

Spectral-based printer modeling and characterization

Silvia Zuffi

Construction Technologies Institute
ITC-CNR
Milano, Italy
E-mail: zuffi@itc.cnr.it

**Raimondo Schettini
Giancarlo Mauri**

DISCo
Università degli Studi di Milano Bicocca
Milano, Italy
E-mail: schettini@disco.unimib.it
E-mail: mauri@disco.unimib.it

Abstract. *Spectral characterization involves building a model that relates the device dependent representation to the reflectance function of the printed color, usually represented with a high number of reflectance samples at different wavelengths. Look-up table-based approaches, conventionally employed for colorimetric device characterization cannot be easily scaled to multispectral representations, but methods for the analytical description of devices are required. The article describes an innovative analytical printer model based on the Yule–Nielsen Spectral Neugebauer equation and formulated with a large number of degrees of freedom in order to account for dot-gain, ink interactions, and printer driver operations. To estimate our model's parameters we use genetic algorithms. No assumption is made concerning the sequence of inks during printing, and the printers are treated as RGB devices (the printer-driver operations are included in the model). We have tested our characterization method, which requires only about 130 measurements to train the learning algorithm, on four different inkjet printers, using different kinds of paper and drivers. The test set used for model evaluation was composed of 777 samples, uniformly distributed over the RGB color space. © 2005 SPIE and IS&T. [DOI: 10.1117/1.1900137]*

1 Introduction

Multispectral imaging is attracting increasing attention, triggered by the appealing advantages of the multispectral approach compared with traditional colorimetric imaging.¹ In image acquisition, the possibility of capturing the reflectance spectra of surfaces in a multispectral image constitutes a powerful tool for detailed and accurate color information regarding objects and scenes. In image reproduction, the multispectral approach can significantly reduce the undesirable effects of metamerism between the original and the corresponding reproduction.²

In the multispectral reproduction of an image, the dimensions of the color space are no longer the three of colo-

rimetric systems employing CIELab or CIEXYZ color spaces, but amount to the number of samples used to describe the spectrum. The color space of image reproduction devices will probably increase as well from three or four to more dimensions, to enhance the capability of these devices to reproduce the reflectance spectra of colors. Conventional reproduction systems, in fact, cannot establish spectral matches between a generic color and an halftone print, due to the limitation in the number of different inks used by the devices on the market nowadays. Printer characterization consists in the definition of the relationship between device dependent and device independent color representations.³ Conventionally, colorimetric characterization is done by printing and measuring a large number of colors ranging over the whole gamut of the device, storing these in look-up tables (LUT), and applying some sort of interpolation method to map intermediate points. LUT-based methods of printer characterization require the measurement of a large number of sample color patches due to the highly complex nonlinear relationship between the device-dependent values and the colorimetric color space.

Spectral characterization involves building a model that relates the device-dependent representation to the reflectance function of the printed color, usually represented with a high number of reflectance samples at different wavelengths. A LUT-based approach for device characterization is consequently not feasible in the context of multispectral reproduction as the size of the LUT to be built would be very large.

Models for the analytical description of printing processes have long been proposed due to the advantage, with respect to LUT-based approaches, of the small number of measurements they require. Most of the methods developed for the spectral-based characterization of binary printers are based on the color mixing model introduced by Hans Neugebauer in 1937. It assumes additive color theory, and com-

Paper 03040 received Mar. 11, 2003; revised manuscript received Mar. 19, 2004; accepted for publication Nov. 18, 2004; published online May 5, 2005.
1017-9909/2005/\$22.00 © 2005 SPIE and IS&T.

putes the reflectance of an halftone print as the sum of each combination of inks weighted by the proportion of paper that it covers. Each ink combination is therefore a primary in the additive model, and is called a Neugebauer primary. The model was unable, however, to predict the reflectance of colored halftones with sufficient accuracy until 1951, when the Yule–Nielsen correction accounting for the scattering of light within the paper was introduced.

The Yule–Nielsen effect of light scattering causes a shadow around the edge of the dots that makes them look larger. The light, entering from points where there is ink, is reflected in a region around the point of incidence due to multiple scattering in the substrate. This makes the reflected image a diffuse version of the incident pattern of light, and as a consequence, the dots on paper appear larger. This phenomenon, referred to as optical dot gain, was modeled by Yule and Nielsen with the introduction of a simple n factor as a power function in the equation of a single color halftone, and successively applied in the Neugebauer color mixing model.

Much has been published on the physical grounds for the Yule–Nielsen n value. It has been observed that values between 1 and 2 are physically meaningful, and values greater than 2 represent various effects, among which the variation in dot density, and are required for high-resolution printers.⁴

In its original formulation, the physical justification of the Neugebauer model is quite intuitive, but it is also true that many phenomena, unpredictable or unknown, take place in printing due to the interaction among inks and of inks with paper. Much research has been devoted to the optimization of the model, whose results indicate that the physical integrity of the mixing model can be sacrificed to maximize its performance of color prediction. It is also our opinion that the adoption of a correction without a clear physical justification is permissible if it improves the model's performance. Hence, the n value is usually treated as a free parameter to improve the model fit; a wavelength-dependent n value has been also used.⁵ Several methods more complex than the simple n value have been successively introduced to model optical dot gain, among which the convolution with a point spread function,⁶ or probabilistic models.^{7,8}

