electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Gernot Kostorz**

Canonical correlation and quantitative phase analysis of microdiffraction patterns in bone-tissue engineering

Antonietta Guagliardi, Cinzia Giannini, Massimo Ladisa, Antonio Lamura, Teresa Laudadio, Alessia Cedola, Stefano Lagomarsino and Ranieri Cancedda

Journal of
Applied Crystallography
ISSN 0021-8898

Received 19 February 2007
Accepted 24 July 2007

Canonical correlation and quantitative phase analysis of microdiffraction patterns in bone-tissue engineering

Antonietta Guagliardi,[a]* Cinzia Giannini,[a] Massimo Ladisa,[a] Antonio Lamura,[b] Teresa Laudadio,[b] Alessia Cedola,[c] Stefano Lagomarsino[c] and Ranieri Cancedda[d]

[a]Istituto di Cristallografia, CNR, via G. Amendola 122/O, I-70126 Bari, Italy, [b]Istituto Applicazioni Calcolo, CNR, sez. Bari, via G. Amendola 122/O, I-70126 Bari, Italy, [c]Istituto di Fotonica e Nanotecnologie, CNR, I-00185 Roma, Italy, and [d]Istituto Nazionale per la Ricerca sul Cancro and Dipartimento di Oncologia, Biologia e Genetica, Università di Genova, L.go R. Benzi 10, 16132 Genova, Italy. Correspondence e-mail: antonella.guagliardi@ic.cnr.it

A novel method is described that combines high-resolution scanning microdiffraction techniques, Rietveld quantitative phase analysis and a statistical method known as canonical correlation analysis (CCA). The method has been applied to a sample taken from a bone-tissue-engineered bioceramic porous scaffold implanted in a mouse for six months. The CCA technique allows the detection of those pixels throughout the investigated sample that best correlate with signal models. Besides the standard usage of this approach, which requires theoretical profiles as signal models, a novel application is presented here, which consists of picking the model spectra out of the experimental data set. Patterns representative of a reasonable range of phase compositions were selected among the huge number of two-dimensional patterns (folded in one-dimensional profiles) to extract quantitative phase fractions. At this stage, the CCA approach was also used to overcome the low Poisson statistic of signal models, so making Rietveld quantitative analysis more reliable. These patterns have been used as profile models for CCA. The final classification map, obtained by assigning the considered pixel to the model spectrum with the highest canonical coefficient, provides the spatial variation of phase concentration.

1. Introduction

The Rietveld method (Rietveld, 1969) is nowadays applied for phase quantification of polycrystalline samples (Hill & Howard, 1987; Bish & Howard, 1988; Madsen *et al.*, 2001; Scarlett *et al.*, 2002). Most of the standard Rietveld packages (*Fullprof*, by Rodriguez-Carvajal, 1992; *GSAS*, by Larson & Von Dreele, 1994; *Maud*, by Lutterotti, 2005) can deal with multiphase specimen diffraction data; some of them can perform the analysis *via* automatic runs, according to simple scripts or as the default choice (like the program *Quanto*, by Altomare *et al.*, 2001). Nonetheless, analysing a huge number of patterns can still be a highly time-consuming process. This is particularly the case when fast area detectors provide enormous amounts of two-dimensional images folded into one-dimensional patterns to be processed. An example is found in microdiffraction experiments performed on composite materials of interest for bone-tissue engineering (Cedola *et al.*, 2006, 2007; Komlev *et al.*, 2006), a very rapidly growing field of research whose final goal is to provide biomaterial scaffolds suitable for repairing bone defects. X-ray powder (micro) diffraction techniques and Rietveld analysis have often been used to investigate bone-tissue samples (Dalconi *et al.*, 2003; Wagermaier *et al.*, 2007) and seem to be very promising to deal with such an issue.

Natural bone tissue is a composite microstructure, consisting of several organic and inorganic parts. The mineral component is essentially made of calcium bioapatite crystals. A large number of different therapeutic approaches have been proposed to repair large bone defects. A very promising solution is 'tissue engineering', *i.e.* the association of suitable scaffolds with osteogenic cells. Studies have been addressed to compare the performances of pure hydroxyapatite-based and tricalcium phosphate-based bone-graft substitutes, especially with respect to their ability to participate in the structure of newly formed bone (Mastrogiacomo *et al.*, 2007), as well as to highlight the role of adult stem cells (Quarto *et al.*, 2001).

Within these studies, a well established model of *in vivo* ectopic bone formation in immunodeficient mice was applied using small implants based on silicon-stabilized tricalcium phosphate (Si-Tcp) (Sayer *et al.*, 2003). Scaffolds were implanted in mice for two and six months, and then analysed by X-ray microtomography and microdiffraction using synchrotron radiation. These studies (for details see Mastrogiacomo *et al.*, 2007) have proved that porous ceramic scaffolds based on Si-Tcp and loaded with bone-marrow stromal

cells before being implanted in test animals, display a progressive scaffold resorption, while new bone tissue is formed. Two-dimensional microdiffraction data were collected on thin slices obtained from small cubic scaffolds implanted in mice; the cubes vary according to the scaffold composition, the presence or absence of osteogenic cells, and the implantation time. A great number of two-dimensional images were obtained following a well defined grid of measurements, suitable to record the spatial compositional variations at the scaffold/bone interface and to investigate the possible coupled mechanism of bone formation and scaffold resorption. Small-angle (SAXS) and wide-angle (WAXS) X-ray scattering data were simultaneously collected at each measured point.

