributions centering them at $\eta = 0$. For a log-normal distribution we find a fixed relation between the possible normalization constants. The average column density is $\langle N\rangle = N_{\rm peak}(\eta)\exp(\sigma^2_{\eta,\rm cloud}/2)$ and the peak of the probability distribution in linear units falls at $N_{\rm peak}(N) = N_{\rm peak}(\eta)\exp(-\sigma^2_{\eta,\rm cloud})$. However, we will show in Sect. 5 that a normalization by the PDF peak is much more stable against selection effects than the mean column density. Therefore, we generally recommend the choice of $N_{\rm peak}$ for the column density normalization.

2.2. Contamination

All effects of contamination by observational noise or other line-of-sight material will be linear in column density N, but not in η. At every point in the map we measure a column density that is given by $N_{\rm tot} = N_{\rm cloud} + N_{\rm contam}$, i.e. on a logarithmic scale we find $\eta_{\rm tot} = \ln(N_{\rm cloud} + N_{\rm contam}) - \ln(N_{\rm peak})$. The resulting PDF of a contaminated map then results from the convolution integral of both distributions on the linear scale

$$\begin{aligned} p_\eta(\eta_{\rm tot}) &= N_{\rm tot}\int_{-\infty}^{\infty} p_{N,\rm contam}(N) \\ &\quad\times\frac{p_{\eta,\rm cloud}[\ln(N_{\rm tot}-N)-\ln(N_{\rm peak})]}{N}dN \; . \end{aligned} \tag{4}$$

We can assume a normal distribution for the contamination with noise

$$p_{N,\rm contam}(N) = \frac{1}{\sqrt{2\pi}\sigma_{\rm noise}} \exp\left(-\frac{N^2}{2\sigma^2_{\rm noise}}\right) \tag{5}$$

or another log-normal distribution for a turbulent foreground or background cloud

$$p_{N,\rm contam}(N) = \frac{1}{\sqrt{2\pi}\sigma_{\rm contam}\times N} \exp\left(-\frac{(\ln N - \ln N_{\rm contam})^2}{2\sigma^2_{\rm contam}}\right) \tag{6}$$

where $N_{\rm contam}$ denotes the most probable column density of the contamination on a logarithmic scale. We assume that the contaminating structures are turbulence-dominated so that they do not show any self-gravitating cores yet and can be described by

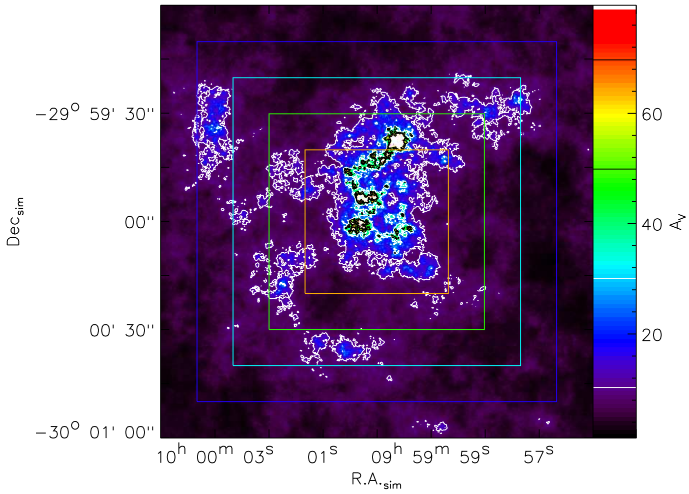

Fig. 1. Test data set representing a fractal cloud with an ideal column density PDF consisting of a log-normal part and a power-law high-density tail. Coordinates are invented to allow for an easy comparison with real observations and to perform simulated ALMA observations in Sect. 6. Contours and colors visualize the same data. The contour levels are indicated in the color bar. Column densities above $A_V = 80$ are saturated in the plot to make the turbulent cloud part better visible. The actual maximum column density is three times larger. As the peak of the log-normal PDF is located at $A_V = 5$, the contours almost cover material in the PDF power-law tail only. According to the low probabilities in the tail, the fraction of green pixels is already small and that of pixels with yellow and red colors is tiny. The four colored squares indicate the subregions discussed in Sect. 5.

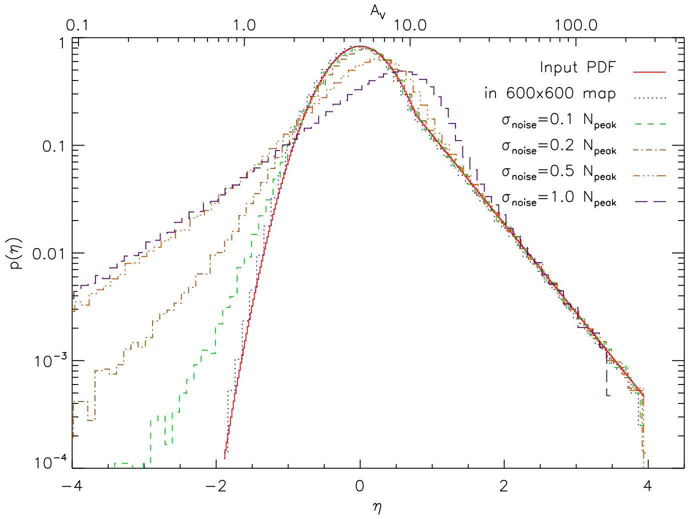

Fig. 2. Impact of observational noise on the PDF measured for the cloud from Fig. 1. The red solid curve shows the analytic description of the input PDF, the blue dotted line is the actually measured PDF of the fractal map with a finite pixel number. The other broken lines show the column density PDF that is measured if the map is affected by different levels of observational noise, characterized by a normal distribution with a standard deviation from 10 % to 100 % of the peak column density.

a pure log-normal distribution without power-law tail. Collapsing parts that produce a power-law tail are easily identified in the maps as small structures so that they can be separated from large-scale contaminations.

In principle, it is possible to compute the original cloud distribution $p_{N,\mathrm{cloud}}(N)$ from the measured distribution $p_N(N_{\mathrm{tot}})$ by inverting the convolution in Fourier space if the contamination PDF, $p_{N,\mathrm{contam}}(N)$, is known

$$p_{N,\mathrm{cloud}}(N) = \mathfrak{F}^{-1}\left(\frac{\mathfrak{F}[p_N(N)]}{\mathfrak{F}[p_{N,\mathrm{contam}}(N)]}\right) \tag{7}$$

where $\mathfrak{F}$ and $\mathfrak{F}^{-1}$ denote the normal and inverse Fourier transform. However, this operation is inherently unstable. It amplifies noise, gridding effects and numerical uncertainties, so that it only works with very well behaving denominators. As Gaussian noise is mathematically well characterized we will test this method in Sect. 3.

2.3. Test data

For a realistic data set that has all properties of a molecular cloud column density map, but no observational limitations we modify a fractional Brownian motion (fBm, see e.g. Stutzki et al. 1998) image to approximate the PDF seen in molecular clouds. The original fBm image has a Gaussian PDF ($p_{N,\mathrm{fBm}}(N)$ given by Eq. 5) and a power-spectral index of 2.8, measured in many molecular clouds (see e.g. Bensch et al. 2001; Falgarone et al. 2004) and consistent with numerical simulations (Kowal et al. 2007; Federrath et al. 2009). Thus it reflects the typical spatial correlations in interstellar cloud maps. To obtain the desired log-normal plus power-law column density PDF with a minimum distortion of the spatial structure we translate the fBm values into cloud column densities by

$$N_{\mathrm{cloud}} = \Pi^{-1}_{\mathrm{cloud}}\left[\Pi_{\mathrm{fBm}}(N)\right] \tag{8}$$

where Π denotes the integral over the normalized PDF

$$\Pi(N) = \int_{-\infty}^{N} p_N(N')dN' \;, \tag{9}$$

varying between zero and unity. The inversion of the function Π^{-1} can only be computed numerically.

For the parameters of the PDF we use values measured by Schneider et al. (2015a) and Schneider et al. (2015c) in several star-forming clouds, e.g. Auriga, Orion B, and Cygnus X. The peak of the PDF N_{peak} is located at $A_V = 5$ and the width of the log-normal part of the PDF in units of the natural logarithm of the column density is $\sigma_{\eta,\mathrm{cloud}} = 0.45$. The transition to the power-law tail occurs at $A_V = 11$ and the tail has an exponent of $s = 1.9$. The PDFs in regions with somewhat lower star-forming activity are sifted towards smaller column densities. Kainulainen et al. (2009); Schneider et al. (2015a) find PDF peaks at $A_V \approx 2$ and deviation points to the power law tail at $A_V = 4\ldots5$. One the η scale, used in all simulations here, this is identical to the PDF of our test data, $p_\eta(\eta)$. Only when translating the results back to an absolute column density scale, the different PDF peak density N_{peak} needs to be applied. We selected a random seed for the fBm that results in a cloud well centered in the map, similar to what an observer might choose.

The resulting map is shown in Fig. 1, where we arbitrarily placed the map on the Southern sky and assigned spatial dimensions in such a way, that it would make sense to observe the cloud with ALMA as simulated in Sect. 6. The PDF of the cloud is shown in Fig. 2.

3. Noise

3.1. Main impact

In a first step we study how the PDF is modified by observational noise, simulated as the superposition of normally distributed random numbers to the map. The result is shown in Fig. 2 where we

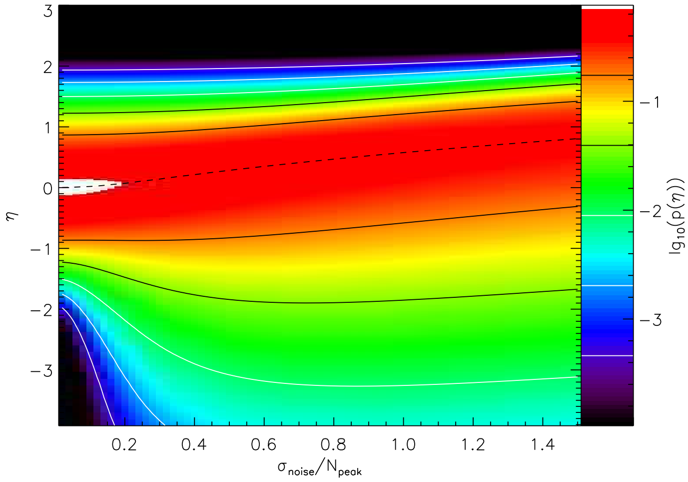

Fig. 3. Distributions obtained from the convolution of a log-normal cloud PDF (Eq. 2 with $\sigma_{\eta,\mathrm{cloud}} = 0.5$) with a Gaussian noise PDF (Eq. 5) as a function of the noise amplitude (σ_{noise}) relative to the peak of the log-normal distribution (N_{peak}). Contours and colors visualize the same data, given on a logarithmic scale. The individual contour levels are marked in the color bar. The left edge of the figure represents the original log-normal PDF. When increasing the noise level σ_{noise} in the right direction, we see the distortion of the PDF mainly at low column densities η. To better follow the shift of the PDF maximum to larger column densities with increasing noise level, the dashed line indicates the peak of the distribution.

varied the noise amplitude, i.e. the root-mean-square (rms) of the noise distribution between 10 % and 100 % of the peak density, N_{peak}, of the cloud PDF. We see three effects:
i) An increasing noise level produces a low-density excess in the PDF, i.e. the log-normal part of the PDF is widened towards lower densities. If the low densities are completely dominated by noise, like in the case of noise amplitudes of 50 % or more of the peak column density, we find a linear behavior of the PDF at low densities $p(\eta) \propto \eta$ due to the factor N in Eq. 1 applied to the Gaussian noise contribution (Eq. 5) when computing the convolution integral (Eq. 4).
ii) For high noise levels, the additional "signal" from the noise also shifts the peak of the distribution towards higher column densities, i.e. may affect the measurement of N_{peak}.
iii) For the power-law tail towards high column densities, however, we find no significant impact. This confirms our earlier findings (Schneider et al. 2015a) and the ones of Lombardi et al. (2015) that the log-normal part of the PDF is easily affected by noise, but that the power-law tail is stable.

