**Astronomy
&
Astrophysics**

Mapping the interstellar dust with near-infrared observations:
An optimized multi-band technique

M. Lombardi[1] and J. Alves[2]

[1] Institüt für Astrophysik und Extraterrestrische Forschung, Auf dem Hügel 71, 53121 Bonn, Germany
[2] European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 Garching bei München, Germany

Abstract. We generalize the technique of Lada et al. (1994) to map dust column density through a molecular cloud (NICE) to an optimized multi-band technique (NICER) that can be applied to any multi-band survey of molecular clouds. We present a first application to a $\sim$625 deg^2 subset of the Two Micron All Sky Survey (2MASS) data and show that when compared to NICE, the optimized NICER technique (i) achieves the same extinction peak values, (ii) improves the noise variance of the map by a factor of 2 and (iii) is able to reach $3\,\sigma$ dust extinction measurements as low as $A_V \simeq 0.5$ magnitudes, better than or equivalent to classical optical star count techniques and below the threshold column density for the formation of CO, the brightest H$_2$ tracer in radio-spectroscopy techniques. The application of the NICER techniques to near-infrared data obtained with a 8 meter-class telescope with a state-of-the-art NIR camera, such as the VLT-ISAAC combination, will be able to achieve dynamic ranges from below 10^{21} protons cm^{-2} to over 10^{23} protons cm^{-2} (A_V in the range $[0.3, 60]$) and spatial resolutions $<10''$, making use of a single and straightforward column density tracer, extinction by interstellar dust.

Key words. ISM: clouds – ISM: dust, extinction – ISM: structure – ISM: individual objects: Orion molecular complex, Mon R2 molecular complex – methods: data analysis

1. Introduction

Dark clouds remain one of the least understood objects in the Universe. Although they are ubiquitous in galaxies like the Milky Way, we still do not have a basic understanding of how they relate to the more diffuse interstellar medium (ISM), of how some of them form stars and planets, or how they vanish. We do know today that they are the coldest objects known ($T \simeq 10$ K), which unfortunately implies that they are also one of the hardest to observe and study.

Before the discovery of emission from molecules in dark clouds (Wilson et al. 1970, after which these clouds became known as molecular clouds), the general technique used to study the density structure of these objects relied on statistical analysis of numbers of dust extincted stars in the background of a cloud. This approach, introduced by Wolf (1923), is known as the star count method and was later refined and intensively applied by Bok (1937, 1956). In this technique, the main mass component of a molecular cloud, H$_2$, is traced by the dust in the cloud, an assumption later proved to be valid by observations. The main feat of the star count method is its straightforward essence and its sensitivity to low column density regions with A_V

in the range $[0.5, 3]$ magnitudes, i.e., from column densities between $\sim 1 \times 10^{21}$ and $\sim 8 \times 10^{21}$ protons cm^{-2} (Rossano 1978; Dickman 1978; Mattila 1986; Gregorio Hetem et al. 1988; see Arce & Goodman 1999 for a multi-technique approach).

Almost everything we know today about molecular cloud structure has been derived from radio observations of molecular tracers (e.g., CO, CS, NH$_3$) of the undetectable H$_2$ (Lada 1996; Blitz & Williams 1999; Myers 1999) and in more recent years through dust thermal emission that can be detected by radio continuum techniques (Andre et al. 2000; Johnstone et al. 2000). Radio spectroscopy and continuum techniques are able to probe deeper into molecular clouds, well past column densities of 10^{22} protons cm^{-2}, an order of magnitude improvement over classical star count techniques. Moreover, radio-spectroscopy offers an unique view of the dynamical structure of molecular clouds (e.g., Rosolowsky et al. 1999), something that none of the absorption or emission continuum techniques can, evidently, assess. However, the interpretation of results derived from radio observations is not always straightforward (e.g. Alves et al. 1999; Chandler & Richer 2000). Several poorly constrained effects inherent to these techniques (e.g., deviations from local thermodynamic equilibrium, opacity variations, chemical

evolution, small-scale structure, depletion of molecules, unknown emissivity properties of the dust, unknown dust temperature) make the construction of an unambiguous picture of the physical structure of these objects a very difficult task. The price of being able to probe deeper into a molecular cloud with radio line and continuum techniques is a complicated data analysis, not free from ambiguities.

It has long been recognized that infrared color excess can be used to obtain a reliable estimate of the extinction through a molecular cloud (see, e.g., Bok 1977; Jones et al. 1980; Frerking et al. 1982; Jones et al. 1984; Casali 1986). The deployment of sensitive, large format infrared array cameras on large telescopes, however, has allowed the direct measurement of line-of-sight dust extinction toward thousands of individual background stars observed through a molecular cloud. Such measurements are free from the complications that plague molecular-line or dust emission data and enable detailed maps of cloud density structure to be constructed. In all fairness, all density tracer methods rely on the assumption of an universal gas-to-dust ratio, a fair assumption on theoretical grounds that still remains to be demonstrated for cold dense molecular clouds. Lada et al. (1994) pioneered a technique, the Near-Infrared Color Excess (NICE) technique, for measuring and mapping the distribution of dust through a molecular cloud using data obtained in large-scale, multiwavelength, infrared imaging surveys. This method combines measurements of near-infrared color excess to directly measure extinctions (as opposed to the statistical measurement in the star count method) and map the dust column density distribution through a cloud. Moreover, the measurements can be made at significantly higher angular resolutions and substantially greater optical depths than previously thought possible. The efficacy of this technique was demonstrated with the study of the dark cloud L 977 (Alves et al. 1998), IC 5146 (Lada et al. 1999), and Barnard 68 (Alves et al. 2001) where through straightforward analysis of nearly 7000 infrared sources background to these clouds produced detailed maps of the extinction to optical depths and spatial resolution an order of magnitude higher than previously possible (A_V in the range $[1, 40]$ magnitudes, spatial resolution down to 10 arcsec).

1.1. The large view on molecular clouds

Nearby galactic molecular cloud complexes represent our best chance to understand molecular cloud structure. These complexes have relatively large sizes (from a few to tens of parsec) and are hence fairly extended, typically stretching several to tens of degrees across the sky. A large view of entire molecular cloud complexes, with sufficient resolution and dynamic range, is the only way to put these clouds in their Galactic context and address the questions on their origin and fate. Census of entire molecular cloud complexes are rare (CO line emission: Dame et al. 2001; Fukui et al. 1999; IRAS opacity

maps: Wood & Myers 1995; optical star counts: Cambrésy 1999), have typical dynamical ranges of 10^{21} protons cm^{-2} to below or about 10^{22} protons cm^{-2}, but suffer from the caveats inherent to the respective used techniques (see above). There is a clear need for a new large scale mapping technique of molecular cloud complexes, that makes use of a more reliable column density tracer, and is able to offer an extended dynamical range.

In this paper we generalize Lada's NICE technique to an optimized multi-band technique, the Near-Infrared Color Excess Revisited technique (NICER). Although inspired by the Two Micron All Sky Survey (2MASS) (Kleinmann et al. 1994), and by the possibility of all sky dust extinction mapping, this technique can be applied to any multi-band survey of molecular clouds (e.g., DENIS, Epchtein et al. 1997; SDSS, York et al. 2000, or any combination of multi-wavelength catalogs). We present a first application to 2MASS data, and show that NICER improves the noise variance of a map by a factor of 2 when compared to NICE, and is able to reach $3\,\sigma$ dust extinction measurements of $A_V \simeq 0.5$ magnitudes, better than or equivalent to optical star count techniques and below the threshold column density for the formation of CO under typical temperatures and densities (van Dishoeck & Black 1988; van Dishoeck 1992). This latter point is highly relevant as CO is the brightest H_2 tracer in radio-spectroscopy techniques. The application of NICER to near-infrared (NIR) data obtained with an 8 meter-class telescope outfitted with a state-of-the-art NIR camera, e.g. the VLT-ISAAC combination, will be able to achieve column density dynamic ranges from below 10^{21} protons cm^{-2} to over 10^{23} protons cm^{-2} (A_V in the range $[0.3, 60]$ magnitudes), on spatial resolutions $<10''$, making use of a single and straightforward column density tracer, extinction by dust.

