

Dimensionality reduction for classification of stochastic fibre radiographs

C.T.J. Dodson¹ and W.W. Sampson²

¹School of Mathematics, ²School of Materials,
University of Manchester, M13 9PL, UK
`ctdodson@manchester.ac.uk`

Abstract. Dimensionality reduction helps to identify small numbers of essential features of stochastic fibre networks for classification of image pixel density datasets from experimental radiographic measurements of commercial samples and simulations. Typical commercial macro-fibre networks use finite length fibres suspended in a fluid from which they are continuously deposited onto a moving bed to make a continuous web; the fibres can cluster to differing degrees, primarily depending on the fluid turbulence, fibre dimensions and flexibility. Here we use information geometry of trivariate Gaussian spatial distributions of pixel density among first and second neighbours to reveal features related to sizes and density of fibre clusters.

Keywords: Dimensionality reduction, fibre networks, fibre clusters, spatial covariance, trivariate Gaussian, radiographic images, simulations

1 Introduction

Near-planar, non-woven stochastic fibre networks are manufactured for a variety of applications such as, at the macroscale for printing, textiles, reinforcing, and filtration and at the nanoscale in medicine. Figure 1 shows a selection of electron micrographs for networks at different scales. Radiography or optical densitometry yield areal density images of the kinds shown in Figure 2.

Much analytic work has been done on modelling of the statistical geometry of such networks and their behaviour in regard to strength, fluid ingress or transfer [1,2,3]. Using complete sampling by square cells, their areal density distribution is typically well represented by a log-gamma or a (truncated) Gaussian distribution of variance that decreases monotonically with increasing cell size; the rate of decay is dependent on fibre and fibre cluster dimensions. Clustering of fibres is well-approximated by Poisson processes of Poisson clusters of differing density and size. A Poisson fibre network is a standard reference structure for any given size distribution of fibres; its statistical geometry is well-understood for finite and infinite fibres. Note that any skewness associated with the underlying point process of fibre centres becomes negligible through the process of sampling by square cells [4].

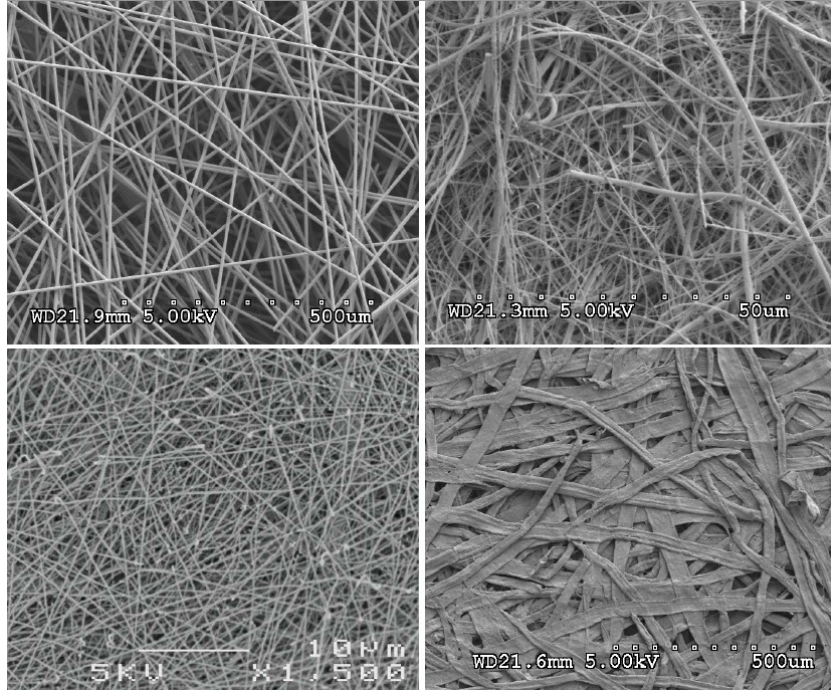


Fig. 1. Electron micrographs of four stochastic fibrous materials. Top left: Nonwoven carbon fibre mat; Top right: glass fibre filter; Bottom left: electrospun nylon nanofibrous network (Courtesy S.J. Eichhorn and D.J. Scurr); Bottom right: paper using wood cellulose fibres—typically flat ribbonlike, of length 1 to 2mm and width 0.02 to 0.03mm.