3.2. Parametric corrections

A correction of the noise effect for the log-normal part of the PDF is, however, possible if the amplitude of the noise contamination is known. To obtain such a correction, we perform a parameter study based on the analytic description of the PDFs in Eqs. (2-5), ignoring the power-law tail that is anyway not sensitive to the noise contamination. Figure 3 shows a first step in such a parameter scan for the noise amplitude. Here, we display all PDFs for the different noise levels in a color-contour plot. To represent the PDFs in the two-dimensional plot equivalent to the curves in Fig. 2 we show the decadic logarithm of $p_\eta(\eta)$ in colors and contour levels. The PDFs are computed from the convolution integral (Eq. 4) using a log-normal cloud PDF with $\sigma_{\eta,\mathrm{cloud}} = 0.5$ and Gaussian noise (Eq. 5) with a varying amplitude. The left edge of the figure represents the original log-normal PDF. With increasing noise level σ_{noise} towards the right part of the plot, we see the increasing distortion of each part of the PDF. This allows us to quantify and correct the distortions based on the noise level. Even for low noise levels there is a low-column-density excess leading to a widening of the PDF. It stays relatively constant for noise levels $\sigma_{\eta,\mathrm{cloud}} \gtrsim 0.5 N_{\mathrm{peak}}$. For noise amplitudes of $\gtrsim 0.2 N_{\mathrm{peak}}$ the PDF peak is also shifted towards higher column densities.

As the noisy PDFs are no longer log-normal they are no longer characterized by two parameters only. We have to discriminate between different methods that can be used for measuring width and peak position in PDFs. The problem is similar to the measurement of the position and width of individual lines measured in a spectroscopic observation. The use of the absolute peak for the position is sensitive to the details of the binning. It can be affected by fluctuations due to the low-number sampling in individual bins for a finite map size. This becomes significant for the submaps that we consider in Sect. 5. In contrast, the moments of the distribution, being independent of the selected binning, are strongly affected by the structure of the wings. This would prevent us from applying the numbers determined here, for the contamination of a purely log-normal distribution, to a molecular cloud structure that shows an additional power-law tail. The compromise, that is often used in the case of spectroscopic observations, is a Gaussian fit to the distribution, i.e. a log-normal fit on our logarithmic η scale. For a log-normal distribution, all three approaches provide the same numbers, but for the noise-contaminated distributions they can deviate from each other.

Figure 4 shows the PDF parameters determined in three different ways as a function of the amplitude of the noise contamination. The upper plot gives the center position measured through the maximum, the first moment, and as the center of a log-normal fit. They agree at zero noise contamination. The position of the peak probability gives the highest column density values for the PDF center, matching the dashed line in Fig. 3. When considering the dotted curve for the PDF centroid, we find a decrease for the center of the distribution at low noise levels compared to the noise-free case due to the widening of the distribution towards small column densities. The center of a log-normal fit shows an intermediate behavior.

The lower plot shows the measured widths of the distribution together with the next higher moments[1]. The width of a log-normal fit to the distribution is very stable. It increases by 10 % only. The standard deviation σ grows much more, by almost a factor of 2. With the formation of the low-density tail due to the noise contamination, the skewness of the distribution quickly drops to about −1.5 and the kurtosis grows. Once the linear dependence seen in Fig. 2 is reached, the skewness remains almost constant and only the steepening of the high-density tail with the shift of the PDF peak is traced through a reduction of the kurtosis. As the Gaussian fit is least affected by the formation of the noise tail and it is also applicable for distributions with a power-law high-density tail, we will continue with this log-normal fit characterization for more extended parameter scans.

When assuming a log-normal cloud PDF, the noise contamination problem is fully described by two parameters, the width of the cloud PDF and the amplitude of the noise distribution relative

[1] The third moment of a distribution is the skewness defined as $S = 1/N \times \sum(\eta - \langle\eta\rangle)^3/\sigma^3$, the fourth moment the kurtosis $K = 1/N \times \sum(\eta - \langle\eta\rangle)^4/\sigma^4$ where the sums run over all N pixels of a map and σ denotes the standard deviation of the distribution in η.

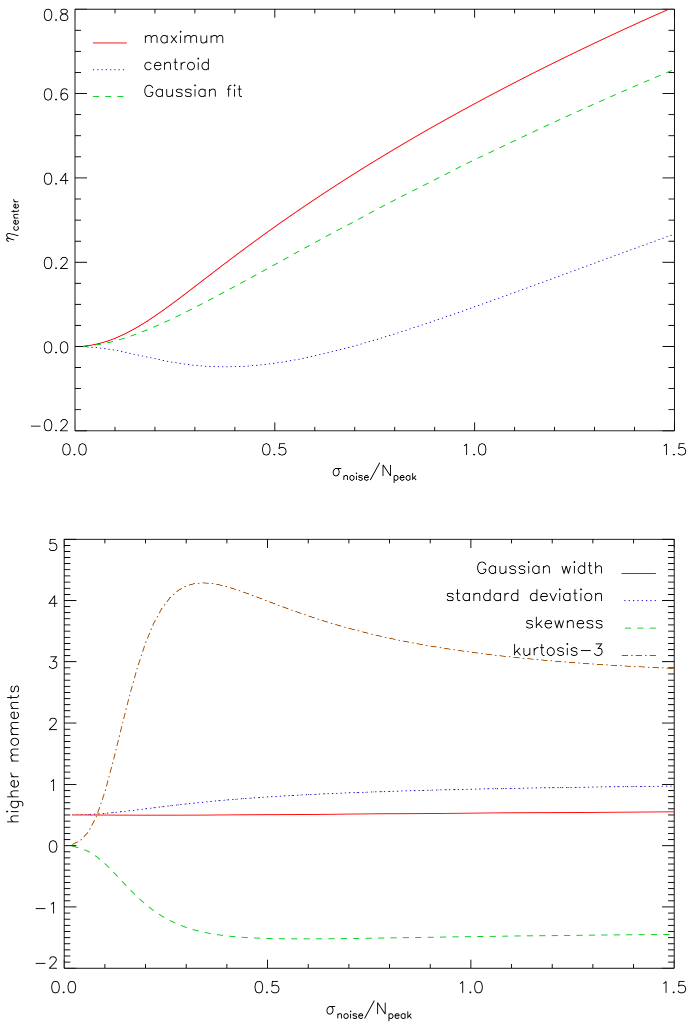

Fig. 4. Change of the measured PDF parameters as a function of the noise amplitude. The upper plot shows the measured center position using three different measures for the position. The red solid line shows the position of the maximum probability, the blue dotted line shows the first moment of the distribution, and the green dashed line gives the center of a Gaussian fit (in η). The lower plot compares the standard deviation with the width of the Gaussian fit, also showing the higher moments.

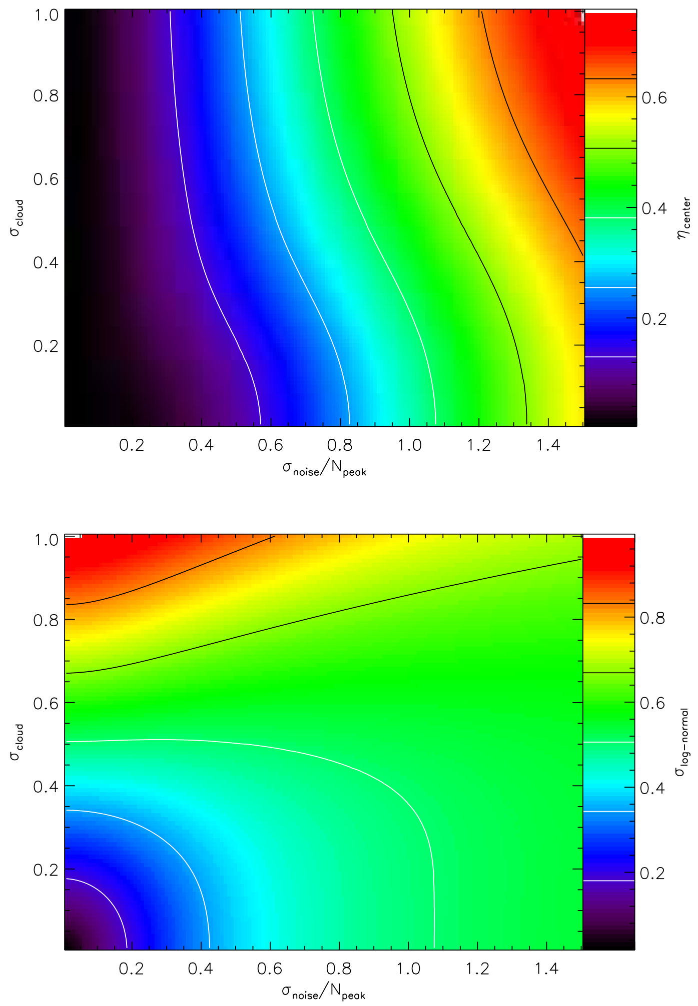

Fig. 5. Variation of the parameters of a log-normal fit to the PDFs of noise-contaminated maps as a function of noise amplitude and the width of the original cloud PDF. The upper plot shows the change of the PDF center, the lower plot the width of the log-normal fit. Without noise contamination $\eta_{\rm center} = \eta_{\rm peak} = 0$ and the measured width matches the PDF width of the cloud, $\sigma_{\rm log-normal} = \sigma_{\rm cloud}$.

to the peak of the cloud PDF. Linear scaling to any cloud observation is straightforward. This allows us to provide a complete picture for the correction of noise effects. Figure 5 shows the change of the PDF peak position and width, determined through the log-normal fit, as a function of the noise contamination amplitude and the width of the original cloud PDF. For all cloud distributions we find an almost linear increase of the observed peak column density with noise amplitude. The highest shift of the peak position occurs for broad cloud distributions. The fitted PDF width reflects the properties of the underlying cloud only for noise amplitudes below about 0.4-0.5 $N_{\rm peak}$. At large noise amplitudes, the measured width becomes almost independent of the input cloud properties, saturating at $\sigma_{\rm log-normal} = 0.5 \ldots 0.7$.

Figure 5 contains all information that is needed to deduce the cloud properties $N_{\rm peak}$ and $\sigma_{\rm cloud}$ from a measured noisy PDF when knowing the noise amplitude. For any value of $\sigma_{\rm noise}/N_{\rm peak}$ in the plot, we can transform the measured peak column density $N_{\rm center}$ into the corresponding logarithmic parameter $\eta_{\rm center} = \ln(N_{\rm center}/\sigma_{\rm noise} \times \sigma_{\rm noise}/N_{\rm peak})$ and compare it with the values given in the the upper plot in Fig. 5. The measured width $\sigma_{\rm log-normal}$ can be directly looked up in the lower plot. Hence, we can deduce the input parameters characterizing the cloud PDF $\sigma_{\rm noise}/N_{\rm peak}$ and $\sigma_{\rm cloud}$ from a fit to the measured parameters $\sigma_{\rm noise}/N_{\rm center}$ and $\sigma_{\rm log-normal}$. The quality of the fit can be quantified in terms of the quadratic deviations of both parameters

$$\chi^2 = \left(\frac{\sigma_{\rm noise}}{N_{\rm center,meas}} - \frac{\sigma_{\rm noise}}{N_{\rm center,Fig.5}}\right)^2 + \left(\sigma_{\rm log-normal,meas} - \sigma_{\rm log-normal,Fig.5}\right)^2 \tag{10}$$

where we assume the same relative accuracy for both parameters.

For any measured pair of $\sigma_{\rm noise}/N_{\rm center,meas}$ and $\sigma_{\rm log-normal,meas}$ we obtain the cloud parameters from the location of the χ^2 minimum in the $\sigma_{\rm noise}/N_{\rm peak} - \sigma_{\rm cloud}$ parameter space[2]. This is visualized in Fig. 6. To save space, we combined

[2] At `http://www.astro.uni-koeln.de/ftpspace/ossk/noisecorrectpdf` we provide an IDL program that uses the precomputed surfaces from Fig. 5 to provide such a fit for any input map and given noise rms. There the Gaussian fit is only performed down to 1/2 of the peak to minimize the effect of distortions by PDF tails and wings.

