THE PROBABILITY DISTRIBUTION FUNCTION OF COLUMN DENSITY IN MOLECULAR CLOUDS

ENRIQUE VÁZQUEZ-SEMADENI1 AND NIEVES GARCÍA2

Received 2000 December 15; accepted 2001 April 25

ABSTRACT

We discuss the probability distribution function (PDF) of column density resulting from density fields with lognormal PDFs, applicable to isothermal gas (e.g., probably molecular clouds). For magnetic and nonmagnetic numerical simulations of compressible, isothermal turbulence forced at intermediate scales (d of the box size), we find that the autocorrelation function (ACF) of the density field decays over relatively short distances compared to the simulation size. We suggest that a “decorrelation length” can be defined as the distance over which the density ACF has decayed to, for example, 10% of its zero-lag value, so that the density “events” along a line of sight can be assumed to be independent over distances larger than this, and the central limit theorem should be applicable. However, using random realizations of lognormal fields, we show that the convergence to a Gaussian is extremely slow in the high-density tail. As a consequence, the column density PDF is not expected to exhibit a unique functional shape, but to transit instead from a lognormal to a Gaussian form as the ratio η of the column length to the decorrelation length (i.e., the number of independent events in the cloud) increases. Simultaneously, the variance of the PDF decreases. For intermediate values of η, the column density PDF assumes a nearly exponential decay. For cases with a density contrast of 10^4, as found in intermediate-resolution simulations, and expected from giant molecular clouds (GMCs) to dense molecular cores, the required value of η for convergence to a Gaussian is at least a few hundred, or, for 10^6, several thousand. We then discuss the density power spectrum and the expected value of η in actual molecular clouds, concluding that they are uncertain since they may depend on several physical parameters. Observationally, our results suggest that η may be inferred from the shape and width of the column density PDF in optically thin line or extinction studies. Our results should also hold for gas with finite-extent power-law underlying density PDFs, which should be characteristic of the diffuse, nonisothermal neutral medium (with temperatures ranging from a few hundred to a few thousand degrees). Finally, we note that for η ≥ 100, the dynamic range in column density is small (less than a factor of 10), but this is only an averaging effect, with no implication on the dynamic range of the underlying density distribution.

Subject headings: ISM: clouds — ISM: molecules

1. INTRODUCTION

In recent years, several studies of the probability density function (PDF) of the density field in numerical simulations of compressible turbulent flows have been advanced as a first step in its full statistical characterization. These studies have shown that the density PDF depends on the effective polytropic exponent γ of the fluid, defined by the expression P ∝ ρ^γ, where P is the pressure and ρ is the gas density. Specifically, for isothermal flows (γ = 1) the PDF is lognormal (Vázquez-Semadeni 1994; Padoan, Nordlund, & Jones 1997; Passot & Vázquez-Semadeni 1998; Scalo et al. 1998; Ostriker, Gammie, & Stone 1999; Ostriker, Stone, & Gammie 2001), while Passot & Vázquez-Semadeni (1998) noted that for γ < 1 [γ > 1], the PDF develops a power-law tail at high [low] densities (see also Scalo et al. 1998; Nordlund & Padoan 1999; and the review by Vázquez-Semadeni et al. 2000). In addition, Gotth & Kraichnan (1993) have reported a power-law tail at high densities for Burgers flows, and Porter, Pouquet, & Woodward (1991) have reported an exponential behavior for adiabatic flows. Passot & Vázquez-Semadeni (1998) explained the lognormal PDF for isothermal flows as a consequence of the central limit theorem (CLT) acting on the distribution of the logarithm of the density field. They assumed that a given density distribution is arrived at by a succession of multiplicative density jumps, which are therefore additive in the logarithm. Since for an isothermal flow the speed of sound is spatially uniform, the density jump expected from a shock of a given strength is independent of the local density, and thus all density jumps can be assumed to follow the same distribution (determined by the distribution of Mach numbers, as studied, for example, by Smith, Mac Low, & Zuvel 2000 and Smith, Mac Low, & Heitsch 2000). Finally, at a given position in space, each density jump is independent of the previous and following ones. Therefore, the CLT, according to which the distribution of the sum of identically-distributed, independent events approaches a Gaussian, can be applied to the logarithm of the density, and the density itself is expected to possess a lognormal PDF.

However, the observationally accessible quantity is not the PDF of the mass (or “volume”) density, but rather that of the column density, i.e., the integral (or sum, for a discrete spatial grid) of the density along one spatial dimension (the “line of sight,” or LOS). Recently, Padoan et al. (2000) and Ostriker et al. (2001, hereafter OSG01) have also discussed this PDF in three-dimensional numerical simulations of isothermal compressible magnetohydrodynamic (MHD)
turbulence, with resolutions up to $256^3$ zones. In particular, OSG01 have found that the column density distribution has essentially the same shape as that of the underlying density field (a lognormal for isothermal gas), although with smaller mean and width. This result is puzzling because, according to the CLT, the PDF of column density should approach a Gaussian shape, since the column density is proportional to the mean density along the LOS. OSG01 attributed the apparent inapplicability of the CLT to the possible presence of intermediate-sized structures in the density field that invalidate the requirement of statistical independence of the individual zones needed for the CLT.