The work presented in this article uses the WAXS data collected on a slice of the aforementioned series (scaffold seeded with bone-marrow stromal cell, six months of implantation time) to focus on an original method of analysis which combines the Rietveld quantitative phase analysis (RQPA), microdiffraction techniques with a high spatial resolution (Rieckel & Davies, 2005; Cedola *et al.*, 2006) and a statistical method known as canonical correlation analysis (CCA) (Johnson & Wichern, 1998), which represents the multivariate variant of the ordinary correlation analysis. CCA has already been successfully applied to biomedical signal processing (Friman, 2003; Laudadio *et al.*, 2005; Ladisa *et al.*, 2007); here it is used to detect those pixels throughout the set of two-dimensional patterns whose one-dimensional profiles best correlate to selected representative pattern models. The standard application of CCA uses theoretical profiles as signal models. In the present work, instead of generating a great number of possible mixture-pattern models, a novel usage of the CCA method has been adopted, by picking up the model spectra from the experimental data set. The RQPA of these signal models results in the achievement of two-dimensional mapping of phase concentration over the whole investigated surface, averaged over the slice thickness. Furthermore, due to the very small illuminated area coupled to the CCD saturation problems produced by large crystals spots, the microdiffraction patterns collected on the investigated sample suffer from a low Poisson statistics. To minimize this problem, CCA is also a useful tool as a classifier: the most correlated signals were summed up to obtain a signal statistics high enough to perform a more reliable RQPA. The whole process relies on fully automatic analysis procedures, to deal with the huge amount of experimental data (about 1000 two-dimensional patterns).

Details of samples and experimental conditions are given in §2. The main features of the CCA method and how it is applied to powder diffraction data can be found in §3. A description of the results obtained on the investigated sample and the discussion are given in §4, followed by some final remarks in §5.

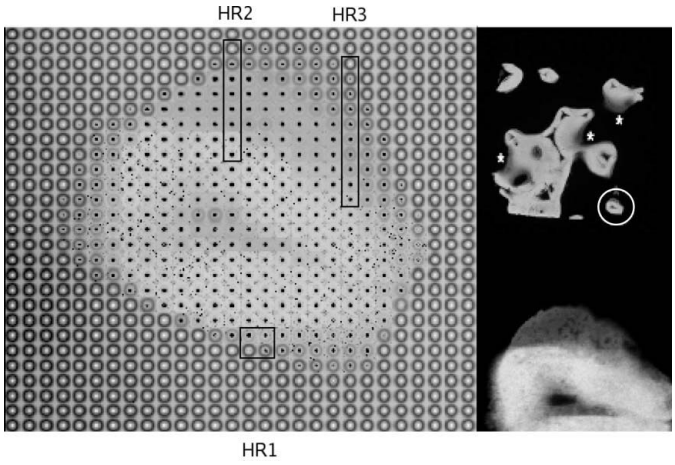

Figure 1
Grid of two-dimensional microdiffraction patterns collected at ID13 on the investigated sample, using a grid step of 25 μm. In the radiographic image, on the top right-hand side of the figure, the sample is inside the circle. It is magnified in the bottom right-hand side. HR1, HR2 and HR3 are the three regions further sampled using a resolution step of 1 μm (see the text for details).

2. Experimental

Bioceramic scaffolds implanted in mice were skelite (Millenium Biologix Corp., Kingston, Canada), a clinically available bone-graft substitute made of a mixture of silicon-stabilized tricalcium phosphate (Si-Tcp) (Sayer *et al.*, 2003), hydroxyapatite (HA) and β-tricalcium phosphate (β-Tcp). Small cubes of $4 \times 4 \times 4$ mm were seeded with bone-marrow stromal cells and implanted subcutaneously in mice for two and six months, then harvested and suitably treated to prepare sections of about 100 μm thickness to be investigated by X-ray microdiffraction. The sections were cut so as to be transversal to the scaffold's pore surface and to preserve, as much as possible, a homogeneous composition along the thickness. Details about scaffold preparation and cell culture have been published by Mastrogiacomo *et al.* (2007). All the procedures were in agreement with protocols approved by the competent ethical authority.

Measurements were carried out at beamline ID13 of the European Synchrotron Radiation Facility (ESRF). A microdiffraction setup provided with a Si 111 monochromator and a Kirkpatrick–Baez mirror system (Rieckel & Davies, 2005) supplied a beam size of 1 μm × 1 μm with a wavelength of 0.976 Å. Patterns were recorded, in transmission mode, using a Mar CCD detector. A data set of NIST standard Si 640c was also collected, to be used as a standard for sample-to-detector distance determination. The investigated area of about (0.7×0.7) mm^2 was scanned following a grid step of 25 μm; three small regions (a square area of about 780 μm^2 and two straight lines of about 180 μm and 220 μm, respectively), located at the scaffold/bone interface, were further sampled at a higher resolution step of 1 μm, as shown in Fig. 1

From the two-dimensional diffraction images of the measured grid, one-dimensional patterns were obtained by using the algorithm developed by Cervellino *et al.* (2006). Skelite powder, as commercially available before implant, was used to collect the diffraction pattern of the scaffold, to be used as a reference and test pattern for crystal structure models. Data were collected using a Rigaku 12 kW Cu rotating anode, a germanium monochromator on the incident beam

and a CPD120 Inel detector. RQPA of this sample was performed using the program *Quanto* (Altomare *et al.*, 2001). Crystal structure models of Si-Tcp (Sayer *et al.*, 2003), β-Tcp (Yashima *et al.*, 2003) and HA (ICSD Nr. 26204) were used as input structures for the automatic run. The phase concentrations [Si-Tcp 78.9 (1)%, β-Tcp 2.1 (2)%, HA 19.0 (2)%] agree well with the compositional range given by Sayer *et al.* (2003). The final best fit is shown in Fig. 2, together with the corresponding profile agreement factor $R_{\rm wp}$ $\left[R_{\rm wp} = \left\{\left[\sum w(I_{\rm o} - I_{\rm c})^2\right]/\left[\sum w(I_{\rm o})^2\right]\right\}^{1/2}\right.$, where $I_{\rm o}$ and $I_{\rm c}$ are the observed and calculated step intensity, respectively, and w is the inverse of the variance of the observation$\left.\right]$, and the goodness of fit (GoF). The weight percentages refer to the crystalline part of the sample; the amorphous component, always occurring in this kind of scaffold according to Sayer *et al.* (2003), was not estimated.

