Information geometry and dimensional reduction for statistical structural features of paper

C.T.J. Dodson
School of Mathematics

W.W. Sampson
School of Materials

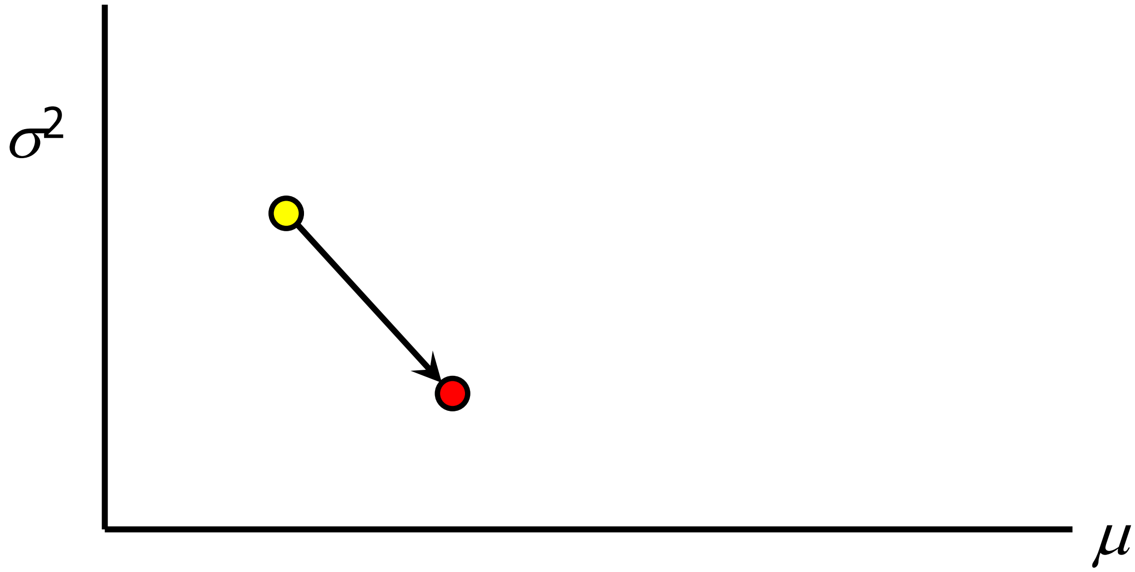
University of Manchester

Overview

- Information geometry: the idea
- Formation
- Information geometry: formation maps
  - Simulations
  - Archive data
- Summary

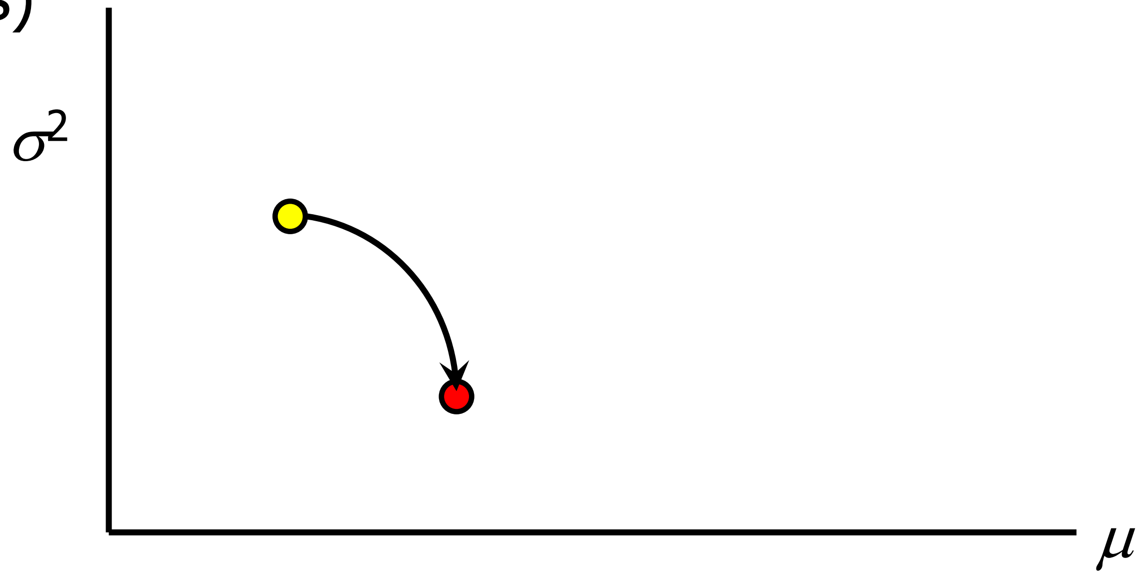
Information geometry: overview

- Application of methods from geometry to problems of statistics;
- Attempts to answer the question, "how far apart are two distributions?"
- *e.g.* Gaussian (μ, σ^2): Euclidian distance between two distributions has no `natural' statistical significance.

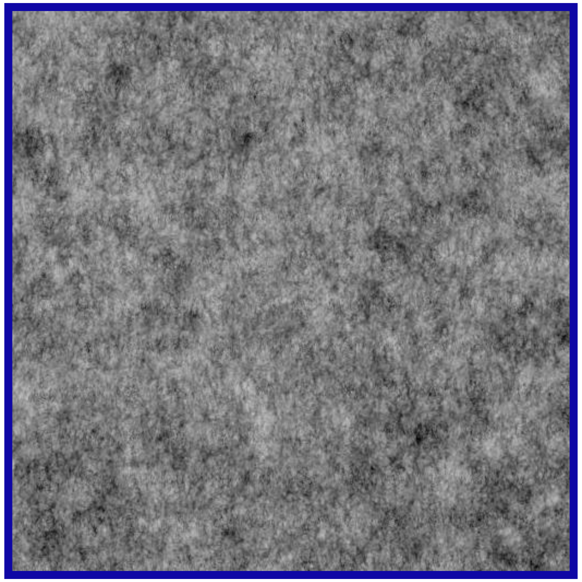


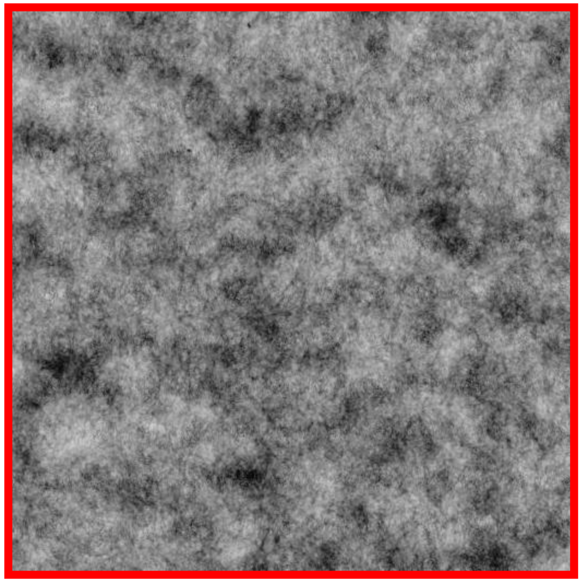
Information geometry: overview

- Information geometry seeks first the shape of the (multidimensional) surface
- Once the surface is known the shortest curve between two points representing the distributions is the 'natural' metric, *i.e.* it has statistical meaning
- For our two Gaussians, the surface is curved. (*cf.* great circles)



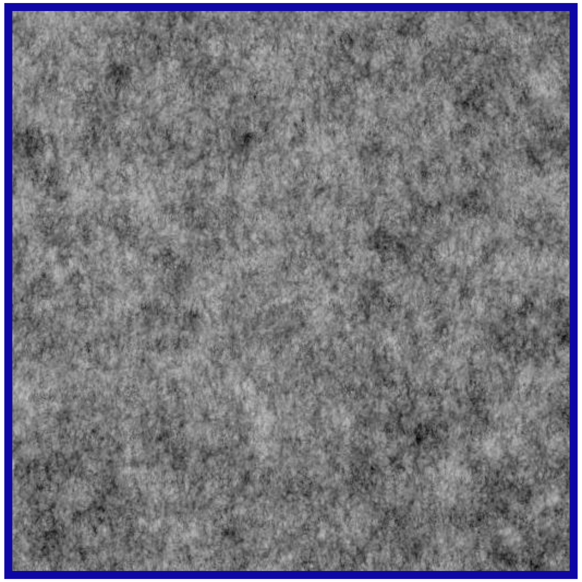
Formation

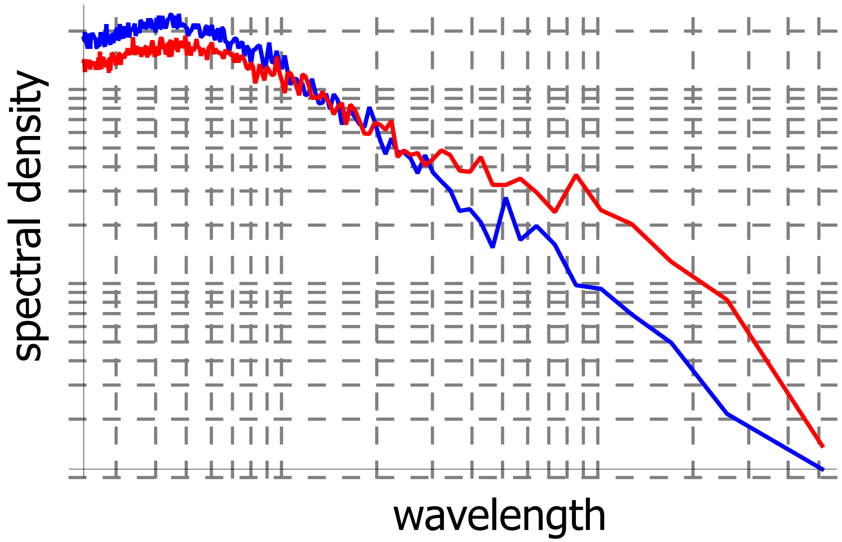


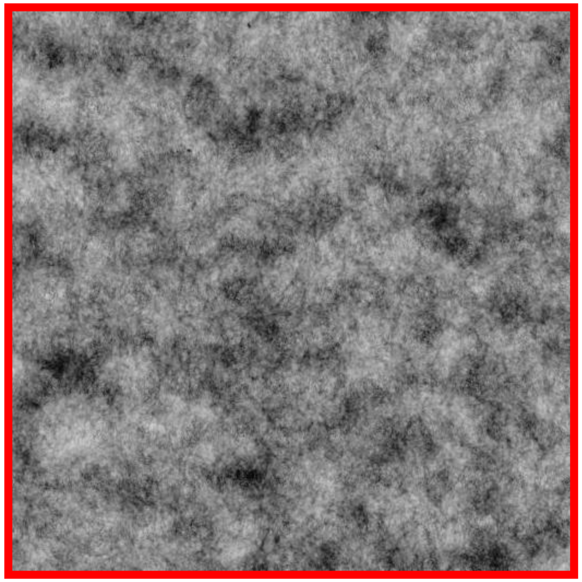


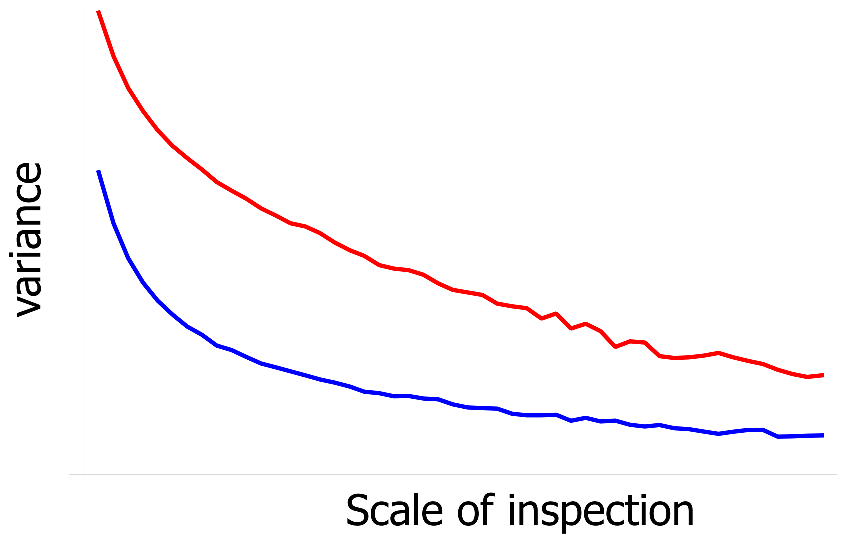
- There are many well-established quantitative measures of formation:
  - variance of local grammage at different scales of inspection
  - power spectrum
  - specific perimeter
- Often comparative quantifiers are used, which compare measured properties with those of a random fibre network.
- Direct mappings exist among all established measure of formation.

Formation




spectral density
wavelength




variance
Scale of inspection

Formation

- The decay of the variance of local grammage with scale of inspection, x, depends on the autocorrelation function for pairs of points separated by a distance r, which we denote $\alpha^*(r)$:

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

- The wavelength power spectrum is given by the Fourier transform of $\alpha^*(r)$.

- For random networks, $\alpha(r)$ and $\sigma^2(\beta)$ are known analytically.