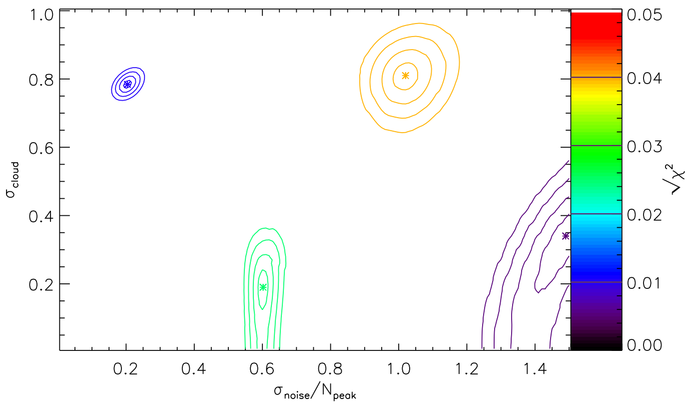

Fig. 6. $\sqrt{\chi^2}$ surfaces for the fit of the measurable parameters $\sigma_{\rm noise}/N_{\rm center}$ and $\sigma_{\rm log-normal}$ through the cloud parameters $\sigma_{\rm noise}/N_{\rm peak}$ and $\sigma_{\eta,\rm cloud}$ for 4 different noise-contaminated clouds with log-normal column-density PDFs. Blue contours stand for a cloud with $\sigma_{\rm cloud} = 0.8$ and $\sigma_{\rm noise} = 0.2N_{\rm peak}$; green contours for $\sigma_{\rm cloud} = 0.2$, $\sigma_{\rm noise} = 0.6N_{\rm peak}$; yellow contours for $\sigma_{\rm cloud} = 0.8$, $\sigma_{\rm noise} = 1.0N_{\rm peak}$; and purple contours for $\sigma_{\rm cloud} = 0.2$, $\sigma_{\rm noise} = 1.4N_{\rm peak}$. They show the square root of the sum of the quadratic deviations of both observables from the measured values (see Eq. 10) at levels $\sqrt{\chi^2} = 0.01, 0.02, 0.03,$ and 0.04.The asterisks mark the minima.

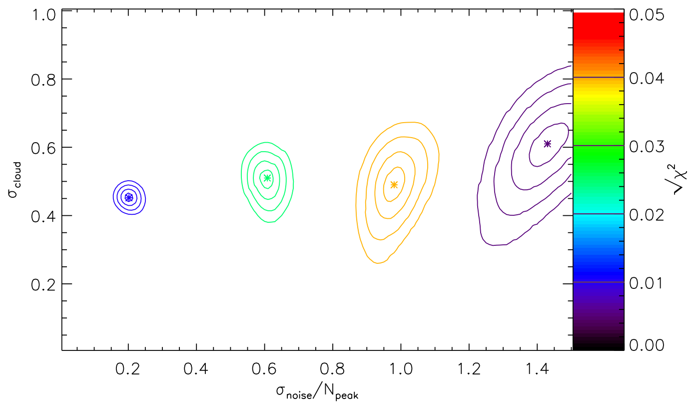

Fig. 7. $\sqrt{\chi^2}$ surfaces for the fit of the measurable parameters $\sigma_{\rm noise}/N_{\rm center}$ and $\sigma_{\rm log-normal}$ through the cloud parameters $\sigma_{\rm noise}/N_{\rm peak}$ and $\sigma_{\eta,\rm cloud}$ for the example cloud from Fig. 1 contaminated with noise levels of 0.2, 0.6, 1.0, and 1.4 $N_{\rm peak}$ (blue, green, yellow, and purple contours). The contours are drawn at $\sqrt{\chi^2} = 0.01, 0.02, 0.03,$ and 0.04. The asterisks mark the minima.

the $\sqrt{\chi^2}$ plots for four different models in one figure. To cover a wide parameter range, we selected two log-normal cloud distributions with $\sigma_{\rm cloud} = 0.8$ and 0.2 and contaminated the first one with noise levels of $0.2N_{\rm peak}$ (blue) and $1.0N_{\rm peak}$ (yellow) and the latter one with noise of $0.6N_{\rm peak}$ (green) and $1.4N_{\rm peak}$ (red). The contours show levels of $\sqrt{\chi^2}$ surfaces from 0.01 to 0.04.

As we fit the observational parameters $\sigma_{\rm noise}/N_{\rm center,meas}$ and $\sigma_{\rm log-normal,meas}$ in the space of the cloud parameters $\sigma_{\rm noise}/N_{\rm peak}$ and $\sigma_{\rm cloud}$, having the same units, we can directly read the accuracy of the parameter determination from the topology of the $\sqrt{\chi^2}$ surfaces. For the example of a low noise contamination ($\sigma_{\rm noise} = 0.2\ N_{\rm peak}$) and broad cloud distributions ($\sigma_{\eta,\rm cloud} = 0.8$), the contours are quite round and the $\sqrt{\chi^2} = 0.04$ contour has a diameter of about 0.08 in terms of the cloud parameters showing that the cloud parameters are recovered with almost the same accuracy to which the observational parameters could be measured. If we increase the noise contamination instead to $\sigma_{\rm noise} = 1.0\ N_{\rm peak}$ the accuracy of the parameter recovery drops by a factor of three, seen by thr three times bigger diameter of the contours. For the model with the narrow cloud distribution, $\sigma_{\eta,\rm cloud} = 0.2$, the parameter retrieval is limited by the degeneracy of the measured cloud width, $\sigma_{\rm log-normal}$, seen as extended green plateau in the lower right corner of Fig. 5. For noise amplitudes $\sigma_{\rm noise} \gtrsim 0.4\ N_{\rm peak}$ the solution in terms of the cloud widths, $\sigma_{\eta,\rm cloud}$, spans a very wide range while the peak position is still well constrained. Unfortunately, this applies to many molecular clouds having narrow widths $0.19 \le \sigma_{\eta,\rm cloud} \le 0.53$ (Berkhuijsen & Fletcher 2008; Hughes et al. 2013; Schneider et al. 2015a). In contrast, for values of $\sigma_{\eta,\rm cloud} > 0.5$ we rather find a gradual decrease of the accuracy of the fit of both parameters.

Figure 7 shows the result for our example cloud from Fig. 1 with $\sigma_{\eta,\rm cloud} = 0.45$ and the additional power-law tail. We show again four different fits in one figure providing the $\sqrt{\chi^2}$ surfaces for models with a noise contamination amplitude of 0.2, 0.6, 1.0, and 1.4 $N_{\rm peak}$, shown as blue, green, yellow, and red contours, respectively.

Here, we are at the edge of the parameter range dominated by the cloud width degeneracy. For a low noise contamination ($\sigma_{\rm noise} < 0.3\ N_{\rm peak}$) the input parameters are accurately recovered. At a noise level of $\sigma_{\rm noise} = 0.6\ N_{\rm peak}$ the accuracy of the parameter determination drops by about a factor of two, but at higher noise levels, the cloud-width degeneracy of the observed PDF also produces an elongated parameter space for the solution, strongly reducing the accuracy of the recovery of the actual cloud PDF width, $\sigma_{\rm cloud}$.

For noise contamination levels $\sigma_{\rm noise} > N_{\rm peak}$ the determination of the cloud parameters becomes very uncertain. An uncertainty of 0.05 in the PDF center position of log-normal width translates into a ten times larger uncertainty of the fitted cloud parameters. Altogether, we can reliably determine $N_{\rm peak}$ from the map and a given noise amplitude if the noise rms is smaller than the PDF peak column density; for a reliable determination of the PDF width, the noise amplitude has to fall below about $0.4 \times N_{\rm peak}$.

3.3. The zero-column-density PDF

The situation is different if the noise level of the observations is a priori not known, e.g. in continuum data with unknown receiver sensitivity. In this case, the PDF may first be used to estimate the noise in the observational data, using the probability of zero intensities/column densities. As a log-normal PDF (Eq. 2) has a zero probability of zero column densities, due to the scaling with N (Eq. 1), one can attribute all measured values of zero column density to the observational noise. Thus we can measure the noise amplitude through inspecting the linear-scale PDF, $p_N(0)$. Establishing the relation between the noise amplitude $\sigma_{\rm noise}/N_{\rm peak}$ and the zero-column PDF, $p_N(0)$, is also required for using $p_N(0)$ in a second step to measure the line-of-sight contamination in Sect. 4.

Figure 8 shows the value of the linear column density PDF for $N = 0$ from the convolution of log-normal cloud PDFs with different widths, $\sigma_{\eta,\rm cloud}$, with a normal noise PDF as a function of the noise amplitude. For very small noise amplitudes and narrow cloud PDFs, $p_N(0)$ still vanishes, but there is a systematic increase with noise amplitude and cloud PDF width.

Closer inspection of the curves from Fig. 8 show that they can be characterized by another log-normal function, providing

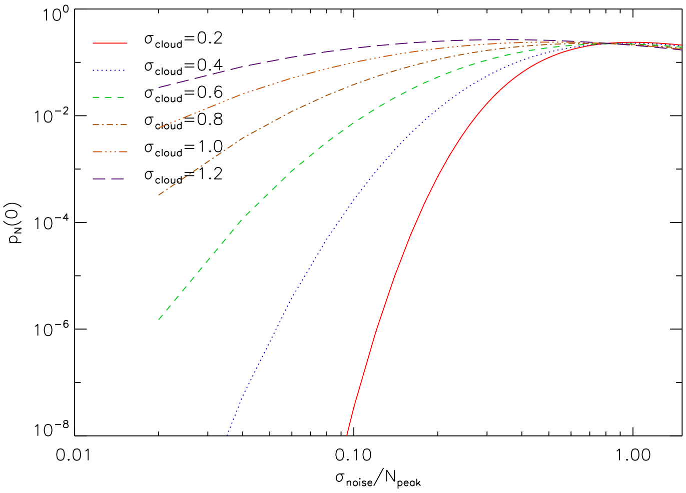

Fig. 8. Values of the linear column density PDF p_N at zero column densities as a function of the observational noise level for different values of the cloud PDF width. The PDF values $p_N(0)$ are given in units of $1/N_{\rm peak}$.

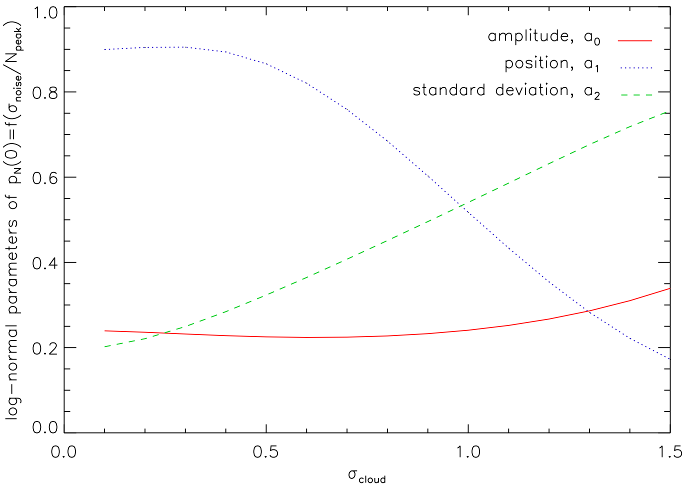

Fig. 9. Parameters of the log-normal description of the zero-column density PDF ($p_N(0)$, Eq. 11) as a function of the noise amplitude. The values give the center, a_1, amplitude, a_0, and width, a_2, of the Gaussian on a logarithmic scale in units of the cloud PDF peak $N_{\rm peak}$.

the zero column PDF as a function of the noise level,

$$p_N(0) = \frac{a_0}{N_{\rm peak}} \exp\left(-\frac{[\ln(\sigma_{\rm noise}/N_{\rm peak}) - \ln(a_1)]^2}{2a_2^2}\right) \tag{11}$$

within a few percent accuracy up to $\sigma_{\rm noise}/N_{\rm peak} = 1$. All curves in the plot show a clear parabola shape[3]. Hence, we can express the information about the zero-column PDFs by the three parameters, a_0, a_1, a_2, for the amplitude, position and standard deviation of these new log-normal curves determined by the cloud width $\sigma_{\eta,\rm cloud}$. For a full parametric description of the zero column PDFs, we show the parameters of the log-normals, $a_0(\sigma_{\eta,\rm cloud}), a_1(\sigma_{\eta,\rm cloud}), a_2(\sigma_{\eta,\rm cloud})$, in Fig. 9 as functions of the cloud PDF width. As visible already in Fig. 8, the parabolas describing the zero-column density PDF become wider with increasing cloud PDF width, the peak position moves to somewhat

[3] The log-normal function for the noisy zero column PDF as a function of the noise level must not be confused with any of the log-normal functions discussed so far, all describing the probability distribution of an (observed) column density.

