CGATS RECOMMENDED INDUSTRY PRACTICE

Color characterization data set development — Procedures for color measurement system process control and inter-lab coordination

Version 1 – March 2007

Notice

NPES The Association for Suppliers of Printing, Publishing and Converting Technologies performs secretariat functions for the Committee for Graphic Arts Technologies Standards (CGATS) and its subcommittees, working groups and task forces. This function includes publishing its documents. NPES it is not the author of this document, nor is it an entity that adopts any standard.

The information in this document was developed by members of CGATS Subcommittee 3 (Metrology). Although its contents were approved by ballot of the voting members of the subcommittee, this document has not been developed under the consensus process of the American National Standards Institute (ANSI), and is not an ANSI consensus document. Nor is it a mandatory standard.

At the Subcommittee's request, NPES is making this document publicly available at no cost for informational and reference purposes only.

This document does not relieve the reader of any burden imposed by law, and the reader should independently investigate and verify the performance, results, safety and efficacy of his or her own product or process, even if it follows any recommendations contained in this document. CGATS, its subcommittees, and NPES expressly disclaim liability for any damages or injury arising out of or related to the use of information contained in this document.

This document may be updated as needed by the developing CGATS subcommittee. Such updates will be identified by version number and date, and will be available on the NPES Standards Workroom.

Comments or questions on this document should be sent to:

CGATS Secretariat
NPES
1899 Preston White Drive
Reston, VA 20191-5468 USA
Tel: 1-703-264-7200
Fax: 1-703-620-0994
E-mail:standards@npes.org

Contents

Introduction

The objective of this CGATS Recommended Industry Practice is to define the key steps necessary to achieve measurement system agreement. It includes information on setting up a measurement laboratory and the instruments within it, developing inter-instrument or lab system agreement, analysis of process control measurement data and reporting of data. It is intended as a reference for persons/organizations responsible for setting up measurement facilities and establishing process control of measurement.

This industry practice is not considered a necessary component of the package of industry practices covering how to develop color characterization data. However, the practice of making consistent high quality measurements is fundamental to the success of creating color characterization data. Since there is a strong link to those industry practices aimed at developing color characterization data a short synopsis of their contents follows.

The key steps necessary to conduct press runs with the purpose of deriving color characterization data is covered in another CGATS Recommended Industry Practice: *Color characterization data set development — Press run guidelines*[21]. That document is useful as a reference and checklist for standards and industry groups developing reference color characterization data, as well as by individual printing organizations as they develop their own internal procedures. The document describes the key steps needed for successful execution of a characterization press run and is organized into sections covering planning, preparation, printing, and postprinting.

A second document, CGATS Recommended Industry Practice: *Color characterization data set development — Analysis and reporting*[22], defines the key steps necessary to analyze press sheets produced for the development of color characterization data. It includes recommendations for preliminary evaluation and sheet selection, data collection, data analysis and data reporting. It is intended as a reference for persons/organizations responsible for preparing color characterization data. This analysis and reporting document relies on making high quality measurements and therefore has a strong relationship to the content covered here in this document.

In addition, this CGATS Recommended Industry Practice covering measurement system agreement incorporates the knowledge gained from lessons learned through the experiences of developing color characterization data sets. Over the last several years both the standards community and various industries trade groups have developed process control aims for various printing conditions. These groups have also been conducting press tests to produce sample material as close to the selected aims as possible. Such tests normally contain the IT8.7/3 or IT8.7/4 data sets, described below, to allow the development of color characterization data. See the bibliography for information on SWOP, SNAP and GRACoL industry trade groups. During this time it has been pointed out that measurement system and laboratory agreement, a part of the protocol used by CGATS SC4 (Process Control), is a very important issue. Measurement agreement between the various laboratories has been improved and documented to the point where the information contained in this industry practice provides a good reference for persons/organizations responsible for setting up measurement facilities and establishing process control of measurement.

Standards relating to color data definition represent a key area of activity in the American National Standards Institute (ANSI) Committee for Graphic Arts Technologies Standards (CGATS). The need for color characterization data is largely driven by the increasing use of electronic data exchange for the movement of print-ready material between prepress and printing coupled with the growing use of color management as a key part of the image preparation process. This requires data that defines the relationship between the CMYK values or non-CMYK values used to prepare the printing form and a colorimetric definition of the printed color produced by a particular printing process, commonly called *color characterization data*. Two standardized

targets are the ANSI IT8/7.3 data set with 928 combinations of CMYK tone value data (patches), and the ANSI IT8.7/4 data set that has 1,617 combinations of CMYK tone value data.

The graphic arts standards community, anticipating the need for color characterization data, also developed colorimetric metrology standards. The colorimetric metrology standard is ANSI CGATS.5, *Graphic technology — Spectral measurement and colorimetric computation for graphic arts images*. ISO 13655 is the equivalent International Standard.

There are many practices in use today for making measurements and computing colorimetric and densitometric values. The measurement techniques often will result in different values being computed for the same sample attribute such as CIE L* or status T red density. Therefore, one may not be able to make valid judgments based on the comparison of data derived under variable measurement conditions.

This CGATS Recommended Industry Practice has been developed to provide a set of procedures that can be applied across the industry to test for and determine measurement system agreement and control and ultimately develop consistent color characterization data.

