



Universiteit  
Leiden  
The Netherlands

## Accessible remote sensing of water

Burggraaff, O.

### Citation

Burggraaff, O. (2022, December 13). *Accessible remote sensing of water*. Retrieved from <https://hdl.handle.net/1887/3497379>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3497379>

**Note:** To cite this publication please use the final published version (if applicable).

# 3 | Standardised spectral and radiometric calibration of consumer cameras

Adapted from

Olivier Burggraaff, Norbert Schmidt, Jaime Zamorano, Klaas Pauly, Sergio Pascual, Carlos Tapia, Evangelos Spyarakos, Frans Snik

*Optics Express*, 27(14), 19075–19101 (2019)

[doi:10.1364/OE.27.019075](https://doi.org/10.1364/OE.27.019075)

Consumer cameras, particularly onboard smartphones and UAVs, are now commonly used as scientific instruments. However, their data processing pipelines are not optimised for quantitative radiometry and their calibration is more complex than that of scientific cameras. The lack of a standardised calibration methodology limits the interoperability between devices and, in the ever-changing market, ultimately the lifespan of projects using them. We present a standardised methodology and database (SPECTACLE) for spectral and radiometric calibrations of consumer cameras, including linearity, bias variations, read-out noise, dark current, ISO speed and gain, flat-field, and RGB spectral response. This includes golden standard ground-truth methods and do-it-yourself methods suitable for non-experts. Applying this methodology to seven popular cameras, we found high linearity in RAW but not JPEG data, inter-pixel gain variations >400% correlated with large-scale bias and read-out noise patterns, non-trivial ISO speed normalisation functions, flat-field correction factors varying by up to 2.79 over the field of view, and both similarities and differences in spectral response. Moreover, these results differed wildly between camera models, highlighting the importance of standardisation and a centralised database.

### 3.1 Introduction

Consumer cameras have seen increasing scientific use in recent years. Their low cost makes them ideal for projects involving large scale deployment, autonomous monitoring, or citizen science. Successful scientific applications include environmental monitoring [56, 94, 121, 162, 276, 285, 317–323], cosmic ray detection [324], vegetation mapping [325–329], colour science [285, 330–335], and biomedical applications [336–343]. However, the use of consumer cameras is made difficult by limited software controls and camera specifications. Inter-calibration of multiple camera models is complex and laborious and the market constantly shifting, and for these reasons many projects are limited to only a few devices. These constraints severely affect both the data quality and the sustainability of projects using consumer cameras.

Smartphones, in particular, have become a common tool for research, thanks to their wide availability and features such as wireless connectivity. Many scientific applications (apps) using smartphone cameras have been developed, across a variety of fields. A recent example is HydroColor, a citizen science tool for remote sensing of water, specifically turbidity and remote sensing reflectance  $R_{rs}$ . These are derived from RGB colour photographs using standard inversion algorithms. Results from this app agree well with professional standard equipment, with mean errors in  $R_{rs}$  and turbidity  $\leq 26\%$  compared to reference sensors. However, due to software constraints, the app uses compressed JPEG data rather than raw sensor data and assumes identical spectral responses for all cameras. These factors severely limit the possible data quality. Nevertheless, HydroColor has already seen significant adoption by the community, and future developments may reduce the aforementioned limitations [121, 276, 318]. Another recent application of smartphone cameras is bioluminescent-based analyte quantitation by smartphone (BAQS), a technique for the detection of bioluminescent bacteria. Using BAQS, flux intensities down to the pW scale can be detected on some smartphone models; however, on others, software constraints and dark noise severely limit its sensitivity [344]. As a final example, Skandarajah et al. used smartphone cameras with conventional microscopes for micron-scale imaging, for example of stained blood samples. Resolutions comparable to scientific cameras were achieved, but intensity and colour measurements were limited by a lack of camera control and factors including non-linearity and white balance [342]. A full review of smartphone science is outside the scope of this work, and we instead refer the reader to a number of extensive reviews by other authors [143, 345–351].

Smartphone spectroscopy is an active field of development [98, 143]. Many spectroscopic add-ons have been developed, including do-it-yourself models costing less than \$10 at Public Lab<sup>9</sup>. One early smartphone spectrometer was iSPEX, a spectropolarimetric add-on for iPhone devices used by >3000 citizen scientists to measure aerosol optical thickness (AOT) in the Netherlands in 2013. iSPEX data were found to agree well with reference sensors, with a correlation coefficient of  $r = 0.81$  between AOT values observed with iSPEX and with the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra satellites [94]. However, the iSPEX data were limited in their polarimetric accuracy (absolute uncertainties in the degree of linear polarisation (DoLP)  $\approx 0.03$ ), preventing quantitative measurements of aerosol compositions and sizes [94]. This relatively large error stemmed from a lack of camera controls, such as the inability to fix the focus of the camera to a controlled and reproducible position. Furthermore, the sustainability of iSPEX in the fast-moving smartphone market was limited by its need for device-specific calibrations.

<sup>9</sup><https://publiclab.org/wiki/spectrometry>

Consumer uncrewed aerial vehicles (UAVs) with RGB cameras have similarly become common scientific instruments. They provide a low-cost, high-resolution replacement for, or complement to, satellite and airplane imagery, especially for environmental monitoring [325–327, 352–354]. UAV data are increasingly being integrated with data from other platforms, such as satellites [355].

However, few scientific consumer camera projects progress past a proof-of-concept on a handful of camera models, which often become obsolete within two years, particularly in the constantly shifting smartphone market. This severely limits the sustainability of projects that require calibrations specific to each camera model. Difficulties in upscaling and future-proofing such calibrations are an oft cited constraint on the combination of multiple camera models [94, 319, 323, 324, 331, 349, 356]. Further complications are introduced by the lack of control over camera hardware and software parameters such as focus and white balance [94, 142, 320, 325, 342, 344]. For example, the dominant smartphone operating systems, Android and iOS, only introduced support for unprocessed (RAW) imagery as recently as 2014 (Android 5.0 ‘Lollipop’) and 2016 (iOS 10). Previously, third-party developers could only use JPEG data, which introduce a number of systematic errors due to their lossy compression and bit-rate reduction [94, 121, 142, 317, 319, 334, 342, 349, 357]. Other common problems in consumer camera data include non-linearity and the gamma correction [94, 121, 162, 285, 317, 322, 329, 332, 334, 338, 342, 357–362], electronic and thermal noise [118, 319, 324, 344, 354, 363, 364], and highly variable (between camera models) spectral response functions which are not provided by manufacturers [121, 317, 323, 331, 333, 334, 349, 351, 354, 358, 365]. These factors limit the accuracy of radiometric measurements done with consumer cameras by introducing systematic errors. Furthermore, the accuracy of colour measurements and their conversion to standard measures, such as the CIE 1931 XYZ and CIELAB colour spaces, is limited by distortions in the observed colours [330] and differences in spectral response functions [331, 333–335].

Extensive (spectro-)radiometric calibrations of consumer cameras are laborious and require specialised equipment and are thus not commonly performed [333, 361, 366]. A notable exception is the spectral and absolute radiometric calibration of a Raspberry Pi 3 V2 webcam by Pagnutti et al. [118], including calibrations of linearity, exposure stability, thermal and electronic noise, flat-field, and spectral response. Using this absolute radiometric calibration, digital values can be converted into SI units of radiance. However, the authors noted the need to characterise a large number of these cameras before the results could be applied in general. Moreover, certain calibrations are device-dependent and would need to be done separately on each device. Spectral and radiometric calibrations of seven cameras, including the Raspberry Pi, are given in [358]. These calibrations include dark current, flat-fielding, linearity, and spectral characterisation. However, for the five digicams included in this work, JPEG data were used, severely limiting the quality and usefulness of these calibrations, as described above.

Spectral characterisations are more commonly published since these are vital for quantitative colour analysis. Using various methods, the spectral responses of several Canon [317, 329, 333, 334, 354, 358, 361, 367], Nikon [317, 323, 333, 334, 361, 365–368], Olympus [333, 334, 358], Panasonic [354], SeaLife [323], Sigma [366], and Sony [317, 333, 354, 358, 367] digital cameras (digicams), as well as a number of smartphones [121, 333], have been measured. Direct comparisons between >2 different camera models are given in [121, 317, 333, 354, 358]. Common features include the peak response wavelengths for the RGB colour filters, approximately 600, 520, and 470 nm, respectively, as well as a roughly Gaussian profile around the

peak. Differences are found especially in the wings, notably the locations of secondary peaks and near-infrared (NIR) and ultraviolet (UV) cut-off wavelengths. These may cause significant differences in observed colours between cameras [331, 333], especially for narrow-band sources.

Camera calibrations in the literature are often limited to a small number of cameras or properties, either to narrow down the scope or because of limitations in time and equipment. Furthermore, calibration data are published in varying formats, locations, and quality, complicating their use by others. Standardised formats exist, such as those for vignetting, bias, and colour corrections described in Adobe's digital negative (DNG) standard [369], but have seen limited adoption. The European Machine Vision Association (EMVA) standard 1288 [370] for characterisation of cameras is extremely thorough, but has also seen limited adoption due to the high-end equipment required [361] and its scope simply being too broad for many practical purposes. Similarly, standardised data sets or databases, for example containing spectral response curves [333, 366], have been created but these are limited in scope and, again, adoption. To our knowledge, no widely adopted standardised methodology or centralised database containing spectral and radiometric calibration data for consumer cameras has been created thus far.

In this work, we present a standardised methodology for the calibration of consumer cameras and a database, SPECTACLE (Standardised Photographic Equipment Calibration Technique And CataLoguE), containing calibration data for the most popular devices. The calibration methodology is focused on simplicity and facilitating measurements by non-experts and those lacking expensive equipment, similarly to [361] but with a broader scope including software, optics, and sensor characteristics. The database is designed with openness and sustainability in mind, focusing on community contributions. Furthermore, we strive to follow previously existing standards, such as DNG [369] and EMVA 1288 [370], where practical. Our focus is on radiometric and photometric measurements but these calibration data can equally be used for colour science purposes, in particular to convert between colour spaces using the measured spectral response curves. We stress that we have no financial nor commercial interests in consumer cameras, and any comparison between devices is purely scientific. The aim of our standardised methodology and the SPECTACLE database is merely to simplify the use of data from consumer cameras, not to cast judgment on their quality.

Section 3.2 contains an overview of hardware and software trends in consumer cameras. We present the standardised calibration methodology in Section 3.3. Section 3.4 contains results from its application to several popular cameras and a description of the SPECTACLE database. Finally, in Section 3.5 we compare our findings with previous work and discuss future perspectives.

## 3.2 Trends in consumer cameras

Consumer cameras can be divided into four categories, namely smartphones, UAVs, digicams (DSLR and mirrorless), and webcams. Despite serving very diverse purposes, these cameras share common characteristics and can be calibrated with the same methods.

CMOS-based sensors now dominate the consumer camera market [98]. These are often not produced in-house by camera manufacturers, but acquired from external parties, such as Sony and Samsung. Different cameras often use the same sensor, such as the Sony IMX298 which is used in 12 smartphone models from 10 different manufacturers.

