THE ASTROPHYSICAL JOURNAL, 536:571–583, 2000 June 20

BAYESIAN PHOTOMETRIC REDSHIFT ESTIMATION

NARCISO BENÍTEZ
Astronomy Department, University of California at Berkeley, 601 Campbell Hall, Berkeley, CA 94720-5030; benitezn@mars.berkeley.edu
Received 1998 November 11; accepted 2000 January 26

ABSTRACT

Photometric redshifts are quickly becoming an essential tool of observational cosmology, although their utilization is somewhat hindered by certain shortcomings of the existing methods, e.g., the unreliability of maximum-likelihood techniques or the limited application range of the "training-set" approach. The application of Bayesian inference to the problem of photometric redshift estimation effectively overcomes most of these problems. The use of prior probabilities and Bayesian marginalization facilitates the inclusion of relevant knowledge, such as the expected shape of the redshift distributions and the galaxy type fractions, which can be readily obtained from existing surveys but are often ignored by other methods. If this previous information is lacking or insufficient—for instance, because of the unprecedented depth of the observations—the corresponding prior distributions can be calibrated using even the data sample for which the photometric redshifts are being obtained. An important advantage of Bayesian statistics is that the accuracy of the redshift estimation can be characterized in a way that has no equivalents in other statistical approaches, enabling the selection of galaxy samples with extremely reliable photometric redshifts. In this way, it is possible to determine the properties of individual galaxies more accurately, and simultaneously estimate the statistical properties of a sample in an optimal fashion. Moreover, the Bayesian formalism described here can be easily generalized to deal with a wide range of problems that make use of photometric redshifts. There is excellent agreement between the ≈ 130 Hubble Deep Field North (HDF-N) spectroscopic redshifts and the predictions of the method, with a rms error of $\Delta z \approx 0.06(1 + z_{\rm spec})$ up to $z < 6$ and no outliers nor systematic biases. It should be remarked that since these results have not been reached following a training-set procedure, the above value of Δz should be a fair estimate of the expected accuracy for any similar sample. The method is further tested by estimating redshifts in the HDF-N but restricting the color information to the *UBVI* filters; the results are shown to be significantly more reliable than those obtained with maximum-likelihood techniques.

Subject headings: galaxies: distances and redshifts — galaxies: photometry — methods: statistical

1. INTRODUCTION

The advent of the new class of 10 m ground-based telescopes is having a strong impact on the study of galaxy evolution. Instruments such as LRIS at the Keck telescopes allow observers to regularly secure redshifts for dozens of $I \approx 24$ galaxies in several hours of exposure. Technical advances in the instrumentation, combined with the proliferation of 10 m class telescopes, guarantees a vast increase in the number of galaxies, bright and faint, for which spectroscopic redshifts will be obtained in the near future. In spite of the progress in the sheer numbers of available spectra, the $I \approx 24$ "barrier" (for reasonably complete samples) is likely to stand for awhile yet, since there are no foreseeable dramatic improvements in the telescope area or detection techniques. This means that most of the galaxies detected in very deep exposures are in practice inaccessible to spectroscopical analysis. The best example is the Hubble Deep Field North (HDF-N; Williams et al. 1996): after several years of intensive efforts by the astronomical community, the spectroscopical sample only comprises $\approx 20\%$ of the $I < 27$ galaxies detected in that field. Very few areas of the sky will receive such a telescope barrage in the near future, so this can almost be considered as the limit of what is currently achievable by spectroscopy. In contrast, surprisingly accurate photometric redshifts were quickly obtained for most of the HDF-N galaxies (notably by Sawicki, Lin, & Yee 1997; see also Lanzetta, Yahil, & Fernández-Soto 1996; Gwyn & Hartwick 1996), although due to the maximum-likelihood methodology employed by these authors, a significant fraction of redshift estimates presented large, "catastrophic" errors (Ellis 1997). Moreover, as will be shown below, using a Bayesian statistical approach it is possible to obtain fast, inexpensive and—more important—highly *reliable* photometric redshifts for $\approx 90\%$ of the $I < 27$ HDF-N galaxies.

In spite of the efforts of Thomas Loredo, who has written stimulating reviews on the subject (Loredo 1990, 1992), Bayesian inference is still far from becoming the standard approach in astrophysics, and is often used as just another tool in the available panoply of statistical methods. However, as any reader of the fundamental treatise by Jaynes (2000)[1] can learn, Bayesian probability theory represents a unified look to probability and statistics, which intends not to complement, but rather to fully replace the traditional "frequentist" statistical techniques (see also Bretthorst 1988, 1990; Sivia 1996; Gelman et al. 1998). The basic aim of this paper is to consider the problem of photometric redshift estimation from the point of view of Bayesian inference. Kodama, Bell, & Bower (1998) also developed a Bayesian classifier for photometric redshifts. There are, however, several differences between their approach and the one followed in this work, perhaps the most significant of which is the treatment of priors (see § 3.3).

The outline of the paper is as follows. Section 2 reviews the current methods of photometric redshift estimation,

[1] Jaynes (2000) is also available at: http://bayes.wustl.edu/etj/prob.html.

pointing out their main sources of error. Section 3 describes in detail how to apply Bayesian probability to photometric redshift estimation. Section 4 compares the performance of traditional statistical techniques, such as maximum likelihood, with Bayesian photometric redshift (BPZ) estimation, by applying both methods to the HDF-N spectroscopic sample and to a simulated catalog. Section 5 briefly summarizes the main conclusions of the paper.

2. PHOTOMETRIC REDSHIFTS: TRAINING SET VERSUS SED-FITTING METHODS

There are two basic approaches to photometric redshift estimation; using the terminology of Yee (1998), they may be called "spectral energy distribution (SED)–fitting" and "empirical training set" methods. The first technique (Koo 1985; Lanzetta et al. 1996; Gwyn & Hartwick 1996; Pelló et al. 1996; Sawicki et al. 1997, etc.) begins by compiling a library of template spectra, empirical or generated with population synthesis techniques. These templates, after being redshifted and corrected for intergalactic extinction, are compared with the galaxy colors to determine the redshift z that best fits the observations. The training-set technique (Brunner et al. 1997; Connolly et al. 1995; Wang, Bahcall, & Turner 1998) starts with a multicolor galaxy sample with apparent magnitudes m_0 and colors C that have been spectroscopically identified. Using this sample, a relationship of the kind $z = z(C, m_0)$ is determined using a multiparametric fit.

It should be said that these two methods are more similar than they are usually thought to be. To understand this, let us look in more detail at how they work. For simplicity, we forget about the magnitude dependence and assume that only two colors, $C = (C_1, C_2)$, are enough to estimate the photometric redshifts. Thus, given a set of spectroscopic redshifts $\{z_{\rm spec}\}$ and colors $\{C\}$, the training-set method will try to fit a surface $z = z(C)$ to the data. This is based on a very strong assumption: that the surface $z = z(C)$ is a *function* defined on the color space, where each value of C corresponds to one and only one redshift. Visually, this means that the surface $z = z(C)$ does not "bend" over itself in the redshift direction. Although this functionality of the redshift/color relationship cannot be taken for granted in the general case, it seems to be a good approximation to the real picture at $z < 1$ redshifts and bright magnitudes (Brunner et al. 1997). A certain scatter around this surface is allowed: galaxies with the same value of (C) may have slightly different redshifts, and it seems to be assumed implicitly that this scatter is the factor limiting the accuracy of the method.

The SED-fitting method is based on the color/redshift relationships generated by each of the library templates T, $C_T = C_T(z)$. A galaxy at the position C is assigned the redshift corresponding to the closest point of any of the C_T curves in the color space. If these C_T functions are inverted, one ends up with the curves $z_T = z_T(C_T)$, which in general are not functions, since they may present self-crossings (and of course intersect each other as well). If we limit ourselves to the region in the color/redshift space in which the training-set method defines the surface $z = z(C)$, for a realistic template set, the curves $z_T = z_T(C_T)$ would be embedded in the surface $z = z(C)$, conforming its "skeleton" and defining its main features.

The fact that the surface $z = z(C)$ is continuous, whereas the template-defined curves are sparsely distributed, does not make a great difference. The gaps may be filled by finely interpolating between the templates (Sawicki et al. 1997), but this is not strictly necessary: usually, the statistical procedure employed to search for the best redshift performs its own interpolation between templates. When the colors of a galaxy do not exactly coincide with one of the spectra, χ^2 or the maximum-likelihood method will assign the redshift corresponding to the nearest template in the color space. This is equivalent to the curves $z_T = z_T(C_T)$ having "influence areas" around them, conforming a sort of step-like surface that interpolates across the gaps and also extends beyond the region limited by them in the color space. Therefore, the SED-fitting method comes with a built-in interpolation (and extrapolation) procedure, and for this reason, the accuracy of the photometric is not dramatically improved by finely interpolating between sparse spectra (see § 4).

The intrinsic similarity of the two photometric redshift methods may explain their comparable performance, especially at $z \lesssim 1$ (Hogg et al. 1998). For magnitude ranges with a relatively simple color-redshift topology, the training-set method should perform better, if only because it avoids the possible systematics due to mismatches between the predicted template colors and the real ones, and also in part because it includes not only the colors of the galaxies but also their magnitudes, which helps to break the color/redshift degeneracies (see below). It should not be forgotten, however, that despite its apparent precision ($\delta z \approx 0.06$ for the HDF-N; Connolly et al. 1997), by its own nature there is not a strong guarantee that such an accuracy will be reached in future samples, even within the same magnitude and redshift ranges. Nevertheless, this could be a good approach for large, low- to moderate-redshift surveys with abundant calibration spectroscopy, such as the Sloan Digital Sky Survey (Gunn & Weinberg 1995).

However, in most cases the training-set approach is impractical or even unfeasible. A basic flaw of this method is the assumption of a single functional form for the color-redshift relationship, since it is obvious that as one goes to higher redshifts and/or fainter magnitudes, the topology of the color-redshift distribution $z = z(C, m_0)$ displays several nasty degeneracies, even if the near-IR information is included. This problem can be somewhat overcome by dividing the redshift/color space into several regions and performing piecemeal fittings within each of them (Wang et al. 1998). Another problem is that because of its empirical and ad hoc basis, it can only be reliably extended as far as the spectroscopic redshift limit. This means that it cannot be applied where it is more needed, to study faint galaxy samples beyond the reach of the spectrograph. Moreover, it is not straightforward to transfer the calibration obtained with a given filter set to a different one. Such an extrapolation can be done with the help of templates, but once the shape of these templates has been determined, it seems more convenient to switch to the more versatile SED-fitting method, especially in its Bayesian version, which will be developed below.

