



Comparison of digital photography and spectrometry for evaluating colour perception in humans and other trichromatic species

Chen Yang^{1,2} · Jie Wang² · Nan Lyu³ · Huw Lloyd⁴

Received: 15 April 2021 / Revised: 17 August 2021 / Accepted: 20 August 2021 / Published online: 8 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Digital photography and spectrometry are widely used for colour measurement, but both methods have a number of advantages and disadvantages. Comparative studies can help determine the most appropriate method for quantifying animal colour perception, but few have attempted to compare them based on colour model conversion. Here we compare colour measurements from digital photography and spectrometry in a controlled standard experimental environment using the three-dimensional colour space model CIE $L^*a^*b^*$ which is designed to approximate colour perception in humans and assess the repeatability and agreement of the two methods. For digital photography, we first extracted RGB values from each colour patch and transferred these to $L^*a^*b^*$ values using colour model conversion. For spectrometry, we measured the spectral reflectance (SR) value and subsequently transferred SR values to $L^*a^*b^*$ values. Using a consensus of correlation analysis, intraclass correlation coefficients, concordance correlation coefficients, and Bland-Altman analysis, we found that although spectrometry showed a slightly higher repeatability than photography, both methods were highly repeatable and showed a strong agreement. Furthermore, we used Bland-Altman analysis to derive the limits of agreement, which can be used as criteria for identifying when photography and spectrometry could be as a suitable alternative for measuring colour perception in humans and other trichromatic species. We suggest that our workflow offers a practical and logical approach that could improve how we currently study colour perception in trichromats.

Significance statement

Measuring colour efficiently and accurately is necessary for investigating the evolutionary biology of colour perception in animals. Digital photography and spectrometry are two methods widely used for colour measurement, but there are benefits and limitations to using either method. Comparative studies based on colour model conversion are therefore critical for helping researchers determine which method is most appropriate. Here we test the repeatability and agreement of the two measuring methods using standard colour patches, as a comparative case study of broader interest in measuring colour perception in humans and similar primates. Our results demonstrate that both methods are highly repeatable, and the two methods may be used interchangeably to measure colour perception in humans under experimental conditions.

Keywords Colour measurement · Digital photography · Spectrometry · CIE $L^*a^*b^*$ · ICC · Bland-Altman analysis

Communicated by K. McGraw

✉ Nan Lyu
nanlyu@bnu.edu.cn

¹ Institute of Qinghai-Tibetan Plateau, Southwest Minzu University, Chengdu 610041, China

² Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China

³ Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, No. 19, XinJieKouWai St., HaiDian District, Beijing 100875, China

⁴ Department of Natural Sciences, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, UK

Abbreviations

<i>SR</i>	Spectral reflectance
<i>ICC</i>	Intraclass correlation coefficient
<i>CCC</i>	Concordance correlation coefficient
<i>LoA</i>	Limits of agreement
<i>CR</i>	Coefficient of repeatability

Introduction

Colouration in biology is defined as the appearance of an organism through reflecting or emitting the light from its surfaces (Osorio and Vorobyev 2008). Colouration is a dynamic and complex characteristic influenced by colour types and distributions of pigments; the shape, position, posture, and movement of the organism; and both the quality and quantity of light source (Langkilde and Boronow 2012; Fairchild 2013). How to measure colour and assess colour differences are key considerations for testing evolutionary hypotheses or colour perception in animals, requiring a method that is fast, simple, and accurate. Consequently, researchers need to compare different methods of colour measurement, such as the performance of a newly developed method with a more established one, to determine whether these methods can be used interchangeably or whether the new method performs better than the established one (Kawchuk and Herzog 1996; Mani et al. 2008). Most empirical studies of animal colour perception have been conducted using a single method (e.g. Akkaynak et al. 2013; Smith et al. 2016; Ligon and McGraw 2018). Direct comparisons of colour measurements from different methods using the same three-dimensional colour space models remain difficult without standardized testing, that could help resolve issues around measuring error and the acceptable range of the different methods, through conducting preliminary experiment to build acceptable conditions in colour measuring.

Accompanied by improvements in measuring accuracy, reduced costs for purchasing and using cameras and spectrometers, and the availability of related colour analysis software in personal computing, the application of conventional spectrometers and digital photography in colour quantification has become more widespread (Villafuerte and Negro 1998; Johnsen 2016; Smith et al. 2016). Spectrometers measure absorption spectra and radiant flux as a function of wavelength, providing more precise intensity distributions of the reflected wavelengths (200–1200 nm), and thus are able to deal with near-ultraviolet, near infrared, and visible light, although not as a function of time (Cageao et al. 2001; Bybee et al. 2012). As a widely used non-invasive and non-contact sampling method, digital photography enables researchers to rapidly and objectively quantify colours (and colour patterns) using

minimal equipment (Stevens et al. 2007; Pike 2011). Spatial information of the sample such as size, shape, area, and test conditions such as focal length, lens, and illumination can be easily controlled (Yam and Papadakis 2004; McKay 2013). Image processing algorithms can be applied to digital images to analyse complex colour patterns such as barring, spotting, or streaking. Furthermore, if the test objects are large, mobile, and heterogeneous, the results from quantitative digital image colour pattern analysis may be more stable and accurate than colour measuring by spectrometer (van den Berg et al. 2020).

Despite this, there are several limitations to using either method, which may have implications for the types of animal colouration perception research they could be utilized for. For example, conventional spectrometers provide only point samples and have relatively low detecting efficiency on small areas (which depends on the size of the probe, which are typically <1 mm²) (Li et al. 2008) and the complex spatial structure of the object surface (Akkaynak et al. 2013). Thus measuring colour perception using spectrometers requires much more information about the spatial relationships between colours reconstructed from the geometry of the sampling array, and the spatial resolution is generally crude (Endler and Mielke Jr 2005; Li et al. 2008). Spectrometers are also considerably more expensive than digital photography equipment, which is more readily available (Bergeron and Fuller 2018). Digital cameras may over-represent certain colours and do not respond linearly to the spectral properties of light (Stevens and Cuthill 2005; Simons et al. 2012). The results of photography also depend on light source, distance to light source, camera settings (resolution, focal length, shutter speed, aperture size), and ambient interference (Montgomerie 2006; Eeva et al. 2008). Furthermore, image processing algorithms may differ among cameras, and numerous assumptions about data acquisition require strict compliance (Troscianko and Stevens 2015; van den Berg et al. 2020). Given the range of advantages and limitations of each method, how to select the most appropriate technique for colour measurement is a timely question in need of more empirical data. Moreover, comparative tests of digital photography and spectrometry for measuring colour perception in trichromatic species may provide some guidelines on standardizing colour measurements of colour perception in other animals.

