Pattern Recognition Approach to Quantitative Description of the Microstructure of Disordered Composites for Estimation of Thermal Conductivity

Ignacio J. Turias[1], Jose M. Gutierrez[2], Pedro L. Galindo[1]

[1]*Department of Computer Science - [1]Research Group "Computational Intelligence Systems"*
[2]*Department of Applied Physics - [2]Research Group "Composite Materials"*
e-mail: [ignacio.turias | josemaria.gutierrez | pedro.galindo]@uca.es
Avda. Ramón Puyol s/n - 11202
EPS of Algeciras, University of Cadiz - Spain

ABSTRACT

This paper explores the application of pattern recognition techniques in the characterization of a multiphase realistic disordered composite. Some descriptors based on Voronoi cells are extracted for different fiber distributions. Principal Components Analysis is used to find the best features to estimate effective thermal conductivity with a Multiple Linear Regression model instead of the Finite Element Method.

Keywords: image analysis, composites, microstructure, thermal conductivity

1. INTRODUCTION

The effective properties of composite materials reinforced by unidirectional fibers play an important role in the design for several applications. Composite materials can improve systems reliability through enhanced thermal and mechanical performance. In this paper, the effective conductivity of composite in normal direction to the fiber axes has been considered. The effective properties of two phase composite material depend on (i) properties of each phase (matrix and reinforcement conductivities), (ii) fiber volume fraction, (iii) geometrical arrangements. This paper focuses on (iii), specifically having considered regular distributions of fibers (rectangular, squared, and hexagonal arrays), exploring realistic composites microstructures generated through computer simulations adding "noise" to the positions of fibers centres of each regular array distribution.

In the literature, the estimation of composite properties has been based typically on the assumption that the microstructure is regular and then it can be represented by a repeating unit cell /1-3/. However, in the composites found in practice, the fibers hardly assume an ordered array, and the predicted conductivity could be considerably different from the actual value. In this work, the application of Voronoi tessellation of a planar two-phase composite has been explored, and then different geometric descriptors to quantify a given morphology have been calculated in an alternative way to that proposed in references /4-7/. Other approaches can be found in the papers by Kim and Torquato, and Smith and Torquato /8,9/. Also Pyrz /10/ has shown how to deal with disordered distributions using fractal description, and Pitchumani /11/ how to construct an equivalent fractal unit cell to estimate effective thermal conductivity.

The first step of the work was the computer generation of simulated artificial microstructures. The conductivity fiber/matrix ratio was fixed at 10. The volume fraction considered is fixed at 25%. Each composite microstructure is limited to a 256x256 pixel resolution. Three different patterns of regular periodic arrangements are considered: squared, rectangular, and hexagonal. For each sample, a set of disordered microstructures has been obtained by the addition of random noise to the coordinates of each fiber centre. The value of the introduced random noise is at maximum 10% of the inter-fiber distance in the original regular array. This random movement of the locations is applied to 100% of the fibers in the distribution. The ANSYS program was used to find out how a given microstructure design works under operating conditions, and then to compute the effective transversal thermal conductivity. ANSYS (copyright by SAS IP, Inc.) is a computer program for analysis and design with Finite Element Method (FEM) available on many kinds of computers.

Several techniques based on automated image pattern recognition have recently been introduced for characterizing composites microstructures. Important contributions in this area have been made by Brockenbrough /12/, Everett /4/, Pyrz /5/, Ghosh /6,7/, and Yang /13/. Dirichlet tessellation has been used as a tool in the characterization. This tessellation allows the discovery of the "natural regions" of immediate influence of each fiber. These regions are called Voronoi cells. This facilitates characterization of an arrangement of fibers, computing some descriptors of the resulting Voronoi cells. These descriptors can be used to distinguish between different distributions and, as we have proposed here, as inputs in a regression model which computes an estimation of the transverse effective thermal conductivity. If a good estimate of some kind of property of the material is wanted, enough data will be required. In a situation where data are scattered, the location of the regressors will be highly sensitive to individual data points (outliers). This tendency to sparseness is referred to as "*the curse of dimensionality*" /14/, and methods such as Principal Component Analysis (PCA) should be used to reduce this effect.

