Recovering Neugebauer colorant reflectances and ink spreading curves from printed color images

Abstract

Spectral reflection prediction models, although effective, are impractical for certain industrial applications such as self-calibrating devices and online monitoring because their calibration requires specific color-constant calibration patches. Using the CMYK Ink Spreading enhanced Yule-Nielsen modified Spectral Neugebauer model (IS-YNSN), we propose a method to recover the colorant reflectances (Neugebauer primaries), the ink spreading curves, and the Yule-Nielsen n-value using only tiles extracted from printed color images. There is no prior knowledge about the reproduction device. Thanks to a set of constraints based on Principal Component Analysis (PCA) and the relationships between composed Neugebauer primaries and the ink transmittances, good approximations of the Neugebauer primaries are achieved. These approximations are then optimized, yielding an accurately calibrated IS-YNSN model comparable to one obtained by classical calibrations. A detailed analysis of these calibrations shows that 25 well-chosen CMYK image calibration tiles are sufficient to accurately recover both the Neugebauer primaries and the ink spreading curves.

Keywords: color prints, color reproduction, dot gain, halftones, ink spreading, prediction model calibration, spectral reflection prediction, image tiles, Neugebauer primaries

1   Introduction

The characterization of printing devices is part of color management systems aiming at ensuring the consistency of colors across different color input and output devices [1]. Characterizing a printing device consists in mapping the control values of the device to colors represented in a device-independent color space. The control values specify the amount of inks to be printed, further referred to as the *nominal ink surface coverages*. The independent color space is usually either the CIELAB or the CIE-XYZ colorimetric space. A spectral reflection prediction model can be used to characterize a printing device [2]. Such a model comprises mathematical equations modeling the interaction between the incident light, the inks, and the paper substrate. For given nominal ink coverages, a spectral prediction model predicts the reflection spectra of printed patches. These reflection spectra can then be converted into CIELAB or CIE-XYZ.

The Yule-Nielsen modified spectral Neugebauer model (YNSN) is a widely used spectral reflection prediction model [2]-[5]. The YNSN model incorporates parameters which must be learned for each combination of paper, inks, and printing device. This step, further referred to as the *calibration* of the model, consists in determining the best values for the different model parameters. The calibration of the YNSN model requires the spectral measurements of predetermined color patches representing the Neugebauer primaries, i.e. the reflectances of paper, of the solid inks, and of superpositions of the solid inks, as well as a few halftones to determine the Yule-Nielsen n-value. The YNSN model actually performs a linear interpolation between the extremities of the space formed by the reflectances of the Neugebauer primaries raised to the power of $1/n$. In order to further improve its accuracy, the YNSN model has been enhanced with an ink spreading model accounting for physical dot gain (IS-YNSN, [6]). The calibration of the IS-YNSN model requires the spectral measurements of a set of predetermined halftone patches.

In the case of a 4 ink IS-YNSN prediction model, 36 spectral measurements of predetermined color-constant halftone and solid patches are required. Since these color-constant patches are not present

within normal printed images, they need to be added, for example in the margins of printed pages. Once measured, they must be removed. Instead of calibrating the IS-YNSN model with predetermined color-constant patches, we propose a calibration procedure relying on the measurements of image tiles extracted from printed color images. In previous contributions of the present authors [7]-[9], the content of printed color images has already been used for the calibration of ink spreading, but not for the YNSN model itself, i.e. the Neugebauer primaries still had to be measured and the Yule-Nielsen n-value determined. Balasubramanian proposed to recalibrate the Neugebauer primaries based on a training set of measured samples, but only as an optimization step based on initially measured primary reflectances [10].

In the current contribution, we show how to recover the Neugebauer primaries from printed color images. We propose a two-step procedure to calibrate both the YNSN model and the ink spreading model using tiles extracted from printed color images without relying on prior knowledge. In the first step, constraints are imposed both on the ink spreading model and on the Neugebauer primaries to prevent large variations when the information contained in the image calibration tiles is not sufficient. In the second step, the Neugebauer primaries are refined to improve the calibration accuracy. We then show how to determine the Yule-Nielsen n-value.

The remainder of this paper is organized as follows. We first detail in Section 2 the Yule-Nielsen modified Spectral Neugebauer model (YNSN) and in Section 3 the ink spreading enhanced YNSN model (IS-YNSN). Section 4 proposes an unconstrained gradient-descent algorithm to fully calibrate the IS-YNSN model. In Section 5, we explain the limitations of the unconstrained approach and introduce a set of constraints to calibrate both the reflection spectra of the colorants (Neugebauer primaries) and the ink spreading curves. In Section 6, the colorant spectra are refined to improve the accuracy of the calibration. The algorithm selecting image calibration tiles from printed color images is described in Section 7. Section 8 describes the measurement setup. Section 9 verifies the accuracy of the fitted colorant reflectances and Section 10 compares the prediction accuracy according to different calibration procedures. The calibration of the Yule-Nielsen n-value is discussed in Section 11 and the importance of accurately calibrating the reflectances of the Neugebauer primaries is underlined in Section 12. The conclusions are drawn in Section 13.

2 The Yule-Nielsen modified Spectral Neugebauer Model (YNSN)

One of the first color prediction models is the Neugebauer model [11]. In its original form, it predicts the RGB values of a color halftone patch. Yule and Nielsen modified this model to account for optical dot gain [3], Yule and Colt applied it for CIE-XYZ tri-stimulus values [12], and Viggiano extended it to spectral reflectances [4]. This model, now known as the Yule-Nielsen modified Spectral Neugebauer model (YNSN), predicts the reflection spectra of color-constant patches whose ink surface coverages are given. The equation used to perform a spectral reflectance prediction is:

$$R(\lambda) = \left(\sum_i a_i R_i(\lambda)^{1/n} \right)^n \tag{1}$$

where $R(\lambda)$ is the predicted reflection spectrum, $R_i(\lambda)$ the reflection spectra of the colorants (Neugebauer primaries), a_i the relative area coverages of the colorants, and n the Yule-Nielsen n-value. In a classical calibration setup, the reflection spectra of the colorants are measured and the scalar n-value is calibrated by relying on a limited number of color-constant halftone patches (Section 11). In the following sections, the n-value is considered to be known.

The relative area coverages of the colorants are computed from the relative ink surface coverages using the Demichel equations. For four inks, the Demichel equations are the following:

$$
\begin{aligned}
a_w &= (1-c)\cdot(1-m)\cdot(1-y)\cdot(1-k) & a_k &= (1-c)\cdot(1-m)\cdot(1-y)\cdot k \\
a_c &= c\cdot(1-m)\cdot(1-y)\cdot(1-k) & a_{ck} &= c\cdot(1-m)\cdot(1-y)\cdot k \\
a_m &= (1-c)\cdot m\cdot(1-y)\cdot(1-k) & a_{mk} &= (1-c)\cdot m\cdot(1-y)\cdot k \\
a_{cm} &= c\cdot m\cdot(1-y)\cdot(1-k) & a_{cmk} &= c\cdot m\cdot(1-y)\cdot k \\
a_y &= (1-c)\cdot(1-m)\cdot y\cdot(1-k) & a_{yk} &= (1-c)\cdot(1-m)\cdot y\cdot k \\
a_{cy} &= c\cdot(1-m)\cdot y\cdot(1-k) & a_{cyk} &= c\cdot(1-m)\cdot y\cdot k \\
a_{my} &= (1-c)\cdot m\cdot y\cdot(1-k) & a_{myk} &= (1-c)\cdot m\cdot y\cdot k \\
a_{cmy} &= c\cdot m\cdot y\cdot(1-k) & a_{cmyk} &= c\cdot m\cdot y\cdot k
\end{aligned}
\tag{2}
$$

where c, m, y and k are the ink surface coverages.

When calibrated, the YNSN model is a function taking the ink surface coverages of a color halftone as input and yielding its reflection spectrum. It is possible to deduce ink coverages from a given spectrum by minimizing a difference metric between predicted and measured reflection spectra as follows:

$$
covs_{opt} = \arg\min_{covs}\left[\sum_k \left(predSpectrum(\lambda_k, covs) - R(\lambda_k)\right)^2\right]
\tag{3}
$$

where $covs = \{c, m, y, k\}$ are the ink coverages, $R(\lambda)$ the measured reflection spectrum, and *predSpectrum* the YNSN reflectance prediction function combining Eqs. (1) and (2).

Due to the physical dot gain, ink coverages deduced using Eq. (3) are usually larger than the corresponding nominal ink coverages. As a consequence, spectral predictions made from nominal ink coverages are not accurate. Therefore, an enhanced YNSN model was proposed which incorporates an ink spreading model [13][6].

3   The Ink Spreading Enhanced YNSN Model (IS-YNSN)

When a reproduction device reproduces a color with given nominal ink coverages by depositing inks on paper, the inks spread out on the paper, resulting in effective coverages usually larger than the requested nominal ink coverages. This ink spreading phenomenon is called *mechanical* or *physical dot gain*. The amount of dot gain depends on whether the ink halftone is printed alone on paper or in superposition with one or more other inks [14]. The goal of the ink spreading model is to compute the *effective ink surface coverages* from the nominal ink surface coverages. These effective ink surface coverages are then used as input to the YNSN model.

3.1   Ink spreading curves

The proposed ink spreading model relies on ink spreading curves (Fig. 1a). An ink spreading curve maps the nominal surface coverages of an ink halftone into its effective surface coverages, i.e. to the surface that the ink halftone effectively covers after being printed. We can approximate an ink spreading curve by a parabola characterized by its mid-point v:

$$
\begin{aligned}
f_{i/jk}(u_i, v_{i/jk}) &= u_i + (4v_{i/jk} - 2)(1 - u_i)u_i \\
&= u_i^2(2 - 4v_{i/jk}) + u_i(4v_{i/jk} - 1)
\end{aligned}
\tag{4}
$$

where u_i is the nominal surface coverage of ink i, $f_{i/jk}$ is the ink spreading curve of ink i superposed with solid inks j and k, and $v_{i/jk} = f_{i/jk}(0.5)$ is the effective surface coverage at 50% nominal surface coverage, also called mid-point of the ink spreading curve [7]. Equation (4) represents a parabola that passes through the three points (0,0), (0.5, $v_{i/jk}$), and (1,1), as shown in Fig. 1a. The parabola is a monotonic increasing function if $v_{i/jk}$ is between 0.25 and 0.75.

There is one ink spreading curve for each ink halftone in each superposition condition. For example, a cyan halftone may be printed alone, *c*; superposed with solid magenta, *c/m*; with solid yellow, *c/y*;
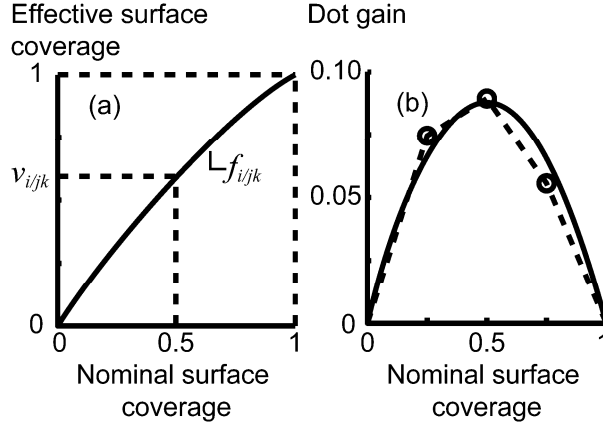
Fig. 1. (a) Example of an ink spreading curve $f_{i/jk}$ characterized by its mid-point $v_{i/jk}$. (b) Corresponding dot gain curve, defined as the difference between effective and nominal surface coverages, with the fitted dot gains of three calibration patches (circles), the linearly interpolated dot gain curve (dashed line), and the corresponding least squares approximated parabolic dot gain curve (solid line).

with solid black, c/k; with solid magenta and solid yellow, c/my; with solid magenta and solid black, c/mk; with solid yellow and solid black, c/yk; and with solid magenta, yellow and black, c/myk. There are 8 different ink spreading curves for each ink, yielding a total of 32 ink spreading curves. However, since any halftone superposed with solid black yields a reflection spectrum very close to the reflection spectrum of solid black, ink spreading curves where one ink halftone is superposed with solid black are discarded [6]. Table I lists the 20 ink spreading curves considered.

Table I. List of the considered ink spreading curve indicia.

Cyan	Magenta	Yellow	Black	
c	m	y	k	k/y
c/m	m/c	y/c	k/c	k/cy
c/y	m/y	y/m	k/m	k/my
c/my	m/cy	y/cm	k/cm	k/cmy

Fig. 1b shows the dot gain curve corresponding to the ink spreading curve of Fig. 1a, where dot gain is defined as the difference between effective and nominal coverages. Each ink spreading curve is calibrated using three calibration patches at 0.25, 0.5, and 0.75 nominal ink surface coverages. The nominal surface coverage of the ink halftone determines the position on the x-axis of the ink spreading curve. The corresponding effective surface coverage is fitted using Eq. (3) and determines the position on the y-axis (Fig. 1a). The parabola is then fitted by minimizing the least-squares difference between the fitted and predicted dot gains. An ink spreading curve can be entirely characterized by the halftone at 50% nominal surface coverage, but as shown is Fig. 1b, 3 calibration patches per ink spreading curve reduce the influence of printing and acquisition inaccuracies.

