Latent Variable and Generative Models

Ian Nabney
Aston University

Outline

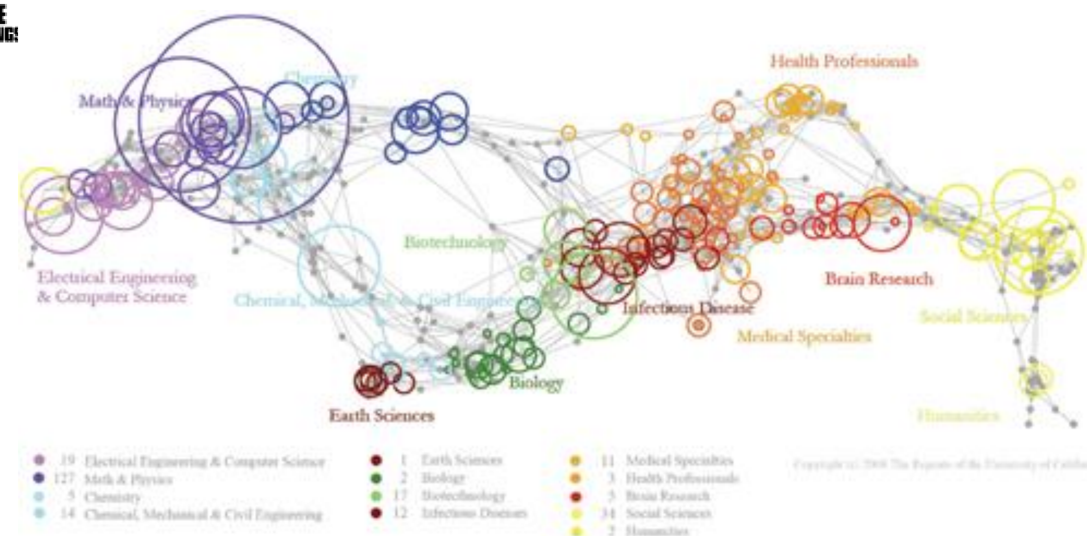
- Comparison of generative and non-generative models
- Principal Component Analysis (PCA) as a generative model
- Generative Topographic Mapping (GTM)
- Leveraging generative models
- Examples/case studies

What is Visualisation?

- Goal of visualisation is to present data in a human-readable way.
- Visualisation is an important tool for developing a better understanding of large complex datasets. It is particularly helpful for users who are not specialists in data modelling.
  - Detection of outliers.
  - Clustering and segmentation.
  - Aid to feature selection.
  - Feedback on results of analysis: seeing what you are doing.
- Two aspects: information visualisation and data projection.
- The goal is to project data to a lower-dimensional space (usually 2d) while preserving as much information or structure as possible.

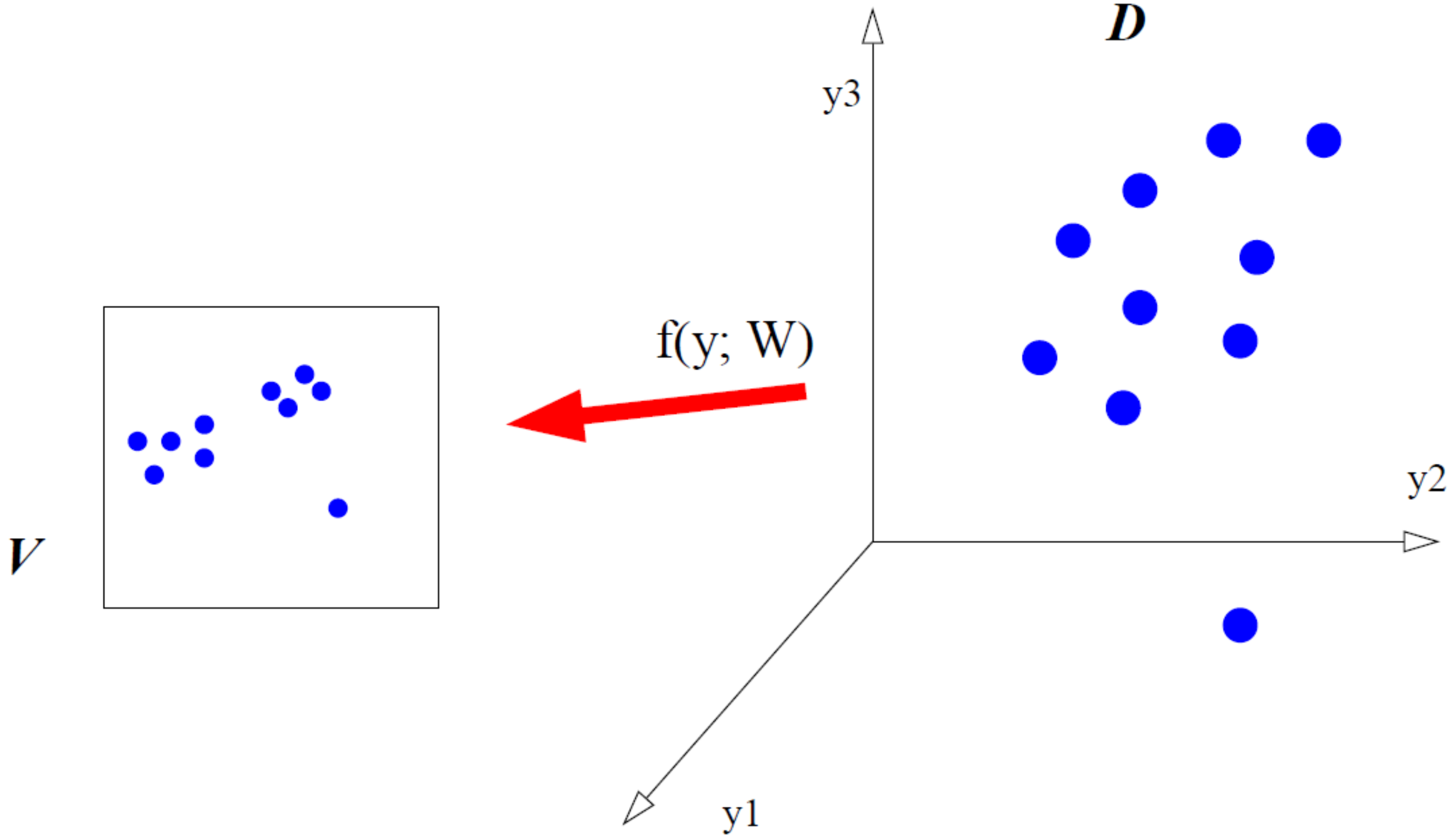
Visualisation Methods

Word Cloud (www.wordle.net)



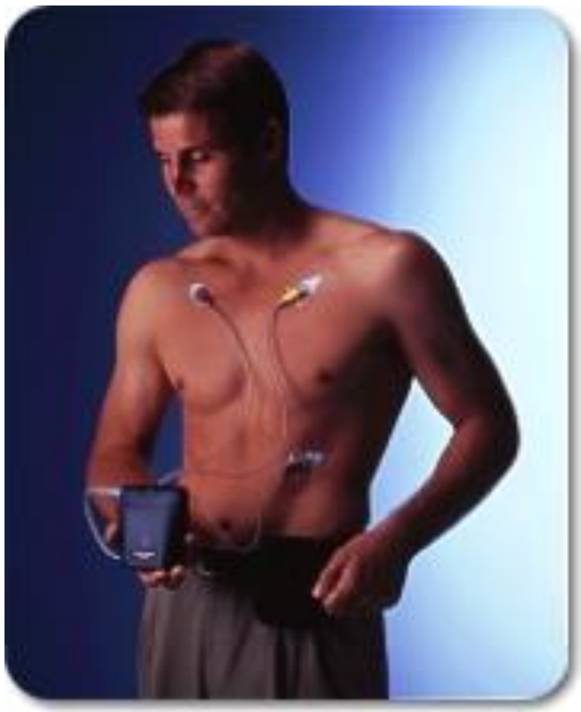
UCSD Map of Science (Zoss and Borner)

