Principal component analysis of reference sites used for calibration and validation of Earth observation satellites

Joaquin Campos¹, Alejandro Ferrero¹, Emma Woolliams², Claire Greenwell², Agnieszka Bialek², Luisa Hernanz¹ and Alicia Pons¹

¹Instituto de Optica (IO-CSIC), Madrid, Spain. ²National Physical laboratory (NPL) Teddington, United Kingdom

Abstract. Determining reflectance factor and its variability across reference sites for Earth observation satellites is a problem involving large amounts of data and measurement time. Principal component analysis (PCA) may be used to simplify this problem by reducing the size of the data and by highlighting spectral features that could be related to physical phenomena. This work presents the results obtained in applying PCA to two reference sites for calibration and validation of Earth observation satellites located at La Crau (France) and Gobabeb (Namibia), respectively.

1. Introduction
The Earth observation community requires the use of terrestrial reference sites to validate and monitor the performance of satellite sensors. Ideally these should be stable long term and if used for calibration, be traceably characterized. The Committee on Earth Observation Satellites (CEOS) has identified an initial set of reference targets for the community: Pseudo Invariant Calibration Sites (PICS) e.g. Saharan deserts, which have long term radiometric stability but are inaccessible to direct measurements by ground teams and a new network of sites which have autonomous instrumentation continually measuring the surface reflectance and atmospheric conditions, called RadCalNet. Because satellites observe with a resolution which is typically 10 m – 300 m and do not have the geometric knowledge to ensure exact co-location spatial, the spatial uniformity of such sites is critical, as is knowledge of the spectral properties of the surface.

To analyze the ground uniformity, measurements are made across the area in a sampling strategy that is consistent with the spatial resolution of the satellites. For each surface element measured we determine the reflectance factor \( \rho(x_i,y_i,\lambda_j) \), where \((x_i,y_i)\) are the ground coordinates of the centre of the surface element and \(\lambda_j\) is the wavelength that characterizes the spectral band analyzed.

Principal Component Analysis (PCA) [1] is a powerful statistical technique that can be used to identify and quantify uncorrelated contributions to the total variance of a collection of data. This method has been successfully applied to different scenarios, providing physical insight about the origin of noise contributions and extracting meaningful signals from a wide variety of situations [2, 3].

2. Principal component analysis
If the number of surface elements of a reference site is \( N \) and the number of spectral values in each spectrum is \( L \), the total number of data points to be handled is \( N \cdot L \). The PCA method produces three
types of elements: $N$ eigenvalues, $\gamma_j$, $N$ eigenvectors, $c_{ij}$, and $N$ eigenspectra, $H_j$. The eigenvalues quantify the importance and the contribution to the total data variance of their associated eigenspectra. The eigenvectors can be seen as the coefficients of the transformation from the correlated variables given by the spectra to a new set of uncorrelated variables expressed by the eigenspectra.

Following PCA, spectral reflectance factor values can be recovered as:

$$\rho_r(x_i, y_i, \lambda_i) = \left(\rho_r(x_i, y_i)\right)_1 \cdot \left[1 + \sum_{j=1}^{M} c_j(x_i, y_i) \cdot H_j(\lambda_i)\right]$$

where $\left(\rho_r(x_i, y_i)\right)_1$ is the spectral average on the surface element located at $(x_i, y_i)$ and $M$ is the number of eigenspectra needed to reproduce the reflectance factor spectrum. Here, $H_j$ and $c_j(x_i, y_i)$ were rescaled so that the standard deviation of $H_j$ is 1. This way, the value of $c_j(x_i, y_i)$ quantifies the contribution of every principal component at the locations. This approach also yields the most significant results for PCA, which performs best in a null-average-value data set.

The dimensionality of the problem will be significantly reduced if pattern spectra can be found underlying the reflectance factor values, making it unnecessary to use all the eigenspectra (i.e. $M << N$). This is likely to be the case for satellite reference sites because they are chosen as areas with relatively uniform ground reflectance. The number $M$ has to be determined by taking into account the contribution to the total variance of each eigenspectra given by $\gamma_j$ and the uncertainty associated with the calibration.

