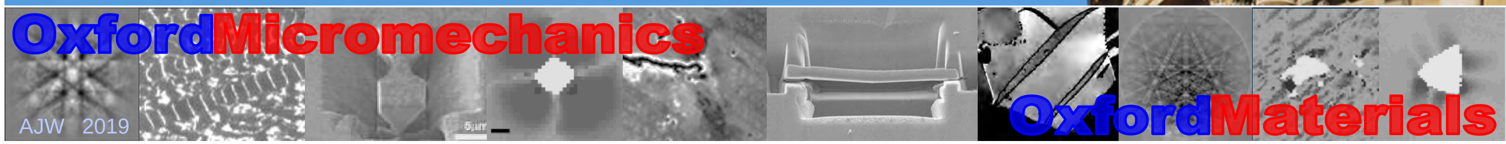
Multivariate Statistical Analysis of EBSD Datasets

Angus J Wilkinson

Royal Microscopical Society

@AngusJWilkinson, #EBSD2019

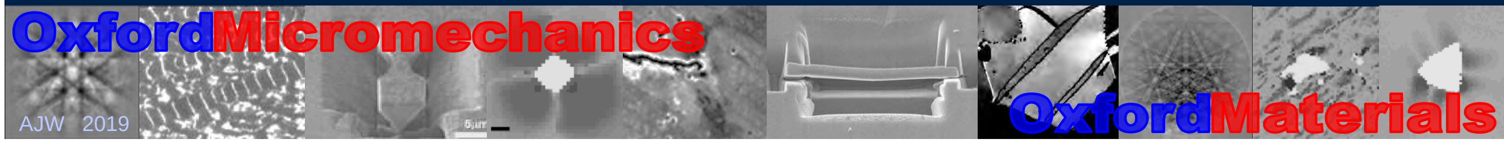
April 2019, NPL, Teddington



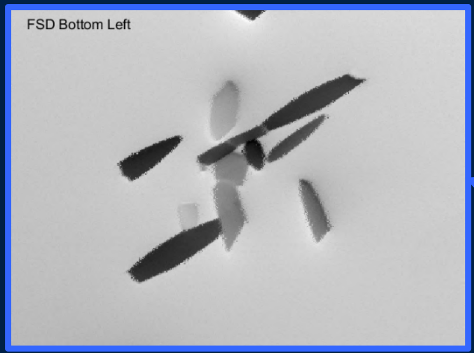
UNIVERSITY OF OXFORD

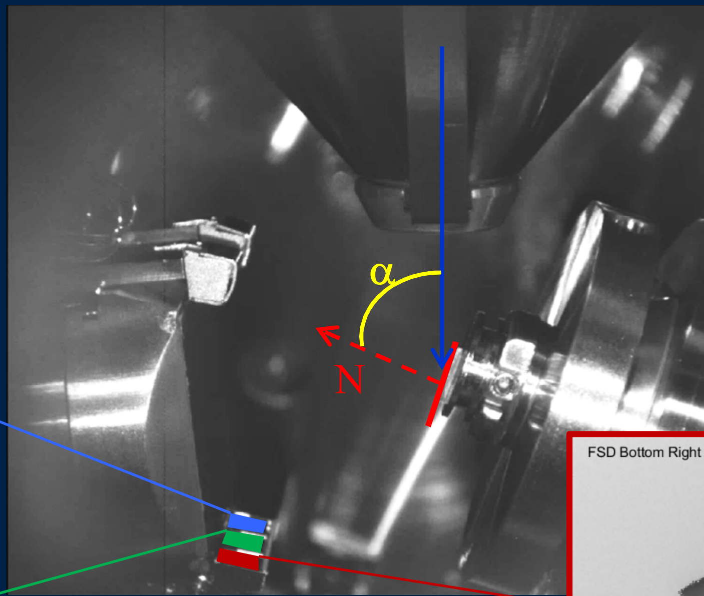
Quantitative Microscopy

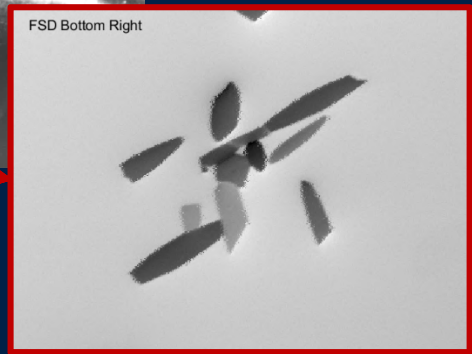
- Segment Domains
  - Identify similar regions (similar how?)
  - Distinguish from dissimilar regions (dissimilar how?)
- Characteristic of Domains
  - chemistry, crystal type, orientation
- Measurements
  - phase fractions, particle/grain sizes, aspect ratios, spacings, …



FSD Imaging





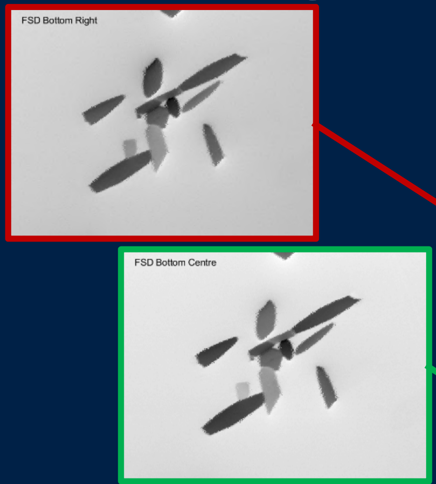


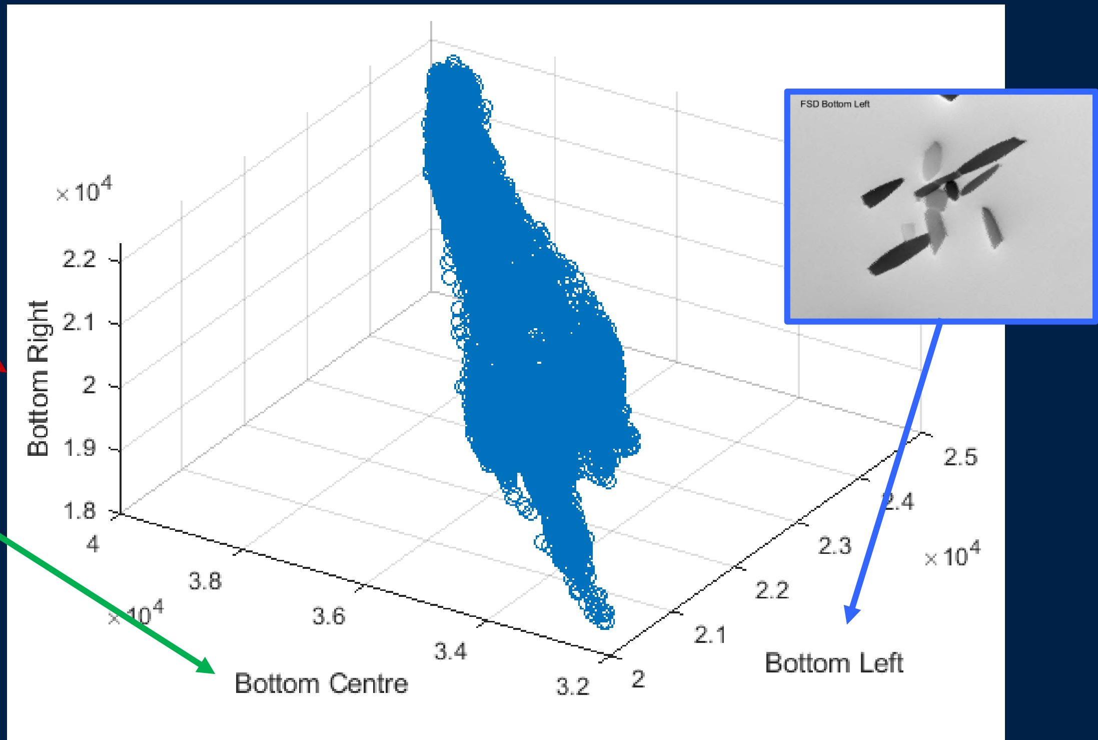
set of 3 images → measurements (here 3 intensities) at series of points on specimen

TiVZrHfTa - High Entropy Alloy

John C Waite – Oxford Micromechanics Group

FSD Imaging





What is the best way to use these measurements to describe the specimen?

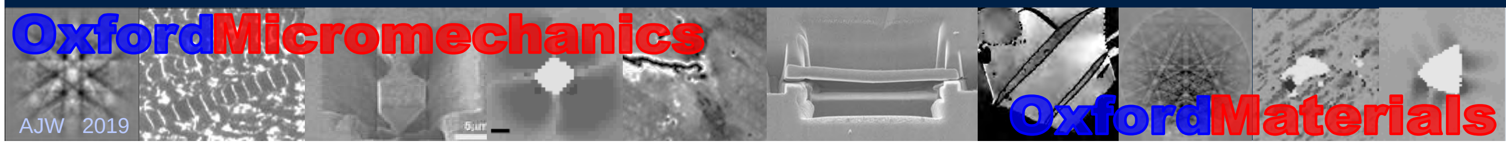
Principal Component Analysis

- PCA is a statistical procedure that converts a set of observations (eg several signals measured at a point on specimen) of a number of variables (eg intensity on a FSD diode) into a set of values of linearly uncorrelated (latent) variables called principal components.
- The first principal component (ie linear combination of the measured variables) is constructed so that it gives the largest possible variance across the observations (ie the spatial map has maximum signal variation).
- Succeeding components must be orthogonal to all preceding components and again have the highest variance possible.
- The principal components form an uncorrelated orthogonal basis set.
- PCA can be used to reduce the dimensionality of a problem by retaining only a subset of the most significant PCs (ie useful for data reduction).

