Principal component analysis of reference sites for Earth observation satellites

Joaquin Campos[1], Alejandro Ferrero[1], Emma Woolliams[2], Claire Greenwell[2], Agnieszka Bialek[2], Luisa Hernanz[1] and Alicia Pons[1]

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

Corresponding e-mail address: joaquin.campos@csic.es

Determining reflectance factor and its variability across reference sites for Earth observation satellites is a problem involving large amounts of data. 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 in La Crau (France) and Gobabeb (Namibia).

INTRODUCTION

Satellite reference sites are used for satellite calibration and validation and are either considered sufficiently stable long-term (pseudo-invariant calibration sites, e.g. deserts), or are instrumented to monitor surface reflectance and atmospheric condition (RadCalNet). Because satellites observe with a resolution which is typically 10 m – 300 m, the spatial uniformity of such sites is critical as is knowledge of the spectral properties of the surface.

To analyse the ground uniformity, measurements are made across the area in a sampling strategy that matches the spatial resolution of the satellites. For each surface element measured we determine the reflectance factor $\rho_r(x_i, y_i, \lambda_l)$.

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].

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 $N \times L$. The PCA method produces three types of elements: N eigenvalues, γ_j, N eigenvectors, $c_{i,j}$, 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:

$$\rho_r(x_i, y_i, \lambda_l) = \langle \rho_r(x_i, y_i) \rangle_\lambda \cdot \\ \cdot \left[1 + \textstyle\sum_{j=1}^{M} c_j(x_i, y_i) \cdot H_j(\lambda_l)\right] \quad (1)$$

where $\langle \rho_r(x_i, y_i) \rangle_\lambda$ 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 $c_j(x,y)$ were rescaled so that the standard deviation of H_j is 1. This way, the value of $c_j(x,y)$ quantifies the contribution of every component at the locations.

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 the ground is generally uniform across them. The number M has to be determined taking into account the significance of each eigenspectra given by γ_j and the uncertainty allowed for in the calibration.

PCA RESULTS FOR REFERENCE SITES

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. The number of eigenspectra needed to reproduce spectral reflectance factors with a residual error lower than 0.1 % is four for La Crau and two for Gobabeb, a significant reduction in the number of spectra in both sites. This implies that the Gobabeb site has less spectral variation between the surface elements.

Spectral features of main eigenspectra (Fig. 1 for La Crau and Fig. 2 for Gobabeb) are also different for the two reference sites, showing more spectral variability La Crau's site than Gobabeb's one. This might be related to different composition of the sites, and may have an influence on the calibration of specific spectral bands of observation instruments.

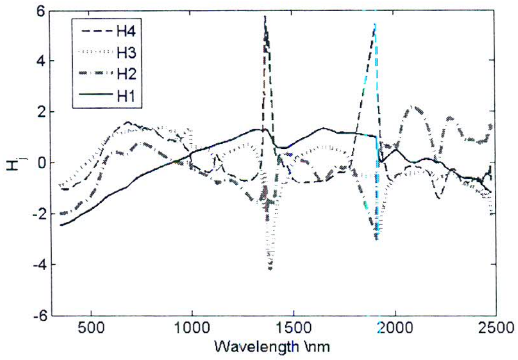

Figure 1. Main eigenspectra for the La Crau site

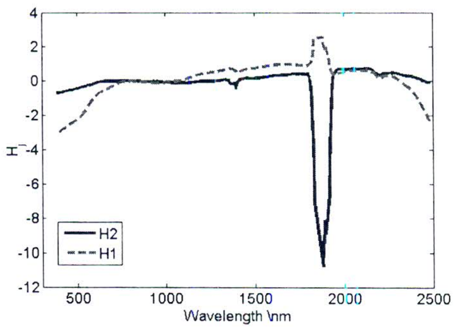

Figure 2. Main eigenspectra for the Gobabeb

The coefficients $c_j(x,y)$ to be used to reconstruct the reflectance factor values at every calibration position, show the contribution to the spectral variance of every component across the site. Fig. 3 shows the value of $c_j(x,y)$ for each of the four principal components at each of the 14 measured locations at La Crau (spatially arranged in a grid pattern), while Fig. 4 shows the coefficients for the two principal components for the 16 surface element at Gobabeb. No data were available for the grid position marked as N/A. From figure 3 it can be concluded that spatial variability at La Crau does not follow any apparent pattern since no relation can be established between the coefficient value at different positions, while it does it at Gobabeb, since steps are seeing in figure 4 in the coefficient of the second eigenspectra, corresponding to points in the same row.

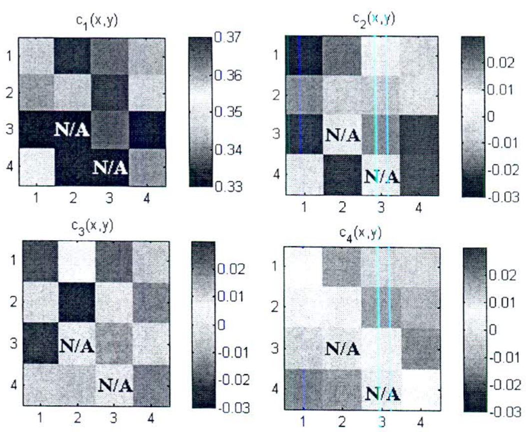

Figure 3. Spatial distribution of $c_j(x,y)$ for the La Crau site (N/A at non-measured locations).

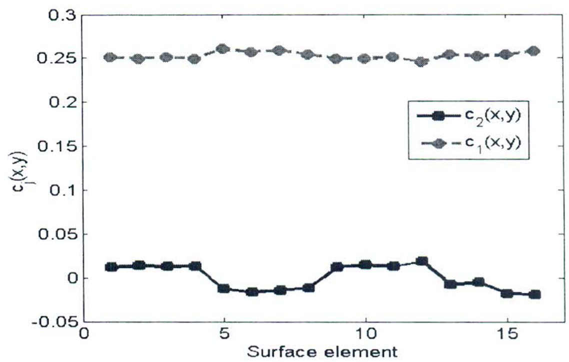

Figure 4. $c_j(x,y)$ for the relevant components at Gobabeb.

CONCLUSIONS

PCA shows patterns hidden in the data set that may be used to analyse 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 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).

ACKNOWLEDGEMENTS

This work has been supported by the European Metrology Research Programme (EMRP) 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.