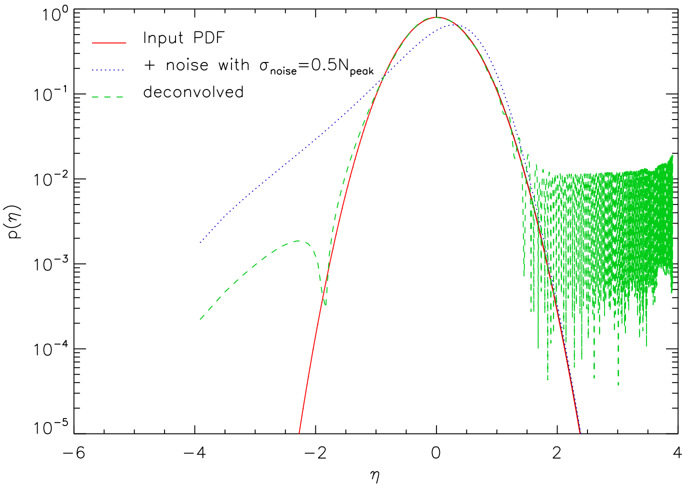

Fig. 10. Result of the numerical deconvolution of a log-normal PDF contaminated by Gaussian noise with $0.5N_{\rm peak}$ amplitude. The red solid line shows the input PDF before the noise contamination, the blue dotted line the contaminated PDFs, and the green dashed line the result of the deconvolution. The deconvolution artifacts at large column densities can be ignored as the measured noisy PDF already reflects the cloud PDF there.

larger columns, but the peak amplitude does not change significantly.

Using the three parameters plotted in Fig. 9 we have a full analytic description of the zero-column PDF as a function of noise level and cloud PDF width that can be used to look up the noise level for any measured $p_N(0)$ and $\sigma_{\eta,\rm cloud}$ or inversely. When using the zero-column PDF to determine $\sigma_{\rm noise}$, an iteration with the fitting program described in the previous section may be needed to correct for the deviation of the actual cloud PDF width $\sigma_{\eta,\rm cloud}$ from the measured width $\sigma_{\rm log-normal}$ in case of large noise amplitudes. Overall we can provide a strategy for determining the parameters of a cloud PDF from a noisy measurement. Power-law tails do not need any correction as they are unaffected by observational noise. If the noise level of the observation is known, the cloud parameters can be fitted from the measured values of $\sigma_{\rm noise}/N_{\rm center,meas}$ and $\sigma_{\rm log-normal,meas}$ using the procedure described in Sect. 3.2. If the noise of the observations is not known, it can be determined first from the zero-column PDF $p_N(0)$ described by log-normal distributions having the parameters given in Fig. 9.

3.4. Deconvolution

As discussed in Sect. 2.2 it is in principle also possible to recover the full original PDF from the measured PDF when knowing the noise contamination distribution. This would also allow to recover the shape of more complex PDFs than the simple log-normal PDF (with power-law tail) used in the computations above[4]. Gaussian noise is the ideal case with an analytic distribution that is well confined in Fourier space allowing to minimize the fundamental instability of the deconvolution process. Figure 10 shows one example for the numerical deconvolution of a noise-convolved PDF for a cloud with $\sigma_{\eta,\rm cloud} = 0.5$ and a noise amplitude of $\sigma_{\rm noise} = 0.5N_{\rm peak}$ sampled in bins of $0.01N_{\rm peak}$. As discussed before, the noise contamination shifts the peak of the distribution and creates a wide low column-density tail, but

[4] Russeil et al. (2013) and Schneider et al. (2015b) show some PDFs with small double-humps.

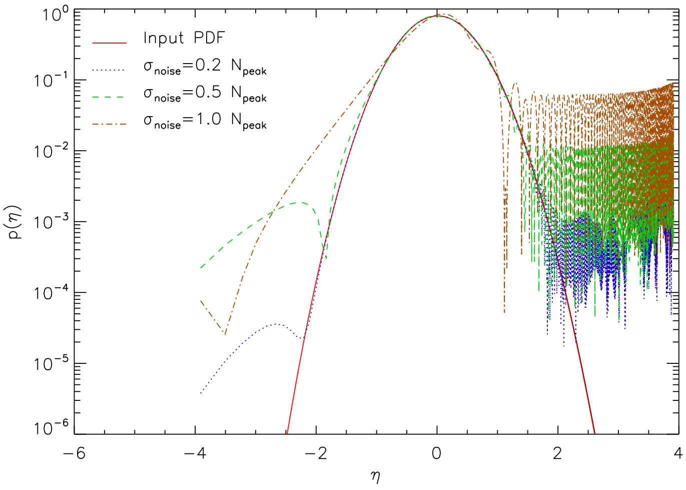

Fig. 11. Result of the numerical deconvolution of a noise-contaminated log-normal PDF for different noise levels. The red solid line shows the input PDF before the noise contamination. The contaminated PDFs are not shown here, but their shape can be taken from Fig. 3.

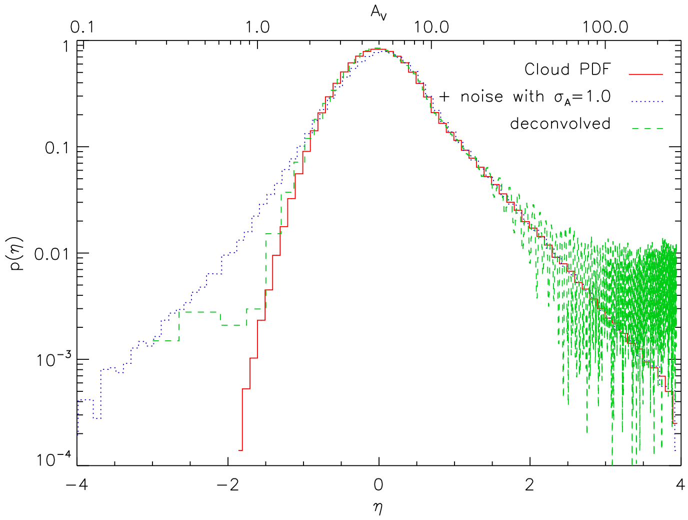

Fig. 12. Result of the numerical deconvolution of the cloud PDF from Fig. 2 for a noise level of $A_V = 1$, i.e. $0.2\,N_{peak}$. The red solid line shows the input PDF before the noise contamination.

leaves the high column-density part unchanged. The deconvolution through Eq. (7) perfectly recovers the central part of the PDF, but creates artifacts at low and high column densities. They result from numerical noise in sampling the wings of the Gaussian in the denominator of Eq. 7, i.e. the fundamental instability in the the division of two very small numbers. To see how the artifacts change with contamination we show the result of the numerical deconvolution of noise-convolved PDFs for three different noise amplitudes in Fig. 11 (see Fig. 3 for the noise-contaminated PDFs).

As the deconvolution has to be performed on a linear scale, the approach always suffers from a low sampling at small column densities. There the original PDF is well recovered for noise amplitudes smaller than the peak column density. We even have a dynamic range of almost two orders of magnitude over which the PDF can be reliably measured. Only when increasing the noise amplitude to the PDF peak column density, the deconvolution results in a too broad PDF after the deconvolution. In all cases there is, however, a floor of deconvolution artifact values at high column densities. Practically, those artifacts at large column densities are not relevant as we can always return to the measured PDF where the noise contamination only changed the PDF at low column densities. At large column densities the measured PDF still reflects the true cloud PDF (see Fig. 2).

As a more practical example, Fig. 12 shows the result of the numerical deconvolution for the full PDF from the map in Fig. 1. The noise amplitude of A_V=1 corresponds to $0.2N_{peak}$, i.e. the blue dotted curve in Fig. 11. We find a very similar behavior to the semi-analytic case. The dynamic range of the recovered PDF is a factor ten smaller than in the case of the ideal log-normal cloud representation at the same noise level, but still a factor of almost 100.

The direct deconvolution needs an accurate knowledge of noise distribution, i.e. works only for perfectly Gaussian noise. If the noise properties are well known, the approach may allow us to recover small distortions and deviations of the PDF from the log-normal shape, but to recover the four main parameters for a log-normal PDF with power-law tail the iterative approach described in Sect. 3.2 is sufficient and more stable.

4. Line-of-sight contamination

Schneider et al. (2015a) simulated the effect of a foreground contamination of a cloud PDF by a constant "screen". In a more realistic scenario, however, the contaminating structure also has turbulent properties leading to a similar PDF as that of the cloud to be observed. We assume that the contaminating cloud is also characterized by a log-normal distribution. This excludes star-forming clouds with a significant gravitationally dominated power-law tail but should represent the typical case where the contaminating structure is more transparent than the main cloud of interest, i.e. the line-of-sight emission is dominated by the studied cloud. The resulting PDF can be computed from the same convolution integral (Eq. 4) as used in the previous section[5].

Figure 13 shows the distribution of PDFs obtained from the convolution when using log-normal distributions for the cloud PDF and the contamination as a function of the contamination level. In this example both distributions have the same width of the log-normal PDF $\sigma = 0.5$. The main effect is the same as from the contamination with a constant foreground. With increasing contamination level, the peak of the PDF is shifted towards higher column densities and the width of the PDF becomes narrower. One should note that the systematic effect on the PDF peak has the same direction for noise and foreground contamination, i.e., both shift the peak artificially to higher column densities, but that the opposite is true for the standard deviation of the PDF, where noise increases the measured standard deviation, while foreground contamination tends to decrease it. The shift of the PDF peak by the line-of-sight contamination is approximately proportional to the contamination level N_{contam}.

This suggests that a correction by a constant offset, as applied for the foreground "screen" by Schneider et al. (2015a), may also work for the contamination with a wider distribution. Figure 14 shows the result after subtracting the column density of the peak of the log-normal contamination. Overall we find a

[5] Correcting for the distortions introduced by material with other PDFs, e.g. a constant gradient can be done in an equivalent way. It requires the convolution with a different shape of the contaminating PDF, $p_{contam}(N)$, in Eq. 4, repeating our quantitative analysis for the new convolution integral.

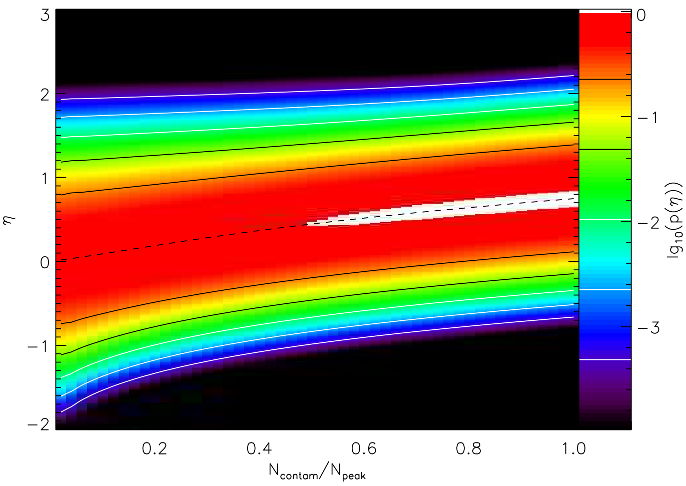

Fig. 13. Distributions obtained from the convolution of a log-normal cloud PDF (Eq. 2, $\sigma_{\eta,\rm cloud} = 0.5$) with a second log-normal PDF (Eq. 6, $\sigma_{\rm contam} = 0.5$) as a function of the typical density of the contaminating structure, $N_{\rm contam}$, relative to the peak of the log-normal cloud distribution, $N_{\rm peak}$. The left edge of the figure represents the original cloud PDF shown in colors of the logarithm of p_η. When increasing the contamination level $N_{\rm contam}$ in the right direction, we see the shift of the PDF towards higher logarithmic column densities, η.

