Principal components analysis on the spectral bidirectional reflectance distribution function of ceramic colour standards

A. Ferrero,* J. Campos, A. M. Rabal, A. Pons, M. L. Hernanz, and A. Corrons

Instituto de Óptica, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28006, Spain

*alejandro.ferrero@csic.es

Abstract: The Bidirectional Reflectance Distribution Function (BRDF) is essential to characterize an object’s reflectance properties. This function depends both on the various illumination-observation geometries as well as on the wavelength. As a result, the comprehensive interpretation of the data becomes rather complex. In this work we assess the use of the multivariable analysis technique of Principal Components Analysis (PCA) applied to the experimental BRDF data of a ceramic colour standard. It will be shown that the result may be linked to the various reflection processes occurring on the surface, assuming that the incoming spectral distribution is affected by each one of these processes in a specific manner. Moreover, this procedure facilitates the task of interpolating a series of BRDF measurements obtained for a particular sample.

© 2011 Optical Society of America

OCIS codes: (290.1483) BSDF, BRDF, and BTDF; (030.5630) Radiometry; (120.5820) Scattering measurements.

References and links


1. Introduction

An object’s reflectance properties are fully characterized in each of its surface points by the Bidirectional Reflectance Distribution Function (BRDF). This function describes the way in which a surface reflects the incoming optical radiation; that is, the amount of radiance per unit of irradiance that is reflected in each direction as a function of the illumination direction [1]. This way, the determination of a sample’s BRDF requires a thorough sampling with different angular configurations. Moreover, the sampling becomes more complex if the object’s reflecting properties are such that symmetry and isotropy simplifications are not possible. In this sense, it would be extremely useful to have physical optics-based analytical models [2–12]. Unfortunately, taking into account the great diversity of existing surfaces, the quest for a unique model is an unfeasible task. Nonetheless, it is possible to provide a set of guidelines and methods that can be applied to each and every object. In this context, when searching for geometric–based or diffraction-based models, it is a common approach to assume that the BRDF is made up of three independent components:

• Specular contribution ($f_{r,sp}$): accounts for mirror-like reflections from the mean plane of the reflecting surface. It can be computed based on Fresnel reflectivity. It is observed when the surface’s roughness is small compared with the wavelength. All the photons coming from a given direction, when they interact with the surface, are reflected in the same specular direction.

• Directional diffuse contribution ($f_{r,dd}$): accounts for the field scattered to the hemisphere but with a certain directional character. It occurs when the photons coming from a given direction, when they interact with the surface, change their direction in a different but correlated way.

• Uniform diffuse contribution ($f_{r,ud}$): accounts for the field diffused uniformly over all the hemisphere. In this case the photons coming from a given direction, when they interact with the surface, change their direction in a different and uncorrelated way.

Consequently, each component originates from a different type of photon-surface interaction. A three-component BRDF is, unquestionably, the result of a phenomenological simplification,
since in reality there should be an undetermined number of potential interactions.

The Principal Components Analysis (PCA), which is a multivariable analysis technique, has proven to be effective for the analysis or characterization of multivariate response data [13–15]. This procedure makes it possible to reduce the multidimensionality of the response data to an intrinsically minimum number. The goal of this work is to show how the PCA technique applied to BRDF data may help to the identification of the different interaction processes occurring on the surface, assuming that each one of these processes has a specific impact on the incoming spectral distribution (the probability of a photon undergoing an interaction of a particular type must be a function of its energy; i.e. its wavelength). The results emerging from this type of procedure allow us to gain a deeper understanding of the individual contributions to the BRDF, and simplify to a great extent the search for models.

The present article starts with a brief review of PCA; afterwards it is shown how, based on experimental data, PCA makes it possible to perform a more detailed analysis of BRDF that could be used as starting point to develop specific models for different surface types.