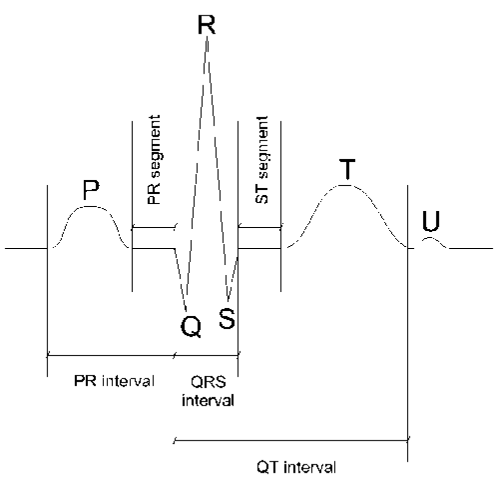
Data Projection

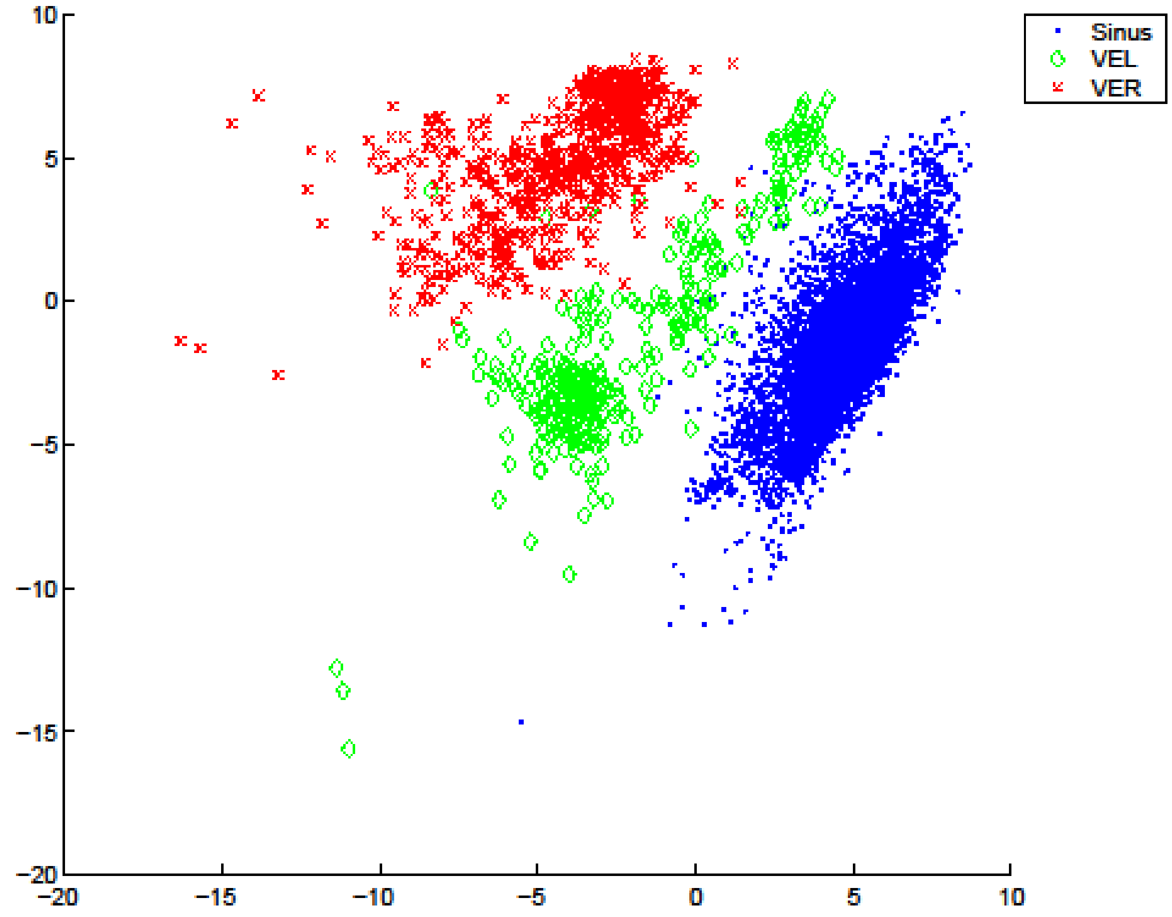


- Define f to optimise some criterion.
- PCA is minimal variance.
- Sammon mapping is minimal stress.

ECG Analysis







What can we learn from this?

Linear Projection

- What is the simplest way to project data? A linear map.
- What is the best way to linearly project data? Want to preserve as much information as possible.
- If we assume that information is measured by variance this implies choosing new coordinate axes along directions of maximal variance; these can be found by analysing the covariance matrix of the data.
- This gives Principal Component Analysis (PCA).
- For large datasets, the end result is usually a circular blob in the middle of the screen.

Principal Component Analysis

- Let S be the covariance matrix of the data so that

$$S_{ij} = \frac{1}{N}\sum_{n}(x_i^n - \bar{x}_i)\left(x_j^n - \bar{x}_j\right)$$

The first *q* principal components are the first *q* eigenvectors w_j of *S*, ordered by the size of the eigenvalues α_j. The percentage of the variance explained by the first *q* PCs is

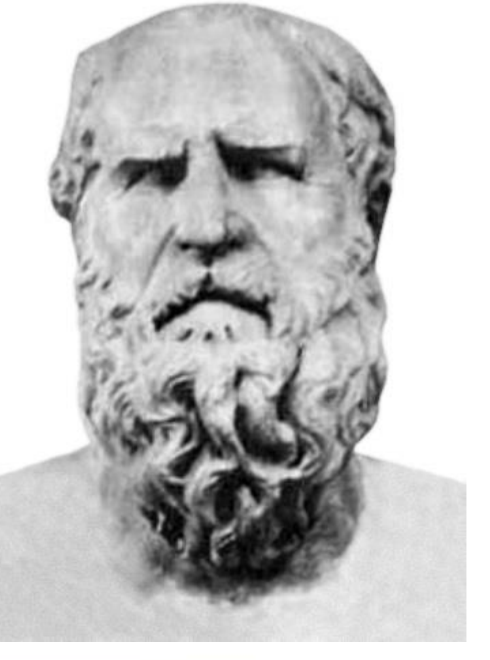
$$\frac{\sum_{j=1}^{q} \alpha_j}{\sum_{j=1}^{d} \alpha_j}$$

where the data dimension is *d*.

- These vectors are orthonormal (perpendicular and unit length). The variance when the data is projected onto them is maximal.

Latent Variable Models

- The projection approach is one way of reducing the data dimensionality.
- An alternative view is to hypothesis how the data might have been generated.
- *A hidden connection is stronger than an obvious one: Heraclitus.*



Latent Variable Models II

- Separate the observed variables and the latent variables. Latent variables generate observations. Use (probabilistic) inference to deduce what is happening in latent variable space.
- Often use Bayes' Theorem:

$$P(L|O) = \frac{P(O|L)P(L)}{P(O)}$$

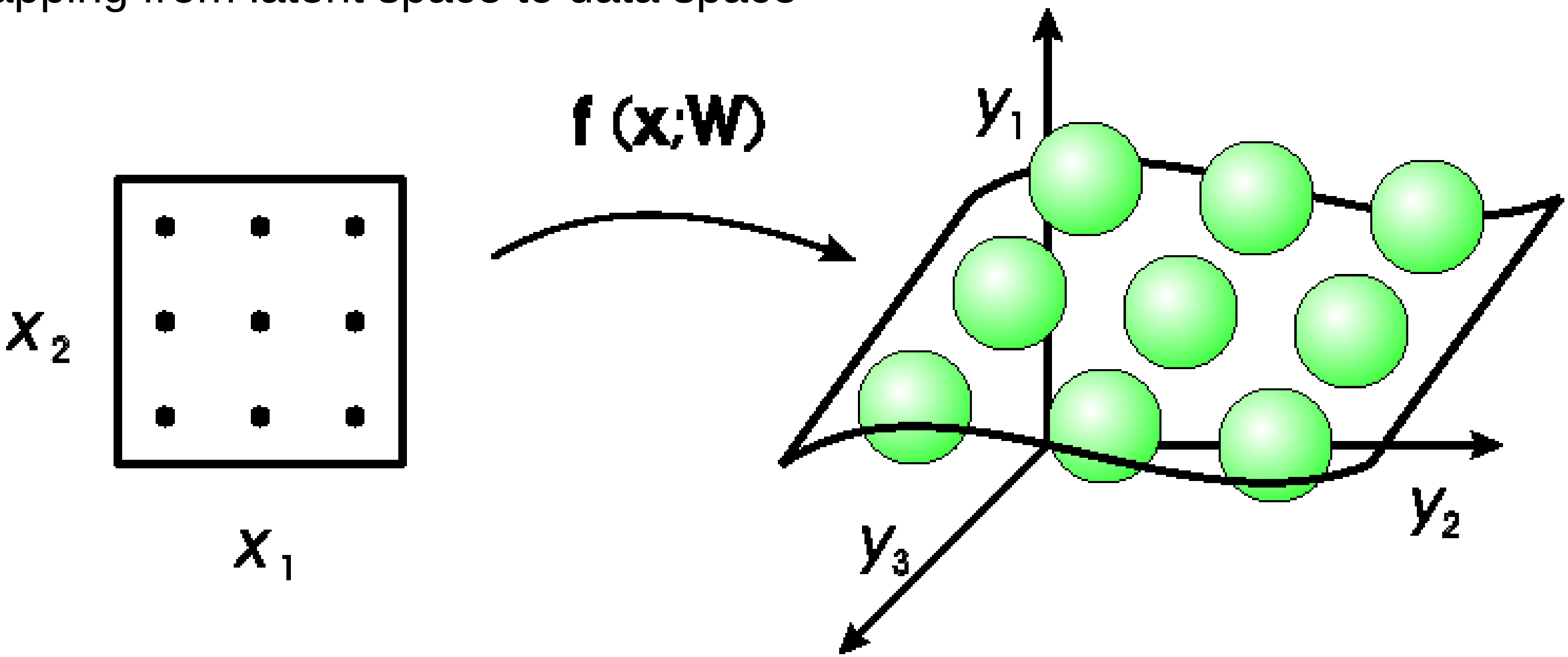
- Simplest case is PCA: *q* latent variables, a linear transformation to observation space and a single Gaussian distribution in latent space.
- Dynamic case:
  - Hidden Markov Models: discrete state space. (Speech recognition).
  - State-Space Models: continuous state space. (Tracking).

Visualisation with Density Models

- Construct a generative model for the data, mapping from a low-dimensional latent space *H* to the data space *D*.
- Maps latent variables *r* to observed variables *x* giving a probability density $p(x|r)$
- To visualise the data we map from observed variables to latent variables using Bayes' theorem.
- Plot a summary statistic of $p(r_i|x_i)$ for each data point x_i: usually the mean.

Generative Topographic Mapping

Mapping from latent space to data space



A thick rubber sheet studded with tennis balls. GTM defines p(y|x;W); use Bayes' theorem to compute p(x|y*;W) for a given point y* in data space.

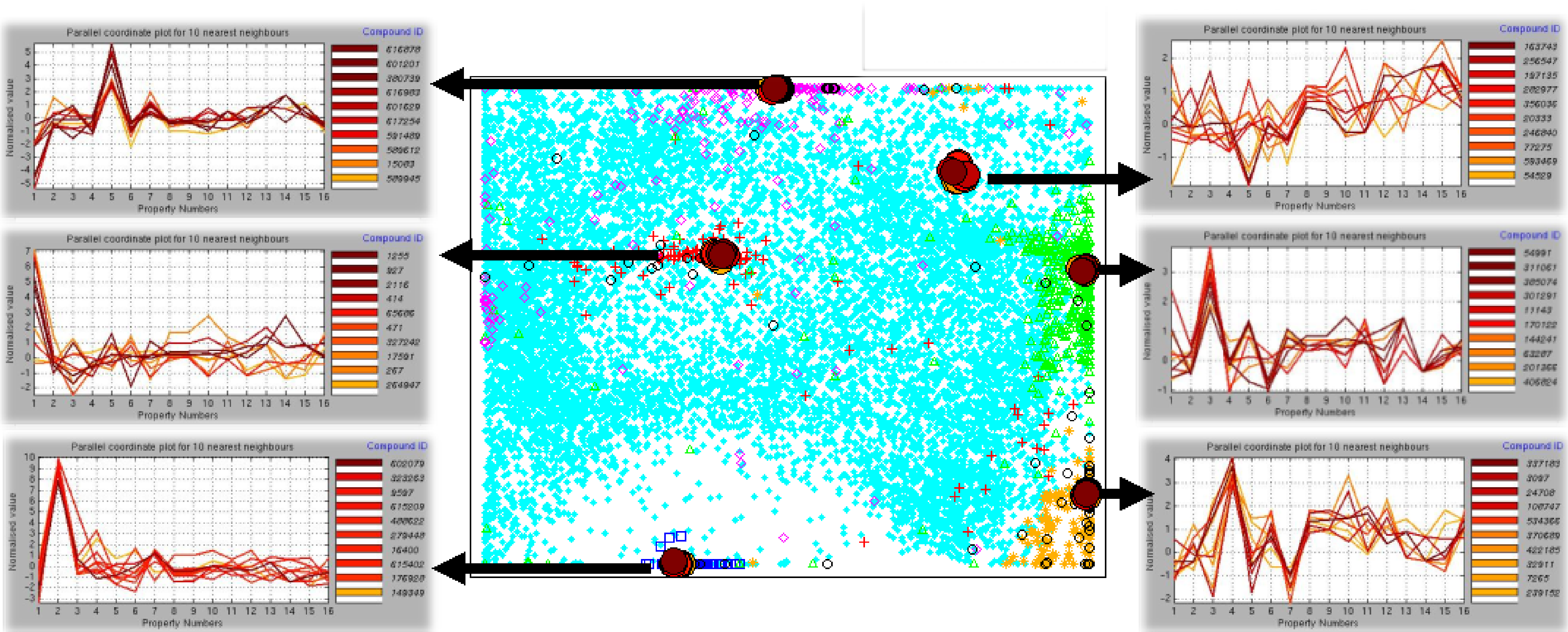
Generative Topographic Mapping II

- GTM (Bishop, Svensen and Williams) is a latent variable model with a non-linear RBF mapping a (usually two-dimensional) latent space *H* to the data space *D*.
- Data doesn't live exactly on manifold, so smear it with Gaussian noise. Introduce latent space density p(x): approximate by a data sample. This is a generative probabilistic model.
- This model assumes that the data lies close to a two-dimensional manifold; however, this is likely to be too simple a model for interesting data.
- We can measure the non-linearity of the sheet and use this to understand the visualisation plot.
- Train the model in maximum likelihood framework using an iterative algorithm (EM).

Enhancements to GTM

- Magnification factors and curvatures give more information about shape of manifold.
- Hierarchy allows the user to drill down into data; either user-defined or automated (MML) selection of sub-model positions.
- Temporal dependencies handled by GTM through Time.
- Discrete data handled by Latent Trait Model (LTM): all the other goodies work for it as well. Mixed data types also handled
- Can cope with missing data in training and visualisation.
- MML methods for feature selection.
- Structured covariance.

