Received: 15 May 2017 | Revised: 19 January 2018 | Accepted: 23 February 2018

DOI: 10.1002/col.22230

RESEARCH ARTICLE

WILEY

An investigation into the variability of skin colour measurements

Mengmeng Wang[1] | Kaida Xiao[1] | Ming Ronnier Luo[1,2] | Michael Pointer[1] | Vien Cheung[1] | Sophie Wuerger[3]

[1] School of Design, University of Leeds, Leeds, UK

[2] Key State Laboratory of Modern Optical Instruments, Department of Optical Engineering, Zhejiang University, Hangzhou, China

[3] Department of Psychological Sciences, University of Liverpool, Liverpool, UK

Correspondence
Kaida Xiao, School of Design, University of Leeds, Leeds, UK.
Email: k.xiao1@leeds.ac.uk

Funding information
Royal Academy of Engineering, Grant/Award Number: RECI033, Engineering and Physical Sciences Research Council (EPSRC), Grant/Award Number: EP/K040057 and EP/L001012

Abstract

This study aimed to investigate the variability of skin colour measurements for two kinds of extensively used instruments, telespectroradiometers (TSR) and spectrophotometers. A Konica Minolta CM700d spectrophotometer and a PhotoResearch PR650 telespectroradiometer were used to measure the forehead and the cheekbone of 11 subjects. The variability was evaluated using different measurement parameters including measurement aperture size and pressure on the facial locations for the spectrophotometer, and measurement distance for the telespectroradiometer. The mean colour difference from the mean was used to define the short-term repeatability; the CIELAB colour difference and colour appearance changes in each perceptual CIELAB attribute between each of two instrument settings were used to evaluate the inter-instrument agreement. The results show that, for the TSR, different measurement distances have identical repeatability but the colour shifts were significant; for the spectrophotometer, the large aperture size of the target masks gave the most repeatable results and the aperture size had more influence on the colour shifts than the measurement pressure. In addition, to investigate the effect of ethnicity and body location on measurement variability, skin colours from additional 151 subjects were measured. The differences between the measurements for different body locations were, in general, larger than the instrument repeatability and the inter-instrument agreement.

KEYWORDS

skin colour, skin colour measurement, skin colour variation

1 | INTRODUCTION

Human skin colour ranges from the darkest brown to the lightest hues. An individual's skin pigmentation is the result of genetics, being the product of the genetic makeup of both of the biological parents of each individual. In evolution, skin pigmentation in human beings evolved by a process of natural selection primarily to regulate the amount of ultraviolet radiation penetrating the skin, controlling its biochemical effects.[1]

The skin colour of darker-skinned people is primarily determined by the pigment melanin, which is produced in cells within the skin. Light skin, on the other hand, is a result of the bluish-white connective tissue under the dermis, the inner of the two layers that make up the skin, and by the haemoglobin, the protein molecule in the red blood cells that carries oxygen circulating in the veins of the dermis. The red colour underlying the skin becomes more visible, especially in the face, when, as a consequence of physical exercise or the stimulation of the nervous system (anger, fear, embarrassment), arterioles dilate. The colour is not entirely uniform across an individual's skin; for example, the skin of the palm of the hand and the sole of the foot is lighter than most other skin, and this is especially noticeable in darker-skinned people.[2]

There is a strong association between the geographic distribution of ultraviolet radiation (UV) and the distribution of indigenous skin pigmentation around the world. Areas that

receive higher amounts of UV, generally located closer to the equator, tend to have darker-skinned populations. Areas that are far from the tropics and hence closer to the poles receive a lower intensity of UV, which is reflected in lighter-skinned populations.[3] Natural skin colour can also darken as a result of tanning due to exposure to sunlight. The leading theory is that skin colour adapts to intense sunlight irradiation to provide partial protection against the ultraviolet fraction of the sunlight that produces damage and thus mutations in the DNA of the skin cells.[4] In addition, it has been observed that adult human females on average are significantly lighter in skin pigmentation than males. Females need more calcium especially during pregnancy and lactation. The body synthesizes vitamin D from sunlight, which helps it absorb calcium. Thus females have evolved to have lighter skin so that their bodies absorb more calcium.[5]

The colour of skin is probably one of the colours that we see most in our daily lives, and it plays an important role in many multidisciplinary applications. Apart from the reproduction of skin colour in general amateur and professional photography,[6] cinematography and printing, these include the photography of skin for medical recording and diagnosis and the potential manufacture of prosthetics,[7–9] skin colour based faced detection for computer vision applications,[10–12] the identification of the skin colour preference for applications in, for example, the cosmetics industries[13–16] and, more recently, skin colour reproduction in 3D printing.[17,18] For all these applications, a reliable technique to quantify the colour of skin objectively, in all its many variations, is of vital importance.

The CIE system of colorimetry[19,20] is widely used for the calculation of appropriate colour-related parameters, for example, CIELAB coordinates, from measurements of the spectral reflectance of a surface using a spectrophotometer or the spectral power emitted from a self-luminous source using a telespectroradiometer.

CIE colorimetry has also been widely used to provide objective measurements of skin colour[21] for multi-disciplinary applications, for example, in industries that rely on paper, printing, pigments, and dyes, as well as information shown on computer or television displays. Comparison of the measurements of skin colour however, can show some variation.[22] Two important contributions to this variation are the fact that the skin tends to be a non-flat, uneven surface and second, skin does not exhibit spatial uniformity over the measurement area: these effects combine to make skin colour measurement difficult and often unreliable. A third reason is that, as described above, the human skin is a complicated multi-layer material that is translucent; some incident light, as well as being reflected by the top surface, is transmitted throough the top layer of the skin and hence penetrates the sub-layers before being reflected which, in turn affects the overall colour appearance. Thus the colour of skin can be changed for many reasons; for example, the applied pressure, the environmental temperature, changes in the blood flow, and so on all impact on the final colour. Consequently, the measurement of skin colour may be affected by these various parameters, as well as additional measurement parameters that include the measurement distance, the instrument aperture size, the pressure applied to the skin by the instrument, as well as the body location selected for measurement and the gender and ethnic origin of the individual being considered.[23-25] Thus it might be expected that measurements made with a spectrophotometer where, usually, the instrument comes into contact with the skin, might yield different results from those made using a telespectroradiometer which is a non-contact instrument.

FIGURE 1 The spectrophotometer location to obtain measurements on the subject's forehead

Despite the importance of reliable skin colour measurements, very little is known about the variability of these measurements and their dependency on the acquisition parameters. The main purpose of this article is to quantify the effect of these factors on the measurement reliability. Knowledge of the instrument settings that produce highest repeatability is useful for other researchers involved in skin measurements, and data on the inter-instrument agreement allows the meaningful comparisons between data sets obtained with different instruments.

2 | METHODOLOGY

The primary aim of this study was to investigate the measurement variability of skin colour using two widely used measurement devices: the Konica Minolta CM700d spectrophotometer (SPM) and the Photo Research PR-650 SpectraScan telespectroradiometer (TSR). The short-term repeatability with different settings was evaluated together with the differences between results from the two instruments. Measurements were made using different body locations and different ethnic groups for both genders.

