JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 13 May 2015
Accepted 22 February 2016

Edited by V. T. Forsyth, Institut Laue–Langevin, France, and Keele University, UK

Keywords: functional data analysis; structural parameters; polymer composites.

Functional data analysis techniques for the study of structural parameters in polymer composites

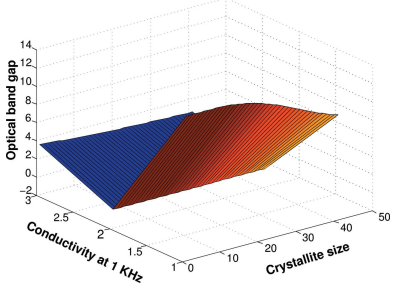
Thejas Gopal Krishne Urs,[a,c] Karthik Bharath,[b] Sangappa Yallappa[c] and Somashekar Rudrappa[a]*

[a]Department of Studies in Physics, University of Mysore, Manasagangotri, Mysore, India, [b]School of Mathematics, University of Nottingham, Nottingham, UK, and [c]Department of Physics, Mangalore University, Mangalore, India. *Correspondence e-mail: rs@physics.uni-mysore.ac.in

This article presents a novel method, based on functional data analysis, to analyse measurements of structural parameters of polymers and polymer composites. The method is demonstrated using newly developed biodegradable conducting polymer composites prepared *via* a solution casting technique. The measurements of the macro- and microstructural parameters that are used in the characterization of these films are obtained using X-ray diffraction, an impedance analyser and a UV–vis spectrometer. A functional representation of the measured values of the parameters at different dopant concentrations is adopted by viewing them as realizations of a continuous-time stochastic process observed with measurement error. This allows one to estimate the mean functional relationship between a parameter and the dopant concentration. A functional version of principal component analysis is performed, by which the major modes of variation are discovered and the correlations of parameter values at different concentrations are estimated. This provides insight into local and global features of the relationship between these parameters. Some comments are made on how the parameters vary as a function of dopant concentration.

1. Introduction

Characterization of polymer composites that are conducting and biodegradable using micro- and macrostructural parameters is an important task in view of the broad applicability of such composites. For instance, a conducting polymer can be used as a cathode material in lithium ion batteries, gas sensors, biosensors (Ates *et al.*, 2012), thin-film transistors (Halls *et al.*, 1995) and supercapacitors. However, reliable statistical analysis of data obtained from measurements of these parameters is hindered by issues of small sample size and measurement error (Ida & Izumi, 2011; Ida, 2011). The overarching objective in such analyses is the accurate estimation of the behaviour of the structural parameters as a function of the dopant concentration. A typical example would be that of the measurement of a parameter across different dopant concentrations for a single preparation or sample; when this is repeated for different samples, the resulting data can be viewed as independent realizations, one for each prepared sample, of the true unknown function representing the relationship between the parameter and the concentration. In other words, the statistical unit is now a functional observation, as opposed to a scalar or a vector, representing values of the parameters collected over an ordered axis (concentration). To this end, in this article, we use techniques from functional data analysis (FDA), which is



© 2016 International Union of Crystallography

becoming an increasingly important area of research within statistics. The framework is based on representing data collected over an ordered variable, such as time, or in this article dopant concentration, as discretely observed trajectories or functions. Such a representation leads to a coherent framework for estimating the true functional relationship between a parameter and dopant concentration by allowing us to borrow strength from within and across preparations or samples.

In this article, we describe the development of a new conducting polymer composite, which is biodegradable, and conduct a statistical study of the experimental parameters using techniques from FDA. Specifically, we use hydroxypropyl methyl cellulose (HPMC) and the metal inorganic salt nickel chloride ($NiCl_2$) hexahydrate to prepare polymer membranes by doping with $NiCl_2$ at various concentrations. These membranes are subjected to different analytical measurements, with the objective of investigating their physico-mechanical properties. We characterize the structural properties of the films using structural parameters obtained *via* measurements from X-ray diffraction, an impedance analyser, a tensile strength measuring unit and a UV–visible spectrometer. Owing to measurement errors, intrinsic properties of the chosen sample and potential defects introduced during the growth of the material, the data corresponding to analytical measurement of the parameters do not, generally, exhibit a linear trend with respect to the dopant concentration; indeed, there is no apparent reason to assume so.

In view of the issues mentioned above, our objective in this article is threefold:

(1) for each sample, predict the functional relationship between the parameter of interest and the dopant concentration, and consequently estimate the average or mean function across the samples;

(2) using the mean function for each parameter, estimate the dependence of its values across dopant concentration;

(3) investigate the functional relationships between parameters as a function of dopant concentration.

The outlined objective involves global and local properties of the underlying functional relationship between a structural parameter and the dopant concentration. The global aspect is captured in the mean function, while the local information needs to be extracted from the individual samples; this involves exploring different modes of variation of the functional relationship through a basis expansion. Therefore we perform a functional version of principal component analysis (FPCA), commonly used in multivariate statistics, to obtain a basis representation of the functional relation between the parameter and dopant concentration in each sample in terms of an orthonormal basis of eigenfunctions. From this, based on the data across dopant concentrations from different samples, we first estimate the average or mean functional relationship; using the estimated mean function, we then estimate the correlation structure amongst the parameter values for different concentration levels. This provides an insight into structural components of the polymer corresponding to changes in the dopant concentrations. Armed with the mean function for each parameter, we then look into their relationships and uncover some interesting properties.

2. Theory and methodology

We review briefly the theoretical aspects of FDA and FPCA relevant for our work. We refer the interested reader to an excellent introduction to FDA in the book by Ramsay & Silverman (2005), and to Yao *et al.* (2005) for the version of FPCA used in this article.

2.1. Functional data

Statistical analysis is based on assuming that data from measurements are realizations of random variables. A collection of random variables over an index set, usually infinite, is referred to as a stochastic process. Functional data consist of a collection of random functions which are viewed as independent and identical realizations of an underlying stochastic process. In our setting, if θ is the structural parameter of interest, and t represents the dopant concentration level measured in percentage, we view $\{\theta(t), t \in \mathbb{R}_{\geq 0}\}$ as a stochastic process, in order to model the uncertainty and variability arising from measurements of θ at different t. Then for one sample, data from N measurements of θ at concentration levels $t_1, \ldots, t_N$ can be represented as $\theta(t_1), \theta(t_2), \ldots, \theta(t_N)$. The key observation is that the set $[\theta(t_1), \theta(t_2), \ldots, \theta(t_N)]$ cannot be viewed as a vector since there is a natural ordering of the set $\mathcal{T}$ arising from the assumed values of t; in other words, in contrast to multivariate data, irrespective of dimension, no meaningful topology on the domain exists. This is illustrated by the fact that one can re-order the components of a multivariate data vector and arrive at exactly the same statistical analysis as for the data vector arranged in the original order; this is certainly not true for functional data. Indeed, there is no requirement for each sample to contain the same number of measurements. For n preparations or samples, if each sample i contains measurements of θ at concentrations $t_1, \ldots, t_{n_i}$, the data can be represented as an $n \times n_i$ array with entries $\{\theta_i(t_j) : i = 1, \ldots, n; j = 1, \ldots, n_i\}$.

At this point, two issues arise. (1) If θ has been observed with a high sampling rate with large n_i, then the mean function can be estimated with point-wise estimates. Unfortunately, this is typically not the case, and the measurements of θ are sparse and irregular: each sample i can contain measurements at different concentration levels, and the number of measurements in each sample is usually low. (2) The current representation does not take into account measurement or experimental errors. There are several approaches to address these issues; for a detailed account, see Ramsay & Silverman (2005). Our approach is based on the representation of the stochastic process $\theta(t)$ in terms of the eigenfunctions of a suitable operator.

