International Commission on Illumination
Commission Internationale de l'Eclairage
Internationale Beleuchtungskommission

PSYCHOPHYSICAL EXPERIMENT TOWARDS A MEASUREMENT SCALE OF SPARKLE

Slembrouck, N., et al.

DOI 10.25039/x051.2025/rjdv5n

This article is also published as part of:

Proceedings of the CIE 2025 Midterm Meeting Vienna, Austria, July 4-11, 2025:
Scientific Conference (July 7-9, 2025)

DOI 10.25039/x051.2025

in

Proceedings of the CIE (International Commission on Illumination)

ISSN no. 3061-015X (print), 3061-0168 (online)

The paper has undergone double-blind peer review and its final version has been presented at the CIE 2025 Midterm Meeting, Vienna, Austria, July 4–11, 2025.

CIE Central Bureau
Babenbergerstrasse 9/9A
A-1010 Vienna, Austria
Tel.: +43 1 714 31 87
e-mail: ciecb@cie.co.at — www.cie.co.at

PSYCHOPHYSICAL EXPERIMENT TOWARDS A MEASUREMENT SCALE OF SPARKLE

Slembrouck, N.[1], Ferrero, A.[2], Perales, E.[3], Basic, N.[4], Audenaert, J.[1], Leloup, F.B.[1]

[1] KU Leuven, Ghent, BELGIUM, [2] Instituto de Óptica "Daza de Valdés", Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, [3] University of Alicante, Alicante, Spain,[4] Federal Institute of Metrology (METAS), Bern-Wabern, Switzerland

Nathan.slembrouck@kuleuven.be

Abstract

The visual appearance effect known as sparkle, defined as observed bright points of light contrasted with a darker surround, is frequently used to enhance modern product finishes. Existing commercial instruments measure sparkle using proprietary methods, lacking standardisation. To address this issue, CIE is working on the recommendation of a sparkle measurement scale based on luminance factor images. This paper presents the results of a study on the correlation between sparkle as optically quantified through luminance factor images and as visually assessed through a magnitude estimation experiment. To this end, 66 samples were selected, obtained from different manufactures and exhibiting different sparkle effects. The study demonstrates that the degree of correlation is strongly dependant on the sample type. This result illustrates that more research is needed for defining a general measurement and or processing procedure that is applicable to a wide variety of sample types.

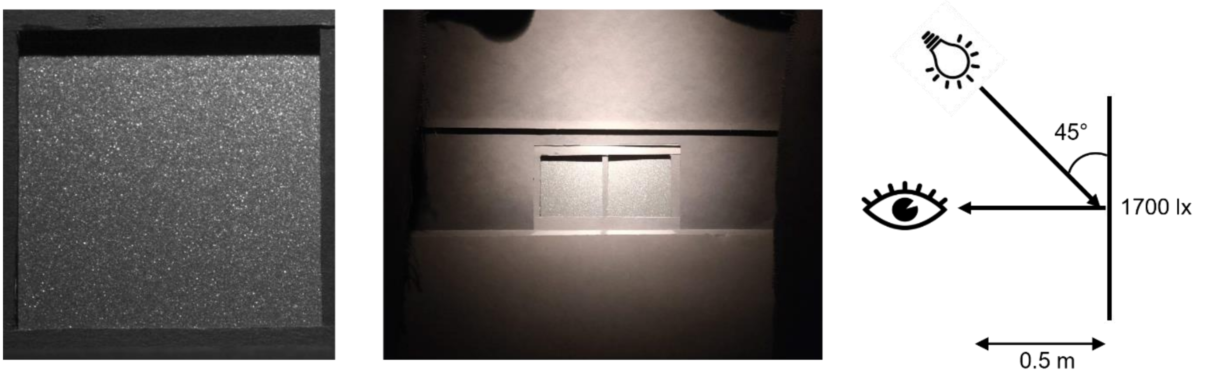
Keywords: Visual Appearance, Soft Metrology, Sparkle

