Third European conference of the
Prognostics and Health Management Society in Bilbao
CONFERENCE TUTORIAL 6 JULY 2016

Multivariate post-processing of big data in Prognostics and Health Management: theory and use cases

Gueorgui Mihaylov

Department of Mathematical Science
Department of Mechanical and Aerospace Engineering

“Since the mathematicians have invaded the theory of relativity I do not understand it myself any more.”

A. Einstein

Aim of the tutorial

- Start from concrete use cases and show the advantages of the application of suitable mathematical tools
- Illustrations of the methods will be provided on different scales

- Single aircraft
- Fleet of aircrafts
- Industrial plant
- International system composed by thousands of plants

Standard “flow” of information

Physical data

Physical or data driven models

Features/indicators
(condensed information)

Often the multivariate post-processing of health, mechanical performance and efficiency indicators reveals relevant hidden information

"Warning about Big data"

- Storage and processing (parallel, distributed, incremental calculus and optimization of the algorithms) of big amounts of data is a quite non-trivial issue which we will not discuss
- Focus on the cases when qualitatively new features appear i.e. when there is some "emergent phenomenon"
- Neural networks are excellent tool able to "mimic" complex emergent behaviour, but less able to "model" it

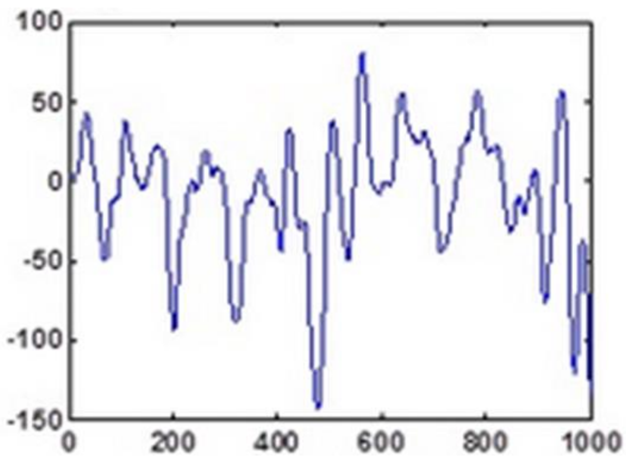
First use case

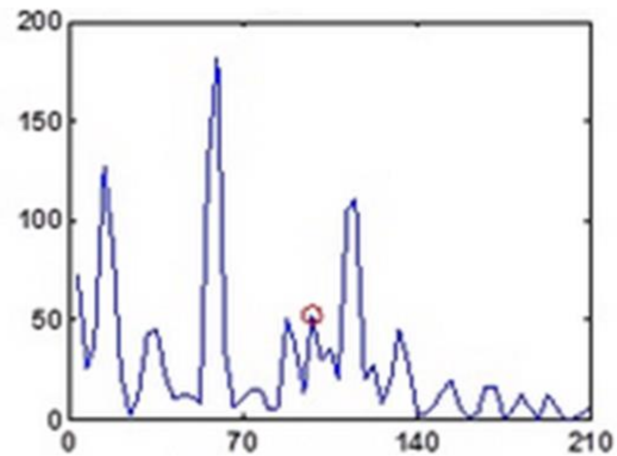
- Joint research project

Agusta Westland + Politecnico di Torino

First use case

- An accelerometric health monitoring system has been installed on helicopters produced by AgustaWestland
- The vibration monitoring methods are based on the analysis of analog signals provided by a set of accelerometers
- Acquisition and remote storage of accelerometric time and frequency (Fourier) domain spectra – first step of the processing





Problem statement

- Second step: scalar Health Indicators are defined by analysis of the local time and frequency domain characteristic of the signal (energy distribution, variability etc.)
- Each health indicator is monitored independently with respect to a priori fixed thresholds. The interpretation of each indicator is **univariate**

Problem statement

- **CRITICAL ISSUE**: the in-service exploitation of the system showed a number false alarms that caused a nuisance in the aircraft operation
- **AIM:** improve the efficiency of the monitoring system
- **SOLUTION:** implementation of an innovative integrated **multivariate** health monitoring system based on “third level analysis” of the same accelerometric features
- **IMPORTANT:** this solution does not require installation of any additional sensors and does not change the indicators!

In service data

AgustaWestland provided a large amount of in-service data:

- 372 aircrafts of the same type
- Period of five years
- Different flight conditions
- Aproximately 500 000 monitored flight hours

In service data

- Our research mainly considered the following set of components of the main rotor power drive:

2-nd Stage Pin RH Bearings

Hangar Ball Bearing

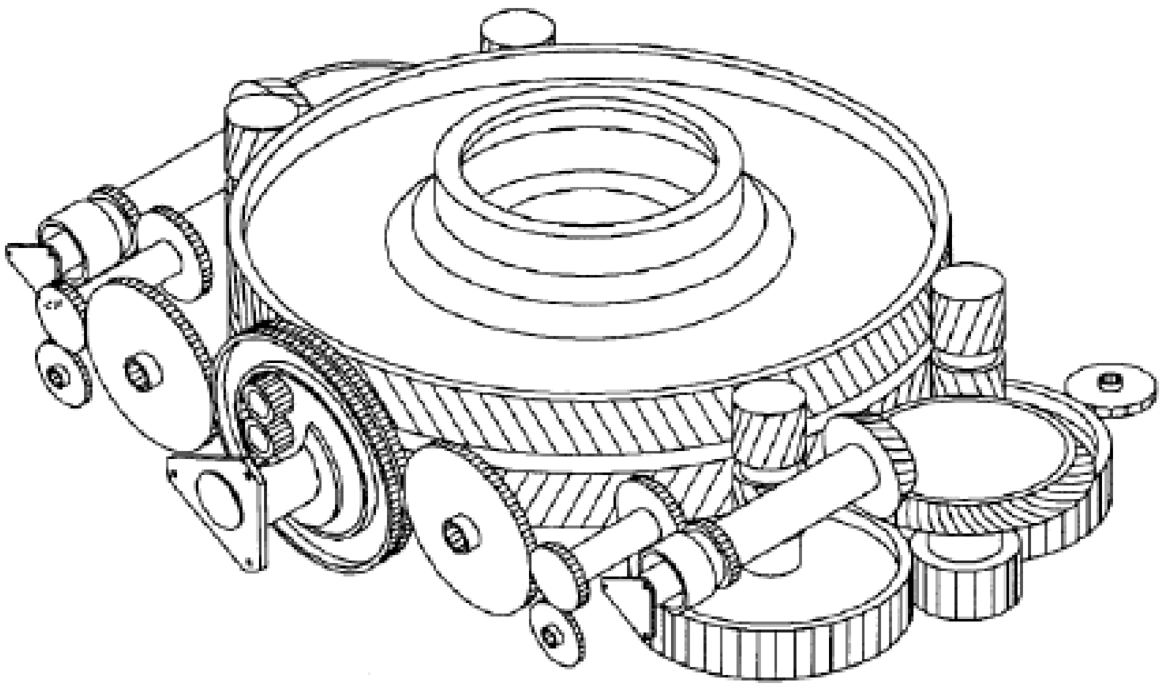
Oil cooler Bearing

TTO Pinion

TGB Gear

IGB Pin

TRDS



Multivariate approach

- Observation: in many cases groups of health indicators react simultaneously in anomalous situations
- **Several multivariate statistical techniques have been applied on VECTORS of health indicators**
- The **variance structure** and the **covariance structure** of the accelerometric health indicators data set have been investigated
- The "**geometric**" approach to multivariate statistics leads to immediate and useful intuitions

Multivariate statistics

- Latent variables
- Statistics-multiple reformulation
- Distances
- Multiple regressions

Latent variables

- Simultaneous measurements of N real random variables
- Variables are interpreted as directions in N-dimensional vector space
- Their magnitudes are the projections along those directions
- “Non-physical” directions can be very relevant - Latent Variables
- Many phenomena are more efficiently described by a suitable set of latent variables
- Many multivariate statistical methods are based on selecting efficient latent variables

The Mahalanobis distance

- Based on the properties of the covariance matrix

$$C_{ij} = Cov(X_i, X_j)$$

- Statistically relevant – “equilibrates” the variance of the variables

$$d(X,Y) = (X - Y)^T C^{-1} (X - Y)$$

- Equidistant points from a given point are ellipsoids

Multivariate Normal distribution

- Generalization of the one-dimensional Gauss distribution

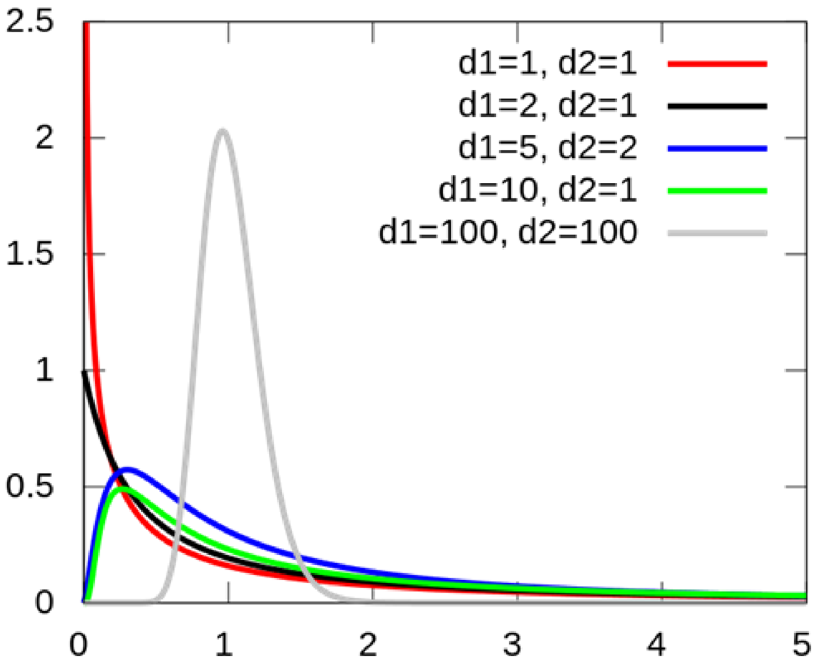
$$f_{\mu,C}(X) = \frac{1}{(\sqrt{2\pi})^n det(C)} \exp[-\frac{1}{2}(X-\mu)^t C^{-1} (X-\mu)]$$

- Level sets are ellipsoids
- The Mahalanobis distances of normally distributed vector from the mean value follows the Hottelling distribution $T_k^2(n)$

Hotelling distribution

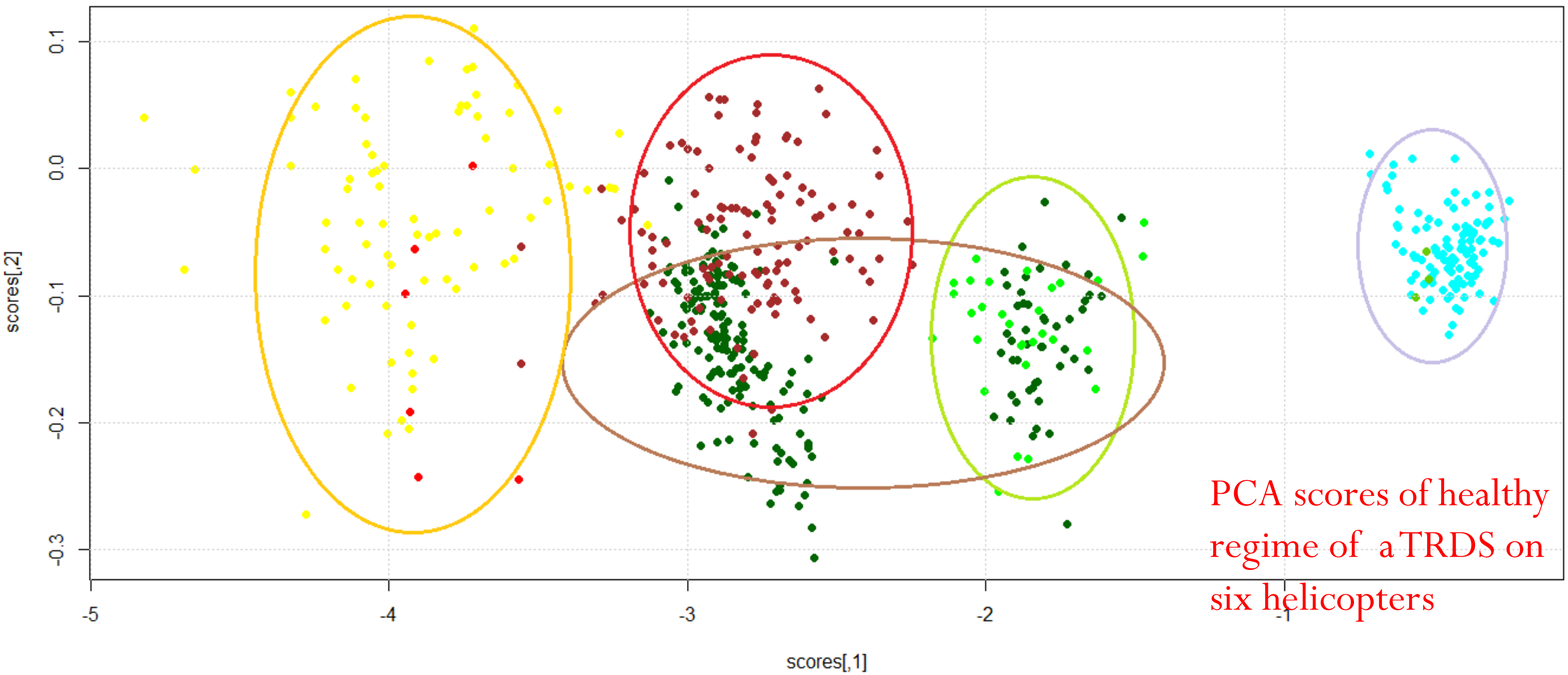
- Hotelling and Fisher-Snedekor

$$\frac{n-k+1}{nk}T_k^2(n) \simeq F_{k,n-k+1}$$



Recalibration procedures

- Values of the indicators, which characterize the healthy regime of a mechanical component, reveal considerable variations for individual helicopters of the same type (huge amount of data difficult to use)



PCA scores of healthy regime of a TRDS on six helicopters

Recalibration procedures

- Available data regarding the flight conditions of each aircraft

Engine 1 Torque, Engine 2 Torque, Rotor Speed, Roll Angle, Pitch Angle, True Airspeed, Radio Altitude, Vertical Speed, Normal Acceleration, Density Altitude, Tail Rotor Torque, Main Rotor Torque, Roll Rate, Pitch Rate, Yaw Rate, Longitudinal Acceleration.