3. Calibration sites studied

PCA analysis has been applied to two reference sites: La Crau (France), where 14 surface elements were selected, and Gobabeb (Namibia), where 16 surface elements were selected. La Crau and Gobabeb are two of the four prototype sites for the new CEOS Radiometric Calibration Network (RadCalNet); La Crau is an established site that has been operated by CNES since 1987 [4], while Gobabeb is a new site being developed by ESA, CNES and NPL [5]. The spatial distribution of the surface elements measured in this work are shown in figure 1, left for La Crau and right for Gobabeb. The measurements at Gobabeb are separated into clusters since the measurements were gathered during the initial characterisation of the site, meaning measurements were needed over a wide area, unlike La Crau, where this was a routine check for an established site. Different colours in the Gobabeb site image correspond to different sets of measurements, taken to represent different parts of the site.

A typical relative spectral distribution of the reflectance factor at a surface element of each site is shown in the Figures 2 and 3. The measured values in the noisy spectral region around 1850 nm were excluded from the analysis described in this paper (this is likely to be due to changes in the atmospheric conditions). The spectral properties are qualitatively different between the two sites. At La Crau, (figure 2) the spectral data decay smoothly from the maximum (excluding the noisy region around 1850 nm)
towards both spectral ends, while at Gobabeb, the spectral reflectance is flatter over a wider spectral region, with a faster spectral decay at both spectral ends.

![Graph of relative spectral distribution of reflectance factor at La Crau (France)](image)

**Figure 2. Typical relative spectral distribution of reflectance factor at La Crau (France)**

![Graph of relative spectral distribution of reflectance factor at Gobabeb (Namibia)](image)

**Figure 3. Typical spectral distribution of reflectance factor at Gobabeb (Namibia)**

### 4. PCA results for reference sites

The first step of PCA analysis is to look at the eigenvalues $\gamma$ of the uncorrelated components and calculate their contribution to the total variability of the data, because this will show how many principal components need to be considered. Table 1 shows these contributions for both calibration sites.

|     | La Crau             | Gobabeb            |
|-----|---------------------|--------------------|
| $\gamma_1$ | 9.93·10^{-01} | 9.97·10^{-01} |
| $\gamma_2$ | 2.93·10^{-03} | 2.97·10^{-03} |
| $\gamma_3$ | 2.15·10^{-03} | 2.46·10^{-03} |
| $\gamma_4$ | 7.75·10^{-04} | 5.03·10^{-05} |
| $\gamma_5$ | 3.19·10^{-04} | 3.50·10^{-05} |
| $\gamma_6$ | 1.37·10^{-04} | 2.34·10^{-05} |
| $\gamma_7$ | 9.87·10^{-05} | 2.22·10^{-05} |
| $\gamma_8$ | 7.03·10^{-05} | 1.72·10^{-05} |
According to Table 1, the number of eigenspectra needed to reproduce spectral reflectance factors with a residual error lower than 0.1 % (contribution lower than $10^{-3}$) is four for La Crau and two for Gobabeb, rather than having a different spectrum for every point, which implies that $M$ (the number of spectra needed to achieve a particular uncertainty) is significantly smaller than the original number of observations. In both sites the first principal component is dominant and other components are relatively minor variations. This is to be expected for relatively uniform sites. The Gobabeb site has less spectral variation between its surface elements than La Crau because a lower number of principal components are needed to achieve a certain residual error level.

The relevant principal components obtained for each calibration site are shown in figures 4 and 5. In both sites the main component is very similar to the typical spectra shown in figures 2 and 3, respectively, except for the peaks.

At the La Crau site (Fig. 4) PC2, PC3 and PC4 are needed to explain the variability across the ground in the visible range, which is likely due to the change in vegetation. Furthermore, PC2 has some spectral features in the NIR that can be related to small changes in chemical composition of the ground. Finally, PC3 and PC4 also have two peaks each that could be related to changes in humidity on the ground (peak about 1400 nm due to water absorption) and the noisy peak around 1850 nm found in the raw spectra. This kind of spectral variability has an influence on the spectral behaviour of the uncertainty in the calibration of Earth observation instruments. Note that vegetation and water content may change in a surface along the time.

