TECHNICAL SPECIFICATION

ISO/TS 23031

First edition
2020-08

Graphic technology — Assessment and validation of the performance of spectrocolorimeters and spectrodensitometers

Reference number
ISO/TS 23031:2020(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 130, *Graphic technology*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Instruments for the measurement of colour and colour difference have been in use since the middle of the 20th century. In the days before digital computers, converting spectral data into CIE tristimulus values was a difficult, manual operation. Additionally, the optics and electronic components were large and difficult to maintain. As a result, every instrument was supplied with a number of reference materials that could be used to assess the performance of the instrument or to adjust the operating parameters. These reference materials included coloured glass filters, rare earth glass filters, neutral density filters and porcelain on steel plaques. Concepts such as accuracy, precision, bias and reproducibility had special and unique applications to these instruments and reference materials.

As the optical and electronic technologies improved, the instruments became smaller, more precise and more affordable. At the same time, the science of metrology matured to the point that the colour-measuring instrument's performance out-paced the ability of the national testing laboratories to produce and certify standard materials suitable for testing. Modern optoelectronic instruments are more precise and more stable than the materials used to assess their performance. Thus, it has become problematic to determine if an instrument is within its factory specification or if two instruments produce results that are in agreement with each other.

Several industries that produce coloured products have chosen to address this situation by adopting and specifying a single brand and design of instrument. The paper and pulp industry have gone so far as to capture one particular design from the 1960s and enshrine it in an International Standard. ISO 2469 describes the optics, the geometry and the operation of an instrument that is ideally suited and specially designed for the measurement of the reflectance and colour of paper and pulp. Additionally, ISO/TC 6, has established a series of authorized laboratories which issue certified reference materials (CRM) for testing and calibrating the performance of an ISO 2469 compliant instrument. This was possible, in part, as the instrument captured in ISO 2469 was widely available on the market and it had no competitive designs and the authorized laboratories market sets of standards which are produced using materials with similar physical and optical properties as production papers or pulps. The authorized laboratories maintain a very close relationship to a single national standards laboratory and to each other. WG3 periodically audits these laboratories to verify that they have calibrated their instruments properly against the scale of radiance factor developed by the national standards laboratory.

In contrast, modern graphic reproduction has moved from the era of artistic interpretation into a time in which the image reproduction is driven by objective numerical assessments. With the availability of modern electro-optics, the number of companies providing instruments and the range of models of different size and capabilities has increased dramatically. Geometries utilized are nominally 45°:0° but may be uniplanar, biplanar, circumferential or annular. While referred to as bidirectional, they are always biconical and the sizes of the influx and efflux cones vary as much as the directionality.

Unfortunately, the national metrology laboratories have not been successful in defining a universally accepted scale of diffuse reflectance factor or diffuse radiance factor for these biconical instruments, especially when the sampling aperture is small. Without artefact standards that closely align with the properties to be measured in the printing industry, the result can easily be a match between two instruments on the reference material that does not correlate to a match on real world materials. As a result, colour-measuring instruments from different manufacturers or with different design intents do not provide adequate agreement on the determination of the colour values or methods for the assessment of the performance of an instrument system relative to its manufacturer declared performance specifications. Further, to make the instruments as simple as possible to operate, the end-user is given little to no access the underlying operation of the instrument. The operator can select an influx spectral quality (M0, M1, M2, M3) but has no way to determine or adjust the spectral quality of the influx. The realization of the scale of 45°:0° reflectance factor or radiance factor is different than that of hemispherical diffuse reflectance factor, even for nearly ideal materials. The operator only has the ability to request that instrument adjust the scale of the instrument using a single reference standard supplied with the instrument. The instrument scale is thus traceable only at the one point. Most do not even offer the ability to set or verify the mid-scale value or the optical null value. Today, optical metrologists refer to this process as standardization, since the instrument is forced to reproduce the values of the one standard.

This document has been prepared to provide the users of portable spectrocolorimeters and spectrodensitometers with guidance on the methods for validation of the performance of those instruments. Since calibration is not possible, the use of a series of certified reference materials (CRM) or a series of stable, idealized reference materials is required. ISO 15790 provides guidance on the development of CRM standards for the scale of optical density. But optical density is a more forgiving measurement than tristimulus colorimetry. Measurement of colour is inherently more complicated than the measurement of optical density, since the logarithmic function compresses the measurement scale and the associated errors. Computing colorimetric tristimulus values from spectral data requires the use of the entire range of reflectance factor values while ISO status density is based on the response of the spectral product. Bright colours, useful for producing a large gamut of colour in image reproduction, possess large differences between the spectral regions of absorption and non-absorption of light but density is only assessing the spectral regions of maximum absorbance. While the human visual system has broad spectral responses, in terms of the cone fundamentals, the post receptor processing allows an observer to perceive hue differences as small as 1 nm. So, the instrumentation for colour assessment needs to have an accuracy several times small than the human visual system.