2.2. Representation and FPCA

The stochastic process $\{\theta(t) : t \in \mathbb{R}_{\geq 0}\}$ is an infinite-dimensional object; for tractable computations the goal here is to

summarize $\theta(t)$ through a finite number of coefficients of a suitable expansion known as functional principal component scores (FPC). The reason for this approach, as described earlier, is first to predict individual sample trajectories $\theta_i(t)$ for all $t \in \mathbb{R}_{\geq 0}$ along with experimental errors through the definition of a formal model, and second to estimate the correlation between values of θ at different t. The two tasks are related through the estimation of the mean or average function, which captures the global behaviour across all samples. To this end, we model $\theta(t)$ as a square-integrable stochastic process with mean function $\mu(t) = E[\theta(t)]$ and covariance function $G(s, t) = \mathrm{Cov}[\theta(s), \theta(t)]$, where the expectation and variance operators are $E(\cdot)$ and $V(\cdot)$, and $\mathrm{Cov}(X, Y)$ is the covariance between random variables X and Y. Analogously to the expansion of a continuous function in terms of Fourier harmonics, an appropriate expansion of $\theta(t)$ will be adopted in terms of the eigenfunctions of the autocovariance operator (Yao *et al.*, 2005):

$$(\mathrm{A}f)(t) = \int_{\mathbb{R}_{\geq 0}} f(s)G(s, t)\,\mathrm{d}s; \tag{1}$$

that is, if $\varphi_k, k = 1, 2, \ldots$ are the eigenfunctions of the autocovariance operator, with associated eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$, then

$$A\varphi_k = \lambda_k \varphi_k. \tag{2}$$

We adopt the following model for $\theta(t)$ based on the Karhunen–Lóeve expansion of a square-integrable stochastic process:

$$\theta(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \varphi_k(t), \tag{3}$$

where A_k are uncorrelated random variables, which are the FPC, and $\mu(t)$ is the deterministic mean function. They satisfy $E(A_k) = 0$, $V(A_k) < \infty$ and have the explicit representation

$$A_k = \int_{\mathbb{R}_{\geq 0}} [\theta(t) - \mu(t)]\varphi_k(t)\,\mathrm{d}t. \tag{4}$$

The main advantage in employing the eigenbasis of the autocovariance operator, as opposed to other bases, lies in the fact that the random coefficients A_k are uncorrelated, facilitating computational implementation. This is not necessarily true for other basis representations, and since the eigenfunctions are obtained from the autocovariance operator, we are implicitly able to capture the dependence structure of the functions across and within samples.

In order to estimate $\hat{\mu}(t)$ of $\mu(t)$ from the data, we need to move from the continuous representation of $\theta(t)$ to the discrete version involving data $\{\theta_i(t_j) : i = 1, \ldots, n; j = 1, \ldots, n_i\}$ and we need to incorporate experimental error; we denote

$$\theta_{ij} = \theta_i(t_j) + \varepsilon_{ij} = \mu(t_{ij}) + \sum_{k=1}^{\infty} A_{ik}\varphi_k(t_{ij}) + \varepsilon_{ij}, \tag{5}$$

where ε_{ij} are independent and identically random variables with $E(\varepsilon_{ij}) = 0$ and $V(\varepsilon_{ij}) < \infty$. From the representations in equations (3) and (5), we observe that $\mu(\cdot)$ is the deterministic component and the sum is the stochastic component. This provides the following interpretation of observed θ as a function of t: what one observes upon measurement is a additive combination of the average behaviour of the parameter with respect to t and the unobserved experimental error, which itself is a function of t. This is a very general setup wherein the amount of experimental error is not assumed to be uniform and can vary depending on the dopant concentration. The advantage of such representations is that $\hat{\mu}(t)$ and the correlation surface can be computed by borrowing strength across samples and between the observations of θ at different concentration levels; this is particularly so in our setting since the measurements are usually sparse and irregular.

2.3. Estimates of mean function and correlation surface

The mean function is estimated by first pooling together all available measurements $\{\theta_{ij}, t_{ij}\}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Noting that this implies a borrowing of information from within and across samples or preparations, we can use a smooth interpolation mechanism to estimate $\mu(t)$. We use the smoothing spline estimate $\hat{\mu}(t)$ which is the unique solution to the optimization problem

$$\mathrm{argmin}_{f \in \mathcal{F}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} [\theta_{ij} - f(t_{ij})]^2 + \lambda \int_{\mathbb{R}_{\geq 0}} [f''(t)]^2\,\mathrm{d}t, \tag{6}$$

where $\mathcal{F}$ is the class of square-integrable functions with at least two derivatives. This is reasonable based on the initial assumption of $\theta(t)$ satisfying square integrability, and it can be shown that as the number of measurements increases without bound $\hat{\mu}(t)$ converges, in an appropriate sense, to $\mu(t)$.

Recall first that correlation between two random variables is a dimensionless quantity, bounded between -1 and 1, defined as the covariance between two random variables scaled by their respective standard deviations. The correlation function between $\theta(s)$ and $\theta(t)$ for $s \neq t$ can be represented as a parametrized surface $\rho : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to [-1, 1]$. We wish to estimate this surface using θ_{ij}. The covariance surface is estimated by taking pair-wise products

$$[\theta_{ij} - \hat{\mu}(t_{ij})][\theta_{ik} - \hat{\mu}(t_{ik})], \quad j \neq k, \tag{7}$$

from which the estimate of the correlation surface $\hat{\rho}$ is obtained. Once the estimate of the covariance has been obtained, estimates of the eigenfunctions $[\hat{\varphi}_k(t)]$, functional principal scores $(\hat{A}_k)$ and eigenvalues $(\hat{\lambda}_k)$ are computed, leading to the required prediction of $\theta_i(t)$, the predicted behaviour of θ as a function of t for the ith sample:

$$\hat{\theta}_i(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{A}_{ik}\hat{\varphi}_k(t), \tag{8}$$

where the infinite sum has been truncated to an appropriate finite constant K. The procedures to estimate the FPC, eigenfunctions and eigenvalues are technically involved; for ease of exposition and brevity, we refer the interested reader to Yao *et al.* (2005) for details.

3. Experimental setup and measurements

We detail the polymer preparation and methods of measurement of the parameters; we then apply the methods described in the preceding section to the data obtained and investigate local and global properties of the relationship between a structural parameter and dopant concentration levels.

3.1. Preparation of polymer composites

Polymer composites were prepared by employing the regular solution casting method. The HPMC used in this study was procured by Loba Chemie, India, and $NiCl_2$ from Merck, India. A stock solution of 5% HPMC was prepared and doped with $NiCl_2$ of the required weight percentage to prepare 1, 2, 3 and 4% $NiCl_2$–HPMC polymer composites. Each solution was stirred thoroughly for over an hour, then poured onto casting plates and allowed to cast for a week. The dried polymer membranes were subsequently peeled off, and care was taken to avoid any mechanical distortions; these films were stored in zip-lock covers to avoid interaction with other contaminants and moisture. Polymer composites prepared in this fashion were found to be flexible and hygroscopic in nature for higher concentrations. The films were dried using a hot-air oven to remove the residual water molecules.

3.2. Microstructural changes observed with X-ray diffraction

Powder X-ray diffraction was recorded for our samples using a Rigaku MiniFlex II desktop X-ray diffractometer to identify the microstructural changes that occurred in the polymer matrix as a result of doping, and these changes were evaluated using the whole powder pattern fitting (WPPF) method (Scardi & Leoni, 2006; Balzar *et al.*, 2004; Koker *et al.*, 2014; Zilahi *et al.*, 2015). Measurements were recorded at an output of 30 kV and 15 mA of Cu $K\alpha$ radiation for a scanning angle between 6 and 60°, with a speed of 5° min^{-1} and a step size of 0.02°. The X-ray diffraction patterns obtained from the samples are shown in Fig. 1. The intensity data were corrected for instrumental broadening (Stokes, 1948) and Lorentz polarization factors as described in our earlier papers (Somashekar & Somashekarappa, 1997; Somashekarappa *et al.*, 2002). We consider the single-order method reported by Somashekar *et al.* (1989) and Hall & Somashekar (1991) to describe the microstructural information of these polymer composites. The efficiency of this single-order method has been confirmed by a recent round robin test. Thus, here we follow the same method and consider single order to explain the broadening of profiles to understand the microstructure (Hall & Somashekar, 1991; Langford *et al.*, 2000; Shahmoradi *et al.*, 2010).

