REPRODUCIBILITY COMPARISON BETWEEN TWO MULTI-GONIO-SPECTOPHOTOMETERS

Elísabet Chorro[1], Esther Perales[1], Victor Navarro[1],
Natividad Alcón[2], Ana Rabal[3], Francisco-Miguel Martínez-Verdú[1]
[1] Colour and Vision Group, Department of Optics, University of Alicante, Spain
[2] Technological Institute of Optics, Colour and Imaging (AIDO), Valencia, Spain
[3] Institute of Applied Physics (CSIC), Madrid, Spain

ABSTRACT

New instruments named multi-gonio-spectrophotometers have appeared to measure and characterize the goniochromatism of special materials like metallic, interference, luster and pearlescent samples. These devices are instruments what can measure the spectrum of the sample with different illumination and observation angles, these angles usually agree with some requirements marked in ASTM and DIN standards related to the gonio colour appearance characterization. On the other hand, the inter-comparison between spectrophotometers, both at repeatability and reproducibility levels, has usually appeared in many contributions in last years. The main purpose of this study is to compare at reproducibility level two multi-gonio-spectrophotometers according to the ASTM E2214-08 guidelines: one of them, with 5 measurement (directional) geometries, highly recommended by the automotive industry, while the other one is a bench-top instrument, with 10 measurement geometries. Therefore, they only have 5 common measurement geometries: 45ºx120º, 45ºx110º, 45ºx90º, 45ºx60º, and 45ºx25º. For the reproducibility comparison we will also make some statistical studies that include a Hotelling's test and a statistical test of inter-comparison to know the confidence interval of the partial colour differences ΔL*, Δa*, Δb* and the total colour difference ΔE_{ab}. It is done using as database a collection of 91 metallic, interference, luster and pearlescent samples, which were measured 20 times without replacement for both instruments. The final findings show that except the 45ºx120º geometry, which is the nearest to specular direction, the reproducibility differences between both instruments are statistically significant. This means that these differences are due to systematic or bias errors (angle tolerances for each geometry, photometric scales, white standards, etc), but not exclusively to random errors. However, both statistical tests used here are not valid to discriminate and quantify the detected bias errors in this inter-instrument comparison.

Keywords: colour measurement, multi-gonio-spectrophotometry, reproducibility, metallic and pearlescent samples

CONTACT

Elisabet.chorro@ua.es

INTRODUCTION

During the last years technologic innovation in all the areas has made it possible, between other things, the appearance of the new materials, as for example the metallic and pearlescent objects. Both of them are very useful for colour quality control in the automotive industry[1], for example. But it is difficult to measure and characterize these kinds of colour samples by conventional colour measuring instruments based on integration sphere.

New instruments appeared to solve this problem, specifically the multi-gonio-spectrophotometers. These devices are instruments which can measure the spectral reflectance of the sample with different illumination and observation angles. In the automotive industry the multi-gonio-spectrophotometer more used is the model X-Rite MA68II meeting the requirements marked in some ASTM and DIN standards related to the gonio colour appearance characterization. But, in last years, other models of multi-gonio-spectrophotometers, like Datacolor FX10 and X-Rite MA98, have incorporated a number of directional geometries higher to those recommended by international standards cited above.

A lot of contributions about the inter-comparison between spectrophotometers have appeared in the last years[2-8], however there are not enough studies with multi-gonio-spectrophotometers for the goniochromatism characterization[9-11]. For this, the purpose of this study is to compare the reproducibility of some multi-gonio-spectrophotometers, specifically the models Datacolor FX10 and X-Rite MA68II, following the ASTM E2214-08 rules[8] for the five common geometries. Finally, since our laboratory has recently acquired one X-Rite MA98, we will complete this reproducibility comparison with this third colour-measuring instrument.

MATERIALS

We have 91 metallic, interference, luster and pearlescent samples, collected by different manufacturers. And every spectral and color measurement has been taken by two gonio-spectrophotometers: Datacolor FX10 and X-Rite MA68II.

The multi-gonio-spectrophotometer MA68II is a portable device and has 5 geometries of illumination/observation, following the ASTM 2194[12] and DIN 6175-2[13] standards. These geometries are summarized in the table 1. This instrument belongs to the Technological Institute of Optics, Color and Imaging (Valencia, Spain).

On the other hand, the multi-gonio-spectrophotometer FX10 is a desktop device and has 10 different geometries of illumination/observation, including the 5 standard geometries of the MA68II. These geometries are also summarized in the table 1. This instrument, joint to one recent MA98, belongs to the Colour & Vision Group of the University of Alicante (Spain).