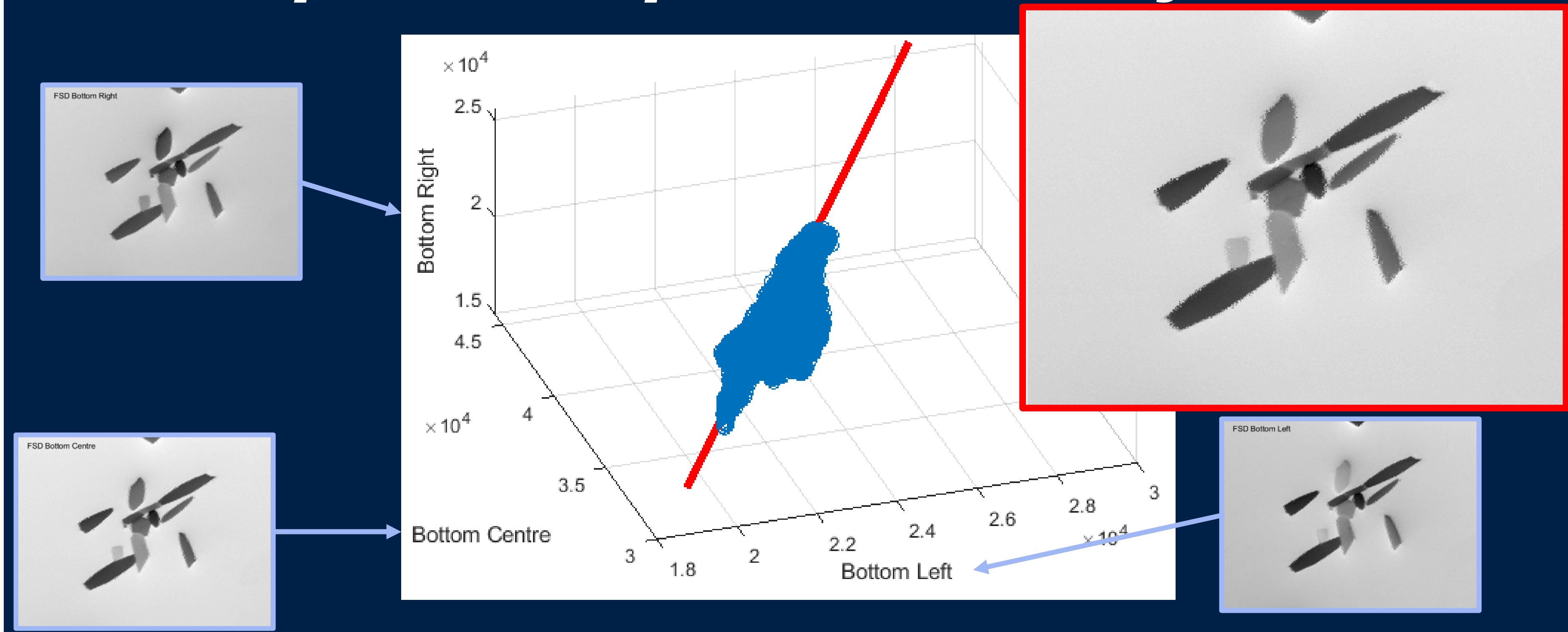
$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\mu(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

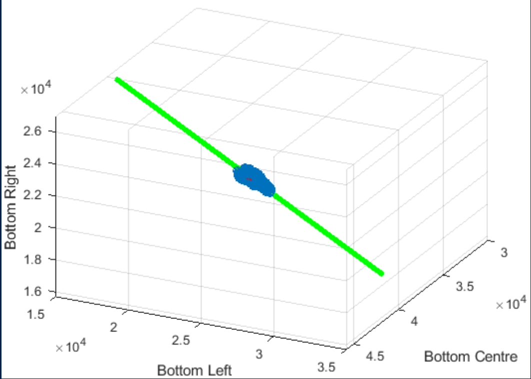
$$Covar(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\big(x_i - \mu(X)\big)\big(y_i - \mu(Y)\big)$$