Some more about autocorrelation

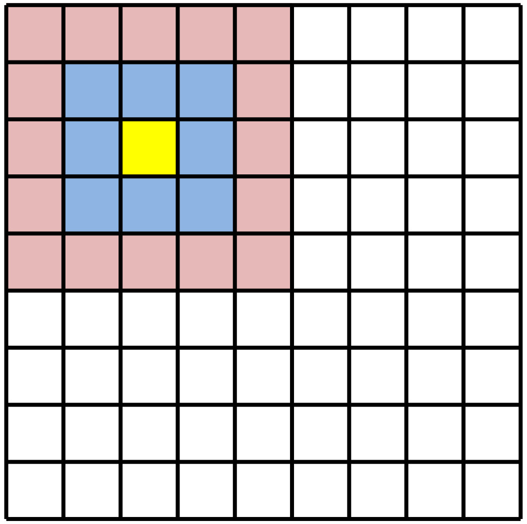
- Autocorrelation is a characteristic of the texture of our grammage map;
- It measures the degree of association of the grammage of each pixel with that of pixels a given distance away:
    - Close pixels are likely to have similar grammage;
    - The grammage of ‘distant’ pixels are independent;
    - Rate of decrease is a characteristic of ‘floc size’
- Autocorrelation is given by the covariance divided by the variance ($0 < \alpha(r) \leq 1$).
- It is a GLOBAL average property.

Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

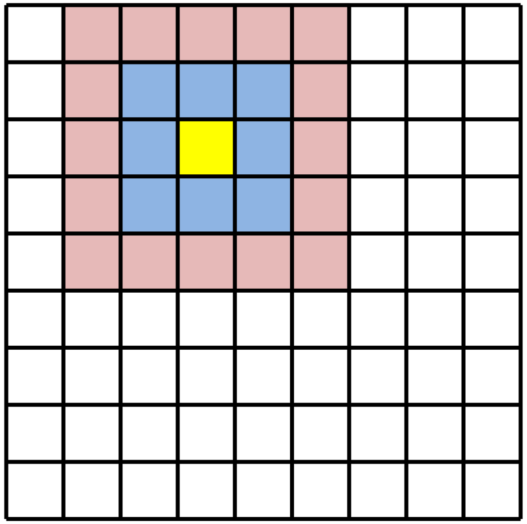


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

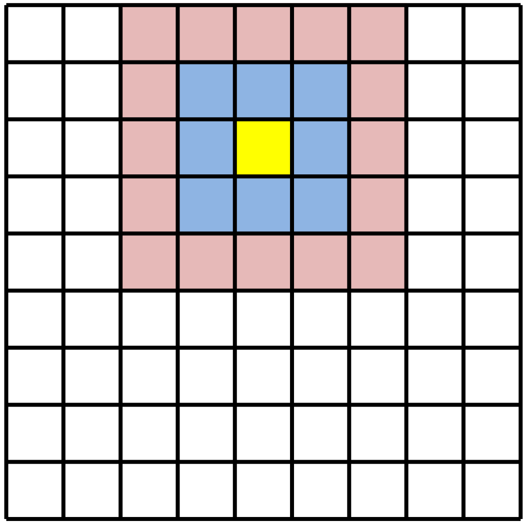


Covariance

- The covariance of a pair of random variables p and q is given by

$$\text{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

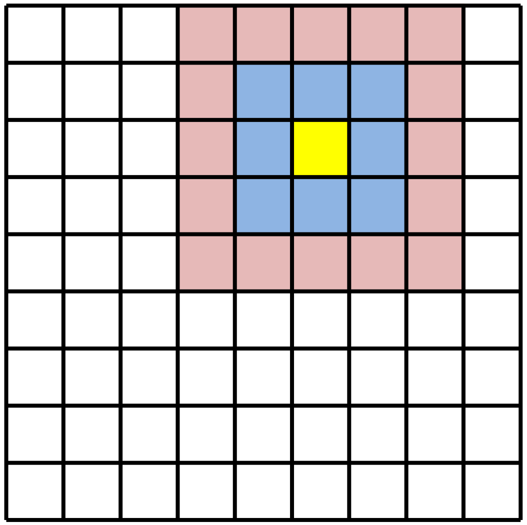


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

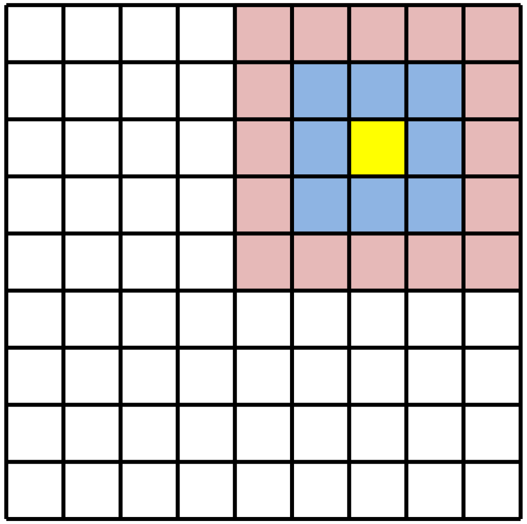


Covariance

- The covariance of a pair of random variables p and q is given by

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

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

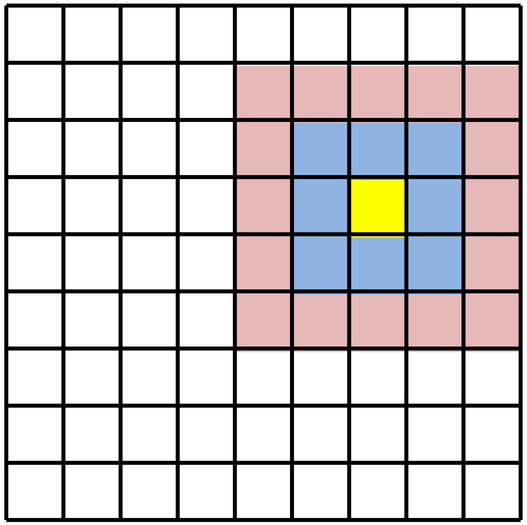


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

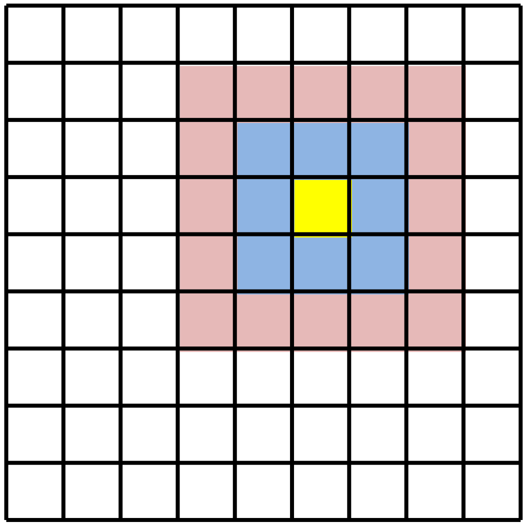


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

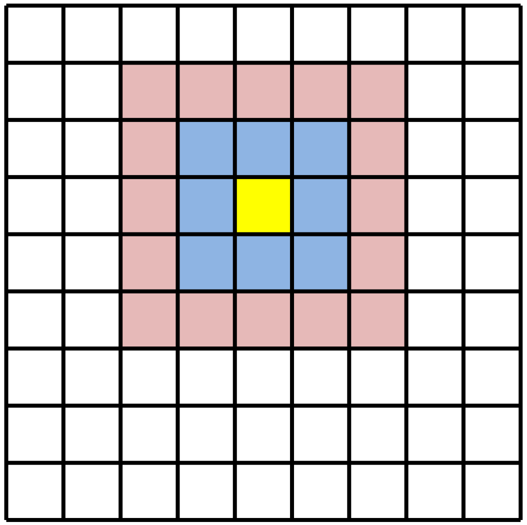


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

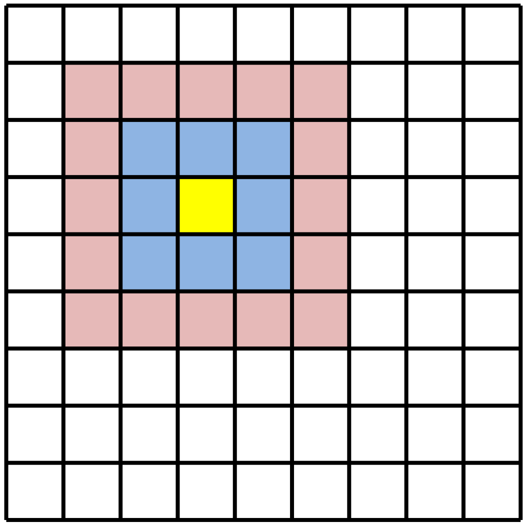


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{pq} - \bar{p}\,\bar{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

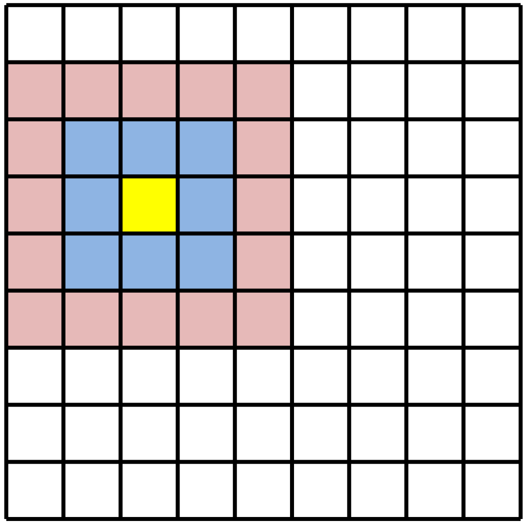


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

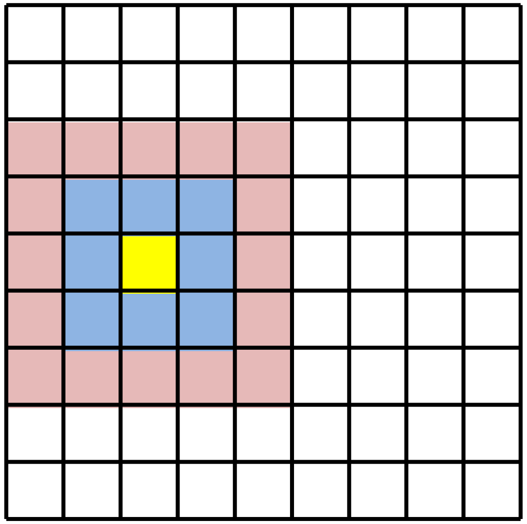


Covariance

- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$

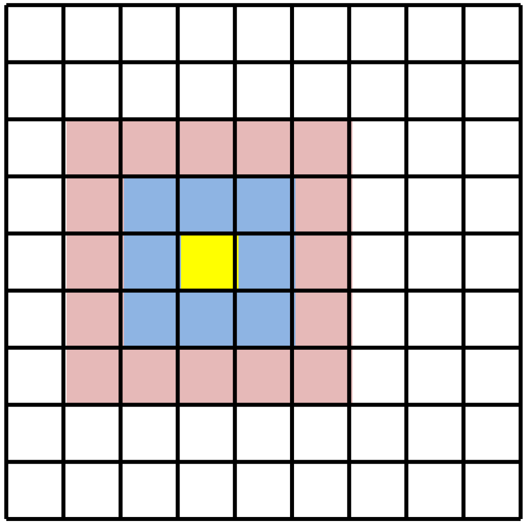


Covariance

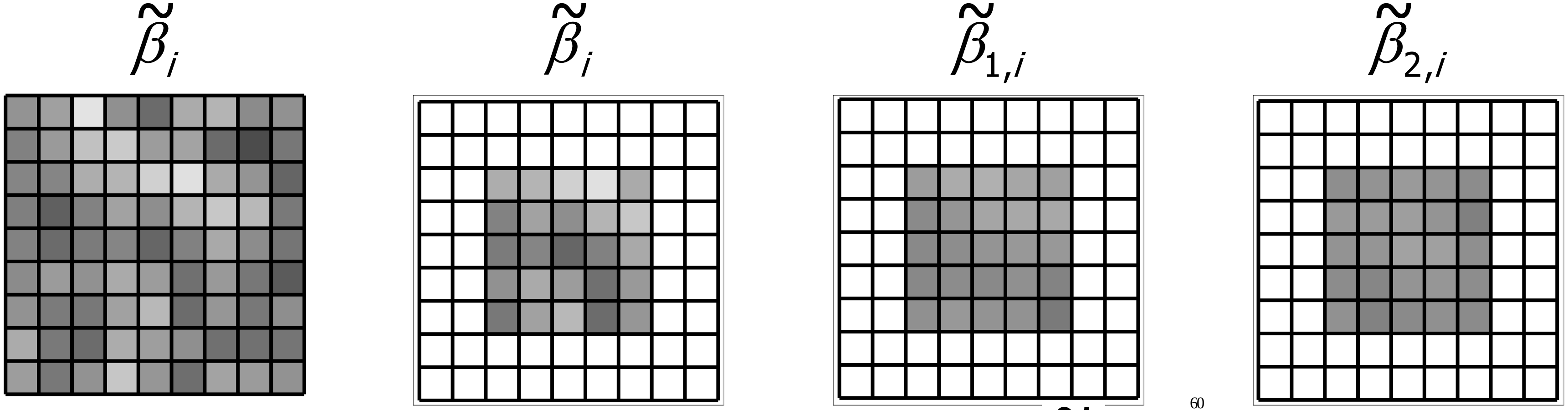
- The covariance of a pair of random variables p and q is given by