Microstructural parameters such as crystal size N and lattice strain g (in %) are usually determined by employing the Fourier method of Warren & Averbach (1952). The intensity of a profile in the direction joining the origin to the centre of the reflection can be expanded in terms of a Fourier cosine series:

$$A(n) = \sum_{s=-s_0/2}^{s=s_0/2} I(s)\cos 2\pi n d(s - s_0), \tag{9}$$

where the coefficients of the harmonics $A(n)$ are functions of the size of the crystallite and disorder of the lattice. Here, s is $\sin(\theta)\lambda^{-1}$, s_0 is the value of s at the peak of the reflection, n is the harmonic number and d is the lattice spacing. The Fourier coefficients are computed as

$$A(n) = A_s(n) * A_d(n), \tag{10}$$

where $*$ is the convolution operator. For various harmonic numbers, an analytical expression for the size distribution yields an expression for crystallite size distribution as

$$A_s(n) = A(0)\frac{\exp[-\alpha(n-p)]}{\alpha N}, \tag{11}$$

and for lattice strain, the expression is

$$A_d(n) = \exp(-2\pi^2 n^2 m g^2). \tag{12}$$

Here the term N represents the number of unit cells participating in the Bragg reflection process, g is the lattice strain, α is the width of the distribution function and p is the smallest number of unit cells in a column; m is the order of reflection, which can be computed using the multidimensional minimization SIMPLEX program 2 (Press *et al.*, 1988). The experimental profile is matched with a simulated profile using the above parameters until a good fit is observed. The goodness of the fit is determined by computing

$$\Delta^2 = \frac{[I_{\rm cal} - (I_{\rm exp} + {\rm BG})]^2}{(npt)}, \tag{13}$$

where BG is the background. We have fitted the whole experimental pattern obtained by the model once in order to evaluate the microstructural parameters. There appears to be agreement within 1% of the mean value.

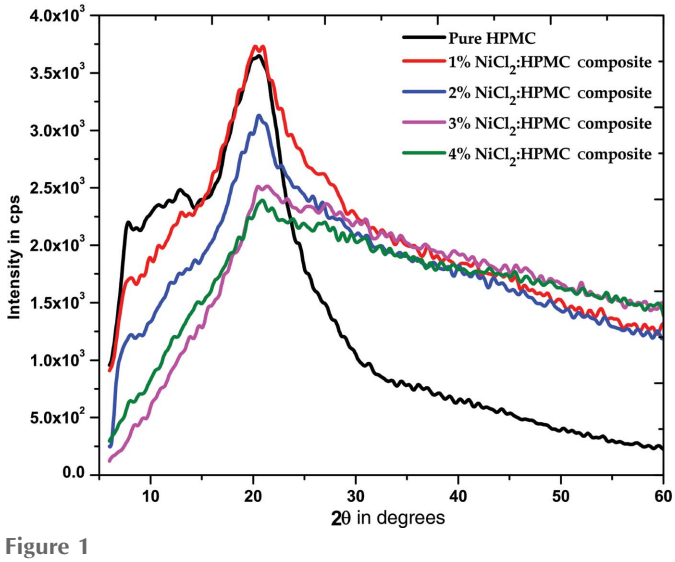


Figure 1
X-ray diffraction profiles obtained for the polymer composites.

Table 1
Microstructural parameters evaluated using WPPF and the other physical parameters evaluated from various other techniques.

Samples	Crystallite area (Å^2)	% Lattice strain	Conductivity at 1 kHz (S cm^{-1})	Conductivity at 10 kHz (S cm^{-1})	Optical band gap (eV)	Tensile strength (MPa)
Pure HPMC	12.87	0.005	4.92×10^{-11}	1.59×10^{-10}	3.56	60.13
	12.84	0.005	1.35×10^{-10}	1.26×10^{-9}	3.55	61.70
	14.9	0.005	1.30×10^{-10}	1.24×10^{-9}	3.54	58.30
	14.9	0.005	7.23×10^{-11}	3.08×10^{-10}	3.56	59.92
	13.10	0.005	5.83×10^{-11}	6.22×10^{-10}	3.54	63.25
	14.9	0.005	1.40×10^{-10}	1.15×10^{-9}	3.57	62.05
1% $NiCl_2$–HPMC	14.10	0.01	8.39×10^{-6}	9.44×10^{-6}	3.02	24.05
	14.9	0.01	8.04×10^{-6}	9.32×10^{-6}	3.00	23.79
	10.9	0.01	7.73×10^{-6}	9.06×10^{-6}	3.01	24.32
	13.46	0.01	7.41×10^{-6}	8.81×10^{-6}	3.01	23.47
	14.60	0.01	6.17×10^{-6}	7.88×10^{-6}	3.03	22.93
	14.41	0.01	5.75×10^{-6}	7.58×10^{-6}	3.02	25.17
2% $NiCl_2$–HPMC	14.27	0.01	5.69×10^{-5}	6.46×10^{-5}	2.82	25.32
	14.47	0.01	5.88×10^{-5}	6.61×10^{-5}	2.81	24.73
	15.32	0.01	5.37×10^{-5}	6.31×10^{-5}	2.80	25.41
	14.41	0.01	4.55×10^{-5}	5.63×10^{-5}	2.82	25.01
	15.07	0.01	4.50×10^{-5}	5.63×10^{-5}	2.81	24.32
	15.25	0.01	4.21×10^{-5}	5.42×10^{-5}	2.82	24.71
3% $NiCl_2$–HPMC	9.24	0.005	1.27×10^{-5}	2.66×10^{-5}	2.66	18.48
	9.7	0.005	1.89×10^{-5}	3.47×10^{-5}	2.65	18.08
	11.8	0.005	1.88×10^{-5}	3.46×10^{-5}	2.67	19.00
	10.34	0.005	1.86×10^{-5}	3.38×10^{-5}	2.66	18.50
	10.51	0.005	1.85×10^{-5}	3.36×10^{-5}	2.64	18.02
	10.00	0.005	1.27×10^{-5}	2.66×10^{-5}	2.65	18.40
4% $NiCl_2$–HPMC	9.81	0.005	5.79×10^{-5}	1.83×10^{-4}	2.56	11.85
	10.03	0.005	6.51×10^{-5}	1.92×10^{-4}	2.57	11.87
	10.12	0.005	6.77×10^{-5}	1.92×10^{-4}	2.56	11.83
	9.81	0.005	6.97×10^{-5}	1.92×10^{-4}	2.58	11.81
	9.54	0.005	7.06×10^{-5}	1.91×10^{-4}	2.56	11.89
	9.76	0.005	4.43×10^{-5}	1.51×10^{-4}	2.58	11.84

3.3. Absorbance investigated through UV–vis spectroscopy

UV–visible absorbance spectra of the samples, indicating macrostructural aspects of the polymer composites, were measured using a Labtronics LT-2800 double beam UV–visible spectrometer to verify the characteristic absorption of the polymer composites with respect to dopant concentration for a wavelength range from 190 to 700 nm. Using this absorption spectrum, the optical band gap was derived, for comparison with the conductivity results. This evaluation of optical band gap was done by employing a Tauc plot.

3.4. AC conductivity measurements

Measurements of AC conductivity were carried out to report the changes in the conducting ability of these samples due to doping. This test was carried out using a Hioki LCR Hi-Tester at room temperature for the frequency range of 50 Hz to 5 MHz. We obtain the conductance of the sample from the instrument, which is then evaluated to get the conductivity for the same respective frequencies using the equation

$$\sigma = \frac{G * \tau}{A}, \tag{14}$$

where σ is the conductivity, G is the obtained transconductance, τ is the thickness and A is the area of the sample.

3.5. Mechanical properties

The tensile strength, percentage elongation at break and other mechanical properties of the polymer composites were examined using a Universal Testing Machine, Zwick Z-2.5, ROELL, Germany. These properties are considerably affected by the addition of $NiCl_2$. The experiments were carried out adopting ASTM standards and for multiple trails. The tensile strength results are reported in Table 1.

4. Results from FDA and discussion

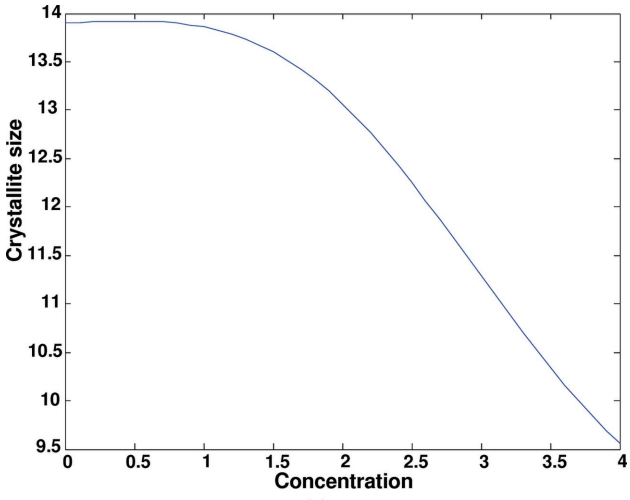
We report, in detail, the results and plots obtained for the microstructural parameter crystalline area; for the other parameters, we note that identical methods were used and hence present only relevant results. Computations were performed using the FPCA package (written in MATLAB; The MathWorks Inc., Natick, MA, USA) available at http://www.stat.ucdavis.edu/PACE/.