Most colour cameras use Bayer filters, on-chip RGB filters arranged in a checkerboard pattern, with two green pixels ( $G$  and  $G_2$ ) for every red or blue one [138]. The spectral responses of these filters differ strongly between cameras and are further modified by fore-optics [331]. Alternate pixelated filter arrangements exist, intended for example to reduce aliasing, but with little adoption so far [371]. Data from the separate  $RGBG_2$  pixels can be recombined through a process known as demosaicking to retrieve an image with interpolated RGB values for each pixel. Many different schemes exist for this [371], ranging from simple bilinear interpolation [118, 330, 354] to complex computational methods [140]. Consumer camera software often includes proprietary demosaicking algorithms [329, 342] which may introduce complex, unpredictable effects. Depending on their implementation, demosaicking schemes typically mix data from different filters and remove their mutual independence, leading to undesirable cross-feed effects [118, 121]. In any case, the added data are fully synthetic and thus do not offer any new radiometric information. It is thus preferable for radiometric applications to treat the  $RGBG_2$  images completely independently [329] and demosaic data for visualisation purposes [118] only.

As discussed previously, the most commonly used digital file formats are JPEG (or JPG) and RAW. In both formats, data are saved on a pixel-by-pixel basis in analogue-digital units (ADU). ADU are alternately referred to as digital numbers (DN) in the literature, but in this work we will use the ADU nomenclature. JPEG (ISO 10918) is based on lossy spatial compression and downsampling to 8-bit values, optimal for small file sizes while maintaining aesthetic qualities. Due to camera-specific processing and compression artefacts, JPEG images lose information and are not recommended for quantitative analysis [94, 121, 142, 317, 319, 329, 334, 342, 349, 357]. While standardisations exist, such as the standard Red Green Blue (sRGB) colour space and gamma curve [369], these are not strictly adhered to and cannot be assumed in data processing [372]. Conversely, RAW files contain relatively unprocessed sensor output, intended for manual post-processing. One factor complicating the reduction of RAW data is their mosaicked nature, due to which they must be demosaicked or treated as multiple independent images, as discussed above. Despite these complications, their unprocessed nature makes RAW data highly preferable for scientific purposes [118, 121, 329, 331, 334, 349].

Available camera controls generally include focus, exposure time, ISO speed (sensitivity), and aperture. Focus and aperture are changed by physical movement of camera optics, though most webcams and smartphones only allow a single, fixed aperture. ISO speed is set by changing the camera gain, through analogue amplification or digital processing. Analogue amplification involves varying the gain of the CMOS amplifiers, which can be done on the level of individual pixels. Conversely, digital gain is implemented in post-processing by simply re-scaling and interpolating measured digital values. Since ISO speed is a measure of the overall sensitivity of the camera, including fore-optics, each camera (and possibly each pixel) has a unique relation between ISO speed and gain. Finally, exposure time may be set by a physical shutter (common in digicams) or an electronic one (common in smartphones). Other parameters like white balance only affect processed imagery and are not relevant to RAW photography.

Many cameras include some built-in calibrations, most notably for non-linearity, dark current, and flat-fielding effects. Non-linearity corrections are typically based on previously measured correction curves [104]. Dark current corrections (autodarkening) are commonly done using unilluminated exposures or permanently dark pixels around the sensor. Finally, flat-fielding (specifically vignetting) is typically corrected using a pre-made correction map.

A variety of methods for generating such maps exists, based for example on computational methods using regular photographs [330, 373–376], simply averaging many exposures [329], and simply imaging white paper [377]. These maps are typically parametrised, for which various methods also exist [118, 329, 330, 369, 374–377], the simplest being the  $\cos^4$  model, a combination of inverse square falloff, Lambert’s law, and foreshortening [374]. Alternately, a pixel-by-pixel map of vignetting correction coefficients may be used. Such maps may be device-specific or generalised for a camera model. Notably, iOS-based smartphones use the seven-parameter parametrisation described in the DNG standard [369] (see Section 3.3.8) while Android-based smartphones use pixel-by-pixel maps.

### 3.2.1 Smartphones

The smartphone market has become remarkably homogeneous in recent years, with virtually all models using the slate form factor, featuring a large touch screen, few buttons, and a camera on either side of the device. The most popular smartphones are all iOS- or Android-based. Both these operating systems now support RAW photography using Adobe’s DNG standard [369], though not on all devices. Hardware properties are rarely released by manufacturers, and are instead often provided by reviewers through disassembly of the smartphone.

Smartphone cameras aim to reproduce the human eye and thus have similar optical properties [121]. Sensors, most commonly from the Sony Exmor series, are compact with 12–16 megapixels and a diagonal of 5–8 mm. Some devices focus on high-resolution imagery with many pixels, while others are optimised for low-light conditions, with fewer but larger pixels.

Smartphones now increasingly have multiple rear cameras. These secondary cameras offer features such as different fixed focal lengths and higher sensitivity, for example with a different lens or a monochromatic sensor. All rear cameras are typically placed in a cluster at the top right or top center of the smartphone.

## 3.3 Methods

In this section we describe the standardised methods for the calibration of consumer cameras. We developed a custom data processing pipeline, implemented in Python scripts available on GitHub<sup>10</sup> and iOS<sup>11</sup> and Android apps<sup>12</sup>.

Section 3.3.1 describes the experimental setups and data processing used in calibration measurements. The methods used to characterise and calibrate the camera responses are given in Sections 3.3.2–3.3.9. Finally Section 3.3.10 describes how consumer camera data are converted into relative radiometric units using the previously described calibration measurements. These units provide a constant scale, independent of exposure parameters and individual device characteristics, for each camera model, a constant factor  $K$  per model away from absolute radiometric units ( $\text{W m}^{-2} \text{sr}^{-1}$ ). Absolute radiometric calibration is outside the scope of this work.

---

<sup>10</sup>[https://github.com/monocle-h2020/camera\\_calibration](https://github.com/monocle-h2020/camera_calibration)

<sup>11</sup>[https://github.com/monocle-h2020/spectacle\\_ios](https://github.com/monocle-h2020/spectacle_ios)

<sup>12</sup>[https://github.com/monocle-h2020/spectacle\\_android](https://github.com/monocle-h2020/spectacle_android)

### 3.3.1 Experimental setup

This section describes the setups used in our golden standard ground-truth experiments. Descriptions of do-it-yourself (DIY) calibration methods are given in the relevant sections. All images from all cameras were taken in RAW format; for the linearity measurements, simultaneous RAW and JPEG images were taken for comparison. As discussed in Section 3.2, demosaicking schemes introduce synthetic data and undesirable cross-feed effects. For this reason, in our data reduction the RAW images were split into separate RGBG<sub>2</sub> images which were analysed individually [329]. Multiple images were taken and stacked for each measurement to improve the signal-to-noise ratio (SNR). On smartphones, the aforementioned iOS and Android apps were used to control the camera and automatically take multiple exposures. Exposure settings, including ISO speeds and exposure times were obtained from camera controls where possible, since EXIF metadata values for these were found (Section 3.4.1) to be unreliable.

The setup for measuring linearity, ISO-gain relations, and inter-pixel gain variations on smartphones is shown in Figure 3.1. A halogen light source (OceanOptics HL-2000-LL) was used, specified by the manufacturer to be stable to 0.15% peak-to-peak and drift <0.3% per hour after a warm-up of 10 minutes. Its light was fed into an optical fiber (Thorlabs M25L02) and collimated using two lenses (Thorlabs AC254-030-A with  $f = 30$  mm and AC508-200-A with  $f = 200$  mm). Two linear polarisers (both Thorlabs LPVISE100-A, with an extinction ratio  $\geq 495$  from 400–700 nm), the first rotatable and the second fixed, were used to attenuate the light beam entering an integrating sphere (Thorlabs IS200). Using Malus's law ( $I = I_0 \cos^2 \theta$ ), the rotation angle between the polarisers could be used to calculate the attenuation. A calibration detector was not necessary since all experiments done with this setup involve relative measurements only. Malus's law was first fitted to a series of exposures over the entire rotation range to determine the reference angle. The rotation angle of the first polariser could be determined visually up to 2° precision, giving a typical uncertainty on the attenuated intensity of 2.5%. Finally, smartphones were placed on top of the integrating sphere, flush against the view-port. The farthest possible focus was used (infinity on Android devices, an arbitrary number on iOS). All experiments done with this setup involved analysis on the individual pixel and (broad-band) filter level, without any spatial averaging. Because of this, differences in illumination due to spectral dependencies in the polariser throughput or the integrating sphere output did not affect any of the experiments.

The linear polarisers can be replaced with alternate methods for attenuation, such as neutral density filters. Attenuation can also be replaced completely by varying exposure times instead, though physical attenuation may be more precise [118]. The integrating sphere may

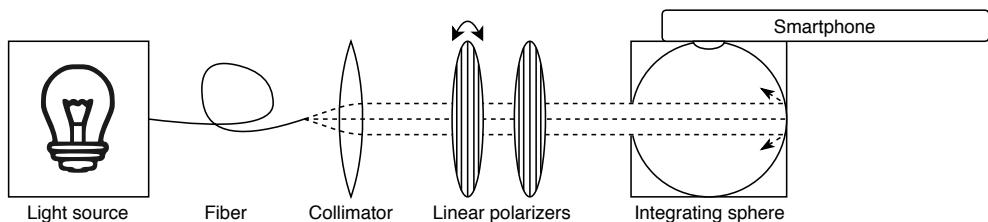


Figure 3.1: Setup used to measure linearity, ISO-gain relations, and inter-pixel gain variations on smartphones. The first linear polariser was rotatable, the second fixed. Smartphones were placed with their camera flush against the view-port at the top of the integrating sphere.

be replaced by another diffuse surface, such as a Spectralon target. If sufficiently wide, the light beam may also be shone directly onto the sensor; such a setup was used for digicams, with the digicam in place of the collimator in Figure 3.1 at a sufficient distance to completely illuminate the sensor. This was done to simplify the alignment process since our digicams had large physical CMOS sizes. Since all measurements were done on the individual pixel level, they were not affected by the added differences in illumination.

Bias, read-out noise, and dark current were measured on all devices by placing the camera flush against a flat surface (such as a table), pointing down, in a dark room. The setups for flat-fielding and spectral characterisation are described in Sections 3.3.8 and 3.3.9, respectively.

### 3.3.2 General properties

General hardware and software properties were retrieved from official specifications and community reviews. A survey across these provided an overview of basic physical and optical parameters of cameras. On Android smartphones, the Camera2 API provides ample information on such parameters, facilitating automatic data collection using an app.

The retrieved device properties included the camera type, manufacturer, product code and internal identifiers, release year, the number of cameras (for smartphones), camera module identifiers, and CMOS sensor models. Sensor properties included physical size, pixel pitch, resolution, orientation with respect to the device, colour filter pattern, and bit depth. Camera optic properties included focal length,  $f$ -number, neutral density filters (for high-end smartphones), and a vignetting model if available. Finally, software and firmware properties included supported software versions, RAW and JPEG support, estimated bias value (see Section 3.3.4), ISO speed range, exposure time range, and the active part of the sensor (accounting for dark pixels, see Section 3.3.5).

