Preferred citation: C.T.J. Dodson and W.W. Sampson. Information geometry and dimensional reduction for statistical structural features of paper. In **Advances in Pulp and Paper Research, Cambridge 2013**, *Trans. of the XVth Fund. Res. Symp. Cambridge, 2013*, (S.J. I'Anson, ed.), pp 55–70, FRC, Manchester, 2018. DOI: 10.15376/frc.2013.1.55.

INFORMATION GEOMETRY AND DIMENSIONALITY REDUCTION FOR STATISTICAL STRUCTURAL FEATURES OF PAPER

C.T.J. Dodson

School of Mathematics
University of Manchester
Manchester, M13 9PL, UK.

W.W. Sampson

School of Materials
University of Manchester
Manchester, M13 9PL, UK.

ABSTRACT

Information geometry provides a metric on spaces of probability density functions. Here we apply it to the space of trivariate Gaussian distributions of joint variation among the areal density variables for pixels and their first and second neighbours, from radiographs and simulations. At a pixel scale of one millimetre these distributions can pick up essential structural features including flocculation intensity and scale. We do this by applying the technique of dimensionality reduction to large mixed data sets of samples and the results show promise for classification, including extraction of groupings that represent different former types. This kind of analysis could be valuable in evaluating trials, comparing different installations of similar formers and for identifying anomalous behaviour.

Keywords: Fibre networks, radiographic images, spatial covariances, simulations, information metric, dimensionality reduction, trial evaluation, former comparison.

1 INTRODUCTION

Much analytic work has been done on modelling of the statistical geometry of heterogeneous fibrous materials such as paper [1, 2]. The inherent non-uniformity of paper influences its mechanical properties [3, 4], fluid ingress or transfer [5, 6, 7] and optical uniformity [8, 9].

The basic property of interest in applications is the uniformity in grammage, or formation. Several quantitative measures of the formation of paper, are reported and widely used in the literature. Experimentally an array of local average grammage values is obtained via a calibrated radiographic technique; for comparison of imaging methods see [10]. Using complete sampling by contiguous square inspection zones, the distribution of local average grammage is typically well represented by a Gaussian distribution of variance that decreases monotonically with increasing inspection zone size [1]; the rate of decay is dependent on fibre and fibre cluster dimensions, see, *e.g.* [11]. Several other techniques have been applied to characterise formation: Norman *et al.* [12, 13, 14] used frequency domain image processing to obtain power spectra allowing quantification of the contribution of structural features of different scales to the overall variability; Jordan, with Nguyen and others [15, 16, 17, 18] used the specific perimeter as a measure of floc size; Bouydain *et al.* [19] and Keller *et al.* [20, 21, 22] use the wavelet transform to capture the spatial variability of non-uniformity within and between regions of a given texture; Farnood *et al.* [23] developed a stochastic decomposition approach to yield characteristic floc sizes and grammages; this approach is developed in the proceedings of this symposium [24]. Direct mappings, albeit in some cases highly non-trivial ones, exist among all these measures [25, 26].

For a random network of fibres, *i.e.* one where the fibre centres are distributed according to a point Poisson process in the plane with uniform distribution of fibre axes to any arbitrary direction, the power spectrum, specific perimeter and variance of local grammage are known analytically, see [27], [15] and [28] respectively. Noting the direct relationships among all these quantifiers of formation, we focus now on the theory for the variance of local grammage, since this provides an appropriate framework for discussion of our subsequent treatment. We denote the local average of a random variable by a tilde ($\widetilde{\ }$) and the global average by a bar ($\overline{\ }$). Dodson [28] derived the spatial covariance for arbitrary rectangular fibres and gave the variance of local grammage, $\sigma_x^2(\widetilde{\beta})$

observed in a random fibre network partitioned into contiguous square zones of side x as

$$\sigma_x^2(\tilde{\beta}) = \sigma^2(\beta) \int_0^{\sqrt{2}x} \alpha(r)\, b(r,x)\, \mathrm{d}r \quad , \tag{1}$$

where $\sigma^2(\beta)$ is the variance of grammage at points, which is readily determined from the underlying Poisson process of coverage, $\alpha(r)$ is the point autocorrelation function for fibres, *i.e.* the probability that a fibre covering a given point in the plane covers also another point separated by a distance r from the first; $b(r, x)$ is the probability density of separation, r, of pairs of randomly chosen points within a square zone of side x. Full derivations of expressions for $\alpha(r)$ and $b(r, x)$ are provided in [29].

At most scales of inspection, x, the variance of local grammage in real paper is greater than that of a random network formed from the same constituent fibres [1]. We have then

$$\sigma_x^2(\tilde{\beta}) = \sigma^{*2}(\beta) \int_0^{\sqrt{2}x} \alpha^*(r)\, b(r,x)\, \mathrm{d}r \quad , \tag{2}$$

where $\sigma^{*2}(\beta)$ is the point variance of the clustered, *i.e.* non-Poissonian, process of fibre coverage, and $\alpha^*(r)$ is the point autocorrelation function for fibre clusters, or *flocs*. Farnood *et al.* [23] modelled flocs as disks with fixed grammage, G and provide expressions for a the autocorrelation function of a random process of disks. For now, pertinent to our subsequent analysis we make three remarks about the autocorrelation function $\alpha^*(r)$:

- it is a global average property of the sampled area;
- it characterises the underlying texture of a grammage map;
- covariance matrices among pixel densities are integrals of $\alpha^*(r)$.

Comparative quantifiers of formation, where the measured formation is compared to that of a corresponding random fibre network, are useful, since they provide natural absolute measures of the extent to which formation might be improved through process interventions [1, 2]. However, all the approaches identified so far seem to rely on second-order statistics. Here, we quantify differences in formation through analysis of distributions that incorporate spatial covariance of local grammage, by means of *information geometry*.

