

# NICEST, a near-infrared color excess method tailored to small-scale structures

M. Lombardi<sup>1,2</sup>

<sup>1</sup> European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 Garching bei München, Germany  
e-mail: mlombard@eso.org

<sup>2</sup> University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

Received 4 July 2008 / Accepted 16 September 2008

## ABSTRACT

Observational data and theoretical calculations show that significant small-scale substructures are present in dark molecular clouds. These inhomogeneities can provide useful clues to the physical conditions inside the clouds, but can also severely bias extinction measurements. We present NICEST, a novel method to account and correct for inhomogeneities in molecular cloud extinction studies. The method, tested against numerical simulations, removes almost completely the biases introduced by sub-pixel structures and by the contamination of foreground stars. We applied NICEST to 2MASS data of the Pipe molecular complex. The map thereby obtained shows significantly higher (up to 0.41 mag in  $A_K$ ) extinction peaks than the standard NICER (Lombardi & Schneider 2001, A&A, 373, 359) map. This first application confirms that substructures in nearby molecular clouds, if not accounted for, can significantly bias extinction measurements in regions with  $A_K > 1$  mag; the effect, moreover, is expected to increase in more distant molecular clouds, because of the poorer physical resolution achievable.

**Key words.** dust, extinction – methods: statistical – ISM: clouds – infrared: ISM – ISM: structure – ISM: individual objects: Pipe molecular complex

## 1. Introduction

Although the details of star and planet formation are little known, there is a general consensus that these objects are created inside dark molecular clouds from the contraction and fragmentation of dense, cold cores. As a result, in the last decades molecular clouds have been studied in detail using many different techniques, from optical number counts (Wolf 1923; Bok 1937), to radio observations of carbon monoxide (CO) molecules (Wilson et al. 1970; Frerking et al. 1982), to near-infrared (NIR) extinction measurements (Lada et al. 1994).

Molecular hydrogen and helium represent approximately 99% of the mass of a cloud, but the absence of a dipole moment in these molecules makes them virtually undetectable at the low temperatures ( $T \sim 10$  K) of these objects. Hence, the (projected) mass distribution of dark clouds is usually inferred from the distribution of relatively rare tracers (such as CO or dust grains) under the assumption that these are uniformly distributed in the cloud.

As pointed out by Lada et al. (1994), the reddening of background stars in NIR bands is a simple and reliable method to study the dust distribution and thus the hydrogen column density. Compared to optical bands, NIR bands are less affected by extinction and are less sensitive to the physical properties of the dust grains (Mathis 1990), and thus their color excess can be directly translated into a hydrogen column density. In a series of papers we reconsidered and improved the NIR color excess (NICE) technique, by generalizing it to use three or more bands (NICER, see Lombardi & Alves 2001) and by taking a maximum-likelihood approach (Lombardi 2005).

Although the NICE and NICER techniques have been very successful in studying molecular clouds (see, e.g. Alves et al. 2001),

they are plagued by two complications: small-scale inhomogeneities and contamination by foreground stars. NIR studies typically assume a uniform extinction over (small) patches of the sky; however, in the presence of substructures such as steep gradients, unresolved filaments, or turbulence-induced inhomogeneities (Lada et al. 1994; Larson 1981; Heyer & Brunt 2004), the background stars used to estimate the cloud extinction would not be uniformly distributed in the patch, but would be preferentially observed in low density regions. Similarly, in the presence of foreground stars, their *relative* fraction increases with the dust column density. Unfortunately, both effects will bias *all* color excess estimators toward low column densities; moreover, the bias will be especially important in the dense regions of molecular clouds, which is where the star formation takes place.

In this paper we describe in detail the repercussions of small scale inhomogeneities and foreground stars in extinction measurements, and propose a simple method to obtain estimates of the column density in the presence of both effects that, under simple working hypotheses, is asymptotically unbiased. Our method does not rely on any (usually poorly verified) assumption regarding the small scale structure of the cloud; moreover, it can be applied to a general class of NIR extinction measurements (marked spatial point processes).

The paper is organized as follows. In Sect. 2 we introduce a general smoothing technique used in all cases to create smooth, continuous maps from the discrete pencil-beam extinction measurements carried out for each background star; we also show that in two typical applications (moving average and nearest neighbour smoothing) a bias is expected. In Sect. 3 we propose a new method that can correct this bias without making any assumption on the underlying form of the cloud substructure. This new method has been tested with numerical simulations, as

described in Sect. 4, and with a real-case application, presented in Sect. 5. We discuss and comment on the results obtained in Sect. 6, and finally we briefly present our conclusions in Sect. 7.

## 2. Cloud substructure

Measurements of the reddening of stars observed through a molecular cloud provide estimates of the cloud column densities in pencil beams characterized by a diameter of the order of a fraction of milliarcsec. However, these high resolution measurements are strongly undersampled and are affected by large uncertainties due to the photometric errors and to the intrinsic scatter of star colors in the NIR. Smooth and continuous extinction maps are normally obtained by interpolating and binning the various pencil beams. Different authors use different recipes to smooth the data (see, e.g. Lombardi 2002, for a description of several interpolation techniques), but in most cases the interpolation follows a standard scheme. Consider  $N$   $K$ -band extinction<sup>1</sup> measurements  $\{\hat{A}_n\}$  obtained, for example, from the color excess of the  $N$  background (see Lada et al. 1994 or Lombardi & Alves 2001). The extinction  $\hat{A}$  at any location in the sky is evaluated using a weighted average

$$\hat{A} = \frac{\sum_n w_n \hat{A}_n}{\sum_n w_n}. \quad (1)$$

The weights  $\{w_n\}$  are usually chosen to be significantly different from zero only for stars angularly close to the given location of the map (for example, Cambr esy et al. 2002 assign a unity weight to the  $N$  nearest neighbors). Clearly, Eq. (1) does not include slightly more complex situations where, for example, one takes the median of the extinctions measured from objects near a given position (e.g. Dobashi et al. 2008); however, most of the discussion carried out for the weighted average actually applies also to median or related estimators.

The binning in Eq. (1) washes out the cloud substructure on scales smaller than the typical size where the  $\{w_n\}$  are significantly different from zero, but this is needed in order to have smooth maps and to increase the signal-to-noise ratio. However, Eq. (1) also introduces a significant bias on the estimated column density. Suppose that, in the region of the sky that we are investigating (i.e., in the area covered by the  $N$  stars used to estimate  $A$ ) the column density has significant variations. Because of this, the local density  $\rho(\mathbf{x})$  of stars is not homogeneous throughout the cloud, but rather follows the scheme

$$\rho(\mathbf{x}) = \rho_0 10^{-\alpha k_\lambda A(\mathbf{x})}, \quad (2)$$

where  $\rho_0$  is the density of stars where no extinction is present,  $\alpha$  is the slope of the number counts, and  $k_\lambda = A_\lambda/A$  is the extinction law in the band  $\lambda$  considered. This effect, which is at the origin of the historical number count method to measure column densities in molecular clouds (Wolf 1923; Bok 1937), also induces a bias in Eq. (1), since regions with smaller extinctions and thus higher density of background stars will, on average, contribute more to the sum in Eq. (1) than regions with large extinctions. Note that since the color excess method requires measurements in at least two different bands, one should replace Eq. (2) with a version that provides the expected density of stars observed in all bands required for the application of the method (e.g.,  $H$  and  $K$ ). For

<sup>1</sup> For simplicity, in this paper we drop the subscript  $K$  normally used in the literature to denote the  $K$ -band extinction  $A_K$ ; it is also obvious that the same method applies to extinction measurements referred to any band (e.g., the visual extinction  $A_V$ ).

equally deep observations in all bands, the result has the same form as Eq. (2), where  $k_\lambda$  is the extinction law of the bluer band considered (e.g.,  $H$ ).

The main issue considered in this paper is the bias introduced by unresolved substructures in Eq. (1). In general, the bias should be defined here as the difference between the expected, mean value of  $\langle \hat{A} \rangle$  and the true column density  $A$  at the same position. However, clearly any smoothing technique introduces a bias because of the smoothing itself, and this bias is normally acceptable if it does not introduces a systematic effect on the *total* column density. In other words, a required property is

$$\int A(\mathbf{x}) d\mathbf{x} = \left\langle \int \hat{A}(\mathbf{x}) d\mathbf{x} \right\rangle = \int \langle \hat{A}(\mathbf{x}) \rangle d\mathbf{x}. \quad (3)$$

Clearly, in order for Eq. (3) to hold, the difference  $\langle \hat{A} \rangle - A$  must be of alternating sign. For our purposes, thus, it is more useful to define the bias as the difference between the expected mean value  $\langle \hat{A} \rangle$  and the same quantity that we would theoretically expect *in the presence of a uniform density of stars*, i.e. by ignoring the dependence of the density of stars on the local extinction described by Eq. (2) (see also below Eq. (7)). This definition is useful, since it isolates the standard effects of a smoothing from the effects introduced by the non-uniform sampling of the cloud structure. In addition, a column density estimator  $\hat{A}(\mathbf{x})$  that is unbiased according to this definition will also be unbiased according to Eq. (3) (provided the weights  $w_n$  are spatially invariant). For, when the density of background stars is uniform,  $\langle \hat{A}(\mathbf{x}) \rangle$  can always be written as a convolution of the true column density map  $A(\mathbf{x})$  with some kernel  $K$  normalized to unity (Lombardi 2002).

In the following, we will calculate explicitly this bias in two common smoothing schemes.

### 2.1. Moving average smoothing

In this section we will consider a simple weighting scheme in Eq. (1) where each weight  $w_n = w(\mathbf{x}_n; \mathbf{x})$  is a simple function of the location  $\mathbf{x}_n$  of the  $n$ th star (typically,  $w(\mathbf{x}_n; \mathbf{x})$  will depend only on the modulus  $|\mathbf{x}_n - \mathbf{x}|$ ).

