Measure of polymer performance based on correlated physical parameters

*G K Gowtham H Somashekarrappa Karthik Bharath R Somashekar**

G K Gowtham
Department of Physics, Yuvaraja's College, University of Mysore,Mysore,India; present address: Department of Physics, Davangere University, Davangere, Karnataka
Email: gowthamgk@live.com
H Somashekarappa
Department of Physics, Yuvaraja's College, University of Mysore,Mysore,India
Email: drhssappa@gmail.com
Karthik Bharath
School of Mathematical Sciences, University of Nottingham, Nottingham, UK
Email: Karthik.Bharath@nottingham.ac.uk
R Somashekar
Department of Materials Science, University of Mysore and RIE, Mysore, India
Email: rs@physics.uni-mysore.ac.in

We introduce a novel measure of performance of polymer composites based on physical parameters whose behaviour depends on levels of dopant concentration used during their preparation. The performance measure is based on a joint analysis of measurements of the physical parameters exhibiting non-trivial correlations that vary across different levels of dopant concentrations. In contrast to traditional multivariate analysis, we treat data from parameter measurements as being obtained from functional parameters, and develop the performance measure based on the joint average function of parameters that encodes the correlation structure. An optimal level of dopant concentration is then ascertained with respect to the performance measure. While the proposed measure is general enough to be applicable to any chosen physical parameters, we demonstrate its utility in the context of assessing performance using microstructural and nonlinear optical parameters. Computing of the measure and optimal dopant concentration are carried out using Monte Carlo sampling, which further facilitates uncertainty quantification.

1 Introduction

There is continued interest in the study of polymers and polymer composites [1]owing to their prominent role in several industrial applications [2], including high intensity lasers, optical computing and photonic devices [3,4]. Typically the study of polymer composites entails detailed experimental analysis of data pertaining to physical parameters that characterise various properties of the polymer (e.g. size, energy gap). Data on the physical parameters is obtained by using doping agents at different concentration levels (e.g. Ammonium diHydrogen Phosphate (ADP)) during polymer preparation. Multivariate analysis based on principal component analysis and clustering on physical parameters data is quite common [5,6]. However, a drawback of this approach is that effect of dopant levels on correlations between parameters is not captured and quantified since dopant information is not explicitly included as part of the data analysis.
Lack of dopant level information affects the wider applicability of polymers, since it is crucial for their affordable use to not only understand their properties through study of the physical parameters but also have knowledge on a narrow range of optimal dopant levels during their preparation. What is required is an objective assessment of performance of a polymer in any given application: how can one quantify the performance of a polymer with respect to the measured physical parameters at various dopant concentration levels? This necessitates undertaking two tasks:

1. understanding dependencies and correlations between structural and non-structural parameters [7] using experimental data;
2. understanding behaviour of these parameters as functions of dopant concentration levels over a range of possible values, including the levels used in the making of polymer composites.

The key challenge in carrying out the tasks is that the experimental parameters exhibit nonlinear correlations, behaviour and magnitude of which vary by level of dopant concentrations [8]. An additional difficulty is that data on the parameters are obtained only at a handful of dopant concentration values from a possible range of values, determined by laboratory conditions and experimental costs. Knowledge of optimal concentration level of the dopant for a given application, at least within a certain range, will enable experimenters to make judicious use of resources while setting up and carrying out experiments with polymer composites.

In this article we propose a novel measure of performance of a polymer composite that enables assessment at any dopant concentration based on data at a small number of concentration levels, and can be profitably used to determine an optimal dopant concentration level. The measure satisfies two complementary desiderata associated with the above mentioned tasks. Construction of the measure is based on viewing data values of parameters at a small number of dopant concentration levels as discrete realisations of a vector-valued function over an interval of concentration values. This perspective enables us to smoothly interpolate between dopant values in a manner that borrows strength between correlated parameter values at neighbouring concentrations, while simultaneously borrowing information across several experimental trials for the parameters.

Specifically, we shall use Multivariate Functional Principal Component Analysis (MFPCA) [9] to estimate a vector-valued function that represents average behaviour of the correlated parameters as a function of dopant concentration levels. This extends our efforts in [10] wherein each parameter was analysed independently of the others at the cost of ignoring correlations between parameters, which is a crucial requirement in the analysis of structural and non-structural parameters [11].

We emphasize that the proposed performance measure is general and can be employed on any set of parameters observed at a handful of dopant concentration levels; this implies that the performance measure can be used to asses *any set of parameters irrespective of their relative importance*. In this article, however, we will focus on its construction and use on data obtained from measurements of structural and non-structural parameters of polymer composites prepared as PVA (Polyvinyl Alcohol) and PVP (Polyvinylpyrrolidone) blended films doped with nonlinear organic salt KDP (Potassium diHydrogen Phosphate) at various concentration levels. In particular, we will consider structural parameters size and energy gap obtained from X-ray diffraction (XRD) and UV-vis patterns respectively, and non-structural parameters characterising nonlinear optical properties of the polymer composite obtain from a Z-scan experiment [12].

Definition of the performance measure as a function of dopant concentration will enable determination of the optimal dopant concentration level at which maximum performance of the polymer can be obtained. To the best of our knowledge, this is the first work in literature to simultaneously capture correlations of parameters across dopant concentration values, and obtain an optimal dopant concentration level within a single unified framework. Summarily our main contributions are:

1. Definition of a functional measure of performance of a polymer composite that accounts for correlations between measured physical parameters as functions of dopant concentration levels;
2. development of stochastic Monte Carlo-baseds method to compute proposed functional measure of performance and optimal dopant concentrations, which further engender uncertainty quantification.

2 Microstructural and nonlinear optical parameters

The parameters in our study were obtained from a polymer composite prepared using a solution casting method based on polyvinyl alcohol (PVA) and polyvinylpyrrolidne (PVP) blend polymer solutions doped with 0.2g, 0.5g, 0.8g, 1g and 1.5g of Potassium dihydrogen phosphate (KDP) (in addition to an undoped blend), resulting in 6 solutions. From each solution measurements of the two microstructural parameters from an X-ray diffraction pattern characterising crystallite size S and optical band gap (energy gap) E_g were recorded, and four nonlinear optical parameters from aperture curves arising from a

Z-scan experiment characterising nonlinear absorption coefficient β and its third order nonlinear susceptibility parameter Im_χ, nonlinear refractive index ν and the corresponding third-order suspectibility Re_χ.
Evidently then the parameter vector $\boldsymbol{\theta} = (\beta, E_g, Im_\chi, Re_\chi, \nu, S)^T \in \mathbb{R}^6$ is measured at five different dopant concentration levels in the interval $D = [0.2, 1.5]$. Preliminary multivariate analysis disregarding dopant information reveals significant correlations between the parameters as shown in Figure 1. In

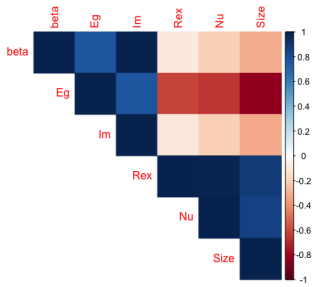

Figure 1: Raw correlations between the parameters.

contrast to standard multivariate analysis, we view the recorded measurements as discrete realisations of a vector-valued function $D \ni t \mapsto \boldsymbol{\theta}(t)$ with $\boldsymbol{\theta}(t) = (\beta(t), E_g(t), Im_\chi(t), Re_\chi(t), \nu(t), S(t))^T \in \mathbb{R}^6$. This allows us to simultaneously smoothly interpolate between dopant values and capture correlations between the parameters.