In order to avoid the ambiguity between the superposition of CMY inks and the black ink, the spectral measurements span both the visible wavelength range (380-730 nm) and a part of the near infrared (NIR) wavelength range (730-850 nm). The NIR wavelength range enables distinguishing the light absorbing pigment-based black ink from the superposition of CMY inks, which are dye-based and do not absorb light in the NIR wavelength range [15].

3.2 Ink Spreading Equations

In order to obtain the effective surface coverages of the ink dots forming a color halftone, we weight the contributions of the different ink spreading curves as follows. The effective coverage of halftone ink i superposed with inks j and k is the weighted average of the ink spreading functions f_i, $f_{i/j}$, $f_{i/k}$, and

$f_{i/jk}$. The weights are computed according to the surface coverages of the underlying colorants (Neugebauer primaries) formed by inks j and k. For CMYK, the ink spreading equations are:

$$
\begin{aligned}
c' &= (1-m') \quad (1-y') \quad f_c(c) & m' &= (1-c') \quad (1-y') \quad f_m(m) \\
&+ \quad m' \quad (1-y') \quad f_{c/m}(c) & &+ \quad c' \quad (1-y') \quad f_{m/c}(m) \\
&+ (1-m') \quad y' \quad f_{c/y}(c) & &+ (1-c') \quad y' \quad f_{m/y}(m) \\
&+ \quad m' \quad y' \quad f_{c/my}(c) & &+ \quad c' \quad y' \quad f_{m/cy}(m) \\[6pt]
y' &= (1-c') \quad (1-m') \quad f_y(y) & k' &= (1-c') \quad (1-m') \quad (1-y') \quad f_k(k) \\
&+ \quad c' \quad (1-m') \quad f_{y/c}(y) & &+ \quad c' \quad (1-m') \quad (1-y') \quad f_{k/c}(k) \\
&+ (1-c') \quad m' \quad f_{y/m}(y) & &+ (1-c') \quad m' \quad (1-y') \quad f_{k/m}(k) \\
&+ \quad c' \quad m' \quad f_{y/cm}(y) & &+ \quad c' \quad m' \quad (1-y') \quad f_{k/cm}(k) \\
& & &+ (1-c') \quad (1-m') \quad y' \quad f_{k/y}(k) \\
& & &+ \quad c' \quad (1-m') \quad y' \quad f_{k/cy}(k) \\
& & &+ (1-c') \quad m' \quad y' \quad f_{k/my}(k) \\
& & &+ \quad c' \quad m' \quad y' \quad f_{k/cmy}(k)
\end{aligned}
\tag{5}
$$

In Eqs. (5), when calculating the effective surface coverage c' of cyan, $(1-m')(1-y')$, $m'(1-y')$, $(1-m')$ y', and $m'y'$ express, respectively, the effective surface coverage of colorants white, magenta, yellow, and red superposed with the cyan halftone layer. Since a cyan, magenta, or yellow halftone over solid black is assumed to yield black, the superposition conditions corresponding to an ink halftone over solid black have been discarded [6]. We have therefore 4 superposition conditions for the cyan, magenta and yellow ink halftones and 8 superposition conditions for the black ink halftone.

We solve Eqs. (5) iteratively, starting by assigning the nominal ink halftone coverages (c, m, y, k) to the effective ink halftone coverages (c', m', y', k') [6]. Four to five iterations ensure sufficient convergence to determine the effective ink halftone surface coverages.

4   Unconstrained calibration of the IS-YNSN model

The IS-YNSN model needs to be calibrated for each combination of printer, halftone screens, inks, and paper. Such a calibration consists in recovering the reflection spectra of the colorants (Neugebauer primaries), fitting the mid-points of the ink spreading curves, and setting the Yule-Nielsen n-value to a suitable scalar value. In a classical calibration, the colorant reflectances are directly measured and the effective surface coverages used for creating the ink spreading curves are fitted using a predetermined set of color-constant halftone calibration patches.

Instead of using predetermined color-constant patches, we show how to calibrate the IS-YNSN model using image calibration tiles extracted from color images and grouped in an image calibration set. The selection of these tiles is described in Section 7. In this section, we propose an unconstrained procedure to calibrate the IS-YNSN model using an image calibration set. This approach relies on a gradient-descent algorithm to find the optimal colorant reflectances and the mid-points of the ink spreading curves. In the present setting, we assume that the Yule-Nielsen n-value is known. Its calibration is discussed in Section 11.

The gradient-descent algorithm uses as objective function the sum of square differences between the measured and predicted reflectance components of the image calibration tiles:

$$
f(C,\vec{v}) = \sum_p \sum_k [predSpectrum(\lambda_k, p, C, \vec{v}) - R_p(\lambda_k)]^2
\tag{6}
$$

where λ_k is the wavelength, p an image calibration tile, C the 48x16 matrix whose columns correspond to the 16 CMYK colorant reflectances sampled each 10 nm between 380 nm and 850 nm, $\vec{v}$ a vector containing the mid-points of the 20 CMYK ink spreading curves, and R_p the measured reflection spectrum of image calibration tile p. Function *predSpectrum* uses the IS-YNSN model to predict the spectrum of image calibration tile p.

The gradient-descent algorithm performs the following minimization using Eq. (6):

$$\{C_{opt} \quad \vec{v}_{opt}\} = \arg\min_{C,\vec{v}}[f(C,\vec{v})] \text{ such that} \begin{cases} R_{\min} \leq R_i \leq R_w \\ 0.25 \leq v_{i/jk} \leq 0.75 \end{cases} \tag{7}$$

where matrix C_{opt} and vector $\vec{v}_{opt}$ contain respectively the colorant reflectances and the mid-points of the ink spreading curves minimizing the sum of square differences between the measured and predicted reflectance spectra of the image calibration tiles. The only constraints imposed on the colorant reflectances and the mid-points of the ink spreading curves are their intrinsic bounds. In respect to the ink spreading curve mid-points, 0.25 and 0.75 are the bounds beyond which the quadratic ink spreading curves are not monotonic increasing functions anymore [Eq. (4)]. In respect to the colorant reflectances, $R_{\min}$ and R_w correspond respectively to the minimal value allowed for a reflectance and the maximal value expressed by the measured paper reflectance. The lower bound for reflectances is actually 0, but $R_{\min}$ is set to 0.001 to avoid searching for negative values. Moreover, on real prints, the darkest black has a reflectance larger than 0.001.

Calibrations of the IS-YNSN model using Eq. (7) are not reliable. Because the colorant reflectances (Neugebauer primaries) are not constrained, they are set to spectra that are not smooth and therefore do not look like real CMYK colorant reflectances. In addition, the relationship between the reflectances of the inks and the reflectances of their superpositions is not preserved. Finally, the image calibration tiles do not always provide enough information to accurately calibrate the colorant reflectances. For example, if the image calibration tiles are primarily composed of cyan ink, it is not possible to calibrate the yellow colorant. In such cases, unconstrained procedures such as Eq. (7) tend to induce large variations in some spectral components in order to influence the difference metric of Eq. (6). As a consequence, the fitted reflectances do not reflect real measurements. Examples of such fitted reflectances are given in Section 9.

5 Constrained calibration of the IS-YNSN model

In order to prevent the calibration procedure from setting the colorant reflectances and mid-points of the ink spreading curves to extreme values, we impose constraints on these parameters. For the ink spreading model, we use the constraints proposed in [7] that were successful in calibrating the ink spreading model by relying on tiles extracted from color images. For the colorants, we propose a set of constraints that forces the calibration procedure to set their reflectances to spectra that are similar to real reflection spectra.

5.1 Ink spreading constraints

The problem faced by the calibration procedure when fitting the mid-points of the ink spreading curves is that the image calibration tiles do not always provide sufficient information about each ink spreading curve in order to set it with high confidence to a sensible value. The goal of the constraints imposed on the ink spreading curves is therefore to restrict the range of values that can be assigned to the mid-points when there is not enough information.

Ink spreading weight $w_{i/jk}$, associated to ink spreading curve $f_{i/jk}$ of halftone ink i superposed with solid inks j and k, indicates if there is enough information to fit the mid-point $v_{i/jk}$ of ink spreading curve $f_{i/jk}$. A weight equal to 0 indicates that no information is available. In this case, the effective surface coverage is made equal to the nominal surface coverage and the mid-point is set to 0.5, i.e. we assume that there is no ink spreading. A weight equal to 1 indicates that the mid-point $v_{i/jk}$ can be freely calibrated between 0.25 and 0.75, the bounds beyond which the ink spreading curves are not monotonic increasing functions anymore. For a weight $w_{i/jk}$ between 0 and 1, the mid-point $v_{i/jk}$ is constrained as follows:

$$0.5 - 0.25 \cdot w_{i/jk} \leq v_{i/jk} \leq 0.5 + 0.25 \cdot w_{i/jk} \tag{8}$$

The ink spreading weights $w_{i/jk}$ are computed from the nominal ink coverages of the image tiles as follows. Given a tile p of the image calibration set, let $u_{i,p}$ be the nominal surface coverage and $u_{i,p}'(v_i, v_{i/j}, v_{i/k}, v_{i/jk})$ be the effective surface coverage of ink i within tile p computed using the ink spreading model. The weight associated to ink spreading curve $f_{i/jk}$ of halftone ink i superposed with solid inks j and k is defined as the following gradient:

$$w_{i/jk,p} = \frac{\partial u_{i,p}'}{\partial v_{i/jk}} \tag{9}$$

Since ink spreading curve $f_{i/jk}$ is fully determined by its mid-point $v_{i/jk}$, the gradient $w_{i/jk,p}$ in Eq. (9) expresses the influence of ink spreading curve $f_{i/jk}$ on the resulting effective surface coverage $u_{i,p}'$ of ink i within tile p.

Since a calibration set is composed of several image tiles, we define the weight associated to ink spreading curve $f_{i/jk}$ for the entire image calibration set as the maximum of the weights of all the tiles:

$$w_{i/jk} = \max_p[w_{i/jk,p}] = \max_p \frac{\partial u_{i,p}'}{\partial v_{i/jk}} \tag{10}$$

We take the maximum of the derivatives across all the tiles because tiles with a high weight have the largest influence on the metric minimized by the ink spreading curve mid-point fitting algorithm [Eq. (6)], i.e. when there is at least one high weight tile and low weight tiles, the ink spreading curve mid-point is mainly fitted by the high weight tile.

By inserting Eq. (4) into Eqs. (5), we can compute the gradients of Eq. (9) for each ink spreading curve. For example, the weight $w_{c/m}$ of the ink spreading curve $f_{c/m}$ of cyan halftones printed on solid magenta is:

$$w_{c/m} = \frac{\partial c'}{\partial v_{c/m}} = m'(1 - y')\frac{\partial f_{c/m}(c)}{\partial v_{c/m}} = m'(1 - y') \cdot 4c(1 - c) \tag{11}$$

where c', m' and y' are the effective surface coverages of the cyan, magenta, and yellow halftones, respectively, and c is the nominal surface coverage of the cyan halftone. Since neither the colorant reflectances nor the ink spreading curves are calibrated yet, the effective coverages m' and y' required in Eq. (11) cannot be computed. We use the nominal coverages m and y instead. This modification influences the weights, but since the difference between effective and nominal coverages is usually small, the weights are not largely affected and the ink spreading mid-point values can be accurately calibrated. This assertion is further discussed in Section 12.

5.2 Constraints in respect to colorant reflectances

The constraints imposed on the ink spreading curves rely on the existence of a default value for the mid-points $v_{i/jk} = 0.5$ that can be used when no information is available. There is no such default value for the colorant reflectances. Instead, we make the distinction between the following three groups of colorants: the paper, the inks, and the ink superpositions. We assume that the paper reflectance representing the white colorant can be measured. The white colorant is therefore known. We then impose different constraints on the ink reflectances and on the ink superposition reflectances.

5.2.1 Ink reflectance constraints

The reflectance of an ink depends on the substrate it is printed on. The first constraint imposed on ink reflectance R_i expresses this dependence by assuming that light reflected by the print is attenuated twice by the ink layer and once by being reflected by the paper:

$$R_i = R_w T_i^2 \tag{12}$$

where R_i is the reflectance of ink i on paper, R_w the paper reflectance, and T_i the transmittance of ink i.

Using Eq. (12), ink reflectances are expressed by the corresponding ink transmittances. We can therefore impose constraints on these transmittances. The spectrum of a transmittance should be smooth and should have a general shape given by the color of the ink. Using principal component analysis (PCA, [16][17]), we can constrain ink transmittances by limiting the number of principal components expressing these transmittances while retaining both smoothness and general shape of their spectra. For each ink color, we assemble a list of ink transmittances based on measurements coming from offset prints, ink jet prints, thermal transfer prints, etc. We then extract the average transmittance as well as the first three principal vectors from these ink transmittances. The principal vectors for ink i are stored in a matrix W_i. W_i has therefore three columns. The transmittance of an ink is then:

$$T_i = W_i \begin{bmatrix} c_{i,1} & c_{i,2} & c_{i,3} \end{bmatrix}^T + \overline{T}_i \tag{13}$$

where T_i is the transmittance of ink i; W_i is the matrix holding the principal vectors for the transmittances of ink i; $\overline{T}_i$ is the average transmittance of ink i; and $c_{i,1}$ to $c_{i,3}$ are three scalars, further referred to as the principal components of ink i. Using Eqs. (12) and (13), the reflectance R_i of ink i can be expressed using only three principal components, i.e. $c_{i,1}$ to $c_{i,3}$. The 12 principal components required for CMYK are listed in Table II. Appendix A provides more details about the accuracy of the implemented PCA.