1 Introduction

Modern product finishes are often enhanced by incorporating interesting visual appearance effects. One of those effects is sparkle, which according to ASTM E284-17 is defined as *"the aspect of the appearance of a material that seems to emit or reveal tiny bright points of light that are strikingly brighter than their immediate surround and are made more apparent when a minimum of one of the contributors is moved"* (ASTM, 2022). An example of the sparkle effect can be found in Figure 1 a. The sparkle effect is similar to the appearance of stars on a night sky, where small bright dots (point light sources) are clearly visible in contrast to a darker surround. Sparkle can only be seen under directional illumination, since the effect originates from mirrorlike reflections caused by small metallic flakes or effect pigments in the material. For this reason, sparkle can be heavily influenced by the orientation of the flakes / pigments when assessed under different illumination and viewing angles.

Measuring sparkle is challenging, since it is a psychophysical quantity that involves a combination of several physical quantities such as the background luminance, size of the sparkle points and their spatial distribution. To reduce measurement complexity, currently only techniques that measure 'static sparkle', where there is no sample or detector movement involved, exist. Currently, three commercially available instruments can measure static sparkle, but each device uses its own proprietary sparkle scale. As a result, information about their working principles is hard to find and standardisation is lacking. Currently, CIE is working on the recommendation of a measurement scale for sparkle in which a procedure is described to calculate the visibility distribution of sparkle spots, based on a luminance factor image of the sample (Ferrero, 2020, 2021).

The presented work aims to study the correlation between the measurement scale and a visual scale of a set of sparkle samples. To this end, a subset of 66 samples was selected from a larger set of 180 samples, obtained from different manufacturers and exhibiting different visual sparkle effects. Luminance images of all test samples were acquired and processed, and a magnitude estimation experiment was performed by 20 observers, from which a visual sparkle scale was derived (ASTM, 2005).



a b c

Figure 1 - Setup of the visual experiment. (a) enlarged image of a sparkle sample in the setup (b) observer point of view (c) setup geometry

2 Experiment setup

2.1 Sample set

Samples were gathered from a variety of coating manufacturers to include different types of sparkle in the study. This lead to the selection of four different industrial sample families being used, denoted as “Standox”, “Luxan”, “GS” and “WF”. Only partial information on the composition of the samples is available. The pigment types are metal flakes or pearlescent pigments, with the effect pigments being made from Al, TiO2 or glass. The pigment shape is either silverdollar (i.e. a coin-looking pigment that contains a thin, smooth and flat surface), or cornflake shaped (i.e. a rougher pigment shape). In general, silverdollar pigments scatter the light towards one preferred direction, while the cornflake pigments scatter the light more uniformly over a larger variety of angles (Wissling, 2006). Some samples have constant pigment sizes, while for others the size varies. The density of the pigment distribution varies across the different samples as well, leading to samples with sparse sparkle points and other samples with a dense sparkle dot distribution. Most samples have low chroma (grey samples), except for three pairs of samples showing a green, blue and subtle red tint.

2.2 Visual experiment

Twenty observers participated in the visual experiment, in which two samples are viewed perpendicularly, side by side (Figure 1 b). The observed area of each sample is 4x4 cm and the observer views the two samples from a distance of 0.5 m. A dimmable, high intensity LED source (Xlamp® XHP35.2 High-Intensity) with a colour rendering index R_a of 90 and a correlated colour temperature of 2700 K illuminates the samples under test. In front of the source, a collimating lens (Gaggione® LLC59N) is mounted to approximate directional illumination under an incidence angle of 45°. The light source is placed at a distance of 0.5 m from the samples, which results in a subtended angle from the left border of the left sample to the right border of the right sample of approximately 8°. The resulting illuminance on the samples, as measured on the sample plane in the centre of the viewing window, is set to 1700 lx (Figure 1 c).

To derive a visual scale, the magnitude estimation procedure (ASTM, 2005) was implemented. A reference sample, assigned an arbitrary sparkle value of 100, is presented continuously during the experiment. Next to the reference sample, a test sample is presented, of which the observer rates the sparkle level. Half of the observers were shown the reference sample in the left window, the other half ware shown the anchor in the right window. Observers were allowed to rate between 0 (no sparkle) and (theoretically) infinity.