The paper is organized as follows: in Sect. 2 we review the NICE technique and discuss its limitations, in Sect. 3 we give a detailed presentation of NICER, in Sect. 4 we apply both techniques to the same 2MASS data subset, and present the conclusions in Sect. 5.

2. Near-infrared color excess

The difference between the observed and the intrinsic color of a background star provides information on the optical depth of the cloud. For example, for the near-infrared and assuming a *normal reddening law* (Rieke & Lebofsky 1985), we have

$$(H - K)^{\text{obs}} = (H - K)^{\text{tr}} + 0.063 \cdot A_V, \qquad (1)$$

where H and K represent the H and K-band (centered at 1.65 μm and 2.2 μm, respectively) the superscripts "obs" and "tr" denote observed and true (intrinsic) quantities, respectively. The visual extinction A_V is often used as a measure of dust column density and will be the main target of our investigations. This quantity is relevant because, as suggested by several studies (Lilley 1955;

Jenkins & Savage 1974; Bohlin et al. 1978), it is directly related to the projected mass density of the cloud. In particular, it has been shown that most clouds have the same gas-to-dust ratio, so that $N(H_2) = A_V \cdot 1.9 \times 10^{21}$ protons cm^{-2} mag^{-1} or, equivalently for the projected two-dimensional density, $\kappa = A_V \cdot 15 \ M_\odot$ pc^{-2}.

Equation (1) would not be particularly useful if the intrinsic star colors (which are generally unknown) would span a large range of values. In reality, infrared colors of stars have a relatively small scatter (e.g., $\sigma(H - K)^{\mathrm{tr}} \simeq 0.08$), so that it is sensible to take the average of Eq. (1), obtaining

$$A_V = 15.87 \cdot \left[\left\langle (H - K)^{\mathrm{obs}} \right\rangle - \left\langle (H - K)^{\mathrm{tr}} \right\rangle \right]. \qquad (2)$$

It is not difficult to estimate $\left\langle (H - K)^{\mathrm{tr}} \right\rangle$ using, for example, a control field where the extinction is negligible. Hence, in principle, we can use Eq. (2) with $\left\langle (H - K)^{\mathrm{obs}} \right\rangle$ replaced by the observed color of each *background* star and obtain an *unbiased* estimate for the cloud column density in the direction of the star. In practice, this method suffers from two main limitations: (i) The use of the observed color of each star in lieu of $\left\langle (H - K)^{\mathrm{obs}} \right\rangle$ introduces a source of noise in the estimate of A_V; (ii) It may *not be easy* to establish if a given star is a background object, i.e. is farther than the cloud. Traditionally, both limitations are dealt with by using two simple techniques. The noise on the measured column density, which is ultimately introduced by the intrinsic scatter on the colors of stars, is reduced by averaging angularly close objects (which, hopefully, are subject to the same amount of extinction; see however below). Thus, normally the extinction is estimated using the equation

$$\hat{A}_V = 15.87 \left[\frac{1}{N} \sum_{n=1}^{N} (H - K)_i^{\mathrm{obs}} - \left\langle (H - K)^{\mathrm{tr}} \right\rangle \right], \qquad (3)$$

where the sum is performed on a set of N angularly close stars. This is the approach taken by the NICE method, originally described by Lada et al. (1994). Regarding the second problem, foreground stars can be recognized from their color, which is significantly bluer than angularly nearby stars (often the number of foreground stars is much smaller than the number of background stars).

It is not difficult to recognize that the two problems discussed above are intimately related. Suppose, for example, that we are investigating a cloud with low column density. Then, because of the intrinsic scatter of star colors and of photometric errors, we could take a foreground star as a background object just because its is observed redder than expected. The situation can be even more difficult for embedded stars. Clearly, a *perfect* discrimination between foreground and background objects is feasible only in the unrealistic case where the column density can be measured *without any error*. This is not possible, but at least we can try to take care of both problems in a uniform way. This is the point of view adopted by our optimized method for extinction estimate.
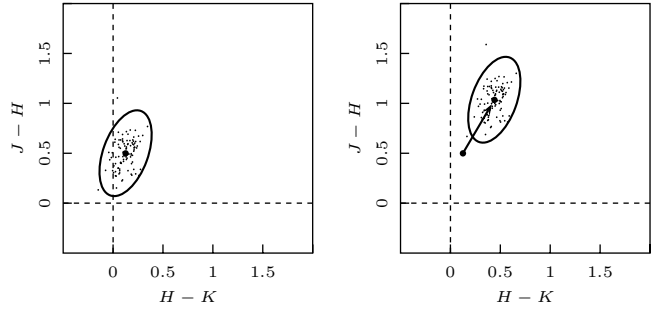
Fig. 1. JHK color-color of one hundred stars (sketch). The left plot shows the colors of stars which are not subject to extinction. The ellipse encloses a 3σ confidence region; note that the ellipse is almost exactly vertical oriented, i.e. $J - H$ correlates only weakly with $H - K$. The right plot shows the same stars as observed through a cloud with $A_V = 5$ magnitudes (a normal reddening law is assumed).

3. The optimized method

In this section we present NICER, the Near-Infrared Color Excess method Revisited, an optimized, multi-band method for extinction estimation based on the near infrared excess of background starlight. The method has been designed by considering separately the two steps needed to make an extinction map:

– Local extinction estimate for each star;
– Spatial smoothing of individual stars estimates.

We describe these two steps in Sects. 3.1 and 3.2.

3.1. Local extinction

As discussed in Sect. 2, the cloud column density A_V is normally measured by comparing the $H - K$ color for stars observed through the cloud with the same color for stars for which no reddening is expected (see Eq. (3)). Clearly, other possibilities are also viable. For instance, we could choose the $J - H$ color and use the expression

$$\hat{A}_V = 9.35 \left[\frac{1}{N} \sum_{n=1}^{N} (J - H)_i^{\mathrm{obs}} - \left\langle (J - H)^{\mathrm{tr}} \right\rangle \right]. \qquad (4)$$

Should we use Eq. (3) or (4) to infer the cloud column density? In both cases, we expect some error on the estimate of A_V. The final error on A_V depends critically on three factors:

1. The scatter of the intrinsic colors: normally, stars present a larger scatter in $J - H$ than in $H - K$ (see Fig. 1);
2. The photometric error on the individual bands: typically errors on J are smaller than errors on K (except along lines-of-sight through dense clouds, where extinction makes stars fainter in J);
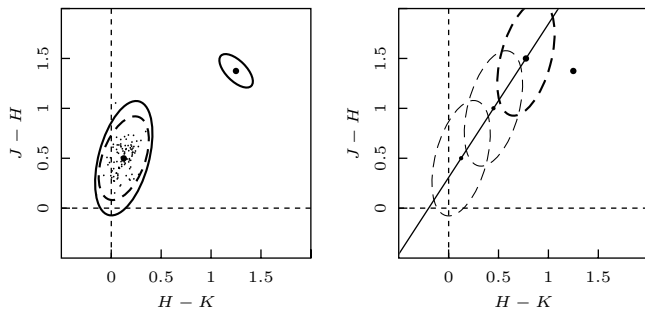
Fig. 2. For each star, NICER obtains an estimate of the extinction A_V using a maximum-likelihood technique. The star colors and the relative covariance (represented in the figure in the left panel by the small ellipse to the top right) are obtained from the observations. Note that correlation of colors is expected (see Eq. (10)). The net effect of photometric errors is a widening (solid ellipse) of the intrinsic color scatter (dashed ellipse) (Eq. (11) and following discussion). To estimate the extinction to the star NICER than finds the ellipse center along the reddening line (see figure to the right) that is closer to the observed colors of the star (Eq. (12)).