2. Principal Components Analysis

The Principal Components Analysis is a powerful statistical technique capable of identifying and quantifying orthogonal contributions to the total variance of a collection of data. The key element is the definition of a multidimensional random variable \( F = \{F_1, F_2, \ldots, F_N\} \), having \( M \) realizations of it. For the BRDF data, this variable is made up of a concatenation of \( N \) spectra, each one corresponding to the measured BRDF for a given angular configuration. The set of spectral values extracted from the different spectra corresponding to a particular wavelength is taken as a particular realization of the multidimensional variable; the number of realizations, \( M \), coincides with the number of wavelengths. In the case tackled in this paper, the collection of data is arranged as a matrix having \( N \) columns and \( M \) rows. Each row contains the BRDF for every angular configuration at a specific wavelength. The covariance matrix of the spectra is obtained as an \( N \times N \) matrix, \( S_F \). This covariance matrix is typically non diagonal, thus revealing the inner correlations between the spectra. The PCA method diagonalizes this covariance matrix and produces three types of elements: \( N \) eigenvalues, \( \gamma_j \), \( N \) eigenvectors, \( e_{i,j} \), and \( N \) eigenspectra, \( A_j \). The eigenvalues are arranged in decreasing order. They quantify the importance and the contribution to the total variance of the data of their associated eigenspectra. Actually, \( \gamma_j \) is the variance of the eigenspectra \( A_j \). 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. These eigenvectors can be arranged as an \( N \times N \) orthogonal matrix, \( E^T = E^{-1} \) (\( ^T \) means transposition). Finally, the eigenspectra, \( A_j \), describe uncorrelated data that provide spectral insight about the different contributions to the BRDF. \( A_j \) is calculated as:

\[
A_j = \sum_{i=1}^{N} e_{i,j} F_i = F e_j^T
\]

where \( F_i \) is obtained by subtracting the mean from the original \( i \)-th spectrum, \( F_i (\bar{F}_i = F_i - \langle F_i \rangle) \).

The importance of a given eigenspectrum, \( A_j \), within the data set is quantified by its associated eigenvalue, \( \gamma_j \). By using the eigenvectors it is possible to migrate from the experimental coordinate system, \( \{F_1, \ldots, F_N\} \), to the eigenspectra coordinate system, \( \{A_1, \ldots, A_N\} \), and vice versa. This transformation can be written as follows:

\[
\bar{F}_i = \sum_{j=1}^{N} e_{i,j} A_j,
\]
From this equation we may obtain a filtered version of the spectra by selecting a customized subset of eigenvectors and their associated eigenspectra. This is easily done by choosing a subset of subscripts within the sum. Then, it is also possible to remove or to select a given collection of contributions characterized by their associated eigenspectra.

An interesting calculation can be done to express the variance of the original spectra as a combination of the variances associated with the eigenspectra. Thus, the variance of the spectrum \( F_i \) is given by:

\[
\sigma^2[F_i] = \sigma^2[\bar{F}_i] = \frac{1}{M-1} \sum_{\lambda=1}^{M} \left( \bar{F}_i^\lambda \right)^2,
\]

(3)

where \( \bar{F}_i^\lambda \) is the value on the wavelength identified as \( \lambda \), in the zero-mean spectrum \( \bar{F}_i \). Equation (2) incorporates the principal component decomposition in the calculation of \( (\bar{F}_i^\lambda)^2 \) as follows:

\[
(\bar{F}_i^\lambda)^2 = \left( \sum_{j=1}^{N} e_{i,j} A_j^\lambda \right)^2 = \sum_{j=1}^{N} \sum_{k=1}^{N} e_{i,j} e_{i,k} A_j^\lambda A_k^\lambda.
\]

(4)

After including the previous equation into the variance calculation, it is possible to go from the sum along the \( M \) wavelengths, labelled with \( \lambda \) super-scripts, to the double sum in \( j \) and \( k \). This yields the following expression:

\[
\sigma^2[F_i] = \sum_{j=1}^{N} \sum_{k=1}^{N} e_{i,j} e_{i,k} \frac{1}{M-1} \sum_{\lambda=1}^{M} A_j^\lambda A_k^\lambda = \sum_{j=1}^{N} e_{i,j}^2 \gamma_j,
\]

(5)

where we applied the fact that the eigenvectors are uncorrelated, which allows us to cancel all the terms having \( j \neq k \). When \( j = k \) the resulting term is the variance of the eigenvectors \( A_j \), which, by definition, is equal to \( \gamma_j \). This previous equation can be written as the following matrix product:

\[
S2 = E2 \cdot G,
\]

(6)

where \( S2 \) is a column vector containing the values of the variance of the original spectra, \( S2_i = \sigma^2[F_i] \); \( G \) is a column vector containing the eigenvalues, \( G_i = \gamma_j \); and \( E2 \) is a matrix whose elements are the squares of the components of the eigenvectors, \( E2_{i,j} = e_{i,j}^2 \).