Interactive Visualisation Tool



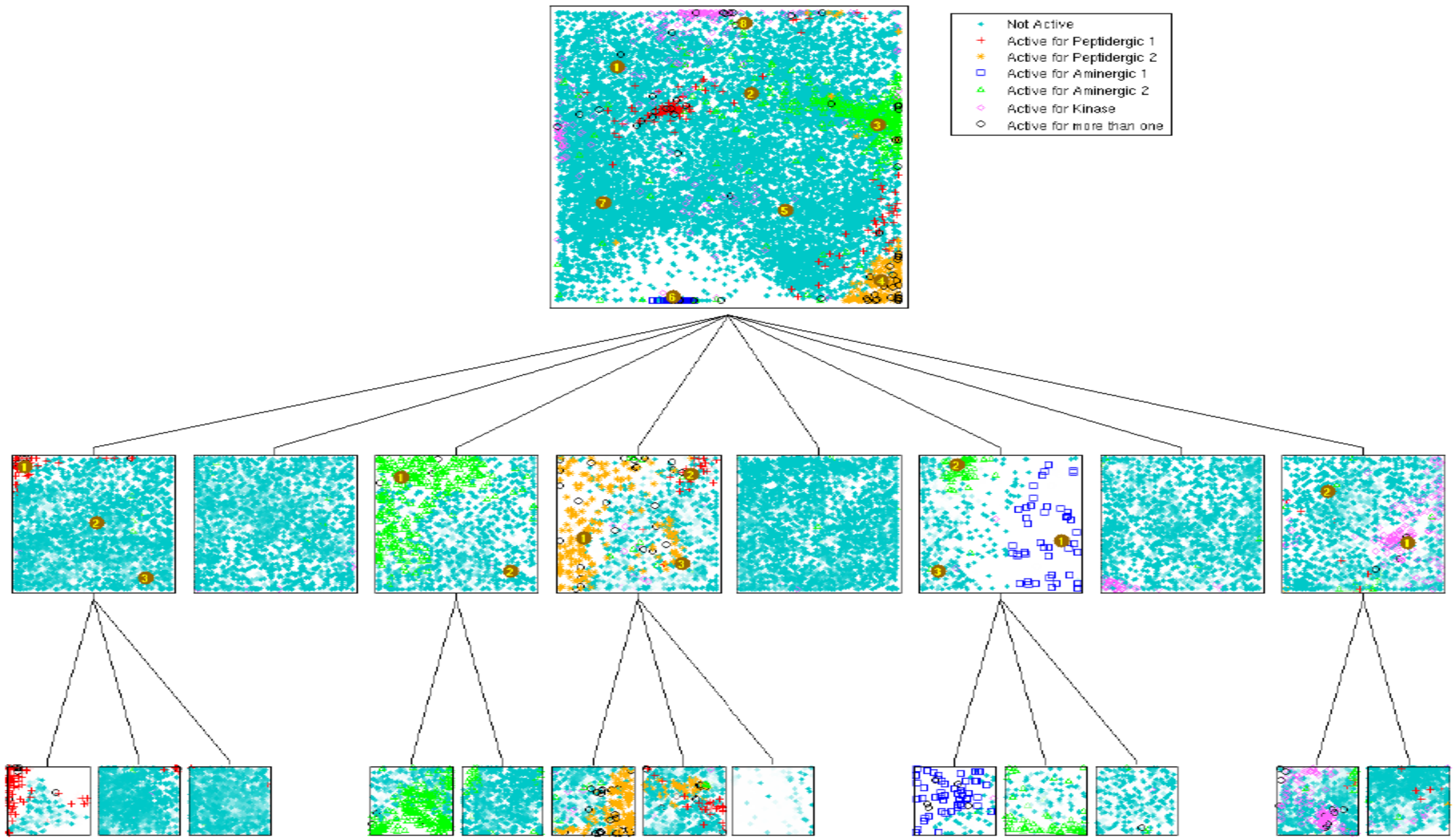
Hierarchical GTM: Drilling Down

- Bishop and Tipping introduced the idea of hierarchical visualisation for probabilistic PCA. We have developed a general framework for arbitrary latent variable models.
- Because GTM is a generative latent variable model, it is `straightforward' to train hierarchical mixtures of GTMs.
- We model the whole data set with a GTM at the top level, which is broken down into clusters at deeper levels of the hierarchy.
- Because the data can be visualised at each level of the hierarchy, the selection of clusters, which are used to train GTMs at the next level down, can be carried out interactively by the user.

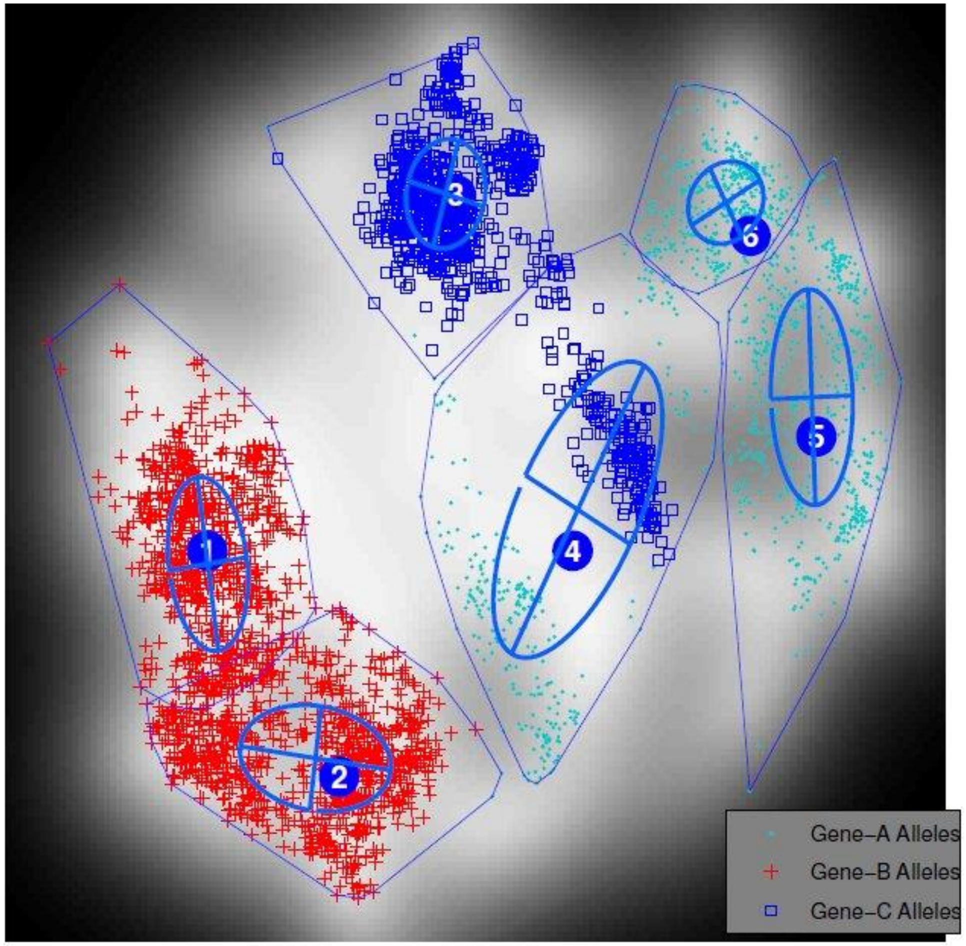
Chemometric Application: HTS Data

- Scientists at Pfizer searching for active compounds can now screen millions of compounds in a day.
- Gain a better understanding of the results of multiple screens through the use of novel data visualisation and modelling techniques.
- Find clusters of similar compounds (measured in terms of biological activity) and using a representative subset to reduce the number of compounds in a screen.
- Build local prediction models.