In this paper we suggest that density “events” along the LOS can be regarded as independent if they are separated by distances larger than some “decorrelation length,” over which the density autocorrelation function (ACF) has decayed by a large enough factor (we use a factor of 10). If the column length is significantly larger than this, then convergence to a Gaussian might be expected. For three-dimensional numerical simulations of magnetic and nonmagnetic isothermal turbulence forced at intermediate to large scales ($\frac{3}{4}$ of the box size), we find that the ACF drops to the 10% level at relatively short separations ($\sim 15\%$ of the box size). Using random realizations of three-dimensional lognormal fields, we show that this convergence is nevertheless very slow because of the large skewness (asymmetry) and kurtosis (wing excess) of the lognormal density PDF. Then we discuss in a speculative way the factors that may determine the shape of the density ACF in molecular clouds, and suggest that its characteristic length can be inferred observationally. In §2, we describe the numerical data we use, both from simulations of isothermal compressible turbulence and from random realizations of lognormal fields. In §3, we discuss the ACFs and PDFs of the projected density fields and, in particular, the LOS lengths required for convergence to a Gaussian. In §4, we discuss the PDF width in simulations and observations, the case of nonisothermal gas, the dependence of the correlation length on physical parameters of the turbulence, and some caveats. Finally, in §5, we summarize our results.

2. NUMERICAL DATA

We use two different sets of data for our analysis. The first comprises two numerical simulations of forced, compressible, isothermal, three-dimensional turbulence, performed at a resolution of $100^3$ grid points. One simulation is nonmagnetic, and the other is magnetic. The numerical method is pseudospectral with periodic boundary conditions, employing a combination of eighth-order hyperviscosity and second-order viscosity that allows larger turbulent inertial ranges than can be attained with second-order viscosity only. A second-order mass diffusion operator is included as well. We refer the reader to Passot, Vázquez-Semadeni, & Pouquet (1995) and Vázquez-Semadeni, Passot, & Pouquet (1996) for details. Here we just mention that for both runs the forcing rises as $k^4$ for $2 \leq k \leq 4$, and decays as $k^{-4}$ for $4 < k \leq 15$, where $k$ is the wavenumber. For the nonmagnetic run the forcing is $100\%$ compressible and has an amplitude of 25 in code units; the hyperviscosity coefficient $\nu = 8 \times 10^{-1}$, the second-order coefficient $\mu = 3.56 \times 10^{-3}$, and the mass diffusion coefficient $\mu_x$ is 0.02. The sound speed is $c = 0.2u_0$, where $u_0$ is the velocity unit. The Mach number has an rms value $\sim 1$, with maximum excursions up to $\sim 3.5$. For the magnetic run, the forcing also peaks at $k = 4$, but is $50\%$ compressible, and has an amplitude of 7.5 in code units; the diffusive coefficients are $\nu = 2 \times 10^{-1}$, $\mu = 3.5 \times 10^{-3}$, and $\mu_x = 0.03$. The sound speed is $c = 0.2u_0$, giving an rms Mach number $\sim 2.5$. A uniform magnetic field is placed initially along the $x$-direction, giving a $\beta$ parameter, defined as the ratio of the mean thermal to magnetic pressures, equal to 0.04, and an rms Alfvénic Mach number $\sim 0.5$. We have chosen this rather strongly magnetized case in order to bring out the effects of the magnetic field clearly. The differences between the two simulations are due to the fact that the magnetic simulation was not originally intended for the present study, but we do not believe this is a concern for our purposes. Our simulations are only mildly supersonic because of limitations of both the numerical scheme and the computational resources available to us, which constrain the resolution to the value mentioned above.

Since at $100^3$ points a projection along one axis gives a square of only $100^2$ points, column density PDFs for one single temporal snapshot contain only 10,000 data points, giving relatively poor statistics. We thus take advantage of the fact that the simulations are statistically stationary (although the maximum density contrast and rms Mach number do fluctuate by about 50% in time), and choose to combine several density snapshots to produce a single column density histogram. Specifically, for the nonmagnetic run we use 19 subsequent snapshots, spaced at an amount $\Delta t = 0.1$ code time units ($\sim 1.6 \times 10^{-2}$ large-scale turbulent crossing times at the rms speed). For the magnetic run we use 18 snapshots, spaced at an amount $\Delta t = 0.2$ code units ($3.2 \times 10^{-2}$ large-scale turbulent crossing times).

In order to overcome the limitations of the numerical simulations, we consider a second set of data consisting of simple realizations of random fields with lognormal PDFs, obtained by generating random numbers $X_i$ with a standard Gaussian distribution (zero mean and unit variance), and by defining a new random variable $p_i = e^{bX_i}$, where $b$ is a parameter that controls the width of the lognormal distribution. We use sequences of these “density” values to fill “cubes” (actually parallelepipeds) with fixed “plane of the sky” (POS) dimensions $\Delta x$ and $\Delta y$, and “LOS” lengths $\Delta z$ ranging from a few tens of grid cells to a few thousands.