Another effect, commonly called mechanical dot gain, determines a color darker than that predicted: due to the spread of the ink on paper, the real size of a dot printed on a substrate is larger than its theoretical size. The nonlinear relationship between the theoretical concentration of ink on paper and the effective concentration (that is, the area the dot actually covers), due to the combination of optical and mechanical dot gain, is usually taken as a parameter in the model fit optimization. Alternative approaches describe the spreading of the ink, as proposed by Emmel and Hersch in a method that models the ink spread by enlarging the drop impact on the basis of the configuration of its neighbors and the state of the surface,^{9,10} or by Gustavson,⁶ who models the physical dot gain with a transmission function defined on a blurred version of the halftone image. The accuracy of the various Neugebauer models has been investigated by Rolleston and Balasubramanian.¹¹ They observed that a considerable improvement may be obtained with a cellular model if broadband measurements are employed, whereas

spectral measurements are less useful. Cellular approaches have also been followed by Agar and Allebach.¹² Balasubramanian has proposed a technique based on a weighted least squares regression to optimize the Neugebauer primaries,¹³ whereas Xia *et al.*¹⁴ has approached Neugebauer optimization with a total least squares regression method that accounts for errors in the measured reflectance for both the primaries and the modeled samples. Berns and his colleagues at the Munsell Color Laboratory have been working on multispectral color reproduction for several years.² Tzeng in his PhD thesis¹⁵ and Tzeng and Berns¹⁶ have presented a spectral-based color separation algorithm exploiting the Yule–Nielsen Spectral Neugebauer model for a six-ink inkjet printer, which takes into account the limit of inktrapping, considering the printer model the union of ten four-ink printer models. The same color separation approach has been used by Taplin and Berns without the ink-limiting constraints.¹⁷ The paper is organized as follows: Section 2 describes the proposed method, Sec. 3 illustrates the experiments performed, and Sec. 4 reports our conclusions.

2 Printer Modeling and Spectral Characterization

We have adopted the Yule–Nielsen modified Spectral Neugebauer (YNSN) model for binary printer characterization. The parameters of the YNSN model are usually computed with regression-based methods. The Yule–Nielsen n value may be derived from an exhaustive procedure of error minimization between the measured and the predicted spectra of a set of colors, where n varies over a limited range of values. Dot gain functions can be estimated before,¹³ or after^{4,15} the n -value optimization. In a context in which the physical meaning of the Yule–Nielsen n value has been lost, the n and dot gain functions represent two strategies for dealing with the effect of dot gain regardless of its origin, optical or mechanical, and should therefore be estimated at the same time. Moreover, our experience shows that the training set of reflectance data does not always exhibit the characteristics of regularity that make it possible to fit the model to the printer simply by using least-square estimated parameters. To solve this problem, we have designed an analytical printer model that can be used regardless of the characteristics of the device considered. Our model is based on the Yule–Nielsen Spectral Neugebauer equation, formulated with a large number of degrees of freedom in order to account for dot gain, ink interaction, and printer-driver operations. In the definition of the whole characterization model, we kept the printer driver and the print process transformations as separate modules. We employ a general but plausible equation to relate RGB to CMYK. Our approach, therefore, would allow an easy adoption of more complex and complete RGB to CMYK models, if they would be available.

To estimate the model's parameters we use genetic algorithms. The basic elements of genetic algorithms are described in Sec. 2.1. Section 2.2 describes the printer model, and Sec. 2.3 the genetic algorithm used.

2.1 Genetic Algorithms

Genetic algorithms (GA) are one of the heuristic search methods that have been proposed and extensively studied to deal with difficult optimization problems, as an alternative

to conventional optimization methods. They are a general method inspired by the mechanisms of evolution in biological systems, based on the Darwinian principle of the survival and reproduction of the fittest, and are appropriate for complex nonlinear models where location of the global optimum is a difficult task (see, e.g., Refs. 18 and 19 for an introduction to GA and their applications).

In the basic genetic algorithm every candidate solution to the optimization problem is represented by a sequence of binary, integer, real, or even more complex values, called an individual, or chromosome. Further, a function assigning a suitable “fitness” value to every individual is defined; in a nonlinear programming scenario, the fitness function will reflect the objective value of the given model.²⁰ Then, a small number n of individuals (with respect to the cardinality of the whole solution space) are randomly generated as an initial population P . The GA then iterates a procedure that produces a new population P' from the current P , until a given “STOP” criterion is satisfied. Each iteration consists in the following steps:

- fitness evaluation: for every individual x in P , the value $f(x)$ is computed;
- selection: $n/2$ pairs of individuals are randomly selected from population P ; the probability of selection is higher for individuals of greater fitness;
- crossover: two new individuals (sons) are generated by cutting the two individuals of each pair (parents) at a randomly chosen point and interchanging the four parts so obtained; and
- mutation: the value of each position of the elements in P' is changed with a given probability p_m .

In this way, since the sons inherit the distinctive features from their parents, the mean fitness of the population increases from a generation to the next one, that is we concentrate our search to regions of the solutions space near to optimal solutions. The GA techniques are particularly useful when dealing with complex nonlinear models that require the simultaneous management of many parameters, and with irregular training data sets. In these situations, globally optimal solutions can be found only by exhaustive search, which is computationally infeasible. So, the returned solution is, in general, a local optimum. However, especially when handling integer variables, GA have some advantages versus conventional gradient search techniques applied to non-linear programming models, as shown for example in Ref. 20. The disadvantages are that genetic algorithms cannot guarantee an optimal solution, and that it is, in general, also difficult to tune their free parameters.

2.2 Printer Modeling

Spectral characterization involves building a model that relates the device dependent representation to the reflectance function of the printed color. In our approach, the device dependent color representation is the RGB values sent to the printer driver by the application. The printer driver operations that implement the conversion from RGB to the CMYK ink amounts to print are therefore included in the model. This strategy is functional to the definition of a

Table 1 Calculation of the area coverage for each Neugebauer primary, given the concentration of inks.