Accelerometric measurements influenced by flight conditions so they were done in specific ranges of the flight parameters

- Residual multi-correlation between accelerometric measures and "environmental" parameters has been hypothesised
- Hypothesis confirmed via **Canonical Correlation Analysis**

Recalibration procedures

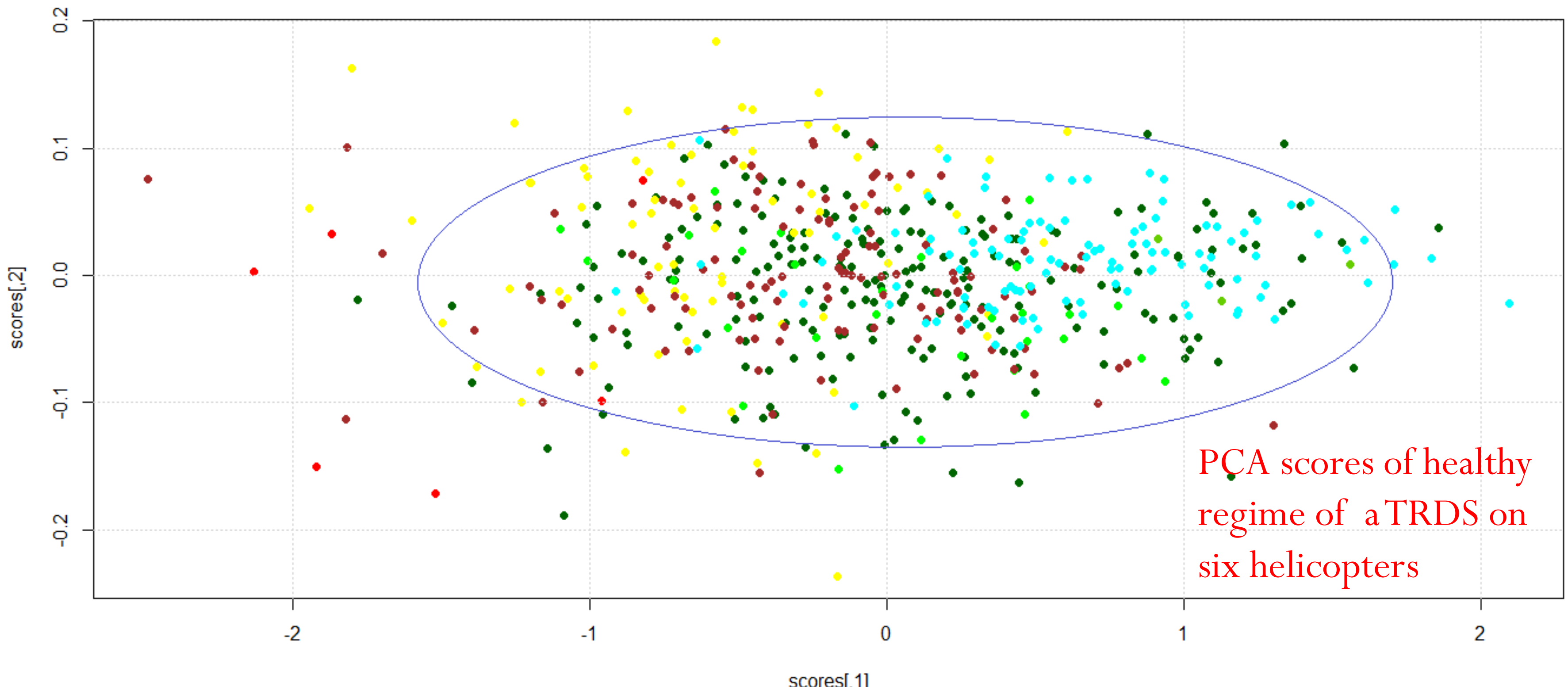
- Linear and quadratic filters, which eliminate the deterministic component of the behavior of the aircraft have been defined

$$f : R^{17} \longrightarrow R^{N}$$

- Effects: the healthy operational regime measured over the fleet becomes much more homogeneous

Recalibration procedures

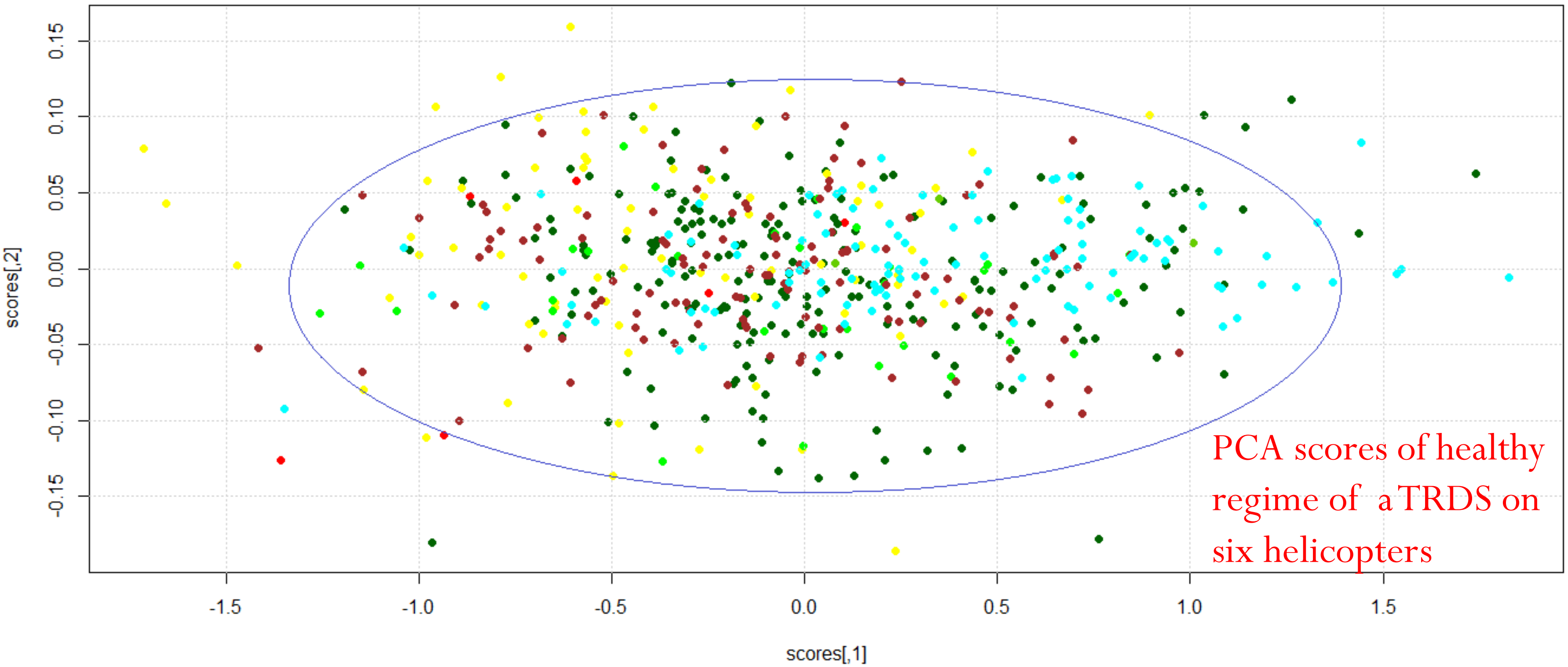
- Linear recalibration is based on a set of multilinear regressions of the health indicators over the set of flight condition parameters



phmsociety

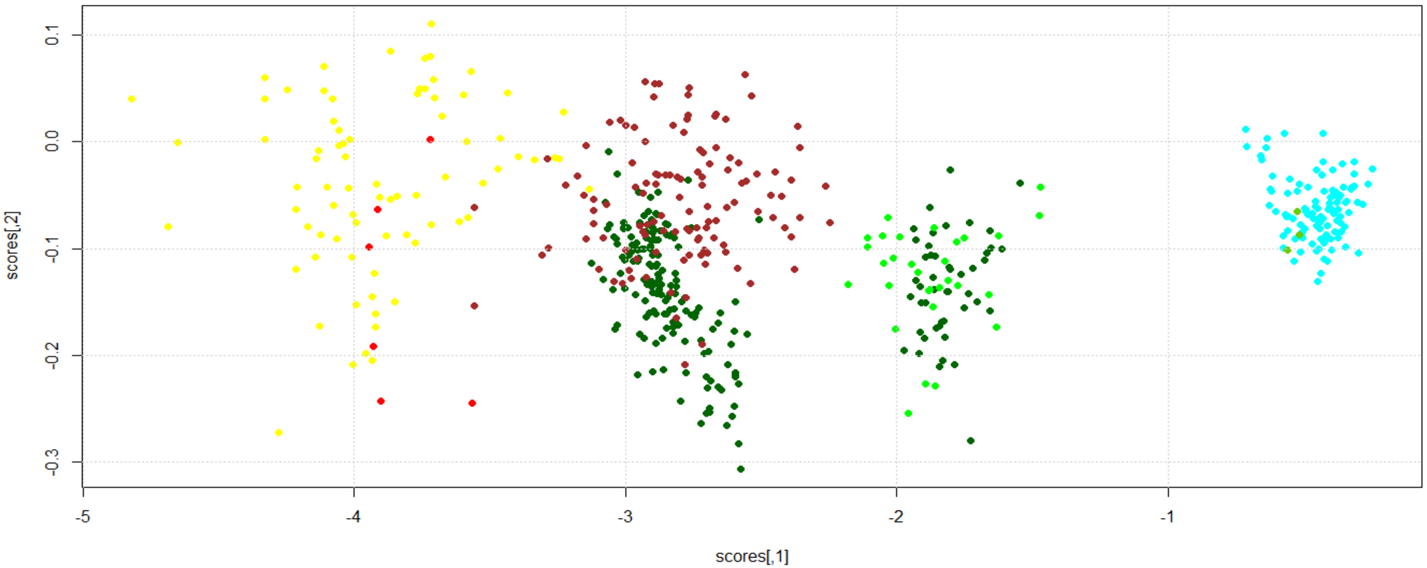
Recalibration procedures

- Quadratic recalibration is based on a set of second-order polynomial regressions of the health indicators over the set of flight condition parameters