4.1. Crystallite size

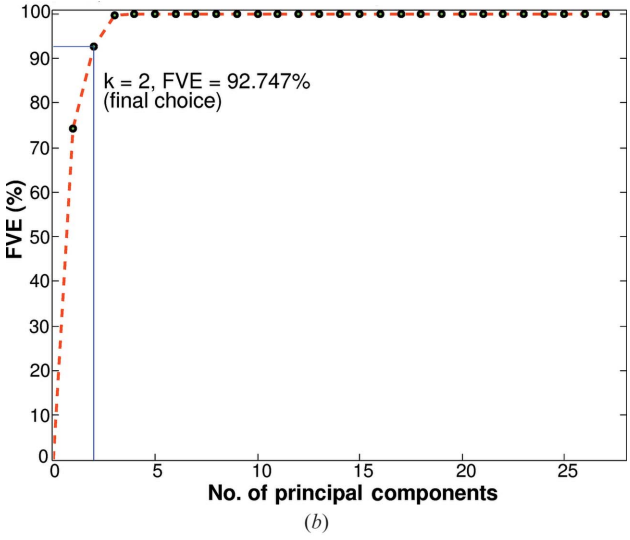
From the X-ray diffraction plots shown in Fig. 1, it is evident that the addition of $NiCl_2$ to the polymer matrix makes the polymer composites more amorphous, *i.e.* the narrow amorphous polymer halo becomes broader and almost vanishes. This is because $NiCl_2$ becomes interstitial in the polymer network and damages the parent polymer structure. The crystallite size values evaluated using the WPPF method support these results.

We observe from Fig. 2(*a*) that crystallite area typically decreases with increasing dopant concentration. The mean function has been estimated by solving the optimization problem given in equation (6) by combining measurements across samples. Fig. 2(*b*) shows a scree plot, which plots the percentage of variance in the measurement values of crystallite area (as a function of dopant concentration) that can be explained by the varying number of eigenfunctions used in the expansion. The plot suggests that only two principal components or eigenfunctions of the autocovariance operator are needed; this is shown in Fig. 2(*c*), which is a plot of the two estimated eigenfunctions. From this we can infer that, since approximately 93% of the variation is expressed in the first two eigenfunctions, the functional relationship between crystallite area and dopant concentration is mostly linear.

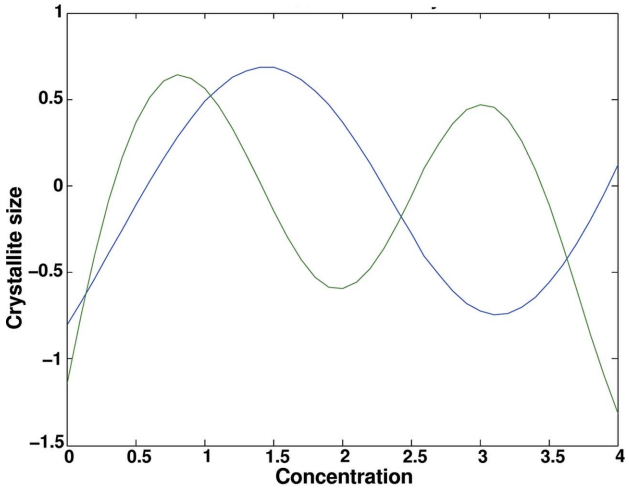
From Fig. 3(*a*), we get some insight into the dependency between crystallite areas at different dopant concentrations; note that this information is unavailable by just observing the mean function. For example, from the correlation plot we note that crystallite areas at 1 and 2.5% are strongly negatively correlated with a correlation close to −1. This informs us that high (low) values of crystallite area at 1% would imply that its area at 2.5% will be low (high). Such information will be particularly helpful in several applications wherein the polymer composites can be profitably used. Fig. 3(*b*) displays



(*a*)



(*b*)



(*c*)

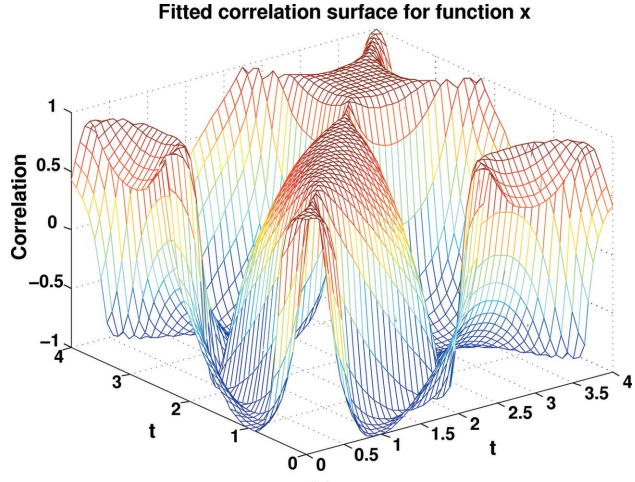
Figure 2
(*a*) Estimated mean function using smoothing splines on pooled data across six samples. (*b*) Scree plot depicting the number of eigenfunctions against % of variation explained. (*c*) Two estimated eigenfunctions displaying modes of variation.

the predicted functional relationship between crystallite area and dopant concentrations for each sample, based on the observed sparse data obtained from measurements. The predicted trajectories, which are mostly linear, further corroborate the fact that only two eigenfunctions explain most of the variability of the samples.

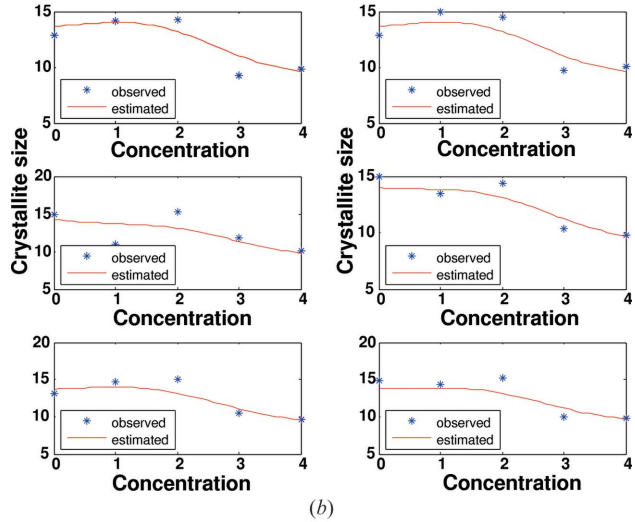
Recall that in order to predict the functions for each sample from the expression in equation (8), in addition to the mean function and eigenfunctions, the principal scores are required; in this case since $k = 1, 2$, corresponding to two eigenfunctions, we estimate A_{i1} and A_{i2} for each sample with trials $i = 1, \ldots, 6$. These estimates are reported in Table 2. For example, the predicted function of crystallite area for sample 1 at a dopant concentration of 2.7% would be computed using equation (8) as

$$\hat{\theta}_1(2.7) = \hat{\mu}(2.7) + 0.4417\hat{\varphi}_1(2.7) - 0.0102\hat{\varphi}_2(2.7), \qquad (15)$$

where $\hat{\mu}(2.7)$ and $\varphi_i(2.7)$ for $i = 1, 2$ are the estimated mean and eigenfunctions at 2.7% concentration, plotted in Figs. 2(*a*)



(*a*)



(*b*)

Figure 3
(*a*) Estimated correlation surface plot of crystallite area. (*b*) Predicted functions for each of the six samples across a fine grid of dopant concentration.

Table 2
The two estimated principal scores corresponding to the eigenfunctions for each sample.