Index (p)	Neugebauer primary	Area coverage (a_p)
0	W	$(1-c)(1-m)(1-y)(1-k)$
1	K	$(1-c)(1-m)(1-y)k$
2	Y	$(1-c)(1-m)y(1-k)$
3	YK	$(1-c)(1-m)yk$
4	M	$(1-c)m(1-y)(1-k)$
5	MK	$(1-c)m(1-y)k$
6	R	$(1-c)my(1-k)$
7	RK	$(1-c)myk$
8	C	$c(1-m)(1-y)(1-k)$
9	CK	$c(1-m)(1-y)k$
10	G	$c(1-m)y(1-k)$
11	GK	$c(1-m)yk$
12	B	$cm(1-y)(1-k)$
13	BK	$cm(1-y)k$
14	CMY	$cmy(1-k)$
15	CMYK	$cmyk$

general method for printer characterization, as in most of the cases the printer driver operations are unknown to the user.

2.2.1 Printer model

The model of the printer we refer to is based on the well-known equation. According to the YNSN model, the spectrum of a N inks halftone print is the weighted sum of 2^N different colors, called Neugebauer primaries, given by all the possible overprints of inks. The weight of each Neugebauer primary is the area it covers in the halftone cell. The YNSN model for a 4-ink halftone print is

$$R_{\text{print},\lambda} = \left[\sum_{p=0}^{15} a_p R_{p,\lambda}^{1/n} \right]^n \quad \lambda = 1, \dots, M, \quad (1)$$

where $R_{\text{print},\lambda}$ is the reflectance of the printed color, n is the Yule–Nielsen factor, $R_{p,\lambda}$ is the reflectance of the p 'th Neugebauer primary, and a_p is the primary area coverage. The area coverage is the percentage of the halftone cell covered by the Neugebauer primary.

A model by Demichel can be used to compute the percentage of the area covered by each primary. The dot overlap is the product of the relative area covered by single inks, computed in a stochastic fashion that assumes the ink dots are randomly arranged. The model is considered valid for random or rotated halftone screen;²¹ it fails in all cases in which there is a singular screen superposition, although the color deviation observed is not excessively large.²² For dot-on-dot printing a different formulation must be considered.²³

We have used the model for a random screen, and computed area coverage according to the equations in Table 1, where $\mathbf{c} = [c, m, y, k]$ represents the concentration of inks for printing a given color.

2.2.2 Printer driver

In this work, as in the majority of practical situations, printer drivers accept only RGB images, and the internal operations that perform the conversion from the input images RGB values to CMYK printer digital counts are unknown. For these reasons, in our work, printers are considered RGB devices, and some reasonable assumptions are made on how the printer driver computes the amounts of CMYK. In the conversion from RGB to CMYK, gray component replacement (GCR), replaces the neutral or gray component of a three-color image with a certain level of black ink. This process might be accomplished by simple techniques up to a sequence of complex operations. Basically, the least predominant of the three primary inks is used to calculate a partial or total substitution by black, and the color components of that image are reduced to produce a print image of nearly equivalent color to the original three-color print.²⁴

A fundamental assumption in GCR is that the gray component of an arbitrary CMY combination is given by the minimum of C, M, and Y. The gray component can then be used as input to the K addition function, which defines the amount of requested black. A simple model of GCR is based on the rationale that the CMY reduction should be proportional to the K addition, usually chosen to meet a desired behavior along the neutral axes.³ The printer driver is therefore modeled according to

$$X = \min(C, M, Y),$$

$$k = X,$$

$$c = C - f(X), \quad (2)$$

$$m = M - f(X),$$

$$y = Y - f(X),$$

where the set of CMY primaries are simply computed as complementary values of RGB, here representing digital counts in the printer color space, and the K addition function is the identity. This simplified model is a basic GCR model, based on an assumption of ideal CMY inks.

The function $f(X)$, that defines the reduction of color components, has the following form: $f(X) = UX$. The coefficient U spans in the range $[0,1]$; its actual value, for the device to be modeled, is estimated during the characterization phase by the genetic algorithm.

2.2.3 Dot gain and ink interaction

If we consider the print of a single ink on paper, the YNSN model is simplified to

$$R_{\text{print},\lambda}^{1/n} = (1 - c_{\text{ink}})R_{w,\lambda}^{1/n} + c_{\text{ink}}R_{\text{ink},\lambda}^{1/n} \quad (3)$$

where c_{ink} is the ink concentration.

If we print and measure a ramp of samples with ink concentrations ranging from 0 to 1, and compute the effective concentration of ink using the above equation with

$n=1.0$, the effective concentration appears to depend on the wavelength (Fig. 1), indicating that the Yule–Nielsen factor may be far from being a constant.

The scattering of light within the substrate is wavelength dependent;¹³ this has been taken into account by Iino and Berns with a wavelength-dependent n factor.⁵ Nevertheless, in our work we follow a different strategy: we refer to wavelength dependence when considering the effective concentration of inks, in order to be able to account for different types of inks in another way. That is, in our model, the complex interaction of inks and substrate determining optical and mechanical dot gains depends upon the wavelength, but also upon the quantity and number, and types of inks in the halftone.

In general, dot gain functions represent an increase in the effective concentration with respect to the theoretical concentration, and peak at around 50% of the latter, we have observed a nonuniform behavior depending on the wavelength, and even a negative gain in some cases (Fig. 1). Consequently, we have looked for a function that could describe both positive and negative dot gain while using the smallest possible number of parameters for the model. The function used, tuned with only one parameter, is

$$c_{\lambda} = \frac{c_t}{(1 - C_{\lambda}) \cdot c_t + C_{\lambda}}, \quad (4)$$

where c_t is the theoretical concentration of ink, computed from RGB values using Eq. (2), and C_{λ} is the wavelength-dependent parameter. The same type of function has been used by Roberts.²⁵

Dot gain functions are commonly used to model the spread of inks on paper, but the spread may be altered when covering a previously deposited ink. Various strategies can be employed to account for this phenomenon.^{26,27,10} We have chosen to account for the interaction of the inks by providing a different dot gain function for any overprinting. Table 2 lists the parameters used to compute the effective concentration of inks, with Eq. (4), to obtain the area coverage for any Neugebauer primary indicated in Table 1.