3. The method

3.1. The canonical correlation analysis

CCA is a statistical technique that quantifies the relationship between two sets of variables. It is a multichannel generalization of ordinary correlation analysis, which quantifies the relationship between two random variables x and y by means of the so-called correlation coefficient

$$\rho = \frac{\mathrm{Cov}\,[x, y]}{(V\,[x]\,V\,[y])^{1/2}}, \tag{1}$$

where Cov and V stand for covariance and variance, respectively. The correlation coefficient is a scalar with a value between -1 and 1 that measures the degree of linear dependence between x and y. For zero-mean variables, equation (1) is replaced by

$$\rho = \frac{E[xy]}{(E[x^2]E[y^2])^{1/2}}. \tag{2}$$

where E stands for expected value. Canonical correlation analysis can be applied to multichannel signal processing as follows: consider two zero-mean multivariate random vectors $\mathbf{x} = [x_1(t), \ldots, x_m(t)]^{\rm T}$ and $\mathbf{y} = [y_1(t), \ldots, y_n(t)]^{\rm T}$, with $t = 1, \ldots, N$, and the superscript T denotes the transpose. The following linear combinations of the components in $\mathbf{x}$ and $\mathbf{y}$ are defined, which respectively represent two new scalar random variables X and Y:

$$X = w_{x_1}x_1 + \ldots + w_{x_m}x_m = \mathbf{w}_x^{\rm T}\mathbf{x}, \tag{3}$$

$$Y = w_{y_1}y_1 + \ldots + w_{y_n}y_n = \mathbf{w}_y^{\rm T}\mathbf{y}. \tag{4}$$

CCA computes the linear combination coefficients $\mathbf{w}_x = [w_{x_1}, \ldots, w_{x_m}]^{\rm T}$ and $\mathbf{w}_y = [w_{y_1}, \ldots, w_{y_n}]^{\rm T}$, called regression weights, so that the correlation between the new variables X and Y is maximized. The solution $\mathbf{w}_x = \mathbf{w}_y = 0$ is not allowed and the new variables X and Y are called canonical variates.

Several implementations of CCA are available in the literature. However, as shown by Laudadio *et al.* (2005), the most reliable and fastest implementation is based on the interpretation of CCA in terms of principal angles between linear subspaces (Hotelling, 1936; Zha, 1993). For further details, the reader is referred to work by Laudadio *et al.* (2005) and references therein; an outline of the aforementioned implementation is provided, for the sake of clarity, in Appendix *A*1.

3.2. CCA applied to diffraction data

When applying correlation analysis to X-ray powder diffraction (XRPD) data, the variables $\mathbf{x}$ and $\mathbf{y}$ need to be specified. Consider the set of one-dimensional spectra associated to the grid in Fig. 1. In ordinary correlation analysis, $\mathbf{x}$ and $\mathbf{y}$ are univariate variables and, specifically, the $\mathbf{x}$ variable consists of the intensity spectrum of the measured signal contained in each pixel, while the $\mathbf{y}$ variable consists of a model intensity spectrum. The correlation coefficient between $\mathbf{x}$ and $\mathbf{y}$ is computed and assigned to the pixel under investigation. Once each pixel has been processed, a new grid, of the same size as the original set of images, is obtained, which contains correlation coefficients instead of XRPD signals. This new grid is called the correlation map.

The difference between ordinary correlation analysis (*i.e.* single-pixel model, see Appendix *A*2) and CCA mainly lies in a different choice of the variables $\mathbf{x}$ and $\mathbf{y}$. In fact, in order to compute the correlation maps, it is possible to exploit the spatial information characterizing the XRPD data set. The variable $\mathbf{x}$ is a multivariate vector with components representing both the intensity spectrum of the considered pixel and the intensity spectra of the neighbour pixels. According to this feature, CCA is more suitable for our purposes with respect to the single-pixel approach, as using the spatial information makes the material detection more reliable. Several spatial models can be adopted for choosing the neighbour pixels, typical examples of which are described by Laudadio *et al.* (2005) and Friman (2003). Some details are given in Appendix *A*2.

The variable $\mathbf{y}$ also consists of a multivariate vector. Its components represent the basis functions of a signal subspace.

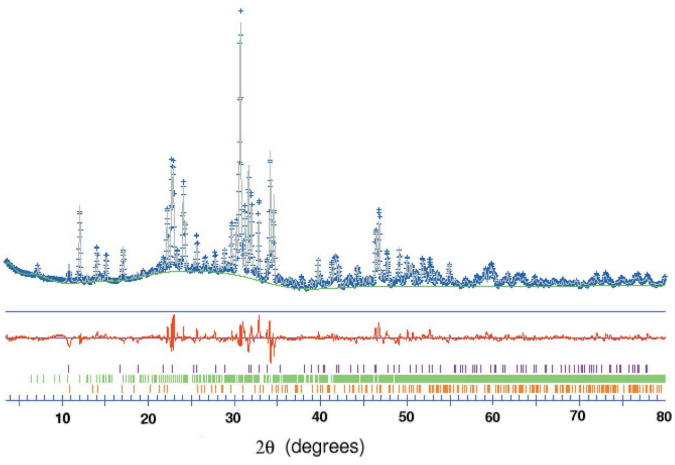

Figure 2
Quanto final best fit of the powder pattern collected on the scaffold, as available before implant. $R_{\rm wp}$ = 5.31, GoF = 4.81.

Several approaches can be adopted; an exhaustive overview is given by Friman (2003). Here, the so-called Taylor model was considered. Accordingly, the first component of the **y** variable is then defined as the patterns chosen as profile models. The second component of **y** was obtained as the first-order derivative of the first component, approximated by first-order finite differences. For the sake of clarity, the procedure to compute the aforementioned subspace model is outlined in Appendix *A*3.

4. Application of CCA to the bone-tissue sample

4.1. Standard usage of CCA

CCA was firstly applied to the investigated problem according to the standard usage of the technique, that is adopting calculated diffraction patterns as signal models. In order to define the first component of the variable **y**, five intensity spectra were generated as profile models: Si-Tcp, β-Tcp and HA (Ha-m hereafter) for the microcrystalline scaffold phases, nanocrystalline HA (Ha-n hereafter) for the new bone, and an amorphous scattering curve. The theoretical patterns have been generated using phase structure models taken from the literature (see §2) and an analytical profile function for peak shape description. Peak width and shape of microcrystalline phases were derived from the experimentally determined peaks of the powder diffraction pattern collected on the original scaffold (see §2). For nanocrystalline HA, it was assumed that their crystal structure is the same as those of HA with crystallites in the micrometre range, with much wider peaks (as estimated from the microdiffraction patterns); the amorphous curve was obtained by fitting a microdiffraction pattern showing a pure amorphous composition.