Following the previous derivation we may conclude that the contribution of a given spectrum, \( A_j \), to the variance of a selected original spectrum, \( F_i \), is given by \( e_{i,j}^2 \gamma_j \). Similarly, the contribution to the standard deviation of the original spectrum \( i \) due only to the eigenspectrum \( j \), is given by:

\[
\sigma_{i,j} = |e_{i,j}| \sqrt{\gamma_j}.
\]

(7)

Due to the fact that \( \sum_{j=1}^{N} e_{i,j}^2 = 1 \), the elements of \( E2 \) represent the relative contribution to the variance of the original spectra of the variance of the eigenspectra obtained by the PCA method.

3. Results

The experimental measurements upon which PCA was applied were obtained by means of a gonio-spectrophotometer GEF E [16]. This instrument consists of a fixed, uniform and collimated light source (Xenon lamp) and a robot-arm that places the sample, making it possible to change automatically and simultaneously the two illumination spherical coordinates (\( \theta_i \), \( \phi_i \)) and the two observation ones (\( \theta_o \), \( \phi_o \)). These spherical coordinates are defined relative to the Sample Coordinate System, whose \( z \) axis is parallel to the sample’s normal direction (Fig. 1). Moreover, thanks to its periscope system with beamsplitter, it is possible to measure under retro-reflection conditions. This configuration yields important information about the BRDF, as
will be shown further on. For the data acquisition a CS-2000 Konica Minolta spectroradiometer is used. This device operates within the visible range [380 nm –780 nm], performing spectral measurements with a 1 nm spectral sampling interval and a 4 nm bandwidth. The relative uncertainty of the BRDF measurement with $k=2$ is less than 0.02 (limited by the linearity correction when the measurement ranges within more than 5 decades). This uncertainty is comparable to other instruments with similar features [17–19].

![Sample Coordinate System showing spherical coordinates for incidence ($\theta_i$, $\phi_i$) and observation directions ($\theta_s$, $\phi_s$).](image)

3.1. BRDF data

The BRDF measurements shown below are from a glossy Ceramic Colour Standard, CCS (Glossy Green 5 cm × 5 cm). The BRDF was sampled at 448 different angular configurations; each one corresponds to a specific combination of $\theta_i$ and $\theta_s$ (which take the values $0^\circ$, $10^\circ$, $20^\circ$, $30^\circ$, $40^\circ$, $50^\circ$, $60^\circ$ and $70^\circ$), $\phi_i$ (which take only the value $0^\circ$) and $\phi_s$ (which take the values $0^\circ$, $30^\circ$, $60^\circ$, $90^\circ$, $120^\circ$, $150^\circ$, $180^\circ$).

Figure 2 shows the BRDF spectra depicted in a semilogarithmic plot. The spectra in the specular region are relatively flatter than those in the diffuse region, and their respective average values differ in more than two orders of magnitude. Moreover, each set of spectra in the specular region corresponds to a particular angle of incidence $\theta_i$. Figures 3 and 4 show that the bigger the angle $\theta_i$, the higher the resulting BRDF (with the exception of the $0^\circ$-configuration). Furthermore, these figures clearly reveal the effect of $\theta_i$ upon the spectral distribution: Whereas the spectral distributions at $40^\circ$, $50^\circ$, $60^\circ$ and $70^\circ$ are very similar, those recorded at $10^\circ$ and $20^\circ$ are slightly different from the rest, and also different from each other. On the other hand, within the diffuse region all spectra are rather similar (Fig. 2). The high frequency noise in the spectra (e.g. between 450 nm and 500 nm and around 770 nm) is an artefact produced by the division of the measured radiance and the measured irradiance on the sample to obtain the BRDF, and it corresponds to the narrow lines of the Xe lamp used in GEFE.