3 Sample preparation and data generation

The requisite samples were prepared by solution casting method, where we dissolved 5g of polyvinyl alcohol (PVA) in 100ml water, preheated at 70°C for 30min. This solution was stirred continuously for a day. Another solution was simultaneously prepared where 2g of Polyvinyl Pyrrolidone (PVP) was dissolved in 100ml water at room temperature and stirred for a day. These solutions were mixed in a ratio of 70:30, ie., 70% of PVA solution and 30% of PVP solution, and were then stirred for a day to prepare the blend solution. Further, 0.2g, 0.5g, 0.8g, 1g and 1.5g of Potassium dihydrogen phosphate (KDP) was doped into the prepared blend solution along with an undoped blend resulting in a set of 6 solutions, which were then stirred for 3hrs and cast onto a casting plate. The casted films were then placed inside the hot air oven for 8hrs and was maintained at 70°C. The dried films were then safely placed inside a zip lock cover for further analysis. Beyond 1.5%, films prepared were brittle in nature and could not be used. X-ray diffraction pattern to study the microstructure property was recorded using Rigaku smartlab with Ni filtered CuKα radiation of wavelength 1.5406 A° with usual standard settings.The FTIR spectra was recorded in transmission mode with spectrophotometer model FTIR-4100 type A, having a resolution of $4cm^{-1}$ in the wavenumber range $600 - 4000cm^{-1}$. Ultra violet–visible spectroscopy of the samples were measured using a Shimadzo UV-1800 UV-Vis Spectrophotometer for a wavelength range from 200nm to 1100nm. The nonlinear optical properties of the polymer composites were studied using Z-scan technique [13]. The experiment was performed at an input peak-intensity of $2 \times 10^9 W/cm^2$ with a Q-switched Nd:YAG laser having 9ns pulse width at 1064nm with a repetition rate of 10Hz. The laser beam waist at the focused spot was estimated to be 31.4μm and the corresponding Rayleigh length is 2.9 mm. The measurements were carried out on polymer thin films of 0.17mm thickness, which is less than the Rayleigh length. Hence, the thin sample approximation is valid.

4 Data on parameters and preliminary analysis

The XRD pattern of the polymer composites with different dopant concentration is shown in the left panel of Figure 2(a). The amorphous nature of the polymer composites can be attributed to the result-

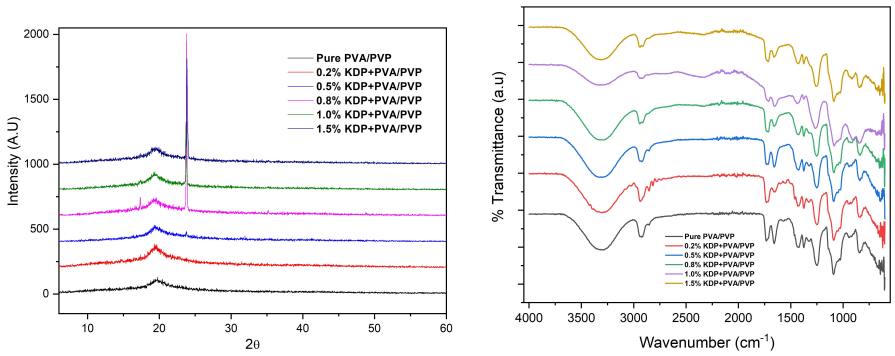

Figure 2: Left: XRD pattern of polymer composites. Right: FTIR Spectra of the polymer composites.

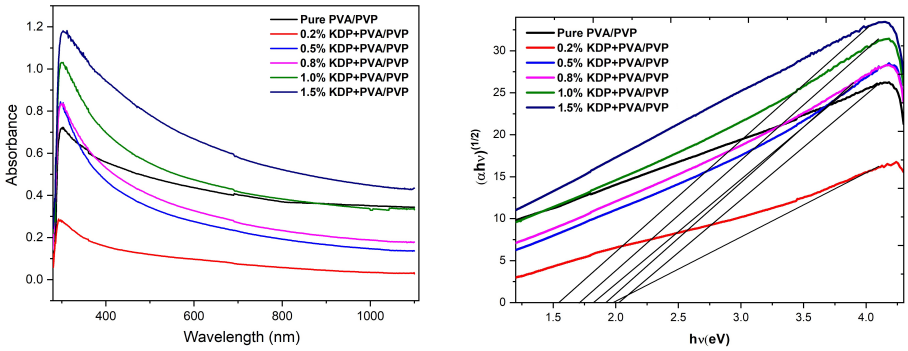

Figure 3: Left:UV-Vis absorption spectra of the polymer composites. Right: Optical Band Gap of Polymer composites.

ing broad XRD peak at $2\theta = 20°$. On increasing the dopant concentration a sharp Bragg peak is observed at $2\theta = 24°$ which is due to added KDP indicating an increase in crystalline region. Microstructural parameters such as crystallite size and strain for the composites were calculated using Williamson Hall (WH) plot [14] using Full Width at Half Maximum [FWHM] and are shown in Table 1.The FT-IR spectra obtained for the polymer composites is shown in right panel of Figure 2(b). The prominent absorption observed at 3100-300 cm^{-1} and 2950-2800 cm^{-1} wavelength corresponds to =C-H stretch and –C-H stretch respectively do change the shape with increase in the concentration of the KDP crystals. FTIR results indicate structural changes in the PVA/PVP polymer network with the addition of KDP. The UV-Vis linear absorption spectra for the polymer composites is shown in the left panel of Figure 2. The spectra shows an absorption band at 300nm which corresponds to n $\rightarrow \pi^*$ electronic transition.The negligible single photon absorption observed at 1064nm IR region indicates that the nonlinear optical measurements carried out in this experiment are under non-resonant excitation.
The optical bandgap of the composites were calculated using Tauc plot method [15]. The plot of $(\alpha h\nu)^{1/2}$ vs hν shown in the right panel of Figure 3; here $r = \frac{1}{2}$ is considered for indirect transition. The band gap values associated with our polymer composites is presented in Table 1. From these results, band gap of the polymer composites decreaes with increase in dopant concentration.
The open aperture Z-Scan curves obtained for the polymer composites are shown in Figure 4. The curves were obtained by fitting (1) to the data using the method of nonlinear least squares, resulting in measurement values for the two-photon nonlinear absorption coeffiicient $\beta_{eff} \equiv \beta$ and linear absorption coefficient α via the relation $L_{eff} = \frac{(1-e^{\alpha L})}{\alpha}$, where L is sample thickness, Z is sample position and Z_0 is Rayleigh length. Specifically, the normalized transmittance as a function of sample position is given by the equation ([16])

$$T(Z) = \frac{ln(1+q_0(Z))}{q_0(Z)}, \quad q_0(Z) = \frac{I_0\beta_{eff}L_{eff}}{(1+x^2)}, \tag{1}$$

where $|q_0(Z)| < 1$ with $x = Z/Z_0$ and I_0 is peak on-axis irradiance at the focus.