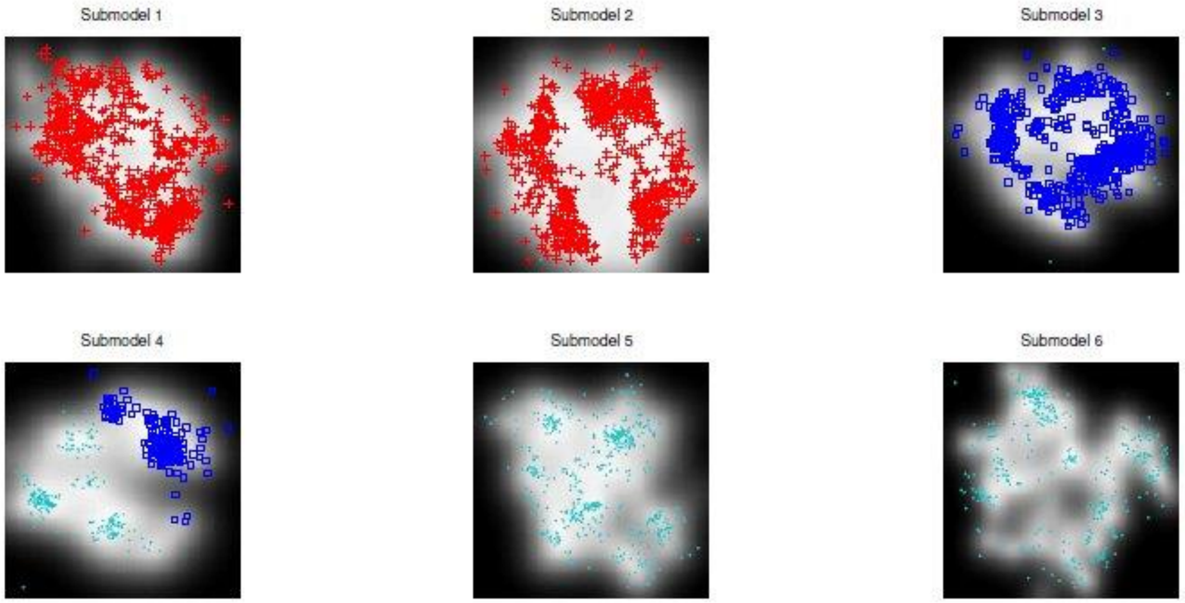
Hierarchical Visualisation



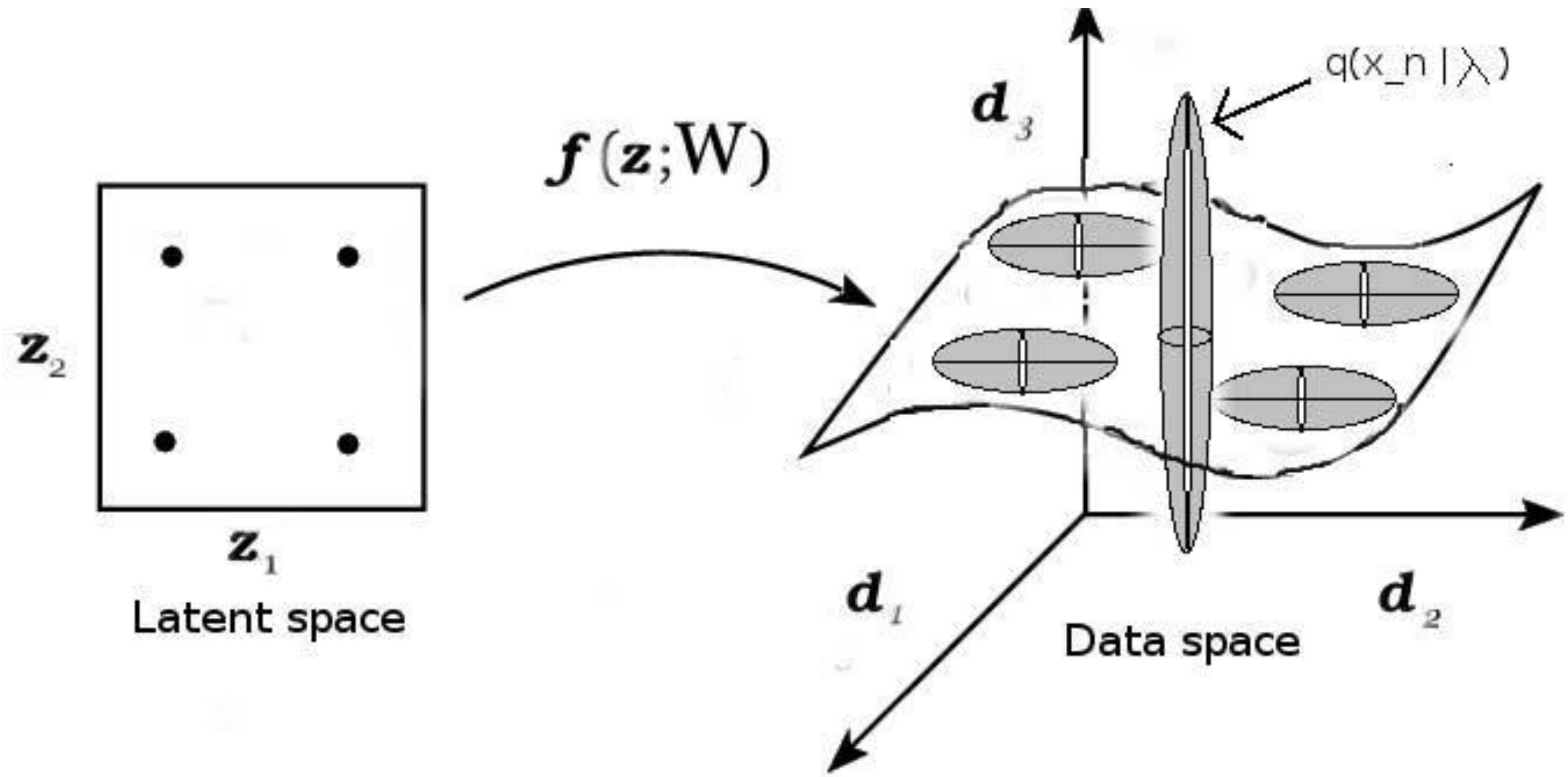
Gaussian Process Latent Variable Model



Replace the RBF mapping with a Gaussian Process.

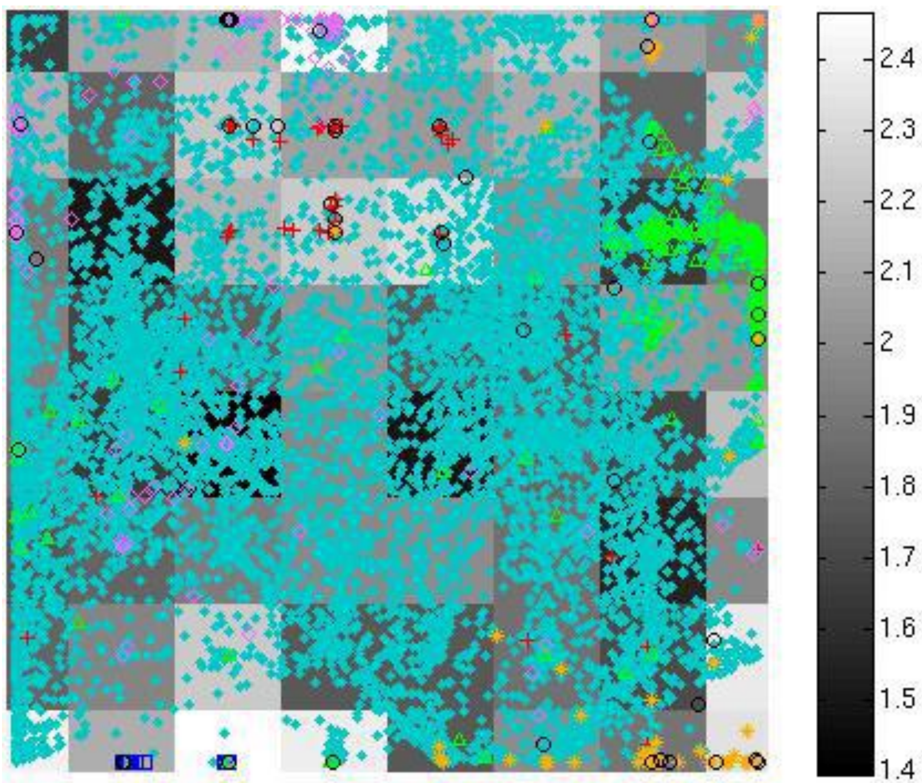


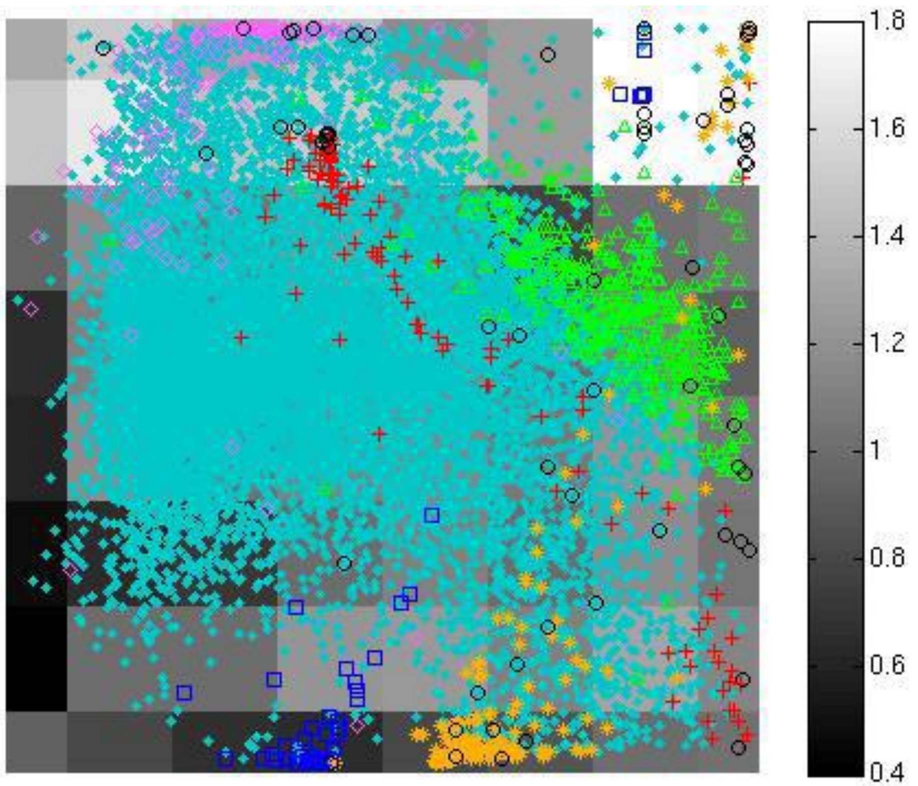
GTM-FS



- $d_1$and d_2 have high saliency; d_3 has low saliency.