Comparative colour measurement tests require not only an analysis of accuracy, but also the degree of reliability within and between both methods under repeat trials. Another key experimental parameter consideration is the selection of an appropriate colour space model, which would, for example, allow researchers to assess the suitability of using camera-derived pixel values as a proxy for

spectrometer-derived values, provided they are mediated by a suitable program. Several three-dimensional colour space models have been applied to quantify the relationships between the spectral profiles of colours to colour perception, including *RGB*, *CMYK*, *HSL*, *HSV*, and *CIE L*a*b** (Schanda 2007; Fairchild 2013). Of these, *CIE L*a*b** is designed to approximate human vision (wavelengths ranging from about 380 to 700 nm), where L^* refers to the lightness of the colour, with 0 producing black and 100 producing a diffused white; a^* refers to the change from greenness to redness (a smaller a^* value corresponds to more greenness, whereas a larger a^* value corresponds to more redness); and b^* refers to the change from blueness to yellowness (a smaller b^* value corresponds to more blueness, and a larger b^* value corresponds to more yellowness). The *CIE L*a*b** colour model is device independent, with the potential to provide consistent colour data for most types of input or output devices such as digital camera, scanner, and monitor (Tkalcić and Tasić 2003). Furthermore, the *CIE L*a*b** colour model divides the colour into two functions (brightness and chromaticity), enabling researchers to change colour easily by adjusting *L*a*b** value in colour control experiments (Wulf and Wise 1999). To our knowledge, few studies have used digital photography and spectrometry data to calculate *CIE L*a*b** values and to compare the outcomes (Ohta and Robertson 2006; Wee et al. 2007).

Here we compare the performance and reliability of spectrometry and digital photography for quantifying the colourations of a standardized colour reference chart. Colour reference patches (e.g. PANTONE colour chart) are widely used for colour measurements and comparisons because they are simple, fast, and widely assumed to be accurate and are universally applied in textile and printing industries (Vik 2017). Although there may be a close match between the targeted colour and a specific reference colour based on the colour matching system, assessment without a quantitative variable for colour measuring should be difficult for rigorously testing hypotheses on the perception of colour in trichromatic species. Our objectives for this study were three-fold: (1) to produce colour measurements using photography and spectrometry and provide details on the transferring process among colour models, (2) to test and compare the repeatability and agreement of colour measurements obtained using both methods, and (3) to give criteria for identifying when the use of photography and spectrometry would be interchangeable. In order to achieve these objectives, we assess the suitability of using camera-derived *CIE L*a*b** pixel values as a proxy for spectrophotometer derived *CIE L*a*b** values, together with the performance of the function of a specific program to translate spectral reflectance (*SR*) of different colour and grey scales into *XYZ* values.

Material and methods

Materials

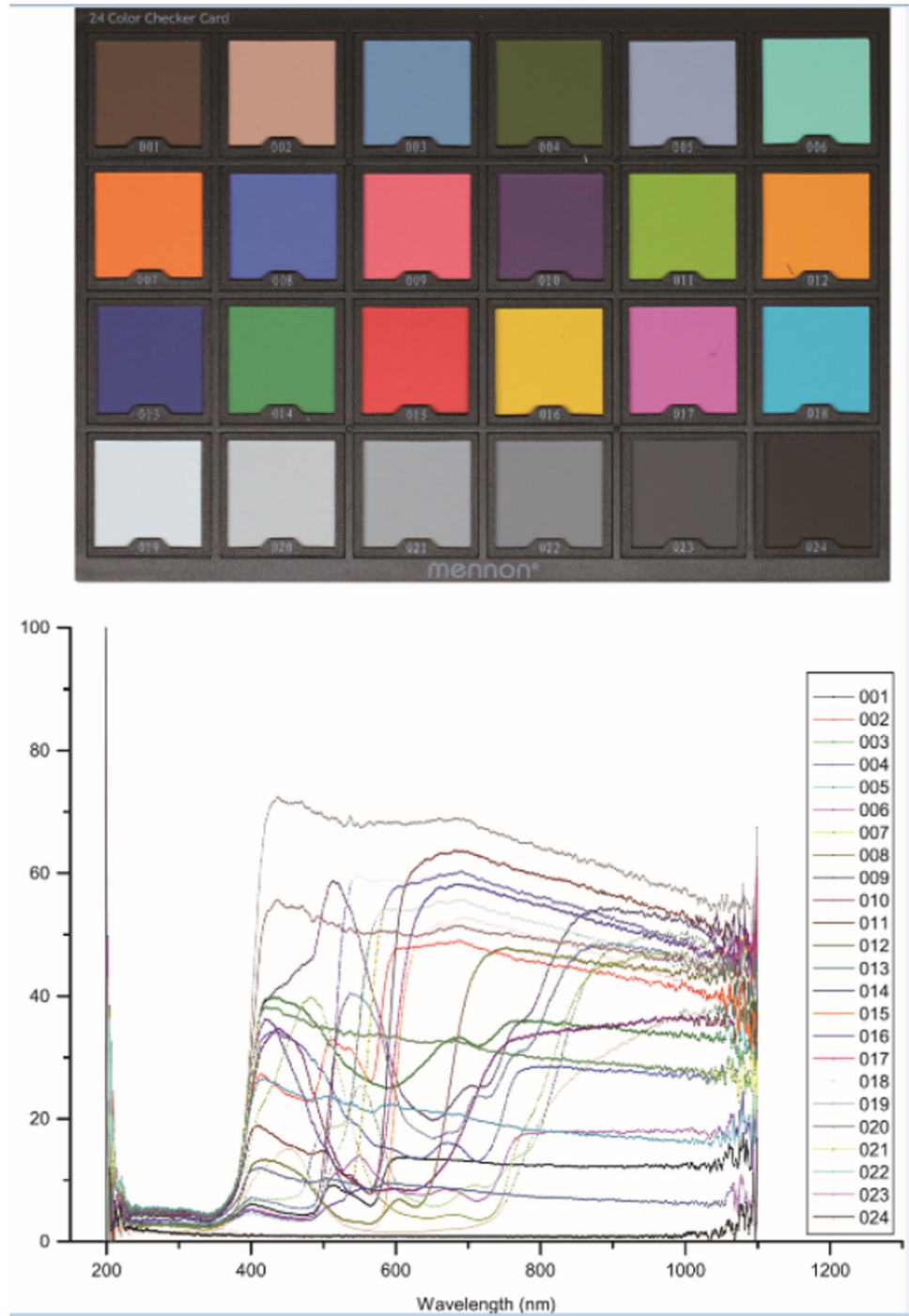
We selected the MENNON' Test Chart with 24-colour test targets, which comprises 6 greyscale tones and 18 coloured patches. The MENNON' Test Chart is a colour test and calibration tool widely used by photographers for colour recognition and evaluation of colour deviations. Among the 18 colour patches, six of them represent natural objects of special interest, such as skin tones, leaf green, and blue sky, six patches represent additive (red-green-blue) and subtractive (cyan-magenta-yellow) primary colours, and six represent a combination of two primary colours (Fig. 1). To minimize observer bias, blinded methods were used when all data were measured and/or analysed.