In the present paper we use information geometry of trivariate Gaussian spatial distributions of pixel density with covariances among first and second neighbours to reveal features related to sizes and density of fibre clusters, which could arise in one, two or three dimensions. For isotropic spatial processes, which we consider here, the variables are means over shells of first and second neighbours, respectively, which share the population mean with the central pixel. For anisotropic networks the neighbour groups would be split into more, orthogonal, new variables to pick up the spatial anisotropy in the available spatial directions. What we know analytically is the geodesic distance between two multivariate Gaussians, A, B , of the same number n of variables in two particular cases [5]:

1. $\mu^A \neq \mu^B, \Sigma^A = \Sigma^B = \Sigma$: $f^A = (n, \mu^A, \Sigma), f^B = (n, \mu^B, \Sigma)$

$$D_\mu(f^A, f^B) = \sqrt{(\mu^A - \mu^B)^T \cdot \Sigma^{-1} \cdot (\mu^A - \mu^B)}. \quad (1)$$

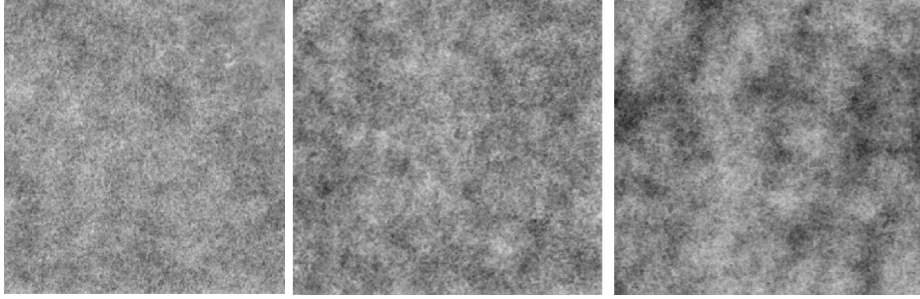


Fig. 2. Areal density radiographs of three paper networks made from natural wood cellulose fibres, with constant mean coverage, $\bar{c} \approx 20$ fibres, but different distributions of fibres. Each image represents a square region of side length 5 cm; darker regions correspond to higher coverage. The left image is similar to that expected for a Poisson process of the same fibres, so typical real samples exhibit clustering of fibres.

$$2. \mu^A = \mu^B = \mu, \Sigma^A \neq \Sigma^B : \quad f^A = (n, \mu, \Sigma^A), f^B = (n, \mu, \Sigma^B)$$

$$D_{\Sigma}(f^A, f^B) = \sqrt{\frac{1}{2} \sum_{j=1}^n \log^2(\lambda_j)}, \quad \text{with } \{\lambda_j\} = \text{Eig}(\Sigma^{A-1/2} \cdot \Sigma^B \cdot \Sigma^{A-1/2}). \quad (2)$$

From the form of $D_{\Sigma}(f^A, f^B)$ in (2) it may be seen that an approximate monotonic relationship arises with a more easily computed symmetrized log-trace function given by

$$\Delta_{\Sigma}(f^A, f^B) =$$

$$\sqrt{\log \left(\frac{1}{2n} \left(\text{Tr}(\Sigma^{A-1/2} \cdot \Sigma^B \cdot \Sigma^{A-1/2}) + \text{Tr}(\Sigma^{B-1/2} \cdot \Sigma^A \cdot \Sigma^{B-1/2}) \right) \right)}. \quad (3)$$

This is illustrated by the plot of $D_{\Sigma}(f^A, f^B)$ from equation (2) on $\Delta_{\Sigma}(f^A, f^B)$ from equation (3) in Figure 3 for 185 trivariate Gaussian covariance matrices. For comparing relative proximity, this is a better measure near zero than the symmetrized Kullback-Leibler distance [8] in those multivariate Gaussian cases so far tested and may be quicker for handling large batch processes.