Chemometric Data





GTM Visualisation

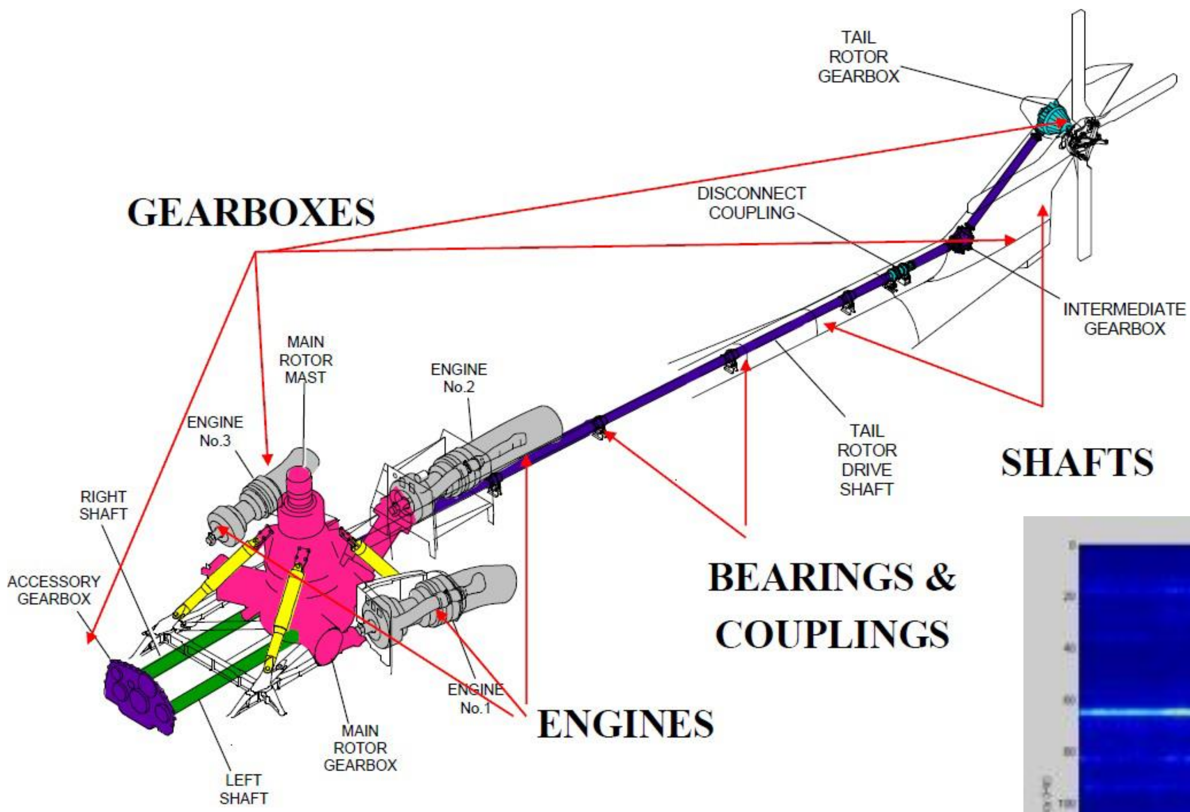
GTM-FS Visualisation

Magnification factors on a log scale

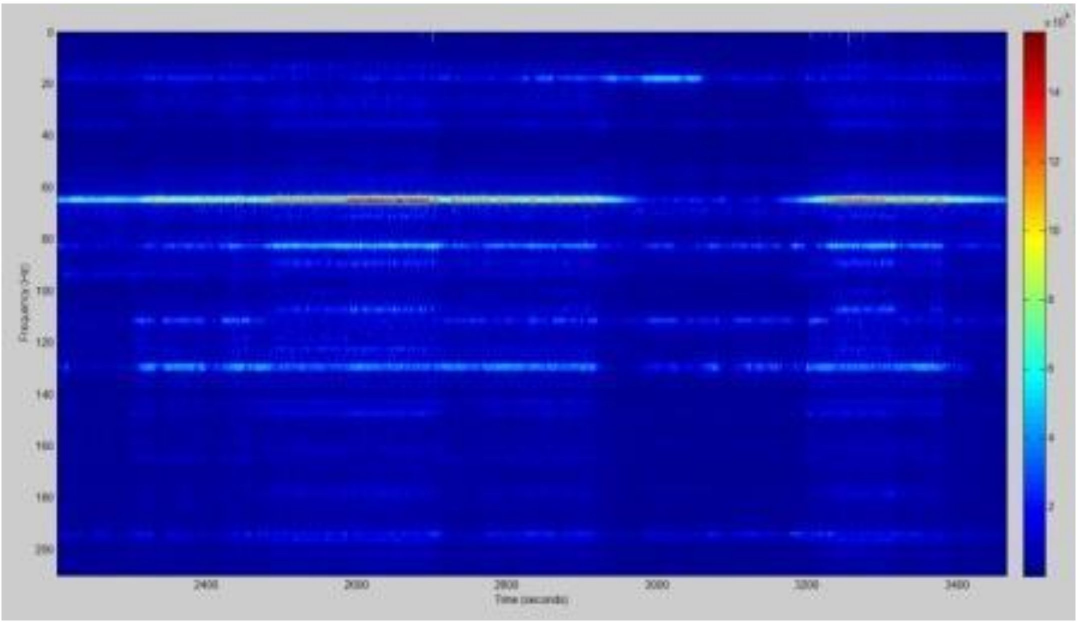
Agusta Westland

- AW has pioneered CVM, the continuous recording of airframe vibration (0-200Hz), to improve the investigation of unusual occurrences and monitor airframe integrity.
- Develop a probabilistic framework for inferring flight mode and key parameters from multiple streams of vibration data.
- Improve indicators of airframe condition: the wavelet transform and kernel entropy to assess the dynamics (i.e. non-stationary characteristics) of the vibration signal.
- Integrated diagnosis based on probabilistic models of normality and using a belief network to model prior knowledge about the domain and interactions between key variables.

Understanding the Data



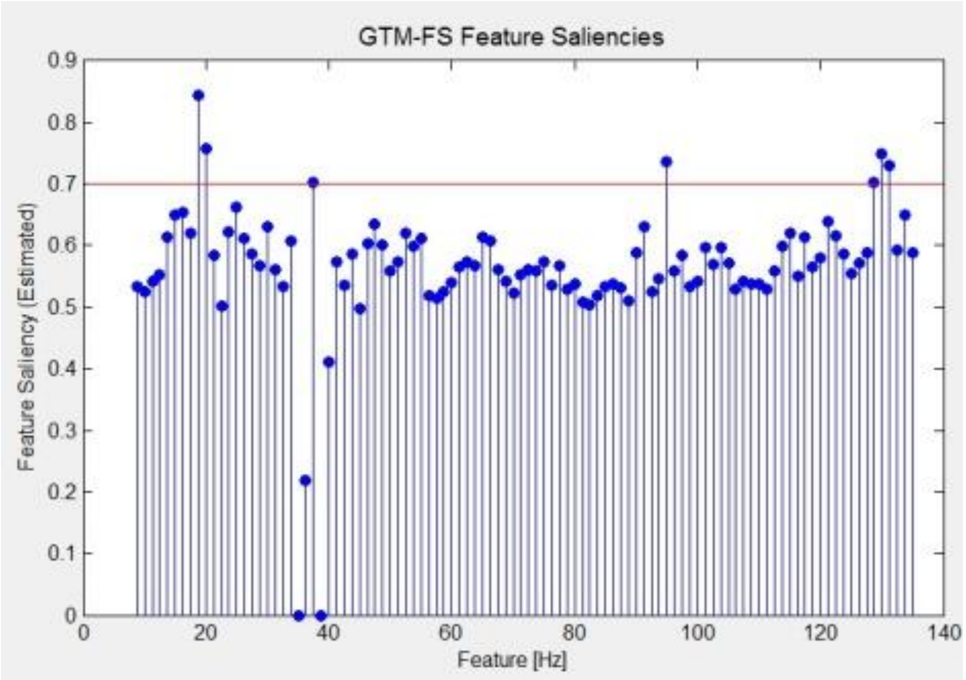
- 8 sensors measuring vibration
- 108 frequency bands (STFFT) for each sensor
- Too much data to build a model from.