Experimental methods

Method 1: Photography

Colour of the sample depends on the part of spectrum reflected from it, so the spectral power distribution of the illumination must be standardized before capturing colour images (Mollon 1999). We selected the D50 (5400 K) for the standard illuminant, which is most similar to the natural daylight. Since the diffuse reflection responsible for the colour occurs at 45° from the incident light, we also placed the angle between the lighting source camera lens at approximately 45° (Stevens et al. 2007). To achieve a uniform light intensity over the whole sample, we set the distance of 30 cm between the light source and the sample in a dark room. We used a Canon D70 camera using a tripod with a 100-mm prime lens by autofocus control and placed vertically at a distance of 22.5 cm from the samples to lens. Flash was off to reduce light interference. To calibrate the digital colour system, 24-colour patches were twice captured together in one picture (see Fig. 1). After capturing the image, we divided the picture into 24 different patches by using the rectangle tool in Software ImageJ 1.53e (64-bit Java 1.8.0_172). Additionally, a 500 × 500 pixel size was randomly selected for each colour chart setting in *ROI* manager and subsequently saved as the testing picture for extracting *RGB* values in tiff format. We then extracted all *RGB* values from the 500 × 500 pixel for each colour patches and calculated mean values of *RGB* using Software MATLAB version 2016a (The Mathworks Inc.) (see Code 1 in supplementary materials). Formulas for calculating the mean *RGB* values (i.e. *r*, *g*, and *b*) for testing picture are shown by the following:

Fig. 1 The 24-colour patches used for colour measuring and spectral reflectance of each colour patches using the Ava Spec-2048 spectrometer. Wavelength ranged from 199 to 1099 nm



$$r = \frac{R_1 + R_2 + \dots + R_n}{n}$$

$$g = \frac{G_1 + G_2 + \dots + G_n}{n}$$

$$b = \frac{B_1 + B_2 + \dots + B_n}{n}$$

where n represents the number of pixels in testing pictures; R_i , G_i , and B_i ($i = 1, 2, \dots, n$) denote the RGB

values for the i th pixel.

Since RGB space could not be transferred to the $L^*a^*b^*$ model directly, we applied the direct model in order to transform RGB to $L^*a^*b^*$ with reference white point, D50 (León et al. 2006; Hunt and Pointer 2011; Westland et al. 2012). The same MENNON' Test Chart was photographed twice, and each digital photo was measured following the previous

procedure. All conversions and extraction data from digital image were carried out in the Software Excel version 2010 (Microsoft Corporation).

Step 1: Standardized *RGB* value based on Gamma function:

$$\text{Step 1} \begin{cases} R = \text{gamma}\left(\frac{r}{255}\right) \\ G = \text{gamma}\left(\frac{g}{255}\right) \\ B = \text{gamma}\left(\frac{b}{255}\right) \end{cases} \text{gamma}(x) = \begin{cases} \left(\frac{x+0.055}{1.055}\right)^{2.4} & (x > 0.0405) \\ \left(\frac{x}{12.92}\right) & (\text{if } x \leq 0.0405) \end{cases}$$

Step 2: Transferred *RGB* value of each image to *xyz* colour space using the following matrix:

$$\text{Step 2} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.436052025 + 0.385081593 + 0.143087414 \\ 0.222491598 + 0.71688606 + 0.06062148 \\ 0.013929122 + 0.097097002 + 0.71418547 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Step 3: Standardized *XYZ* value using three coefficients ($x_n = 96.4211$; $y_n = 100$; $z_n = 82.5211$):

$$\text{Step 3} \begin{cases} X = \frac{x}{x_n} = \frac{x}{96.4221} \\ Y = \frac{y}{y_n} = \frac{y}{100} \\ Z = \frac{z}{z_n} = \frac{z}{82.5211} \end{cases}$$

Step 4: Converted *XYZ* value by function $f(t)$:

$$\text{Step 4} \quad f(t) = \begin{cases} t^{1/3} & (t > 0.008856) \\ 7.787 * t + 0.13791 & (\text{if } t \leq 0.008856) \end{cases}$$

Step 5: Converted *XYZ* to $L^*a^*b^*$ by following formulas:

$$\text{Step 5} \begin{cases} L^* = \begin{cases} 116 * Y^{1/3} & (Y > 0.008856) \\ 903.3 * Y & (\text{if } Y \leq 0.008856) \end{cases} \\ a^* = 500 * [f(X) - f(Y)] \\ b^* = 200 * [f(Y) - f(Z)] \end{cases}$$

Method 2: Spectrometry

We also measured the colour patches using a *CIELAB* colourimeter, equipped with Ava Spec-2048 spectrometer and Software AvaSpec 7.6 (The Avantes Inc.). The spectrometer was standardized with an Argon-Mercury lamp (200–1200 nm). Quantification of colour was achieved by measuring the optical density across the range of sampled wavelengths. We first used a standard diffused whiteboard to measure and save the present spectrum in a light and dark environment as a reference. The integration time for contact

sample measurement was 24.5 ms, and the integrated area amounted to 3 mm². To reduce error in each test, measuring geometry (45°), angle of view (2°), and Illuminant (D50) was set to standardize the test condition. We first captured the reflectance (%) for each colour chart ranging from 199 to 1099 nm (Fig. 1) and then converted the reflectance value to *XYZ* colour space (Steps 1 and 2) based on the method written by Zheng and Zhou (2013). We finally transferred *XYZ* to $L^*a^*b^*$ (Steps 3, 4, and 5). The same MENNON® Test Chart was scanned twice in the spectrometer, and the same procedure was used for calculating $L^*a^*b^*$ values. Demographics, detail observation value, and conversion functions using the two methods are shown in Tables S1 and S2.