Table 1.Illumination and observation angles of the measures. (In bold you can see the common geometries).

Geometries	**ASTM**						**ASTM/DIN**			
Influx (incident) angle	25º	25º	45º	**45º**	75º	75º	**45º**	**45º**	**45º**	**45º**
Efflux (detection) angle (aspecular)	170º (-15º)	140º (+15º)	150º (-15º)	**120º (+15º)**	120º (-15º)	90º (+15º)	**110º (+25º)**	**90º (+45º)**	**60º (+75º)**	**25º (+110º)**

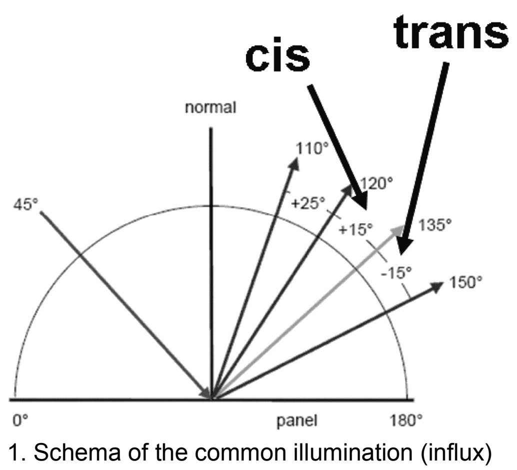
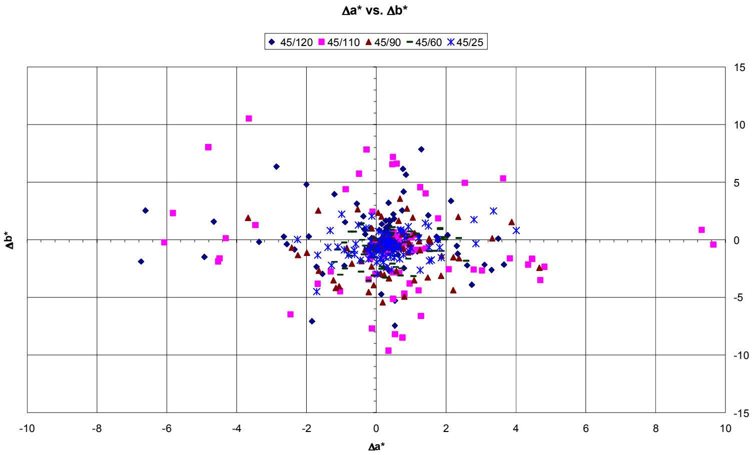


Fig 1. Schema of the common illumination (influx) and observation (efflux) angles.

As it was said earlier, current ASTM and DIN geometries are enough for characterization of metallic samples, but they are not enough for characterization of other special-effect pigments like pearlescent one or colour with other optic effects as glitter, sparkle, etc. However, in this study the purpose is to compare the results of the two or three multi-gonio-spectrophotometers, because of this, we only use the common geometries between both of them: 45ºx120º, 45ºx110º, 45ºx90º, 45ºx60º and 45ºx25º, which is in bold in the table 1 and are represented in the figure 1.

METHODOLOGY

The 91 goniochromatic samples were measured 20 times, without replacement, by the two multi-gonio-spectrophotometers, after a long time of stand-by (higher to 20 min). The temporal interval between the measurements done in each laboratory was not longer to 2 months. Relative reflectance factors from each respective matte white standard and colorimetric coordinates under illuminant D65 and standard observer CIE-1964 XYZ are the colour data obtained from these measurements for each one of these samples. From this first step, with the mean values of 20 measurements for each sample, we calculated the partial and total colour differences in CIELAB colour space measured in both instruments. If the reproducibility level was ideal, all colour differences should be zero.