phmsociety

Recalibration procedures

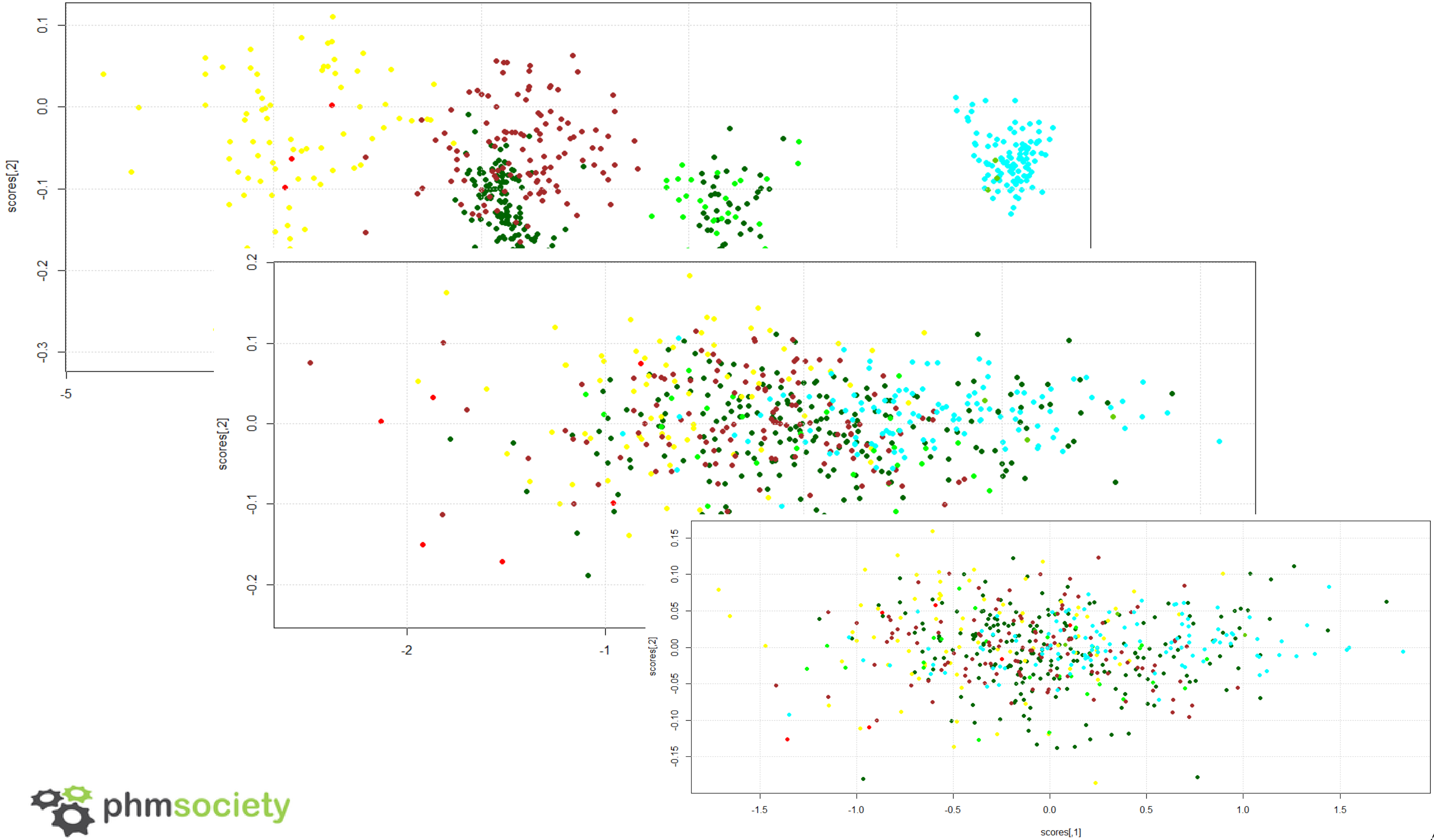


phmsociety

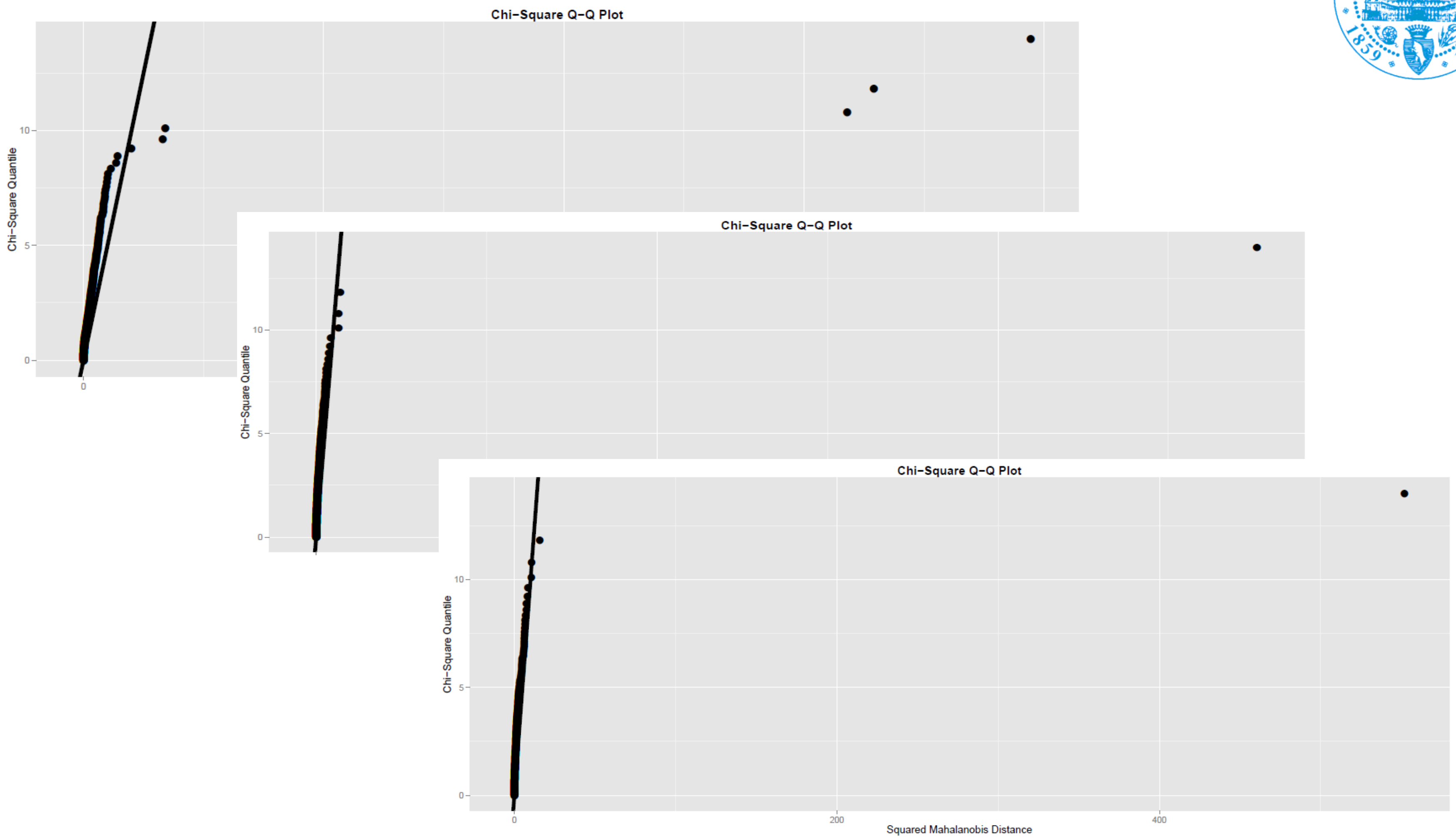
Recalibration procedures

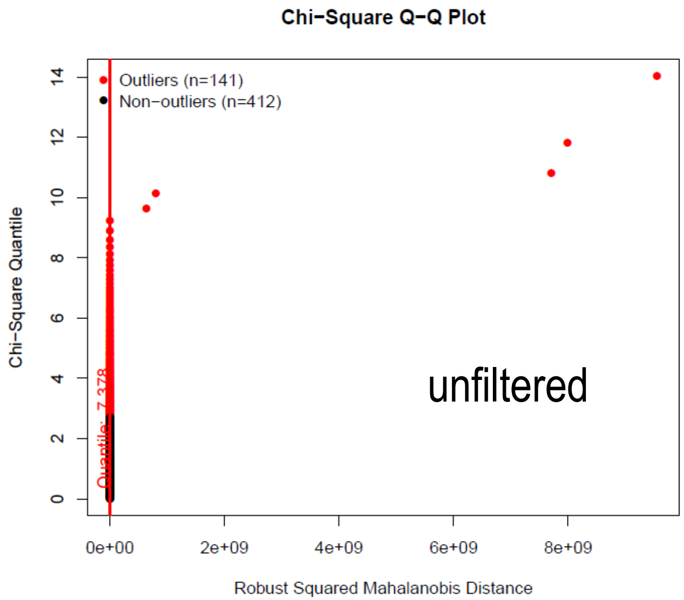
phmsociety

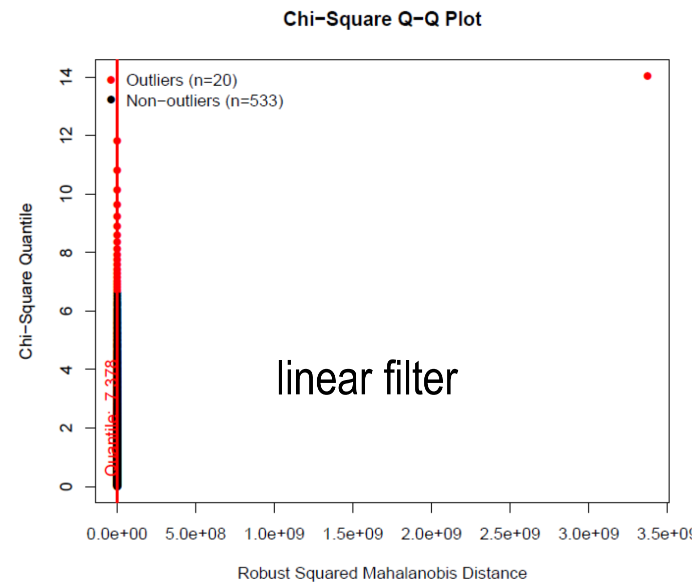
Recalibration procedures

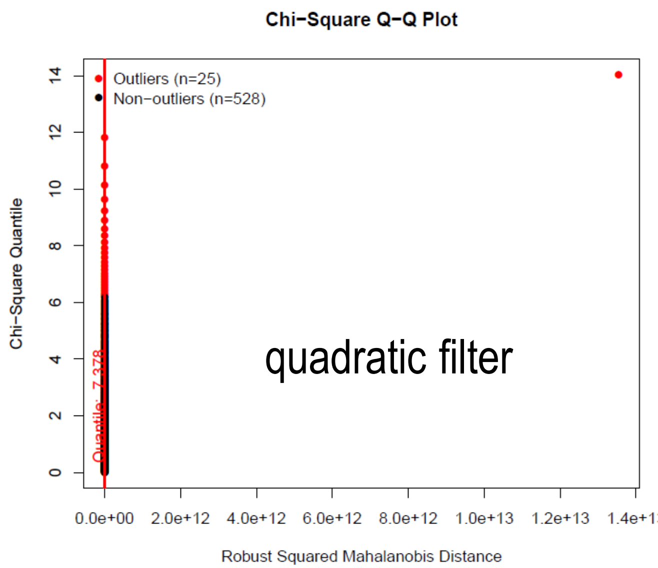


phmsociety


Chi-Square Q-Q Plot
Chi-Square Quantile
10
5
0
0
Chi-Square Q-Q Plot
Chi-Square Quantile
10
5
0
Chi-Square Q-Q Plot
Chi-Square Quantile
10
5
0
0
200
400
Squared Mahalanobis Distance

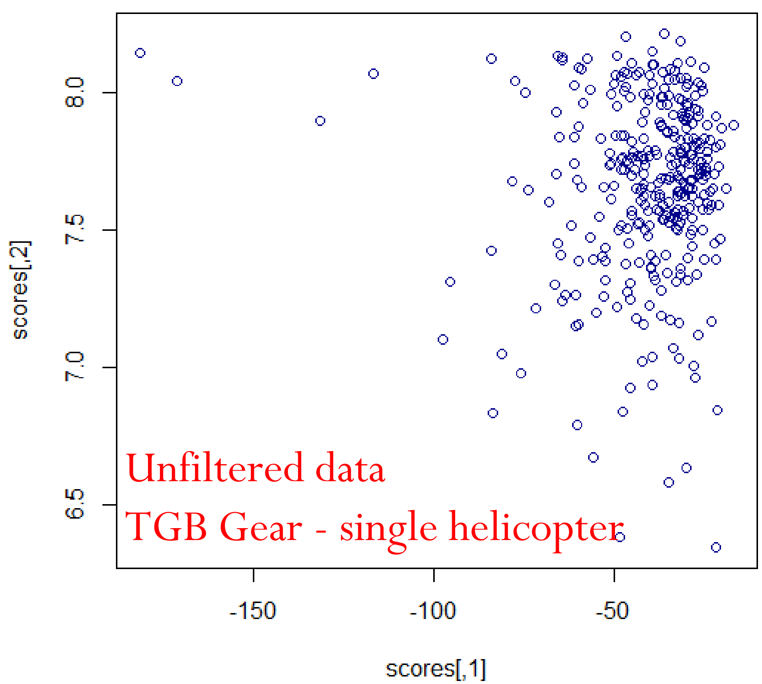

Chi-Square Q-Q Plot
Outliers (n=141)
Non-outliers (n=412)
unfiltered
Quantile: 7.378
Chi-Square Quantile
0
2
4
6
8
10
12
14
0e+00
2e+09
4e+09
6e+09
8e+09
Robust Squared Mahalanobis Distance

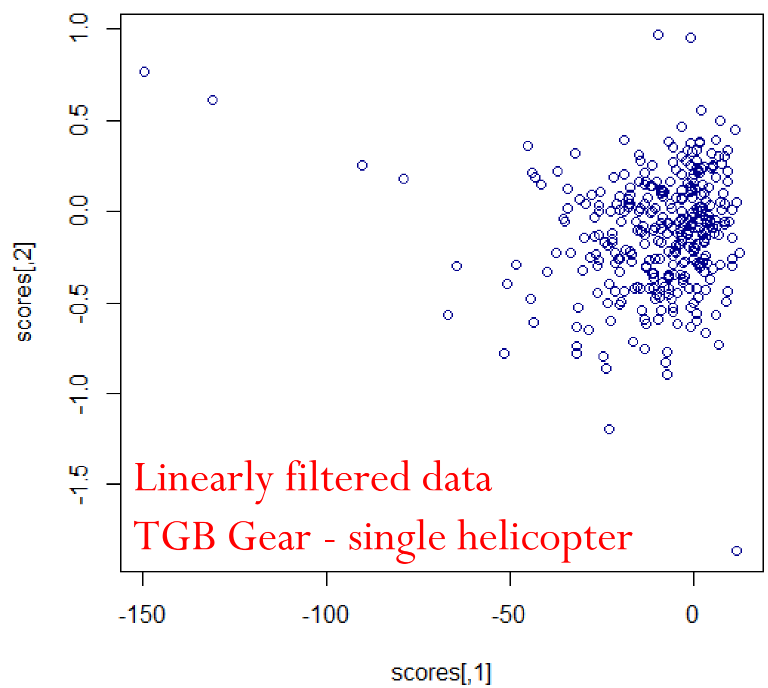

Chi-Square Q-Q Plot
Outliers (n=20)
Non-outliers (n=533)
linear filter
Quantile: 7.378
Chi-Square Quantile
0
2
4
6
8
10
12
14
0.0e+00
5.0e+08
1.0e+09
1.5e+09
2.0e+09
2.5e+09
3.0e+09
3.5e+09
Robust Squared Mahalanobis Distance


Chi-Square Q-Q Plot
Outliers (n=25)
Non-outliers (n=528)
quadratic filter
Quantile: 7.378
Chi-Square Quantile
0
2
4
6
8
10
12
14
0.0e+00
2.0e+12
4.0e+12
6.0e+12
8.0e+12
1.0e+13
1.2e+13
1.4e+13
Robust Squared Mahalanobis Distance