2.3 Sparkle measurement scale

A luminance image was captured of each sample in a 45°:0° (illumination:viewing) geometry, using the MARS (Multi-Angle Reflectance Setup) instrument available at the Swiss Federal Institute of Metrology METAS (Ferrero *et al.*, 2020). From these luminance images, the distribution of the sparkle dot visibility of a given sample was determined according to the method described in (Ferrero *et al.*, 2021). The visibility of a sparkle spot is thereby defined as the ratio of local contrast to the visibility contrast threshold. The visibility distribution can be expressed in terms of percentiles or quantiles (see Figure *2*). In this study, the percentiles 50, 75 and 95 of visibility (further denoted as P50, P75 and P95, respectively) were used. From Figure 2, it can be observed that the “Luxan” samples (sample names that start with “x”) exhibit a sparkle visibility distribution which is more skewed than other samples and that they contain a small set of sparkle spots with an extremely high contrast, as illustrated by their large P95 value. Moreover, it can be observed that the width of the visibility distribution, here defined as the range between P25 and P75, varies as well between the types of samples. In general, samples exhibiting a higher degree of visual sparkle have a larger spread of sparkle dot visibility.

DOI: 10.25039/x051.2025/rjdv5n

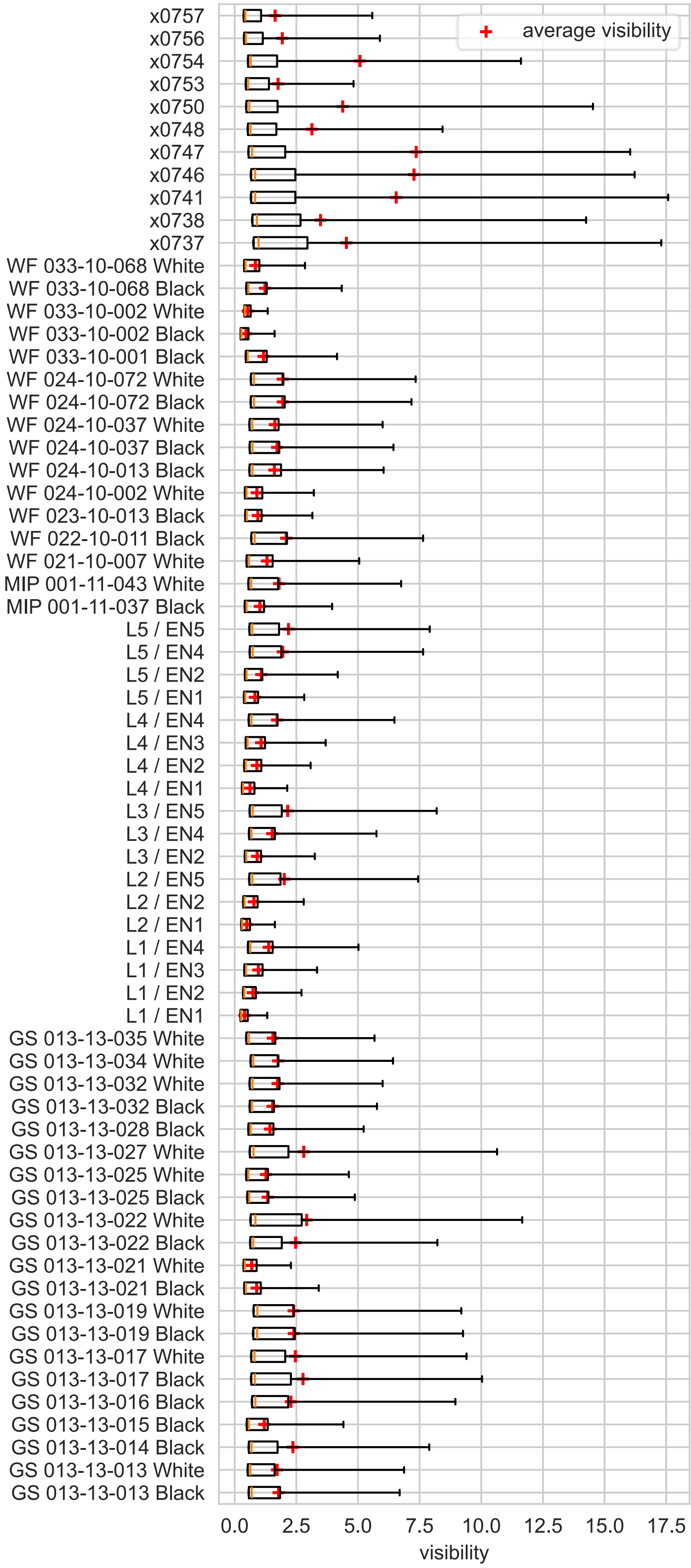


Figure 2 – Boxplots based on percentiles of the measured sparkle visibility distribution, for a virtual aperture size of 1 pixel. The upper whiskers represent P95.

A crucial parameter for measuring a correct sparkle distribution is the virtual aperture size, which sets the size of the window that is used when the image is being scanned for sparkle dots. This window size should match the size of the sparkle dots. If it is too small, one sparkle dot might be counted as multiple points, while if it is too large, multiple sparkle dots might be measured as one big sparkle point. Therefore, the measurements were processed and analysed for different virtual apertures of a width of 1, 3, 5 and 7 pixels, respectively, in order to select the correct aperture size.

3 Results

3.1 Observer variability

The intra and inter-observer variability was evaluated using the standardised STRESS index (Melgosa *et al.*, 2011). The average STRESS value for the intra-observer variability numbers 28.0, with the lowest STRESS value being 16.4 and the highest being 40.0. The calculated STRESS values for inter-observer variabilities range from 20.18 to 43.9 STRESS units, with an average of 30.3. All STRESS values are presented in Figure 3.

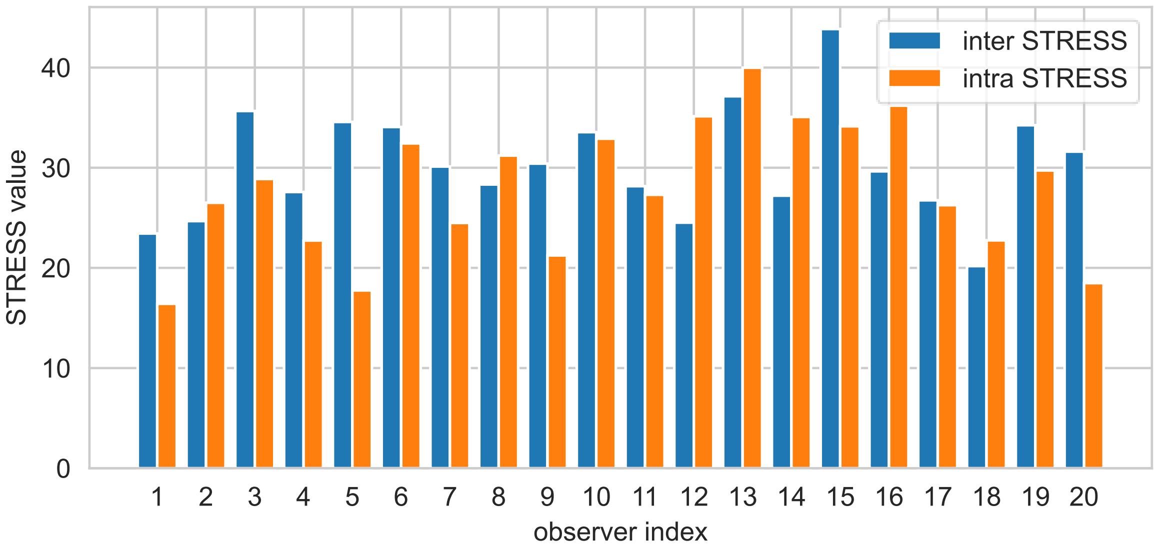


Figure 3 - Intra and inter-observer variability, characterized by the STRESS values for all 20 observers.