Although the SED-fitting method is not affected by some of these limitations, it also comes with its own set of problems. Several authors have analyzed in detail the main sources of errors affecting this method (Sawicki et al. 1997; Fernández-Soto et al. 1999), which can be divided into two broad classes: color/redshift degeneracies and template incompleteness.

Figure 1 (*left*) shows $V-I$ versus $I-K$ for the morphological types employed in § 4 and $0 < z < 5$. The color/redshift degeneracies happen when the line corresponding to a single template self-intersects or when two lines cross each other at a point corresponding to different redshifts [these cases correspond to "bendings" in the redshift/color relationship $z = z(C)$]. It is obvious that the frequency of such crossings will rise with the extension of the considered redshift range and with the number of templates included. Moreover, the presence of color/redshift degeneracies is also increased by random photometric errors, which can be visualized as a blurring or thickening of the $C_T(z_T)$ relationship (Fig. 1, *right*): each point of the curves in the left panel of Figure 1 is expanded into a square of size δC, the error in the measured color. The first consequence of this is a "continuous" [$\delta z \approx (\partial C/\partial z)\delta C$] increase in the rms of the "small-scale" errors in the redshift estimation. Worse still, the overlaps in the color-color space become more frequent, with a corresponding rise in the number of "catastrophic" redshift errors. Multicolor information may often be redundant, so increasing the number of filters does not necessarily break the degeneracies. For instance, by applying a simple PCA analysis to the HDF-N photometric sample, it can be shown that the information contained in the seven *UBVIJHK* filters for the HDF galaxies can be condensed using only three parameters, the coefficients of the principal components of the flux vectors (see also Connolly et al. 1995). Therefore, if the photometric errors are large, increasing the number of filters cannot totally eliminate the color/redshift degeneracies, which makes them almost unavoidable in faint galaxy samples. The training-set method somewhat alleviates this problem by introducing an additional parameter in the estimation: the magnitude, which in some cases breaks the degeneracy. However, color/redshift degeneracies may also affect galaxies with the same magnitude, and the training-set method does not even contemplate their possibility.

The SED-fitting method at least allows for the existence of this problem, although it is not very efficient in dealing with it, especially with noisy data. Its choice of redshift is exclusively based on the goodness of fit between the observed colors and the templates. In cases such as the one described above, where two or more redshift/morphological type combinations have nearly the same colors, the value of the likelihood $\mathscr{L}$ would have two or more approximately equally high maxima at different redshifts (see Fig. 2). Depending on the random photometric error, one maximum would prevail over the others, and a small change in the flux could involve a catastrophic change in the estimated redshift (see Fig. 2). However, in many cases there is additional information, discarded by the maximum likelihood (ML) method, that could potentially help to

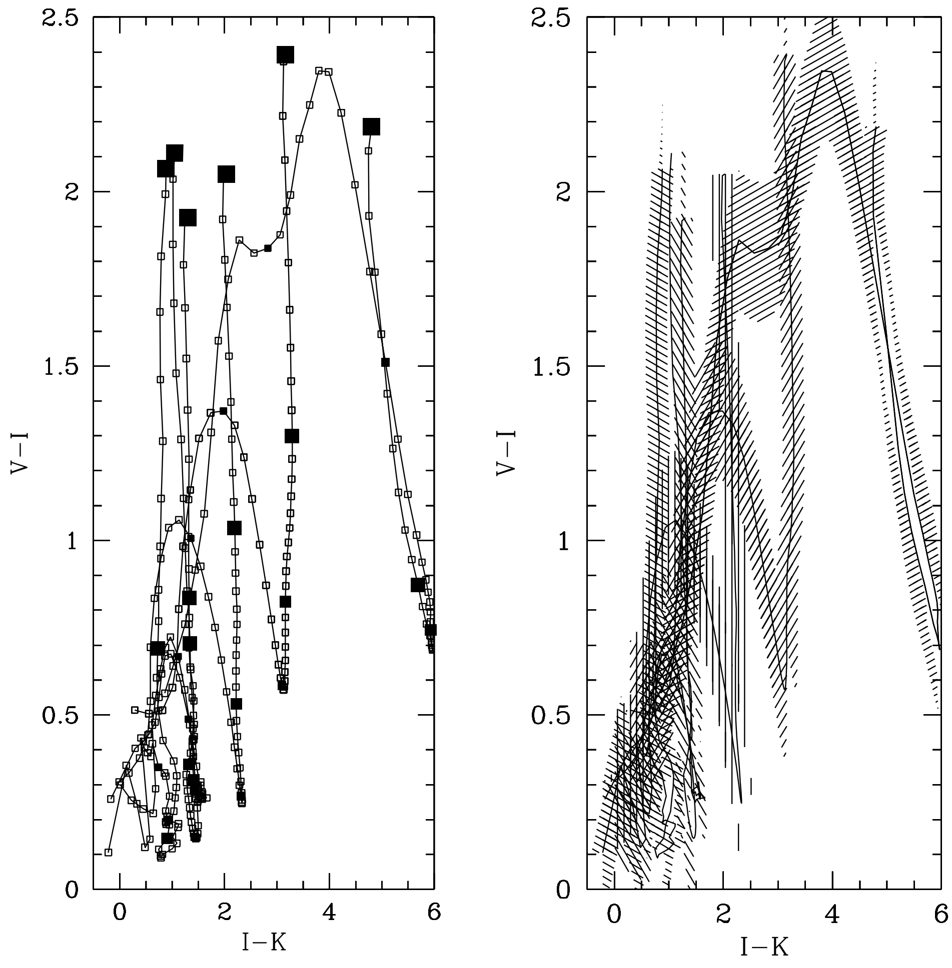

FIG. 1.—*Left*: $V-I$ vs. $I-K$ for the templates used in § 4 in the interval $1 < z < 5$. The size of the filled squares grows with redshift, from $z = 1$ to $z = 5$. If these were the only colors used for the redshift estimation, every crossing of the lines would correspond to a color/redshift degeneracy. *Right*: The same color-color relationships, "thickened" by a 0.2 photometric error. The probability of color/redshift degeneracies increases greatly.

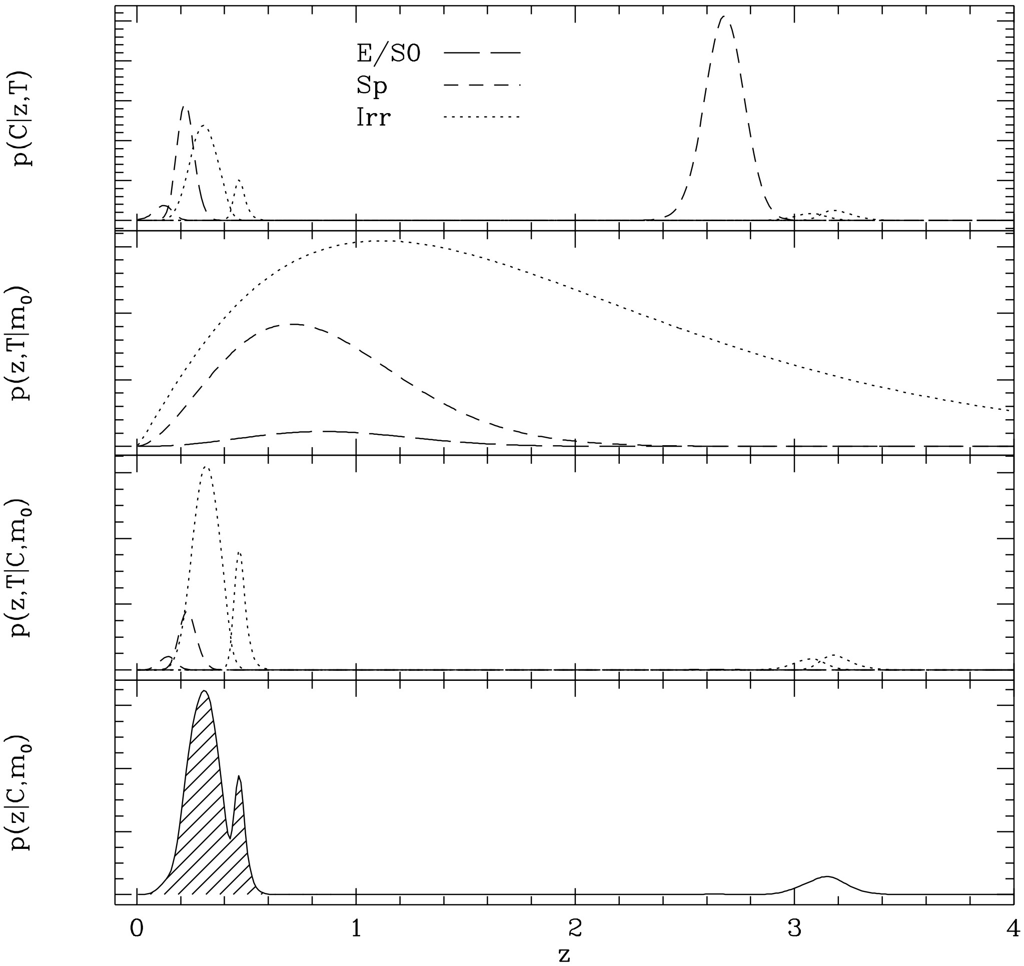

FIG. 2.—Example of the main probability distributions involved in BPZ for a galaxy at $z = 0.28$ with an Irr spectral type and $I \approx 26$, to which random photometric noise is added. From top to bottom: (*a*): Likelihood functions $p(C \mid z, T)$ for the different templates used in § 4. Based on ML, the redshift chosen for this galaxy would be $z_{\rm ML} = 2.685$, and its spectral type would correspond to a spiral. (*b*): Prior probabilities, $p(z, T \mid m_0)$, for each of the spectral types (see text). Note that the probability of finding a spiral spectral type with $z > 2.5$ and a magnitude $I = 26$ is almost negligible. (*c*) Probability distributions, $p(z, T \mid C, m_0) \propto p(z, T \mid m_0)p(C \mid z, T)$, that is, the likelihoods in the top plot multiplied by the priors. The high-redshift peak due to the spiral has disappeared, although there is still a small chance of the galaxy being at high redshift if it has a Irr spectrum, but the main concentration of probability is now at low redshift. (*d*) Final Bayesian probability, $p(z \mid C, m_0) = \sum_T p(z, T \mid C, m_0)$, which has its maximum at $z_b = 0.305$. The shaded area corresponds to the value of $p_{\Delta z}$, which estimates the reliability of z_b and yields a value of ≈ 0.91.

solve such conundrums. For instance, it may be known from previous experience that one of the possible redshift/type combinations is much more likely than any other, given the galaxy magnitude, angular size, shape, etc. In that case, and since the likelihoods are not informative enough, Bayesian probability states that the best option would be the one more likely a priori. This is plain common sense, but it is not easy to implement using ML; at best, one can modify the redshift of the problematic objects by hand or devise ad hoc solutions for each case. In contrast, Bayesian probability theory allows one to include this additional information in a rigorous and consistent way, effectively dealing with this kind of error (see § 3).

