# Accessible remote sensing of water

Burggraaff, O.

# 4 | Accuracy and reproducibility of above-water radiometry with calibrated smartphone cameras using RAW data

Adapted from

Olivier Burggraaff, Mortimer Werther, Emmanuel Boss, Stefan Simis, Frans Snik

*Frontiers in Remote Sensing, 3, 940 096 (2022)*

doi:10.3389/frsen.2022.940096

Consumer cameras, especially on smartphones, are popular and effective instruments for above-water radiometry. The remote sensing reflectance $R_{rs}$ is measured above the water surface and used to estimate inherent optical properties and constituent concentrations. Two smartphone apps, HydroColor and EyeOnWater, are used worldwide by professional and citizen scientists alike. However, consumer camera data have problems with accuracy and reproducibility between cameras, with systematic differences of up to 40% in intercomparisons. These problems stem from the need, until recently, to use JPEG data. Lossless data, in the RAW format, and calibrations of the spectral and radiometric response of consumer cameras can now be used to significantly improve the data quality. Here, we apply these methods to above-water radiometry. The resulting accuracy in $R_{rs}$ is around 10% in the red, green, and blue (RGB) bands and 2% in the RGB band ratios, similar to professional instruments and up to 9 times better than existing smartphone-based methods. Data from different smartphones are reproducible to within measurement uncertainties, which are on the percent level. The primary sources of uncertainty are environmental factors and sensor noise. We conclude that using RAW data, smartphones and other consumer cameras are complementary to professional instruments in terms of data quality. We offer practical recommendations for using consumer cameras in professional and citizen science.

## 4.1 Introduction

The remote sensing reflectance $R_{rs}(\lambda)$ is an apparent optical property that contains a wealth of information about the substances within the water column [2]. In above-water radiometry, $R_{rs}$ is measured using one or more (spectro)radiometers deployed above the water surface [120]. The absorption and scattering coefficients and concentrations of coloured dissolved organic matter (CDOM), suspended particulate matter, and prominent phytoplankton pigments such as chlorophyll-a (chl-a) can be determined from $R_{rs}$ [234]. Due to spectral range and long-term stability requirements, the equipment necessary for accurate measurements of $R_{rs}$ is often expensive. High costs limit the uptake and, therefore, impact of these instruments.

Consumer cameras have long been seen as a low-cost alternative or complement to professional instruments [102]. Work in this direction has mostly focused on hand-held digital cameras, which measure the incoming radiance in red-green-blue (RGB) spectral bands typically spanning the visible range from 390–700 nm [323]. Uncrewed aerial vehicles (UAVs or drones) and webcams have similar optical properties, often contain the same sensors, and are also increasingly used in remote sensing [281]. Consumer cameras have been used to retrieve CDOM, chl-a, and suspended mineral concentrations through above-water radiometry [34, 323]. They are particularly useful for measuring at small spatial scales, short cadence, and over long time periods [386, 387].

Smartphones are especially effective as low-cost sensing platforms thanks to their wide availability, cameras, and functionalities including accelerometers, GPS, and wireless communications. They are already commonly used in place of professional sensors in laboratories [56, 388]. However, what smartphones truly excel at is providing a platform for citizen science in the field [94, 389]. There is a vibrant ecosystem of applications (apps) using the smartphone camera for environmental citizen science purposes [273]. Some use additional fore-optics to measure hyperspectrally [284, 390], while most use the camera as it is [121,274,391]. Smartphone science apps are also commonly used for educational purposes and in professional research [318, 392, 393].

Two apps are currently widely used for above-water radiometry, namely HydroColor [121] and EyeOnWater [274]. HydroColor measures $R_{rs}$ in the RGB bands using the Mobley protocol [209], guiding the user to the correct pointing angles with on-screen prompts. Through an empirical algorithm based on the red band of $R_{rs}$, the app estimates the turbidity, suspended matter concentration, and backscattering coefficient of the target body of water. EyeOnWater uses the WACODI algorithm [285] to determine the hue angle $\alpha$ of the water, representing its intrinsic colour. From $\alpha$ it also estimates the Forel-Ule (FU) index, a discrete water colour scale with a century-long history [306]. $\alpha$ and the FU index are reasonable first-order indicators of the surface chl-a concentration and optical depth [10].

While these apps and other consumer camera-based methods provide useful data, improvements to the accuracy and reproducibility are necessary to derive high-quality end products. Validation campaigns have consistently found the radiance, $R_{rs}$ in the RGB bands, and hue angle from consumer cameras to be well-correlated with reference instruments, but often with a wide dispersion and a significant bias. For $R_{rs}$, the mean difference between smartphone and reference match-up data is typically $\geq 0.003$ sr$^{-1}$ or $\geq 30\%$, but varies wildly between studies [121, 276, 391, 394]. As an extreme example, Malthus et al. found no correlation at all between HydroColor and reference $R_{rs}$ data [275], albeit under challenging observing conditions. The typical accuracy in $\alpha$ is around 10° or 1–2 FU [275,285,391,395]. Differences in $R_{rs}$ between smartphones can be as large as 40% [276]. The uncertainties,

as well as the differences between smartphones and reference instruments, in observed optical properties and derived water constituent concentrations are often even greater than 40% [275, 396, 397], although this may be explained in part by differences in inherent optical properties and observing conditions between study sites.

A major source of uncertainty in existing methods is the use of the JPEG data format. Until recently JPEG was the only format available to third-party developers on most smartphones and other consumer cameras. JPEG data are irreversibly compressed and post-processed for visual appeal, at the cost of radiometric accuracy and dynamic range. Most importantly, they are very non-linear, meaning a 2× increase in radiance does not cause a 2× increase in response [281]. Instead, in a process termed gamma correction or gamma compression, the radiance is scaled by a power law. The non-linearity of JPEG data is a significant contributor to the uncertainty in $R_{rs}$ obtained from consumer cameras and apps such as HydroColor [275, 281, 394]. Some approaches, including WACODI, attempt to correct for non-linearity through an inverse gamma correction [285, 394]. This inverse correction cannot be performed consistently because the smartphone JPEG processing differs between smartphone brands, models, and firmware versions [281].

A secondary source of uncertainty are the spectral response functions (SRFs) of the cameras. Because exact SRF profiles are laborious to measure and are rarely provided by manufacturers, it is often necessary to use simplified SRFs and assume them to be device-independent [121, 285]. However, the SRFs of different cameras actually vary significantly [281].

The quality of consumer camera radiometry can be improved significantly by using lossless data, in the RAW format, and camera calibrations. RAW data are almost entirely unprocessed and thus are not affected by the uncertainties introduced by the JPEG format. Furthermore, through calibration and characterisation of the radiometric and spectral response, consumer cameras can be used as professional-grade (spectro)radiometers [281].

In this work, we assess the uncertainty, reproducibility, and accuracy of calibrated smartphone cameras, using RAW data, for above-water radiometry. By comparing in situ observations from two smartphone cameras and two hyperspectral instruments, we test the hypothesis that the new methods decrease the uncertainty and increase the reproducibility and accuracy of data from consumer cameras. To our knowledge, this is the first time that the new methods have been applied or assessed in a field setting.

Section 4.2 describes the data acquisition and processing as well as the performed experiments. The results are presented in Section 4.3. In Section 4.4, we discuss the results, compare them to the literature, and present some recommendations for projects using smartphones. Finally, the conclusions of the analysis are presented in Section 4.5.

## 4.2 Methods

Smartphone and reference data were gathered on and around Lake Balaton, Hungary, from 3–5 July 2019. Lake Balaton is the largest (597 $km^2$) lake in central Europe, with a mean depth of only 3.3 m, and is well-studied. It has a high concentration of suspended mineral particles and appears very bright and turquoise (bluish-green) to the eye (Figure 4.1, further discussed in Section 4.2.1). Due to inflow from the Zala river, the western side of the lake is richer in nutrients than the eastern side. The adjacent Kis-Balaton reservoir is hypereutrophic with chl-a concentrations up to 160 mg $m^{-3}$. More detailed descriptions of this site are given

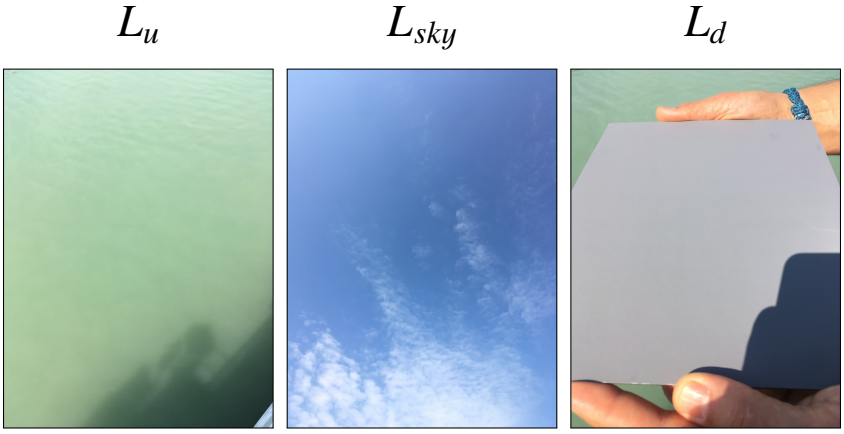


Figure 4.1: Example iPhone SE images of $L_u$, $L_{sky}$, and $L_d$, taken at Lake Balaton on 3 July 2019 at 07:47 UTC. Little wave motion is visible on the water surface in $L_u$, while $L_{sky}$ shows patchy cloud coverage. The conditions seen here were representative for the entire campaign.

in [27, 264].

Two smartphones were used, an Apple iPhone SE and a Samsung Galaxy S8, and two hyperspectral spectroradiometer instruments were used as references. The reference instruments were a set of three TriOS RAMSES instruments mounted on a prototype Solar-tracking Radiometry (So-Rad) platform [123] to maintain a favorable viewing geometry throughout the day, and a hand-held Water Insight WISP-3 spectroradiometer [126]. The spectral and radiometric calibration of the smartphones is described in [281]; manufacturer calibrations were used for the So-Rad and WISP-3.

Data processing and analysis were done using custom Python scripts based on the NumPy [312], SciPy [398], and SPECTACLE [281] libraries, available from GitHub[15]. The smartphone data processing pipeline supports RAW data from most consumer cameras. The processing of the reference and smartphone data is further discussed in Sections 4.2.2–4.2.4, the analysis in Sections 4.2.5 and 4.2.6.

### 4.2.1 Data acquisition

Observations were performed on 3 July 2019 from the Tihany-Szántód ferry on eastern Lake Balaton, performing continuous transects around 46°53'00"N 17°53'43"E, facing southwest before 10:00 UTC (12:00 local time) and northeast afterwards. Data were also acquired on 4 July in the Kis-Balaton reservoir at 46°39'41"N 17°07'45"E and on 5 July on western Lake Balaton at 46°45'15"N 17°15'09"E, 46°42'25"N 17°15'53"E, 46°43'59"N 17°16'34"E, and 46°45'04"N 17°24'46"E. The So-Rad, which was mounted on the ferry, was only used in the morning on 3 July; the two smartphones and WISP-3 were used at all stations. All data, including a detailed station log, are available from Zenodo[16].

The upwelling radiance $L_u$, sky radiance $L_{sky}$, and either downwelling radiance $L_d$ (smartphones) or downwelling irradiance $E_d$ (references) were measured. The So-Rad and WISP-3 data were hyperspectral, the smartphones multispectral in different RGB bands [281]. A

[15] https://github.com/burggraaff/smartphone-water-colour
[16] https://dx.doi.org/10.5281/zenodo.4549621

Brandess Delta 1 18% gray card was used to measure $L_d$, which is discussed in Section 4.2.3. The observations on 3 and 5 July were done under a partially clouded sky (Figure 4.1), which introduced uncertainties in $L_{sky}$ and $R_{rs}$ by increasing the variability of the sky brightness and causing cloud glitter effects on the water surface [209]. Simultaneous measurements from different instruments were affected in the same way, meaning an intercomparison was still possible. However, for measurements taken farther apart in time and space, the match-up error may be significant. On 4 July, the sky was overcast.

Following standard procedure [120, 209], the smartphone observations were performed pointing 135° away from the solar azimuth in the direction furthest from the observing platform and 40° from nadir ($L_u$, $L_d$) or zenith ($L_{sky}$). The smartphones were taped together and aligned in azimuth by eye and in elevation using the tilt sensors in the iPhone SE, to approximately 5° precision. Example smartphone images are shown in Figure 4.1. The same viewing geometry is used in HydroColor, but not EyeOnWater [275]. The reference observations were performed in the same way, following standard procedure for the respective sensors [90, 126].

The So-Rad and WISP-3 each recorded $L_u$, $L_{sky}$, and $E_d$ simultaneously while the smartphones took sequential $L_u$, $L_d$, and $L_{sky}$ images within one minute. Using the SPECTACLE apps for iOS and Android smartphones [281], the iPhone SE took one RAW image and one JPEG image simultaneously, and the Galaxy S8 took 10 sequential RAW images per exposure. The exposure settings on both smartphones were chosen manually to prevent saturation and were not recorded, but were kept constant throughout the campaign.

In total, 304 and 453 sets of WISP-3 and So-Rad spectra, respectively, and 28 sets each of iPhone SE and Galaxy S8 images were obtained. For the WISP-3, one set of spectra (5 July at 10:35:51 UTC) was manually removed because it appeared excessively noisy. Six sets of smartphone data were discarded due to saturation.