The subscripts in the dot gain parameters refer to the inks present in the Neugebauer primary. For example, $C_{\text{cy},\lambda}$ is the parameter used to compute dot gain for the effective concentration of cyan when computing the area coverage of the Neugebauer primary G (green), which is an overprint of cyan and yellow. The same parameter is used to compute the amount of cyan in the Neugebauer primary Y (yellow) because if any amount of cyan is present, the dot gain must account for the presence of both yellow and cyan inks.

There are 32 different dot gain parameters in the model; each is a vector, the dimension of which is the number of wavelengths considered. The area of paper covered is computed as the difference between the sum of the area coverage of the inks and their overprints, with the constraint to be positive:

$$a_0 = 1 - \sum_{p=1}^{15} a_p \quad (5)$$

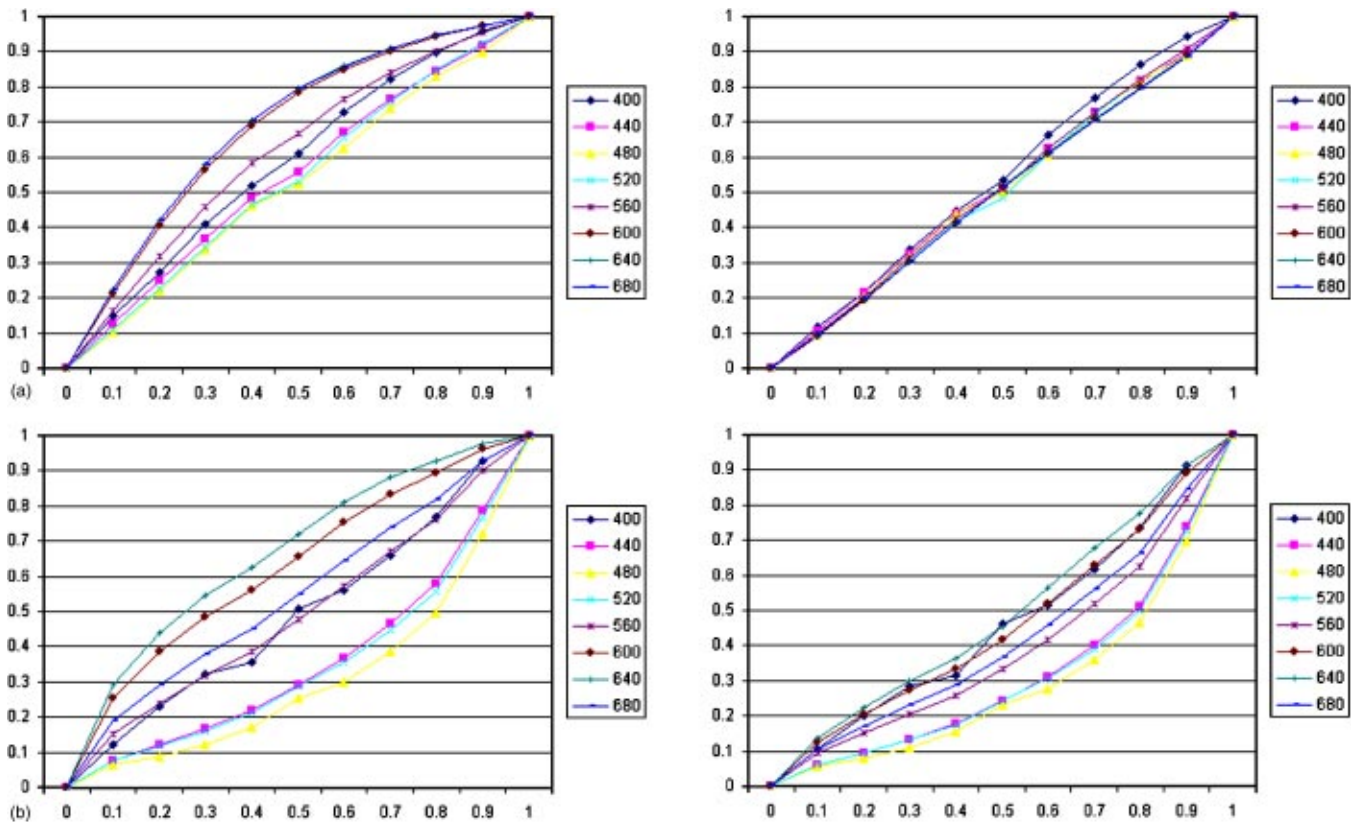


Fig. 1 Effective concentrations of cyan ink when computed with Eq. (3) (x axis: theoretical concentration and y axis: effective concentration): (a) effective concentration with $n=1$ and $n=10.3$ for the Epson Stylus Color printer. (b) effective concentration with $n=1$ and $n=7.3$ for the HP 2000C printer. Each curve is referred to a wavelength. The values of n have been numerically estimated to minimize the CIELAB ΔE_{ab}^* under the D65 illuminant between the measured and computed reflectance spectra of the cyan ramp. The plot shows that the wavelength dependence of ink concentration for the HP printer cannot be reduced by setting an optimum value for n .

Table 2 Dot gain parameters for computing effective ink concentration using Eq. (4).