Step 1: Calculated *K* value for D50 Illuminant when *Y* value is 100 based on *XYZ* CIE1931, where $S(\lambda)$ is D50 relative spectral distribution, $R(\lambda)$ spectral reflectance for D50, and \bar{y} standard spectral tristimulus values for D50.

$$\text{Step 1} \quad K = \frac{100}{\sum S(\lambda) * \bar{y}(\lambda) * \Delta\lambda}$$

Step 2: Transferred spectral reflectance (*SR*) of each image to *XYZ* colour space, where $S(\lambda)$: D50 relative spectral distribution $R(\lambda)$ is spectral reflectance for samples, and \bar{x} , \bar{y} , \bar{z} are standard spectral tristimulus values for D50.

$$\text{Step 2} \quad \begin{cases} X = K \int_{\lambda} S(\lambda) R(\lambda) \bar{x}(\lambda) d\lambda \\ Y = K \int_{\lambda} S(\lambda) R(\lambda) \bar{y}(\lambda) d\lambda \\ Z = K \int_{\lambda} S(\lambda) R(\lambda) \bar{z}(\lambda) d\lambda \end{cases} \Rightarrow \begin{cases} X = K \sum S(\lambda) R(\lambda) \bar{x}(\lambda) \Delta\lambda \\ Y = K \sum S(\lambda) R(\lambda) \bar{y}(\lambda) \Delta\lambda \\ Z = K \sum S(\lambda) R(\lambda) \bar{z}(\lambda) \Delta\lambda \end{cases}$$

Steps 3, 4, and 5: Repeated same steps in photography to finish the conversion from *XYZ* to $L^*a^*b^*$.

Statistical analysis

In this study, all $L^*a^*b^*$ values represented continuous variables, so we selected a consensus of four statistical methods to estimate repeatability and agreement of the colour measurement results: correlation analysis, intraclass correlation coefficient (*ICC*), concordance correlation coefficients (*CCC*), and Bland-Altman scatter plots. Here, we first conducted linear regressions to assess the relationships between the two trials of either photography or spectrometry and between the two methods. Furthermore, to examine the repeatability between the two trials for either photography or spectrometry, we then implemented the *ICC* using a two-way random linear regression model with 95% confidence intervals (95% CI). To examine the agreement between the two methods, we firstly calculated the mean values of the two trials for each method and then derived the *ICC* using a two-way linear model. For both models, *ICC* can vary between 0 and 1.0 theoretically,

where an *ICC* of 0 indicates no reliability and an *ICC* value of 1.0 indicates perfect reliability (Weir 2005). In practice, an *ICC* > 0.75 would suggest that both photography and spectrometry are reliable methods for colour measurement, but an *ICC* < 0.4 would suggest poor reliability (Koo and Li 2016). Since regression and paired *t*-tests may conflate between-sample variability, we also calculated *CCC* to estimate the reproducibility between two colour measuring methods (Altman and Bland 1983; Lawrence 1989). The *CCC* is a widely accepted criterion of score consistency in biology and medicine (Barchard 2012). When the value of *CCC* is closer to 1, the two measurements show a better agreement.

The Bland-Altman analysis (Bland and Altman 1986; Bland and Altman 2007) is a popular and widespread method used in medical sciences for analysing the level of agreement between different methods or medical instruments (Braždžionytė and Macas 2007; Carkeet 2015; Giavarina 2015; Gerke 2020). This analysis is a graphical method based on visualization of differences in the measurements from the different methods, using a scatter plot to represent the difference against the mean of the measurements. Specifically, the limits of agreement (*LoA*) would be drawn as the mean difference plus and minus 1.96 times the standard deviation (SD) of the differences (Bland and Altman 1986; Bland and Altman 1999; Labiris et al. 2009; Donekal et al. 2013). According to a priori biologically and analytically relevant criteria, people can see if these limits (i.e. *LoA*) are exceeded or not, which determines whether the two methods can be used interchangeably (Bland and Altman 1986). In this study, we applied the Bland-Altman analysis to plot the differences of the $L^*a^*b^*$ values of each standard colour patch between the two methods (i.e. spectrometry and digital photography). As 95% of the differences are expected to fall within the *LoA*, we therefore suggested that people can use these limits (i.e. *LoA*) as pre-specified criteria for identifying whether the photography and spectrometry can be used interchangeably in other studies under different environments. Moreover, we used the Bland-Altman analysis to assess the repeatability between the two trials for spectrometry and digital photography. The coefficient of repeatability (*CR*) was calculated as 1.96 times the standard deviation (SD) of the differences between the two measurements (d_2 and d_1) (Bland and Altman 1986):

$$CR = 1.96 \times \sqrt{\frac{\sum(d_2 - d_1)^2}{n}}$$

Finally, we calculate the colour difference (ΔE^*) between the two measuring methods for each colour patch using the following formula:

$\Delta E^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$, where ΔL^* , Δa^* , and Δb^* represent the differences in colour coordinates between the two methods.

All results are shown as mean \pm SE, and all tests were two-tailed. All statistical analyses, including linear regressions (ordinary least squares (OLS)), *ICC*, *CCC*, and Bland-Altman analysis, were performed in MedCalc v.18.2.1 (MedCalc Software Ltd.).