2 Dimensionality reduction of spatial density arrays

We follow the methods described by Carter et al. [6,7] for datasets of pixel density arrays from complete sampling of density maps of simulated networks. Elsewhere we shall describe the results of such studies on a large collection of radiographs from commercial networks.

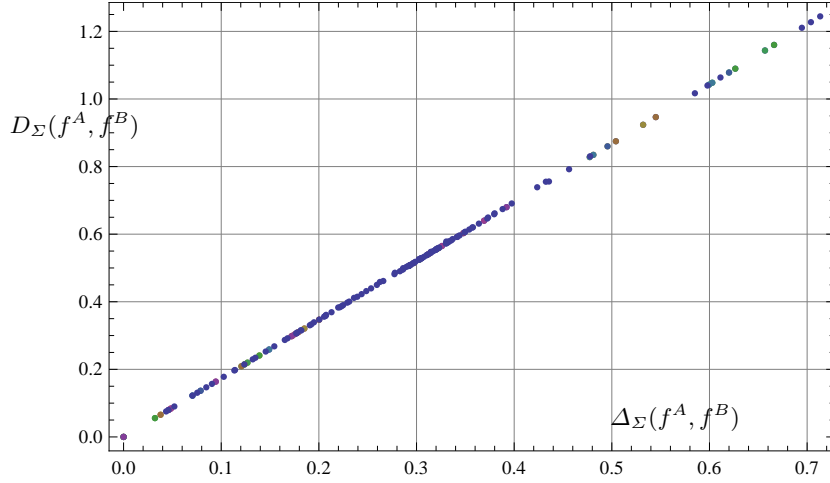


Fig. 3. Plot of $D_{\Sigma}(f^A, f^B)$ from (2) on $\Delta_{\Sigma}(f^A, f^B)$ from (3) for 185 trivariate Gaussian covariance matrices.

Our three spatial variables are central pixel, mean of its first neighbours, and mean of its second neighbours. Figure 4 gives analyses for spatial arrays of pixel density differences from Poisson networks. It shows a plot of $D_{\Sigma}(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for geodesic information distances among 16 datasets of 1mm pixel density differences between a Poisson network and simulated networks made from 1mm fibres. Each network has the same mean density but with different scales and densities of clustering; thus the mean difference is zero in this case. Second row: Dimensionality reduction embedding of the same data grouped by numbers of fibres in clusters and cluster densities. Using pixels of the order of fibre length is appropriate for extracting information on the sizes of typical clusters. The embedding reveals the clustering features as orthogonal subgroups.

Next, Figure 5 gives analyses for pixel density arrays of the clustered networks. It shows the plot of $D_{\Sigma}(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among the 16 datasets of 1mm pixel densities for simulated networks made from 1mm fibres, each network with the same mean density but with different clustering. In this case the trivariate Gaussians all have the same mean vectors. Second row: Dimensionality reduction embedding of the same data grouped by numbers of fibres in clusters and cluster densities; the solitary point is a Poisson network of the same fibres.