Index (p)	Neugebauer primary	Dot gain parameters			
1	K	$C_{ck,\lambda}$	$M_{mk,\lambda}$	$Y_{yk,\lambda}$	$K_{k,\lambda}$
2	Y	$C_{cy,\lambda}$	$M_{my,\lambda}$	$Y_{y,\lambda}$	$K_{ky,\lambda}$
3	YK	$C_{cyk,\lambda}$	$M_{myk,\lambda}$	$Y_{yk,\lambda}$	$K_{ky,\lambda}$
4	M	$C_{cm,\lambda}$	$M_{m,\lambda}$	$Y_{ym,\lambda}$	$K_{km,\lambda}$
5	MK	$C_{cmk,\lambda}$	$M_{mk,\lambda}$	$Y_{ymk,\lambda}$	$K_{km,\lambda}$
6	R	$C_{cmy,\lambda}$	$M_{my,\lambda}$	$Y_{ym,\lambda}$	$K_{kmy,\lambda}$
7	RK	$C_{cmyk,\lambda}$	$M_{myk,\lambda}$	$Y_{ymk,\lambda}$	$K_{kmy,\lambda}$
8	C	$C_{c,\lambda}$	$M_{mc,\lambda}$	$Y_{yc,\lambda}$	$K_{kc,\lambda}$
9	CK	$C_{ck,\lambda}$	$M_{mck,\lambda}$	$Y_{yck,\lambda}$	$K_{kc,\lambda}$
10	G	$C_{cy,\lambda}$	$M_{cmy,\lambda}$	$Y_{yc,\lambda}$	$K_{kcy,\lambda}$
11	GK	$C_{cyk,\lambda}$	$M_{cmyk,\lambda}$	$Y_{yck,\lambda}$	$K_{kcy,\lambda}$
12	B	$C_{cm,\lambda}$	$M_{mc,\lambda}$	$Y_{cmy,\lambda}$	$K_{kcm,\lambda}$
13	BK	$C_{cmk,\lambda}$	$M_{mck,\lambda}$	$Y_{cmyk,\lambda}$	$K_{kcm,\lambda}$
14	CMY	$C_{cmy,\lambda}$	$M_{cmy,\lambda}$	$Y_{cmy,\lambda}$	$K_{kcmk,\lambda}$
15	CMYK	$C_{cmyk,\lambda}$	$M_{cmyk,\lambda}$	$Y_{cmyk,\lambda}$	$K_{kcmk,\lambda}$

2.2.4 Training set

To effectively tune the model, the training set must be customized to employ all the parameters. At the same time we want to have as few color patches to print and measure for printer characterization as possible. The training set consists therefore of ramps of eleven patches, ranging from the absence of ink to full ink coverage of cyan, magenta, yellow, red, green, blue, black, cyan with black, magenta with black, yellow with black, red with black, green with black and blue with black, for a total of 143 samples (Fig. 2).

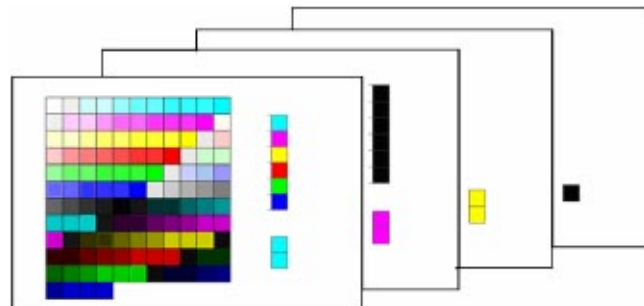


Fig. 2 Specimens (133 patches) for the training set and for the measurement of the Neugebauer primaries.

2.3 Estimation of the Printer Model Parameters Using Genetic Algorithms

The choice of the representation of candidate solutions (individuals) and the fitness function used to evaluate individuals are crucial factors in the effectiveness of this genetic approach. The genetic material, called the genome, of each individual must consist of the minimum amount of data required to represent a solution to the problem.

In the printer model we have introduced:

- the Yule–Nielsen factor [Eq. (1)],
- U for the printer driver model [Eq. (2)], and
- and 32 parameters for the dot gain functions (Table 2).

Reflectance spectra have been sampled at intervals of 40 nm in the range from 400 to 680 nm, producing 8 samples. The dot gain parameters are wavelength dependent, giving us a total of 258 real numbers.

Since the genome is an array of real numbers, a range must be specified. And since we consider that the dot gain functions do not alter the theoretical value of ink concentration by more than some 30%, consequently the range for real parameters has been set at [0.3;3.0]. The theoretical value of the Yule–Nielsen factor ranges from 1.0, corresponding to the absence of scattering, to 2.0, corresponding to Lambertian or perfectly diffused scattering, with the assumption that the dots are rectangular in cross section. In reality the dots have soft transitions and, in cases of high frequency rotated screens, or error diffusion, much of the paper is covered by transitory regions. In these cases, the Yule–Nielsen factor experimentally computed, may exceed the theoretical limit of 2.²⁸ In our experiment, we have considered a range of [1.0;12.0] for the Yule–Nielsen factor. The fitness function is computed as

$$\text{fitness} = \frac{1}{S} \left(\sum_{s=1}^S \frac{1}{\Gamma} \left(\sum_{\lambda=1}^{\Gamma} (R_{\text{print},\lambda,s} - R_{\text{meas},\lambda,s})^2 \right) \right), \quad (6)$$

where S is the number of elements in the training set ($S = 143$), Γ is the number of wavelengths ($\Gamma = 8$), and $R_{\text{print},\lambda}$ is computed with Eq. (1).