### 3.3.3 Linearity

Sensor linearity was quantified by measuring the camera response to varying exposures, either by attenuating a light source or by varying the exposure time, as discussed in Section 3.3.1. We used the setup shown in Figure 3.1 with two linear polarisers to attenuate the light for smartphones, since exposure times on those are not completely trustworthy (Section 3.4.1). Conversely, for digicams, exposure times are reliable [361, 366] and thus were used instead of physical attenuation to simplify the setup. A third method, varying the physical aperture, changes the distribution of light on the sensor [374] and thus cannot be used to measure linearity.

Two common types of non-linearity exist, either across the entire intensity range or only at high intensities. The former is common in JPEG imagery due to the gamma correction [329, 342] while the latter is expected in both JPEG and RAW data. We only investigated the former since it has the largest impact on data quality, as described in Section 3.1. Non-linearity at high intensities is easily negated by discarding data above a threshold value; we use a threshold of  $\geq 95\%$  of the maximum digital value.

The linearity of each pixel was expressed through the Pearson correlation coefficient  $r$ , a measure of the linear correlation between intensity and camera response. Pixels were analysed individually to negate differences in illumination and vignetting effects (Section 3.3.8). Simulated responses of a perfectly linear camera with a mean error of 5% in the incoming intensity simulating, for example, errors in exposure parameters or polariser alignment in the

setup described in Section 3.3.1, as well as Poisson noise ( $\sigma_N = \sqrt{N}$ ) and 10 ADU read noise in the response were analysed. This included simulated measurements at 15 different exposures, averaged over 10 images per exposure. These simulated data resulted in a mean value of  $r = 0.996 \pm 0.002$  and a lower 0.1 percentile  $P_{0.1}(r) = 0.985$ . To account for unforeseen measurement errors, we set the cut-off for linearity at  $r \geq 0.980$ .

Additionally, the JPEG data were compared to sRGB-like profiles to determine whether gamma inversion [285] is possible. The sRGB-like profiles are described by Equation (3.1), with  $J_C$  the JPEG response (0–255) in band  $C$ ,  $n$  a normalisation factor,  $\gamma$  the gamma correction factor and  $I$  the incoming intensity in arbitrary units. The JPEG response of each pixel was individually fit to Equation (3.1) with  $n$  and  $\gamma$  as free parameters. Additionally, profiles with standard  $\gamma$  values (2.2 and 2.4 [285]) were fit to the JPEG data (with  $n$  free) to determine the accuracy of these standards.

$$J_C = 255 \times \begin{cases} 12.92nI & \text{if } nI < 0.0031308 \\ 1.055(nI)^{1/\gamma} - 0.055 & \text{otherwise} \end{cases} \quad (3.1)$$

### 3.3.4 Bias & read-out noise

Bias (or ‘black level’) and read-out noise (RON) were measured by stacking short dark exposures. The bias and RON in individual pixels are given by the mean and variance, respectively, of their values in each stack. Many (>50) images per stack are required to distinguish bias variations from RON. Temporal variations were probed by repeating this process several times. While EXIF metadata often contain a bias value, this is only an estimate and should be validated by measurement.

### 3.3.5 Dark current

Dark current (thermal noise) was measured by taking dark exposures with different lengths and fitting a linear relation between exposure time and camera response to determine the dark current in ADU  $s^{-1}$ . For cameras that have autodarkening (see Section 3.2), the residual dark current was characterised instead. Depending on the autodark precision, the exposure-response relation may be non-linear in this case.

### 3.3.6 ISO speed

The relation between camera sensitivity and ISO speed was measured by taking identically exposed images at different ISO speeds. These were bias-corrected and pixel values were divided by those at the lowest ISO speed. A relation between ISO speed and normalisation factor was then fitted. Like the linearity measurements (Section 3.3.3), this was done individually per pixel to negate illumination differences and vignetting effects.

This relation may be any combination of linear and constant functions, depending on the implementation of ISO speed ratings. Linear relations correspond to analogue gain, while digital gain may result in linear or constant relations, as described in Section 3.2.

### 3.3.7 Gain variations

Inter-pixel and inter-filter gain variations were characterised using the mean-variance method [378], which exploits the Poissonian nature of photo-electrons in a sensor. We applied this

method to individual pixels rather than averaging over the sensor, to measure inter-pixel variations and remove the need for flat-fielding prior to this calibration. The response of a digital camera to incoming light is given by Equation (3.2), with  $M$  the mean response in ADU,  $I$  the exposure in photo-electrons,  $D$  the dark current in  $e^-$ ,  $B$  the bias in ADU, and  $G$  the gain in ADU/ $e^-$ . Both  $I$  and  $D$  are integrated over the exposure time.

$$M = IG + DG + B \quad (3.2)$$

The variance in the response of a pixel is a combination of shot noise on the photo-electrons and dark current, and read noise. The shot noise follows a Poissonian distribution with a standard deviation  $\sigma_I = \sqrt{I}$  and thus a variance  $V_I = I$ . The total variance in the response is expressed in Equation (3.3), with  $V$  the variance in ADU<sup>2</sup> and  $RON$  the read noise in ADU.

$$V = IG^2 + DG^2 + RON^2 \quad (3.3)$$

After correcting for bias and dark current, and assuming  $DG^2$  is negligible, a linear relation between mean and variance is found, shown in Equation (3.4).

$$V = GM_{\text{cor}} + RON^2 \quad (3.4)$$

Equation (3.4) was fitted to mean and variance values from several image stacks taken under different illumination conditions. Within each stack, all images were exposed identically, while the illumination varied between stacks. A large amount of data (>10 stacks of >50 images each) was necessary to constrain the fitted gain values sufficiently (typical relative errors in individual pixels <15%). ISO normalisation functions derived in Section 3.3.6 may be used to extrapolate measured values to different ISO speeds.

### 3.3.8 Flat-field correction

Flat-fielding was performed by imaging a uniform light source. Unlike telescopes, most consumer cameras have fields-of-view (FoV) too large to use the twilight sky for this. Instead, a large integrating sphere was used to create an isotropic light field, as described in [118]. We used a LabSphere HELIOS USLR-D12L-NMNN lit by three halogen lamps with a specified luminance uniformity of  $\pm 1.0\%$ , sequentially placing each camera before its aperture.

Any significant chromatic differences in the flat-field response were measured automatically, since all filters were exposed simultaneously. The RGBG<sub>2</sub> images were split out and each normalised to their maximum value, then recombined and smoothed with a Gaussian filter ( $\sigma = 10$  pixels); both the individual RGBG<sub>2</sub> images and the recombined image were analysed. Since vignetting, often the dominant flat-field component, is caused by the camera aperture, the flat-field response changes and must be measured again when varying the aperture [374].

Vignetting can be parametrised in a number of different ways, as discussed in Section 3.2. For consistency, we used the DNG seven-parameter ( $k_0 \dots k_4, \hat{c}_x, \hat{c}_y$ ) model, also used internally in iOS smartphones, for the flat-field correction factor  $g(x, y)$ , expressed in Equation (3.5), with  $r$  the normalised Euclidean distance from pixel  $(x, y)$  to the optical center  $(\hat{c}_x, \hat{c}_y)$ .

$$g(x, y) = 1 + k_0 r^2 + k_1 r^4 + k_2 r^6 + k_3 r^8 + k_4 r^{10} \quad (3.5)$$

Three simpler, alternate methods were also tested. The first involved imaging an overcast sky, the second imaging the sun with a piece of paper taped onto the camera as a diffuser similarly to the Hukseflux Pyranometer app<sup>13</sup>. For the final method, the camera, again with a paper diffuser, was held flush against a computer monitor displaying a white screen, somewhat similarly to [361]. In all three cases, the camera was dithered and rotated 360° during measurements to average out anisotropies. Data from all three methods were processed in the same way as the integrating sphere data, to compare their efficacy.

### 3.3.9 Spectral response

The spectral response of a camera, which is a product of the individual spectral responses of its fore-optics, filters, and sensor, was measured in two ways. The first method, using a monochromator, is simple processing-wise as the data are simply a series of images at different wavelengths with known intensities [118, 354, 368]. It also allows for the measurement of inter-pixel variations in spectral response. The second, a spectrometer add-on such as iSPEX [94], is more accessible than monochromators but its spectral data are more difficult to calibrate and process.

We used a double monochromator (OL 750-M-D) at the NERC Field Spectroscopy Facility to scan a wavelength range of 390–700 nm. This wavelength range was chosen because no significant response was found outside it on any of the test cameras. The effective spectral resolution (half bandwidth) of the monochromator was 4 nm, calculated from the grating (1200 grooves/mm) and slits (2.5 mm entrance/exit and 5.0 mm central slit) used. The wavelength range was critically sampled at 2 nm intervals. A laser-driven light source (Energetiq EQ-99X) was used, and its spectral output calibrated using a silicon photodiode (Gooch & Housego OL DH-300C with a Hamamatsu S1337-1010BQ sensor). The system was NIST-traceably calibrated in 2012 and is described in more detail in [354].

Spectral characterisation was also done using a modified (removing polarisers and retarders) iSPEX add-on [94]. iSPEX has a slit consisting of two parts, one 0.4 mm ('broad') and the other 0.17 mm ('narrow') wide and a 1000 grooves/mm transmission grating foil (Edmund Optics #52-116). Using this foil, a similar spectrometer can be built for any other camera.

The reflection of sunlight on a piece of white paper was measured using the iSPEX on an iPhone SE. iSPEX projects a spectrum onto the sensor, so the pixel responses must be corrected for bias, dark current, and flat-field to obtain a quantitative spectrum. The 436.6, 544.5, and 611.6 nm spectral lines of a commercial fluorescent lamp were used for the wavelength calibration, fitting a quadratic relation between pixel position and wavelength. A stray light correction was done by subtracting the mean pixel value per column above and below the spectrum from the narrow and broad slit spectra, respectively. Two theoretical reference spectra were used to normalise the observed spectra, namely a 5777 K black body (approximating the Sun) and a diffuse solar irradiance spectrum generated using the Simple Model for the Atmospheric Radiative Transfer of Sunshine (SMARTS2) [379, 380] and smoothed to the 5 nm resolution of narrow-slit iSPEX spectra. For the latter, the location and time of the iSPEX measurements as well as the built-in urban aerosol and ground albedo models were used instead of default parameters. The models differed significantly (RMS 34%) due to the diffuse sky irradiance factored into the SMARTS2 model. Finally, the observed spectra

<sup>13</sup><http://www.hukseflux.com/product/pyranometer-app>

were corrected for the transmission of the iSPEX optics, determined by measuring the zero-order transmission using a halogen lamp and spectrometer (OceanOptics HL-2000-LL and USB2000+, respectively).

Instead of the sun, a previously calibrated commercial lamp may be used. For example, the LICA-UCM database<sup>14</sup> contains spectra of common commercial lamps which can be used as standard light sources for spectral response measurements [381]. This method has the advantage of independence from weather conditions and higher reproducibility compared to solar measurements. Combined with the new version of iSPEX we are currently developing, featuring a universal smartphone hardware interface, this enables volunteer measurements of smartphone camera spectral responses.

The spectral curves  $R_C(\lambda)$  thus derived were normalised to the global maximum transmission in all bands and used for calibration of spectral measurements and in the radiometric correction of imaging data (Section 3.3.10) to calculate effective spectral bandwidths  $\Lambda_C$ . These are defined as  $\Lambda_C = \int_C R'_C(\lambda) d\lambda$ , with  $R'_C(\lambda)$  the spectral response  $R_C(\lambda)$  normalised to the maximum in band  $C$  [99, 118]. This integral was calculated using the composite trapezoid method, implemented in the NumPy function `numpy.trapz` [312].