Results

Repeatability analysis between repeated trials of photography and spectrometry

Regression analysis revealed significant relationships between the first and the second replicated trials using photography (L^* : $r = 0.997$, $P < 0.01$; a^* : $r = 0.995$, $P < 0.01$; b^* : $r = 0.997$, $P < 0.01$, $n = 24$). Values of L^* , a^* , b^* also indicated significant relationships between the two spectrometry tests (L^* : $r = 0.999$, $P < 0.001$; a^* : $r = 0.995$, $P < 0.001$; b^* : $r = 0.999$, $P < 0.001$, $n = 24$). Intraclass correlation coefficients also revealed quite high repeatability between the two photography (L^* , a^* , b^* : $ICC = 0.999$) and spectrometry trials (L^* , a^* , b^* : $ICC = 0.999$). Results of Bland-Altman analysis for L^* , a^* , b^* values between the two trials for photography and spectrometry are shown in Table 1. The mean differences of $L^*a^*b^*$ values of photography (-0.769 , 0.208 , -0.065) were higher than those from spectrometry (-0.012 , 0.025 , 0.029) (Fig. 2; Table 1). Coefficients of repeatability (*CR*) for $L^*a^*b^*$ values were also lower in spectrometry images than those from photographic images (Table 1). Nonetheless, the absolute values of the mean difference for all variables were < 1 (much smaller than the mean values of L^* , a^* , b^*), and the coefficients of repeatability were < 2 (Table 1). Collectively, these data suggest the $L^*a^*b^*$ values generated by photography and spectrometry were both highly repeatable.

Agreement analysis between photography and spectrometry

The regressions of $L^*a^*b^*$ values measured by photography and spectrometry showed significant relationships (L^* : $r = 0.960$, $P < 0.01$; a^* : $r = 0.992$, $P < 0.01$; b^* : $r = 0.974$, $P < 0.01$, $n = 24$; see Table 2). The *ICC* analysis for L^* , a^* , b^* values also showed good agreement in quantifying colour between the two methods (L^* : $ICC = 0.979$; a^* : $ICC = 0.995$; b^* : $ICC = 0.987$). Moreover, the concordance correlation coefficients showed a high degree of concordance between the two methods (L^* : $CCC = 0.860$; a^* : $CCC = 0.987$; b^* : $CCC = 0.972$) (Table 2). Therefore, $L^*a^*b^*$ values generated by photography and spectrometry should be in high agreement under the standard conditions set in this study. The results of

Table 1 Bland-Altman analysis for measuring colour between two repeated trials for quantifying colour for both photography and spectrometry

Methods	Variables	Mean difference	95% CI of mean difference	LoA (mean difference± 1.96 SD)	Coefficient of repeatability
Photography	L^*	-0.769	-0.925 to -0.613	(-1.492, -0.046)	1.665
	a^*	0.208	-0.053 to 0.468	(-1.002, 1.417)	1.252
	b^*	-0.065	-0.325 to 0.195	(-1.271, 1.140)	1.187
Spectrometry	L^*	-0.012	-0.045 to 0.020	(-0.163, 0.138)	0.150
	a^*	0.025	-0.028 to 0.078	(-0.221, 0.272)	0.246
	b^*	0.029	-0.001 to 0.059	(-0.110, 0.168)	0.147

Notes: *CI* confidence interval, *LoA* limits of agreement, defined as the mean difference ± 1.96 SD of the differences

the Bland-Altman analysis for $L^*a^*b^*$ values between the two methods (see Table 2) revealed that the mean differences (i.e. 8.785, 1.980, and - 2.173, respectively) and the coefficients of repeatability for $L^*a^*b^*$ values (i.e. 20.060, 8.113, and 16.277, respectively) between the two methods were obviously higher than those between the two trials (see Table 1). The *LoA* for $L^*a^*b^*$ values between the two methods ranged from -1.727 to 19.297 (L^*), from -9.258 to 5.299 (a^*), and from -18.221 to 13.874 (b^*), respectively (Fig. 3; Table 2). The colour differences between the two methods (i.e. ΔE^*) ranged from 2.973 to 38.086 for different patches (see Table S3).

Discussion

Here we developed frameworks for quantifying the colour perception in humans and other trichromatic species with digital photography and spectrometry data. We note that the colour differences (ΔE^*) between the two measuring methods (i.e. photography and spectrometry) may be relatively high under the human’s perceptibility (e.g. Douglas et al. 2007; Buchelt and Wagenführ 2012). However, using a consensus of three statistical methods (i.e. correlation analysis, *ICC*, and *CCC*), we demonstrated high reliability on $L^*a^*b^*$ values both within and between the two measuring methods. We therefore