5.2.2 Ink superposition reflectance constraints

The reflectance of superposed inks depends on the substrate on which they are printed as well as on the individual ink transmittances. However, a simple extension of Eq. (12) to multiple inks is not appropriate. Hersch *et al.* have shown that the relative thickness of a printed solid ink layer depends on the presence of other superposed inks [18]. Beer's law enables expressing these ink thickness variations. We adapt Eq. (12) to account for modified relative ink thicknesses:

$$R_{ij} = R_w T_i^{2d_{Ij}} T_j^{2d_{iJ}} \tag{14}$$

$$R_{ijk} = R_w T_i^{2d_{Ijk}} T_j^{2d_{iJk}} T_k^{2d_{ijK}} \tag{15}$$

$$R_{ijkl} = R_w T_i^{2d_{Ijkl}} T_j^{2d_{iJkl}} T_k^{2d_{ijKl}} T_l^{2d_{ijkL}} \tag{16}$$

where R_{ij}, R_{ijk}, and R_{ijkl} are the reflectances of the superposition of 2, 3, and 4 solid inks, respectively; R_w is the paper reflectance; T_i, T_j, T_k, and T_l are the transmittances of inks i, j, k, and l computed by Eq. (13); and d_{Ij} and d_{iJ} are the scalar ink thickness variation factors of ink i superposed with ink j and of ink j superposed with ink i. For four inks, we have 6 colorants that are composed of two inks, 4 colorants composed of 3 inks, and a single colorant composed of all 4 inks. Based on Eqs. (14) to (16), only the 28 scalar ink thickness variation factors listed in Table II are required to model the 11 colorant reflectances composed of 2 or more inks.

Table II. List of the 12 principal components and 28 ink thickness variation factors required for CMYK prints. All these parameters are scalar values.

Ink	Principal components	Ink thickness variation factors
Cyan	$c_{c,1}$, $c_{c,2}$, $c_{c,3}$	d_{Cm}, d_{Cy}, d_{Ck}, d_{Cmy}, d_{Cmk}, d_{Cyk}, d_{Cmyk}
Magenta	$c_{m,1}$, $c_{m,2}$, $c_{m,3}$	d_{cM}, d_{My}, d_{Mk}, d_{cMy}, d_{cMk}, d_{Myk}, d_{cMyk}
Yellow	$c_{y,1}$, $c_{y,2}$, $c_{y,3}$	d_{cY}, d_{mY}, d_{Yk}, d_{cmY}, d_{cYk}, d_{mYk}, d_{cmYk}
Black	$c_{k,1}$, $c_{k,2}$, $c_{k,3}$	d_{cK}, d_{mK}, d_{yK}, d_{cmK}, d_{cyK}, d_{myK}, d_{cmyK}

5.3 Constrained gradient-descent calibration of the IS-YNSN model

Using the constrained colorant reflectances expressed by Eqs. (12) to (16), we can modify the objective function of Eq. (6) as follows:

$$f(\vec{v},\vec{c},\vec{d}) = \sum_p \sum_k [predSpectrum(\lambda_k, p, \vec{v}, \vec{c}, \vec{d}) - R_p(\lambda_k)]^2 \qquad (17)$$

where we have added $\vec{c}$, the vector containing the principal components of the inks, and $\vec{d}$, the vector containing the ink thickness variation factors of the ink superpositions.

Finally, using Eq. (17) and the ink spreading constraints of Eq. (8), the calibration procedure becomes:

$$\{\vec{v}_{opt},\vec{c}_{opt},\vec{d}_{opt}\} = \underset{\vec{v},\vec{c},\vec{d}}{\arg\min}[f(\vec{v},\vec{c},\vec{d})] \text{ such that } \begin{cases} v_{i/jk} \leq 0.5 + 0.25 \cdot w_{i/jk} \\ v_{i/jk} \geq 0.5 - 0.25 \cdot w_{i/jk} \\ R_{min} \leq R_i \leq R_w \end{cases} \qquad (18)$$

where $\vec{v}_{opt}$, $\vec{c}_{opt}$, and $\vec{d}_{opt}$ are respectively the optimal mid-points of the ink spreading curves, optimal principal components of the ink transmittances, and optimal ink thickness variation factors. These optimal values minimize the sum of square differences between the measured and predicted spectra. Note that we maintain the intrinsic bounds for the colorants expressed by R_{min} and R_w, i.e. the minimal value allowed for a reflectance and the maximal value expressed by the paper white.

The colorant reflectances recovered from the optimal principal components and optimal ink thickness variation factors fitted in Eq. (18) are close to the measured colorant reflectances if there is enough information to fit them. However, these recovered colorant reflectances are subject to strong constraints and cannot exactly match the measured colorant reflectances, even when there would be enough information in the image calibration set to achieve such a match. We therefore propose a second calibration step to improve the recovered colorant reflectances.

6 Surface coverage based optimization of the colorant reflectances

The constraints imposed on the colorant reflectances in Eqs. (12)-(16) enable the first calibration step in Eq. (18) to recover colorant reflectances similar to real reflectances even when there is not enough information in the image calibration set. However, because of the PCA, the recovered colorant

reflectances are restricted to a limited set of reflection spectra. If the real reflectances lay outside of this set, it is not possible to optimally fit the colorant reflectances to match these real reflectances and the IS-YNSN model will not be able to offer the most accurate predictions. To improve the match between the recovered and real colorant reflectances, we perform a second calibration step in which only the colorant reflectances are modified. The mid-points of the ink spreading curves are considered to be accurately calibrated by the minimization performed by Eq. (18) and are not modified in this second step.

6.1   Surface coverage based colorant reflectance constraints

Thanks to the first calibration step performed in Eq. (18), an approximation of the colorant reflectances is available and we can use a similar approach as in Section 5.1. If no information is available in the image calibration set to fit reflectance R_i of colorant i, the colorant reflectance recovered in the first calibration step R_i^{old} is not modified. The more information is available, the more freedom is given to fit colorant reflectance R_i. *Colorant weight* ω_i, associated to colorant i, indicates the amount of information available to fit colorant reflectance R_i. A weight equal to 0 indicates that no information is available. In this case, the colorant reflectance is set to R_i^{old}, the spectrum recovered in Eq. (18). A weight equal to 1 indicates that the colorant reflectance can be freely calibrated between the minimal value allowed for a reflectance and the paper reflectance. For colorant weight ω_i between 0 and 1, colorant reflectance R_i is constrained as follows:

$$\begin{cases} R_i(\lambda) \le (1-\omega_i)R_i^{old}(\lambda) + \omega_i \cdot R_w(\lambda) \\ R_i(\lambda) \ge (1-\omega_i)R_i^{old}(\lambda) + \omega_i \cdot R_{min} \end{cases} \tag{19}$$

where R_i is the considered colorant reflectance, R_i^{old} the corresponding colorant reflectance recovered in Eq. (18), R_w the measured paper reflectance, and R_{min} the minimal value allowed for a reflectance.

The colorant weights ω_i are computed from the nominal ink coverages of the image tiles as follows. For each tile p of the image calibration set, we compute the effective ink coverages using Eq. (5). We then input these effective ink coverages to the Demichel equations to obtain the effective colorant coverages [Eq. (2)]. An effective colorant coverage $a_{i,p}$ indicates the area effectively covered by colorant i of reflectance R_i in tile p. The larger the area covered by colorant i, the more information about that colorant is available. We therefore consider the effective colorant coverage $a_{i,p}$ as a metric indicating the amount of information about colorant i contained in image tile p. We also define the weight ω_i associated to colorant i for the entire image calibration set as the maximum effective colorant coverage among the tiles:

$$\omega_i = \max_p [a_{i,p}] \tag{20}$$

Note that in contrast to the ink spreading weights, the colorant weights are computed from effective coverages instead of nominal coverages. This is possible because the ink spreading curves are calibrated in the first calibration step performed in Eq. (18). Moreover, a weight $\omega_i = 1$ indicates that within the measured reflectances of the calibration set, there is also the measured colorant reflectance R_i. This is always true for the white colorant since the paper reflectance is measured. When a measured colorant reflectance is present, that colorant is removed from the set of unknown colorant reflectances to be optimized.

The colorant weights ω_i define the bounds within which the colorant reflectances can be set. However, because each colorant reflectance is a spectrum composed of many elements, usually 36 or 48, the

Table III. List of variables used to express matrix and vector sizes.

Variable	Description	Typical value
n_c	Number of colorants	16 in the case of CMYK
n_i	Number of colorants satisfying $\omega_i \neq 1$	15 if only paper is measured
n_p	Number of calibration patches	50 in Sections 8 through 11
n_k	Number of wavelength components composing reflectances	48 when sampled every 10 nm between 380 nm and 850 nm

optimization procedure has a tendency to let not only high weight tiles influence the fitted colorant reflectances, but also middle weight tiles. Since information contained in middle weight tiles is not as relevant as information contained in high weight tiles, the influence of middle weight tiles must be reduced. We achieve this by defining weights ω_p associated to the image tiles. An image tile is considered relevant for the calibration if at least one of its effective colorant coverages is large. *Image tile weight ω_p* associated to image tile p is set to the maximum effective colorant surface coverage present:

$$\omega_p = \max_{\{i|\omega_i \neq 1\}} [a_{i,p}] \tag{21}$$

Note that the maximum only considers the effective coverages of colorants that must be optimized. Since the sum of the effective colorant coverages of a tile is equal to 1, a tile cannot have both middle and high colorant coverages. Either it has a single large colorant coverage and the remaining colorant coverages are small, or it has two or three middle colorant coverages and the remaining colorant coverages are also small. In the first case, its associated image tile weight ω_p is large and the influence of this tile within the optimization procedure is not reduced. In the second case, its associated image tile weight ω_p is small and its influence on the optimization procedure is reduced.

Using both image tile weights ω_p and colorant weights ω_i, we show in Section 6.2 how to design an optimization procedure ensuring respectively that (1) an image tile contributes to the computation of the colorant reflectances only if it provides valuable information, and (2) a given colorant reflectance is modified only when valuable information is available.

6.2 Obtaining the optimized colorant reflectances using a least-squares method

The optimization procedure used for the second calibration step is based on a weighted and constrained least-squares approach derived from Eq. (1). The matrix and vector sizes have the values shown in Table III. Let us first rewrite Eq. (1) in a linear form [19]:

$$Q_p(\lambda_k) = R_p(\lambda_k)^{1/n} = \sum_i a_{i,p} R_i(\lambda_k)^{1/n} = \sum_i a_{i,p} Q_i(\lambda_k) \tag{22}$$

where $Q = R^{1/n}$ is a n_k x 1 reflectance raised to the power $1/n$, further referred to as an *n*-modified reflectance. Since each wavelength λ_k is independent, we can rewrite Eq. (22) as follows:

$$q_{p,k} = \sum_i a_{i,p} q_{i,k} = \vec{a}_p^T \vec{\theta}_k \tag{23}$$

where $q_{p,k}$ is the kth component of the *n*-modified reflectance of image tile p, $a_{i,p}$ the effective coverage of colorant i for image tile p, $q_{i,k}$ the kth component of the *n*-modified reflectance of colorant i, $\vec{a}_p$ the n_c x 1 vector containing the effective coverages of all the colorants for image tile p, and $\vec{\theta}_k$ the n_c x 1 vector containing the kth components of all the colorant reflectances. If some colorant reflectances are

known, i.e. $\omega_i = 1$, we adjust Eq. (23) by determining the reflectance part $\tilde{q}_{p,k}$ of the unknown colorant reflectances:

$$\tilde{q}_{p,k} = q_{p,k} - \sum_{\{i|\omega_i=1\}} a_{i,p} q_{i,k} = \sum_{\{i|\omega_i \neq 1\}} a_{i,p} q_{i,k} = \vec{a}_p'^T \vec{\theta}_k' \tag{24}$$

where $\vec{a}_p'$ and $\vec{\theta}_k'$ are both n_i x 1 vectors. Eq. (24) can be written for each of the p image calibration tiles. By combining the equations of all the image calibration tiles, we obtain the following equation:

$$\vec{q}_k = \begin{bmatrix} \tilde{q}_{1,k} \\ \vdots \\ \tilde{q}_{n_p,k} \end{bmatrix} = \begin{bmatrix} \vec{a}_1'^T \\ \vdots \\ \vec{a}_{n_p}'^T \end{bmatrix} \vec{\theta}_k' = A \vec{\theta}_k' \tag{25}$$

where A is a n_p x n_i matrix whose elements correspond to the effective colorant coverages $a_{i,p}$. Since the effective colorant coverages $a_{i,p}$ can be computed using the ink spreading curves and since the reflectances $R_p(\lambda_k)$ of the image calibration tiles are measured, colorant coverage matrix A and vector $\vec{q}_k$ of size n_p x 1 are known. Based on Eq. (25), we formulate the following typical least-squares problem:

$$\vec{\theta}_k' = \arg\min_{\vec{\theta}} \left[(\vec{q}_k - A\vec{\theta})^T (\vec{q}_k - A\vec{\theta}) \right] = LSQ(\vec{q}_k, A) \tag{26}$$