### 4.2.2 Reference data processing

$R_{rs}$ spectra were calculated from the WISP-3 and So-Rad data (Figure 4.2). For the WISP-3, the Mobley method [209] shown in Equation (4.1), with a sea surface reflectance factor of $\rho = 0.028$, was used. Wavelength dependencies are dropped for brevity. The value of $\rho = 0.028$ was chosen for the WISP-3 and smartphone data processing (Section 4.2.3) to enable a direct comparison to HydroColor, which uses the same value [121]. Given the brightness of Lake Balaton, the relative magnitude of $\rho L_{sky}$ compared to $L_u$ was small (typically <5% in the WISP-3 data) for any value of $\rho$ around 0.03, and thus the effect of a small difference in $\rho$ on $R_{rs}$ was negligible. The So-Rad data, having a wider spectral range, were processed using the three-component (3C) method, which subtracts an additional glint term $\Delta$ and determines $\rho$ empirically from a spectral optimisation [210, 399].

$$R_{rs} = \frac{L_u - \rho L_{sky}}{E_d} \tag{4.1}$$

The general appearance of the reflectance spectra (Figure 4.2) is that of a broad peak around 560 nm. On the short wavelength side of this peak, absorption by phytoplankton and CDOM suppresses $R_{rs}$ to approximately 25% of the peak amplitude. Towards longer wavelengths, the effects of increasing absorption by water are clearly seen around 600 nm and beyond 700 nm, and $R_{rs}$ reaches near-zero amplitude at the edge of the visible spectrum. The reflectance is ultimately skewed towards blue-green wavelengths, giving the water a turquoise appearance. A minor absorption feature of chl-a and associated accessory pigments is visible

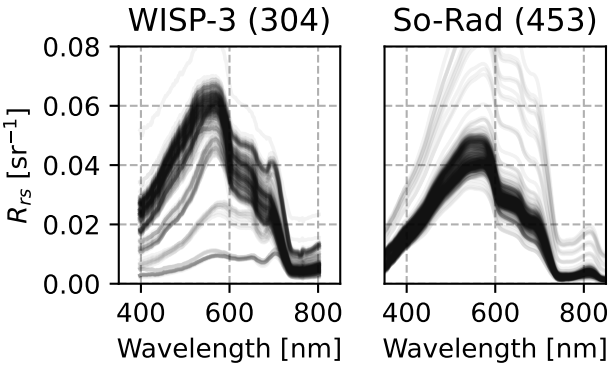


Figure 4.2: Reference $R_{rs}$ spectra derived from measurements on and around Lake Balaton. There is a difference in normalisation between the two data sets, which is discussed in Section 4.4.3.

around 675 nm. Sun-induced chl-a fluorescence is visible at 680–690 nm in the WISP-3 spectra taken on 4 and 5 July, but not the WISP-3 or So-Rad spectra taken on 3 July.

### 4.2.3 Smartphone data processing

The RAW smartphone images were processed using a SPECTACLE-based [281] pipeline (Figure 4.3). The images were first corrected for bias or black level, which shifts the pixel values in each image by a constant amount. On the Galaxy S8, the nominal black level was 0 analogue-digital units (ADU), while on the iPhone SE it was 528 ADU or 13% of the dynamic range, as determined from the RAW image metadata and validated experimentally [281]. Next, a flat-field correction was applied, correcting for pixel-to-pixel sensitivity variations. The sensitivity varies by up to 142% across the iPhone SE sensor [281], although in the central $100 \times 100$ pixels, the variations are only 0.2% on the iPhone SE and 1.6% on the Galaxy S8. A central slice of $100 \times 100$ pixels was taken to decrease the uncertainties introduced by spatial variations across the image [121]. The central pixels were then demosaicked into separate images for the $RGBG_2$ channels, where $G_2$ is the duplicate green channel present in most consumer cameras [281]. The $RGBG_2$ images were flattened into lists of 10 000 samples per channel and normalised by the effective spectral bandwidths of the channels, determined from the SRFs [281]. The mean radiance was calculated per channel, after which the G and $G_2$ channels, which have identical SRFs, were averaged together. Finally, $R_{rs}$ was calculated from $L_u$, $L_{sky}$, and $L_d$ using Equation (4.2) [209]. Like for the WISP-3 (Section 4.2.2) and in HydroColor, a constant $\rho = 0.028$ was used. $R_{ref}$ is the gray card reference reflectance, nominally 0.18.

$$R_{rs} = \frac{L_u - \rho L_{sky}}{\frac{\pi}{R_{ref}} L_d} \tag{4.2}$$

For $R_{ref}$, a Brandess Delta 1 18% gray card was used by manually holding it horizontal in front of the camera. The nominal reflectance of $R_{ref} = 18\%$ was verified to within 0.5 percent point in the smartphone RGB bands by comparing spectroradiometer measurements of $L_d$ on a similar gray card to cosine collector measurements of $E_d$. Angular variations in $R_{ref}$ were found to be $\lesssim$1 percent point for nadir angles of 35°–45° in a laboratory experiment

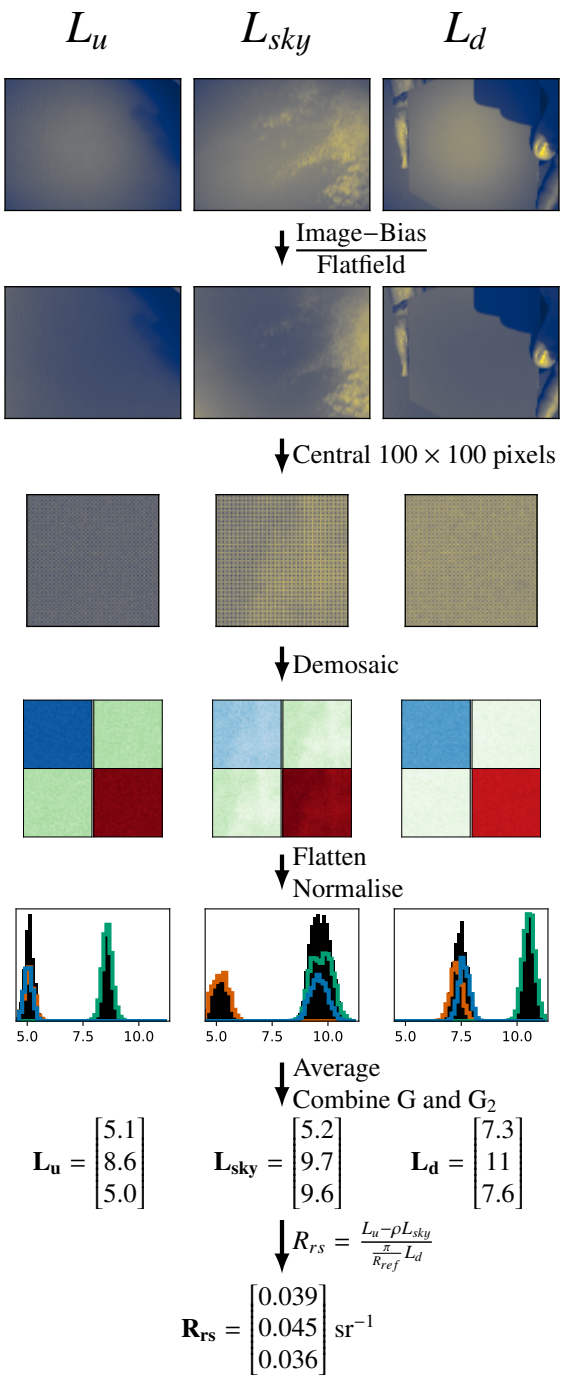


Figure 4.3: Smartphone data processing pipeline, from RAW images to multispectral $R_{rs}$. The example input images are those from Figure 4.1. Some processing steps have been combined for brevity. The histograms show the distribution of normalised pixel values in the central 100 × 100 pixels for the RGB$G_2$ channels separately (coloured lines, G and $G_2$ combined) and together (black bars). The order of elements in **L** and $\mathbf{R_{rs}}$ is RGB.

with the iPhone SE. This value is similar to previous characterisations of different consumer-grade gray cards [400]. To account for these factors as well as fouling, an uncertainty of $\sigma_{R_{ref}} = 0.01$, or 1 percent point, was used in our data processing. This does not account for systematic errors (Section 4.4.3).

Unlike EyeOnWater, which selects multiple sub-images from different parts of each image, our pipeline only used a central slice of $100 \times 100$ pixels. The use of sub-images was not necessary since all images were manually curated and sub-imaging has been shown to have little impact on the data quality [275]. The $100 \times 100$ size was chosen to minimise spatial variations, but a comparison of box sizes from 50–200 pixels showed that the exact size made little difference. For example, the mean radiance typically varied by <0.4%, less than the typical uncertainty on the radiance estimated from each image (Section 4.3.1). Furthermore, the signal-to-noise ratio (SNR) varied by <3% for $L_u$ and $L_d$ but up to 19% in $L_{sky}$ due to the patchy cloud coverage.

The iPhone SE JPEG data were processed using a simplified version of the RAW pipeline, lacking the bias and flat-field corrections and G-$G_2$ averaging. Smartphone cameras perform these three tasks internally for JPEG data [281]. The processing was repeated with an additional linearisation step, like in WACODI and EyeOnWater, to determine whether linearisation improves the data quality. Following WACODI, the default sRGB inverse gamma curve was used, although this curve has already been shown to be poorly representative of real smartphones [281].

The uncertainties in the image data, determined from the sample covariance matrix of the 10 000 pixels per channel per image, were propagated analytically as described in Section 4.A. The pixel values were approximately normally distributed (Figure 4.3). Significant correlations between the RGB$G_2$ channels were found. For example, the iPhone SE $L_{sky}$ image from 3 July 2019 at 07:47 UTC had a correlation of $r_{RG} = 0.68$ between R and G, while in the 08:01 image this was only $r_{RG} = 0.09$. The observed correlations were likely due to spatial structures in the images [140], such as patchy clouds for $L_{sky}$ and waves for $L_u$. In larger data sets, the presence of strong correlations could be used as a means to filter out images that are not sufficiently homogeneous. The propagated uncertainties in $R_{rs}$ were typically 5–10% of the mean $R_{rs}$ and similarly correlated between channels. For example, the 07:47 data had correlations in $R_{rs}$ of $r_{RG} = 0.67$, $r_{RB} = 0.57$, and $r_{GB} = 0.72$.

### 4.2.4 Colour

In addition to absolute $R_{rs}$ in the RGB bands, several relative colour measurements were investigated, namely RGB band ratios, hue angle, and FU index.

The band ratios were calculated as specific combinations of $R_{rs}$ bands. For simplicity in notation, the ratios are expressed as, for example, G/R instead of $R_{rs}(G)/R_{rs}(R)$. Following the literature, the numerators and denominators were chosen as G/R, B/G, and R/B. The G/R ratio is sensitive to water clarity and optical depth [391]. B/G is sensitive to the chl-a concentration [323], at least in water types where phytoplankton covaries with other absorbing substances. Finally, the R/B ratio is particularly sensitive to broad features such as CDOM absorption, as well as the concentration of scatterers (turbidity, suspended matter concentrations), as described in [34, 323].

To calculate the hue angle, the data were first transformed to the CIE XYZ colour space. CIE XYZ is a standard colour space representing the colours that a person with average colour vision can experience [101]. The reference data were spectrally convolved with the XYZ

colour matching functions [401]. The spectral convolution was applied directly to $R_{rs}$, since $R_{rs}$ represents the true colour of the water [283]. For the smartphone data, transformation matrices calculated from the smartphone camera SRFs (Section 4.B) were used [402, 403]. These matrices are given in Equations (4.3) and (4.4). The uncertainties on the matrix elements were not included since this would require a full re-analysis of the raw SRF data [404], which is outside the scope of this work. The resulting colours were relative to an E-type (flat spectrum, $x = y = 1/3$) illuminant.

$$\mathbf{M}_{RGB\to XYZ}^{iPhone\ SE} = \begin{bmatrix} 0.5709 & 0.2452 & 0.1839 \\ 0.3760 & 0.4346 & 0.1894 \\ 0.0439 & 0.0913 & 0.8648 \end{bmatrix} \tag{4.3}$$

$$\mathbf{M}_{RGB\to XYZ}^{Galaxy\ S8} = \begin{bmatrix} 0.5611 & 0.1451 & 0.2938 \\ 0.3944 & 0.2391 & 0.3666 \\ 0.0231 & 0.0416 & 0.9353 \end{bmatrix} \tag{4.4}$$

From XYZ, the chromaticity $(x, y)$ and hue angle $\alpha$ were calculated as shown in Equations (4.5) and (4.6). Chromaticity is a normalisation of the XYZ colour space that removes information on brightness [101]. The FU index was determined from $\alpha$ using a look-up table [10, 306]. The uncertainties in $R_{rs}$ were propagated analytically into XYZ and $(x, y)$, as described in Section 4.A. However, further propagation into $\alpha$ was not feasible, since the linear approximation of Equation (4.6) breaks down near the white point $(x, y) = (1/3, 1/3)$, especially with highly correlated $x$ and $y$ [405].

$$x = \frac{X}{X + Y + Z} \qquad y = \frac{Y}{X + Y + Z} \tag{4.5}$$

$$\alpha = \arctan2\,(y - 1/3, x - 1/3) \mod 2\pi \tag{4.6}$$

### 4.2.5 Replicate analysis

The Galaxy S8 data were taken in sets of 10 sequential replicates per image (Section 4.2.1). The variability between these replicates was analysed to assess the uncertainty in smartphone data.

The processing chain described in Section 4.2.3 was applied to each image from each set, resulting in 10 measurements per channel of $L_u$, $L_{sky}$, and $L_d$. $R_{rs}$ was calculated from each combination of images, resulting in 1 000 values. From these, the band ratios, $\alpha$, and FU were calculated.

The relative uncertainty in $L_u$, $L_{sky}$, $L_d$, $R_{rs}$, and the band ratios was estimated through the coefficient of variation $\frac{\sigma}{\mu}$, $\sigma$ being the standard deviation and $\mu$ the mean value. Because $\alpha$ and FU have arbitrary zero-points, relative uncertainties are not applicable to them, and $\sigma$ was instead used to estimate the absolute uncertainty.

### 4.2.6 Match-up analysis

Simultaneous data taken with the various sensors were matched up and compared. There were 27 pairs of iPhone SE and Galaxy S8 images, taken on average 50 s apart. On the ferry, which

had an average speed of 8 km/h, a 50 s delay corresponded to a distance along the transect of approximately 120 m. The smartphone images were also matched to reference spectra taken within a 10-minute time frame, resulting in 1–41 reference spectra per match-up. The reference $R_{rs}$ spectra were convolved to the smartphone RGB bands by first convolving the reference radiances [283]. For match-ups with multiple reference spectra per smartphone image, the median value of each variable in the reference spectra was used, with the standard deviation as an estimate for the uncertainty. For match-ups with a single reference spectrum per smartphone image, the uncertainty was instead estimated as the median uncertainty on the multiple-spectrum match-ups, for each variable. Match-up reference spectra with large uncertainties, for example relative uncertainties of >10% in $R_{rs}$, were not discarded because these represent common measurement scenarios.