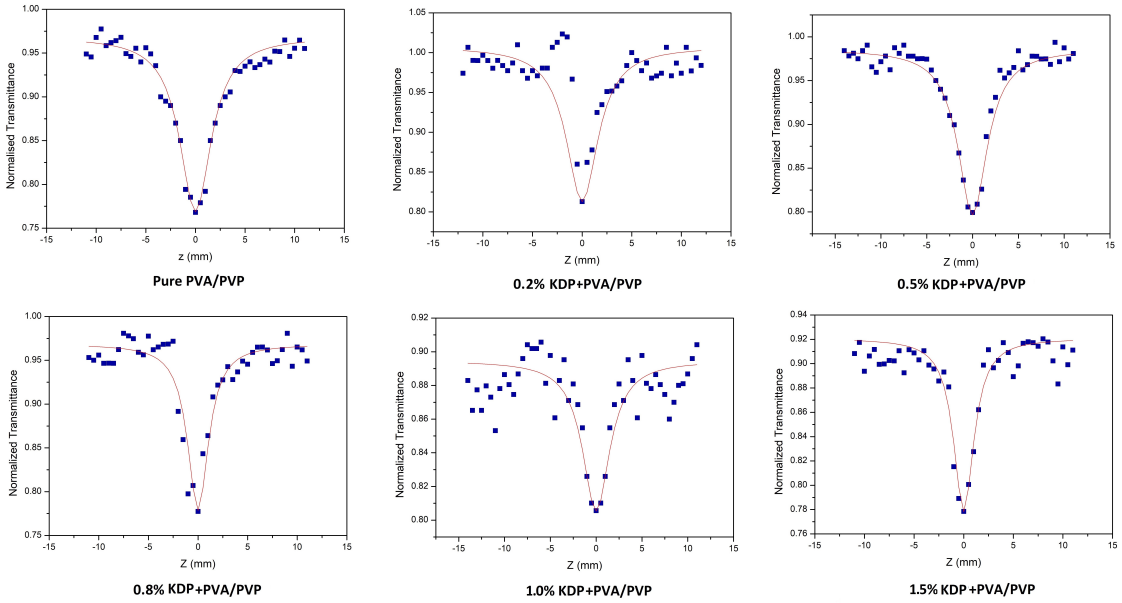

Figure 4: Open aperture curve from Z-Scan technique for pure polymer and polymer composites.

The shape of the open aperture curve indicates that the sample exhibits two photon absorption, where the transmission is symmetric with respect to the focus (Z=0) and minimum. The nonlinear absorption coefficient β is related to the imaginary part of third order nonlinear optical susceptibility through the equation, $Im\chi^{(3)} = n_0^2\varepsilon_0 c\lambda\beta\pi$, where n_0 is linear refractive index, ε_0 is permitivity of free space, c is speed of light in vaccum.

By placing an aperture infront of the detector in open aperture setup the closed aperture experiment is performed. Here again, the curves for the closed aperture Z-scan were obtained by fitting (1) to the data by estimating the relevant parameters in (1) using nonlinear least square and is shown in Figure 5.

We observe that as the sample is moved away from the lens the closed aperture curve shows a self defocusing nonlinearity (-ve refractive index) for the polymer composites, which is a peak followed by a valley in the normalized transmittance. This is due to positive lensing in the sample [17]. When the sample is placed before the focus the positive lensing reduces the farfield divergence allowing for a larger aperture transmittance. In contrast, with the sample placed after the focus the same positive lensing moves the focal position closer to the sample resulting in greater farfield divergence and reduced aperture transmittance.

The normalized transmittance for pure nonlinear refraction is given by

$$T(Z) = 1 - \frac{4x\Delta\phi}{[(x^2+9)(x^2+1)]}, \quad x = Z/Z_0, \tag{2}$$

where $\Delta\phi$ is on axis phase change given by

$$\Delta\phi = \frac{\Delta T_{P-V}}{0.406(1-S)^0.25}, \quad \text{for } |\Delta\phi| \leq \pi;$$

here ΔT_{P-V} is the change in transmittance between peak and valley, S is the linear aperture transmittance. The nonlinear refractive index is given by

$$\nu = \frac{\Delta\phi\lambda}{2\pi L_{eff} I_0}(m^2/W).$$

On the other hand, the nonlinear refractive index n_2(esu) is related to $\nu(m^2/W)$ as $n_2(esu) = \frac{cn_0}{40\pi}\nu(m^2/W)$. Indeed, the real part of third order nonlinear susceptibility is given by $Re\chi^{(3)} = 2n_0^2\varepsilon_0 cn_2(esu)$. The parameters derived from both open aperture and closed aperture are tabulated in Table 1. We could not observe NLO activity in the closed aperture for pure polymer material.

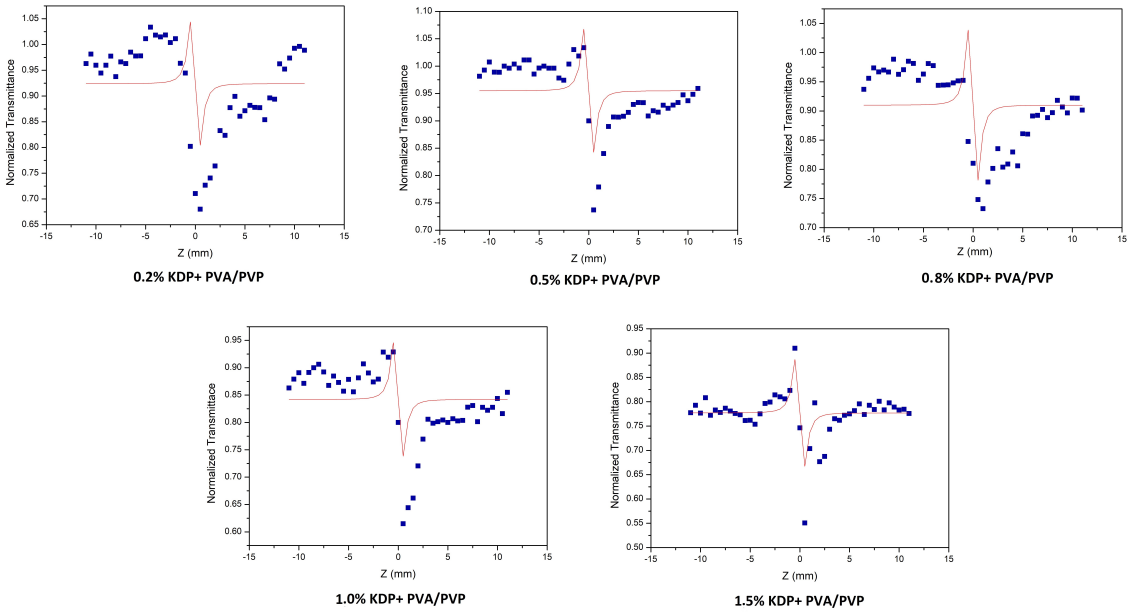

Figure 5: Closed aperture trace for doped polymer composites.