Let us now introduce the image tile weights ω_p to weight the least-squares residuals and reduce the influence of calibration tiles that do not provide valuable information. We first construct a diagonal matrix W of size n_p x n_p with the following diagonal elements:

$$W_{pp} = \omega_p \tag{27}$$

Using matrix W, we formulate the following weighted least-squares problem [20]:

$$\begin{aligned} \vec{\theta}_k' &= WLSQ(\vec{q}_k, A, W) \\ &= \arg\min_{\vec{\theta}} \left[(\vec{q}_k - A\vec{\theta})^T W (\vec{q}_k - A\vec{\theta}) \right] \\ &= \arg\min_{\vec{\theta}} \left[(W^{1/2}\vec{q}_k - W^{1/2}A\vec{\theta})^T (W^{1/2}\vec{q}_k - W^{1/2}A\vec{\theta}) \right] \\ &= LSQ(W^{1/2}\vec{q}_k, W^{1/2}A) \end{aligned} \tag{28}$$

The final least-squares problem becomes:

$$\vec{\theta}_k' = LSQ(W^{1/2}\vec{q}_k, W^{1/2}A) \text{ such that} \begin{cases} \theta_{i,k} = Q_i(\lambda_k) \leq \left((1-\omega_i) R_i^{old}(\lambda_k) + \omega_i \cdot R_w(\lambda_k) \right)^{1/n} \\ \theta_{i,k} = Q_i(\lambda_k) \geq \left((1-\omega_i) R_i^{old}(\lambda_k) + \omega_i \cdot R_{min} \right)^{1/n} \end{cases} \tag{29}$$

where $Q_i(\lambda_k)$ is the kth component of the reflectance of the ith colorant raised to the power $1/n$, ω_i the colorant weight associated to colorant reflectance R_i, $R_i^{old}(\lambda)$ the reflection spectrum of colorant i recovered in Eq. (18), $R_w(\lambda)$ the reflection spectrum of the paper substrate, and R_{min} the minimum value allowed for a reflectance. This least-squares problem is solved for each wavelength λ_k to obtain the entire optimized colorant reflectances.

7   Selection of image calibration tiles

As seen in previous works [7]-[9], we should use uniform image tiles to calibrate the IS-YNSN model. The algorithm used in this contribution to extract such uniform tiles from color images is similar to the optimized selection algorithm described in [9]. The algorithm creates two different sets of image tiles. The first set, further referred to as the *image calibration set*, is used to calibrate the IS-YNSN model and is composed of 50 image tiles. The second set, further referred to as the *image test set*, is used to test the accuracy of the calibrated model and is composed of 30 image tiles.

Two concepts, non-uniformity and proximity in the CMYK color space, are used to select tiles. The non-uniformity value of a tile located within a small sub-domain of the considered color image is computed using its c, m, y, and k pixel values as follows:

$$u = \sqrt{\left(\sigma(c)^2 + \sigma(m)^2 + \sigma(y)^2 + \sigma(k)^2 \right) / 4} \tag{30}$$

where u is the non-uniformity value; c, m, y, and k are the cyan, magenta, yellow, and black pixel values of the considered image tile ranging from 0 to 1; and the σ function is the standard deviation. In addition to selecting uniform tiles, the algorithm ensures that the tiles are not too close to other already selected tiles. We define *not too close* using the Euclidian distance (norm) in the CMY or CMYK color space. For a CMY tile to be selected, it must not have any tile within a CMY norm of 0.5 and no more than one other tile within a CMY norm of 1.5. For a CMYK tile to be selected, it must not have any tile within a CMYK norm of 0.3 and no more than one other tile within a CMYK norm of 0.9. When a tile respects these constraints, it is said to be not too close to the already selected tiles.

The first phase of the algorithm consists in listing the candidate tiles from which calibration and test tiles are selected. Each image is scanned horizontally and vertically in 2 mm steps. Successive 5 mm large square tiles form the candidate tiles. Each tile is associated with a non-uniformity value according to Eq. (30). Tiles with a non-uniformity value above 0.1 are discarded.

The second phase consists in the actual selection of tiles. The selection algorithm not only selects uniform tiles, but also tiles best suited to calibrate the model. Because it is more important to accurately calibrate the colorant reflectances than the ink spreading curves, we define the best tiles as the tiles maximizing the colorant weights ω_i computed in Eq. (20). To maximize the colorant weights ω_i, the selection algorithm needs to find among the candidate tiles the tiles that both have large effective colorant coverages a_i and are uniform. To account for both aspects, the effective colorant coverages are adjusted according to their non-uniformity value u as follows:

$$\tilde{a}_{i,p} = a_{i,p} \cdot e^{-(u_p / 0.05)^4} \tag{31}$$

where $\tilde{a}_{i,p}$ is the uniformity-adjusted effective colorant coverage of tile p, $a_{i,p}$ is the effective colorant coverage of tile p, and u_p is the non-uniformity value of tile p computed in Eq. (30). If the given tile is uniform, i.e. its non-uniformity value u_p is low, the adjustment function is close to unity and the effective colorant coverage is not modified. If the tile is non-uniform, its non-uniformity value is high and the adjustment function is close to 0. In that case, the tile is discarded. For each colorant, the selection algorithm chooses the tiles (two for CMY images and one for CMYK images) having the maximum uniformity-adjusted effective colorant coverage $\tilde{a}_{i,p}$ and that are not too close to already selected tiles according to the CMY or CMYK norm. Since image tiles having a large maximum uniformity-adjusted effective colorant coverage $\tilde{a}_{i,p}$ also have a large colorant weight ω_i, the proposed selection algorithm ensures that the colorant weights ω_i computed in Eq. (20) are optimal. Once the 16

tiles maximizing the colorant weights ω_i are selected, the algorithm completes the image calibration set by selecting the 34 most uniform image calibration tiles that are not too close to already selected tiles. The final calibration set is therefore composed of 50 image tiles. In Section 12, we examine if smaller sets can be used.

The image test set created by the selection algorithm is composed of the 30 most uniform tiles not part of the image calibration set and not too close to already selected test tiles according to the CMY or CMYK norm.

8   Setup of the experiments

In order to analyze how the different calibration procedures perform, we use the 21 different sRGB images shown in Appendix B. The first three images come from a digital camera and the other eighteen from different sources providing standard test images [21]-[23]. The sizes of the images have not been modified, but they have been resampled to 600 dpi. Moreover, they have been converted to two different color spaces: CMY and CMYK. Both conversions have been performed using Photoshop and the *U.S Web coated (SWOP) v2* CMYK profile. The conversions to CMY and CMYK have been performed without GCR (Gray Component Replacement) and with medium GCR, respectively.

The image calibration and test tiles selected using the algorithm described in Section 7 are shown in Appendix C for the CMY images and in Appendix D for the CMYK images. They are printed on a Canon PixmaPro 9500 inkjet printer at 600 dpi using classical rotated halftone screens. The CMY images are printed at 150 lpi with CMY halftone screen angles at 75°, 15°, and 45°, respectively. The CMYK images are printed at 100 lpi with CMYK halftone screen angles at 75°, 15°, 0° and 45°, respectively. The size and shape of the image tiles are adapted to the Datacolor MF45IR spectrophotometer. The tiles are conceived as disks with a diameter of 3 mm. Since this device illuminates the sample using a directed source (45°d:0° geometry), we measure each sample four different times, each time rotating it by 90° and taking the average. This reduces errors due to the positioning of the device and irregularities of the illumination geometry. The non-uniformity value of a given tile is computed on an area slightly larger than the actual tile, i.e. a 5 mm disk instead of a 3 mm disk. Ensuring that the area surrounding the actual tile is also uniform further reduces positioning errors.

To compare the prediction accuracy of the different calibrations on color-constant patches well distributed across the printable gamut, we use four test sets composed of color-constant patches. The first test set, further referred to as the *117 CMY set*, is composed of 125 color-constant CMY patches with all possible combinations of cyan, magenta, and yellow at 0%, 25%, 50%, 75%, and 100% surface coverages from which we exclude the 8 CMY colorants. The second set, further referred to as the *64 CMY halftone set*, is composed of 64 color-constant CMY patches with all possible combinations of cyan, magenta, and yellow at 20%, 40%, 60%, and 80% surface coverages. The third and fourth sets, further referred to as the *117 UCR set* and *64 UCR halftone set*, respectively, are composed of 117 and 64 color-constant CMYK patches with nominal surface coverages obtained by converting the 117 CMY set and 64 CMY set to CMYK using the standard under color removal algorithm described above.

9   Accuracy of the recovered colorant reflectances

For each of the 21 image calibration sets extracted from the 21 sRGB images, we perform the following three calibrations: the *unconstrained calibration* described in Section 4, the *constrained calibration* described in Section 5, and the surface coverage based *optimized image calibration* described in Section 6. We then compare the recovered colorant reflectances with the measured

reflectances. Figures 2 to 6 focus on five different cases: the yellow colorant in the *Hats* image (Fig. 2), the red colorant in the *Yacht* image (Fig. 3), the blue colorant in the *Hats* image (Fig. 4), the green colorant in the *Lighthouse II* image (Fig. 5), and the black colorant in the *Handa Island* image (Fig. 6). In all the figures, the light gray region indicates the unconstrained colorant bounds introduced in Eq. (7), i.e. the area between R_{min} and R_w. The dark gray region indicates the optimized colorant bounds determined by the colorant weights ω_i in Eq. (19). The colorant reflectances recovered using the unconstrained (thin solid line) and constrained (dotted line) image calibrations are restricted to the light gray region, whereas the colorant reflectances recovered using the optimized image calibration (dashed line) are restricted to the dark gray region. The measured colorant reflectances (thick solid line) are also shown.

As seen in Figs 2 to 6, the recovered unconstrained reflectances (thin solid lines) do not match the measured reflectances (thick solid lines) and show a CIELAB ΔE_{94} error close to or above 20. Moreover, the unconstrained red colorant in Fig. 3 reaches the lower bound of the light gray region, which creates acute angles in the spectrum, angles that are not found in real reflectances. Such reflectances are not suitable for predictions and show the limits of the unconstrained calibration procedure.

The recovered constrained reflectances (dotted lines) fit much better the measured reflectances, but the shapes of the reflectances are not always accurate. For example, in Fig. 2, the transition at 500 nm is not as sharp as the transition of the measured reflectance. In Fig. 3, the bump at 500 nm is not present. Conversely, in Fig. 4, a hollow is present at 570 nm in the constrained reflectance and not in the measured reflectance. Although not always perfectly aligned with the corresponding measured reflectances, the constrained reflectances have shapes similar to the measured reflectances and display an average ΔE_{94} error of 5.79 as shown in Table IV.

The largest ΔE_{94} errors occur for the recovered reflectances of colorants incorporating the black ink, e.g. the colorant shown in Fig. 6. Because the GCR algorithm uses the black ink only to replace large percentages of cyan, magenta, and yellow, it is not possible for an image tile to have a large weight for a colorant including black. Since there is not much information about these colorants, their fitted reflectances are less accurate than the reflectances of colorants that do not include black. If we consider only the 7 CMY colorants, the average ΔE_{94} error between the measured colorant reflectances and those recovered using the constrained calibration algorithm drops to 4.34. By comparison, the average ΔE_{94} error is 4.16 for CMY prints (Table IV). It is slightly better than the 7 CMY colorant average in CMYK prints because the GCR algorithm reduces the surface coverage of some colorants, especially the CMY colorants.