$$\mathrm{Cov}(p,q) = \overline{p\,q} - \overline{p}\,\overline{q}$$

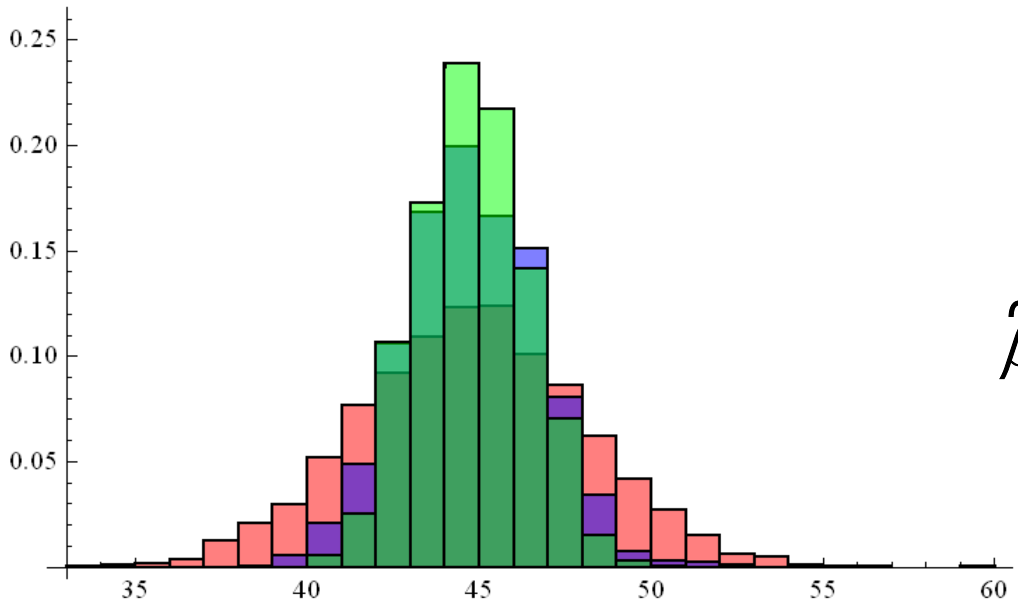
- From our array of local grammage values, $\tilde{\beta}$ we obtain the local average grammage of the first and second neighbours, $\tilde{\beta}_1$ and $\tilde{\beta}_2$



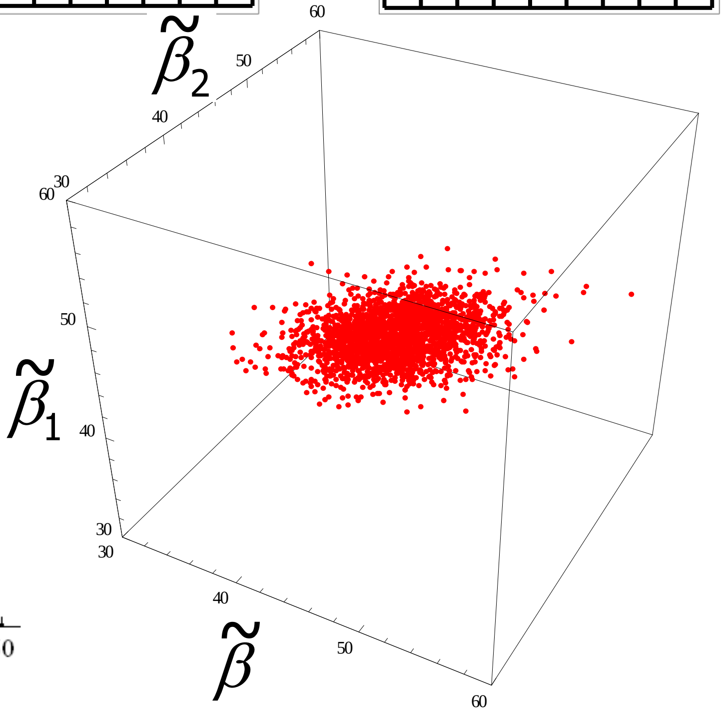
Covariance



$\tilde{\beta}_i$
$\tilde{\beta}_i$
$\tilde{\beta}_{1,i}$
$\tilde{\beta}_{2,i}$



0.25
0.20
0.15
0.10
0.05
35
40
45
50
55
60



$\tilde{\beta}_2$
$\tilde{\beta}_1$
$\tilde{\beta}$
30
40
50
60

Covariance

- So, from the distribution of local grammages, we obtain three approximately Gaussian distributions.
- The random variables, $\tilde{\beta}$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are not independent; they have covariances:

$$\mathrm{Cov}(\tilde{\beta}, \tilde{\beta}_1) \qquad \mathrm{Cov}(\tilde{\beta}, \tilde{\beta}_2) \qquad \mathrm{Cov}(\tilde{\beta}_1, \tilde{\beta}_2)$$

- The covariance matrix characterises the trivariate Gaussian distribution:

$$\Sigma = \begin{pmatrix} \sigma^2\left(\tilde{\beta}\right) & \mathrm{Cov}\left(\tilde{\beta}, \tilde{\beta}_1\right) & \mathrm{Cov}\left(\tilde{\beta}, \tilde{\beta}_2\right) \\ \mathrm{Cov}\left(\tilde{\beta}, \tilde{\beta}_1\right) & \sigma^2\left(\tilde{\beta}_1\right) & \mathrm{Cov}\left(\tilde{\beta}_1, \tilde{\beta}_2\right) \\ \mathrm{Cov}\left(\tilde{\beta}, \tilde{\beta}_2\right) & \mathrm{Cov}\left(\tilde{\beta}_1, \tilde{\beta}_2\right) & \sigma^2\left(\tilde{\beta}_2\right) \end{pmatrix}$$

Information distance

- For a pair of trivariate Gaussian distributions, *A* & *B*, with
  - common mean vector, $\mu_A = \mu_B = \mu$
  - different covariance matrices, $\Sigma^A \neq \Sigma^B$

the information distance is known and is given by

$$D_\Sigma\left(f^A, f^B\right) = \sqrt{\frac{1}{2}\sum_{j=1}^{3}\log^2\left(\lambda_j\right)}$$

where

$$\left\{\lambda_j\right\} = \mathrm{Eig}\left(\Sigma^{A^{-1/2}} \cdot \Sigma^B \cdot \Sigma^{A^{-1/2}}\right)$$

Information distance by example

Inputs:

- Grammage, $\bar{\beta}$
- Fibre properties:
  - Length, λ
  - Coarseness, δ
  - Width, ω
- Mean floc radius, r_f
- Floc intensity, $0 \leq I \leq 1$
- Expected number of fibres per cluster, $\overline{n_c}$

Simulation:

- Number of fibres per cluster, n_c, is a Poisson variable with mean, $\overline{n_c}$
- Mean grammage, G, of each cluster is assumed constant (*cf.* Farnood *et al.* 1995)

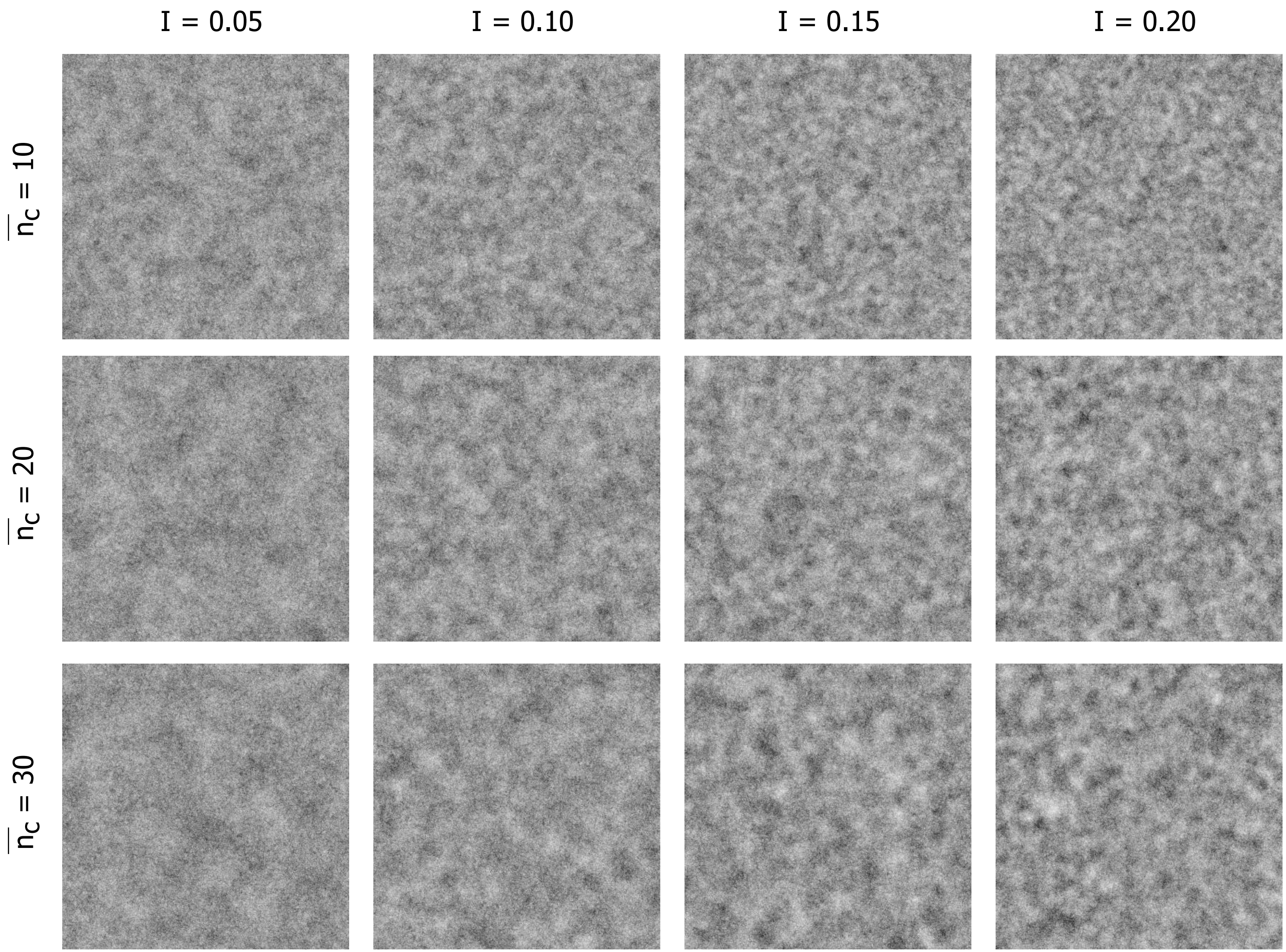
$$G = I\,\beta_{\text{fib}} = \frac{I\,\delta}{\omega}$$

- Radius of each cluster is

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

- n_c fibre centres deposited within circles of radius r.
- For each fibre, contribution to mass of each pixel calculated.

Information distance by example



Information distance by example

16 samples

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5		✓			✓		✓
	10		✓	✓	✓	✓	✓	✓
	20		✓			✓		✓
	30		✓			✓		✓

Information distance by example

- From each of these 16 samples, we compute $\tilde{\beta}$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ and their covariance matrices

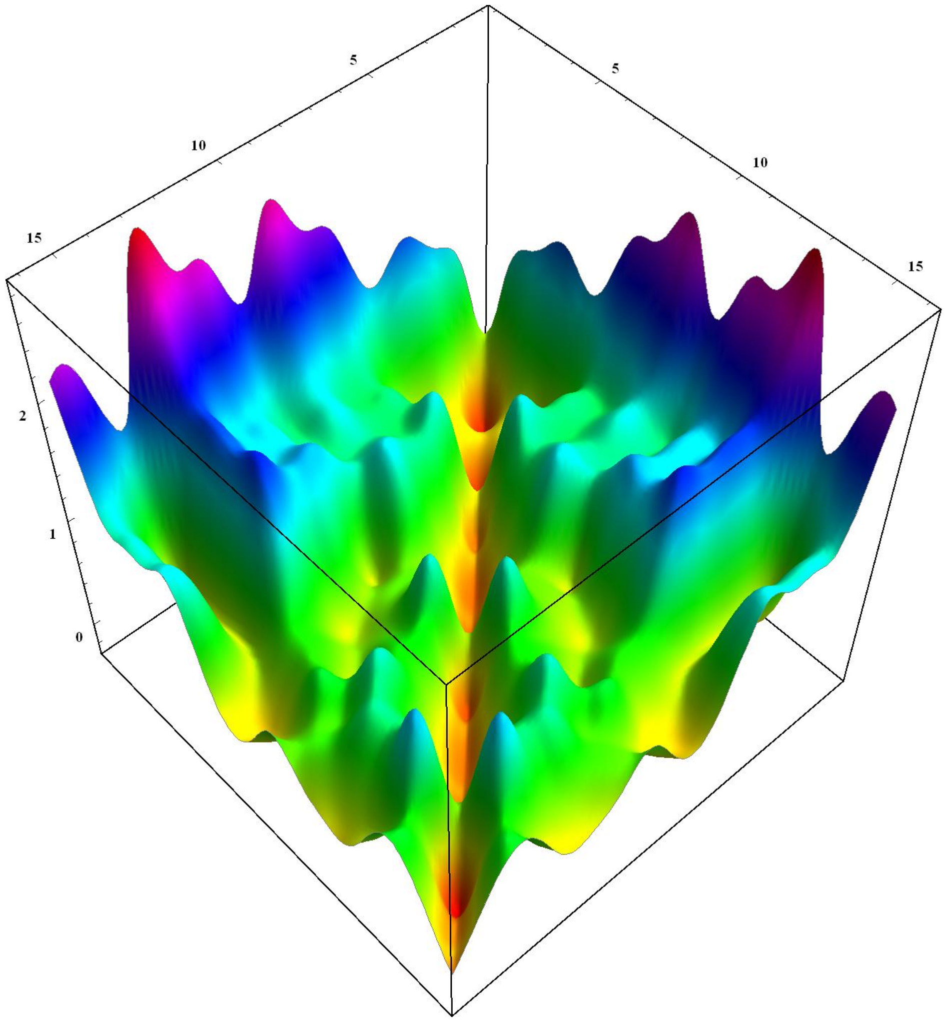
$$\left\{ \begin{pmatrix} 7.1 & 0.5 & 0.0 \\ 0.5 & 1.1 & 0.1 \\ 0.0 & 0.1 & 0.5 \end{pmatrix}, \begin{pmatrix} 11 & 3.3 & 1.4 \\ 3.3 & 3.4 & 1.4 \\ 1.4 & 1.4 & 1.7 \end{pmatrix}, \ldots, \begin{pmatrix} 50 & 27 & 7.9 \\ 27 & 21 & 9.2 \\ 7.9 & 9.2 & 9.0 \end{pmatrix} \right\}$$