Recalibration procedures

- Once filtered the impact of the flight conditions, the intrinsic variability of the healthy operational regime of each component can be modelled over a random noise process





phmsociety

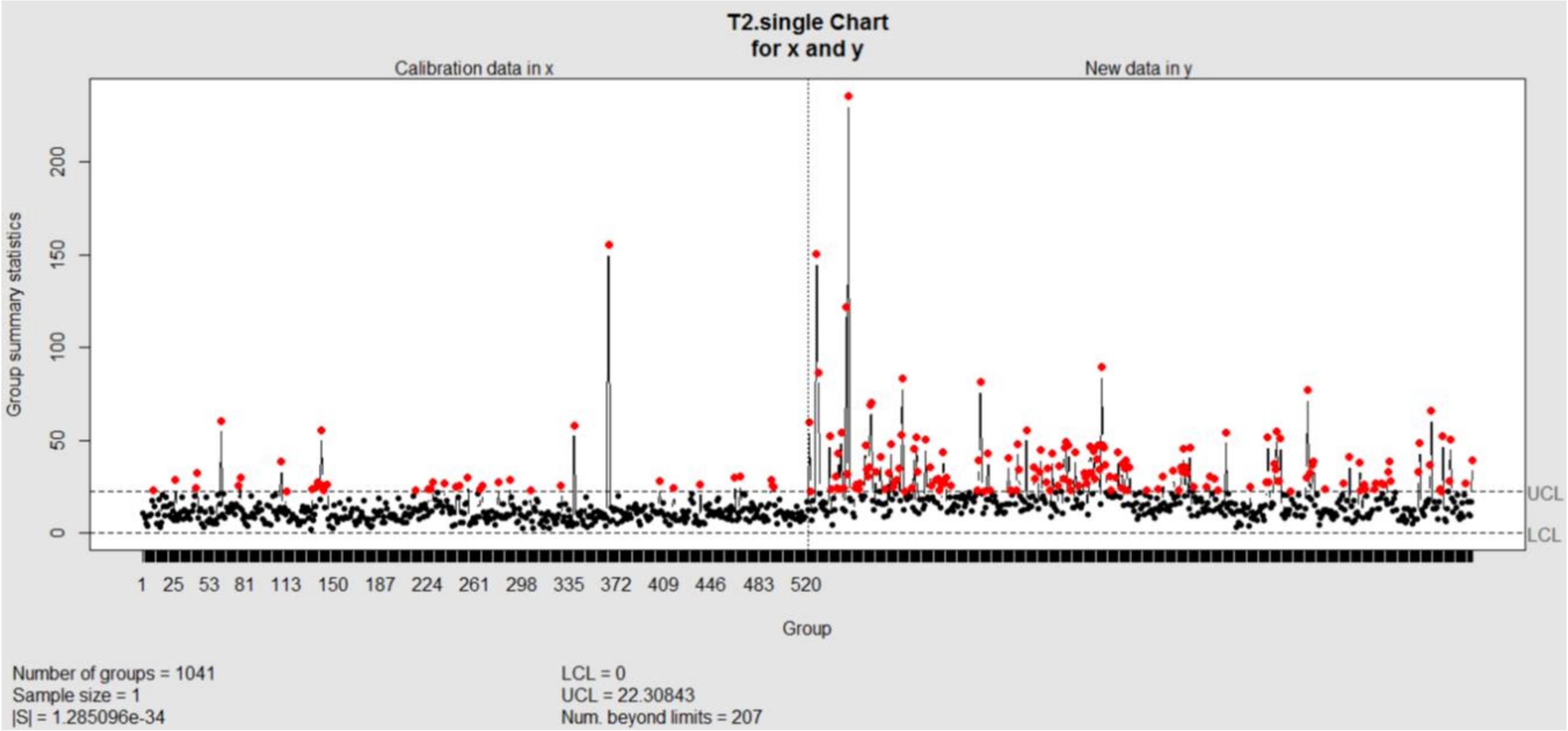
Anomaly detection

- The Shewhart control chart is a standard anomaly detection tool based on the statistically relevant Mahalanobis distance
- The filtered healthy operational regime follows the multivariate Gauss distribution
- Substitute the individual HI thresholds by a self-learning Shewhart control chart, which calibrates individually on each component of each aircraft

ADVANTAGEOUS SOLUTION

Eliminates a specific problem

Shewhart control chart



phmsociety

Anomaly detection via CCA

- **Canonical Correlation Analysis** has been tested as a **consistency based anomaly detection method**

A THEORETICAL HYPOTHESIS:

The failure modes of a component are uncorrelated with the environmental data, i.e. is a manifestation of an inconsistency with a valid multilinear model

- Results: Some cases in which the canonical correlations HI - flight condition parameters decrease considerably in case of anomaly!

Anomaly classification

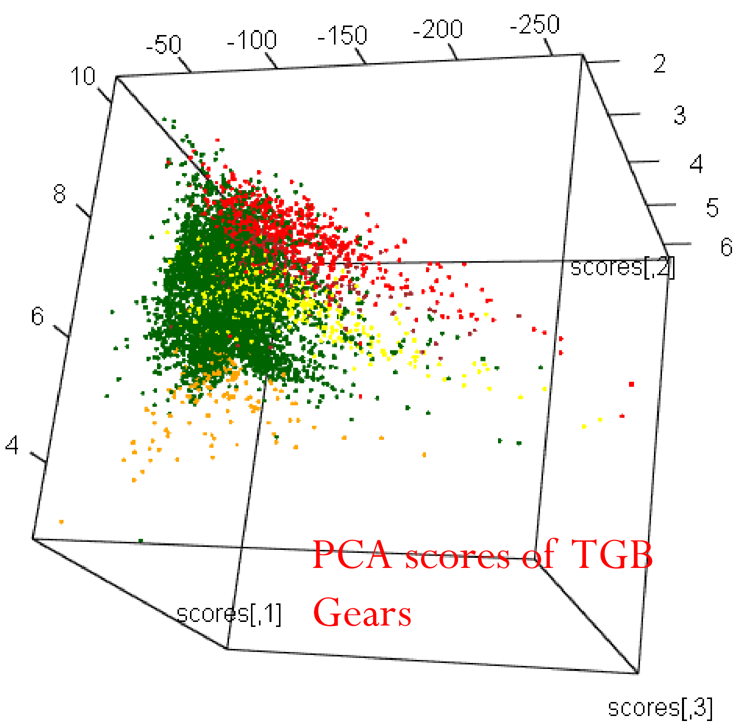
- Failures are rare
- Calibration, validation and application of any statistical model for monitoring purposes **on a single aircraft** is extremely **unrealistic**
- The recalibration procedures make the healthy operational regimes of the same component on individual helicopters

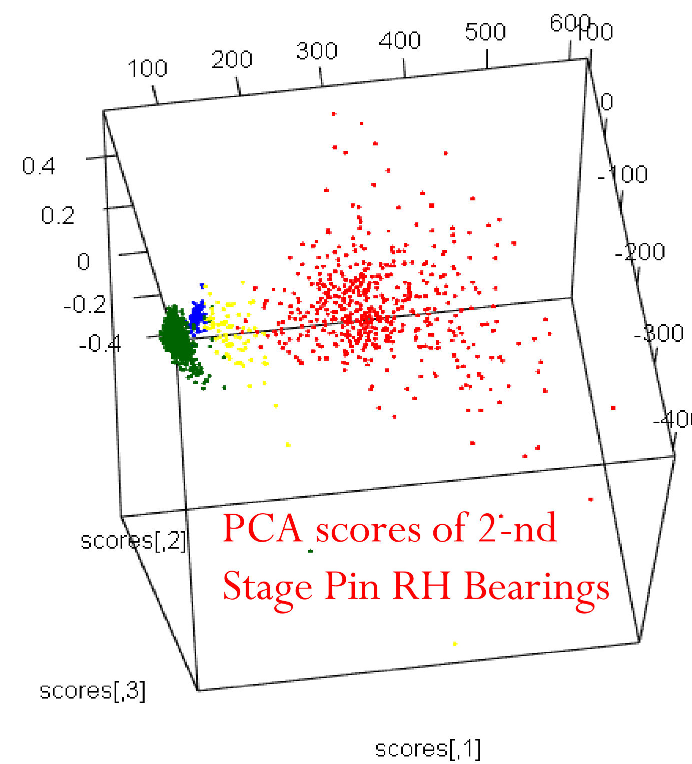
COMPARABLE!

- Powerful statistical discriminant tools can be calibrated on a fleet of helicopters of the same type

Anomaly classification

- **Principal Component Analysis** highlights the intrinsic variability of the experimental data set
- Latent variables are the eigenvectors of C

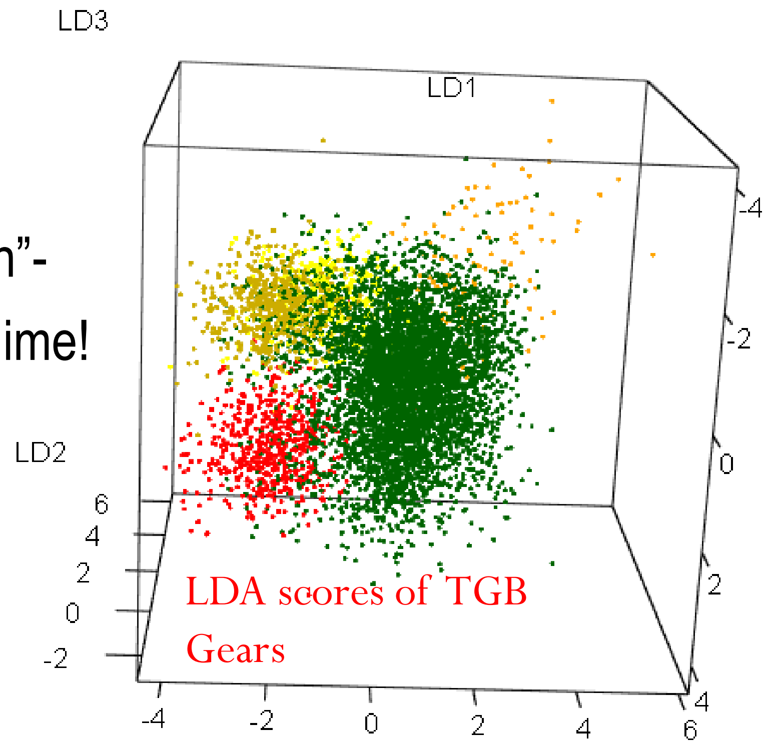




Anomaly classification

- **Linear Discriminant Analysis** exploits projections in the directions which maximize the variance between groups
- Latent variables are eigenvectors of $D = C^{-1}C^{(b)}$
- Linear affine decision boundaries

Observe that anomalies cannot be interpreted as “in the tail of the distribution”- different phenomena from the healthy regime!



Anomaly classification

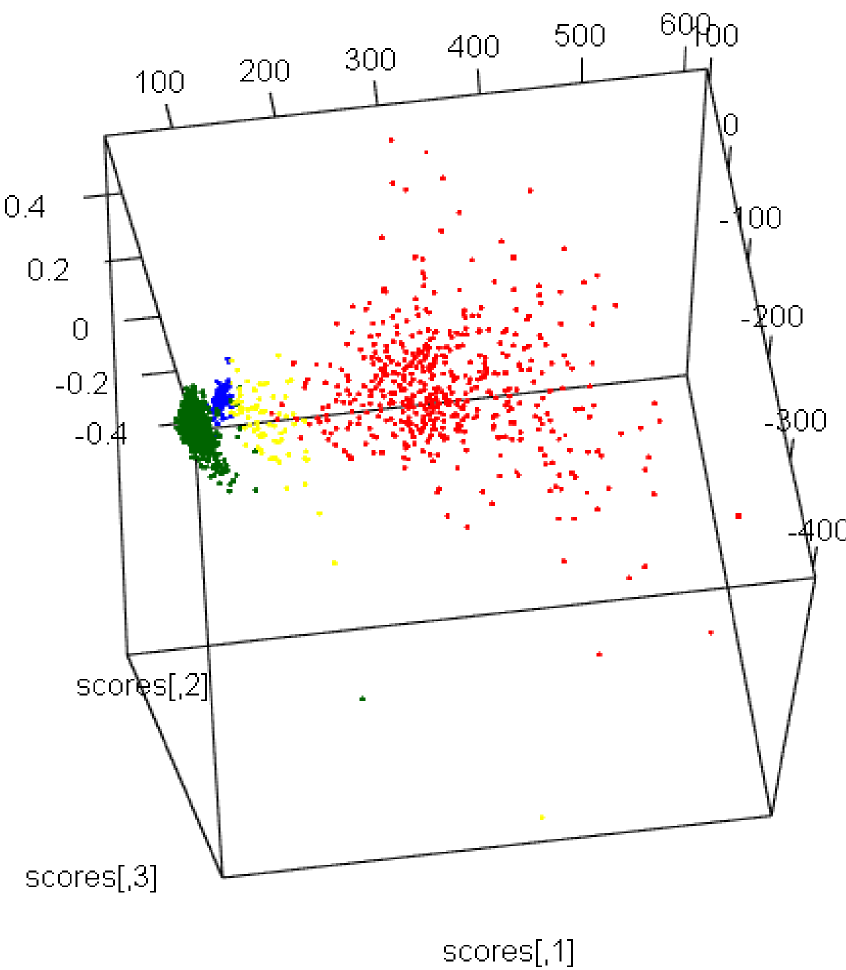
- **Quadratic Discriminant Analysis** exploits a quadratic classifier based on the Mahalanobis distance

$$q(X) = -\tfrac{1}{2}(X - (\bar{X})_j)^t\, C^{-1}(X - (\bar{X})_j) - \tfrac{1}{2}\log(\det C) + \log \pi_j$$

- Quadratic decision boundaries

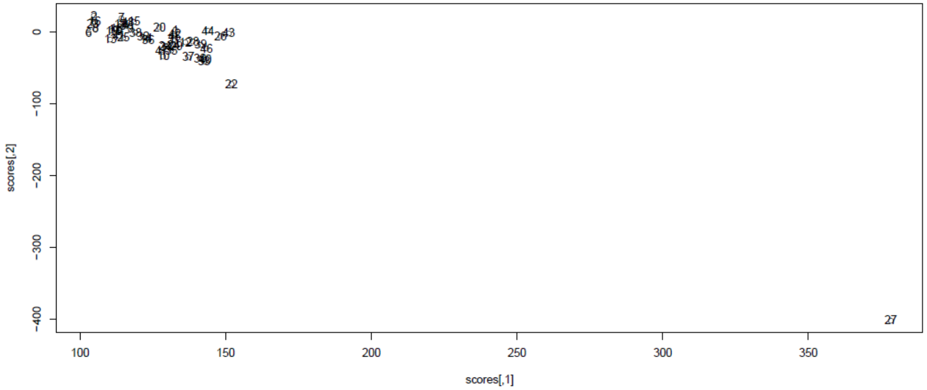
Combined application of these methods classifies anomalies with very high level of statistical confidence