2.1 | Spectrophotometer measurements

A spectrophotometer measures the reflectance factor, the ratio of the radiant flux reflected by the sample into a defined

FIGURE 2 The four masks used to make measurements using the spectrophotometer. Upper left: medium aperture/low pressure (MA/LP). Lower left: small aperture/low pressure (SA/LP). Upper right: medium aperture/high pressure (MA/HP). Lower right: small aperture/high pressure (SA/HP)

cone, to the amount of radiant flux similarly reflected by the perfect diffuser. The "defined cone" is usually assumed to be fairly small—it would have to be negligibly small to give radiance factor exactly. The measurement is independent of the spectral power distribution of the light source used in the instrument. The perfect diffuser is approximated by a calibrated white surface, usually a ceramic or enamel white tile.

It is usually required that the measurement aperture of the SPM be in contact with the surface so that the measurement is not affected by ambient illumination. The Konica Minolta spectrophotometer used was a portable instrument such that it could be taken to the body location on the subject to be measured, rather than requiring the body location to be

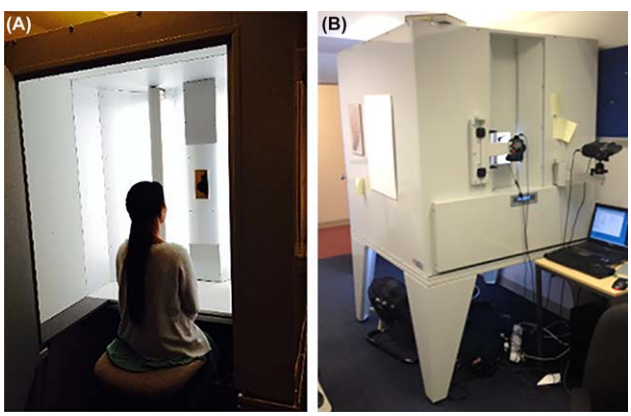

FIGURE 3 The subject sitting in the front of the viewing cabinet (left) and the telespectroradiometer mounted at the back of the viewing cabinet (right)

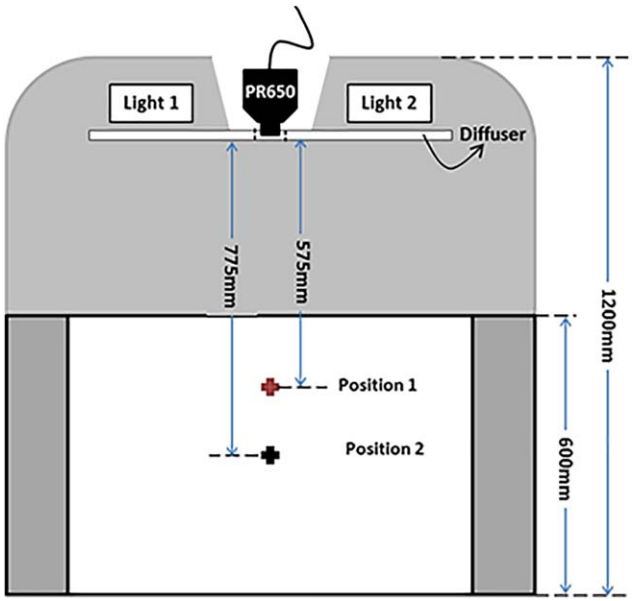

FIGURE 4 The viewing cabinet used to measure skin colour with the telespectroradiometer. The position of the subject provided two different measurement field sizes

presented to a measurement aperture on a bench instrument, Figure 1. Also the instrument uses a pulsed xenon lamp with a UV cut-off filter so the measurement time is extremely short at approximately 1 second. This should serve to minimize the possible effects of movement during the measurement period bearing in mind that the instrument is being held at the required location on the subject's body. The instrument is also able to include or exclude in the measurement the specular component of the reflected light: in this study, specular measurements were included.

The Konica Minolta spectrophotometer comes with CM-SA skin analysis software appropriate for skin colour measurement. Four different aperture masks were used to vary the size of the measurement area and the pressure applied to the skin surface: a medium aperture (MA) with a diameter of 8 mm and a small aperture (SA) with a diameter of 3 mm. Each aperture size was coupled with two different target masks: one with a plate in the front to reduce pressure by actual contact over a larger area, the low pressure (LP) mask, and the other without the plate, the high pressure (HP) mask, respectively, Figure 2. The illumination/viewing geometry of

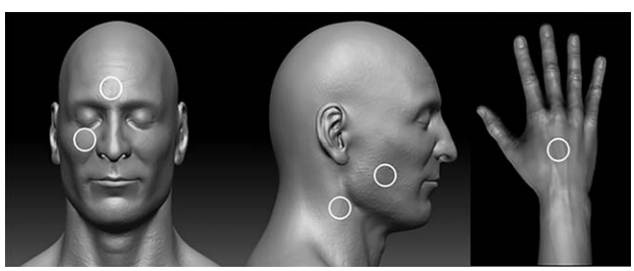

FIGURE 5 The body locations used for the skin colour measurements. Left upper, forehead; left lower, cheek bone; middle upper, cheek; middle lower, neck; right, back of hand

TABLE 1 The short term repeatability of the SPM and the TSR measurements with different instruments settings

		SPM							
		Repeatability		Pressure		Aperture size		TSR	
Mean MCDM (ΔE^*_{ab})		CT (continuous)	CS (consecutive)	LP	HP	MA	SA	P1 (near)	P2 (far)
Human skin	Max.	0.87	1.82	1.56	1.82	0.94	1.82	0.97	1.19
	Min.	0.08	0.28	0.08	0.18	0.08	0.20	0.24	0.14
	Mean	0.37	0.63	0.34	0.40	0.35	0.39	0.51	0.53
PANTONE SkinTone™ Guide	Max.	0.15						0.10	
	Min.	0.02						0.08	
	Mean	0.07						0.09	

the SPM was such that the illumination was diffuse with viewing at 8° from the surface normal: thus the geometry could be designated di:8° using CIE terminology.[20] The wavelength range measured was from 400 to 700 nm with a measurement interval of 10 nm; the half-bandwidth of the instrument is stated by the manufacturer as being approximately 10 nm.

2.2 | Spectroradiometer measurements

A spectroradiometer measures the absolute radiant flux emitted by a source of radiation and, to achieve these absolute measurements it must be calibrated against a standard lamp of known spectral radiant power, usually provided by the instrument manufacturer or a national standardization laboratory. The units measured are watts per steradian per square meter per nanometer (W st^{-1} m^{-2} nm^{-1}). In many situations, knowledge of the absolute power is not required and the measurements can be compared with the known relative spectral power distribution of a standard source. If the surface to be measured is a reflecting surface then it must be illuminated by a suitable light source and the reflected light focused onto the detector in the spectroradiometer, usually by a telescope appropriately attached to the instrument: hence it becomes a telespectroradiometer.