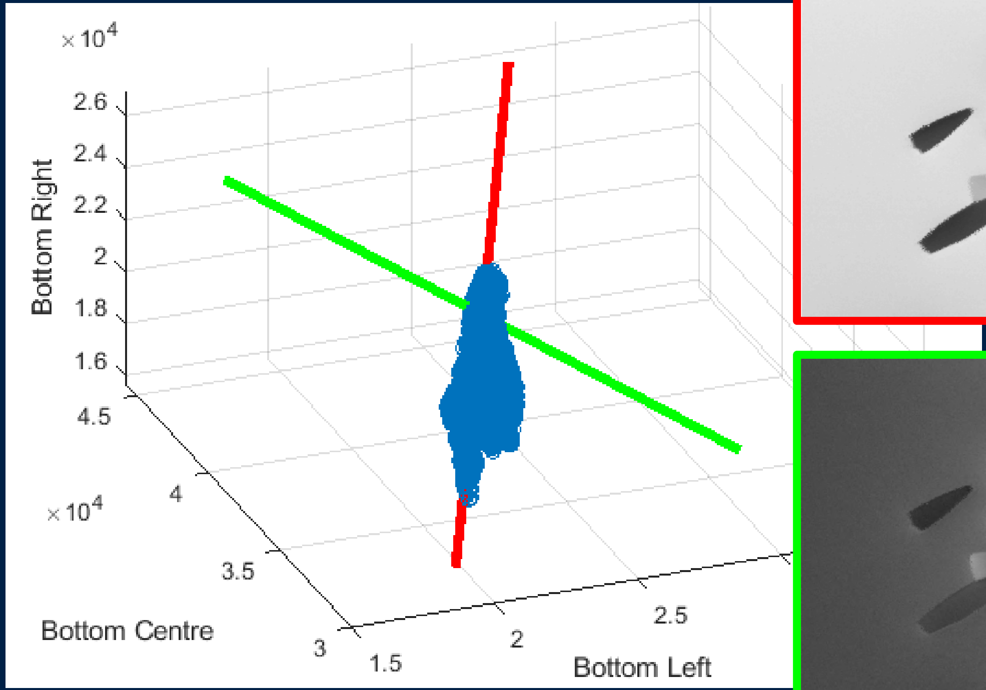


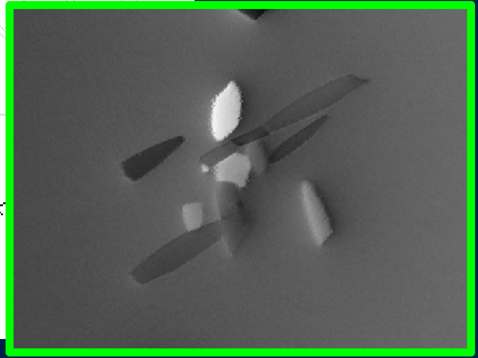
Principal Component Analysis



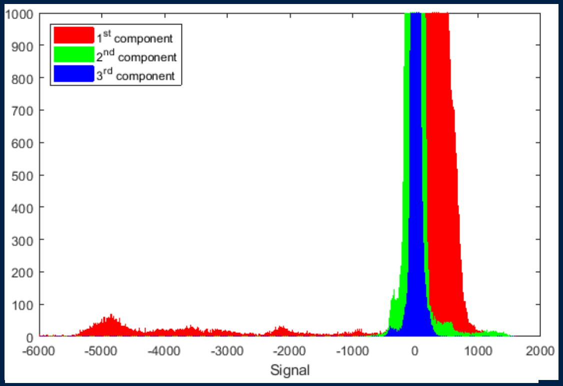
Principal Component Analysis

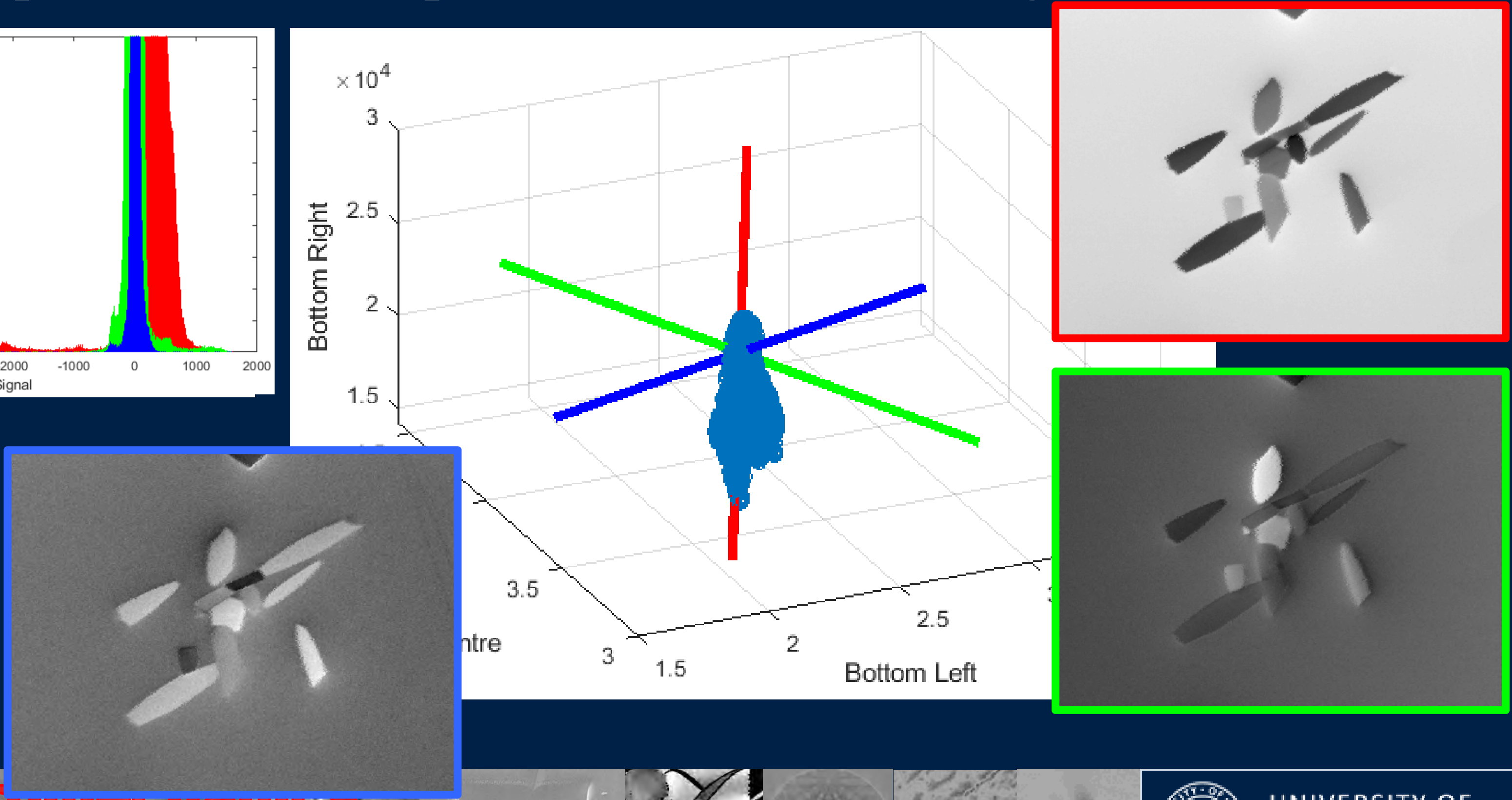




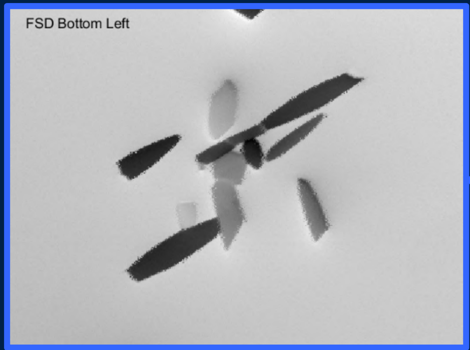


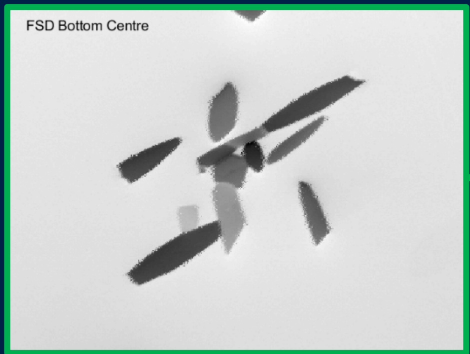
Principal Component Analysis

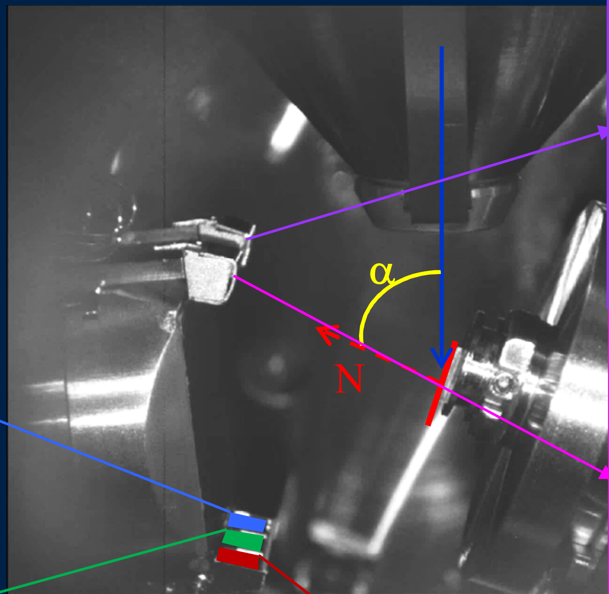


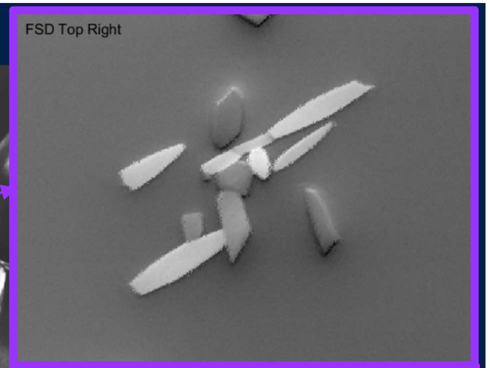


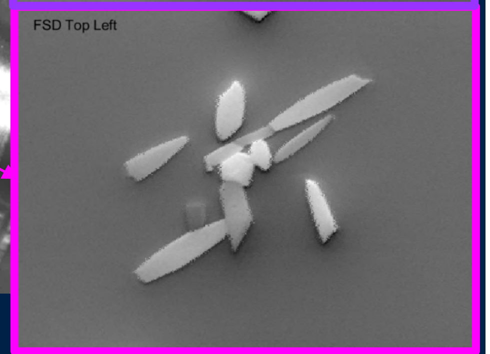
FSD Imaging











More than 3 input variables?

Same idea but can't easily plot/visualise the process – 5D plots needed

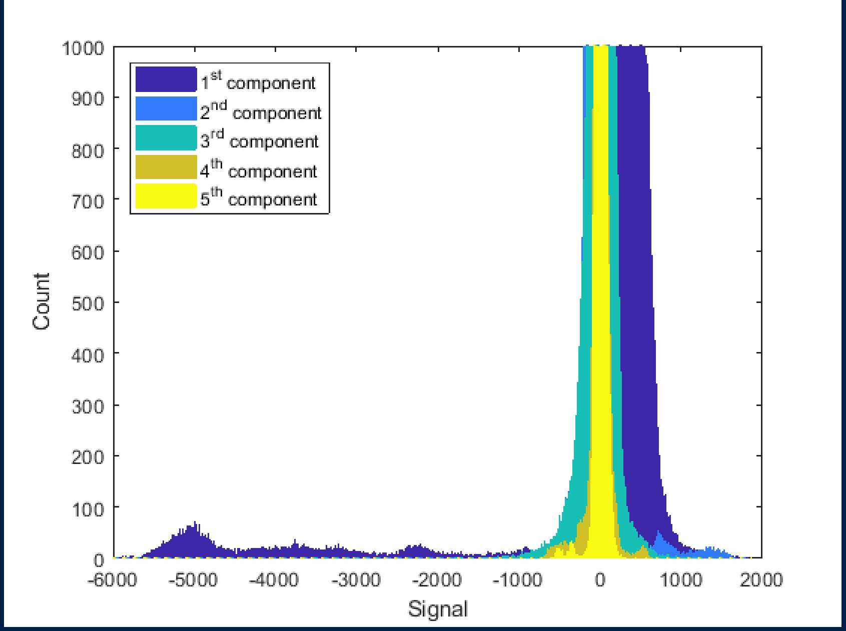
FSD Imaging

Using 5 FSD signals

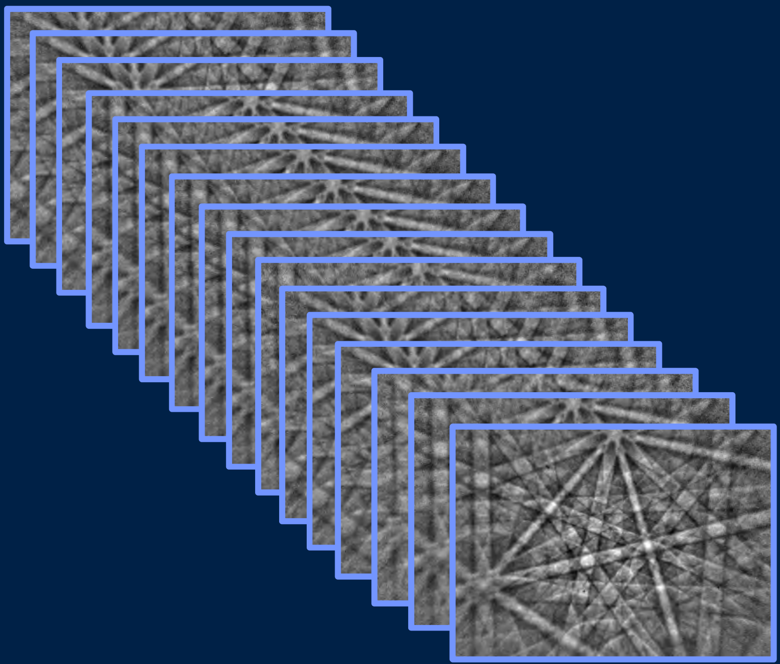
Same idea but can't easily plot/visualise the process – 5D scatter plots needed

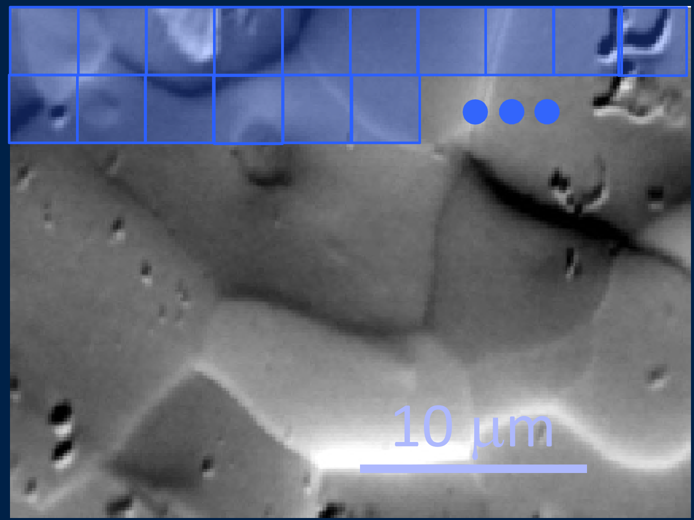
Note little new information added by last few components