At Gobabeb’s site (Fig. 5) the spectral variability across the ground is much less. Only 2 principal components are relevant. The ground is highly uniform in the NIR where only the peak at about 1850 nm seems to have more variability. PC2 is needed to give explanation for that variability, and for changes in wavelengths below 700 nm, which could also be due to varying vegetation levels. The variation is
smaller than for La Crau (the vegetation is sparser and drier than at La Crau), so this site has several benefits for spectral calibration compared to the La Crau site.

![Figure 5](image_url)

Figure 5. Relevant principal components for the calibration site at Gobabeb

Principal Component Analysis determines coefficients $c_j(x_i, y_i)$ to be used to reconstruct the reflectance factor values at each calibration position across the site. The variation of these calibration factors shows the contribution to the spectral variance of each component across the site. Fig. 6 shows $c_j(x_i, y_i)$ for each of the four principal components at each of the 14 measured locations at La Crau (spatially arranged in a grid pattern, similarly to what is shown in Figure 1, but rotated to the left); no data were available for the grid positions marked as N/A. From figure 6 it can be concluded that spatial variability at La Crau does not follow any apparent spatial pattern since no relation can be established between the coefficient values at different positions. This is consistent with the conclusion that the origin of the spectral variability is variations in vegetation levels, which is a variability on a small spatial scale on that site.

![Figure 6](image_url)

Figure 6. Spatial distribution of the coefficients for every relevant principal component at La Crau

The coefficients for the two relevant principal components at Gobabeb are shown in figure 7. In this site the coefficient of PC2 has a spatial pattern: surface elements grouped under the colour names (see Fig. 2) have the same sign. The correlation is stronger for the groups “blue”, “green” and “red” (colours...
relate to those in Fig. 1), because the coefficient is almost the same for every surface element in the group. This shows up larger-spatial-scale variability across the site, from one group to the next, (albeit at a smaller absolute level of variation compared to that at La Crau).

Figure 7. Coefficients for every relevant principal component at Gobabeb

5. Conclusions
PCA shows patterns hidden in the data set that may be used to analyze spectral and spatial variability between the surface elements of the reference site and to get the number of independent spectra involved in the problem. Spectral features varying across the site may be relevant to evaluate calibration uncertainty values.

PCA may help to identify physical processes influencing the reflectance factor of surface elements by associating spectral features to physical phenomena. It may help to study temporal evolution of identified phenomena if the measurements are repeated over time.

References
[1] J. L. Simonds, “Application of characteristic vector analysis to photographic and optical response data,” J. Opt. Soc. Am. 53, 968–971 (1963).
[2] J. M. Lopez-Alonso, J. Alda, and E. Bernabeu, “Principal-component characterization of noise for infrared images,” Appl. Opt. 41, 320–331 (2002).
[3] A. Ferrero, J. Alda, J. Campos, J. M. López-Alonso, and A. Pons, “Principal components analysis of the photoresponse nonuniformity of a matrix detector”, Appl. Opt. 46, 9-17 (2007).
A. Ferrero, J. Campos, A.M. Rahal, A. Pons, M. L. Hernanz, A. Corróns, “Principal Components Analysis on the spectral Bidirectional Reflectance Distribution Function of Ceramic Colour Standards”, Opt. Exp. Vol. 19, No. 20, pag. 19199-19211 (2011).
[4] A. Meygret, R. Santer, B. Berthelot, “ROSAS: A Robotic Station for Atmosphere and Surf face characterization dedicated to on-orbit calibration”, Proc of SPIE, vol 8153, p 815311-1, (2011).
[5] A. Bialek, CL. Greenwell, M. Lamare, A. Meygret, S. Marcq, S. Lacherade, ER. Woolliams, B. Berthelot, M. Bouvet, M. King, C. Underwood, N. Fox, “New radiometric calibration site located at Gobabeb, Namib desert”, IGARSS 2016, p. 6094-6097, doi:10.1109/IGARSS.2016.7730592, (2016)

Acknowledgements
To the European Metrology Research Program (EMRP), supporter of this work within the joint research project ENV53 “European metrology for Earth observation and climate” (MetEOC2). The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. We
thank CNES for access to La Crau and their support with the measurements at both La Crau and Gobabeb. The IO-CSIC authors are also grateful to Comunidad de Madrid for funding the program SINFOTON-CM: S2013/MIT-2790.