3

### 3.3.10 Relative radiometric calibration

The calibrations described in the previous section are used to convert digital values to radiance. Following the methods described in [99, 118, 382, 383], a digital value  $M$  (in ADU) in band  $C$  (RGBG<sub>2</sub> for Bayer filters) can be converted to effective radiance  $L_C$ , in units of  $\text{W m}^{-2} \text{sr}^{-1}$ .

Since absolute radiometric calibration is outside the scope of this work, we instead determined the relative effective radiance  $L'_C = L_C/K$ , in relative radiometric units (RRU)  $\text{m}^{-2} \text{sr}^{-1}$ , with  $K$  an extra factor accounting for the absolute quantum efficiency and transmission of the lens. Measuring these requires a previously calibrated light source with a known radiance.

The expression for converting  $M$  to  $L'_C$  is given in Equation (3.6). The advantage of the piece-wise calibration given in Equation (3.6) over a black-box approach containing all calibration components is its adaptability when a small subset of parameters are changed, such as due to firmware updates or manufacturing changes. This way, calibration data can be re-used rather than requiring a full re-calibration with every change.

$$L'_C = hc \frac{1}{A_d \Lambda_C} g \left[ \frac{4(f\#)^2}{\pi \tau N} \right] (M - B - D\tau) \quad (3.6)$$

First, the bias  $B$  (ADU; Section 3.3.4) and dark current  $D\tau$  (with  $D$  in  $\text{ADU s}^{-1}$  and  $\tau$  the exposure time in seconds; Section 3.3.5) are subtracted. Linearisation of digital values [383] is not necessary since we only used sufficiently linear pixels ( $r \geq 0.980$ ; Section 3.3.3).

Next, the image is corrected for the exposure parameters, dividing by the exposure time  $\tau$ , ISO speed normalisation factor  $N$  (Section 3.3.6), and aperture, approximated as  $\pi/4(f\#)^2$ , with  $f\#$  the  $f$ -number of the camera [118]. This approximation causes a systematic error of 4% at  $f/2.0$  [118]; for fixed-aperture systems like smartphones, this error is not relevant. For systems with adjustable apertures, an exact solution may be preferable if operating at very low  $f$ -numbers. These corrections yield a response in normalised  $\text{ADU s}^{-1} \text{sr}^{-1}$ .

The third step is the flat-field correction. The response is multiplied by the flat-field correction  $g$  (unitless; Section 3.3.8). The flat-fielding methods used here account for both

<sup>14</sup>[https://guaix.fis.ucm.es/lamps\\_spectra](https://guaix.fis.ucm.es/lamps_spectra)

optical and electronic variations in sensitivity, so a separate correction for inter-pixel gain variations (Section 3.3.7) is not necessary. Since absolute transmission and quantum efficiency were not measured, this step yields a response in relative counts  $\text{s}^{-1} \text{sr}^{-1}$ , proportional to the number of photo-electrons  $\text{s}^{-1} \text{sr}^{-1}$ .

Next, sensor properties are corrected for. The response is divided by the pixel size  $A_d$  ( $\text{m}^2$ ; Section 3.3.2) to give a response in relative counts  $\text{s}^{-1} \text{m}^{-2} \text{sr}^{-1}$ . It is then divided by the effective spectral bandwidth of band  $C$ ,  $\Lambda_C = \int_C R'_C(\lambda) d\lambda$  (Section 3.3.9).

Finally, the result is converted to a relative radiance by multiplying by a factor  $hc$ , with  $h$  Planck's constant and  $c$  the speed of light. This yields  $L'_C$  in  $\text{RRU m}^{-2} \text{sr}^{-1}$ .

For specific applications, Equation (3.6) may be simplified or adjusted. For example, inter-pixel bias and dark current variations are typically negligible in bright conditions. In those cases,  $B$  and  $D$  may be approximated by constants, and inter-pixel variations incorporated in the error budget. For spectroscopic applications, a relative spectral radiance  $L'_{C,\lambda}$  in  $\text{RRU m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$  is measured, which is not averaged over band  $C$ . In this case, the energy per photon is simply  $hc/\lambda$  and only the transmission at wavelength  $\lambda$ ,  $R_C(\lambda)$  is relevant; furthermore, the result must be divided by the wavelength coverage of each pixel  $\Delta\lambda$ . This is expressed in Equation (3.7).

$$L'_{C,\lambda} = \frac{hc}{\lambda} \frac{1}{A_d R_C(\lambda) \Delta\lambda} g \left[ \frac{4(f\#)^2}{\pi\tau N} \right] (M_\lambda - B - D\tau) \quad (3.7)$$

## 3.4 Results

The methodology described in Section 3.3 was applied to three iOS smartphones (Apple iPhone SE, 6S, and 7 Plus), two Android devices (Samsung Galaxy S6 and S8), one digi-cam (Nikon D5300), and one UAV camera (DJI Phantom Pro 4). This section contains an overview of results from these various calibration steps. Results for all devices are included in the SPECTACLE database further described in Section 3.4.9.

### 3.4.1 General properties

General hardware and software properties were retrieved from the survey described in Section 3.3.2, with a specific focus on smartphones using the previously described Android app. Little variation was found in these general properties, especially for smartphones. For example, virtually all main cameras on smartphones have apertures of  $f/2.4$ – $f/1.5$ , focal lengths of 3.8–4.5 mm, and sensors of  $3.4$ – $6.7 \times 2.7$ – $4.7$  mm, giving fields-of-view (FoVs) of  $60^\circ$ – $75^\circ \times 45^\circ$ – $55^\circ$ .

It was found from test images that EXIF metadata from some cameras are inaccurate. For example, the iPhone SE can use unrounded exposure times of 1/3.0 s and 1/3.9 s but records both as simply 1/3 s in metadata. Assuming the recorded exposure time of 1/3 s for a real exposure of 1/3.9 s would lead to photometric errors up to 30%. To counteract this, exposure parameters such as ISO speed and exposure time should be recorded separately from default EXIF metadata, for example with custom EXIF tags or extra files.

### 3.4.2 Linearity

The linearity of two smartphones (iPhone SE and Galaxy S8) and one digicam (Nikon D5300) was measured using the methods described in Section 3.3.3 and the setup described in Section 3.3.1 and shown in Figure 3.1. The smartphones were analysed using rotating linear polarisers while the D5300 was analysed by varying exposure times. Simultaneous RAW and JPEG images were taken on each device (using the Fine JPEG setting on the D5300) to compare their responses. JPEG images were taken with a fixed white balance.

The Pearson  $r$  coefficients of the RAW and JPEG responses of all pixels were calculated and their histograms are shown in Figure 3.2. The JPEG responses of all pixels in all cameras were well below the linearity threshold ( $r \geq 0.980$ ), showing again that JPEG data are highly non-linear. Conversely, nearly all RAW responses were well within the bounds for linearity, with 99.9% of  $r$  values  $\geq 0.997$  (iPhone SE),  $\geq 0.996$  (Galaxy S8), and  $\geq 0.999$  (D5300). The Galaxy S8 was the only camera with RAW responses having  $r < 0.980$ , though only in 56 pixels.

The JPEG and RAW responses of individual pixels in the iPhone SE and Galaxy S8 cameras are shown in Figure 3.3. The JPEG responses are visibly non-linear ( $r = 0.956, 0.918$ ) while the RAW responses are linear within measurement errors ( $r = 0.999, 0.998$ ). Furthermore, the dynamic range of the JPEG data is much smaller than that of the RAW data. These differences highlight the advantages of RAW data.

Finally, Figure 3.4 shows the best-fitting  $\gamma$  for the JPEG response per pixel as well as the accuracy of two standard values ( $\gamma = 2.2$  and  $2.4$ , expressed in RMS relative difference  $(1 - \text{data}/\text{fit})$ ). Large inter-pixel, inter-filter, and inter-device differences in best-fitting  $\gamma$  exist, indicating an sRGB gamma inversion with a single  $\gamma$  value is not possible. Furthermore, the  $\gamma = 2.2$  and  $2.4$  models are both clearly very inaccurate for all cameras. For the  $\gamma = 2.2$  and  $2.4$  cases respectively, 99.9% of pixels had RMS relative differences between observations and the sRGB model of  $>7\%$  and  $>10\%$  (iPhone SE),  $>13\%$  and  $>15\%$  (Galaxy S8), and  $>19\%$  and  $>21\%$  (Nikon D5300).

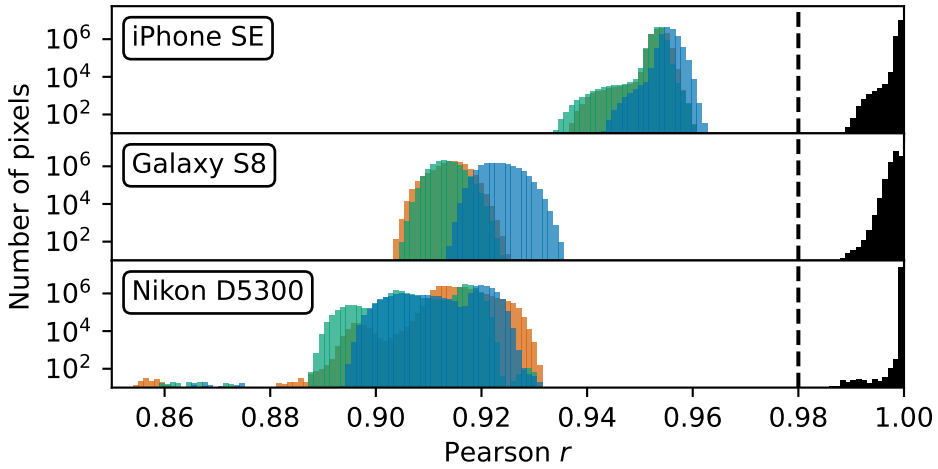


Figure 3.2: Histogram of Pearson  $r$  coefficients for RAW (black, all filters combined) and JPEG (red/green/blue) responses. The  $r \geq 0.980$  cut-off is shown with a dashed black line. Note the logarithmic vertical scale.