There is a need to use a set of 10 to 20 physical standards to sample the visible spectrum with materials possessing both high and low reflectance levels and that transition between the two extremes over a very small range of wavelengths. Those materials are stable and nearly opaque to avoid the problems of lateral diffusion observed when the sampling aperture are small. The procedures described here have been shown to provide end-users with methods to quantify the performance of spectrocolorimeters on the day it arrives from the manufacturer or distributor until the day it is retired from service. The methods may also be used to validate the instrument system against manufacturer's specifications and against the requirements for product quality.

National measurement laboratories (NML) continue to develop new scales and new methods of assessing artefacts with the goal of providing certified standard materials for establishing the level of traceability and reproducibility of commercial instruments. Unfortunately, these standards have historically been too expensive for routine use. Only recently have the NMLs began developing automated methods for characterizing reference colours or even user supplied materials. Currently, only large corporations or instrument makers can afford to own such materials. Practical users rely on secondary laboratories and reference standards designed specifically for the end use case. In the graphic arts, that should be some form of printed material with a relatively short duty lifetime.

Finally, even after the CRM has been obtained, the methods for assessing the measurement data are not well described. A spectral reflectance factor curve should include 31, 36, 40 or more measurements. Trying to assign values, tolerances and uncertainties to the individual wavelengths is a challenge. For example, it is possible that measurements of an artefact are consistent for 28 wavelengths and inconsistent at 3 others. Should these instruments be considered as acceptable or failures? Converting the measured data to colorimetric values (XYZ or L*a*b*) improves the situation slightly, but the dilemma of comparing 3 individual readings from one lab or instrument to 3 individual values from another lab, remains a problem not conveniently described in the standards literature. It is the intent of this document to document and describe objective ways of assessing and comparing the performance of a colour-measuring instrument with the ultimate goal of identifying an optimum method for application in the graphic reproduction workflow.

Graphic technology — Assessment and validation of the performance of spectrocolorimeters and spectrodensitometers

1 Scope

This document describes procedures for the assessment and validation of the performance of an optical spectrometer intended for use in capturing the spectral reflectance factor or the spectral radiance factor of printed areas comprised of non-fluorescent or fluorescent materials, respectively. While it does not describe the application to transmitting materials directly, many of the procedures can be applied to transmitting systems by backing them with a reflective white backing material.

This document does not address spectral measurements appropriate to other specific application needs, such as those used during the production of materials (e.g. printing paper and proofing media), which are well described by ISO standards under the jurisdiction of ISO/TC 6. It does not describe the special requirements for testing instruments that make in-process or online colour measurements.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 13655:2017, *Graphic technology — Spectral measurement and colorimetric computation for graphic arts images*

ISO 15790:2004, *Graphic technology and photography — Certified reference materials for reflection and transmission metrology — Documentation and procedures for use, including determination of combined standard uncertainty*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
accuracy
closeness of agreement between a test result and an accepted reference value

Note 1 to entry: The qualitative term accuracy, when applied to a set of observed values, is a combination of a random precision component and a systematic error or bias component. Since, in routine use, random components and bias components cannot be completely separated, the reported “accuracy” is interpreted as a combination of these two elements.

[SOURCE: ASTM E 284]

3.2
bandwidth
width of the spectral response function of the instrument, measured between the half-power points often termed full width at half maximum (FWHM)

3.3
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

Note 1 to entry: Contrary to a common usage, calibration is not the process of adjusting a measurement system such that it produces values that are believed to be correct. Calibration permits either the assignment of values of measurands to the indications (creating a reference table) or the decision to reset or adjust the device.

Note 2 to entry: Following the resetting or adjusting of the device, a calibration needs to be verified to ensure that the new device setting(s) provide indications within the accepted values. Verification of a measuring device requires determination of the uncertainty of the calibration.

[SOURCE: ISO/IEC Guide 99:2007:2.39, modified — The definition has been editorially revised and the original Notes to entry have been replaced.]

3.4
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

[SOURCE: ISO 15790:2004, 3.1.2]

3.5
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 variance or covariance of these other quantities weighted according to how the measurement result varies with changes in these quantities

[SOURCE: ISO/IEC Guide 98-3:2008, 2.3.4]

3.6
coverage factor
k
numerical factor used as a multiplier of the *combined standard uncertainty* (3.5) in order to obtain an expanded uncertainty

Note 1 to entry: The coverage factor is chosen based on the level of confidence desired. This coverage factor, k, is typically in the range of 2 to 3. A coverage factor (k) of 2 generally results in a level of confidence of approximately 95 %, and a coverage factor of 3 generally results in a level of confidence of approximately 99 %. This association of confidence level and coverage factor is based on assumptions regarding the probability distribution of measurement results. For a more thorough explanation, see the Guide to the Expression of Uncertainty in Measurement[13].

[SOURCE: ISO/IEC Guide 98-3:2008, 2.3.6, modified — Note 1 to entry has been elaborated.]

3.7
CRM reference value
value of the certified property of a Certified Reference Material (CRM), reported in the documentation supplied with it

3.8
expanded uncertainty
U
quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of values that could reasonably be attributed to the measurand

Note 1 to entry: Expanded uncertainty is the product of the combined standard uncertainty (u_c) and the chosen coverage factor (k).

[SOURCE: ISO/IEC Guide 98-3:2008, 2.3.5, modified — Notes to entry 1 and 2 have been omitted.]