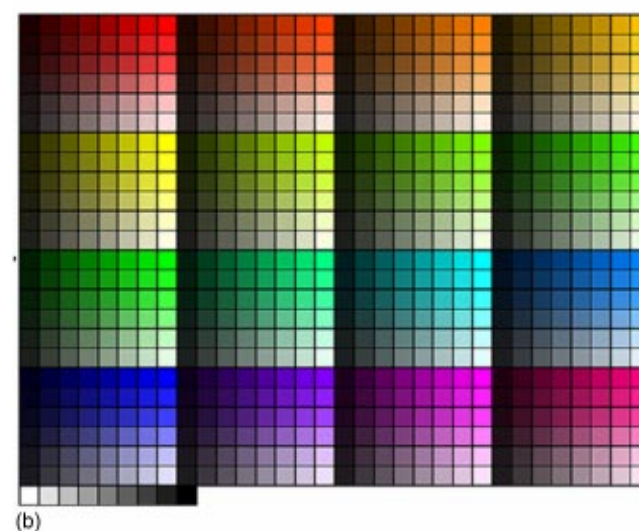
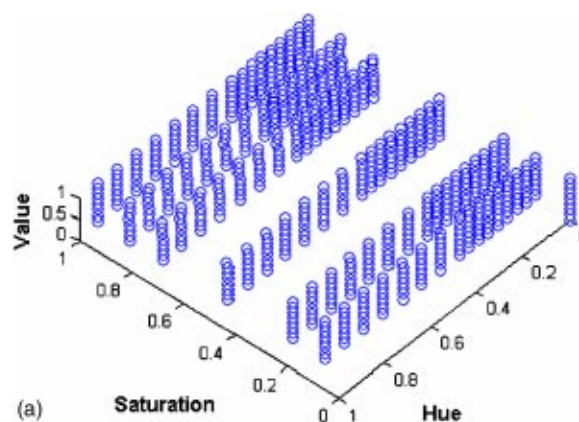


Fig. 3 HSV diagram and image of the 777 colors in test set specimen.

We have used the “simple” genetic algorithm in the Galib library.²⁹ It employs non-overlapping populations: at each generation the algorithm creates a new population of individuals by selecting parents from the previous population, and then mating to produce the new offspring. The

Table 3 Statistics of color distances and spectra differences for the training set (m =mean, M =maximum, sdv =standard deviation). rms is the root mean square error between measured and computed reflectance spectrum.

	Epson 890 photo quality paper Linux driver			Epson Stylus Color photo quality paper Linux driver			HP 2000C plain paper Windows driver			Epson Stylus C80 plain paper Windows driver		
	m	M	sdv	m	M	sdv	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.30	7.63	1.80	1.49	6.47	1.15	2.29	8.02	1.67	1.67	5.93	1.21
ΔH	0.93	4.61	1.02	0.70	3.67	0.74	0.87	3.29	0.82	0.78	3.33	0.71
ΔL	0.96	6.25	1.01	0.52	3.27	0.56	0.79	3.56	0.73	0.92	5.24	1.11
ΔE_{94}^*	1.85	6.95	1.43	1.18	4.85	0.91	1.71	5.77	1.25	1.47	5.85	1.18
CIEDE2000	1.59	5.60	1.25	1.02	3.69	0.87	1.49	5.39	1.15	1.30	6.31	1.15
rms	0.86	3.43	0.73	0.65	2.34	0.41	1.12	3.23	0.82	0.96	4.69	0.68
Fitness		1.26			0.59			1.92			1.38	

Table 4 Statistics of the color distances and spectra differences for the Test Set (m =mean, M =maximum, sdv =standard deviation). rms is the root mean square error between measured and computed reflectance spectrum.

	Epson 890 photo quality paper Linux driver			Epson Stylus Color photo quality paper Linux driver			HP 2000C plain paper Windows driver			Epson Stylus C80 plain paper Windows driver		
	m	M	sdv	m	M	sdv	m	M	sdv	m	M	sdv
ΔE_{ab}^*	3.21	11.0	2.05	1.92	7.99	1.22	5.4	18.5	3.12	2.74	8.51	1.45
ΔH	1.36	7.87	1.26	0.87	6.31	0.97	2.01	12.2	1.88	0.97	6.00	0.76
ΔL	1.70	7.97	1.64	0.9	3.41	0.71	1.74	6.62	1.31	1.74	7.98	1.56
ΔE_{94}^*	2.78	8.54	1.72	1.64	6.03	0.95	4.09	11.0	1.95	2.51	8.45	1.42
CIEDE2000	2.41	7.22	1.39	1.50	6.48	0.89	3.41	8.87	1.56	2.30	7.24	1.35
rms	1.42	4.49	0.85	0.90	5.82	0.56	2.24	12.2	1.34	1.72	5.37	1.15

best individual from each generation is also carried over into the next generation (elitism). The probability of cross-over and mutation are set at 0.9 and at 0.002, respectively. Selection is based on the “roulette wheel” method. The initial population of 12 individuals is randomly selected; only the initial value of parameter U in the conversion from RGB to CMYK [Eq. (2)] is initialized at 1. The stopping criterion is the number of iterations performed: we have set this at 4000 iterations.

3 Experiment and Results

We have applied our model in the characterization of four printers:

- Epson Stylus Color,
- HP 2000C,
- Epson Stylus C80, and
- Epson 890.

For the Epson Stylus Color and the Epson 890 a Linux driver was used, which employed Floyd Steinberg dithering and Epson Photo Quality paper; for the Epson Stylus C80 and the HP 2000C, the drivers were those supplied by the

printer manufacturers for Windows 2000, disabling any color management or color enhancement, and using plain paper.

The characterization procedure started with the printing and measurement of the Neugebauer primaries and the training set (Fig. 2). The Neugebauer primaries were obtained by measuring the printed inks at full coverage, and their overprints, by successive prints on the same sheet. The test set consisted of 777 samples, uniformly distributed in the RGB color space (Fig. 3).

The spectra were measured with a Gretag Spectrolino, considering values in the wavelength range from 400 to 680 nm with a step of 40 nm. Reflectance spectra lie in the range of [0;100]. The results are reported in terms of color difference in CIELAB ΔE_{ab}^* , CIELAB ΔE_{94}^* ,³⁰ and CIEDE2000,³¹ hue and lightness difference,³² and root-mean-square error between measured and computed reflectance samples, in Tables 3 and 4.

We note that the ΔE_{94}^* values are small and quite similar for all four printers for the training set of data, indicating that the genetic algorithm can provide a good solution for training in diverse configurations. Results for the test set are more varied. The best performance was obtained with the Epson Stylus Color printer, employing the Linux driver

Table 5 Statistics for the Epson 890 printer using photo quality paper, with both Windows and Linux driver (m =mean, M =maximum, sdv =standard deviation). rms is the root mean square error between measured and computed reflectance spectrum.

