

*Dimensionality Reduction for Information  
Geometric Characterization of Surface  
Topographies*

Dodson, CTJ and Mettanen, M and Sampson, WW

2017

MIMS EPrint: **2016.58**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary  
School of Mathematics  
The University of Manchester  
Manchester, M13 9PL, UK

ISSN 1749-9097

# Dimensionality Reduction for Information Geometric Characterization of Surface Topographies

C.T.J. Dodson<sup>1</sup>, M. Mettänen<sup>2</sup>, and W.W. Sampson<sup>3</sup>

<sup>1</sup> School of Mathematics, University of Manchester, Manchester, M13 9PL, UK.

<sup>2</sup> Department of Automation Science & Engineering, Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland.

<sup>3</sup> School of Materials, University of Manchester, Manchester, M13 9PL, UK.

**Abstract.** Stochastic textures with features spanning many length scales arise in a range of contexts in physical and natural sciences, from nanostructures like synthetic bone to ocean wave height distributions and cosmic phenomena like inter-galactic cluster void distributions. Here we used a data set of 35 surface topographies, each of  $2400 \times 2400$  pixels with spatial resolution between  $4 \mu\text{m}$  and  $7 \mu\text{m}$  per pixel, and fitted trivariate Gaussian distributions to represent their spatial structures. For these we computed pairwise information metric distances using the Fisher-Rao metric. Then dimensionality reduction was used to reveal the groupings among subsets of samples in an easily comprehended graphic in 3-space. The samples here came from the papermaking industry but such a reduction of large frequently noisy spatial data sets is useful in a range of materials and contexts at all scales.

**Keywords:** Dimensionality reduction, information metric, surface topography, trivariate Gaussian

## 1 Introduction

Stochastic textures with features spanning many length scales arise in a range of contexts in physical and natural sciences. Whereas the features of interest may differ when considering cosmological scale data for galactic density distributions, from those at the global scale representation of oceanographic temperatures or nanoscale features such as the surface topography of synthetic bone, the common format for the data is as a two-dimensional array, which is typically rendered as an image. In general, the challenge is the extraction of the features of interest which may be obscured within an inherently noisy data set. Since paper is made as a web from the continuous filtration of a stochastic dispersion of cellulose fibres, there is a standard reference structure which can be used: a planar isotropic Poisson process of the given fibres, for which the structure is known [1].

Here, we use information geometry and dimensionality reduction to bypass the extraction of features from textures and instead make a direct assessment of

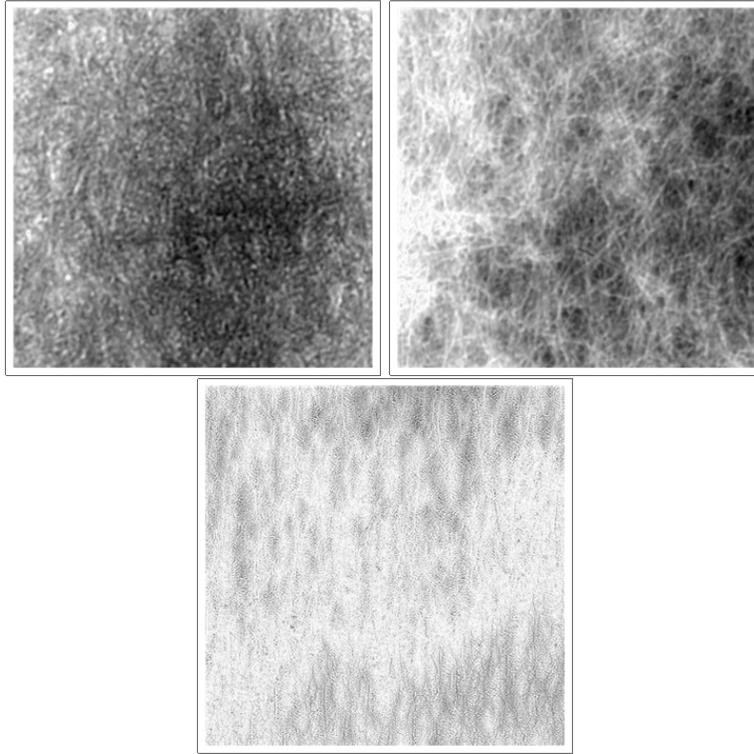
whether they are different or not. We illustrate our approach using the example of two-dimensional stochastic textures arising from measurements of the surface topography of different grades of paper. Whereas paper represents a convenient source of data available with a wide range of surface topographies, it turns out that there is genuine interest in the papermaking industry in characterizing this structural feature of the material and its influence on product performance [2].

Information geometry uses the Fisher information metric on smoothly parametrized families of probability density functions to provide a natural distance structure. Gaussians parametrized by mean and standard deviation yield a 2-dimensional curved surface and bivariate Gaussians yield a 5-dimensional curved space, *cf.* Amari [3]. Thus, the information metric gives an arc length function along any curve between two probability density functions in the given family. The geometry of commonly occurring families of probability density functions is well-known, see [4] for relevant examples. The technical algorithmic difficulty is that, in the curved space of probability density functions, the true information distance between two points is the infimum of arc length taken over all curves joining the points. This infimum is the length of the geodesic between the points.