Information geometry uses a natural distance structure, the Fisher information metric [30], on smoothly parametrized families of probability density functions. Gaussians parametrized by mean and standard deviation yield a 2-dimensional

curved surface, bivariate Gaussians yield a 5-dimensional curved space and trivariate Gaussians yield a 9-dimensional curved space. 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 [30] for relevant examples and some applications. 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 does exist and is the length of the shortest curve, called a geodesic, between the points; on a sphere the geodesic (literally 'divides the Earth') is a great circle.

Accordingly, information geometry can be readily applied to the smooth families of the distributions that arise in characterizing paper, to illustrate and metrize the evolution of structure during the manufacturing process under changes of conditions or constituent mixtures of fibres and their clustering properties. Further, it provides a means to quantify the proximity of a process to a natural reference state in the family of probability density functions—*e.g.* the corresponding Poisson process.

In what follows we shall illustrate the differences of features in given data sets obtained from the distribution of local grammage of real samples and simulated paper structures. In such cases there is benefit in mutual information difference comparisons of samples in the set but the difficulty is often the large number of samples in a set of interest—perhaps a hundred or more. The problem is how to present the information contained in the whole data set, each sample yielding a 3×3 covariance matrix Σ and mean μ. The optimum presentation is to use a 3-dimensional plot, but the question is what to plot on the axes.

To solve this problem we 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 correspond to the particular characteristics; in our context we have different former types, grades and differing intensities of flocculation. Such a methodology has particular value in the quality control for processes with applications that frequently have to study large data sets of samples from a trial or through a change in conditions of manufacture or constituents. Moreover, it can reveal anomalous behaviour of a former or unusual deviation in a product. The raw data of one sample from a study of formation might typically consist of a spatial array of 250×250 pixel density values, so what we solve is a problem in classification for stochastic image textures. The method, which we introduced in a preliminary report [33], depends on extracting the three largest eigenvalues and their eigenvectors from a matrix of mutual information distances among distributions representing the samples in the data set. The number in the data set is unimportant, except for the computation time in finding eigenvalues.

2 BACKGROUND THEORY

The covariance of a pair of random variables, p and q is a measure of the degree of association between them and is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \bar{p}\,\bar{q}\ . \tag{3}$$

In particular, the covariance of a variable with itself is its variance. From the array of local average grammage values $\widetilde{\beta}_i$, we generate two numbers associated with each: the average grammage of the 6 first-neighbour pixels, $\widetilde{\beta}_{1,i}$ and the average grammage of the 16 second-neighbour pixels, $\widetilde{\beta}_{2,i}$. Thus, we have a trivariate distribution of the random variables ($\widetilde{\beta}_i$, $\widetilde{\beta}_{1,i}$, $\widetilde{\beta}_{2,i}$) with $\overline{\beta}_2 = \overline{\beta}_1 = \overline{\beta}$. Figure 1 provides an example of a typical data set obtained from a radiograph of a commercial newsprint sample; the histogram and three-dimensional scatter plot show data obtained for pixels of side 1 mm. From the Central Limit Theorem, we expect the marginal distributions of $\widetilde{\beta}_i$, $\widetilde{\beta}_{1,i}$ and $\widetilde{\beta}_{2,i}$ to be well approximated by Gaussian distributions. For the example in Figure 1, these Gaussians are represented by the solid lines on the histogram; this Gaussian approximation holds for all samples investigated in this study. Accordingly, the approach we apply here is to 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, *i.e.* flocs.

What we know analytically is the geodesic distance between two multivariate Gaussians, A, B, with probability density functions f^A,f^B mean vectors μ^A, μ^B and covariance matrices Σ^A, Σ^B of the *same* number n of variables in two particular cases [34]:

- **Common covariance matrix, different mean vectors:** $\mu^A \neq \mu^B$, $\Sigma^A = \Sigma^B = \Sigma$: $f^A = (n, \mu^A, \Sigma), f^B = (n, \mu^B, \Sigma)$

$$D_\mu(f^A, f^B) = \sqrt{(\mu^A - \mu^B)^T \cdot \Sigma^{-1} \cdot (\mu^A - \mu^B)}. \tag{4}$$

- **Common mean vector, different covariance matrices:** $\mu^A = \mu^B = \mu$, $\Sigma^A \neq \Sigma^B$: $f^A = (n, \mu, \Sigma^A), f^B = (n, \mu, \Sigma^B)$

$$D_\Sigma(f^A, f^B) = \sqrt{\frac{1}{2}\sum_{j=1}^{n} \log^2(\lambda_j)}, \quad \text{with } \{\lambda_j\} = \mathrm{Eig}(\Sigma^{A^{-1/2}} \cdot \Sigma^B \cdot \Sigma^{A^{-1/2}}). \tag{5}$$

From the form of $D_\Sigma(f^A, f^B)$ in (5) and recalling that the trace is the sum of the eigenvalues, it may be seen that an approximate monotonic relationship

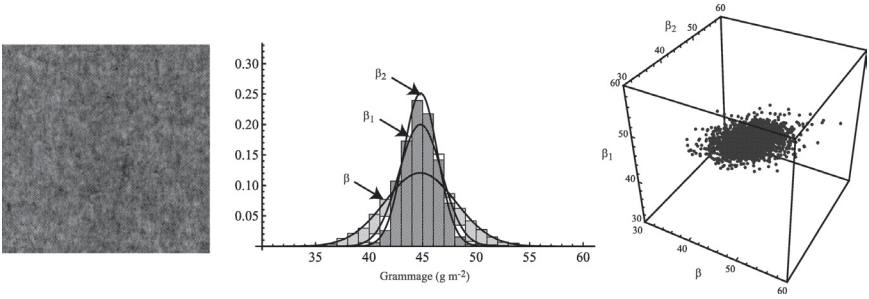

Figure 1. Trivariate distribution of grammage values for newsprint sample. Left: source grammage map; centre: histogram of $\widetilde{\beta}_{i}$, $\widetilde{\beta}_{1,\,i}$ and $\widetilde{\beta}_{2,\,i}$; right: 3D scatter plot of $\widetilde{\beta}_{i}$, $\widetilde{\beta}_{1,\,i}$ and $\widetilde{\beta}_{2,\,i}$.