By dispensing with the spectral information (i.e., by representing only the value for a particular wavelength or the average value), the BRDF’s angular distribution can be plotted in greater detail. For this purpose, diagrams such as those shown in Fig. 5 are used. This figure represents the BRDF data for $\lambda = 500$ nm measured within the incidence plane. This plane comprises two half-planes, each of which is characterized by a particular angle $\phi_s$. We will say that, by defini-
Fig. 2. Spectral BRDF components of a glossy ceramic colour standard.

Fig. 3. Specular spectral BRDF measurements of a glossy ceramic colour standard at $\theta_i = \theta_s = \theta = 0^\circ$, 10$^\circ$, 20$^\circ$ and 30$^\circ$, and $\phi_i = 0^\circ$ and $\phi_s = 180^\circ$. 
Fig. 4. Specular spectral BRDF measurements of a glossy ceramic colour standard at $\theta_i = \theta_s = \theta = 40^\circ$, $50^\circ$, $60^\circ$ and $70^\circ$, and $\phi_i = 0^\circ$ and $\phi_s = 180^\circ$.

tion, the half-plane containing the incoming light is $\phi_s = 0^\circ$, whereas the half-plane containing the specularly reflected light is $\phi_s = 180^\circ$.

A very interesting effect can be observed in Figs. 3 and 5. In Fig. 3, the specular BRDF spectrum at $\theta_i = 0^\circ$ has a higher value than those spectra at $\theta_i = 10^\circ$, $20^\circ$ and $30^\circ$. On the other hand, an increase of the BRDF at retro-reflection conditions is observe in Fig. 5. Both effects might be explained by the coherent backscattering of light (CBS) [20, 21].

Since Fig. 5 represents the incidence plane, we will call this figure 0:180, according to the notation (first half-plane: second half-plane). Each plot in this figure represents the measured BRDF as a function of the observation polar angle $\theta_s$ for a specific illumination polar angle $\theta_i$. A negative $\theta_s$ indicates, by definition, that the observation is carried out in the first half-plane.

Figure 6 shows the resulting BRDF curve ($\lambda = 500$ nm) when observing from outside the incidence plane (half-planes 30:150), for $\theta_i = 20^\circ$, $30^\circ$, $40^\circ$, $50^\circ$, $60^\circ$ and $70^\circ$. This figure reveals more clearly the presence of the BRDF's diffuse component.

As the reader will observe, the diffuse BRDF decreases considerably when either the illumination ($\theta_i$) or the observation ($\theta_s$) angle is large (grazing angles), whereas for smaller angles it remains relatively constant. There is also a certain rise in BRDF with $\theta_s$ within the 150-half-plane (close to the half-plane that contains the specular direction); in fact, the larger $\theta_i$ is, the steeper the increase is. Therefore, it can be concluded that the diffuse BRDF is more uniform within the small-angle range (applicable both for illumination ($\theta_i$) as well as for observation ($\theta_s$) angles). The weak point of this type of representation is that it is not possible to assess the spectral dependency with the different angular configurations. We will see that this type of representations can be used to represent and highlight precisely this spectral variation, resorting for this purpose to PCA.

3.2. PCA on BRDF data

PCA was applied to BRDF experimental data, as described in section 2 ($N = 401$ wavelengths and $M = 448$ BRDF spectra, with a $448 \times 448$ covariant matrix). The resulting eigenspectra are shown in Fig. 7. These are the spectra associated with the four highest eigenvalues (in descending order) which are, consequently, the most relevant ones when it comes to computing the total variance of the data (their joint contribution accounts for 99.6% of this variance: eigenvalues 0.9864, 0.0066, 0.0018 and 0.0013 for these four eigenspectra and 0.0003 for the first
Fig. 5. Angular representation of the BRDF measurements at 500 nm in the incidence plane (0:180). Every plot shows BRDF values versus observation angles and corresponds to a different polar incidence angle ($\theta_i$).

Fig. 6. BRDF measured at 500 nm outside the incidence plane (half-planes 30:150) versus observation angle ($\theta_s$) for different incidence angles ($\theta_i$).
neglected eigenspectrum). We can see that the eigenspectrum PC1 (Fig. 7(a)) looks very similar to the specular spectra in Figs. 3 and 4. Similarly, there seems to be a large similarity between PC2 (Fig. 7(b)) and diffuse spectra in Fig. 2.