Fault detection



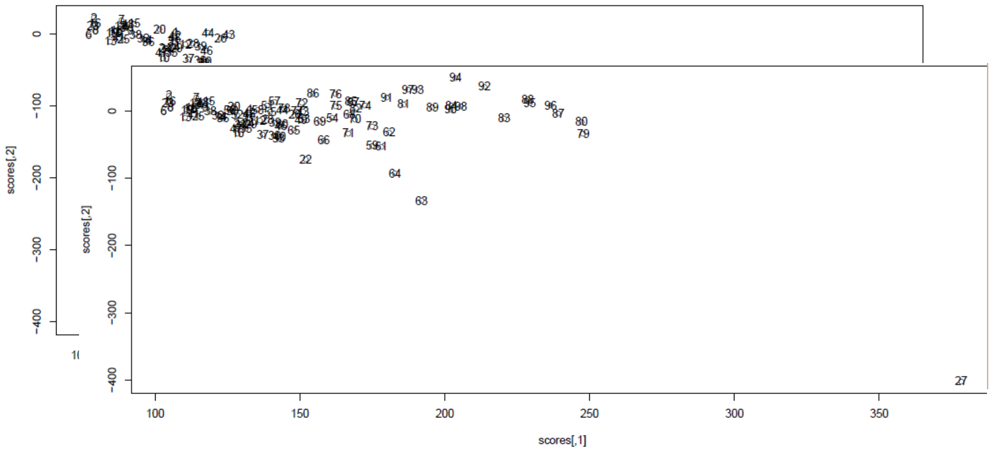
Fault into failure progression occurs in a specific direction

Fault detection



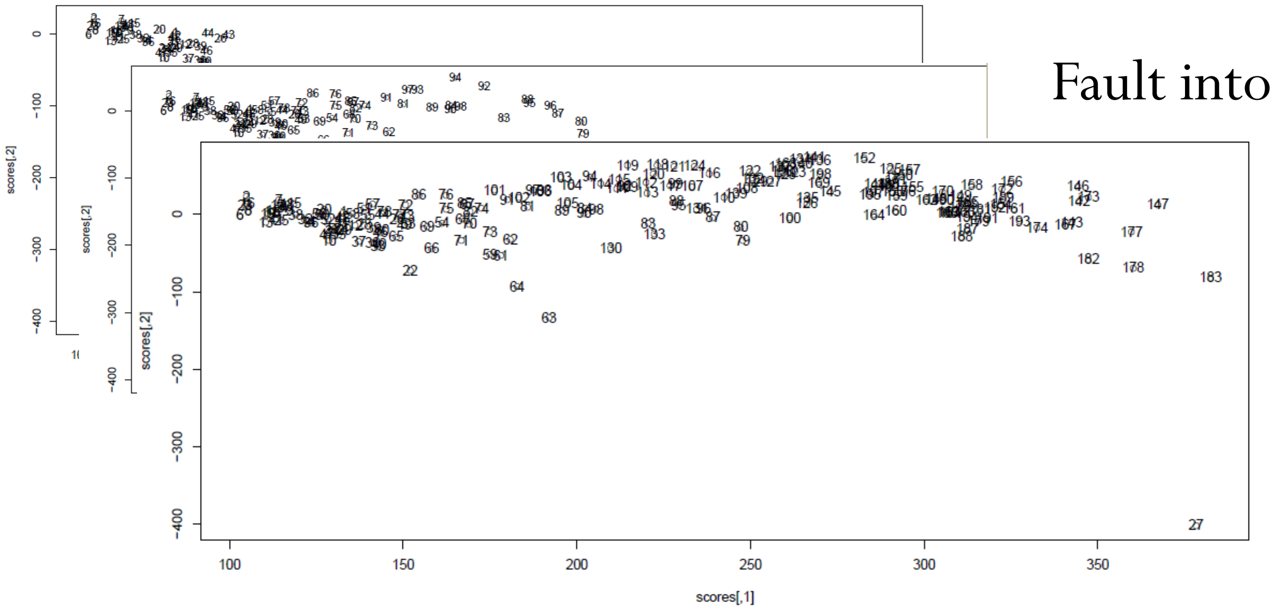
Fault into failure progression

Fault detection



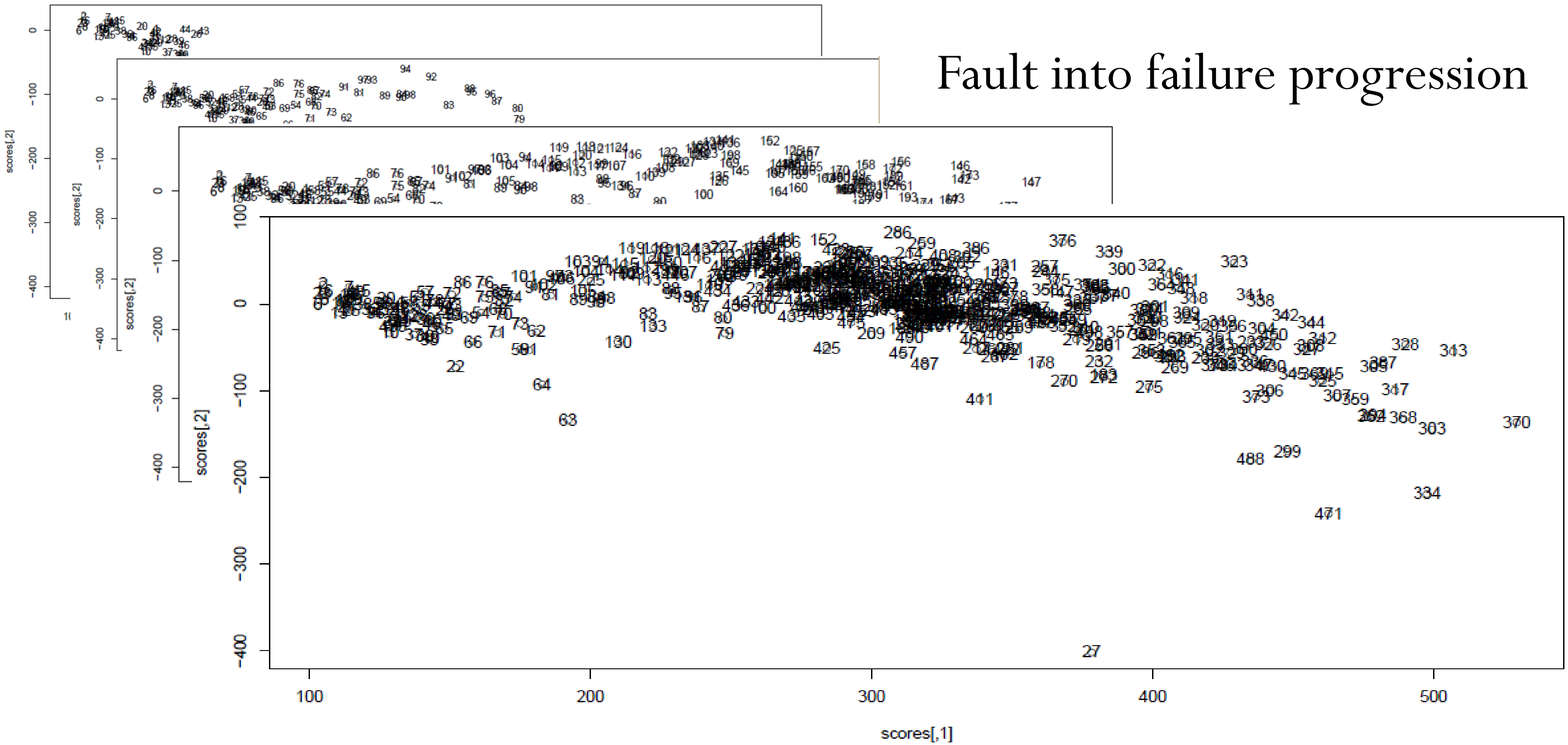
Fault into failure progression

Fault detection



Fault into failure progression

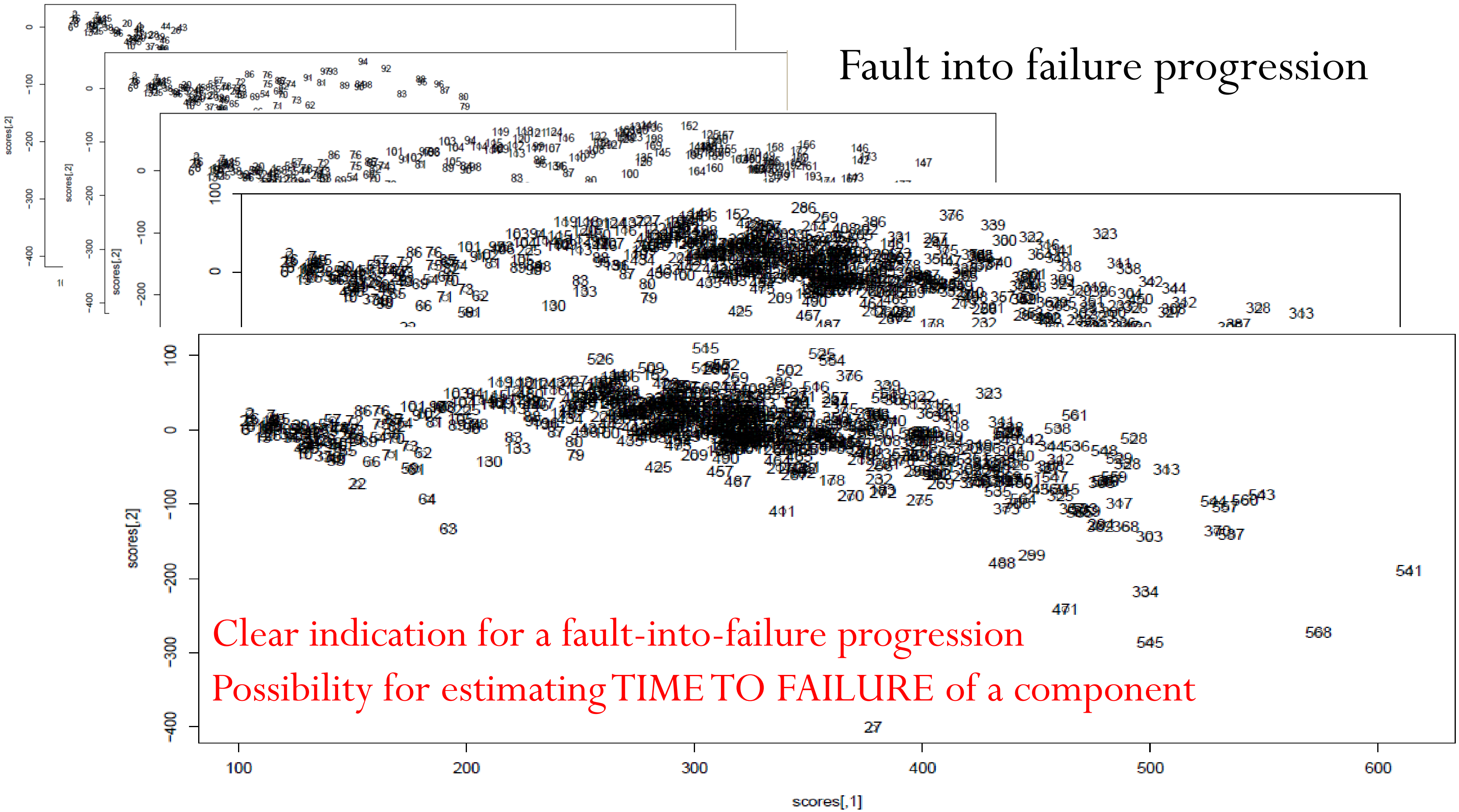
Fault detection



Fault into failure progression

Anomaly classification

Fault into failure progression



Clear indication for a fault-into-failure progression

Possibility for estimating TIME TO FAILURE of a component

Anomaly classification

A THEORETICAL HYPOTHESIS:

Since he healthy regime is modelled over random noise

If in case of true failure different health indicators react simultaneously in a consistent and correlated way

THEN in case of true failure the **covariance structure** of the data set is modified

Anomaly classification

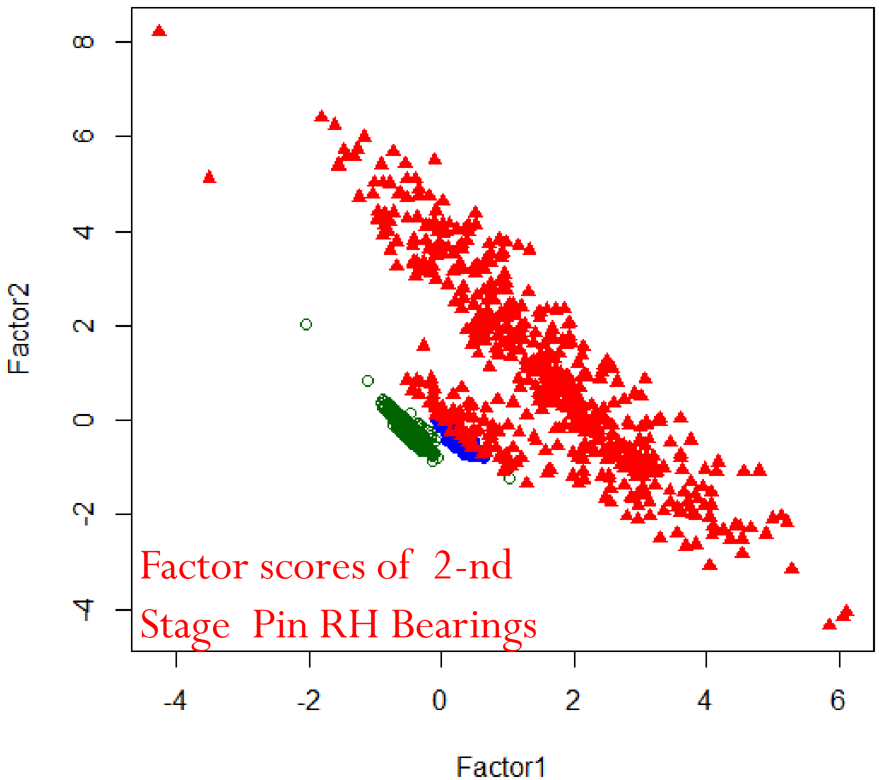
- The **covariance structure** of accelerometric data has been investigated by means of Principal Factor Analysis