The detailed analysis of the 7 CMY colorants for CMYK prints is given in Table V. It confirms that the recovered reflectances that best match the measured reflectances are the optimized reflectances. However, this holds only when the colorant weights ω_i are large enough. When the weight is low, as in Figs. 5 and 6, the optimized colorant reflectances may show a larger ΔE_{94} error than the constrained colorant reflectances. In such cases, the dark gray regions may not always entirely cover the measured spectra and prevent the optimization algorithm from achieving better matches. However, the dark gray regions prevent the recovered optimized colorant reflectances from deviating too much from the constrained colorant reflectances due to lack of information. For example, in Fig. 5, the dark gray region prevents a good match in the near-infrared region above 750 nm, but also prevents a larger deviation between 550 nm and 600 nm.
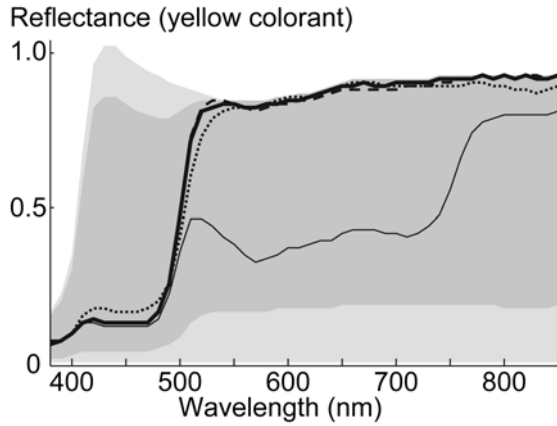
Fig. 2. Measured reflectance (thick solid line), unconstrained (thin solid line), constrained (dotted line), and optimized (dashed line) recovered reflectances for the yellow colorant of the *Hats* image (colorant weight ω_{yellow}=0.80). The ΔE_{94} error between the measured reflectance and the unconstrained, constrained, and optimized fitted reflectances are respectively 26.52, 2.66, and 0.56.
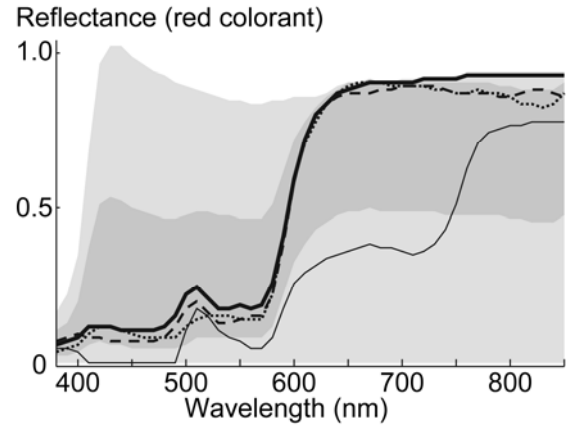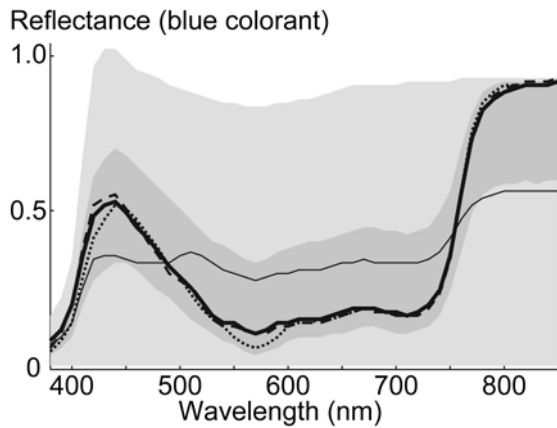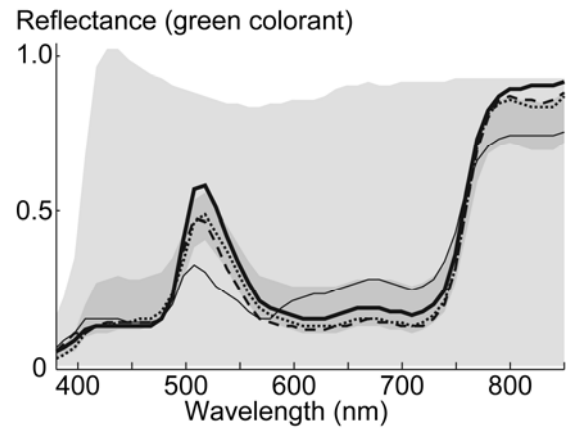

Fig. 3. Measured reflectance (thick solid line), unconstrained (thin solid line), constrained (dotted line), and optimized (dashed line) recovered reflectances for the red colorant of the *Yacht* image (colorant weight ω_{red}=0.46). The ΔE_{94} error between the measured reflectance and the unconstrained, constrained, and optimized fitted reflectances are respectively 25.42, 4.72, and 3.74.

Fig. 4. Measured reflectance (thick solid line), unconstrained (thin solid line), constrained (dotted line), and optimized (dashed line) recovered reflectances for the blue colorant of the *Hats* image (colorant weight ω_{blue}=0.35). The ΔE_{94} error between the measured reflectance and the unconstrained, constrained, and optimized fitted reflectances are respectively 34.82, 4.08, and 1.98.

Fig. 5. Measured reflectance (thick solid line), unconstrained (thin solid line), constrained (dotted line), and optimized (dashed line) recovered reflectances for the green colorant of the *Sea* image (colorant weight ω_{green}=0.17). The ΔE_{94} error between the measured reflectance and the unconstrained, constrained, and optimized fitted reflectances are respectively 19.95, 5.48, and 8.25.
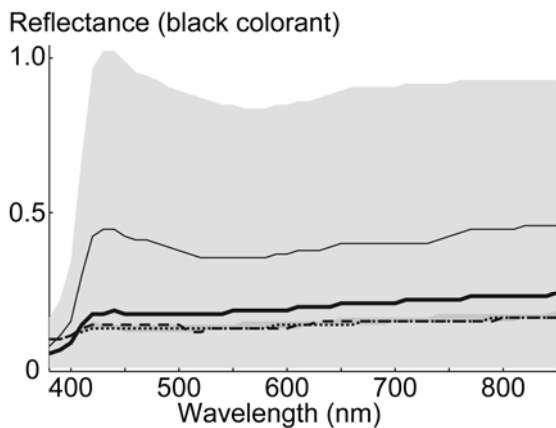

Fig. 6. Measured reflectance (thick solid line), unconstrained (thin solid line), constrained (dotted line), and optimized (dashed line) recovered reflectances for the black colorant (black ink) of the *Handa Island* image (colorant weight ω_{black}=0.02). The ΔE_{94} error between the measured reflectance and the unconstrained, constrained, and optimized fitted reflectances are respectively 19.06, 7.23, and 8.23.

Average reduction in ΔE_{94} error
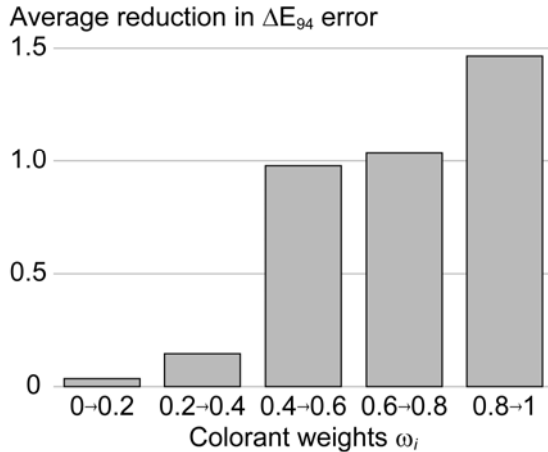
Fig. 7. Average reduction in ΔE_{94} error between recovered optimized and constrained colorant reflectances according to colorant weights ω_i. A reduction of 1 ΔE_{94} indicates that the optimized reflectances are on average more accurate than the constrained reflectances by 1 ΔE_{94}.
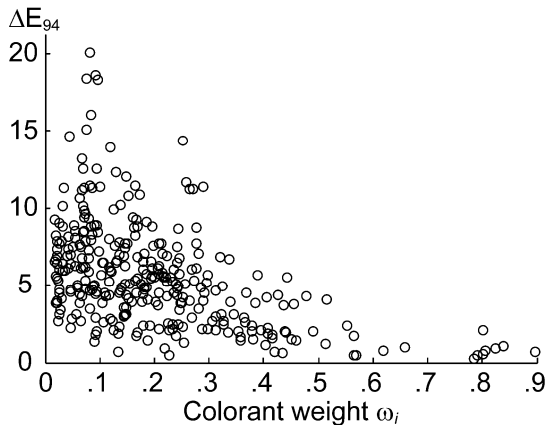
ΔE_{94}

Fig. 8. ΔE_{94} error between measured and recovered optimized colorant reflectances according to colorant weights ω_i. The figure shows that the higher colorant weight ω_i, the more accurate the recovered optimized colorant reflectance.

Table IV. Average ΔE_{94} error between recovered and measured colorant reflectances according to different calibrations. For the constrained and optimized calibrations, the average is computed from the 21 different calibrations performed on the 21 images. The averages do not include the white colorant which is measured.

	Average ΔE_{94}		
	Constrained	Optimized	Optimized distributed 117 patches
CMY prints			
7 CMY colorants	4.16	3.63	1.34
CMYK prints			
15 CMYK colorants	5.79	5.61	3.83
7 CMY colorants	4.34	4.02	2.19

Table V. Average ΔE_{94} error between the measured and 21 recovered colorant reflectances for the constrained and optimized image calibrations, as well as the average and standard deviation of the corresponding colorant weights.

	Colorants						
	C	M	CM	Y	CY	MY	CMY
Average ΔE_{94}							
Constrained	3.45	3.54	4.46	2.59	4.42	3.46	8.45
Optimized	4.72	3.10	4.99	1.62	4.12	2.44	7.14
Colorant weights							
Average	0.27	0.16	0.18	0.45	0.32	0.45	0.20
Std. dev.	0.10	0.12	0.10	0.19	0.11	0.26	0.05

According to Table IV, the average ΔE_{94} error between measured and optimized colorant reflectances is 5.61, or respectively 4.02 if the colorants containing black are excluded. However, as shown in Fig. 7, colorants whose weights ω_i are above 0.4 induce a large accuracy improvement. When the colorant weight is lower, the accuracy remains stable, confirming that the constraints imposed on the optimized colorant reflectances successfully prevent the optimized calibration procedure from setting the colorant reflectances to spectra unrelated to the measured spectra because of lack of information. Finally, Fig. 8 shows that the higher the colorant weight ω_i, the lower the ΔE_{94} error between the measured and the recovered optimized colorant reflectances.

The last column of Table IV shows the average ΔE_{94} error between the fitted and measured colorants for a calibration further referred to as the *optimized distributed 117 patch calibration*, i.e. the optimized calibration according to Section 6 performed using the 117 UCR set instead of an image calibration set. Such a calibration set is optimal since the tiles are evenly distributed across the entire printer gamut and do not contain any colorant. In this case, the average ΔE_{94} error is further reduced to 3.83 for CMYK prints, respectively to 2.19 when the colorants containing black are excluded, and
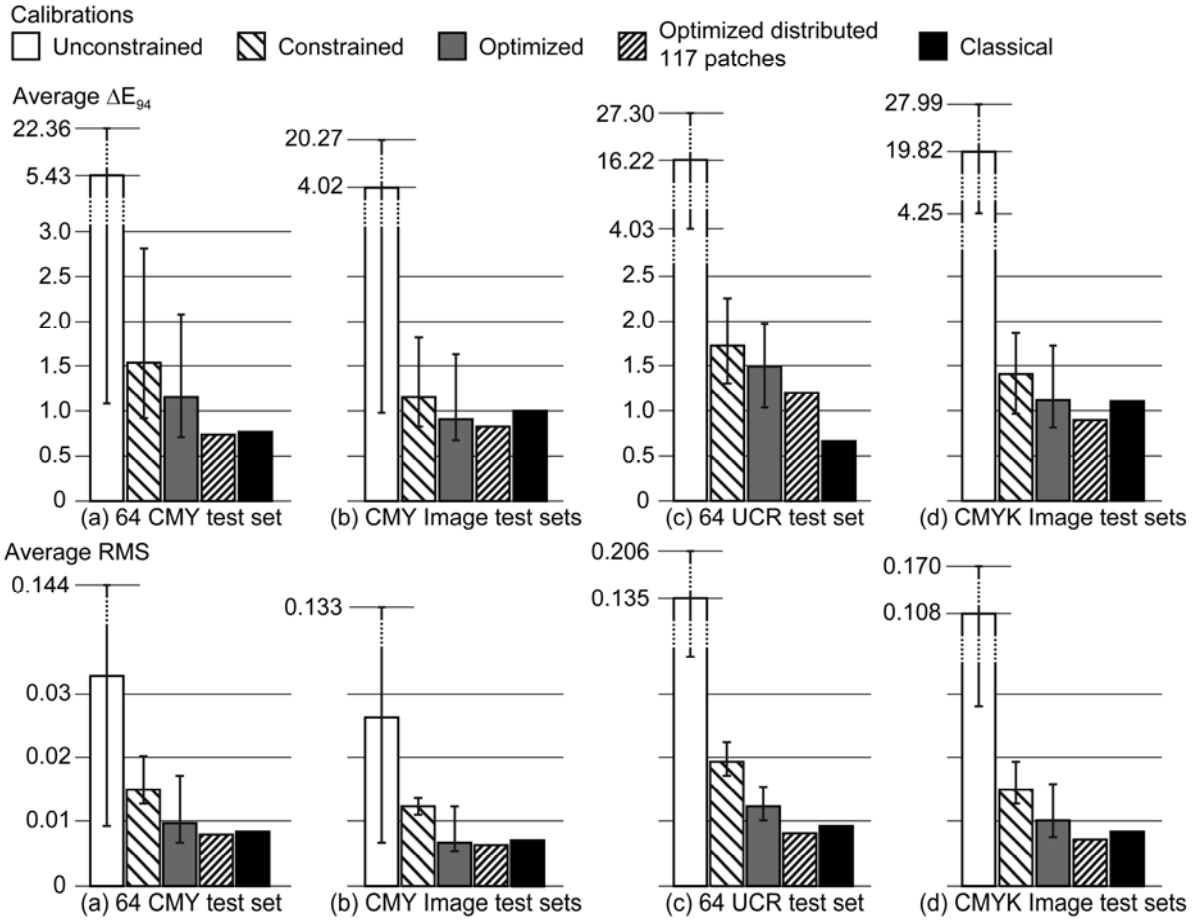
Fig. 9. Comparison of the accuracy of 5 different types of calibrations. For the first three types of calibrations, the bar corresponds either to the average ΔE_{94} (above) or to the average root mean square error (RMSE, below) and the error bar to the best and worst among the 21 calibrations. The test sets predicted are (a) the 64 CMY test set, (b) the 21 merged CMY image test sets, (c) the 64 UCR test set, and (d) the 21 merged CMYK image test sets.