Materials scientists study the interdependence of the structure and properties of materials and how these may be influenced by manufacturing processes. Typically, the properties of the material are the product specifications for end-use and employed for quality control; examples include mechanical behaviour, thermal or electrical conductivity, permeability, *etc.* Our focus here is identification of differences in structure that may be difficult to identify using conventional data handling methodologies.

We illustrate our approach using measurements of the surface topography of paper. This is a particularly convenient material to study: almost all grades of paper have principally the same chemical structure – they consist of natural cellulose fibres with length of order a millimeter or two and width a few tens of micrometers; sheets may be filled or coated with minerals such as clay or calcium carbonate. Structural variability in paper is observed at scales corresponding to the fibre dimensions and above and, importantly, depends on the fibre dimensions and the manufacturing processes employed; these dependencies are discussed in detail in, *e.g.* [5,6]. Papermakers control global average structural properties to influence the sheet properties for a given application, so, for example, other than weight per unit area, the principle difference between newsprint and bathroom tissue is the density of the sheet. Local average variability in such properties is far more difficult to characterize and control because of the underlying stochastic variability [5] arising from the finite length fibres and influencing the autocovariance function of the planar ensemble.

The stochastic variability of mass of paper, *i.e.* the distribution of local averages of areal density in the plane of the sheet is a fundamental structural property of paper and characterizes the extent to which fibres are clustered. Recently, through analysis of simulated textures representing the distribution of mass, we demonstrated that information geometry could be used to discriminate variability arising from the size and intensity of clusters [7,8].



**Fig. 1.** Graphical representations of sample height data. Top left: coated board; top right: uncoated packaging paper; bottom: bathroom tissue. Each image represent a square of side 1500 pixels with approximate resolution 5  $\mu\text{m}$  per pixel

In what follows we illustrate the differences of features in given data sets obtained from measurements of the surface topography of different paper samples. For each sample, our source information is an array of surface heights, which we process to yield a  $3 \times 3$  covariance matrix  $\Sigma$  and mean vector  $\mu$  arising from pixels and their first and second neighbours. We proceed to use dimensionality reduction to extract the three most significant features from the set of samples so that all samples can be displayed graphically in a 3-dimensional plot. The aim is to reveal groupings of data points that distinguish among and within grades of paper. The method depends on extracting the three largest eigenvalues and their eigenvectors from a matrix of pairwise information distances among distributions representing the samples in the data set. The number of samples in the data set is unimportant, except for the computation time in finding eigenvalues.

## 2 Data sets

Data was acquired as local height values from the surfaces of paper samples using a photometric stereo device; details of the measurement technique are provided in [2]. Data were acquired at different times under subtly different optical conditions; though in all cases we handled arrays of at least  $2400 \times 2400$  pixels with spatial resolution between  $4 \mu\text{m}$  and  $7 \mu\text{m}$  per pixel, which is smaller than the expected width of the constituent fibres. All measurements were made on industrially manufactured paper samples. Measurements were made on 3 groups of samples:

**Group 1: Packaging and printing grades.** Ten samples: coated packaging paper and cardboard; uncoated packaging paper and cardboard; uncoated wrapping grades. Measurements made on both sides of each sample.

**Group 2: Tissue.** Five samples of two-ply bathroom tissue. Measurements made on one side only.

**Group 3: Printing, writing and sack grades.** Five samples: one high quality coated grade and three utility grades for printing and writing; one grade for making paper sacks. Measurements made on one side only; two measurements made of each sample.

Graphical representations of three examples of the surface height distribution are provided in Figure 1. These show three very different surfaces: a coated board surface, an uncoated packaging paper surface and the surface of a bathroom tissue. In the figure, dark regions correspond to low height and *vice versa*; each image represents a square of side 1500 pixels.

## 3 Information geometry model

Each of our source data sets consists of a two-dimensional array of local average height values  $\tilde{h}_i$ . From each of these, we generate two numbers: the average height of the 8 first-neighbour pixels,  $\tilde{h}_{1,i}$  and the average height of the 16 second-neighbour pixels,  $\tilde{h}_{2,i}$ . Thus, we have a trivariate distribution of the random variables  $(\tilde{h}_i, \tilde{h}_{1,i}, \tilde{h}_{2,i})$  with  $\bar{h}_2 = \bar{h}_1 = \bar{h}$  and the marginal distributions of  $\tilde{h}_i$ ,  $\tilde{h}_{1,i}$  and  $\tilde{h}_{2,i}$  are well approximated by Gaussian distributions.

The geodesic distance between two multivariate Gaussians,  $A, B$ , with probability density functions  $f^A, f^B$  mean vectors  $\mu^A, \mu^B$  and covariance matrices  $\Sigma^A, \Sigma^B$  of the *same* number  $n$  of variables is known analytically in two particular cases [9]:

**Common covariance matrix, different mean vectors:**

$\mu^A \neq \mu^B, \Sigma^A = \Sigma^B = \Sigma; f^A = (n, \mu^A, \Sigma), f^B = (n, \mu^B, \Sigma)$

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

**Common mean vector, different covariance matrices:**

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

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

Here we shall take the simplest choice and sum the two components (1) and (2) to give a net measure of distance between two arbitrary n-variate Gaussians  $f^A, f^B$

$$D(f^A, f^B) = \frac{1}{2} (D_{\mu}(f^A, f^B) + D_{\mu}(f^B, f^A)) + D_{\Sigma}(f^A, f^B) \quad (3)$$

where we have to take the average of (1) using  $\Sigma^A$  and  $\Sigma^B$  so (3) gives an upper bound on the true distance.