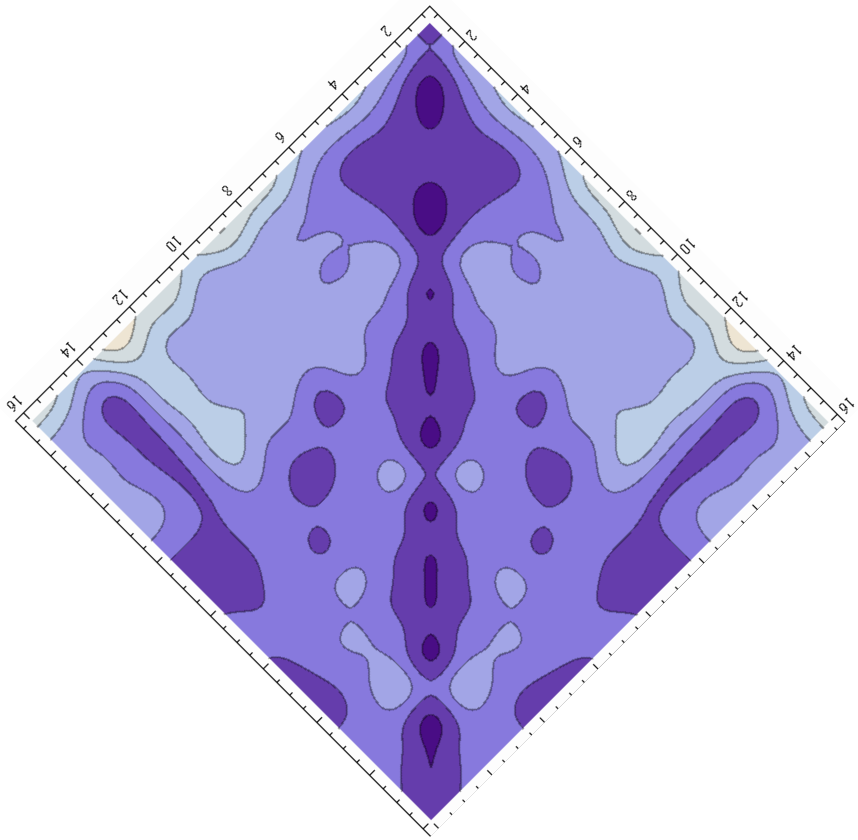
- We then compute the mutual distances between all pairs of covariance matrices.

Information distance by example

0.	1.1	1.4	1.6	1.5	1.6	2.1	2.2	2.5	1.9	2.4	2.5	2.8	1.6	2.	2.2
1.1	0.	0.3	0.6	0.5	0.8	1.1	1.2	1.4	1.1	1.5	1.5	1.8	0.5	1.	1.3
1.4	0.3	0.	0.3	0.2	0.8	1.	1.1	1.2	1.1	1.4	1.4	1.6	0.4	0.8	1.2
1.6	0.6	0.3	0.	0.2	0.9	1.	1.	1.1	1.1	1.4	1.4	1.5	0.4	0.8	1.2
1.5	0.5	0.2	0.2	0.	0.9	1.1	1.1	1.2	1.1	1.5	1.5	1.7	0.5	0.9	1.3
1.6	0.8	0.8	0.9	0.9	0.	0.6	0.8	1.1	0.4	0.8	0.9	1.3	0.6	0.6	0.6
2.1	1.1	1.	1.	1.1	0.6	0.	0.3	0.6	0.6	0.5	0.4	0.7	0.6	0.2	0.2
2.2	1.2	1.1	1.	1.1	0.8	0.3	0.	0.4	0.8	0.6	0.5	0.6	0.7	0.3	0.4
2.5	1.4	1.2	1.1	1.2	1.1	0.6	0.4	0.	1.	0.8	0.7	0.5	0.9	0.6	0.7
1.9	1.1	1.1	1.1	1.1	0.4	0.6	0.8	1.	0.	0.5	0.7	1.1	0.8	0.7	0.5
2.4	1.5	1.4	1.4	1.5	0.8	0.5	0.6	0.8	0.5	0.	0.3	0.7	1.1	0.7	0.3
2.5	1.5	1.4	1.4	1.5	0.9	0.4	0.5	0.7	0.7	0.3	0.	0.5	1.1	0.6	0.3
2.8	1.8	1.6	1.5	1.7	1.3	0.7	0.6	0.5	1.1	0.7	0.5	0.	1.3	0.8	0.7
1.6	0.5	0.4	0.4	0.5	0.6	0.6	0.7	0.9	0.8	1.1	1.1	1.3	0.	0.4	0.8
2.	1.	0.8	0.8	0.9	0.6	0.2	0.3	0.6	0.7	0.7	0.6	0.8	0.4	0.	0.4
2.2	1.3	1.2	1.2	1.3	0.6	0.2	0.4	0.7	0.5	0.3	0.3	0.7	0.8	0.4	0.

Information distance by example





Dimensionality reduction by example

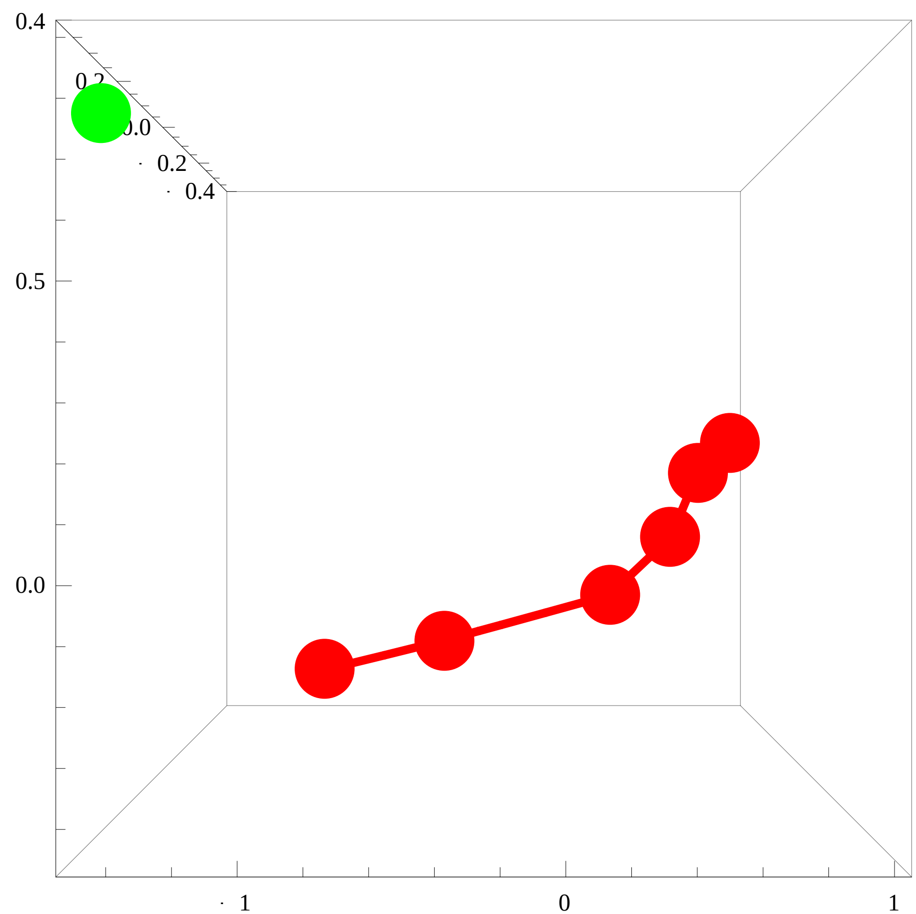
- We seek to visualize our N (=16) samples on a 3D surface.
- We employ the 'dimensionality reduction' or 'multi-dimensional scaling' approach of Carter *et al.* (2009):
  1. Centralize the matrix of DS by subtracting row and column means and adding grand mean;
  2. Compute the N eigenvalues and $N \times N$-dimensional eigenvectors of the resultant matrix;
  3. Make a 3 × 3 matrix, A, of the three largest eigenvalues; make a 3 x N matrix, B, of corresponding eigenvectors;
  4. The transpose of the product $A \cdot B$ is an $N \times 3$ matrix which gives N coordinates in 3-space.

Dimensionality reduction by example

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5		✓			✓		✓
	10		✓	✓	✓	✓	✓	✓
	20		✓			✓		✓
	30		✓			✓		✓

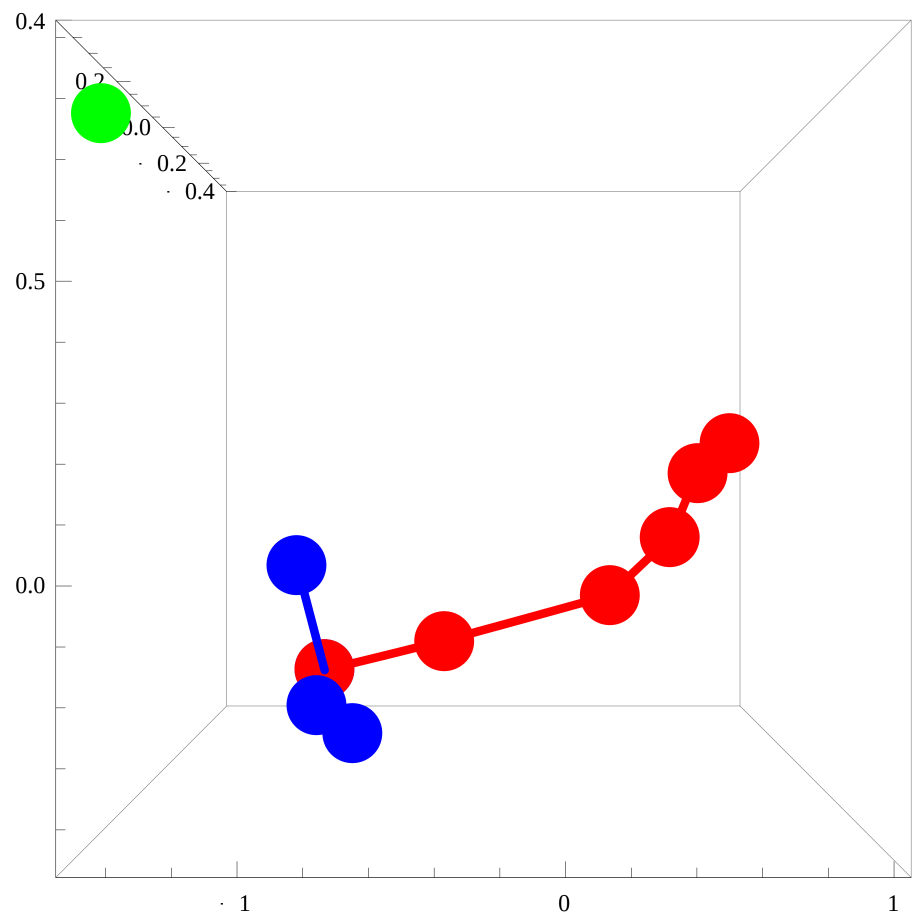
Dimensionality reduction by example

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5							
	10		✓	✓	✓	✓	✓	✓
	20							
	30							