Although in some cases the spectrum of a galaxy has no close equivalents in the template library, it will be assigned by ML the redshift corresponding to the nearest template in the color/redshift space, no matter how distant it is from the observed color (and from the real redshift) in absolute terms. The solution to this problem seems obvious: to include more templates in the library until all the possible galaxy types are considered. However, since all the templates have equal status in ML, doing this increases the number of color/redshift degeneracies. Bayesian inference is much less affected by this problem, since it weights each template by its prior probability, and therefore templates corresponding to relatively uncommon types, such as, e.g., AGNs, etc., can be included without unduly disturbing the redshift estimation for normal galaxies.

As explained above, the SED-fitting techniques perform their own "automatic" interpolation and extrapolation, so once the main spectral types are included in the template library, the results are not greatly affected if one finely interpolates among the main spectra. The effects of using a correct but incomplete set of spectra are shown in § 4.

Both sources of errors described above are exacerbated for high-redshift galaxies, which are usually faint, and therefore have large photometric errors. Moreover, the color/redshift space has a very extended range in z, and thus degeneracies are more likely to appear; in addition, the template incompleteness is worse, since there are few or no empirical spectra with which to compare the template library.

The effectiveness of any photometric redshift method is established by contrasting its output with a sample of galaxies with spectroscopic redshifts. It should be kept in mind, however, that the results of this comparison may be misleading, since the available spectroscopic samples are almost by definition especially well suited to photometric redshift estimation, being relatively bright (and thus with small photometric errors) and often filling a privileged niche in the color-redshift space, far from degeneracies (e.g., Lyman-break galaxies). Thus, it is risky to extrapolate the accuracy reached by current methods as estimated from spectroscopic samples (this also applies to BPZ) to fainter magnitudes. This is especially true for the training-set methods, which deliberately minimize the difference between the spectroscopic and photometric redshifts.

3. BAYESIAN PHOTOMETRIC REDSHIFTS (BPZ)

Within the framework of Bayesian probability, the problem of photometric redshift estimation can be posed as finding the probability $p(z \mid D, I)$, i.e., the probability of a galaxy having redshift z given the data $D = \{C, m_0\}$, *and* the prior information I, which includes any knowledge relevant to the hypothesis under consideration not already contained in the data D. Although some authors recommend that the term $\mid I$ should not be dropped from the expressions of probability, here the rule of simplifying the mathematical notation whenever there is no danger of confusion will be followed, and from now on $p(z)$ will stand for $p(z \mid I)$, $p(D \mid z)$ for $p(D \mid z, I)$, etc.

As a trivial example, let us consider just one template in our library. Applying Bayes' theorem,

$$p(z \mid C, m_0) = \frac{p(z \mid m_0)p(C \mid z)}{p(C)} \propto p(z \mid m_0)p(C \mid z) . \quad (1)$$

Here the expression $p(C \mid z) \equiv \mathscr{L}(z)$ is simply the redshift likelihood: the probability of observing the colors C if the galaxy has redshift z. The probability $p(C)$ is a normalization constant, and usually there is no need to calculate it.

The first factor, the *prior* probability, $p(z \mid m_0)$, is the redshift distribution for galaxies with magnitude m_0. This function allows us to include information such as the existence of upper or lower limits on the galaxy redshifts, the presence of a cluster in the field, etc. The effect of the prior $p(z \mid m_0)$ on the estimation depends on how informative it is. It is obvious that for a constant prior (all redshifts equally likely a priori), the estimate obtained from equation (1) will exactly coincide with the ML result. This is also roughly true if the prior is "smooth" enough and does not present significant structure. However, in other cases, values of the redshifts that are considered very improbable from the prior information would be "discriminated;" i.e., they must fit the data much better than any other redshift in order to be selected.

Note that rigorously, one should write the prior in equation (1) as

$$p(z \mid m_0) \propto \int d\hat{m}_0\, p(\hat{m}_0)p(m_0 \mid \hat{m}_0)p(z \mid \hat{m}_0) , \quad (2)$$

where $\hat{m}_0$ is the "true" value of the observed magnitude m_0, $p(\hat{m}_0)$ is proportional to the number counts as a function of the magnitude m_0, and $p(m_0 \mid \hat{m}_0) \propto \exp[(m_0 - \hat{m}_0)^2/2\sigma_{m0}^2]$, i.e., the probability of observing m_0 if the true magnitude is $\hat{m}_0$. The above convolution accounts for the uncertainty in the value of the magnitude m_0, which has the effect of slightly "blurring" and biasing the redshift distribution $p(z \mid m_0)$. To simplify our exposition, this effect will not be consider hereafter; just $p(z \mid m_0)$ and its equivalents will be used.

3.1. *Bayesian Marginalization*

It may seem from equation (1) (and it is unfortunately quite a common misconception) that the only difference between Bayesian and ML estimates is the introduction of a prior; in this case, $p(z \mid m_0)$. However, there is more to Bayesian probability than priors.

The galaxy under study may belong to different morphological types, represented by a set of n_T templates. This set is considered to be *exhaustive*, i.e., including all possible types, and *exclusive*: the galaxy cannot belong to two types at the same time. In that case, using Bayesian marginalization, the probability $p(z \mid D)$ can be "expanded" into a basis formed by the hypothesis $p(z, T \mid D)$ (the probability of the galaxy redshift being z *and* the galaxy type being T). The sum over all these "atomic" hypothesis will give the total probability, $p(z \mid D)$. That is,

$$p(z \mid C, m_0) = \sum_T p(z, T \mid C, m_0) \propto \sum_T p(z, T \mid m_0)p(C \mid z, T) , \quad (3)$$

where we have applied Bayes' theorem in the second step. Here $p(C \mid z, T)$ is the probability of the data C given z and T (where it is assumed that it does not depend on the magnitude m_0). The prior $p(z, T \mid m_0)$ can be developed using the product rule,

$$p(z, T \mid m_0) = p(T \mid m_0)p(z \mid T, m_0) , \quad (4)$$

where $p(T \mid m_0)$ is the galaxy type fraction as a function of magnitude, and $p(z \mid T, m_0)$ is the redshift distribution for galaxies of a given spectral type and magnitude.

Equation (3) and Figure 2 clearly illustrate the main differences between the Bayesian and ML methods. ML would just pick the highest maximum over all the likelihoods $p(C \mid z, T)$ as the best redshift estimate, without looking at the plausibility of the corresponding values of z or T. On the contrary, Bayesian probability averages over all the likelihoods after weighting them by their prior probabilities, $p(z, T \mid m_0)$. In this way, the estimation is not affected by spurious likelihood peaks caused by noise (Fig. 2; see also the results of § 4). Of course, in an ideal situation with noiseless observations—and a nondegenerate color/redshift space—the results obtained with ML and Bayesian inference would be the same.

It is straightforward to extend equation (3) to a spectral library that depends on a set of continuous parameters, e.g., synthetic templates that depend on the metallicity Z, the dust content, the star formation history, etc., or a set of a few empirical spectra that are expanded using the principal component analysis (PCA) technique (Sodré & Cuevas 1997). In general, if the spectra are characterized by n_S possible parameters $S = \{s_1 \ldots s_{ns}\}$ (which may be physical characteristics of the models or just PCA coefficients), the probability of z given C and m_0 can be expressed as

$$p(z \mid C, m_0) = \int dS p(z, S \mid C, m_0)$$

$$\propto \int dS p(z, S \mid m_0)p(C \mid z, S) . \quad (5)$$

One situation in which the use of redshift priors most clearly reveals its advantages is the study of galaxy cluster fields, especially when near-IR photometry is not available. Since the 4000 Å break in the spectra of intermediate-redshift cluster members is difficult to distinguish from the Lyman break of higher redshift galaxies, for many objects the redshift likelihood will present two or more peaks, making ML photometric redshift estimation unfeasible for a major fraction of the sample. Using a redshift prior modeled as a smooth background component with a "spike" at the cluster redshift strongly reduces the number of objects with undetermined redshifts even with limited color information (N. Benítez & T. Broadhurst, in preparation).

3.2. *The Redshift Likelihood*

The redshift likelihood was written above as $p(C \mid z, T)$, assuming that it only depends on z and T. However, the redshift likelihood usually employed by ML photometric redshift techniques also depends on a, the template normalization factor:

$$-\log(\mathscr{L}) + \text{const} \propto \chi^2(z, T, a) = \sum_\alpha \frac{(f_\alpha - a f_{T\alpha})^2}{\sigma^2_{f_\alpha}} , \quad (6)$$

where $\{f_\alpha\}$, with $\alpha = 0, n_c$, are the observed galaxy fluxes, and $f_{T\alpha}(z)$ are the fluxes of the set of n_T templates. The expression for χ^2 in the previous equation can be rewritten as

$$\chi^2(z, T, a) = F_{OO} - \frac{F^2_{OT}}{F_{TT}} + (a - a_m)^2 F_{TT} , \quad (7)$$

where

$$a_m = \frac{F_{OT}}{F_{TT}} \quad (8)$$

is the value of a that minimizes equations (6) and (7), and

$$F_{OO} = \sum_\alpha \frac{f^2_\alpha}{\sigma^2_{f_\alpha}} , \quad F_{TT} = \sum_\alpha \frac{f^2_{T\alpha}}{\sigma^2_{f_\alpha}} , \quad F_{OT} = \sum_\alpha \frac{f_\alpha f_{T\alpha}}{\sigma^2_{f_\alpha}} . \quad (9)$$

In Bayesian Probability, the nuisance parameter a should be introduced from the beginning in the full expression for the redshift probability:

$$\begin{aligned} p(z \mid C, m_0) &= \int da \sum_T p(z, T, a \mid C, m_0) \\ &\propto \sum_T p(z, T \mid m_0) \int da\, p(a \mid m_0) \\ &\quad \times p(C \mid z, T, a) , \end{aligned} \quad (10)$$

where we have assumed that z and T do not depend on a once m_0 is known. It is obvious by comparison with equation (3) that under these assumptions,

$$p(z \mid C, T) \propto \int da\, p(a \mid m_0) p(C \mid z, T, a) . \quad (11)$$

In the absence of information about the shape of $p(a \mid m_0)$, a safe approach in this particular case is to assume a flat prior, $p(a \mid m_0) = \text{const}$. Integrating over a, the likelihood defined using equations (6) and (7), we find

$$p(C \mid z, T) \propto F_{TT}(z)^{-1/2} \exp\left[-\frac{\chi^2(z, T, a_m)}{2}\right] , \quad (12)$$

i.e., the same expression that would be reached using ML, except for the normalization factor $F_{TT}^{-1/2}(z)$.