Sample	$\hat{A}_1$	$\hat{A}_2$
1	0.4417	−0.0102
2	0.4641	0.0067
3	−0.1102	−0.1725
4	0.1226	−0.0968
5	0.3406	−0.0009
6	0.2738	−0.0931

and 2(*c*). This is the clear advantage in being able to predict functions from discrete measurements at only a few dopant concentrations – we are able to explore the parameter's behaviour even in the region of unobserved concentrations.

A comparison of the mean functions at the observed dopant concentrations by FPCA and point-wise methods is shown in Fig. 4. The discrepancy at 3% concentration is striking; the FPCA mean is constrained by the smoothness conditions imposed in the penalty term in the optimization problem in equation (6), and hence discontinuities are discouraged.

4.2. UV–visible spectra

The UV–visible absorption spectroscopy carried out for these samples showed that the absorbance of these composites has increased with the concentration of $NiCl_2$ doping. Such an increase in absorbance indicates that there is a variation in the band gap of the material, and thus we have derived the optical band gap for these samples using the Tauc plot method by considering indirect transitions and deriving absorption coefficients (Urs *et al.*, 2014). This plot of $(\alpha h\nu)^{1/2}$ against $h\nu$ is shown in Fig. 5. The optical band gap has decreased with the dopant concentration, which in turn increases the conductivity of these samples. The conductivity results obtained are in accordance with these results and show an increase in conductivity of the samples. The obtained optical band gap values are given in Table 1.

4.2.1. Optical band gap. The optical band gap values are derived from the six individual measurements of UV–vis absorbance recorded for each sample (Figs. 6 and 7). Fig. 6(*a*) shows the average values of these band gaps, which decrease with concentration. These polymer composites exhibit a linear increase in absorbance of UV–vis wavelength with concentration, as a result of excitation of molecules. This indicates that the $NiCl_2$ dopant decreases the band gap of the polymer composites, making them more semi-conductive. The nickel ions are embedded in a polymer matrix of dielectric constant ϵ and hence act as several Bohr atoms in the dielectric medium, which results in different energy levels for free electrons in Ni. This enhances the interaction with the polymer network, thereby reducing the optical band gap with increasing concentration.

4.3. Conductivity

It is noteworthy that the addition of $NiCl_2$ to the HPMC polymer has increased the conductivity of the samples by up to several orders (six), and this indicates that the metal ions in the network actively participate in electrical conduction. This is feasible and supported by the amorphous nature of the polymer, which increases with the dopant concentration. That is, at the higher concentrations of $NiCl_2$, both the number of free metal ions and the amorphousness of the network make the polymer more conductive. Thus $NiCl_2$ proves to be a good metal ion dopant for a polymer that is to be tailored as conductive. The conductivity results as a function of concentrations at 1 and 10 kHz are presented in this section.

Fig. 8 shows the averaged variation of the conductivity of the polymer composites at 1 kHz with concentration. It is seen that the conductivity of these samples increases with the increase in concentration. The free metal ions contribute to the ionic conductivity of the matrix, which is supported by the

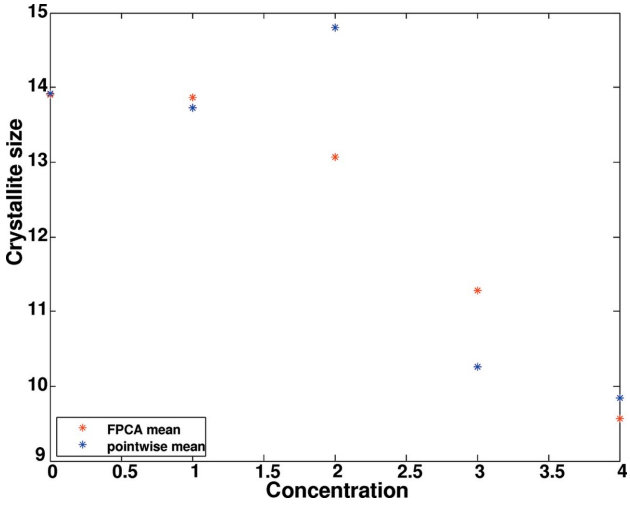


Figure 4
The point-wise mean taken across samples at the five observed concentrations is plotted in blue; the estimated mean function from the FPCA procedure at the five observed concentrations is plotted in red.

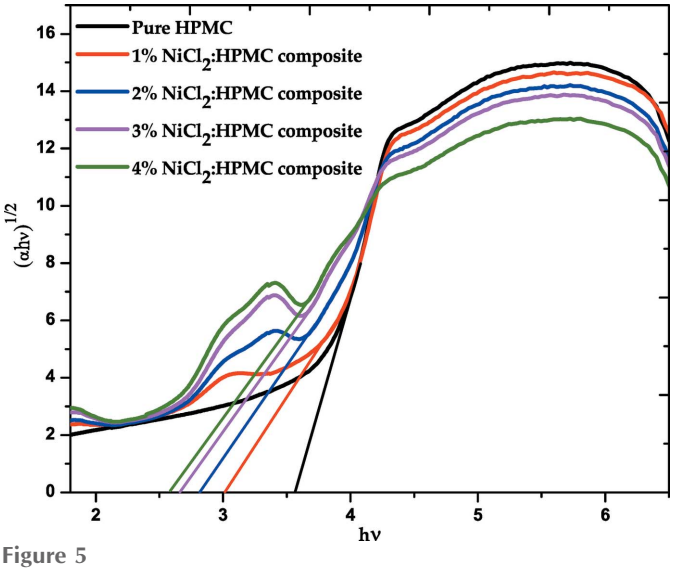


Figure 5
Tauc plot derived for the composites using UV–visible spectra.

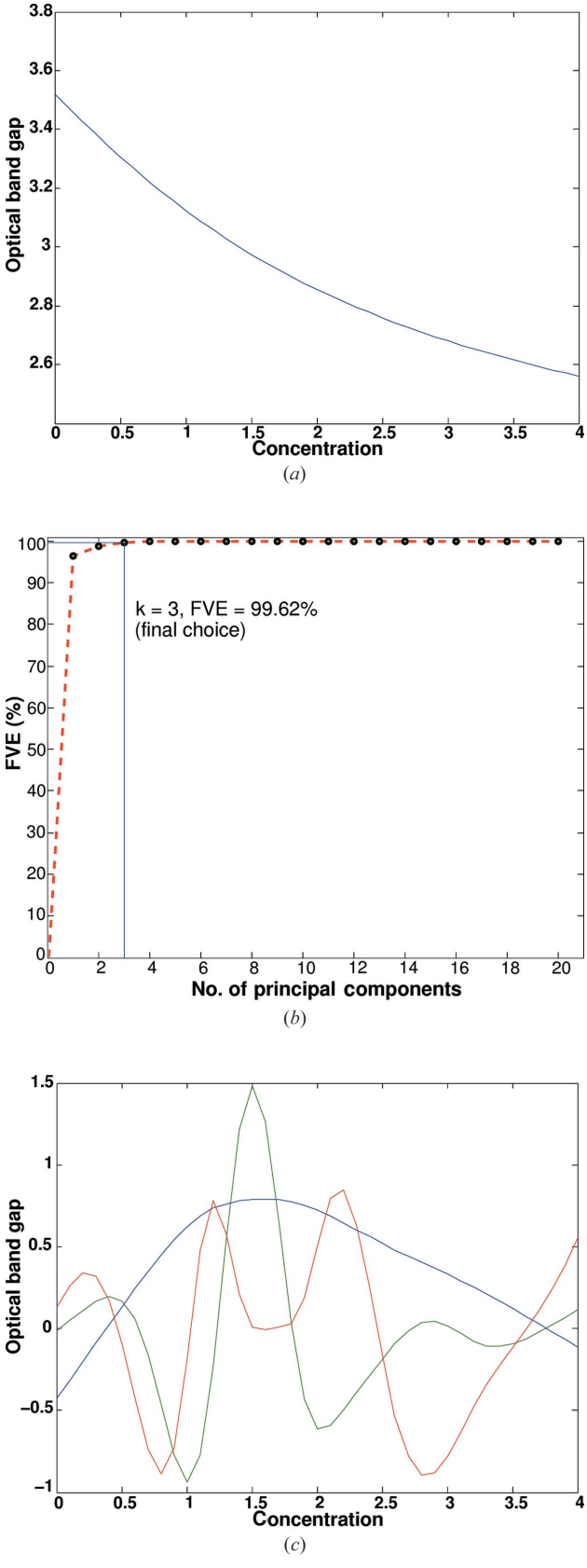


Figure 6
(*a*) Estimated mean function using smoothing splines on pooled data across six samples. (*b*) Scree plot depicting the number of eigenfunctions against % of variation explained. (*c*) Three estimated eigenfunctions displaying modes of variation.

amorphous nature of the polymer. The interstial $NiCl_2$ molecules in the matrix exhibit a hopping mechanism for conduction, and hence there is a linear increase in conductivity with concentration.