3.9
inter-instrument agreement
expected level of reproducibility between two or more instruments of exactly the same design and manufacturer

3.10
inter-model agreement
expected level of reproducibility between two or more instruments of different designs, models or manufacturer

3.11
manufacturer's calibration reference material
physical device or material, certified or non-certified, supplied by the instrument manufacturer, which is used to standardize a specific instrument to the manufacturer's scale calibrated to a reference material

3.12
mean colour difference from the mean
MCDM
measure of the dispersion of the results of a series of colour measurements

Note 1 to entry: The MCDM quantifies the average colour difference between each reading and the mean of the group of readings

Note 2 to entry: MCDM is a better single number indicator of the dispersion of a set of colour readings than is the standard deviation of the colour difference (ΔE). This is because the distribution of colour difference is not Normally distributed.

3.13
measurand
particular quantity subject to measurement

Note 1 to entry: Examples are: density, lightness, transmittance, reflectance factor.

[SOURCE: ISO/IEC Guide 99:2007, 2.3, modified — The notes to entry have been deleted.]

3.14
measurement uncertainty
parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the *measurand* (3.13)

Note 1 to entry: Each component of the uncertainty is assumed to have a normal distribution. For cases where this assumption may not be valid, users follow the concepts and rules shown in the Guide to the Expression of Uncertainty in Measurement[13].

Note 2 to entry: The result of a measurement is only an approximation or estimate of the value of the measurand and thus is complete only when accompanied by a statement of the uncertainty of that estimate (see 3.3 and ISO 15790:2004, 7.1.6).

Note 3 to entry: Colour values are three dimensional variables. The uncertainties of a colour are derived from the propagation of the uncertainty from the spectral readings. The method for this process has been documented in the literature[1].

[SOURCE: ISO/IEC Guide 99:2007, 2.26, modified — The definition has been slightly modified and the Notes to entry have been replaced.]

3.15
precision
closeness of agreement between test results obtained under prescribed conditions

[SOURCE: ASTM E 284]

3.16
radiance factor
ratio of the radiance from a point on a specimen, in a given direction, to that from the perfect reflecting or transmitting diffuser, similarly irradiated and viewed

Note 1 to entry: For fluorescent media, the radiance factor is the sum of two quantities, to that from the perfect reflecting or transmitting diffuser, similarly irradiated and viewed.

[SOURCE: ASTM E 284]

3.17
reference material
material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the *calibration* (3.3) of an apparatus, the assessment of a measurement method, or for assigning values to other materials

[SOURCE: ISO Guide 30:2015, 2.1.1, modified — Notes to entry have been omitted.]

3.18
reflectance factor
ratio of the radiant or luminous flux reflected in the directions delimited by the given cone to that reflected in the same directions by a perfect reflecting diffuser identically irradiated or illuminated

Note 1 to entry: The industry commonly, but incorrectly, uses the term reflectance rather than reflectance factor.

Note 2 to entry: It is important to specify the geometry that establishes the given conditions of measurement. See CIE Publication 176.

[SOURCE: IEC 60050-845-04-64]

3.19
repeatability
<of results of measurements> closeness of the agreement between the results of successive measurements on that single specimen using a single instrument by the same operator, in the same location and in a short period of time

3.20
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

Note 1 to entry: Reproducibility is distinct from repeatability. Conditions of measurement may include operator, specimen — including repositioning the same material standard, longer spans of time between readings — including hours, days, weeks, etc.

[SOURCE: ISO International Vocabulary of Basic and General Terms in Metrology]

3.21
spectrocolorimeter
spectrometer, one component of which is a dispersive element (such as a prism, grating, or interference filter or wedge or tunable or discrete series of monochromatic sources) that is normally capable of producing as output the colorimetric data (such as tristimulus values and derived coordinates) in addition to the underlying spectral data from which colorimetric data are derived

[SOURCE: ASTM E 284]

3.22
spectrodensitometer
spectrometer, one component of which is a dispersive element (such as a prism, grating, or interference filter or wedge or tunable or discrete series of monochromatic sources) that is normally capable of producing as output the ISO 5 status density data (such as Status E, Status T or Status NB) in addition to the underlying spectral data from which spectral product data are derived

[SOURCE: ASTM E 284]

3.23
spectrometer
instrument for measuring a specified property as a function of a spectral variable

Note 1 to entry: In optical radiation measurements, the spectral variable is wavelength or wavenumber and the measured property is (or is related to) absorbed, emitted, reflected, or transmitted radiant power.

Note 2 to entry: This term defines a wide class of instruments designed to measure optical radiation either emitted by a source (radiometry) or redirected and absorbed by a material (reflectometry or transmissometry). The collection of the radiant power requires a number of geometric constraints and the transformation from optical power to physical characteristics (irradiance, luminance, reflectance, transmittance, absorbtance or appearance) involves mathematical functions.