Sample	**crystallite size**($\AA$)	**Band Gap** E_g**(eV)**	β $\times 10^{-7}\frac{m}{W}$	**Im**χ $\times 10^{-15}$	$\nabla\phi$ $\nabla\phi$	$\mathbf{n^2}$ $\times 10^{-8}(esu)$	**Re**χ $\times 10^{-9}$	ν $\times 10^{-16}(\frac{m^2}{W})$
pure	44.56;41.97	2.09;1.96;	1.13;1.06	1.72;1.62;	-	-	-	-
	45.43;41.10	2.13;1.92	1.15;1.04	1.75;1.58;	-	-	-	-
	46.73;43.27	2.19;2.03;	1.18;1.10;	1.80;1,67	-	-	-	-
0.2%	42.65; 40.16	2.09; 1.91	1.04; 0.97	1.58; 1.49	0.88; 0.83	4.08; 3.80	7.35; 6.92	2.94; 2.77
	43.48; 39.33	2.06; 1.87	1.06; 0.95	1.61; 1.46	0.90; 0.81	4.16; 3.77	7.49; 6.73	3.00; 2.71
	44.72; 41.41	2.12; 1.97	1.09; 1.01	1.66; 1.54	0.92; 0.86	4.28; 3.97	7.71; 7.14	3.08; 2.86
0.5%	47.58; 44.81	1.97; 1.86	1.03; 0.97	1.56; 1.47	0.80; 0.75	3.70; 3.49	6.87; 6.46	2.66; 2.51
	48.51; 43.89	2.01; 1.82	1.05; 0.95	1.59; 1.44	0.82; 0.74	3.78; 3.42	7.00; 6.33	2.71; 2.46
	49.89; 46.20	2.07; 1.92	1.08; 1.00	1.64; 1.52	0.84; 0.78	3.88; 3.60	7.20; 6.67	2.79; 2.59
0.8%	60.66; 57.13	1.87; 1.76	1.06; 0.99	1.60; 1.51	0.96; 0.91	4.45; 4.20	8.01; 7.54	3.21; 3.02
	61.84; 55.95	1.91; 1.73	1.08; 0.97	1.63; 1.48	0.98; 0.89	4.54; 4.11	8.16; 7.39	3.27; 2.96
	63.61; 58.90	1.96; 1.82	1.11; 1.03	1.68; 1.56	1.01; 0.94	4.67; 4.33	8.40; 7.78	3.36; 3.12
1.0%	46.93;44.20	1.76; 1.65	0.46; 0.43	0.68; 0.65	0.84; 0.79	3.89; 3.66	6.99; 6.58	2.80; 2.63
	47.84; 43.29	1.79; 1.62	0.47; 0.42	0.71; 0.64	0.86; 0.77	3.96; 3.59	7.12; 6.45	2.85; 2.58
	49.21; 45.57	1.84; 1.71	0.48; 0.45	0.73; 0.68	0.88; 0.82	4.08; 3.78	7.33; 6.79	2.93; 2.72
1.5%	69.39; 65.34	1.58; 1.49	0.78; 0.73	1.18; 1.11	0.96; 0.91	4.45; 4.20	8.01; 7.54	3.21; 3.02
	70.73; 64.00	1.62; 1.43	0.79; 0.72	1.20; 1.09	0.98; 0.89	4.54; 4.11	8.16; 7.39	3.27; 2.96
	72.75; 67.37	1.66; 1.54	0.82; 0.76	1.24; 1.15	1.01; 0.94	4.67; 4.33	8.40; 7.78	3.36; 3.12

Table 1: Data from parameter measurements derived from both open aperture and closed aperture experiments.

At each dopant level $t \in D$ the nonlinear absorption coefficient $\beta(t)$ is related to the imaginary part of third order nonlinear optical susceptibility through the equation,

$$Im_\chi(t) = n_0^2 \varepsilon_0 c \lambda \pi \beta(t),$$

where n_0 is linear refractive index, ε_0 is permitivity of free space, c is speed of light in vaccum. Relatedly, nonlinear refractive index $\nu(t)$ is related to the real part of the third order nonlinear optical suspectibility as

$$Re_\chi(t) = \frac{1}{40\pi} 2 n_0^3 \varepsilon_0 c^2 \nu(t).$$

We thus expect $\beta(t) \propto Im_\chi(t)$ and $\nu(t) \propto Re_\chi$ for each $t \in D$ since n_0, ϵ_0 and c do not change with dopant levels.
On the other hand, raw correlations in Figure 1 indicate the need to study the parameters *jointly* when assessing the performance of a polymer. Moreover, correlations between optical parameters $\beta(t), \nu(t)$ and microstructural parameters $S(t), E_g(t)$ are typically known to be nonlinear and complex [12], which change with dopant levels t. Since polymer preparation and recording measurements is an expensive and time-consuming procedure, quantifying such nonlinear dependence is crucial for an experimentalist wishing to ascertain the 'right' level of dopant to obtain a 'good' polymer composite. Our main goal is to mathematically define and compute a performance measure for a polymer, which then allows one to identify an optimal dopant level.

5 Multivariate functional data analysis for joint analysis of parameters

Quantifying correlations and dependencies between the parameters as functions of the concentration values is not possible if the parameters are treated individually, and not jointly, as done in our earlier work [10] using univariate Functional Principal Component Analysis (FPCA). For joint analysis we consider Multivariate Functional Principal Component Analysis (MFPCA)[18-20], and carry out analysis of the experimental parameters, explicitly taking into account their complex dependence structure.
MFPCA constitutes the key first step towards constructing a performance measure of a polymer based on dependent parameter measurements. We describe the methodology entirely in the context of the data on parameters obtained above. For a general exposition and details on the precise MFPCA technology used in this paper, we refer to [9].

5.1 The multivariate Karhunen-Loéve representation

Data presented in the Supplementary Materials forms the basis for joint statistical analysis of these parameters. Starting from discrete measurements of each parameter at concentrations t =0.2, 0.5,0.8,1,1.5 we obtain the corresponding function $\boldsymbol{\theta}(t)$ by linearly interpolation between the points. We do not employ any smoothing procedures (for e.g. splines) owing to the sparse nature of the measurements (only 5 measurements of the function). Joint analysis of the parameters depends on being able to express $\boldsymbol{\theta}(t)$ in terms of an orthonormal set of basis functions which decompose the observed variation in the data. Such a decomposition is based on the multivariate Karhunen-Loéve representation of a stochastic process, which we now review.
Let $\mathbb{L}^2(D)$ denote the set of real-valued square-integrable random functions with respect to the Lebesgue measure on D. Let the set of such functions $\boldsymbol{\theta}(t), t \in D$, representing the data, reside in a (product) Hilbert space $\mathcal{H}$. Assume that $\boldsymbol{\theta}$ has a continuous mean function $\boldsymbol{\mu}(t) := (\mu_1(t), \ldots, \mu_6(t))^T$ with $\mu_k(t) = E(\theta_k(t)), k = 1, \ldots, 6$ representing the mean of the kth parameter, and covariance function $\boldsymbol{G}(u, v) := \{G_{jk}(u, v); (u, v) \in D \times D; j, k = 1, \ldots, 6\}$ where $G_{jk}(u, v) := Cov(\theta_j(u), \theta_k(v))$ is the cross-covariance functions between parameter j and k. When $j = k$, we get the usual covariance function of parameter j. When $j \neq k$, $G_{jk}(\cdot, \cdot)$ is interpreted as the covariance function between the two parameters θ_j and θ_k.

Note that for every $(u,v) \in D \times D$, $\boldsymbol{G}(u,v)$ is a 6×6 positive definite matrix, and $\boldsymbol{G}(u,v)$ is hence a 6×6 matrix-valued function indexed by $D \times D$. That is, for $(u,v) \in D \times D$,

$$\boldsymbol{G}(u,v) = \begin{pmatrix} G_{1,1}(u,v) & G_{1,2}(u,v) & \dots \\ \vdots & \ddots & \\ G_{6,1}(u,v) & & G_{6,6}(u,v) \end{pmatrix}.$$

In other words, $G(u,v)$ captures the covariance (and hence correlation) between the six parameters at each combination of dopant concentrations in D. When $u = v$, $G(u,u)$ captures covariance between parameters at a single concentration value u.
Evidently, $\boldsymbol{G}(u,v) = \boldsymbol{G}(u,v)^T$ and hence symmetric for every $(u,v) \in D \times D$. The function $\boldsymbol{\mu}(t)$ represents the joint mean function of of the six parameters $\boldsymbol{\theta}(t)$ and the function $\boldsymbol{G}(\cdot,\cdot)$ represents the cross-covariance between the covariance functions $\{G_{j,k}(\cdot,\cdot)\}$ corresponding to the individual $\theta_k, k = 1,\dots,d$. The space $\mathbb{L}^2(D)$ is equipped with the usual inner product $\langle \theta_1, \theta_2 \rangle := \int_D f\theta_1(t)\theta_2(t)dt$ with norm $\|f\| := \langle \theta_1, \theta_2 \rangle^{1/2}$. From this an inner product between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ on $\mathcal{H}$ can be defined as $\langle \boldsymbol{\theta}, \boldsymbol{\eta} \rangle_{\mathcal{H}} := \sum_{k=1}^{6} \langle \theta_k, \eta_k \rangle$ with norm $\|\boldsymbol{\theta}\|_{\mathcal{H}} := \langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle_{\mathcal{H}}^{1/2}$. The multivariate KL representation requires an appropriate integral covariance operator $\mathbb{G} : \mathcal{H} \to \mathcal{H}$ based on the covariance kernel $\boldsymbol{G}(u,v)$. Note that $\boldsymbol{G}$ can represented as $\boldsymbol{G}(u,v) = (\boldsymbol{G}_1(u,v), \dots, \boldsymbol{G}_6(u,v))^T$, where we can view $\boldsymbol{G}_j(u,v) = (G_{j1}, \dots, G_{j6})^T, j = 1,\dots,6$ as the jth row of the 6×6 matrix $\boldsymbol{G}(u,v)$ for every $(u,v) \in D \times D$. Observe that for $\boldsymbol{\eta}(t) = (\eta_1(t), \dots, \eta_6(t)^T$, $\langle \boldsymbol{G}_j(u,\cdot), \boldsymbol{\eta} \rangle_{\mathcal{H}} = \sum_{k=1}^{6} \langle G_{jk}(u,\cdot), \eta_k \rangle$, and we can hence define the covariance operator $\mathbb{G}$ for a given $\boldsymbol{\eta} \in \mathcal{H}$ as