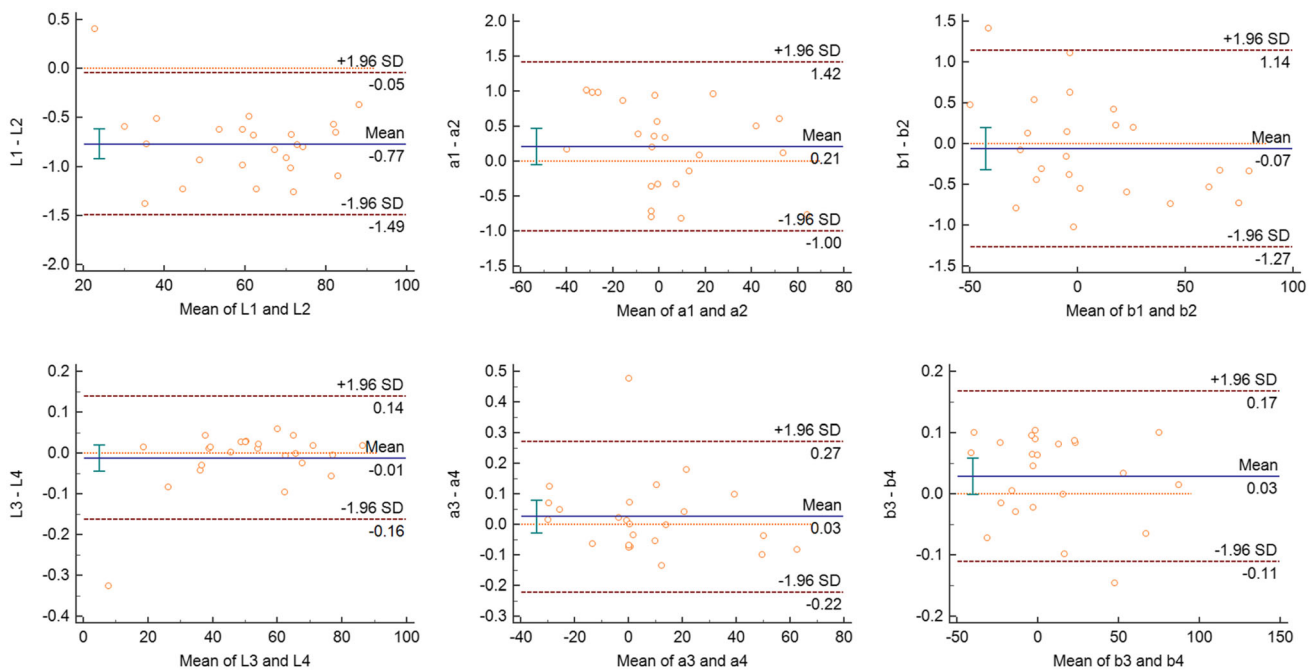


Fig. 2 Bland-Altman plot showing the difference between two trials for photography and spectrometry mean values of $L^*a^*b^*$ in the 24-colour patches. Photography trial 1: L^*_1, a^*_1, b^*_1 and photography trial 2: L^*_2, a^*_2, b^*_2 ; Spectrometry trial 1: L^*_3, a^*_3, b^*_3 and spectrometry trial 2: L^*_4, a^*_4, b^*_4 . Orange-dotted line represents the line of equality (difference =

0) for detecting a systematic difference. Blue solid line represents the mean difference between two trials. Error bars for mean difference are ± SE. Purple dotted line represents the limits of agreement (*LoA*: mean difference ± 1.96 SD)

Table 2 Consensus of three statistical analyses used to estimate the agreement for measuring colour in humans using photography and spectrometry

Variables	Correlation coefficient		Intraclass correlation coefficient (<i>ICC</i>)		Concordance correlation coefficients (<i>CCC</i>)		Bland-Altman analysis		Coefficient of repeatability	
	<i>r</i>	<i>p</i>	<i>ICC</i>	95% CI (for <i>ICC</i>)	<i>CCC</i>	95% CI for <i>CCC</i>	Mean difference	95% CI for the mean difference <i>LoA</i> (mean difference ± 1.96 SD)		
<i>L</i> *	0.960	< 0.01	0.979	(0.951, 0.991)	0.860	(0.744, 0.926)	8.785	(6.520, 11.050)	(-1.727, 19.297)	20.060
<i>a</i> *	0.992	< 0.01	0.995	(0.988, 0.998)	0.987	(0.973, 0.994)	-1.980	(-3.548, -0.412)	(-9.258, 5.299)	8.113
<i>b</i> *	0.974	< 0.01	0.987	(0.969, 0.994)	0.972	(0.936, 0.988)	-2.173	(-5.631, 1.284)	(-18.221, 13.874)	16.277

Notes: *ICC* intraclass correlation coefficient, *CCC* concordance correlation coefficients, *CI* confidence interval, *LoA* limits of agreement, defined as the mean difference ± 1.96 SD of the differences

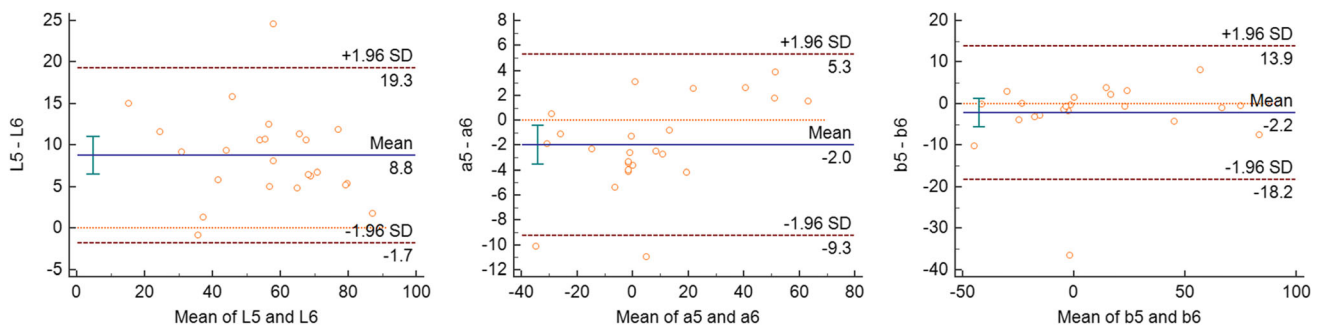


Fig. 3 Bland-Altman plot showing the difference between digital photography and spectrometry mean values of $L^*a^*b^*$ in the 24-colour patches. Method 1 (photography): L^*_5, a^*_5, b^*_5 and method 2 (spectrometry): L^*_6, a^*_6, b^*_6 . Orange-dotted line represents the line of equality (difference = 0) for detecting a systematic difference. Blue solid line represents the mean difference between two methods. Green bar represents 95% CI of mean difference. Purple-dotted lines represent the limits of agreement (LoA: mean difference \pm 1.96 SD)

highlight that photography can be used as a suitable alternative to spectrometry for quantifying colour measurements in trichromatic species. Other authors have also concluded that colour measurement data collected from digital photographs correlates with results obtained from spectrometry through correlation analysis (Simons et al. 2012; Fairhurst et al. 2014).

Generally, the mean $L^*a^*b^*$ differences between spectrometer trials were lower than those from digital photography (Fig. 2, Table 1), demonstrating smaller random errors from the spectrometer trials than those from digital photography trials. This may be related to the capacity to create more stable experimental testing conditions for the spectrometer (e.g. stable wavelength of light source, light intensity, fixed angle and distance from the specimen, test duration, background). Even though, our results show that colour measurements from digital photography can also be highly repetitive (Fig. 2, Table 1)—a key requirement for any robust colour quantification method (Bergeron and Fuller 2018). Furthermore, as digital images could be divided into numerous independent pixels equally (with the number of pixels based on the resolution of the imaging device) and $L^*a^*b^*$ values tested and evaluated for each pixel, this would help resolve issues around irregular spatial structures, surfaces, and colour heterogeneity. However, the results of colour measurement from photography also strongly depend on the camera set up and other interference light sources (Montgomerie 2006; Eeva et al. 2008). Therefore, $L^*a^*b^*$ values obtained from digital images may vary with experimental conditions, and greater efforts to standardize parameters are necessary to control for these factors (Giraudeau et al. 2015; Troscianko and Stevens 2015; Peneaux et al. 2021).