	Epson 890—Training Linux driver			Epson 890—Training Windows driver			Epson 890—Test Linux driver			Epson 890—Test Windows driver		
	m	M	sdv	m	M	sdv	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.30	7.63	1.80	3.23	8.97	2.45	3.21	11.0	2.05	4.85	11.41	2.15
ΔH	0.93	4.61	1.02	1.43	5.26	1.35	1.36	7.87	1.26	2.42	10.25	1.77
ΔL	0.96	6.25	1.01	1.41	7.35	1.49	1.70	7.97	1.64	2.11	9.14	1.74
ΔE_{94}^*	1.85	6.95	1.43	2.57	7.87	2.06	2.78	8.54	1.72	4.10	9.67	1.90
CIEDE2000	1.59	5.60	1.25	2.16	8.53	2.00	2.41	7.22	1.39	3.65	8.68	1.89
rms	0.86	3.43	0.73	1.14	4.16	0.75	1.42	4.49	0.85	1.70	5.32	0.76
Fitness		1.26			1.87							

Table 6 Statistics for the Epson C80 printer using a Windows 2000 driver, with both quality and plain paper (m =mean, M =maximum, sdv =standard deviation). rms is the root mean square error between measured and computed reflectance spectrum.

	Epson Stylus C80—Training photo quality paper			Epson Stylus C80—Training plain paper			Epson Stylus C80—Test photo quality paper			Epson Stylus C80—Test plain paper		
	m	M	sdv	m	M	sdv	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.19	5.47	1.49	1.67	5.93	1.21	3.79	13.5	2.04	2.74	8.51	1.45
ΔH	0.93	4.71	0.86	0.78	3.33	0.71	2.04	10.3	1.86	0.97	6.00	0.76
ΔL	0.84	3.32	0.86	0.92	5.24	1.11	1.48	5.11	1.22	1.74	7.98	1.56
ΔE_{94}^*	1.70	4.17	1.14	1.47	5.85	1.18	3.17	9.48	1.61	2.51	8.45	1.42
CIEDE2000	1.59	4.67	0.97	1.30	6.31	1.15	2.86	7.86	1.39	2.30	7.24	1.35
rms	0.92	2.66	0.55	0.96	4.69	0.68	1.47	4.45	0.61	1.72	5.37	1.15
Fitness		1.15			1.38							

and photo quality paper. The worst is with the HP 2000C printer, employing its Windows 2000 driver and plain paper.

The model's performance, we see, can be strongly influenced by the printer driver, to the point that it may be impossible for the training set to provide a full description of the printer's behavior. We report in Table 5 the statistics for the Epson 890 printer with photo quality paper and both Windows and Linux drivers. Results indicate that the performance of the proposed method may be substantially influenced by the strategy adopted to perform the conversion from RGB to inks amounts. But the type of paper also plays a role. We report in Table 6 the statistics for the Epson C80 printer using a Windows 2000 driver with both photo quality and plain paper. The results are again quite similar for the training set, but worse for the test set when photo quality paper is used.

4 Conclusions

We have experimented an innovative approach to the spectral characterization of binary printers. Our objective was to see whether it is feasible to approach the complex problem of printer spectral modeling by introducing a set of parameters into the mathematical framework of the Yule–Nielsen Spectral Neugebauer equation. Considering the complex interaction between these parameters, we also tested to see whether stochastic optimization could be used to produce an acceptable solution. If this were so, with our method the spectral characterization of a generic binary printer would require only some 130 measurements performed manually by an operator, without any of the equipment needed to measure the large number of samples required for colorimetric printer characterization based on interpolation. Our results indicate that characterization is possible, with a mean ΔE_{94}^* ranging from 1.64 to 4.09, and a maximum hue difference of 12.2.

Some key aspects of our work should be noted. First, we have assumed that any overprinting of inks can be predicted on the basis of model optimization performed on a small set of samples. If this method were not correct, there would be large shifts in hue in the test set prediction, especially when using a stochastic method for parameter estimation. Second, we have made no assumption concerning the sequence of printing. Third, the printers have been

treated as RGB devices, and we have therefore included printer-driver operations in our model. In fact, it is often impossible to directly command the printer in terms of CMYK ink amounts, moreover in many cases the actual conversion algorithm between the input RGB digital counts and the CMYK amounts is not known. To address these problems we employed a simple, but plausible, model to relate RGB to CMYK in which the only free parameter is simultaneously estimated with the actual printer model parameters during the training phase. The work presented can also be seen as a combination of analytical and empirical modeling. The high number of degrees of freedom of the proposed model gives a great flexibility to the characterization method; its modular structure would allow an easily adoption of more complex and complete RGB to CMYK models, if they would be available.