It is important to note that we have two different sets of “samples” in this problem. One is the set of points along the LOS, whose number is given by $\Delta z$ (for simplicity, $\Delta z$ is measured in grid cells, so that it is numerically equal to the number of contributing cells). The density is effectively averaged along the LOS. The other set of samples is the lines of sight in the POS, whose number is given by the product $\Delta x \Delta y$. This equals the number of data points in the column density PDFs. We emphasize that the number of points in a PDF is completely independent of the LOS length $\Delta z$, so that we can have PDFs with the same number of data points, but with different values of $\Delta z$. Increasing the number of points in the POS allows us to improve the “signal-to-noise” ratio for the PDF, especially at the wings. However, the functional shape of the PDF is expected to depend only on the number of points in the LOS. Indeed, the column density is equivalent to the sample mean (along the LOS) in sampling theory, and it is well known that the statistics of the sample mean depend on the sample size (again, the sample along the LOS). In other words, the column density PDFs are histograms of the sample means, of which there is one for each LOS.
To improve the PDF signal-to-noise ratio, we consider many parallelepipeds (actually, 50 in all cases, each with \( \Delta x = \Delta y = 50 \)) for each set of parameters \((b, \Delta z)\), although this is exactly equivalent to having a single larger parallelepiped with 125,000 data points in the "plane of the sky," because of the statistical independence of the data, and we only keep track of the individual parallelepipeds for analogy with the procedure of combining several temporal snapshots used in the case of the numerical simulations. However, in practice, the only relevant value in this sense is the number of data points that each PDF contains, since the projected "shape" of the parallelepiped on the POS is completely irrelevant (for example, it may be a square, or a straight line). Thus, the total number of grid cells in the larger parallelepipeds (i.e., their total volume), is 125,000\( \Delta z \).

We consider two subsets of data, obtained from using two different values of \( b \), namely \( b = 1 \) and \( b = 1.5 \).

For both the simulation and the random data sets, we first normalize the lognormal density data as required by the CLT by defining a new variable \( \rho' = (\rho - \langle \rho \rangle)/\sigma_\rho \), where \( \langle \rho \rangle \) is the mean density, \( \sigma_\rho \) is the standard deviation, and \( i \) counts the pixel position along the LOS. For the random data, the mean and variance of the \( \rho' \) distribution are related to those of the Gaussian variable \( X \) by \( \langle \rho \rangle = \exp(\langle X \rangle + \frac{1}{2}\sigma_X^2) \) and \( \sigma_\rho^2 = [\exp(\sigma_X^2) - 1] \exp(2\langle X \rangle + \sigma_X^2) \) (see, e.g., Peebles 1987, Appendix F). For the simulation data, the mean density is 1, but \( \sigma_\rho \) is not known a priori, and attempting to measure it gives large errors both because of the relatively high frequency of high-density events and because it is not constant over time. We find empirically that the values of \( \sigma_\rho \) necessary to bring the column density to near unit variance (see below) are approximately 2 and 3 for the nonmagnetic and magnetic runs, respectively.

We then project (sum) the normalized density along the \( z \)-axis to obtain its associated normalized (i.e., of zero-mean and unit-variance) "column density" \( \zeta \), defined by Peebles (1987, § 4.7) as

\[
\zeta \equiv \sqrt{\frac{\sum \rho_i' / \Delta z}{\Delta z}} \tag{1}
\]

where the sum extends over all grid cells along the LOS. In the next section, we discuss the PDFs of \( \zeta \). Figure 1 shows the underlying density PDFs for the numerical simulation data (left) and for the random lognormal data (right) before normalization. The density fields are seen to be exactly lognormal in the case of the random data, and approximately so in the simulation data. The PDF of the nonmagnetic simulation exhibits an excess at small densities, but since we will be focusing mostly on the high-density side, and our main conclusions will be drawn from the random data, we do not consider this excess to be a problem. Note also that the nonmagnetic run has a wider density PDF even though it has a smaller mean Mach number than its magnetic counterpart. This is probably due to the fact that in the latter the forcing is only 50% compressible, and of smaller amplitude. The density PDFs for the random data are seen to span dynamic ranges of \( 10^4 \) and \( 10^6 \) for \( b = 1 \) and \( b = 1.5 \), respectively.

### 3. THE ACFS AND COLUMN DENSITY PDFS

Figure 2 shows, in log\( y \), linear\( x \) form, the time-integrated (i.e., adding several temporal snapshots into the same histogram) normalized column density (\( \zeta \)) PDFs for the magnetic and nonmagnetic numerical simulations. In this graph format, a Gaussian is a parabola, and an exponential is a straight line. In the two runs, a nearly exponential decay is apparent at moderately high \( \zeta \), although the very large \( \zeta \) tail clearly exhibits an excess from this trend in the nonmagnetic case and a defect in the magnetic one. This may be an effect of the lesser extent of the underlying

![Fig. 1.](image-url)
density PDF in the magnetic case. As already pointed out by OSG01, with respect to their nearly lognormal column density PDFs, these results are puzzling; one would expect the \(\zeta\)-PDF to be Gaussian, since the column density is essentially a sum (or equivalently, an average) of the density events along each LOS, the distribution of which should approach a Gaussian by virtue of the CLT. As mentioned in §1, OSG01 interpreted the deviation from a Gaussian in terms of a violation of the statistical independence requirement of the CLT, due to the existence of intermediate-size correlations in the density field.