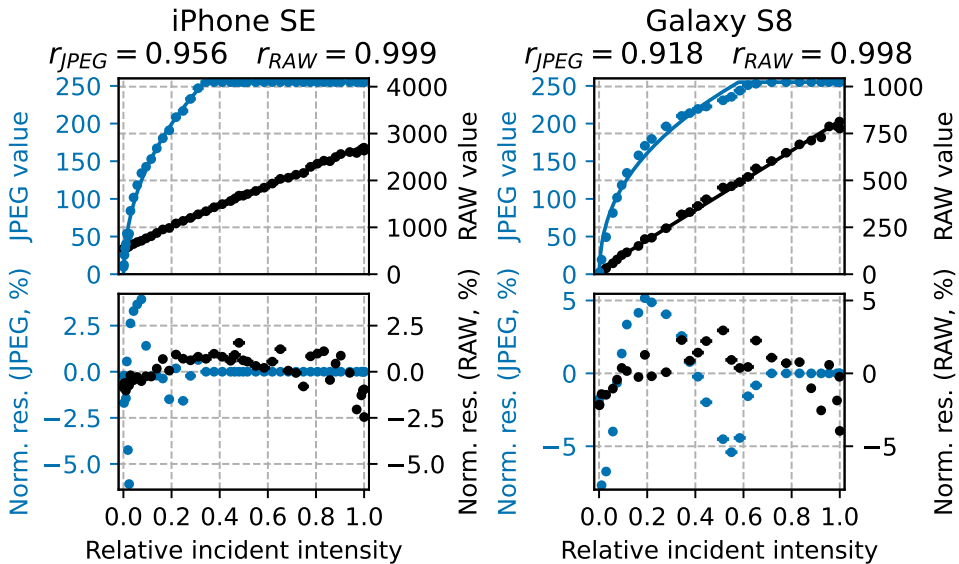


Figure 3.3: JPEG (blue, left vertical axis) and RAW (black, right axis) response of a single B pixel in the iPhone SE (left) and Galaxy S8 (right) rear cameras, under varying incident intensities. Each point represents the mean of a stack of 10 images at the same exposure. Vertical error bars are smaller than the dot size. The black and blue lines represent the best-fitting linear (RAW) and sRGB-like (JPEG) profiles, respectively. The lower row shows the residuals, normalised to the dynamic range.

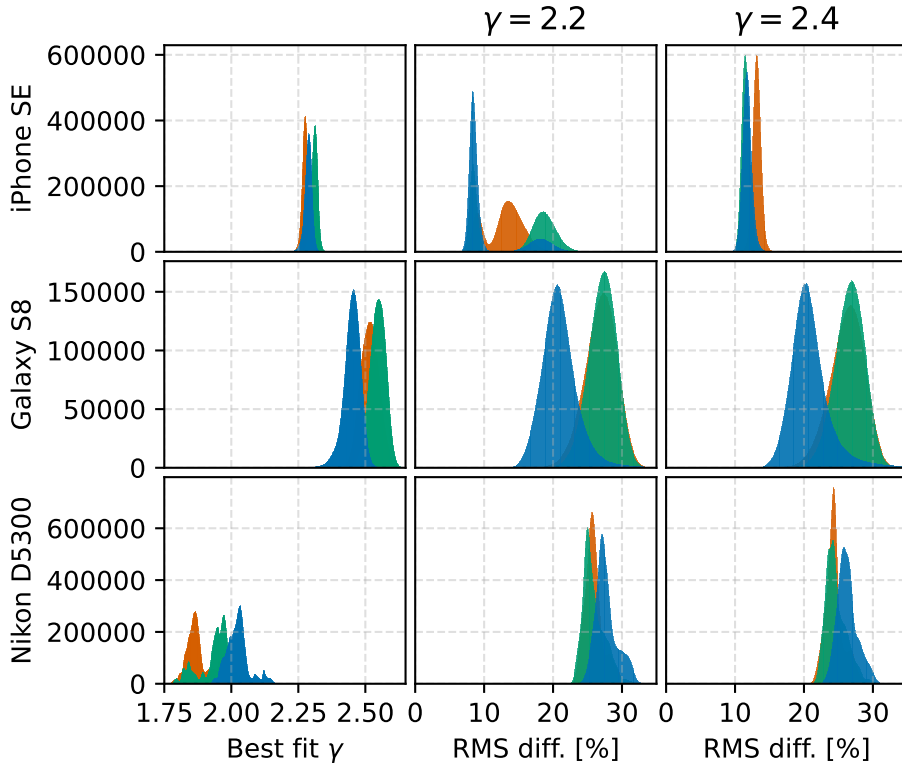


Figure 3.4: Histogram of best-fitting  $\gamma$  and RMS relative difference between JPEG data and fit (for models with  $\gamma = 2.2$  and  $2.4$ ) in the RGB bands.

### 3.4.3 Bias & read noise

Bias and read noise variations in four smartphone cameras (iPhone SE and 7 Plus, Galaxy S6 and S8), one digicam (Nikon D5300), and one UAV camera (Phantom Pro 4) were analysed using the methods from Section 3.3.4.

Bias values in all cameras deviated systematically from the EXIF values by  $<1$  ADU on average, with standard deviations also  $<1$  ADU. However, large outliers were found, such as some pixels in our Galaxy S6 which even saturated in bias frames. Phantom Pro 4 data are scaled up from 12-bit (its native bit depth) to 16-bit, increasing the observed bias variations. Scaled down to 12 bits, its bias variations are similar to those in the other cameras.

Typical observed RON values were distributed similarly to inter-pixel bias variations. The smartphones and D5300 show RON distributions consisting of one or two main components  $<3$  ADU, which correlate with inter-pixel gain variations (Section 3.4.6), and a long but shallow tail towards RON values  $>20$  ADU. As with the bias variations above, the Phantom Pro 4 showed a comparatively high mean RON (14 ADU at ISO speed 100) in 16-bit (scaled-up) data but a comparable value (1.8 ADU) when scaled down to its native bit depth of 12 bits.

Large-scale patterns in inter-pixel and inter-filter bias and RON variations were observed

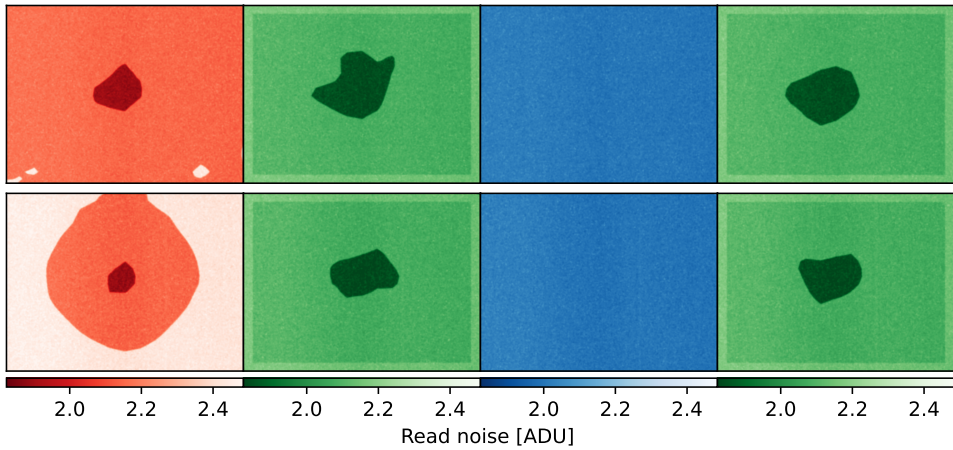


Figure 3.5: Read-out noise per pixel of two iPhone SE devices (top and bottom) at ISO speed 23, in the RGB<sub>2</sub> filters from left to right. Darker colours correspond to lower read-out noise. A two-dimensional Gaussian filter ( $\sigma = 5$  pixels) has been applied to better visualise large-scale variations. The G image shows similar patterns to Figure 3.7.

in several cameras, most prominently in the smartphones. Figure 3.5 shows the RON per pixel in the sensors of two iPhone SE devices. The RON and bias patterns on each device are strongly correlated, suggesting a common origin. The RMS difference in bias between these two devices was 0.31 ADU, larger than the standard deviation on either device (0.24 and 0.21 ADU). The large-scale patterns persisted over time scales of months, indicating that they are systematic.

Both bias variations and RON decreased with ISO speed when normalised (Section 3.3.6). This may be a result of better amplifier or ADC performance at a higher gain. Similarly, large-scale patterns such as those in Figure 3.5 become less distinct at high ISO speeds.

Either a map of mean bias per pixel at a given ISO speed  $B(x, y, ISO)$  or a mean value  $B$  is used in Equation (3.6). For low-light applications such as dark-sky measurements [320] or spectroscopy, a detailed map is necessary since a single ‘bad’ pixel with an abnormally high output may cause a significant systematic error. Being manufacturing defects, bad pixels are in different locations even on two cameras of the same model, and thus a map is required for each device. Conversely, for bright conditions, the bias variations are not significant and thus a mean value can be used. Similarly, RON values can be incorporated in the error budget separately for individual pixels or using the RMS value as an ensemble estimate.

### 3.4.4 Dark current

The methods described in Section 3.3.5 were applied to two smartphones (iPhone SE and Galaxy S8) to measure their dark current properties. Both cameras have built-in dark current calibrations (autodark; see Section 3.2). Measurements were done at room temperature, with short breaks between differently exposed stacks to prevent overheating the sensor. However, sensor temperatures were not obtained from the camera software.

A separate data set consisting of 96 images taken with 4 seconds between each on the iPhone SE, during which the entire device palpably warmed up, was analysed to identify

thermal effects. Pearson  $r$  correlations between response and time stamps (as a proxy for temperature) were calculated for the individual pixels. These  $r$  values were well-described by a normal distribution with  $\mu = 0.00$  and  $\sigma = 0.10$ , indicating that no strong relation exists between temperature and residual dark current. However, we note that again no direct sensor temperatures could be obtained.

In both cameras, a small residual (positive or negative) dark current signal was observed. Most pixels in both cameras had little dark current (RMS  $< 2 \text{ ADU s}^{-1}$ , 99.9th percentile of absolute values  $< 6 \text{ ADU s}^{-1}$ ), though notable outliers were found, such as  $> 300$  pixels in our Galaxy S8 with dark current  $> 50 \text{ ADU s}^{-1}$ . The residual dark current decreased at higher ISO speeds, similar to RON and bias variations (Section 3.4.3), but showed no large-scale patterns.

These results show that autodarkening accurately corrects most pixels, but is inadequate for outliers. Since autodarkening is built into camera chips, it cannot be disabled. For outliers and in low-light conditions, it should be augmented with a manual dark current correction. As with bias variations, the dark current map  $D(x, y, ISO)$  is used in Equation (3.6) for low-light conditions, but an approximation is adequate for bright conditions. For autodarkened cameras like the ones tested here, a mean value of  $D = 0 \text{ ADU s}^{-1}$  is assumed, and the RMS variation incorporated into the error budget. Outliers may be masked in either case.

### 3.4.5 ISO speed

The normalisation of data at different ISO speeds was measured using the methods from Section 3.3.6 on two smartphones (iPhone SE and Galaxy S8) and one digicam (Nikon D5300).

The measured and best-fit normalisation curves are depicted in Figure 3.6. The Nikon D5300 and Galaxy S8 were best fit with a single linear relation, while the iPhone SE curve is clipped at ISO 184. This clipping is not due to image saturation, as none of the pixels in any image reached saturation. The linear part of the iPhone SE relation passes through the origin, while the Nikon D5300 and Galaxy S8 curves do not, instead showing significant ( $> 5\%$ ) systematic errors when using the simplest mathematical model (zero offset and slope  $1/\text{minimum ISO speed}$ ).

The clipping of the iPhone SE curve likely corresponds to a transition from purely analogue to purely digital gain. However, data from the Camera2 API on the Galaxy S8 indicated that it too uses digital gain, at ISO speeds  $> 640$ . This suggests that there are different implementations of gain for RAW photography.

The large observed differences in ISO speed normalisation can lead to significant systematic errors when combining data taken at different ISO speeds, if not adequately calibrated. Data are normalised by dividing by  $N$ , as expressed in Equation (3.6).

### 3.4.6 Gain

The methods from Section 3.3.7 were used to characterise inter-pixel gain variations in two smartphones (iPhone SE and Galaxy S8).