The EBSD detector measures intensities at many pixels → extend process further…



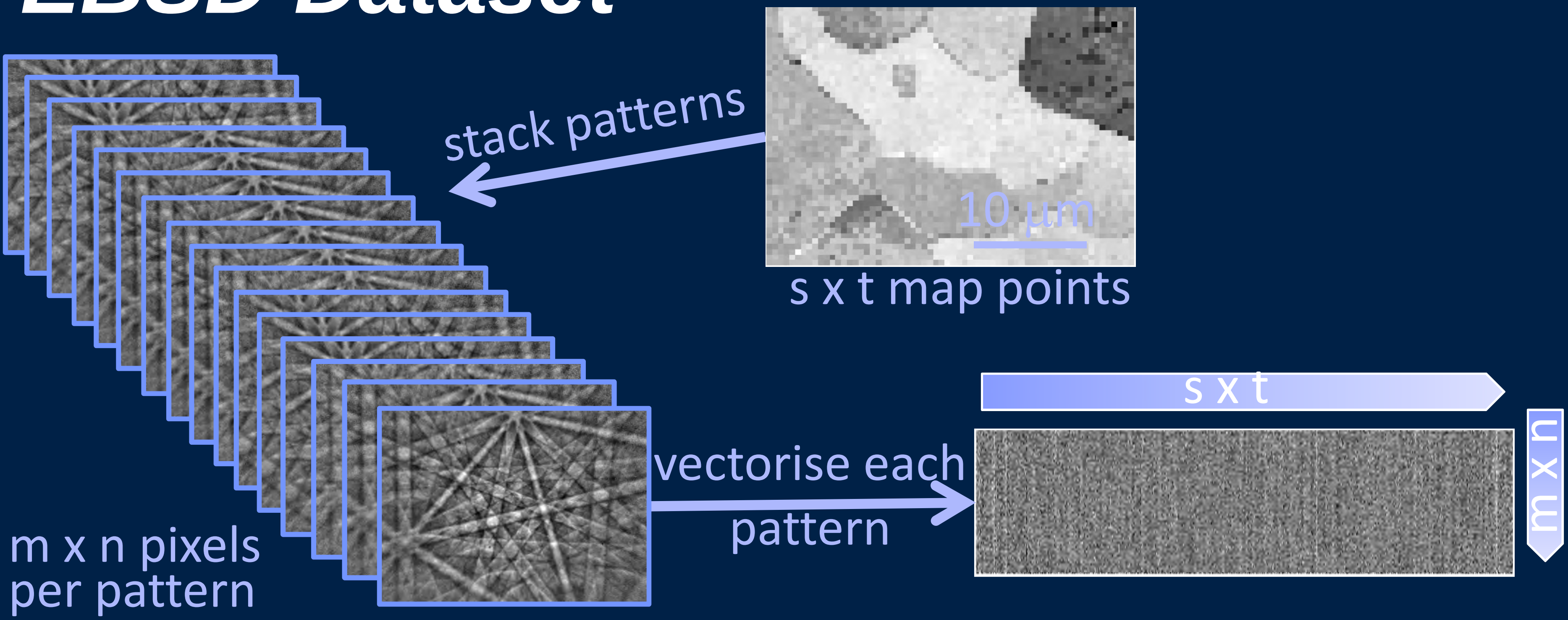
EBSD Dataset





Ferritic Steel Sample – David M Collins

EBSD Dataset

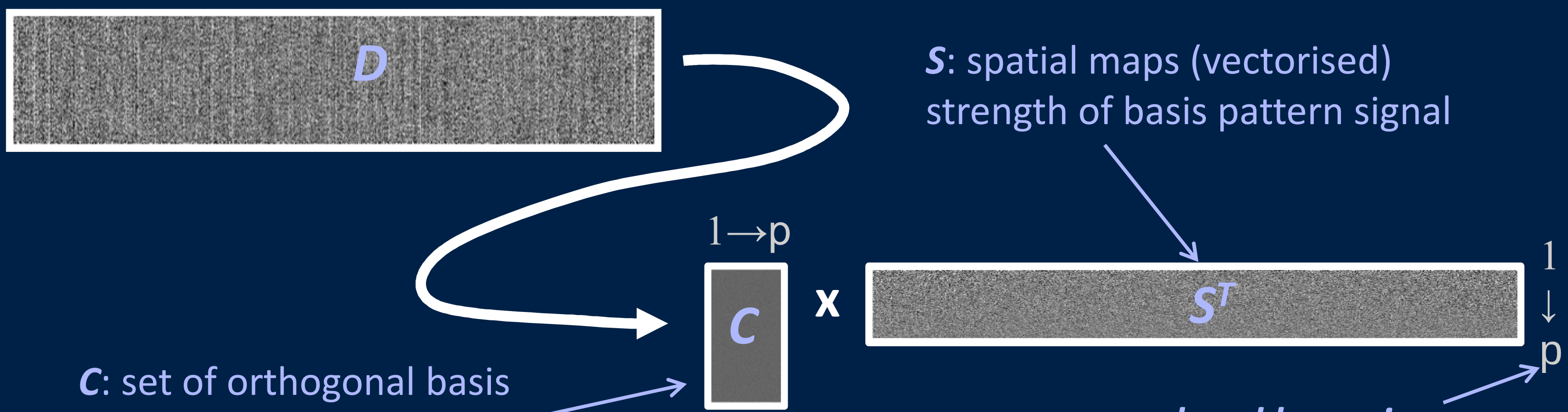


Applications of Multivariate Statistical Methods and Simulation Libraries to Analysis of EBSD and TKD Datasets
Wilkinson, Collins, Zayachuk, Korla, Vilalta-Clemente (2019) Ultramicroscopy **196** 88-98

PCA of EBSD/TKD Patterns

Principal Component Analysis

$$\boldsymbol{D}_{[\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]} = \boldsymbol{C}_{[\mathrm{m}\times\mathrm{n},\mathrm{p}\leq\mathrm{m}\times\mathrm{n}]}\, \boldsymbol{S}^{T}_{[\mathrm{p}\leq\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]}$$



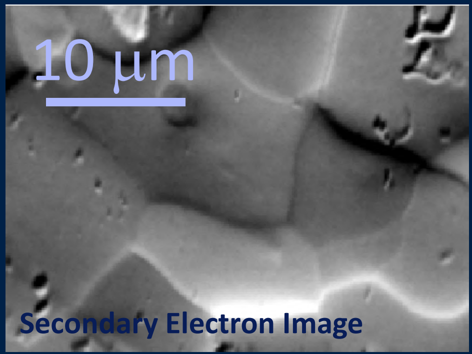
S: spatial maps (vectorised)
strength of basis pattern signal

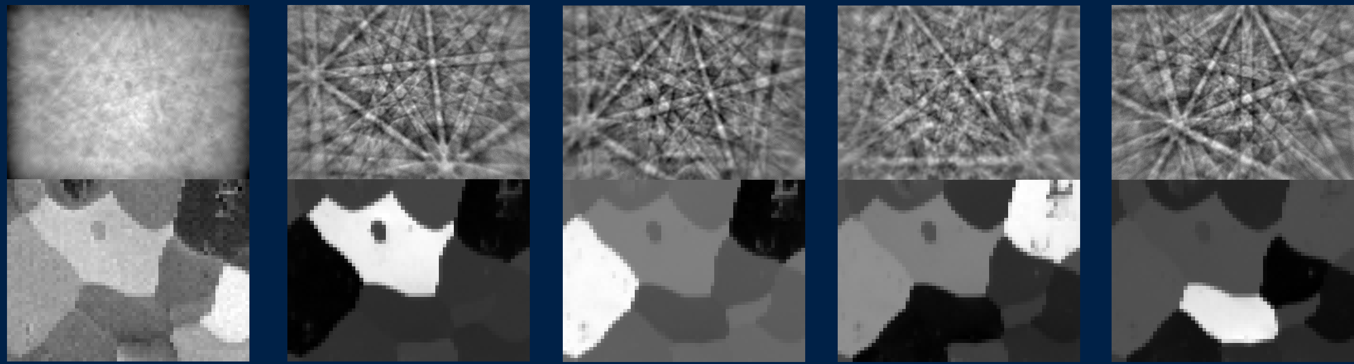
C: set of orthogonal basis patterns (vectorised)

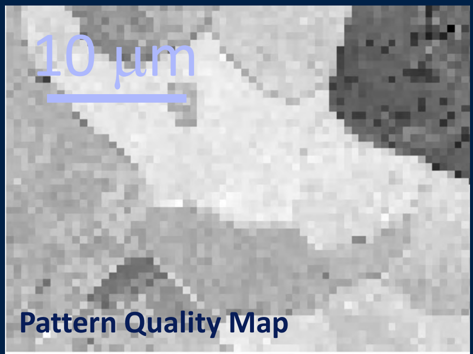
ordered by variance

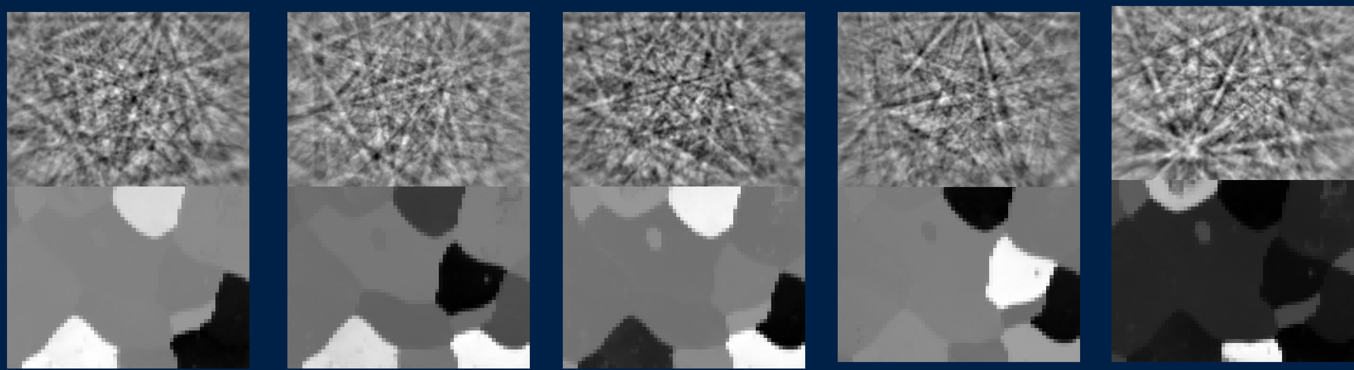
Applications of Multivariate Statistical Methods and Simulation Libraries to Analysis of EBSD and TKD Datasets
Wilkinson, Collins, Zayachuk, Korla, Vilalta-Clemente (2019) Ultramicroscopy **196** 88-98