3. The numerical coefficient used to convert color excess into A_V extinction: this coefficient is larger for Eq. (3) (15.87 vs. 9.35).

Hence, on the one hand we are encouraged to use Eq. (3) because of the smaller scatter in colors; on the other hand we want to avoid this estimator because of the larger numerical coefficient and larger photometric errors. In conclusion, this Heuristic discussion does not lead to a definitive answer regarding the choice between the two estimators (3) and (4). This suggests that a more quantitative approach is needed to study the problem and decide which estimator should be used. Actually, the analysis we are going to carry out will provide a much more interesting result and will show that a clever use of all available infrared bands produces significantly more accurate results than the individual estimators (3) or (4). [In the following we will consider the typical case of three bands J, H, and K available; the generalization to more bands is trivial.]

With the three bands at our disposal we can make two independent colors, $c_1 = J - H$ and $c_2 = H - K$. For each c_i (with $i = 1, 2$), we can write the relationship between the observed color c_i^{obs} and the true one c_i^{tr} (i.e., the color that would be observed if no extinction were present) as

$$c_i^{\mathrm{obs}} = c_i^{\mathrm{tr}} + k_i A_V + \epsilon_i. \tag{5}$$

Here $k_i = E_i/A_V$ is the ratio between the color excess on the band i and the extinction on the V band; for the colors considered here we have $k_1 = 1/9.35$ and $k_2 = 1/15.87$. The extra term ϵ_i above represents the noise on the colors, i.e. the result of photometric error on the estimate of $J-H$ and $H-K$. Let us restrict the discussion to estimators of A_V which are linear in the observed colors c_i^{obs}, i.e. of the form

$$\hat{A}_V = a + b_1 c_1^{\mathrm{obs}} + b_2 c_2^{\mathrm{obs}}. \tag{6}$$

The coefficients a, b_1, and b_2 need to be determined so as to satisfy two conditions (see Fig. 2):

1. The estimator is *unbiased*, i.e. its expected value is the true extinction A_V;
2. The estimator has *minimum variance*.

The first condition is equivalent to the equations

$$b_1 k_1 + b_2 k_2 = 1, \qquad a + b_1 \langle c_1^{\mathrm{tr}} \rangle + b_2 \langle c_2^{\mathrm{tr}} \rangle = 0. \tag{7}$$

The variance of the estimator $\hat{A}_V$ can be evaluated from the expression

$$\hat{A}_V - \langle \hat{A}_V \rangle = \sum_i b_i \big(c_i^{\mathrm{tr}} - \langle c_i^{\mathrm{tr}} \rangle \big) + \sum_i b_i \epsilon_i. \tag{8}$$

As a result, we can immediately write

$$\mathrm{Var}\big(\hat{A}_V\big) = \sum_{i,j} b_i b_j \mathrm{Cov}_{ij}(c^{\mathrm{tr}}) + \sum_{i,j} b_i b_j \mathrm{Cov}_{ij}(\epsilon). \tag{9}$$

The covariance matrices introduced in this equation represent two different sources of noise. The first matrix, $\mathrm{Cov}_{ij}(c^{\mathrm{tr}})$ is the intrinsic scatter of star colors, which clearly makes the determination of the extinction A_V uncertain (for example, if a star is redder than expected we will overestimate A_V). This matrix, obviously independent of the star considered, can easily be determined by using a control field, where the extinction can be neglected, and measuring the scatters of star colors.

The second matrix, $\mathrm{Cov}_{ij}(\epsilon)$, is related to photometric errors, and thus changes for each star. This matrix can be easily calculated provided that an estimate of the photometric errors for each star is available. In the typical case of the three bands J, H, and K considered here, we can write

$$\mathrm{Cov}_{ij}(\epsilon) = \begin{pmatrix} \sigma_J^2 + \sigma_H^2 & -\sigma_H^2 \\ -\sigma_H^2 & \sigma_H^2 + \sigma_K^2 \end{pmatrix}. \tag{10}$$

Note that, while the photometric errors on different bands (namely, σ_J, σ_H, and σ_K) can be taken to be uncorrelated, errors on colors are not.

We can now minimize Eq. (9) with the requirement that Eq. (7) holds using Lagrange's multipliers. This way we reduce our problem to the solution (b_1, b_2) of the linear system

$$\begin{pmatrix} C_{11} & C_{12} & -k_1 \\ C_{21} & C_{22} & -k_2 \\ -k_1 & -k_2 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \tag{11}$$

where the matrix $C_{ij} = \mathrm{Cov}_{ij}(c^{\mathrm{tr}}) + \mathrm{Cov}_{ij}(\epsilon)$ is the sum of the two covariances. The parameter λ in the previous equation is Lagrange's multiplier and, for our purposes, its value is uninteresting. The solution of this system is then, in matrix notation,

$$\boldsymbol{b} = \frac{C^{-1} \cdot \boldsymbol{k}}{\boldsymbol{k} \cdot C^{-1} \cdot \boldsymbol{k}}, \tag{12}$$

where $\boldsymbol{b} = (b_1, b_2)$ and $\boldsymbol{k} = (k_1, k_2)$. The final expression for the optimal estimator is thus

$$\hat{A}_V = b_1 \big[c_1^{\mathrm{obs}} - \langle c_1^{\mathrm{tr}} \rangle \big] + b_2 \big[c_2^{\mathrm{obs}} - \langle c_2^{\mathrm{tr}} \rangle \big], \tag{13}$$

with $\boldsymbol{b}$ given by Eq. (12). Its variance is given by Eq. (9), or equivalently $\mathrm{Var}(\hat{A}_V) = \boldsymbol{b} \cdot C \cdot \boldsymbol{b}$.

A nice feature of this technique is that it can be applied without significant modifications in the case where one of the bands is not observed. Suppose, for example, that the J band is not available. Then we can assume an arbitrary value for this band and set the corresponding error σ_J to an extremely large value. If we "blindly" apply Eqs. (12) and (13), we obtain a matrix C^{-1} with only a single non-vanishing element at the position $(2,2)$. In other words, the use of a large error on J automatically suppresses the use of this band in the evaluation of $\hat{A}_V$. Interestingly, the same technique can be used if the missing band is H, which contributes to both colors $c_1 = J - H$ and $c_2 = H - K$: in this case the estimator will be composed only of the combination $c_1 + c_2 = J - K$, thus avoiding the use of the missing H band.

3.2. Spatial smoothing

So far we have considered only a single star. In order to obtain a smooth extinction map with high signal-to-noise ratio, NICER applies a *spatial smoothing* to angularly close stars. The smoothing is performed using three different techniques, described below in individual subsections.

The choice of the smoothing technique is important for two main reasons: (i) The final map has a signal-to-noise ratio that strongly depends on the smoothing used; (ii) The smoothing is responsible for the selection of background stars. As already pointed out, NICER tries to deal with both problems at the same time.