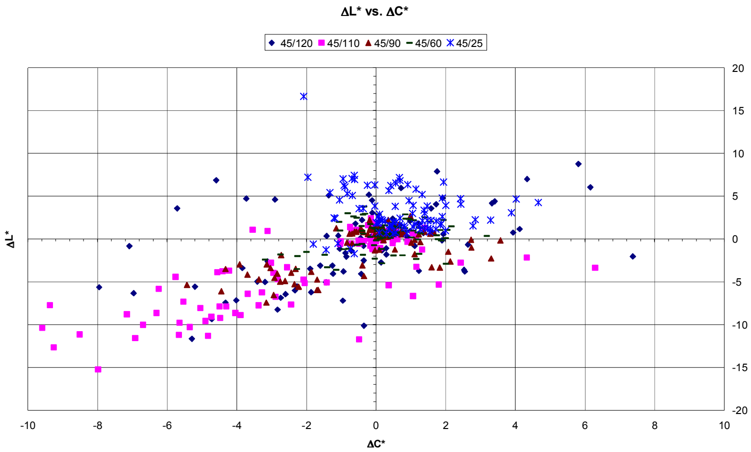


Fig 2. Diagrams Δa* vs Δb* and ΔC* vs ΔL* of the 91 samples to show its distribution

As we said before, the average and mean square deviation of the colorimetric values was made for the statistical studies of the reproducibility comparison between devices. These statistical studies include Hotelling's test and a statistical test of inter-comparison to know the confidence interval of the partial colour differences ΔL*, Δa*, Δb* and the total colour difference ΔE_{ab} between both colour-measuring instruments. The Hotelling's test and the inter-comparison-test have been used earlier by other authors[8] and they are really easy implementing it. The equations (1) to (5) describe the necessary process to calculate the Hotelling's parameter and the critical value $t_{\Delta E}$.

$$S = \begin{bmatrix} \mathrm{var}(\Delta L^*) & \mathrm{cov}(\Delta L^*, \Delta a^*) & \mathrm{cov}(\Delta L^*, \Delta b^*) \\ \mathrm{cov}(\Delta L^*, \Delta a^*) & \mathrm{var}(\Delta a^*) & \mathrm{cov}(\Delta a^*, \Delta b^*) \\ \mathrm{cov}(\Delta L^*, \Delta b^*) & \mathrm{cov}(\Delta a^*, \Delta b^*) & \mathrm{var}(\Delta b^*) \end{bmatrix} \qquad (1)$$

$$T^2 = n \cdot [\Delta L^* \quad \Delta a^* \quad \Delta b^*]^T \cdot S^{-1} \cdot [\Delta L^* \quad \Delta a^* \quad \Delta b^*] \tag{2}$$

$$\alpha = \frac{mean(\Delta L^*)}{mean(\Delta E_{ab})}, \quad \beta = \frac{mean(\Delta a^*)}{mean(\Delta E_{ab})}, \quad \gamma = \frac{mean(\Delta b^*)}{mean(\Delta E_{ab})} \tag{3}$$

$$g_E = g_{11}\alpha^2 + g_{22}\beta^2 + g_{33}\gamma^2 + 2g_{12}\alpha\beta + 2g_{23}\beta\gamma + 2g_{13}\alpha\gamma \tag{4}$$

$$t_{\Delta E} = \sqrt{\frac{\chi_3^2}{n \cdot g_E}} \quad , \quad n = 91 \tag{5}$$

where n is the total of measurements and χ^2 is the chi-square value for 3 degree of freedom. This critical value is very important in this study, because it fits the limit that it let us establish if the total colour differences ΔE_{ab} are statistically significant, that is, if it is likely to have occurred by chance or they are not. Specifically if the average is higher than critical value, mean (ΔE_{ab}) > $t_{\Delta E}$, the difference is significant, that is, that for directional geometry the measurement data are unlikely to have occurred by chance. This would be that differences between both instruments are due to systematic or bias errors (angle tolerances for each geometry, photometric scales, white standards, etc), but not exclusively to random errors.

RESULTS

Firstly, it is shown the relative reflectance factors with the five common geometries obtained by both devices for one of the measured sample, which commercial name is Mearlin Micro Blue 6303M A. These spectra are represented in the figure 3.

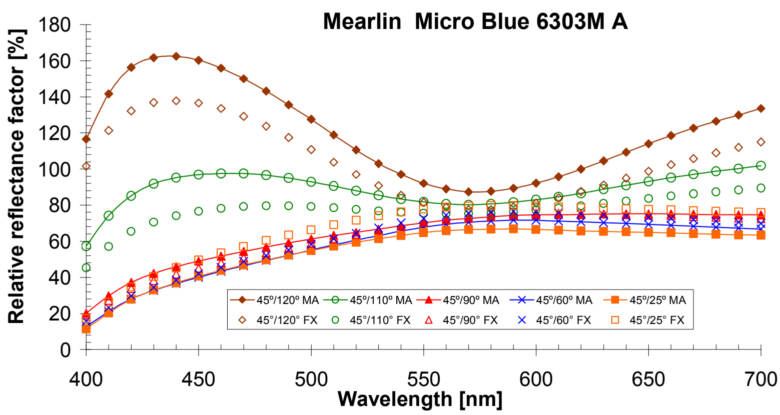


Fig 3. Relative reflectance factors obtained by both multi-gonio-spectrophotometers, MA68II and FX10, for their five common measurement geometries.