In this study the measurements were made in a purpose-built viewing cabinet supplied by Verivide® (Figure 3). This viewing cabinet was 1200 mm wide × 1200 mm deep × 2000 mm high which was big enough to allow subjects to sit inside it, Figure 4. It was painted a neutral matt colour inside (Munsell Value N7) and contained a D65 fluorescent simulator (CCT = 6726 K, (x, y)=(0.3063, 0.3290, CIE colour rendering index = 98)) to provide diffuse illumination. The TSR was installed at the back of the cabinet and behind the light sources such that no light could be directly measured. The location of the TSR was fixed during the measurement. Different facial locations of each subject were measured by adjusting the height of the chair and the sitting position for each subject. Two measurement distances (distance between the face of the subject and the instrument) were used: 575 and 775 mm, indicated as Position 1 (P1) and Position 2 (P2) in Figure 4. The collection angle of the instrument was fixed at 1° resulting in measurement field sizes with a diameter of 10.0 and 13.5 mm, respectively.

TABLE 2 Values ofcolour difference, ΔE^*_{ab}, obtained either within an instrument (TSR; SPM) or between different instruments (TSR vs. SPM)

Instrument(s)	Cross-comparisons	Mean colour difference (ΔE^*_{ab})
TSR	Position: P1 and P2	2.79
SPM	MAV with different pressure: HP and LP	0.88
	SAV with different pressure: HP and LP	1.82
	LP with different aperture size: MA and SA	2.48
	HP with different aperture size: MA and SA	3.04
TSR versus SPM	TSR versus SPM (SA/LP)	3.49
	TSR versus SPM (MA/LP)	2.57
	TSR versus SPM (SA/HP)	4.02
	TSR versus SPM (MA/HP)	2.69
	TSR versus SPM: PANTONE SkinTone™ Guide	0.88

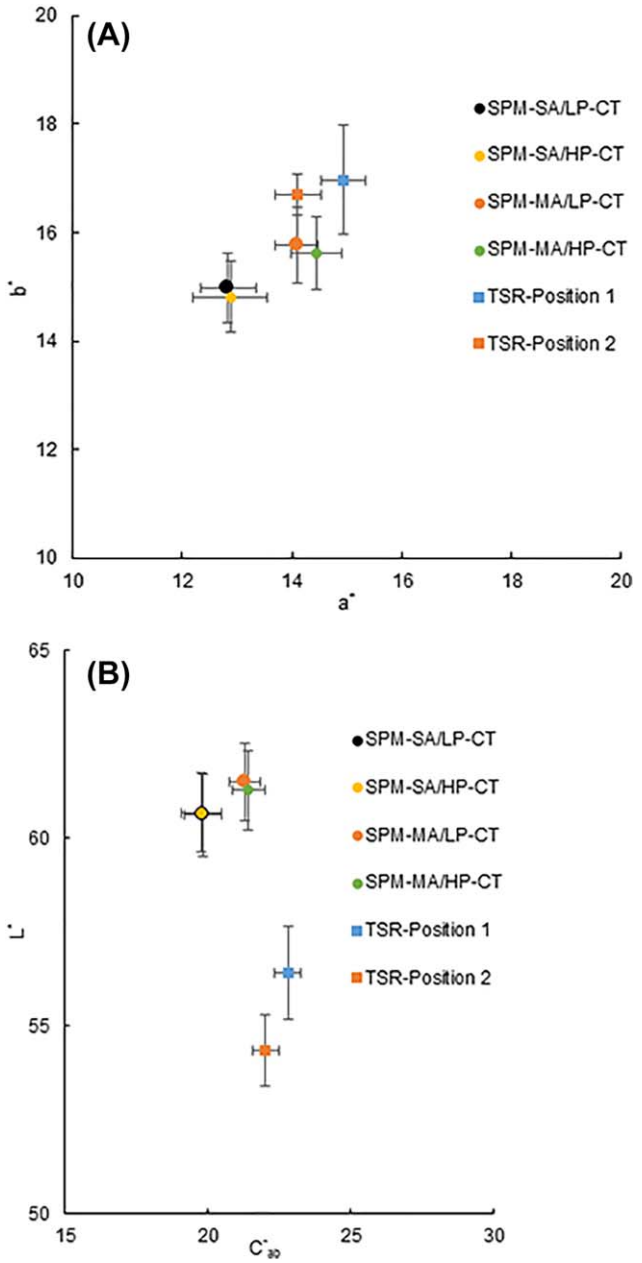

FIGURE 6 CIELAB values for different settings of the SPM and the TSR in the (A) a^*b^* plane and (B) in the $L^*C^*_{ab}$ plane. Appropriate values of the standard deviation of the measurements are also shown. MA, SA, small, medium aperture; LP, HP, low, high pressure; CT, continuous measurements

The PhotoResearch PR650 telespectroradiometer was calibrated by a national standardization laboratory. The wavelength range measured was from 380 to 780 nm with a measurement interval of 4 nm: the half-bandwidth of the instrument is stated by the manufacturer as being approximately 8 nm.

2.3 | Colorimetric calculations

The measurement results of the TSR and the SPM were interpolated into 1 nm using a recommended CIE interpolation method.[26] Then, the CIE 2 degree colour matching function tabulated at 1 nm intervals were used to calculate the CIE tristimulus values. Then the CIE tristimulus values was converted into CIELAB values by using the tristimulus values of the D65 simulator in the viewing cabinet which were gained by using the TSR to measure a reference white placed in the light booth at P1 and P2, respectively. The measurement results from SPM used the CIEXYZ from the P1 in the Conversion. For the TSR, the measurement results from P1 and P2 used the tristimulus values of D65 samulator at P1 and P2 in the conversion, respectively.

2.4 | Short-term repeatability

To investigate measurement short-term repeatability for the two separate instruments, two facial locations, the forehead (FH), and the cheekbone (CB), of each of the eleven subjects, (seven Caucasians, three Chinese, and one South Asian), were measured with both the SPM and the TSR. For the SPM measurements, each of the four masks was used to measure at each skin location. Two measurement methods were used: a continuous and consecutive method. For the continuous repeatability measurement (CT), the target was measured five times continuously without removing the instrument from the subject. For consecutive repeatability measurement (CS), the target was also measured five times, but the instrument was removed and replaced between each measurement. For the TSR measurement, each skin location was measured five times, at each of two viewing distances. Thus a total of 1100 spectra were measured: 880 for the spectrophotometer and 220 for the telespectroradiometer.

To assess the consistency within and between the instruments, the measurement results were recorded in terms of CIELAB coordinates using CIE illuminant D65 and the CIE two degree standard observer. From these coordinates, it was possible to calculate the mean of the colour differences from the mean value of colour difference (MCDM) between the five repeat measurements which is reported as a value of ΔE^*_{ab}, as defined in Equation 1 below: a large MCDM value reflects poor repeatability.

$$\mathrm{MCDM} = \frac{\sum_{i=1}^{n} \Delta E^*_{ab,i}}{n} \tag{1}$$

where

$$\Delta E^*_{ab,i} = \sqrt{\left(L^*_i - L^*_m\right)^2 + \left(a^*_i - a^*_m\right)^2 + \left(b^*_i - b^*_m\right)^2}$$

$$L^*_m = \frac{\sum_{i=1}^{n} \Delta L^*_i}{n},\ a^*_m = \frac{\sum_{i=1}^{n} \Delta a^*_i}{n},\ b^*_m = \frac{\sum_{i=1}^{n} \Delta b^*_i}{n}$$

And $n = 5$ is the total number of repeat measurements taken in each group, L^*_i, a^*_i, b^*_i are the CIELAB values of

TABLE 3 Short-term repeatability as measured by values of MCDM of skin colour measurements for four ethnic groups and five body locations: FH, forehead; CB, cheek bone; CK, cheek; NK, neck; BH, back of hand

MCDM (ΔE^*_{ab})	TSR						SPM					
	FH	CB	CK	NK	BH	Mean	FH	CB	CK	NK	BH	Mean
Chinese	0.33	0.76	0.87	0.98	0.63	0.71	0.42	0.38	0.36	0.40	0.24	0.43
Caucasian	0.47	0.80	1.04	1.13	1.01	0.89	0.42	0.46	0.42	0.55	0.28	0.48
South Asian	0.40	0.66	0.99	0.72	1.04	0.76	0.40	0.76	0.34	0.41	0.26	0.43
African	0.75	0.55	2.30	0.78	0.62	1.00	0.17	0.18	0.23	0.23	0.19	0.25
Mean	0.49	0.69	1.30	0.90	0.82	**0.84**	0.35	0.44	0.34	0.40	0.24	**0.40**

each measurement and L^*_m, a^*_m, b^*_m are the mean CIELAB values of each group of five measurements.