3.2.1. Weighted mean

The simplest way to obtain a smooth mass map from individual estimates for A_V is to use a weighted mean. More precisely, we use the values of $\hat{A}_V$ obtained for stars close to a direction $\boldsymbol{\theta}$ on the sky to estimate the extinction on $\boldsymbol{\theta}$. Calling $\boldsymbol{\theta}^{(n)}$ the position on the sky of the nth star, $\hat{A}_V^{(n)}$ its estimated extinction, and $\mathrm{Var}(\hat{A}_V^{(n)})$ the corresponding estimated variance, we write

$$\hat{A}_V(\boldsymbol{\theta}) = \frac{\sum_{n=1}^{N} W^{(n)} \hat{A}_V^{(n)}}{\sum_{n=1}^{N} W^{(n)}}, \qquad (14)$$

where $W^{(n)}$ is the weight for the nth star, given by

$$W^{(n)} = \frac{W\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(n)}\right)}{\mathrm{Var}\left(\hat{A}_V^{(n)}\right)}. \qquad (15)$$

Note that, in contrast to Eq. (3), here the index n runs over *all* N observed stars in the field. The weight $W^{(n)}$ is the combination of two different factors: (i) a spatial term, described by a *weight function* W; (ii) an error weight, proportional to $1/\mathrm{Var}(\hat{A}_V)$. The weight function is usually chosen to be a Gaussian with appropriate width. We stress that the choice of the characteristic length of the weight function is a fundamental step in the reconstruction process, as shown by a detailed statistical study of the statistical properties of the smoothed map. This has already been done in a different context (Lombardi & Bertin 1998; see also Lombardi & Schneider 2001), and thus we just state the main results obtained. The size of the weight function directly determines the *correlation length* of the final extinction map, or, equivalently, sets the scale of the smallest details we can hope to observe on the map (every feature smaller than the characteristic size of the weight function will be washed out). Moreover, the larger the width of the weighting function, the larger the *effective* number of stars used for each point on the map, and the smaller the error on the extinction. In other words, we can decide whether we want a high signal-to-noise ratio map with low resolution, or a more noisy map with higher resolution. A local estimate for the error $\sigma_{\hat{A}_V}$ is given by

$$\sigma_{\hat{A}_V}^2(\boldsymbol{\theta}) = \frac{\sum_{n=1}^{N} \left(W^{(n)}\right)^2 \mathrm{Var}\left(A_V^{(n)}\right)}{\sum_{n=1}^{N} \left(W^{(n)}\right)^2}, \qquad (16)$$

where $\mathrm{Var}(A_V^{(n)})$ is the expected variance for the nth star (see note after Eq. (13)). Finally, we note that the choice (15) for the coefficients of the stars maximizes the signal-to-noise ratio.

So far we have assumed that all stars are background objects with respect to the cloud. In reality, some of them could be in the foreground and could thus contaminate our extinction estimate. If the number of foreground stars is not too large, we can ignore them and proceed *as if* all stars were in the background. Because of the linearity of the estimator (14), we can correct for the dilution of the signal a posteriori by multiplying the map $\hat{A}_V(\boldsymbol{\theta})$ by the factor N/N_{back}, where N_{back} is the estimated number of background stars. Actually, this simple technique is often not very efficient and can also introduce biases for dense clouds (see below Sect. 3.2.4). For this reasons, it might be preferable to use other smoothing techniques described below.

3.2.2. Sigma-clipping

Foreground stars do not contribute to the signal, still contribute to the noise. Hence, it is desirable to exclude in the average (14) those stars that are suspected to be in the foreground because of their low extinction. This can be easily obtained by using a sigma-clipped or by a median estimator (described below in Sect. 3.2.3).

In order to implement sigma-clipping, we make a first estimate of A_V using Eq. (14) and calculate the expected variance from Eq. (16). Then we repeat the entire process using only those stars for which $\hat{A}_V^{(n)}$ does not differ by more than a factor $3\sigma_{\hat{A}_V}$ from the first estimate of A_V evaluated at $\boldsymbol{\theta}^{(n)}$, i.e. $\hat{A}_V(\boldsymbol{\theta}^{(n)})$. This procedure quickly converges to a pair of values, the measured extinction

$\hat{A}_V(\boldsymbol{\theta})$ and the related error, which we take as fiducial values.

In order to verify the goodness of our procedure, we can evaluate the *observed* variance on the data:

$$\hat{\sigma}^2_{\hat{A}_V}(\boldsymbol{\theta}) = \frac{\sum_{n=1}^{N} \left(W^{(n)}\right)^2 \left(\hat{A}_V^{(n)} - \hat{A}_V\right)^2}{\sum_{n=1}^{N} \left[W^{(n)}\right]^2}. \tag{17}$$

Note that this variance is defined in a different way with respect to Eq. (16). Assuming that the photometric errors on the various bands have been correctly evaluated, we expect $\hat{\sigma}_{\hat{A}_V} \geq \sigma_{\hat{A}_V}$, the two variances being equal only if all stars are subject to the same extinction, i.e. if there are no foreground stars and the cloud is perfectly homogeneous. Actually, the difference between $\hat{\sigma}_{\hat{A}_V}$ and $\sigma_{\hat{A}_V}$ can be used to obtain precious information on the substructure of the molecular cloud (see, e.g. Juvela 1998).

3.2.3. Weighted median

Alternatively, we can use a *weighted median* in order to make the extinction map. This quantity is defined as the solution $\hat{A}(\boldsymbol{\theta})$ of the non-linear equation

$$\sum_{n=0}^{N} W^{(n)} \mathrm{sgn}\left(\hat{A}(\boldsymbol{\theta}) - \hat{A}^{(n)}\right) = 0, \tag{18}$$

where sgn is the sign function ($\mathrm{sgn}\,x = \{+1, 0, -1\}$ depending on the sign of x). Since in general the condition (18) cannot be satisfied exactly, a linear interpolation is used (see Lombardi et al. 2000 for a similar example of application of a weighted median).

The use of the median has several advantages with respect to the simple mean. Probably, the most interesting is the fact that the median is a *robust* estimator, since it is basically unaffected by outliers. This property is particularly useful in our case, since it provides an efficient method to remove foreground objects from our map. On the other hand, the median has the significant disadvantage of having noise properties difficult to study. We also note that the implementation of the median is not as efficient as the mean, but this should be considered a minor problem.

3.2.4. Discussion

In the previous subsections we have described three simple smoothing techniques used by NICER. An important point to note is that there is not really a single *best method* to perform the spatial smoothing. In some conditions, several subtle problems, in fact, may make each technique described above inappropriate, noisy, or even biased.

As already noted by Thoraval et al. (1997), the simple or weighted mean (as described in Sect. 3.2.1) can be severely biased in high-extinction regions. In fact, because of extinction, the density of background stars decreases as the cloud column density increases, while the density of foreground stars is left unchanged. As a result, in the dense regions foreground stars represent a larger fraction of the local number of stars with respect to less dense regions. This way, we could under-estimate the column density in the central regions of clouds. Such a bias can generally be ignored for very nearby clouds. In other cases, the bias might be reduced using the median or the sigma-clipping methods.

On the other hand, substructure on the scale of the weighting function can also be a problem. In case of the simple weighted mean, the smoothing procedure is straightforward to understand (at least at the simple level kept in the discussion here; see however Lombardi & Schneider 2001): the final map obtained is expected to be the original, true extinction of the cloud convolved with the weight function used. In the case of sigma-clipping or median smoothing, things are much more complicated. Suppose, for example, that the cloud does not present any significant substructure on the scale of the weighting function, except for small "holes" with very low absorption. In this case, we would identify stars observed through these holes as foreground objects, although they are in reality in the background. As a result, we would completely miss the holes, and thus overestimate the cloud column density. Similarly, for a cloud with substructures consisting of dense globules, we would underestimate the column density (note that both sigma-clipping and median are "symmetric," in the sense that they discard outliers with both large or small A_V).

In summary, each smoothing technique has advantages and disadvantages (this is the ultimate reason to include all of them in NICER). A key role is played by the fraction of foreground stars (and thus the distance of the cloud), and by the substructure of the cloud. It is in any case always advisable to use all smoothing techniques and to check the consistency of the maps obtained.

4. Nicer at work: The Orion region from 2MASS data

The Two Micron All Sky Survey (2MASS) offers a unique opportunity to test our algorithm. Conceived about a decade ago (Kleinmann et al. 1994), the 2MASS project is an all-sky-survey obtained from two 1.3-m telescopes in the J (1.25 μm), H (1.65 μm), and K_s band (2.17 μm). Although the surveys has been completed, only about 47% of the sky is at present publicly available in the Second Incremental Release[1].