$$\mathbb{G}\boldsymbol{\eta}(u) := \int \boldsymbol{G}(u,v)\boldsymbol{\eta}(v)dv = \begin{pmatrix} \langle \boldsymbol{G}_1(u,\cdot), \boldsymbol{\eta} \rangle_{\mathcal{H}} \\ \vdots \\ \langle \boldsymbol{G}_6(u,\cdot), \boldsymbol{\eta} \rangle_{\mathcal{H}} \end{pmatrix}.$$

The covariance $\boldsymbol{G}$ is symmetric and non-negative definite, and hence by the multivariate version of Mercer's theorem [21], the following decomposition exists:

$$\langle \boldsymbol{G}_j(u,\cdot), \boldsymbol{\phi}_l \rangle_{\mathcal{H}} = \sum_{k=1}^{6} \langle G_{jk}(u,\cdot), \phi_l \rangle = \lambda_l \phi_{jl}(u), \tag{3}$$

where $\{\boldsymbol{\phi}_l = (\phi_{1l}, \dots, \phi_{6l})^T, l = 1,2,\dots\}$ is a set of orthonormal basis functions in $\mathcal{H}$ satisfying $\langle \boldsymbol{\phi}_l, \boldsymbol{\phi}_m \rangle_{\mathcal{H}} = \delta_{lm}$, and λ_l is the l-th eigenvalue in non-increasing order with corresponding eigenfunction $\boldsymbol{\phi}_l(u)$. For random functions $\boldsymbol{\theta} \in \mathcal{H}$, the decomposition above permits the multivariate KL representation based on the covariance kernel $\boldsymbol{G}$ [22],

$$\boldsymbol{\theta}(t) = \boldsymbol{\mu}(t) + \sum_{l=1}^{\infty} \xi_l \boldsymbol{\phi}_l(t), \quad t \in D, \tag{4}$$

where the random coefficients are $\xi_l = \langle \boldsymbol{\theta} - \boldsymbol{u}, \boldsymbol{\phi}_l \rangle_{\mathcal{H}}$ with $E\xi_l = 0$ and $E(\xi_l \xi_m) = \lambda_l \delta_{lm}$. Explicitly, for each $t \in D$, the representation in eq. (4) is

$$\begin{pmatrix} \theta_1(t) \\ \vdots \\ \theta_6(t) \end{pmatrix} = \begin{pmatrix} \mu_1(t) \\ \vdots \\ \mu_6(t) \end{pmatrix} + \xi_1 \begin{pmatrix} \phi_{11}(t) \\ \vdots \\ \phi_{61}(t) \end{pmatrix} + \xi_2 \begin{pmatrix} \phi_{12}(t) \\ \vdots \\ \phi_{62}(t) \end{pmatrix} + \cdots$$

The eigenvalues λ_j represent the amount of variability in $\boldsymbol{\theta}$ explained by the $\mathbb{R}^6$-valued multivariate functional principal component $\boldsymbol{\phi}_j$, while the multivariate functional principal component score ξ_j serve as the weights for $\boldsymbol{\phi}_j$ in the above representation. Thus the $\{\boldsymbol{\phi}_j, j = 1,2,\dots\}$ capture the dominant modes of the variations in the vector-valued random function $\boldsymbol{\theta}$,and the scores $\{\xi_j\}$ (along with the eigenvalues $\{\lambda_j\}$) characterize the correlations between the individual component functions $\theta_k, k = 1\dots,6$. In practice, a truncated version of the KL representation

$$\boldsymbol{\theta}_M(t) = \boldsymbol{\mu}(t) + \sum_{k=1}^{M} \xi_k \boldsymbol{\phi}_k(t), \quad t \in D, \tag{5}$$

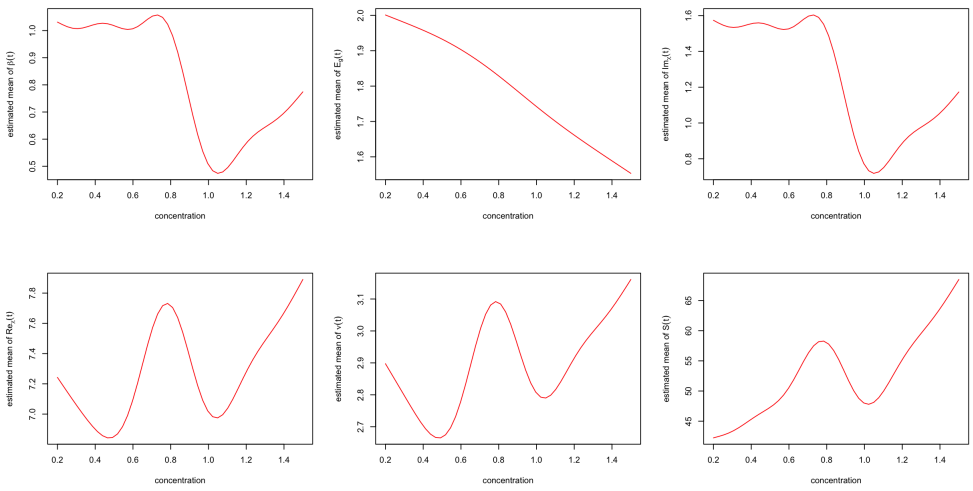

Figure 6: Mean functions of $\beta(t), E_g(t), Im_\chi(t)$ (top row), and $Re_\chi(t), \nu(t), S(t)$ (bottom row) following MFPCA.

for some reasonably large positive integer M. The joint eigenfunctions $\boldsymbol{\phi}_j$ are the multivariate analogues of the univariate eigenfunctions in the KL expansion presented in our earlier paper [10]. They represent the *joint* modes of variation in the set of six parameters under consideration.