With respect to the choice of spatial models, we tested all those reported in Appendix *A*2. The best performance was obtained by using the symmetric 3×3 model without corner pixels (for details see Ladisa *et al.*, 2007).

Five different **y** variables were defined, one for each phase. CCA was applied between the experimental one-dimensional data set and the aforementioned **y** variables. We obtained five different correlation maps, and by comparing, pixel by pixel, the five canonical correlation coefficients, we built a classification map (the so-called 'nosologic' image in De Edelenyi *et al.*, 2000) by assigning the considered pixel to the phase with the highest canonical coefficient.

In Fig. 3, the correlation maps provided by CCA are visualized for amorphous, HA-n, HA-m, Si-Tcp and β-Tcp phases. The last frame in Fig. 3 shows the classification image, where the grey tones denote the tissue component as follows: the black denotes the amorphous, the dark grey indicates the Si-Tcp, the grey corresponds to HA-n, the light grey denotes HA-m.

As can be observed, CCA gives a classification image that resembles the pattern of Fig. 1 quite well. We can see that the CCA approach is able to determine which tissue component is dominant in each scattering pixel, giving a direct visualization without any user interaction. For the investigated sample,

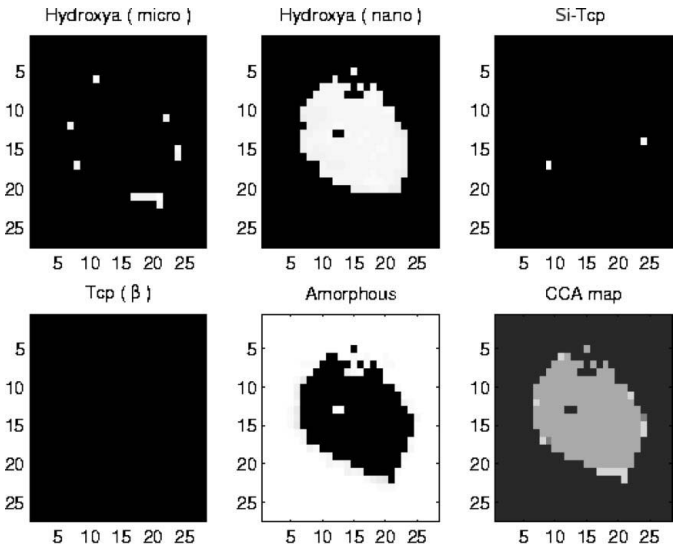

Figure 3
Correlation maps provided by CCA for HA-m, HA-n, Si-Tcp, β-Tcp and amorphous phases. The last frame shows the classification image, where the grey tones denote the tissue component: amorphous (black), Si-Tcp (dark grey), HA-n (grey), HA-m (light grey).

amorphous and HA-n produce the highest intensity signals, while the crystalline components of the original scaffold are vanishing almost everywhere.

However, even though the classification map obtained thus far can be considered an excellent result, it does not provide any quantitative information about the investigated patterns, about the concentration and spatial variations of the phases involved; the agreement among the considered model spectra and the experimental patterns has to be regarded as a 'guide to the eye', with no claim of a quantitative analysis of the investigated sample.

4.2. Combining CCA and RQPA

Phase concentrations and spatial distribution of phases are aspects of great interest for the sample under study and, in general, for bone-tissue engineering, to evaluate the spatial evolution of the materials belonging to the original scaffold and the deposition of the new mineral bone tissue. It is worth noting that, in order to provide a final accurate description of the distribution of phases throughout the investigated sample, the Rietveld quantitative phase analysis of each single pattern of the measured grid should be performed. Two main reasons discouraged this approach: the long computational time necessary to analyse the 756 patterns shown in Fig. 1 and the fact that only an approximate analysis was provided, taking into account that the patterns show quite low Poisson statistics (mainly due to the CCD saturation problems produced by large-crystal spots), very often coupled to a non-statistical representation of all crystal orientations (mainly related to the preferred orientation of the newly formed crystals), as required for an ideal powder. As a matter of fact, we argued that CCA could offer a different approach to this kind of

problem with respect to the standard RQPA. Indeed, rather than the pure phase patterns, theoretical mixture profiles can also be used as model signals for CCA. However, this choice raises some problems. In fact, because quantitative phase information is not *a priori* available, a huge number of phases fractions combinations is theoretically possible. To avoid wasting time by including unsuitable pattern models or, on the other hand, possible arbitrary choices of a limited compositional range, we decided to pick the set of model spectra out of the experimental high-resolution XRPD data collected on the same sample, from which we extracted phase combinations by means of a RQPA. As already mentioned in §2, three different regions of the investigated sample, located at the bone/scaffold interface, were scanned using a resolution step of 1 µm. A few patterns were selected for each region, by applying an 'adaptive binning' (see Sanders & Fabian, 2001, and references therein) criterium. On the first iteration, the first pattern corresponds to the region centre, while the next four correspond to the corners of a full size square centred on the first one; on the second iteration, a half-size square is centred on each point resulting from the first iteration, and another four patterns are picked up at the corners falling within the region of interest, and so on (see Fig. 4). Following such a criterium, nine, five and five patterns were selected for the three regions, respectively. Nine patterns corresponds to a two-iterations procedure on a square region, while five patterns corresponds to a two-iterations procedure on a straight line region (eight patterns shrink to four). To overcome the problem of low counting statistics of the 19 experimental patterns so far obtained, the CCA method was applied, as a classifier, to look for the 20 patterns that were best correlated to each of them within the corresponding region. In fact, it is both experimentally and computationally observed that the higher is the canonical correlation among the one-dimensional profiles, the closer are the two-dimensional patterns in the involved area (see Ladisa *et al.*, 2007), and since we might expect that the material changes do not occur on space scales larger than a few micrometres, the nearest patterns should be relevant within the CCA models involving information from neighbours. Accordingly, these patterns were summed to obtain signal statistics high enough to perform a more reliable RQPA. The classifying power of CCA is seen in Fig. 5, where both the high-resolution signals selected by CCA as the most correlated and the final signal are shown. The agreement is satisfactory. The plot at the bottom of Fig. 5 highlights some small differences among the 20 summed patterns, which are likely to depend also on small compositional variations; these are, however, difficult to estimate with good accuracy due to the low counting statistics of each profile.