arises with a more easily computed symmetrized log-trace function given by $\Delta_{\Sigma}(f^{A},f^{B}) =$

$$\sqrt{\log\left(\frac{1}{2n}\left(Tr(\Sigma^{A^{-1/2}}\cdot\Sigma^{B}\cdot\Sigma^{A^{-1/2}})+Tr(\Sigma^{B^{-1/2}}\cdot\Sigma^{A}\cdot\Sigma^{B^{-1/2}})\right)\right)}. \quad (6)$$

This is illustrated by the plot of $D_{\Sigma}(f^{A}, f^{B})$ from (5) on $\Delta_{\Sigma}(f^{A}, f^{B})$ from (6) in Figure 2 for 185 trivariate Gaussian covariance matrices. For comparing relative proximity, this is a better measure near zero than the traditional symmetrized Kullback-Leibler approximate information distance [37] in those multivariate Gaussian cases so far tested and would be quicker than equation (5) for handling large batch processes.

3 DIMENSIONALITY REDUCTION FOR DATA SETS

It is common in situations where large data sets are analyzed to compare among themselves the objects, typically digital images or documents, in order to identify clustering into groups, trends in prominent features or anomalies. The data set of interest here is a family of trivariate Gaussian distributions obtained from textures characterising the distribution of local grammage from each sample; the trends or features are, for example, the mean grammage, intensity and size of flocs. Although current analyses and simulations permit larger arrays and higher resolutions, we constrain our analysis to 250×250 pixel arrays of local grammage values with spatial resolution around 200 μm per pixel. This is what was generated in the

University of Toronto Archives [31], which provides data sets of 200 samples made on a wide range of formers. First, we illustrate the approach by analysis of a family of simulated textures.

Our formation simulator was described in detail at the previous symposium [32], so here we summarize the main aspects of the algorithm. The simulator generates a grammage map as an array of square pixels, which can be considered to correspond to the pixels obtained from a calibrated scanned image of a contact β-radiograph.

The code works by dropping clusters of fibres within a circular region where the centre of each cluster is distributed as a point Poisson process in the plane and the number of fibres per cluster, n_c, is a Poisson distributed random variable. The size of each cluster is determined by an intensity parameter, $0 < I \leq 1$ such that the mean mass per unit area of the cluster is constant and less than the grammage of a fibre. Denoting the length and width of a fibre by λ and ω respectively, the radius of a cluster containing n_c fibre centres is

$$r = \sqrt{\frac{n_c \lambda \omega}{\pi I}} \,. \tag{7}$$

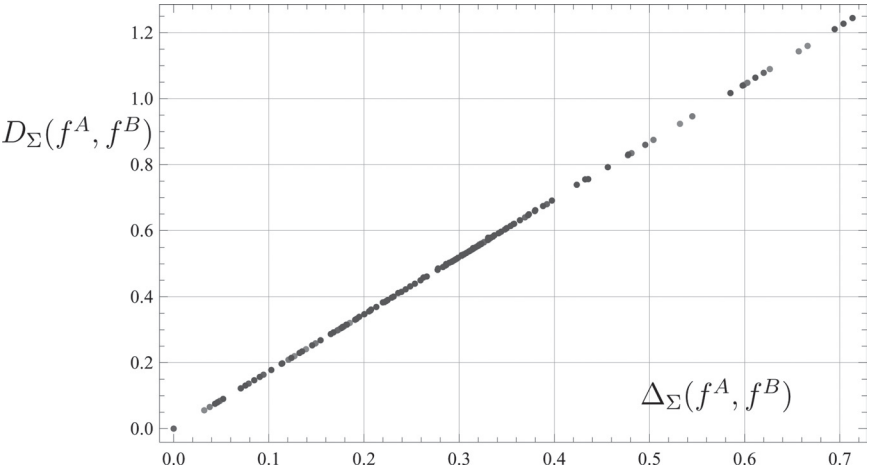

Figure 2. Plot of information distance $D_\Sigma(f^A,f^B)$ from (5) on $\Delta_\Sigma(f^A,f^B)$ from (6) for 185 trivariate Gaussian covariance matrices. This shows an almost linear relationship but the approximation $\Delta_\Sigma(f^A,f^B)$ is quicker to compute and may be useful in large analyses.

Figure 3 shows examples of density maps generated by the simulator. We observe textures that increase in 'cloudyness' with n_c and increase in 'graininess' with I.

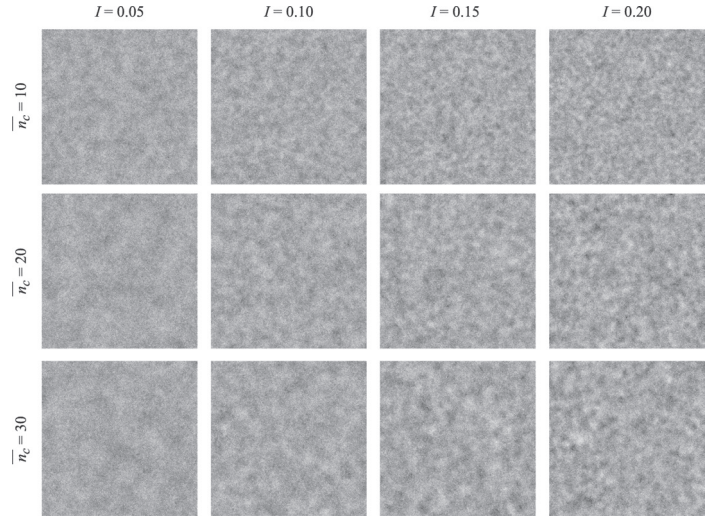

Figure 3. Simulated grammage maps each representing a 4 cm × 4 cm region with mean grammage 60 g m^{-2} formed from fibres with length $\lambda = 1$ mm, linear density $\delta = 2 \times 10^{-7}$ kg m^{-1} and width $\omega = 20$ μm.