5.2 Estimation of joint mean function and modes of variation

While (5) provides the theoretical basis for the required decomposition, joint analysis of parameters with data requires estimating each quantity in the expression. In our earlier work [10] we had detailed the estimation procedure for FPCA for each parameter within $\boldsymbol{\theta}$. In the interests of brevity, it suffices here to note that the estimation under the univariate setup can be adapted and extended to the multivariate setting upon combining the univariate functions to obtain $\boldsymbol{\theta}$. There are several ways of carrying out the extension (see [19.20], for example). In this paper, we will adopt the one proposed in [9]. For our purposes we merely note that using eq. (5), for a fixed M, the function $\boldsymbol{\theta}(t)$ can be represented in the estimated (truncated) joint orthonormal basis $\{\hat{\boldsymbol{\phi}}_l, l = 1, \ldots, M\}$ as

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\mu}}(t) + \sum_{l=1}^{M} \hat{\xi}_l \hat{\boldsymbol{\phi}}_l(t), \quad t \in D, \tag{6}$$

where $\hat{\boldsymbol{\mu}}, \{\hat{\xi}_j\}$ and $\{\hat{\boldsymbol{\phi}}_j\}$ are estimates of $\boldsymbol{\mu}, \{\xi_j\}$ and $\{\boldsymbol{\phi}_j\}$ respectively. Consequently, the jth parameter $\theta_j, j = 1, \ldots, d$ assumes the representation

$$\hat{\theta}_j(t) = \hat{\mu}_j(t) + \sum_{l=1}^{M} \hat{\xi}_l \hat{\phi}_{jl}(t), \quad t \in D.$$

We refer the interested reader to [9] for details on the estimation procedure. The key point to note here is that starting from discrete observations, through the use of MFPCA, we are able to capture and quantify correlations between the experimental parameters as functions of concentration values. Figure 6 shows the estimated mean functions of the parameters using (6) with $M = 2$. We choose $M = 2$ since most of the variation in the sample of functions are almost fully captured in the first eigenfunction $\hat{\boldsymbol{\phi}}_1(t)$ shown in Figure 7. The corresponding eigenvalue is 8.79 and the second eigenvalue is 0.0000245 offering further evidence that most of the information is present in the first eigenfunction. Theoretically, we have that $\beta(t) \propto Im_\chi(t)$ and $\nu(t) \propto Re_\chi$, and this behaviour is observed in the estimated mean functions as well. Another perspective of the correlation between the parameters can be obtained by plotting 2D curves corresponding to the estimated mean functions of the individual parameters: $[0.2, 1.5] \ni t \mapsto (\hat{\mu}_i, \hat{\mu}_j)^T, i, j =$

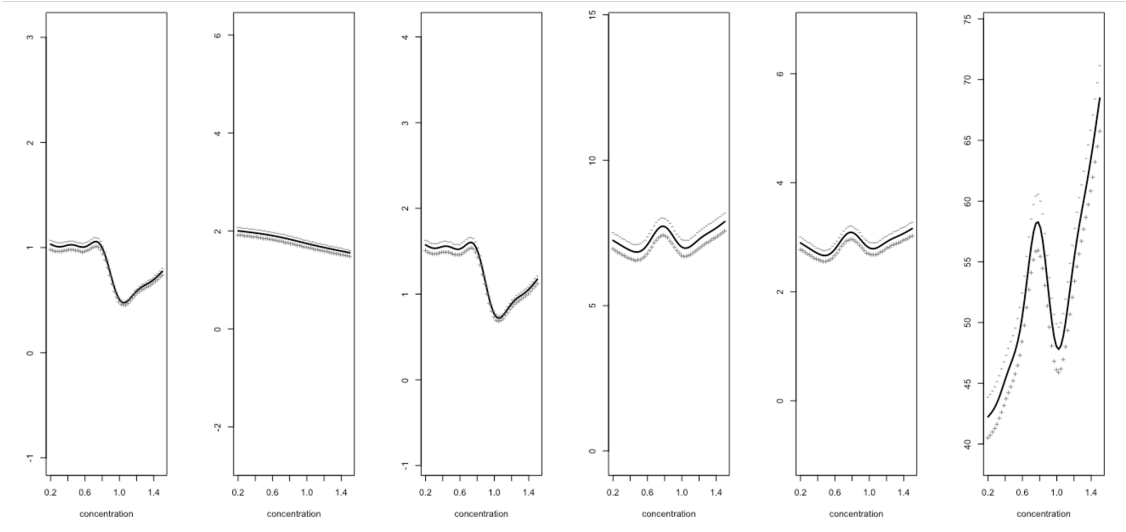

Figure 7: Individual eigenfunctions of the parameters $\beta(t), E_g(t), Im_\chi(t), Re_\chi(t), \nu(t), S(t)$ (from left to right) in the first Principal Component direction (6-dimensional vector-valued eigenfunction). Plotted are the mean function plus/minus a constant factor times each eigenfunction.

$1, \ldots, 6, i \neq j$. Figure 8 provides more evidence of the correlations, and clearly demonstrates the advantages associated with joint estimation of the mean functions of the parameters under the MFPCA framework.

6 Performance measure and optimal dopant level

The MFPCA procedure provides us with an estimate $\hat{\boldsymbol{\mu}}(t) = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_6)^T$ of the joint mean function $\boldsymbol{\mu}(t)$ of $\boldsymbol{\theta}(t) = (\beta(t), E_g(t), Im_\chi(t), Re_\chi(t), \nu(t), S(t))^T$ at any dopant concentation level in $t \in D = [0.2, 1.5]$ that accounts for correlations (linear or nonlinear) between microstructural and nonlinear optical parameters, as evidenced in Figures 6 and 8. This provides information on how the parameters, on an average, vary with one another at any dopant level. Thus an overall measure of performance of the polymer at any dopant level can be define using $\hat{\boldsymbol{\mu}}(t)$, from which the optimal dopant concentration level can be determined.

Our definition of such a performance measure will be based on the component estimated mean functions $\hat{\mu}_k(t), k = 1, \ldots, 6$. Since the parameter values in the data are all positive and on different scales (for example, $S(t) \in [39, 71]$ for any dopant level, significantly larger when compared to the other parameters with a combined range $[0, 8]$; see Supplementary Materials for data), we first normalise each mean function to obtain

$$\hat{\mu}_k^n(t) := \frac{\hat{\mu}_k(t)}{\sup_{t \in D} \hat{\mu}_k(t)}, \quad k = 1, \ldots, 6,$$

that provides an estimated joint normalised mean function $\hat{\boldsymbol{\mu}}^n(t) = (\hat{\mu}_1^n(t), \ldots, \hat{\mu}_6^n(t))^T$. We consider a weighted convex combination

$$\boldsymbol{w}^T \hat{\boldsymbol{\mu}}^n(t) = \sum_{k=1}^{6} w_k \hat{\mu}_k^n(t),$$

as a measure of performance, where $\boldsymbol{w} = (w_1, \ldots, w_6)^T$, where for each $k = 1, \ldots, 6$, $w_k \geq 0$ with $\sum_{k=1}^6 w_k = 1$. The weights can represent prior beliefs of the experimenter (for e.g. one may wish to have crystallite size to have maximum influence on the measure of performance), or can be derived from experimental conditions. However, the arbitrariness with the choice of weights $\boldsymbol{w}$ implies that the weighted combination $\boldsymbol{w}^T \hat{\boldsymbol{\mu}}^n(t)$ can significantly change in shape depending on the weights. In order to circumvent this issue, we define a measure of performance known as the *performance index* function of a poly-