The match-up data were compared using the metrics shown in Equations (4.7)–(4.10). Here $P, Q$ are any two data sets with elements $p_i, q_i$; $\mathrm{cov}(P, Q)$ is their covariance; $\sigma_P, \sigma_Q$ are the standard deviations in $P$ and $Q$, respectively; $\mathrm{Med}_i$ is the median evaluated over the indices $i$; and sgn is the sign function. The RGB channels were treated as separate samples, as were the three band ratios.

$$r = \frac{\mathrm{cov}(P, Q)}{\sigma_P \sigma_Q} \tag{4.7}$$

$$\mathcal{M} = \mathrm{Med}_i \left( |q_i - p_i| \right) \tag{4.8}$$

$$\zeta = \exp \left[ \mathrm{Med}_i \left( \left| \ln \frac{q_i}{p_i} \right| \right) - 1 \right] \tag{4.9}$$

$$\mathcal{B} = \mathrm{sgn} \left[ \mathrm{Med}_i \left( \ln \frac{q_i}{p_i} \right) \right] \times \left[ \exp \left( \left| \mathrm{Med}_i \left( \ln \frac{q_i}{p_i} \right) \right| \right) - 1 \right] \tag{4.10}$$

The Pearson correlation $r$ and median absolute deviation $\mathcal{M}$ are well-known [406, 407]. The median symmetric accuracy $\zeta$ and signed symmetric percentage bias $\mathcal{B}$, both expressed as a percentage, are recent introductions, which we chose to use for their robustness, symmetry, and ability to span multiple orders of magnitude in the data [406]. $r$ expresses the degree of linear correlation between variables, from −1 to 1, but is sensitive to outliers and the data range. $\mathcal{M}$ and $\zeta$ measure the typical random error or dispersion between variables in absolute and relative terms, respectively. Both are robust to outliers. $\mathcal{B}$ is similar to $\zeta$ but measures the bias towards over- or underestimation. The covariance, standard deviations, and median calculated in $r$ and $\mathcal{M}$ were weighted by $w_i = \frac{1}{\sigma_{p_i}^2 + \sigma_{q_i}^2}$. $\zeta$ and $\mathcal{B}$ are unweighted.

The FU indices were also compared by the number of matches [395, 407], considering both full ($\Delta\mathrm{FU} = 0$) and near-matches ($\Delta\mathrm{FU} \leq 1$). The typical uncertainty on human observations is 1 FU [280].

5–95% confidence intervals (CIs) on the metrics were estimated by bootstrapping over pairs of $(p_i, q_i)$, and $w_i$ if applicable. Bootstrapping involves randomly resampling the data with replacement, mimicking the original sampling process [408]. This was necessary to account for the relatively small size of our data set, which increases the effects of outliers, even on robust metrics like $\mathcal{M}$ or $\zeta$. The bootstraps were evaluated with 9 999 resamples, sufficient to obtain consistent results matching the analytical formula for CIs on unweighted $r$ to 4 decimals [408].

Some data were also compared through a linear regression ($y = ax + b$ with parameters $a, b$), to convert data to the same units or account for normalisation differences. The regres-

sion was done through the `scipy.odr` function for orthogonal distance regression, which minimises differences and accounts for weights on both axes. The same process was used to fit a power law ($y = ax^b$) in the JPEG data comparison (Section 4.3.4).

## 4.3 Results

### 4.3.1 Replicate analysis

The Galaxy S8 replicate analysis showed that among the radiances, $L_u$ had the largest relative variability with a quartile range (QR, the 25–75% percentile range of variability among the sets of replicate observations) of 1.8–5.8%, followed by $L_{sky}$ with 1.1–3.4%, and $L_d$ with 0.4–1.2% (Figure 4.4). $L_u$ and $L_{sky}$ were affected primarily by cloud and wave movement, shaking of the camera, and movement of the ferry on 3 July. Therefore, the variability in $L_u$ and $L_{sky}$ was largely methodological in nature, as discussed further in Section 4.4.1. Since $L_d$ was measured on a bright, stable gray card, it was not affected by the above factors, and its variability best represented the radiometric stability of the smartphone camera.

The RGB $R_{rs}$ varied by 1.9–8.1%, while the $R_{rs}$ band ratios only varied by 0.5–1.9%. The difference can be explained by correlations between channels. For example, wave movements between successive images affected all three RGB channels of $L_u$ equally, changing the individual $R_{rs}$ values, but having little effect on their ratios. The same held true for other environmental variations and camera stability issues.

Finally, there was a variability in hue angle $\alpha$ of 2.1°–6.8° and in FU index of 0.19–0.62 FU. The variability distributions of $\alpha$ and FU index did not have the same shape because the hue angle difference between successive FU indices varies greatly.

The variability between replicates represents the typical uncertainty associated with random effects on our data. However, there are some caveats. First, systematic effects such as an error in $R_{ref}$ would affect successive measurements equally, and not cause random variations. Second, the uncertainty in individual images may be larger due to spatial structures, which the uncertainty propagation described in Section 4.2.3 does account for. Both of these issues explain differences between the replicate and propagated uncertainties in our data. For

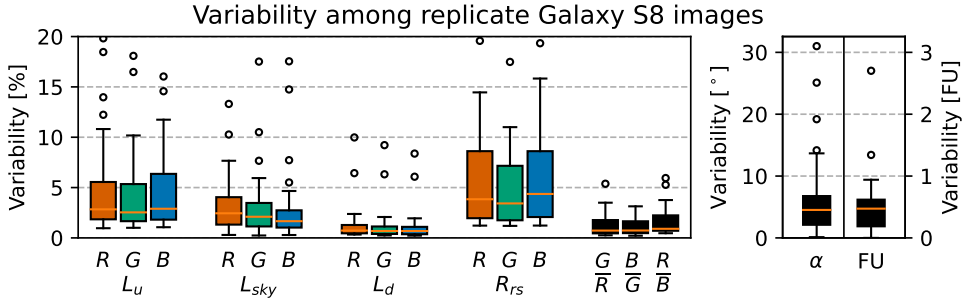


Figure 4.4: Variability in radiance, $R_{rs}$, and colour between replicate Galaxy S8 images. The boxes show the distribution, among 27 individually processed sets of 10 replicates, of the variability between replicate images. The orange lines indicate the medians, the boxes span the quartile range (QR), the whiskers extend to 1.5 times the QR, and circles indicate outliers. Up to two outliers per column fell outside the y-axis range.

example, the propagated uncertainty in individual images was 6.6–9.0% for RGB $R_{rs}$ and 4.5–7.0% for the band ratios. While the exact uncertainties will differ between campaigns, sites, and even smartphones, the trends seen here can be generalised.

As a point of comparison, the uncertainty QRs for the spectrally convolved WISP-3 data in the Galaxy S8 match-up (Section 4.3.3), were 4.2–38% in $L_u$, 4.8–14% in $L_{sky}$, 2.5–30% in $E_d$, 2.6–7.2% in RGB $R_{rs}$, 0.7–2.9% in $R_{rs}$ band ratios, 0.4°–2.8° in $\alpha$, and 0–0.46 in FU. While the Galaxy S8 and WISP-3 variability cannot be compared 1:1 due to differences in data acquisition and processing and in the uncertainty estimation, the order of magnitude of the uncertainties in the Galaxy S8 and WISP-3 reference data was the same.

### 4.3.2 Smartphone comparison

There was a strong correlation, $r = 0.94$ (CI 0.90, 0.96), between the iPhone SE and Galaxy S8 radiances (Figure 4.5). Due to differences in exposure settings, both cameras measured radiance in different, arbitrary units (a.u.). After re-scaling the Galaxy S8 data through a linear regression (Section 4.2.6), the median absolute deviation was $\mathcal{M} = 0.39$ (CI 0.29, 0.52) in iPhone SE units and the median symmetric accuracy was $\zeta = 6.9\%$ (CI 5.1%, 8.7%). The value of $\zeta$ was comparable to the variability between replicate images (Section 4.3.1).

The $R_{rs}$ match-ups between the two smartphones, in both RGB (Figure 4.6) and band ratios (Figure 4.7), showed excellent agreement. The data were strongly correlated, with $r = 0.98$ (CI 0.95, 0.99) for RGB and $r = 0.99$ (CI 0.99, 1.00) for band ratio $R_{rs}$. The typical difference in RGB $R_{rs}$ was $\mathcal{M} = 0.0010$ (CI 0.0005, 0.0013) $\text{sr}^{-1}$ or $\zeta = 5.5\%$ (CI 3.8%, 8.2%). For band ratios, the typical difference was $\mathcal{M} = 0.032$ (CI 0.026, 0.035), unitless, and $\zeta = 2.9\%$ (CI 2.3%, 3.7%). Both values of $\zeta$ are consistent with Section 4.3.1, as is the observation that band ratios are more reproducible than RGB $R_{rs}$. Finally, the signed symmetric percentage bias in RGB $R_{rs}$, $\mathcal{B} = -2.7\%$ (CI −7.0%, −1.8%), was smaller than the typical uncertainty. There was no significant offset in the band ratios, with $\mathcal{B} = -1.1\%$ (CI −1.8%, +0.7%).

The agreement in $\alpha$ and FU was poorer but still similar to the expected uncertainties (Figure 4.8). The typical difference was $\mathcal{M} = 8.3°$ (CI 5.0°, 11°) in $\alpha$ and $\mathcal{M} = 1$ (CI 0, 2) in FU index. 33% (CI 15%, 48%) of the match-up pairs had the same FU index, 59% (CI 37%, 74%) had a difference $\Delta\text{FU} \leq 1$. The wide CIs are due to the relatively small number ($N = 27$) of match-ups. The data did not span the full range of $\alpha$, but were mostly concentrated into two clusters, around 50° (FU 14–16, greenish brown) and 90° (FU 8–9, bluish green). Interestingly, while the 90° cluster was centred roughly on the 1:1 line, the 50° cluster fell entirely underneath it. However, due to the small $N$ and the uncertainties on the data, it is difficult to say whether this was significant.

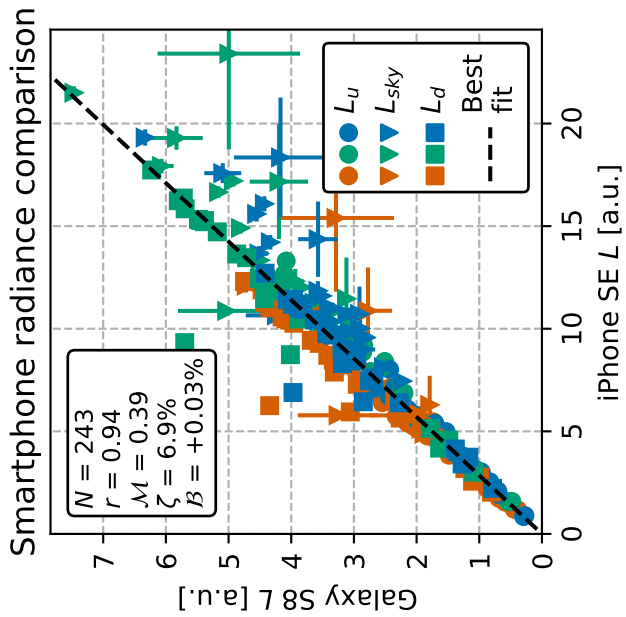


Figure 4.5: Comparison between iPhone SE and Galaxy S8 radiance measurements. The axes are in different units due to differences in exposure settings. The RGB channels are shown in their respective colours, with different symbols for $L_u$, $L_{sky}$, and $L_d$. The statistics in the text box are relative to the regression line.

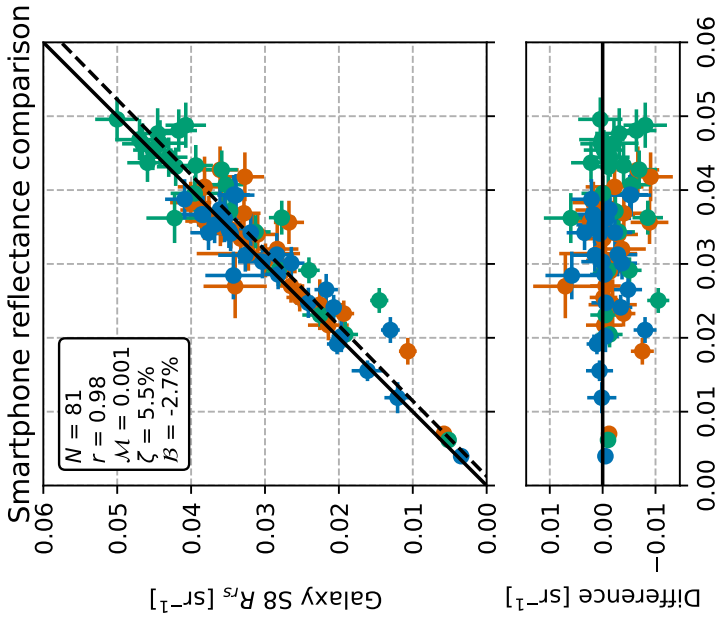


Figure 4.6: Comparison between iPhone SE and Galaxy S8 $R_{rs}$ measurements in the RGB bands. The solid line corresponds to a 1:1 relation, the dashed line is the best-fitting linear regression line. The statistics in the text box are based on a 1:1 comparison, as are the differences in the lower panel.

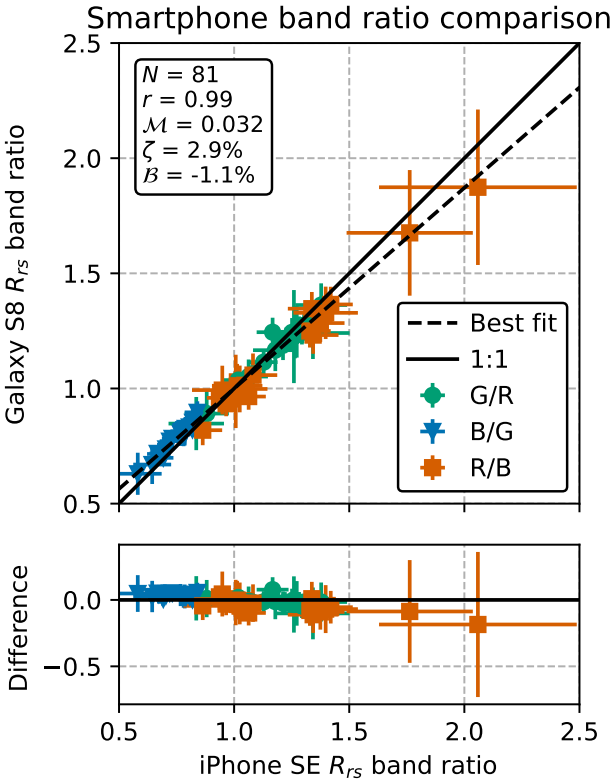


Figure 4.7: Comparison between iPhone SE and Galaxy S8 $R_{rs}$ band ratios. The solid line corresponds to a 1:1 relation, the dashed line is the best-fitting linear regression line. The statistics in the text box are based on a 1:1 comparison, as are the differences in the lower panel.