Now, a family of 100 samples would give us a 100×100 symmetric positive definite matrix of mutual 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 100 to fit into a 3-dimensional image. Human brains can do this rather well, since we have enormous numbers of optical sensors that stream information from the eyes into the brain but the result is a 3-dimensional reduction which serves to help us 'see' the external environment. We want to see our whole data set organised in such a way that natural groupings are revealed and quantitative dispositions among groups are preserved. There is a convenient and widely applied mathematical method that can help us achieve what we want, it is called multi-dimensional scaling, or dimensionality reduction, see Carter *et al.* [35, 36] for a detailed presentation with examples. Briefly, the series of computational stages is as follows:

1. Obtain mutual 'information distances' $D(i, j)$ among the members of the data set of textures $X_1, X_2, \ldots, X_N$ each with 250×250 pixel density values.
2. The array of $N \times N$ differences $D(i, j)$ is a symmetric positive definite matrix with zero diagonal. 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 positive definite with diagonal zero. We compute its N eigenvalues $ECD(i)$, which are necessarily real, and find the N corresponding N-dimensional eigenvectors $VCD(i)$.
4. Make a 3×3 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.

4 RESULTS

First we illustrate with univariate Gaussian distributions of pixel densities in a small number of simulated networks of straight fibres. Figure 4, left, shows the dimensional reduction embedding of 12 data sets of 250×250 pixels of side 0.2 mm from simulated fibre networks with differing levels and intensities of clustering at the fixed mean 60 g m^{-2}. Here the source data for the analysis consisted of the matrix of areal density pixel differences from a Poisson fibre network with no clustering. These individual distributions of pixel differences

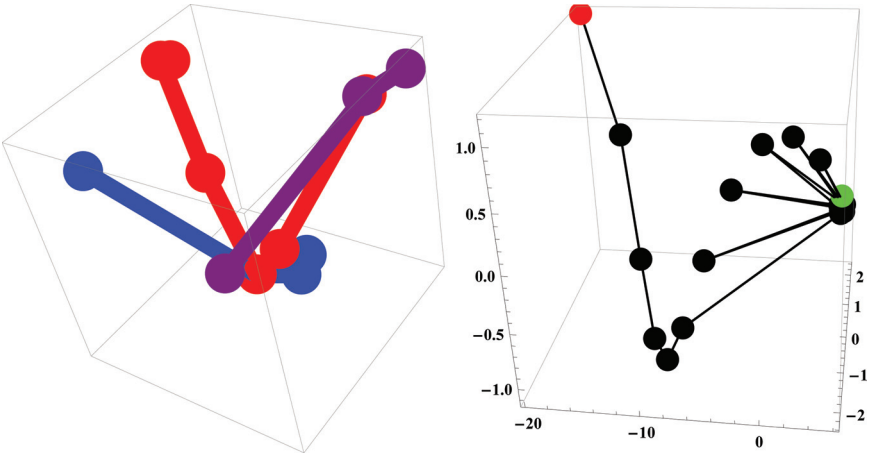

Figure 4. Embedding in 3 dimensions via univariate Gaussian Fisher distances. Left: data set of 0.2mm pixels for 12 simulated networks made from 1mm fibres with differing levels and intensities of clustering at the same mean areal density. Right: the same clustered networks, together with 12 unclustered Poisson fibre networks with differing mean areal densities using univariate Gaussian information distances. The clustered networks are those in a fan from the green point.

were all very close to univariate Gaussians centred on zero and we used distances in the Gaussian Fisher metric. The right diagram in Figure 4 uses data from the same clustered networks, together with 12 unclustered Poisson fibre networks with differing mean areal densities from 5 g m^{-2} up to 60 g m^{-2} at the red point. The clustered networks are those in a fan from the green point.

Figure 5 uses a trivariate Gaussian fitting to the pixel differences from a Poisson network. It shows a 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 data sets of 1 mm pixel differences between a Poisson network and simulated networks made from 1 mm 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.

Next, Figure 6 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 16 data sets of 1 mm pixels for simulated networks made from 1 mm fibres, each network with the same mean density but with different clustering. Using pixels of the order of fibre length is appropriate for extracting information on the sizes of typical clusters. Second row: Dimensionality reduction embedding of the same data grouped by numbers of fibres in clusters and cluster densities; the solitary point is an unclustered Poisson network.

Figure 7 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 1 mm fibres, with different mean density, using pixels at 1 mm scale. Second row: Dimensionality reduction embedding of the same Poisson network data, showing the effect of mean network density. The benefit from this kind of analysis is the representation of the important structural features of number of fibres per cluster and cluster density, by almost orthogonal sequences in the embedding.

Figure 8 shows a 3-dimensional embedding for a data set from [31] including 182 paper samples from gap formers, handsheets, pilot machine samples and hybrid formers. We see that to differing degrees the embedding separates these different and very disparate forming methods by assembling them into subgroups. This kind of discrimination could be valuable in evaluating trials, comparing different installations of similar formers and for identifying anomalous behaviour.