Significant inter-pixel variations were observed, as shown in Figure 3.7 for the G pixels in both cameras. Since the measurement protocol is applied on the individual pixel level, the observed variations are only due to differences in gain, rather than external factors such as vignetting effects. The iPhone SE showed small variations, with higher gain values at the edges and lower values in the center. This pattern is similar to that seen in Figure 3.5, suggesting

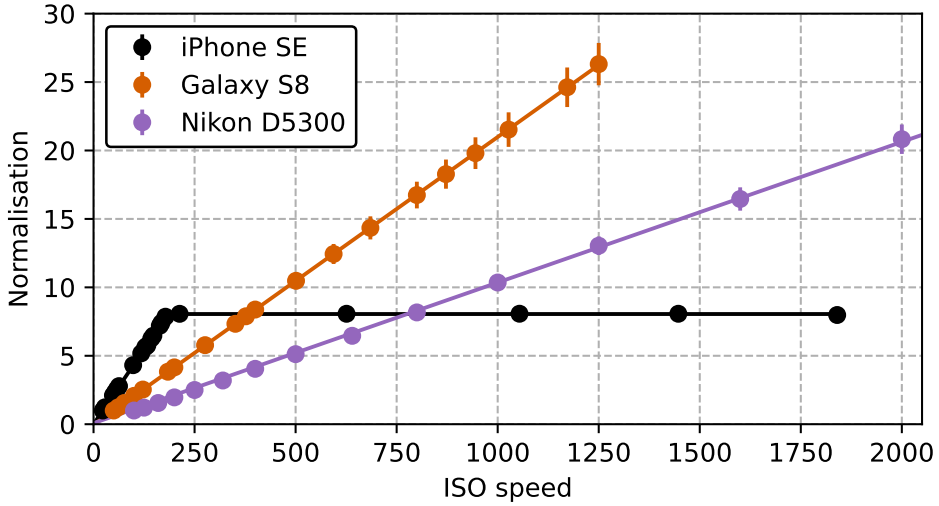


Figure 3.6: ISO speed normalisation for the iPhone SE, Samsung Galaxy S8, and Nikon D5300. Dots indicate means of image stacks divided by the mean value per pixel at the lowest ISO speed. Lines indicate the best-fitting relationships.

a common origin. Conversely, on the Galaxy S8 a concentric pattern with a very wide range is clearly visible, likely intended as a first-order vignetting correction. Both showed similar ranges in gain (0.58–2.74 and 0.59–2.45  $\text{ADU}/e^-$ , respectively), though on the iPhone SE most variations were on small scales and thus are not visible in the smoothed image.

Histograms of gain values for both cameras are shown in Figure 3.8. Inter-filter differences are small in the Galaxy S8 but obvious in the iPhone SE. In particular, the R, G, and B distributions in the latter clearly have different mean values and widths (means and standard deviations of  $1.97 \pm 0.24$ ,  $1.78 \pm 0.29$ , and  $1.73 \pm 0.30$   $\text{ADU}/e^-$ , respectively). Furthermore, the G distribution is bimodal while both others are unimodal; no significant differences be-

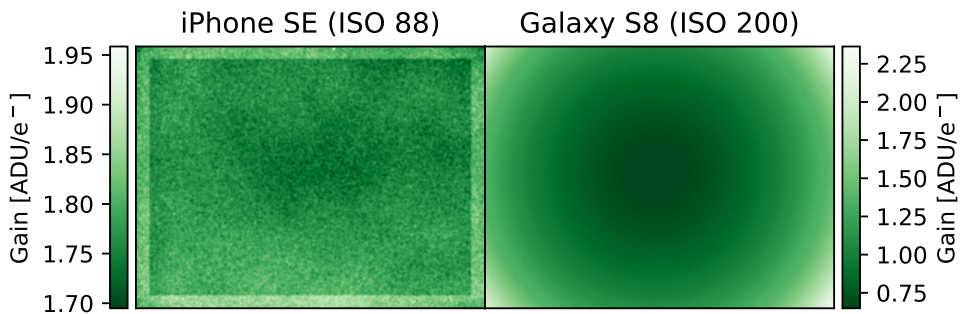


Figure 3.7: Gain values of G pixels in the iPhone SE (left; ISO speed 88) and Galaxy S8 (right; ISO speed 200) sensors. Darker colours indicate lower gain values. A two-dimensional Gaussian filter ( $\sigma = 5$ ) has been applied to better visualise large-scale fluctuations. The iPhone SE patterns are similar to the read noise shown in Figure 3.5.

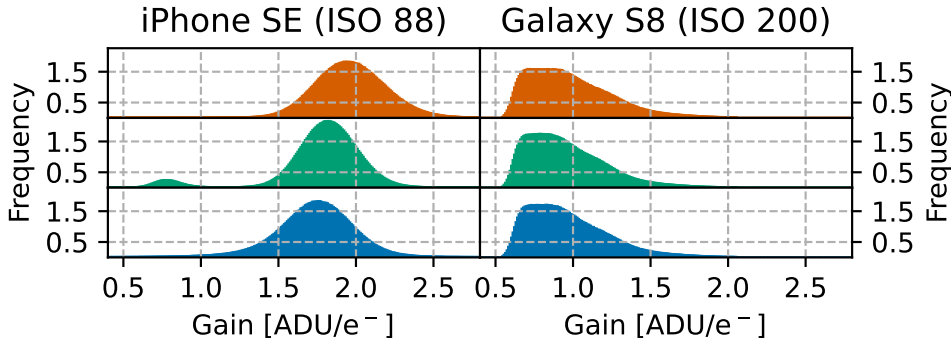


Figure 3.8: Histogram of gain values in the R (top), G and  $G_2$  (middle), and B pixels (bottom) in the iPhone SE (left; ISO speed 88) and Galaxy S8 (right; ISO speed 200) sensors. The vertical axes were normalised to account for the different amounts of pixels.

3

tween the  $G$  and  $G_2$  gain distributions were found, so this is not the cause for the observed bimodality.

The observed gain variations are significant and provide insights into camera design and fabrication, specifically the origins of some of the phenomena seen in flat-field data (Section 3.4.7). However, they are not necessary in the radiometric calibration of data, since our flat-field correction method (Sections 3.3.8 and 3.4.7) accounts for these inter-pixel gain variations as well as all other differences in sensitivity, such as vignetting, as discussed in Section 3.3.10.

### 3.4.7 Flat-field correction

Three smartphones (iPhone SE and 6S, and Galaxy S8) and one UAV (DJI Phantom Pro 4) were flat-fielded using an integrating sphere as described in Section 3.3.8. An aperture of  $f/2.8$  was used for the Phantom Pro 4, and on each device the maximum focus was used. 300 images were taken with the iPhone SE and Galaxy S8, 224 with the Phantom Pro 4, and 30 with the iPhone 6S. The latter was flat-fielded using a different set-up, with a Newport 819D-SF-4 integrating sphere and taking only 30 images as this was sufficient for an SNR  $>3$  in  $>99\%$  of its pixels.

Significant vignetting was found in all cameras. The observed correction factors of the iPhone SE, the best-fitting model, and residuals between the two are shown in Figure 3.9. The smooth pattern suggests optical vignetting is the main flat-field component; the same is true in the iPhone 6S and Galaxy S8. The Phantom Pro 4 data showed an additional steep cut-off near the corners, suggesting mechanical vignetting. To counteract the latter, the outermost 250 pixels on all sides of the images from all cameras were removed prior to further analysis. Correction factors up to 2.42 (iPhone SE), 2.03 (iPhone 6S), 1.43 (Galaxy S8), and 2.79 (Phantom Pro 4) were observed. No significant chromatic differences were found, so the recombined data were used instead of separate RGBG<sub>2</sub> data.

As seen in Figure 3.9, the DNG model fitted the data well with only small residuals remaining. The RMS of the residuals, normalised to the unsmoothed observed values, was 1.5% (iPhone SE), 1.4% (Galaxy S8), 3.1% (iPhone 6S), and 2.0% (Phantom Pro 4). These differences drop to  $\leq 0.7\%$  on all cameras when using the spatially smoothed data, implying

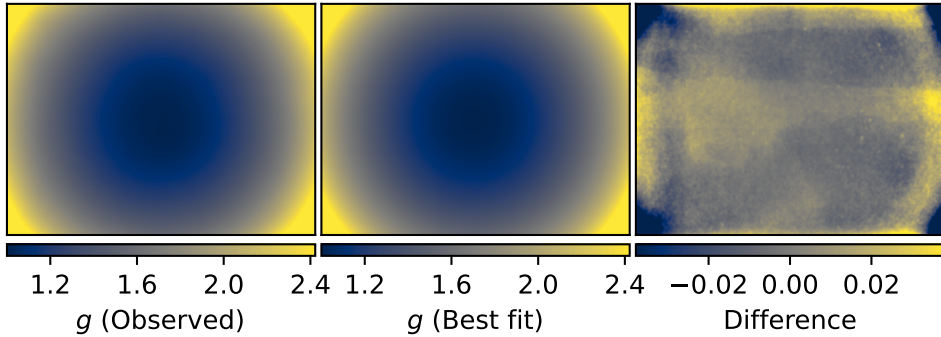


Figure 3.9: Flat-field correction factor  $g$  for the iPhone SE camera. From left to right: observed values (inverse of observed relative sensitivity), best-fitting DNG model, and residuals.

that they are mostly due to pixel-by-pixel variations and noise in the observations. These small residuals show that the DNG model is an adequate approximation for most applications; a pixel-by-pixel map per device is necessary only if sub-percent precision is required. Estimated errors in the model were  $<0.01$  on the polynomial coefficients and  $<10^{-5}$  on the optical center (in relative coordinates) for all cameras. Anomalous dots can be seen throughout the difference image in Figure 3.9, possibly due to dust particles or inter-pixel gain variations (Section 3.4.6).

Since iOS also uses the DNG model for its internal vignetting correction, a direct comparison between correction models for the iPhone SE was made. The RMS relative residual between our smoothed data and the internal model was 5.9%, more than 10 times that of our model (0.5%). While the iOS model is symmetric ( $\hat{c}_x = \hat{c}_y = 0.5$ ), ours had a slight offset ( $\hat{c}_x = 0.494226(1)$  and  $\hat{c}_y = 0.503718(2)$ ). The polynomial coefficients all differed by  $>400\sigma$ , with  $\sigma$  the standard error on our model derived by the fitting routine. Finally, the RMS difference between the models per pixel was 5.7%.

The three alternate methods described in Section 3.3.8 were tested on the Galaxy S8. 40 images of the overcast sky were taken, as well as 40 of the sun and 50 of a monitor with a paper diffuser. The Galaxy S8 was used because its integrating sphere data show a large asymmetry ( $\hat{c}_x = 0.449391(5)$ ,  $\hat{c}_y = 0.426436(9)$ ), providing a simple comparison metric. The RMS difference between the smoothed data from the integrating sphere and alternative methods relative to the sphere data were 4%, 4%, and 5%, respectively. The best-fitting optical centers of all three data sets differed significantly both from the sphere data and from each other ( $\hat{c}_x = 0.53447(1)$ ,  $0.501989(4)$ ,  $0.490794(4)$  and  $\hat{c}_y = 0.38837(2)$ ,  $0.449426(7)$ ,  $0.477590(7)$ , for the sky, sun, and monitor methods, respectively). This causes a typical systematic error on the order of 5% in all three cases. Finally, six replicate measurement sets (50 images each) were taken using the monitor method to assess the effects of nonuniformities in the paper diffusers, generating a correction model for each set. The typical variation, expressed as the RMS of the standard deviation per pixel relative to the mean value per pixel, was 3%, smaller than the typical deviations between the do-it-yourself methods and ground truth data. The effect of paper nonuniformities thus does not significantly impact the quality of do-it-yourself data.