Colourful visual displays of animals are ubiquitous and remarkably diverse in nature, serving as vital communication signals (Osorio and Vorobyev 2008; Caro et al. 2017). Unsurprisingly, measuring colour perception has

therefore become a model system for understanding evolutionary processes and testing predictions in behavioural ecology studies (Endler and Mappes 2017). The ability to obtain accurate and comparable colour measurements for assessing the functions of different colours is therefore critical, particularly in light of the rapid development of new technologies now available for colour measurement. Here, we have compared two commonly used but yet-to-be-compared techniques and measure a wide range of colours from a textile colour chart, since these are ideal for calibration and for standardizing light conditions on variable surfaces when comparing methods (e.g. Potash et al. 2020). We acknowledge that the actual textures and colours of these charts are far more homogeneous than those found in nature, and such natural variations could affect differences in measurements using spectrometers and cameras. Our study offers a logical and whole sequential workflow to better inform such studies of trichromatic vision, since dealing with the biological variation that would otherwise affect how both devices would perform (as well as controlling for ambient conditions) requires that both methods be repeatable and interchangeable within any given range of colours. Our workflow shows that calibrating digital imagery is essential, since this has the potential to provide thousands of colour measurements and could capture the entire range of chromatic gradients (van den Berg et al. 2020). Concurrently, our workflow demonstrates the importance of conducting a series of preliminary experiments under different conditions to estimate the corresponding differences with the relatively stable spectrometry results. Quantifying the limits of agreement between the two methods, e.g. using Bland-Altman analysis, under standardized conditions, can be used as a threshold for future trichromatic studies, such that if the differences of $L^*a^*b^*$ values between the two methods calculated in pre-experiments are located within

these limits, they can be considered in good agreement and may be used interchangeably in subsequent experiments under the same conditions prescribed in the preliminary experiments.

Conclusions

Spectrometry and digital photography are both widely used for colour measurement, but spectrometry is generally more portable, precise, has less risk from disturbance, and has a greater spectrum range than photography. Based on our reliability analysis (correlate, *ICC*, and *CCC*), we confirm that photography is also highly repetitive under the standardized measurement and conversion procedure. We therefore suggest that digital photography can be used as a suitable alternative method for non-contact and continuous colour measurement technique over a long-time scale, due to its simplicity, versatility, and lower cost. In this study, we also provide a standard procedure for comparing colours between different measuring methods. We recommend researchers conduct pre-experiments to find appropriate conditions that can ensure an acceptable error range (e.g. using the Bland-Altman method) in colour measuring before conducting their primary experiments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00265-021-03071-8>.

Acknowledgements We would like to extend our sincere thanks to Yupei Yan for their help on the sample collection. We thank Dr. Guangzhan Fang for his support on data analysis. We thank Dr. Raul E. Diaz for his help in improving the English writing of this manuscript. We also sincerely thank two anonymous referees for reviewing the manuscript and their helpful comments.

Author contribution CY and NL conceived the ideas and designed methodology; CY and JW collected the data; CY and JW analysed the data; and CY, JW, NL, and HL led the writing of the manuscript. All authors contributed to drafts and gave final approval for publication.

Funding This research was supported by the National Natural Science Foundation of China (Grant Nos. 31500316 to CY and 32070448 to NL). The funders had no role in the design of the study, the collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials The datasets generated and/or analysed during the current study are available in the *Baidu Netdisk* repository [<https://pan.baidu.com/s/14udE-1K0odDSiAYE0wqMMg>] (Password: 8nsm).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Akkaynak D, Allen JJ, Mähger LM, Chiao C-C, Hanlon RT (2013) Quantification of cuttlefish (*Sepia officinalis*) camouflage: a study of color and luminance using in situ spectrometry. *J Comp Physiol A* 199:211–225
- Barchard K A (2012) Examining the reliability of interval level data using root mean square differences and concordance correlation coefficients. *Psychol Methods* 17(2):294–308. <https://doi.org/10.1037/a0023351>
- Bergeron ZT, Fuller RC (2018) Using human vision to detect variation in avian coloration: how bad is it? *Am Nat* 191:269–276
- Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *J Roy Stat Soc D-Sta* 32:307–317
- Bland JM, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327:307–310
- Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160
- Bland JM, Altman DG (2007) Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 17: 571–582
- Braždžionytė J, Macas A (2007) Bland–Altman analysis as an alternative approach for statistical evaluation of agreement between two methods for measuring hemodynamics during acute myocardial infarction. *Medicina* 43:208
- Buchelt B, Wagenführ A (2012) Evaluation of colour differences on wood surfaces. *Eur J Wood Prod* 70:389–391
- Bybee SM, Yuan F, Ramstetter MD, Llorente-Bousquets J, Reed RD, Osorio D, Briscoe AD (2012) UV photoreceptors and UV-yellow wing pigments in heliconius butterflies allow a color signal to serve both mimicry and intraspecific communication. *Am Nat* 179:38–51
- Cageao RP, Blavier J-F, McGuire JP, Jiang Y, Nemtchinov V, Mills FP, Sander SP (2001) High-resolution Fourier-transform ultraviolet-visible spectrometer for the measurement of atmospheric trace species: application to OH. *Appl Opt* 40:2024–2030
- Carkeet A (2015) Exact parametric confidence intervals for Bland-Altman limits of agreement. *Optom Vis Sci* 92:e71–e80
- Caro T, Stoddard MC, Stuart-Fox D (2017) Animal coloration research: why it matters. *Philos Trans R Soc B* 372:20160333
- Donekal S, Ambale-Venkatesh B, Berkowitz S, Wu CO, Choi EY, Fernandes V, Yan R, Harouni AA, Bluemke DA, Lima JAC (2013) Inter-study reproducibility of cardiovascular magnetic resonance tagging. *J Cardiovasc Pharmacol* 15:37
- Douglas RD, Steinhauer TJ, Wee AG (2007) Intraoral determination of the tolerance of dentists for perceptibility and acceptability of shade mismatch. *J Prosthet Dent* 9:200–208
- Eeva T, Sillanpää S, Salminen J-P, Nikkinen L, Tuominen A, Toivonen E, Pihlaja K, Lehikoinen E (2008) Environmental pollution affects the plumage color of great tit nestlings through carotenoid availability. *EcoHealth* 5:328–337
- Ender JA, Mielke PW Jr (2005) Comparing entire colour patterns as birds see them. *Biol J Linn Soc* 86:405–431
- Ender JA, Mappes J (2017) The current and future state of animal coloration research. *Philos Trans R Soc B* 372:20160352
- Fairchild MD (2013) Color appearance models. John Wiley & Sons, Hoboken
- Fairhurst GD, Dawson RD, van Oort H, Bortolotti GR (2014) Synchronizing feather-based measures of corticosterone and