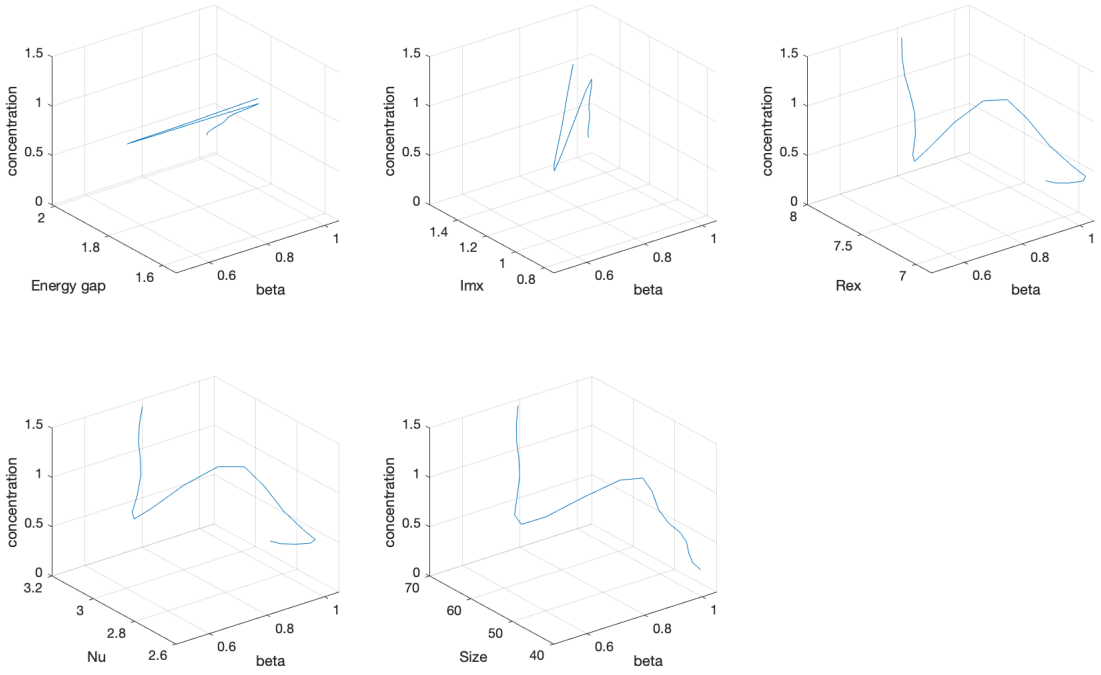

Figure 8: Two-dimensional estimated mean functions of parameters against concentration values. Top row (left to right): β and E_g, β and Im_χ, β and Re_χ; bottom row (left to right): β and ν, β and S.

mer composite as

$$\mathcal{P}(t) = \int_{\Delta_5} \boldsymbol{w}^T \hat{\boldsymbol{\mu}}^n(t) d\boldsymbol{\lambda}, \quad t \in D,$$

where $\Delta_5 = \{\boldsymbol{w} \in \mathbb{R}^6 : 0 \leq w_k \leq 1, \sum_k w_k = 1\}$ is the five-dimensional unit simplex $\mathbb{R}^6$, and $d\boldsymbol{\lambda}$ is the uniform distribution (volume measure) on Δ_5. An equivalent definition is

$$\mathcal{P}(t) = \sqrt{6} \int_{\Sigma_5} \boldsymbol{x}_0^T \hat{\boldsymbol{\mu}}^n(t) d\boldsymbol{x},$$

where $d\boldsymbol{x}$ in the Lebesgue measure on $\mathbb{R}^5$, $\Sigma_5 = \{\boldsymbol{x} = (x_1, \ldots, x_5)^T \in \mathbb{R}^5 : x_i \geq 0, \sum_i x_i \leq 1\}$, and $\boldsymbol{x}_0 = (1 - \sum_{i=1}^5 x_i, x_1, \ldots, x_5)^T$. Armed with the performance index $\mathcal{P}(t)$, expressed as a function of dopant level $t \in D$, the optimal dopant level then is defined as

$$t_{opt} := \text{argmax}_{t \in D} \mathcal{P}(t)\,.$$

Note that by virtue of its definition $0 \leq \mathcal{P}(t) \leq 1$ for all $t \in D$. Since t_{opt} remains invariant to scale transformations of $\mathcal{P}(t)$, it is easy to see that t_{opt} would be unaffected by the normalisation $\hat{\boldsymbol{\mu}}^n(t)$ of the estimated mean function $\hat{\boldsymbol{\mu}}(t)$.
The performance index $\mathcal{P}(t)$ (and consequently t_{opt}) cannot be computed in closed form since the estimated (normalised) mean function $\hat{\boldsymbol{\mu}}^n(t)$ is computed from the data numerically. The integral with respect to the Lebesgue measure over the set Σ_5 can be computed numerically in principle. Alternatively, we note that $\mathcal{P}(t)$ can be expressed as the expectation $E(\boldsymbol{w}^T \hat{\boldsymbol{\mu}}^n(t))$ taken with respect to the uniform distribution on the simplex Δ_5. We thus approximate the integral using Monte Carlo (MC) by sampling from the uniform distribution $d\boldsymbol{\lambda}$ on the simplex Δ_5 through the mapping $\boldsymbol{w} \mapsto \mathcal{P}(t)$. The Dirichlet distribution on Δ_5 with parameters $\alpha_1, \ldots, \alpha_6$ is the given by the probability density function

$$f(\boldsymbol{w}) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_{i=1}^6 \Gamma(\alpha_i)} \prod_{i=1}^6 w_i^{\alpha_i - 1}, \quad \boldsymbol{w} \in \Delta_5$$

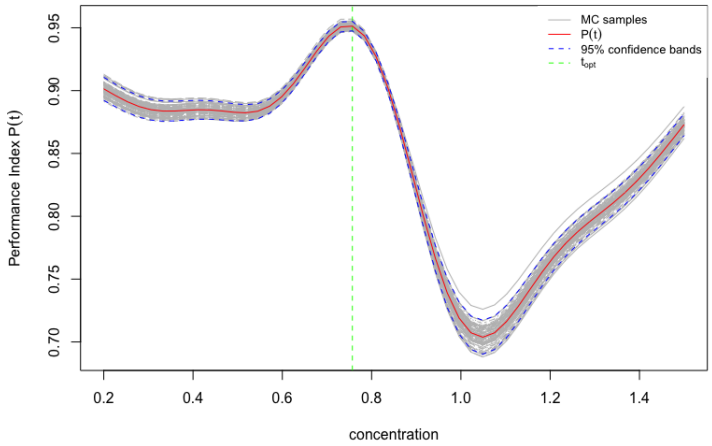

Figure 9: MC estimate of performance index $\mathcal{P}(t)$ (solid red) with 95% confidence bands (dashed blue) with green dashed line identifying $t_{opt} = 0.7571$. The MC samples are in grey.

which when $\alpha_1 = \cdots = \alpha_6 = 1$ reduces to the uniform distribution on Δ_5. It is evidently easier to sample uniformly from Δ_5 as opposed to the set Σ_5.

6.1 Monte Carlo estimation and uncertainity quantification of $\mathcal{P}(t)$ and t_{opt}

The MC approximation of $\mathcal{P}(t)$ is given by $\mathcal{P}_{mc}(t) := \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{w}_i^T \hat{\boldsymbol{\mu}}^n(t)$, where $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ is a random sample from the uniform distribution on Δ_5. From well-known properties of the MC method, we note that $E(\mathcal{P}_{mc}(t)) = \mathcal{P}(t)$ for every $t \in D$ and is hence an unbiased estimator. From the Strong Law of Large Numbers (SLLN) we also note that $\mathcal{P}_{mc}(t) \to \mathcal{P}(t)$ as $N \to \infty$ almost surely, and consequently the approximation error of the MC method measured using the root mean squared error is

$$\sqrt{E\left([\mathcal{P}_{mc}(t) - \mathcal{P}(t)]^2\right)} = O(N^{-1/2}),$$

which decays to zero fairly quickly with increasing MC sample size N.
A second level of MC sampling can now be carried out to obtain uncertainty or distributional estimates of the MC estimator $\mathcal{P}_{mc}(t)$. That is, $\mathcal{P}_{mc}(t)$ can again be computed J times using further MC sampling of $\boldsymbol{w}$ from uniform distribution on Δ_5 to obtain $\mathcal{P}_{mc}^{(j)}(t), j = 1, \ldots, J$.
Let $\bar{\mathcal{P}}_{mc}(t) = \frac{1}{J}\sum_{j=1}^{J} \mathcal{P}_{mc}^{(j)}(t)$, which will be quite close to $\mathcal{P}_{mc}(t)$. From the Central Limit Theorem, as $J \to \infty$ for each $t \in D$, $\bar{\mathcal{P}}_{mc}(t)$ approximately follows a univariate Gaussian distribution with mean $\mathcal{P}_{mc}(t)$ and variance $\sigma_J = \frac{1}{J-1}\sum_{i=1}^{J}(\mathcal{P}_{mc}^{(j)}(t) - \bar{\mathcal{P}}_{mc}(t))$. However, since $\mathcal{P}_{mc}(t) \to \mathcal{P}(t)$, by another application of SLLN we also have $\bar{\mathcal{P}}_{mc}(t) \to \mathcal{P}(t)$ almost surely as $\min\{N, J\} \to \infty$.
This enables us to construct approximate pointwise 95% confidence bands for $\mathcal{P}(t)$ using MC samples with

$$CI_{95,J}(t) = \left[\bar{\mathcal{P}}_{mc}(t) - 1.96\frac{\sigma_J}{\sqrt{J}}\,,\, \bar{\mathcal{P}}_{mc}(t) + 1.96\frac{\sigma_J}{\sqrt{J}}\right].$$