In order to test this hypothesis, we have computed the ACF of the density field in the numerical simulations at time \(t = 2.8\) for the nonmagnetic run, and at \(t = 3.2\) for the magnetic run (\(\sim 0.45\) and \(0.51\) large-scale turbulent crossing times, respectively). These are shown in Figure 3 as a function of spatial separation (“lag”) \(r\) in grid cells. Note that we show lags only up to half the simulation size, since the periodic boundary conditions imply that the ACF is symmetric about this value. It is seen that the ACF has decreased to half its maximum (zero-lag) value at separations of only about 7 cells, and to 10% at lags of only \(\sim 14\) cells. We can effectively consider the latter to be a “decorrelation length” for the simulations. Note that the presence of the magnetic field does not seem to have an important effect on the decorrelation length. For distances significantly larger than this decorrelation length, the effects of density autocorrelation should be negligible, and the CLT should be applicable (see §4 for a discussion of possible caveats). We do not choose the more familiar \(1/e\) criterion for the decorrelation length for two reasons. First, the \(1/e\) criterion is only truly meaningful for exponential decay laws, but in general the ACF does not decay in this form, and in fact crosses zero at a finite lag in the nonmagnetic run. Second, we are interested in lags at which the ACF has become effectively negligible compared to its zero-lag value, so that events separated by this length can be assumed to be independent, and a factor of \(1/e\) seems more appropriate for this purpose than a factor of \(1/e\). For these reasons, we also have chosen to refer to this as a “decorrelation” length, rather than a correlation one. In any case, this choice is essentially arbitrary. In what follows, we denote the ratio of the column (or cloud) length to the decorrelation length by \(\eta\).

Our simulations clearly do not have a large enough number of independent events along the LOS for the column density PDF to approach a Gaussian, since \(\eta \sim 7\) in the simulation box. We have thus chosen to study this problem using simple random realizations of lognormal fields, sacrificing the realistic hydrodynamic origin of actual density data in favor of the ability to control \(\eta\) precisely, and to generate much longer LOSs than can be attained with even the largest currently available computational resources in numerical hydrodynamical simulations. This approach has been used in the past for simulating turbulent velocity fields without the numerical expense of actual hydrodynamical simulations (e.g., Dubinsky, Narayan, & Phillips 1995; Klessen 2000; Brunt & Heyer 2000). The main feature that is lost by doing this is the spatial correlation that is inherently present in actual mass density fields, because of the continuum nature of real flows. In any case, random, spatially uncorrelated fields should constitute a best-case scenario for studying the convergence to a Gaussian PDF. The presence of correlations of a certain size in grid cells should increase the required path lengths for convergence by a factor equal to this size, making convergence even slower. For the random lognormal realizations, one decorrelation length can be thought of as a single cell, so that the integration length along the line of sight, \(\Delta z\), equals \(\eta\) for the random data.
We study the convergence of the PDF to a Gaussian as a function of two parameters: the width of the underlying lognormal density PDF, given by $b$, and $\Delta z$. Figure 4 shows the PDFs of $\zeta$ for three realizations, with $\Delta z = 10$, 50, and 500 grid cells. It can be seen that at $\Delta z = 10$, the PDF of $\zeta$ appears to decay exponentially for $0 < \zeta < 4$, but develops a concavity at larger $\zeta$. At $\Delta z = 50$, the high-$\zeta$ side of the PDF is almost a perfect exponential, but at $\Delta z = 500$, no exponential segment is left, and the curve begins to approach a Gaussian.

Figure 5 shows a sequence similar to that in Figure 4, except with $b = 1.5$. In this figure we show realizations with $\Delta z = 200$, 500, 2000, and 4000. Again, a transition from concavity to convexity is seen to occur at high $\zeta$ as $\Delta z$ increases, although in this case, even at $\Delta z = 4000$ an excess is seen at the largest values of $\zeta$, so the convergence is not yet complete at this path length for $b = 1.5$. Indeed, it is known that for very asymmetric distributions with important wing excesses, the convergence to a Gaussian is fastest near the middle of the PDF, and slowest at the tails (Peebles 1987, § 4.7). We conclude that, even for completely uncorrelated data, convergence to a Gaussian occurs very slowly at the high-$\zeta$ tail if the skewness, kurtosis, and dynamic range of the underlying density data are large. In addition, we can expect that as more LOSs are included in the column density PDF, the extreme-$\zeta$ tail will reach higher $\zeta$ values, and will thus require larger lengths of $\Delta z$ to converge.

4. DISCUSSION

4.1. What is the Value of $\eta$ in Real Molecular Clouds?

The convergence studied in the last section refers to completely uncorrelated random data, so that the correlation length is effectively one grid cell. As mentioned in § 3, the presence of a finite decorrelation length in the density data should cause the convergence to be even slower with path length, since sufficiently independent “events” are expected to be separated by lags of the order of the decorrelation length. Note that if the decorrelation length is a sizeable fraction of the column (cloud) size, i.e., if $\eta$ is not very large, then full convergence to a Gaussian is not expected.