1.34 for CMY prints. These averages show the importance of measuring image tiles well distributed within the printer gamut.

10 Comparison of different model calibrations

Five different calibrations are compared. The first is the *unconstrained calibration* described in Section 4. The second is the *constrained calibration* described in Section 5. The third is the surface coverage based *optimized image calibration* described in Section 6. These first three calibrations are performed for each of the 21 image calibration sets. There are therefore 21 unconstrained calibrations, 21 constrained calibrations, and 21 optimized image calibrations. The fourth calibration is the *optimized distributed 117 patch calibration* introduced in Section 9. It is the optimized calibration described in Section 6, performed on the 117 CMY or respectively for CMYK on the 117 UCR set instead of the image calibration set. The fifth calibration, used as reference, is the classical calibration of the IS-YNSN model relying on the measurements of the colorant reflectances and of predetermined color-constant halftone patches.

Two test sets are used to compare the accuracy of the CMY calibrations: the 64 CMY halftone set and the test set merging all the 21 CMY image test sets. Since each image test set is composed of 30 image tiles, this merged test set is composed of 630 image tiles. The following CMYK test sets are used to compare the accuracy of the CMYK calibrations: the 64 UCR set and the test set merging the 21 CMYK image test sets.

In Fig. 9, each calibration is used to predict each test set. We use the average ΔE_{94} as a measure of accuracy. A lower average ΔE_{94} prediction error implies a more accurate calibration. Since there are 21 different calibrations for the first three types of calibrations, the first three bars represent the mean average ΔE_{94} and the vertical lines indicate the minimum and maximum average ΔE_{94}. We first observe that the unconstrained calibration is not able to accurately calibrate the IS-YNSN model. The average ΔE_{94} prediction error is 4 to 5 times larger in the CMY case (Figs. 9a,b) and up to 20 times larger in the CMYK case (Figs. 9c,d) compared with the classical calibration based on color-constant patches. When constraining the calibration, the average ΔE_{94} prediction error is on average reduced below 2. Optimizing the colorant reflectances further reduces the average ΔE_{94} to the point where even the worst optimized image calibration has an average ΔE_{94} of 2. The surface coverage based optimized image calibration is therefore the best method to calibrate the IS-YNSN model using tiles extracted from color images.

By comparing the performance of the optimized image calibrations to the performance of the optimized distributed 117 patch calibration, we see that the accuracy of the optimized distributed 117 calibration corresponds to the best optimized image calibration. It is therefore important for the image calibration tiles to be well distributed in the printer gamut to achieve the most accurate calibration of the model. Moreover, a classical calibration may perform better than an optimized image calibration when predicting color-constant patches (Figs. 9a,c), but not when predicting image tiles (Figs. 9b,d). Classical calibrations are the most accurate when predicting color-constant patches, but optimized image calibrations are on average more accurate when predicting tiles extracted from color images.

The average ΔE_{94} error achieved by the optimized image calibrations may seem surprisingly low when compared with the larger ΔE_{94} difference between the measured and fitted colorant reflectances shown in Tables IV and V of Section 9. This difference in prediction accuracy is explained as follows. The IS-YNSN model performs predictions by interpolating between the colorants. The colorants are therefore the extreme points of the interpolation domain. In a classical calibration, these extreme points are measured, but are not the best references because they do not correspond to the normal halftoning operation of the printing device. Conversely, by recovering the reflectances of the colorants from tiles extracted from color images, we rely on information that better represents these normal operational conditions. The resulting IS-YNSN model may not be the most accurate when predicting reflectances close to the colorants, but is highly accurate when predicting reflectances of halftones and image tiles.

11 Calibration of the Yule-Nielsen *n*-value

In the preceding sections, the Yule-Nielsen *n*-value is set to a fixed scalar value. Let us investigate how to fit this *n*-value for both the constrained and the optimized image calibrations. In a classical calibration, the *n*-value is fitted as follows. A list of candidate *n*-values is established, for example between 1 and 10 in 0.5 steps. For each *n*, the IS-YNSN is calibrated and used to predict the calibration set. The average root mean square error (RMSE) between the predicted and measured calibration patch reflectances is recorded for each *n*. The chosen *n*-value is the one yielding the lowest average RMSE. We apply the same selection procedure for both the constrained and the optimized image calibrations.

The results are shown in Fig. 10. The average RMSE of the classical calibration [line (1)] decreases smoothly when *n* increases. *n* is therefore set to 10, but the difference is not significant between *n* = 7 and higher values of *n*. When performing constrained image calibrations, the *n*-value is usually set to 10, except for the *Child*, *Peppers*, and *Sailboat III* image calibrations where *n* is set to 9.5 (circles on
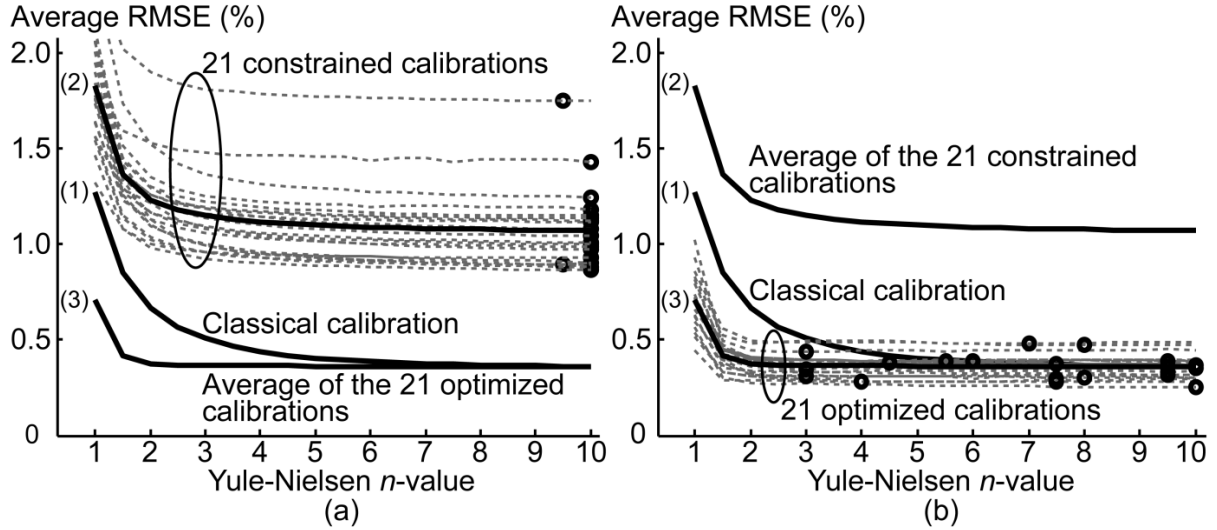
Fig. 10. Average RMSE between the measured and predicted calibration patches, or image calibration tiles, according to different Yule-Nielsen n-values. Solid line (1) corresponds to the classical calibration, the dashed lines in (a) to each of the 21 constrained image calibrations, solid line (2) to the average of the dashed lines in (a), the dashed lines in (b) to each of the 21 optimized image calibrations, and solid line (3) to the average of the dashed lines in (b). The circles correspond to the n-value along each dotted line yielding the minimum average RMSE.

the left). Generally, the average RMSE of such calibrations also decreases when n increases (dashed lines on the left), but some small local increases lead to a few n-values being set to values lower than 10. However, since the average RMSE of the 21 constrained image calibrations behaves on average in the same manner as the average RMSE of the classical calibration [line (2)], the selected n-values correspond to the n-value chosen for the classical calibration.

When optimized image calibrations are used to select n, the general shape of the average RMSE is on average a sharp drop between 1 and 2 followed by a straight line [line (3)], providing no useful information about the best n-value. As seen in Fig. 10b, many different optimal n-values are selected. However, since the lines are nearly flat, any n-value of 2 or above could be chosen. The flexibility introduced by the optimal calibration seems to make the n-value a redundant parameter. Nevertheless, we rely on the constrained image calibrations to choose the best n-value since this n-value corresponds to the choice for the classical calibration. Once the best n-value is chosen, we can optimize the colorants of the corresponding constrained image calibration to obtain an IS-YNSN model fully calibrated from color printed images without relying on prior knowledge about colorant reflectances or ink spreading parameters.

12 Importance of accurately calibrating the colorant reflectances

In the previous sections, we have shown how to calibrate the IS-YNSN model using a set of 50 image calibration tiles maximizing the colorant weights ω_i defined in Eq. (20). In this section, we show the importance of maximizing the colorant weights ω_i and provide more insights about the required number of image calibration tiles.

In order to study the influence of the colorant weights ω_i, we perform the following experiment. We first merge the 1050 available calibration tiles selected in the 21 different test images (see Appendices C and D). From this pool of image calibration tiles, we create calibration sets by randomly selecting tiles one after another and adding the chosen tile to the calibration set only if it increases the sum s of the colorant weights ω_i:

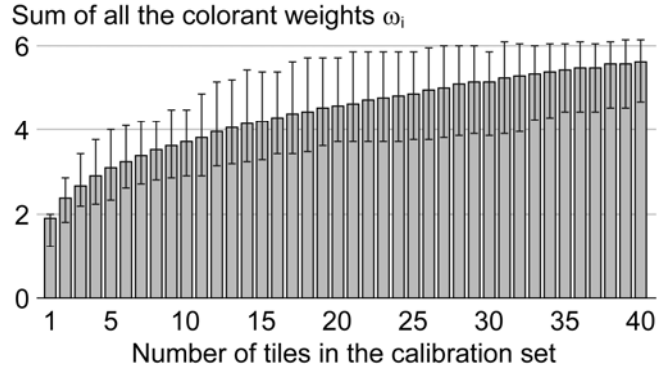$$s = \sum_{\omega_i} \omega_i = \sum_i \max_p [a_{i,p}] \tag{32}$$

Fig. 11. Sums of the CMYK colorant weights ω_i according to the number of image tiles in the iterative calibration sets. The bars represent the average sums over 100 different iterative calibration sets. The error bars represent the highest and lowest sums among the 100 iterative calibration sets.

where the colorant weights ω_i are computed with the current calibration set according to Eq. (20). Such image calibration sets are further referred to as *iterative calibration sets*. We create 100 different CMY iterative calibration sets containing each 25 image calibration tiles and 100 different CMYK iterative calibration sets containing each 40 image calibration tiles. Fig. 11 shows how the sum s of the colorant weights ω_i increases with each additional calibration tile. The sum of the weights may slightly decrease when a tile is added because the colorant weights are computed based on the effective colorant coverages [Eq. (20)], relying themselves on ink spreading curves recalibrated each time a new calibration tile is added. As a reference, the CMYK image calibration sets created in Section 7 and shown in Appendix D are composed each of 50 image tiles and have sums between 3.14 (*Sea* image) and 5.37 (*Hats* image) with an average of 4.0. An average sum of colorant weights of 4 corresponds to iterative calibration sets composed of 12 image tiles.

For each iterative calibration set, we perform as many optimized calibrations as tiles in the calibration set. The first calibration uses only the first tile of the set, the second calibration the first 2 tiles, the third calibration the first 3 tiles, etc. The last calibration therefore uses all the tiles of the calibration set. Based on an iterative calibration set, we calibrate the IS-YNSN model using the optimized calibration algorithm described in Section 6. However, instead of setting the starting point of the gradient-descent algorithm to the default values, i.e. setting the mid-points of the ink spreading curves to 0.5, the principal components of the ink transmittances to 0, and the ink thickness variation factors to 1, we use the values obtained by the previous calibration, i.e. the calibration performed with one calibration tile less. By using the values of the previous calibration, the ink spreading curves are already partly calibrated and the effective ink coverages can be computed to determine the ink spreading weights $w_{i/jk}$ (Section 5.1).