## 4 Dimensionality reduction

Now, our family of 35 data sets gives us a  $35 \times 35$  symmetric matrix of pairwise information distances between pairs of samples, each sample represented by a trivariate Gaussian distribution. Graphically, we can comprehend a 3-dimensional representation of features so we need a method to reduce the feature representation in our data set of 35 to fit into a 3-dimensional image. Accordingly, we follow the methods described by Carter *et al.* [10,11] to reduce the dimensionality of our data sets and hence identify clustering of data sets with similar topographies through 3-dimensional rendering of the resultant plots. Briefly, we follow a series of computational steps:

1. Obtain pairwise ‘information distances’  $D(i, j)$  among the members of the dataset of textures  $X_1, X_2, \dots, X_N$  characterised by pixel arrays representing height values.
2. The array of  $N \times N$  distances  $D(i, j)$  is a symmetric matrix with diagonal zero. This is centralized by subtracting row and column means and then adding back the grand mean to give  $CD(i, j)$ .
3. The centralized matrix  $CD(i, j)$  is again symmetric with diagonal zero. We obtain its  $N$  eigenvalues  $ECD(i)$ , which are necessarily real, and the  $N$  corresponding  $N$ -dimensional eigenvectors  $VCD(i)$ .
4. Make a  $3 \times 3$  diagonal matrix  $A$  of the first three eigenvalues of largest absolute magnitude and a  $3 \times N$  matrix  $B$  of the corresponding eigenvectors. The matrix product  $A \cdot B$  yields a  $3 \times N$  matrix and its transpose is an  $N \times 3$  matrix  $T$ , which gives us  $N$  coordinate values  $(x_i, y_i, z_i)$  to embed the  $N$  samples in 3-space.

Of course, any pairwise divergence matrix could be used in this situation and might yield different numerical values. However, the qualitative effect will be the same due to a one-to-one monotonic relationship.

## 5 Results

We illustrate the effectiveness of the approach for the surface textures of the three groups of samples described in §2. We consider first an application of the theory from samples in cases when both  $\mu$  and  $\sigma$  are known and then when  $\mu$  is disregarded. We proceed to show the reproducibility of the approach for discrimination among samples and their position when embedded in 3-space. Finally, we examine the influence of applying the algorithm to data subjected to a simple high-pass filter, as applied in conventional image processing of such data.

### 5.1 Sensitivity to mean vector $\mu$

The top row of Figure 2 shows the plot of  $D(f^A, f^B)$  from equation (3) as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among our 35 data sets<sup>4</sup>. On first inspection it is clear that there is structure in the assembled information and the three groups of data can be readily identified from these graphics. The resultant 3-dimensional embedding is shown on the bottom row of Figure 2; here we observe that the data from Groups 1 and 3 occupy a different region from those for Group 2 which is consistent with the observed surface texture of tissue being manifestly different from those of printing, writing and packaging grades of paper. The first 10 eigenvalues arising from the dimensionality reduction are plotted as a bar chart in Figure 3, showing clearly that the majority of the spectral information is captured by the 3-largest eigenvalues, in this case approximately 75 %.

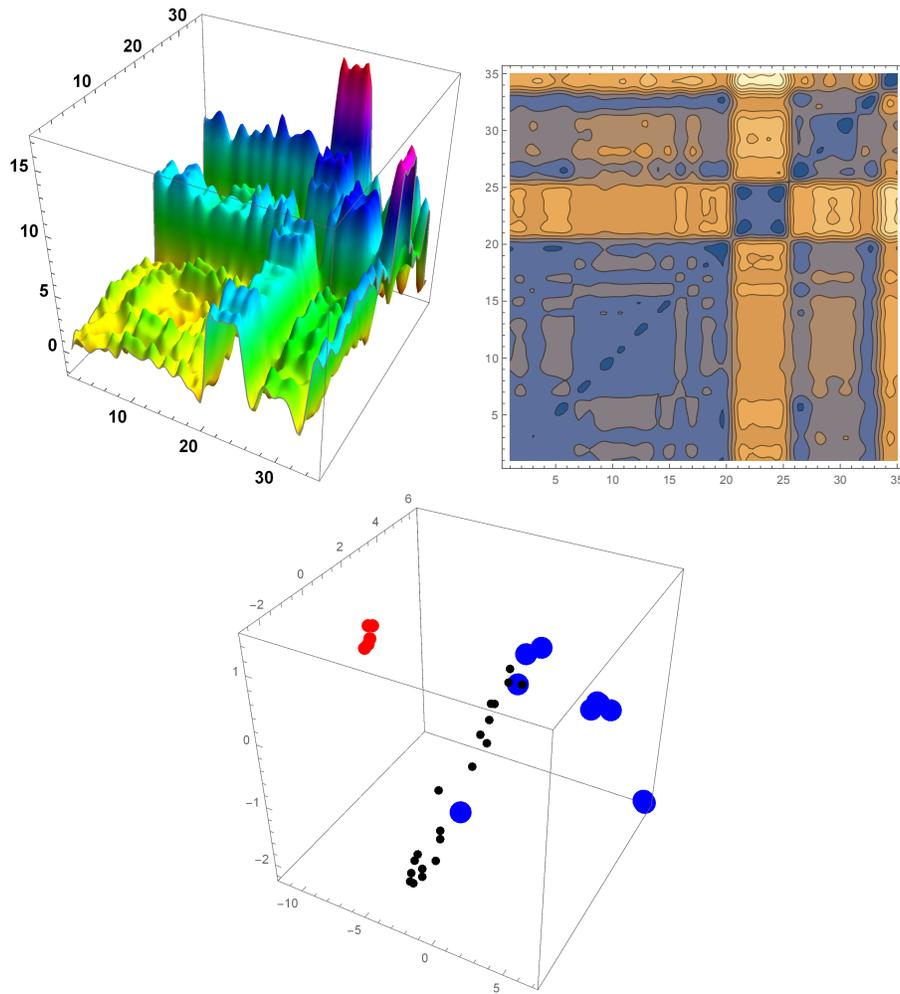
Figure 4 shows graphics corresponding to those in Figure (2) but computed using only the covariances to estimate distances  $D_{\Sigma}(f^A, f^B)$  from equation 2 among samples. We see that for this data set the influence on information distance of the changes in mean are rather minor compared with those of the covariances.