As discussed above, our aim is to compare the expected column density estimate with the one that we would obtain in the absence of any selection effect in the number of background stars. We need thus to evaluate two averages: (i)  $\langle \hat{A}(\mathbf{x}) \rangle$ , where  $\hat{A}$  is calculated according to Eq. (1) and the ensemble average is taken over all positions of stars with the density given by Eq. (2) and over all individual column density measurements  $\hat{A}_n$ ; and (ii) the average  $\bar{A}(\mathbf{x})$ , which is basically the same quantity evaluated with a constant density of background stars  $\rho_0$ .

In principle, both averages can be calculated analytically using the method described by Lombardi & Schneider (2001; see below Sect. 3); in practice, for the purposes of this section a simple approximation can be used provided that in the weighted average of Eq. (1) a relatively large number of column densities are used with weights significantly different from 0. In this case we find for the first average

$$\langle \hat{A}(\mathbf{x}) \rangle = \frac{\int A(\mathbf{x}') w(\mathbf{x}'; \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}'}{\int w(\mathbf{x}'; \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}'}. \quad (4)$$

As shown by Eq. (2),  $\rho(\mathbf{x}')$  is a simple function of  $A(\mathbf{x}')$ . This suggests that the equation can be recast in a simpler form by

defining  $p_A(A; \mathbf{x})$ , the (weighted) probability that a point with extinction  $A$  is used to evaluate  $\hat{A}(\mathbf{x})$ , the column density at  $\mathbf{x}$ :

$$p_A(A; \mathbf{x}) = \frac{\int w(\mathbf{x}'; \mathbf{x}) \delta(A - A(\mathbf{x}')) d\mathbf{x}'}{\int w(\mathbf{x}'; \mathbf{x}) d\mathbf{x}'} . \quad (5)$$

By using Eq. (5) in Eq. (4) we find then

$$\langle \hat{A}(\mathbf{x}) \rangle = \frac{\int A p_A(A; \mathbf{x}) \rho(A) dA}{\int p_A(A; \mathbf{x}) \rho(A) dA} = \frac{\int A p_A(A; \mathbf{x}) 10^{-\alpha k_\lambda A} dA}{\int p_A(A; \mathbf{x}) 10^{-\alpha k_\lambda A} dA} . \quad (6)$$

In presence of a uniform distribution of stars,  $\rho(\mathbf{x}) = \rho_0$ , we would instead obtain simply

$$\bar{A}(\mathbf{x}) = \frac{\int A(\mathbf{x}') w(\mathbf{x}'; \mathbf{x}) d\mathbf{x}'}{\int w(\mathbf{x}'; \mathbf{x}) d\mathbf{x}'} = \int A p_A(A; \mathbf{x}) dA . \quad (7)$$

The two values presented in Eqs. (6) and (7) do not differ significantly provided the scatter of  $A$  in the patch considered around  $\mathbf{x}$  is much smaller than  $(\alpha k_\lambda \ln 10)^{-1}$ . For example, if we consider the  $H$  band of a typical region close to the Galactic plane with a standard reddening law (Rieke & Lebofsky 1985), we have  $\alpha \approx 0.34 \text{ mag}^{-1}$  and  $k_H \approx 1.55$  (Indebetouw et al. 2005), so that the maximum scatter allowed in  $A_K$  to have a negligible bias is  $\ll 0.82 \text{ mag}$ . Hence, clearly we need to be concerned by this bias in regions from moderate to large extinction.

It is also interesting to evaluate the bias  $\langle \hat{A}(\mathbf{x}) \rangle - \bar{A}(\mathbf{x})$  in the presence of small variations of  $A$  within the region considered. In this case, the probability distribution  $p_A(A; \mathbf{x})$  is peaked around the mean value of Eq. (7), and we can expand to second order the two exponential functions present in the numerator and denominator of Eq. (6). After a few manipulations we obtain then

$$\begin{aligned} \langle \hat{A}(\mathbf{x}) \rangle - \bar{A}(\mathbf{x}) &\approx -\alpha k_\lambda \ln 10 \int p_A(A; \mathbf{x}) [A - \bar{A}(\mathbf{x})]^2 dA \\ &= -\beta \frac{\int w(\mathbf{x}'; \mathbf{x}) \Delta^2(\mathbf{x}'; \mathbf{x}) d\mathbf{x}'}{\int w(\mathbf{x}'; \mathbf{x}) d\mathbf{x}'} , \end{aligned} \quad (8)$$

where  $\beta \equiv \alpha k_\lambda \ln 10$  and  $\Delta(\mathbf{x}'; \mathbf{x}) \equiv A(\mathbf{x}') - \bar{A}(\mathbf{x})$ . Hence, the difference between the two values is proportional to a weighted average of the scatter of  $A$  around its mean value  $\bar{A}(\mathbf{x})$  at  $\mathbf{x}$ , a quantity that, as we will see below, can be directly estimated from the data. Note that the bias is, to first order, quadratic on  $\Delta(\mathbf{x}'; \mathbf{x})$  and thus will be particularly severe when steep gradients are present in the underlying column density  $A(\mathbf{x})$ .

## 2.2. Nearest neighbor(s)

Cambr sy et al. (2002) suggest using a different prescription to make smooth extinction maps from the individual column density measurements. For each point in the map, they take the average extinction of the  $N$  angularly closest stars (nearest neighbours interpolation). As argued by Cambr sy et al. (2005, 2006), this method can potentially alleviate the bias introduced by the varying background density of stars described by Eq. (2), because measurements from low density regions will mostly appear isolated and will thus be used for relatively large patches of the sky.

Interestingly, using the technique described by Lombardi (2002) it is possible to obtain an analytical estimate for the average of  $\hat{A}$  in the smoothing scheme proposed:

$$\langle \hat{A}(\mathbf{x}) \rangle = \int A(\mathbf{x}') K(\mathbf{x}'; \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}' , \quad (9)$$

where the linear kernel  $K(\mathbf{x}'; \mathbf{x})$  is given by

$$K(\mathbf{x}'; \mathbf{x}) = \frac{e^{-\mu(\mathbf{x}'; \mathbf{x})}}{N} \sum_{n=0}^{N-1} \frac{[\mu(\mathbf{x}'; \mathbf{x})]^n}{n!} . \quad (10)$$

In this equation,  $N$  is the number of stars used in the nearest neighbors interpolation and  $\mu(\mathbf{x}'; \mathbf{x})$  is the average number of stars observed in  $B(\mathbf{x}'; \mathbf{x})$ , the disk centered on  $\mathbf{x}$  and of angular radius  $|\mathbf{x} - \mathbf{x}'|$ :

$$\mu(\mathbf{x}'; \mathbf{x}) = \int_{B(\mathbf{x}'; \mathbf{x})} \rho(\mathbf{x}'') d\mathbf{x}'' . \quad (11)$$

As shown by Lombardi (2002), the kernel  $K(\mathbf{x}'; \mathbf{x})$  is normalized, i.e.

$$\int K(\mathbf{x}'; \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}' = 1 , \quad (12)$$

an obvious result if one considers the measurement of the extinction in a uniform cloud where  $A(\mathbf{x}) = \text{const}$ . In our case, this property can also be proved directly from Eqs. (10) and (11). First, note that  $\mu(\mathbf{x}'; \mathbf{x}) = \mu(r; \mathbf{x})$  depends only on  $\mathbf{x}$  and on  $r = |\mathbf{x} - \mathbf{x}'|$ , and thus the same applies to  $K(\mathbf{x}'; \mathbf{x}) = K(r; \mathbf{x})$ . This suggests that we can recast the integral of Eq. (12) as

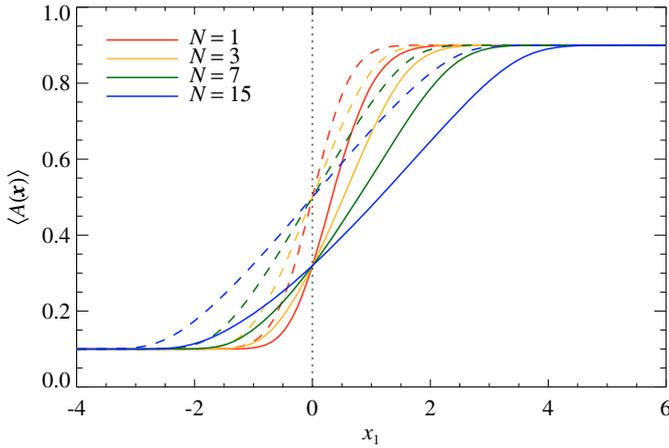
$$\begin{aligned} \int_0^\infty K(r; \mathbf{x}) \frac{\partial \mu(r; \mathbf{x})}{\partial r} dr \\ = \sum_{n=0}^{N-1} \frac{n!}{N} \int_0^\infty e^{-\mu(r; \mathbf{x})} [\mu(r; \mathbf{x})]^n \frac{\partial \mu(r; \mathbf{x})}{\partial r} dr = 1 . \end{aligned} \quad (13)$$

The fact that  $K(\mathbf{x}'; \mathbf{x})$  only depends on  $\mathbf{x}$  and  $r = |\mathbf{x} - \mathbf{x}'|$  (and not on the direction of the vector  $\mathbf{x} - \mathbf{x}'$ ) also implies that the nearest neighbors interpolation suffers from the bias discussed in this paper. Indeed, in the integral of Eq. (9) the kernel  $K$  gives the same ‘‘weight’’ to all points on circles centered on  $\mathbf{x}$ , irrespective of the specific value of  $\rho$  on the various points of the circles (what matters here is only the *average* value of  $\rho$  on a circle, and not the specific value at the various points). Hence, we do not expect that this technique can solve the bias introduced by the correlation between the extinction  $A$  and the density of background stars  $\rho$  expressed in Eq. (2), especially if steep gradients are present in the intrinsic cloud column density. We also expect that the bias present in the nearest neighbours interpolation increases with the number of neighbours  $N$ . Indeed, when  $N$  increases, the ‘‘size’’ of the kernel  $K(r; \mathbf{x})$ , i.e. the range in  $r$  where the kernel is significantly different from 0, also increases because of the effects of the polynomial terms in Eq. (10). This, as shown by Eq. (9), implies that the kernel averages out more distant regions in the sky, where differences among the local densities can be significant.