CGATS RECOMMENDED INDUSTRY PRACTICE

Color characterization data set development — Procedures for color measurement system process control and inter-lab coordination

1 Scope

The purpose of this document is to provide guidance in setting up both a reflection measurement system and the process control necessary to ensure that the data reported is meaningful. Data produced by such a measurement system might include spectral reflectance, colorimetry, densitometry, and data derived from such measurements.

Guidance is also provided for those situations where data measured in multiple facilities is to be combined into a single database.

2 Definitions

2.1
box and whisker plot
nonparametric data analysis diagram that illustrates the 25%, 50%, and 75% cumulative distribution of values in a data set (the box) and the expected range of values, defined by distance outside the box ends; see also **whisker (2.20)**, and Figure B.1

2.2
calibration
set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards

[ISO *International vocabulary of basic and general terms in metrology* [2]]

2.3
certified reference material (CRM)
reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence

[ANSI CGATS/ISO 15790[11]]

2.4
combined standard uncertainty, u_c
standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities

[ISO *Guide to the expression of uncertainty in measurement*[1]]

2.5
extreme value
single reading, selected from a series of readings, whose value is so significantly different from the rest of the readings that it is most certainly anomalous. See also **outlier (2.9)**

2.6
hinges
25% and 75% cumulative distribution points of a box and whisker plot of a series of readings taken during a measurement

NOTE 1 Hinges represent the values in which 25% of the readings are less than the lower hinge and 75% of the readings are less than the upper hinge. See also **hinge length (2.7)**.

NOTE 2 Hinges are sometimes called the lower (Q_1) and upper (Q_3) quartile values.

2.7
hinge length, *H*
range of values between the lower and upper hinges of a box and whiskers plot of a series of readings taken during a measurement; see also **box and whisker plot (2.1)**

NOTE The hinge length is sometimes called the box width or the interquartile range Q_3 to Q_1.

2.8
metrology
science of making measurements

[ISO *International vocabulary of basic and general terms in metrology*[2]]

2.9
outlier
single reading, selected from a series of readings, whose value is so significantly different from the rest of the readings that it is suspect but not considered to be an extreme value; see also **extreme value (2.5)**

2.10
reproducibility (of results of measurements)
closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement

[ISO *International vocabulary of basic and general terms in metrology*[1]]

NOTE Reproducibility is distinct from repeatability. Repeatability is the closeness of the agreement between the results of successive measurements on that single sample using a single instrument by the same observer, in the same location and in a short period of time.

2.11
sampling number, *n*
number of multiple measurements, or number of multiple samples, required to reduce the variability of color or color-difference measurement to a desired level

2.12
standard deviation of measurement, *s*
estimate of the standard deviation of the value, p_i, being considered; this includes variability due to the measuring instrument as well as due to the sample itself

$$s = \sqrt{\frac{\sum_{i=1}^{n}(p_i - \overline{p})^2}{n-1}} \quad (1)$$

where

p_i is the i[th] measurement taken with repositioning of the instrument

$$\overline{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \quad (2)$$

n is the number of replicate measurements made

2.13
standard deviation of instrument, s_i
standard deviation of the value due to instrument variability alone:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n}(q_j - \overline{q})^2}{n-1}} \quad (3)$$

where

q_j is the j[th] measurement without repositioning the instrument

2.14
standard error of the estimated mean, s_e
standard deviation of the average of *n* measurements of the same sample; it is equal to the standard deviation of measurements divided by the square root of the sampling number:

$$s_e = \frac{s}{\sqrt{n}} \quad (4)$$

2.15
standard error goal, s_g
maximum acceptable value for standard error

2.16
standardization
adjustment of instrument output to correspond to a previously established calibration using one or more homogeneous samples or reference materials

[ASTM E 284, *Standard Terminology of Appearance*[15]]

2.17
tolerance
upper tolerance limit minus the lower tolerance limit; the total allowable range of the color scale or color difference scale value considered

2.18
traceability
property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties

[ISO *International vocabulary of basic and general terms in metrology*[2]]

2.19
verification
assessment of the overall reliability and accuracy of an instrument or method of measurement by use of material standards for which the measurable quantities have accepted values

[ASTM E 284, *Standard Terminology of Appearance*[15]]

2.20
whisker
in a **box and whisker plot (2.1)**, the lines extending out from the box ends to the largest and smallest observations lying within 1.5 times the hinge length, measured from the box ends

3 Measurements

3.1 Laboratory requirements

In order for a laboratory to assure that measurements are reliable, certain standard operating procedures must be followed. This section is an overview of the requirements.

3.1.1 Standardization and calibration

Standardization is what many people are most familiar with and may simply call calibration. Standardization is the process of adjusting a measurement system such that it produces values that are believed to be correct. This process requires the measurement of a physical reference in order to normalize the device. This physical reference is generally intended for a specific instrument or instrument type and must have values traceable to a standards laboratory, such as the National Institute of Standards and Technology (NIST).

Contrary to common usage, calibration is defined as a set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material and the corresponding values realized by standards. (See *ISO International vocabulary of basic and general terms in metrology*[2])

The end user rarely performs calibration. Calibration is generally a factory or laboratory procedure. In normal practice, standardization is done by using one or more reference standards, such as a white plaque, a black plaque or a black light trap to force the readings from the instrument to agree with the manufacturer's supplied calibrated scale. Once a white and/or black standardization has been performed, it should also be verified. This requires a Certified Reference Material.