### 5.2 Reproducibility

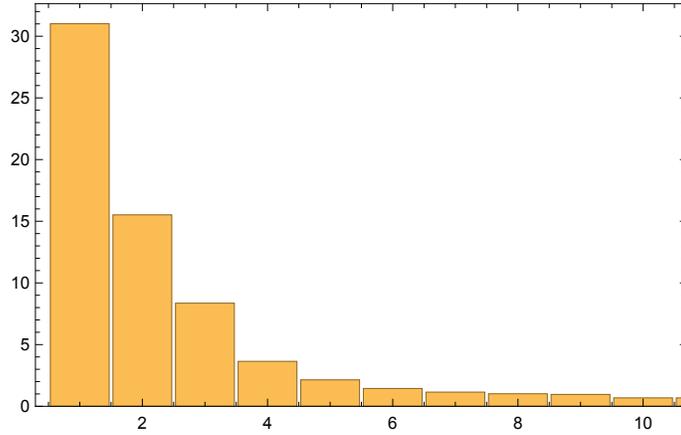
A potential application of the methods we present is the on-line monitoring of change in manufacturing processes. For such applications, repeated sampling and computation of the information distance will yield a surface representing the operating region of the process. Through qualitative and quantitative calibration processes, we might anticipate that data sampled when the process is manufacturing on-specification product would yield embedded data that populate a well-defined region that surface, such that when data fall outside this

---

<sup>4</sup> The small positive values visible in the diagonal in these and subsequent contour plots are an artefact arising from the cubic interpolation.



**Fig. 2.** Top row: Plot of  $D(f^A, f^B)$  from equation (3) as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 35 datasets of surface heights capturing all three data groups. Axes numbering corresponds to data sets: 1–20, Group 1; 21–25, Group 2; 26–35, Group 3. Bottom row: Dimensionality reduction embedding of the same data. Group 1 (small black), Group 2 (medium red), Group 3 (large blue)



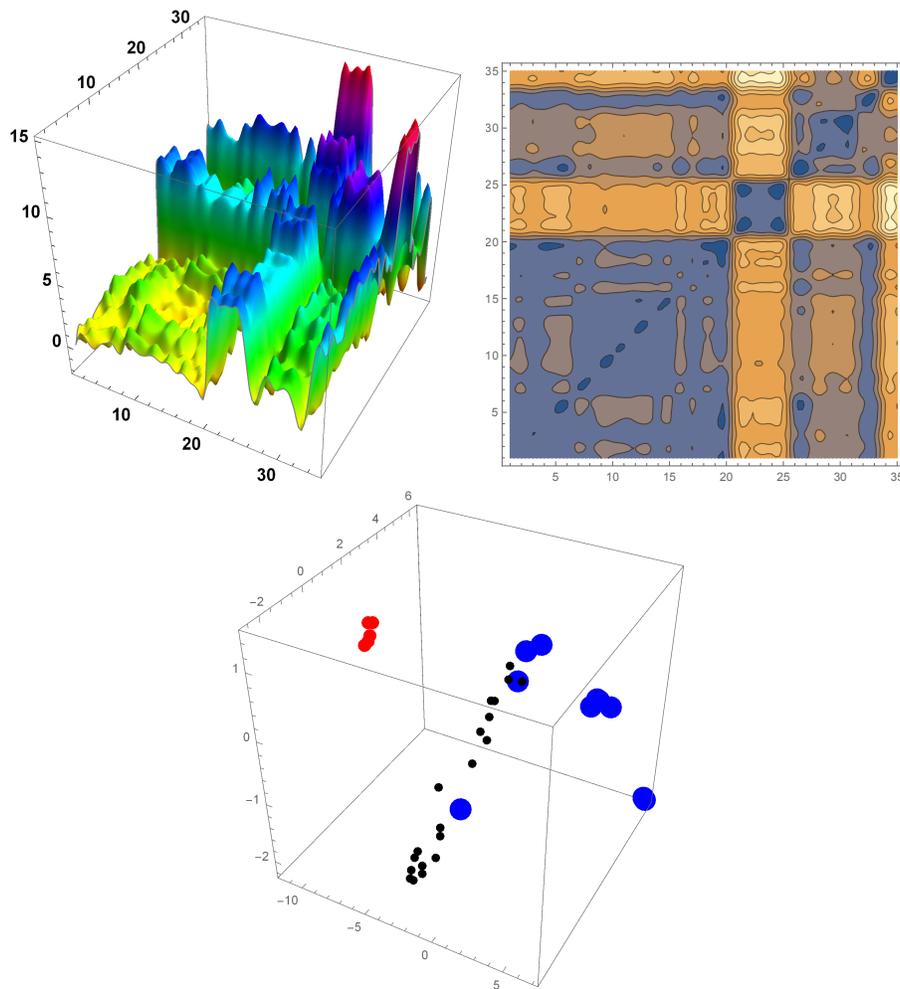
**Fig. 3.** Bar chart of first 10 eigenvalues arising from dimensionality reduction shown in Figure 2. Approximately 75 % of the information is captured by the 3-largest eigenvalues