In order to better understand the points mentioned in the previous paragraph, it is useful to look at a simple example. Consider a cloud characterized by a Heaviside column density:

$$A(\mathbf{x}) = \begin{cases} A_1 & \text{if } x_1 < 0, \\ A_2 & \text{if } x_1 \geq 0. \end{cases} \quad (14)$$

In this case we can carry out all the calculations analytically, and obtain for each value of  $x_1$  the expected average measured extinction  $\langle \hat{A} \rangle$ . We do not report here the relevant equations, but rather show the results obtained in a typical case in Fig. 1. As shown by this plot, the nearest neighbours are clearly biased: for



**Fig. 1.** The bias of the nearest neighbors interpolation on a model where the true extinction is  $A(x) = 0.1$  mag for  $x_1 < 0$  and 0.9 mag for  $x_1 \geq 0$ . The solid lines show the average measured extinction of the nearest neighbours method, calculated using Eq. (9) for various numbers of neighbours  $N$ . The dashed lines show the extinction that one would measure if the density of stars were uniform in the whole field. Note that the solid lines are always below the dashed lines except at the extremes of this plot (where all estimators agree with the true value of  $A$  there). For this plot we set  $\alpha = 0.34$ ,  $k = 1.55$ , and  $\rho_0 = 1$ .

example, the various curves are not symmetric around  $x_1 = 0$ , and in addition

$$\langle \hat{A} \rangle = \frac{A_1 10^{\alpha k_1 A_1} + A_2 10^{\alpha k_1 A_2}}{10^{\alpha k_1 A_1} + 10^{\alpha k_1 A_2}} < \frac{A_1 + A_2}{2}. \quad (15)$$

In order to better understand the bias, we also plot in Fig. 1 the expected results in the case of a uniform density, obtained assuming  $\alpha = 0$  in Eq. (2) (dashed lines): in this case the curves are perfectly symmetric around  $x_1 = 0$  and  $\langle \hat{A} \rangle = (A_1 + A_2)/2$ . A comparison of the solid and dashed lines also confirms that curves with large  $N$ , being less steep, are biased over a larger interval around  $x_0 = 0$ . This is expected, since as shown by Eq. (10) the intrinsic size of the smoothing kernel  $K$  increases with  $N$ .

One might wonder whether the bias decreases by using a median estimator instead of a simple mean, as done by [Cambr esy et al. \(2005, 2006\)](#). These authors still identify, for each sky position  $\mathbf{x}$ , the  $N$  nearest neighbour stars, and then evaluate a simple median of the relative extinction measurements. Luckily, it is possible to evaluate the exact statistical properties of the median estimator (see [Lombardi 2002](#), Appendix A). For this purpose, it is useful to evaluate the probability distribution  $p_N(A; \mathbf{x})$  that one of the  $N$  neighbours around the position  $\mathbf{x}$  has a column density  $A$ :

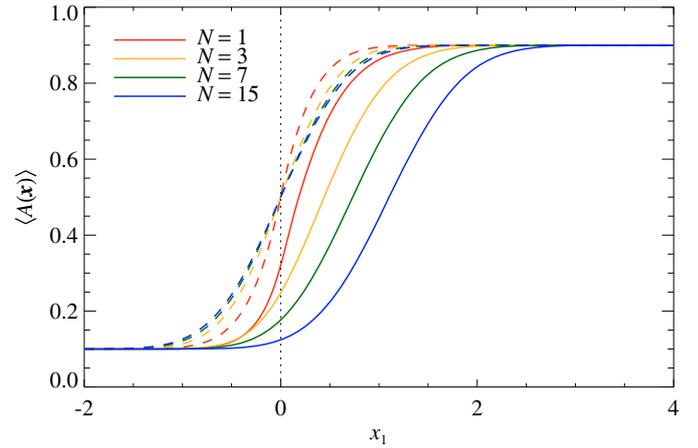
$$p_N(A; \mathbf{x}) = \int d\mu e^{-\mu} \frac{\mu^{N-2}}{N!} \int_{B(\mu; \mathbf{x})} dx' \rho(x') \delta(A - A(x)), \quad (16)$$

where  $B(\mu; \mathbf{x})$  is the disk centered at  $\mathbf{x}$  and with radius  $r$  defined by  $\mu = \mu(r; \mathbf{x})$  (note that since  $\mu(r; \mathbf{x})$ , for  $\mathbf{x}$  fixed, is a monotonic function of  $\mathbf{x}'$  this definition is unique). The cumulative distribution associated with  $p_N(A; \mathbf{x})$  is given by

$$P_N(A; \mathbf{x}) = \int d\mu e^{-\mu} \frac{\mu^{N-2}}{N!} v(\mu, A; \mathbf{x}). \quad (17)$$

In this equation,  $v(\mu, A; \mathbf{x})$  is the integral of  $\rho(x)$  carried out over all points of  $B(\mu; \mathbf{x})$  where  $A(x') < A$ :

$$v(\mu, A; \mathbf{x}) \equiv \int_{B(\mu; \mathbf{x}) \cap \{x' | A(x') < A\}} \rho(x') dx'. \quad (18)$$



**Fig. 2.** Similarly to Fig. 1, but for the median-nearest neighbors interpolation. Again, the solid lines, representing the expected mean measurements, are always below the dashed lines, representing the same measurements in the ideal case where the density of background stars is uniform within the whole field. Note that because of the properties of the median, for large values of  $N$  the “transition” from  $A = 0.1$  mag to  $A = 0.9$  mag is faster than in Fig. 1.

Note that  $v(\mu, A; \mathbf{x}) \rightarrow \mu$  as  $A \rightarrow \infty$ . Finally, the median of the  $N = 2k - 1$  nearest neighbours is then provided by the expression

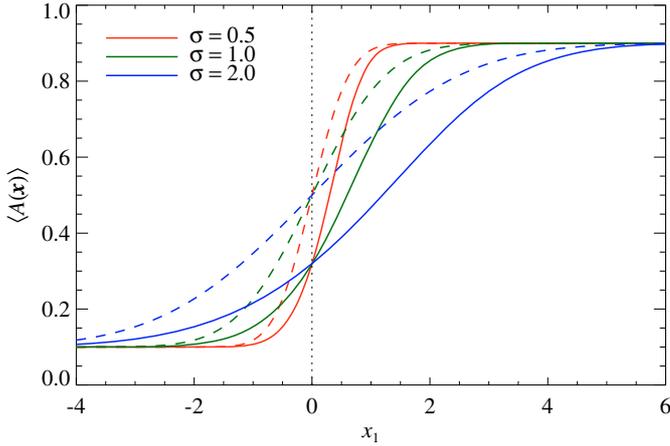
$$p_m(A) = k \binom{N}{k} p_N(A; \mathbf{x}) [P_N(A; \mathbf{x})]^{k-1} [1 - P_N(A; \mathbf{x})]^{k-1}. \quad (19)$$

Although it is difficult to obtain a general expression for the bias introduced by the median-nearest neighbours estimator, it is clear that to a first approximation the same arguments discussed above for the simple mean apply (note also that, trivially, when  $N = 1$  the two estimators are identical). However, for the simple example considered above, we can actually carry out the calculations using Eqs. (16–19) and show that the bias is still significant. Figure 2 summarizes the results obtained in the model described by Eq. (14), and proves that no real improvement is obtained from the use of a median estimator (cf. Fig. 1). In particular, a severe bias is still present, and its amount increases with  $N$ . For comparison, in Fig. 3 we also show the results obtained from the simple moving average smoothing discussed in Sect. 2.1: interestingly, the plot is very similar to the one for the simple average nearest neighbors interpolation (Fig. 1). This is not surprising because the expected average values  $\langle A(x) \rangle$  measured with these two techniques can be described in terms of a simple convolution with some appropriate kernels (cf. Eqs. (4) and (10); see also [Lombardi 2002](#)).

In summary, a careful analysis of the nearest neighbors shows that this interpolation technique, for the specific case of extinction measurements, is still affected by a significant bias. At least for the case of a simple mean estimator, the origin of the problem can be traced to the different behaviour of the kernel  $K(\mathbf{x}; \mathbf{x}')$  with respect to the density  $\rho(\mathbf{x}')$  in Eq. (9): in particular, the kernel  $K$  does not respond directly to variations of  $\rho$ , and instead depends only  $\mu(\mathbf{x}'; \mathbf{x})$ , i.e. on the total “mass” within a disk of radius  $r = \langle \mathbf{x} - \mathbf{x}' \rangle$ .

### 3. NICEST: over-weighting high column-density measurements

As shown by Eq. (8), the bias discussed here strongly depends on the small-scale structure of the molecular cloud through the



**Fig. 3.** Similarly to Fig. 1, but for the moving average smoothing with a Gaussian window function with different dispersion parameters  $\sigma$ . As for Figs. 1 and 2, the solid lines, representing the expected mean measurements, are always below the dashed lines, representing the same measurements in the ideal case where the density of background stars is uniform within the whole field.

probability distribution  $p_A(A)$ . Although on large scales many models predict a log-normal distribution for  $p_A$ , clearly we should expect significant deviations from the theoretical expectations on the small scales often investigated in NIR studies. In addition, even at large scales the superposition of different cloud complexes, or a significant “thickness” of a cloud along the line of sight, can produce significant deviations from a log-normal distribution (Vázquez-Semadeni & García 2001; see also Lombardi et al. 2006, for a case where the log-normal distribution is not a good approximation). Hence, *we need to be able to correct for the substructure bias in a model independent way.*

### 3.1. Statistical properties of the moving average smoothing

In order to better understand, and then fix, the bias present in Eq. (1) in the case of the moving average smoothing, we use the theory developed in Lombardi & Schneider (2001, 2002, 2003). The key equations to obtain the ensemble average of  $\hat{A}(\mathbf{x})$  are summarized below:

$$Q(s; \mathbf{x}) = \int [e^{-sw(\mathbf{x}'; \mathbf{x})} - 1] \rho(\mathbf{x}') d\mathbf{x}' \quad (20)$$

$$Y(s; \mathbf{x}) = \exp[Q(s; \mathbf{x})], \quad (21)$$

$$C(w; \mathbf{x}) = \frac{1}{1 - P(\mathbf{x})} \int_0^\infty e^{-ws} Y(s; \mathbf{x}) ds, \quad (22)$$

$$\langle \hat{A}(\mathbf{x}) \rangle = \int A(\mathbf{x}') w(\mathbf{x}'; \mathbf{x}) C(w(\mathbf{x}'; \mathbf{x}), \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}'. \quad (23)$$

We briefly comment on this set of equations:

- The equations are invariant upon the transformation  $w(\mathbf{x}'; \mathbf{x}) \mapsto qw(\mathbf{x}'; \mathbf{x})$ , with  $q$  a positive constant. This is expected, since an overall multiplicative constant in the weight function does not effect Eq. (1). For simplicity, in the following we will assume, without loss of generality, that the weight function is normalized according to

$$\int w(\mathbf{x}'; \mathbf{x}) \rho(\mathbf{x}') d\mathbf{x}' = 1. \quad (24)$$

- The quantity  $Q(s; \mathbf{x})$  can be shown to be related to the Laplace transform of  $p_w(w; \mathbf{x})$ , the probability distribution for the values of the weight function  $w(\mathbf{x}'; \mathbf{x})$  with  $\mathbf{x}$  fixed.