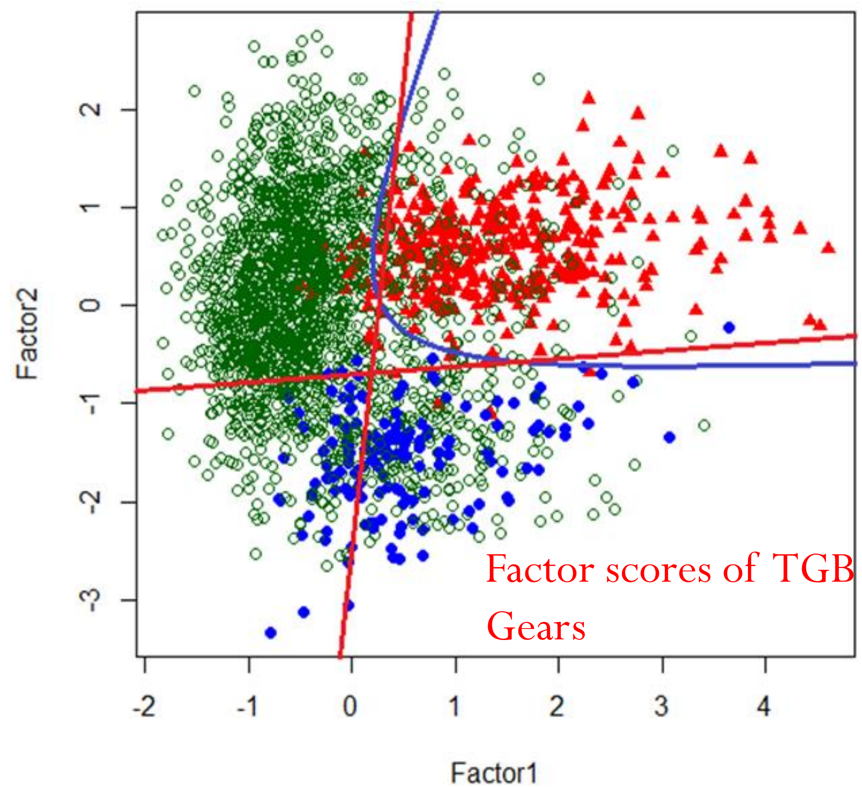
- Factor analysis aims to describe the structure of the correlations of a set of variables by means of a small number of underlying uncorrelated latent variables called factors

$$X = \mu + \Lambda F + U$$

Anomaly classification

- The **factor scores** of vector states give rise to further discriminant conditions



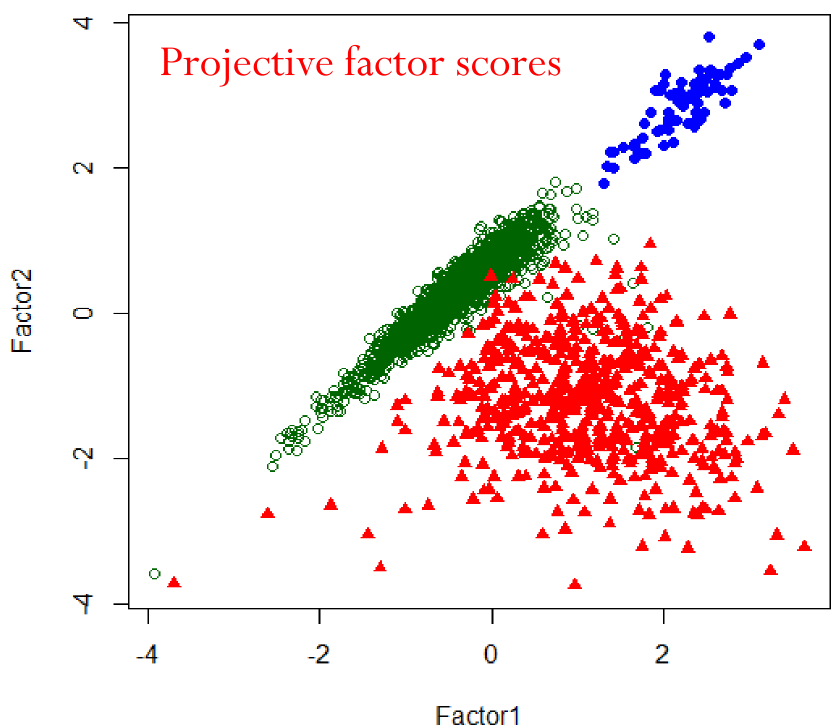


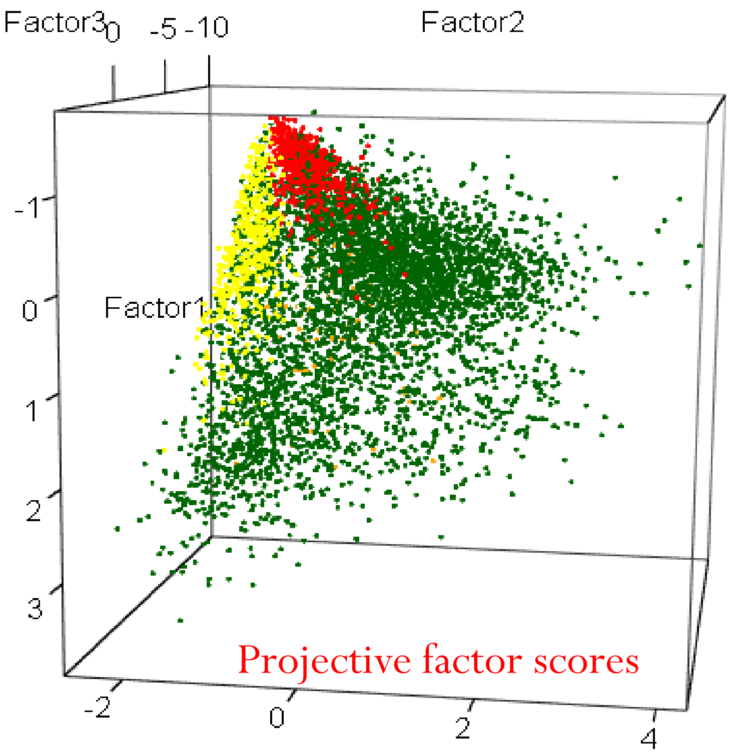
phmsociety

Anomaly classification

- Unusual “experiment” inspired by geometric considerations:

Let us consider the **projective structure** of the data set



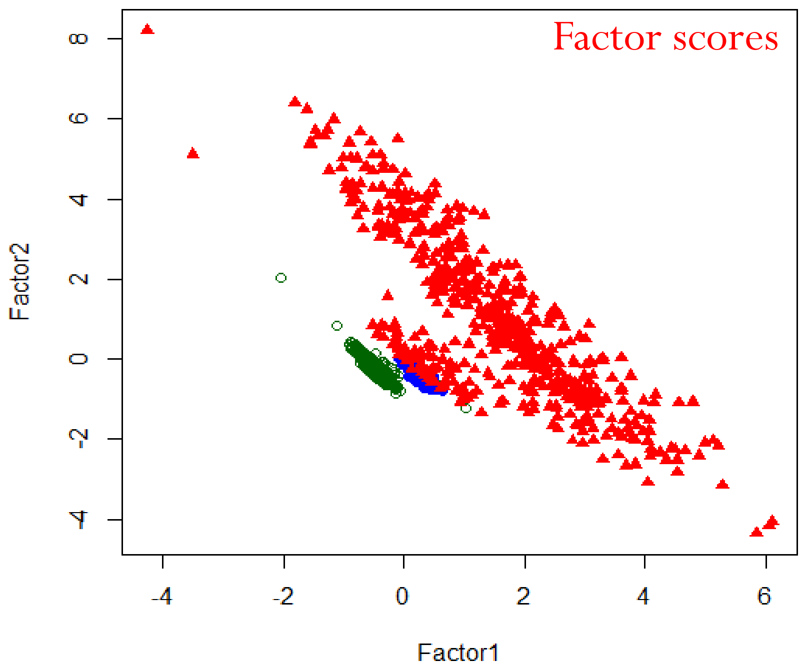


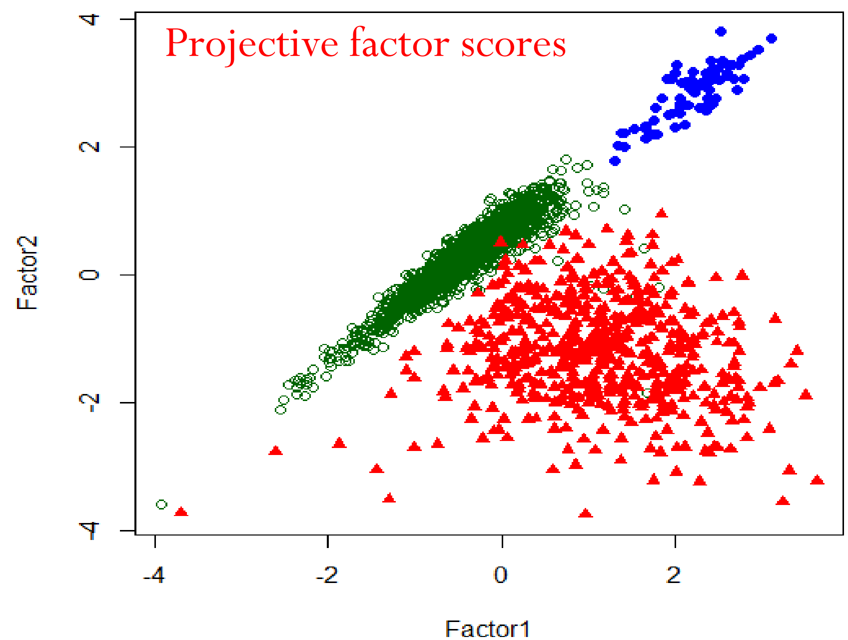
phmsociety

Anomaly classification

Useful effects of projectivisation:

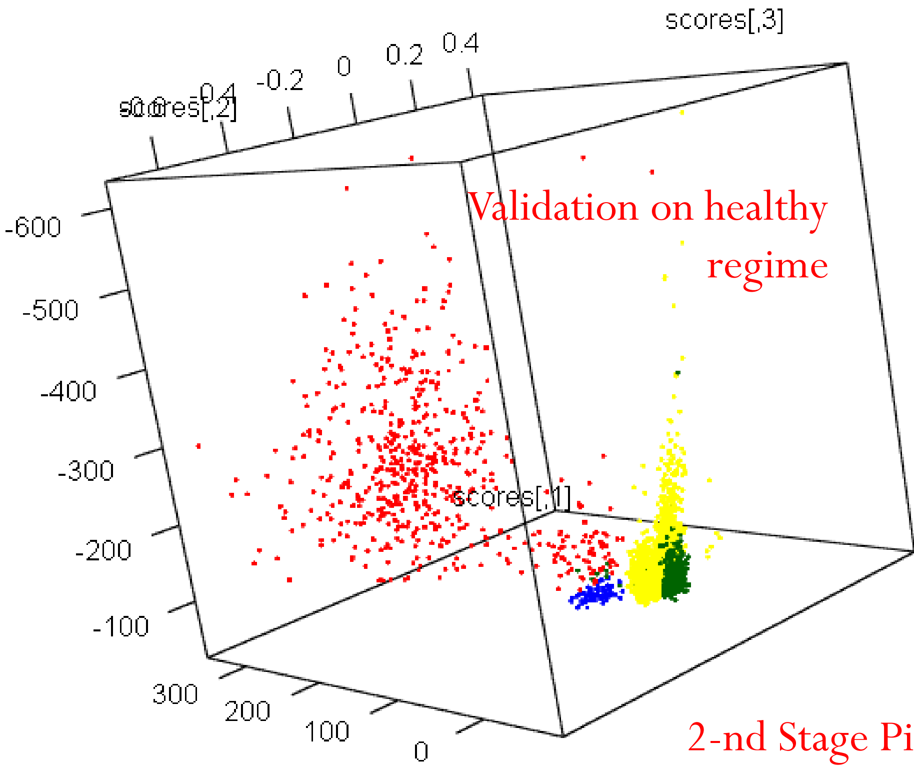
- Compactification of the clusters
- Cases in which the separation of true and false improves

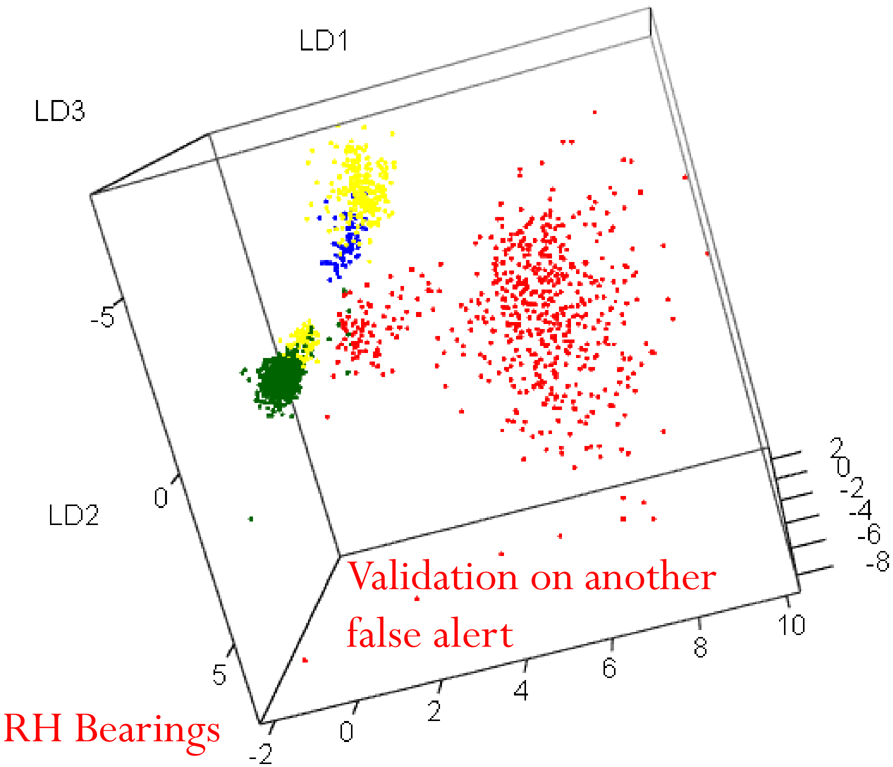




Anomaly classification

- **Validation** performed by means of calibration/validation data sets and leave-one-out procedures