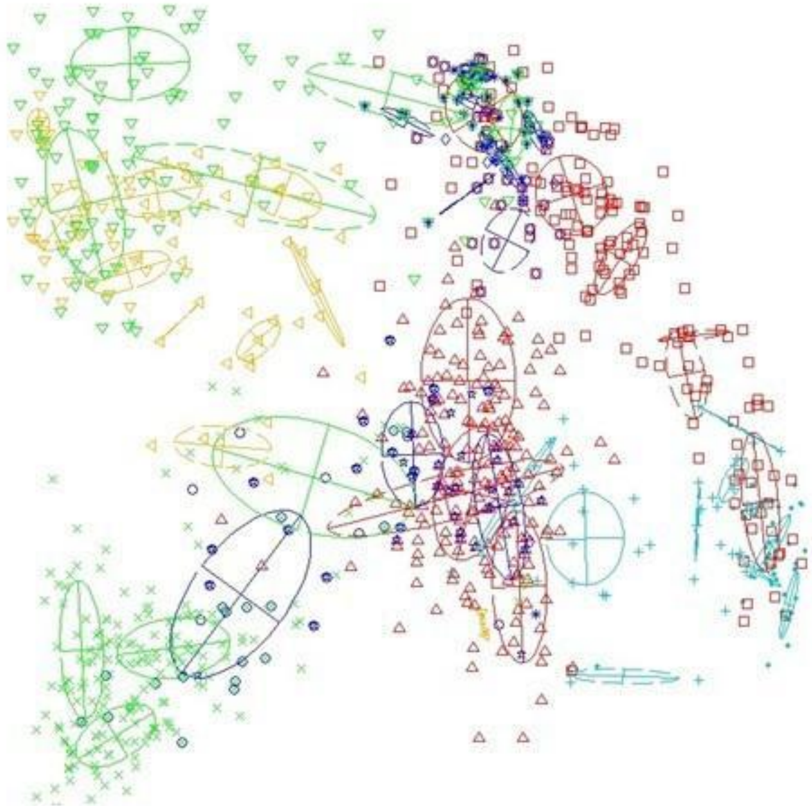


Feature Selection

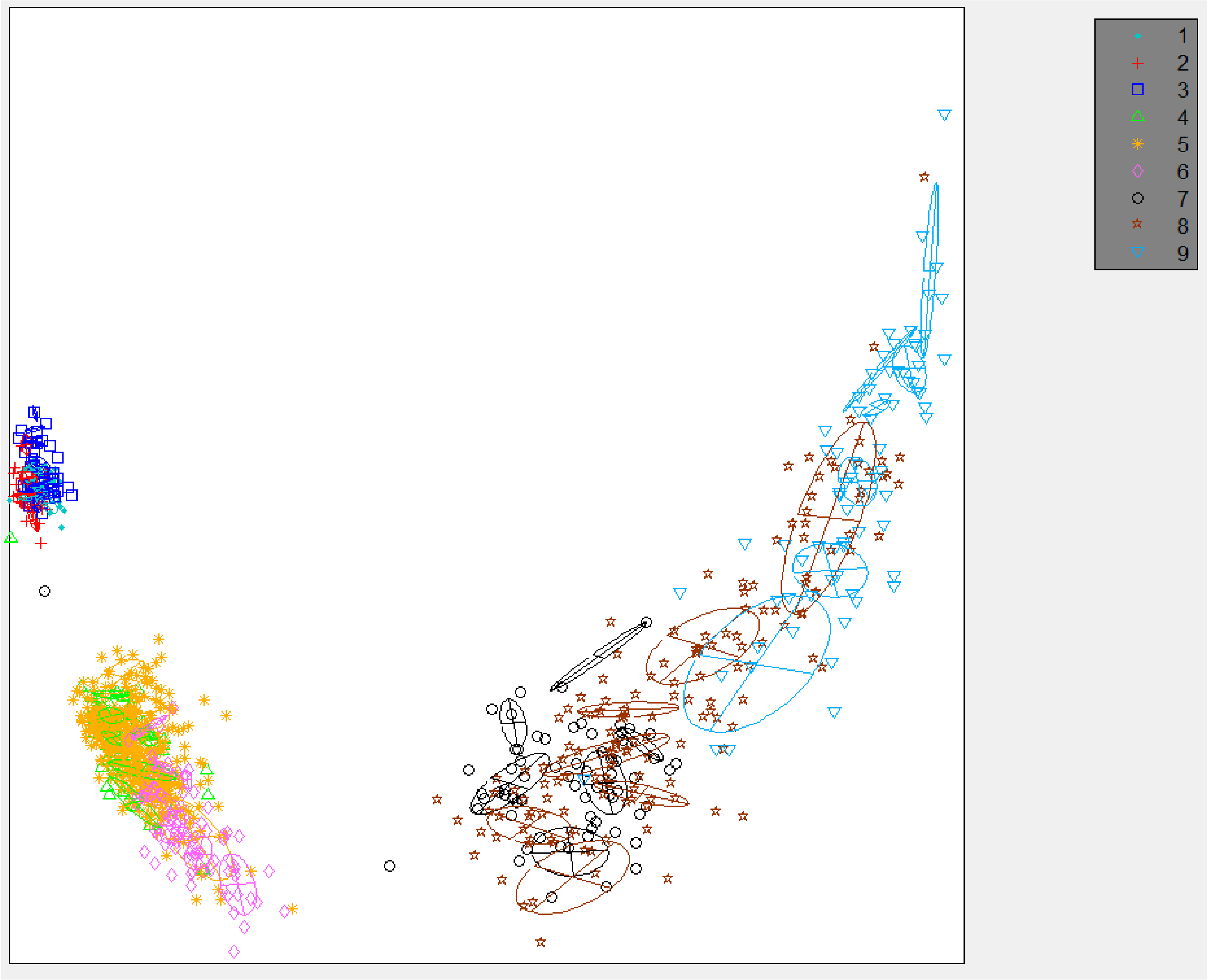
Features are selected using GTM with Feature Saliencies.

Sensors are selected by comparing inter-class separation in different plots.