- $P(\mathbf{x})$  is the probability that no single star is present within the support  $\pi_w(\mathbf{x})$  of  $w(\mathbf{x}'; \mathbf{x})$ , defined as  $\pi_w(\mathbf{x}) \equiv \{\mathbf{x}' | w(\mathbf{x}'; \mathbf{x}) > 0\}$ . Hence,  $P(\mathbf{x})$  can be simply evaluated from

$$P(\mathbf{x}) = \exp \left[ - \int_{\pi_w(\mathbf{x})} \rho(\mathbf{x}') d\mathbf{x}' \right]. \quad (25)$$

Note that  $P(\mathbf{x}) = 0$  for weight functions with infinite support (such as a Gaussian).

- $C(w; \mathbf{x})$  is a simple Laplace transform of  $Y(s; \mathbf{x})$ . This is the key quantity that enters the final result (23) together with the original smoothing function  $w(\mathbf{x}'; \mathbf{x})$  and the star density  $\rho(\mathbf{x}')$ .
- A key parameter in the calculation of  $C(w)$ , and thus in the final result  $\langle \hat{A}(\mathbf{x}) \rangle$ , is the so-called *weight number of objects*  $\mathcal{N}(\mathbf{x})$ , defined as

$$\mathcal{N}(\mathbf{x}) \equiv \left[ [1 - P(\mathbf{x})] \int [w(\mathbf{x}'; \mathbf{x})]^2 \rho(\mathbf{x}') d\mathbf{x}' \right]^{-1}, \quad (26)$$

where as usual  $w(\mathbf{x}'; \mathbf{x})$  is taken to be normalized according to Eq. (24). Informally,  $\mathcal{N}(\mathbf{x})$  counts the number of stars that contribute (with a weight significantly different from zero) to the average  $\hat{A}(\mathbf{x})$ .

- In the limit  $\mathcal{N}(\mathbf{x}) \gg 1$  it is possible to obtain a simple expression for  $C(w; \mathbf{x})$ , which for  $P(\mathbf{x}) = 0$  takes the form

$$C(w; \mathbf{x}) \simeq \frac{1}{1 + w} + \frac{[\mathcal{N}(\mathbf{x})]^{-1}}{(1 + w)^3}. \quad (27)$$

This expression shows that to lowest order  $C \simeq 1 - w + \mathcal{N}^{-1}$ , where both terms  $w$  and  $\mathcal{N}^{-1}$  introduce a *relative* correction of order  $\mathcal{N}^{-1}$  in the final result  $\langle \hat{A}(\mathbf{x}) \rangle$ .

The results summarized above rigorously prove the statements of Sect. 2.1. In particular, in the limit of a large weight number of objects  $\mathcal{N}(\mathbf{x})$ , we have  $C(w) \simeq 1$ , and thus Eq. (23) together with the normalization (24) reduces to Eq. (4). In practice, numerical simulations show that Eq. (4) is already accurate for relatively small weight numbers (of the order of  $\mathcal{N} \sim 10$ ; see Fig. 3 of Lombardi & Schneider 2001).

### 3.2. A fix for the inhomogeneities bias

The bias of Eq. (1) is essentially due to a change in the density of extinction measurements as a function of  $A$  (Eq. (2)). In the framework of the moving average smoothing, it is thus reasonable to try to “compensate” for this effect by including in the weights of the estimator (1) a factor  $10^{\alpha k_i A}$ :

$$w_n = w(\mathbf{x}_n; \mathbf{x}) 10^{\alpha k_i \hat{A}_n}. \quad (28)$$

With this modification, we expect  $\hat{A}$  to be unbiased provided that  $\mathcal{N}$  is large (so that we can take  $C \simeq 1$ ) and that measurement errors on  $\hat{A}_n$  can be neglected. Instead, in the presence of significant errors on the individual column density estimations, we expected  $\hat{A}$  to be biased toward high extinctions. This happens because of the non-linearity of the introduced factor in  $\hat{A}_n$ : for example,  $\hat{A}_n$  is symmetrically distributed around  $A_n$ ,  $10^{\alpha k_i \hat{A}_n} \hat{A}_n$  will be skewed against values greater than  $10^{\alpha k_i A_n} A_n$ . In summary, the modified weight of Eq. (28) removes (to a large degree) the bias due to the cloud substructure, but introduces a new bias related to the scatter of the individual extinction measurements. While the former is unpredictable, the latter can be estimated accurately provided we know the expected errors on  $\hat{A}_n$ .

In order to better understand the properties of the estimator (1) with the new weight, we evaluate the average  $\langle \hat{A} \rangle$ . For simplicity, initially we will ignore measurement errors and focus on the effects introduced by a non-uniform density. We first note that Eq. (28) is equivalent to the use of a weight  $w_n = w'(\mathbf{x}_n; \mathbf{x})$ , where

$$w'(\mathbf{x}'; \mathbf{x}) \equiv w(\mathbf{x}'; \mathbf{x})/\rho(\mathbf{x}'). \quad (29)$$

We require the new weight  $w'(\mathbf{x}'; \mathbf{x})$  to be normalized according to Eq. (24), which implies

$$\int w(\mathbf{x}'; \mathbf{x}) d\mathbf{x}' = 1. \quad (30)$$

Interestingly, the normalization condition (30) for  $w$  does not depend on the density  $\rho$  anymore, and it is thus possible to choose for  $w$  a spatially invariant function  $w(\mathbf{x}'; \mathbf{x}) = w(\mathbf{x}' - \mathbf{x})$  (a typical choice would be a Gaussian,  $w \propto e^{-|\mathbf{x}' - \mathbf{x}|/2\sigma^2}$ ). We have then

$$\begin{aligned} \langle \hat{A}(\mathbf{x}) \rangle &= \int A(\mathbf{x}') w'(\mathbf{x}'; \mathbf{x}) [1 - w'(\mathbf{x}'; \mathbf{x}) + \mathcal{N}^{-1}(\mathbf{x})] \rho(\mathbf{x}') d\mathbf{x}' \\ &= \int A(\mathbf{x}') w(\mathbf{x}'; \mathbf{x}) [1 - w(\mathbf{x}'; \mathbf{x})/\rho(\mathbf{x}') + \mathcal{N}^{-1}(\mathbf{x})] d\mathbf{x}'. \end{aligned} \quad (31)$$

After a few manipulations we can rewrite this expression in the form

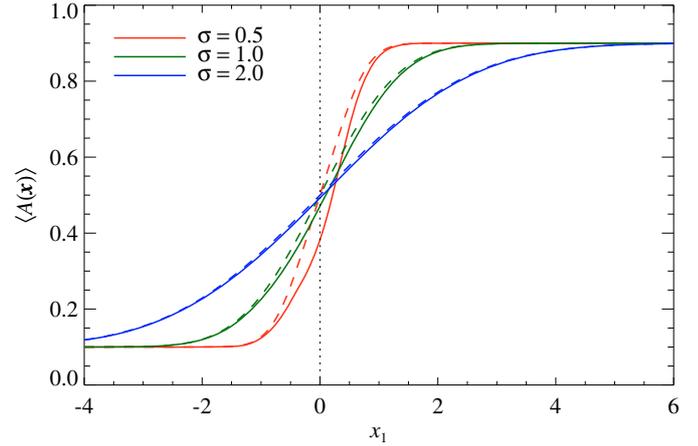
$$\begin{aligned} \langle \hat{A}(\mathbf{x}) \rangle &= \bar{A}(\mathbf{x}) - \frac{1}{\rho_0} \int w^2(\mathbf{x}'; \mathbf{x}) 10^{\alpha k_\lambda \bar{A}(\mathbf{x}')} [A(\mathbf{x}') - \bar{A}(\mathbf{x})] d\mathbf{x}' \\ &\simeq \bar{A}(\mathbf{x}) - \frac{10^{\alpha k_\lambda \bar{A}(\mathbf{x})}}{\rho_0} \int w^2(\mathbf{x}'; \mathbf{x}) [\Delta(\mathbf{x}'; \mathbf{x}) \\ &\quad + \beta \Delta^2(\mathbf{x}'; \mathbf{x})] d\mathbf{x}, \end{aligned} \quad (32)$$

where in the second equality we have expanded the exponential term to first order using the definitions following Eq. (8). In summary, the bias  $\langle \hat{A}(\mathbf{x}) \rangle - \bar{A}(\mathbf{x})$  of the proposed estimator is composed of two terms:

- a weighted average of  $\Delta(\mathbf{x}'; \mathbf{x})$ . Since the weighted average uses  $w^2(\mathbf{x}'; \mathbf{x})$  as the weight, this term is not expected to vanish identically unless the weight is a top-hat function [cf. Eq. (8), where instead the weighted average involves  $w(\mathbf{x}'; \mathbf{x})$  and therefore no linear contribution in  $\Delta$  is present]. Still, we do not expect any systematic bias related to this term, because on average  $\Delta(\mathbf{x}'; \mathbf{x})$  by construction has an alternating sign;
- a weighted average of  $\Delta^2(\mathbf{x}'; \mathbf{x})$ . This term is manifestly negative and therefore introduces a bias in the result. The bias is very similar to the original bias discussed in Eq. (8): it is proportional to  $\beta \equiv \alpha k_\lambda \ln 10$ , and depends on the scatter  $\Delta(\mathbf{x}'; \mathbf{x}) \equiv A(\mathbf{x}') - \bar{A}(\mathbf{x})$  of the local column density with respect to its average value. However, there is a significant difference: the bias is inversely proportional to  $\rho_0 10^{-\alpha k_\lambda \bar{A}(\mathbf{x})}$ , i.e. to the local average density within the weight function; moreover, the average is taken over  $w^2$  instead of  $w$ . These two differences, together, indicate that the bias of Eq. (32), compared to that of Eq. (8), is smaller by a factor  $\sim \mathcal{N}(\mathbf{x})$ .