A principal component analysis was used to evaluate the potential multi-dimensionality of the visual sparkle scale. From this analysis, the explained variance from the first principal component is 81%, with the second component only explaining 4% of the total variance. When plotting the first and second principal components, it becomes clear that a subset of samples correlate less with the first principal component (Figure 4). These are the “Luxan” samples, which often contain a sparse but strongly contrasting sparkle dot distribution. In a previous experiment, these samples also showed the largest deviations when a commercial sparkle instrument was compared against a visual scale from a different visual experiment (Gómez *et al.*, 2016). Only the first principal component seems convincingly correlated with the visual scale, explaining 81% of the variation in the visual data. This suggests that all observers grade sparkle in a similar way.

DOI: 10.25039/x051.2025/rjdv5n

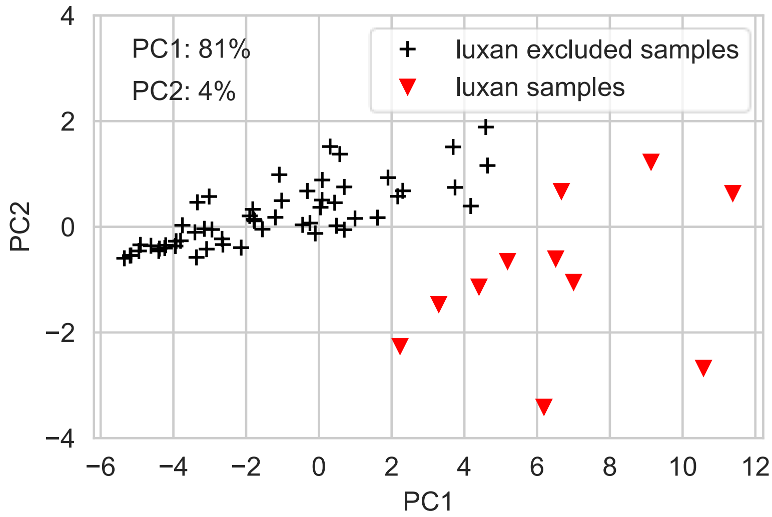


Figure 4 - First and second principal components for all samples of the visual scale.

3.2 Initial correlation between visual scale and measurement scale

The Pearson R^2 correlation coefficients between the visual scale and candidate measurement scales based on the sparkle visibility distribution were calculated. The candidate measurement scales are: sparkle density, average sparkle dot visibility and the P50, P75 and P95 visibility percentiles. In addition, this evaluation was performed for all provided virtual aperture sizes (1,3,5 and 7 pixels). Best results were achieved for a virtual aperture size of 7 pixels in width. From Figure 5 it can be observed that the average sparkle dot visibility provides a good fit, with a Pearson R^2 coefficient of 0.81. The alternative parameters provide a measurement scale that is less convincing, however.

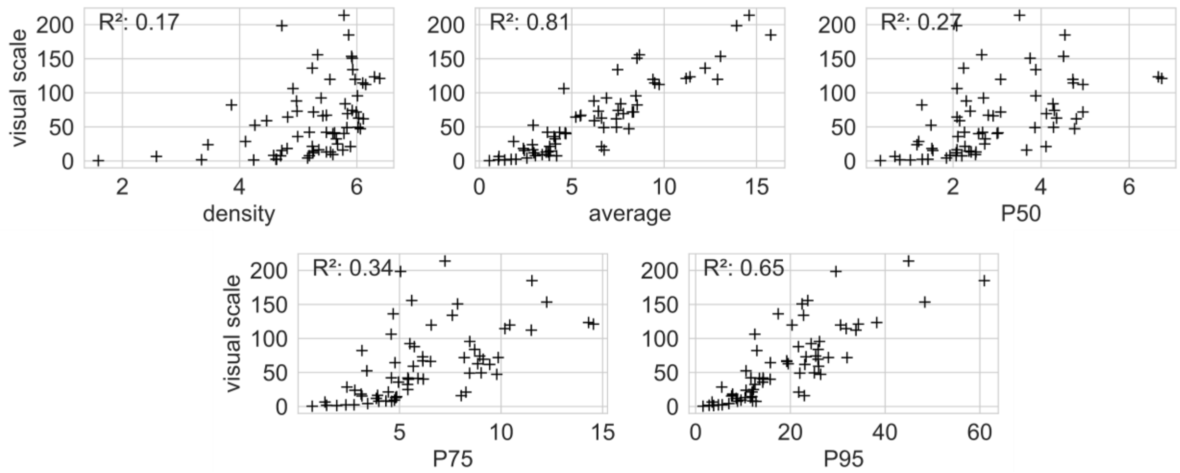


Figure 5 - Evaluation of the visual scale against candidate measurement scales.