In this figure two details attract our attention. The first detail is that the relative reflectance factors for the 45ºx120º geometry are greater than 100 %. It is so due to the existence of metallic and/or pearlescent particles and the fact that the observation angle 120º is near the specular angle (135º) for an illumination angle of 45º.

The other detail you can see in the figure 3 is that the relative reflectance factors that have worst reproducibility (or coincidence) are the corresponding to 45ºx120º and 45ºx110º geometries, and it is due to the fact that the observation angle is too near the specular angle. For the other geometries, 45ºx90º, 45ºx60º and 45ºx25º, you can see that the measured reflectance factors by the devices are very similar. In fact, considering the results, it seems that the observation angle is further from specular angle, the agreement between relative reflectance factors is better.

Secondly, the colorimetric values CIE-L*a*b*, which have been obtained from these spectra, are shown in the next figure:

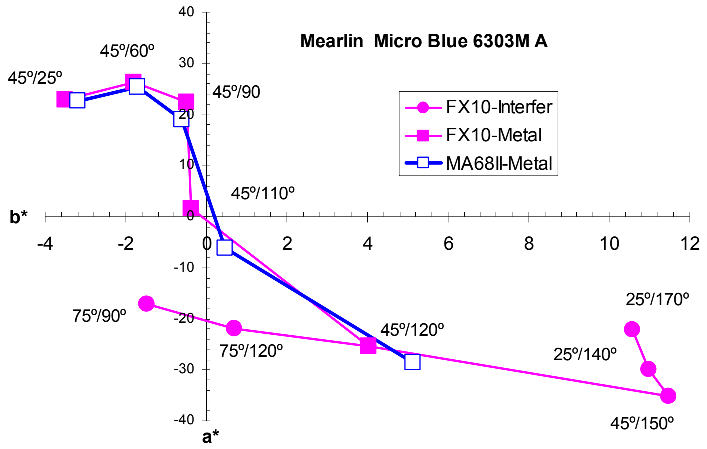


Fig 4. CIE-L*a*b* colorimetric coordinates obtained by the both multi-gonio-spectrophotometers, MA68II and FX10, for all their geometries

In the figure, you can see the colorimetric coordinates obtained by the two multi-gonio-spectrophotometers with all the measured geometries, ten by the FX10 and five by the MA68II. The metallic samples present a characteristic T-shape[9,14], with a part associated to metallic effect (same incidence angle 45º, but different observation angles) and other part associated to interference effect (different incidence angle, but the same observation angle, the aspecular angle ±15º, see figure 1). Again you can see the slight discrepancy between the measurements by the two devices, overcoat for nearby angles to specular angle.

Next we made the multivariate statistical analysis of the differences between the colour coordinates measured by the two gonio-spectrophotometers for the 91 special samples. The result from this study is given in the next table:

Table 2. Hotelling's analysis T^2 for colour differences of 91 samples measured by both spectrophotometers (FX10 and MA68II) with a confidence interval of 95%.

Geometry	Sample size	Variables	T^2	χ^2	df	P
45°x120°	91	3	3.2568	3.2568	3	0.3537
45°x110°	91	3	62.7269	62.7269	3	0.0000
45°x90°	91	3	29.5718	29.5718	3	0.0000
45°x60°	91	3	92.3155	92.3155	3	0.0000
45°x25°	91	3	237.5260	237.5260	3	0.0000

where T^2 is Hotelling's T-Squared statistic and P is the probability that null H_0 were true, performed using a Matlab routine[15] freely downloaded from Internet. Considering the results we can say that the probability of the colour differences for the 45ºx120º geometry was due to at random is statistically no significant. And therefore, we can conclude that the difference is due to at random and is not due to systematic or bias errors. Consequently, for the 45ºx120º geometry both devices are statistically equivalent. On the other hand, for the rest of the geometries, the differences are significant, i.e. there are statistical evidences that the colour differences are due to systematic errors.