[SOURCE: ASTM E 284]

3.24
spectrophotometer
spectrometer, one component of which is a dispersive element (such as a prism, grating, or interference filter or wedge or tunable or discrete series of monochromatic sources) that is normally capable of producing as output the spectral reflectance factor or spectral transmittance of a material specimen

Note 1 to entry: A spectrophotometer is essentially a reflectance or transmittance spectrometer, utilizing either a bidirectional or a hemispherical optical measuring system. The suffix photometer derives from the time the light transducer used was the human eye. It is now almost always superseded by an optoelectronic receiver system.[SOURCE: ASTM E 284]

3.25
standardization
process of forcing or adjusting a measurement system to produce readings that correspond to a previously established *calibration* (3.3) using one or more homogeneous specimens, the *manufacturer's calibration reference material* (3.11) or certified reference materials

Note 1 to entry: As defined here, standardization is normally carried out by an instrument user as described in ISO 15790. The adjusting of an instrument to align with a stored calibration scale using only a single measurement of a single reference standard constitutes a standardization since it is not possible to determine the uncertainty of the process from a single measurement.

[SOURCE: ISO 13655:2017, 3.15, modified — Note 1 to entry has been modified.]

3.26
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

[SOURCE: ISO/IEC Guide 99:2007, 2.41, modified — The term "metrological" was removed. Notes to entry have been omitted.]

3.27
uncertainty of CRM
U_{CRM}
measurement uncertainty that is attributed to the reported value of a CRM in the certificate supplied with it, often expressed as an expanded uncertainty with a coverage factor

4 Known practices for instrument characterization

4.1 Repeatability

4.1.1 General

Optical instruments for the characterization of graphic image reproductions are designed to report values that are clearly consistent from reading to reading, otherwise, process control of the image reproduction would not be possible. Optical instruments for the measurement of optical power, however, have a unique property compared to other types of fundamental metrology. The noise in an optical signal is a function of the signal, the noise in the measurements tends to increase monotonically with the signal. Replicate measurements from a spectrocolorimeter or spectrodensitometer may vary due to electronic noise which is relatively unchanged over time and drift which gradually changes over time. Before beginning any measurement sequence, the instrument should be standardized using the manufacturer's calibration reference material to set the scale that was assigned to the instrument at the instrument maker's factory. The schedule for the frequency in which this step is performed is established by the manufacturer. Repeating this too frequently increases the magnitude of random error.

Thus, to assess the form of precision known as repeatability, one collects many readings from a standard material as quickly as is reasonable[2][3][4]. Since the noise is a function of the signal, it therefore is common practice to use a highly reflecting or transmitting reference material, with little or no spectral modulation, for this determination, that is, a white or transparent reference material.

4.1.2 Procedures

Repeatability of spectral reflectance is normally assessed using a white material that does not contain fluorescent brightening agents. This is usually a ceramic tile, such as the instrument's white standard, a painted white panel or a white paper. The actual material is not so important, so long as the spectral reflectance factors are at least 85 % at all wavelength greater than 420 nm. For this test, a fresh piece of the backing material defined in ISO 13655:2017, A.3 shall be used for this purpose.

The instrument is generally standardized per the manufacturer's directions and the manufacturer's calibration plaque is removed and the reference material is positioned into the measurement position of the instrument. Without moving the instrument or material, at least 30 readings are normally taken with at least 5 s between readings but usually no more than 30 seconds between readings. The readings are always saved as spectral data.

The average of the spectral readings is calculated, and that average imported back into the test software. Normally, the spectral data are averaged (instead of the CIELAB values) because the transformation from XYZ to L*a*b* involves a nonlinear transformation which distorts the means. If the spectral reflectance values are not available, then the procedure is to perform the average in XYZ. The XYZ values for each CIELAB measurement are computed, the collection of XYZ values is averaged,

and finally, this average XYZ is used to compute the average CIELAB values. This yields the same result as averaging the spectra.

Two comparisons are required for assessing the repeatability. First, the colour values of the first reading are compared to the colour values of the remaining 29 readings. The comparison requires the computation of a colour difference. For the first comparison, there should be 29 CIELAB colour differences (ΔE^*_{ab}). The average difference and the maximum differences are noted. In the second comparison method, each of the colour readings of the test are compared to the values of the average all the readings. There should be 30 colour differences. The usual practice is to keep only the average value of the 30 differences. This is known as the Mean Colour Difference from the Mean (MCDM)[2]. This value is widely used as an indicator of dispersion for precision since the sample or population standard deviation is not well suited for the statistics of non-Normal data. See for example, Billmeyer and Saltzman[1], Billmeyer and Alessi[2] or Wyble and Rich[3].

Although ISO/TC 130 recommends the use of the modern colour tolerance equation, CIEDE2000, there is no strong reason to favour CIEDE2000 over ΔE^*_{ab} for this application. Historically, ΔE^*_{ab} has been used to assess repeatability, and ΔE^*_{ab} is easier to compute. It is therefore the recommendation of this document to use ΔE^*_{ab} to compute repeatability. See Formula (1).

$$MCDM = \frac{1}{N}\sum\nolimits_N \Delta E\left(C_\mathrm{i}, C_\mathrm{m}\right) \tag{1}$$

where

N is the number of readings;

C_i are the CIELAB colour coordinates of the i^{th} readings;

C_m are the CIELAB colour coordinates of the average reflectance of all readings.