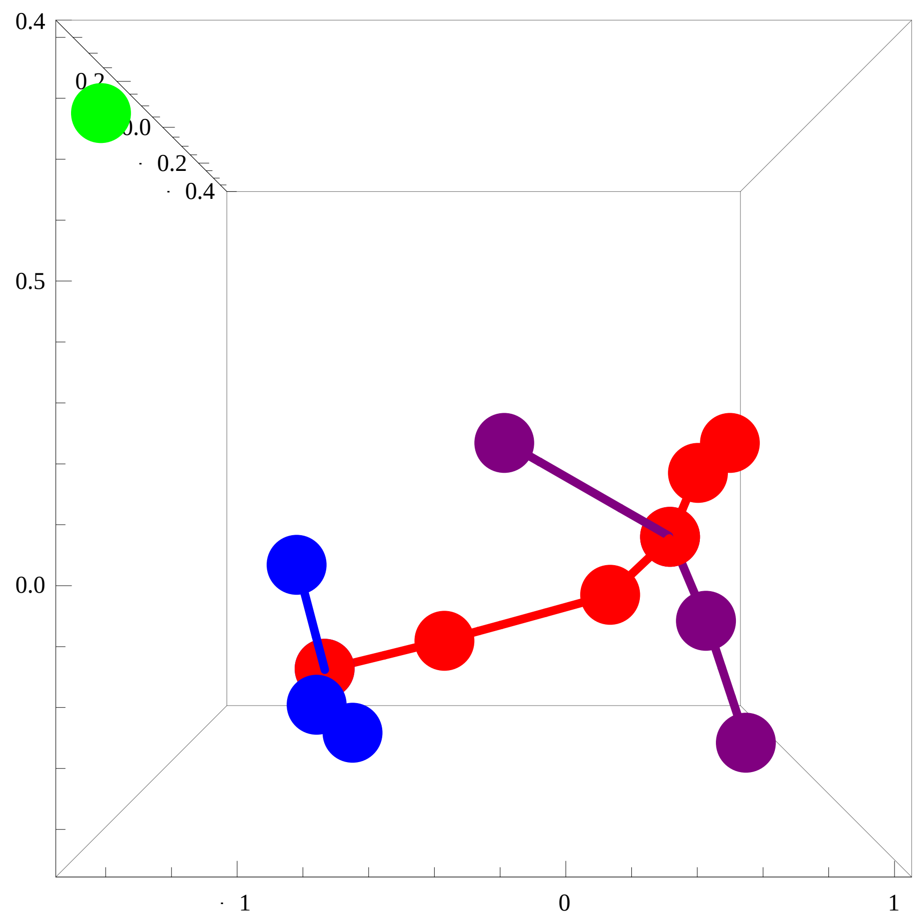
Dimensionality reduction by example

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5		✓					
	10		✓	✓	✓	✓	✓	✓
	20		✓					
	30		✓					



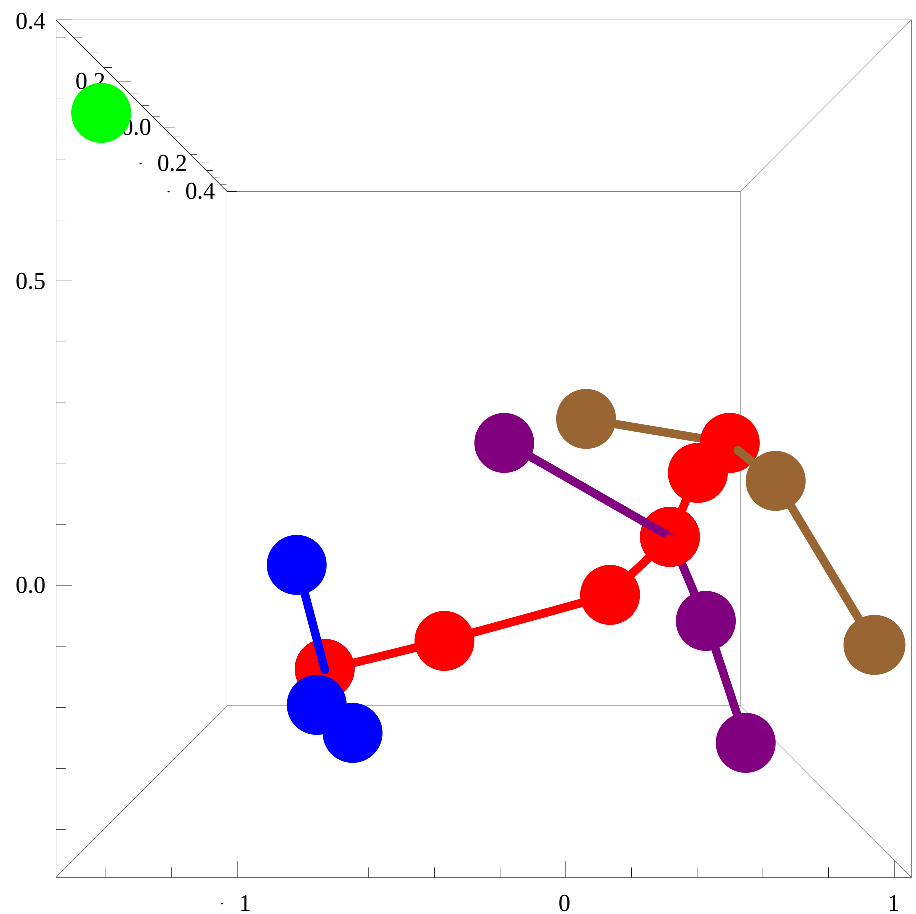
Dimensionality reduction by example

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5		✓			✓		
	10		✓	✓	✓	✓	✓	✓
	20		✓			✓		
	30		✓			✓		



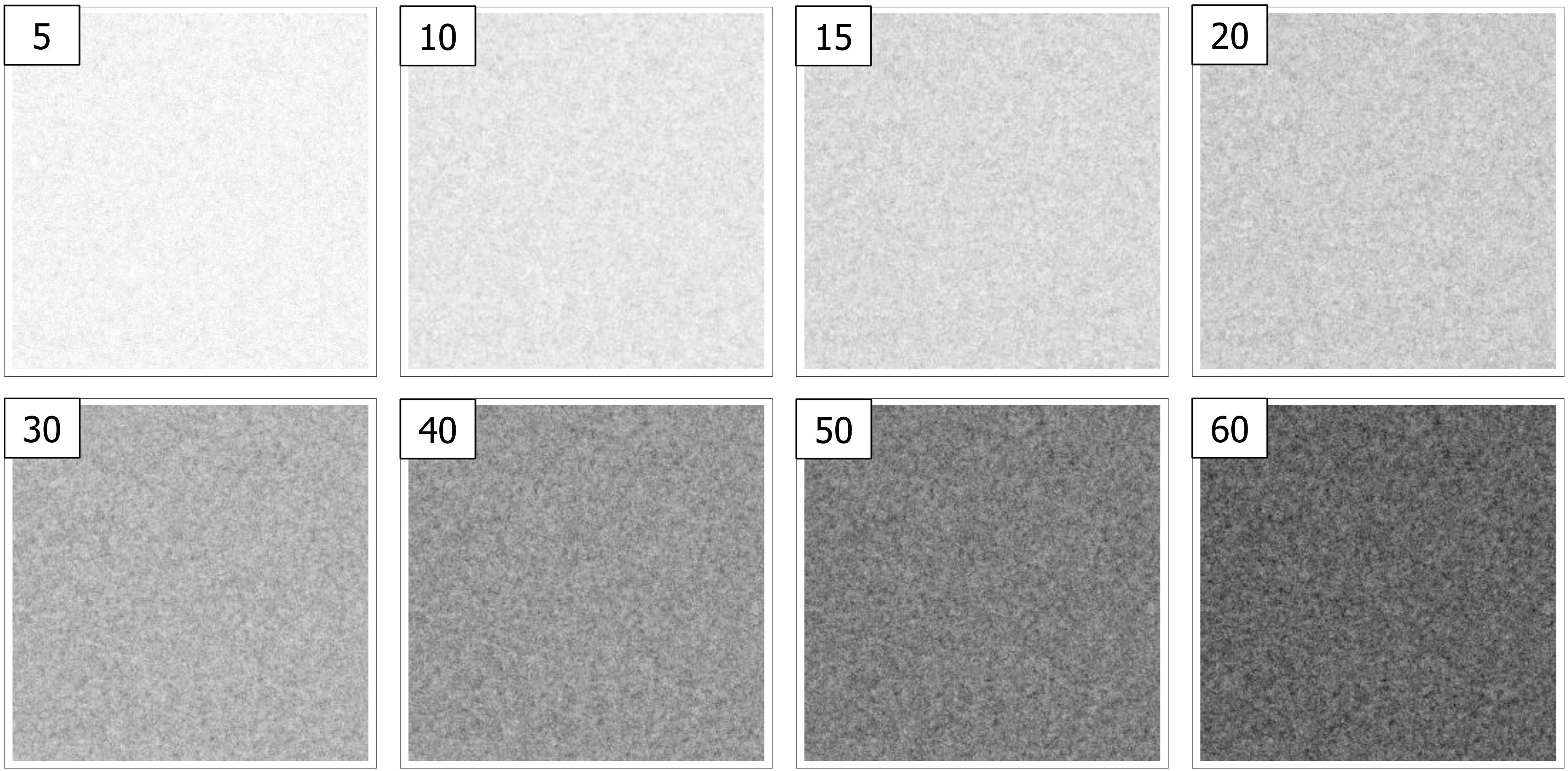
Dimensionality reduction by example

		l						
		–	0.010	0.020	0.035	0.050	0.075	0.100
n_c	1	✓						
	5		✓			✓		✓
	10		✓	✓	✓	✓	✓	✓
	20		✓			✓		✓
	30		✓			✓		✓