Finally, the new set of 19 signal models (each is now the sum of 20 patterns) was further shrunk to nine final signals. Two main reasons determined such a choice. The first reason is that some of them are highly correlated and, indeed, appear to be very similar to each other (see Fig. 6); additionally, they show pretty close phase concentrations, as estimated by a RQPA (all the 19 patterns were analysed). The second reason is the convenience of reducing the number of pattern models, so making the final classification map more easily interpretable (even though this sacrifices some finer details).

The nine final patterns are then considered as representative of the majority (within a small range of compositional variability) throughout the sample.

The program *Quanto* was used to perform the analysis. A batch file to launch an automatic run of the program for each of the nine inputs was prepared. The outputs were available after a few minutes on a Linux PC with an Intel Xeon 3.06 GHz processor.

The results are given in Table 1, together with the profile agreement factor R_{wp} and the goodness of fit (GoF). It can be observed that some patterns do contain only the amorphous component mixed with (models 5, 8 and 9) or without (model 6) HA-n. The best fit of the pattern model number 1 is also

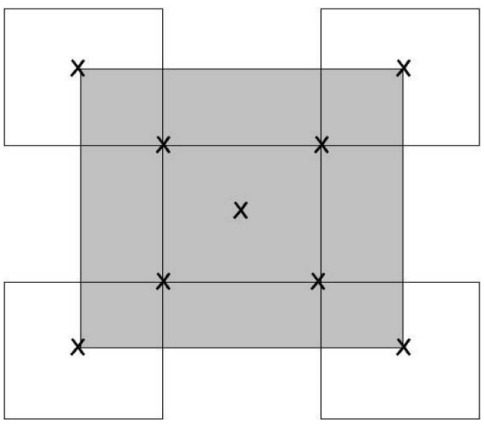

Figure 4
Adaptive binning criterium applied to the square region of Fig. 1 to pick up some model patterns. The shaded area is the region of interest.

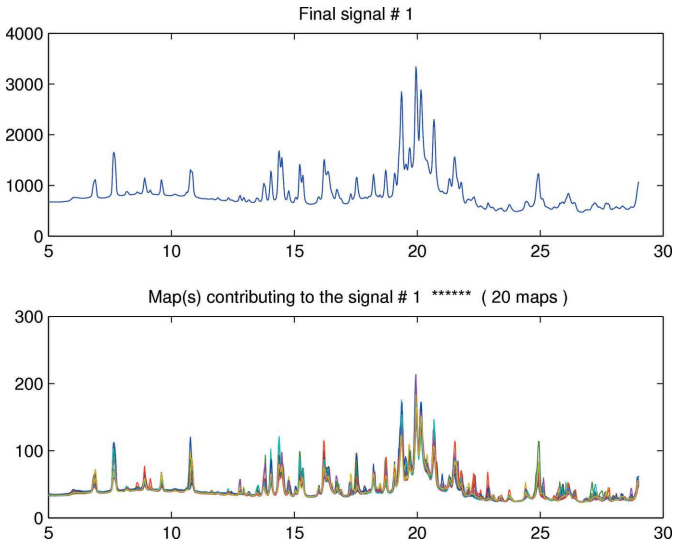

Figure 5
Plot of the 20 high-resolution one-dimensional patterns best correlated to that of the signal model (lower) according to the CCA approach, and the final pattern obtained as a sum of the previous 20 (upper).

Table 1
Weight percentages of the crystalline phases as estimated for the nine pattern models.

The profile agreement index (R_{wp}) and goodness of fit (GoF) are given in the last two columns.

Model No.	HA-n	HA-m	Si-Tcp	β-Tcp	R_{wp}	GoF
1	44.4 (7)	19.1 (4)	31.7 (7)	4.8 (2)	6.22	1.83
2	73.9 (9)	18.1 (9)	8.0 (1.7)	0.0	1.59	1.51
3	48.8 (4)	13.9 (3)	33.6 (5)	3.7 (2)	5.05	1.46
4	68.4 (5)	13.4 (5)	16.0 (5)	2.2 (3)	1.89	1.57
5	100.0 (–)	–	–	–	5.96	1.77
6	–	–	–	–	–	–
7	87.7 (4)	8.0 (2)	2.4 (2)	1.9 (2)	5.41	1.59
8	100.0 (–)	–	–	–	4.68	1.36
9	100.0 (–)	–	–	–	2.22	1.51

Table 2
Weight percentages of the four crystalline phases given in Table 1 for the nine profile models after rescaling with respect to the total sample volume.

An estimate of the amorphous fraction (Amor) is given.

Model No.	Amor	HA-n	HA-m	Si-Tcp	β-Tcp
1	32.0 (–)	30.2 (5)	13.0 (3)	21.5 (5)	3.3 (1)
2	95.7 (–)	3.17 (4)	0.78 (4)	0.30 (7)	0.0 (–)
3	48.2 (–)	25.3 (2)	7.2 (2)	17.4 (3)	1.9 (1)
4	92.2 (–)	5.33 (4)	1.05 (4)	1.25 (4)	0.17 (2)
5	12.4 (–)	87.6 (–)	–	–	–
6	100.0 (–)	–	–	–	–
7	16.8 (–)	73.0 (3)	6.6 (2)	2.0 (2)	1.6 (2)
8	36.3 (–)	63.7 (–)	–	–	–
9	85.8 (–)	14.2 (–)	–	–	–

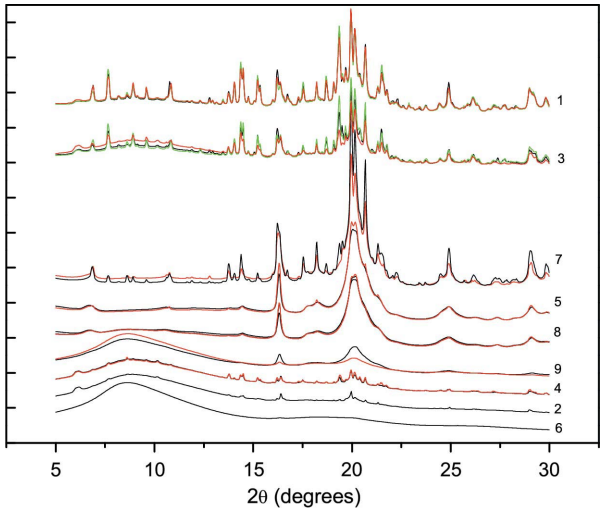

Figure 6
Plot of all the original 19 profiles provided by the 'adaptive binning' scheme grouped according to the final nine pattern models. The nine models are shifted on purpose (each with highly correlated patterns), along the vertical axis, in order to be clearly distinguishable in the plot; for each model, the identification number corresponds to that given in Table 1.