Elsewhere we shall report on analyses using finer structure than 1 mm pixels and more detailed interpretation of flocculation behaviour.

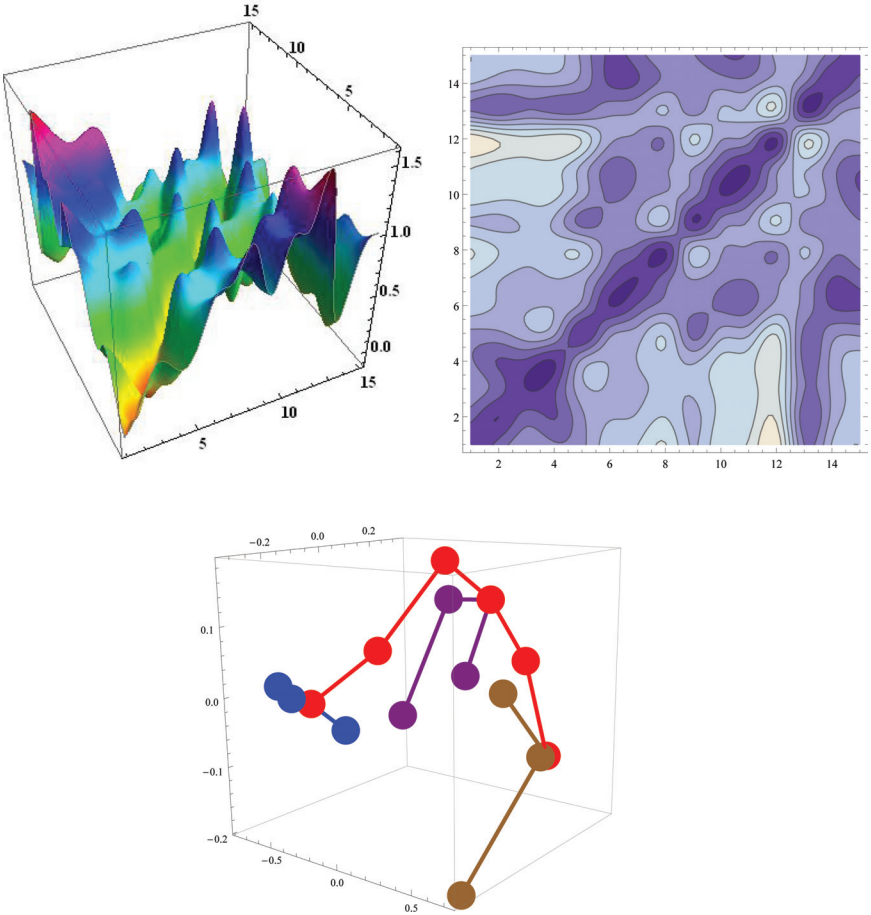

Figure 5. Top row: plot of information distance $D_{\Sigma}(f^{A}, f^{B})$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 data sets of 1mm pixel differences between a Poisson network and simulated networks made from 1mm fibres, 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.

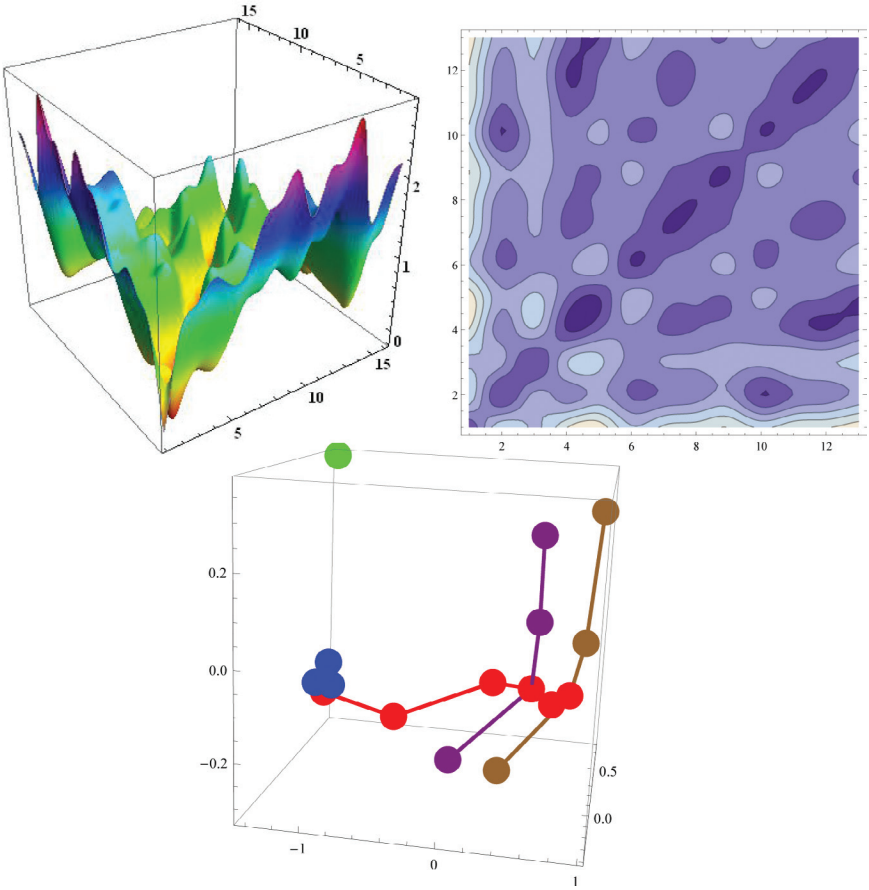

Figure 6. Top row: plot of information distance $D_{\Sigma}(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 data sets of 1mm pixels for simulated networks made from 1mm fibres, each network with the same mean density 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.