In other words, the modification of the weight operated by Eq. (28) reduces the bias present in the original simple moving weight average by a factor  $\mathcal{N}(\mathbf{x})$ , which in typical cases is significantly greater than unity<sup>2</sup>. On the other hand, the fact that

<sup>2</sup> In reality, a detailed calculation shows that the key parameter here is  $\mathcal{N}_{\text{eff}}(\mathbf{x})$ , which is defined analogously to  $\mathcal{N}(\mathbf{x})$  but with  $w$  replaced by  $w_{\text{eff}} \equiv wC(w)$ . As shown in Lombardi (2002),  $\mathcal{N}_{\text{eff}} \geq 1$ , so the bias never increases.



**Fig. 4.** Similar to Fig. 3, but for the NICEST moving average smoothing (see Sect. 3). In this case a significant bias is observed only for  $\sigma = 0.5$ , which however corresponds to a very small weight number  $\mathcal{N} \simeq 2$ .

the method is ineffective when  $\mathcal{N}$  is low is evident from the definition of  $w_n$ : in particular, as  $\rho \rightarrow 0$  we expect that only a single star contributes to the averages of Eq. (1) with a weight  $w_n = w(\mathbf{x}_n; \mathbf{x}) 10^{\alpha k_\lambda \hat{A}_n}$  significantly different from zero, but then trivially  $\hat{A} = \hat{A}_n$  and no correction is effectively performed<sup>3</sup>.

In Fig. 4 we show the effects of the proposed smoothing technique on the model discussed in Sect. 2.2. A comparison with Fig. 3 shows that the NICER interpolation is very effective in reducing the bias of the simple moving average interpolation, especially for relatively large values of the smoothing factor  $\sigma$  (and thus of the weight number of objects  $\mathcal{N}$ ).

We now consider the effects of measurement errors on the modified estimator. For simplicity, we consider the limit of a large number of stars ( $\mathcal{N} \gg 1$ ), so that  $C(w) \simeq 1$  (in any case, as shown above, the technique proposed is inefficient when  $\mathcal{N}$  is of the order of unity). In this case, a simple calculation shows that measurement errors introduce a bias  $B_{\text{err}}$  of the order of

$$B_{\text{err}} \simeq \beta \frac{\sum_n w_n \sigma_n^2}{\sum_n w_n}, \quad (33)$$

where  $\{\sigma_n^2\}$  are the variances on  $\{\hat{A}_n\}$ . In other words, the method proposed in this paper introduces a small bias toward large extinction values. In order to estimate the order of magnitude of  $B_{\text{err}}$ , we note that the median error in  $K$  band extinctions on NICER column density estimates from the 2MASS catalog is typically<sup>4</sup>  $\sigma \simeq 0.13$  mag, and that less than 1% of stars have  $\sigma > 0.2$  mag. Taking  $\sigma_n = 0.13$  mag for all stars in Eq. (33), we find a bias  $\langle \hat{A} \rangle - A \simeq 0.02$  mag (always in the  $K$  band). Interestingly, although very small, this bias, in contrast to the one described by Eq. (32), can be corrected to first order because its expression involves only known quantities.

In summary, we propose to replace the estimator (1) with

$$\hat{A} = \frac{\sum_n w_n \hat{A}_n}{\sum_n w_n} - \beta \frac{\sum_n w_n \sigma_n^2}{\sum_n w_n}, \quad (34)$$

<sup>3</sup> In this example  $\mathcal{N} \rightarrow 0$  while  $\mathcal{N}_{\text{eff}} \rightarrow 1$ , which agrees with the previous footnote.

<sup>4</sup> The expected error on  $\hat{A}_n$  is calculated using the standard NICER techniques, and thus includes both the typical 2MASS photometric uncertainties and the star color scatters as measured in control fields where the extinction is negligible. Hence, this error includes the effects of different stellar populations on the scatter in the intrinsic star colors, as long as the various stellar populations are represented in the control field.

where  $w_n \equiv w(\mathbf{x}'; \mathbf{x})10^{\alpha k_\lambda \hat{A}_n}$ . This new estimator, that we name NICEST, significantly reduces the bias due small-scale inhomogeneities in the extinction map and is (to first order) unbiased with respect to uncertainties on  $\hat{A}_n$ .

If one is ready to accept a small bias on  $\hat{A}$  for extinction measurement uncertainties, it is also possible to use Eq. (34) without the last term. In this respect, we also note that the correcting factor present in Eq. (34) is independent of the local extinction, and only depends on the individual errors on the measurements (these typically are a combination of the photometric uncertainties and of the intrinsic scatter of colors of the stellar population considered). As a result, it is not unrealistic to expect that this factor is approximately constant within the field of observation, and that it is equally present in the control field used to fix the zero of the extinction.

Finally, we note that in NICEST a central role is played by the slope of the number counts  $\alpha$ , operationally defined through Eq. (2). In reality, however, it clear that Eq. (2) cannot strictly hold, because there is an upper limit on the (un-reddened) magnitude of stars. Hence, depending on the limiting magnitude of the observations, there is an upper limit to the measurable extinction of stars. As a result, in regions where the extinction is greater than this limit, the density of background stars vanishes, with the result that in these cases even NICEST cannot be unbiased. The exact value of the upper measurable extinction depends on many factors, including the waveband  $\lambda$  and limiting magnitude  $m_{\max}$  of the observations, and the distance  $d$  of the molecular cloud. An approximate relation for the maximum extinction  $A_{\max}$  is

$$A_{\max} \sim \frac{1}{k_\lambda} \left[ m_{\max} - M_{\max} - 5 \log_{10} \left( \frac{d}{10 \text{ pc}} \right) \right], \quad (35)$$

where  $M_{\max}$  is the maximum absolute magnitude of stars in the band considered. The maximum extinction  $A_{\max}$  should be evaluated in each specific case; for example, for typical 2MASS observations, using  $m_{\max} \simeq 15$  mag,  $k_H = 1.55$  (Indebetouw et al. 2005), and  $M_{\max} \simeq -5$  mag we obtain  $A_{\max} \simeq 10$  mag for a cloud located at 100 pc.

### 3.3. Foreground stars

Foreground stars do not contribute to the extinction signal, but do contribute to the noise of the estimators (1) and (34), and thus whenever possible they should be excluded from the analysis. Usually, foreground stars are easily identified in high ( $A_K > 1$  mag) column density regions, but are almost impossible to distinguish in low and mid extinction regions. So far we assumed that all stars are background to the cloud, but clearly in real observations we should expect a fraction  $f$  of foreground objects. In principle, if this fraction is known, we could correct the column density estimate  $\hat{A}$  into  $\hat{A}/(1-f)$ : in this way, we would obtain an unbiased estimator of the column density at the price of increased noise (due to foreground objects). In practice, the fraction of foreground stars is itself an increasing function of the local column density, i.e.  $f = f(A)$ , because of the decreased density of background stars in highly extinguished regions. Hence, the simple scheme proposed above is not easily implemented.

Interestingly, NICEST perfectly adapts to the presence of foreground stars. Suppose that in regions *with negligible extinction* a fraction  $f_0 \equiv f(A=0) < 1$  of stars is foreground to the cloud: then, in our notation, we can rephrase this by assigning a non-vanishing probability to  $p_A(A=0)$ , i.e. by treating foreground stars as a special case of substructure (much like ‘‘holes’’ in the

cloud). As we consider higher density regions, the fraction  $f(A)$  of foreground stars increases but, because of the correcting factor in  $w_n$ , we still expect to measure  $(1-f_0)A$ : in other words, the estimator  $\hat{A}/(1-f_0)$ , with  $\hat{A}$  given by Eq. (34) is expected to be unbiased both for small-scale inhomogeneities and for foreground contamination. Note that the correcting factor  $(1-f_0)^{-1}$  is usually very close to 1 for nearby molecular clouds, and can be easily evaluated by comparing the density of foreground stars (easily measured in highly extinguished regions) with the total density of stars in regions free from extinction.

## 4. Simulations

In order to test the effects of substructures and the reliability of the method proposed here in a real case scenario we performed a series of simulations. We started from a 2MASS/NICER map of the Pipe nebula (Lombardi et al. 2006), a nearby complex of molecular clouds seen in projection to the Galactic bulge, an ideal case for extinction studies (see also Alves et al. 2008; Onishi et al. 1999). We took this map as the *true* dust column density of a cloud complex, and simulated a set of background stars. In order to show the effects of substructures, we used a very low density of background stars (25 stars  $\text{deg}^{-2}$  instead of the original  $\sim 9.4 \times 10^5$  stars  $\text{deg}^{-2}$  used to build the 2MASS/NICER map). This, effectively, corresponds to a linear downsizing of the structures of the Pipe nebula by a factor of  $\sim 60$ .