It is seen from Fig. 9(*b*) that all six samples exhibit an almost linear relationship with the concentration of $NiCl_2$ in the polymer network. This is because Ni^+ ions are embedded in the polymer matrix with the dielectric constant of the polymer matrix, resulting in an increase in the conductivity.

4.4. Tensile strength

The interstial $NiCl_2$ breaks the polymer network and leads to the weakening of chemical bonds. This results in a decrease in the mechanical strength of these polymer composites, and thus the test of their tensile strength (Figs. 10 and 11) shows a decrease with the dopant concentration. Fig. 11(*a*) shows such a decreasing trend of tensile strength and Fig. 11(*b*) plots the eigenfunctions, which explain the main modes of variation.

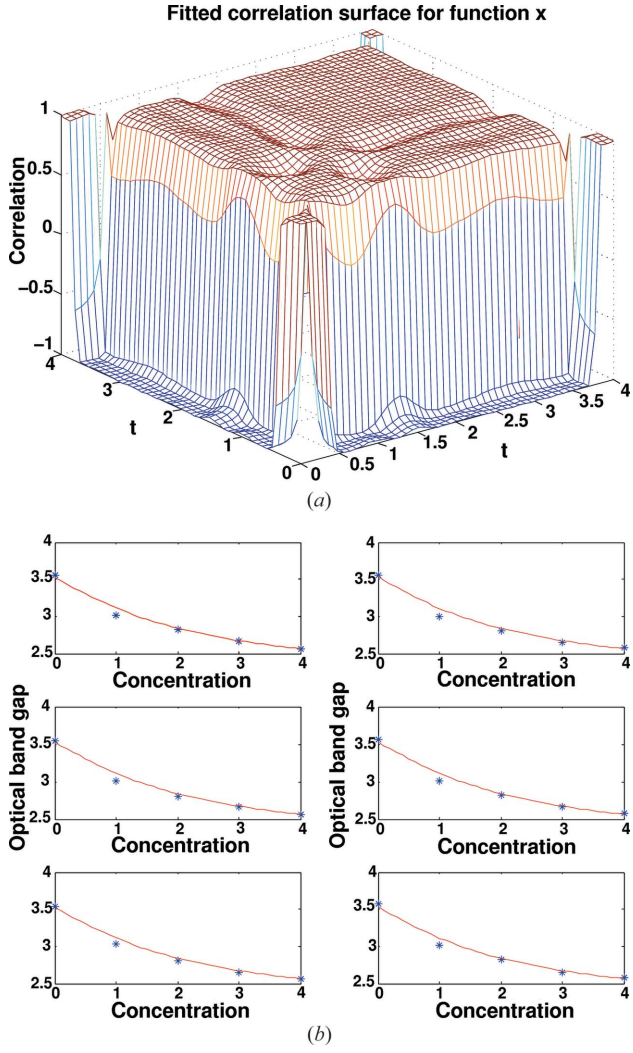
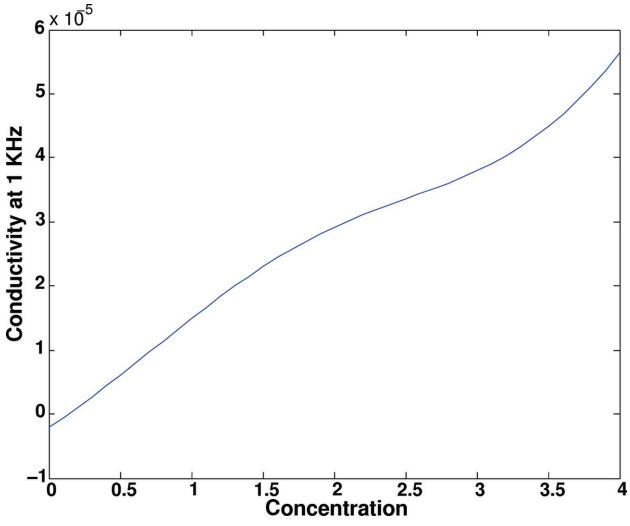
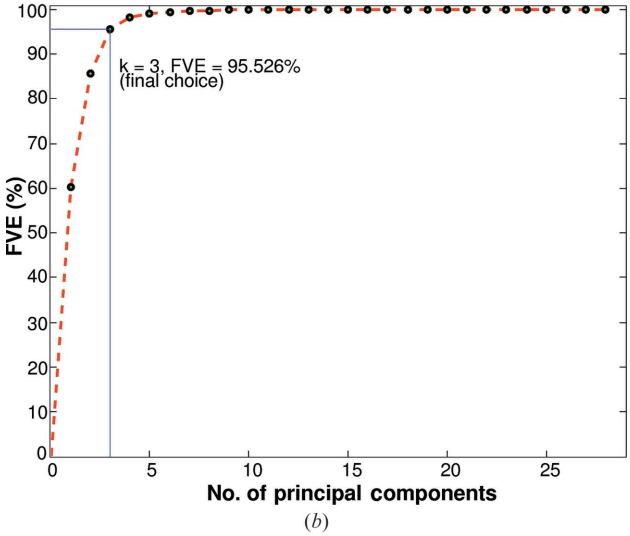


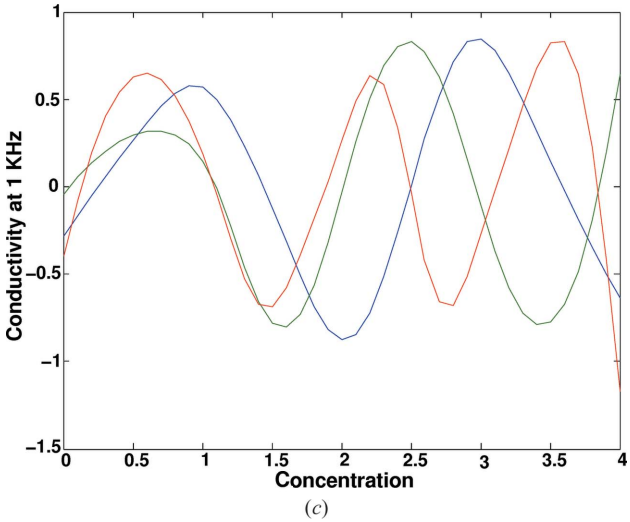
Figure 7
(*a*) Estimated correlation surface plot of optical band gap. (*b*) Predicted functions for each of the six samples across a fine grid of dopant concentration.



(*a*)



(*b*)

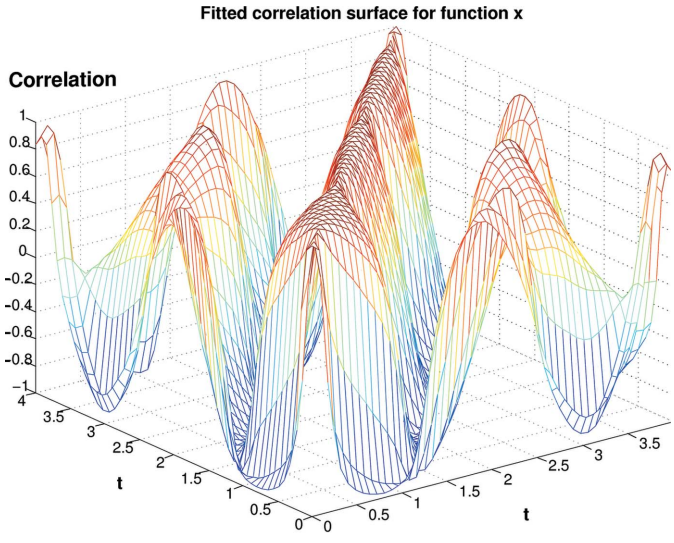


(*c*)

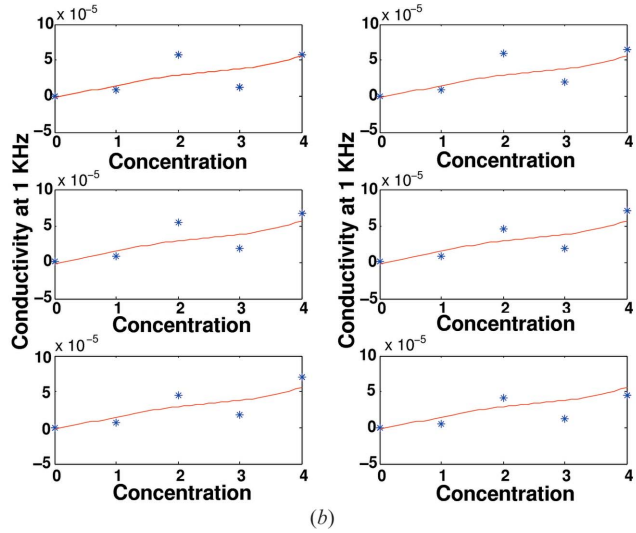
Figure 8
(*a*) Estimated mean function using smoothing splines on pooled data across six samples. (*b*) Scree plot depicting the number of eigenfunctions against % of variation explained. (*c*) Three estimated eigenfunctions displaying modes of variation.