2.5 | Inter-instrument agreement

Inter-instrument agreement is measured in terms of colour difference and changes of colour appearance in each perceptual attribute between measurement results, when the skin colour is measured at the same body location of the same subject but using different instruments with different instrument settings. In this study, the value of CIE colour difference, ΔE^*_{ab} between corresponding measurements made using the two instruments is used to represent the inter-instrument agreement.

2.6 | Subjects and samples measured

To investigate whether body location and ethnicity affect the variability in the skin colour measurements, the skin colour of each of 151 subjects (sampled from four ethnic groups: 69 Chinese subjects, 64 Caucasian subjects, 10 South Asian subjects, and 8 African subjects—in an ideal world, each ethnicity should be represented by an equal-sized group, but in the real word this is difficult to achieve) was measured at five body locations (forehead, cheek, cheekbone, neck, and the back of the hand) using both the SPM and the TSR (Figure 5) with those instrument settings that gave the most repeatable results (see Table 1). For the TSR, the skin patches were placed at Position 1 and measured five times continuously; for the SPM, the same skin patches were measured with MA/LP mask five times continuously.

To determine how much of the measurement variability is due to the nature of human skin (uneven surface, inhomogeneity), additional measurements were taken using flat, two-dimensional surfaces of the PANTONE SkinTone™ Guide, created by scientifically measuring thousands of actual skin tones across many human skin types and made by colour painting on to a thick paper surface.

3 | RESULTS AND DISCUSSION

3.1 | Short-term repeatability

The values of the mean MCDM for measurements made using the SPM and the TSR, with different instrument settings, are shown in Table 1. For the TSR, the mean MCDM values for both measurement distances were approximately 0.50 ΔE^*_{ab}, which shows that the short-term repeatability of the TSR is not affected by distance. The smallest variability was found for the large aperture size and low-pressure mask, 0.34 ΔE^*_{ab}. The mean MCDM values for the masks with and without the pressure plate were 0.34 and 0.40 ΔE^*_{ab}; the mean MCDM values for the MA and SA aperture sizes were 0.35 and 0.39 ΔE^*_{ab}, respectively. While different measurement field sizes and pressures do not affect repeatability (ranging between 0.30 and 0.40), continuous repeatability (CT) is much better than consecutive repeatability (CS) by almost a factor of 2 (0.63 vs. 0.37 ΔE^*_{ab}). However, all of the mean MCDM values calculated from the SPM and the TSR measurements were less than 1.0 ΔE^*_{ab} which is very low.

When the PANTONE SkinTone™ Guide was used as a control, both SPM (MA/LP) and TSR (P1) repeatability is very high (MCDM < 0.09 ΔE^*_{ab}). This demonstrates that

TABLE 4 Inter-instrument agreement as measured by values of MCDM, for four ethnic groups and five body locations: FH, forehead; CB, cheek bone; CK, cheek; NK, neck; BH, back of the hand

ΔE^*_{ab}	FH	CB	CK	NK	BH	Mean	STDEV
Chinese	2.27	3.32	3.85	6.35	3.05	3.77	1.55
Caucasians	3.60	4.08	4.60	6.00	3.96	4.45	0.94
South Asian	2.97	5.94	4.81	6.81	3.63	4.83	1.59
African	3.82	3.86	3.76	4.49	2.73	3.73	0.63
Mean	3.17	4.30	4.26	5.91	3.34	**4.20**	1.09
STDEV	0.70	1.14	0.53	1.00	0.56	0.54	