The two calculations, described above, are similar in that a mean colour difference is reported, but the repeatability is now a direct measure of instrument drift. This instrument repeatability or drift is similar to the assessment of "deviation tolerance" of a print run, as described in ISO 12647-1. The drift analysis is also applied to time frames up to 24 h, taking a reading every 30 min. Many instruments have the capability to take "timed" readings in this fashion. Instead of averaging the colour differences, one may plot the differences as a function of the time of the reading. This drift plot is useful for determining the standardization interval by noting the time at which the colour difference versus the first reading has drifted unacceptably far from the initial point.

4.2 Reproducibility

4.2.1 General

Reproducibility is usually assessed for multiple reference materials, multiple placements of the reference materials, longer time intervals and multiple operators. Typically, at least four (4) reference materials are used, including bright red, bright green, deep blue and medium grey. Some instrument makers supply a single green tile for use in assessing the long-term repeatability, a form of reproducibility in which only the time between readings is changed, but the time is determined in days, weeks and months, and is usually a much longer time than the interval between standardizations. While this "green tile" test can monitor the day-to-day or week-to-week drift in an instrument, it is not a sufficiently robust test of the instrument's true reproducibility. The minimum number of placements (replicate readings) is three (3) and the minimum number of operators is three (3). The number 3 has been reported in several studies in the literature of modern colour-measuring instruments as being the number of replications at which the coefficient of variation stabilizes to a nearly constant value[3]. The time interval depends on the reason for the testing, but the minimum time between replicates is

generally one (1) week. Thus, a year's testing might be 26 readings, every two weeks. This is a typical sampling interval for products like X-Rite's NetProfiler® product[1].

4.2.2 Determination of temporal reproducibility

The reference materials are presented to the instrument, one at a time in a random order. It has been reported that photodiode detection systems have a small hysteresis and thus they "remember" the signal level of the last measured specimen. If one reads the reference materials in a consistent order, then there can be a systematic bias in the readings. Similarly, each set of reference materials are read three (3) times, changing the order between readings, if possible[2]. Scanning instruments and reference materials that are permanently fixed into an array of coloured patches are not be able to be randomized but it is possible to rotate the test chart and scan the patches in a different order.

The spectral data of the three readings on each standard material are averaged and the colour difference between the current average and a reference reading determined.

The reference reading may one of the following:

— a single, initial reading collected on the first week of use after the instrument has been installed (this form of reproducibility is sometimes termed "long-term" repeatability or temporal reproducibility), or

— a set of reference values assigned by the manufacturer (this form of reproducibility is sometimes termed instrument drift), or

— the average colour taken by a number of instruments.

Temporal reproducibility is best analysed using a trend chart, showing the colour difference versus individual readings, in much the same manner as discussed for repeatability drift. A clear, statistical description of the drift is obtained from the trend chart by using a multivariate chart, treating each of the CIELAB coordinates (L*, a*, b*) as a normally distributed random variable. It is also possible to plot the CIELAB colour difference ΔE^*_{ab} versus time, as was described previously. Figure 1 illustrates both types of plot. The Hotelling's T^2 statistics is the multivariate equivalent of the Student's t statistic. See Formulae (2) and (3).

$$T \approx T^2_{p,n-1} = \frac{p(n-1)}{n-p} F_{p,n-p} \tag{2}$$

where

p is the number of colour coordinates (always 3);

n is the number of readings in the test.

$$T = (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) \tag{3}$$

where

$\bar{x}$ and μ are 3 × 1 vectors;

Σ is a 3 × 3 variance-covariance matrix.

If one uses the CIELAB coordinates as the normally distributed variables, then the following steps document how to compute the statistical parameters.

1) Netprofiler® is a product of X-Rite, Incorporated.. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same results.

For variance, see Formula (4):

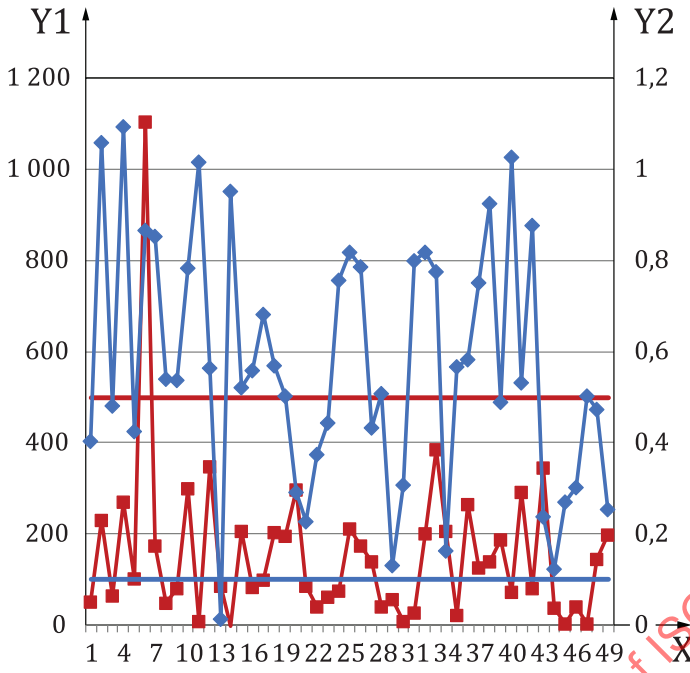
$$\frac{1}{n-1}\sum_{i=1}^{n}\left(a_i^*-\overline{a^*}\right)\left(a_i^*-\overline{a^*}\right)$$
$$\frac{1}{n-1}\sum_{i=1}^{n}\left(b_i^*-\bar{b}\right)\left(b_i^*-\bar{b}\right) \quad (4)$$
$$\frac{1}{n-1}\sum_{i=1}^{n}\left(L_i^*-\bar{L}\right)\left(L_i^*-\bar{L}\right)$$