We selected from the 2MASS archive a large area around the Orion and MonR2 nebulae, with galactic coordinates

$$200° < l < 230°, \qquad -30° < b < -5°. \tag{19}$$

---

[1] See http://www.ipac.caltech.edu/2mass/
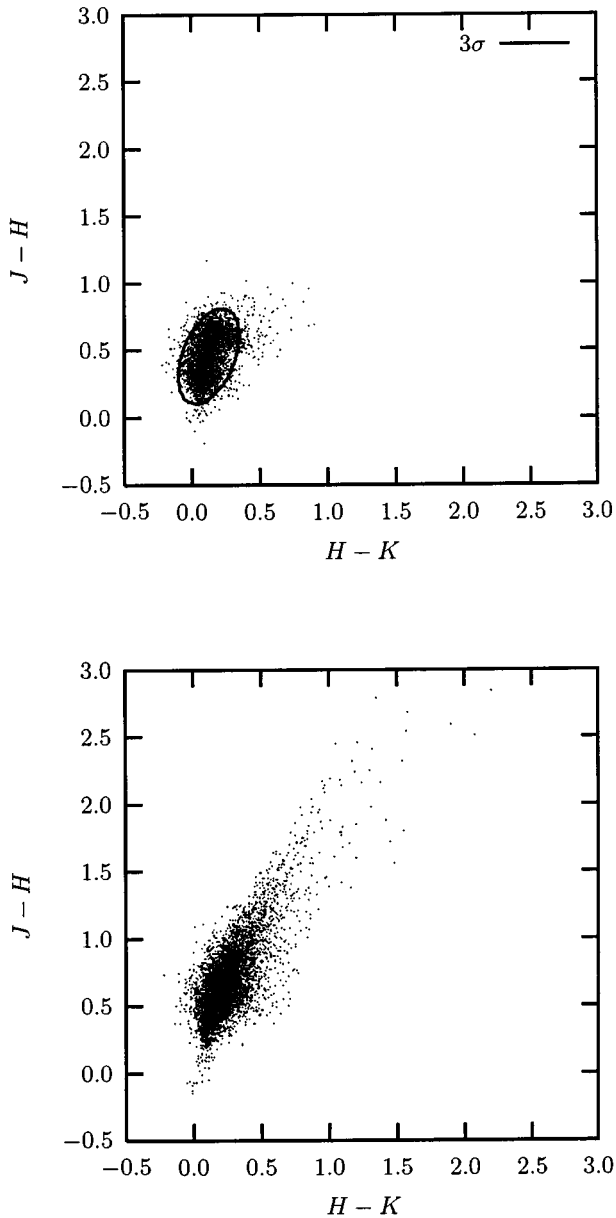
Fig. 3. Top: color-color diagram for stars in the control field, made from the colors of 6 000 stars. The ellipse represents a 3σ confidence region. **Bottom:** color-color diagram for stars in the central region of the cloud. The reddening due to the dust in the cloud significantly stretches the locus of stars in the plot.

The Orion region is a good candidate for our study for several reasons. The density of background stars, although not particularly high, is enough to perform a good analysis. Moreover, the 2MASS second incremental release covers this region very well, with only a small number of "holes" (except for the Nord-West part). Finally, there are already a number of studies in the literature which can be used to test our results (e.g., Carpenter 2000 uses the same data set to investigate the spatial distribution of young stars in these clouds and presents averaged $J - K_{\rm s}$ color maps of this region).

For each point source in this coordinate range we retrieved the magnitude on the three bands (fields x_mag), the relative total errors (x_msigcom), and some additional flags used to classify the goodness and the reliability of the detection (cc_flg and rd_flg). Possible artifacts (objects with cc_flg $\neq$ 0) were excluded from the analysis. We then applied our dedicated pipeline, as described in the following sections.

### 4.1. Preliminary analysis

Given the area on the sky and the number of stars, the pipeline "suggested" to perform the analysis using a grid of 820×700 points, with scale $2.4'$ per pixel, and with Gaussian smoothing characterized by $FWHM = 4.8'$. By default, the pipeline uses a $FWHM$ of 2 pixels, and about 15 "effective" stars per point (i.e., about 15 stars contribute to the signal for each point). We then made a preliminary analysis using standard values for the average colors and color scatters, thus obtaining a first extinction map for the region. The map obtained was used only as a first check for the basic parameters chosen, to identify a control field, and to obtain the photometric properties of stars in our sample. In particular, inside the boundaries Eq. (9), we found a region in the South-East of the map which does not show any significant extinction and which thus has been selected as control field (see Fig. 4). We then used the color of stars in this control field to set the photometric parameters needed in our algorithm, namely the average colors of stars (i.e., $\langle J - H \rangle$ and $\langle H - K \rangle$) and the scatter in colors (or, more precisely, the covariance matrix, which describes also correlations between colors). This step was performed in our pipeline by flagging stars inside the control region (which can be defined using the standard SAOimage format for region files) and by making a statistical analysis on the photometric data of flagged stars. Note that here only the 2MASS photometric data were used, and not the preliminary map (this map was used only to select the control field). As a result, an inaccurate choice for the initial constants of Eq. (6) to make the preliminary map had no influence on the statistical parameters obtained.

We observe that the use of a large survey such as 2MASS has the significant advantage with respect to normal observations of letting us study in detail the halos of molecular clouds. As a result, the choice for the control field is very robust, in the sense that even regions with extremely low extinction can be avoided. We also stress that this is possible because of the availability of data on large areas. Figure 3 shows the color dispersion of stars in the control field, together with the 3σ elliptical confidence region.