2. QUANTITATIVE CHARACTERIZATION

Given m data points in n-dimensional space, a Dirichlet tessellation (also known as Voronoi diagram) is the partition of n-dimensional space into m polyhedral regions, one region for each data point. Such a region is called a Voronoi cell. A Voronoi cell satisfies the condition that it contains all points that are closer to its data point than any other data point in the set. In that way, the microstructure is covered with a unique cell for each fiber, defined as the smallest convex polygon surrounding it. In this work, an original region-growing algorithm based on an incremental rhomboidal growth of each fiber centre has been used /15/. This allows us to directly calculate which cell each individual pixel belongs to in the microscopic image.

Tessellation of a microstructure has considerable importance in generating descriptors to quantify a given composite. It naturally identifies regions of immediate influence for each fiber. In related papers /4-7/, the main contributions concern how to characterize fiber clustering. This paper is an attempt to highlight how to distinguish different kinds of arrangements of real non-ordered microstructures in order to estimate thermal conductivity as a pattern recognition approach alternative to descriptions found in the literature.

3. DESCRIPTION

Pattern Recognition techniques categorize or analyse objects based upon some measurements made on those objects /16/. The features must have sufficient information to uniquely characterize a kind of object. In our work, this description must be invariant to scale, since two microstructures with different numbers and diameters of fibers can share the same volumetric fraction. Of course, it must be invariant to location, but not invariant to rotation because we want to predict a directional property of the composite material. In case of ordered microstructure, which contains a periodic set of fibers, it is possible to construct a repeating unit cell with only one fiber. However, realistic samples should include several fibers in a representative window and take the mean value of features (see Figure 1). Three different sets of features to describe a given object were used:

i) *Basic Geometric Features* are the simplest descriptors from a computational point of view, since the required effort is less. However, it is easily understood that such

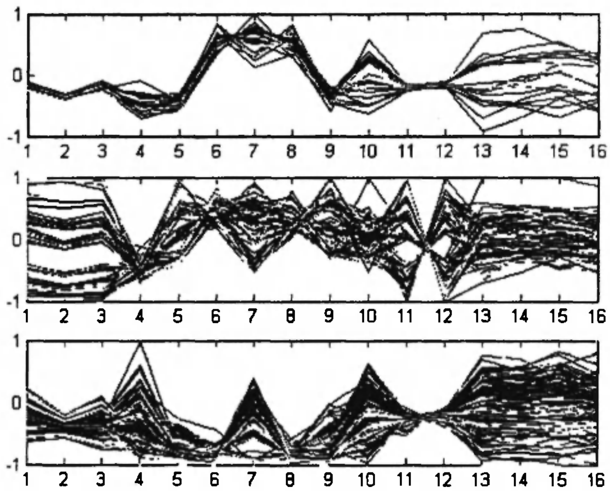


Fig. 1: Set of the 16 extracted features to realistic squared, rectangular and hexagonal arrangements. Numbers 1-3 are basic geometric features, 4-9 are Fourier descriptors *df(2)-df(7)* and, 10-16 are the invariant moments to location and scale. Feature numbers 5, 6, 8, 9 (Fourier descriptors) can help to distinguish between the set of rectangular-squared arrangement patterns and the hexagonal set, simply by using a threshold. Basic geometric features, and features labelled numbers 11 and 12, allow to classify between hexagonal-squared set and rectangular one, but this is not the objective of this study.

descriptors must be used carefully. In this case one can select (Figure 1, features 1-3):

- *height/width*: its value is unity for squared shapes.
- *ix/iy*: it has one greater order, it takes inertias with respect principal axes.
- *length/width*: redefinition of PI number. For a square shape its value is four.

ii) *Fourier descriptors* are another way to identify objects with the aid of the boundary points. The best and most complete introduction can be found in the paper by Wallace /17/. Firstly the boundary must be scanned and stored in a counter clockwise direction. Each point *(a,b)* will be addressed as a complex number $z-a+jb$ and the sequence of these N numbers as a complex function $f(z)$. The resulting transform $F(u)$ gives the Fourier description by means of coefficients $(df(0),df(1),...,df(N-1))$. These values need to be normalized for location and size. Element $df(0)$ contains the centroid of the boundary. Normalisation for location is done by setting this value to 0. Coefficient $df(1)$ had the largest magnitude when the boundary was scanned in counter clockwise direction. Normalisation by size is therefore achieved by dividing all Fourier Descriptors by $df(1)$. A further normalisation to rotation is not necessary. Generally, these descriptors are only used to describe the overall shape without the need for too many details; therefore it is sufficient to set the number of used coefficients to 8 (Figure 1, features 4-9).

iii) *Invariant Moments* are based on continuous moments of a multidimensional function (m_{pq}) [16]. Central moments of order $pq(\mu_{pq})$, are invariants to location. In order to get scale invariance normalized central moments η_{pq} can be used:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y)$$

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}; \; \gamma = \frac{p+q}{2} + 1$$

A complete overview is given by Hu /18/. Commonly, it is only necessary to select the set of values $pq = \{11,20,02,12,21,30,03\}$ to normalized central moments (Figure 1, features 10-16). Additional invariant moments can be derived, which are invariant to rotation, location and scale, but they are not necessary in this application.