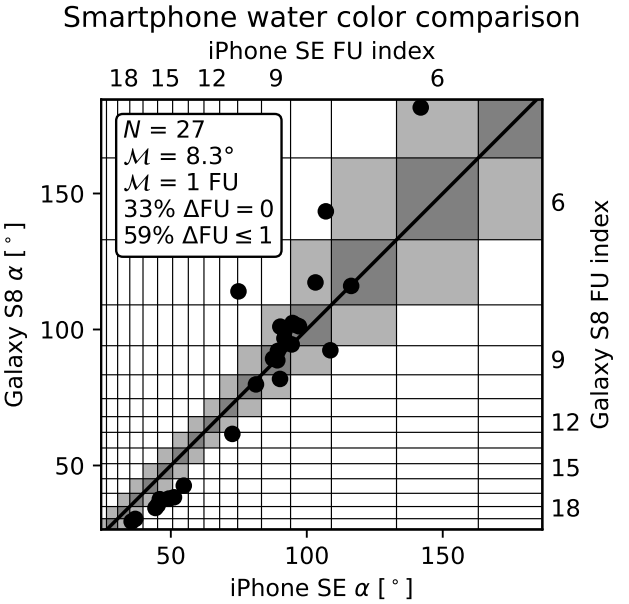


Figure 4.8: Comparison between iPhone SE and Galaxy S8 measurements of hue angle and FU index. The solid line corresponds to a 1:1 relation. The dark gray squares indicate a full FU match, the light gray ones a near-match. Accurate uncertainties on individual points could not be determined (Section 4.2.4). The statistics in the text box are based on a 1:1 comparison.

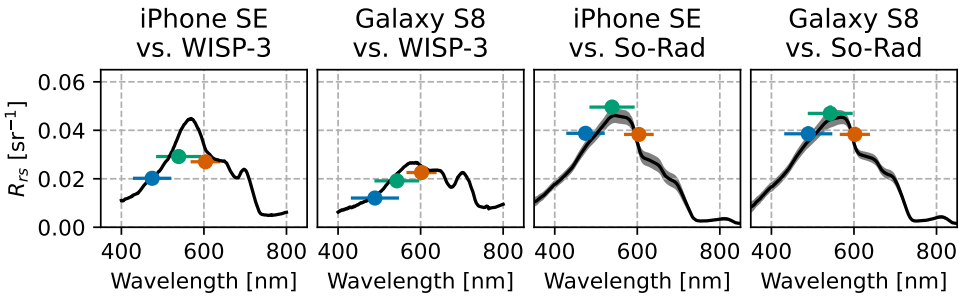


Figure 4.9: Examples of smartphone vs. reference $R_{rs}$ match-ups at different stations. The solid lines show the reference spectrum, with uncertainties in gray. The RGB dots show the smartphone data, with error bars indicating the effective bandwidth (horizontal) and $R_{rs}$ uncertainty (vertical). In some panels, the vertical error bars are smaller than the data point size.

### 4.3.3 Smartphone vs. reference comparison

A total of 72 pairs of smartphone vs. reference match-up spectra were analysed, four of which are shown in Figure 4.9. There were 27 match-ups between the WISP-3 and each smartphone and 9 between the So-Rad and each smartphone. Except for the normalisation difference that was also present between the So-Rad and WISP-3 (Figure 4.2, discussed in Section 4.4.3), there was good agreement between the instruments (Figure 4.9).

The full statistics of the match-up analysis are given in Table 4.1. The correlation between smartphone and reference radiance was $r \geq 0.71$ in all pairs of instruments (Figure 4.10). The median symmetric accuracy $\zeta$ ranged between 12–19%, larger than the typical uncertainties and the value from the smartphone vs. smartphone comparison. This larger difference in observed radiance is not surprising, since the smartphone vs. reference match-ups typically differed more in time and location than the smartphone vs. smartphone match-ups. No significant differences in the match-up statistics between the individual RGB bands were found.

The RGB $R_{rs}$ data were strongly correlated between smartphone and reference sensors ($r \geq 0.94$ for the WISP-3) and showed a relatively small dispersion, although with a normalisation difference in the WISP-3 comparisons (Figure 4.11), similar to that between the WISP-3 and So-Rad data (Figure 4.2). To negate the normalisation issue, the smartphone data were re-scaled based on a linear regression (Section 4.2.6) for the smartphone vs. WISP-3 RGB $R_{rs}$ comparison. The So-Rad and smartphone data were compared 1:1. The typical differences in $R_{rs}$, then, were on the order of $10^{-3}$ sr$^{-1}$ for the So-Rad and $10^{-4}$ sr$^{-1}$ for the WISP-3, differing mostly due to their different ranges. The difference in range of $R_{rs}$ also decreased the correlation coefficient $r$ for the So-Rad comparisons. In the four smartphone vs. reference $R_{rs}$ comparisons, $\zeta$ was between 9–13%, twice the value seen in the smartphone vs. smartphone comparison but similar to the differences between smartphone and reference radiances.

The agreement between smartphone and reference $R_{rs}$ band ratios was better than the agreement in RGB $R_{rs}$ (Figure 4.12). In all four band ratio comparisons, the correlation was near-perfect ($r \geq 0.97$), and the typical differences ($1.1\% \leq \zeta \leq 3.8\%$) were consistent with the uncertainties in the data. The WISP-3 normalisation difference did not affect this comparison since it divided out.

The agreement in $\alpha$ and FU was not as good as that in $L$ and $R_{rs}$, like in the smartphone

| | | $N$ | $r$, $\mathcal{M}(\alpha)$ | $\mathcal{M}$ | $\zeta$, $\Delta$FU = 0 | $\mathcal{B}$, $\Delta$FU $\leq$ 1 |
|---|---|---|---|---|---|---|
| | | **WISP-3** | | | | |
| **iPhone SE** | $L$ | 162 | 0.71 | 0.009 | 19% | −7.8% |
| | | | (0.57, 0.80) | (0.007, 0.010) | (15%, 24%) | (−11%, +0.21%) |
| | $R_{rs}$ | 81 | 0.97 | 0.0014 sr$^{-1}$ | 22% | −22% |
| | | | (0.91, 0.99) | (0.0001, 0.0063) | (18%, 27%) | (−28%, −20%) |
| | $R_{rs}$ | 81 | 0.97 | 0.0004 sr$^{-1}$ | 9.7% | +1.9% |
| | (regr.) | | (0.90, 0.99) | (0.0001, 0.0025) | (6.7%, 13%) | (−1.1%, +4.8%) |
| | Band | 81 | 0.98 | 0.013 | 1.9% | +0.25% |
| | ratios | | (0.97, 0.99) | (0.009, 0.019) | (1.2%, 2.7%) | (−0.61%, +0.87%) |
| | $\alpha$, FU | 27 | 9.4° | 1 FU | 26% | 59% |
| | | | (6.3°, 12°) | (1, 2) | (7.4%, 41%) | (37%, 74%) |
| **Galaxy S8** | $L$ | 162 | 0.75 | 0.009 | 19% | −3.2% |
| | | | (0.66, 0.83) | (0.007, 0.011) | (15%, 24%) | (−7.5%, +2.5%) |
| | $R_{rs}$ | 81 | 0.94 | 0.0025 sr$^{-1}$ | 31% | −31% |
| | | | (0.75, 0.98) | (0.0013, 0.0072) | (24%, 35%) | (−36%, −25%) |
| | $R_{rs}$ | 81 | 0.93 | 0.0007 sr$^{-1}$ | 13% | +5.9% |
| | (regr.) | | (0.70, 0.97) | (0.0005, 0.0041) | (9.2%, 14%) | (+3.2%, +11%) |
| | Band | 81 | 0.98 | 0.010 | 1.7% | +0.04% |
| | ratios | | (0.96, 0.99) | (0.007, 0.012) | (1.1%, 2.0%) | (−0.59%, +0.68%) |
| | $\alpha$, FU | 27 | 16° | 2 FU | 19% | 48% |
| | | | (11°, 21°) | (1, 4) | (3.7%, 33%) | (26%, 63%) |
| | | **So-Rad** | | | | |
| **iPhone SE** | $L$ | 54 | 0.87 | 0.006 | 13% | −4.7% |
| | | | (0.75, 0.94) | (0.004, 0.007) | (8.4%, 16%) | (−12%, −0.01%) |
| | $R_{rs}$ | 27 | 0.70 | 0.004 sr$^{-1}$ | 13% | +12% |
| | | | (0.49, 0.86) | (0.003, 0.005) | (9.9%, 16%) | (+6.5%, +14%) |
| | Band | 27 | 0.97 | 0.013 | 3.8% | +0.82% |
| | ratios | | (0.95, 0.98) | (0.006, 0.042) | (0.98%, 5.6%) | (−1.3%, +1.6%) |
| | $\alpha$, FU | 9 | 11° | 1 FU | 11% | 89% |
| | | | (6.1°, 13°) | (1, 1) | (0%, 33%) | (34%, 100%) |
| **Galaxy S8** | $L$ | 54 | 0.83 | 0.005 | 12% | −5.4% |
| | | | (0.69, 0.93) | (0.003, 0.007) | (9.2%, 16%) | (−12%, +1.4%) |
| | $R_{rs}$ | 27 | 0.75 | 0.003 sr$^{-1}$ | 8.5% | +6.4% |
| | | | (0.49, 0.87) | (0.002, 0.005) | (5.5%, 13%) | (−0.08%, +9.1%) |
| | Band | 27 | 0.99 | 0.004 | 1.1% | +0.36% |
| | ratios | | (0.98, 0.99) | (0.003, 0.009) | (0.42%, 2.4%) | (−0.35%, +0.45%) |
| | $\alpha$, FU | 9 | 16° | 1 FU | 11% | 56% |
| | | | (12°, 23°) | (1, 2) | (0%, 33%) | (11%, 78%) |

Table 4.1: Summary of the smartphone vs. reference match-up analysis. The values between parentheses indicate the 5–95% CI determined from bootstrapping. $N$ is the number of matching observations; the other metrics are described in Section 4.2.6. $\mathcal{M}(L)$ is in units of W m$^{-2}$ nm$^{-1}$ sr$^{-1}$. For the WISP-3, $R_{rs}$ was compared 1:1 and with a linear regression (regr.).

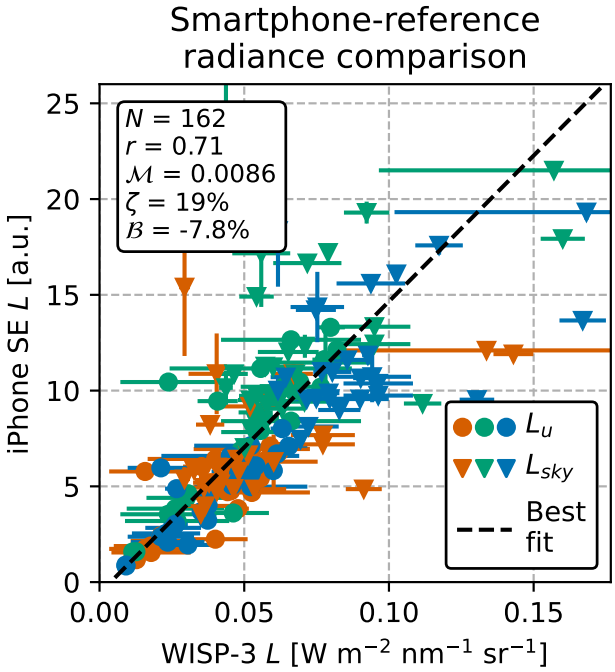


Figure 4.10: Comparison between iPhone SE and spectrally convolved WISP-3 radiance measurements. The RGB channels are shown in their respective colours, with different symbols for $L_u$ and $L_{sky}$. The statistics in the text box are relative to the regression line. We note that this regression line cannot be used as a general absolute radiometric calibration for the iPhone SE due to the arbitrary choice of exposure settings.

intercomparison (Section 4.3.2). For each smartphone, there were only $N = 27$ WISP-3 match-ups and even fewer So-Rad ones, making the CIs wide and the interpretation difficult. The difference between the WISP-3 and iPhone SE was slightly larger than in the smartphone comparison, at $\mathcal{M}(\alpha) = 9.4°$ (CI 6.3°, 12°) and $\mathcal{M}$(FU) = 1 (CI 1, 2). The Galaxy S8 and WISP-3 differed more, at $\mathcal{M}(\alpha) = 16°$ (CI 11°, 21°) and $\mathcal{M}$(FU) = 2 (CI 1, 4). The cause for this difference is unclear but may simply be an artifact of the small number of match-ups; the Galaxy S8 also differed more in RGB $R_{rs}$ but not in the band ratios. Both smartphones performed similarly in the FU match-ups, with 19–26% of the match-ups agreeing fully and 48–59% to within 1 FU, although these figures had particularly wide CIs.

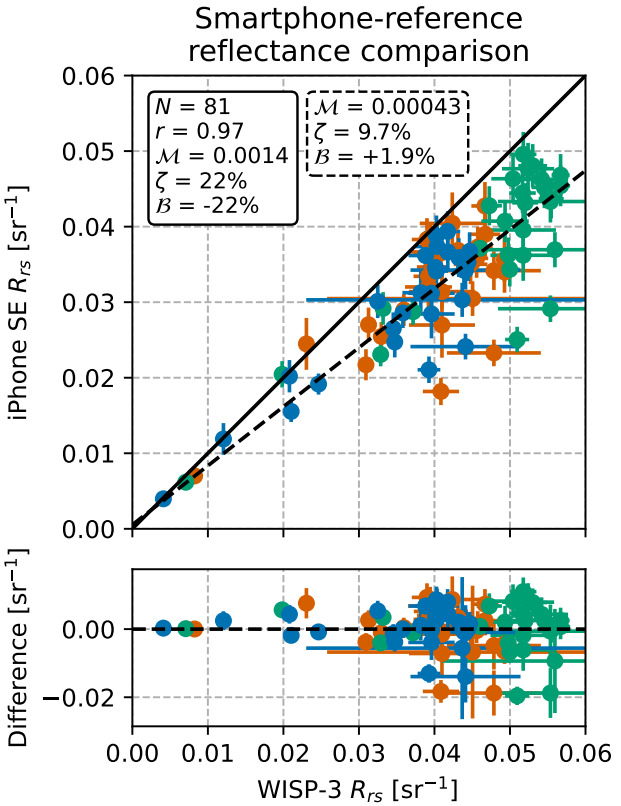


Figure 4.11: Comparison between iPhone SE and spectrally convolved WISP-3 $R_{rs}$ measurements in the RGB bands. The solid line corresponds to a 1:1 relation, the dashed line is the best-fitting linear regression line. The statistics in the solid-outline text box are based on a 1:1 comparison, those in the dashed-outline text box are based on the regression line. The differences in the lower panel are based on the regression line.