3.1.2 Certified Reference Materials (CRMs)

One additional tool to use to supplement the manufacturer's documentation is the use of certified reference materials (CRMs). The graphic arts standard that describes CRMs and also provides an understanding of uncertainty of measurement is ANSI CGATS/ISO 15790[11]. Use of certified reference materials can assist in verifying the performance of measurement systems.

In many areas CRMs are not readily available. ANSI CGATS/ISO 15790 also provides guidance in such circumstances by showing how to determine the reproducibility of results of measurement, even in the absence of CRMs. Using CRMs as part of quality assurance activities is essential for verification and calibration of measurement systems and can increase confidence in data obtained from measurement instruments.

NOTE The manufacturer's calibration reference material is designed to work with a specific instrument and in a specific application. A manufacturer's calibration reference material is a CRM if it conforms to Clause 6 of ANSI CGATS/ISO 15790, which requires the following information be affixed to, or uniquely associated with, the CRM: manufacturer's name, product identification, serial number, certification date, and expiration date or useful life.

3.1.3 Laboratory standard operating procedures

Setting up a measurement system requires control of many elements regarding the care and use of the instrument and reference materials.

The instrument should be stored and maintained according to manufacturer's instructions and recommendations.

Some degree of control or monitoring of ambient temperature and humidity is important, since the behavior of the electronics of the instrument may be dependent upon the temperature and some materials have been found to be thermochroic, that is, their reflectance changes with temperature.

It is important that the appropriate standards for measurement (such as ANSI CGATS.5[9] and ANSI CGATS/ISO 15790[11]) be followed. For example, the measurement instrument geometry and aperture size need to be appropriate for the sample being measured. The sample should be backed with the appropriate backing material as specified in ANSI CGATS.5[9].

As indicated in 3.2.2, all colorimetric computations are based on D50 illuminant. Where the materials being measured are known to fluoresce, the instrument light source should also be a reasonable simulation of the spectral power distribution of D50. It is recognized that instruments presently do not have a measurement source that matches this illuminant. Therefore, measurements of fluorescent materials should be considered relative measurements and only be compared between instruments when it has been established that the spectral power distribution of the illumination of the instruments are similar.

Standardization should follow the manufacturer's recommendations. Calibration through the use of CRMs should be performed: 1) when the instrument is first received, 2) if the instrument has been repaired or recertified, and 3) at regular intervals.

Since the characteristics of a reference material may change over time, reference materials should be replaced or recertified when damaged or as per the recommendation of the CRM manufacturer.

NOTE If a secondary or derivative reference is used, procedures should be established for its maintenance and replacement.

3.1.4 Uncertainty

For calibration (though not for a standardization) to be meaningful, it is necessary to determine the combined standard uncertainty of the result of the measurement.

In order to determine this combined standard uncertainty, one first estimates the uncertainties of each component of the process. (These estimates may be determined by statistical or other methods.) This requires knowledge of all of the sources of uncertainty, including reproducibility, CRM uncertainty (that is, an estimation of the inexactness of the CRM reference value), and other known sources of measurement uncertainty. The uncertainty of a CRM (U_{CRM}) is provided by the manufacturer of the CRM.

See ANSI CGATS/ISO 15790[11] for more details on determining combined standard uncertainty. Information on methods for determining instrument performance is found in ISO 14807[7].

The combined standard uncertainty should be reported along with any measurements to assist the end-user of the data in understanding some of the components of overall imaging variability.

NOTE 1 The uncertainty of a measurement result determined for the measurement of a CRM might not be equal to the uncertainty of results of measuring other materials, even though they may be evaluated with the same measurement process. For example, the measured reflectance of a very glossy or very matte sample is not as sensitive to changes in illumination or detection geometry as a sample with intermediate gloss. Therefore, an instrument that differs slightly from 0:45 geometry may have a fairly low uncertainty with the measurement of a glossy CRM, but will have a higher uncertainty when measuring a moderately glossy sample. In such a case, the effect of gloss should be determined, and combined with other sources of uncertainty. See Annex A for characteristics of CRMs that may also need to be considered when evaluating measurement uncertainties.

NOTE 2 There is a common practice of applying a correction factor to bring two instruments into agreement (calibration). This has been done, for example, by adding an offset to measurements or by multiplying by a constant. This practice is not recommended unless the source of the discrepancy has been diagnosed and completely understood so that the form of the correction can be established.

For example, if one instrument has an imperfect standardization for the absolute black level, it may be appropriate to add or subtract a correction factor to the reflectance. If, on the other hand, the offending instrument has been improperly standardized against a white reference, then it may be appropriate to multiply the reflectance values by some correction factor. Other sources of discrepancy may introduce other types of error. These errors may be sample dependent.

If it is established that a discrepancy arises from a recognized effect, and if the effect can be quantified, a correction can be applied to compensate for the effect. After correction the expected value of the error arising from the systematic effect is zero. Unknown systematic errors cannot be eliminated, but can often be reduced.

NOTE 3 The establishment of uncertainty is not required for standardization. Nor is it required for the establishment of repeatability or reproducibility or for estimating the relative level of inter-instrument agreement. These are all relative forms of measurement system analysis. However, knowledge of the measurement system uncertainty is critical for assessing conformance to or creating absolute numerical specifications for color applications.

3.1.5 Statistical process control

Measurement system process control should be established as part of the standard operating procedures of the facility. Combined standard uncertainty of the measurement system should first be determined, and the measurement tolerance goals required for the facility should be established. Operator techniques can introduce additional variation into the measurement process; therefore, a procedure is required to quantify the additional variation.