region, operators may be alerted that the process may have altered to give out-of-specification product.

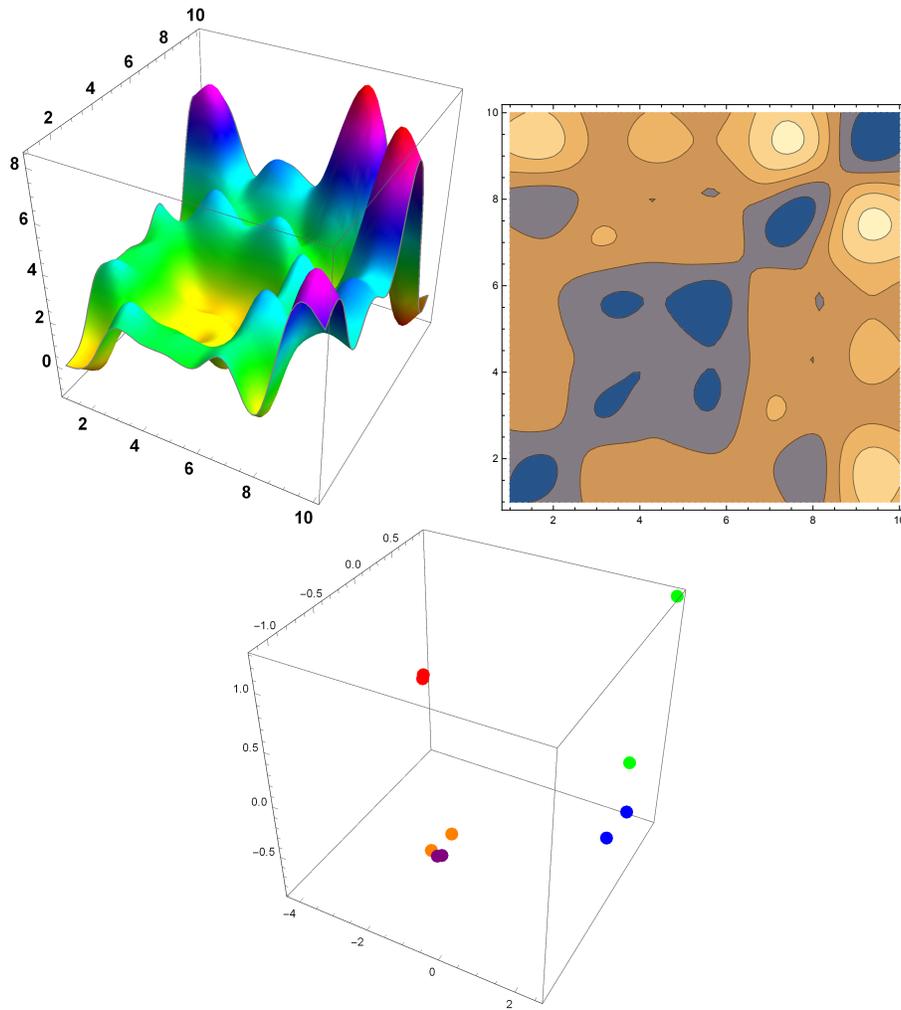
Recall that the data in Group 3 consist of duplicated measurements made on the same side of five different paper specimens. Further, since paper is an inherently stochastic material, we expect some variability from region to region when sampling its surface textures. Accordingly, we use the paired data within Group 3 to investigate the reproducibility of the measurements made on nominally identical samples, which is a prerequisite for on-line monitoring processes of the type proposed.

The plot of  $D(f^A, f^B)$  as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among the 10 data sets representing duplicate measurements from five samples is given in the top row of Figure 5; the resultant 3-dimensional embedding is shown at the bottom of the figure. Again, the first three eigenvalues captured about 75 % of the information

Discrimination among different paper samples within the embedded space is clear and it is noteworthy that the three utility printing and writing grades occupy a different region of the plot from the other grades, which are themselves clearly differentiated. Note that the surface uniformity of these grades are also very different and this manifests itself in the reproducibility of the paired data: for the utility printing and writing grades the two points representing each pair are close; for the high quality coated grade the surface is very smooth and the pair of red points are almost coincident; finally, the structure of sack grade packaging paper is highly non-uniform due to the long fibres used to achieve high mechanical strength and in this case the data points exhibit the greatest separa-



**Fig. 4.** Plot of  $D_{\Sigma}(f^A, f^B)$  from equation (2) as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian covariance only information distances among 35 datasets of surface heights capturing all three data groups. Axes numbering corresponds to data sets: 1–20, Group 1; 21–25, Group 2; 26–35, Group 3. Bottom row: Dimensionality reduction embedding of the same data. Group 1 (small black), Group 2 (medium red), Group 3 (large blue)



**Fig. 5.** Top row: plot of  $D(f^A, f^B)$  from equation (3) as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 10 datasets of surface heights arising from duplicated measurements of the five samples in Group 3. Bottom: Dimensionality reduction embedding of the same data. High quality coated paper (red), sack paper (green), three utility printing & writing grades (orange, purple, blue)

tion, though still occupy a manifestly different region from the those representing the other grades.