The simulations were carried out using the following simple technique. We generated a set of background stars uniformly distributed on the field of view. The stars were characterized by exponential number counts with exponent  $\alpha = 0.34$  in the  $K$  band and by intrinsic color and color scatters similar to the ones observed in the 2MASS catalog (see Table 2 of Lombardi 2005, for a complete list of the parameters used). We added to each star magnitude the local value of the 2MASS/NICER extinction and also some random measurement errors:

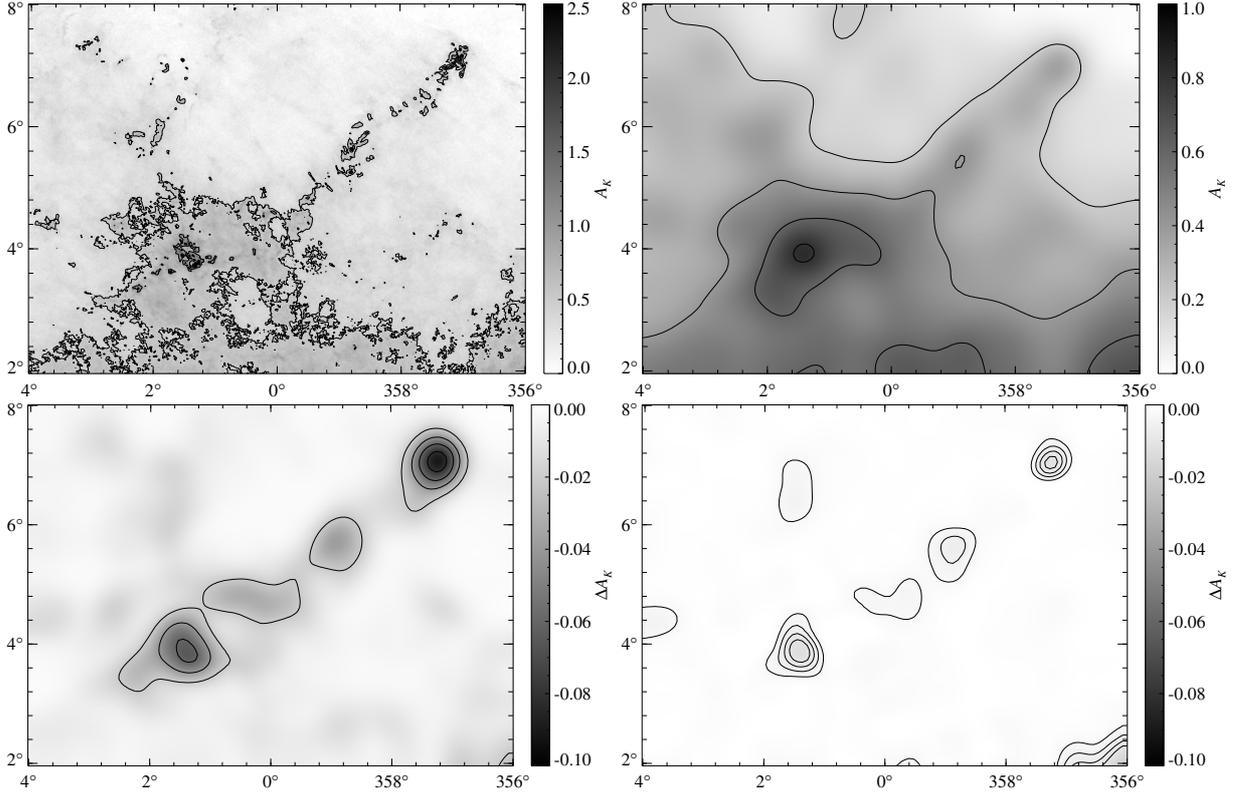
$$\hat{J}_n = J_n + k_K A_K(\mathbf{x}_n) + \varepsilon_n^{(J)}, \quad (36)$$

$$\hat{H}_n = H_n + k_H A_K(\mathbf{x}_n) + \varepsilon_n^{(H)}, \quad (37)$$

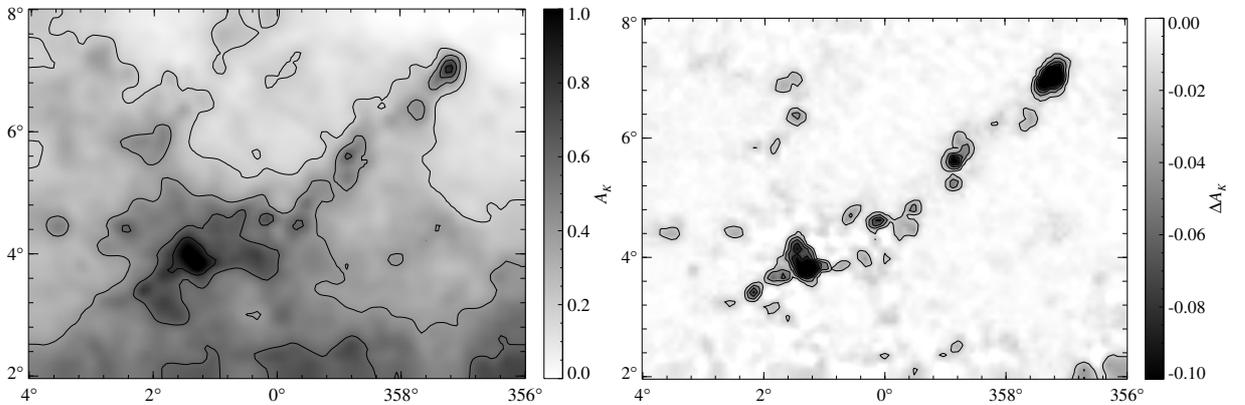
$$\hat{K}_n = K_n + A_K(\mathbf{x}_n) + \varepsilon_n^{(K)}, \quad (38)$$

where  $\varepsilon_n^{(J,H,K)}$  denote random variables used to model the photometric errors of the stars, and where as usual we used the hat to denote measured quantities. For simplicity, here, we took the errors as normal distributed random variables with constant variance 0.05 mag in all bands (the typical median variance of 2MASS magnitudes); moreover, we assumed a hard completeness limit at 15mag in all bands (i.e. if a magnitude exceeded 15 we took the star as not detected in the corresponding band). We then applied the NICER algorithm to these data, thus obtaining, for each star, its measured extinction and related error. Finally, we produced smooth extinction maps by interpolating the individual extinction measurements with three different techniques: (1) a simple moving average using Eq. (1); (2) a modified moving average using NICEST (Eq. (34)); (3) a nearest neighbour interpolation with a single star ( $N = 1$ ).

We repeated the whole procedure 1000 times, using each time a different set of random background stars, and took the average maps  $\hat{A}(\mathbf{x})$  obtained with the three techniques. We then compared these average maps with similar maps obtained from uniformly distributed stars (in order to produce such maps, we applied the completeness limit to un-extinguished magnitudes). Figures 5 and 6 show the results obtained, which support NICEST. In particular, both the simple moving average and the nearest



**Fig. 5.** *Top left:* the original 2MASS/NICER map of the Pipe nebula used in the simulations. *Top right:* the mean reconstructed map that one would obtain if the density of background stars were uniform. This is basically a smoothed version of the original map on the left. Levels are spaced at 0.2 mag. *Bottom left:* the difference between the average reconstructed maps, calculated from background stars following the density of Eq. (2), and the average map at *top right*. Levels are spaced at 0.02 mag. *Bottom right:* same map as *bottom left*, but using the improved method presented in this paper. Levels are spaced at 0.0025 mag.



**Fig. 6.** *Left:* average reconstructed map using a 1-star ( $N = 1$ ) Voronoi method, assuming a uniform distribution of stars. Levels are spaced at 0.2 mag. *Right:* difference between the average reconstructed Voronoi map, obtained from background stars following the density of Eq. (2), and the average true map to the left. Levels are spaced at 0.02 mag.

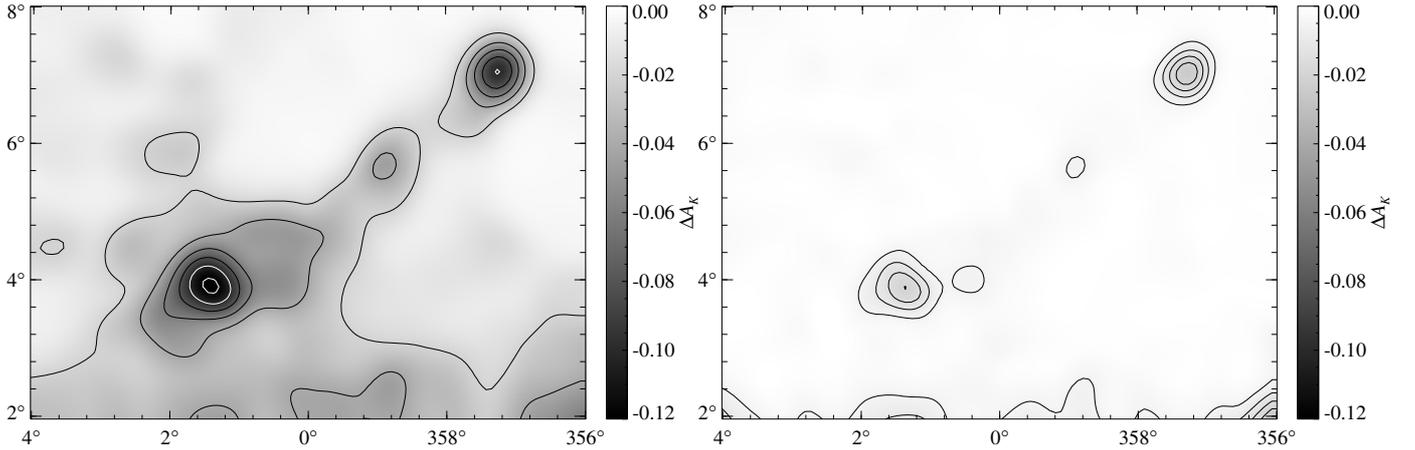
neighbor suffer from significant biases, up to  $A_K \approx 0.1$  mag or above, particularly in the most dense regions of the Pipe nebula. As shown in various papers (e.g. Lada et al. 1994; Lombardi et al. 2006, 2008), the amount of small scale structure present in dark molecular clouds increases with the column density, and thus it is not surprising that larger biases are observed there. In contrast, the method presented in this paper reduces the bias by  $\sim 10$ , a factor comparable to the weight number of stars  $N$  (which in our simulations is  $N \approx 10$  in the higher column density regions).