Instead of fluxes, it may be more convenient to work with colors, normalizing the total fluxes in each band by the flux in a "base" filter, e.g., the one corresponding to the band in which the galaxy sample was selected and is considered to be complete. Here the colors, $C = \{c_i\}$, are defined as $c_i = f_i/f_0$ $(i = 1, n_c)$, where f_0 is the base flux. The exact way in which the colors are defined is not relevant; other combinations of filters are equally valid. Introducing the following definitions,

$$C_{OO} = \sum_i \frac{c^2_i}{\sigma^2_{c_i}} , \quad C_{OT} = \sum_i \frac{c_i c_{Ti}}{\sigma^2_{c_i}} , \quad C_{TT} = \sum_i \frac{c^2_{Ti}}{\sigma^2_{c_i}} , \quad (13)$$

where $\sigma_{c_i} = \sigma_{f_i}/f_0$ and $c_{Ti} = f_{Ti}/f_{T0}$, one has

$$\chi^2(z, T, a_m) = \sigma_0^{-2} + C_{OO} + \frac{(\sigma_0^{-2} + C_{OT})^2}{\sigma_0^{-2} + C_{TT}} , \quad (14)$$

$$F_{TT} = a^2_0(\sigma_0^{-2} + C_{TT}) , \quad (15)$$

where $\sigma_0 = \sigma_{f_0}/f_0$ and $a_0 = f_{TO}/f_0$. Equations (3) and (12)–(15) will be used below in all tests and practical applications.

3.3. *Prior Calibration*

In those cases where the a priori information is vague and does not allow us to choose a definite expression for the prior probability, Bayesian inference allows us to "calibrate" the prior, if necessary using the very sample under consideration.

Let us suppose that the distribution $p(z, T, m_0)$ is parametrized using n_λ continuous parameters λ. These may be the coefficients of a polynomial fit, a wavelet expansion, etc. In that case, including λ in equation (3), the probability can be written as

$$\begin{aligned} p(z \mid C, m_0) &= \int d\lambda \sum_T p(z, T, \lambda \mid C, m_0) \\ &\propto \int d\lambda\, p(\lambda) \sum_T p(z, T, \mid m_0, \lambda) \\ &\quad \times p(C \mid z, T) , \end{aligned} \quad (16)$$

where $p(\lambda)$ is the prior probability of λ, and $p(z, T \mid m_0, \lambda)$ is the prior probability of z, T and m_0 as a function of the parameters λ. The latter has not been included in the likelihood expression, since C is completely determined once the values of z and T are known.

Now let us suppose that the galaxy belongs to a sample containing n_g galaxies. Each jth galaxy has a base magnitude m_{0j} and colors C_j. The sets $\mathbf{C} \equiv \{C_j\}$ and $\mathbf{m}_0 \equiv \{m_{0j}\}$, $(j = 1, n_g)$ contain the colors and magnitudes, respectively, of all the galaxies in the sample. Then the probability of the ith galaxy having redshift z_i, given the full sample data $\mathbf{C}$ and $\mathbf{m}_0$, can be written as

$$p(z_i \mid \mathbf{C}, \mathbf{m}_0) = \int d\lambda \sum_T p(z_i, T, \lambda \mid C_i, m_{0i}, \mathbf{C}', \mathbf{m}'_0) . \quad (17)$$

The sets $\mathbf{C}' \equiv \{C_j\}$ and $\mathbf{m}'_0 \equiv \{m_{0j}\}, j = 1, n_g, j \neq i$ are identical to $\mathbf{C}$ and $\mathbf{m}_0$, except for the exclusion of the data C_i and m_{0i}. Applying Bayes' theorem, the product rule, and simplifying,

$$\begin{aligned} p(z_i \mid \mathbf{C}, \mathbf{m}_0) &\propto \int d\lambda\, p(\lambda \mid \mathbf{C}', \mathbf{m}'_0) \\ &\quad \times \sum_T p(z_i, T \mid m_{0i}, \lambda) p(C \mid z_i, T) , \end{aligned} \quad (18)$$

where, as before, the likelihood of C_i only depends on z_i, T, and the probability of z_i and T only depends on $\mathbf{C}'$ and $\mathbf{m}'_0$ through λ. The expression we obtain is very similar to equation (16), only now the shape of the prior is estimated from the data $\mathbf{C}'$, $\mathbf{m}'_0$. This means that even if one starts with a very sketchy idea of the shape of the prior, the very galaxy sample under study can be used to determine the value of the parameters λ, and thus to provide a more accurate estimate of the individual galaxy characteristics. Assuming that the data $\mathbf{C}'$ (as well as $\mathbf{m}'_0$) are independent among themselves,

$$p(\lambda \mid \mathbf{C}', \mathbf{m}'_0) \propto p(\lambda) p(\mathbf{C}' \mid \mathbf{m}'_0, \lambda) = p(\lambda) \prod_{j, j \neq i} p(C_j \mid m_{0j}, \lambda) , \tag{19}$$

where we have taken into account that the parameters λ do not depend on the set $\mathbf{m}'_0$ [since they describe the redshift/type prior probability distribution of a galaxy given its magnitude m_0, but do not contain any information about the general number counts distribution, $n(m_0)$]:

$$\begin{aligned} p(C_j \mid m_{0j}, \lambda) &= \int dz_j \sum_{T_j} p(z_j, T_j, C_j \mid m_{0j}, \lambda) \\ &= \int dz_j \sum_{T_j} p(z_j, T_j \mid m_{0j}, \lambda) p(C \mid z_j, T_j) . \end{aligned} \tag{20}$$

If the number of galaxies in our sample is large enough, it can be reasonably assumed that the prior probability $p(\lambda \mid \mathbf{C}', \mathbf{m}'_0)$ will not change appreciably with the inclusion of the data C_i, m_{0i} belonging to a single galaxy. In that case, a time-saving approximation is to use as a prior the probability $p(\lambda \mid \mathbf{C}, \mathbf{m}_0)$, calculated using the whole data set, instead of calculating $p(\lambda \mid \mathbf{C}', \mathbf{m}'_0)$ for each galaxy. In addition, it should be noted that $p(\lambda \mid \mathbf{C}, \mathbf{m}_0)$ represents the Bayesian estimate of the parameters that define the shape of the redshift distribution.

3.3.1. *Physical Priors: Galaxy Evolution and Redshift Clustering*

In the next section, a semiempirical parameterization λ is chosen to describe the redshift prior, $p(z, T \mid m_0)$. This is done for convenience, but nothing prevents the use of a more physical parameter set, particularly if we consider that the parameters λ characterize the joint magnitude-redshift–morphological type galaxy distribution, which contains important information about galaxy evolution and the fundamental cosmological parameters. If it is assumed that all the galaxies in a sample can be classified as belonging to a few morphological types, the joint redshift–magnitude-type distribution would be

$$n(z, m_0, T) \propto \frac{dV(z)}{dz} \phi_T(m_0) , \tag{21}$$

where $V(z)$ is the comoving volume as a function of redshift, which depends on the cosmological parameters Ω_0, Λ_0, and H_0, and ϕ_T is the Schechter luminosity function for each morphological type T, with the absolute magnitude M_0 replaced by the apparent magnitude m_0 (a transformation that depends on the redshifts, cosmological parameters, and morphological type). Schechter's function also depends on M^*, α, and ϕ^*, and on the evolutionary parameters ϵ, such as the merging rate, the luminosity evolution, etc., and therefore, the prior probability of z, T, and m_0 depends on the parameters $\lambda_C = \{\Omega_0, \Lambda_0, H_0\}$, $\lambda_* = \{M^*, \phi^* \alpha\}$, and ϵ. These parameters can therefore be directly estimated from a large enough multicolor sample using equation (19).

Another instance in which equation (19) can be used beyond its original purpose of aiding the redshift estimation is in the detection of galaxy clusters. Although the presence of a high-z cluster may only produce a negligible effect in the number counts normalization, it is usually signaled by a conspicuous spike in the redshift distribution. The latter can be parametrized as a combination of a smooth "field" component and a sharp Gaussian at the cluster redshift, z_c, which, together with the number of galaxies in the cluster, n_c, can be left as free parameters and determined using equation (19). This procedure, which has already been successfully applied to search for possible galaxy clusters in deep NICMOS/VLT observations (Benítez et al. 1999), will be developed and tested in detail in a forthcoming paper (N. Benítez, in preparation).

4. A PRACTICAL TEST FOR BPZ

The Hubble Deep Field North (HDF-N; Williams et al. 1996) has become *the* benchmark in the development of photometric redshift techniques. In this section, BPZ will be applied to the HDF-N and its performance contrasted with the results obtained with the standard "frequentist" (in the Bayesian terminology) approach, the procedure most frequently applied to the HDF-N (Gwyn & Hartwick 1996; Lanzetta et al. 1996; Sawicki et al. 1997, etc.). The photometry used for the HDF-N is that of Fernández-Soto, Lanzetta, & Yahil (1999), which, in addition to magnitudes in the four *Hubble Space Telescope* (*HST*) *UBVI* filters, also includes *JHK* magnitudes from the observations of M. Dickinson et al. (in preparation). Here I_{814} is chosen as the base magnitude, m_0. The colors are defined as described in § 3.2. Mark Dickinson has kindly provided a recent compilation of ($\sim$130) HDF-N spectroscopic redshifts drawn from Cohen et al. (1996), Steidel et al. (1996), Lowenthal et al. (1997), Dickinson (1998), Weymann et al. (1998), Spinrad et al. (1998), Hogg et al. (1998), and Barger et al. (1999), together with some unpublished redshifts from the Steidel and Spinrad groups. From this catalog, we excluded the stars and object 2-256.0, whose spectroscopic redshift is problematic (M. Dickinson 1999, private communication).