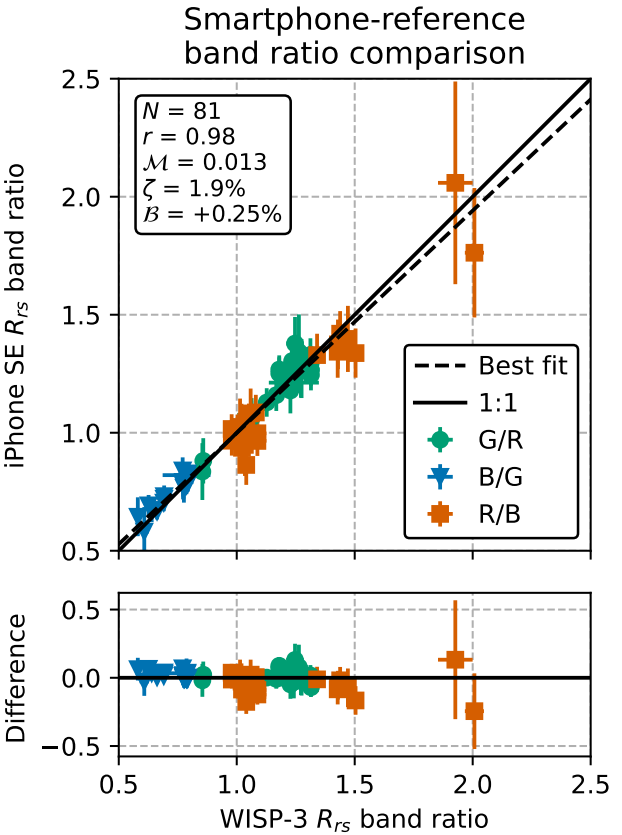


Figure 4.12: Comparison between iPhone SE and spectrally convolved WISP-3 $R_{rs}$ band ratios. The solid line corresponds to a 1:1 relation, the dashed line is the best-fitting linear regression line. The statistics in the text box are based on a 1:1 comparison, as are the differences in the lower panel.

### 4.3.4 JPEG data

28 sets of JPEG images from the iPhone SE, taken simultaneously with the RAW images, were analysed and compared to the RAW and reference data.

The relationship between JPEG and RAW radiances was highly non-linear (Figure 4.13). Each RGB channel had a different best-fitting power law, with exponents ranging from 0.477±0.005 for B to 0.949±0.013 for R. Due to differences between the RAW and JPEG data processing, the power law exponents are not equivalent to sRGB gamma exponents [281]. Figure 4.13 also shows the significant dispersion of the data around the power law curves. Comparing the RAW and re-scaled JPEG data yielded $\zeta$ ranging from 8.9% (CI 7.5%, 11%) for B to 38% (CI 29%, 43%) for R.

The JPEG vs. RAW $R_{rs}$ match-ups agreed better, particularly in the band ratios. The RGB $R_{rs}$ were strongly correlated, with $r = 0.92$ (CI 0.84, 0.97), but the JPEG data showed a large, consistent overestimation of $\mathcal{B} = +52\%$ (CI +39%, +59%). Comparing $R_{rs}$ through a linear regression removed this offset, although a significant dispersion of $\zeta = 15\%$ (CI 12%, 21%) remained. Conversely, the $R_{rs}$ band ratios were more similar with $r = 0.97$ (CI 0.95, 0.98), $\mathcal{M} = 0.033$ (CI 0.023, 0.042), and $\zeta = 4.9\%$ (CI 3.6%, 6.8%).

Finally, the agreement in $\alpha$ and FU was similar to the smartphone vs. smartphone and smartphone vs. reference comparisons. $\mathcal{M}$ was 11° (CI 3.6°, 14°) in $\alpha$ and 1 (CI 0, 2) in FU. 39% (CI 18%, 54%) of match-up pairs had the same FU index, while 61% (CI 39%, 75%) agreed to within 1 FU.



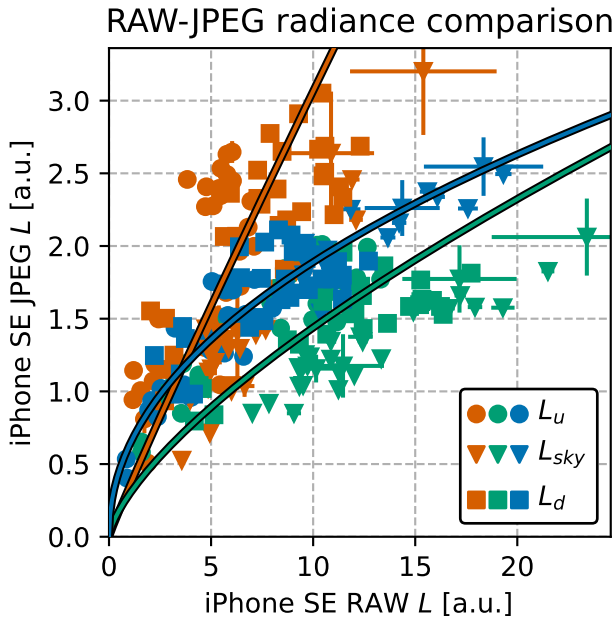


Figure 4.13: Comparison between RAW- and JPEG-based iPhone SE radiance measurements. The axes are in different units due to differences in exposure settings and normalisation. The RGB channels are shown in their respective colours, with different symbols for $L_u$, $L_{sky}$, and $L_d$. The coloured lines show the best-fitting power law for each channel.

The agreement between JPEG and reference data was notably worse than between RAW and reference data. While the JPEG vs. reference radiance match-ups appeared to follow a single linear relationship, rather than the multiple power laws seen in the JPEG vs. RAW comparison, they were only weakly correlated, with $r = 0.39$ (CI 0.22, 0.52) in the JPEG vs. WISP-3 comparison. The dispersion around the regression line was $\zeta = 31\%$ (CI 26%, 41%), 1.6× larger than for the RAW data.

The JPEG data consistently overestimated $R_{rs}$ compared to the references, and were widely dispersed. In the JPEG vs. WISP-3 comparison, $\mathcal{B} = +17\%$ (CI +10%, +19%), although this was reduced to $\mathcal{B} = +1.1\%$ (CI −7.3%, +5.8%) when comparing to a regression line instead of the 1:1 line, as in Section 4.3.3. However, the dispersion remained significant at $\mathcal{M} = 0.0039$ (CI 0.0018, 0.0047) $\text{sr}^{-1}$ or $\zeta = 21\%$ (CI 12%, 24%), with $\mathcal{M}$ 9× as much as for the RAW data, and $\zeta$ 2.1×.

The JPEG band ratios deviated from the WISP-3 by >2.5× as much as the RAW data, with $\mathcal{M} = 0.032$ (CI 0.023, 0.041) and $\zeta = 5.5\%$ (CI 3.7%, 6.4%). The So-Rad comparison showed a similarly stark difference. However, while this represents a serious reduction in performance, a typical difference of 5.5% is still relatively small.

It was only in $\alpha$ and FU that the JPEG vs. reference and RAW vs. reference agreements were similar. $\mathcal{M}(\alpha)$ in the JPEG vs. WISP-3 comparison was even marginally better at 7.1° (CI 5.0°, 11°); in the JPEG vs. So-Rad comparison it was 13° (CI 3.8°, 16°), almost identical to Table 4.1. $\mathcal{M}$(FU) and the fraction of FU matches were also similar, at $\mathcal{M}(\text{FU}) = 1$ (CI 1, 2), with 26% (CI 7.4%, 41%) full and 59% (CI 37%, 74%) partial FU matches between the JPEG and WISP-3 data. The agreement between JPEG and reference $\alpha$ and FU is discussed further in Section 4.4.3.

The effectiveness of an sRGB linearisation applied to the JPEG data, like in WACODI, was also investigated (Section 4.2.3). In $\alpha$ and FU, the main outputs from WACODI, the linearisation had very little effect. In the JPEG vs. WISP-3 comparison, $\mathcal{M}(\alpha)$ changed from 7.1° (CI 5.0°, 11°) originally to 7.0° (CI 5.4°, 9.4°) with linearisation. In radiance and $R_{rs}$, the linearisation made all comparison metrics significantly worse.

## 4.4 Discussion

### 4.4.1 Uncertainty

The uncertainty of the smartphone data as derived from replicate measurements (Section 4.3.1) is comparable to that of professional spectroradiometers. This was shown by the comparison with WISP-3 replicate measurements, which had a variability similar to, and in some cases larger than, the Galaxy S8. In general, the uncertainty from instrumental effects, excluding environmental factors and photon noise, in professional spectroradiometer data is around 1% [92]. In field data, the typical uncertainty is 1–7% [110]. The Galaxy S8 replicate variability, which was 0.4–1.2% ($L_d$), 1.1–3.4% ($L_{sky}$), and 1.8–5.8% ($L_u$), falls within this range.

The same is true for the smartphone $R_{rs}$ uncertainty, both in RGB (1.9–8.1%) and in band ratios (0.5–1.9%). $R_{rs}$ is typically measured with an uncertainty of 5% at blue and green wavelengths [88] and this is the target for satellites like PACE [86]. The 5% target also applies to narrower bands than the smartphone SRFs and to waters considerably darker than Lake Balaton, which increases the influence of sensor noise. The reduced uncertainty in band ratios is well-known and can be attributed to correlated uncertainties dividing out [146].

Propagated into the mineral suspended sediment (MSS) algorithm described in [34], for R/B ranging from 1.0–1.4, a 2% uncertainty in R/B results in a relative MSS uncertainty of only 1%. In the chl-a algorithm from [323], a 2% uncertainty in B/G induces a relative chl-a uncertainty of 9%. This level of uncertainty is well within the desired limits for many end users [88].

Finally, the uncertainty of the Galaxy S8 $\alpha$ (2.1°–6.8°) and FU index (0.19–0.62 FU) estimates is similar to the uncertainty of satellite and human measurements as well as the existing EyeOnWater app. Through propagation from $R_{rs}$, Pitarch et al. found uncertainties on SeaWiFS-derived $\alpha$ of 6°–18° [10], although it is difficult to compare these values due to the vastly different water types examined. Furthermore, propagated and replicate-based uncertainty estimates may vary significantly due to differences in sensitivity to various factors (Section 4.3.1). A more representative comparison point is the standard deviation of 3.15° among replicate EyeOnWater observations by Malthus et al. [275], which falls squarely within the range found in this work. The similarity in uncertainty is interesting because EyeOnWater is based on JPEG data, not RAW. However, since we did not take replicate JPEG images, a direct comparison in uncertainty between JPEG and RAW could not be made. The accuracy of JPEG and RAW data, including $\alpha$ and FU index, is compared in Section 4.4.3. The uncertainty of 0.19–0.62 FU is 5.3–1.6× better than human measurements, which have a typical uncertainty of 1 FU with perfect colour vision [280].

Since the use of RAW data eliminates virtually all smartphone-specific sources of uncertainty [281], the primary remaining sources are those that apply to all (spectro)radiometers as well as environmental factors. For a thorough overview of the former, we refer the reader to [106, 110]; for the latter, to [88]. Read-out noise, thermal dark current, and digitisation noise are negligible for well-lit smartphone images [281]. Since $L_d$ was measured on a stable target, its variability of 0.4–1.2% between replicates can be ascribed mostly to sensor noise (Section 4.3.1). Sensor noise scales with the square root of the number of photons, so the induced uncertainty will be larger in darker conditions such as overcast days, highly absorbing waters, and low solar elevation angles. In practice, smartphone observations under dark conditions will require longer exposure times or multiple images to attain similar levels of uncertainty. The impact of sun glint, which is estimated from $L_{sky}$, on the uncertainty in $R_{rs}$ is also larger for darker waters. The sensitivity of smartphone cameras to temperature variations and polarisation is unknown, although the latter is expected to be negligible unless special fore-optics are used [284]. Because our data were gathered in a single 3-day campaign, long-term sensor drift is unlikely to have had any effect; in general, sensor drift does not affect relative measurements like $R_{rs}$ and $\alpha$. Environmental factors, such as the patchy clouds that were present during our campaign (Figure 4.1), likely contributed the bulk of the uncertainty in $L_{sky}$ and $L_u$. These environmental factors also affected the reference measurements and are inherent to above-water radiometry.



### 4.4.2 Reproducibility

As there are hundreds of different smartphone models, reproducibility between devices is key. This is a major problem with HydroColor, as reported to us directly by users and as reported in the literature. For example, HydroColor measurements of $R_{rs}$ with different smartphones regularly differ by as much as 50% or 0.005 $sr^{-1}$ [121, 276]. This is largely due to the use of JPEG data, which are processed differently on every smartphone model, leading to a wide variety of errors and uncertainties that cannot be reliably corrected [281]. On the other

hand, Goddijn et al. reported smaller differences (4% ± 4%) between JPEG data from two high-quality digital cameras [323], suggesting that some of the problems may be specific to smartphones.

In Section 4.3.2, we showed that with RAW data and camera calibrations, excellent agreement and thus reproducibility between smartphones can be achieved. Near-simultaneous iPhone SE and Galaxy S8 measurements of radiance and $R_{rs}$ were nearly perfectly correlated ($r \geq 0.94$), and their dispersion could be explained by the uncertainties in the individual measurements. The typical difference in $R_{rs}$ was 0.0010 (CI 0.0005, 0.0013) $sr^{-1}$ or 5.5% (CI 3.8%, 8.2%), both major improvements over HydroColor. In fact, the dispersion in radiance between the two smartphones, $\zeta = 6.9\%$ (CI 5.1%, 8.7%), is only slightly larger than that between professional instruments in a similar experiment [92].

On the contrary, the smartphone JPEG processing algorithm was found to be poorly constrained and highly inconsistent between the RGB channels (Section 4.3.4). Moreover, the internal JPEG processing in the smartphone is re-tuned every time a camera session is started [281]. Combined, the differences between channels and between sessions highly limit the reproducibility of JPEG-based measurements of radiance and $R_{rs}$. As discussed below, white-balancing further reduces the reproducibility of JPEG-based $R_{rs}$ band ratios and hue angles. Finally, the JPEG processing algorithms differ between manufacturers, further reducing the reproducibility of JPEG data between devices [281]. Due to limitations in the SPECTACLE app in 2019, we did not collect Galaxy S8 JPEG data in this study, meaning a direct comparison between the RAW vs. RAW and JPEG vs. JPEG reproducibility could not be performed. Reproducing JPEG data from the RAW data was not possible, due to the aforementioned proprietary smartphone algorithms.