4.2. Final map

The final extinction map, shown in Fig. 4, was performed using the method discussed in this paper. Stars in the
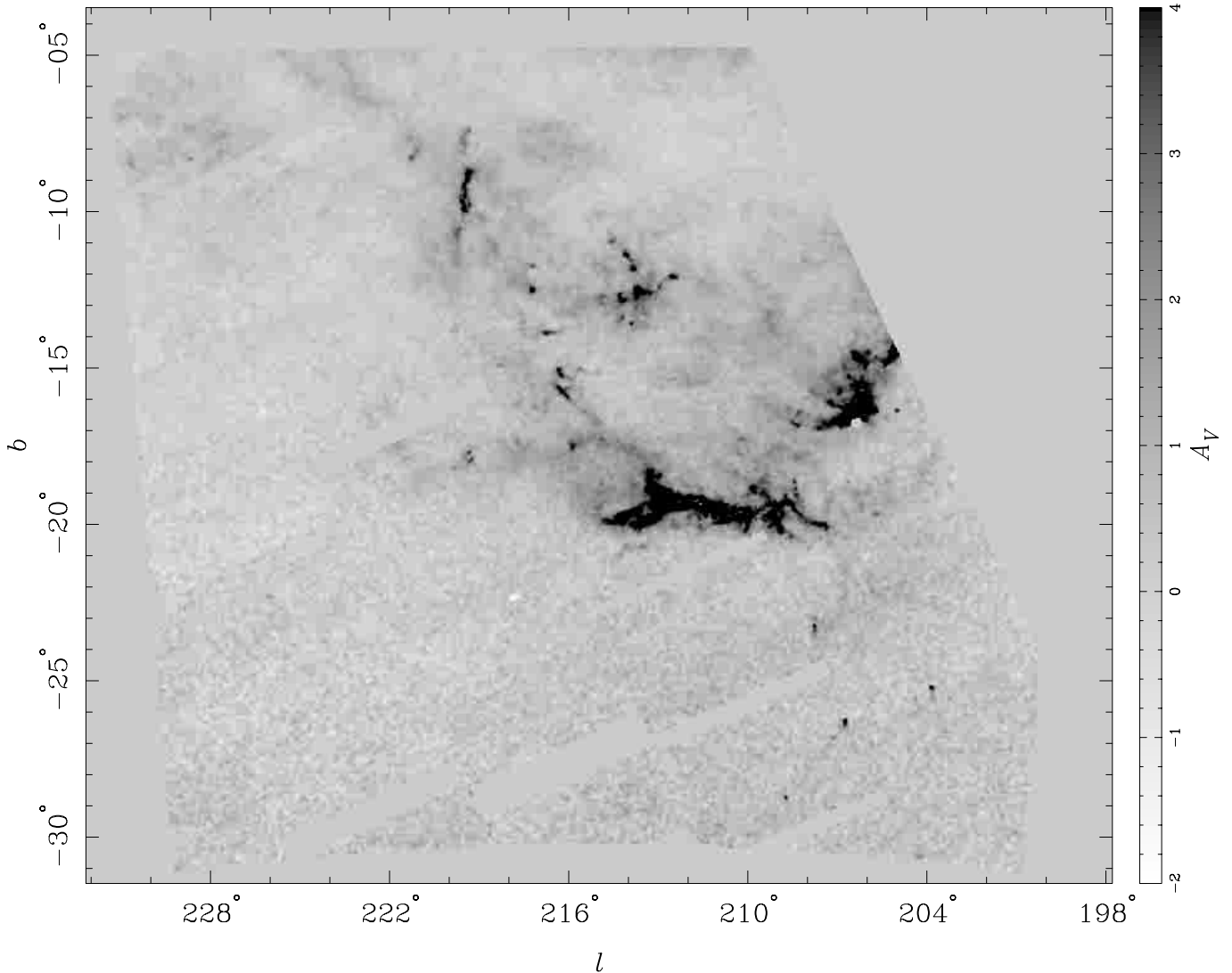
Fig. 4. The total region mapped using NICER. The cuts used in A_V emphasize the faint, diffuse halos around the main cores. The low noise of this map allows us to detect a $A_V = 0.5$ extinction with $3\,\sigma$ confidence level. The higher noise observed in the southern part is due to the smaller star density.

control field was used to set the intrinsic colors and color dispersions of (i.e., to evaluate the quantities $\langle c_i \rangle$ and $\mathrm{Cov}_{ij}(c^{\mathrm{tr}})$ described in Sect. 3.1). For each star in our region, we evaluated the column density using Eqs. (12) and (13). Finally, column densities for close stars was averaged out using three different schemes, namely simple (weighted) mean, sigma clipping, and median.

Figures 4 and 5 show the result obtained using a Gaussian smoothing characterized by $FVHM = 5'$ with σ-clipping (other smoothing schemes gives similar results, and differences cannot generally be seen by eye). Note that use of our optimized method, combined with the quality of the 2MASS data, allows us to detect at an unprecedented level of detail the faint, extended halos on which the main cores are embedded. At the smoothing size used, we reach the excellent sensitivity $A_V = 0.13$ at $1\,\sigma$ level.

Since the focus of this paper is mainly on the optimized reconstruction method, we will defer a discussion of the results obtained on this and on similar cloud complexes to a future paper. However, we want to point out a couple of peculiar features of the map obtained. First, we detect a single region with significantly bluer color (see Fig. 8, right). This region corresponds to the open cluster NGC 2204, whose blue stars contaminate our map thus simulating a "negative" extinction. Note also the cometary structures observed south to the Orion cloud (Fig. 8, left). In these condensations we observe extinction $A_V > 6$ on scales close to our smoothing length. Clearly, it would be worthwhile to study these regions at higher resolution.

In order to better appreciate the advantages obtained using the optimized technique described in this paper, we also carried out the analysis on the same field using only
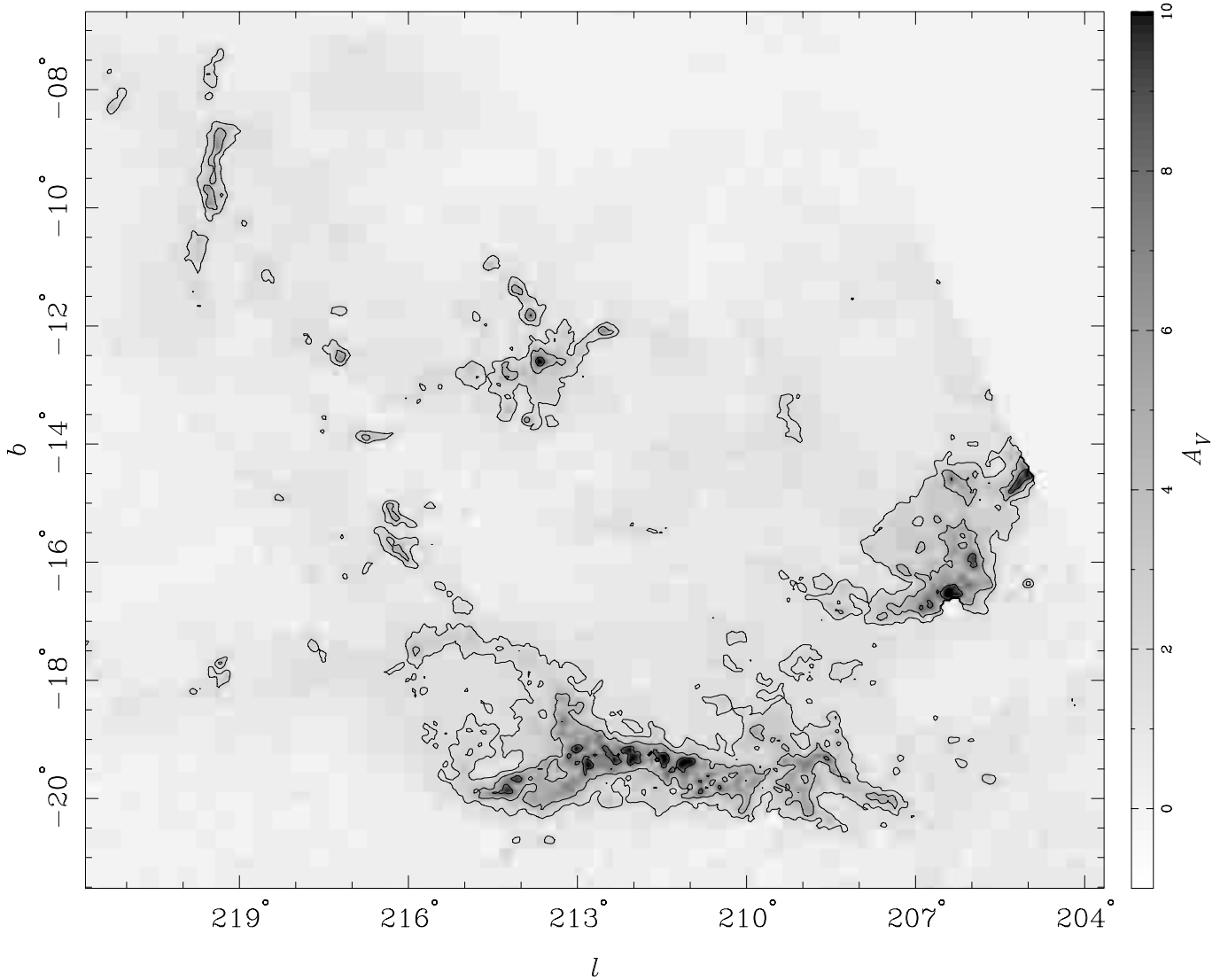
Fig. 5. Zoom of the central region of the Nicer extinction map. Cuts in this map emphasize the faint, diffuse halos around the main cores. Contours are displayed at $A_V = \{2, 4, 8\}$ magnitudes. Note that the maximum value obtained for A_V is 17.7 magnitudes.

the weighted average on the $H - K$ color for stars with relatively low photometric errors (we required errors on H and K to be smaller than 0.15 magnitudes). Figure 7 shows the result obtained with this standard method, and should be compared with Fig. 5.