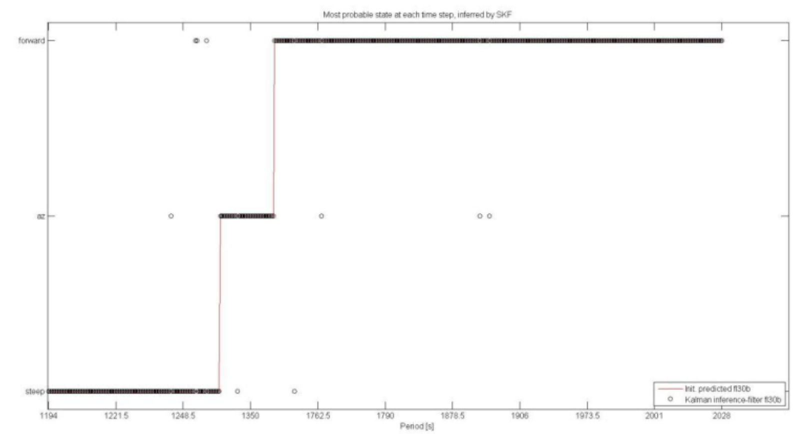




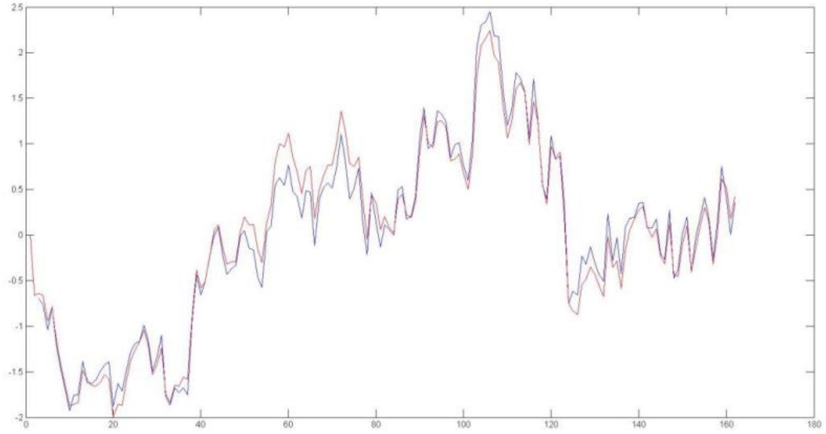
Flying through the visualisation



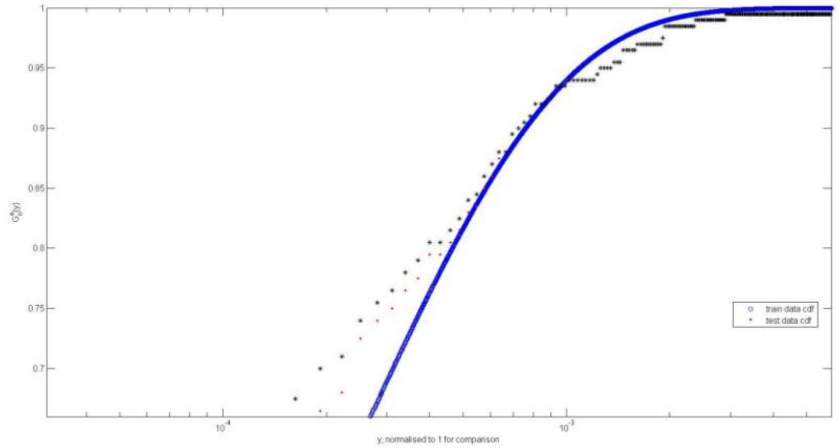
Novelty Detection



Inference of flight state



Non-linear signal prediction



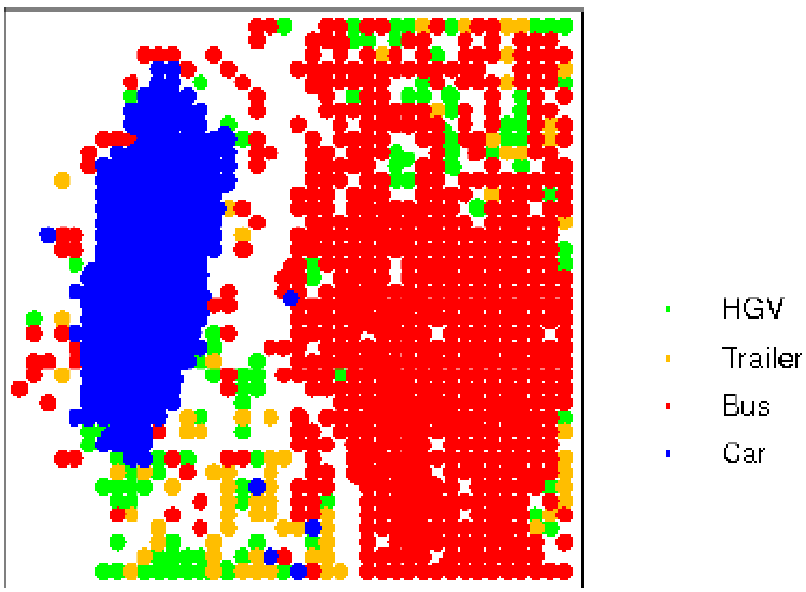
Outlier detection using Extreme Value Theory

Wheelright

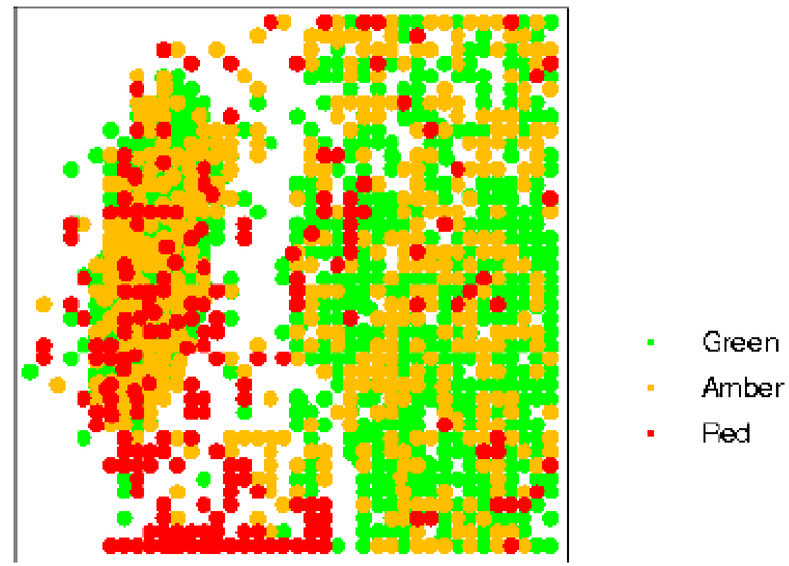
- Typically vehicle tyres are 10% underinflated.
- This wears the tyres 8% faster.
- £500M per year in additional fuel costs

Tyre pressure 'read as you drive'

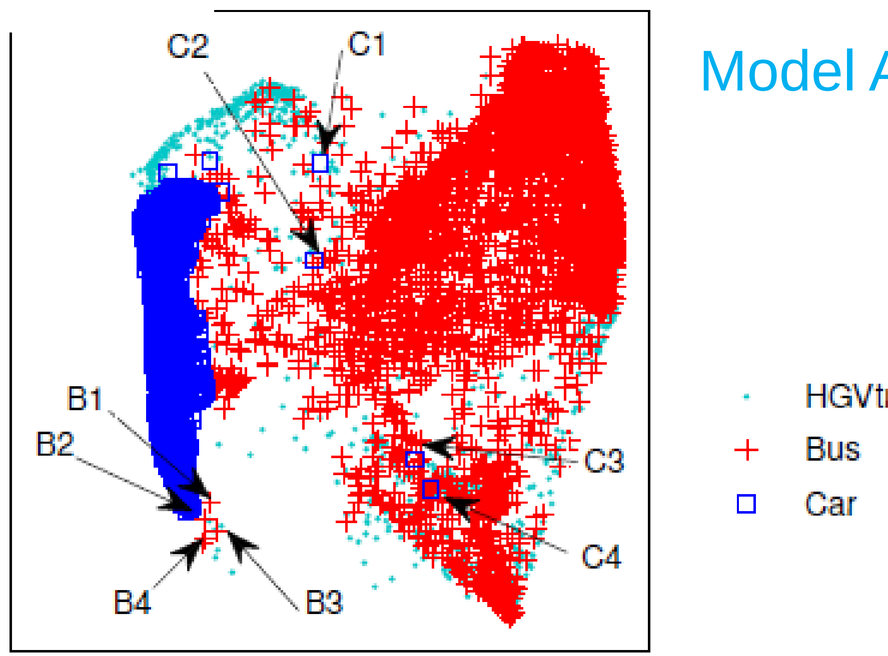
Visualisation of Results



Vehicle Classification



Model Accuracy



Outliers

Block GTM

- Include prior information about the correlations of variables into a GTM by using a full covariance matrix in the noise model and enforcing a block structure.
- This results in a reasonably sparse covariance matrix and keeps the number of unknown parameters low. The additional flexibility allows the model to fit the data more closely.
- The extension of the learning algorithm is straightforward and the only changes occur in the computation of responsibilities in the E-step and of Σ in the M-step.

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_p \end{bmatrix}$$

Conclusions

- Visualisation is an important tool for all types of user; the domain expert must be involved in the process.
- Interaction with the plots allows the user to query the data more effectively.
- Probabilistic (latent) models allow the user to do visualisation+ in a single coherent framework.
- Presenting the data in the right way is key. Feature selection is a very important tool.
- Accounting for known structure (e.g. block covariance) improves results.