4. PRINCIPAL COMPONENT ANALYSIS

The goal of PCA is to reduce the dimension of the working space, preserving as much as possible the relevant information maximising the variance. This linear dimensionality reduction procedure is also called Karhunen-Loéve transformation. PCA is discussed at length in Jollife /19/. The aim is to construct a new space, where no one component is correlated with any other components. The diagonal form of the final covariance matrix implies that the variance of a variable with itself will be maximised whereas the covariance with any other variable will be nil. The amount of information of each new component is given by its eigenvalue. PCA has been calculated onto our composites database. For four new transformed components, the explained variance is 93.98 percent of total, if the number of reduced features decrease to 3 or 2, variance also decrease to 84.28 and 64.87 percent, respectively. In such a case some principal components can be dropped because they explain only a small amount of the data. Figure 2 shows the first four eigenvectors obtained.

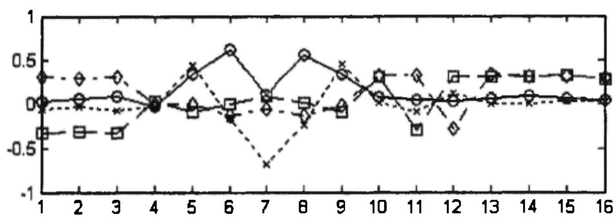
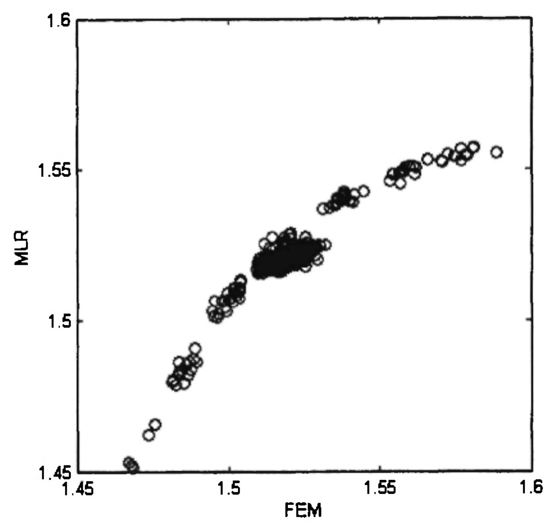


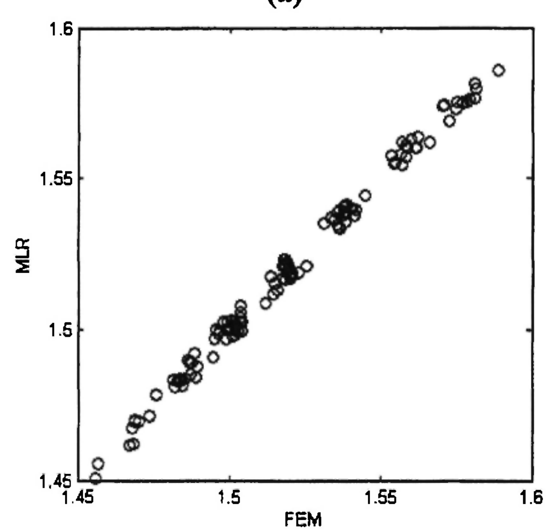
Fig. 2: The new basis vectors are linear combination of the 16 original features. Only the values for the first four new principal components are shown since they retained most of the variance of the data (1 – solid+circle, 2 – square+dashed, 3 – diamond+ dashdotted, 4 – mark+dotted)