An example of a standard operation procedure to determine the variability of the measurement system (instrument and measurement technique) for a particular application is to produce a sample of a set of typical colorants on a substrate and record values such as CIELAB, density, tone value, Tone Value Increase (TVI), as

well as other relevant parameters. This sample is measured, typically by multiple operators of that measurement system, five times a day for a period of five days. For each measurement, the measurement device will be repositioned so as to incorporate the variability of the sample itself. The data is then analyzed using standard practices for statistical process control.

From this baseline the measuring system variability can be determined, checked against the combined standard uncertainty of the system and applied as tolerances to future measurements.

The colorants on a substrate may change over time. New measurement system process control samples should be produced and measured on a regular schedule, dependant on knowledge of substrate and colorant fading. Sample-to-sample variation should be controlled by crossover measures between samples. Where possible, samples should be kept in a cool, dry, dark environment except when needed for process control measurements. Some materials are not optically stable, even in the absence of light. Samples produced from these materials should not be maintained in their finished form but will need to be made up fresh before each test.

Additional information on methods for evaluating instrument performance can be found in ISO 14807[7].

3.2 Measurement procedures

The orientation of the instrument with respect to the surface of the measured material, and the backing material must be according to manufacturer specifications when taking measurements. With automated instruments, this is generally established. With a portable instrument, it is important that the sample should lie on a flat surface while being measured. The portable instrument's aperture and the sample should also lie in the same plane.

3.2.1 Multiple measurements

Multiple measurements should be taken of each sample target area, either on the same sheet or across multiple sheets. Multiple measurements are beneficial for several reasons. First, they provide a means to detect anomalies in the data being collected. It is suggested that a minimum of three measurements be made to detect anomalies.

Second, multiple measurements allow for the reduction in variation. Where the standard error goal is not met through the use of three measurements, additional averaging may be necessary to reduce the standard error to an acceptable level.

A third reason for collecting multiple measurements is to allow for detection of outliers using the mean and standard deviation test (see B.2.4). For this test, at least 10, and preferably 20, measurements should be taken. Annex B describes other methods for determining outliers that may require fewer measurements.

The number of samples required (n) depends upon the relationship between the standard error of the instrument (its reproducibility) and the goal for the standard error. In order to determine the minimum number of samples to average, the following formula is used:

$$n = \max\left(3, \left(\frac{s_e}{s_g}\right)^2\right) \tag{5}$$

where

s_e is standard error of the estimated mean, and

s_g is the standard error goal.

Multiple measurements should be made by first collecting one set of measurements for all the patches, then collecting a second set of measurements, then a third, and so on. This way, it is more likely that a flawed measurement can be identified.

3.2.2 Computation of colorimetric and/or densitometric values

If the measurements are spectral data, all sets of spectral data should be processed through either the instrument's software or computations based on the weighting functions of ANSI CGATS.5. (Colorimetric computations are to be based on CIE illuminant D50 and the 2° standard observer.)

Calculation of status density values should use the weighting factors provided in ISO 5-3[5]. Other graphic arts densitometry–based computations such as tone value and tone value increase are given in ANSI CGATS.4[8].

3.2.3 Identification and potential elimination of outliers

It is important to note that with a large numbers of samples, there is the potential for measurement errors. Errors may be due to misalignment of the instrument's aperture with respect to the sample, an unnoticed defect or wrinkle in the sample, or a transient malfunction of the instrument. Since an outlier may significantly distort results, outliers must be investigated.

To identify outliers, the data for each patch are then grouped. Annex B describes a number of methods which may be used for the identification of outliers.

If it is determined that an outlier exists, all data from that measurement should be removed from the data set. If for example, the L* value for a given measurement of a given patch is determined to not be acceptable, then all remaining data from that measurement of that patch should be removed; e.g. the spectrum, the a* and b* values as well as any other derived values, such as C* or ΔE.

Where data is removed, maintaining a record of the location and values of the data removed allows for further analysis.

3.2.4 Averaging of data

Once any outliers have been eliminated, averaging is to be performed on all remaining measurements of each patch. It is recommended that averaging be performed on spectral data when spectral data is available. If spectral data is not available, averaging of tristimulus (CIEXYZ) data is preferred over averaging of L*a*b* values.

If the measurement variation is fairly small, averaging of spectral, XYZ and L*a*b* values will give very close agreement. On the other hand, for somewhat larger variation, the computations involved in nonlinear color transformations (i.e. L*a*b*) may introduce small differences in the averages. Under typical conditions, these differences are not large enough to cause appreciable errors, but differences are a potential source of confusion.

As an extreme example, consider the calculation of L* value from two normalized tristimulus values of 0.01 and 0.02. If the L* value is computed from these values (to arrive at L* values of 8.99 and 15.49) and the average is computed, the resulting average is 12.24. If on the other hand, the average is computed first and L* computed from the average, the resulting value is 12.61. Note that a 6.5 ΔE in the original values (which would be considered fairly large) led to a discrepancy of only 0.37 ΔE between the two methods.

Which of the two approaches is optimal depends upon the shape of the distribution of errors. Averaging is optimally done in whatever units have the least skew. Since reflectance values generally have a distribution that is close to normal, it is usually preferable that averaging be done on spectral reflectance values.

Therefore, in the absence of statistical analysis that recommends a different type of averaging, averaging should be performed on the spectral data.