- carotenoid-dependent signals: what relationships do we expect? *Oecologia* 174:689–698
- Gerke O (2020) Reporting standards for a Bland–Altman agreement analysis: a review of methodological reviews. *Diagnostics* 10:334
- Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med* 25:141–151
- Girardeau M, Chavez A, Toomey MB, McGraw KJ (2015) Effects of carotenoid supplementation and oxidative challenges on physiological parameters and carotenoid-based coloration in an urbanization context. *Behav Ecol Sociobiol* 69:957–970
- Hunt RWG, Pointer MR (2011) *Measuring colour*, 4th edn. Wiley, New York
- Johnsen S (2016) How to measure color using spectrometers and calibrated photographs. *J Exp Biol* 219:772–778
- Kawchuk G, Herzog W (1996) A new technique of tissue stiffness (compliance) assessment: its reliability, accuracy and comparison with an existing method. *J Manip Physiol Ther* 19:13–18
- Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chromatogr Sci* 15: 155–163
- Labiris G, Gkika M, Katsanos A, Fanariotis M, Alvanos E, Kozobolis V (2009) Anterior chamber volume measurements with Visante optical coherence tomography and Pentacam: repeatability and level of agreement. *Clin Exp Ophthalmol* 37:772–774
- Langkilde T, Boronow KE (2012) Hot boys are blue: temperature-dependent color change in male eastern fence lizards. *J Herpetol* 46:461–465
- León K, Mery D, Pedreschi F, León J (2006) Color measurement in L*a*b* units from RGB digital images. *Food Res Int* 39:1084–1091
- Li G, Song G, Yao Z, Jian H (2008) Fiber optical spectrometer and its applications in on-line color measurement. *Guangdong Chem Ind* 35:112–116
- Lawrence IKL (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268
- Ligon RA, McGraw KJ (2018) A chorus of color: hierarchical and graded information content of rapid color change signals in chameleons. *Behav Ecol* 29:1075–1087
- Mani SA, Naing L, John J, Samsudin AR (2008) Comparison of two methods of dental age estimation in 7-15-year-old Malays. *Int J Parallel Prog* 18:380–388
- McKay BD (2013) The use of digital photography in systematics. *Biol J Linn Soc* 110:1–13
- Mollon JD (1999) Specifying, generating and measuring colours. In: Robson J, Carpenter R (eds) *A practical guide to visual research*. Oxford University Press, Oxford, pp 106–128
- Montgomerie R (2006) Analysing colors. In: Hill GE, McGraw KJ (eds) *Bird coloration*. Harvard University Press, Boston, pp 90–147
- Ohta N, Robertson A (2006) *Colorimetry: fundamentals and applications*. John Wiley & Sons, Chichester
- Osorio D, Vorobyev M (2008) A review of the evolution of animal colour vision and visual communication signals. *Vis Res* 48:2042–2051
- Peneaux C, Hansbro PM, Griffin AS (2021) The potential utility of carotenoid-based coloration as a biomonitor of environmental change. *Ibis* 163:20–37
- Pike TW (2011) Using digital cameras to investigate animal colouration: estimating sensor sensitivity functions. *Behav Ecol Sociobiol* 65: 849–858
- Potash AD, Greene DU, Foursa GA, Mathis VL, Conner LM, McCleery RA (2020) A comparison of animal color measurements using a commercially available digital color sensor and photograph analysis. *Curr Zool* 66:601–606
- Schanda J (2007) *Colorimetry: understanding the CIE system*. John Wiley & Sons, Hoboken
- Simons MJP, Briga M, Koetsier E, Folkertsma R, Wubs MD, Dijkstra C, Verhulst S (2012) Bill redness is positively associated with reproduction and survival in male and female zebra finches. *PLoS One* 7: e40721
- Smith KR, Cadena V, Endler JA, Porter WP, Kearney MR, Stuart-Fox D (2016) Colour change on different body regions provides thermal and signalling advantages in bearded dragon lizards. *Proc R Soc B* 283:20160626
- Stevens M, Cuthill IC (2005) The unsuitability of html-based colour charts for estimating animal colours – a comment on Berggren and Merilä (2004). *Front Zool* 2:14
- Stevens M, Párraga CA, Cuthill IC, Partridge JC, Troscianko TS (2007) Using digital photography to study animal coloration. *Biol J Linn Soc* 90:211–237
- Tkalcic M, Tasic JF (2003) *Colour spaces: perceptual, historical and applicational background*. Institute of Electrical and Electronics Engineers. Ljubljana, Slovenia
- Troscianko J, Stevens M (2015) Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. *Methods Ecol Evol* 6:1320–1331
- van den Berg CP, Troscianko J, Endler JA, Marshall NJ, Cheney KL (2020) Quantitative colour pattern analysis (QCPA): a comprehensive framework for the analysis of colour patterns in nature. *Methods Ecol Evol* 11:316–332
- Vik M (2017) *Colorimetry in textile industry*. VÚTS, a.s., Svárovska
- Villafuerte R, Negro JJ (1998) Digital imaging for colour measurement in ecological research. *Ecol Lett* 1:151–154
- Wee AG, Lindsey DT, Kuo S, Johnston WM (2007) Color accuracy of commercial digital cameras for use in dentistry. *J Prosthet Dent* 97: 178
- Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19:231–240
- Westland S, Ripamonti C, Cheung V (2012) *Computational colour science using MATLAB*. John Wiley & Sons, Chichester
- Wulf DM, Wise JW (1999) Measuring muscle color on beef carcasses using the L*a*b* color space. *J Anim Sci* 77:2418–2427
- Yam KL, Papadakis SE (2004) A simple digital imaging method for measuring and analyzing color of food surfaces. *J Food Eng* 61: 137–142
- Zheng YL, Zhou SS (2013) *Printing chromatics (Third Edition)*. Cultural Development Press, Beijing

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.