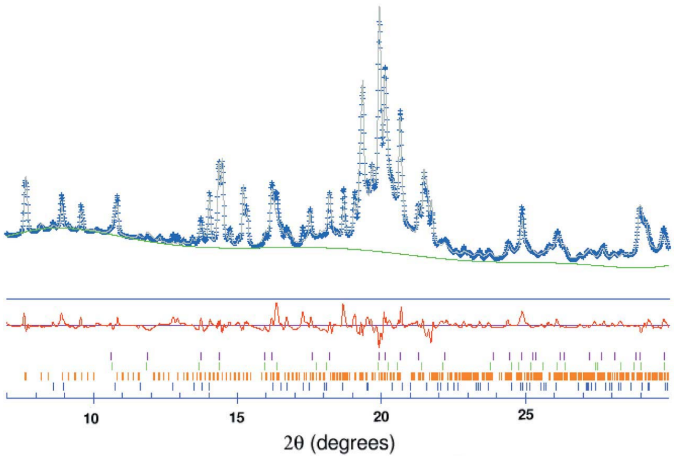

Figure 7
Quanto final best fit corresponding to the profile model number 1 in Table 1. R_{wp} = 6.22, GoF = 1.87.

shown in Fig. 7. For the same profile model, the RQPA of each of the 20 component patterns was performed. The corresponding mean and standard deviation of each phase weight percentage have been calculated and can be compared with the corresponding results of the pattern sum in Table 1. They are (in %): 45.3 ± 6.1 for HA-n, 17.4 ± 4.9 for HA-m, 32.6 ± 3.9 for Si-Tcp and 4.8 ± 2.5 for β-Tcp.

However, as the amorphous component is not included, the results shown in Table 1 do not allow an exhaustive description of the investigated problem, with the additional difficulty that, sometimes, the amorphous phase becomes the only detected phase. A reasonable way of deriving this information was to extract it (even though only approximately) from its contribution to the background curve, modelled by a Chebyshev polynomial with adjusted coefficients. After subtracting a constant noise level, the total integral area under the background and the experimental pattern were computed at the end of the refinement process. These values have been used to separate the volume fractions of amorphous and crystalline components and, then, to rescale the estimated weight percentages (and the corresponding standard deviations) of the crystalline phases of Table 1 to the total sample volume. Volume fractions of the amorphous component can be seen as corresponding mass fractions, within an acceptable approximation, provided that the average density of the amorphous component is not very different from that of the crystalline one. These new values are given in Table 2.

The nine intensity spectra so far analysed represent our profile models for the final step, where the CCA approach is applied to the whole sample following the same procedure described in the previous paragraph. The computation time was about 2 min, using *MatLab7* routines on a machine with an Intel Xeon 2.80 GHz processor. The resultant classification map is shown in Fig. 8, where each of the grey tones corresponds to one of the signal models, identified in the legend by the numbers given in Table 1. It is worth noting the resemblance of this map with that in Fig. 1. Furthermore, a quantitative characterization of the analysed tissue is provided for each pixel. A detail is given in Fig. 9, where the trend of the mass fraction for each phase along column 17 of Fig. 8 is supplied.

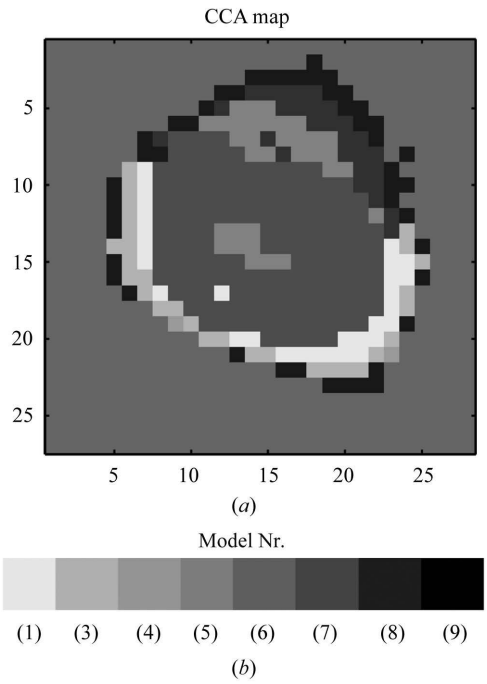

Figure 8
Nosologic map corresponding to the nine profile models selected for the bone tissue. The grey tones are related to the nine models according to the numbers given in Table 1.

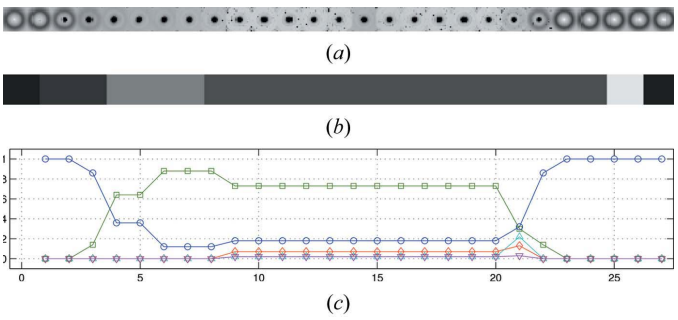

Figure 9
Details of the analysis along column 17 of the investigated grid: set of the corresponding two-dimensional patterns (*a*); classification map (*b*); mass fraction of each bone-tissue component [amorphous (circles), HA-n (squares), HA-m (rhombi), Si-Tcp (triangles), β-Tcp (stars)] (*c*).