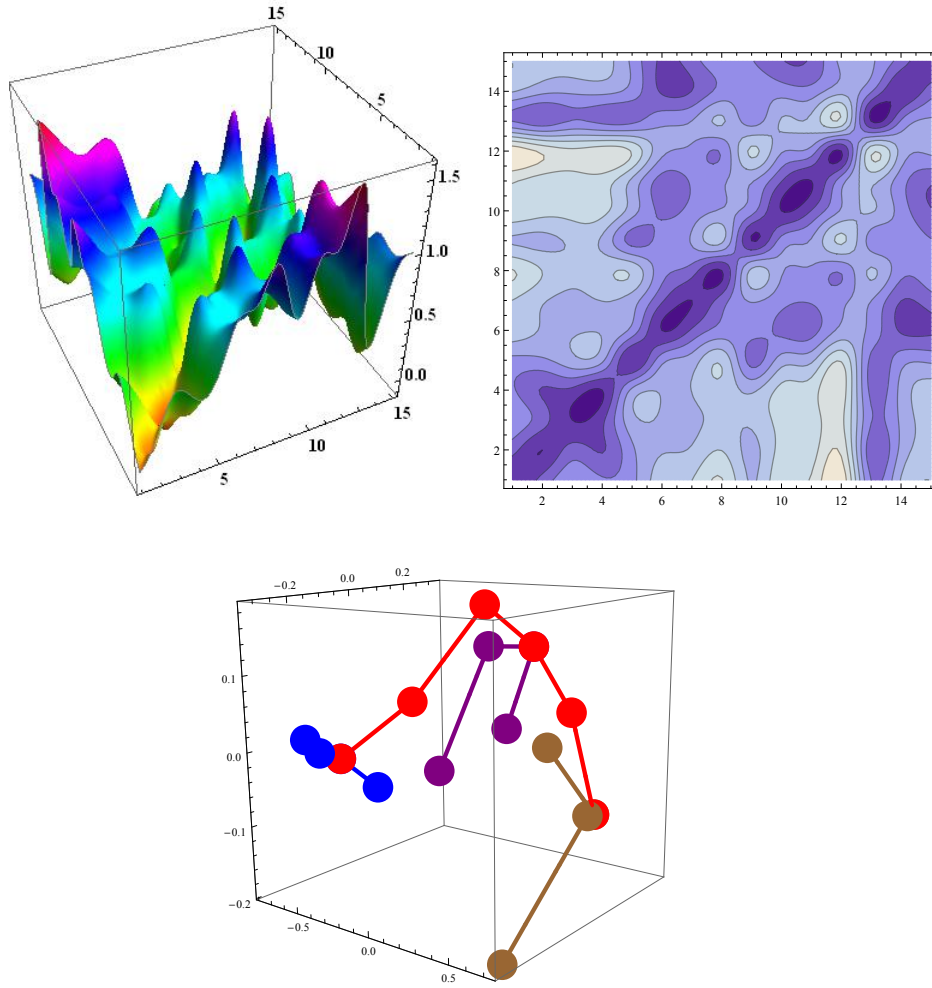


Fig. 4. Pixel density differences from Poisson networks. Top row: plot of $D_{\Sigma}(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 datasets of 1mm pixel density differences between a Poisson network and simulated networks made from 1mm fibres, each network has the same mean density but with different clustering. Second row: Embedding of the same data grouped by numbers of fibres in clusters and cluster densities.