### 5.3 Filtered data

It is common in conventional image processing of stochastic data to apply a high-pass filter to two-dimensional data prior to analysis to characterise features of interest, in this case the surface roughness relevant to, *e.g.* printing. Indeed, such processing was applied to the textures described in [2] which were similar to those analysed here. Accordingly, we have applied our treatment to the data in Group 3 after application of a high-pass filter.

Plots corresponding to those in Figure 5 are shown in Figure 6 and we observe a similar quality of discrimination among samples and a similar level of reproducibility, with similar eigenvalue distribution.

Figure 7 shows the effect of using filtered data in comparison to unfiltered data. On the top, the plot combines the embeddings shown at the bottom of Figures 5 and 6; on the bottom the figure shows the embedding obtained by combining the filtered and unfiltered data sets for all sampled to yield a group of 20 arrays (2 filter-states  $\times$  2 repeats  $\times$  5 samples) and computing the information distance  $D(f^A, f^B)$  from equation (3). Note that although the different processes yield different embeddings, each discriminates well between samples and yields good reproducibility, indicating excellent potential for the use of raw, unfiltered and noisy data in on-line monitoring by application of the approach described.

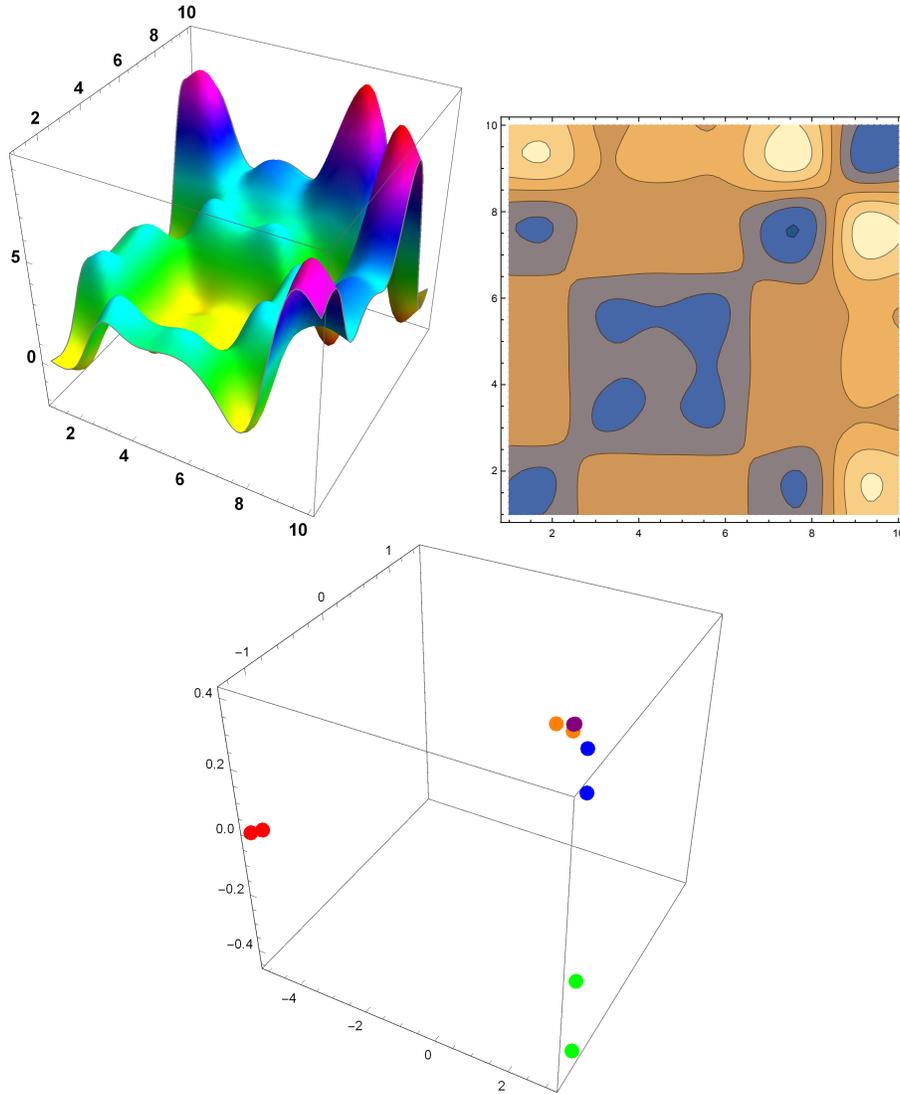
## 6 Random fibre networks

A natural choice of reference structure for the surface of heterogeneous fibrous web-like materials such as paper is a thin network of fibres with uniform orientation and with centres distributed according to a planar Poisson point process [1,5,12]. In [8] §3 we outlined for such structures the analytic derivation of the spatial variance function for local averages  $\tilde{c}$  of the coverage by fibres all of length  $\lambda$  and width  $\omega$ , which tends to a Gaussian random variable. For sampling of the process using, say square inspection pixels of side length  $x$ , the variance of their density  $\tilde{c}(x)$  is the expectation of the point autocorrelation function  $\alpha$