3.3 Reporting of data

3.3.1 Data format

When data is reported in electronic form it should conform to either the ASCII keyword value file format or the XML format as defined in ANSI CGATS.17[10].

3.3.2 Recommended data content

When data are reported, they should be accompanied by the following information:

a) statement that the measurements and computations are in conformance with ANSI CGATS.5[9], or identifications of any deviations from CGATS.5;

b) originator of the data;

c) date of creation of data;

d) a description of the purpose or contents of the data being exchanged;

e) a description of the instrumentation used, including, but not limited to, the brand and model number, light source, filters, and wavelength interval used;

f) when density data are reported, the spectral products weighting function (status or type response) used should also be identified;

g) the sample backing (black, white, self-backing or N/A). The spectral reflectance of the white backing or the stack of self-backing material should also be reported;

h) when reporting color difference, report the metric used (e.g. CIELAB ΔE^*_{ab}, CIE94 or CMC) and the appropriate parameters;

i) identification of traceability and uncertainty of the data.

4 Procedures for combining data from multiple facilities

4.1 Introduction and general procedure

Before measurements by the individual locations or laboratories are made, measurement system agreement should be checked using a suitable reference material. Ideally this validation of inter-lab agreement can be combined with the actual measurement procedure to minimize the measurements and time required.

The procedure described in 4.2, 4.3 and 4.4, which is used by CGATS to test for and achieve inter-instrument/lab agreement and to complete a measurement data set, is offered as an example. In this example it is assumed that one site will be a coordinating site and two or more additional sites will participate in the data measurement/collection process for a particular characterization data set.

4.2 Preparation of comparison samples

Using the sheet selection procedure described in Clause 4 of the CGATS Recommended Industry Practice, *Color characterization data set development — Analysis and reporting*[22], select at least four sheets for each of the measurement sites participating in the data collection process. The coordinating site should then make a preliminary measurement of one of the sheets selected for each of the participating sites. Depending on the

schedule and preference of the coordinating site this preliminary measurement can be either the complete target or a sample subset.

The sample subset used by CGATS for comparison purposes is listed in Table 1. Multiple measurements (at least five) should be taken and averaged to provide a reasonable estimate of the average value and uncertainty for each of the patches of the sample subset. The individual samples that contribute to each average should be evaluated to be sure that there are no significant outliers that might bias the data.

These samples should be printed with the same ink and on the same paper as the test samples to be measured. The most expeditious way to accomplish this is by integrating the comparison and validation with the actual measurement process.

After the reference samples are measured at the coordinating site, the sheets to be measured are shipped to the individual participating sites. The participating sites then measure at least the sample subset of the reference sheet previously measured at the coordinating site. Again five measurements are taken, and the average and variability determined.

Table 1 — Patches used for inter-instrument agreement

#	IT8.7/3 ID	IT8.7/4 ID	C	M	Y	K
1	1	73	100	0	0	0
2	2	9	0	100	0	0
3	3	649	0	0	100	0
4	4	81	100	100	0	0
5	5	721	100	0	100	0
6	6	657	0	100	100	0
7	7	729	100	100	100	0
8	11	41	40	40	0	0
9	12	329	0	40	40	0
10	13	365	40	40	40	0
11	14	361	40	0	40	0
12	19	1262	0	100	0	100
13	20	1278	0	0	100	100
14	21	1268	100	100	0	100
15	22	1284	100	0	100	100
16	23	1280	0	100	100	100
17	24	1286	100	100	100	100
18	25	1260	0	0	0	100
19	26	1	0	0	0	0
20	37	10	10	0	0	0
21	50	2	0	10	0	0
22	63	82	0	0	10	0
23	76	1362	0	0	0	10

4.3 Evaluation of comparison samples and individual uncertainty of measurement

The data measured at the participating site are then compared with the data measured at the coordinating site and evaluated for compatibility. There are several methods by which this can be accomplished. The easiest and

simplest is to compute the ΔE between the two sets of average data. If the average ΔE is well under a value of 1.0 and the maximum ΔE is less than 1.0, then it is reasonable to say that the measurement systems agree and to proceed with the measurement of the full set of samples.

Alternatively, the means and distributions of the two sets of CIE X, Y and Z data can be compared and statistical estimates made of the agreement between the two measurement sites.

NOTE 1 The determination of agreement between multiple participating sites is a judgment that must be made by the participants involved.

NOTE 2 If the agreement between participating sites is not deemed satisfactory, techniques are available to improve instrument agreement, and are described in a white paper by Dr. Danny Rich[26].

4.4 Uncertainty of measurement of combined data

It must be remembered that ΔE is a one-sided distribution (only positive values) and the ideal value is zero, therefore the distribution is not normal. As such, traditional statistics (standard deviations) do not provide a meaningful insight.

One technique for the evaluation of the variability of the data that creates a characterization data set and comes from multiple measurements of printed samples is the use of cumulative probability plots of ΔE. Such plots are created as follows:

a) Compute the mean of L*, a* and b* for each patch in the data set.

b) Compute ΔL*, Δa*, and Δb* between the individual measurements of each patch and the mean value for that patch. Use these values to compute a ΔE value for each of the individual measurements.

c) Plot cumulative probability of ΔE. The simple approach is to rank order the ΔE values, and number them 1 to n. Then plot the rank order divided by n as the ordinate and ΔE as the abscissa of the plot.

More sophisticated techniques may be appropriate for generating a histogram.