PCA – Example: Ferritic Steel

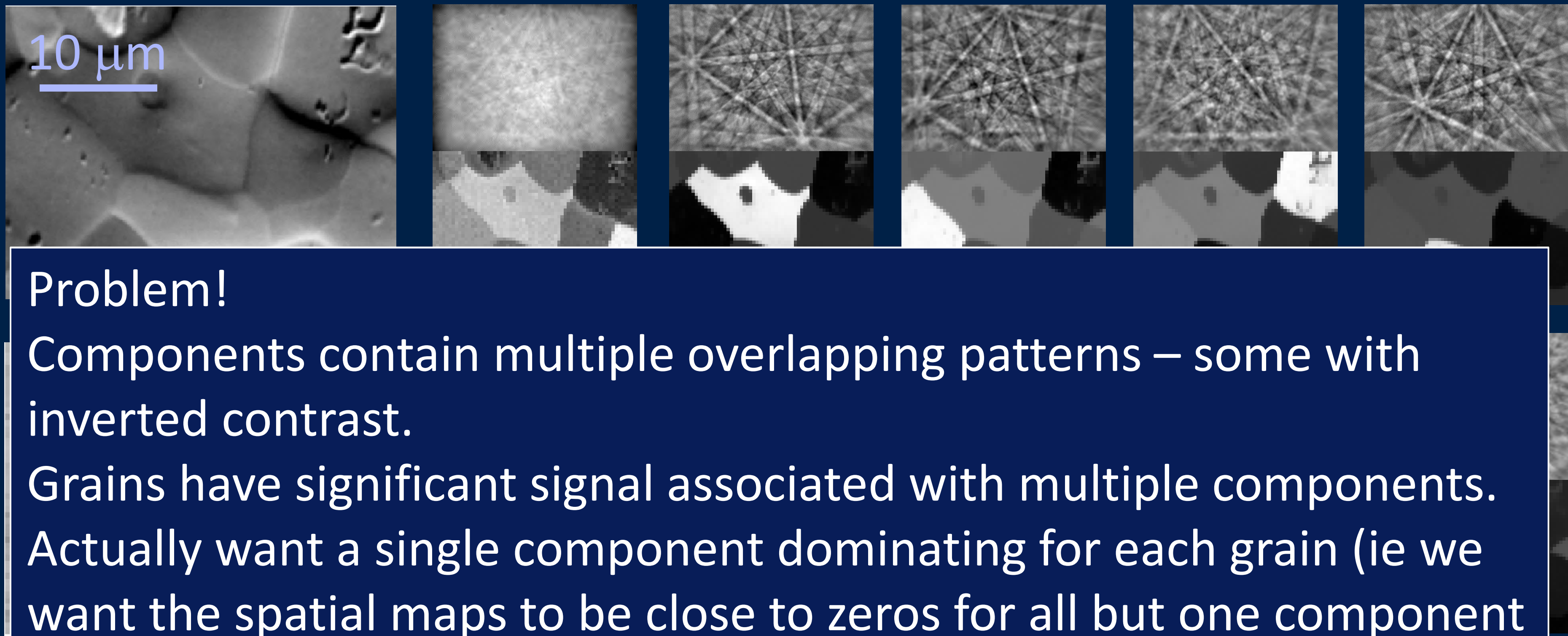








PCA – Example: Ferritic Steel



Problem!

Components contain multiple overlapping patterns – some with inverted contrast.

Grains have significant signal associated with multiple components.

Actually want a single component dominating for each grain (ie we want the spatial maps to be close to zeros for all but one component – some times called simplicity)

Varimax Basis Vectors

Principal Component Analysis: $\boldsymbol{D}_{[\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]} = \boldsymbol{C}_{[\mathrm{m}\times\mathrm{n},\mathrm{p}\leq\mathrm{m}\times\mathrm{n}]}\, \boldsymbol{S}^{T}_{[\mathrm{p}\leq\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]}$

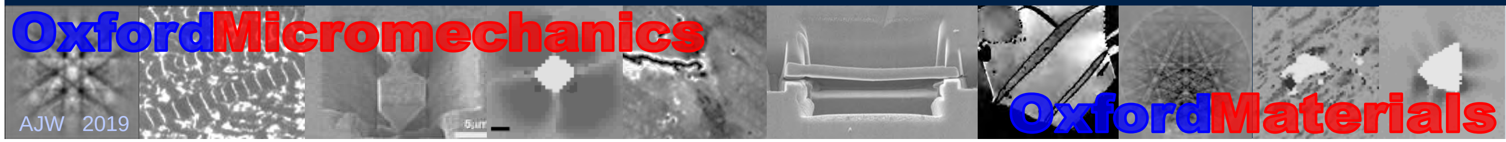
- Apply (hyper-) rotation to basis vectors
- the '*Varimax*' rotation transforms the basis vectors so that each observation (ie EBSD map point) tends to be dominated by a single basis vector (ie pattern).

$$\boldsymbol{D}_{[\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]} = \underbrace{\boldsymbol{C}_{[\mathrm{m}\times\mathrm{n},\mathrm{p}\leq\mathrm{m}\times\mathrm{n}]}\, \boldsymbol{R}_{[\mathrm{p},\mathrm{p}]}}_{\textbf{Patterns}}\ \underbrace{\boldsymbol{R}^{T}_{[\mathrm{p},\mathrm{p}]}\, \boldsymbol{S}^{T}_{[\mathrm{p}\leq\mathrm{m}\times\mathrm{n},\mathrm{s}\times\mathrm{t}]}}_{\textbf{Maps}}$$

- ***R*** maximizes the sum of squared correlations between variables and factors

$$R_{\mathrm{VARIMAX}} = \arg\max_{R} \left(\frac{1}{p} \sum_{j=1}^{k} \sum_{i=1}^{p} (\Lambda R)^4_{ij} - \sum_{j=1}^{k} \left(\frac{1}{p} \sum_{i=1}^{p} (\Lambda R)^2_{ij} \right)^2 \right)$$

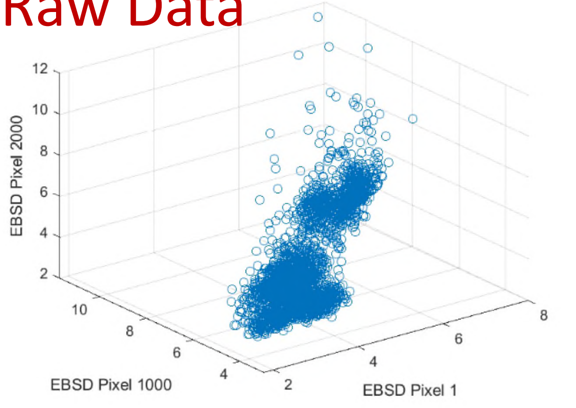
Applications of Multivariate Statistical Methods and Simulation Libraries to Analysis of EBSD and TKD Datasets
Wilkinson, Collins, Zayachuk, Korla, Vilalta-Clemente (2019) Ultramicroscopy **196** 88-98