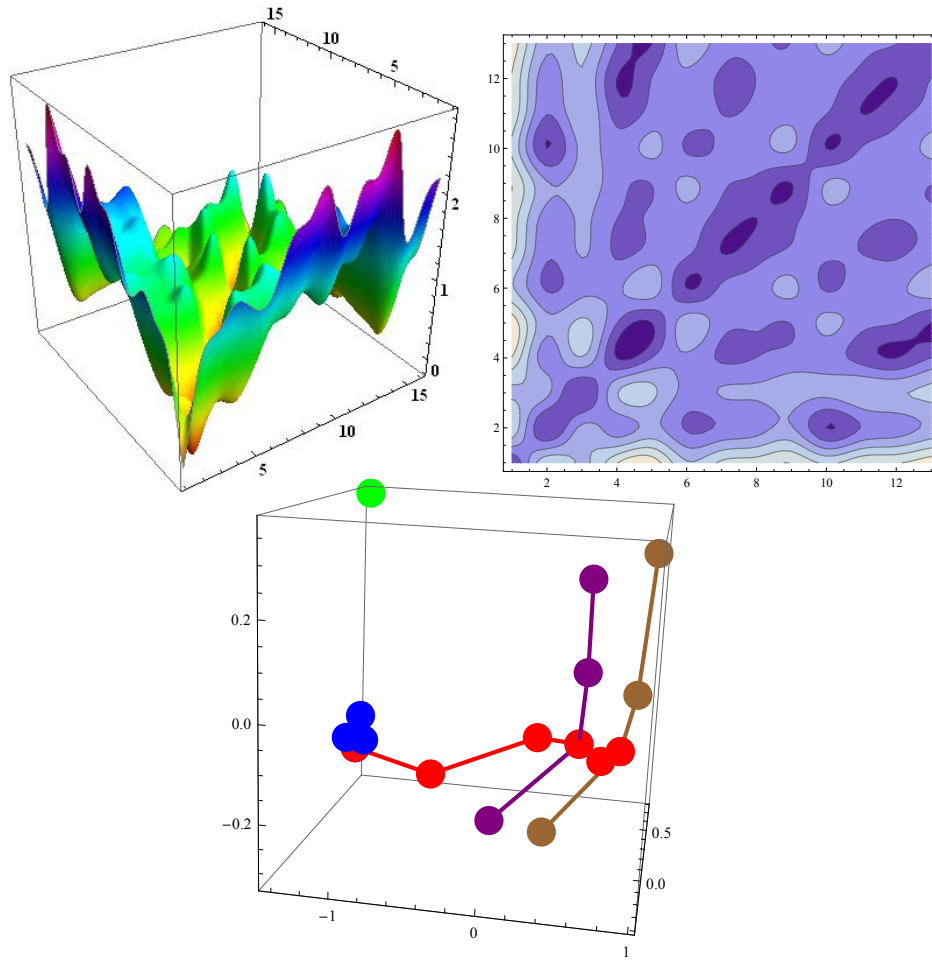


Fig. 5. Pixel density arrays for clustered networks: Top row: plot of $D_{\Sigma}(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 datasets of 1mm pixel density arrays for simulated networks made from 1mm fibres, each network with the same mean density but with different clustering. Second row: Embedding of the same data grouped by numbers of fibres in clusters and cluster densities; the solitary point is an unclustered Poisson network.