From this plot, one can easily identify both the ΔE at various percentiles as well as the relative shape of the probability distribution. This latter function is particularly valuable when comparing between different measurement sets.

Annex A

Characteristics of CRMs

Use of CRMs having characteristics that differ from those of materials whose properties are to be measured may yield erroneous measurement results. Consideration should be given to these characteristics when selecting and using a CRM in order that unintended effects are minimized. Documentation should be provided regarding any properties that can adversely affect instrument calibration.

The following is a non-exhaustive list of characteristics of CRMs that can affect the calibration of measuring instruments:

gloss	color
translucency	stability (with time, light exposure, temperature, humidity, etc.)
density	dots
permeability	size
structure (layers)	area
uniformity	reflectance
sharpness	opacity
shape	spectral characteristics (of reflectance, transmittance)
transmittance	fluorescence
background	texture
polarization	

Annex B

Assessment of outliers

B.1 General

Before the data evaluation for inter-instrument or laboratory agreement is finalized, it should be evaluated for potential outliers. In this evaluation, data is searched for outliers and, if an outlier is found, the anomalous data should be analyzed to determine the cause of the deviation. If possible, any patches with outlying measurements should be re-measured.

When re-measurement is not possible, one must consider carefully whether outliers should be dropped or retained. Outliers should be dropped if those readings are not considered to be part of the desired dataset by whatever consistent criteria are accepted. Extreme values should be dropped.

Outlier analysis may be performed using data appropriate to the application (e.g. density, CIEXYZ, or CIELAB data). One procedure is to test each computed value (e.g. L*, a* and b*) from each patch. If it is determined to eliminate one of the three, then all three computed values should be eliminated before computing the average spectrum or CIEXYZ values.

Several methods for testing for outliers are described below. Any point that lies outside a defined region about the patch average is considered an outlier and should be rejected.

In any normally distributed population, there will be members that potentially range from minus to plus infinity. Theoretically, one should include any member of the population in any sample based on estimates of the population parameters. Practically, including a member that is found far from the mean within a small sample, most members of which are found near the mean, will introduce a systematic bias into the estimate of the population parameters (mean, standard deviation, standard error). Such a bias is in direct contrast with the goal of this practice, namely, to reduce the effects of variability of measurement.

B.2 Techniques for identification of outliers and extreme values

The following four sections describe four different methods for identifying outliers and extreme values. If an outlier has been identified, one must decide whether to eliminate the data point from the data set or leave the data point in. When an extreme value has been identified, it should be removed from the data set.

If a data point is eliminated, the measurement of that sample may be repeated, the measurement may be removed and subsequent computations will be based on one less sample, or the eliminated data point may be replaced by the average of the remaining data points.

If it is deemed necessary to eliminate a data point, only a single data point should be eliminated at a time and the procedure repeated with the remaining data. Care should be taken so as to monitor the number of data points that are eliminated within a data set. A record should be kept of the identification of any eliminated data to allow additional error analysis if the number of data points eliminated becomes excessive or shows any pattern.

B.2.1 Use of histograms and cumulative probability plots

A histogram including cumulative probability distribution of variations in the data (e.g. ΔE^*_{ab}, ΔL^*, Δa^*, or Δb^*) can be created to examine the trend in the data to look for outlying points. Any points that are larger than some established percentage determined by the application should be examined to detect if potential errors exist or if

the variation can be explained by noise in the samples or measurement process. It is important to note that histograms and cumulative probability plots can be used on data from individual patches or on data that has been pooled from multiple patches.

Data that is obviously in error is identified at this point to allow it to be excluded during the final stage of data summations.

NOTE If the histogram is comprised of only a few elements then that histogram is not particularly useful. In this case pooling of data from multiple patches should be considered.

B.2.2 Use of ΔE

The common metric that most users are familiar with in evaluating data in variations in color data is the color difference ΔE. Unfortunately ΔE is a one-sided distribution (only positive values) and the ideal value is zero, therefore the distribution is not normal. As such, traditional statistics (standard deviations) do not provide a meaningful insight.

A technical presentation by Dr. Friedrich Dolezalek[23] suggested that the ΔE distribution of printed samples is represented by the chi-squared function with three degrees of freedom. This approach uses the average of the standard deviations (s-avg) of L*, a* and b* as a single parameter to characterize the probability distribution. The quantity "ΔE/s-avg" when squared follows the chi-squared distribution. This provides a convenient estimate of the distribution of ΔE which is more realistic than the use of Gaussian statistics.

For reference, the chi-squared distribution indicates the relationship between s-avg and probability as shown in Table B.1.

Table B.1 — Relationship between s-avg and probability

ΔE	Probability
1 X s-avg	0.211
2 X s-avg	0.749
3 X s-avg	0.973
3.35 X s-avg	0.990

To create the data necessary for evaluation using cumulative probability plots and evaluation of outliers using the chi-squared parameter, the following procedure is recommended.

1. The individual L*, a* and b* measurements for each patch are first averaged.

2. The ΔL*, Δa*, and Δb* are then computed between the average and the individual measurements for each patch.

3. The ΔE are computed using these Δ values and are combined into one series.

4. The ΔL*, Δa*, and Δb* values are pooled between all patches and the sample standard deviations computed for ΔL*, Δa*, and Δb*, s_L, s_a, and s_b.