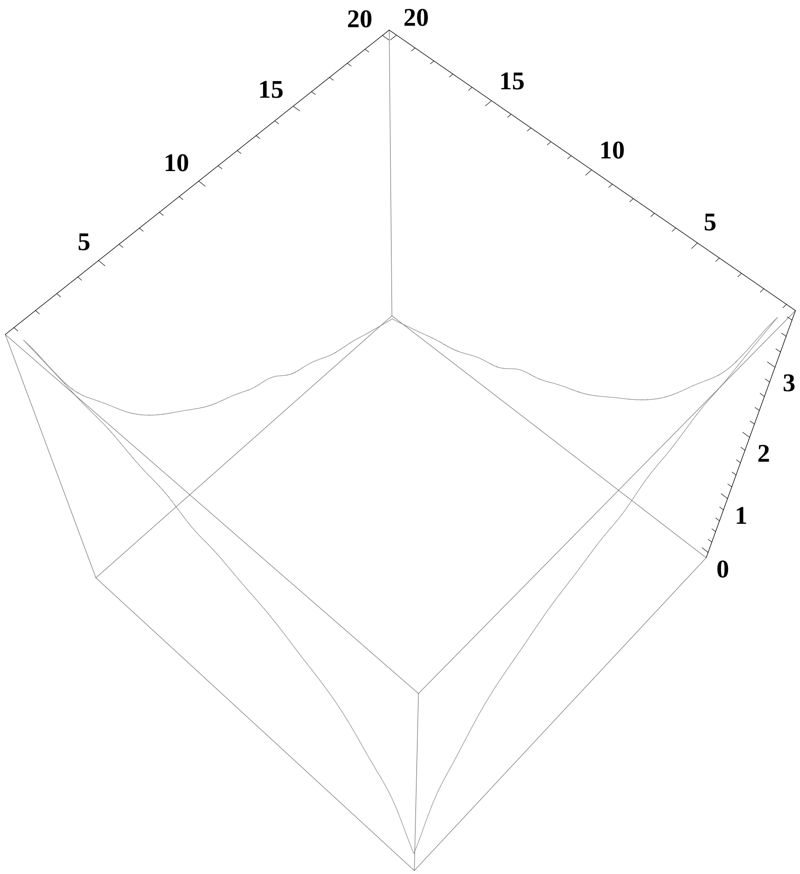


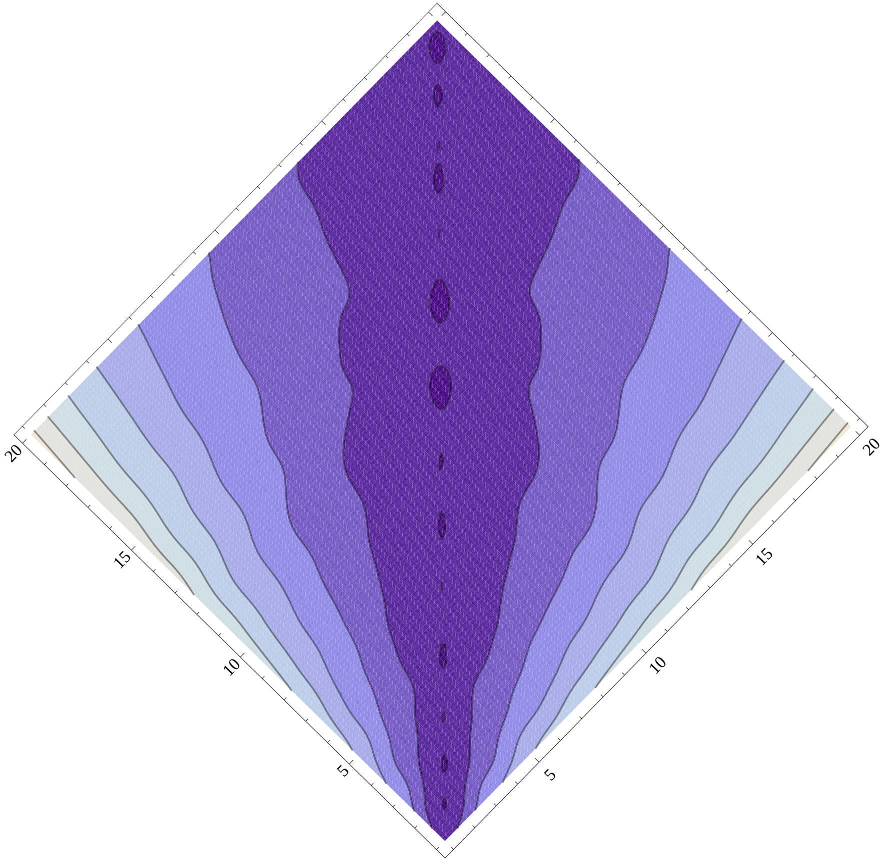
Effect of grammage; random

- Simulated networks with grammage 5, 10, 15, ..., 100 g m^{-2}

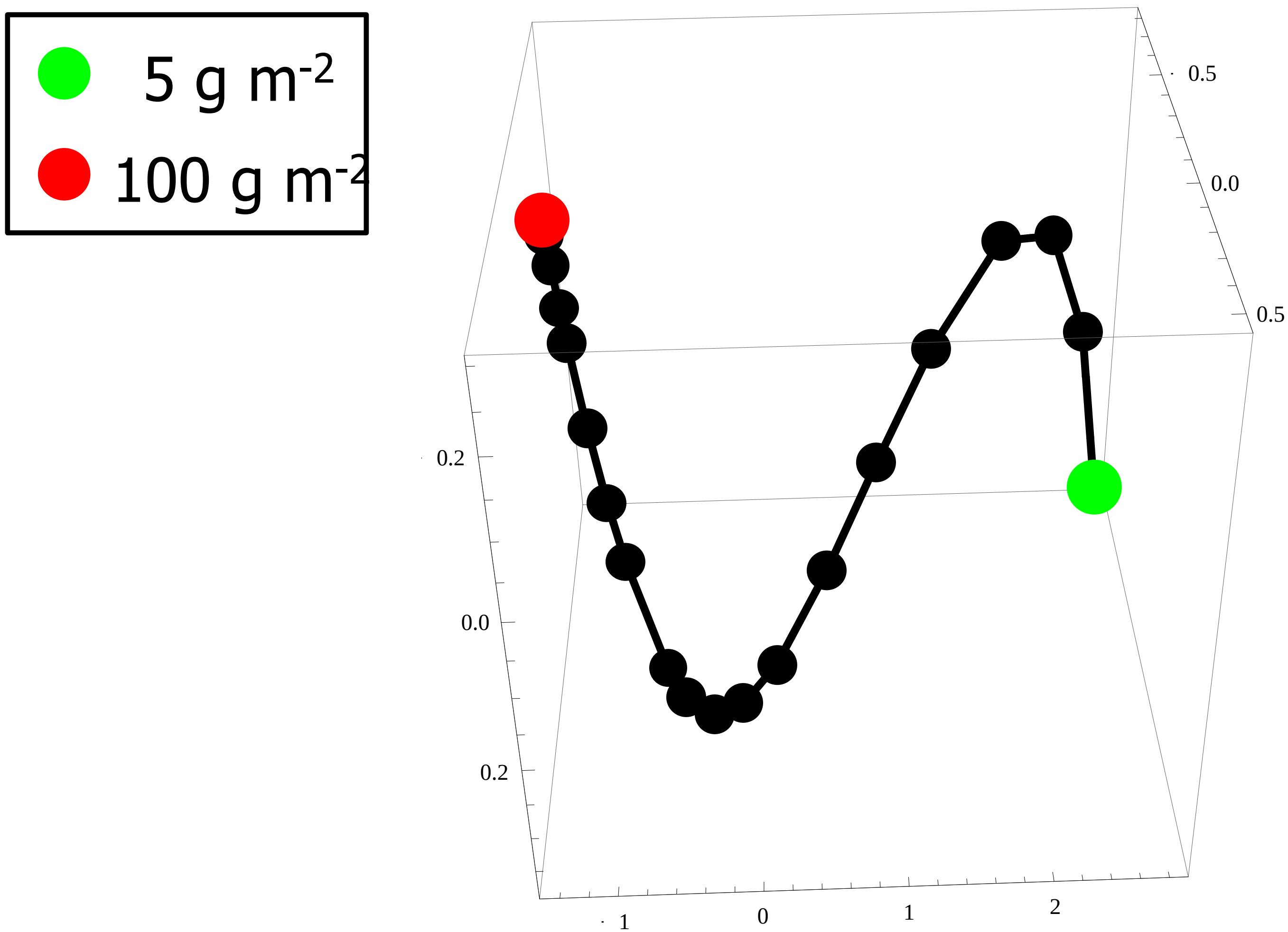


Effect of grammage; random



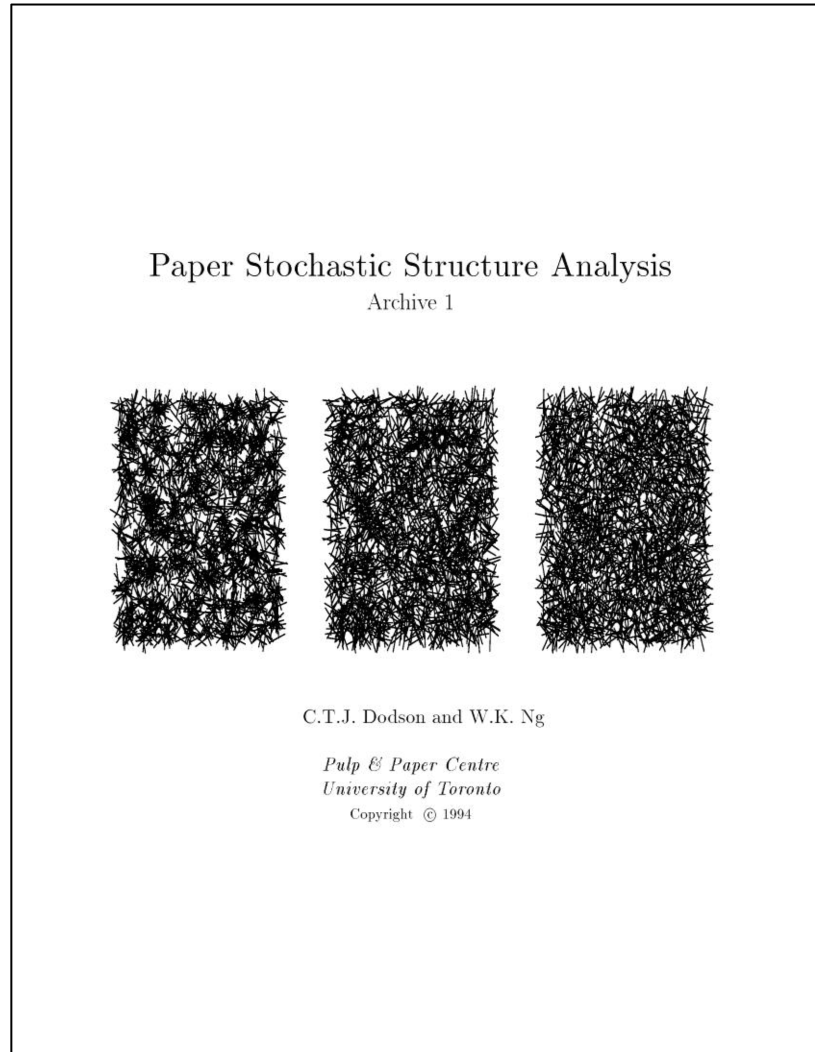


Effect of grammage; random



UofT archive

- 'Archive 1'
- Radiographs of 182 samples
- Handsheets
  - Headbox
  - Couch trim
  - Settling experiments
  - Miscellaneous
- Pilot machines
- Gap formers
- Hybrid formers