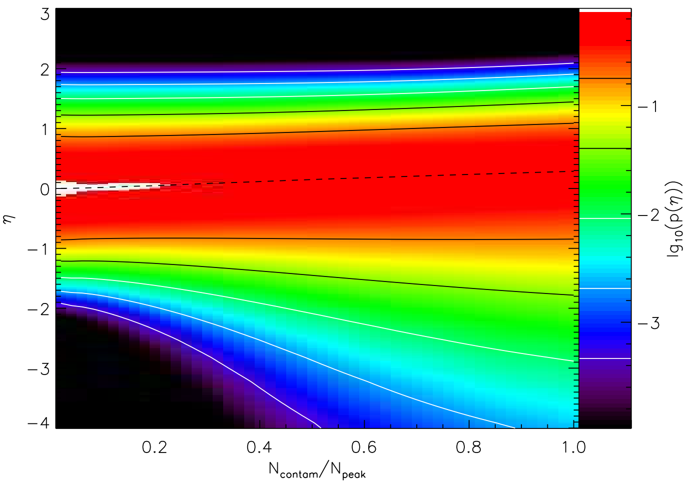

Fig. 14. PDFs of contaminated clouds corrected for the contamination by the subtraction of a constant offset of the peak of the contaminating structure PDF. The PDFs are plotted as a function of the density of the contamination, $N_{\rm contam}$, using a fixed width of the contamination PDF of $\sigma_{\rm contam} = 0.5$ matching the width of the main cloud PDF. The corrected distributions are represented through colors showing the logarithm of the PDF.

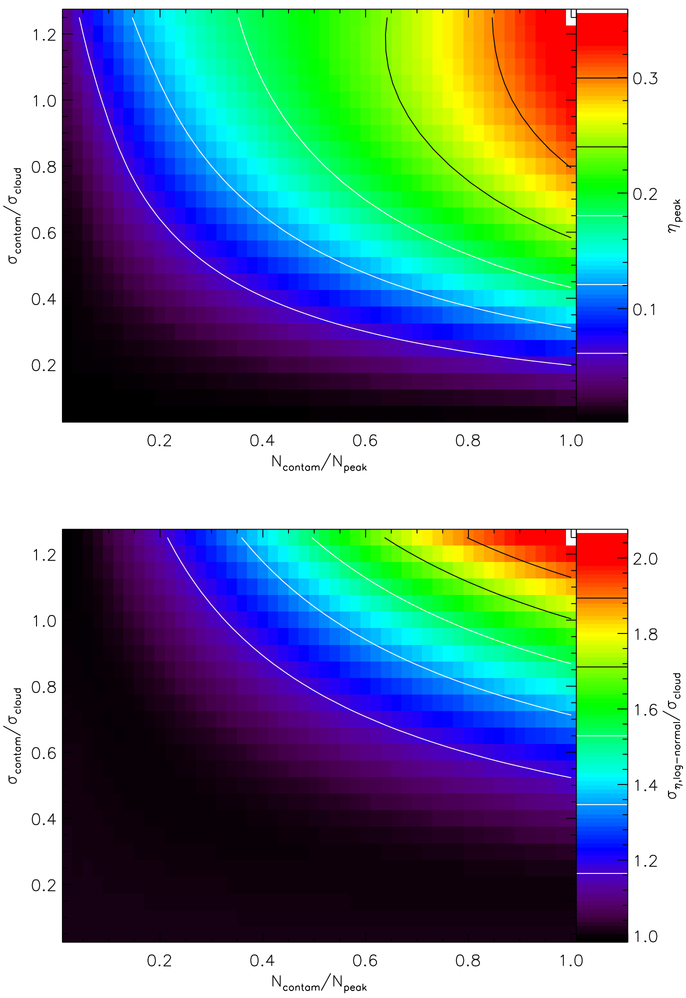

Fig. 15. Parameters of the PDFs of contaminated clouds corrected for the contamination through the subtraction of a constant offset given by the peak of the PDF of the contaminating structure. The upper plot shows the position of the PDF peak on the logarithmic η scale, i.e. a value of 0 represents the correct peak position and a value of 0.3 stands for a 35 % overestimate of the peak density. The lower plot shows the width of the corrected cloud PDF relative to the original cloud PDF. In both plots we varied the amplitude of the contamination in horizontal direction and the width of the contaminating PDF in vertical direction.

quite good reproduction of the central part of the original PDF even for contaminations that have the same amplitude as the cloud structure itself. The PDF peak position is recovered within $\Delta\eta = 0.2$. There is, however, a residual broadening of the distribution, in particular towards lower column densities, due to the "overcompensation" of contamination contributions with lower than typical column densities. The low column density wing of the corrected PDF appears too shallow relative to the original cloud PDF. One has to take into account, however, that the plot is given in logarithmic units, i.e. the deviations occur at levels of less than 1 % of the PDF peak.

To systematically quantify the residual effects after the correction, we run a parameter scan over the full range of free parameters covering the relative amplitude of the contamination relative to the cloud investigated, i.e. the relative distance of the two log-normals on the η scale, and the relative width of the PDF of the contaminating structure compared to the one of the cloud studied. This provides two dimensional plots that are shown in Fig. 15. The upper plot quantifies the accuracy of the recovery of the PDF peak, the lower plot shows the width of the corrected cloud PDF relative to the original cloud PDF. Both quantities are computed from a Gaussian fit to the corrected PDFs. The figure shows that the cloud PDF is accurately recovered when the width of the contaminating PDF is narrow ($\sigma_{\rm contamin} \lesssim 0.5\sigma_{\eta,\rm cloud}$) or its column density is small ($N_{\rm contamin} \lesssim 0.2N_{\rm peak}$). Fortunately, the first condition is usually met if the contamination stems from diffuse H I clouds, having typical PDF widths $\sigma_{\rm contam} \approx 0.2$ (Berkhuijsen & Fletcher 2008, 2015). Significant deviations occur if the properties of the contaminating structure are similar to those of the observed cloud. Then the peak column density is overestimated by 35 % and the width by a factor 1.7. If the width of the contamination PDF gets even wider, this propagates directly into the width of the corrected PDF, i.e. the recovered PDF can be more than twice as wide as the original PDF.

When knowing the properties of the contamination, we can use Fig. 15 to correct this effect in a parametric way as demonstrated above for the noise contamination. The typical level of the contamination usually can be measured by inspecting pixels at the boundaries of the cloud that are representative of the contamination only. However, one has hardly ever the same statistics for the contamination as for the whole map so that already the width of the contaminating PDF may not be known with sufficient accuracy. One last resort to infer the width is the inspection of the zero-column-density of the offset-corrected PDF, equivalent to our approach of measuring the noise in Sect. 3.3. As both effects are usually superimposed onto each other, we first have to subtract the noise contribution to the linear PDF at zero column densities using the parameters from Fig. 9 to derive the width of the contaminating PDF in a second step.

The procedure for this derivation is discussed in detail in Appx. A. It shows, however, that small errors in the determination of the amplitude of the contamination lead to large errors in the measured contamination PDF width. Hence, it is practically very difficult to use the zero-column PDF after the line-of-sight correction to estimate the width of the PDF of the contaminating structure. A direct measurement from a field tracing only the contamination seems to be always preferable if such a region can be observed.

The direct deconvolution of the contamination from the measured PDF is practically irrelevant as we can never know the properties of the contamination accurately enough to allow for the computation of the deconvolution integral (Eq. 7) without significant errors. Altogether, we find that the correction of line-of-sight contamination from a turbulent cloud can be easily corrected in terms of subtracting a constant screen if the width of the PDF of the contaminating cloud is at most half of the width of the cloud PDF to be investigated or if its column density is lower by a factor of 4-5. If both conditions are not met we can still measure the PDF peak column density with an accuracy better than 35 %, but have no good handle on the PDF width of the main cloud.

5. Boundary effects

A typical limitation of interstellar cloud observations is given by the finite map size. From multiwavelength observations it is known since long that there is no absolute definition of an interstellar cloud boundary. Depending on tracer and observational sensitivity, i.e. molecular lines, extinction, or dust continuum, the extent of a cloud appears different and can just be approximated by the lowest closed contour. For convenience - and to restrict the clouds e.g. to their molecular gas content - certain thresholds in extinction are proposed, such as $A_V = 1$ or 2 (Lada et al. 2010; Heiderman et al. 2010). However, in most cases maps are anyway only centered on prominent peaks and the total area that is mapped is limited due to observational constrains. These finite size maps obviously truncate the statistics of the cloud PDF. Schneider et al. (2015a,b) discussed the impact of incomplete sampling of the PDF by considering a truncation of the statistics at various contour levels. This only modifies the normalization of the PDF but has no effect on the measured shape. As an extreme example for this selection effect, they showed that the PDFs of Infrared Dark Clouds only consist of a power-law tail. Lombardi et al. (2015) showed that the PDF width is reduced when mapping smaller areas. In real observations, however, there is often no information about the structure outside of the mapped area.

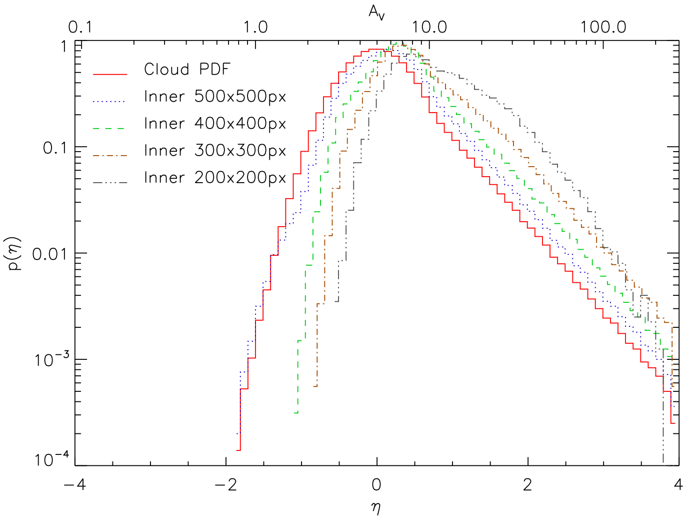

Fig. 16. Result of partial observations focusing on the center of the cloud. PDFs are measured for the four submaps of different size indicated in Fig. 1. Up to a map size of 300×300 pixels, all high-column-density spots are included in the map. A smaller map truncates some parts of the high-density center of the map leading to a noticeable distortion of the power-law tail. The submap sizes correspond to 69 %, 44 %, 25 %, and 11 % of the original map area.

In contrast such a truncation can also be intended. Studying PDFs of selected subregions within an interstellar cloud can be a valuable tool to differentiate between the physical processes determining the column density structure. In this way Schneider et al. (2012); Russeil et al. (2013); Tremblin et al. (2014) showed that in the same cloud the PDFs of quiescent subregions have a log-normal shape, denser cloud parts exhibit a power-law tail at high column densitites, and compressed shells at the interface to HII-regions provoke a second PDF peak. Sadavoy et al. (2014) and Stutz & Kainulainen (2015) linked different PDF slopes seen in different subregions of the Perseus and Orion A clouds to the content of young stellar objects and proposed an evolutionary sequence within the cloud. Hence, it is important to statistically quantify the impact of the finite size maps on the properties of the measured cloud PDF following the typical observer's approaches.

This is simulated in Fig. 16 where we study submaps of Fig. 1 with different size, placed around the central column density peak. In this simulation the spatial structure of the cloud starts to become relevant while in the previous sections the convolution integrals were actually independent of the cloud shape. Details of the resulting PDFs may depend on the exact shape of the mapped structure, but the general behavior is the same for all cases if we follow a typical observational approach. Our submaps try to mimic the observational strategy for mapping an interstellar cloud structure around the central density peaks with a finite array in a limited observing time.