After a few tests with the HDF-N spectroscopic subsample, a template library similar to that of Sawicki et al. (1997) was chosen. It contains the four Coleman, Wu, & Weedman (1980) templates (E/S0, Sbc, Scd, and Irr), plus the spectra of two starbursting galaxies from Kinney et al. (1996; Sawicki et al. 1997 used two very blue SEDs from GISSEL). All the spectra were extended to the UV using a linear extrapolation and a cutoff at 912 Å, and to the near-IR using GISSEL synthetic templates. The spectra are corrected for intergalactic absorption following Madau (1995).

It might seem in principle that a synthetic template set that takes galaxy evolution into account (at least tentatively) is more appropriate than a low-redshift empirical library extrapolated to very high redshifts. However, simple tests (see also Yee 1998) show that the extended CWW set offers much better results than the GISSEL models (Bruzual & Charlot 1993). Since the synthetic models do not seem to work well even within the relatively bright magnitude range corresponding to the HDF-N spectroscopic sample, there are few reasons to suppose that their performance will improve at fainter magnitudes.

A crucial point is to ensure that the template library covers the main characteristics of all the spectral types present in the data. Figure 3 illustrates the effects of template incompleteness on the redshift estimation. The top plot displays the results obtained using ML (§ 3.2) redshift estimation using only the four CWW templates (this plot is similar to the z_{phot}-z_{spec} diagram shown in Fernández-Soto et al. 1999). In the bottom panel of Figure 3, the results shown also use ML (no BPZ yet), but include two more templates, SB2 and SB3 from Kinney et al. (1996). It can be seen that the new templates have little affect in the low-redshift range, but the changes at $z > 2$ are quite dramatic; the sagging of the CWW-only diagram disappears, and the general scatter of the diagram decreases by 25%.

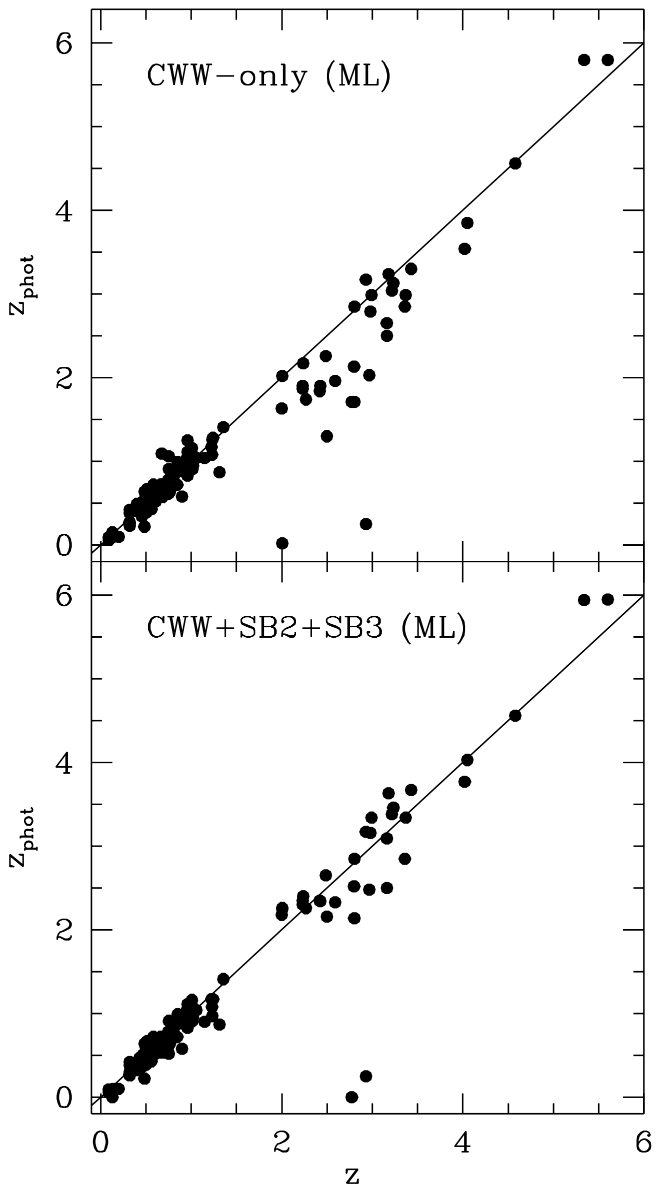

FIG. 3.—*Top*: Photometric redshifts obtained by applying our ML algorithm to the HDF-N spectroscopic sample using a template library that contains only the four CWW main types: E/SO, Sbc, Scd, and Irr. These results show characteristics similar to those of Fernández-Soto et al. (1999). *Bottom*: Shows the significant improvement (without using BPZ) obtained by just including two of the Kinney et al. (1996) spectra of starburst galaxies, SB2 and SB3, in the template set. The sagging or systematic offset between $1.5 < z < 3.5$ is eliminated, and the general scatter of the relationship decreases from $\Delta z/(1 + z_{spec}) = 0.09$ to $\Delta z/(1 + z_{spec}) = 0.07$.

The next step in the application of BPZ is choosing the shape of the prior probabilities. Because of the depth of the HDF-N, there is practically no reliable information about the expected redshift distribution. This is therefore a good example of a situation in which the prior calibration procedure described in § 3 should be applied. It will be assumed that the early types (E/S0) and spirals (Sbc, Scd) have a spectral type prior (eq. [4]) of the form

$$p(T \mid m_0) = f_t \, e^{-k_t(m_0 - 20)} , \tag{22}$$

with $t = 1$ for early types and $t = 2$ for spirals. The fraction of irregulars (the remaining three templates; $t = 3$) is automatically set to complete the galaxy mix, so there is no need to parametrize it. The spectral fractions at $m_0 = 20$ are E/SO 35%, spirals 50%, and Irr 15%. The results are very robust with respect to changes in these initial values, since the free parameters k_1 and k_2 provide enough leeway to set the correct fractions from the data at fainter magnitudes. The shape chosen for the redshift prior is

$$p(z \mid T, m_0) \propto z^{\alpha_t} \exp\left\{-\left[\frac{z}{z_{mt}(m_o)}\right]^{\alpha_t}\right\} . \tag{23}$$

This tentatively reproduces the exponential cutoff at high redshifts present in spectroscopical redshift surveys, and has the advantage of a great flexibility: depending on the value of the parameter α_t, it can roughly approximate almost any reasonable unimodal redshift distribution, from very narrow, concentrated ones for $\alpha_t \gg 2$ to practically flat ones at $\alpha_t \ll 1$. In this way, the prior distribution to be estimated from the data is as little biased as possible by the functional shape chosen in equation (23). The "median" redshift, z_m, is chosen to have a simple, linear dependence on magnitude,

$$z_{mt}(m_0) = z_{0t} + k_{mt}(m_0 - 20) . \tag{24}$$

In total, there are 11 free parameters ($\lambda = \{\alpha_t, z_{0t}, k_{mt}, k_t\}$) to be determined using the calibration procedure.

The HDF-N data set used for the prior calibration procedure is formed by 737 galaxies with $20 < I < 27$. For the $\sim$130 objects with spectroscopic redshifts, the likelihood $p(C \mid z, T)$ was built as a delta function located at the galaxy redshift and type T. The calibration sample was beefed up at bright magnitudes by including the CFRS catalog (Lilly et al. 1995; Crampton et al. 1995), 591 galaxies with $20 < I < 22.5$ that were spectrally classified using their $V - I$ colors. This ensures the presence of enough galaxies at all magnitude ranges for a meaningful determination of the parameters λ.

Table 1 shows the best values, $\hat{\lambda}$, of the parameters in equations (23) and (24), found by maximizing the probability in equation (19) using the subroutine POWELL, a direction-set minimization method (Press et al. 1992). To estimate the errors on these parameters, the region around the maximum is approximated as a multidimensional Gaussian,

$$p(\lambda \mid D) \approx p(\hat{\lambda} \mid D) \exp\left[-\tfrac{1}{2}(\lambda - \hat{\lambda}) I (\lambda - \hat{\lambda})\right] .$$

The Fisher information matrix, I, is defined as

$$I = \frac{\partial^2 \ln [p(\hat{\lambda} \mid D)]}{\partial^2 \lambda} ,$$

TABLE 1

PARAMETERS OF THE PRIORS, $p(z \mid T, m_0)$

Spectral Type	α_t	z_{0t}	k_{mt}	k_t
E/S0	2.46 ± 0.22	0.431 ± 0.030	0.091 ± 0.017	0.147 ± 0.013
Sbc, Scd	1.81 ± 0.10	0.390 ± 0.024	0.0636 ± 0.0090	0.450 ± 0.036
Irr	0.91 ± 0.05	0.063 ± 0.013	0.123 ± 0.012	...

and calculated by numerical differentiation. The 1 σ errors of the parameters, which can be estimated by inverting I, are shown in Table 1.

Figure 4 shows the full prior in redshift $p(z \mid m_0)$, found by summing over T:

$$p(z \mid m_0) = \sum_T p(T \mid m_0) p(z \mid T, m_0) . \tag{25}$$

Once the priors are determined, we can proceed to the redshift estimation using equation (16). The results for individual galaxies are very robust to variations in the prior parameters within the errors shown in Table 1, and thus the multiplication by the probability distribution $p(\lambda)$ and the integration over $d\lambda$ will be skipped; the additional computational effort of performing an 11 dimensional integral is not justified.

There are several options for converting the continuous probability, $p(z \mid C, m_0)$, to a point estimate of the "best" redshift, z_b. Here the mode of the final probability is chosen, although taking the median value of z, corresponding to 50% of the cumulative probability, or even the average, $\langle z \rangle \equiv \int dz\, z p(z \mid C, m_0)$, could also be valid.