References

1. B. Hill, "Color capture, color management and the problem of metamerism: does multispectral imaging offer the solution?" *IS&T/SPIE Conf. on Color Imaging: Device-Independent Color, Color Hardcopy and Graphics Arts V*, SPIE (2000).
2. R. S. Berns, F. H. Imai, P. D. Burns, and D. Tzeng, "Multispectral-based color reproduction research at the Munsell Color Science Laboratory," *Proc. SPIE* **3409**, 14–25 (1998).
3. R. Bala, "Device characterization," Chap. 5, *Digital Color Imaging Handbook*, G. Sharma, Ed., CRC Press, Boca Raton, Florida (2003).
4. D. R. Wyble and R. S. Berns, "A critical review of spectral models applied to binary color printing," *Color Res. Appl.* **25**(1), 4–19 (2000).
5. K. Iino and R. S. Berns, "Building color management modules using linear optimization I. Desktop color system," *J. Imaging Sci. Technol.* **42**, 79–94 (1998).
6. S. Gustavson, "The color gamut of halftone reproduction," *The Fourth IS&T/SID Color Imaging Conference*, 19–22 (1996).
7. J. S. Arney, T. Wu, and C. Blehm, "Modeling the Yule–Nielsen effect on color halftones," *Proc. of The Fifth IS&T/SID Color Imaging Conference*, 62–65 (1997).
8. J. S. Arney, T. Wu, and C. Blehm, "Modeling the Yule–Nielsen effect on color halftones," *J. Imaging Sci. Technol.* **42**(4), 335–340 (1998).
9. P. Emmel and R. D. Hersch, "A unified model for color prediction of halftoned print," *J. Imaging Sci. Technol.* **44**(4), 351–385 (2000).
10. P. Emmel and R. D. Hersch, "Modeling ink spreading for color prediction," *J. Imaging Sci. Technol.* **46**(3), (2002).
11. R. Rolleston and R. Balasubramanian, "Accuracy of various types of Neugebauer model," *Proc. of the First IS&T/SID Color Imaging Conference* (1993).
12. A. U. Agar and J. P. Allebach, "An iterative cellular YNSN method for color printer characterization," *Proc. of the Sixth IS&T/SID Color Imaging Conference*, 197–200 (1998).
13. R. Balasubramanian, "Optimization of the spectral Neugebauer model for printer characterization," *J. Electron. Imaging* **8**(2), 156–166 (1999).

14. M. Xia, G. Sharma, and A. M. Tekapl, "End-to-end color printer calibration by total least square regression," *IEEE Trans. Image Process.* **8**(5), 700–716 (1999).
15. D. Tzeng, "Spectral-based color separation algorithm development for multiple-ink color reproduction," PhD Thesis, R.I.T., Rochester, New York (1999).
16. D. Tzeng and R. S. Berns, "Spectral-based six-color separation minimizing metamerism," *The Eighth IS&T/SID Color Imaging Conference* (2000).
17. L. A. Taplin and R. S. Berns, "Spectral color reproduction based on a six-color inkjet output system," *Proc. of Ninth Color Imaging Conference: Color Science and Engineering, Systems, Technologies and Applications*, IS&T, Springfield, Massachusetts, 209–213 (2001).
18. M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA (1996).
19. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, Boston, MA (1989).
20. S. Mardle and S. Pascoe, "An overview of genetic algorithms for the solution of optimisation problems," *Comput. Higher Educ. Econ. Rev.* **13**(1), (1999).
21. G. L. Rogers, "Neugebauer revisited: Random dots in halftone screening," *Color Res. Appl.* **23**(2), 104–113 (1998).
22. I. Amidror and R. D. Hersch, "Neugebauer and Demichel: Dependence and independence in n-screen superpositions for colour printing," *Color Res. Appl.* **25**(4), 267–277 (2000).
23. R. Balasubramanian, "A printer model for dot-on-dot halftone screens," *Proc. SPIE* **2413**, 356–364 (1995).
24. H. R. Kang, *Color Technology for Electronic Imaging Devices*, SPIE, Bellingham, Washington (1997).
25. P. Roberts, "A mathematical model for inkjet printer characterization," *The Eighth IS&T/SID Color Imaging Conference* (2000).
26. E. J. Stollnitz, V. Ostromoukhov, and D. H. Salesin, "Reproducing color images using custom inks," *Proc. of SIGGRAPH'98, Computer Graphics Proceedings, Annual Conference Series*, 267–274 (1998).
27. K. Iino and R. S. Berns, "A spectral based model of color printing that compensates for optical interactions of multiple inks," *Proc. 8th Congress International Colour Association* (1997).
28. J. A. Stephen Viggiano, "Modeling the color of multi-colored halftones," *Proc. TAGA*, 44–62 (1990).
29. Galib: A C++ Library of Genetic Algorithm Components, <http://lancet.mit.edu/ga/>
30. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, Massachusetts (1998).
31. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.* **26**(5), 340–350 (2001).
32. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., Wiley, New York (1982).

been a research scientist in the Construction Technologies Institute, CNR, Milan, Italy. Her current research interests include cross-media color reproduction, multispectral imaging, and spectral-based printer characterization.



Raimondo Schettini is an associate professor at DISCo, University of Milano Bicocca where he is in charge of the Imaging and Vision Lab. He has been associated with Italian National Research Council (CNR) since 1987. He has been team leader in several research projects and published more than 140 refereed papers on image processing, analysis, and reproduction, and on image content-based indexing and retrieval. He is an associate editor of the Pattern Recognition Journal. He was a co-guest editor of three special issues about internet imaging (Journal of Electronic Imaging, 2002), color image processing and analysis (Pattern Recognition Letters, 2003), and color for image indexing and retrieval (Computer Vision and Image Understanding, 2004). He was general co-chairman of the 1st Workshop on Image and Video Content-based Retrieval (1998), of the First European Conference on Color in Graphics, Imaging and Vision (2002), and of the EI Internet Imaging Conferences (2000 to 2005).



Giancarlo Mauri is a full professor of computer science at the University of Milano-Bicocca. His research interests are mainly in the area of theoretical computer science, and include: formal languages and automata, computational complexity, computational learning theory, soft computing techniques, cellular automata, bioinformatics, and molecular computing. On these subjects, he has published more than 150 scientific papers in international journals, contributed volumes, and conference proceedings.



Silvia Zuffi was born in Rimini, Italy, on July 3, 1969. She graduated magna cum laude in 1995, with the Laurea in Electronic Engineering at University of Bologna, Italy. From 1995 through 1997 she was a Computer Software Engineer, then she joined the Movement Analysis Laboratory at the Istituti Ortopedici Rizzoli, Bologna, Italy. From May 1999 through December 2001 she was with the Italian National Research Council (CNR) in the Institute for Multimedia Information Technologies in Milan, Italy. Since 2002 she has