5. Conclusions

An original method has been presented to show how the canonical correlation analysis technique can be used to describe the spatial variation of phase concentration over a grid of microdiffraction patterns collected on a sample of interest within the field of bone-tissue engineering. The phase quantification is provided by the Rietveld quantitative analysis of a few patterns, chosen as signal models to represent a reasonable range of compositions for the investigated sample. Advanced microdiffraction techniques allowed the sample to be scanned with a high spatial resolution, while a fast CCD detector made it possible to collect a huge number of two-dimensional images in acceptable time. As far as we are aware, this is the first time that CCA is applied to a grid of powder diffraction spectra in combination with a Rietveld quantitative phase approach. Indeed, the standard usage of CCA requires theoretical profiles to be used as signal models, instead of patterns picked up from the same experimental data set. When using computed patterns of the pure phases occurring in the investigated sample, the final classification image, obtained by combining the correlation maps corresponding to each signal model (*i.e.* by assigning the considered pixel to the profile model with the highest canonical coefficient), has a rather qualitative content, even though it is provided without any user intervention.

Selecting experimental profiles as signal models is a novelty for the CCA approach. It has the advantage that quantitative information is directly extracted from the measured data and, then, extended to other portions of the investigated sample, thanks to the high degree of correlation detected by the CCA. Furthermore, the final classification map includes hundreds of diffraction patterns, the RQPA of which requires a very long computational time, even if performed with automatic tools. On the other hand, it is worth outlining that the classification map cannot offer a spatial variation as accurate as that achievable by a standard RQPA performed on every single pattern.

An additional advantage of the CCA method is that it can also be applied to improve the Poisson statistics of experimental patterns, by summing up those spectra correlated by the technique, as has been done for the signal models of the present application. Needless to say that such an application is useless provided sufficiently high statistics over the set of XRPD patterns.

The approach described in this article explored the spatial variation of phase concentration in a bone-tissue-engineered bioceramic scaffold. This is a feature of great interest when investigating the bone-formation process related to biomaterials that are being developed as bone-graft substitutes. Indeed, it allows the supervision of the mineralization process with respect to the selective resorption of the original scaffold phases. This aspect has been discussed in a recent paper (Mastrogiacomo *et al.*, 2007), as a function of the implantation time, the scaffold composition and the cell dependency of the scaffold resorption itself. The information available through the classification map of the investigated sample confirms that in the contact areas between the newly deposited bone and the ceramic scaffold, the TCP component of the ceramic decreased much faster than the HA component. Note that Si-TCP ceramics are essentially insoluble in biological media, but are reabsorbed when acted upon by osteoclasts (large multinucleated cells that function in bone resorption). Such information can help to clarify a possible mechanism in the complex and dynamic bone-modelling process.

Finally, it is worth highlighting the potential of the canonical correlation analysis when combined with microdiffraction scanning techniques: in principle, such a combination can be used for investigating the spatial variations of other sample features (such as the crystal size distribution of the new bone) besides the phase concentration shown in the present article.

APPENDIX *A*
CCA in terms of principal angles

CCA in terms of principal angles is the most reliable and fastest implementation of canonical correlation analysis. An outline follows.

A1. CCA by computing principal angles

Given the zero-mean multivariate random vectors $\mathbf{x} = [x_1(t), \ldots, x_m(t)]$ and $\mathbf{y} = [y_1(t), \ldots, y_n(t)]$, with $t = 1, \ldots, N$, perform the following steps.

Step 1

Consider the matrices $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$, defined as follows

$$\widetilde{\mathbf{X}} = \begin{bmatrix} x_1(1) \cdots x_m(1) \\ \vdots\vdots \\ x_1(N) \cdots x_m(N) \end{bmatrix}, \quad \widetilde{\mathbf{Y}} = \begin{bmatrix} y_1(1) \cdots y_n(1) \\ \vdots\vdots \\ y_1(N) \cdots y_n(N) \end{bmatrix}. \quad (5)$$

Step 2

Compute the QR decompositions (Golub & Van Loan, 1996) of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ as $\widetilde{\mathbf{X}} = \mathbf{Q}_{\tilde{X}}\mathbf{R}_{\tilde{X}}$ and $\widetilde{\mathbf{Y}} = \mathbf{Q}_{\tilde{Y}}\mathbf{R}_{\tilde{Y}}$, where $\mathbf{Q}_{\tilde{X}}$ and $\mathbf{Q}_{\tilde{Y}}$ are orthogonal matrices and $\mathbf{R}_{\tilde{X}}$ and $\mathbf{R}_{\tilde{Y}}$ are upper triangular matrices.

Step 3

Compute the SVD (Golub & Van Loan, 1996) of $\mathbf{Q}_{\tilde{X}}^{T}\mathbf{Q}_{\tilde{Y}}$ as $\mathbf{Q}_{\tilde{X}}^{\mathrm{T}}\mathbf{Q}_{\tilde{Y}} = \mathbf{USV}^{\mathrm{T}}$, where $\mathbf{S}$ is a diagonal matrix and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. The cosines of the principal angles are given by the diagonal elements of $\mathbf{S}$.

Step 4

Set the canonical correlation coefficients equal to the diagonal elements of the matrix $\mathbf{S}$ and compute the corresponding regression weights as $\mathbf{w}_{\tilde{X}} = \mathbf{R}_{\tilde{X}}^{-1}\mathbf{U}$ and $\mathbf{w}_{\tilde{Y}} = \mathbf{R}_{\tilde{Y}}^{-1}\mathbf{V}$.

The computation of the principal angles yields the most robust implementation of CCA, since it is able to provide reliable results even when the matrices $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ are rank-deficient.

A2. Choice of the spatial model

As already mentioned in §3, several spatial models can be chosen when applying CCA. As a particular case, the ordinary correlation analysis can be considered as a single-pixel model. The performance of the following spatial models was investigated.

(i) The single-pixel model.