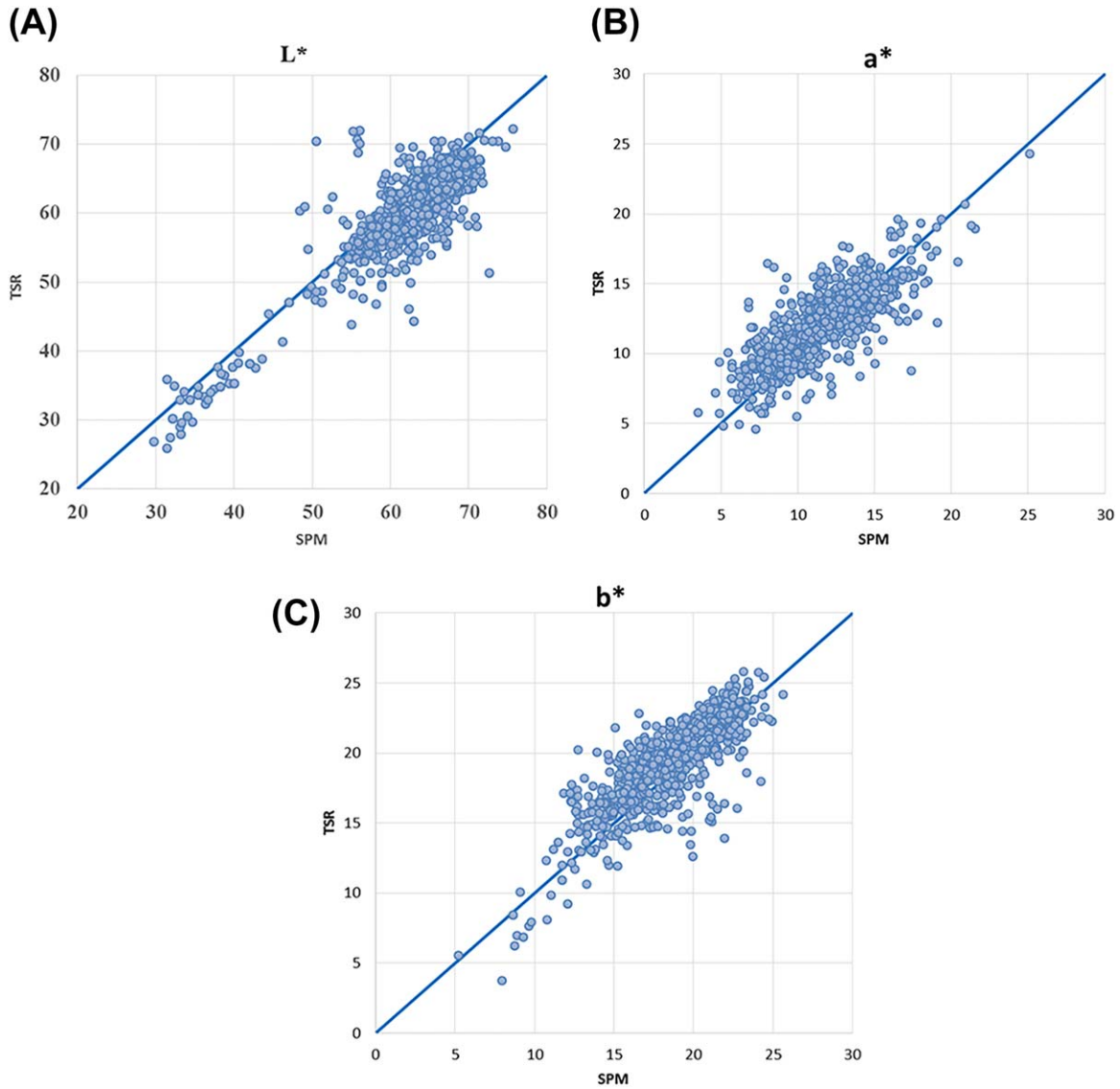

FIGURE 7 Relationship between skin colours measured by two instruments (SPM vs. TSR) for the three colour appearance attributes, L^*,a^*,b^*. The coeffcients of determination are as follows. (A) for L^*: $r^2 = 0.89$; (B) for a^*: $r^2 = 0.78$; (C) for b^*: $r^2 = 0.82$

both the spectroradiometer and the spectrophotometer are able to provide extremely constant readings for uniform and flat surfaces and the variability is due to the nature of the skin patches.

3.2 | Comparisons of the colour shifts between different instrument settings

Mean values of colour differences between different settings on the same instrument and between the two instruments are shown in Table 2. Here, the measurement results for the SPM (CS) setting were used to investigate the colour difference. For the TSR, the mean colour difference between the near (P1) and far measurements (P2) was 2.79 ΔE^*_{ab}. For the SPM, the mean colour differences between HP and LP measurements, were 0.88 and 1.82 ΔE^*_{ab}, respectively, for the medium (MA) and small (SA) apertures. The mean colour differences between the two measurement apertures were 2.48 and 3.04 ΔE^*_{ab} for the low and HP. This result showed that the measurement results of the two measurement field sizes of the TSR and the SPM were greater than 2.0 ΔE^*_{ab}. For the SPM, the greater effect of field size on the colour shift compared with that of pressure is interesting. These imply that reasonable pressure on the skin does not significantly affect the measured colour values, but the measurement field size had the greater impact on the colour shift in the measurements.

Then, based on the above results, the colour shift between the two instruments was investigated by comparison of the measurements between the TSR with measurement distance of P1 and the SPM with all four masks, since the measuring field size of the TSR (P1) is similar to the measurement field size of the SPM (MA).

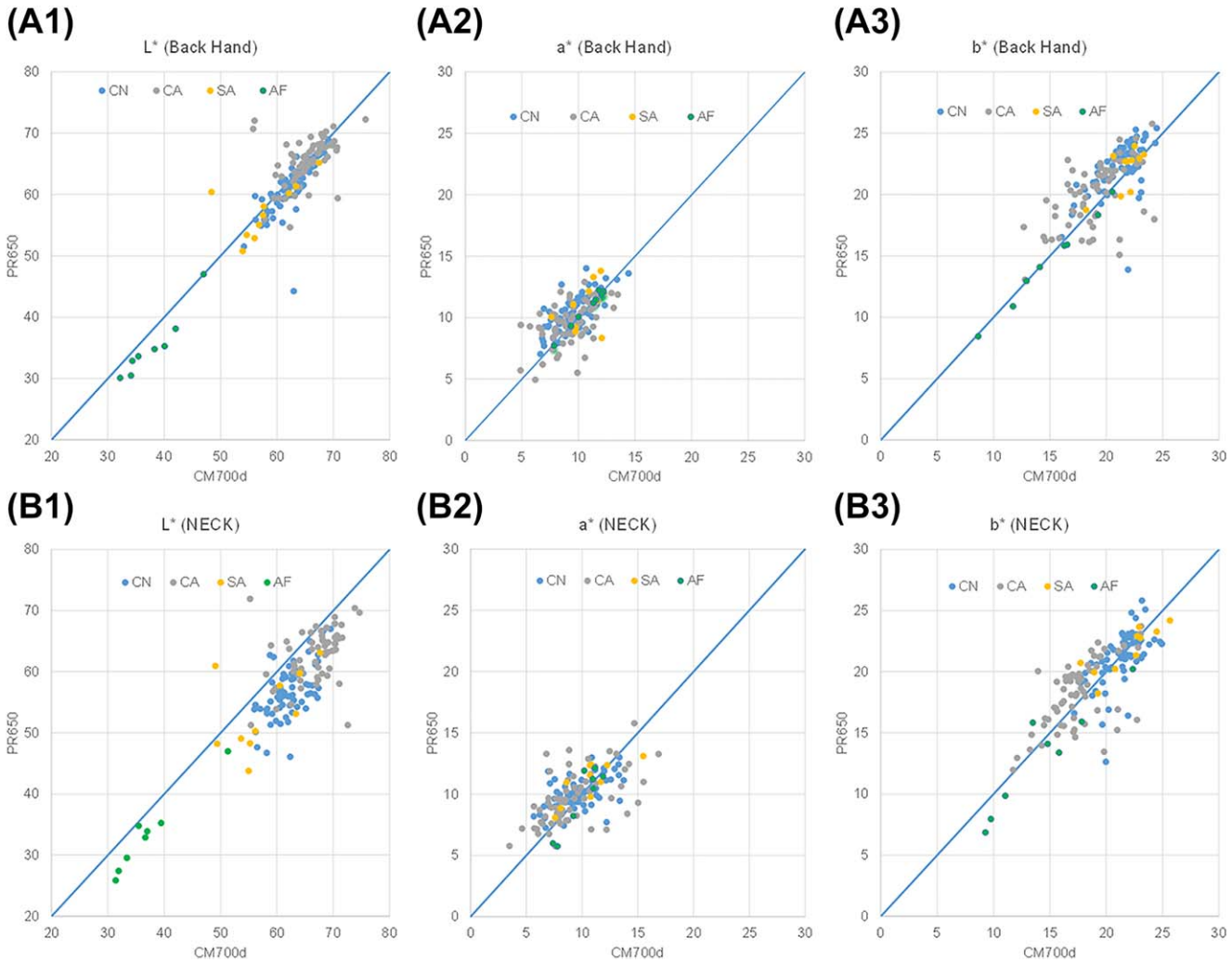

FIGURE 8 Same as Figure 7, but data are plotted for each body location and ethnicity separately

The mean colour shift between the SPM and the TSR (P1) instrument for the PANTONE SkinTone™ Guide was 0.88 ΔE^*_{ab}. When human skin is measured using the two different instruments, the colour difference is in the range 2.57-4.02 ΔE^*_{ab} depending on the instrument measurement settings. The colour shift observed with the Pantone SkinTone Guide may be generated by the difference of measurement geometry, illumination uniformity or the spectral interval of the measurements as well as the inherent measurement uncertainty in the calibration of the instruments which is likely to be greater for the TSR than the SPM. A much larger colour shift is observed for real skin, as expected, due to skin texture, non-uniformity, and so on.