Since Bayesian theory works with full probabilities, it offers a way to characterize the accuracy of the redshift estimation not available to ML estimates. For instance, a 1 σ error can be defined using an interval that contains 66% of the integral of $p(z \mid C, m_0)$ around z_b, etc. The indicator of redshift reliability chosen here is the quantity $p_{\Delta z}$, the probability of $|z - z_b| < \Delta z$, where z is the galaxy redshift. When the value of $p_{\Delta z}$ is low, we are warned that the redshift probability is spread over a large range in redshift, and therefore the prediction is likely to be unreliable. As demonstrated below, $p_{\Delta z}$ efficiently picks out the galaxies with catastrophic redshift errors, usually those with multimodal or very diluted redshift likelihoods.

The photometric redshifts resulting from applying BPZ to the HDF-N spectroscopic sample are plotted in Figure 5. Some interpolation (three intermediate spectra between each pair of our template library) has been performed, which only has the effect of slightly (10%) reducing the small-scale scatter. A finer interpolation did not appreciably improve the scatter and considerably increased the computational burden. Only one galaxy is discarded after applying the $p_{\Delta z} < 0.95$, $\Delta z = 0.2 \times (1 + z)$ threshold, not by coincidence one of the two outliers appearing in Figure 3. The other outlier is assigned a correct redshift by BPZ, and in general it is evident from Figure 5 that the agreement is very good at all redshifts. The residuals, $\Delta z_b = z_b - z_{\rm spec}$, have $\langle \Delta z_b \rangle = 0.01$. The rms of the quantity $\Delta z_b/(1 + z_b)$ is only 0.059, and there are no appreciable systematic effects.

Comparing the bottom panel of Figure 3 with Figure 5, it may seem that, apart from the elimination of two outliers, there is little advantage in using BPZ rather than ML. This is not surprising in the particular case of the HDF-N spectroscopic sample, formed of relatively bright galaxies, which

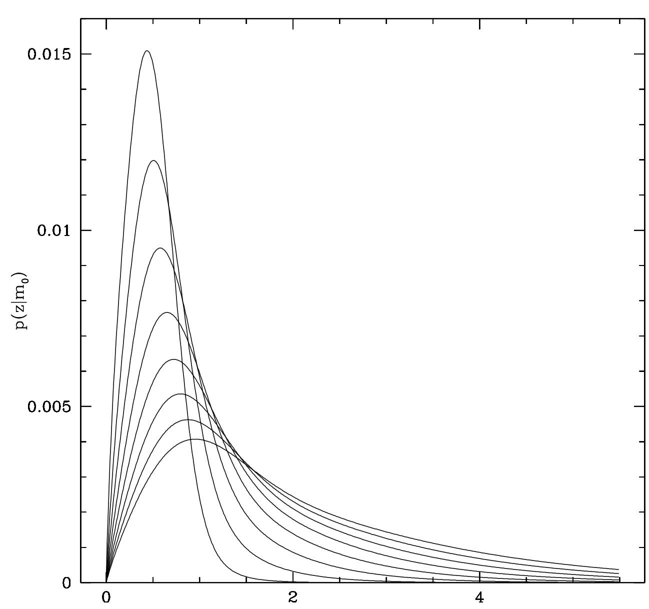

FIG. 4.—Prior in redshift, $p(z \mid m_0)$, estimated from the HDF-N data using the prior calibration procedure described in § 4, for different values of the magnitude, m_0 ($I_{814} = 21$–28)

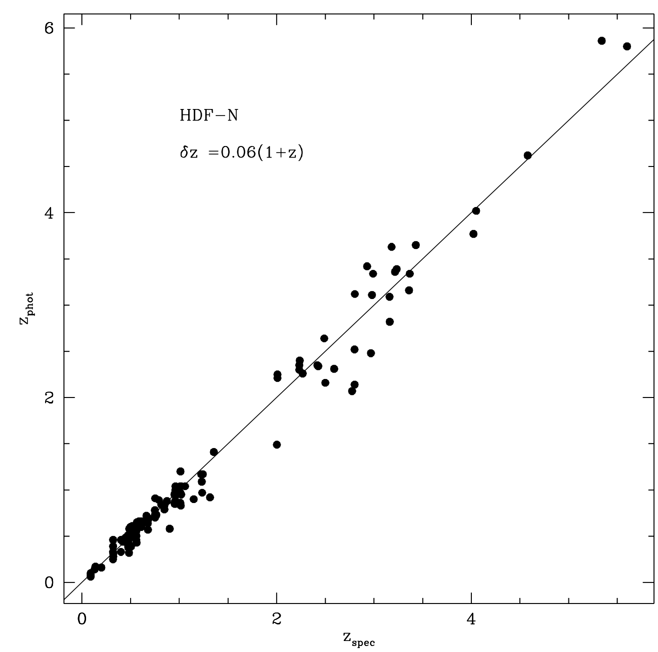

FIG. 5.—BPZ photometric redshifts, plotted against the spectroscopical ones for the HDF-N. Some interpolation (three intermediate spectra) is performed between the main template spectra mentioned in the text, which slightly reduces (by 10%) the small-scale scatter. One of two outliers in the bottom panel of Fig. 3 is assigned a correct redshift by BPZ. The other is the only object discarded when a $p_{\Delta z} < 0.95$ threshold is applied (see text). The rms scatter around the continuous line is $\Delta z_b/(1 + z_b) = 0.059$.

often occupy privileged regions in the color space, and which consequently have sharp likelihood peaks, little affected by smooth prior probabilities. To better illustrate the effectiveness of BPZ, especially when working under worse than ideal conditions, the photometric redshifts for the spectroscopic sample are estimated again using ML and BPZ, but restricting the color information to the *UBVI HST* filters. The results are plotted in Figure 6. The ML redshift diagram displays six catastrophic errors ($\Delta z \gtrsim 1$). Note that these are the same kind of errors pointed out by Ellis (1997) in the first HDF-N photometric redshift estimations. BPZ assigns correct redshifts to four of these outliers, and setting a $p_{\Delta z} > 0.95$ threshold (which discards a total of three galaxies) eliminates the other two. This is a clear example of the capability of BPZ (combined with an adequate template set) to obtain reliable photometric redshift estimates. Note that the ML estimates shown in Figure 3 presented outliers, which shows that applying BPZ to UV-only data may yield results more reliable than those obtained with ML including near-IR information.

The advantages of BPZ can also be illustrated with a simulated sample. These can be generated using the procedure described in Fernández-Soto et al. (1999). Each galaxy in the HDF-N is assigned a redshift and type using BPZ, and a mock catalog is created containing the colors that correspond to the best-fitting redshifts and templates. To simulate the photometric errors, a random photometric noise of the same amplitude as the real photometric error is added to each object. The bottom panel of Figure 7 shows the ML estimated redshifts for the mock catalog ($I < 28$) against the "true" redshifts; although in general the agree-

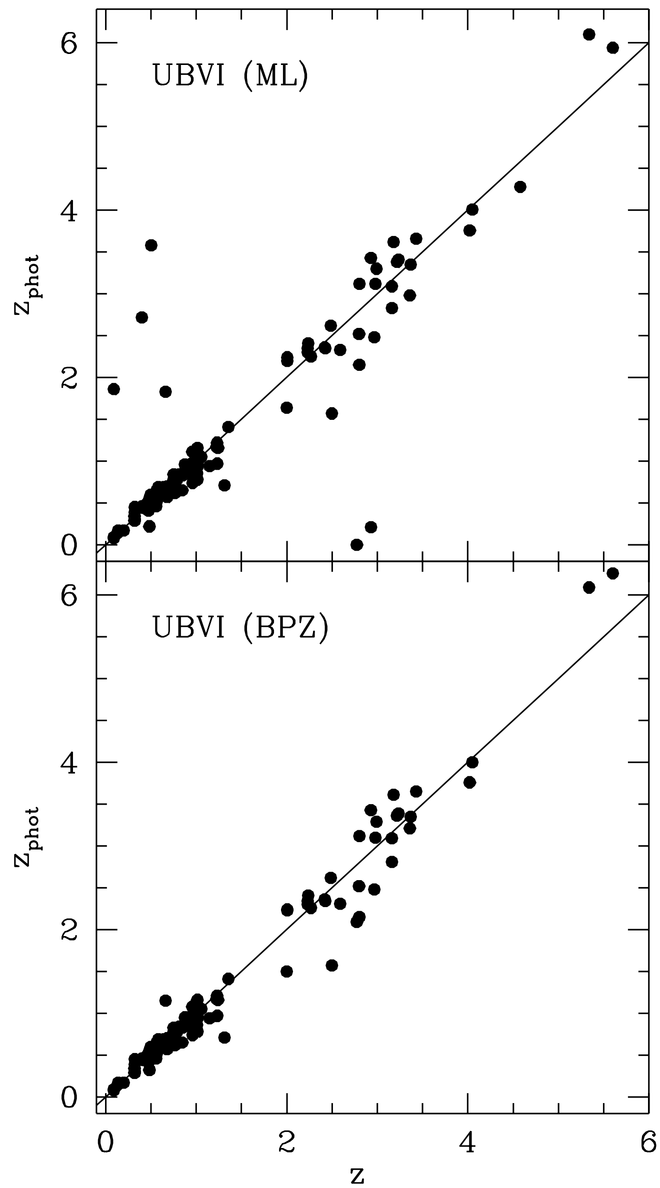

FIG. 6.—*Top*: Results of applying ML to the HDF-N spectroscopic sample using only the four UBVI *HST* bands. *Bottom*: Effects of applying BPZ to the same sample. Four of the outliers are assigned correct redshifts, and setting a $p_{\Delta z} > 0.95$ threshold (which discards three galaxies) eliminates the other two. Compare with Fig. 3 (*bottom*), which also includes the near-IR photometry of Dickinson et al. (1998), in preparation. Even with fewer filters, BPZ is more reliable than ML.