As long as a sufficiently large part of the cloud is covered, the truncation at the boundaries only removes low column density gas from the statistics, leaving the power-law tail unchanged – except for the modified normalization. The removal of low-density pixels also shifts the peak of the log-normal part to larger column densities. The effect is small as long as less than half of the pixels are removed, but already when going from 600×600 pixels to 400×400 pixels (44 % of the area) the truncation of the low-density statistics becomes significant so that we significantly overestimate the peak position. In that step one can also notice a measurable reduction of the PDF width that continues

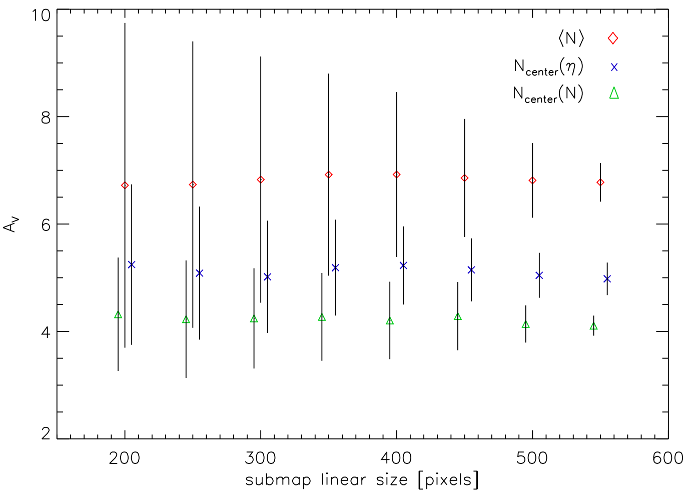

Fig. 17. Parameters of the column density distribution of randomly selected submaps of the map from Fig. 1 as a function of their size. The symbols show the mean from an ensemble of 50 submaps and the error bars represent the standard deviation. Red diamonds show the average map density, blue crosses give the PDF peak on the logarithmic scale, and green triangles the peak on the linear density scale. The symbols and error bars for the two PDF peaks are displaced by ±5 pixels relative to the actual submap size for a better visibility in the plot.

when going to smaller submap sizes. At a submap size of 11 % of the original map, we also start to loose some of the high-density structures. In this case, even the power-law tail is affected.

The impact is much stronger when considering mean quantities. This is a well know effect to most observers: increasing the mapping size around the main emission peaks tends to drastically lower the average column density as more "empty" regions are added to the statistics. For our example in Fig. 16 we find that the measured peak of the PDF in logarithmic bins $N_{\rm center}$ changes by a factor of 1.6 when reducing the field size from 600×600 to 200×200 pixels (11 % of the area), the same factor applies for the peak in linear units, but that the average column density increases by a factor of 2.3 when reducing the map size[6].

To statistically quantify this selection effect we investigate ensembles of 50 randomly selected submaps of different size and measure their average and most probable column densities. Figure 17 shows the ensemble mean and standard deviation for the average column density in the submaps and the PDF peak position $N_{\rm center}(\eta)$ on the logarithmic scale. For completeness, we also computed the center of a Gaussian PDF on the linear column density scale $N_{\rm center}(N)$. In contrast to the submap selection centered on the main peak in Fig. 16, the random selection provides for all three parameters a mean ensemble density that does not depend on the map size. The mean average density and PDF peak positions of the input map are recovered, but we find that the uncertainty of the parameters grows drastically towards smaller submap sizes. While the uncertainty of the PDF peak positions grows up to 25 % for 200×200 pixel submaps, the uncertainty of the average column density grows to 45 %.

This bigger uncertainty of the average density is to be expected as the average is more easily affected by low-probability outliers and they are more easily missed or hit in a random selection than values with high probabilities. This explains why the use of the PDF peaks for normalization purposes (see Sect. 2.1) always provides more stable results than the use of the average density.

[6] As our test cloud also includes a power-law tail of high densities, the average column density is not only 10 % higher than $N_{\rm peak}$ as expected for a log-normal distribution (see Sect. 2.1), but 44 % higher for the original map.

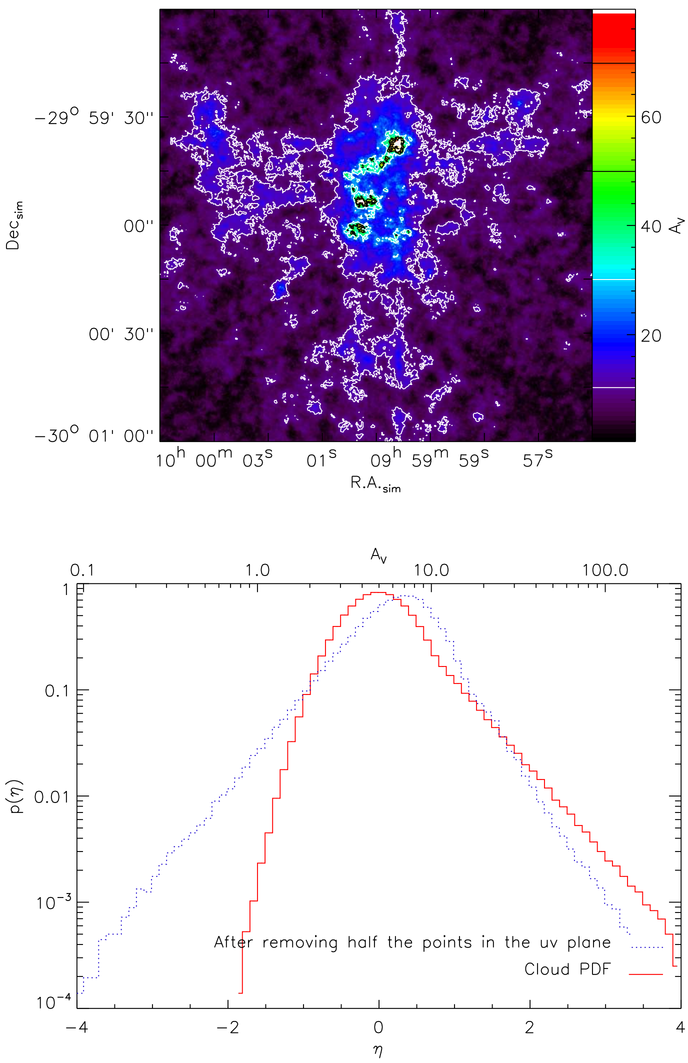

Fig. 18. Result of incomplete sampling in the uv plane. Map (upper panel) and corresponding PDF (lower panel) obtained from randomly replacing half of the points in the uv plane by zeros.

6. Interferometric observations

A special source of observational uncertainty results from the incomplete sampling of the uv plane that is unavoidable in interferometric observations. Rathborne et al. (2014) discussed the shape of the low-column density tail of a PDF obtained from combining ALMA observations with Herschel data, but they did not quantify to what degree that tail is determined by the observational limitations of the interferometric observations.

To get a first impression we simulate this effect in a very general way, without any assumptions about the actual observations, by transforming the cloud from Fig. 1 into the Fourier domain and randomly removing half of the points in Fourier space. Figure 18 shows the result after the inverse transform. The upper panel gives the map and the lower panel the corresponding PDF. While the map recovers most of the large-scale characteristic cloud structure we see already many changes in details of individual contours. The PDF shows a strong low-column density excess, very similar to the observational noise effect shown

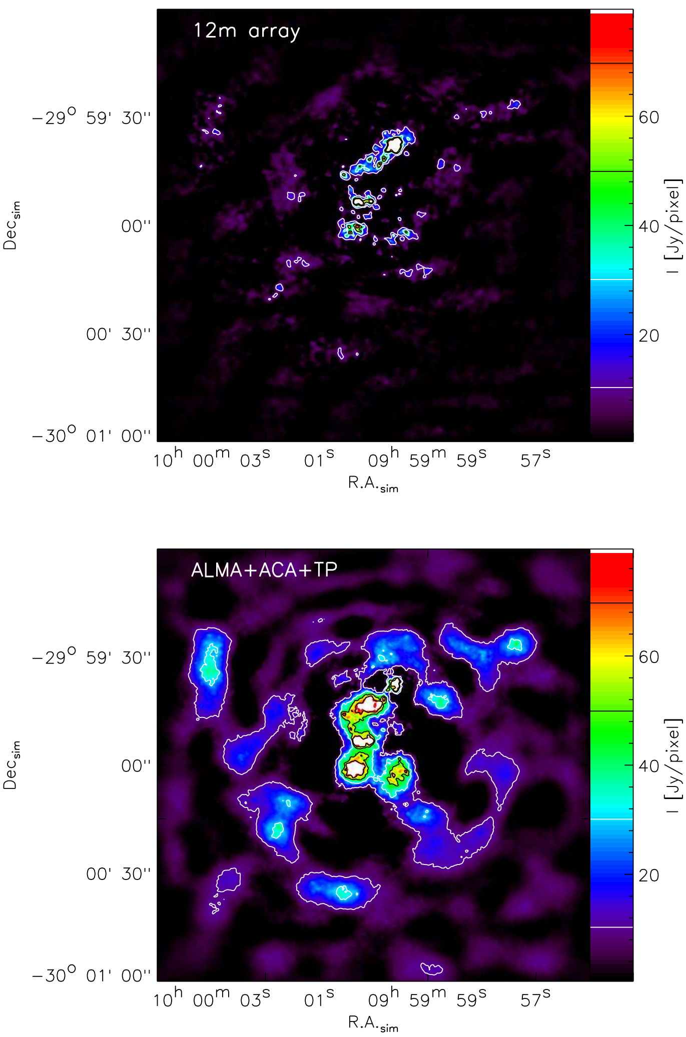

Fig. 19. Result of a simulated 2h ALMA observation. The upper plot shows the map obtained when using the 12m array without short-spacing correction. The lower plot is the map obtained after combining the 12m array data with the data from the compact array and single-dish zero spacing. The intermediate map obtained from combining only the ALMA 12m and ACA results is visually hardly distinguishable from the map in the lower plot so that it is not shown here.

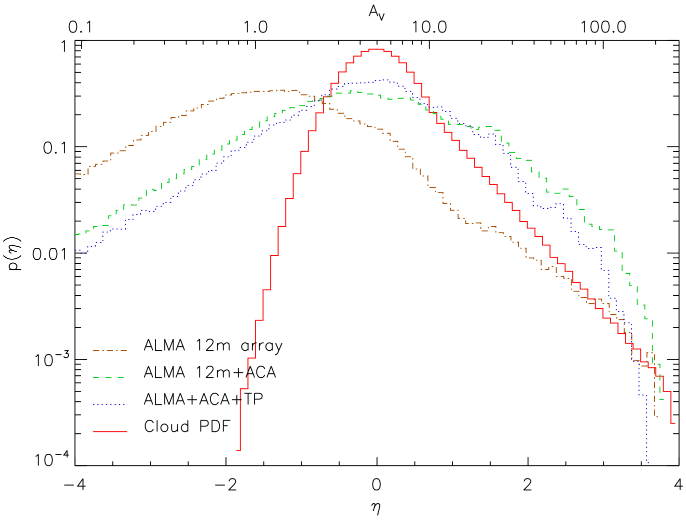

Fig. 20. Comparison of the original cloud PDF with the PDFs of the maps obtained from the 2h ALMA observations (Fig. 19).

in Fig. 2. On top of this, we also notice a change of the high-column-density statistics. Part of the highest column densities are transformed to lower values resulting in a steepening of the power-law tail and a shift of the PDF peak towards larger column densities. This makes a recovery of the underlying cloud PDF very difficult. Removing half of the points in the uv plane has a much stronger impact on the determination of the PDF than removing half of the points in normal space as demonstrated in Sect. 5.

In a more realistic scenario, we have to take into account that the coverage of the uv plane is not random, but in continuous telescope tracks given by the baselines of the interferometer and the rotation of the sky above the array. In particular with ALMA a better instantaneous uv coverage is given by the unprecedentedly large number of baselines. We simulated this case using the ALMA simulator in CASA 4.4[7] with a configuration providing a resolution of 0.6”. We used the default pipeline setup based on the standard CLEAN algorithm (Högbom 1974) that would be used by most observers. To convert our column density map into observable quantities we treated the input map as Jy/pixel units, used 230 GHz as center frequency, and a 7.5 GHz bandwidth corresponding to standard continuum observations. We used the mosaicking mode to map the total field of view of $(2')^2$, and combined the 12m array and ACA observations matching what ALMA provides for continuum observations (ALMA+ACA). The choice of these parameters should not have any major impact on our results as we were not limited by sensitivity or resolution as discussed later. To obtain a complete picture for the capabilities and limitations of the approach, we also performed the step of the single-dish total power correction (ALMA+ACA+TP). Total power zero spacing is not provided for continuum observations at ALMA, only for line observations, but including this steps provides us with a more complete picture in terms of the fundamental capabilities of interferometric observations. In a separate test, we added thermal noise to the simulations but found that it does not have any significant impact, indicating that the results are not limited by sensitivity but, as intended, by uv coverage.