The average sparkle dot visibility scale was further examined in detail. When calculating the correlation for each of the four sample types separately, it becomes clear that this correlation is strongly dependent on the sample type. Two sample types have a strong correlation, while for the other two types of samples the correlation is very weak (see Table 1). The "Standox" sample set has the highest R^2 coefficient of 0.93. Interestingly, these are also the samples that were used when the measurement procedure was originally being developed (Ferrero *et al.*, 2020). The "Luxan" sample set, on the contrary, shows a very weak R^2 coefficient of 0.47.

Table 1 - Pearson R^2 correlation coefficients between the visual scale and the measurement scale based on the average sparkle dot visibility, for different sample sets.

Sample set	Standox	GS	WF	Luxan
Pearson R^2	0.93	0.86	0.57	0.47

3.3 Correlation for different percentiles

The large differences in correlation between the visual scale and measurement scale based on the average sparkle dot visibility for the different types of samples, might originate from their different shapes of sparkle visibility distributions. Therefore, utilizing different percentiles could increase the performance of the measurement procedure. A second measurement series gathered from luminance images taken from the point of view of the observers in the visual experiment setup was used to gather the sparkle visibility distributions. From this, a series of percentiles, starting from 0 to 100 in steps of 0.1 %, was plotted against its correlation with the visual scale. As can be seen in Figure 6, in general the correlation seems to increase with the sparkle visibility percentile value. However, this increase is not equal for each sample type. Especially for the "Luxan" samples, a relative good correlation is obtained only for higher percentiles larger than P90, which suggests that for these types of samples, observers judged sparkle based on the few very bright sparkle dots that these samples exhibit. For other sample types, a more constant increase in correlation can be observed with larger percentiles. Interestingly, for the "GS" samples, a sparkle visibility percentile threshold value can be defined above which the correlation starts to drop (i.e. P85). Crucially, for the "Luxan" samples a very poor correlation is obtained at this P85 threshold. This ultimately means that we cannot generalise the image analysis procedure for all four types of samples. While for the "GS" samples the best correlation is obtained with a P75 sparkle visibility percentile, the P95 threshold provides the best correlation for the three other types of samples