Differences in smartphone SRFs set some minor fundamental limits on the reproducibility between different cameras [331]. However, since most natural waters have broad and smooth spectra, this should only lead to minor differences. In theory, JPEG data do not have this problem because they are always in the sRGB colour space [285], but in practice the various proprietary colour algorithms cause larger differences in JPEG data than in RAW [281]. Furthermore, to account for illumination differences, JPEG data are white-balanced, changing the relative intensity of each channel. The re-normalisation directly reduces the accuracy of band ratio and hue angle measurements and is difficult to correct post-hoc [281, 391]. The white-balance setting may be locked between exposures [121, 323], but this does not guarantee consistency between different devices. Finally, due to differences in field-of-view between cameras, the central slice of 100 × 100 pixels does not always subtend the same solid angle. In future work, it may be advisable to use a constant solid angle rather than a constant pixel slice [121].

### 4.4.3 Accuracy

In Section 4.3.3, we compared smartphone and reference data to determine the accuracy of the smartphone data, but this comes with important caveats. While each instrument measured $L_u$ and $L_{sky}$, they did not do so in exactly the same way, having differences in field of view, spectral response, spectral resolution, and time and location. While the smartphones measured $L_d$ on a gray card, the references measured $E_d$ with a cosine collector. Due to these differences, the true 'ground truth' value of each measurand is not known [106, 409]. The reference data can be used to approximate the true values and achieve closure [234], but one must be aware of the uncertainties and systematic errors that may be present. Additionally,

one must exercise caution when comparing different metrics, such as the median symmetric accuracy $\zeta$ and the mean percentage deviation, which measure the same quantity but are calculated differently and on different data.

The WISP-3 and So-Rad $R_{rs}$ spectra were similarly shaped, but differently normalised (Section 4.2.2). Both were similar to spectra from previous work in shape, with the So-Rad more similar in magnitude [27, 264]. Normalisation differences and offsets have been seen in previous comparisons between the WISP-3 and other instruments [92, 126], so we felt confident in using a linear regression to re-scale $R_{rs}$ in the smartphone vs. WISP-3 comparisons. In fact, since each smartphone $R_{rs}$ measurement was based on three images from the same camera, rather than from three separate sensors like the WISP-3, and the gray card reference was independently verified, we can be more confident in the normalisation of the smartphone $R_{rs}$ than that of the WISP-3, at least for the particular unit and calibration settings used during our campaign. These results suggest that smartphones and other low-cost cameras could be used to provide closure when there is tension between data from professional instruments (Section 4.4.5).

Considering the above, the level of closure between smartphone and reference data was comparable to intercomparisons between professional radiometers to within a factor of 2–3. The dispersion $\zeta$ in radiance was relatively large at 12–19%, 2–3× that reported in a comparison of hyperspectral instruments on a single, stable platform [92], but as discussed previously, our radiance measurements were particularly affected by environmental factors and were taken at slightly different times and positions between instruments. Patchy clouds can increase the dispersion in radiance match-ups by a factor of 10 or more [126]. In $R_{rs}$, the typical difference was on the order of $10^{-4}$–$10^{-3}$ $sr^{-1}$ or 9–13%. Comparing hyperspectral radiometers, Tilstone et al. found mean differences between sensors on the order of $10^{-3}$ $sr^{-1}$ or 1–8%, with outliers up to 13% [114]. A comparison between WISP-3 and RAMSES sensors under cloudy conditions, similar to ours, found differences in $R_{rs}$ of 20–30% [126].

Most importantly, the smartphone and reference measurements of $R_{rs}$ band ratios agreed to within 2% in three out of four comparisons. The difference was only larger in the iPhone SE vs. So-Rad comparison, at 3.8%. Since band ratios are what most inversion algorithms for inherent optical properties and constituent concentrations are based on, it is the band ratio accuracy that determines the usefulness of smartphones as spectroradiometers. An accuracy and uncertainty of around 2% is well within most user requirements (Section 4.4.1).

The accuracy of the JPEG data was considerably worse (Section 4.3.4). In $R_{rs}$, the dispersion in the JPEG vs. WISP-3 comparison was 0.0039 (CI 0.0018, 0.0047) $sr^{-1}$ or 21% (CI 12%, 24%), which is in line with previous validation efforts for HydroColor [121, 276] and other JPEG-based methods [391, 394]. At $\mathcal{M} = 0.032$ (CI 0.023, 0.041) and $\zeta = 5.5\%$ (CI 3.7%, 6.4%), the same is true for the $R_{rs}$ band ratios [276]. The RAW data performed better on each of these metrics, most notably by 9× for the RGB $R_{rs}$ and 2.5× for the band ratios. These results do not completely invalidate previous JPEG-based methods nor HydroColor specifically [275], but demonstrate the significant increase in accuracy and decrease in uncertainty obtained by using RAW data.

The results for the hue angle $\alpha$ and FU index were less conclusive. While at first glance the dispersion of approximately 10° or 1 FU appears to be in line with previous studies [275, 285, 395], our measurement protocol (Section 4.2.1) did not follow the EyeOnWater protocol exactly, so the results cannot be compared directly to the aforementioned validation efforts. Additionally, our data only contained 27 smartphone vs. WISP-3 match-ups and even fewer for the So-Rad, with little diversity. Lastly, hue angles derived from narrow-band

multispectral satellite data have been shown to differ systematically by several degrees, up to 20° in extreme cases, compared to hue angles derived from hyperspectral data [10, 410]. This effect may also be present in the smartphone data and a correction term in the hue angle algorithm may be necessary [411]. This work used the original hue angle algorithm, which is based only on the SRFs [402], to enable a comparison between RAW and JPEG data and between the current study and previous works, particularly the WACODI algorithm [285]. We recommend that future work be done to investigate the magnitude of the hue angle bias in consumer camera data. Interestingly, there was little difference in accuracy between the RAW- and JPEG-derived hue angles and FU indices. It is unclear whether this is because the method is inherently robust to JPEG-induced errors [285], although Gao et al. have suggested that it is not [391]. More data, from more diverse waters, will be necessary to compare the accuracy of RAW- and JPEG-based hue angles and FU indices.

A potentially important source of systematic error is the 18% gray card. While the gray card used here did not deviate significantly from $R_{ref} = 18\%$ (Section 4.2.3), this may not be true in general. Since many smartphone radiometry projects are aimed at citizen scientists, who may purchase a wide variety of gray cards and may not always use them correctly, this presents an important possible source for error. Even a small difference in $R_{ref}$ can significantly bias $R_{rs}$. One possible solution to this problem is to issue or recommend standardised gray cards [391]. Characterising the most popular gray cards is another possibility [400], which may itself be done through citizen science. The use of relative quantities like band ratios negates this problem.



### 4.4.4 Recommendations

Based on previous work and the results discussed above, several recommendations can be made. Some are specific to smartphones, but most apply in general to above-water radiometry with consumer cameras since the cameras in most smartphones, digital cameras, UAVs, and webcams are extremely similar [281].

RAW data provide professional-grade radiometric performance and should be used whenever possible. Most consumer cameras now support this natively and many smartphone apps provide this capacity. Within the MONOCLE[17] project, a universal smartphone library for RAW acquisition and processing is in development. In the future, apps like HydroColor may simply import this library and use RAW data without further work from the user. The SPECTACLE Python library (Section 4.2.3) provides this functionality on PCs.

Few calibration data are necessary for above-water radiometry. Our processing pipeline contains bias and flatfield corrections, demosaics the data to the $RGBG_2$ channels, and normalises by the SRF spectral bandwidths (Figure 4.3). RAW files from virtually all cameras contain metadata describing the bias correction and demosaicking pattern. The flatfield correction requires additional data, which can be obtained through do-it-yourself methods [281], but may also be neglected at little cost in accuracy because its effect is typically small (0.2% for the iPhone SE and 1.6% for the Galaxy S8) in the central $100 \times 100$ pixels. The flat-field correction is more important in approaches that require a wider field-of-view like the multiple gray card approach [391]. The bandwidth normalisation divides out in the calculation of $R_{rs}$ and thus is only necessary to obtain accurate radiances. The SRFs are also required to accurately calculate $\alpha$ and convolve hyperspectral data in validation efforts, but may be approximated by standard profiles [121]. Low-cost smartphone spectrometers and other novel

[17] https://monocle-h2020.eu/

methods will soon enable on-the-fly SRF calibrations [284, 412].

As discussed in [281], it is important to accurately record exposure settings. In the current study, the exposure settings were not recorded, so it is not possible to combine our data with data from other studies, taken with different settings. The most important exposure settings are ISO speed and exposure time, which strongly affect the observed signal, but are not recorded accurately in the image metadata (EXIF). The settings must therefore be recorded by the user or the app. Since ISO speed does not affect the signal-to-noise ratio (SNR), a constant value maybe used. Longer exposure times increase the SNR but run the risk of saturation. Ideally, an automatic exposure time is determined and recorded for each image; if this is not possible, a single value may be used.

Algorithms to retrieve inherent optical properties from smartphone-based $R_{rs}$ measurements are best based on band ratios since they are the most precise, reproducible, and accurate. Algorithms based on absolute $R_{rs}$ in RGB [121, 391] are more susceptible to uncertainty and systematic errors. Because the RGB SRFs are broad and overlapping, some narrow spectral features like pigment absorption peaks cannot be distinguished, and retrieval algorithms require tuning to specific sites [34]. In edge cases where spectral features fall on wavelengths where SRFs vary significantly between devices, the reproducibility of retrieval algorithms between devices may also vary. For example, the iPhone SE and Galaxy S8 B-band SRFs differ greatly between 550–600 nm [281]. Algorithms that use spectrally distinct peaks, for example to retrieve chl-a concentrations, should be unaffected. Distinguishing between chl-a and CDOM, which both absorb in the B and G bands, may require a three-band algorithm that also estimates the backscattering coefficient $b_b$ from the R-band [228]. Alternative colour spaces like relative RGB [34, 386], hue-saturation-intensity [388], and CIE L*a*b* [230] are also worth exploring. Potential algorithms may be identified through spectral convolution of archival $R_{rs}$ spectra [283].



### 4.4.5 Outlook

The findings presented in this work extend to other methods for smartphone (spectro)radiometry and to most consumer cameras. This study was performed as a precursor to the field validation for the iSPEX 2 smartphone spectropolarimeter [284]. The uncertainty, accuracy, and reproducibility of iSPEX 2 data will be comparable to what was found in this study, although longer exposure times will be necessary to attain similar photon counts. The low uncertainty and high accuracy of the $R_{rs}$ band ratios is particularly promising since iSPEX 2 will measure hyperspectrally across the visible range, enabling many such algorithms. Also applicable to iSPEX 2 are some of the limitations found in this work, primarily the dependence on a gray card and the question of sensitivity in low-light conditions.

There is also potential for low-cost cameras, like webcams and UAV cameras, to augment professional spectroradiometers. Removal of the direct sun glint remains challenging, requiring assumptions about the spectrum and wave statistics [120, 210]. Low-cost camera images, taken simultaneously with the spectra, could be used to determine the wave statistics akin to [413] but for individual exposures. A similar system, which flags spectra if the associated image has saturated pixels, was already demonstrated in [414], and there are further opportunities for image-based anomaly detection. Finally, low-cost cameras can serve as simple validation checks for other sensors, for example to identify normalisation problems.

## 4.5 Conclusions

In this work, we have assessed the performance of smartphones as multispectral above-water radiometers. We have extended the existing smartphone-based approaches by using RAW data, processed through the SPECTACLE method for calibration of consumer cameras [281]. Using field data gathered under realistic observing conditions on and around Lake Balaton, we have analysed the uncertainty, reproducibility, and accuracy of above-water radiometry data taken with smartphone cameras. Furthermore, by comparing RAW and JPEG data, we have determined to what extent our new method improves upon existing work.

The uncertainty of the smartphone data, determined from replicate observations, was on the percent level and was comparable to professional radiometers. The typical uncertainty on $R_{rs}$ band ratios was 0.5–1.9%, leading to percent-level uncertainties in retrieved inherent optical properties and constituent concentrations. This level of uncertainty falls within the desired limits for many end users.

The reproducibility between smartphones was excellent, representing a significant improvement over existing methods, in some cases nearly tenfold. Any differences in the data between smartphones could be explained by measurement uncertainties.

The accuracy of smartphone data, as determined from match-ups with reference instruments, was comparable to professional instruments. The typical difference between smartphone and reference instruments was $10^{-4}$–$10^{-3}$ $\mathrm{sr}^{-1}$ or 9–13% in RGB $R_{rs}$, and 0.004–0.013 or 1.1–3.8% in $R_{rs}$ band ratios. These differences were an improvement of 9× and 2.5×, respectively, over JPEG data.

Based on the findings of this study, we recommend the use of RAW data for above-water radiometry with smartphones by professional and citizen scientists alike. We further recommend that retrieval algorithms be based on $R_{rs}$ band ratios rather than absolute RGB $R_{rs}$. Potential algorithms may be identified through spectral convolution of archival hyperspectral data. The conclusions and recommendations described above extend to other consumer cameras and to hyperspectral approaches like iSPEX 2. Future work should focus on determining the limitations of consumer cameras, primarily in terms of sensitivity, and exploring opportunities for complementary use of consumer cameras and professional spectroradiometers.

## 4.A Uncertainty propagation

As discussed above, significant correlations were found between channels in the smartphone data, as well as between images within one data set. To account for this, the inter-channel and inter-image covariances were incorporated into the uncertainty propagation.