Clearly, in the case of real molecular clouds the concept of “grid cell” disappears, and the natural unit for measuring the path length should be the decorrelation length itself. Thus, the ratio $\eta$ serves as a measure of the path length. In our simulations, $\eta \sim 7$, and according to the results of § 3, this is too small a length to produce a full convergence to a Gaussian column density PDF even at moderate underlying density contrasts. However, in real molecular clouds the actual value of $\eta$ is essentially unknown, and convergence to a Gaussian column density PDF is plausible if $\eta$ is large enough. Thus, it becomes important to assess this possibility.

The value of $\eta$ must be related to some characteristic scale of the density-fluctuation power spectrum. If the spectrum has a self-similar (power-law) dependence on wavenumber over some range, then there are no characteristic scales in this range, and the only natural characteristic scales are those where the power-law range ends at high and low wavenumbers (analogous to the “inner” and “outer” scales of the turbulent energy spectrum). In order to investigate this dependence, we use a spectrum-modifying algorithm introduced by Lazarian et al. (2001), which allows us to modify the spectrum of any physical field without modifying its spatial distribution. This amounts to only changing the “contrast” of the physical field (Armi &

---

We thank Thierry Passot for pointing this out.
Since the power spectrum depends only on the Fourier amplitudes of a field, but not on its Fourier phases, the modification is accomplished by Fourier transforming the physical field, and then replacing the Fourier amplitudes by others that satisfy the desired spectral shape, without altering the phases. We refer the reader to Lazarian et al. (2001) for details of the algorithm.

We apply the spectrum-modifying algorithm to the density field of the nonmagnetic simulation, and impose on it the two spectra shown in Figure 6a. In both cases, the spectrum rises as $k^3$ for $k \leq k_p$ and then decreases as $k^{-3}$ for $k > k_p$, where $k$ is the wavenumber in units of the inverse box length, and $k_p$ is a “peak” wavenumber. One case has $k_p = 3$ (solid line) and the other has $k_p = 7$ (dotted line). These spectra produce the density ACFs shown in Figure 6b. It is seen that the 10% level of the ACF occurs at a lag $r \sim 1/(3k_p)$, with suggesting that the decorrelation length is related to the “outer scale” of the density power spectrum.

What determines the shape of the density fluctuation power spectrum, and, in consequence, of the density ACF in highly compressible turbulence is, to our knowledge, an open problem. It should most likely be related to the energy spectrum and the forcing (energy injection) spectrum, but the actual forms of all these spectra in molecular clouds are unknown. For example, the temporally and spatially intermittent energy injection in molecular clouds from embedded stellar sources and passing shocks differs significantly from the standard random forcing scheme used in most numerical simulations, which is applied everywhere in space and continuously in time (see, e.g., Norman & Ferrara 1996 and Avila-Reese & Vázquez-Semadeni 2001 for related discussions at larger scales in the ISM). In particular, if the energy is injected at small scales, the energy spectrum in a cloud complex may peak at scales quite smaller than the size of the complex, and possibly drive the density field to a similar spectral shape. Moreover, note that if the energy spectrum is dominated by shocks, then its form ($k^{-2}$) is of geometrical, rather than dynamical, origin (see, e.g., Vázquez-Semadeni et al. 2000), and in this case the density power spectrum need not have the same outer scale as the energy spectrum. Self-gravity may be an important ingredient, too. In summary, the actual shapes of all the relevant spectra in molecular clouds, and thus the value of $k_p$, remain unknown, and deserve to be studied systematically.

Observationally, several workers have looked at correlation lengths in molecular gas. In a pioneering study, Kleiner & Dickman (1984) investigated the ACF of column density in the Taurus region, and from their plots one infers a correlation length of a few pc. This is not too short a distance compared to the size of the complex, but note, however, that this correlation length refers to the projected intensity data rather than to the underlying three-dimensional density field. Most other observational correlation studies have focused on the ACF of the line velocity centroid distribution, and are not directly applicable to our purposes. In any case, they have either reported correlation lengths of fractions of a parsec (e.g., Scalo 1984; Kleiner & Dickman 1985) or else find them difficult to determine unambiguously (e.g., Miesch & Bally 1995).

In this respect, our results suggest that the column density PDF provides us with a means of observationally measuring the ratio of the cloud size to the decorrelation length $\eta$ when optically thin transitions or extinction data are used: the observed column density PDF should transit from a lognormal to an exponential and then on to a Gaussian as $\eta$ increases. Unfortunately, we do not know the path length a priori, but if it can be estimated by some other means, at least in some cases, then the decorrelation length, $r_D = 1/(3k_p)$, is a good estimator of the actual decorrelation length.

We thank E. Ostriker for noting this point.