5. The $\bar{s}$ value is computed from these standard deviations: $\bar{s} = \frac{s_L + s_a + s_b}{3}$.

6. The series of ΔE values is rank ordered and any values that exceed the prescribed limit are considered for removal. A recommended prescribed limit is the .99 probability point, which is $3.35\bar{s}$ (see Table B.1).

B.2.3 Box and whisker test

While ASTM E 178[14] deals with outliers, it unfortunately does not include definitions or procedures relating to the box and whisker test. This method performs well even if the distribution of data is skewed, and it is not affected by a few anomalous points.

This test is best carried out by computer. Many programs for the box and whisker technique are available.

To perform the box and whisker test, start by sorting the readings in order from lowest (1) to highest value (*n*). This ordering of the data is used to determine the lower quartile, the median and the upper quartile.

The reading at position *n*/2 is the median. The reading at position n/4 is the lower quartile Q_1, and the reading at position 3*n*/4 is the upper quartile Q_3. The inter-quartile difference, *H* is the difference in value between upper and lower quartiles.

A value is considered an outlier if its value is smaller than $Q_1 - 1.5H$, or larger than $Q_3 + 1.5H$. A value is considered an extreme value if it is outside the range from $Q_1 - 3H$ to $Q_3 + 3H$.

See Figure B.1 for a graphical representation. *Exploratory Data Analysis*[29] provides further discussion on the box and whisker test.

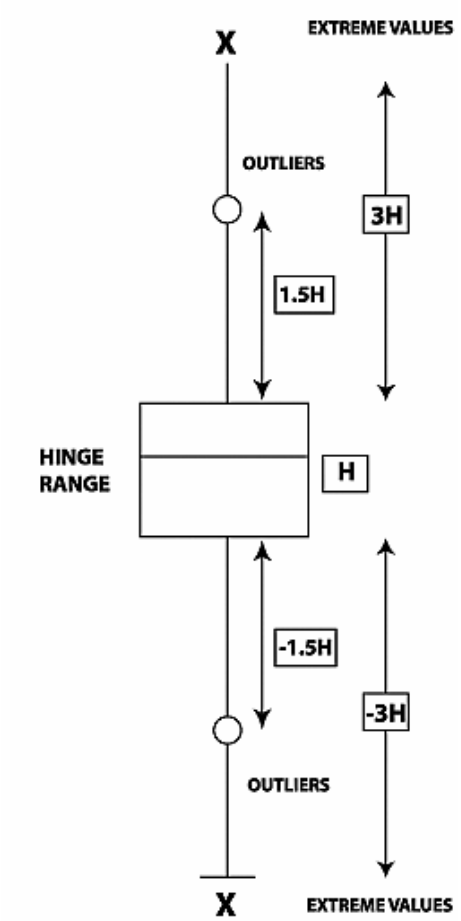

Figure B.1 — Box and whisker plot

B.2.4 Mean and standard deviation test

The mean and standard deviation test is appropriate if it can be established that the variation of the data is normally distributed (i.e. "bell shaped" distribution). If, for example, the variation is bimodal (two humps), this technique is not useful for determining outliers. As noted previously, ΔE is not a normal distribution.

The test for outliers in ASTM E 178[14] is based on the sample mean $\bar{x}$, and the sample standard deviation *s*. The readings are first ordered from the lowest value x_1 to highest value x_n. Calculate the following two statistics, T_1 for the lowest value, and T_n for the highest value in a set of *n* ordered readings as follows:

$$T_1 = \frac{\bar{x} - x_1}{s} \tag{B.1}$$

$$T_n = \frac{x_n - \bar{x}}{s} \tag{B.2}$$

Compare the values of T_1 and T_n to critical values in B.2. If T_1 or T_n is larger than the critical value for *n* readings at the 1% level of significance, then reading 1 or *n* may be considered an outlier. If T_1 or T_n is larger than the critical value for *n* readings at the 0.1% level of significance, then reading 1 or *n* may be considered an extreme value.

NOTE Table B.2 contains critical values for a variety of series lengths at 0.1% and 1% significance levels. For other significance levels and smaller or larger datasets, see Table 1 of ASTM E 178[14].

Table B.2 — Critical values for T (one-sided)

Number of observations, *n*	Upper 0.1% significance level	Upper 1% significance level
10	2.606	2.410
20	3.230	2.884
50	3.789	3.336
100	4.084	3.600

B.2.5 Using a known uncertainty of measurement

The "known uncertainty of measurement test" is appropriate if it can be established that the variation of the data is normally distributed (i.e. "bell-shaped" distribution). If, for example, the variation is bimodal (two humps), this technique is not useful for determining outliers. As mentioned before, ΔE data does not fit a normal distribution.

For many applications, it may be prohibitive to collect a large number of measurements of a single sample (e.g. cyan solid measured on a single sheet, or over multiple sheets). Furthermore, it may have been determined by Equation 5 that only a small number of measurements need be taken. If the number of replicate measurements is less than 10, the methods for determining outliers described in B.2.1, B.2.3, and B.2.4 may not be reliable.

However, it may be possible to obtain a reasonable estimate of the standard deviation by characterization of the instrument and of the process; that is, including the variability from sheet to sheet. Ten to twenty measurements

are needed to yield a reliable estimate of the standard deviation of a population, so a good estimate of the standard deviation cannot be drawn from the data itself.