5. Relationship between structural parameters

Having obtained mean estimates of the behaviour of each structural parameter with respect to dopant concentration, we are in a position to examine how the parameters vary with respect to each other, again as a function of dopant concentration. It has typically been assumed, at least for data analysis, that the structural parameters, both micro and macro, behave independently of each other. However, the FDA techniques used in this paper, through the consistent estimation of the mean function, caution us against such an assumption. The plots in Fig. 12 illustrate the pairwise relationships between the parameters. The surface plot in Fig. 13 between crystallite area, optical band gap and conductivity at 1 kHz, at identical dopant concentrations, uncovers some interesting properties. For instance, the plot of mean crystallite area against mean optical band gap, as functions of dopant concentration, in

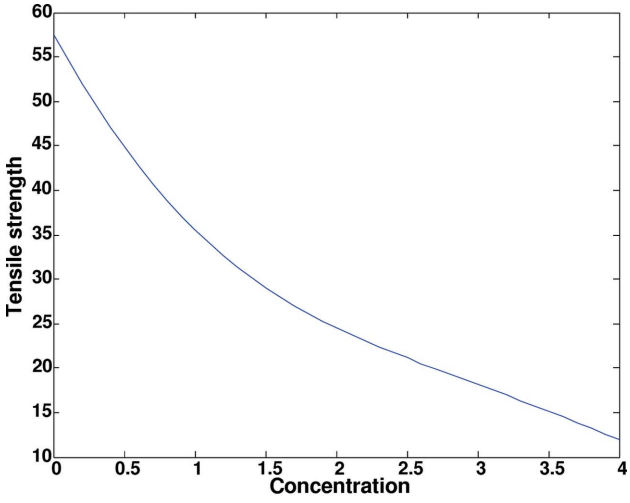


(*a*)

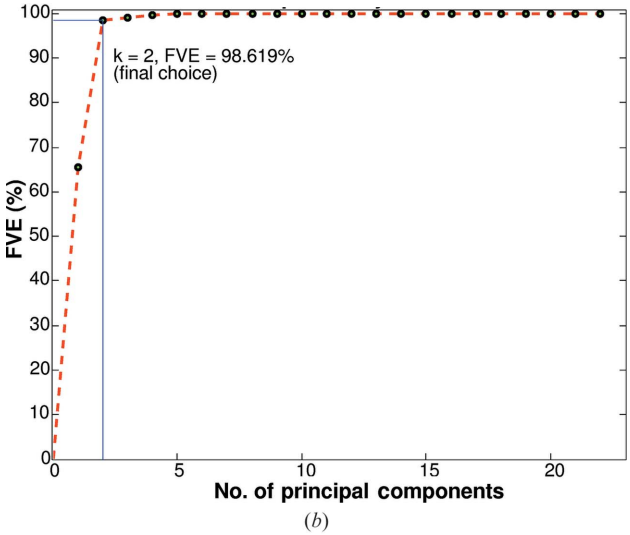


(*b*)

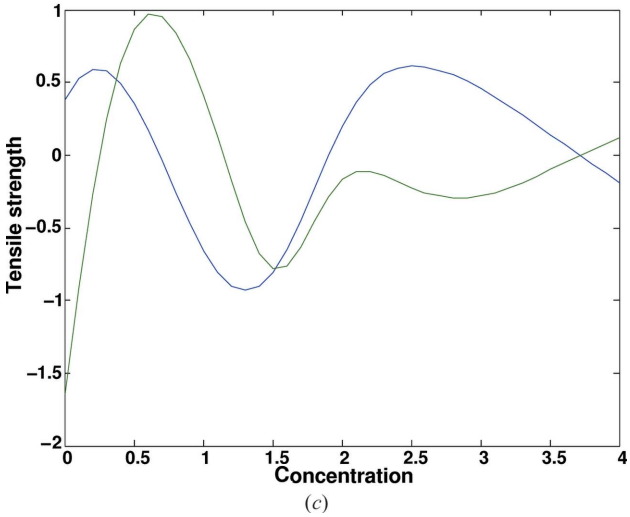
Figure 9
(*a*) Estimated correlation surface plot of conductivity at 1 kHz. (*b*) Predicted functions for each of the six samples across a fine grid of dopant concentration.



(*a*)



(*b*)



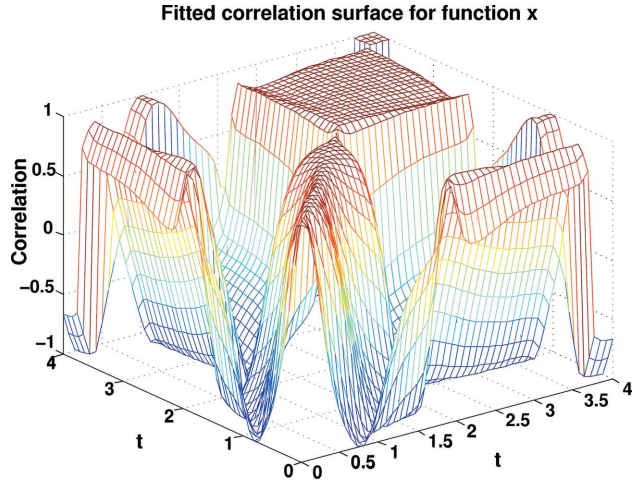
(*c*)

Figure 10
(*a*) Estimated mean function using smoothing splines on pooled data across six samples. (*b*) Scree plot depicting the number of eigenfunctions against % of variation explained. (*c*) Two estimated eigenfunctions displaying modes of variation.

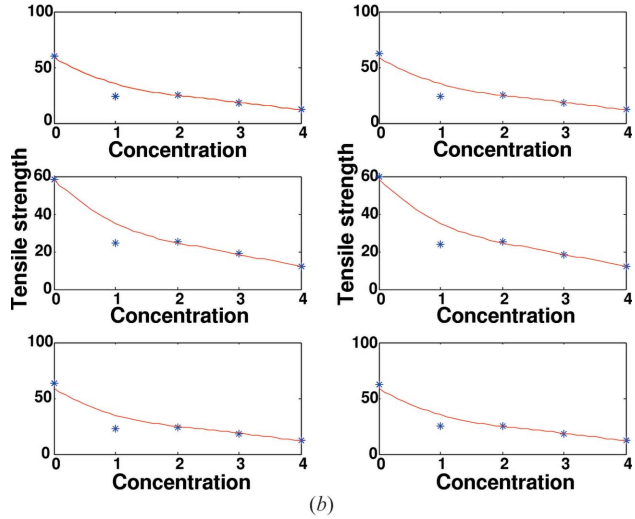
Fig. 12 indicates that their relationship is an increasing function, for a fixed value of conductivity at 1 kHz; however, we can observe from the surface plot that this is not true for all dopant concentrations, when conductivity is not held fixed. This informs us that the parameters perhaps ought not to be treated as independent of one another. While such a behaviour could be attributed to sampling variability, the consistency of the FPCA procedure and the mean function obtained as a solution to the optimization problem in equation (6) imply that the observed behaviour in the surface plot will not change drastically for other samples.