The following statistical test were the ASTM inter-comparison test to know the confidence interval of the partial colour differences ΔL*, Δa*, Δb* and the total colour difference ΔE_{ab} between the measures from both spectrophotometers. After averaging out at about 20 measures of each of the 91 samples and computing the colour differences between devices we obtained the results shown in the tables 3 and 4.

Table 3. Averages maximum and minimum values of the partial colour differences obtained for every measurement geometry in the multi-gonio-spectrophotometers FX10 and MA68II.

	45ºx120º			**45ºx110º**			**45ºx90º**			**45ºx60º**			**45ºx25º**		
	ΔL*	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***
Average	-0.68	0.11	0.22	-3.63	0.50	-0.47	1.09	0.39	-0.73	0.41	0.34	-0.66	3.21	0.34	-0.54
|Max|	11.66	6.74	7.85	15.25	9.65	10.53	7.43	4.67	5.43	3.77	2.57	3.53	16.66	4.01	4.48
|Min|	0.11	0.01	0.05	0.03	0.02	0.01	0.10	0.00	0.00	0.03	0.02	0.00	0.51	0.02	0.01

Specifically in the table 3 you can see some results calculated from the 91 samples for each of the geometries. Firstly at all, you can see the average of the partial colour differences, ΔL*, Δa* and Δb* and the maximum and minimum values of these partial colour differences. Then, it is obvious that the colour differences shown here are clearly higher to perceptibility limits, in many cases passing usual industrial colour tolerances or acceptability limits.

After that, the covariance matrix S and its inverse have also been calculated (following equations from 1 to 4). And finally, the critical value $t_{\Delta E}$ has been obtained from the inverse covariance matrix and the values α, β y γ. This critical value is very important in this study, because it fits the limit that it let us establish if the total colour differences ΔE_{ab} are statistically significant, that is, if it is likely to have occurred by chance or they are not.

In the next table, you can see the total colour differences ΔE_{ab} and the critical value $t_{\Delta E}$ calculated for each measurement geometry between the instruments FX10 and MA68II.

Table 4. Averages and critical values of the total colour differences ΔE_{ab} obtained for every common measurement geometry in the multi-gonio-spectrophotometers FX10 and MA68II.

	45ºx120º			**45ºx110º**			**45ºx90º**			**45ºx60º**			**45ºx25º**		
	ΔL*	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***	**ΔL***	**Δa***	**Δb***
S	18.22	-0.34	-0.52	19.67	-1.10	3.51	6.46	-0.11	2.23	2.64	0.19	1.00	6.93	-0.05	0.87
	-0.34	3.34	-0.36	-1.10	5.78	-0.98	-0.11	1.41	0.01	0.19	0.57	0.14	-0.05	1.01	0.32
	-0.52	-0.36	6.18	3.51	-0.98	12.46	2.23	0.01	3.21	1.00	0.14	1.32	0.87	0.32	1.35
α,β,γ	-0.15	0.03	0.05	-0.67	0.09	-0.09	-0.39	0.13	-0.26	0.20	0.17	-0.33	0.85	0.09	-0.14
g_E	0.0018			0.0232			0.0413			0.2473			0.1824		
$t_{\Delta E}$	6.85			1.92			1.44			0.59			0.69		
Mean(ΔE_{ab})	4.42			5.45			2.80			2.03			3.78		

After making comparison between the critical value, $t_{\Delta E}$, and the average of the total colour differences, shown in the earlier table, we can establish what differences are statistically significant or are not. In our case, FX10 vs. MA68II, only the 45ºx120º geometry is not significant, the rest of the geometries are statistically significant because the averages are higher than critical values, mean(ΔE_{ab}) > $t_{\Delta E}$, that is, these geometries are unlikely to have occurred by chance. These results agree with the results previously obtained by the Hotelling test for colour differences.

Finally, although without enough space to describe a complete reproducibility comparison, we have got all results for the pair comparisons FX10 vs. MA98 and MA68II vs. MA98. We may advance the main results as follows:

- FX10 vs. MA98: the smallest *mean*(ΔE_{ab}), equals to 0.98 jnd, is for the geometry 45ºx90º, but it is not enough to pass the inter-comparison test ($t_{\Delta E}$ = 0.23). Therefore, the five analyzed common geometries are statistically significant. However, this reproducibility comparison is not complete because both instruments share other measurement geometries as 75ºx120º, 75ºx90º, etc. Consequently, this reproducibility comparison should be described with more space for the final presentation of this contribution.