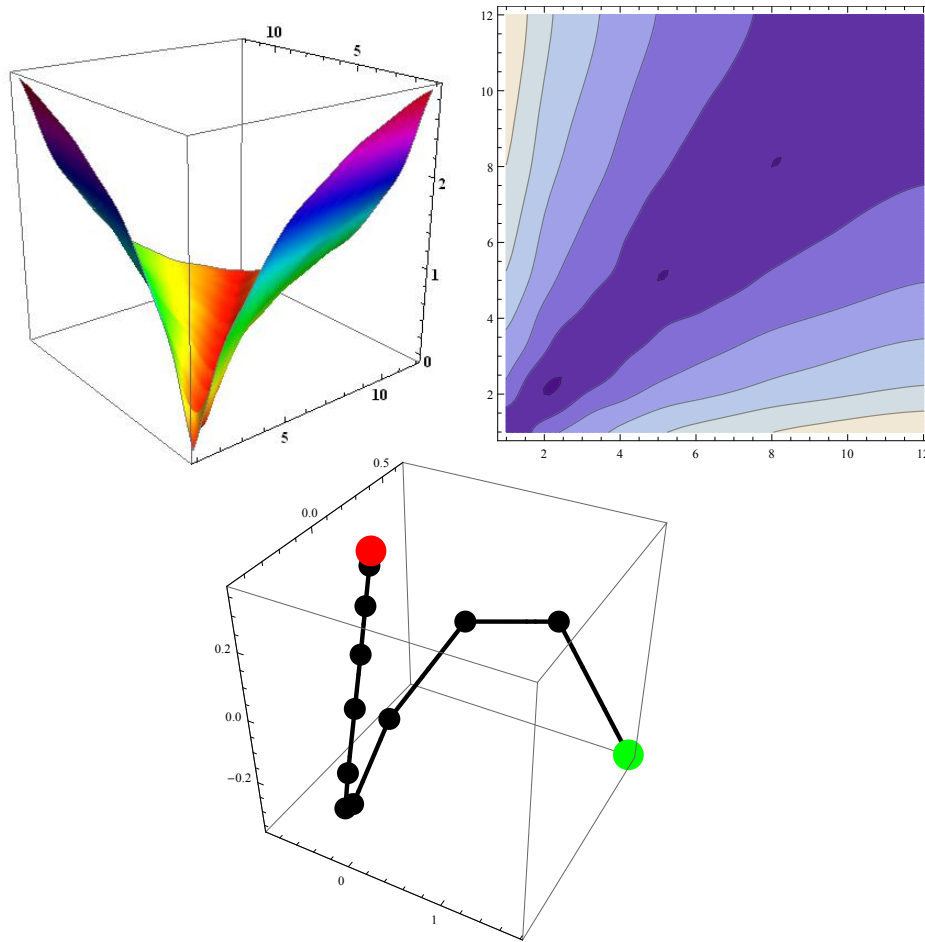


Fig. 6. Pixel density arrays for Poisson networks of different mean density: Top row: plot of $D(f^A, f^B) = D_\mu(f^A, f^B) + D_\Sigma(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 simulated Poisson networks made from 1mm fibres, with different mean density, using pixels at 1mm scale. Second row: Embedding of the same Poisson network data, showing the effect of mean network density.

Figure 6 gives analyses for pixel density arrays for Poisson networks of different mean density. It shows the plot of $D(f^A, f^B) = D_\mu(f^A, f^B) + D_\Sigma(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 simulated Poisson networks made from 1mm fibres, with different mean density, using pixels at 1mm scale. Second row: Dimensionality reduction embedding of the same Poisson network data, showing the effect of mean network density.

The benefit from these analyses is the representation of the important structural features of number of fibres per cluster and cluster density, by almost orthogonal subgroups in the embedding.

References

1. K. Arwini and C.T.J. Dodson. **Information Geometry Near Randomness and Near Independence**. Lecture Notes in Mathematics. Springer-Verlag, New York, Berlin, 2008, Chapter 9 with W.W. Sampson, Stochastic Fibre Networks pp 161-194.
2. M. Deng and C.T.J. Dodson. **Paper: An Engineered Stochastic Structure**. Tappi Press, Atlanta, 1994.
3. W.W. Sampson. **Modelling Stochastic Fibre Materials with Mathematica**. Springer-Verlag, New York, Berlin, 2009.
4. W.W. Sampson. Spatial variability of void structure in thin stochastic fibrous materials. *Mod. Sim. Mater. Sci. Eng.* 20:015008, (2012), pp13.
doi:10.1088/0965-0393/20/1/015008
5. C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: Indian Journal of Statistics* 48, A, 3 (1981) 345-365.
6. K.M. Carter, R. Raich and A.O. Hero. Learning on statistical manifolds for clustering and visualization. In 45th Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, 2007.
<https://wiki.eecs.umich.edu/global/data/hero/images/c/c6/Kmcarter-learnstatman.pdf>
7. K.M. Carter **Dimensionality reduction on statistical manifolds**. PhD thesis, University of Michigan, 2009. <http://tbayes.eecs.umich.edu/kmcarter/thesis>
8. F. Nielsen, V. Garcia and R. Nock. Simplifying Gaussian mixture models via entropic quantization. In Proc. 17th European Signal Processing Conference, Glasgow, Scotland 24-28 August 2009, pp 2012-2016.