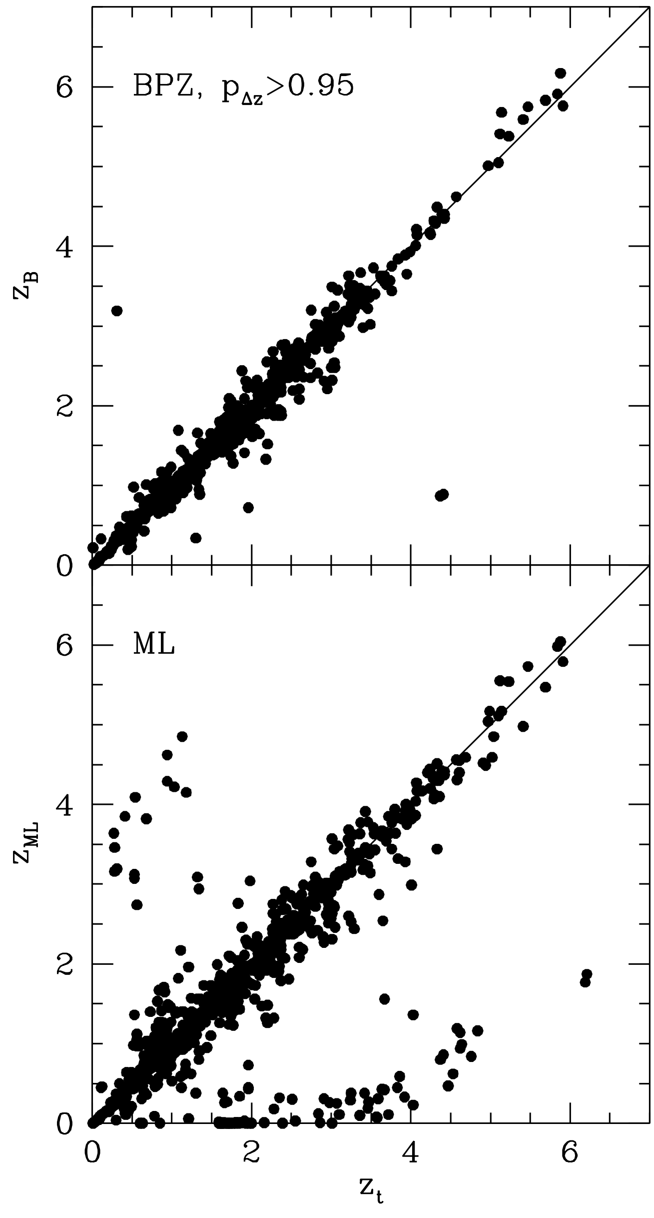

FIG. 7.—*Top*: BPZ photometric redshifts, z_b, for the $I < 28$ HDF-N mock catalog, plotted against the "true" redshifts, z_t (see text). A threshold of $p_{\Delta z} > 0.95$ has been applied, which eliminates 20% of the objects. *Bottom*: Results obtained by applying ML to the same mock sample. The fraction of outliers is 7%.

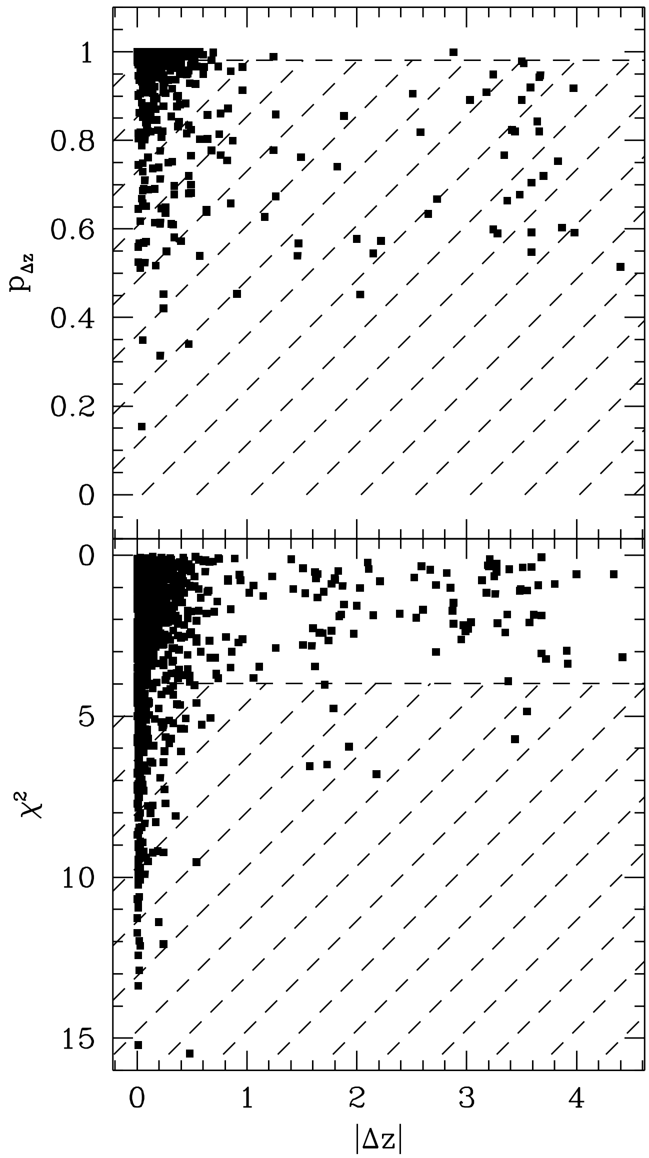

FIG. 8.—*Top*: Probability, $p_{\Delta z}$, plotted against the absolute value of the difference between the "true" redshift (z_t) and the one estimated using BPZ (z_b) for the mock sample described in § 4. The higher the value of $p_{\Delta z}$, the more reliable the redshift should be. The shaded region shows the low quartile in the value of $p_{\Delta z}$. Most of the outliers are at low values of $p_{\Delta z}$, which allows us to eliminate them by setting a suitable threshold of $p_{\Delta z}$ (see text and Fig. 7, *bottom*). *Bottom*: Plot showing that it is not possible to do something similar using ML redshifts and χ^2 as an estimator. The value of χ^2 of the best ML fit is plotted against the error in the ML redshift estimation, $|z_t - z_{\rm ML}|$. The shaded region shows the high quartile in the values of χ^2. One would expect that low values of χ^2 (and therefore better fits) would correspond to more reliable redshifts, but this obviously is not the case. This is not surprising: the outliers in this figure are all due to color/redshifts degeneracies, like the one displayed in Fig. 1, which may give an extremely good fit to the colors, *C*, but a totally wrong redshift.

ment is not bad (as could be expected), there is a considerable fraction of outliers (7%), whose positions illustrate the main source of color/redshift degeneracies: high-z galaxies that are erroneously assigned $z \lesssim 1$ redshifts and vice versa. This shortcoming of the ML method is analyzed in detail in Fernández-Soto et al. (1999). In contrast, the top panel of Figure 7 shows the results of applying BPZ with a threshold of $p_{\Delta z} > 0.95$, which eliminates 20% of the initial sample (almost one-third of which have catastrophically wrong redshifts), but reduces the number of outliers to a remarkable 1%.

It is not clear how to define a reliability estimator analogous to $p_{\Delta z} > 0.95$ within the ML framework. The obvious choice, χ^2, is practically useless as a criterion to pick out the outliers. Although one would naively expect that low values of χ^2 (and therefore better fits) would correspond to more reliable redshifts, the bottom panel of Figure 8 shows almost the opposite to be true. This figure plots the value of χ^2 versus the ML redshift error for the mock catalog, with the shaded region representing the upper quartile (25%) in χ^2; most of the outliers are above it, at lower χ^2. This is not difficult to understand: these outliers are caused by color/redshift degeneracies (Fig. 1), which may produce excellent fits to the colors *C*, but catastrophically wrong redshifts. In stark contrast, the top panel of Figure 8 plots the errors in the BPZ redshifts versus the values of $p_{\Delta z}$. The shaded region representing the lower quartile contains practically all the outliers. Thus, by setting an appropriate threshold, one can virtually eliminate the catastrophic errors.

Figure 9 shows the numbers of galaxies above a given $p_{\Delta z}$ threshold in the HDF-N as a function of magnitude and redshifts. It shows how the reliability of photometric redshifts quickly decreases for faint $I \gtrsim 27$ objects as the fraction of objects with possible catastrophic errors steadily grows with magnitude.

There is one caveat regarding the use of $p_{\Delta z}$ or similar quantities as a reliability estimator. They provide a safety check against the color/redshift degeneracies, since basically they tell us whether there are other probability peaks comparable to the highest one, but they do not insure against template incompleteness. If the template library does not contain any spectra similar to the one corresponding to the galaxy, there is no statistical indicator that can warn us about the unreliability of the prediction.

Finally, Figure 10 shows the redshift distributions for the HDF-N galaxies with $I < 27$. No objects have been removed on the basis of $p_{\Delta z}$, so the values of the histogram bins should be taken with caution. The overplotted continuous curves are the distributions used as priors (see text). The results obtained from the HDF-N and HDF-S will be analyzed in more detail using a revised photometry elsewhere.

5. CONCLUSIONS

Although spectroscopical techniques have experienced remarkable progress in the last few years, photometric redshifts are becoming increasingly important in many areas of observational cosmology. This is not surprising, as the example of the HDF-N (Williams et al. 1996) shows: after several years, the best efforts of the astronomical community have only provided spectroscopic redshifts for $\approx 20\%$ of the $I < 27$ galaxy sample. In contrast, reasonably accurate photometric redshifts for most of its galaxies were obtained within a few months by several groups, mostly using the SED-fitting method with a ML statistical approach. However, the redshifts estimated following this procedure present a considerable number ($\sim 10\%$) of catastrophic errors ($\Delta z \gtrsim 1$), which excludes their use for many practical applications (Ellis 1997). The other approach to photometric redshift estimation, the training-set method, is apparently more reliable, but since it cannot be extended beyond the spectroscopical limit, it is of limited use for very faint samples, such as the HDF-N.