It is important a proper choice of the scale of the data, since different scales return different eigenvalues. In our specific case, choosing absolute spectra as input variables to the PCA returns a better result than choosing relative ones (for instance, normalizing the spectra to their variances). The reason is that specular spectra have a smaller relative variance than diffuse spectra and, in addition, there are much more diffuse spectra than specular spectra in the data. In consequence, applying PCA with a relative scaling would give a smaller weight to the specular components. This consideration should be done before applying PCA to a similar kind of data.

The values on the Y-axis have a relative importance, since it must be recalled that, according to Eq. (2), when reconstructing the original variance the eigenspectrum is multiplied by the coordinate of the corresponding eigenvector, which ranges between -1 and 1. Even though the eigenspectra that are represented in the figures are all oriented in a natural way for the sake of clarity (to make interpretation easier), the eigenspectra obtained following PCA are often inverted (multiplied by -1). In order to determine the true orientation of a given eigenspectrum $A_j$, the simplest way is to locate an spectrum $i$ where this eigenspectrum has a substantial contribution to the total variance; then one only needs to look at the sign of $e_{i,j}$. If the sign is negative then the eigenspectrum $A_j$ has to be inverted in the representation, but not in subsequent calculations.

![Eigenspectra](image)

**Fig. 7.** Eigenspectra of the BRDF measurements of a glossy green ceramic colour standard: (a) PC1, (b) PC2, (c) PC3 and (d) PC4.

How the contribution of an eigenspectrum to the variance of a given spectrum can be computed is explained in the paragraph that follows Eq. (7). Figures 8 and 9 show the relative contribution to the spectral variance of the first four principal components in a semilogarithmic plot. These plots make use of an angular representation, and they correspond to the half-planes
No systematic tendency outside the specular region could be extracted from the data of the contributions. It is worth pointing out that, in each configuration, the relative contributions of the four components must add up to nearly one. If this was not the case for a particular scenario, it would imply that a non-negligible component was being left out; in that situation the corresponding eigenspectrum would have to be added. The impact of leaving out these excluded eigenspectra will be rigorously analyzed and assessed in Section 4.

Fig. 8. Relative contributions to the spectral variance of the first four principal components (half-plane 0:180) in a semilogarithmic plot.

4. Discussion

4.1. Physical interpretation of the eigenspectra of the BRDF

In the light of Figs. 8 and 9, one can made the following analysis about the physical meaning of each eigenspectrum:

- Eigenspectrum PC1. This component always makes a contribution in specularreflection scenarios (Fig. 8). It never contributes significantly in other configuration scenarios (Fig. 9). It is the only component that always contributes significantly in specular conditions; that is why it can be stated that eigenspectrum PC1 is influenced by Fresnel reflection.

- Eigenspectrum PC2. It predominates over the other components, except for the case where specular reflection is involved. It represents the BRDF’s diffuse component.

- Eigenspectrum PC3. It contributes too to the BRDF’s diffuse component. The reason why PCA renders them as two separate components is that, under specular reflection conditions at low incidence angles, eigenspectrum PC3’s contribution to the variance is clearly more relevant than that of eigenspectrum PC2. Taking into account that they have a rather different behavior under specular reflection conditions, it can be said that the
spectral distributions associated to eigenspectra PC2 and PC3 originate from different physical phenomena.

- Eigenspectrum PC4. Even though this component’s contribution clearly exceeds that of eigenspectrum PC1 under non-specular observation conditions, its contribution is nonetheless much smaller than that of eigenspectra PC2 and PC3. It is under specular conditions with $\theta_i = 10^\circ$ and $\theta_s = 20^\circ$ where this component’s contribution becomes really significant (see also Fig. 3). Perhaps this could be explained by the presence in the material of a second flat layer, where small incidence angles could be reflected and transmitted, but where large incidence angles would be trapped due to total reflection.

The idea behind this analysis is that it is possible to use the resulting information for identifying the uncorrelated components of the BRDF to interactions in different layers of the surface. A model of layers can be proposed from the PCA information as starting point for interpolation with physical models. It is necessary to notice that uncorrelated variables are not always independent. To make sure that the number of layers is not overestimated, it is possible to apply an algorithm of Independent Component Analysis to the PCA result.