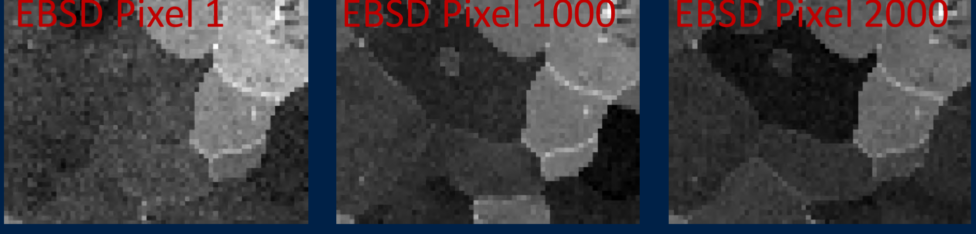


UNIVERSITY OF OXFORD

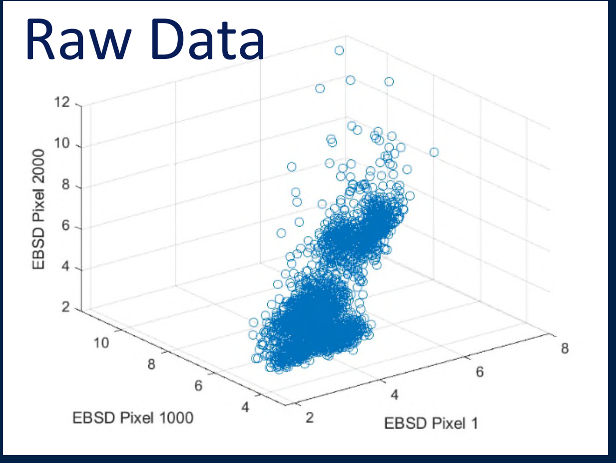
PCA of EBSD Data

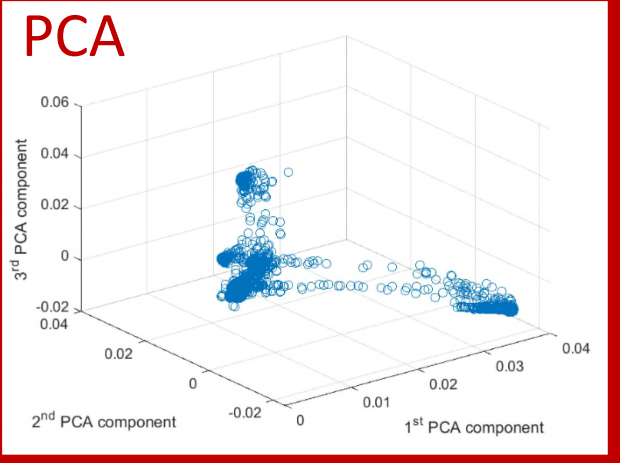
Raw Data

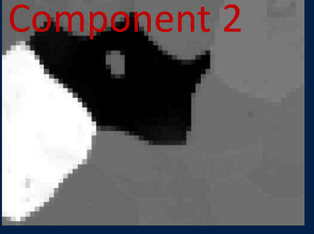




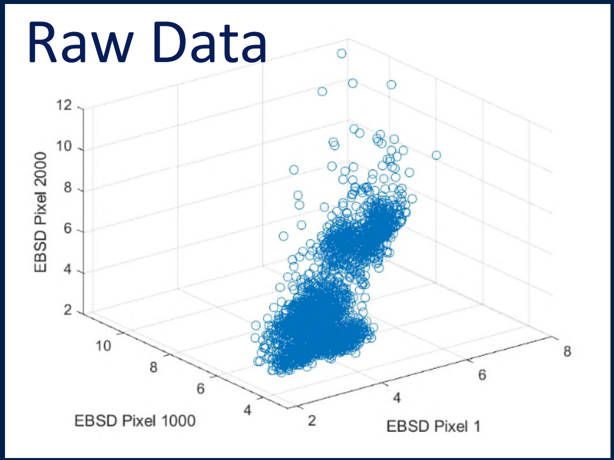
PCA of EBSD Data

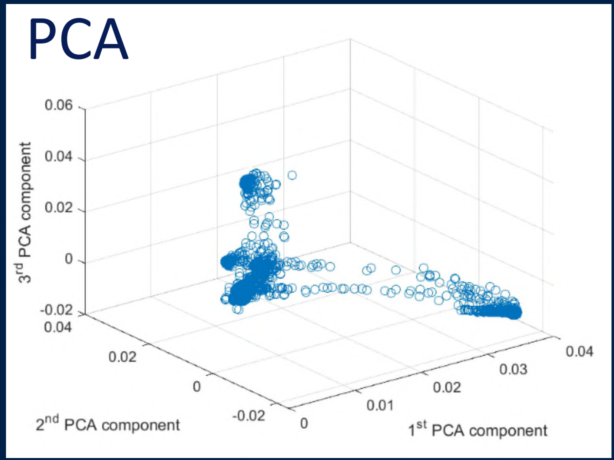


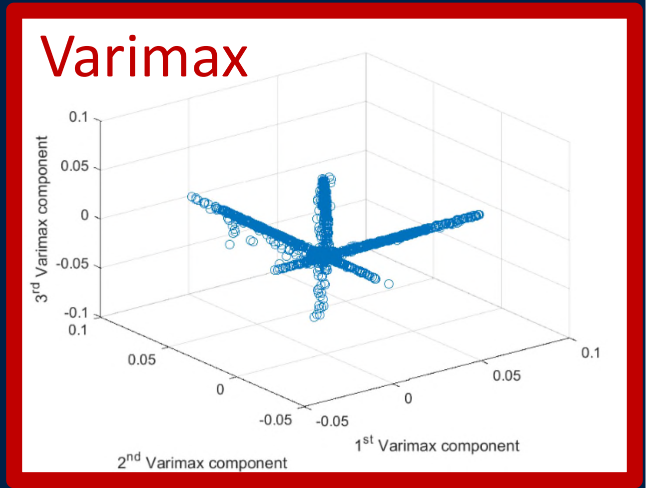




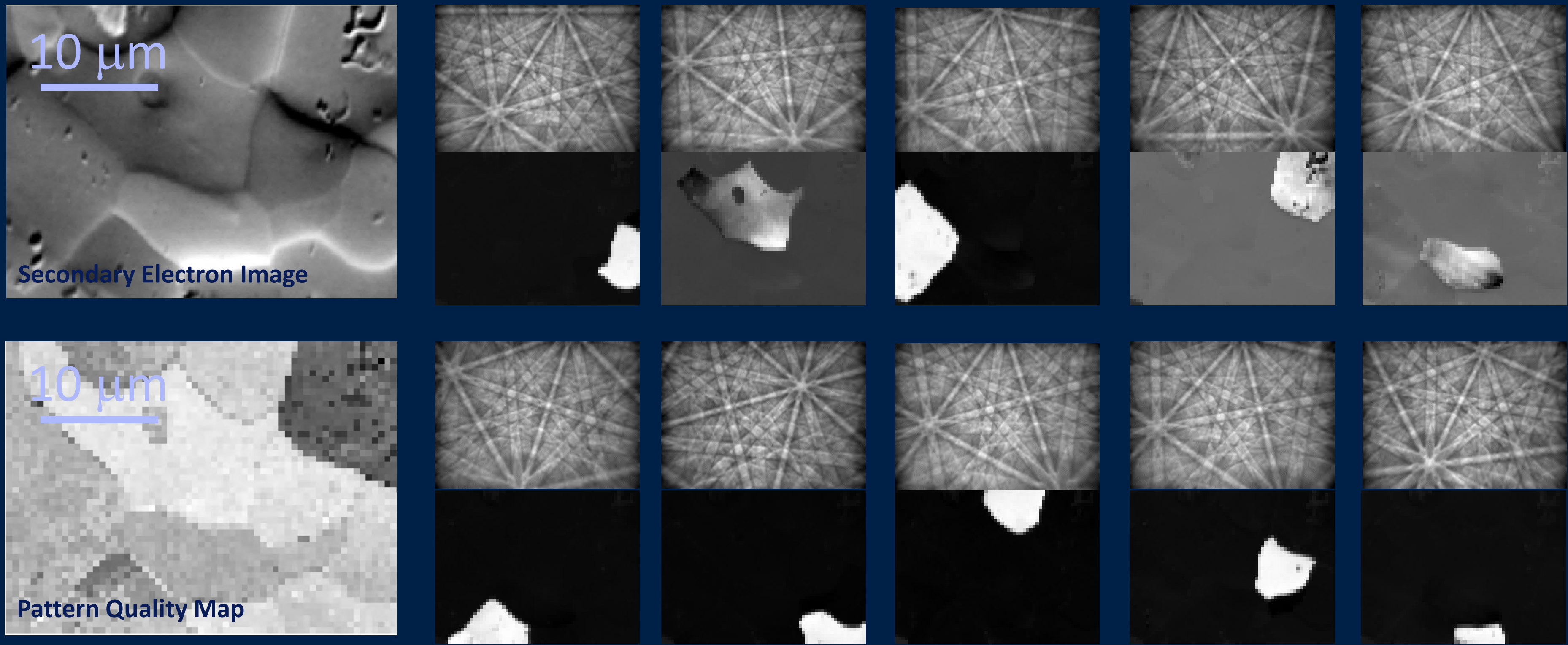
PCA of EBSD Data



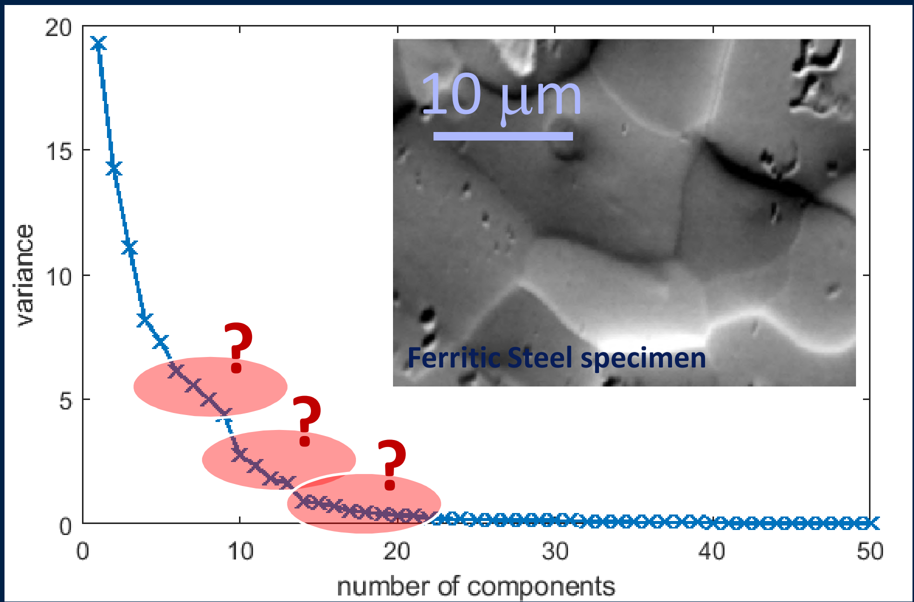




Varimax – Example: Ferritic Steel

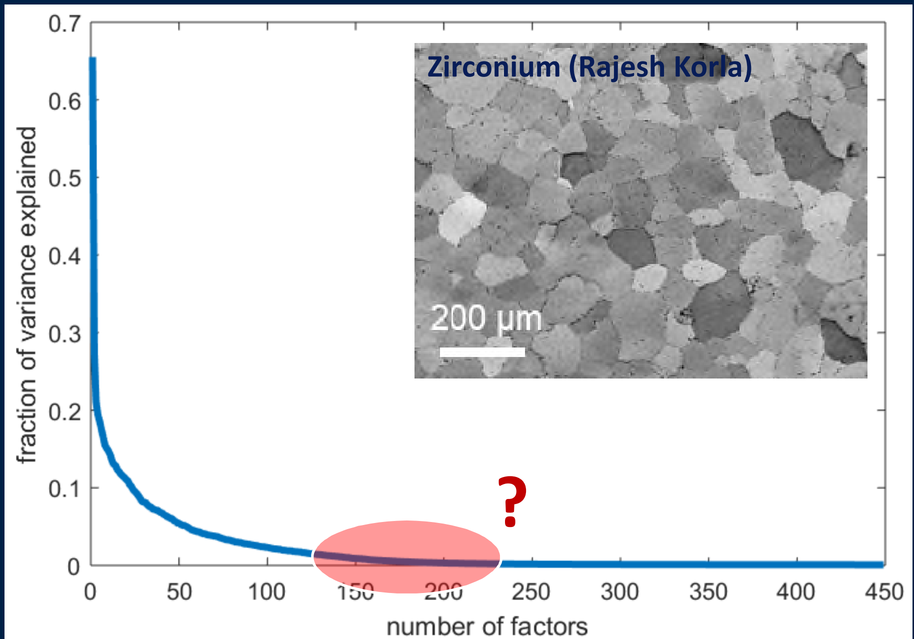


How many components?