### 4.A.1 Covariance and correlation

First, the mean radiances were combined into a single vector $\mathbf{L}$, containing 12 elements, corresponding to each channel in each image. The radiance vector $\mathbf{L}$ for the 07:47 data set is given in Equation (4.11). The elements of $\mathbf{L}$ are in analogue-digital units (ADU) $\mathrm{nm}^{-1}$. To save space, $\mathbf{L}$ is shown in its row vector form $\mathbf{L}^T$.

$$\begin{aligned}\mathbf{L}^T &= \begin{bmatrix} L_u(R) & L_u(G) & L_u(B) & L_u(G_2) & L_{sky}(R) & L_{sky}(G) & L_{sky}(B) & L_{sky}(G_2) & L_d(R) & L_d(G) & L_d(B) & L_d(G_2) \end{bmatrix} \\ &= \begin{bmatrix} 5.10 & 8.58 & 5.05 & 8.59 & 5.19 & 9.71 & 9.60 & 9.73 & 7.28 & 10.53 & 7.63 & 10.54 \end{bmatrix}\end{aligned} \tag{4.11}$$

The radiance vector $\mathbf{L}$ has a corresponding covariance matrix $\mathbf{\Sigma_L}$. The diagonal elements of $\mathbf{\Sigma_L}$ contain the variances of individual elements of $\mathbf{L}$, namely $\sigma^2_{L_u(R)}, \sigma^2_{L_u(G)}, \ldots$, while the off-diagonal elements contain the covariances, namely $\sigma_{L_u(R)L_u(G)}, \sigma_{L_u(R)L_u(B)}, \ldots$, all in units of ADU$^2$ nm$^{-2}$. The covariance matrix corresponding to Equation (4.11) is given in Equation (4.12). The rows and columns are in the same order as the elements of $\mathbf{L}$, with $\sigma^2_{L_u(R)}$ and $\sigma^2_{L_d(G_2)}$ in the top left and bottom right corners, respectively.

$$\mathbf{\Sigma_L} = \begin{bmatrix} 0.043 & 0.003 & 0.002 & 0.003 & 0.006 & 0.007 & 0.007 & 0.007 & 0.003 & 0.002 & 0.002 & 0.004 \\ 0.003 & 0.050 & 0.002 & 0.004 & 0.002 & 0.006 & 0.004 & 0.001 & 0.004 & 0.009 & 0.003 & 0.003 \\ 0.002 & 0.002 & 0.033 & 0.002 & -0.002 & -0.004 & 0.004 & -0.004 & 0.003 & 0.003 & 0.007 & 0.003 \\ 0.003 & 0.004 & 0.002 & 0.047 & 0.004 & 0.007 & 0.005 & 0.009 & 0.003 & 0.003 & 0.003 & 0.010 \\ 0.006 & 0.002 & -0.002 & 0.004 & 0.114 & 0.101 & 0.075 & 0.108 & -0.003 & -0.007 & -0.006 & -0.007 \\ 0.007 & 0.006 & -0.004 & 0.007 & 0.101 & 0.195 & 0.103 & 0.151 & -0.002 & -0.003 & -0.007 & -0.013 \\ 0.007 & 0.004 & 0.004 & 0.005 & 0.075 & 0.103 & 0.142 & 0.107 & -0.004 & -0.007 & 0.001 & -0.005 \\ 0.007 & 0.001 & -0.004 & 0.009 & 0.108 & 0.151 & 0.107 & 0.206 & -0.004 & -0.010 & -0.008 & -0.008 \\ 0.003 & 0.004 & 0.003 & 0.003 & -0.003 & -0.002 & -0.004 & -0.004 & 0.066 & 0.007 & 0.007 & 0.008 \\ 0.002 & 0.009 & 0.003 & 0.003 & -0.007 & -0.003 & -0.007 & -0.010 & 0.007 & 0.068 & 0.007 & 0.013 \\ 0.002 & 0.003 & 0.007 & 0.003 & -0.006 & -0.007 & 0.001 & -0.008 & 0.007 & 0.007 & 0.056 & 0.010 \\ 0.004 & 0.003 & 0.003 & 0.010 & -0.007 & -0.013 & -0.005 & -0.008 & 0.008 & 0.013 & 0.010 & 0.073 \end{bmatrix} \tag{4.12}$$

It is often easier to think in terms of correlation $r$, which is dimensionless and ranges from $-1$ to 1. The correlation between two quantities is simply their covariance normalised by their individual uncertainties: $r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$. $r = 0$ implies no correlation is present, while $r = 1$ and $r = -1$ imply a perfect positive and negative correlation, respectively. When applied to a covariance matrix, this yields a correlation matrix $\mathbf{R}$. The matrix $\mathbf{R_L}$ corresponding to Equation (4.12), is shown in Equation (4.13). The diagonal elements of $\mathbf{R_L}$ are always 1 and the matrix is always symmetric.

$$\mathbf{R_L} = \begin{bmatrix} 1.00 & 0.05 & 0.06 & 0.08 & 0.09 & 0.07 & 0.09 & 0.08 & 0.06 & 0.04 & 0.04 & 0.07 \\ 0.05 & 1.00 & 0.06 & 0.09 & 0.02 & 0.06 & 0.05 & 0.01 & 0.07 & 0.16 & 0.06 & 0.05 \\ 0.06 & 0.06 & 1.00 & 0.05 & -0.03 & -0.05 & 0.06 & -0.05 & 0.06 & 0.06 & 0.16 & 0.07 \\ 0.08 & 0.09 & 0.05 & 1.00 & 0.06 & 0.07 & 0.06 & 0.09 & 0.05 & 0.05 & 0.07 & 0.16 \\ 0.09 & 0.02 & -0.03 & 0.06 & 1.00 & 0.68 & 0.59 & 0.71 & -0.03 & -0.08 & -0.07 & -0.08 \\ 0.07 & 0.06 & -0.05 & 0.07 & 0.68 & 1.00 & 0.62 & 0.75 & -0.02 & -0.03 & -0.07 & -0.11 \\ 0.09 & 0.05 & 0.06 & 0.06 & 0.59 & 0.62 & 1.00 & 0.63 & -0.04 & -0.07 & 0.01 & -0.05 \\ 0.08 & 0.01 & -0.05 & 0.09 & 0.71 & 0.75 & 0.63 & 1.00 & -0.04 & -0.09 & -0.07 & -0.07 \\ 0.06 & 0.07 & 0.06 & 0.05 & -0.03 & -0.02 & -0.04 & -0.04 & 1.00 & 0.11 & 0.11 & 0.12 \\ 0.04 & 0.16 & 0.06 & 0.05 & -0.08 & -0.03 & -0.07 & -0.09 & 0.11 & 1.00 & 0.11 & 0.19 \\ 0.04 & 0.06 & 0.16 & 0.07 & -0.07 & -0.07 & 0.01 & -0.07 & 0.11 & 0.11 & 1.00 & 0.15 \\ 0.07 & 0.05 & 0.07 & 0.16 & -0.08 & -0.11 & -0.05 & -0.07 & 0.12 & 0.19 & 0.15 & 1.00 \end{bmatrix} \tag{4.13}$$

### 4.A.2 Averaging the G and $G_2$ channels

The G and $G_2$ channels in each image were averaged because these bands have the same spectral response function. This was done by multiplying $\mathbf{L}$ with a simple transfer matrix as shown in Equation (4.14); $\mathbf{M}$ is shown in Equation (4.15).

$$\mathbf{L}_{RGB} = \mathbf{M}\mathbf{L} \tag{4.14}$$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{4.15}$$

Since Equation (4.14) is a simple linear transformation, the propagation of $\mathbf{\Sigma}_{\mathbf{L}}$ to $\mathbf{\Sigma}_{\mathbf{L}_{RGB}}$ is a simple matrix multiplication with $\mathbf{M}$ and its transpose, as shown in Equation (4.16). In the special case that all covariances are 0, this reduces to the well-known sum-of-squares equation.

$$\mathbf{\Sigma}_{\mathbf{L}_{RGB}} = \mathbf{M}\mathbf{\Sigma}_{\mathbf{L}}\mathbf{M}^T \tag{4.16}$$

For illustration, the resulting radiance vector $\mathbf{L}_{RGB}$ and correlation matrix $\mathbf{R}_{\mathbf{L}_{RGB}}$ are shown in Equations (4.17) and (4.18), respectively.

$$\begin{aligned} \mathbf{L}_{RGB}^T &= \begin{bmatrix} L_u(R) & L_u(G) & L_u(B) & L_{sky}(R) & L_{sky}(G) & L_{sky}(B) & L_d(R) & L_d(G) & L_d(B) \end{bmatrix} \\ &= \begin{bmatrix} 5.10 & 8.59 & 5.05 & 5.19 & 9.72 & 9.60 & 7.28 & 10.54 & 7.63 \end{bmatrix} \end{aligned} \tag{4.17}$$

$$\mathbf{R}_{\mathbf{L}_{RGB}} = \begin{bmatrix} 1.00 & 0.09 & 0.06 & 0.09 & 0.08 & 0.09 & 0.06 & 0.07 & 0.04 \\ 0.09 & 1.00 & 0.07 & 0.05 & 0.09 & 0.07 & 0.08 & 0.19 & 0.09 \\ 0.06 & 0.07 & 1.00 & -0.03 & -0.05 & 0.06 & 0.06 & 0.09 & 0.16 \\ 0.09 & 0.05 & -0.03 & 1.00 & 0.74 & 0.59 & -0.03 & -0.10 & -0.07 \\ 0.08 & 0.09 & -0.05 & 0.74 & 1.00 & 0.67 & -0.03 & -0.10 & -0.07 \\ 0.09 & 0.07 & 0.06 & 0.59 & 0.67 & 1.00 & -0.04 & -0.08 & 0.01 \\ 0.06 & 0.08 & 0.06 & -0.03 & -0.03 & -0.04 & 1.00 & 0.15 & 0.11 \\ 0.07 & 0.19 & 0.09 & -0.10 & -0.10 & -0.08 & 0.15 & 1.00 & 0.17 \\ 0.04 & 0.09 & 0.16 & -0.07 & -0.07 & 0.01 & 0.11 & 0.17 & 1.00 \end{bmatrix} \tag{4.18}$$

### 4.A.3 Remote sensing reflectance

The remote sensing reflectance $R_{rs}$, in units of $\mathrm{sr}^{-1}$, was calculated from $L_u$, $L_{sky}$, and $L_d$ using Equation (4.2). This equation applies to each band individually, meaning that for example $R_{rs}(R)$ only depends on $L_u(R)$, $L_{sky}(R)$, and $L_d(R)$. However, since there were non-zero covariances between bands and between images, these were propagated through to $R_{rs}$.

$$R_{rs} = \frac{L_u - \rho L_{sky}}{\frac{\pi}{R_{ref}} L_d} \qquad \text{(4.2 rev.)}$$

While the reference reflectance $R_{ref}$ was assumed to be a constant $R_{ref} = 0.18$ in each band, it had its own uncertainty $\sigma_{R_{ref}} = 0.01$, which was also propagated through to $R_{rs}$. This was done by appending its uncertainty to $\mathbf{\Sigma}_{\mathbf{L}_{RGB}}$, as shown in block matrix form in Equation (4.19).

$$\mathbf{\Sigma}_{\mathbf{L}_{RGB}}^{R_{ref}} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{L}_{RGB}} & \mathbf{0} \\ \mathbf{0} & \sigma_{R_{ref}}^2 \end{bmatrix} \tag{4.19}$$

Since Equation (4.2) is not a simple linear transformation, a linear approximation was used in the uncertainty propagation, as shown in Equation (4.20). This is similar to Equation (4.16), but using the Jacobian matrix of Equation (4.2). This approximation is valid for functions that are locally well-approximated by a linear function.

$$\mathbf{\Sigma}_{\mathbf{R_{rs}}} = \mathbf{J}_{\mathbf{R_{rs}}} \mathbf{\Sigma}_{\mathbf{L}_{RGB}}^{R_{ref}} \mathbf{J}_{\mathbf{R_{rs}}}{}^{T} \tag{4.20}$$

The Jacobian matrix $\mathbf{J}_{\mathbf{R_{rs}}}$ contains all first-order derivatives of $R_{rs}$ in each band, as shown in Equation (4.21).

$$\mathbf{J}_{\mathbf{R_{rs}}} = \begin{bmatrix} \frac{\partial R_{rs}(R)}{\partial L_u(R)} & \frac{\partial R_{rs}(R)}{\partial L_u(G)} & \frac{\partial R_{rs}(R)}{\partial L_u(B)} & \frac{\partial R_{rs}(R)}{\partial L_{sky}(R)} & \frac{\partial R_{rs}(R)}{\partial L_{sky}(G)} & \frac{\partial R_{rs}(R)}{\partial L_{sky}(B)} & \frac{\partial R_{rs}(R)}{\partial L_d(R)} & \frac{\partial R_{rs}(R)}{\partial L_d(G)} & \frac{\partial R_{rs}(R)}{\partial L_d(B)} & \frac{\partial R_{rs}(R)}{\partial R_{ref}} \\ \frac{\partial R_{rs}(G)}{\partial L_u(R)} & \frac{\partial R_{rs}(G)}{\partial L_u(G)} & \frac{\partial R_{rs}(G)}{\partial L_u(B)} & \frac{\partial R_{rs}(G)}{\partial L_{sky}(R)} & \frac{\partial R_{rs}(G)}{\partial L_{sky}(G)} & \frac{\partial R_{rs}(G)}{\partial L_{sky}(B)} & \frac{\partial R_{rs}(G)}{\partial L_d(R)} & \frac{\partial R_{rs}(G)}{\partial L_d(G)} & \frac{\partial R_{rs}(G)}{\partial L_d(B)} & \frac{\partial R_{rs}(G)}{\partial R_{ref}} \\ \frac{\partial R_{rs}(B)}{\partial L_u(R)} & \frac{\partial R_{rs}(B)}{\partial L_u(G)} & \frac{\partial R_{rs}(B)}{\partial L_u(B)} & \frac{\partial R_{rs}(B)}{\partial L_{sky}(R)} & \frac{\partial R_{rs}(B)}{\partial L_{sky}(G)} & \frac{\partial R_{rs}(B)}{\partial L_{sky}(B)} & \frac{\partial R_{rs}(B)}{\partial L_d(R)} & \frac{\partial R_{rs}(B)}{\partial L_d(G)} & \frac{\partial R_{rs}(B)}{\partial L_d(B)} & \frac{\partial R_{rs}(B)}{\partial R_{ref}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{R_{ref}}{\pi L_d(R)} & 0 & 0 & \frac{-\rho R_{ref}}{\pi L_d(R)} & 0 & 0 & \frac{-R_{rs}(R)}{L_d(R)} & 0 & 0 & \frac{R_{rs}(R)}{R_{ref}} \\ 0 & \frac{R_{ref}}{\pi L_d(G)} & 0 & 0 & \frac{-\rho R_{ref}}{\pi L_d(G)} & 0 & 0 & \frac{-R_{rs}(G)}{L_d(G)} & 0 & \frac{R_{rs}(G)}{R_{ref}} \\ 0 & 0 & \frac{R_{ref}}{\pi L_d(B)} & 0 & 0 & \frac{-\rho R_{ref}}{\pi L_d(B)} & 0 & 0 & \frac{-R_{rs}(B)}{L_d(B)} & \frac{R_{rs}(B)}{R_{ref}} \end{bmatrix} \tag{4.21}$$