2-nd Stage Pin RH Bearings

phmsociety

Implementation

- Implementation

COMPLETE INTEGRATED ENGINEERING TOOL

- Monitoring and control functions
- Calibration procedures of the statistical models

Implementation

- Structure of the integrated control process

Linear/quadratic recalibration procedure

↓

Detection - self-learning control chart - **individual**

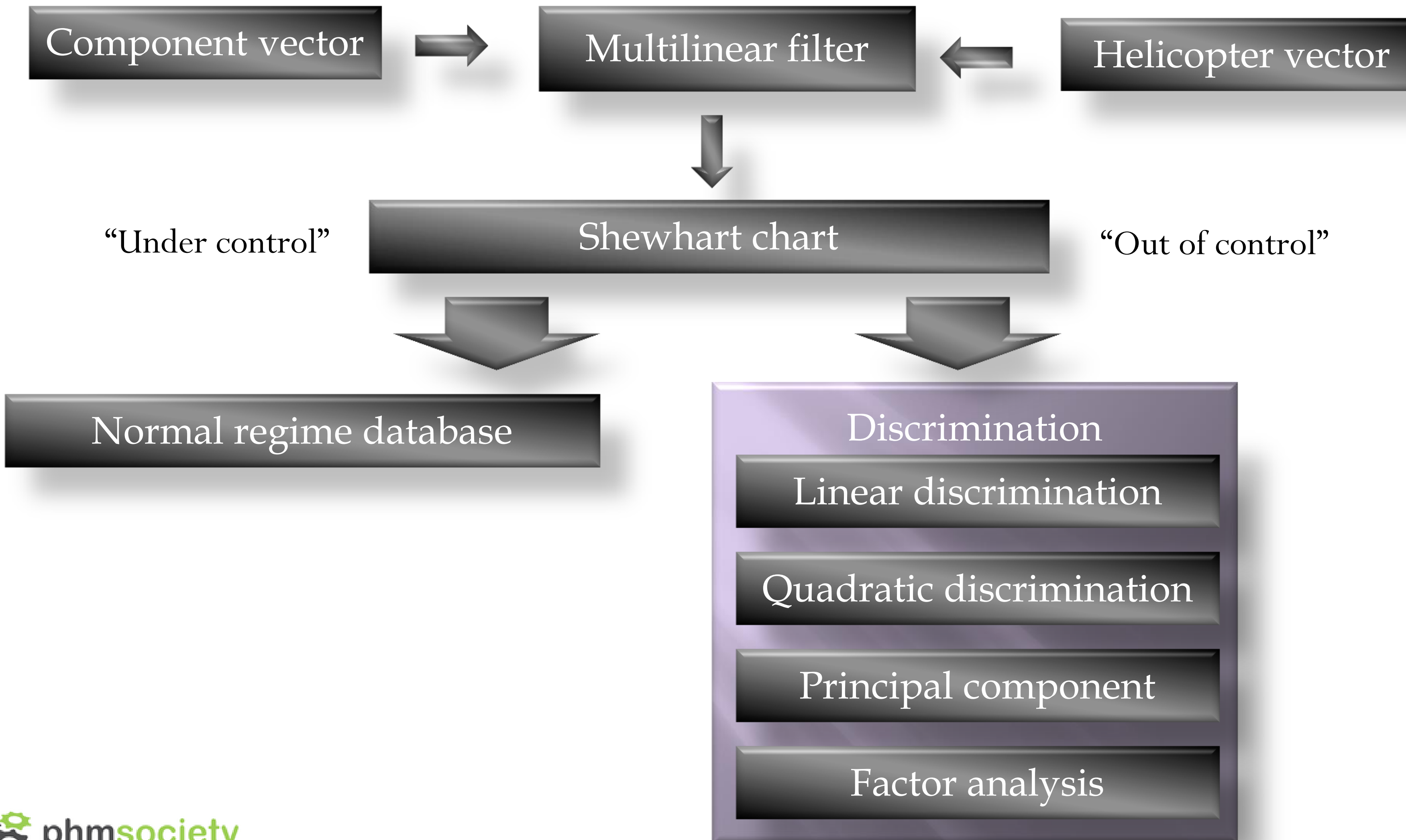
↓

Classification by PCA LDA QDA etc. - **fleet**

↓

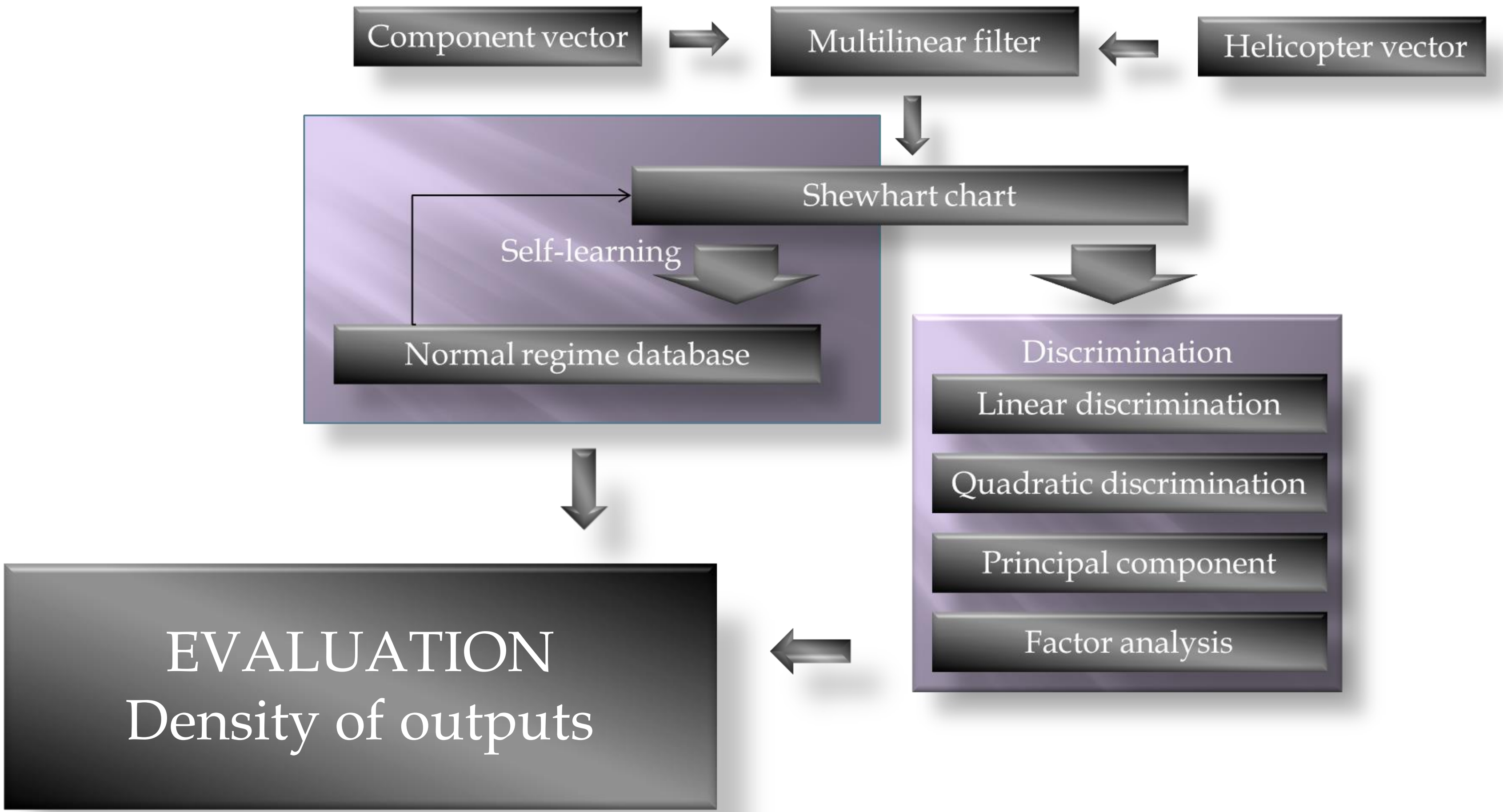
Evaluation by overall maximization of efficiency

Implementation



phmsociety

Implementation



phmsociety

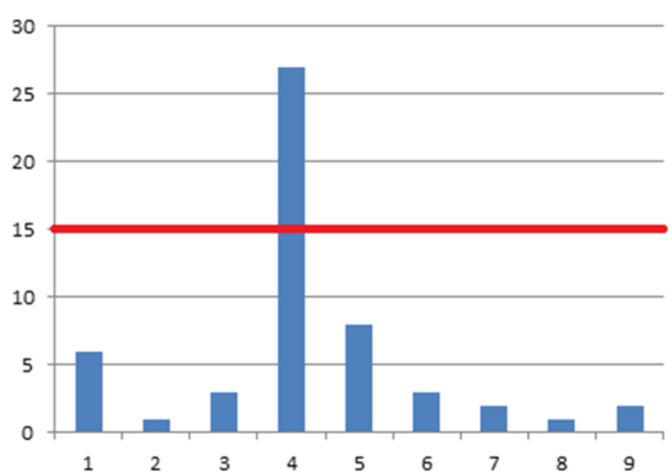
Effects

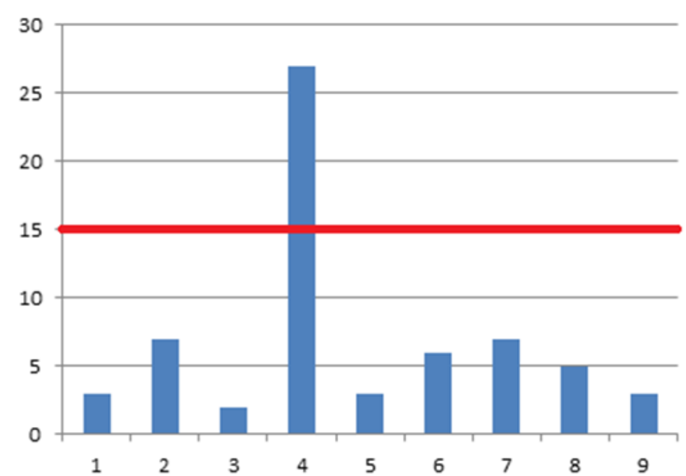
- The integrated control system has been tested several times on different sets of real in-service data: 50-60% less false alerts!
- Proved the loss of relevant information caused by the univariate processing

Even in those cases in which the univariate interpretation provides misleading results, the multivariate system separates true failures from false anomalies with very high level of statistical confidence

Possible development

- Diagnostic application of the multivariate profile
- Implementation of prognostic algorithms (directions etc.)
- A posteriori validation procedures for a diagnostic systems i.e. evaluation of the efficiency of given health indicators – mathematically non trivial problem
- Observability question: How much relevant information is lost because of the univariate processing of the indicators?





phmsociety

Second use case

- Joint research project

Company producer of industrial equipment + Politecnico di Torino

Problem statement

- Industrial equipment of the Company is installed worldwide in thousands of industrial plants owned by different customer companies
- Equipment operates in extremely heterogeneous productive contexts
- Factors: specific country, the company running the plant, market, climatic, seasonal and etc.

Problem statement

- For many issues the **industrial plant** is the relevant level of organization in this system:

  - equipment, material and spare parts supply

  - productive efficiency

  - maintenance services and contracts

  - safety

- **AIM**: accurate characterization of the "technical behaviour" at plant level: definition of efficient descriptive features, clustering and classification
- **BENEFITS:** strategic decisions, large scale anomaly detection, elaboration of specific services and prognostics

In-service data

- Advanced productive line monitoring system provides information on both the health and the mechanical efficiency of the productive lines
- High number of technical parameters (pressures, gear and bearing state indicators etc.) is measured and collected continuously
- The total amount of in-service data collected worldwide from all the monitored machines is very big

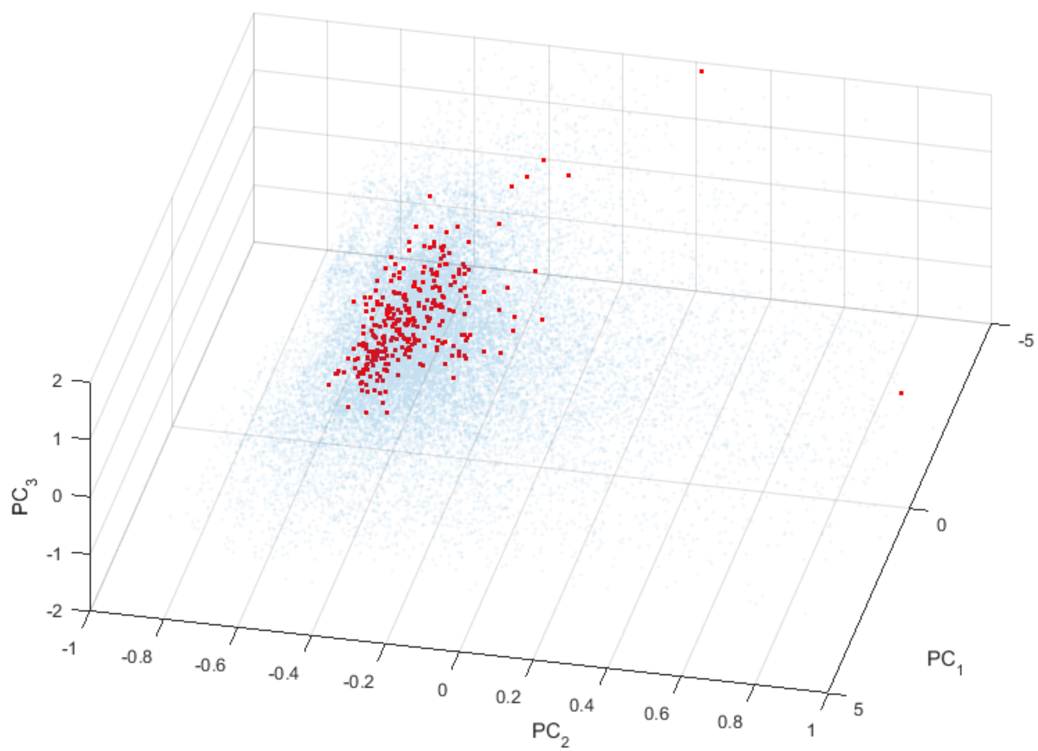
In-service data

- Information is condensed into a set of indicators to monitor the **mechanical performances** of each machine
- Different time scales relevant for different phenomena
- Set of ten mechanical performance indicators computed monthly for each productive line
- Take into account produced items, production time, occurrence of failures, maintenance stops and fault emergency stops, duration, recovery procedures etc.