(ii) The 3×3 model:

$$\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_9].$$

(iii) The 3×3 model without corner pixels:

$$\mathbf{x} = \left[\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_8\right]^{\mathrm{T}}.$$

(iv) The symmetric 3×3 model:

$$\mathbf{x} = \left[\mathbf{x}_5, \frac{\mathbf{x}_1 + \mathbf{x}_9}{2}, \frac{\mathbf{x}_2 + \mathbf{x}_8}{2}, \frac{\mathbf{x}_3 + \mathbf{x}_7}{2}, \frac{\mathbf{x}_4 + \mathbf{x}_6}{2}\right]^{\mathrm{T}}.$$

(v) The symmetric 3×3 model without corner pixels:

$$\mathbf{x} = [\mathbf{x}_5, (\mathbf{x}_2 + \mathbf{x}_8)/2, (\mathbf{x}_4 + \mathbf{x}_6)/2]^{\mathrm{T}}.$$

The best performance was obtained by applying the symmetric 3×3 model without corner pixels (for further details, see Ladisa *et al.*, 2007).

A3. The Taylor subspace model

Concerning the choice of the $\mathbf{y}$ variable, the so-called Taylor model was considered in order to define the proper signal subspace able to model the characteristic spectra of the investigated sample and their possible variations.

Step 1

Choose the profile model $P(n)$, $n = 1, \ldots, N$, where $N = 1024$, corresponding to the considered material.

Step 2

Set the components of the variable $\mathbf{y}$ as $y_1(n) = P(n)$, $y_2(n) = [P(n+1) - P(n)]/\Delta\theta$, where $n = 1, \ldots, 1024$, and $\Delta\theta = 0.012^{\circ}$ is the sampling angle.

For the single-pixel approach, only one component was considered and set equal to the first component of the Taylor subspace model, namely $\mathbf{y} = \mathbf{y}_1$.

References

Altomare, A., Burla, M. C., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Rizzi, R. (2001). *J. Appl. Cryst.* **34**, 392–397.

Bish, D. L. & Howard, S. A. (1988). *J. Appl. Cryst.* **21**, 86–91.

Cedola, A., Mastrogiacomo, M., Burghammer, M., Komlev, V., Giannoni, P., Cancedda, R., Rustichelli, F., Favia, A. & Lagomarsino, S. (2006). *Phys. Med. Biol.* **51**, N109–N116.

Cedola, A., Mastrogiacomo, M., Lagomarsino, S., Cancedda, R., Giannini, C., Guagliardi, A., Ladisa, M., Burghammer, M., Rustichelli, F. & Komlev, V. (2007). *Spectrochim. Acta B*, **62**, 642–647.

Cervellino, A., Giannini, C., Guagliardi, A. & Ladisa, M. (2006). *J. Appl. Cryst.* **39**, 745–748.

Dalconi, M. C., Meneghini, C., Nuzzo, S., Wenk, R. & Mobilio, S. (2003). *Nucl. Instrum. Methods Phys. Res. B*, **200**, 406–410.

De Edelenyi, F. S., Rubin, C., Estève, F., Grand, S., Décorps, M., Lefournier, V., Le Bas, J. F. & Rémy, C. (2000). *Nat. Med.* **6**, 1287–1289.

Friman, O. (2003). PhD thesis, Department of Biomedical Engineering, Linkpings University, Sweden.

Golub, G. H. & Van Loan, C. (1996). *Matrix Computations*. New York: Johns Hopkins University Press.

Hill, R. J. & Howard, C. J. (1987). *J. Appl. Cryst.* **20**, 467–474.

Hotelling, H. (1936). *Biometrika*, **28**, 321–377.

Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th ed. New Jersey: Prentice Hall.

Komlev, V., Peyrin, F., Mastrogiacomo, M., Cedola, A., Papadimitropoulos, A., Rustichelli, F. & Cancedda, R. (2006). *Tissue Eng.* **12**, 3449–3458.

Ladisa, M., Lamura, A. & Laudadio, T. (2007). *EURASIP J. Adv. Sign. Proc.* **2007**, 19260(1–8).

Larson, A. C. & Von Dreele, R. B. (1994). Los Alamos National Laboratory Report LAUR 86–748, New Mexico, USA.

Laudadio, T., Pels, P., De Lathauwer, L., Van Hecke, P. & Van Huffel, S. (2005). *Magn. Reson. Med.* **54**, 1519–1529.

Lutterotti, L. (2005). *CPD Newsl.* **32**, 53–55.

Madsen, I. C., Scarlett, N. V. Y., Cranswick, L. M. D. & Lwin, T. (2001). *J. Appl. Cryst.* **34**, 409–426.
Mastrogiacomo, M., Papadimitropoulos, A., Cedola, A., Peyrin, F., Giannoni, P., Pearce, S. G., Alini, M., Giannini, C., Guagliardi, A. & Cancedda, R. (2007). *Biomaterials*, **28**, 1376–1384.
Quarto, R., Mastrogiacomo, M., Cancedda, R., Kutepov, S. M., Mukhachev, V. & Lavroukov, A. (2001). *N. Engl. J. Med.* **344**, 385.
Rieckel, C. & Davies, R. J. (2005). *Curr. Opin. Colloid Interface Sci.* **9**, 396–403.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Rodriguez-Carvajal, J. (1992). *Physica B*, **192**, 55.
Sanders, J. S. & Fabian, A. C. (2001). *Mon. Not. R. Astron. Soc.* **325**, 178.
Sayer, M., Stratilatov, A. D., Reid, J., Calderin, L., Stott, M. J., Yin, X., MacKenzie, M., Smith, T. J. N., Hendry, S. D. & Langstaff, S. D. (2003). *Biomaterials*, **24**, 369–382.
Scarlett, N. V. Y., Madsen, I. C., Cranswick, L. M. D., Lwin, T., Groleau, E., Stephenson, G., Aylmore, M. & Agron-Olshina, N. (2002). *J. Solid State Chem.* **175**, 272–277.
Yashima, M., Sakai, A., Kamiyama, T. & Hoshikawa, A. (2003). *J. Solid State Chem.* **175**, 272–277.
Wagermaier, W., Gupta, H. S., Gourrier, A., Paris, O., Roschger, P., Burghammer, M., Riekel, C. & Fratzl, P. (2007). *J. Appl. Cryst.* **40**, 115–120.
Zha, H. (1993). PhD thesis, Pennsylvania State University, USA.