The simulation performed here allows us also to evaluate the average quadratic difference between the extinction map

obtained for the various methods  $\hat{A}(\mathbf{x})$  and the true (smoothed) map  $\bar{A}(\mathbf{x})$ , i.e. the quantity

$$\langle [\hat{A}(\mathbf{x}) - \bar{A}(\mathbf{x})]^2 \rangle = \text{Var}[\hat{A}(\mathbf{x})] + \langle [\hat{A}(\mathbf{x}) - \bar{A}(\mathbf{x})]^2 \rangle. \quad (39)$$

As shown by the above equation, the quantity considered here can be written as the sum of the variance and the square of the bias of the extinction map. Our simulations show that, although NICEST, as expected, has a slightly greater variance than NICER, there is still a gain by a factor of at least two in the squared difference of Eq. (39). In other words, the significant reduction in the bias performed by NICEST largely compensates for the



**Fig. 7.** Same as Fig. 5 (bottom), but with a fraction of  $f = 0.05$  of foreground stars. The increases of relative number of foreground stars in the denser regions of the cloud makes the bias more pronounced. Still, NICEST performs very well. Contours levels are spaced at 0.02 mag in the NICER plot (left) and at 0.005 mag in the NICEST plot (right); contours at  $-0.1$  mag or below are plotted in white.

increase in the scatter introduced by this method. A naive comparison between the results obtained from NICEST and from the Voronoi method would provide even larger differences in favour of NICEST, but we stress that since the Voronoi method interpolates the extinctions over a fixed number of stars (typically smaller than the one employed by NICEST in our simulations), a direct comparison is not possible.

Figure 7 shows the result of simulation carried out with the presence of foreground stars. In particular, we generated stars following the same prescriptions described in the previous paragraph, but allowing for a fraction  $f_0 = 0.05$  of foreground objects. Because of the effects of extinction, the effective fraction of foreground objects increases in the most dense regions of the cloud, which usually are also the ones with the larger substructures: the two biases thus add up and can be particularly severe. As shown by Fig. 7, NICEST is effectively able to cope with relatively large fractions of foreground stars while still providing a virtually unbiased column density estimate.

## 5. A sample application

The method presented in this paper was finally applied to the whole 2MASS point source catalog for the Pipe nebula region. We used the 4.5 million stars of the 2MASS catalog located in the window

$$-4^\circ < l < 4^\circ, \quad +2^\circ < b < +8^\circ. \quad (40)$$

The analysis was carried out following the prescriptions of Lombardi & Alves (2001), but using the modified estimator (34) to evaluate  $\hat{A}$ . In particular, we generated the final extinction map, shown in Fig. 8, on a grid of about  $1000 \times 750$  points, with scale of 30 arcsec per pixel, and with a Gaussian smoothing characterized by  $FWHM = 1$  arcmin. The slope of the number counts was estimated to be  $\alpha = 0.32 \pm 0.02$  in the  $H$  band.

The final, *effective* density of stars of about 8 stars per pixel guarantees that the approximation used to derive the unbiased estimator (34) is valid, and that a significant improvement over the standard NICER method can be expected. The highest extinction was measured close to Barnard 59, where  $\hat{A} \approx 2.68$  mag (a value that is 0.41 mag higher than what was obtained in Lombardi et al. 2006).

Figures 9 and 10 show a comparison between dense regions mapped using NICER and NICEST. We note that, as expected, the

two methods are equivalent in low-density regions, while the new one consistently estimates higher column densities as  $A$  increases. The same effect can be appreciated more quantitatively from Fig. 11, where we plot the relationship between the average estimates  $\hat{A}$  obtained in Lombardi et al. (2006) and here.

Finally, we argue that the plot of Fig. 11 is strongly related to Fig. 9 of Lombardi et al. (2006) where we show the increase in the scatter of  $\hat{A}_n$  for the different stars used at each point of the extinction map as a function of the average  $\hat{A}$  at the point. This  $A$ - $\sigma$  relation is most likely due to unresolved substructures in the cloud that are expected to be more prominent in high-density regions. Recently, we reconsidered the  $A$ - $\sigma$  relation and defined a quantity called  $\Delta^2(\mathbf{x})$  (Lombardi et al. 2008). This quantity is simply related to the local scatter of measured extinctions, but also properly takes into account the contribution of measurement errors to the observed scatter. Interestingly, this quantity can be defined in terms of simple observables, and can be shown to be directly related to the local scatter of column densities:

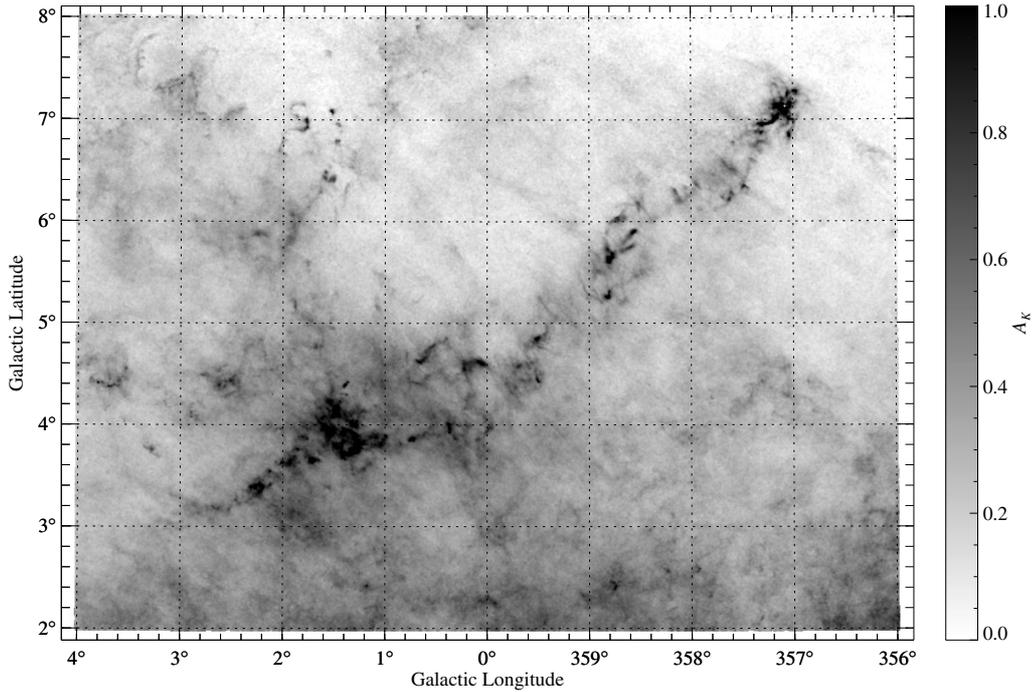
$$\Delta^2(\mathbf{x}) = \frac{\sum_n w_n \Delta_n^2}{\sum_n w_n}, \quad (41)$$

where  $\Delta_n \equiv A(\mathbf{x}_n) - \bar{A}(\mathbf{x})$  is the difference between the column extinction at the position of the  $n$ -th star and the local average column extinction. A comparison of Eq. (41) with Eq. (8) clearly shows that, except for a numerical factor  $\beta$ , the bias expected in the standard NICER method is simply proportional to the  $\Delta^2(\mathbf{x})$  of Lombardi et al. (2008).

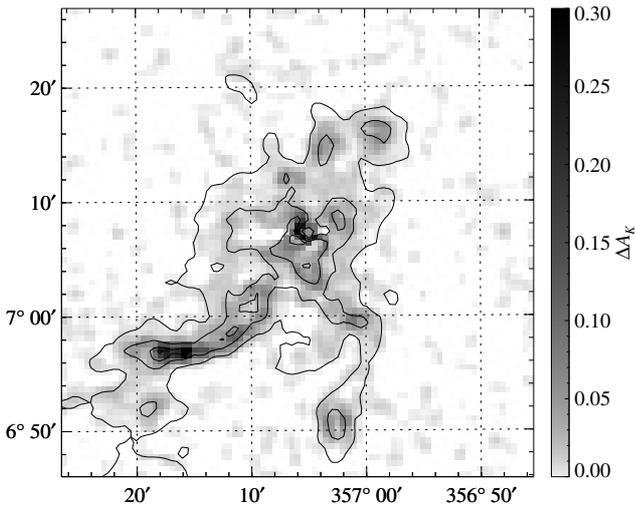
## 6. Discussion

The numerical simulations and a first sample application of NICEST have shown that the method presented in this paper can significantly alleviate the bias introduced by small-scale structures and by foreground stars in extinction studies. Of course, NICEST too has some limitations, which however are largely unavoidable (and thus inherent to any extinction-based method).

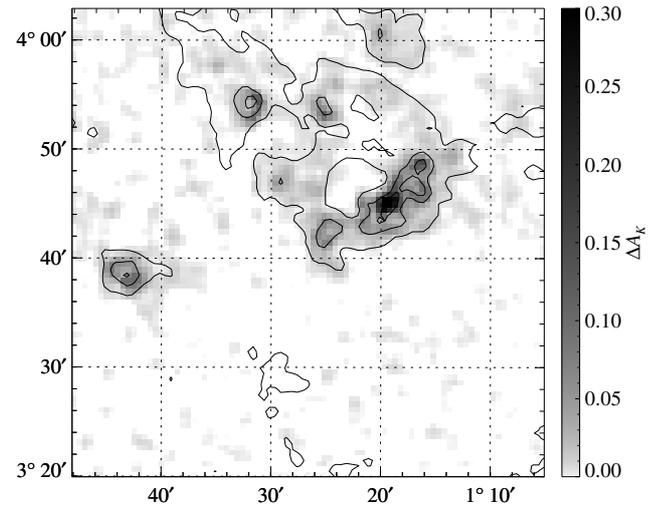
First, we note that in order for NICEST to work effectively, the weight number of background stars  $\mathcal{N}$  must be significantly greater than unity: as shown in Sect. 3.2,  $\mathcal{N}$  is directly related to the reduction in the bias provided by NICEST with respect to NICER, and thus having  $\mathcal{N} \sim 1$  does not give any benefit. This point is also important when correcting for foreground star contamination. For example, if  $\mathcal{N} \sim 1$  and the local fraction  $f$  of



**Fig. 8.** The NICEST extinction map of the Pipe nebula, using the modified estimator (28).



**Fig. 9.** The difference between the modified extinction estimates (Eq. (28)) and the standard ones [Eq. (1)] around Barnard 59. The contour levels are at  $A = \{0.5, 1.0, 1.5, 2.0\}$  mag of the map of Fig. 8.