6. Conclusions

We have proposed some novel methods from functional data analysis for analysing various analytical results obtained for polymer composite data, which can be extended further to any material system. The main aim behind this work is to develop a data-driven approach to investigating the relationships



(*a*)



(*b*)

Figure 11
(*a*) Estimated correlation surface plot of tensile strength. (*b*) Predicted functions for each of the six samples across a fine grid of dopant concentration.

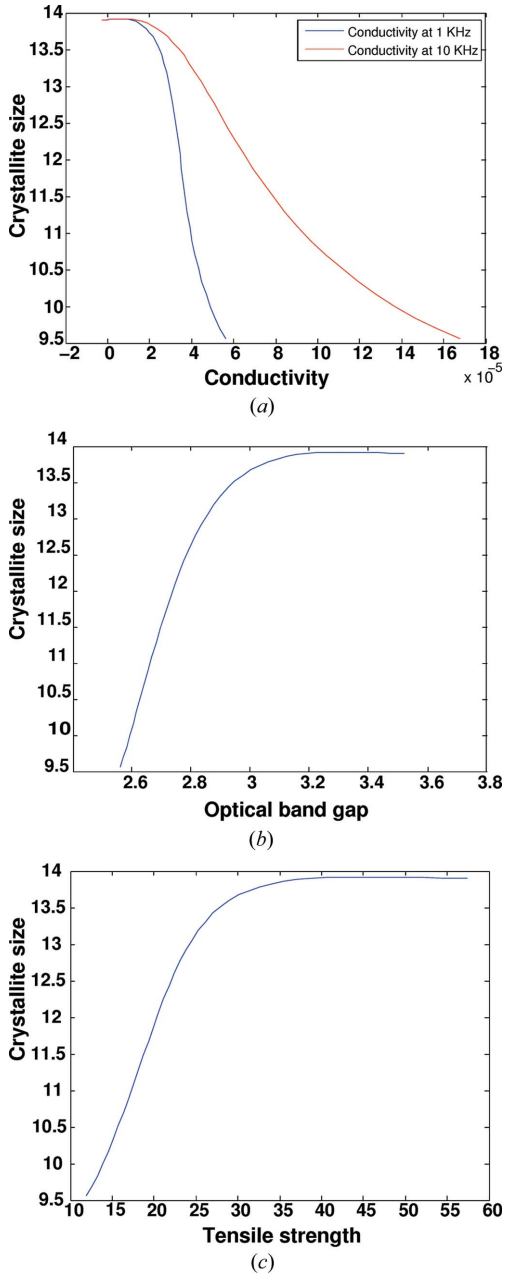


Figure 12
Plots of estimated mean functions of the parameters against each other.

between the structural parameters of a material and their influence on other physical properties. From this perspective, the statistical nature of the article is unavoidable and in fact offers good insight into the complex dependencies between the structural parameters and their effect on the physical properties of a material, as a function of the dopant concentration. Our framework allows for the study of functional relationships manifesting in an X-ray scattering pattern in a unified manner under the FDA framework, wherein discretization effects are accounted for.

We have used the WPPF procedure to compute microcrystalline parameters such as crystallite area and strain for polymer composites for various percentage concentrations of

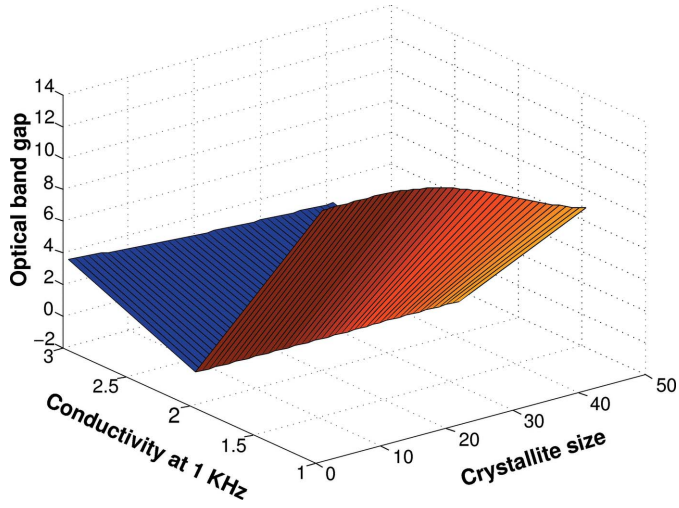


Figure 13
Surface plot of the relationship, at identical dopant concentrations, between crystallite size, optical band gap and tensile strength.

$NiCl_2$. Six trials in each sample were used to obtain measurements which were analysed using techniques from functional data analysis, leading to the estimation of the functional relationship between the parameters and the percentage concentration of $NiCl_2$, in conjunction with details of their global and local characteristics through a suitable basis representation. From this a maximum crystallite area of 15.04 $Å^2$ corresponding to the percentage of $NiCl_2$ of 1.4% in the polymer composites was estimated. Similar analyses were carried out for the other physical parameters, such as conductivity, tensile strength and energy band gap, and maximum values and corresponding percentage concentrations were determined: a maximum energy gap of 3.55 eV is attained at a concentration of 0.0% of $NiCl_2$; a maximum tensile strength of 61 MPa is attained at a concentration of 0.0% $NiCl_2$; and a maximum conductivity of 6.25×10^{-5} S cm^{-1} is attained at a concentration of 2.5% $NiCl_2$. These estimates are useful for designing materials for a particular application in industry. The results of our analyses caution against viewing the different parameters as independent quantities. The proposed FDA framework can be used to investigate the relationships between the structural parameters in a functional regression setting; this will be pursued in future work.

Acknowledgements

The authors thank UGC, New Delhi, India, for projects UPE and CPEPA awarded to the University of Mysore.

References

Ates, M., Karazehir, T. & Sarac, S. (2012). *Curr. Phys. Chem.* **2**, 224–240.
Balzar, D., Audebrand, N., Daymond, M. R., Fitch, A., Hewat, A., Langford, J. I., Le Bail, A., Louër, D., Masson, O., McCowan, C. N.,

Popa, N. C., Stephens, P. W. & Toby, B. H. (2004). *J. Appl. Cryst.* **37**, 911–924.
Hall, I. H. & Somashekar, R. (1991). *J. Appl. Cryst.* **24**, 1051–1059.
Halls, J. J. M., Walsh, C. A., Greenham, N. C., Marseglia, E. A., Friend, R. H., Moratti, S. C. & Holmes, A. B. (1995). *Nature*, **376**, 498–500.
Ida, T. (2011). *J. Appl. Cryst.* **44**, 911–920.
Ida, T. & Izumi, F. (2011). *J. Appl. Cryst.* **44**, 921–927.
Koker, M. K. A., Welzel, U. & Mittemeijer, E. J. (2014). *J. Appl. Cryst.* **47**, 391–401.
Langford, J. I., Louër, D. & Scardi, P. (2000). *J. Appl. Cryst.* **33**, 964–974.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1988). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.
Ramsay, J. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer.
Scardi, P. & Leoni, M. (2006). *J. Appl. Cryst.* **39**, 24–31.
Shahmoradi, B., Ibrahim, I. A., Sakamoto, N., Ananda, S. & Somashekar, R. (2010). *J. Environ. Sci. Health. Part A*, **45**, 1248–1255.
Somashekar, R., Hall, I. H. & Carr, P. D. (1989). *J. Appl. Cryst.* **22**, 363–371.
Somashekar, R. & Somashekarappa, H. (1997). *J. Appl. Cryst.* **30**, 147–152.
Somashekarappa, H., Annadurai, V., Sangappa, Y., Subramanya, G. & Somashekar, R. (2002). *Mater. Lett.* **53**, 415–420.
Stokes, A. R. (1948). *Proc. Phys. Soc.* **61**, 382–391.
Warren, B. E. & Averbach, B. L. (1952). *J. Appl. Phys.* **23**, 497.
Yao, F., Müller, H. G. & Wang, J. L. (2005). *J. Am. Stat. Assoc.* **100**, 577–590.
Urs, G. T., Hurkadli, R. V., Basvaraj, R. V., Niranjana, M., Manjunath, A. & Somashekar, R. (2014). *Progr. Cryst. Growth Charact. Mater.* **60**, 87–93.
Zilahi, Gy., Ungár, T. & Tichy, G. (2015). *J. Appl. Cryst.* **48**, 418–430.