For covariance, see Formula (5):

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(a_i^*-\overline{a^*}\right)\left(b_i^*-\overline{b^*}\right)$$
$$\frac{1}{n-1}\sum_{i=1}^{n}\left(a_i^*-\overline{a^*}\right)\left(L_i^*-\overline{L^*}\right) \quad (5)$$
$$\frac{1}{n-1}\sum_{i=1}^{n}\left(b_i^*-\overline{b^*}\right)\left(L_i^*-\overline{L^*}\right)$$

The variance-covariance matrix is shown as Formula (6):

$$\begin{bmatrix} var\left(a^*\right) & covar\left(a^*,b^*\right) & covar\left(a^*,L^*\right) \\ covar\left(a^*,b^*\right) & var\left(b^*\right) & covar\left(b^*,L^*\right) \\ covar\left(a^*,L^*\right) & covar\left(b^*,L^*\right) & var\left(L^*\right) \end{bmatrix} \quad (6)$$



Key

X measurement number[1]

Y1 Hotelling's T^2 statistic

Y2 CIELAB ΔE*

tri-variate normal statistic

colour difference from the mean

statistical control limit

colour difference control limit

Figure 1 — Trend chart showing CIELAB ΔE^*_{ab} and Hotelling's T^2 statistic

It has been reported[6] that in monitoring large sets of production data, using multivariate statistical analysis or using total colour difference (ΔE) can hide the direction in colour space of the product defect or it may result in a false positive. However, this behaviour has never been reported in the literature for instrument testing, where there are far fewer degrees of freedom for variation than in a random mixture of colorants in printing system. Nor has this been reported as a problem in the engineering plastics or the automotive coatings industries where large numbers of individual moulded and painted parts are evaluating against production guide panels with very tight tolerances.

Additionally, the Hotelling's T^2 method has the advantage that it is a unitless statistical parameter and as such, its upper limit is set from tables of the F distribution with 3, N-2 degrees of freedom, where N is the number of readings in the test. The test statistic is shown as Formula (7):

$$\frac{N-3}{3(N-1)}T \sim F_{3,N-3} \tag{7}$$

4.2.3 Determination of instrument reproducibility

There are two common forms of instrument reproducibility. They are commonly referred to as, Inter-Instrument Agreement and Inter-Model Agreement. Which type of reproducibility is to be assessed depends on the location and purpose of the determination. Some locations have a variety of instruments in use in varying aspects of the workflow. There may be bench-top instruments, hand-held instruments, strip scanning instruments, area scanning instruments and page reading instruments. If all of these instruments are used to assess the colour of a product at some point in the production of the

final reproduction, then the level of agreement between each model can be determined. This is the type of reproducibility known as Inter-Model Agreement.

In a large printing shop, there may be multiple presses and each press may have an inspection station. Thus, there may be a series of identical instruments, one placed at each station. These instruments are tested to be certain that the assessments of the colour of the products are in agreement from one press to the next. Since all instruments are similar, except perhaps in terms of the date in which they were acquired, it is expected for there to be better agreement between the similar instruments than between instruments with different designs. This is the type of reproducibility known as Inter-Instrument Agreement.

In each scenario described above, the determination of the reproducibility is the same. The series of materials standards are read on each instrument and the colour differences between the instruments on each specimen are determined for each standard in the test set, be it 4, 12, 24 or 42 specimens. The colour differences (ΔE^*_{ab}) can be compiled into the average difference, 95 % cumulative difference and even MCDM dispersion, using the average difference of all specimens and instruments as the mean. This use of the MCDM is the one exception to the rule of only using reflectance values to compute the mean colour. Here, the variable of interest is the difference between two instruments. Specimen 1 has a difference, specimen 2 has a difference, specimen 3 has a difference, etc. The series of differences for a series of measurement points with an expected value of 0. The average (mean) of those differences is the expected difference between the instruments and MCDM is the dispersion of those values around the average. While the absolute colour values determined from the measured spectral reflectance factor or spectral radiance factor readings from each instrument may be different, the colour differences between pairs of readings on each instrument should be similar to the readings from the other instrument. While the more traditional standard deviation could be computed, if the differences between instruments are small, then there is a strong possibility that the standard deviation may represent negative colour differences. It has been shown in the literature that it is less likely that the MCDM produces this undesirable property.