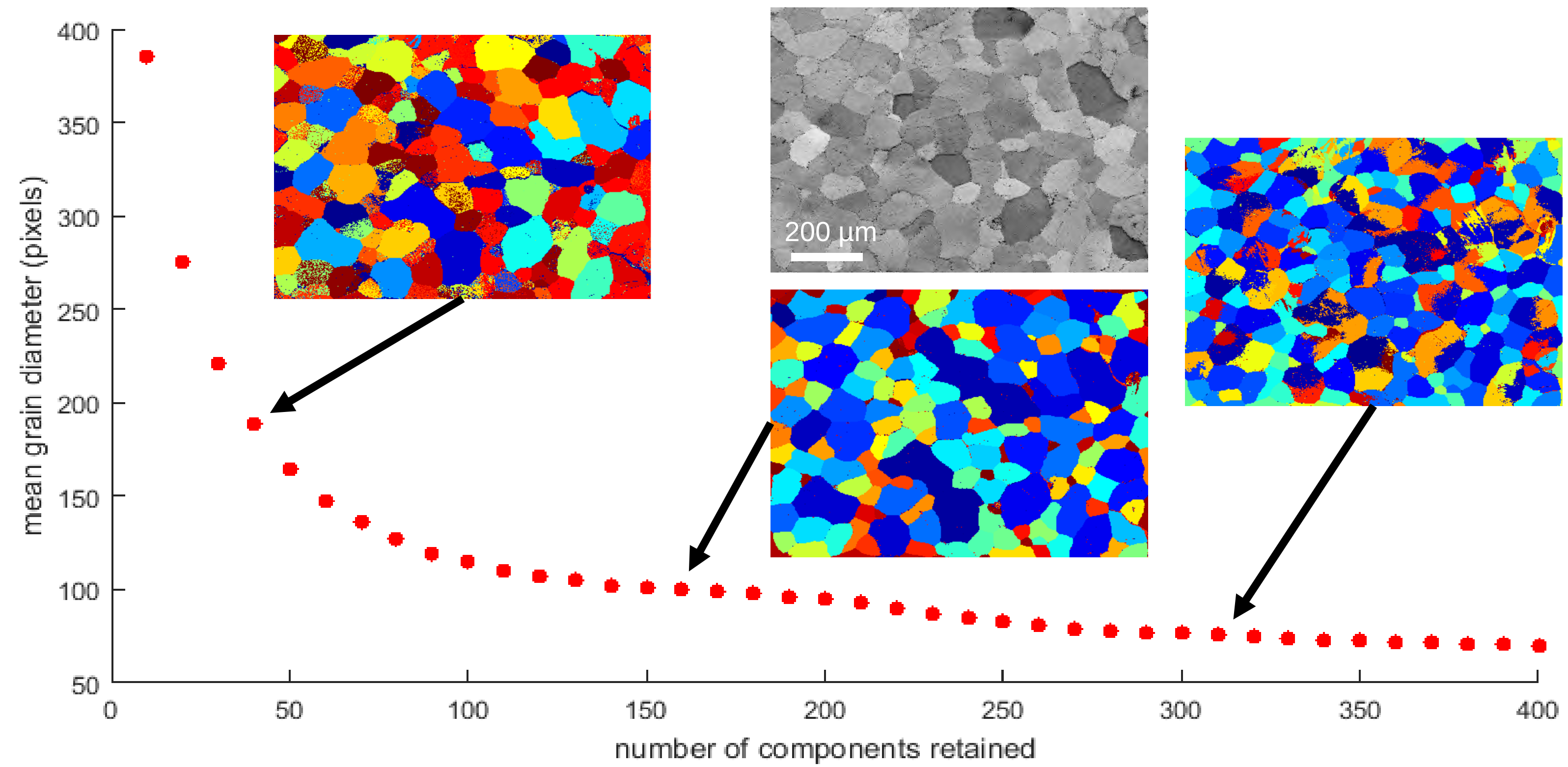
- Scree plot
  - plot variance associated with successive components
  - ‘elbow’ shows transition from signal to noise
- but
  - Subjective
  - Multiple elbows
  - Smooth transition

How many components?

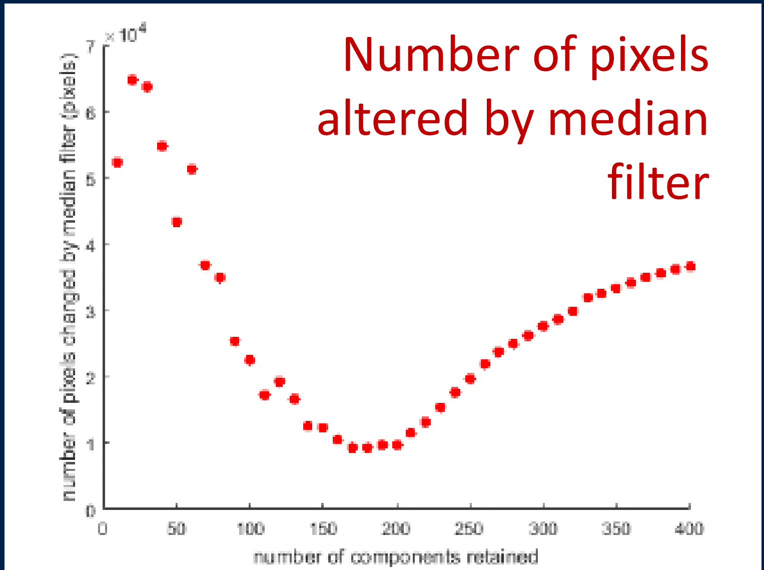


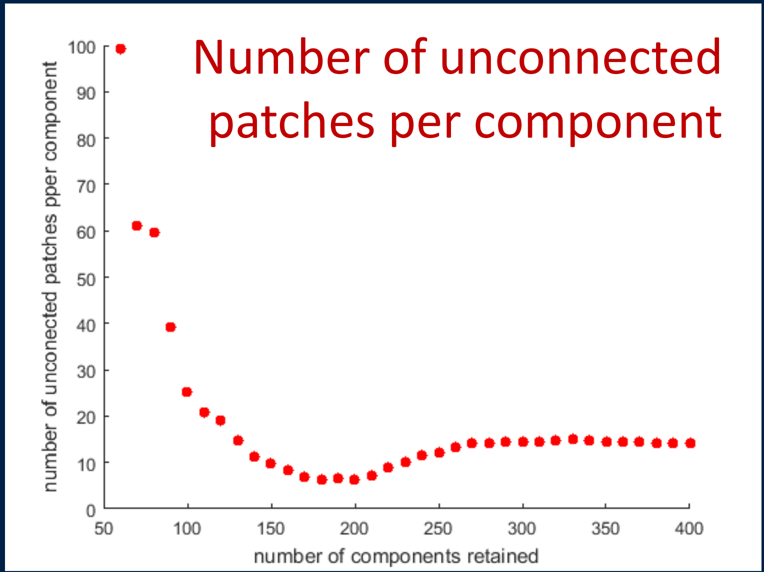
- Scree plot
  - plot variance associated with successive components
  - 'elbow' shows transition from signal to noise
- but
  - Subjective
  - Multiple elbows
  - Smooth transition

How many components?



How many components?



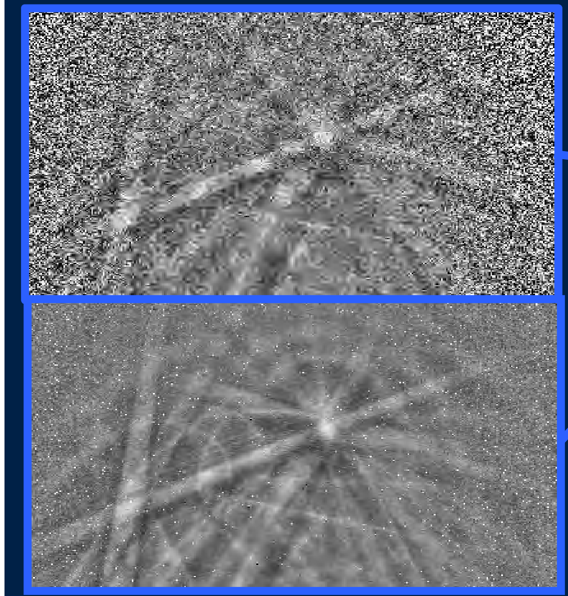


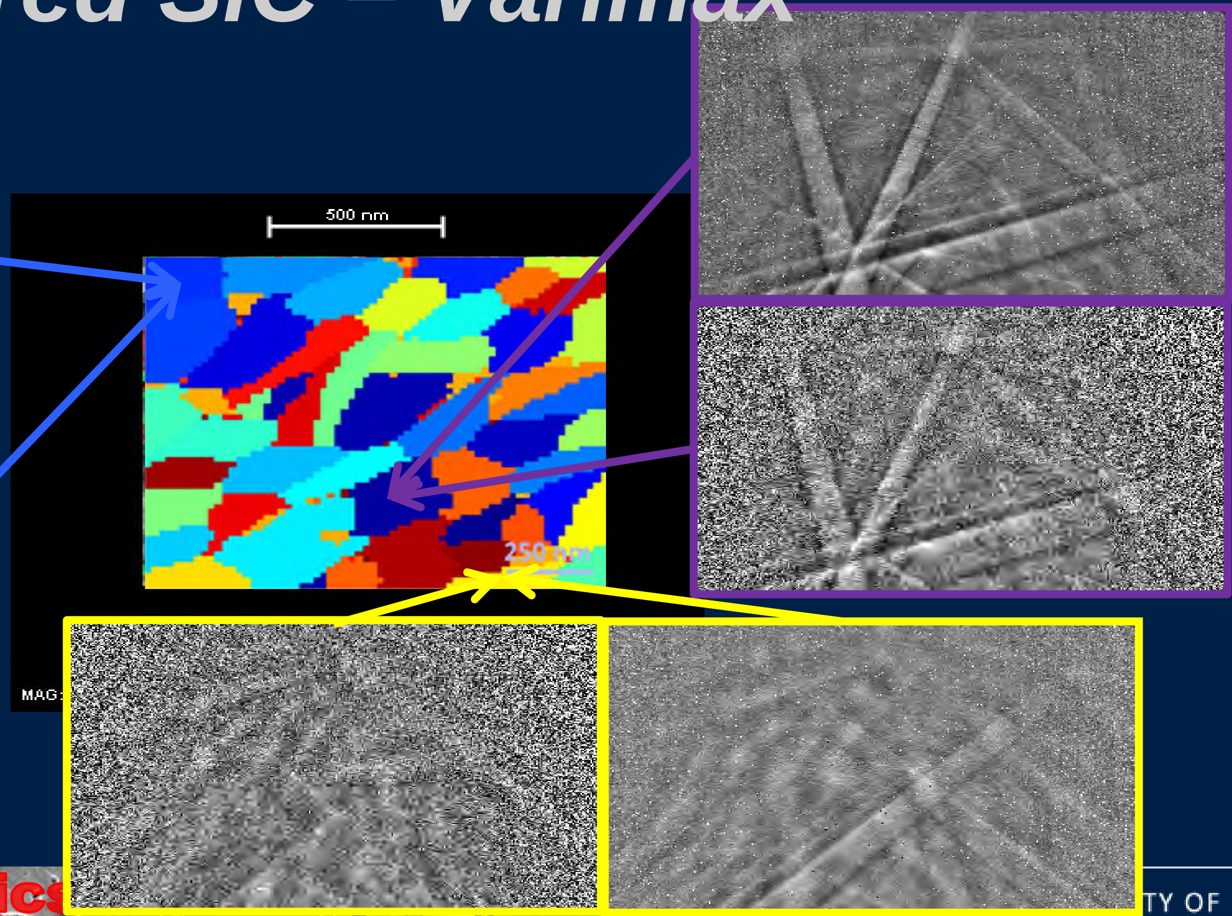
- Use more specific measures reflecting expected form of grain structure
- Aim to allow more components than stitch back together through indexing