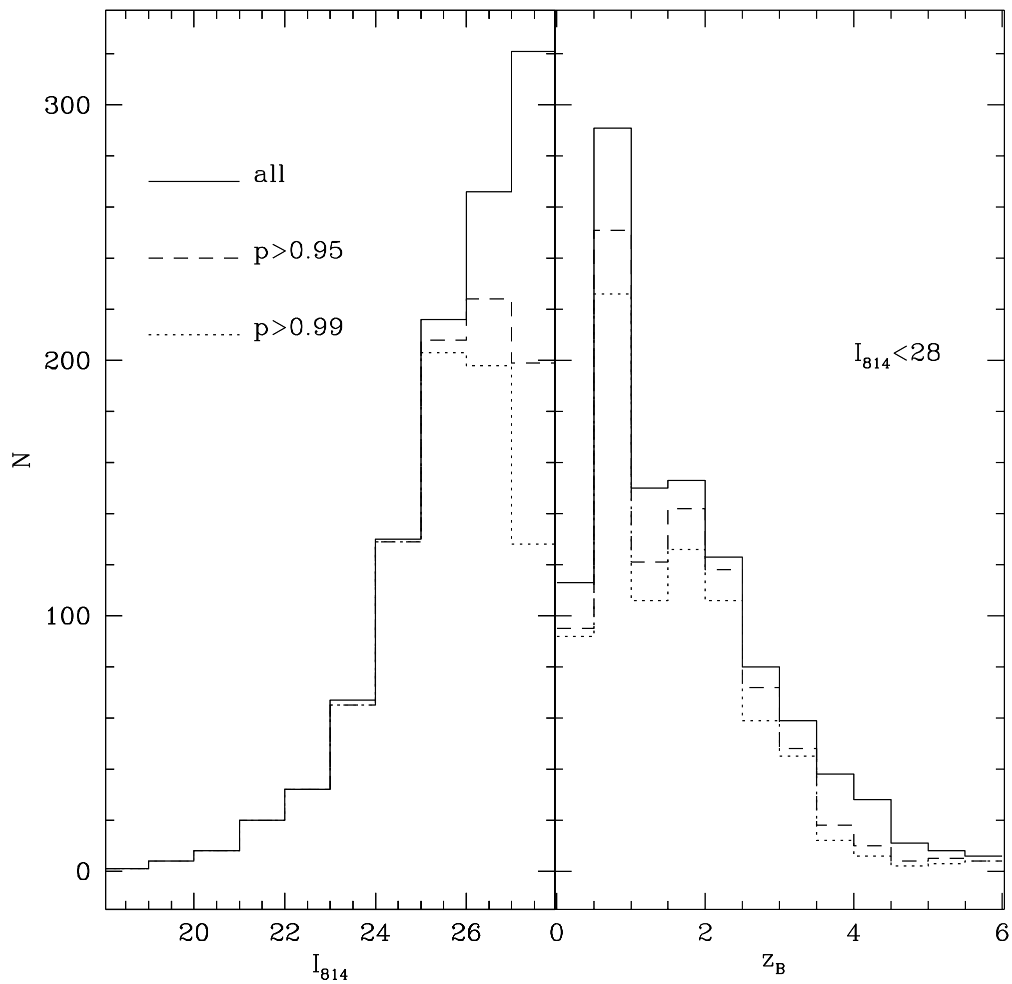

FIG. 9.—*Left*: Histograms showing the number of galaxies over $p_{\Delta z}$ thresholds of 0.95 and 0.99 as a function of magnitude. It can be seen that the number of galaxies with reliable photometric redshifts quickly decreases with magnitude. *Right*: Same as left panel, but as a function of redshift.

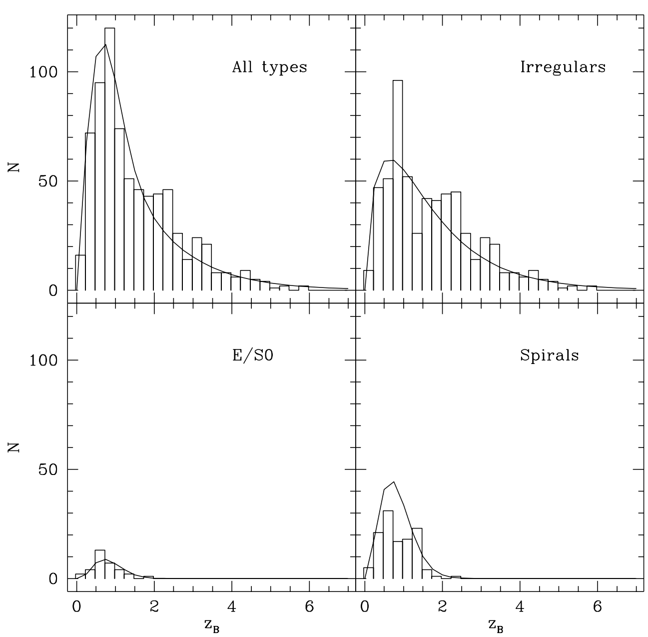

FIG. 10.—The z_b redshift distributions for the $I < 27$ HDF-N galaxies, divided by spectral types. Solid lines represent the corresponding $p(z, T)$ distributions estimated with the HDF-N and CFRS samples using the prior calibration method described in the text.

The application of Bayesian inference to photometric redshift estimation effectively overcomes most of the drawbacks of the ML and training-set methods. The use of prior probabilities and Bayesian marginalization facilitates the inclusion of relevant knowledge, such as the expected shape of the redshift distributions and the galaxy type fractions, which can be readily obtained from existing surveys but are often ignored by other methods. If this previous information is lacking or insufficient (for instance, because of the unprecedented depth of the observations), the corresponding prior distributions can be calibrated using even the data sample for which the photometric redshifts are being obtained. An important advantage of Bayesian statistics is that the accuracy of the redshift estimation can be characterized in a way that has no equivalents in other statistical approaches, enabling the selection of galaxy samples with extremely reliable photometric redshifts. In this way, it is possible to determine more accurately the properties of individual galaxies and simultaneously estimate the statistical properties of a sample in an optimal fashion. Moreover, the Bayesian formalism described here can be easily generalized to deal with a wide range of problems that make use of photometric redshifts.

There is an excellent agreement between the ≈ 130 HDF-N spectroscopic redshifts and the predictions of the method, with a rms error $\Delta z \approx 0.06(1 + z_{\rm spec})$ up to $z < 6$ and no outliers nor systematic biases. It should be remarked that these results have not been reached following a training-set procedure; since the template library is empirical, the above value of Δz should be a fair estimate of the expected accuracy for any similar sample. The method is further tested by estimating redshifts in the HDF-N but restricting the color information to the *UBVI* filters; the results are shown to be significantly more reliable than those obtained with maximum-likelihood techniques.

I would like to thank Tom Broadhurst and Rychard Bouwens for carefully reading the manuscript and making valuable comments, Alberto Fernández-Soto and collaborators for kindly providing the HDF-N photometry and filter transmissions, Mark Dickinson for making available his HDF-N spectroscopic redshift compilation, and Brenda Frye for help with the intergalactic absorption correction. The referee, Prasenjit Saha, made valuable comments which helped to improve the paper, in particular the suggestion to use a flat prior in a for the redshift likelihood. The author acknowledges a Basque Government postdoctoral fellowship and financial support from the NASA grant LTSA NAG-3280 and the Advanced Camera for Surveys project, NASA contract NAS5-32865.

REFERENCES

Barger, A. J., Cowie, L. L., Trentham, N., Fulton, E., Hu, E. M., Songaila, A., & Hall, D. 1999, AJ, 117, 102
Benítez, N., Broadhurst, T., Bouwens, R., Silk, J., & Rosati, P. 1999, ApJ, 515, L65
Bretthorst, L. 1988, Bayesian Spectrum Analysis and Parameter Estimation (Berlin: Springer)
———. 1990, J. Magnetic Resonance, 88, 552
Brunner, R. J., Connolly, A. J., Szalay, A. S., & Bershady, M. A. 1997, ApJ, 482, L21
Bruzual, A., G., & Charlot, S. 1993, ApJ, 405, 538
Cohen, J. G., Cowie, L. L., Hogg, D. W., Songaila, A., Blandford, R., Hu, E. M., & Shopbell, P. 1996, ApJ, 471, L5
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393
Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., & Munn, J. A. 1995, AJ, 110, 2655
Connolly, A. J., Szalay, A. S., Dickinson, M., Subbarao, M. U., & Brunner, R. J. 1997, ApJ, 486, L11
Crampton, D., Le Fevre, O., Lilly, S. J., & Hammer, F. 1995, ApJ, 455, 96
Dickinson, M. 1998, in The Hubble Deep Field, ed. M. Livio, S. M. Fall, & P. Madau, (Cambridge: Cambridge Univ. Press), 219
Ellis, R. S. 1997, ARA&A, 35, 389
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. 1998, Bayesian Data Analysis, (London: Chapman & Hall)
Gunn, J. E., & Weinberg, D. H. 1995, in Wide-Field Spectroscopy and the Distant Universe, ed. S. J. Maddox & A. Aragón-Salamanca (3d ed.; Singapore: World Scientific)
Gwyn, S. D. J., & Hartwick, F. D. A. 1996, ApJ, 468, L77
Hogg, D. W., et al. 1998, AJ, 115, 1418
Jaynes, E. T. 2000, Probability Theory: The Logic of Science (Cambridge: Cambridge Univ. Press), in press
Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., & Schmitt, H. R. 1996, ApJ, 467, 38
Kodama, T., Bell, E. F., & Bower, R. G. 1998, MNRAS, 302, 152
Koo, D. C. 1985, AJ, 90, 418
Lanzetta, K. M., Yahil, A., & Fernández-Soto, A. 1996, Nature, 381, 759
Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, ApJ, 455, 50
Loredo, T. 1990, in Maximum Entropy and Bayesian Methods, ed. P. Fougere (Dordrecht: Kluwer), 81
———. 1992, in Statistical Challenges in Modern Astronomy, ed. E. D. Feigelson & G. J. Babu (New York: Springer), 275
Lowenthal, J. D., et al. 1997, ApJ, 481, 673
Madau, P. 1995, ApJ, 441, 18
Pelló, R., Miralles, J. M., Le Borgne, J.-F., Picat, J.-P., Soucail, G., & Bruzual, G. 1996, A&A, 314, 73
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in FORTRAN: The art of Scientific Computing (Cambridge: Cambridge Univ. Press)
Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
Sivia, D. S. 1996, Data Analysis: A Bayesian Tutorial (Oxford: Oxford Univ. Press)
Sodré, L., & Cuevas, H. 1997, MNRAS, 287, 137
Spinrad, H., Stern, D., Bunker, A., Dey, A., Lanzetta, K. M., Yahil, A., Pascarelle, S., & Fernández-Soto, A. 1998, AJ, 116, 2617
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, AJ, 112, 352
Wang, Y., Bahcall, N., & Turner, E. L. 1998, AJ, 116, 2081
Weymann, R. J., Stern, D., Bunker, A., Spinrad, H., Chaffee, F. H., Thompson, R. I., & Storrie-Lombardi, L. J. 1998, ApJ, 505, L95
Williams, R. E., et al. 1996, AJ, 112, 1335
Yee, H. K. C. 1998, preprint (astro-ph/9809347)