An additional advantage of the MC method is that for each MC sample $\mathcal{P}_{mc}^{(j)}(t)$ we can obtain the optimal dopant level $t_{opt}^{(j)}, j = 1, \ldots, J$. This enables us to quantify uncertainty of the optimal dopant concentration level as well. Left pnael of Figure 9 and 10 illustrate the MC estimate $\mathcal{P}_{mc}(t)$ along with pointwise 95% confidence bands, in addition to an estimate of t_{opt}; right panel shows the density plot of $t_{opt}^{(j)}$. The bands are quite narrow suggesting that our estimate of $\mathcal{P}(t)$ is not highly variable and is a good approximation the integral used to define the performance index. Our estimate of the optimal dopant concentration level t_{opt} is 0.7571. Using the MC estimates we assess its variability and note that it is mostly

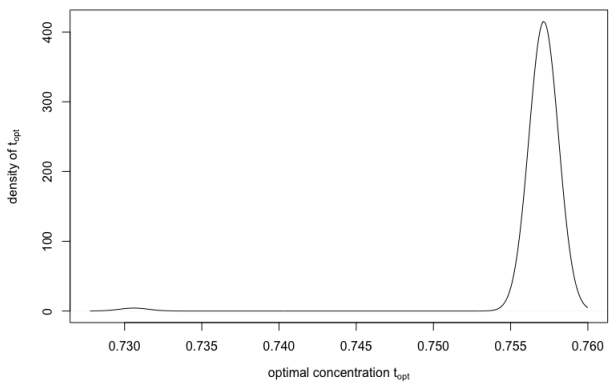

Figure 10: Estimated probability density of t_{opt} from MC samples.

restricted to the interval $[0.73, 0.76]$. An estimate of the distribution of t_{opt} is provided in Figure 10; the left skew nature of the density is due to about 10 samples (out of J=1000) of t_{opt} being close to 0.73, but the rest are concentrated around 0.757. Thus, excluding the 10 slightly outlying ones, the density is effectively symmetric around 0.757. We have used $N = J = 10000$ in order to obtain the MC estimates.

7 Concluding remarks

The proposed measure of performance is based on Multivariate Functional Principal Component Analysis (MFPCA) of data pertaining to microstructural and optical parameters of a polymer composite. For the polymer composite under consideration, the proposed performance measure suggests an optimal dopant concentration level of 0.76%, which is quite close to level of 0.8% used during polymer preparation. Beyond 1.5% the composite films become brittle and unsuitable for applications. It is however important to note that the *performance measure (and MFPCA) can be used on data from any set, containing any number, of physical parameters* of interest in studies of polymer composites.
The definition of performance measure $\mathcal{P}(t)$ as an integral over the weights $\boldsymbol{w}$ ensures that it is invariant to any particular choices of weights. If indeed the experimenter is compelled to view certain parameters as more important than others then the integral can be dispensed with by using $\mathcal{P}(t) = \boldsymbol{w}^T \hat{\boldsymbol{\mu}}^n$ as an alternative measure of performance for a particular choice of weights $\boldsymbol{w}$.
The use of Monte Carlo techniques based on sampling uniform weights from the unit simplex Δ_5 is key to quantifying uncertainty in estimating the performance measure and optimal dopant level. Such uncertainty arises since data measurements can be recorded with measurement error, and moreover, the $\mathcal{P}(t)$ and t_{opt} change with changing data composition. This aspect of the methodology proposed can be a valuable data analytic tool in the field of polymer composites.
Computation was carried out in `MATLAB` and statistical software `R` using the package `MFPCA` [9]. Source code can be made available on request.

8 Acknowledgements

KB's research is partially supported by grants NSF DMS 1613054, NSF DMS 2015374 and NIH R37-CA214955.

9 References

[1] K. Janardhana, V. Ravindrachary, P. Rajesh Kumar, and Ismayil, Polymer Engineering and Science (2013), 53, 1958.
[2] P. Poornesh, S. Shettigar, G. Umesh, K. Manjunatha, K. Prakash Kamath, B. Sarojini, and B. Narayana, Optical Materials (2009),31, 854.
[3] L. Beecroft and C. Ober, Chemistry of Materials (2007),9(6),1302.
[4] V. Scardaci, A. Rozhin, F. Hennrich, and W. Milne, Physica E Low-dimensional Systems and Nanostructures (2007), 36(1), 115.
[5] R. M. T. Madiona, D. A. Winkler, B. W. Muir, and P. J. Pigram, Applied Surface Science. (2019) 478, 465.
[6] M. Bonta and A. Limbeck, Journal of Analytical Atomic Spectroscopy (2018), 33, 1631.
[7] S. Shettigar, G. Umesh, K. Chandrasekharan, and B. Kalluraya, Synthetic Metals (2007),157, 142.
[8] L. Yang, R. Dorsinville, Q. Z. Wang, W. K. Zou, P. P. Ho, N. L. Yang, R. R. Alfano, R. Zamboni, R. Danieli, G. Ruani, and C. Taliani,(1989),J. Opt. Soc. Am. B 6, 753.
[9] C. Happ and S. Greven, Journal of the American Statistical Association (2018),113, 649.
[10] T. G. K. Urs, K. Bharath, S. Yallappa, and S. Rudrappa,Journal of Applied Crystallography (2016), 49, 594.
[11] G. Zhao, H. Ni, S. Ren, and G. Fang, Polymers (2018),10, 290 .
[12] H. Saadon, Optical and Quantum Electronics (2016),48(1), 40.
[13] M. Sheik-Bahae, A. A. Said, T. H. Wei, D. J. Hagan, and E. W. V. Stryland, IEEE Journal of Quantum Electronics (1990), 26, 760.
[14] G.K Williamson and W.H Hall, Acta Metallurgica (1953), 1,22.
[15] H. Kaneko, S. Nishimoto, K. Miyake, and N. Suedomi, Journal of Applied Physics (1986),59, 2526.
[16] Nagappa, J. Mahadeva, R. Somashekar, C. V. Yalemaggad, G. Umesh, and K. B. Manjunatha, Molecular Crystals and Liquid Crystals(2011), 540, 88.
[17] E. W. Van Stryland and M. Sheik-Bahae,Characterization tech-niques and tabulations for organic nonlinear materials (1998),18, 655.
[18] J. Ramsay and B. Silverman, Functional Data Analysis (2nd ed.) (Springer, NY, 2005).
[19] J. Jacques and C. Preda, Computational Statistics and Data Analysis (2014),71, 92.
[20] J. M. Chiou, Y. F. Yang, and Y. T. Chen, Statistica Sinica (2014),24, 1571.
[21] C. S. Withers, Bulletin of the Australian Mathematical Society (1974), 11, 373.
[22] A. Balakrishnan, Journal of Mathematical Analysis and Applications (1960),1, 386.