---

**Fig. 6.** Left: Two power spectra imposed on the density field of the nonmagnetic simulation for studying their effect on the decorrelation length. In both cases the spectrum rises as $k^3$ until $k_p$ and then decreases as $k^{-3}$, where $k$ is the wavenumber in units of the inverse box length. Solid line: $k_p = 3$. Dotted line: $k_p = 7$. Right: Resulting density autocorrelation functions for the power spectra shown in the left panel. The line type matches that of the corresponding power spectrum.
and consequently the density power spectrum outer scale, can be derived. This suggests that it is necessary to investigate numerically how the decorrelation length depends on parameters of the flow, such as the forcing parameters, self-gravity, the energy and magnetic spectra, etc.

4.2. The Width of the Column Density PDF

Another implication of the results from §3 is that, at large \( \eta \), the column density dynamic range becomes small. Figure 7 shows the PDFs of the mean density (i.e., the unnormalized column density divided by the path length) for all LOSs for the two sets of random density fields. It is seen that while the underlying density PDFs discussed here have density contrasts of up to \( 10^6 \), the column density PDFs typically have dynamic ranges of at most a factor of 20, and, for very large \( \eta \), of factors of only a few. This is actually a trivial result, since in the limit \( \eta \to \infty \), all LOSs would give exactly the same column density (i.e., the sample mean asymptotically approaches the distribution mean), and the column density PDF would collapse to a Dirac delta function, independently of the dynamic range of the underlying density distribution. This suggests that if \( \eta \) is large in actual clouds, then nearly constant column densities are expected, but this tells little about the dynamic range of the actual density field. In this case, the density-size relation of Larson (1981), \( \rho \sim R^{-1} \), which implies constant column density, could simply be an observational averaging effect along the LOS (J. Scalo 2000, private communication). On the other hand, a relatively large observational column density range would point toward relatively small values of \( \eta \).

Observational studies of extinction (proportional to column density; e.g., Lada et al. 1994; Kramer et al. 1998; Cambresy 1999) typically report extinction dynamic ranges of about a factor of 10. Comparing with the mean-density PDFs of Figure 7, these ranges are consistent with \( \eta \sim 10 \) and 100 for underlying density ranges of \( 10^4 \) and \( 10^6 \), respectively. For comparison, the column densities reported by Padoan et al. (2000) from numerical simulations of MHD turbulence at a resolution of \( 128^3 \), with underlying density fields with a dynamic range of \( 10^6 \), span 3 orders of magnitude, suggesting that in actual molecular clouds, \( \eta \) may be significantly larger than in those simulations. On the other hand, OSG01 have compared the column density cumulative distribution from their simulations with that of visual extinction in cloud IC 5146 (Lada, Alves, & Lada 1999), finding that the overall curve width in both data sets is comparable. Unfortunately, in order to determine whether similar PDF widths imply similar values of \( \eta \), it is necessary to know whether the dynamic ranges of the underlying density distributions are also comparable. Moreover, because of the poorly sampled nature of the observational data, OSG01 had to present the distributions in cumulative form, and in linear plots, rather than semilogarithmic. In this format, it would be difficult to distinguish, for example, between lognormal and exponential PDFs that have similar values at moderate column densities. In any case, the results are promising, and indicate that properties of molecular cloud turbulence such as the decorrelation length can indeed be determined from the column density PDF and the dynamic range of the density field.

Finally, note also that our results imply that there should exist a relationship between the functional shape of the column density PDF and its width, i.e., between its skewness and its variance. We plan to quantify this relation in future work.

4.3. The Case of Nonisothermal Gas

In this paper we have restricted the analysis to lognormal underlying density PDFs, in part for simplicity and in part...
in order to relate our results on PDFs to those from
numerical simulations of compressible isothermal MHD
turbulence (e.g., OSG01; Padoan et al. 2000). Isothermal
flows are normally considered as representative of the flow
within molecular clouds. However, it is possible that mole-
cular clouds are really only close to being isothermal in the
density range $10^3 \leq n/cm^{-3} \leq 10^4$ (see the discussion by
Scalo et al. 1998). Moreover, diffuse gas in the ISM, either
neutral or ionized, is in general nonisothermal, and in this
case, if the flow behaves approximately barotropically ($P \propto \rho^\gamma$, $\gamma \neq 1$), a power-law range is expected to appear in
the PDF (Passot & Vázquez-Semadeni 1998). In this case
the CLT does not necessarily apply. Indeed, let us consider
of
The CLT to apply (J. Scalo 2000, private communication).
possible that the flow is not sufficiently decorrelated for the
PDF (Passot & Vázquez-Semadeni 1998). In this case
the CLT does not necessarily apply. Indeed, let us consider
a power-law range of the form $f(\rho) = C\rho^{-\alpha}$, where $C$ is a
constant. If the range extends to arbitrarily large and/or
small values, the variance does not exist, and therefore the
CLT does not apply. If the power law is truncated at low
densities, and $\alpha > 1$, then the column density PDF becomes
a gamma distribution (Adams & Fatuzzo 1996). However, if
the power-law range has a finite extension, and beyond it
the PDF drops rapidly, such as the PDFs reported by Scalo
et al. (1998) for nonisothermal numerical simulations of the
ISM, and by Passot & Vázquez-Semadeni (1998) for poly-
tropic flows with $\gamma \neq 1$, then the variance should still exist
and the CLT should apply. We expect this to be the case for
observational PDFs of diffuse gas.