4.2. PCA’s usefulness for the development of BRDF models

The task of finding an analytical model that fits the highly complex BRDF data can be simplified to a large extent if the contributions of the different reflections occurring on the sample can be identified, isolated and analyzed separately. Generally this is not the case, since samples have a complex surface structure that is not known down to the detail. Since PCA yields orthogonal components, it is worth considering whether or not it would be possible to reconstruct the BRDF from a relatively small number of PCA components and a few fitting parameters. This way, an empirical model would be available, allowing us to interpolate the BRDF for any given
illumination/observation geometry and to compute the sample’s reflectance or reflectance factor with greater accuracy.

In the framework of the principal components, the reconstruction of the BRDF data is carried out by using only those components \((\bar{F}_{i,j})\) that interest us:

\[
\bar{F}_{i,j} = e_{i,j}A_j
\]

where \(\bar{F}_{i,j}\) is the resulting reconstruction of \(\bar{F}_i\) when only \(A_j\) is used. If we recall that, as stated in a previous section, \(\bar{F}_i = F_i - \langle F_i \rangle\), then it can be inferred that the \(F_i\) corresponding to the principal component \(j\) can be computed as \(F_i = \bar{F}_i + \kappa_{i,j}\langle F_i \rangle\), where \(\kappa_{i,j}\) represents the relative contribution associated with \(A_j\) to the \(i\) spectrum of the BRDF. If we assume that this fraction is proportional to this component’s contribution, then under this approximation it can be said that: \(\kappa_{i,j} = \mathbf{E}_{2i,j}\). This approximation can be used to perform the model’s fitting, taking \(\kappa_{i,j}\) as a free parameter with an initial value \(\mathbf{E}_{2i,j}\).

Strictly speaking, the number of principal components that is required to explain a BRDF spectrum must be the one that is able to reproduce the original data (within a given uncertainty). Figure 10 represents, for the \((0:180)\) half-planes, the relative average error \(\varepsilon_{r,i}\) resulting from the reconstruction of the angular configurations \(i\) that relies, respectively, on the four and five eigenspectra associated with the highest eigenvalues. The average error is computed as follows:

\[
\varepsilon_i = \frac{1}{M} \sum_{\lambda=1}^{M} | \bar{F}^\lambda_{i,r} - \bar{F}^\lambda_{i,i} | \tag{9}
\]

where \(\bar{F}_{i,r}\) represents the reconstruction of \(\bar{F}_i\).

Fig. 10. Angular representations of the relative average error \(\varepsilon_r\) in the incidence plane \((0:180)\).

The figure reveals, as was to be expected, that the more eigenspectra are used for the reconstruction, the smaller the resulting average error. For the four-component case considered in
this work, the relative average error is around 1%, and clearly falls below this value when one 
additional eigenspectra are used (around 0.5%).

5. Conclusions

The PCA method has been applied to the BRDF that was experimentally measured on a sample 
glossy green ceramic colour standard. The analysis leads to the conclusion that only 4 principal 
components are responsible for most of the variance of the BRDF. The relative weight of each 
one of these components is linked to the specific geometry of the reflection. Thus, component 
1 is more significant (i.e., plays a more significant role) when specular geometry is considered, 
whereas 2, 3 and 4 are more significant in those geometries where there is diffuse reflection. 
What makes these components 2, 3 and 4 different is the relative weight that they have in the 
calculation of the variance of the spectral distribution observed in the specular direction. Conse-
quently, there is a relationship between the principal components and the physical phenomena 
that govern reflection in this sample. In this context, a model having a physical meaning can be 
set out based on these principal components.

It has been demonstrated that this principal-component-based model can explain very well 
the BRDF spectra even when only 4 components are included. With PCA it is possible to repre-
sent in a simple way and simultaneously both the spectral and the angular nature of the BRDF. 
As a result, PCA becomes a powerful tool to undertake the complex task of interpreting this 
kind of data. Moreover, PCA facilitates the task of interpolating the various BRDF measures-
ments corresponding to a particular sample. The reason lies in its underlying physical model: 
in PCA, each of the resulting principal components corresponds to a particular component of 
the BRDF of different nature; also, each principal component can be fitted independently from 
the rest.

Acknowledgments

The authors are grateful to the “Plan Nacional de Física” for funding this work (FIS2007-
66671-C02-01), and to the CSIC’s JAE Program for providing the group with a training re-
searcher.