Paper Stochastic Structure Analysis

Archive 1

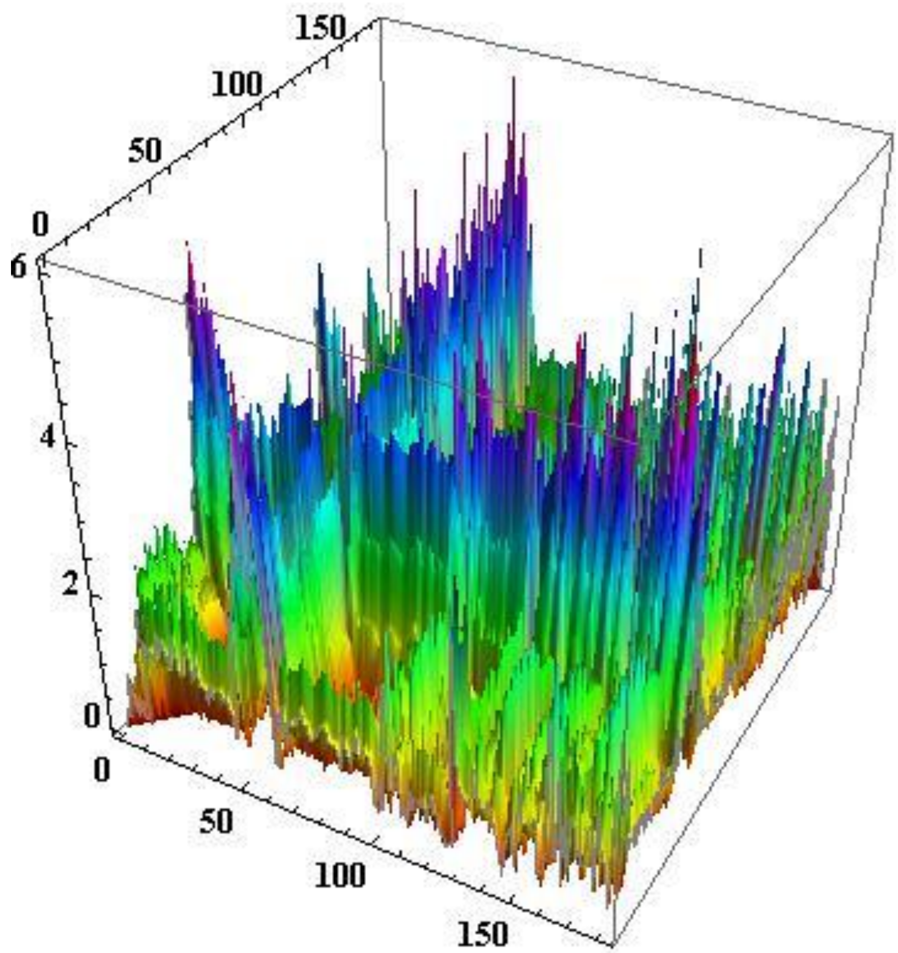
C.T.J. Dodson and W.K. Ng

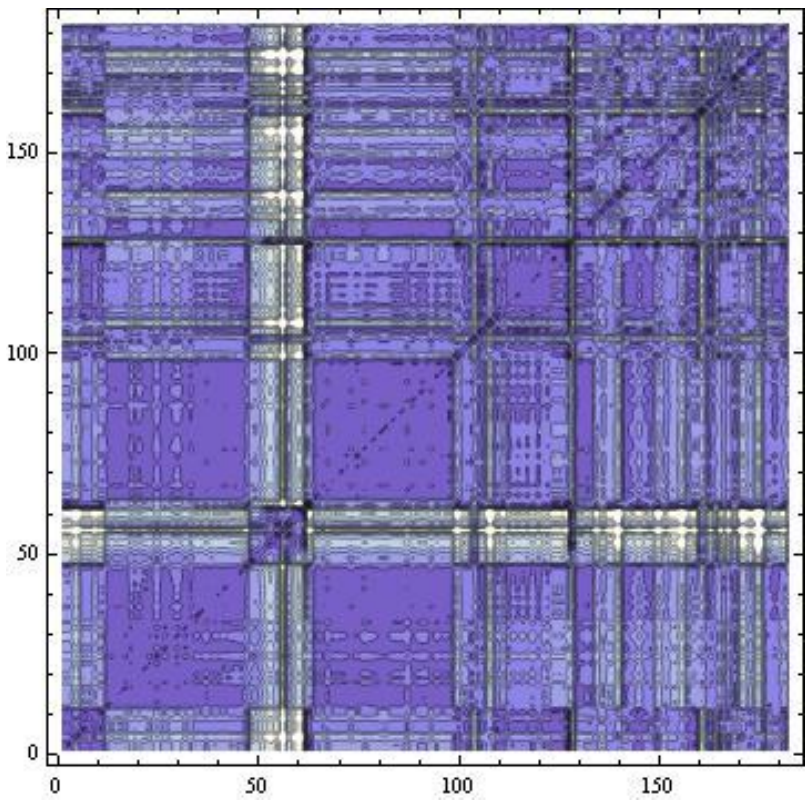
Pulp & Paper Centre

University of Toronto

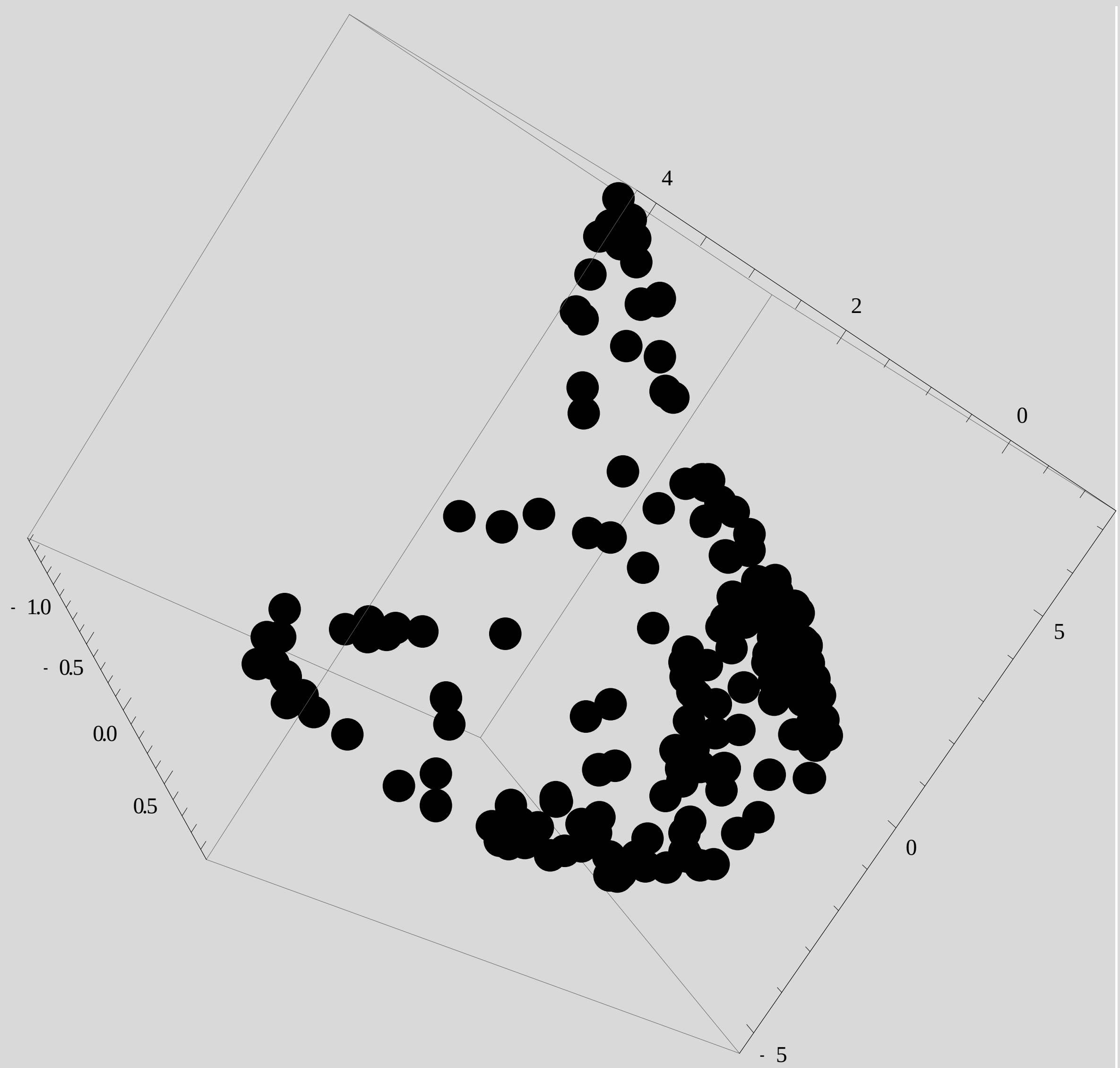
Copyright © 1994

UofT archive



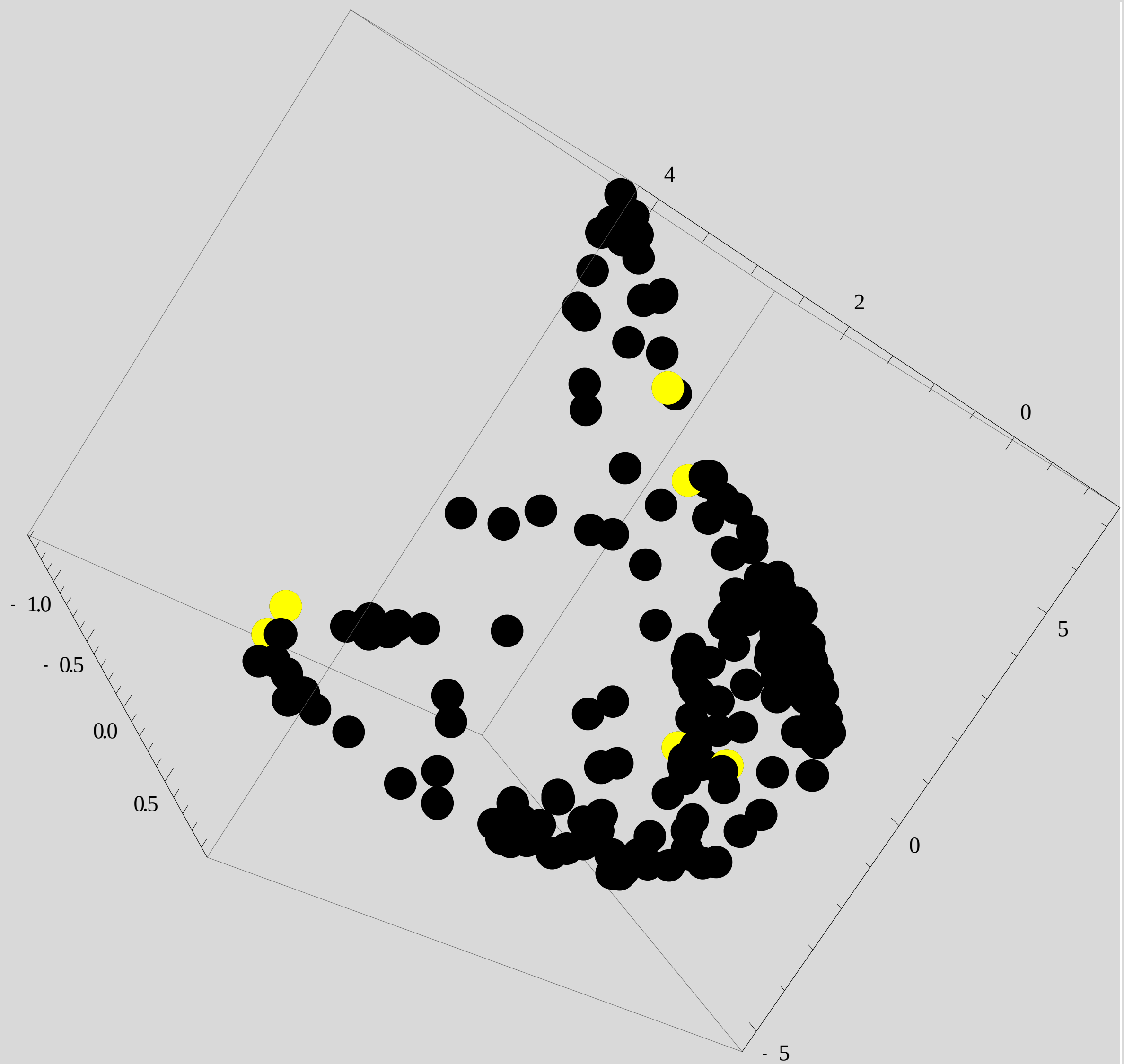


UofT archive



UofT archive

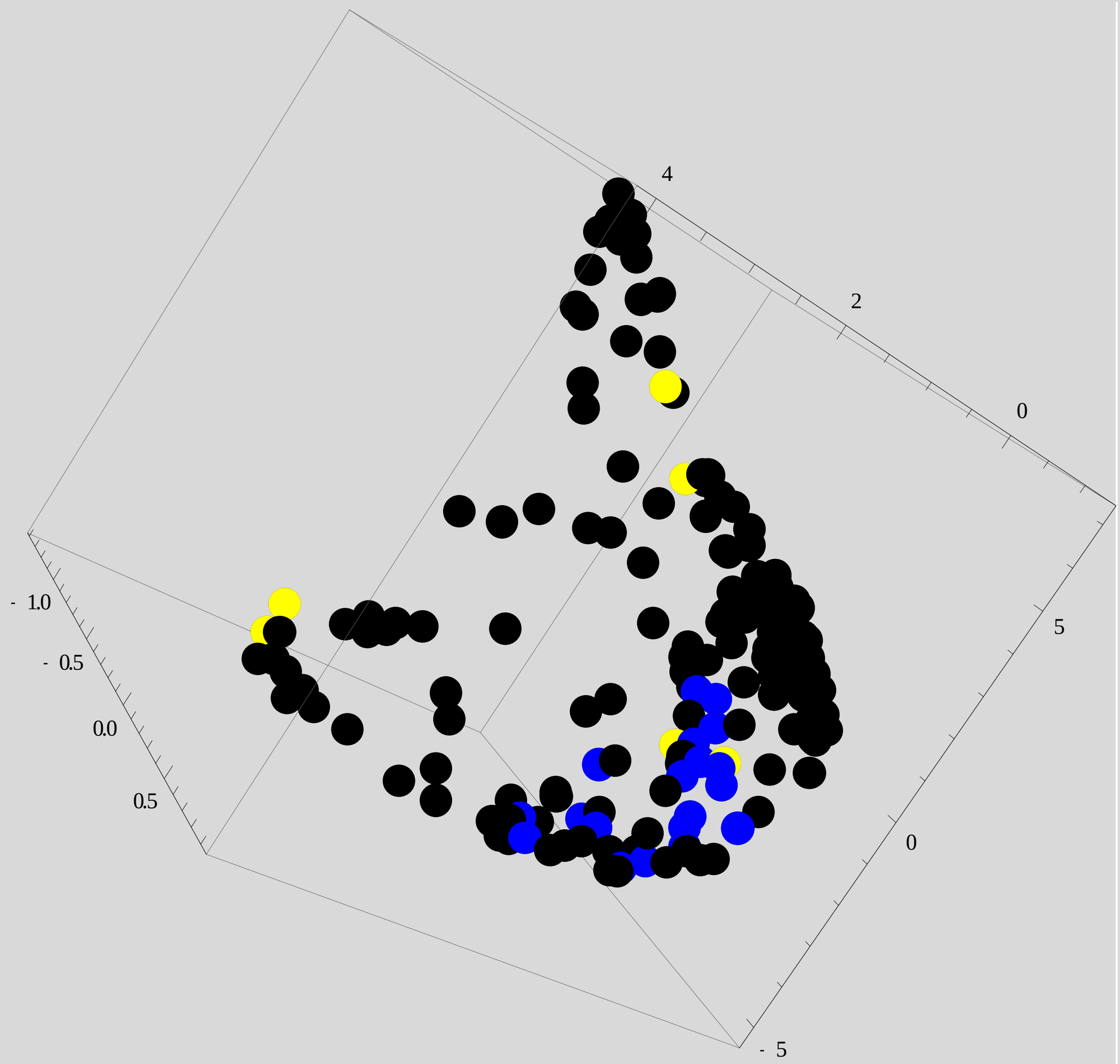
Fourdrinier



UofT archive

Fourdrinier

Gap

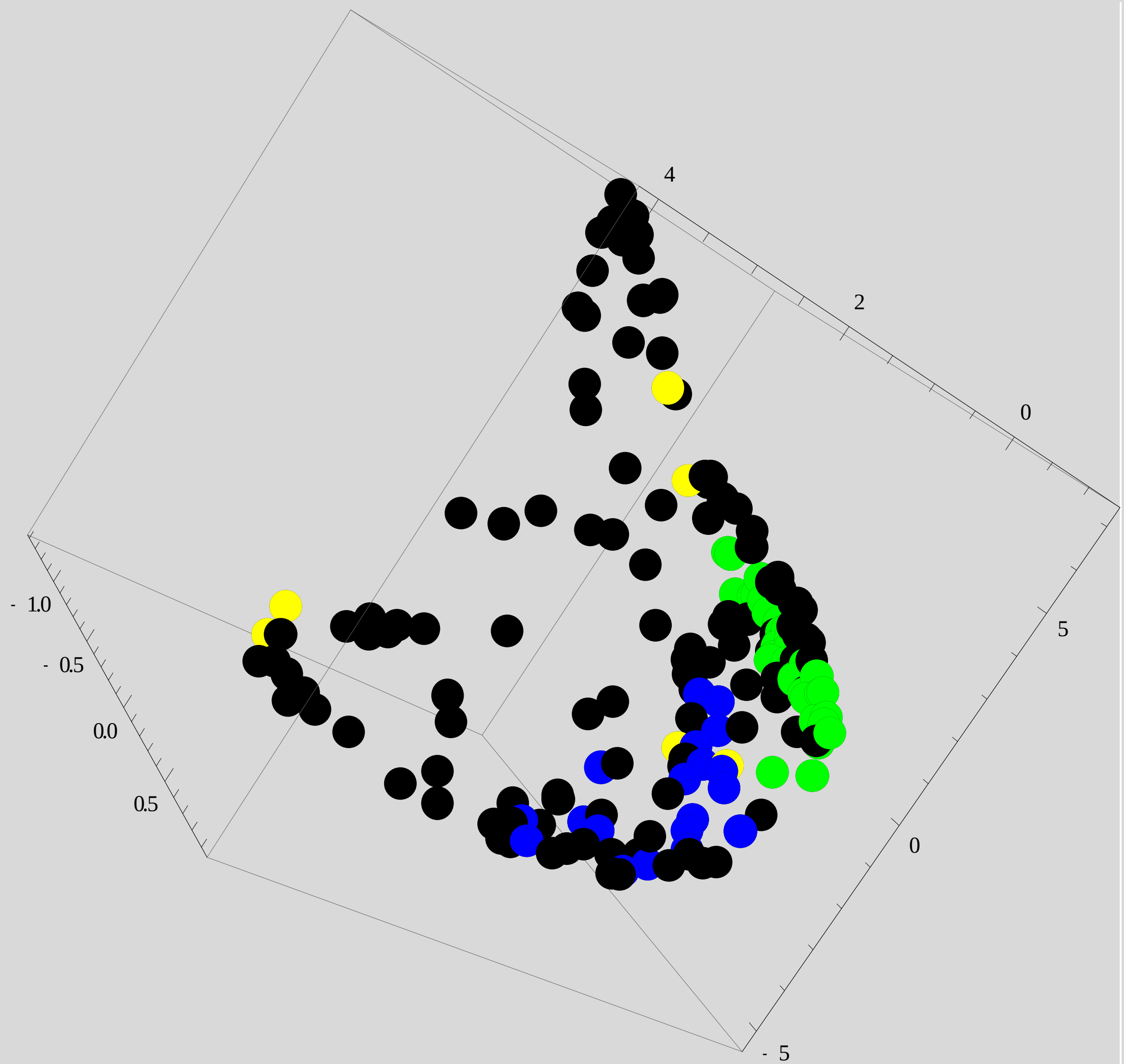


UofT archive

Fourdrinier

Gap

Hybrid



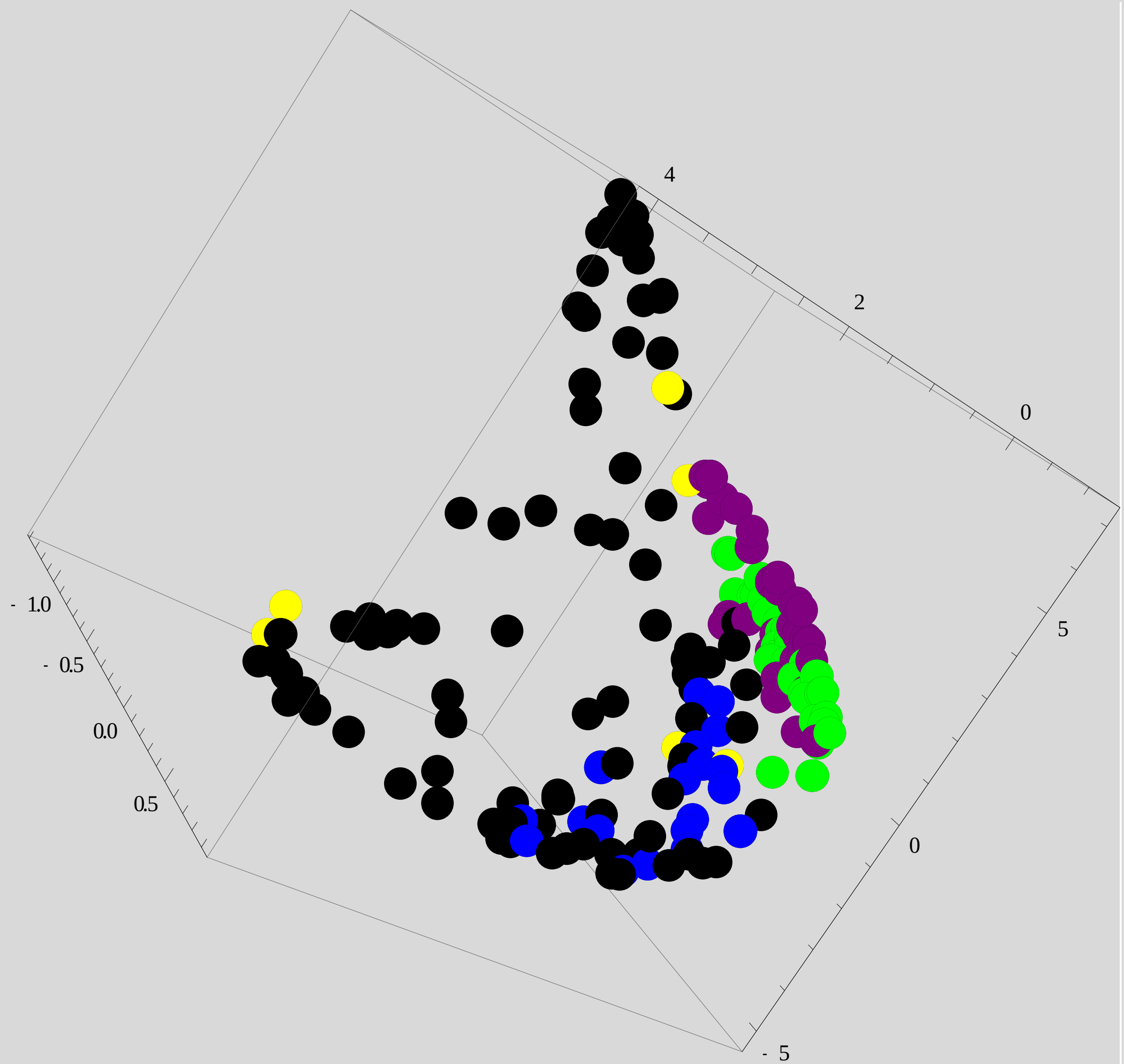
UofT archive