4.4. Caveats

Although the results of this paper are rather straightfor-
ward, a number of possible complications should be kept in
mind. First, it is possible that the ACF fails to capture
long-range correlations because the short-range ones may
mask them, since small-scale structures are generally much
denser. So, even in cases where the 10% level of the ACF is
reached over lags much smaller than the cloud size, it is
possible that the flow is not sufficiently decorrelated for the
CLT to apply (J. Scalo 2000, private communication). Numerical simulations of turbulent flows with large values of $\eta$ are necessary to test for this possibility. Second, in cases in which the Jeans length is close to the system size, self-
gravity may promote the formation of large-scale structures,
countering the possible action of small-scale energy injection sources, and tending to reduce the value of $\eta$. In this case, column densities closer to logarithmic shapes, and with rather large variances, might be expected. High-
resolution numerical experiments with self-gravity and real-
istic stellar-like forcing, even if just in two dimensions,
similar to those of Passot et al. (1995) or of Vázquez-
Semadeni, Ballesteros-Paredes, & Rodríguez (1999), but
with cooling functions appropriate for molecular clouds,
may help resolve this issue.

Finally, we have suggested that a small column density
dynamic range should be taken as an indication of large
values of $\eta$. Unfortunately, small column density dynamic
ranges may also arise from limitations in the sensitivity of the
observations and saturation effects. Thus, the best
suited observations for testing the above results are those in
which these limitations are minimized.

5. SUMMARY

Our results can be summarized as follows:

1. We have proposed that the relevant parameter deter-
miming the form of the column density PDF in molecular
clouds is the ratio $\eta$ of the cloud size to the decorrelation
length of the density field, with the latter operationally
defined in this paper as the lag at which the density autoco-
relation function (ACF) has decayed to its 10% level. Assuming
that density "events" along the LOS are essen-
tially uncorrelated for $\eta \geq 1$, large values of this ratio imply
that the central limit theorem (CLT) can be applied to those
events, and a Gaussian PDF should be expected for large
enough values of $\eta$. This parameter is essentially the number
of independent events (the "sample size") along the LOS,
and the column density is equivalent to the "sample mean"
along the LOS.

2. We have measured $\eta$ in two three-dimensional nume-
rical simulations of isothermal turbulence forced at interme-
diate scales, one magnetic and one nonmagnetic. In both
cases we find $\eta \sim 7$, suggesting that at least partial con-
vergence to a Gaussian PDF should occur. The column
density PDFs for both runs are approximately exponential.

3. Using simple random realizations of uncorrelated,
lognormally-distributed fields, we have shown that the
PDF of the normalized (i.e., with zero mean and unit
variance) column density $\zeta$ indeed converges to a Gaussian
shape as $\eta$ increases, as dictated by the CLT, albeit very
slowly, because of the large dynamic range, skewness, and
kurtosis of the density lognormal distribution. For cases in
which the underlying data have a dynamic range ("density
contrast") of $\sim 10^4$, convergence to a Gaussian requires an
$\eta$ of several hundred. For density dynamic ranges of $\sim 10^6$,
the required sample size is several thousand events. In addi-
tion, the width (variance) of the column density PDF also
decreases as $\eta$ increases, as expected for the distribution of the
"sample mean". Specifically, for $\eta \sim 10$ and an under-
lying density dynamic range of $10^6$, the column density
dynamic range is $\sim 20$, and has decreased to a factor of a
few for an $\eta$ of a few hundred.

4. We have discussed the turbulent parameters that deter-
mine $\eta$. Using a spectrum-modifying algorithm, we have
shown that the 10%-level decorrelation length appears to
be given approximately by $1/(3k_p)$, where $k_p$ is the wave-
number at which the density fluctuation power spectrum
peaks. Thus, the decorrelation length appears to be very
close to the "outer scale" of the density power spectrum.
However, we believe that what determines the shape of the
density spectrum in molecular cloud turbulence is still
an open problem requiring much further work.

5. We have suggested that the slow convergence of the
column density PDF, which transits from lognormal (or a
power-law, if the underlying gas behaves polytropically, but
is not isothermal), to exponential and on to nearly Gaussian
shapes as $\eta$ increases, can be used to observationally deter-
mine the latter in molecular clouds. This would provide a
direct observational diagnostic of this fundamental prop-
erty of the turbulence in molecular clouds. In addition,
since the variance of the column density PDF decreases
with increasing $\eta$, a functional relationship between the
variance and skewness of the PDF is expected to exist.

6. The decrease of the PDF variance with increasing $\eta$
suggests that if $\eta$ turns out to be large in real molecular
clouds, the density-size relation of Larson (1981), which
implies roughly constant column density, could be simply a
result of this averaging along the LOS (J. Scalo 2000,
private communication). Conversely, wide, skewed PDFs
may be an indication that the clouds are not very large
compared to the turbulent density decorrelation length, and
Larson’s relation might then be the result of limited observational dynamic range (Kegel 1989; Scalo 1990; Vázquez-Semadeni et al. 1997).