- MA68II vs. MA98: the smallest *mean*(ΔE_{ab}), equals to 1.11 jnd, is for the geometry 45ºx25º (retro-reflection), being enough to pass inter-comparison test ($t_{\Delta E}$ = 1.27). Therefore, except this measurement geometry, and in spite of the fact that both instruments belong to the same manufacturer, the rest of four analyzed common geometries are statistically significant.

CONCLUSION

Finally the conclusion is that most of the geometries are statistically significant. This means that these differences are due to systematic or bias errors (angle tolerances for each geometry, photometric scales, white standards, etc), but not exclusively to random errors. However, both statistical tests used here are not valid to discriminate and quantify the detected bias errors in this inter-instrument comparison. For pair comparison FX10 vs. MA68II, only the nearest measurement geometry, 45ºx120º, to the specular direction (135º), with a priori a large photometric scale, shows a pure statistical deviation in both instruments. These results also show the intrinsic difficulty to find efficient methods for comparing the reproducibility in multi-gonio-spectrophotometers, even between models belonged to same manufacturer. For instance, for the comparison MA98 vs. MA68II, only the measurement geometry 45ºx25º (retro-reflection) passed the statistical inter-comparison test.

ACKNOWLEDGEMENTS

This research was supported by the Spanish Ministry for Education and Science by means of the grant number DPI2005-08999-C02-02, and, the Spanish Ministry of Science and Innovation by the grant number DPI2008-06455-C02-02.

REFERENCES

1. H.J. Streitberger and K.F. Dössel, *Automotive Paints and Coatings*, 2nd Edn, Wiley-VCH: Weinheim, Germany, 2008.
2. J. Campos, M. Rubiño, F. Castillo and A. Pons, "Comparación de instrumentos de medida del color". *Optica Pura y Aplicada*,Vol. **37**, pp. 113–118, 2008.
3. J. Campos, A. Pons and J. Martín, "Reflectance changes in white reflectance standards measured in different instruments with 0/d geometry". Proceeding 10th Congress of the International Colour Association - AIC, pp. 41–44, 2005.
4. A.R. Hanson, M.R. Pointer and P.J. Clarke, "The Uncertainty of surface colour measurements made with commercially available instrumentation and recommended calibration protocols", *NPL Report DQL-OR 005,* National Physical Laboratory: Teddington, UK, 2004.
5. J. Hiltunen, *Accurate Color Measurement*. MSc Dissertation, Faculty of Science, University of Joensuu, Finland, 2002.
6. J. Taylor, A.R. Hanson, P.J. Clarke, J. Campos, A. Pons, W. Czepluch, T. Jaaskelainen, J. Hiltunen, J. Gaudemer, O. Soares, J.L.C. Costa and L. Liedquist, "Good practice guide to surface colour measurement". *Publicación Europea SMT4-CT96-2140 EUR 19552EN*, 2000.
7. D.R. Wyble and D.C. Rich, "Evaluation of methods for verifying the performance of color measuring instruments. Part I: repeatability". *Color Research Application*, **32**, pp. 166–175, 2007.
8. D.R. Wyble and D.C. Rich, "Evaluation of methods for verifying the performance of color measuring instruments. Part II: inter-instrument reproducibility", *Color Research Application*, **32**, pp. 176–194, 2007.
9. W.R. Cramer and P. Gabel, "Measuring special effects", *Paint and Coating Industry*, **29**, pp. 36–46, 2001.
10. E. Chorro, F.M. Martínez-Verdú, J. Campos and A. Pons, "On the importance of rotating metallic and pearlescent samples around the normal in multi-gonio-spectrophotometric measurements", *CIE Expert Symposium 2008 on Advances in Photometry and Colorimetry*, CIE Expert: Wien, Austria, pp. A1–A6, 2008.
11. V. Navarro, *Análisis de la incertidumbre en la medida del goniocromatismo*, MSc Dissertation, Faculty of Sciences, University of Alicante, 2009.
12. ASTM, "Standard Practice for Specifying and Verifying the Performance of Color-Measuring Instruments", *Standard ASTM E2214-08*, West Conshohocken, 2008.
13. ASTM, "Standard Terminology of Appearance", *Standard ASTM E284-08*, West Conshohocken, PA, 2008.
14. W.R. Cramer, "Examples of interference and color pigment mixtures green with red and red with green", *Color Research Application*, **27**, pp. 276–281, 2002.
15. http://www.mathworks.com/matlabcentral/fileexchange/2844 .