The large SPM aperture size resulted in better agreement with the TSR measurements: the mean colour difference between the measurement results for the TSR and two aperture sizes of the SPM were 3.49 and 2.57 ΔE^*_{ab} at LP, and 4.02 and 2.69 ΔE^*_{ab} at HP, respectively. The average measuring field size of the TSR at the two measuring distances was 12 mm which is closer to the size of the SPM MA with a diameter of 8 mm.

The mean CIELAB values for both instruments are plotted in Figure 6, in both the a^*b^* plane (a) and the $L^*C^*_{ab}$ plane (b). The standard deviation of each appropriate CIELAB value is also plotted as an error bar or the TSR, the measurements at the shorter distance (P1) resulted in a higher lightness L^* and chroma C^*_{ab} (Figure 6B) and also appeared redder (Figure 6A) than the measurements at a longer distance (P2). The lightness and chroma of the near (P1) is higher than the far (P2) about 2 and 1, respectively. The hue angle of the far (P2) is higher than the near (P1) about 1°. For the SPM, the pressure has little effect on lightness and chroma. On the other hand, different apture sizes change the colour appearance: MA measurements gave higher chroma and lightness compared with SA measurements. In summary, the different field sizes result in systematic colour shifts, primarily in the lightness direction and to a smaller extent in chromaticity. From the standard deviation of each data series it can be seen that the variation in the a^*b^* plane (Figure 6A) was not affected by the measurement methods whereas the standard deviations in the lightness values were larger than those for the chroma (Figure 6B).

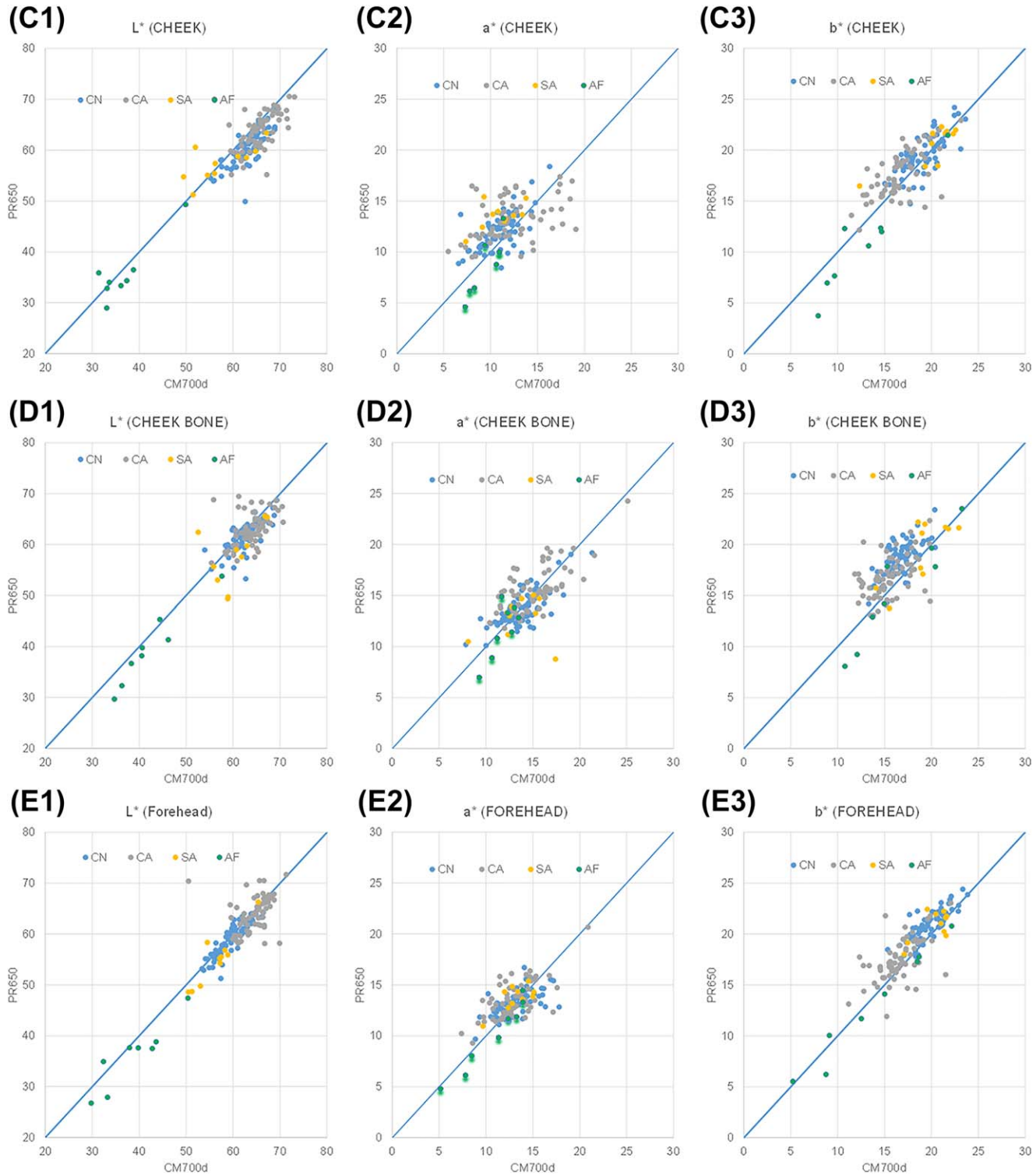

FIGURE 8 Continued

3.3 | Effect of body location and ethnicity

To investigate whether measurement repeatability depends on ethnicity and body location, measurements from an additional 151 subjects were obtained with the parameters that resulted in the best repeatability (see section 2 for details) for the TSR and the SPM.

3.3.1 | Short term repeatability

Measurements for 151 subjects, from each of five body locations, using the SPM and the TSR and the mean values of the colour difference from the mean value (MCDM) are shown in Table 3. For the TSR, the MCDM ranges from 0.71 to 1.00 ΔE^*_{ab} across ethnicities, whereas a larger range

is observed across body locations, from 0.49 to 1.30 ΔE^*_{ab}. This poor repeatability across body locations could be caused by the difference in measurement angle. Best repeatability was observed for the chinese group (lowest MCDM: 0.71); highest MCDM for the African group (1.00). Repeatability was on average best for the Forehead (FH: 0.49) and worst for the cheekbone (CB: 1.3), in particular for the African group (MCDM: 2.3). The poor repeatability in the African group is probably caused by the longer integration time required for dark samples. Also the TSR is a non-contact instrument and thus it is hard to fix the measurement location precisely compared to using the SPM.

For the SPM, the MCDM for skin colour measurements between the different ethnic groups range from 0.25 to 0.48 ΔE^*_{ab}, while the MCDM between different body locations range from 0.35 to 0.44 ΔE^*_{ab}. On average, repeatability of the SPM exceeds that of the TSR by a factor or 2; the exception is the African group, which shows superior repeatability when the SPM is used (0.25 compared to 1.00). This implies that the variability in the measurements was caused not only by the skin itself, but also by the method of measurement, and that these two factors might interact.