- We first note that, apart from the different noise level, the two maps basically give the same values for the column density. This proves that the new method described in this paper is reliable and unbiased;
- The simple $H - K$ map is significantly noisier than the one obtained using the optimized method. Note that the resolution, set by the scale of the Gaussian smoothing, is the same in both maps. A quantitative analysis shows that with the use of the optimized method we gain a factor of two on the variance of the map. In other words, the new technique is able to reach the

same signal-to-noise ratio of the standard method using only half of the stars;
- Because of the reduced noise level, in Fig. 4 we are able to distinguish regions with extremely low column density ($A_V \simeq 0.5$ magnitudes), which would be otherwise undetectable.

We note that the exact gain obtained with the optimized method *strongly* depends on the individual intrinsic scatter of stars on each color. This last point depends on several factors, such as the depth of observations, the galactic latitude of the field, and the colors used. On the other hand, given the excellent results obtained with the 2MASS archive, we are confident that Nicer can produce extinction maps with significantly higher signal-to-noise ratio for any infrared data. Finally, we stress that although the method has been presented here for the
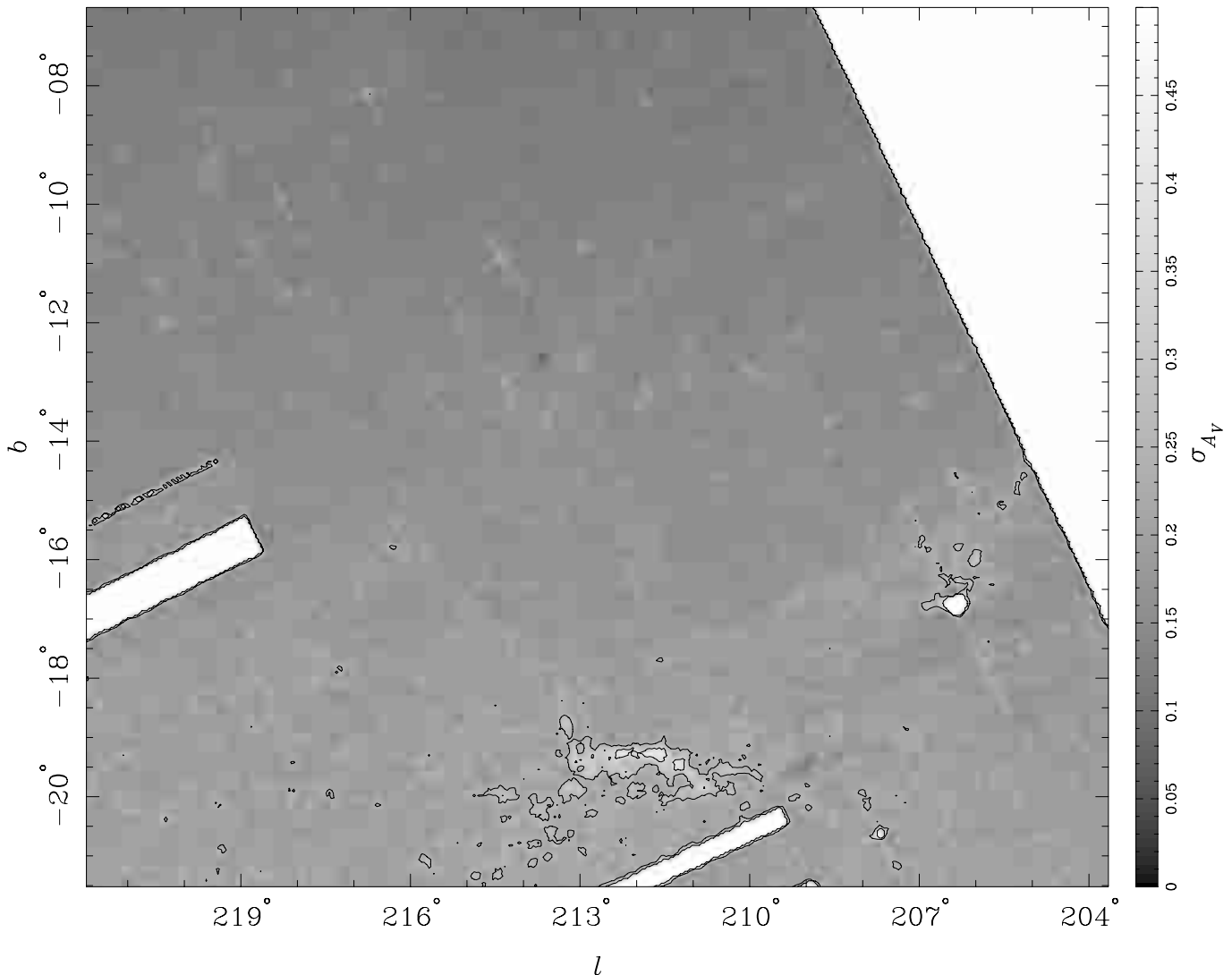
Fig. 6. The noise estimate $\sigma_{\hat{A}_V}$ (cf. Eq. (16)) for the central region shown in Fig. 5. The average noise ranges from about 0.15 to the top of the figure to about 0.25 to the bottom (where the density of background stars is smaller because of the increase in $|b|$); contours are plotted at $\sigma_{\hat{A}_V} = \{0.25, 0.35\}$. In the white regions the error diverges because of the lack of data. Increase in noise is observed in high absorption regions (because of the decrease in density of stars), and around bright stars in the top of the figure (which have been masked by the 2MASS pipeline). The white dot on the right is due to bright star zeta Ori ($V = 1.79$ mag). Except for these features, the overall noise level is rather uniform.

J, H, and K bands, it can actually be applied to any set of bands.