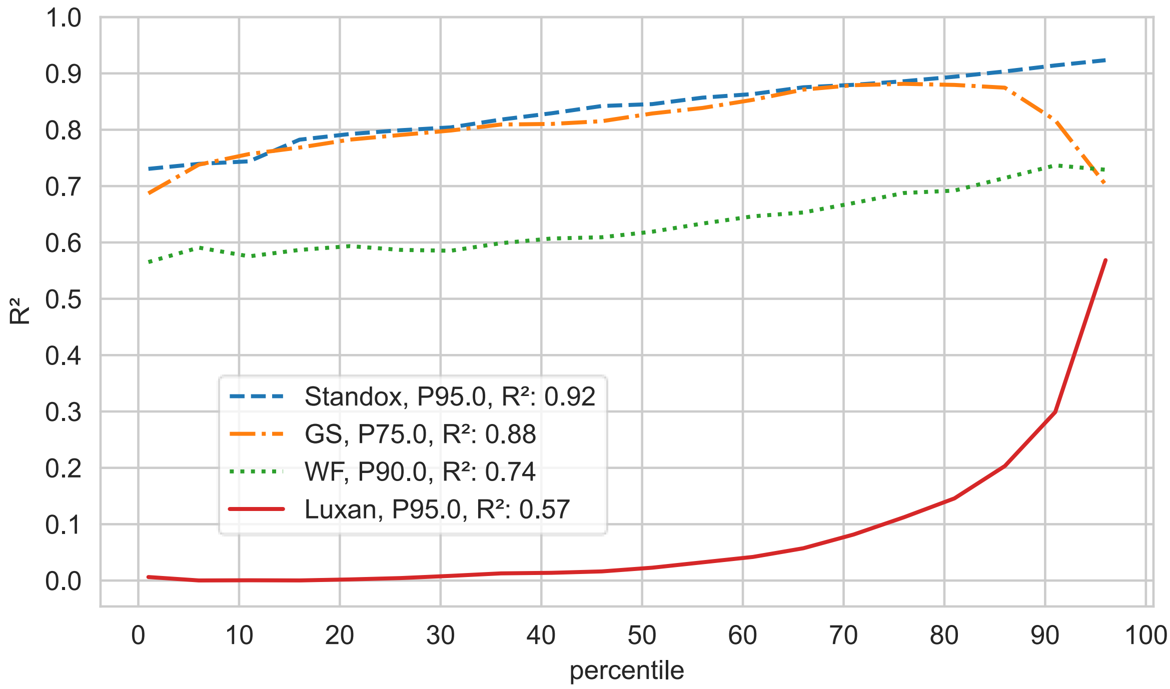


Figure 6 - Pearson R^2 correlation in function of the sparkle visibility percentiles. The text in the legend indicates the percentile for which the best correlation was found and its corresponding best R^2 value.

4 Conclusion

A visual experiment was performed to investigate the correlation between the visual perception of sparkle and a new sparkle measurement scale being developed. The relation between the visual data and measured density, average, P50, P75 and P95 percentiles of the sparkle visibility distribution was investigated. In addition, other percentiles were examined to investigate if they could improve the correlation.

The use of percentiles to describe the sparkle visibility distribution enables a more in-depth characterisation of the sample. Samples with a small amount of high-contrasting sparkle points, for example, could be easily distinguished from samples exhibiting lower contrasting sparkle dots. While this more detailed analysis looks promising to further increase the correlation between a measurement model and a visual sparkle scale, so far the exact formula to do this has not been found.

At first, a good correlation (R^2 0.89) between the average visibility and the visual data was observed. However, when different sample types were examined separately, it was demonstrated that the correlation strongly depends on the sample type. For each type of samples, the correlation could be improved by including different sparkle dot visibility percentiles. Unfortunately, a general rule of thumb to determine a threshold value for the sparkle dot percentile is missing and the influence of including one or another threshold value differs across the various sample types. Therefore, a general method for determining a measurement scale that is robust to the type of sample has not yet been found. To conclude, more research is needed to examine better ways of adapting the method to cover more types of sparkle samples.

Funding. This study has been partially supported by KU Leuven Internal Funds; Special Research Fund (BOF) - STG/22/008

References

ASTM 2005. ASTM E1697-05 Standard Test Method for Unipolar Magnitude Estimation of Sensory Attributes, ASTM international. https://doi.org/10.1520/E1697-05R20.

ASTM 2022. ASTM E284-17 Standard Terminology of Appearance, ASTM international. https://doi.org/10.1520/E0284-17.

FERRERO, A. *ET AL.* 2020. An insight into the present capabilities of national metrology institutes for measuring sparkle. *Metrologia*, 57(6), 065029. https://doi.org/10.1088/1681-7575/abb0a3.

FERRERO, A. *ET AL.* 2021. Preliminary measurement scales for sparkle and graininess. *Optics Express*, 29(5), 7589. https://doi.org/10.1364/OE.411953.

GOMEZ, O. *ET AL.* 2016. Visual and instrumental correlation of sparkle by the magnitude estimation method. *Applied Optics*, 55(23), 6458. https://doi.org/10.1364/AO.55.006458.

MELGOSA, M. *ET AL.* 2011. Notes on the application of the standardized residual sum of squares index for the assessment of intra- and inter-observer variability in color-difference experiments. *Journal of the Optical Society of America A*, 28(5), 949. https://doi.org/10.1364/JOSAA.28.000949.

WISSLING, P. 2006. *Metallic effect pigments: Fundamentals and Applications*. Hannover: Vincentz Network.