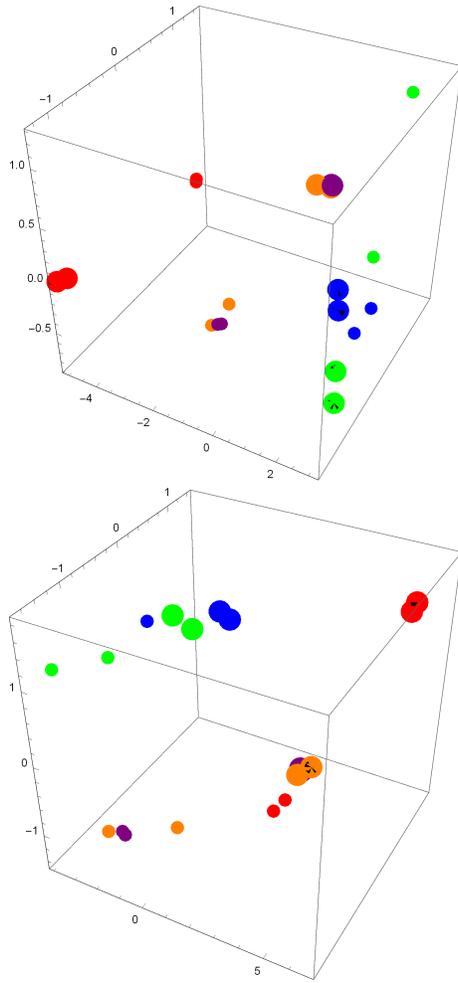
$$Var(\tilde{c}(x)) = Var(c(0)) \int_0^{\sqrt{2}x} \alpha(r, \omega, \lambda) b(r) dr \quad (4)$$

where  $b$  is the probability density function for the distance  $r$  between two points chosen independently and at random in the given type of pixel; it was derived by Ghosh [13].

For practical variance computations we usually have the case of sampling using large square pixels of side  $m x$  say, which themselves consist of exactly  $m^2$  small square pixels of side  $x$ . The variance  $Var(\tilde{c}(m x))$  is related to  $Var(\tilde{c}(x))$



**Fig. 6.** Top row: plot of  $D(f^A, f^B)$  from equation (3) as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 10 datasets of surface heights arising from duplicated measurements of the five samples in Group 3 subjected to a high-pass filter. Bottom: Dimensionality reduction embedding of the same data. High quality coated paper (red), sack paper (green), three utility printing & writing grades (orange, purple, blue)



**Fig. 7.** Combined dimensionality reduction embedding using  $D(f^A, f^B)$  from equation (3) for unfiltered data (small circles) and data subjected to a high-pass filter (large circles). High quality coated paper (red), sack paper (green), three utility printing & writing grades (orange, purple, blue). Top: embedding shown in Figure 5 superimposed on that shown in Figure 6; bottom: information distances & embedding calculated for combined data set of filtered and unfiltered data from Group 3 ( $2 \times 2 \times 5$ ).

through the covariance  $Cov(x, mx)$  of  $x$ -pixels in  $mx$ -pixels [1]:

$$Var(\tilde{c}(mx)) = \frac{1}{m^2}Var(\tilde{c}(x)) + \frac{m^2 - 1}{m^2}Cov(x, mx). \quad (5)$$

As  $m \rightarrow \infty$ , the small pixels tend towards points,  $\frac{1}{m^2}Var(\tilde{c}(x)) \rightarrow 0$  so  $Var(\tilde{c}(mx))$  admits interpretation as  $Cov(0, mx)$ , the covariance among points inside  $mx$ -pixels, the intra-pixel covariance, precisely  $Var(\tilde{c}(mx))$  from equation (4).

Then by rearranging equation 5 the fractional between pixel variance for  $x$ -pixels is

$$\tilde{\rho}(x) = \frac{Cov(0, x)}{Var(c(0))} = \frac{Var(\tilde{c}(x))}{Var(c(0))} \quad (6)$$

which increases monotonically with fibre length  $\lambda$  and with fibre width  $\omega$  but decreases monotonically with  $mx$ , see Deng and Dodson [5] for more details. In fact, for a Poisson process of such rectangles the variance of coverage at points is precisely the mean coverage,  $Var(c(0)) = \bar{c}$ , so if we agree to measure coverage as a fraction of the mean coverage then equation (4) reduces to the integral