The accuracy of the calibrations based on the iterative calibration sets is shown in Fig. 12 when predicting the 64 CMY halftone set and the 21 merged CMY image test sets, and in Fig. 13 when predicting the 64 UCR set and the 21 merged CMYK image test sets. We compare these predictions with the predictions of the optimized (*) and classical (‡) calibrations performed in Section 10. For both the CMY and CMYK test sets, the prediction error decreases with each additional calibration tile. This decrease correlates perfectly with the increase of the sum of the colorant weights shown in Fig. 11. Sets composed of 10 CMY or 16 CMYK image calibration tiles are sufficient to achieve an accuracy similar to the calibration sets used in Section 10 (*). The prediction error keeps decreasing slowly with each additional tile, but not significantly once the image calibration set contains 20 CMY or 30 CMYK tiles. Calibrations using the iterative calibration sets therefore show significant improvement over the optimized calibrations performed in Section 10 (*).
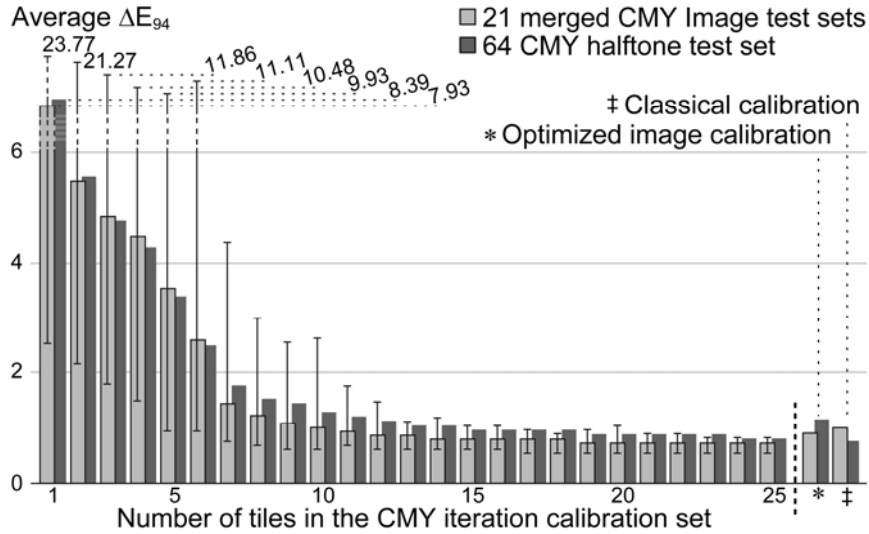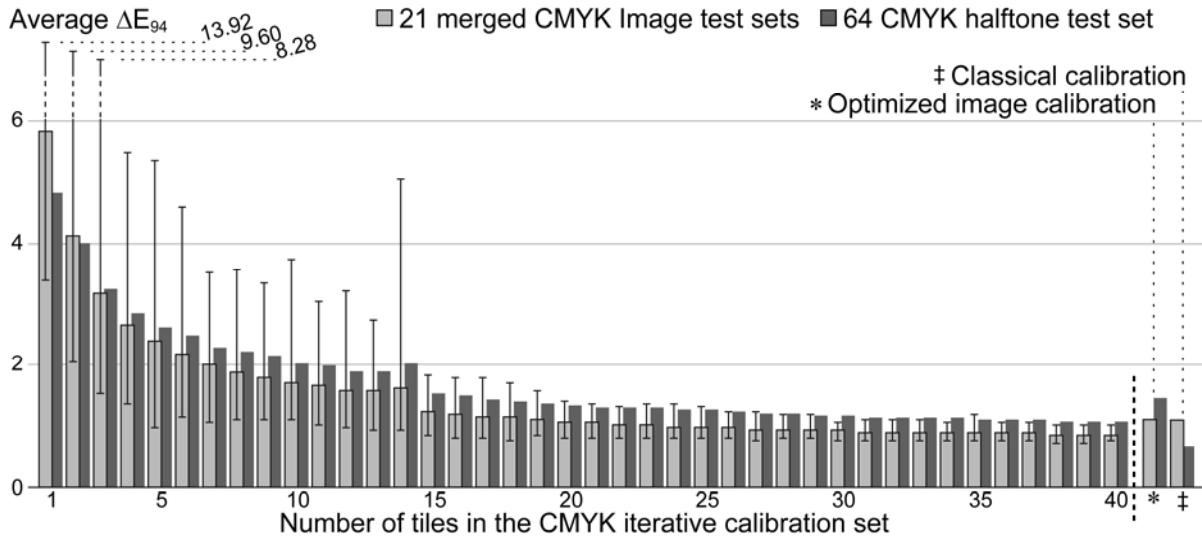
Fig. 12. Average ΔE_{94} between the measured and predicted image tiles of the 21 merged CMY image test sets (light gray) and the CMY 64 halftone test set (dark gray) according to the number of tiles in the iterative calibration sets. The iterative image calibrations are compared with the optimized image calibration performed in Sections 9 and 10 (*) and the classical calibration (‡). The bars indicate the mean average $\Delta E94$ of all the 100 iterative calibrations. The error bars indicate the average ΔE_{94} of the best and worst calibrations among the 100 iterative calibrations performed. Since the error bars are similar for both test sets, they are shown only once.

Fig. 13. Average ΔE_{94} between the measured and predicted image tiles of the 21 merged CMYK image test sets (light gray) and the CMYK UCR 64 halftone test set (dark gray) according to the number of tiles in the iterative calibration sets. The iterative image calibrations are compared with the optimized image calibration performed in Sections 9 and 10 (*) and the classical calibration (‡).The bars indicate the mean average ΔE_{94} of all the 100 iterative calibrations. The error bars indicate the average ΔE_{94} of the best and worst calibrations among the 100 iterative calibrations performed. Since the error bars are similar for both test sets, they are shown only once.

In respect to the classical calibration (‡), calibrations using the iterative calibration sets are more accurate when predicting the image test tiles and almost as accurate when predicting the CMY color-constant halftone patches. Iterative optimized image calibrations are however not as accurate as a classical calibration when predicting the CMYK color-constant halftone patches. These results show the importance of maximizing the colorant weights ω_i. Since these weights indicate the amount of information available to calibrate the colorant reflectances, large colorant weights ω_i improve the accuracy of the recovered colorant reflectances, and therefore the accuracy of the IS-YNSN model.

In order to ensure that the least-squares optimization of Section 6.2 is not rank deficient, the number of image calibration tiles should not be less than the number of colorant reflectances to calibrate. If the white colorant is the only measured colorant, there should be no less than 7 CMY or respectively 15 CMYK image calibration tiles. We can see in Figs. 12 and 13 that calibrations using an insufficient number of calibration tiles are not stable and induce large variations in accuracy. In such cases, we advise using the constrained calibration (Section 5) instead of the optimized calibration (Section 6).

13   Conclusion

Spectral reflection prediction models are important tools for the characterization of color reproduction devices. Among such models, due to its simplicity and accuracy, the Yule-Nielsen model is the most widely used. However, many calibration patches are required for its calibration, patches that are usually not available within the printed document. In order to simplify the calibration of the prediction model and to prevent printing special color-constant halftone and solid calibration patches, we use image tiles extracted from printed color images. The model used in this contribution is the ink spreading enhanced Yule-Nielsen model (IS-YNSN model), a variant of the Yule-Nielsen model including an ink spreading model accounting for physical dot gain.

Image tiles have previously been used by the authors to calibrate the ink spreading curves, but not to recover the reflectances of the Neugebauer primaries. In this contribution, we propose a method to recover the colorant reflectances (Neugebauer primaries), the ink spreading curves, and the Yule-Nielsen n-value using only tiles extracted from printed color images. There is no prior knowledge about the reproduction device. The first calibration step introduces constraints for colorant reflectances derived from the principal component analysis of a large number of ink transmittances and surface coverage related constraints for the ink spreading curves. Such a constrained calibration is performed for a list of candidate Yule-Nielsen n-values. The n-value yielding the most accurate calibration is selected. In a second step, to improve the accuracy of the calibration, the colorant reflectances are optimized by a weighted least-squares approach where the weights are set according to the colorant surface coverages.

The proposed optimized image calibration is almost as accurate as a classical calibration when predicting color-constant halftone patches and more accurate when predicting image tiles extracted from color images. Moreover, this high accuracy is achieved with sets composed of less than 20 well-chosen image calibration tiles. By entirely eliminating the need to print predetermined patches, these calibration procedures broaden the field of use of spectral reflection prediction models. Such models can now be used when it is impossible or too expensive to print extra patches or when the reproduction device is not available. It also enables the real-time monitoring of printing devices without modifying their print jobs.

In respect to the optimized image calibration, in which colorant reflectances are relaxed from their initial constraints and further optimized, the Yule-Nielsen n-value seems to become a redundant parameter. Further research is needed to establish the relationship between colorant reflectance optimization and n-value. Future work also includes applying the presented approach for the real-time monitoring of reproduction devices.

References

[1]   Phil Green, L.W. MacDonald, "Colour Engineering: Achieving Device Independent Colour", *New York: John Wiley and Sons*, 2002, 482 p;

[2]   R. Balasubramanian, "Optimization of the spectral Neugebauer model for printer characterization," *Journal of Electronic Imaging*, Vol. 8, No. 2, 1999, pp. 156-166;

[3] J.A.C. Yule, W.J. Nielsen, "The penetration of light into paper and its effect on halftone reproductions," Proc. TAGA, Vol. 3, 1951, pp. 65-76;

[4] J.A.S. Viggiano, "Modeling the color of multi-colored halftones," Proc. TAGA, vol. 42, 1990, pp. 44-62;

[5] A. U. Agar and J. P. Allebach, "An Iterative Cellular YNSN Method for Color Printer Calibration," Proc. of the 6th IS&T/SID Color Imaging Conference, Scottsdale AZ, 1998, pp. 197-200;

[6] T. Bugnon, M. Brichon, R.D. Hersch, "Simplified Ink Spreading Equations for CMYK Halftone Prints", *Color Imaging XIII: Processing, Hardcopy, and Applications*, *SPIE* Vol. 6807, pp. 680717-1 to 680717-12, January 2008;

[7] T. Bugnon, R.D. Hersch, "Constrained Acquisition of Ink Spreading Curves from Printed Color Images", *IEEE Trans. Image Process.*, Vol. 20, No. 2, February 2011;

[8] R.D. Hersch, M. Brichon, T. Bugnon, M. Hébert, "Deducing ink spreading curves from reflection spectra acquired within printed color images," *Journal of Imaging Science and Technology*, Vol. 53, No.3, 2009, paper 030502, pp. 1-7;

[9] T. Bugnon, R.D. Hersch, "Optimized Selection of Image Tiles for Ink Spreading Calibration," *Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications*, *SPIE* Vol. 7866, pp. 786612-1 to 786612-16, January 2011;

[10] Raja Balasubramanian, "The Use of Spectral Regression in Modeling Halftone Color Printers," Proc. IS&T/OSA Annual Conference, Optics & Imaging in the Information Age, pp. 372-375, 1996;

[11] H.E.J. Neugebauer, "Die theoretischen Grundlagen des Mehrfarbendrucks", Zeitschrift fuer wissenschaftliche Photographie, Vol. 36, 1937, pp. 36-73, translated by D. Wyble and A. Kraushaar in "The theoretical basis of multicolor letterpress printing," Color Res. Appl., Vol. 30, 2005, pp. 323-331;

[12] J.A.C. Yule, R. Colt, "Colorimetric investigation in multicolor printing," Proc. TAGA, Vol. 3, 1951, pp. 77-82;

[13] R.D. Hersch, F. Crété, "Improving the Yule-Nielsen modified spectral Neugebauer model by dot surface coverages depending on the ink superposition conditions," *Color Imaging X: Processing, Hardcopy, and Applications*, *SPIE* Vol. 5667, 2005, pp. 434-445;

[14] R.D. Hersch, P. Emmel, F. Collaud, F. Crété, "Spectral reflection and dot surface prediction models for color halftone prints," *Journal of Electronic Imaging*, Vol. 14, No. 3, 2005, pp. 033001-12;

[15] Th. Bugnon, M. Brichon and R.D. Hersch, "Model-Based Deduction of CMYK Surface Coverages from Visible and Infrared Spectral Measurements of Halftone Prints," *Color Imaging XII: Processing, Hardcopy, and Applications*, *SPIE* Vol. 6493, 2007, pp. 649310-1 to 649310-10;

[16] H.S. Fairman, and M.H. Brill, "The principal components of reflectances," *Color Research and Application*, Vol. 29, No. 2, October 2004, pp. 104–110;

[17] D.-Y. Tzeng, and R.S. Berns, "A review of principal component analysis and its applications to color technology," *Color Research and Application*, Vol. 30, No. 2, April 2005, pp. 84–98;

[18] R.D. Hersch *et al.*, "Deducing ink thickness variations by a spectral prediction model," *Color Research and Application*, Vol. 34, No. 6, October 2009, pp. 432–442;

[19] P. Urban, R.-R. Grigat, "Spectral-based color separation using linear regression iteration," *Color Research and Application*, Vol. 31, No. 3, June 2006, pp. 229–238;

[20] Å. Björk, "Numerical Methods for Least Squares Problems," SIAM press, Philadelphia, PA, 1996, pp. 165-171;

[21] http://r0k.us/graphics/kodak/

[22] http://sipi.usc.edu/database/database.cgi?volume=misc

[23] http://www.hlevkin.com/TestImages/additional.htm

Appendix A

In Section 5.2.1, principal component analysis (PCA) is used to constrain the transmittances of the inks. Given a set of ink transmittances sampled at k different wavelengths, a PCA yields a list of k principal components that allow the reconstruction of the ink transmittances such that the first component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

By limiting the number of principal components to the first m components, only m scalars are required to reconstruct a transmittance. Moreover, as the first m principal components account for most of the variability of the data, we ensure that the reconstructed transmittances keep the intrinsic properties of real transmittances, i.e. their smoothness and characteristic shape. Note that one PCA is performed separately for each of the 4 basic c, m, y, and k inks.