For illustration, the resulting $\mathbf{R_{rs}}$, its covariance matrix $\mathbf{\Sigma}_{\mathbf{R_{rs}}}$, and its correlation matrix $\mathbf{R}_{\mathbf{R_{rs}}}$ are shown in Equations (4.22) and (4.23), respectively. Evidently, there are strong correlations between the RGB channels in $\mathbf{R_{rs}}$. The uncertainty in $R_{ref}$ is particularly important; halving its value to $\sigma_{R_{ref}} = 0.005$ approximately halves the off-diagonal elements of $\mathbf{R}_{\mathbf{R_{rs}}}$.

$$\mathbf{R_{rs}}^{T} = \begin{bmatrix} R_{rs}(R) & R_{rs}(G) & R_{rs}(B) \end{bmatrix} = \begin{bmatrix} 0.039 & 0.045 & 0.036 \end{bmatrix} \tag{4.22}$$

$$\mathbf{\Sigma}_{\mathbf{R_{rs}}} = \begin{bmatrix} 9.0 & 5.5 & 4.4 \\ 5.5 & 7.5 & 5.1 \\ 4.4 & 5.1 & 6.6 \end{bmatrix} \times 10^{-6} \qquad \mathbf{R}_{\mathbf{R_{rs}}} = \begin{bmatrix} 1.00 & 0.67 & 0.57 \\ 0.67 & 1.00 & 0.72 \\ 0.57 & 0.72 & 1.00 \end{bmatrix} \tag{4.23}$$

### 4.A.4 Band ratios

The calculation of the G/R, B/G, and R/B $R_{rs}$ band ratios was straightforward and the Jacobian matrix approximation was used again. This is shown in Equation (4.24). We assumed

this approximation to be valid here due to the small range of our data and relatively small uncertainties in $\mathbf{R_{rs}}$.

$$\mathbf{J}_{\mathbf{R_{rs,br}}} = \begin{bmatrix} \frac{\partial(G/R)}{\partial R} & \frac{\partial(G/R)}{\partial G} & \frac{\partial(G/R)}{\partial B} \\ \frac{\partial(B/G)}{\partial R} & \frac{\partial(B/G)}{\partial G} & \frac{\partial(B/G)}{\partial B} \\ \frac{\partial(R/B)}{\partial R} & \frac{\partial(R/B)}{\partial G} & \frac{\partial(R/B)}{\partial B} \end{bmatrix} = \begin{bmatrix} -G/R^2 & 1/R & 0 \\ 0 & -B/G^2 & 1/G \\ 1/B & 0 & -R/B^2 \end{bmatrix} \tag{4.24}$$

The resulting band ratio vector $\mathbf{R_{rs,br}}$ for the 07:47 UTC data and its covariance and correlation matrices $\mathbf{\Sigma}_{\mathbf{R_{rs,br}}}$ and $\mathbf{R}_{\mathbf{R_{rs,br}}}$ are shown in Equations (4.25) and (4.26).

$$\mathbf{R_{rs,br}}^T = \begin{bmatrix} G/R & B/G & R/B \end{bmatrix} = \begin{bmatrix} 1.2 & 0.79 & 1.1 \end{bmatrix} \tag{4.25}$$

$$\mathbf{\Sigma}_{\mathbf{R_{rs,br}}} = \begin{bmatrix} 4.5 & -0.54 & -3.5 \\ -0.54 & 1.6 & -1.7 \\ -3.5 & -1.7 & 5.6 \end{bmatrix} \times 10^{-3} \qquad \mathbf{R}_{\mathbf{R_{rs,br}}} = \begin{bmatrix} 1.00 & -0.20 & -0.69 \\ -0.20 & 1.00 & -0.57 \\ -0.69 & -0.57 & 1.00 \end{bmatrix} \tag{4.26}$$

### 4.A.5 Chromaticity and hue angle

Since the colour space transformation from RGB to XYZ was a simple linear transformation, so was the uncertainty propagation, analogous to Equation (4.16). As discussed in the main paper, the uncertainties on the elements of the transformation matrices were ignored here. The results for the 07:47 data are shown in Equations (4.27) and (4.28). The resulting correlations were very strong, particularly for $R_{rs}(X)$ and $R_{rs}(Y)$, due to the spectral overlap between the XYZ colour-matching functions amplifying the existing correlations in $\mathbf{R_{rs}}(RGB)$.

$$\mathbf{R_{rs,XYZ}}^T = \begin{bmatrix} R_{rs}(X) & R_{rs}(Y) & R_{rs}(Z) \end{bmatrix} = \begin{bmatrix} 0.040 & 0.041 & 0.037 \end{bmatrix} \tag{4.27}$$

$$\mathbf{\Sigma}_{\mathbf{R_{rs,XYZ}}} = \begin{bmatrix} 6.5 & 6.3 & 5.2 \\ 6.3 & 6.2 & 5.3 \\ 5.2 & 5.3 & 6.2 \end{bmatrix} \times 10^{-6} \qquad \mathbf{R}_{\mathbf{R_{rs,XYZ}}} = \begin{bmatrix} 1.00 & 0.98 & 0.81 \\ 0.98 & 1.00 & 0.85 \\ 0.81 & 0.85 & 1.00 \end{bmatrix} \tag{4.28}$$

The resulting uncertainties were further propagated analytically into $(x, y)$ chromaticity, again using the Jacobian matrix approximation as shown in Equations (4.29) and (4.30).

$$\mathbf{R_{rs,xy}} = \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} \frac{X}{X+Y+Z} & \frac{Y}{X+Y+Z} \end{bmatrix} \tag{4.29}$$

$$\mathbf{J}_{\mathbf{R_{rs,xy}}} = \begin{bmatrix} \frac{Y+Z}{(X+Y+Z)^2} & \frac{-X}{(X+Y+Z)^2} & \frac{-X}{(X+Y+Z)^2} \\ \frac{-Y}{(X+Y+Z)^2} & \frac{X+Z}{(X+Y+Z)^2} & \frac{-Y}{(X+Y+Z)^2} \end{bmatrix} \tag{4.30}$$

The results for the 07:47 data are shown in Equation (4.31) and (4.32)

$$\mathbf{R_{rs,xy}}^T = \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 0.34 & 0.35 \end{bmatrix} \tag{4.31}$$

$$\mathbf{\Sigma}_{\mathbf{R_{rs,xy}}} = \begin{bmatrix} 2.7 & 1.2 \\ 1.2 & 1.3 \end{bmatrix} \times 10^{-5} \qquad \mathbf{R}_{\mathbf{R_{rs,xy}}} = \begin{bmatrix} 1.00 & 0.65 \\ 0.65 & 1.00 \end{bmatrix} \tag{4.32}$$

The uncertainty propagation from $(x, y)$ to hue angle $\alpha$ is problematic. $\alpha$ itself is calculated using Equation (4.6), giving $\alpha = 71°$ for the 07:47 data. The associated Jacobian matrix is given in Equation (4.33) and can be applied, giving an uncertainty of $\sigma_\alpha = 4°$ in the example.

$$\alpha = \arctan2\left(y - 1/3, x - 1/3\right) \mod 2\pi \qquad \text{(4.6 rev.)}$$

$$\mathbf{J}_\alpha = \begin{bmatrix} \frac{-(y-1/3)}{(x-1/3)^2+(y-1/3)^2} & \frac{x-1/3}{(x-1/3)^2+(y-1/3)^2} \end{bmatrix} \tag{4.33}$$

However, for $(x, y) \to (1/3, 1/3)$, the assumption underpinning this method, namely that the transformation can be locally approximated by a linear one, breaks down. This can result in extremely large uncertainty estimates, such as $\alpha = (89 \pm 59)°$ for the 2019-07-03 12:32 UTC Galaxy S8 data, where $(x, y) = (0.3334, 0.3409)$. For this reason, the uncertainty in $\alpha$ is better estimated from replicate observations rather than this analytical propagation.

## 4.B Smartphone RGB to XYZ transformation

The RGB to XYZ transformation matrices for the iPhone SE and Galaxy S8 were calculated by determining the locations of the RGB primary vectors in XYZ space, following https://www.ryanjuckett.com/rgb-color-space-conversion/. This process was implemented in Python and integrated into the SPECTACLE module, available from https://github.com/monocle-h2020/camera_calibration.

The transformation matrix $\mathbf{M}_{RGB\to XYZ}$ had the RGB primary vectors, $\mathbf{r}_{XYZ}, \mathbf{g}_{XYZ}, \mathbf{b}_{XYZ}$, as its columns, as shown in block matrix form in Equation (4.34).

$$\mathbf{M}_{RGB\to XYZ} = \begin{bmatrix} \mathbf{r}_{XYZ} & \mathbf{g}_{XYZ} & \mathbf{b}_{XYZ} \end{bmatrix} \tag{4.34}$$

First, the RGB SRFs were convolved with the CIE XYZ colour matching functions (CMFs) to give $\mathbf{r}'_{XYZ}, \mathbf{g}'_{XYZ}, \mathbf{b}'_{XYZ}$. This is shown in Equation (4.35), where $S_{R,G,B}$ is the SRF for the RGB bands, and $\bar{x}, \bar{y}, \bar{z}$ are the CMFs. $\lambda$ dependencies are dropped for clarity. The integrals were evaluated numerically from 390–700 nm.

$$\mathbf{r}'_{XYZ} = \begin{bmatrix} \int S_R \bar{x} \mathrm{d}\lambda \\ \int S_R \bar{y} \mathrm{d}\lambda \\ \int S_R \bar{z} \mathrm{d}\lambda \end{bmatrix} \qquad \mathbf{g}'_{XYZ} = \begin{bmatrix} \int S_G \bar{x} \mathrm{d}\lambda \\ \int S_G \bar{y} \mathrm{d}\lambda \\ \int S_G \bar{z} \mathrm{d}\lambda \end{bmatrix} \qquad \mathbf{b}'_{XYZ} = \begin{bmatrix} \int S_B \bar{x} \mathrm{d}\lambda \\ \int S_B \bar{y} \mathrm{d}\lambda \\ \int S_B \bar{z} \mathrm{d}\lambda \end{bmatrix} \tag{4.35}$$

The normalisation of $\mathbf{r}'_{XYZ}, \mathbf{g}'_{XYZ}, \mathbf{b}'_{XYZ}$ was arbitrary, and they needed to be re-normalised to the desired E-type illuminant white point $\mathbf{w}_{XYZ} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$. First, the corresponding vectors in $xyz$ chromaticity were calculated as shown in Equation (4.36).

$$\mathbf{r}_{xyz} = \frac{\mathbf{r}'_{XYZ}}{\sum_{XYZ} \mathbf{r}'_{XYZ}} \qquad \mathbf{g}_{xyz} = \frac{\mathbf{g}'_{XYZ}}{\sum_{XYZ} \mathbf{g}'_{XYZ}} \qquad \mathbf{b}_{xyz} = \frac{\mathbf{b}'_{XYZ}}{\sum_{XYZ} \mathbf{b}'_{XYZ}} \tag{4.36}$$

Similarly, the following held for the column vectors of $\mathbf{M}_{RGB\to XYZ}$:

$$\mathbf{r}_{XYZ} = \mathbf{r}_{xyz} \sum_{XYZ} \mathbf{r}_{XYZ} \qquad \mathbf{g}_{XYZ} = \mathbf{g}_{xyz} \sum_{XYZ} \mathbf{g}_{XYZ} \qquad \mathbf{b}_{XYZ} = \mathbf{b}_{xyz} \sum_{XYZ} \mathbf{b}_{XYZ} \tag{4.37}$$

Combining Equations (4.34) and (4.37) led to the following:

$$\mathbf{M}_{RGB\to XYZ} = \begin{bmatrix} \mathbf{r}_{xyz} \sum_{XYZ} \mathbf{r}_{XYZ} & \mathbf{g}_{xyz} \sum_{XYZ} \mathbf{g}_{XYZ} & \mathbf{b}_{xyz} \sum_{XYZ} \mathbf{b}_{XYZ} \end{bmatrix} \tag{4.38}$$

$$= \begin{bmatrix} r_x & g_x & b_x \\ r_y & g_y & b_y \\ r_z & g_z & b_z \end{bmatrix} \begin{bmatrix} \sum_{XYZ} \mathbf{r}_{XYZ} & 0 & 0 \\ 0 & \sum_{XYZ} \mathbf{g}_{XYZ} & 0 \\ 0 & 0 & \sum_{XYZ} \mathbf{b}_{XYZ} \end{bmatrix} \tag{4.39}$$

The unknowns in the diagonal matrix on the right were determined by defining the white point $\mathbf{w}$ to be an eigenvector of $\mathbf{M}_{RGB\to XYZ}$ with eigenvalue 1, as shown in Equation (4.40).

$$\mathbf{M}_{RGB\to XYZ}\mathbf{w} = \mathbf{w} \tag{4.40}$$

$$\begin{bmatrix} r_x & g_x & b_x \\ r_y & g_y & b_y \\ r_z & g_z & b_z \end{bmatrix} \begin{bmatrix} \sum_{XYZ} \mathbf{r}_{XYZ} & 0 & 0 \\ 0 & \sum_{XYZ} \mathbf{g}_{XYZ} & 0 \\ 0 & 0 & \sum_{XYZ} \mathbf{b}_{XYZ} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{4.41}$$

$$\begin{bmatrix} r_x & g_x & b_x \\ r_y & g_y & b_y \\ r_z & g_z & b_z \end{bmatrix} \begin{bmatrix} \sum_{XYZ} \mathbf{r}_{XYZ} \\ \sum_{XYZ} \mathbf{g}_{XYZ} \\ \sum_{XYZ} \mathbf{b}_{XYZ} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{4.42}$$

$$\begin{bmatrix} \sum_{XYZ} \mathbf{r}_{XYZ} \\ \sum_{XYZ} \mathbf{g}_{XYZ} \\ \sum_{XYZ} \mathbf{b}_{XYZ} \end{bmatrix} = \begin{bmatrix} r_x & g_x & b_x \\ r_y & g_y & b_y \\ r_z & g_z & b_z \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{4.43}$$

Since all elements on the right-hand side of Equation (4.43) were known, the elements on the left-hand side could be calculated. Finally, plugging these back into Equation (4.39) resulted in the RGB to XYZ transformation matrix. The resulting matrices for the iPhone SE and Galaxy S8 are provided in the main paper.

## Funding

European Commission Horizon 2020 programme (grant nr. 776480, MONOCLE).

## Acknowledgements

The authors would like to thank Tom Jordan, Victor Martínez-Vicente, Aser Mata, Caitlin Riddick, Norbert Schmidt, and Anna Windle for their help in the data acquisition and processing, and Thomas Leeuw, Sanjana Panchagnula, and Hans van der Woerd for valuable discussions relating to this work.

The datasets generated for this study are available from Zenodo (https://dx.doi.org/10.5281/zenodo.4549621).