The use of brightly coloured reference materials, such as those described in Clause 5, introduces a new issue. All coloured materials are subject to a phenomenon known as thermochromism. The physical causes of absorption and hence of the colour stimulus function are all susceptible to changes in the ambient temperature. Thus, when taking readings on coloured reference materials, care is taken to control the ambient temperature of the surface of the material standard. In addition, some material standards are also sensitive to the ambient relative humidity. Critical testing can only be performed under strict environmental controls in a measurement laboratory. National measurement laboratories and instrument manufacturer's calibration and testing laboratories utilize precise environmental controls keeping ambient temperature within ±1° and relative humidity within 5 %.

The more difficult and complex problem is that of Inter-Model Agreement. This comparison can present many challenges in the task of obtaining consistent colour readings from all instruments. It can be the result of using a higher performance instrument in the materials lab and a lower cost, lower performance instrument on the production floor. It can also be the result of the production company using instrument 1 from supplier A and the print buyer using instrument 2 from supplier B. Finally, it can be an estimate of the disagreement between the same instrument in two locations, each with different sizes of sampling aperture, which produces varying amounts of lateral diffusion error in two readings Spooner[7] or with different geometric properties of the influx and efflux optics as specified in CIE Publication 176. Averaging of measurements across instruments in an inter-model assessment is not recommended as the assumption that the two instruments are assessing the same property may not be valid.

4.2.4 Data collection and analysis

Each material standard is read three or more times on each instrument. With inter-instrument agreement, the colour difference between the average of the three readings is compared to either, the average of each of the other instruments (a form of MCDM) or to a contrast comparison, similar to those made in an Analysis of Variance (ANOVA) or Gage R&R study. When using the individual colour coordinate values, this analysis is performed using multivariate analysis of variance, treating the CIELAB coordinates (L*a*b*) as trivariate normal random variates. The individual colour differences

would then be averaged and reported. Thus, if there were 5 instruments, the average reading on each material standard would be compared.

EXAMPLE Instrument-1 to Instrument-2, Instrument-1 to Instrument-3, Instrument-1 to Instrument-4, Instrument-1 to Instrument-5, Instrument-2 to instrument-3, Instrument-2 to Instrument-4, Instrument-2 to Instrument-5, etc. ending with Instrument-4 to Instrument-5.

There would then be 10 contrasts for each tile. Since the contrasts represent differences for the same material, the contrasts may be averaged, and the average difference reported for that material standard. For example, a pale grey might have a very small average difference, but a bright red colour might have a much larger difference. Finally, all of the differences for the material standards may be averaged and an expected difference number be reported as follows: "The average difference between the determinations of Instrument 1-5 is 0.xx ΔE units", where the colour difference metric may be either CIELAB units or CIEDE2000 units. This is what is often shown in manufacturer's specification tables for their inter-instrument agreement. The simple average differences, sometimes, the more formal engineering root-mean square differences.

4.3 Accuracy

Normally, accuracy is a critical factor in the assessment of a gauge. One does not want the readings of a 6 mm part to be incorrect by 0,5 mm. The assessment of the accuracy of a meter depends on the availability of the material or artefact standards against which the scale of the meter may be assessed.

Most of the critical measurement gauges have a fundamental or physical basis for their scale. Time is based on the frequency of radiations emitted by a cesium atom, meter is the length of the path travelled by light in a vacuum during a time interval of 1 / 299 792 458 of a second, electric current is based on the Ampere which is related to the current required to produce a force equal to 2×10^{-7} newton per meter of a set of parallel plate conductors. Similar definitions exist for the other four fundamental units.

But there is no physical, fundamental unit of bidirectional or biconical reflectance factor. National standards laboratories supply certified reference materials (CRM) for the high end of the scale of reflectance. Even then, the uncertainty of the CRM that reflects 95 % of the incident light may be as high as 0,4 %. Some standards laboratories and some third-party commercial standards laboratories offer certified, ceramic coloured materials but the uncertainty assigned to such secondary coloured reference materials can be very large, as much as 1 %. Most users concerned with the critical use of spectrocolorimeters, leave the assessment of accuracy of spectral diffuse reflectance factor to the instrument manufacturers, who can afford to purchase the traceable primary standards from a national measurement institution.

However, if accuracy is to be determined, then the assessor should be prepared to estimate the full combined uncertainty of the reference materials and the assessment procedures[1]. Concomitant variables in these assessments include, but are not limited to, ambient temperature at the surface of the test material, relative humidity during the testing, translucency of the upper surface of the test material, stability of the spectral scale and spectral window of the spectrometer, linearity of the radiation detectors. Last, but not least, there is the question of the exact geometry of the instrument. Small, portable instruments often make compromises in the optical design to achieve the lightweight, compact formfactor. The interested reader is referred to CIE Publication 176 for more information on the geometric specifications and tolerances for high accuracy colorimetry.