Figure 19 shows the maps and Fig. 20 contains the corresponding PDFs from an observation that can be executed in a 2h observing block (2h for 12m array, 4h for ACA, 8h for TP).

The simulated observations with the 12m array only provide a reasonable reproduction of the position and shape of the high-intensity contours, but we recover basically no information about structures with an intensity below 20 Jy/pixel, i.e. only the high end of the PDF power-law tail ($\eta > 3$) is retained. After combining the data with the short-spacing information, some low-intensity structure is recovered appearing in some spiral shapes, probably produced by the shape of the uv tracks. But the shape of the high-intensity peaks is more heavily distorted than in the pure 12m array data. The intermediate map obtained when combining the data from the 12m array with the ACA data is very similar to the map obtained when including all short-spacing information.

The visual impressions from the interferometric maps are confirmed by the PDFs of the maps. When sticking to the pure interferometric data, the CLEAN algorithm manages to reproduce the high-intensity peaks above $\eta = 3$ so that the high-intensity end of the power-law tail of the PDF is recovered. At

[7] Spot tests with CASA 3.3 and CASA 4.5 showed no significant differences.

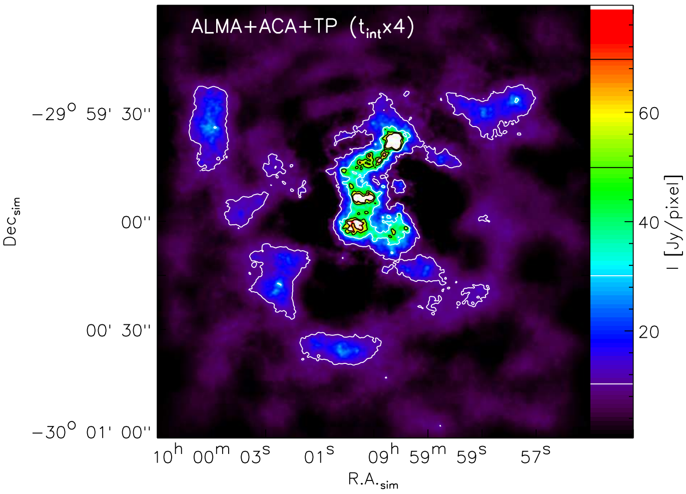

Fig. 21. Result of a simulated 8h ALMA observation. The map shows the final product from combining the 12m array data with the data from the compact array and single-dish zero spacing.

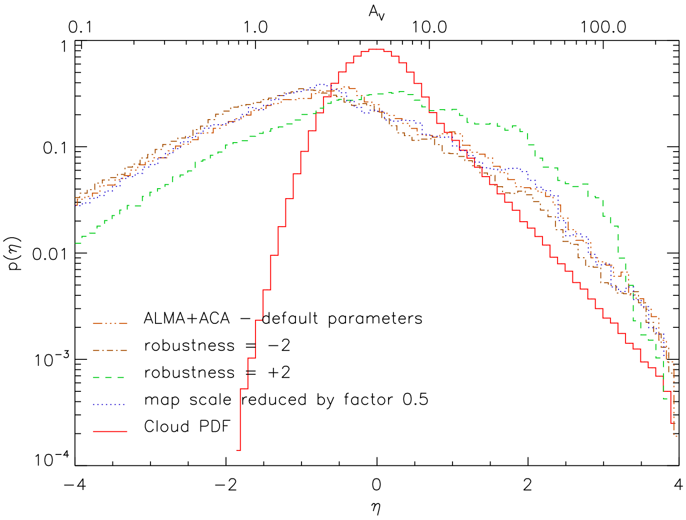

Fig. 23. Impact of different parameters on the PDFs from the ALMA+ACA maps obtained in the simulated 8h ALMA observations.

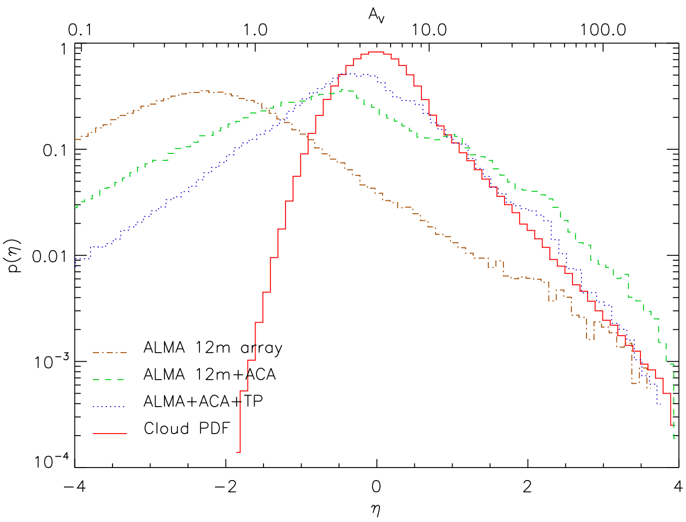

Fig. 22. PDFs of the maps obtained from the 8h ALMA observations (Fig. 21).

lower intensities, however, there is hardly any correspondence of the map PDF to the underlying cloud PDF. The typical values of $A_V \doteq 5$ Jy/pixel are strongly underabundant and the PDF peak falls at a five times lower value. In contrast, when combining the interferometric data with the short-spacing information, the high-density tail is even more distorted, but the location of the peak of the PDF is reasonably recovered. There is still a high excess at low intensities, but five times lower compared to the 12m array data.

An obvious approach to improve the mapping results is the use of longer integration times, automatically providing a better uv plane coverage due to the rotation of the earth. Figures 21 and 22 show the results when quadrupling the observing time. Here, the simulations used a total continuous observing time of 8h with the 12m array. This corresponds to 40h of total observatory time (8h for 12m array, 2 × 8h for ACA, and 4 × 8h for TP) an amount that rarely would be granted. It is, moreover, an idealized case because usually 2h observing blocks are executed randomly by the observatory, potentially leading to overlaps in the uv coverage.

When considering the map from the 12m array we do not find a big improvement compared to the shorter integration time. Again only the high intensity end ($\eta > 3$) of the cloud PDF is recovered. More low-intensity material is detected leading to an additional shift of the PDF maximum to even lower intensities. Adding short-spacing information does no longer provide the correct PDF peak position, but when including the single-dish zero-spacing, all the high-intensity PDF power-law tail is well recovered now. This is also visible in the map in Fig. 21 which shows a much better resemblance to the input map (Fig. 1), reproducing all bright structures. Here the combination of a very long integration time with all zero-spacing information allows to actually recover a significant part of the underlying cloud PDF. Unfortunately, this is currently not possible in real observations as ALMA does not provide the single-dish total power information for continuum data yet. Moreover, we can conclude that the need for a wide uv coverage is not well satisfied by single observing blocks of 2h. For the cloud statistics, it would be better to distribute the same observing time in smaller chunks over the whole visibility window of a celestial object, at least if one is not limited by sensitivity as in our example.

Even in this case the PDF still does not contain any reliable information on the log-normal part. It shows the same large excess of low intensities as in the case of the shorter integration time. The excess appears very similar to the low-intensity PDF wing reported by Rathborne et al. (2014) for their interferometric data. Comparing the outcome with Fig. 18 suggests that the CLEAN algorithm manages to better recover the properties of the high-intensity peaks than the simple Fourier transform while it performs worse for the majority of the pixels providing the central part of the PDF. This outcome is expected as CLEAN is designed to trace individual bright sources in the map while the Fourier transform should be statistically more stable.

In the combination of the data from the 12m array and ACA only, the longer integration time already leads to an improvement in the recovery of the high-intensity tail compared to the 2h ALMA observation, but the deviation of the measured PDF from the cloud PDF is still significant. One can only conclude that there must be a power-law tail. Measuring its parameters is still unreliable. To see whether one can better recover the cloud PDF based on the data provided by ALMA, we performed some additional tests varying the parameters of the CLEAN algorithm. The results are shown in Fig. 23. In two of the curves we modified the robustness parameter of the "Briggs" weighting. A value of −2 corresponds to uniform weighting, cleaning with a weighting robustness of +2 corresponds to natural weighting (see CASA User Reference & Cookbook[8]). The figure shows that natural weighting clearly deteriorates the recovery of the map PDF, there is no significant difference between the standard weighting (robustness of 0) and the uniform weighting. In a second test we used a mask to help the convergence of the CLEAN algorithm. This should provide a better data reduction strategy. However, we found no significant difference to the outcome using the default parameters so that the corresponding curve is not added in the plot, being almost identical to the brown curve. Finally, we tested the impact of the map size by changing the pixel size of our test data set by a factor of 0.5 resulting in a map size of $1'^2$. In this way we should quantify the pure resolution effect. The blue curve in Fig. 23 shows, that this has also no significant impact on the recovery of the cloud PDF. As long as we avoid the natural Briggs weighting, the combined ALMA+ACA maps are equally good or bad for measuring the cloud PDF.

A reliable partial recovery of the cloud PDF is only possible when adding single dish information in a very long integration that provides a very wide uv coverage. This asks for an extension of the current capabilities of ALMA. The sparse sampling of the uv plane unavoidably leads to strong distortions of the PDFs as a general problem of interferometric observations. None of our experiments simulating interferometric observations allowed to recover the low-intensity part of the PDF including the log-normal distribution. Therefore, the interpretation of PDFs from interferometric data should be done with extreme care.

7. Conclusions

When measuring interstellar cloud column density PDFs, the observational effects have to be carefully treated. Otherwise, all conclusions on the PDF parameters can be wrong by large factors. We provide tools to derive the cloud parameters from the measured PDFs. Our focus lies on on the log-normal part of the PDF because this part is most strongly affected by observational noise, contamination by fore- or background material, and the selection of the map boundaries. In single-dish observations the power-law tail in the PDF of self-gravitating clouds is typically only marginally influenced by the effects discussed here (see also Lombardi et al. 2015). This is different for interferometric observations.

Noise can be treated mathematically as a convolution integral. If the noise distribution is perfectly Gaussian the original PDF can be recovered through direct deconvolution. For general single-dish data, the PDF parameters can be computed from the measured PDF and a known noise level using the algorithm discussed in Sect. 3. This allows for an accurate measurement of all parameters of the log-normal part if the noise level falls below 40 % of the typical column density in the map. It even works for noise levels up to the typical column density in the map in case of clouds with a broad PDF ($\sigma_{\eta,\mathrm{cloud}} > 0.3$). For clouds with narrower PDFs and high noise contaminations, we can only retrieve the position of the PDF peak, but no longer its width. If the noise level in a map is a priori not known it can be deduced from the zero-column PDF.

Line-of-sight contamination can be corrected by subtracting a constant "screen" with the typical column density of the contamination from the measurement. This works well if the PDF of the contamination is either a factor of two narrower or a factor of four to five weaker than that of the cloud to be characterized. If width and strength of the contamination are well known, one can also retrieve the parameters of the cloud PDF through a fit to the measured parameters, similar to the noise correction. However, in the general case of an unknown broad contamination with a column density comparable to that of the cloud studied, we find a strong ambiguity between the impacts of the two PDF widths so that it turns out to be impossible to reliably derive the cloud PDF width from the measurement. We can only recover the PDF peak position of the studied cloud with an accuracy of about 35 %.

The effect of a limited field of view due to map boundaries cannot be easily corrected as we cannot "invent" information that was not measured. We found, however, that the effect is small if the observations cover at least 50 % of the structure to be measured. Smaller maps will underestimate the PDF width and overestimate the peak position. The width is slightly more stable against selection effects than the position. The sensitivity to boundary selection effects clearly discourages from the use of mean quantities to characterize the properties of a map. The characterization in terms of the PDF peak is more stable.