5. Conclusions

In this paper we have presented an optimized technique to produce highly accurate extinction maps from multi-band near-infrared photometric data. The method, which is a natural generalization of the near infrared color excess method of Lada et al. (1994), is able to produce significantly less noisy (and thus more accurate) extinction maps taking advantage of all bands available. A first example of application of this new technique to 2MASS data has shown an improvement with respect to the standard NICE algorithm of a factor 2 on the noise variance. This way, we have been able to detect extended diffuse halos down to $A_V \simeq 0.5$ magnitudes.
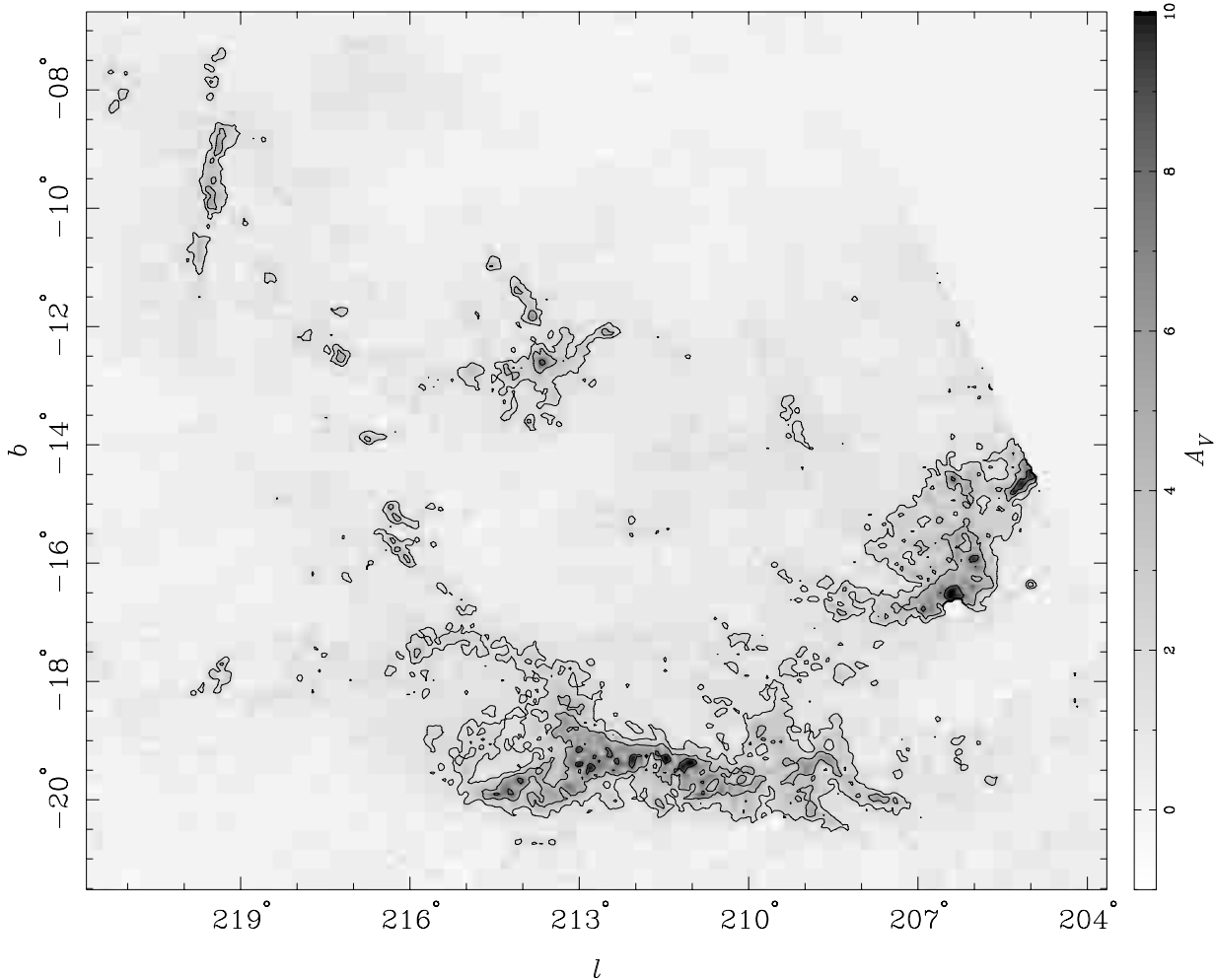
Fig. 7. The same region shown in Fig. 5 but with the extinction map obtained using the standard method (Lada et al. 1994). A quantitative analysis of the noise, carried out in regions with the lowest column density, shows that the variance of this map is about a factor two larger than the one shown in Fig. 5. Note also that, although the contours levels are the same as in Fig. 5, they appear much less smooth. This is another indication of the higher noise level of this map (rather than of intrinsic cloud structure; note that the resolution is the same for all images).
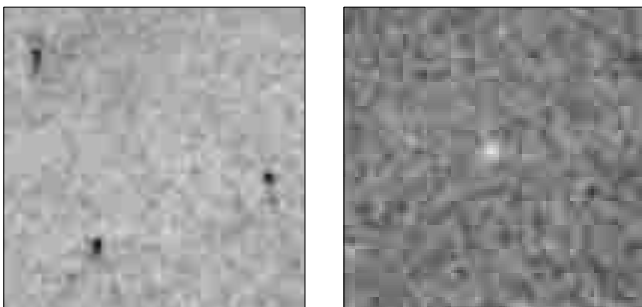
Fig. 8. Left: peculiar cometary structures are observed south of the main cloud. The size of the image is $5 \times 5 \ \mathrm{deg}^2$. **Right:** the only significant "negative" extinction corresponds to the open cluster NGC 2204, an open cluster in the halo (Hawarden 1976), whose average color is bluer than the average galactic field color. The size of the image is $4 \times 4 \ \mathrm{deg}^2$.

References

Alves, J., Lada, C. J., & Lada, E. A. 1999, ApJ, 515, 265

Alves, J., Lada, C. J., & Lada, E. A. 2001, Nature, 409, 159

Alves, J., Lada, C. J., Lada, E. A., Kenyon, S. J., & Phelps, R. 1998, ApJ, 506, 292

Andre, P., Ward-Thompson, D., & Barsony, M. 2000, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russell (University of Arizona Press, Tucson), 59

Arce, H. G., & Goodman, A. A. 1999, ApJ, 517, 264

Blitz, L., & Williams, J. P. 1999, in NATO ASIC Proc. 540: The Origin of Stars and Planetary Systems, 3

Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132

Bok, B. J. 1937, The distribution of the stars in space (Chicago: University of Chicago Press)

Bok, B. J. 1956, AJ, 61, 309

Bok, B. J. 1977, PASP, 89, 597

Cambrésy, L. 1999, A&A, 345, 965

Carpenter, J. M. 2000, AJ, 120, 3139

Casali, M. M. 1986, MNRAS, 223, 341

Chandler, C. J., & Richer, J. S. 2000, ApJ, 530, 851

Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792

Dickman, R. L. 1978, AJ, 83, 363

Epchtein, N., de Batz, B., Capoani, L., et al. 1997, The Messenger, 87, 27

Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, ApJ, 262, 590

Fukui, Y., Mizuno, N., Yamaguchi, R., et al. 1999, PASJ, 51, 745

Gregorio Hetem, J. C., Sanzovo, G. C., & Lepine, J. R. D. 1988, A&AS, 76, 347

Jenkins, E. B., & Savage, B. D. 1974, ApJ, 187, 243

Johnstone, D., Wilson, C. D., Moriarty-Schieven, G., et al. 2000, ApJ, 545, 327

Jones, T. J., Hyland, A. R., & Bailey, J. 1984, ApJ, 282, 675

Jones, T. J., Hyland, A. R., Robinson, G., Smith, R., & Thomas, J. 1980, ApJ, 242, 132

Juvela, M. 1998, A&A, 338, 723

Kleinmann, S. G., Lysaght, M. G., Pughe, W. L., et al. 1994, Experimental Astronomy, 3, 65

Lada, C. J. 1996, in CO: Twenty-Five Years of Millimeter-Wave Spectroscopy IAU Symp., 170, 387

Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, ApJ, 429, 694

Lilley, A. E. 1955, ApJ, 121, 559

Lombardi, M., & Bertin, G. 1998, A&A, 335, 1

Lombardi, M., Rosati, P., Nonino, M., et al. 2000, A&A, 363, 401

Lombardi, M., & Schneider, P. 2001, A&A, accepted [astro-ph/0104132]

Mattila, K. 1986, A&A, 160, 157

Myers, P. C. 1999, in NATO ASIC Proc. 540: The Origin of Stars and Planetary Systems, 67

Rieke, G. H., & Lebofsky, M. J. 1985, ApJ, 288, 618

Rosolowsky, E. W., Goodman, A. A., Wilner, D. J., & Williams, J. P. 1999, ApJ, 524, 887

Rossano, G. S. 1978, AJ, 83, 241

Thoraval, S., Boisse, P., & Duvert, G. 1997, A&A, 319, 948

van Dishoeck, E. F. 1992, in Astrochemistry of Cosmic Phenomena, IAU Symp., 150, 143

van Dishoeck, E. F., & Black, J. H. 1988, ApJ, 334, 771

Wilson, R. W., Jefferts, K. B., & Penzias, A. A. 1970, ApJ, 161, L43

Wolf, M. 1923, Astronomische Nachrichten, 219, 109

Wood, D. O. S., & Myers, P. C. 1995, ApSS, 224, 199

York, D. G., Adelman, J., Anderson, J. E., et al. 2000, AJ, 120, 1579