An investigation of the impact of the absolute colour values on the short-term repeatability was made. In all cases the association between the absolute colour value and the MCDM is weak when expressed by the coefficient of determination (r^2): in all cases less than 3% of the variance is explained.

3.3.2 | Inter-instrument agreement

To assess the inter-instrument agreement, the colour difference of measurements made at the same body location measured using both the TSR and the SPM for each ethnic group at each body location are listed in Table 4. The mean colour difference of 151 subjects at the five body locations was approximately 4.20 ΔE^*_{ab}, which was larger than the previous test, (2.57 ΔE^*_{ab}; Table 2, for 11 subjects and two body locations). This may be due to additional measurements of the Neck (NK) which was not included in the first set of measurements.

Average inter-instrument differences ranged from 3.73 ΔE^*_{ab} (African group) to 4.83 ΔE^*_{ab}(South Asian). The variation across ethnicity, however, was smaller than the effect of body location, where inter-instrument differences were in the range 3.17-5.91 ΔE^*_{ab}. The forehead (Chinese, Caucasians, and South Asian) and the back of the hand for the African group yielded the best instrument agreement, whereas the neck had the worst variation for all the different ethnic groups. This variation across body location may be attributed to the nature of these surfaces: the back of the hand and the forehead are relatively flat surfaces compared to the neck. To compare variations between body locations and ethnicities, the standard deviation of the overall mean value for each body location (bottom row, Table 4) and that in each ethnic group (right hand column, Table 4) was calculated. The variation as a function of body location was 1.09 ΔE^*_{ab}, approximately twice as large as the variation due to ethnicity (0.54 ΔE^*_{ab}). This confirms that inter-instrument variation is due primarily to body location, not ethnicity.

TABLE 5 Coefficients of best-fit line for different body locations for each colour attribute

Coefficients	Back of hand	Cheek	Cheek bone	Neck	Forehead
k (L^*)	0.99	0.97	0.98	0.93	0.99
k (a^*)	1.04	1.06	0.99	1.01	1.01
k (b^*)	1.04	1.03	1.08	1.00	1.04

To quantify the inter-instrument agreement, the colour attributes L^*, a^*, b^* derived from both instruments are plotted (Figure 7). The best fitting line was determined for each colour attribute, with the constraint that the line passes through zero, since each instrument is bound to have zero output for a black sample. The slopes of the best-fit lines are 0.97, 1.02, and 1.04 for colour attributes L^*, a^*, b^* (Figure 7A-C).

Figure 8 shows the inter-instrument association for each ethnicity and body location separately. Rows represent body locations: back of the hand (BH), neck (NK), cheek (CK), cheek bone (CB), and forehead (FH); the colour in each subplot represent the four different ethnic groups Chinese (CH), Caucasian (CA), South Asian (SA), and African (AF). As expected from Figure 8, the data points are in general clustered around the 45° line indicating a good agreement between the TSR and the SPM. There are some systematic inter-instrument differences, for example, for the neck measurements, the SPM consistently yields a higher lightness value than the TSR, as the complicate geometry of the neck location.

The slope (k) of the best-fitting lines passing through the origin are shown Table 5.

Applying this linear transformation to the SPM measurements, will reduce the observed differences between the SPM and TSR measurements. The colour differences between these predicted results and original measurements are given in Table 6.

Comparing the results shown in Tables 4 and 6, it can be seen that the overall mean colour difference is reduced from 4.20 to 3.50 ΔE^*_{ab}. Specifically, the colour difference becomes smaller or equal for almost all data points, except for the African Cheek results where it is 0.10 ΔE^*_{ab} larger.

TABLE 6 The colour differences between the predicted TSR results and the measured TSR results together with the associated value of the standard deviation (STDEV)

ΔE^*_{ab}	FH	CB	CK	NK	BH	Mean	STDEV
Chinese	2.03	2.59	2.70	3.66	2.44	**2.68**	0.60
Caucasian	3.32	3.66	4.20	4.42	3.78	**3.88**	0.44
South Asian	2.63	5.27	4.59	4.39	3.23	**4.02**	1.07
African	3.80	3.91	3.86	2.79	2.65	**3.40**	0.63
Mean	**2.94**	**3.86**	**3.84**	**3.82**	**3.02**	**3.50**	0.47
STDEV	0.78	1.10	0.82	0.77	0.60	0.60	

The standard deviation of the mean results for each different body location is reduced from 1.09 to 0.47 ΔE^*_{ab}, indicating that the impact of body location on instrument agreement is largely reduced. All these results demonstrate that instrument agreement can be enhanced by linear correction.

4 | CONCLUSIONS

In this study, the variability of skin colour measurements was investigated for two spectral measurement instruments, a Konica Minolta CM700d spectrophotometer and a PhotoResearch PR650 telespectroradiometer. Skin colour measurements for 11 subjects using two facial areas, the forehead and the cheekbone, were performed and the variability was evaluated using different measurement parameters. As expected, we find that different measurement field sizes and different pressure applied during measurements affect the short-term repeatability. The short-term repeatability of the spectrophotometer measurements is greatly affected by the measurement method: continuous measurements, without removing the instrument from the body location yield more repeatable results than consecutive measurements where the instrument is removed and replaced between each measurement. For the telespectroradiometer, significant colour shifts were found for different measurement distances. These colour shifts were comparable with the differences between the measurements from the spectrophotometer and the telespectroradiometer.

In addition, a large number of skin colour reflection spectra were measured and the effect of ethnicity and body location on measurement variability was investigated. The differences between the measurements for different body locations were, in general, larger than the instrument repeatability and the inter-instrument agreement. A linear best-fit procedure applied to the individual CIELAB coordinates served to improve this situation.

The analysis provides useful guidance for the definition of a protocol for skin colour measurement and the establishment of a skin colour spectral database. It should also be useful when comparing data sets obtained with different measurement instruments.

ACKNOWLEDGMENTS

This work was supported by the Royal Academy of Engineering (Grant No. RECI033) and the Engineering and Physical Sciences Research Council (EPSRC) (EP/K040057, EP/L001012). We would also like to thank Verivide Limited for kindly providing the viewing cabinet.

ORCID

Kaida Xiao http://orcid.org/0000-0001-7197-7159
Sophie Wuerger http://orcid.org/0000-0003-0080-5813

REFERENCES

[1] Muehlenbein MP. *Human Evolutionary Biology*. Cambridge University Press; 2010;192–214.

[2] Milburn PB, Sian CS, Silvers DN. The color of the skin of the palms and soles as a possible clue to the pathogenesis of acral-lentiginous melanoma. *Am J Dermatopathol.* 1982;*4*:429–434.

[3] Webb AR. Who, what, where and when—influences on cutaneous vitamin D synthesis. *Prog Biophys Mol Biol.* 2006;*92*:17–25.

[4] Jablonski NG, Chaplin G. Human skin pigmentation as an adaptation to UV radiation. *Proc Nat Acad Sci.* 2010;*107*(Supplement_2):8962–8968.

[5] Jablonski NG, Chaplin G. The evolution of human skin coloration. *J Hum Evol.* 2000;*39*:57–106.

[6] Imai FH, Tsumura N, Haneishi H, Miyake Y. Principal component analysis of skin color and its application to colorimetric color reproduction on CRT display and hardcopy. *J Image Sci Technol.* 1996;*40*:422–430.