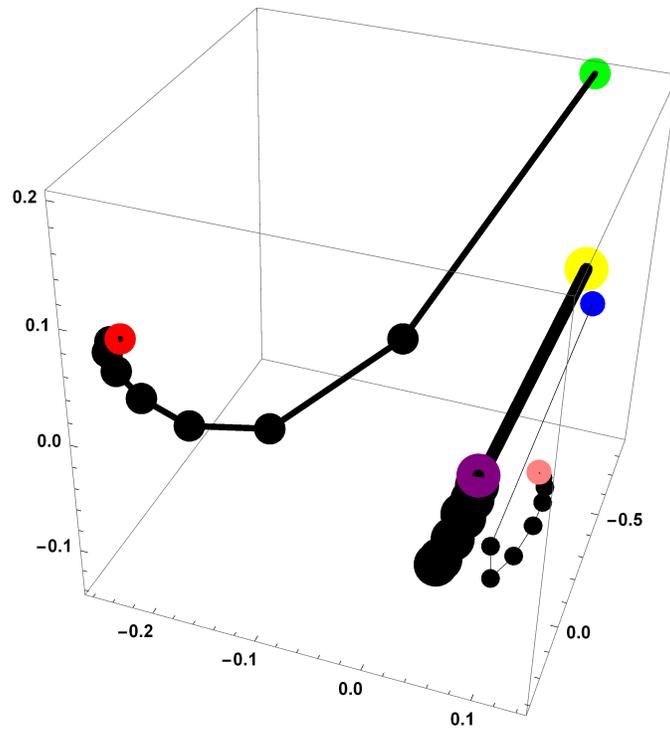
$$\frac{Var(\tilde{c}(x))}{\bar{c}} = \int_0^{\sqrt{2x}} \alpha(r, \omega, \lambda) b(r) dr = \tilde{\rho}(x). \quad (7)$$

The covariance among points inside  $mx$ -pixels,  $Cov(0, mx)$ , is the expectation of the covariance between pairs of points separated by distance  $r$ , taken over the possible values for  $r$  in an  $mx$ -pixel; that amounts to the integral in equation (4). By this means we have continuous families of  $2 \times 2$  covariance matrices for  $x \in \mathbb{R}^+$  and  $2 < m \in \mathbb{Z}^+$  given by

$$\begin{aligned} \Sigma^{x,m} &= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} Var(\tilde{c}(x)) & Cov(x, mx) \\ Cov(x, mx) & Var(\tilde{c}(mx)) \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\rho}(x) & \tilde{\rho}(mx) \\ \tilde{\rho}(mx) & \tilde{\rho}(x) \end{pmatrix}. \end{aligned} \quad (8)$$

which encodes information about the spatial structure formed from the Poisson process of fibres, for each choice of fibre dimensions  $\omega \leq \lambda \in \mathbb{R}^+$ .

The embedding generated by applying equation (2) to compute  $D_\Sigma$  from analytic autocovariance matrices from equation 8 for a planar Poisson process of fibres having width  $\omega = 0.1$ , lengths  $\lambda = 0.5, 1.0, 1.5$ , for square pixels of side length  $x = 0.1, 0.2, \dots, 1.0$ . is shown in Figure 8. We see that fibre length separates the three sets, and in each set the decreasing covariance with increasing pixel size separates the points.



**Fig. 8.** Dimensionality reduction embedding for coverage autocovariances for planar Poisson processes of fibres of lengths  $\lambda = 1.0, 1.5, 2.0$  with width  $\omega = 0.1$ , for sampling with square pixels of side length  $x = 0.1, 0.2, \dots, 1.0$ . For the three increasing fibre lengths, the embeddings have respectively the endpoints, blue to pink, green to red, and yellow to purple, with points and line thicknesses in increasing size.

## References

1. C.T.J. Dodson. Spatial variability and the theory of sampling in random fibrous networks. *J. Roy. Statist. Soc. B* **33**(1):88-94, 1971.
2. M. Mettänen and U. Hirn. A comparison of five optical surface topography measurement methods. *Tappi J.* **14**(1):27-37, 2015
3. S. Amari. **Information Geometry and Its Applications**. *Appl. Math. Sci.* **194**, Springer, Japan, 2016.
4. K. Arwini and C.T.J. Dodson. **Information Geometry Near Randomness and Near Independence**. Lecture Notes in Mathematics. Springer-Verlag, New York, Berlin, 2008.
5. M. Deng and C.T.J. Dodson. **Paper: An Engineered Stochastic Structure**. Tappi Press, Atlanta, 1994.
6. W.W. Sampson. Materials properties of paper as influenced by its fibrous architecture. *Int. Mater. Rev.* **54**(3):134-156, 2009.
7. C.T.J. Dodson and W.W. Sampson. Information geometry and dimensionality reduction for statistical structural features of paper. In **Advances in Pulp and Paper Research, Cambridge, 2013**. *Trans. XVth Fund. Res. Symp.* (S.J. P'Anson, ed.), pp55-69, FRC, Manchester, 2013.
8. C.T.J. Dodson and W.W. Sampson. Dimensionality reduction for classification of stochastic texture images. In **Geometric Theory of Information**, (F. Nielsen, ed.), Signals and Communication Technology Series, Springer International Publishing, Switzerland, 2014.
9. C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: Indian Journal of Statistics* **48A**(3):345-365, 1981.
10. 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>
11. K.M. Carter **Dimensionality reduction on statistical manifolds**. PhD thesis, University of Michigan, 2009.  
<http://tbayes.eecs.umich.edu/kmcarter/thesis>
12. W.W. Sampson. **Modelling Stochastic Fibre Materials with Mathematica**. Springer-Verlag, New York, Berlin, 2009.
13. B. Ghosh. Random distances within a rectangle and between two rectangles. *Calcutta Math. Soc.* **43**(1):17-24, 1951.