A suitable estimate of the standard deviation of readings from an instrument can be obtained by taking at least ten and preferably twenty measurements of each of the patches which are a sub-sampling of the entire set of patches. The patches should be chosen to include a representative gamut (e.g. white paper, solid black and 50% chromatic). The determination of the sample standard deviation must be made under similar conditions (e.g. similar stock and similar colorants) to what is being measured. The instrument should be repositioned with each measurement.

NOTE It is expected that the standard deviation may depend on the reflectance of the sample. The measurement of halftone dots (regarding the relationship between aperture size and screen ruling) also introduces a source of variation that may be significant in terms of colorimetry.

If an estimate of the standard deviation σ can be arrived at without using the data points themselves, then the following calculations are used to generate the statistics T_1' and T_n', similar to the previous section

$$T_1' = \frac{\overline{x} - x_1}{\sigma} \tag{B.3}$$

$$T_n' = \frac{x_n - \overline{x}}{\sigma} \tag{B.4}$$

Table B.3 is then used to determine which values for T' are to be considered outliers. If the value of T_1' or T_n' is larger than the critical value from the table at a significance level of 1%, then it is an outlier. If the significance level is at 0.5%, then it is considered an extreme value.

NOTE Table B.3 contains critical values for a variety of series lengths at 0.5% and 1% significance levels. For other significance levels and smaller or larger datasets, see Table 13 of ASTM E 178.

Table B.3 — Critical values of *T'* (one-sided)

Number of observations, *n*	Upper 0.5% significance level	Upper 1% significance level
3	2.40	2.22
5	2.76	2.57
10	3.12	2.93

Bibliography

[1] ISO *Guide to the expression of uncertainty in measurement*, 1995

[2] ISO *International vocabulary of basic and general terms in metrology*, Second Edition, 1993

[3] ISO 5-1, *Photography (sensitometry) — Density measurements — Terms, symbols and notation*

[4] ISO 5-2, *Photography (sensitometry) — Density measurements — Geometric conditions for transmission densitometry*

[5] ISO 5-3, *Photography (sensitometry) — Density measurements — Spectral conditions*

[6] ISO 5-4, *Photography (sensitometry) — Density measurements — Geometric conditions for reflection densitometry*

[7] ISO 14807, *Photography — Transmission and reflection densitometers — Method for determining performance*

[8] ANSI CGATS.4, *Graphic technology — Graphic arts reflection density measurements — Terminology, equations, image elements and procedures*

[9] ANSI CGATS.5, *Graphic technology — Spectral measurement and colorimetric computation for graphic arts images*

[10] ANSI CGATS.17, *Graphic technology — Exchange format for color and process control data using XML or ASCII text*

[11] ANSI CGATS/ISO 15790, *Graphic technology and photography — Reflection and transmission metrology — Certified reference materials — Documentation and procedures for use, including determination of combined standard uncertainty*

[12] ANSI IT8.7/3, *Graphic technology - Input data for characterization of 4-color process printing*

[13] ANSI IT8.7/4, *Graphic technology – Input data for characterization of 4-color process printing – Expanded data set*

[14] ASTM E 178, *Standard Practice for Dealing with Outlying Observations*

[15] ASTM E 284, *Standard Terminology of Appearance*

[16] ASTM E 1164, *Practice for Obtaining Spectrophotometric Data for Object-color Evaluation*, American Society for Testing and Materials, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959

[17] ASTM E 1345, *Standard Practice for Reducing the Effects of Variability of Color Measurement by Use of Multiple Measurements*

[18] ASTM E 1708, Practice For Interchange of Color And Appearance Data, American Society for Testing and Materials, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959.

[19] CIE Publication 15, *Colorimetry, 3rd Edition*, Central Bureau of the CIE, P. O. Kegelgasse 27, A-1033 Vienna, Austria

[20] CIE Technical Report 116 *Industrial Colour-difference Evaluation*, available from USNC/CIE c/o TLA-Lighting Consultants, Inc., 7 Pond St., Salem, MA 01970; or Central Bureau of the CIE, P. O. Kegelgasse 27, A-1033 Vienna, Austria

[21] CGATS Recommended Industry Practice: *Color characterization data set development — Press run guidelines*, available from www.npes.org/standards/toolsric.html

[22] CGATS Recommended Industry Practice: *Color characterization data set development — Analysis and reporting*, available from www.npes.org/standards/toolsric.html

[23] Dolezalek, F., *Appraisal of Production Run Fluctuations from Color Measurements in the Image*, TAGA Proceedings - 1994, p. 154-164

[24] Hsia, J.J., *Optical Radiation Measurements: The Translucent Blurring Effect - Method of Evaluation and Estimation*, NBS Technical Note 594-12, Oct. 1976.

[25] Rich, Danny C. and Martin, D., “Improved model for improving the inter-instrument agreement of spectrocolorimeters”, *Analytica Chimica Acta*, 380, 263-276, (1999)

[26] Rich, Danny C., CGATS white paper, *Graphic technology — Improving the inter-instrument agreement of spectrocolorimeters*, available from www.npes.org/standards/toolsric.html

[27] Sigg, Franz, "Errors in Measuring Halftone Dot Areas", *Journal of Applied Photographic Engineering*, Vol. 9, No. 1, Feb. 1983, pp. 27-32.

[28] Spooner, D. L., "Translucent Blurring Errors in Small Area Reflectance Spectrophotometer and Densitometer Measurements", *TAGA Proceedings* - 1991, p. 130-143

[29] Tukey, John, *Exploratory Data Analysis*, Addison-Wesley, pp. 39-43, 1977