Fourdrinier

Gap

Hybrid

Pilot



UofT archive

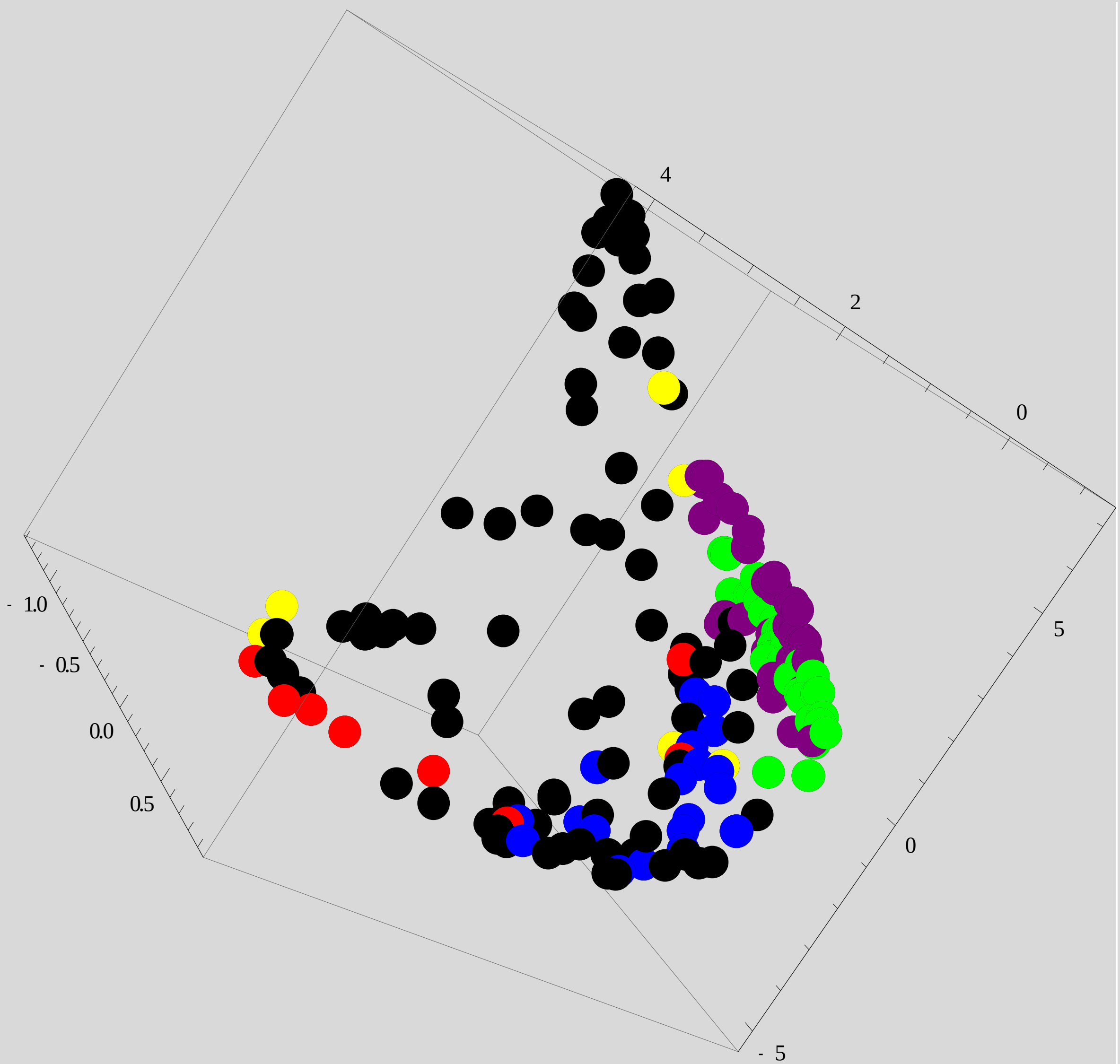
Fourdrinier

Gap

Hybrid

Pilot

Handsheets (gsm)



UofT archive

Fourdrinier

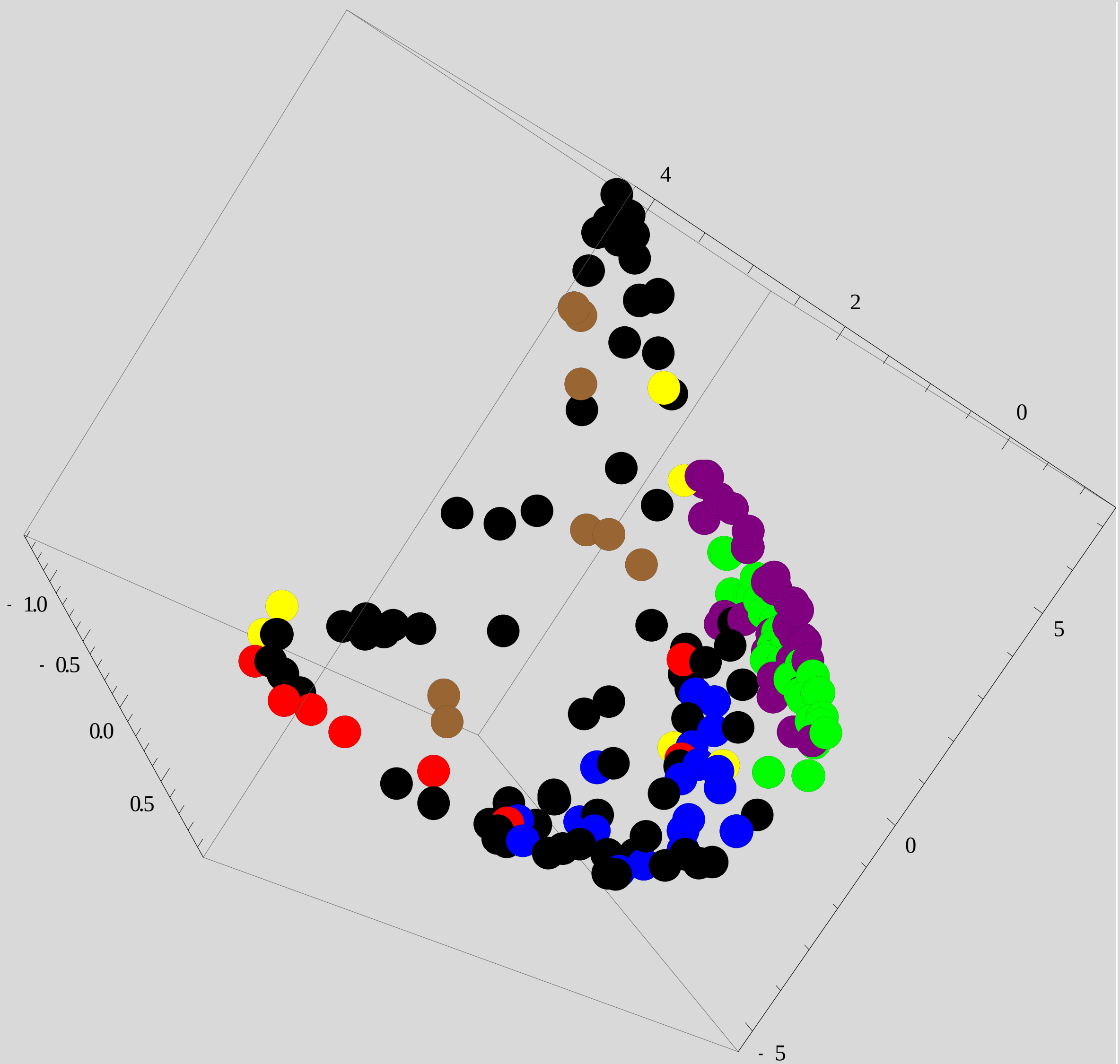
Gap

Hybrid

Pilot

Handsheets (gsm)

Settled handseets (gsm)



Conclusions

- Information geometry provides a natural metric to discriminate among formation textures
- Discrimination among simulated textures is consistent with the parameters used to generate them
- Sheets formed by different forming methods exhibit clustering according to forming conditions