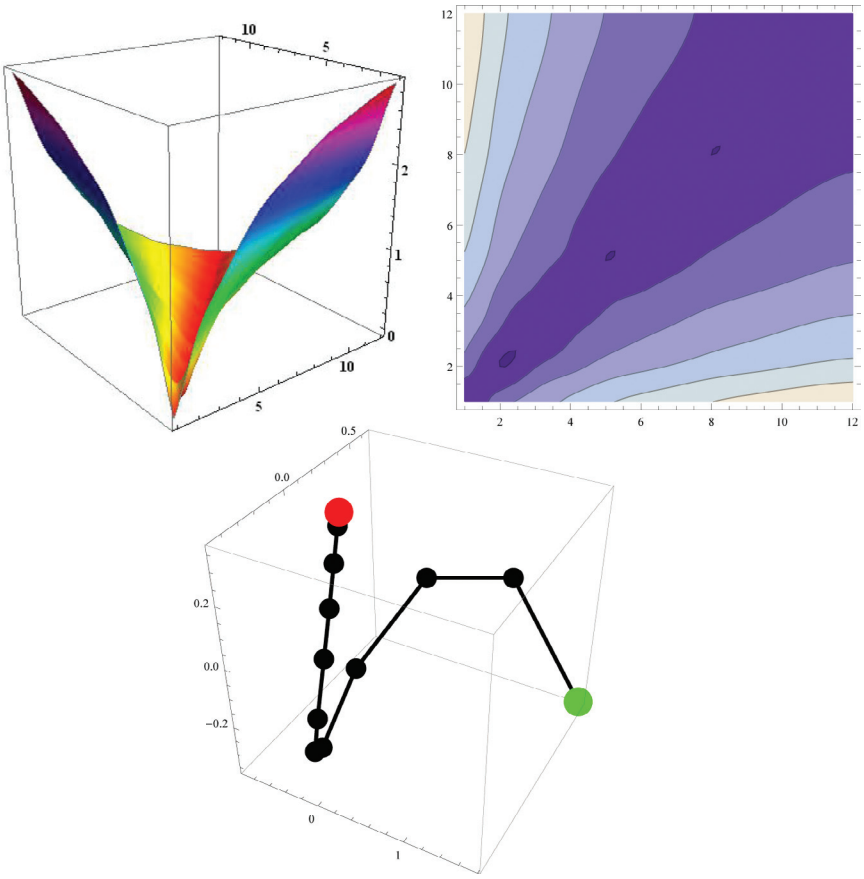

Figure 7. Top row: plot of information distance $D(f^A, f^B) = D_\mu(f^A, f^B) + D_\Sigma(f^A, f^B)$ as a cubic-smoothed surface (left), and as a contour plot (right), for trivariate Gaussian information distances among 16 simulated Poisson networks made from 1 mm fibres, with different mean density, using pixels at 1 mm scale. Second row: Embedding of the same Poisson network data, showing the effect of mean network density increasing from the green point to the red point.

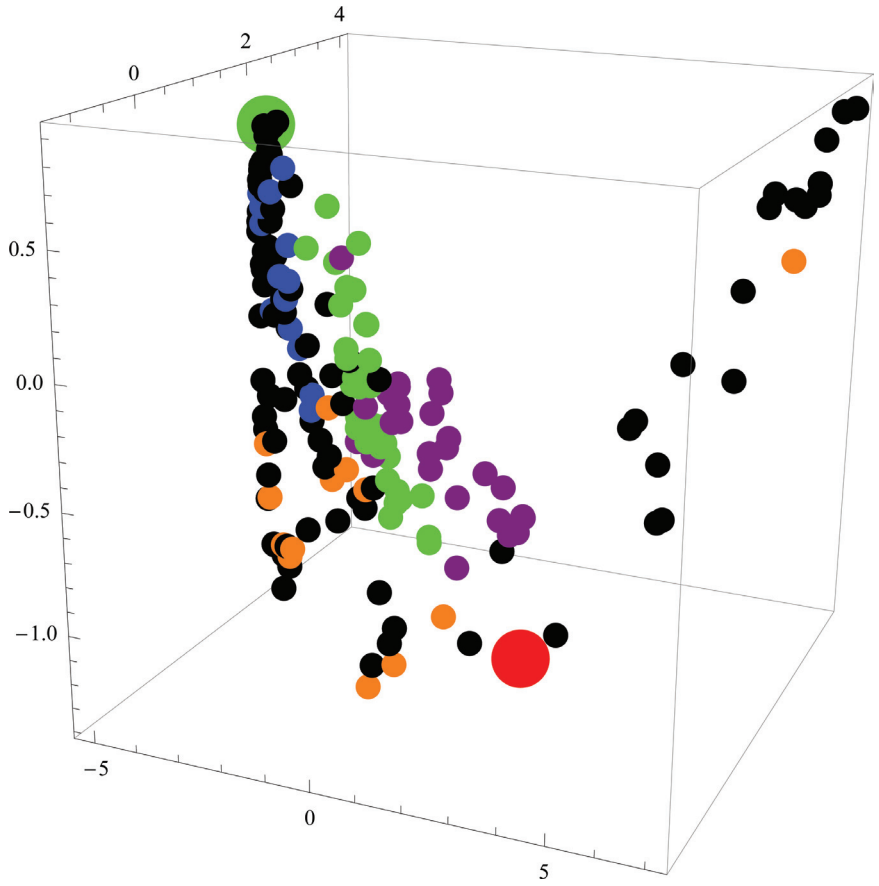

Figure 8. Embedding using 182 trivariate Gaussian distributions for samples from the data set [31]. Blue points are from gap formers; orange are various handsheets, purple are from pilot paper machines and green are from hybrid formers. The embedding separates these different forming methods into subgroups.