5. THERMAL CONDUCTIVITY ESTIMATION

The system considered was designed to take a microscopic image of a composite and estimate the value of its effective thermal conductivity. Thus, the overall system can be viewed as a mapping from a set of input features, X, to an output variable Ke_y /14/. In general, it is not possible to determine a suitable form for the mapping, and we have to work with a set of examples, called the training set. The mathematical form of the mapping is determined with the help of the data. Of course, we need to build a system capable of making good predictions on unseen data. In order to measure this generalization capability, two-fold cross-validation with another set of samples called the test set is used. Computing regression parameters by directly inverting matrix X of independent variables is really dangerous. There is a vast array of methods to solve Multiple Linear Regression (MLR) problems: Gauss-Jordan, LU, or QR decomposition. However, the most effective method is Singular Value Decomposition which handles all problems that may arise, such as singularities or ill-conditioning /20/. It is very interesting to point out that PCA results in a set of features which are independent and uncorrelated. This fact permits the resolution MLR using a method such as QR. Tables 1-3 can help to clarify the results obtained when PCA is used versus the outcome of take all of the original features. Besides, the sizes of the training and test sets were changed as well as the number of final principal features. Figure 3 shows how similar were the

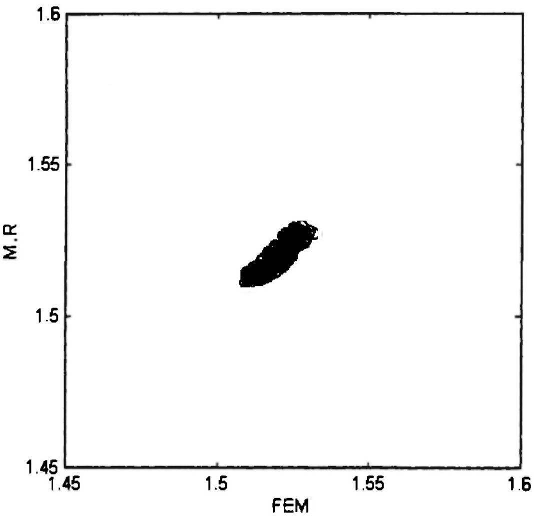


(a)



(b)

continued...



(c)

Fig. 3: FEM vs. MLR estimations. The range of values of effective conductivities in the case of samples belong to squared-rectangular class are bigger than the values for hexagonal samples which are less sensitive to orientation and noise (a) multiple linear regression (MLR) model for both of disordered squared-rectangular and hexagonal arrangements. The correlation coefficient R between FEM and MLR estimations was 0.724 for the hexagonal samples and 0.94 for squared-rectangular class (b) MLR model for disordered squared-rectangular samples. R = 0.997 (c) MLR model for disordered hexagonal arrangement class. R = 0.88. These values of correlation coefficient R clarify that design two different MLR models improves the results obtained with only one.

results obtained with FEM and MLR model, whether only one MLR model was designed for both kinds of arrangements or different MLR models were created for each separate class.

6. CONCLUSIONS

The results suggest that this work provides an effective way of computing conductivities on real-disordered arrangements, obtaining very similar results to those of FEM with a MLR model, as well as characterization based on the features explained above was adequate to reach that estimation. However, as is clearly shown in Figure 3.a, with only one regression model for all the samples the outcome of the fitting is not good enough since a clear curvature can be observed in the correlation between FEM and MLR estimations. Furthermore, the set of disordered hexagonal patterns does not fit as well as rectangular and squared arrangements, as can be seen in the values of correlation coefficients. To avoid this problem, one regression model was built for each class of considered patterns, improving the results in such a way as Figures 3.b-c point out, where a much better correlation between FEM and MLR is displayed. The numeric values of the root mean square errors can be found in Tables 1-3. It is worth mentioning that the regression model for hexagonal class of patterns produces the best results in the test stage, much better than those obtained with only one model for both classes. In the MLR model designed for class of squared-rectangular disordered arrangements, it is difficult to reach those results, since the samples do not cover entirely the space of possible values of effective conductivity and this clear tendency to sparseness is impossible to avoid with the available simulated data, as could happen in real life.

Eigenvectors from PCA transformation, especially the first one, show that the Fourier descriptors are the strongest weighted, and therefore, the most relevant original features in order to perform a regression analysis for such kinds of fiber distributions.

PCA improves the results obtained when all of the features are used over test sets. As can be seen in Table 1-3, increasing the size of the sets produces greater error in the training stage, since it is more difficult to find the parameters of the regression model. In the test stage, the size of the test set does not matter so much, since results were very similar with the different sizes considered. Although the use of all of the features makes the error of the training stage smaller, PCA gives better results in the testing stage.