4.4 Quality of the influx spectrum

Many substrates used in graphic reproduction are produced containing some level of an optical brightening agent (OBA). An OBA is an additive that absorbs short wavelength (380 nm to 420 nm) and UVA radiations (320 nm to 380 nm). It emits optical radiations in the visible range (400 nm to 460 nm). The physical process is known as luminescence or specifically, in this case, fluorescence. The process of fluorescence is highly linear and predictable from a knowledge of the relative spectral power distribution of the excitation source. ISO 3664[8] was revised to meet this need and the assessment of the spectral distribution of radiation in a lighting cabinet is documented there. ISO 13655 followed the lead of ISO 3664 and created a specification for instrument sources that is equivalent to the viewing

booth requirements. The primary source for all measurements in graphic reproduction is CIE D50 and is identified in the International Standards as M1. In other industries, the primary source is CIE D65 and there are material standards available to be purchased and used to either verify the quality of the instrument source or to be used to adjust the source to come into conformance with industry requirements.

Verification of the M1 conformance of portable spectrocolorimeter requires two reference standards, one with a low level of OBA and one with a high level of OBA. Since fluorescence is normally quite linear, two levels of emission shall be adequate to demonstrate conformance. The measurement of luminescence has been shown to be more variable than the measurement of reflectance or transmittance. Therefore, the number of replicate readings shall be increased beyond the three (3) readings recommended above.

If using reference papers as material standards, then each paper should be read six (6) times and the total spectral radiance factors averaged. The colour values (CIELAB) and the CIE Whiteness[9] are computed from the average. Then the colour difference and whiteness difference between the nominal assigned values for the papers and the measured values for the papers shall be determined. In the spirit of ISO 3664, the CIELAB colour differences (ΔE^*_{ab}) between the nominal values and the measured values can be compared to the letter values (A < 0,5, B < 1,0, C < 1,5, etc.). ISO 3664, allows colour difference on the UV activated white metamers to be C or better, which means the maximum colour difference is less than 1,5 CIELAB ΔE^*_{ab} units

5 Reference materials for assessment of performance

5.1 Reference materials for comparison to the manufacturer's specifications

The best way to assess the performance of an instrument or group of instruments relative to the manufacturers published specification is to use a set of reference materials similar or identical to those used by the manufacturer. Most instrument vendors assess their instruments against a set of 12 ceramic tiles known as the Ceramic Colour Standards — Series II[10], from the United Kingdom. These are custom fired ceramic tiles produced in 100 mm square format and may be purchased cut down into a 50 mm square format. These tiles are approximately 10 mm thick at the edges and the glaze is not flat, but slightly convex, being slightly thicker in the middle than on the edge. These reference materials can be purchased with a secondary calibration of the reflectance or colour coordinates, so that the tiles become Certified Reference Materials (CRM), each tile supplied with its CRM reference values. But secondary calibration materials are quite expensive, and an end-user should be strongly committed to metrology to invest in the set of certified tiles. The tiles are relatively easy to use with a benchtop instrument but can be challenging for a hand-held instrument. One should create a jig to position the tiles under the instrument's measurement port. A flat plastic plate or wooden board of the correct thickness with a 100-mm notch in one end should suffice. The tile is positioned into the notch and the ambient temperature recorded. All coloured materials are sensitive to ambient temperature[11], a process known as thermochromism. Many of the ceramic standards change colour by about 0,1 CIELAB unit for each degree Celsius the ambient temperature changes. There is a table of temperature coefficients on the Lucideon website[2)] which gives the temperature coefficients. The Lucideon tiles are also available from other sources including some national laboratories (e.g. NIST, NPL, etc.) and some standards laboratories (e.g. Avian Technologies, LLC, Boson Technology Co., LTD) and from instrument manufacturers (e.g. Datacolor, Konica Minolta).

The coloured glazes are translucent, so they are very sensitive to differences in aperture sizes and aperture alignment[7]. Special precautions are taken to get reliable readings of these ceramic standards using a portable spectrophotometer.

Fortunately, Lucideon has been focused on the printing industry in recent years and has developed some new sets of tiles in a format that is suitable for use with hand-held instruments. Each tile set is equipped with a liquid crystal thermometer to determine the ambient temperature in the area around the tiles. The new tiles have semi-matte finishes, lowering the gloss and as a result, the chroma, are applied to

2) Thermochromic coefficients can be found on the following web page: https://www.lucideon.com/materials-technologies/colour-standards/thermochromism-data.

a single, thin ceramic substrate to make it easier to collect reflectance data with a small, hand-held instrument and are surrounded by black to minimize lateral diffusion of light into the measurement field. See Figures 2 and 3.

Figure 2 — Lucideon business card target

Figure 3 — Lucideon printing standards in a case with flat plastic bed to support the hand-held instrument and position tiles at the correct level

One other method for assessing the performance of instruments is through an organized N-way intercomparision study. This is sometimes referred to as a round-robin study. There are some organizations who can plan, organize and execute these studies for one or more organizations who wish to demonstrate the level of agreement between their different laboratories. These organizations include CEPI-CTS in France[3], IGT in the Netherlands[4] and the Collaborative Testing Services in the United States[5].

3) Available at: http://www.webctp.com/cepi-cts-service-testing-laboratories~D_gb~T_centre-technique-du-papier-fckPages~I_85.centre-technique-du-papier]. In the website, select LABORATORIES \ CEPI-CTS Testing Service.

4) Available at: https://www.igt.nl/laboratory-testing/optical-properties-and-means-of-visual-assesment/.

5) Available at: http://www.collaborative-testing.com