REFERENCES

1. M. Deng and C.T.J. Dodson. **Paper: An Engineered Stochastic Structure**. Tappi Press, Atlanta, 1994.
2. W.W. Sampson. The structural characterisation of fibre networks in papermaking processes – A review. In **The Science of Papermaking**, *Trans. XIIth Fund. Res. Symp.* (ed. C.F. Baker), pp1205–1288, Pulp and Paper Fundamental Research Society, Bury, 2001.

3. R. Wathén and K. Niskanen. Strength distributions of running paper webs. *J. Pulp Pap. Sci.* **32**(3):1–8, 2006.
4. D.T. Hristopulos, T. Uesaka. Structural disorder effects on the tensile strength distribution of heterogeneous brittle materials with emphasis on fiber networks. *Phys. Rev. B.***70**(6):064108, 2004.
5. Y. Termonia. Permeability of sheets of nonwoven fibrous media. *Chem. Eng. Sci.* **53**(6):1203–1208, 1998.
6. E. Schweers and F. Löffler. Realistic modelling of the behaviour of fibrous filters through consideration of filter structure. *Powder Tech.* **80**(3):191–206, 1994.
7. V. Koivu, M. Decain, C. Geindreau, K. Mattila, J. Alaraudanjoki, J.-F. Bloch and M. Kataja. Flow permeability of fibrous porous materials. In **Advances in Pulp and Paper Research, Oxford 2009**. *Trans. XIVth Fund. Res. Symp.* (S.J. I'Anson, ed.), pp437–454, FRC, Manchester, 2009.
8. A. Komppa and K. Ebeling. Correlation between the areal mass and optical densities in paper. In **The Role of Fundamental Research in Papermaking**, *Trans. VIIth Fund. Res. Symp.* (J. Brander, ed.), pp603–633, Mechanical Engineering Publications, London, 1983.
9. N. Dooley and W.W. Sampson. Opportunities for improved light transmission formation analysis. *Pap. Tech.* **43**(8):31–36, 2002.
10. D.S. Keller and J.J. Pawlak. Analytical technique for the comparison of paper formation imaging methods. *J. Pulp Pap. Sci.* **27**(5):171–176, 2001.
11. C.T.J. Dodson and W.W. Sampson. In **Advances in Pulp and Paper Research, Oxford 2009**. *Trans. XIVth Fund. Res. Symp.* (S.J. I'Anson, ed.), pp665–691, FRC, Manchester, 2009.
12. B. Norman and D. Wahren. A comprehensive method for the description of mass distribution in sheets and flocculation and turbulence in suspensions. *Svensk Papperstidn.* **75**(20):807–818, 1972.
13. L. Nerelius, B. Norman and D. Wahren. Measurement of flocculation characteristics by light reflection. *Tappi J.* **55**(4):574–580, 1972.
14. B. Norman and D. Wahren. Mass distribution and sheet properties of paper. In **The Fundamental properties of Paper Related to its Uses**, *Trans. Vth Fund. Res. Symp.* (F. Bolam, ed.), pp7–70, BPBIF, London, 1973.
15. B.D. Jordan and N.G. Nguyen. Specific perimeter – a graininess parameter for formation and print mottle textures. *Pap. ja Puu* **68**(6–7):476–482, 1986.
16. B.D. Jordan and N.G. Nguyen. Graininess and formation ranking. *J. Pulp Pap. Sci.* **14**(2):J42–J43, 1988.
17. M. Laleg and N. Nguyen. The specific perimeter of fibre flocs – A criterion of paper formation. *J. Pulp Pap. Sci.* **21**(10):J356–J361, 1995.
18. R.J. Trepanier, B.D. Jordan and N.G. Nguyen. Specific perimeter: a statistic for assessing formation and print quality by image analysis. *Tappi J.* **81**(10):191–196, 1998.
19. M. Bouydain, J.F. Colom and J. Pladellorens. Using wavelets to determine paper formation by light transmission analysis. *Tappi J.* **82**(7):153–158, 1999.
20. D.S. Keller, J. Lewalle and P. Luner. Analysis of paper variability using the continuous wavelet transform. *Pap. ja Puu* **81**(6):440–446, 1999.

21. D.S. Keller, J. Lewalle and P. Luner. Wavelet analysis of simulated paper formation. *Pap. ja Puu* **81**(7):499–505, 1999.
22. M. Kellomäki, J.J. Pawlak, Y.-J. Sung and D.S. Keller. Characterization of nonstationary structural uniformities in paper. In **The Science of Papermaking**, *Trans. XIIth Fund. Res. Symp.* (C.F. Baker, ed.), pp1313–1342, Pulp and Paper Fundamental Research Society, Bury, 2001.
23. R.R. Farnood, C.T.J. Dodson and S.R. Loewen. Modelling flocculation. Part I: Random disk model. *J. Pulp Pap. Sci.* **21**(10):J348–J356, 1995.
24. R.R. Farnood, D. Ngai, and N. Yan.Revisiting the random disk model: Decomposition of paper structure using using randomly deposited disks with arbitrary size distributions. *Trans. XVth Fund. Res. Symp., Cambridge, 2013*, FRC, Manchester, 2013.
25. I.M. Kajanto, A. Komppa and R.K. Ritala. How formation should be measured and characterised. *Nord. Pulp Pap. Res. J.* **4**(3):219–228, 1989.
26. S.J. I'Anson and W.W. Sampson. Determination of spatial domain formation statistics using the Fast Fourier Transform. *Pap. ja Puu.* **85**(7):403–408, 2003.
27. L. Haglund, B. Norman and D. Wahren. Mass distribution in random sheets – theoretical evaluation and comparison with real sheets. *Svensk Papperstidn*. **77**(10):362–370, 1974.
28. 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.
29. W.W. Sampson. **Modelling Stochastic Fibrous Materials with *Mathematica***. Springer-Verlag, London, 2009.
30. K. Arwini and C.T.J. Dodson. **Information Geometry Near Randomness and Near Independence.** Lecture Notes in Mathematics. Springer-Verlag, New York, Berlin, 2008.
31. C.T.J. Dodson, W.K. Ng and R.R. Singh. **Paper: Stochastic Structure Analysis Archive**. Pulp and Paper Centre, University of Toronto 1995, 3 CDs.
32. C.T.J. Dodson and W.W. Sampson. Structural invariance of stochastic fibrous networks. In Proc. 14th Fundamental Research Symposium, Oxford, 13–18 September 2009, ISBN 97809545272 66 pp 665–691.
33. C.T.J. Dodson and W.W. Sampson. Dimensionality reduction for classification of stochastic fibre radiographs. Proc. **GSI2013 Geometric Science of Information**, Paris, 28–30 August 2013.
34. C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: Indian Journal of Statistics* **48**(A3):345–365, 1981.
35. 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
36. K.M. Carter **Dimensionality reduction on statistical manifolds**. PhD thesis, University of Michigan, 2009. http://tbayes.eecs.umich.edu/kmcarter/thesis
37. F. Nielsen, V. Garcia and R. Nock. Simplifying Gaussian mixture models via entropic quantization. In Proc. 17*th* European Signal Processing Conference, Glasgow, Scotland 24–28 August 2009. pp 2012–2016.
38. H. Corte. Formation and the distribution of mass density in paper. *The Paper Maker* International Number:81–85, 1971.