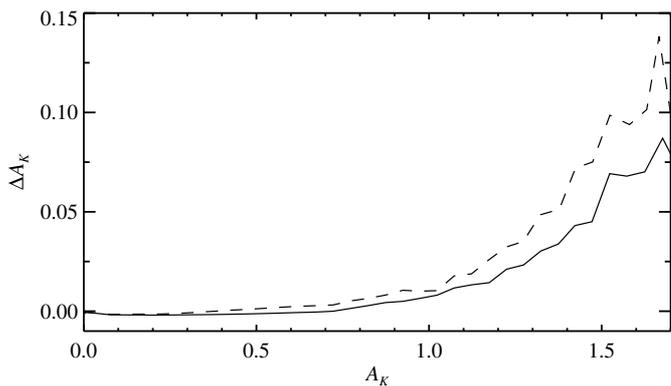
**Fig. 10.** The difference between the modified extinction estimates (Eq. (28)) and the standard ones (Eq. (1)) around the peak ID 3 of Lombardi et al. (2006). The contour levels are at  $A = \{0.5, 1.0, 1.5, 2.0\}$  mag of the map of Fig. 8.

foreground stars is large (e.g., because we are in a particularly dense core), on average we do not expect any background stars, and we will thus consistently measure a vanishing extinction. Hence, NICEST, like any other extinction-based method, only works if there are a sizable number of background stars that can be used for reliable extinction measurements.

The above point might be interpreted as an exceedingly stringent requirement for the smoothing window used in NICEST. In fact, if a weight function  $w(\mathbf{x}'; \mathbf{x}) = w(\mathbf{x}' - \mathbf{x})$  invariant upon translation is used, this function should be taken broad enough to guarantee that the weight number of background stars  $\mathcal{N}(\mathbf{x})$  is significantly larger than unity everywhere. This, in turn, would imply that  $\mathcal{N} \gg 1$  in the intermediate extinction regions, i.e. that the extinction map in these regions has a poor resolution, well below the limits imposed by the density of background stars. In

reality, the whole derivation of the statistical properties of the NICEST technique is still valid for weight functions which are not spatially invariant. Hence, one does not need to use a fixed window size for  $w(\mathbf{x}'; \mathbf{x})$ , but rather this function could be taken to change shape for different locations  $\mathbf{x}$ . For example, a simple scheme could be the use of a Gaussian shape for  $w(\mathbf{x}'; \mathbf{x})$  with the typical scale chosen according to a local estimate of the density of background stars, in a way such that  $\mathcal{N}(\mathbf{x}) \sim \text{const} \sim 10$ . This choice would guarantee an optimal resolution everywhere, and would still allow one to make use of the NICER technique to reduce the inhomogeneity bias.

Another point to keep in mind is that NICEST gives a large weight to red sources. This has two potential unwanted consequences. First, it introduces a small bias, that has been corrected



**Fig. 11.** The average difference  $\Delta A$  between the estimators (28) and (1) as a function of the estimated column density  $A$ . The lower, solid line is relative to the Pipe nebula map at  $FWHM = 1$  arcmin resolution; the upper dashed line is at for a resolution  $FWHM = 1.5$  arcmin. The latter plot is essentially what we would obtain at a  $FWHM = 1$  arcmin if the distance of the Pipe nebula increases by a factor 1.5.

to first order in Eq. (34). Second, as shown also by the numerical simulations described in Sect. 5, it slightly increases the noise in the final map. Again, since the noise in final map  $\hat{A}(x)$  is proportional to  $1/N$ , the solution is to make sure that  $N$  is sufficiently large. We believe that a small increase in the noise is a fair price to pay for the significant reduction in the bias provided by NICEST over NICER (see also Eadie et al. 1971).

A potentially more severe problem can be young stellar objects (YSOs) with infrared excess. These sources can be present in the most dense star-forming cores of molecular clouds, where they can severely bias our method. We note, however, that since these sources have peculiar colors and are not usually present in the control field used for the calibration, they represent a problem for all extinction based methods, including NICER and the nearest neighbour ones. A median estimator is in principle safer to use in these cases, as long as the YSOs present in the core are a minority of the total number of background stars observed in the region; unfortunately, the tendency of YSOs to appear in clusters does not help here (note also that a median is usually more noisy than the simple average). The obvious solution is thus to exclude as much as possible YSOs from the list of sources used in the extinction maps.

Finally, let us briefly mention alternative possibilities to reduce the bias considered in this paper. Since the substructure bias is due to a relationship between the local density  $\rho(x)$  and the extinction map  $A(x)$  (Eq. (2)), one could try to use a tracer of the local density to perform the correction. We already discussed a method based on this concept, the nearest neighbours interpolation. However, as shown in Sect. 4, this method fails (see also Sect. 2.2). In general, a problem with methods based on the local estimate of the density of background stars is that one needs to be able to obtain  $\rho(x)$  on scales smaller than the ones that characterize the weight  $w(x'; x)$ , or otherwise it is not possible to give different weights to the different stars that contribute significantly to the weighted average of Eq. (1). However, if the density of background stars is large enough to allow accurate measurements of the local density  $\rho(x)$ , one can disregard substructures, because it is always possible to make extinction maps at the same resolution as the density map. Hence, the only effective way to deal with substructures involves the use of the local extinction measurements, as done in NICEST. Note also that any density-based correction will be completely ineffective with foreground stars, in contrast to the method presented here.

## 7. Conclusions

The main results obtained in this paper can be summarized in the following points:

- We discussed the effects of small scale inhomogeneities in the NIR extinction maps based on color excess methods, and showed that large inhomogeneities can significantly bias standard extinction maps toward low column densities.
- We proposed a new estimator for  $\hat{A}$ , NICEST, and we showed that it is (i) equivalent to the usual estimator in low-density regions; (ii) *unbiased* and *robust* for any value of  $A$ , i.e. insensitive to the actual form and amount of substructure present in the cloud.
- We showed that the new estimator is also suitable in the presence of contamination by foreground stars.
- We tested NICEST against numerical simulations, and showed that it effectively reduces by a large factor the biases due to substructures and to foreground stars. We also tested an alternative approach, the nearest neighbor, and showed that the results obtained from this interpolation are still severely biased.
- We applied NICEST to the Pipe nebula and showed a few preliminary properties of the resulting extinction map. We also noted a direct connection between the bias of the NICER and the  $\Delta^2$  map defined by (Lombardi et al. 2008).

*Acknowledgements.* We thank J. Alves, C. J. Lada, and G. Bertin for stimulating discussions and suggestions, and the referee, L. Cambresy, for helping us improve the paper. This research has made use of the 2MASS archive, provided by NASA/IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

## References

- Alves, J., Lada, C. J., & Lada, E. A. 2001, *Nature*, 409, 159
- Alves, J., Lombardi, M., & Lada, C. 2008, in *Handbook of low mass star formation regions*, ed. B. Reipurth, ASP Conf. Ser.
- Bok, B. J. 1937, *The distribution of the stars in space* (Chicago: University of Chicago Press)
- Cambr esy, L., Beichman, C. A., Jarrett, T. H., & Cutri, R. M. 2002, *AJ*, 123, 2559
- Cambr esy, L., Jarrett, T. H., & Beichman, C. A. 2005, *A&A*, 435, 131
- Cambr esy, L., Petropoulou, V., Kontizas, M., & Kontizas, E. 2006, *A&A*, 445, 999
- Dobashi, K., Bernard, J.-P., Hughes, A., et al. 2008, *A&A*, 484, 205
- Eadie, W., Drijard, D., James, F., Roos, M., & Sadoulet, B. 1971, *Statistical Methods in Experimental Physics* (Amsterdam, New-York, Oxford: North-Holland Publishing Company)
- Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, *ApJ*, 262, 590
- Heyer, M. H., & Brunt, C. M. 2004, *ApJ*, 615, L45
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, *ApJ*, 619, 931
- Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, *ApJ*, 429, 694
- Larson, R. B. 1981, *MNRAS*, 194, 809
- Lombardi, M. 2002, *A&A*, 395, 733
- Lombardi, M. 2005, *A&A*, 438, 169
- Lombardi, M., & Alves, J. 2001, *A&A*, 377, 1023
- Lombardi, M., & Schneider, P. 2001, *A&A*, 373, 359
- Lombardi, M., & Schneider, P. 2002, *A&A*, 392, 1153
- Lombardi, M., & Schneider, P. 2003, *A&A*, 407, 385
- Lombardi, M., Alves, J., & Lada, C. J. 2006, *A&A*, 454, 781
- Lombardi, M., Lada, C. J., & Alves, J., 2008, *A&A*, 489, 143
- Mathis, J. S. 1990, *ARA&A*, 28, 37
- Onishi, T., Kawamura, A., Abe, R., et al. 1999, *PASJ*, 51, 871
- Rieke, G. H., & Lebofsky, M. J. 1985, *ApJ*, 288, 618
- V azquez-Semadeni, E., & Garc a, N. 2001, *ApJ*, 557, 727
- Wilson, R. W., Jefferts, K. B., & Penzias, A. A. 1970, *ApJ*, 161, L43
- Wolf, M. 1923, *Astron. Nachr.*, 219, 109