Characteristic Patterns

- Enhanced signal to noise for characteristic patterns (important for specimens that are difficult to prepare or suffer beam damage…). 'Smart averaging' over similar microstructural domains identified direct from data.
- Subtle effects can be discriminated (eg polarity)
- Reduced pattern numbers for indexing
  - Hough-based → try AstroEBSD
  - Template matching → significant reduction of computational time (stay tuned for next talk from Katharina Marquardt)

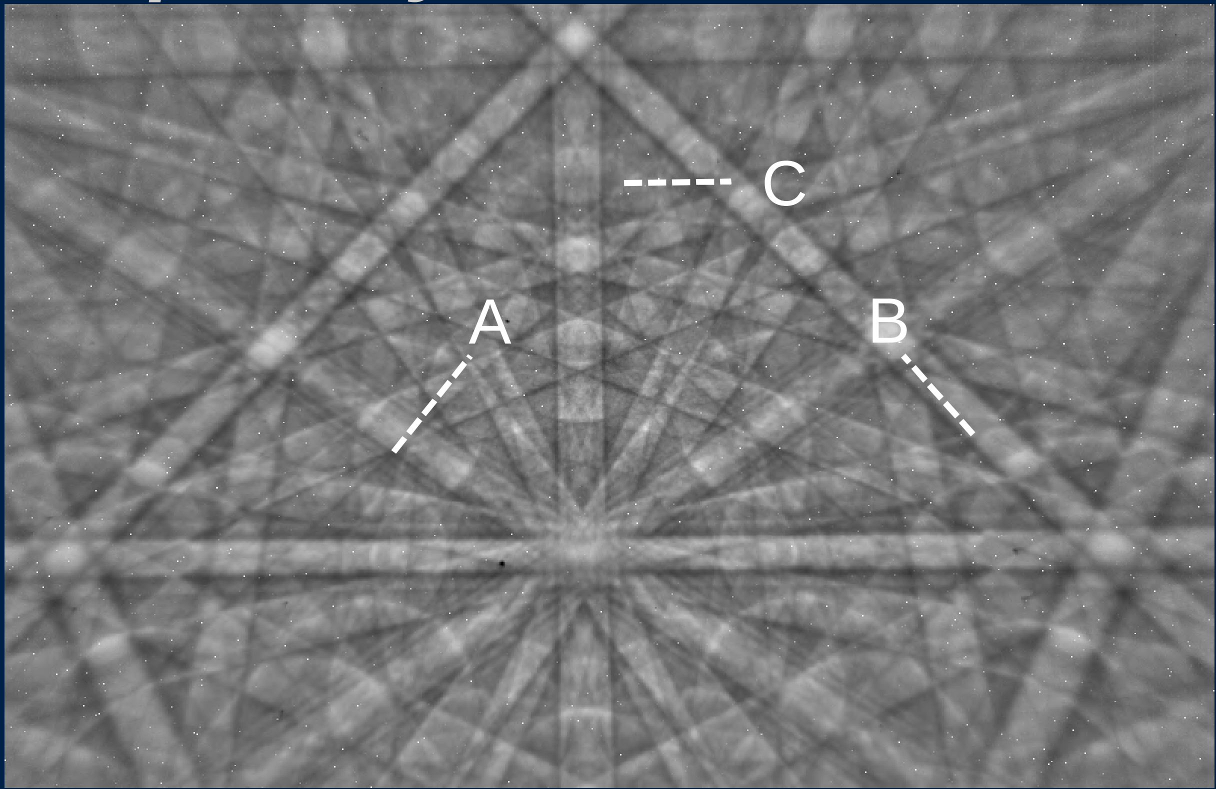
Nanostructured SiC – Varimax

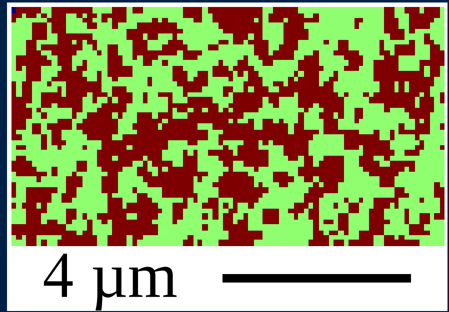


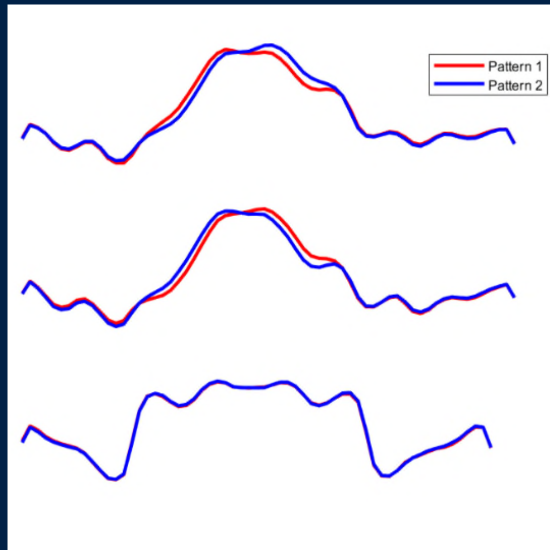


TKD data from
Yevhen Zayachuk

GaP polarity +/- Domains

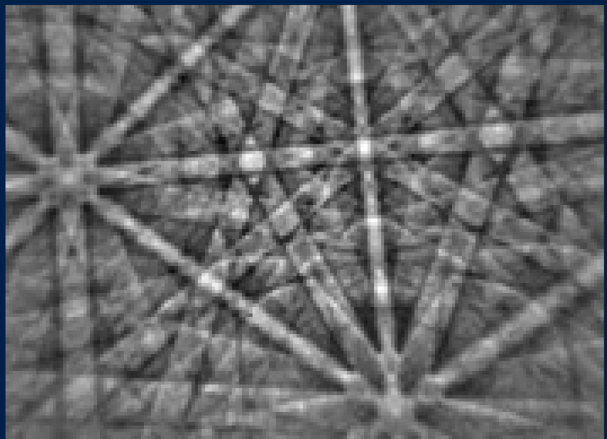




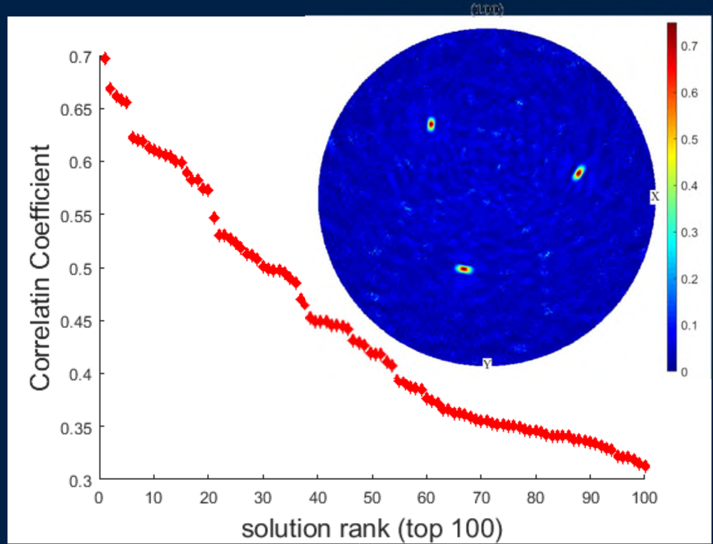


Arantxa Villata-Clemente

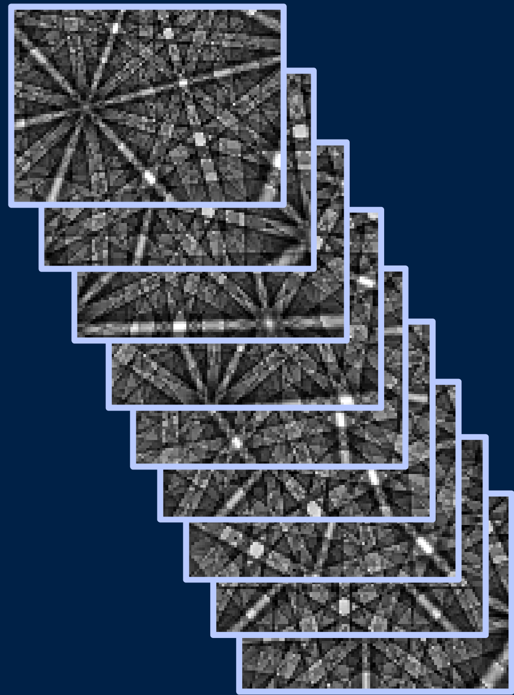
Template Matching for Indexing EBSD



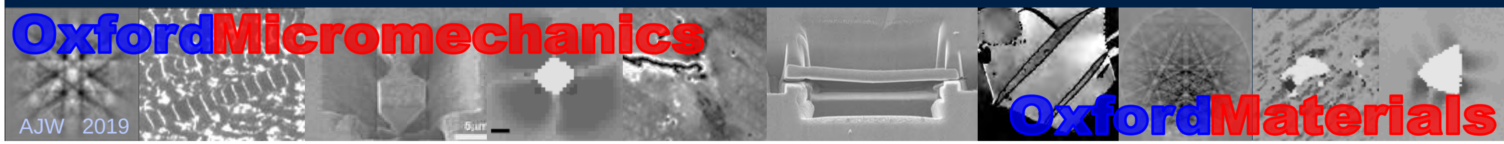
Characteristic Pattern from multivariate statistical analysis



Normalised Cross-Correlation



Simulated Library



UNIVERSITY OF OXFORD