Future research will address the application of other kinds of regression models such as artificial neural networks.

Table 1

One multiple linear regression model (MLR) was designed for the patterns belong to all kinds of disordered arrangements (squared-rectangular and hexagonal). Root Mean Square Errors (RMSE) found in regression over training and test sets. These results were obtained as means of 30 times resampling random procedure.

Training/Test Sets Sizes	RMSE PCA-MLR 2 reduced features		RMSE PCA-MLR 4 reduced features		RMSE MLR 16 original features	
	Training	Test	Training	Test	Training	Test
50 / 50	7.92 E-05	1.14 E-03	1.85 E-05	1.22 E-03	5.82 E-06	1.23 E-03
25 / 25	6.80 E-05	1.14 E-03	1.52 E-05	1.29 E-03	2.70 E-06	1.25 E-03

Table 2

MLR model designed for realistic squared-rectangular arrangements. RMSE in regression over training and test sets. These results were obtained as means of 30 times resampling random procedure.

Training/Test Sets Sizes	RMSE PCA-MLR 2 reduced features		RMSE PCA-MLR 4 reduced features		RMSE MLR 16 original features	
	Training	Test	Training	Test	Training	Test
50 / 50	2.28 E-05	2.35 E-03	4.96 E-06	2.32 E-03	1.94 E-06	2.43 E-03
25 / 25	2.01 E-05	2.42 E-03	4.27 E-06	2.53 E-03	9.31 E-07	2.35 E-03

Table 3

MLR model designed for realistic hexagonal arrangements. RMSE in regression over training and test sets. These results were obtained as means of 30 times resampling random procedure.

Training/Test Sets Sizes	RMSE PCA-MLR 2 reduced features		RMSE PCA-MLR 4 reduced features		RMSE MLR 16 original features	
	Training	Test	Training	Test	Training	Test
50 / 50	1.42 E-05	5.70 E-05	1.20 E-05	5.34 E-05	8.24 E-06	6.08 E-05
25 / 25	1.46 E-05	5.32 E-05	1.06 E-05	6.08 E-05	3.96 E-06	6.47 E-05

REFERENCES

1. E. Behrens. *J. Compos. Mater.*, **2** (1),2-17 (1968).
2. L.S. Han and A.A. Cosner. *J. Heat Transfer*, **103**, 387-392 (1981).
3. J.M. Gutierrez. Ph.D. Thesis, University of Cádiz, Spain,1996.
4. R.K. Everett and J.H. Chu. *J. Compos. Mater.*, **27** (11), 1128-1144 (1993).
5. R. Pyrz. *Comp. Sci. Techn.*, **50**, 197-208 (1994).
6. S. Ghosh, Z. Nowak and K. Lee. *Acta Metall. Mater.*, **45** (6), 2215-2235 (1997).
7. S. Ghosh, Z. Nowak and K. Lee. *Comp. Sci. Techn.*, **57**, 1187-1210 (1997).
8. I.C. Kim and S. Torquato. *J. Appl. Phys.*, **68** (8), 3892-3903 (1990).
9. P.A. Smith and S. Torquato. *J. Appl. Phys.*, **65** (3), 893-900 (1989).
10. R. Pyrz. *Sci. Eng. Compos. Mat.*, **6** (3), 141-149 (1997).

11. R. Pitchumani. *J. Heat Transfer, Trans. ASME*, **121**, 163-166 (1999).
12. J.R. Brockenbrough *et al., Acta Metall. Mater.*, **39** (5), 735-752 (1991).
13. S. Yang, A. Tewari and A.M. Gokhale, *Acta Mater.*, **45** (7), 3059-3069 (1997).
14. C.M. Bishop. *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press, 1995.
15. I. Turias and I. Lloret. *II Regional Workshop in Computer Systems*, University of Cádiz, Spain, 1996; I; 320-325.
16. R.C. Gonzalez and P. Wintz. *Digital Image Processing*, 2nd Edition, Reading, MA, Addison-Wesley Publishing Company, 1987.
17. T.P. Wallace and P.A. Wintz. *Comput. Graph. Image Proc.*, **13**, 99-126 (1980).
18. M.K. Hu. *IRE Trans. Info. Theory*, **IT-8**, 179-187 (1962).
19. I.T. Jollife. *Principal Component Analysis*, New York: Springer-Verlag, 1986.
20. W.H. Press *et al.*, *Numerical Recipes in C*, New York, Cambridge University Press, 1992.