[7] Xiao K, Zhu Y, Li C, Connah D, Yates JM, Wuerger S. Improved method for skin reflectance reconstruction from camera images. *Opt Exp.* 2016;*24*:14934–14950.

[8] Chung DH, Sapiro G. Segmenting skin lesions with partial-differential-equations-based image processing algorithms. *IEEE Trans Med Imag.* 2000;*19*:763–767.

[9] Hirose M, Akaho R, Maita C, Sugawara M, Tsumura N. Optimization of spectral sensitivities of mosaic five-band camera for estimating chromophore densities from skin images including shading and surface reflections. *Opt Rev.* 2016;*23*:544–548.

[10] Bilal S, Akmeliawati R, Salami MJE, Shafie AA. Dynamic approach for real-time skin detection. *J Real Time Image Process.* 2015;*10*:371–385.

[11] Brancati N, Frucci MD, Pietro G, Gallo L. Dynamic clustering for skin detection in YCBCR colour space. *Pattern Recogn Inform Process.* 2016;49–53.

[12] Jones MJ, Rehg JM. Statistical color models with application to skin detection. *Int J Comput Vis.* 2002;*46*:81–96.

[13] Fink B, Grammer K, Matts PJ. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evol Hum Behav.* 2006;*27*:433–442.

[14] Kikuchi K, Masuda Y, Yamashita T, Kawai E, Hirao T. Image analysis of skin color heterogeneity focusing on skin chromophores and the age-related changes in facial skin. *Skin Res Technol.* 2015;*21*:175–183.

[15] Lefevre CE, Perrett DI. Fruit over sunbed: carotenoid skin colouration is found more attractive than melanin colouration. *Q J Exp Psychol.* 2015;*68*:284–293.

[16] Shuster S, Fisher G, Harris E, Binnell D. The effect of skin disease on self image. *Br J Dermatol.* 1978;*99*(s16):18–19.

[17] Xiao K, Zardawi F, van Noort R, Yates JM. Developing a 3D colour image reproduction system for additive manufacturing of facial prostheses. *Int J Adv Manuf Technol.* 2014;*70*(9–12): 2043–2049.

[18] Xiao K, Zardawi F, Van Noort R, Yates JM. Color reproduction for advanced manufacture of soft tissue prostheses. *J Dent.* 2013;*41*(Suppl. 5):e15–e23.

[19] Hunt RWG, Pointer MR. *Measuring Colour*. John Wiley & Sons; 2011.

[20] Colorimetry C. CIE Publication 15: 2004. Commission Internationale de lEclairage, Vienna. 2004.

[21] Weatherall IL, Coombs BD. Skin color measurements in terms of CIELAB color space values. *J Investigat Dermatol.* 1992;*99*: 468–473.

[22] Harrison GA. Differences in human pigmentation: measurement, geographic variation, and causes. *J Investigat Dermatol.* 1973; *60*:418–426.

[23] Xiao K, Yates JM, Zardawi F, et al. Characterising the variations in ethnic skin colours: a new calibrated data base for human skin. *Skin Res Technol.* 2017;*23*:21–29.

[24] Xiao K, Liao N, Zardawi F, et al. Investigation of Chinese skin colour and appearance for skin colour reproduction. *Chinese Opt Lett.* 2012;*10*:083301.

[25] Wang Y, Luo MR, Wang M, Xiao K, Pointer M. Spectrophotometric measurement of human skin colour. *Color Res Appl.* 2007;*42*:764–774.

[26] International Commission on Illumination. CIE Publ. 167:2005, Recommended practice for tabulating spectral data for use in colour computations. Vienna: CIE Central Bureau; 2005.

AUTHOR BIOGRAPHIES

MENGMENG WANG received a B.Eng. in electronic engineering from the Birmingham City University, UK (2011) and an M.S. in imaging communication and signal processing from the University of Bristol, UK (2012). She received a Ph.D. in colour and image technology at the University of Leeds, U.K. She works as a research assistant to develop a 3D facial imaging system in the Department of Psychological Science in the University of Liverpool. She is also a member of CIE TC1-92 skin colour database. Her current work focuses on researching the colour measurement of skin and the affective design of the facial images.

DR. KAIDA XIAO worked in Samsung Advanced Institute of Technology in Korea, at Tru-Colour Ltd., the University of Sheffield, the University of Liverpool and is now a University Academic Fellow in the University of Leeds and a Visiting Professor in the University of Science and Technology Liaoning, China. His research interests are related to 3D colour image reproduction, 3D colour printing, medical image capture and analysis, colour appearance modelling and image quality enhancement.

MING RONNIER LUO is a Global Expertise Professor at the College of Optical Science and Technology, Zhejiang University (China), a Visiting Professor of Colour Science and Imaging, University of Leeds (UK) and a Chair professor at the National Taiwan University of Science and Technology, Chinese Taipei. He is a Fellow of the Society for Imaging Science and Technology (IS&T), and the Society of Dyers and Colourists (SDC). He has received numerous awards for his research in colour science and technology including the recent AIC 2017 Judd Award.

DR. MICHAEL R. POINTER received his PhD from Imperial College, London working with David Wright. He then worked in the Research Division of Kodak Limited on fundamental issues of colour science applied to the photographic system. After periods at the University of Westminster, and the National Physical Laboratory, he is now a Visiting Professor at the University of Leeds and Technical Advisor to the Color Engineering Laboratory at Zhejiang University, Hangzhou, China. 1997, he received the Fenton Medal, The Royal Photographic Society's award for services to the Society. In 2004, he received a Silver Medal from the Society of Dyers and Colourists for "contributions to colour science." He is also an Honorary Member of the Colour Group of Great Britain and was awarded the 2014 Newton Medal by the Group. He is a Fellow of The Royal Photographic Society and UK Associate Editor of the journal Color Research & Application.

DR. VIEN CHEUNG currently holds the academic post of Associate Professor in Colour and Imaging Science at the University of Leeds. Her research interests are colour vision, spectral imaging, colour reproduction and colour, all as applied to the art and design disciplines. Dr Cheung is active in several professional bodies including The Colour Group Great Britain (CGGB), the International Colour Association (AIC), the Society of Dyers and Colourists (SDC) and the Society for Imaging Science and Technology (IS&T). She was awarded the Selwyn Award from The Royal Photographic Society in 2008 for her research in colour imaging and imaging technology, a Silver Medal from the Society of Dyers and Colourists in 2011 in recognition of her contributions to education, the Society and in the interests of the allied industries, and the

2016 Service Award from the IS&T for her contributions in organising IS&T's color conference.

SOPHIE WUERGER is a professor in vision science at the University of Liverpool and is interested in understanding how the brain processes visual information, in particular how the signals from the eye map into high-order colour appearance mechanisms. To gain an understanding of the neural visual mechanisms she uses behavioural methods, EEG and functional brain imaging. Her work has both clinical and technological application, including device-independent display calibration and the development of computational appearance models.

How to cite this article: Wang M, Xiao K, Luo MR, Pointer M, Cheung V, Wuerger S. An investigation into the variability of skin colour measurements. *Color Res Appl*. 2018;43:458–470. https://doi.org/10.1002/col.22230