The flat-field correction is incorporated in the radiometric correction expressed in Equation (3.6) as the factor  $g = g(x, y)$ . For cameras with a fixed aperture, such as smartphones,

one measurement is sufficient; otherwise,  $g$  varies with aperture. This corrects for the systematic error induced by flat-fielding effects but pixels at the edges still receive fewer photons than those in the center. The former correspondingly have a smaller SNR due to shot noise, scaling as  $SNR \propto g^{-1/2}$ . Therefore, objects of interest are preferably imaged near the optical center of the camera.

### 3.4.8 Spectral response

Two smartphones (iPhone SE and Galaxy S8) and one UAV (DJI Phantom Pro 4) were spectrally calibrated using a monochromator, and the iPhone SE using iSPEX, as described in Section 3.3.9.

Figure 3.10 shows the normalised spectral response curves derived from the monochromator data, calibrated to the spectral throughput of the monochromator and spectral irradiance of the light source. This calibration was done by measuring its output under the same conditions as during the measurements, using a pre-calibrated silicon photodiode. Parts of the spectra were measured with different exposure settings and monochromator filters; these were first calibrated and then normalised and averaged on overlapping sections. The peak response wavelengths and effective bandwidths of the RGBG<sub>2</sub> filters in the different cameras are given in Table 3.1.

Some similarities and differences between the cameras are immediately obvious from Figure 3.10 and Table 3.1. Notably, no significant differences between G and G<sub>2</sub> were found in any camera (RMS differences  $\leq 0.003$ ); the different peak wavelength for the Phantom Pro 4 is likely due to noise. The peak response wavelengths are very similar or even identical between cameras, as are the effective bandwidths, with two notable exceptions. The Galaxy S8 B filter is significantly broader than the others, with a comparatively high response at  $\lambda > 500$  nm. Conversely, the Phantom Pro 4 has a relatively narrow R filters due to its NIR cut-off around 670 nm rather than 680 nm. Moreover, the R filters in all three cameras show a secondary peak around 535 nm and nearly identical responses between 570–650 nm.

The spectral response curves measured with iSPEX, shown in Figure 3.11, were simi-

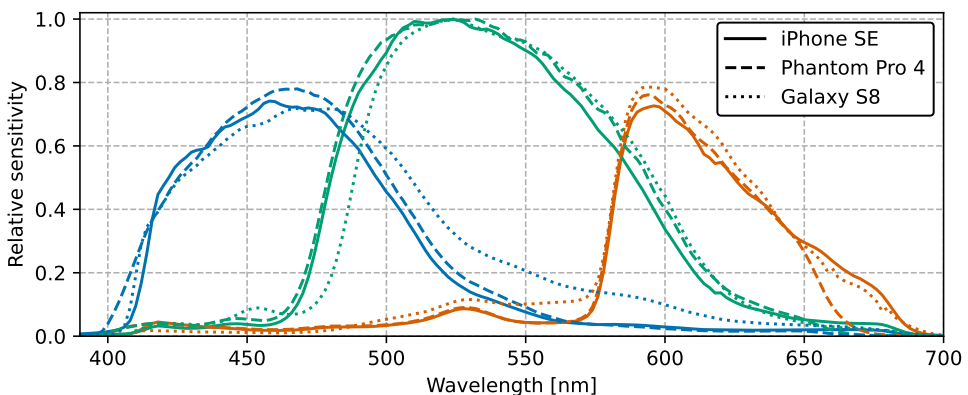


Figure 3.10: Spectral response curves of the iPhone SE, Galaxy S8, and Phantom Pro 4, derived from monochromator data. The responses are normalised to the global maximum per camera, giving relative sensitivities. G is the average of the G and G<sub>2</sub> responses over the wavelength axis, since no significant differences were found. RMS errors are  $\leq 0.005$ .

Camera	$\lambda_{P,R}$	$\Delta_R$	$\lambda_{P,G}$	$\Delta_G$	$\lambda_{P,G2}$	$\Delta_{G2}$	$\lambda_{P,B}$	$\Delta_B$
iPhone SE	596	72	524	110	524	109	458	93
Galaxy S8	594	73	524	109	524	109	468	117
Phantom Pro 4	594	65	524	115	532	116	468	94

Table 3.1: Peak response wavelength  $\lambda_{P,C}$  and effective spectral bandwidth  $\Delta_C$  of each filter in the three cameras, derived from monochromator measurements. All values are in nm.

lar to those derived from the monochromator data but showed small though significant systematic differences. No significant differences were found between narrow- and broad-slit spectra, so these were averaged. RMS differences between iSPEX- and monochromator-derived responses were 0.04, 0.02, and 0.02 (SMARTS2 normalisation) and 0.12, 0.10, and 0.10 (black-body normalisation), in RGB respectively. The black-body under-estimated the irradiance  $<500$  nm and over-estimated it  $>500$  nm compared to the SMARTS2 model, resulting in large deviations in the retrieved spectral response. The RMS difference between the monochromator-derived and black-body-normalised iSPEX-derived spectral responses could be reduced to 0.05, 0.10, and 0.04 by multiplying each filter with an empirical constant. However, systematic differences  $>0.2$  remained in the G filter at wavelengths of 500–600 nm. Conversely, the SMARTS2-normalised iSPEX-derived spectral responses only showed a significant systematic difference compared to monochromator data at wavelengths  $>650$  nm, the origins of which are unclear.

The observed differences between devices have important implications for RGB colour measurements and spectroscopy, for example for colour measurements as discussed in Section 3.1. The effective spectral bandwidths are incorporated into the radiometric calibration of imaging data as described in Section 3.3.10. Furthermore, smartphone spectrometers naturally require calibration for the spectral response of the camera, as expressed in Equation (3.7).

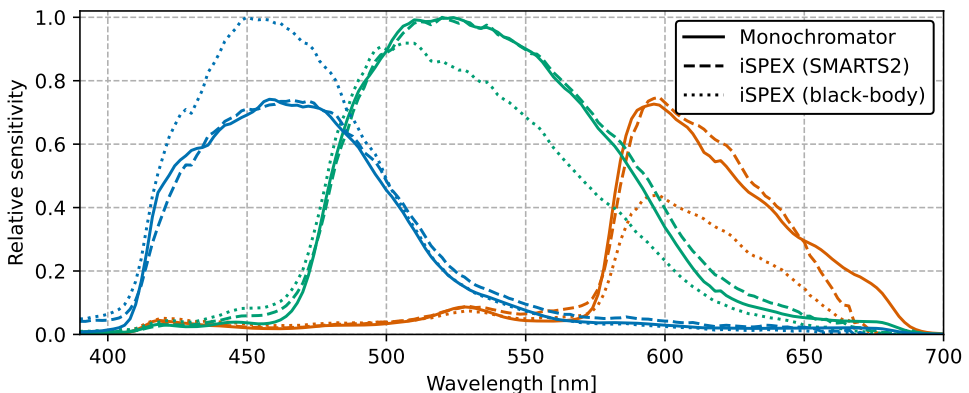


Figure 3.11: Comparison of the iPhone SE spectral response curves measured with the monochromator and iSPEX. iSPEX data are normalised using a 5777 K black-body and a SMARTS2 model, as described in Section 3.3.9.

### 3.4.9 SPECTACLE database

To facilitate the use of consumer cameras in scientific projects and improve future compatibility, we have created the SPECTACLE (Standardised Photographic Equipment Calibration Technique And CataLoguE) database. It includes the calibration data required for radiometric corrections (Section 3.3.10), for the most popular cameras. The data are given in standardised formats, split into three categories (device, camera, and software) to minimise the amount of data required. For example, two devices using the same camera module have the same spectral response curves and flat-field response, while software parameters such as bias and ISO speed settings vary. The former can thus be combined while keeping the latter separate. Since the properties of a camera may change with firmware updates or changes in manufacturing, database entries may be split according to device version, rather than assuming devices of the same model are clones. Finally, given calibration data for multiple identical devices, statistics on variations within a camera model may be included. The open design of the SPECTACLE database, based on the Parse platform, allows anyone to use or contribute data, particularly using the calibration apps we have developed. Submitted data are currently curated by the authors to ensure their quality. As the database grows, community curation or automated curation based on outlier analysis may become preferable. SPECTACLE can be accessed at <http://spectacle.ddq.nl/>.

## 3.5 Discussion & conclusions

In this work, we have presented a standardised calibration methodology for the most important factors limiting the quality of consumer camera data, the first to our knowledge. Furthermore, we have developed the SPECTACLE database, containing calibration data for the most popular devices. The standardised methodology and the SPECTACLE database have the potential to improve the sustainability of projects using these cameras, by simplifying their calibration and the use of multiple camera models.

The main difference between our approach and those in much of the literature is the use of RAW data. Software constraints previously forced the use of JPEG data, which are compressed and heavily processed, introducing systematic effects that negatively affect the data quality and are difficult to calibrate [94, 121, 142, 285, 319, 334, 342, 349, 357]. The desire to use RAW data has been expressed widely in the literature [118, 121, 285, 329, 331, 334, 349, 351], and their superiority is clearly demonstrated by the highly linear response and larger dynamic range found in Section 3.4.2. The former is especially notable since non-linearity and the associated gamma correction are among the most cited problems of JPEG data [94, 121, 162, 285, 317, 322, 329, 332, 334, 338, 342, 357–360, 362]. While JPEG non-linearity corrections exist, either fully empirical or based on the sRGB standard [285, 342, 358], the wide (1.7–2.6) variations in gamma and large (>30%) deviations from sRGB profiles shown in Section 3.4.2 and Figure 3.4 indicate that these are inaccurate and difficult to generalise. The highly linear nature of RAW data was previously demonstrated in [118, 361, 366] and may be a result of internal linearity corrections in the CMOS chip [104]. Furthermore, RAW data are not affected by white balance, a colour correction in JPEG processing which severely affects colourimetric measurements, is difficult to calibrate, and differs strongly between measurements and cameras [56, 94, 285, 317, 318, 323, 342, 349, 351, 376]. This variable gamma correction and white balance make it impossible to invert the JPEG algorithm and recover RAW data. However, RAW data are no panacea, since they still require further calibrations.

Furthermore, not all consumer cameras support RAW imagery, especially low-end smartphones; hence the low adoption rate in literature until now. Still, we consider the linearity, larger dynamic range, and lack of unknown post-processing affecting the data worth relying on RAW data, especially in a market trending towards broader support.