The situation is very different for interferometric observations with a sparse sampling in the uv plane. We found no way to reliably measure the statistics for the majority of the pixels in the map forming the center and low-density wing of the PDF. The removal of information in the uv plane has a much stronger impact than the direct removal of pixels in the map. An interferometric determination of the high-density PDF tail is, however, possible, but asks for an approach that goes beyond the current capabilities of ALMA. Combining single-dish zero spacing with a very long integration time or with many small chunks of 12m array observations over a long period, that widely populate the uv plane in principle allow to retrieve the high-density wing of a cloud PDF.

Acknowledgements. This work was supported by the German *Deutsche Forschungsgemeinschaft, DFG* via the SPP (priority programme) 1573 'Physics of the ISM'. V.O. and N.S. acknowledge supported by the DFG project number OS 177/2-2. C.F. acknowledges funding provided by the Australian Research Council's Discovery Projects (grants DP130102078 and DP150104329). RSK acknowledges support from the DFG via projects KL 1358/18-1 and KL 1358/19-2 and via the SFB 881 "The Milky Way System" (sub-projects B1, B2 and B8). In addition, RSK thanks for funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) via the ERC Advanced Grant "STARLIGHT" (project number 339177).

References

Alves, J., Lombardi, M., & Lada, C. J. 2014, A&A, 565, A18
Bensch, F., Stutzki, J., & Ossenkopf, V. 2001, A&A, 366, 636
Berkhuijsen, E. M. & Fletcher, A. 2008, MNRAS, 390, L19
Berkhuijsen, E. M. & Fletcher, A. 2015, MNRAS, 448, 2469
Brunt, C. M., Federrath, C., & Price, D. J. 2010a, MNRAS, 403, 1507
Brunt, C. M., Federrath, C., & Price, D. J. 2010b, MNRAS, 405, L56
Burkhart, B., Ossenkopf, V., Lazarian, A., & Stutzki, J. 2013, ApJ, 771, 122
Druard, C., Braine, J., Schuster, K. F., et al. 2014, A&A, 567, A118
Falgarone, E., Hily-Blant, P., & Levrier, F. 2004, Ap&SS, 292, 89
Federrath, C. & Banerjee, S. 2015, MNRAS, 448, 3297
Federrath, C. & Klessen, R. S. 2013, ApJ, 763, 51
Federrath, C., Klessen, R. S., & Schmidt, W. 2008, ApJ, 688, L79
Federrath, C., Klessen, R. S., & Schmidt, W. 2009, ApJ, 692, 364
Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81
Froebrich, D. & Rowles, J. 2010, MNRAS, 406, 1350
Ginsburg, A., Federrath, C., & Darling, J. 2013, ApJ, 779, 50
Girichidis, P., Konstandin, L., Whitworth, A. P., & Klessen, R. S. 2014, ApJ, 781, 91
Heiderman, A., Evans, II, N. J., Allen, L. E., Huard, T., & Heyer, M. 2010, ApJ, 723, 1019
Hily-Blant, P., Falgarone, E., & Pety, J. 2008, A&A, 481, 367

[8] http://casa.nrao.edu/docs/cookbook/

Högbom, J. A. 1974, A&AS, 15, 417
Hughes, A., Meidt, S. E., Schinnerer, E., et al. 2013, ApJ, 779, 44
Kainulainen, J., Beuther, H., Banerjee, R., Federrath, C., & Henning, T. 2011, A&A, 530, A64
Kainulainen, J., Beuther, H., Henning, T., & Plume, R. 2009, A&A, 508, L35
Kainulainen, J., Federrath, C., & Henning, T. 2014, Science, 344, 183
Kainulainen, J. & Tan, J. C. 2013, A&A, 549, A53
Klessen, R. S. 2000, The Astrophysical Journal, 535, 869
Konstandin, L., Girichidis, P., Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 149
Kowal, G., Lazarian, A., & Beresnyak, A. 2007, ApJ, 658, 423
Kritsuk, A. G., Norman, M. L., & Wagner, R. 2011, ApJ, 727, L20
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Lombardi, M., Alves, J., & Lada, C. J. 2011, A&A, 535, A16
Lombardi, M., Alves, J., & Lada, C. J. 2015, A&A, 576, L1
Lombardi, M., Lada, C. J., & Alves, J. 2010, A&A, 512, A67
Miesch, M. S., Scalo, J., & Bally, J. 1999, ApJ, 524, 895
Miesch, M. S. & Scalo, J. M. 1995, ApJ, 450, L27
Molina, F. Z., Glover, S. C. O., Federrath, C., & Klessen, R. S. 2012, MNRAS, 423, 2680
Nolan, C. A., Federrath, C., & Sutherland, R. S. 2015, MNRAS, 451, 1380
Nordlund, Å. K. & Padoan, P. 1999, in Interstellar Turbulence, ed. J. Franco & A. Carraminana, 218
Padoan, P., Jones, B. J. T., & Nordlund, Å. P. 1997, ApJ, 474, 730
Padoan, P. & Nordlund, Å. 2011, ApJ, 730, 40
Passot, T. & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4501
Price, D. J., Federrath, C., & Brunt, C. M. 2011, ApJ, 727, L21
Rathborne, J. M., Longmore, S. N., Jackson, J. M., et al. 2014, ApJ, 795, L25
Russeil, D., Schneider, N., Anderson, L. D., et al. 2013, A&A, 554, A42
Sadavoy, S. I., Di Francesco, J., André, P., et al. 2014, ApJ, 787, L18
Schneider, N., Csengeri, T., Hennemann, M., et al. 2012, A&A, 540, L11
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Schneider, N., Ossenkopf, V., Csengeri, T., et al. 2015a, A&A, 575, A79
Schneider, N., Csengeri, T., Klessen, R. S., et al. 2015b, A&A, 578, A29
Schneider, N., Bontemps, S., Motte, F., et al. 2015c, ArXiv e-prints [arXiv:1509.01082]
Stutz, A. M. & Kainulainen, J. 2015, A&A, 577, L6
Stutzki, J., Bensch, F., Heithausen, A., Ossenkopf, V., & Zielinsky, M. 1998, A&A, 336, 697
Tofflemire, B. M., Burkhart, B., & Lazarian, A. 2011, ApJ, 736, 60
Tremblin, P., Schneider, N., Minier, V., et al. 2014, A&A, 564, A106
Vazquez-Semadeni, E. 1994, ApJ, 423, 681

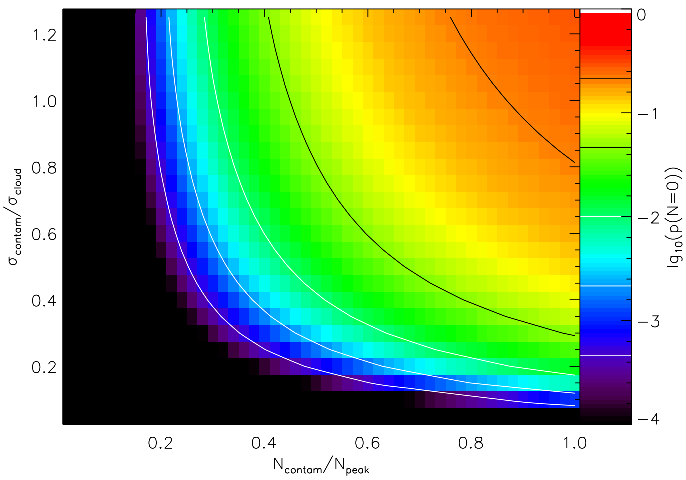

Fig. A.1. Linear column density PDF after offset correction at zero column, $p_N(0)$, as a function of the amplitude of the contamination and the width of the contaminating PDF. Colors and contours are given on a logarithmic scale.

Appendix A: Recovering the PDF width of the contaminating cloud

The typical level of the contamination usually can be measured by inspecting pixels at the boundaries of the cloud that are representative of the contamination only. However, the number of those pixels will be small compared to the whole map so that higher moments and even the width of the contaminating PDF may not be known with sufficient accuracy. However, in principle one can infer the width by inspecting the zero-column-density of the offset-corrected PDF, equivalent to our approach of measuring the noise in Sect. 3.3. The colors at the lower edge of Fig. 14 demonstrate that for high contamination levels, i.e. when the reconstructed PDF is too wide compared to the cloud PDF, it extends to low column densities including $p_N(0)$. A relation between contamination width, $\sigma_{\rm contam}/\sigma_{\eta,{\rm cloud}}$ and the zero-column PDF $p_N(0)$ then could be used to deduce the unknown contamination width.

In Fig. A.1 we show the linear column density $p_N(0)$ measured after the subtraction of the typical contamination level as a function of the relative amplitude and width of the contaminating PDF. We can accurately measure the width of the contaminating PDF from the plot if it shows a steep gradient of $p_N(0)$ in the direction of the contamination width, i.e. in the vertical cut through the plot given by the known contamination amplitude $N_{\rm contam}$. Unfortunately, we find such a steep gradient only in case of relatively narrow PDFs of the contaminating structure of less than $\sigma_{\rm contam} < 0.5\sigma_{\eta,{\rm cloud}}$, i.e. in a regime where the recovery of the cloud PDF by the constant offset correction anyway does not need any correction, but a relatively shallow behavior for broader contaminations.

Moreover, this approach depends on the accurate knowledge of the contamination level, $N_{\rm contam}$. Any "overcorrection" of the contamination will shift the PDF towards lower densities, strongly increasing the zero column density level. A small shift can lead to a significant change of the measured probability $p_N(0)$. In Fig. A.2 we show the resulting zero-column PDF when assuming a 20 % error in the amplitude of the contamination correction. When comparing the two plots in Fig. A.2 we see that the zero-column PDF can change by a factor of 100 in the sensitive range, while it still changes by a factor of five in the range of the weak dependence from the width of the contaminating PDF. Hence, it is practically very difficult to use the zero-column PDF after the line-of-sight correction to estimate the width of the PDF of the contaminating structure. A direct measurement from a field tracing only the contamination is always preferable.

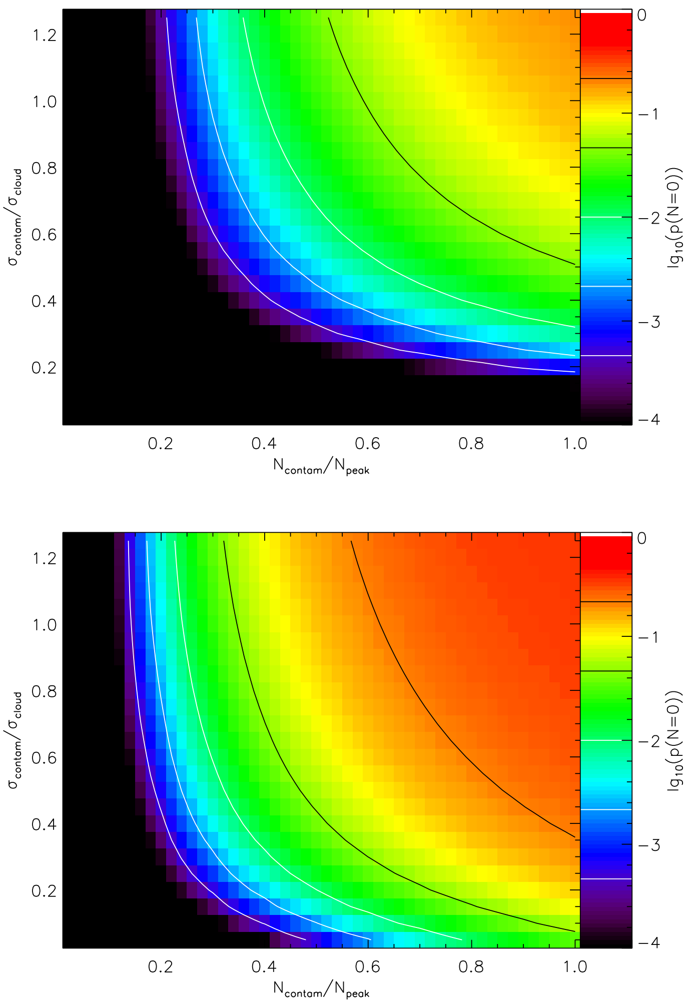

Fig. A.2. Same as Fig. A.1 but using an imperfect subtraction of the typical contamination level. In the upper plot we assume that the contamination level is underestimated by 20 %, in the lower plot that it is overestimated by 20 %.