The data used for the PCAs is composed of the spectral reflectance measurements of the cyan, magenta, yellow, and black colorants as well as the substrate on which these colorants are printed. We measured the ink reflectances of 25 CMY and 37 CMYK reproduction devices, comprising a wide range of technologies such as ink jet, thermal transfer, and electrophotographic printers; offset presses; and proofing devices. Note that the Canon PixmaPro 9500 printer used in the present experiments was not used to build the principal vectors of the PCA. Once the reflectance measurements are available, the transmittances of the inks are computed using Eq. (12) and a PCA is performed on these transmittances. For different values of the number of components m, we reconstruct the transmittances and compute the reconstructed colorant reflectances using Eq. (12). We then calculate the following error metrics between the measured and reconstructed colorant reflectances: root mean square error (RMS), ΔE_{94} average, 95 percentile (95%), and maximum ΔE_{94}. We also indicate the score of each principal component, i.e. the percentage of variability explained by the considered principal component, as well as the total percentage of variability explained by the m selected principal components.

Tables VI to IX show the accuracy of the reconstructed transmittances according to the number m of retained principal components for the cyan, magenta, yellow, and black inks, respectively. Except for yellow, the first two principal components already account for more than 90% of the variance (score). Using three principal components, we are able to accurately reconstruct all the colorants with a maximum ΔE_{94} of 2.97 and a 95 percentile ΔE_{94} of 0.89. This accuracy is sufficient and provides us with a method to express the ink reflectances with only 3 scalars per ink.

Table VI. Accuracy of the PCA for the **cyan** transmittances according to the number of principal components used (#PC).

#PC	RMS	ΔE_{94}			Score	
		Avg	95%	Max	Current	Total
1	0.012	0.59	1.08	3.78	69.32	69.32
2	0.007	0.53	0.96	2.96	24.31	93.63
3	0.005	0.49	0.89	2.97	3.99	97.62
4	0.003	0.42	0.83	2.87	1.02	98.64
5	0.003	0.42	0.85	2.75	0.75	99.39
6	0.001	0.28	0.65	1.75	0.43	99.81
7	0.001	0.04	0.09	0.11	0.14	99.95
8	0.001	0.04	0.08	0.10	0.02	99.97
9	0	0.02	0.05	0.08	0.01	99.98
10	0	0.02	0.05	0.06	0.01	99.99

Table VIII. Accuracy of the PCA for the **yellow** transmittances according to the number of principal components used (#PC).

#PC	RMS	ΔE_{94}			Score	
		Avg	95%	Max	Current	Total
1	0.010	1.08	2.19	4.60	88.13	88.13
2	0.006	1.03	2.09	4.96	6.13	94.26
3	0.004	0.37	0.91	1.86	3.84	98.09
4	0.003	0.25	0.54	0.64	1.27	99.36
5	0.002	0.11	0.24	0.49	0.36	99.72
6	0.001	0.09	0.23	0.45	0.16	99.88
7	0.001	0.04	0.09	0.13	0.09	99.97
8	0	0.03	0.07	0.08	0.01	99.98
9	0	0.02	0.04	0.08	0.01	99.99
10	0	0.01	0.02	0.02	0	99.99

Table VII. Accuracy of the PCA for the **magenta** transmittances according to the number of principal components used (#PC).

#PC	RMS	ΔE_{94}			Score	
		Avg	95%	Max	Current	Total
1	0.014	0.83	2.34	4.54	61.94	61.94
2	0.008	0.69	2.11	4.14	24.92	86.86
3	0.004	0.32	0.84	1.71	9.18	96.05
4	0.003	0.28	0.80	1.60	2.08	98.13
5	0.002	0.23	0.56	1.13	1.23	99.36
6	0.001	0.07	0.18	0.28	0.44	99.80
7	0.001	0.04	0.09	0.11	0.11	99.90
8	0.001	0.02	0.05	0.07	0.04	99.95
9	0	0.02	0.05	0.07	0.03	99.97
10	0	0.01	0.03	0.04	0.01	99.98

Table IX. Accuracy of the PCA for the **black** transmittances according to the number of principal components used (#PC).

#PC	RMS	ΔE_{94}			Score	
		Avg	95%	Max	Current	Total
1	0.008	1.11	2.78	17.98	74.63	74.63
2	0.001	0.36	0.79	1.65	24.74	99.37
3	0.001	0.34	0.73	1.40	0.59	99.96
4	0	0.13	0.28	0.89	0.03	99.99
5	0	0.08	0.21	0.28	0.01	100.00
6	0	0.07	0.12	0.29	0	100.00
7	0	0.04	0.09	0.16	0	100.00
8	0	0.02	0.06	0.07	0	100.00
9	0	0.02	0.05	0.07	0	100.00
10	0	0.01	0.02	0.03	0	100.00

Appendix B: Test images (scale 1:5)

N°1: Handa Island

N°2: Beach

N°3: Flower I

N°5: Goldhill

N°6: House I
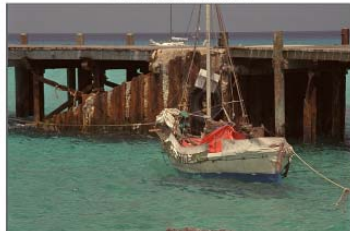

N°7: Hats

N°4: Fruits

N°8: Woman

N°9: Sailboat I

N°10: Flower II

N°12: Sailboat II
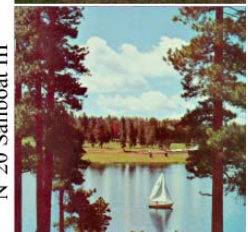

N°13: Child

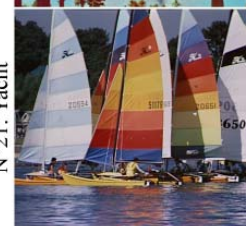N°15: Lighthouse I

N°11: Boat race

N°14: Sea

N°16: Lighthouse II
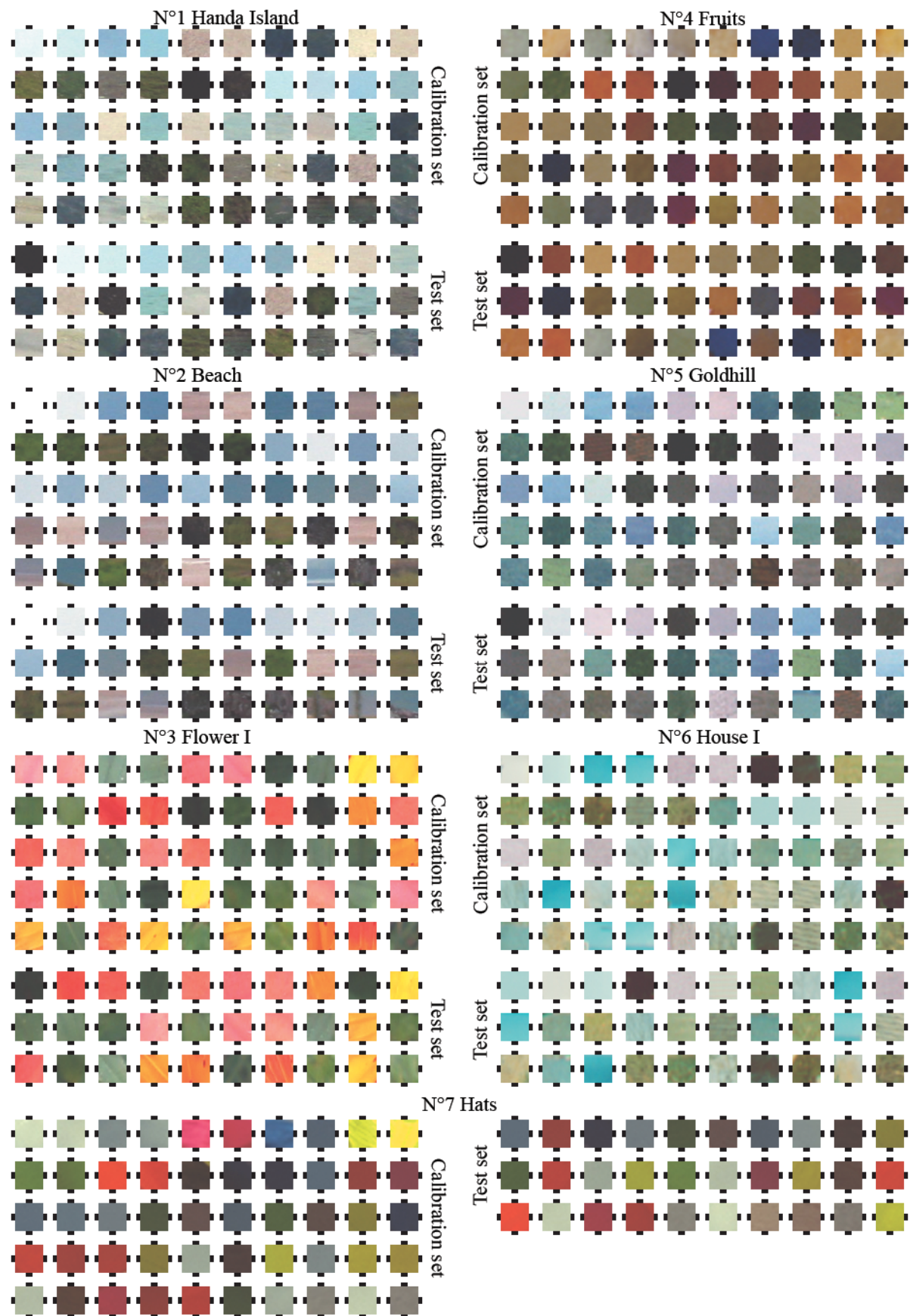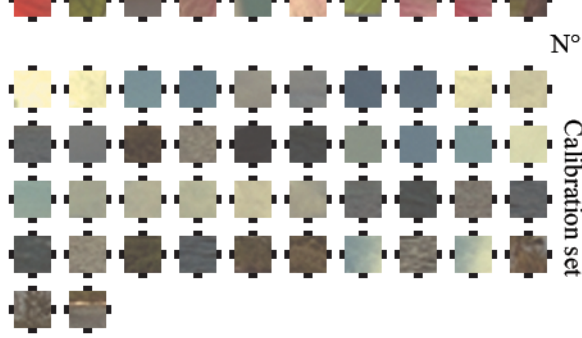

N°20 Sailboat III

N°17: House II

N°18: Parrots

N°19 Peppers

N°21: Yacht

Appendix C: CMY calibration and test image tiles

N°1 Handa Island

Calibration set

Test set

N°4 Fruits

Calibration set

Test set

N°2 Beach

Calibration set

Test set

N°5 Goldhill

Calibration set

Test set

N°3 Flower I

Calibration set

Test set

N°6 House I

Calibration set

Test set

N°7 Hats

Calibration set

Test set

N°8 Woman

N°11 Boat race

N°9 Sailboat I

N°12 Sailboat II

N°10 Flower II

N°13 Child

N°14 Sea

N°15 Lighthouse I

N°18 Parrots

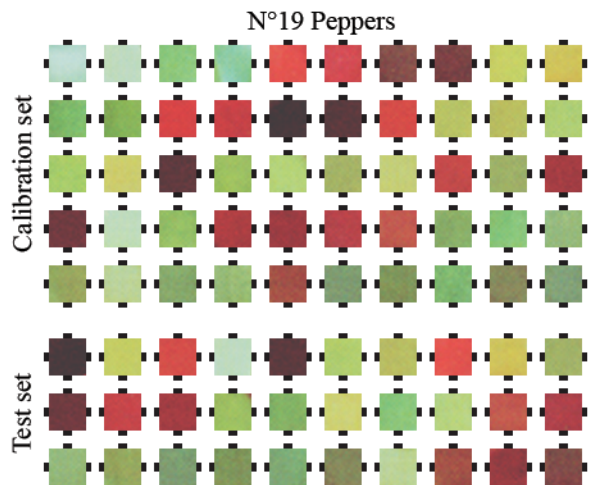N°16 Lighthouse II

N°19 Peppers

N°17 House II

N°20 Sailboat III

N°21 Yacht

Appendix D: CMYK calibration and test image tiles

N°1 Handa Island

Calibration set

Test set

N°2 Beach

Calibration set

Test set

N°3 Flower I

Calibration set

Test set

N°4 Fruits

Calibration set

Test set

N°5 Goldhill

Calibration set

Test set

N°6 House I

Calibration set

Test set

N°7 Hats

Calibration set

Test set

N°8 Woman

N°11 Boat race

N°9 Sailboat I

N°12 Sailboat II

N°10 Flower II

N°13 Child

N°14 Sea

32/33

N°15 Lighthouse I

N°16 Lighthouse II

N°17 House II

N°21 Yacht

N°18 Parrots

N°19 Peppers

N°20 Sailboat III