Transcription of Discussion

INFORMATION GEOMETRY AND DIMENSIONALITY REDUCTION FOR STATISTICAL STRUCTURAL FEATURES OF PAPER

C.T.J. Dodson[1] *and W.W. Sampson*[2]

[1] School of Mathematics, University of Manchester, Manchester, M13 9PL, UK
[2] School of Materials, University of Manchester, Manchester, M13 9PL, UK

Steve I'Anson FRC Chairman (from the chair)

In figure 8 in the proceedings, most points are identified by coloured markers. There are colours for different formers, forming methods etc. What do the black points represent?

Bill Sampson

These are the data that we grouped as "miscellaneous". There are some data which are not well classified: odd pairs of handsheets, a pair of vertiformer samples, that kind of thing. They're not sufficiently grouped to see a cluster; many are pairs of points only and it would be nice to drill down further into it. One of the problems with the archive that these data come from[5] is that, although we know a lot about what is there, we don't know any detail. For example, many are labelled as "pilot machines" but it would be very interesting to know which pilot machine. Which sample was from the Manchester machine? Which one from Grenoble, which from FEX in Stockholm? They are probably all in there and that's actually not captured in the archive.

[5]C.T.J. Dodson, W.K. Ng and R.R. Singh, "Paper: Stochastic Structure Analysis Archive", Pulp and Paper Centre, University of Toronto 1995, 3 CDs.

Discussion

Kit Dodson University of Manchester (co-author)

Bill, can I start by answering the question that you raised, if we go back to the density level and figure 7 in the paper? The reason the points pack closely on those bends is that they are lying on a cubic surface and that is where the curvature is highest. The information geometry extracts important features and groups like qualities.

The other point I wanted to make was to emphasize one of your statements. This isn't just a very exciting piece of theory and quite new technique coming from a thesis completed only four years ago, it is used, for example, also in radar work for tracking of aircraft. It is used in meteorology, wave height analysis and so on, and so it has a great variety of applications in many areas. But in the contexts that interest people here, it can be a diagnostic tool, as Bill said, but it can also allow you to identify anomalies in the behaviour, for example, of a particular former or of a process. There are other ways of applying it, and one would be to surface texture which has similar statistical features that can be analysed in the same way. The real trick, and why it's used for big data in very large scale analyses, is that you can reduce to view in three dimensions any number of points. In a sense this is what the eyes do; we have tens of thousands of sensors picking up light, telling us things about the external world, and the brain very cleverly gives us a three-dimensional picture. The process is something like a dimensional reduction.

Pierre Caulet Munksjö

You did not mention fillers. Did you work on purely cellulose sheets?

Bill Sampson

There will be fillers in the commercial sheets. The algorithm is not sensitive to what gives rise to a given texture, instead it seeks to characterize it. So, it does not matter if the grammage distribution is arising from non-uniform distribution of filler or non-uniform distribution of fibre, or indeed from the mass distribution of a coated sheet. If you wanted to look at whether you had a change in your process that affected the uniformity of distribution of filler in the plane of the paper then this technique should be able to detect it.

Ramin Farnood University of Toronto

Thank you Bill for a very interesting talk. You might have mentioned it in your presentation, but did you examine the effect of fibre dimensions and, if so, what

would you expect would be the change in the diagram which shows the random sheet behaviour in this three dimensional space?

Bill Sampson

No, we've not done that yet. Part of the issue is that here we've analysed mainly the simulations that we already had, although some were specifically performed for this project. We originally developed the simulations to look at second order statistics. We ran our simulator to look at the effect of fibre dimensions to ensure that we matched the analytical results, which we did. Thereafter we focused on clustering of fibres with one morphology. Given time, and a pair of hands, we will look at this again using the new technique.

Wolfgang Bauer Graz University of Technology

Did you try to compare the method to information obtained with spectra derived from wavelets or FFT? And did you use your method on other materials?

Bill Sampson

We have been using it on other materials. We do not need to consider other spectra because the algorithm works on the raw 2D grammage map. So the distribution of local grammages at the base pixel resolution is our information, and then we look for the distances among pairs of those values. We have looked at other materials, and other stochastic sequences, and the preliminary results are very encouraging.