7. We also discussed briefly the case of power-law underlying density PDFs, expected when the gas is not isothermal. In this case, the CLT is only expected to apply if the power laws are truncated at both low and high densities, although the convergence to a Gaussian may be even slower if the power-law range is very extended, since power laws have even higher tails than a lognormal distribution.

8. Finally, we have mentioned several possible caveats, specifically: (1) the possibility that the large-scale correlations are masked in the density ACF because they involve lower-density structures, (2) the fact that self-gravity may possibly increase the decorrelation length, and (3) the fact that sensitivity and saturation problems with the observations limiting their dynamic range may incorrectly be taken to mean large values of $\eta$.

We thank Laurent Cambresy for discussing his data with us, and Eve Ostriker, Thierry Passot, Luis Rodriguez, and John Scalo for sharp comments and/or a careful reading of the manuscript. In particular, John Scalo provided us with important comments about statistical distributions, limitations of the various statistical methods, and interesting implications of this work. Remarks from Eve Ostriker helped us uncover a misconception in an earlier version of this paper. The turbulence simulations were performed on the Cray Y-MP 4/64 of DGSCA, UNAM. This work has made extensive use of NASA’s Astrophysics Data System Abstract Service, and received partial funding from Conacyt grant 27752-E to E. V.-S.

REFERENCES

Adams, F. C., & Fatuzzo, M. 1996, ApJ, 464, 256
Armi, L., & Flament, P. 1985, J. Geophys. Res., 90, 11779
Avila-Reese, V., & Vázquez-Semadeni, E. 2001, ApJ, 553, 645
Brunt, C., & Heyer, M. H. 2000, ApJ, in press (preprint astro-ph/0011200)
Cambresy, L. 1999, A&A, 345, 965
Dubinsky, J., Narayan, R., & Phillips, T. G. 1995, ApJ, 448, 226
Gotoh, T., & Kraichnan, R. H. 1993, Phys. Fluids A, 5, 445
Kegel, W. H. 1989, A&A, 225, 517
Kleiner, S. C., & Dickman, R. L. 1984, ApJ, 286, 255
——. 1985, ApJ, 295, 466
Klessen, R. S. 2000, ApJ, 535, 869
Kramer, C., Alves, J., Lada, C. J., Lada, E. A., Sievers, A., Ungerechts, H., & Walmsley, M. 1998, A&A, 329, L33
Lada, C. J., Alves, J., & Lada, E. A. 1999, ApJ, 512, 250
Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, ApJ, 429, 694
Larson, R. B. 1981, MNRAS, 194, 809
Lazarian, A., Pogosyan, D., Vázquez-Semadeni, E., & Pichardo, B. 2001, ApJ, 555, 130
Miesch, M. S., & Bally, J. 1994, ApJ, 429, 645
Nordlund, Å., & Padoan, P. 1999, in Interstellar Turbulence, ed. J. Franco & A. Carramiñana (Cambridge: Cambridge Univ. Press), 218
Norman, C. A., & Ferrara, A. 1996, ApJ, 467, 280
Ostriker, E. C., Gammie, C. F., & Stone, J. M. 1999, ApJ, 513, 259
Ostriker, E. C., Stone, J. M., & Gammie, C. F. 2001, ApJ, 546, 980 (OSG01)

Padoan, P., Juvela, M., Bally, J., & Nordlund, Å. 2000, ApJ, 529, 259
Padoan, P., Nordlund, Å., & Jones, B. J. T. 1997, MNRAS, 288, 145
Passot, T., & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4301
Passot, T., Vázquez-Semadeni, E., & Pouquet, A. 1995, ApJ, 455, 536
Peebles, P. Z., Jr. 1987, Probability, Random Variables, and Random Signal Principles, 2d ed. (New York: McGraw-Hill)
Porter, D. H., Pouquet, A., & Woodward, P. R. 1991, in Large-Scale Structures in Hydrodynamics and Theoretical Physics, ed. J. D. Fournier & P. L. Sulem (Berlin: Springer), 105
Scalo, J. 1984, ApJ, 277, 556
Scalo, J. 1990, in Physical Processes in Fragmentation and Star Formation, ed. R. Capuzzo-Dolcetta, C. Chiosi, & A. Di Fazio (Dordrecht: Kluwer), 151
Scalo, J., Vázquez-Semadeni, E., Chappell, D., & Passot, T. 1998, ApJ, 504, 835
Smith, M. D., Mac Low, M.-M., & Heitsch, F. 2000, A&A, 362, 333
Smith, M. D., Mac Low, M.-M., & Zuev, J. M. 2000, A&A, 356, 287
Vázquez-Semadeni, E. 1994, ApJ, 423, 681
Vázquez-Semadeni, E., Ballesteros-Paredes, J., & Rodriguez, L. F. 1997, ApJ, 474, 292
Vázquez-Semadeni, E., Ostriker, E. C., Passot, T., Gammie, C., & Stone, J. 2000, in Protostars & Planets IV, ed. V. Mannings, A. Boss & S. Russell (Tucson: Univ. Arizona Press), 3
Vázquez-Semadeni, E., Passot, T., & Pouquet, A. 1996, ApJ, 473, 881