Inter-pixel and inter-device bias variations and read noise were found to be small in general ( $\sigma < 1$  for bias variations, mean RON  $< 3$  ADU), though with large outliers (Section 3.4.3). These distributions are similar to those found in several smartphones [324] and a Raspberry Pi camera [118], though neither work distinguishes between bias variations, read noise, and dark current. The large-scale patterns seen in Figure 3.5 were not found in the literature. Their cause is unclear, though correlations with inter-pixel gain variations (Section 3.4.6) suggest a common origin. Ultimately, since both phenomena are small, for most applications these patterns are merely a curiosity and an estimate in the error budget and masking of outliers is sufficient for further radiometric calibrations (Section 3.3.10).

While dark current has been implicated in the literature as a major noise source [118, 319, 324, 344, 354, 363, 364], the results presented in Section 3.4.4 indicate that it is actually typically quite minor. The RMS dark current in the iPhone SE and Galaxy S8 ( $< 2$  ADU  $s^{-1}$ ) is similar to values found in [319, 358, 363, 364], while we found larger outliers, such as  $> 300$  pixels with  $> 50$  ADU  $s^{-1}$  in our Galaxy S8. Similarly to [364], no significant relationship was found between temperature and residual dark current, though this experiment should be repeated under more controlled conditions and using internal sensor temperatures to draw strong conclusions. In general, a quantitative comparison with the literature is difficult, since those studies used JPEG data, not RAW. While our sample of two cameras is insufficient to draw broad conclusions, these results suggest that dark current is less important than previously thought. As discussed in Section 3.4.4 and similarly to the aforementioned bias and RON variations, extensive characterisation of the dark current in individual pixels is necessary for low-light applications and spectroscopy as these are significantly affected by a few ‘bad’ pixels. Conversely, for bright-light conditions the dark response is typically negligible and an ensemble estimate in the error budget and masking of outliers are sufficient.

ISO speed normalisation is typically done by simply dividing digital values by the ISO speed [121, 367], but the results presented in Section 3.4.5 and Figure 3.6 contradict the validity of this method. This discrepancy was also identified in [368]. Observed relations differ significantly from the naïve linear model in shape, offset and slope. For example, differences between the two models of  $> 5\%$  were found in the Galaxy S8. More extremely, the expected and observed normalisation factor at ISO speed 1840 on the iPhone SE differ by a factor of 10. Moreover, Android documentation suggests that more complex curves with mixed analogue and digital gain may also be in use. Thus, to prevent similar systematic errors, either a single ISO speed per device must be used or these relations must be calibrated.

Significant inter-pixel gain variations were found in Section 3.4.6, as shown in Figures 3.7 and 3.8. The Galaxy S8 showed a strong radial pattern, likely intended as a first-order vignetting correction; this was not seen in the iPhone SE. Conversely, gain values in the latter differed significantly between colour filters. This may be a colour correction called analogue white balance, which is described in the DNG standard [369]; however, in this case it is not clear why significant inter-pixel variations exist. No previous discussion of such variations in gain in a consumer camera was found in the literature. Typically, an equal gain in all pixels is assumed in absolute radiometric calibrations [118, 368] but the variations found here cast doubt on the generality of this assumption.

Strong flat-field effects were found in Section 3.4.7, with correction factors up to 2.79.

Similarly large correction factors have been found for other cameras, for instance approximately 2.8 in a Canon EOS 400D [329] and 650D [330], 4 in a Raspberry Pi camera [118], 1.8 in a Canon EOS 10D [374], and 1.5 in a Nikon E775 [375]. It should be noted that vignetting is highly aperture-dependent and thus these correction factors will change with varying apertures [374]. Interestingly, we did not find the large chromatic differences described in [118, 329]. Notably, the Galaxy S8 showed a much weaker vignetting effect ( $g_{max} = 1.43$ ) than the other cameras ( $g_{max} > 2$ ), likely due to the aforementioned inter-pixel gain variations. These may also explain the strong asymmetry ( $\hat{c}_x = 0.449391(5)$ ,  $\hat{c}_y = 0.426436(9)$ ) seen in the Galaxy S8, due to the main symmetrical component having been corrected already.

The 7-parameter vignetting model described in the DNG standard [369] fits our data very well (RMSE  $\leq 3.1\%$  for raw data,  $\leq 0.7\%$  for smoothed data), without significant systematic differences. Since the typical difference between observed and modeled corrections is small, pixel-by-pixel flat-fielding is necessary only for applications requiring sub-percent precision. For those, a flat-field map would be made for each individual device, rather than using the same map for multiple devices of the same model. Flat-field measurements of the latter could be used to quantify typical variations in flat-field response among identical devices and further determine when pixel-by-pixel or modeled flat-field corrections are preferable. The DNG model is also used for internal vignetting correction in iOS. While this correction is sometimes considered a major advantage of JPEG data over RAW data, the internal model of the iPhone SE was shown to be significantly less accurate (RMSE = 5.9%) than one based on our data (RMSE = 0.5%), contradicting this notion. Moreover, residual vignetting effects up to 15% have been observed in JPEG data [358]. A comparison to the internal correction data in Android smartphones, consisting of pixel-by-pixel look-up tables, has not yet been done since these data are relatively difficult to access.

Finally, three simpler alternative flat-fielding methods were tested, namely imaging the sky, the sun, and a computer monitor, as described in Section 3.3.8. Applied on the Galaxy S8, data from these methods differed from the integrating sphere data by  $\leq 5\%$  RMS. These errors mainly result from a difference in the location of the optical center. The cause of these discrepancies is unclear, though insufficiently isotropic light sources are an obvious explanation. Nevertheless, the RMS difference of  $\leq 5\%$  is small compared to the overall flat-field correction of up to 179% and better than the internal correction of the iPhone (RMS 5.9%). These methods thus serve as a useful first estimate for the flat-field correction in the absence of integrating sphere data. As discussed in Section 3.2, many further alternative flat-fielding methods exist [329, 330, 373–377]. Our data may be useful as a ground truth for a thorough comparison of such methods akin to [330, 361].

The spectral responses found in Section 3.4.8 and shown in Figure 3.10 agree well with those found in the literature [118, 121, 317, 323, 329, 333, 334, 354, 358, 361, 365–368], with the RGB curves centered around 600, 520, and 470 nm, respectively. Notably, the strong secondary peaks seen in [121, 358] were not found in our data and may be JPEG artefacts. Differences are mainly found in the wings, such as the NIR cut-offs [329, 354] and harmonics. The comparatively high response of the Galaxy S8 B filter at wavelengths  $>500$  nm is also seen in the Nokia N900 [333] and Sony A7SII [317], and to a lesser extent the Galaxy S5 [121], but is otherwise uncommon. The early NIR cut-off of the Phantom Pro 4 appears to be similarly uncommon but not unique [121, 317, 333, 354]. These differences again show the importance of spectral characterisation for normalising smartphone spectrometer data. Furthermore, the significant variations show that the common assumption of sRGB responses [285, 332] does not hold, as has been suggested previously [331], and characteri-

sation of the spectral response is necessary to convert observed colours to colour spaces such as CIE 1931 XYZ or CIELAB [333, 335]. However, colour measurements still depend on the incident light spectrum [335]; hyperspectral measurements, for example with iSPEX [94], and characterisation of common light sources [317, 381] may provide valuable additional information. Finally, while no significant response was found at wavelengths <390 or >700 nm on our test cameras, it may be worthwhile in the future and the SPECTACLE database to use a spectral range of 380–780 nm to follow colourimetric standards [335, 359, 363].

Spectral response measurements done with the iSPEX smartphone spectrometer [94] agreed well (RMS differences  $\leq 0.04$ ) with the monochromator measurements (Section 3.4.8 and Figure 3.11). The only systematic difference was an under-estimation at wavelengths >650 nm, though it is unclear what causes this. The good agreement shows that iSPEX measurements are an adequate replacement for monochromator data if the latter are not available. This will be especially useful with the new iSPEX we are developing, which will also feature universal smartphone hardware interface. One downside of this method is that it requires an accurate solar reference spectrum. We used one generated with SMARTS2 [379, 380]; this model matches observed solar spectra very well but it is not very portable or user-friendly for non-expert users. A 5777 K black-body approximation was also used but reproduced the SMARTS2 spectrum poorly (RMSE of 34%) and accurate spectral response curves could not be retrieved this way. A more portable model or set of standard spectra could improve the user-friendliness of this calibration method.

Further alternative methods for spectral response characterisation include those based on multispectral measurements using computational methods to enhance their resolution [335, 372, 384, 385] or those using a linear variable edge filter [334]. However, the former are not sufficiently accurate [366] while the latter is not necessarily more accessible than a monochromator. Our data may be used as a ground-truth for testing other methods akin to [366] but with the advantage of smartphones being more accessible than the cameras used therein.

Finally, we have created the SPECTACLE database containing the calibration data described above. The aim of this database is to facilitate the use of consumer cameras in scientific projects by reducing the labor required for calibration. Data sets containing spectral responses [333, 366] and extensive calibrations of single cameras [118] have been published before but to our knowledge SPECTACLE is the first comprehensive, centralised spectral and radiometric calibration database. It is designed with community participation in mind, relying on volunteer contributions to become and remain complete in the rapidly evolving camera market. This will require a critical mass of users to maintain it, which is easier if more accessible calibration methods, like those discussed previously, can be used. We have kick-started this process with the calibrations done in this paper and will continue this while developing iSPEX.

Though extensive, our calibration methodology is not complete. The two most prominent missing components are geometric distortions and absolute radiometric calibration. The former are a well-known phenomenon with a large impact on image quality but relatively simple to measure and correct [322, 326–328, 353, 377]. A parametric model for distortion is given in the DNG standard [369] and a comparison between measured distortions and the internal correction models of different cameras, similar to that done in Section 3.4.7 for vignetting corrections, may be used to determine the accuracy of the latter. Absolute radiometric calibration is extremely valuable for quantitative measurements, as described in Section 3.3.10. In principle, our methods and calibration data contain most of the information

required for this, bar a constant  $K$ . Absolute radiometric calibration of consumer cameras has been demonstrated before, notably in the Raspberry Pi camera [118], and Nikon D300 and Canon 40D [368], though only for a small number of devices. Another notable example is the Hukseflux Pyranometer app (Section 3.3.8) for measurements of solar irradiance, though it is intended for education and entertainment rather than scientific measurements. Finally, most of our calibrations were done on a single device, and differences between devices may exist, as shown in Figure 3.5. Calibration of multiple devices per camera model would allow the characterisation of these differences and the associated errors when using multiple devices. Additionally, differences may be introduced by changes in manufacturing or camera software. Characterisation of different generations of the same model camera will be necessary to characterise these, which may result in separate entries in the SPECTACLE database for each camera version being necessary. However, the modular design of the SPECTACLE database makes it simple to extend. The simple, standardised calibration methods described in this work and the SPECTACLE database have the potential to greatly improve the data quality and sustainability of future scientific projects using consumer cameras.

## 3

## Funding

European Commission Horizon 2020 programme (grant nr. 776480, MONOCLE and grant nr. 824603, ACTION).

## Acknowledgements

The authors wish to thank Molly and Chris MacLellan of the NERC Field Spectroscopy Facility for experimental help and invaluable insights in the flat-field and spectral response measurements. Figure 3.1 was drawn using *draw.io*. Data analysis and visualisation were done using the AstroPy, ExifRead, Matplotlib, NumPy, RawPy, and SciPy libraries for Python. Finally, the authors wish to thank the two anonymous reviewers for their thorough and constructive reviews.