Wish to find some hidden information

Multivariate statistics

- monitoring record of a plant is the record of all installed machines



Multivariate statistics

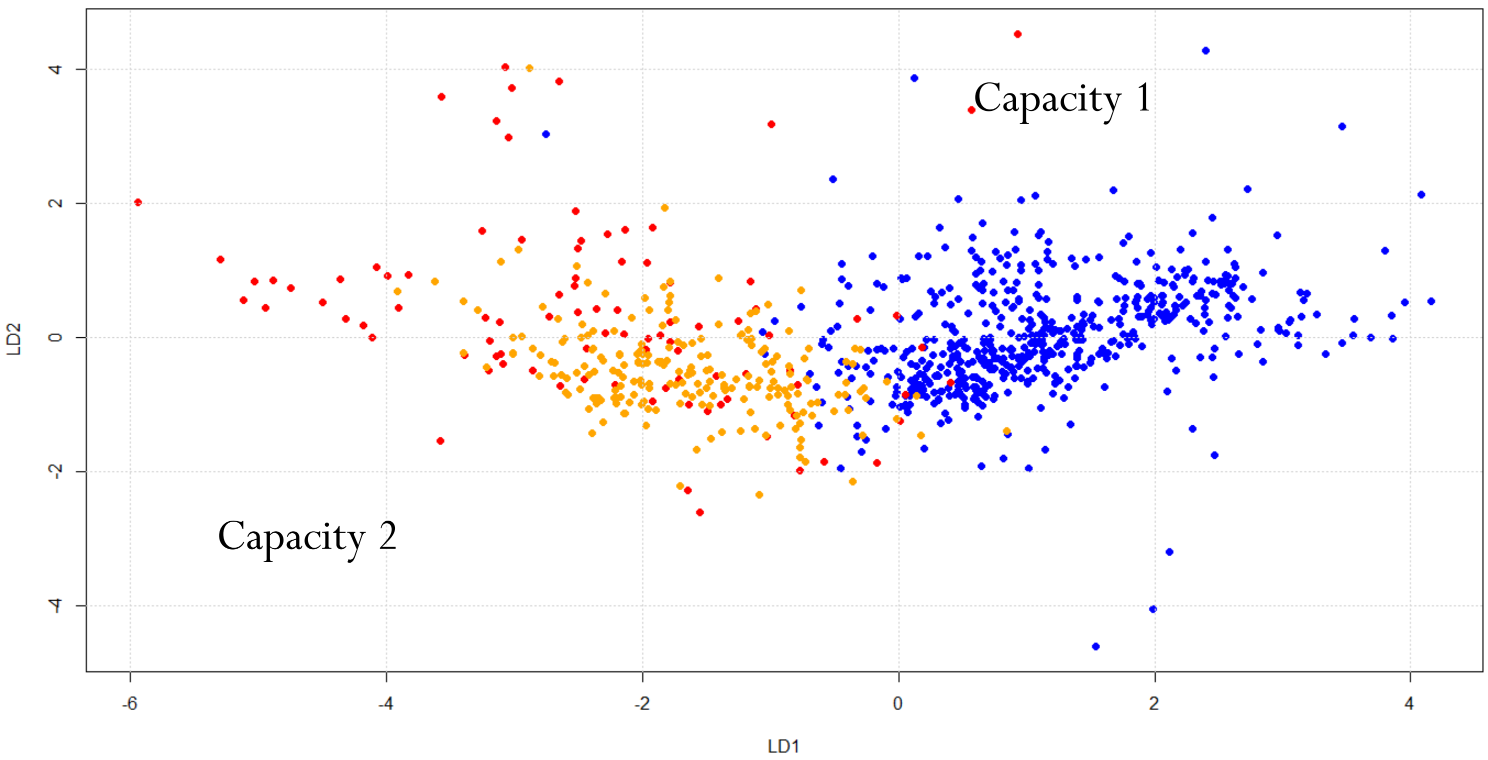
- Spread in huge portions of the parameter space
- Poor results of standard hierarchic or k-means clustering

More analysis

- The monitoring record is influenced by a series of qualitative factors:
  - Specific machine system
  - Characteristics of the produced items
  - Geographic (logistics)
  - Climatic
  - Corporative (company policies)
  - Social-economic factors (level of qualification)
- The impact of the factors is easy to detect in “homogeneous conditions”

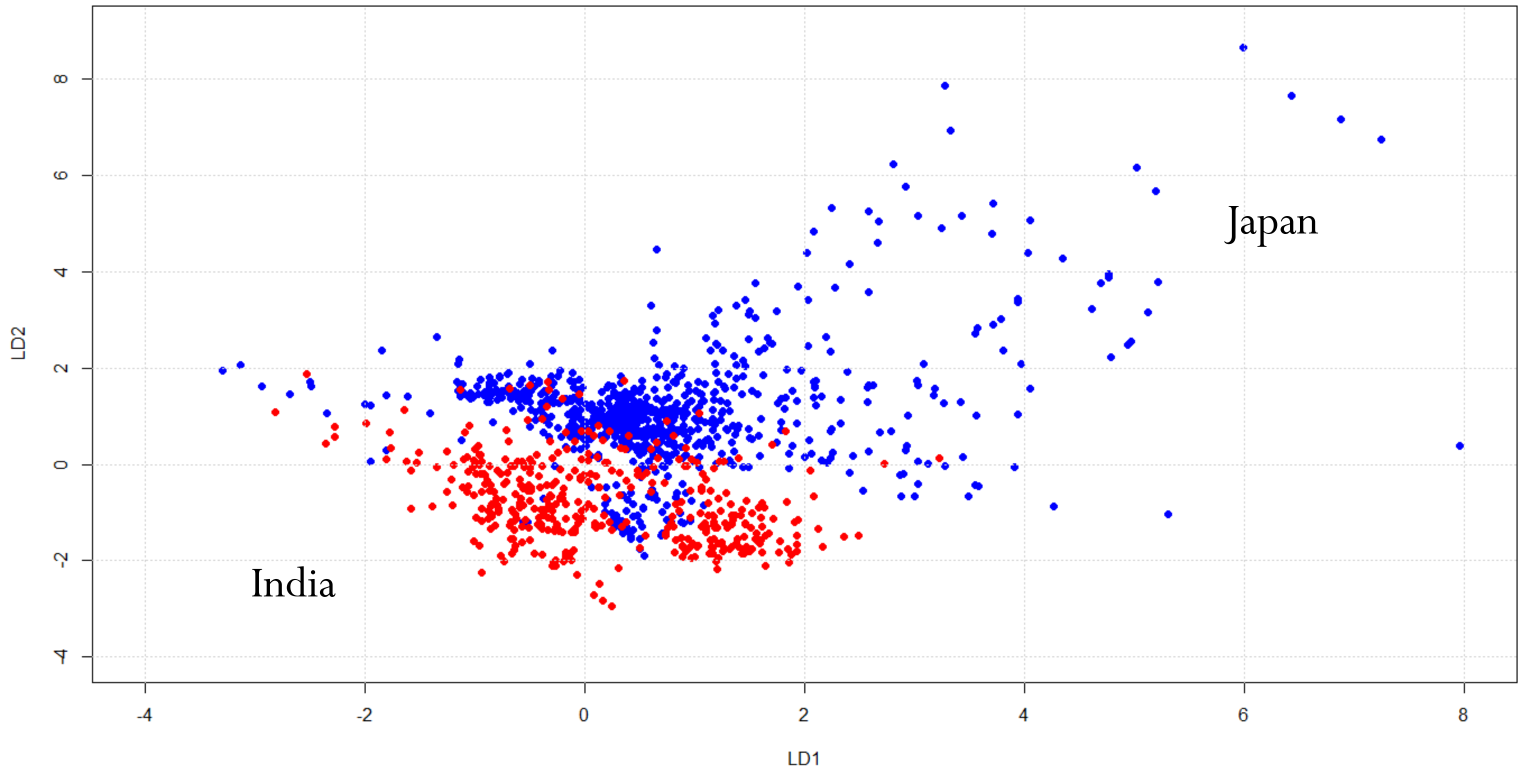
Homogeneous condition

- Impact of a specific item characteristics on the monitoring record



Homogeneous conditions

- Impact of the country on the monitoring record



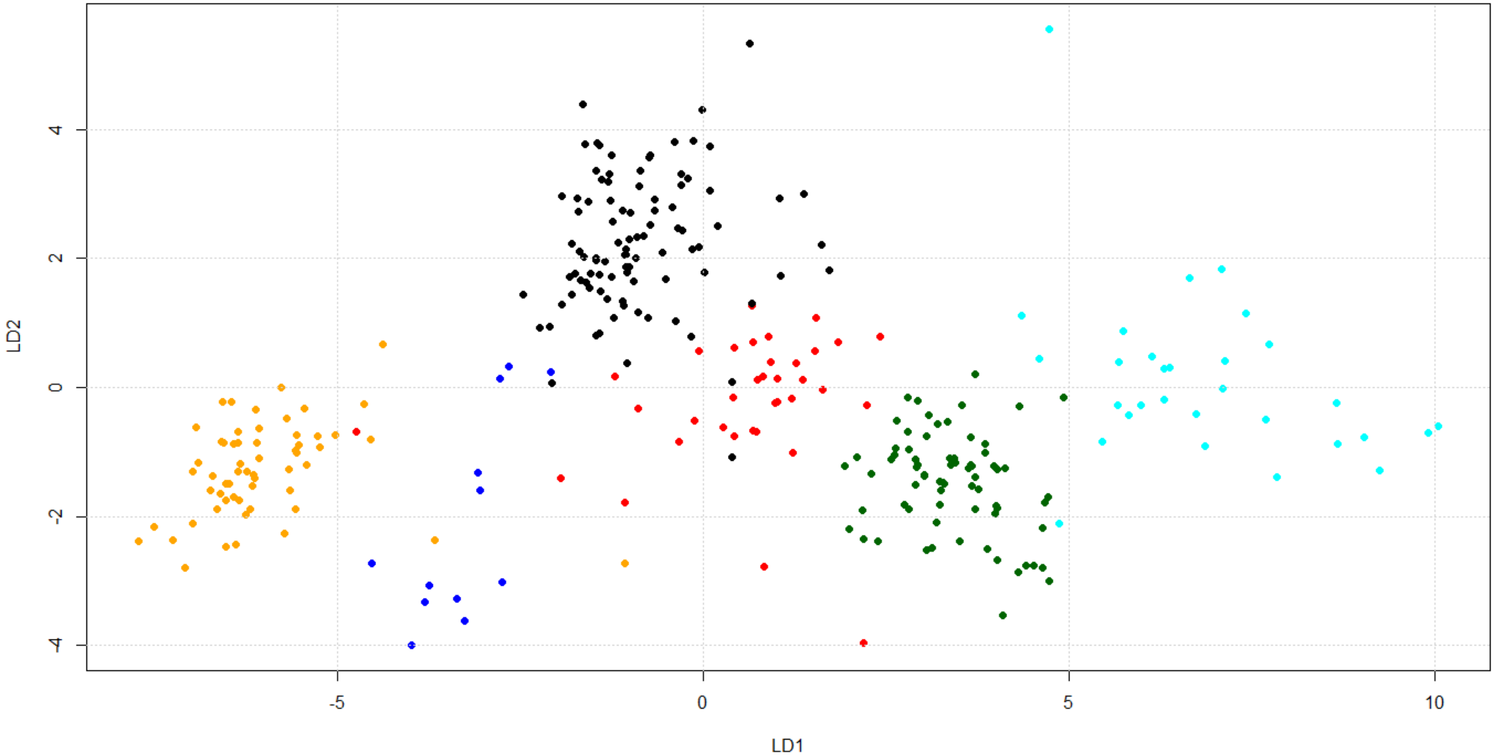
Heterogeneous conditions

Very high level of confusion!

What to do?

- To each plant has been associated its “fingerprint” i.e. its monitoring record divided in groups which are homogeneous with respect to the qualitative criteria
- Each homogeneous group is approximately normally distributed with small number of outliers (multivariate tests)
- This filtering procedure eliminates the deterministic impact of the qualitative factors

A typical plant fingerprint



IDEAS

- We develop two parallel ideas:
- IDEA1: Compare in terms of global features the shapes of plant fingerprints and classify them in segments of analogous objects
- IDEA 2: The fingerprint of a plant depends on some relations between its productive lines

(the key point is that in an appropriate mathematical formalism these two interpretations appear to be equivalent)

Complex adaptive systems

- Consider a system composed by agents and relations between agents
- Large scale phenomena, hardly modelled by local degrees of freedom or local interactions of agents or subsystems are called emergent
- Examples: large scale fluctuations, shock wave propagation, complex collective molecular motions, phase transitions etc.
- **Leading paradigm:** Emergent phenomena appear when local phenomena cannot be extended to global ones

Emergent phenomena

- Opposite tendencies in the local dynamical behaviour of the agents
- "Frustration" makes impossible the "synchronization" of the states of agents on large scale in a complex system
- Classical dynamical systems whose normal modes are global degrees of freedom as the vibrating string (harmony), phonon propagation etc.

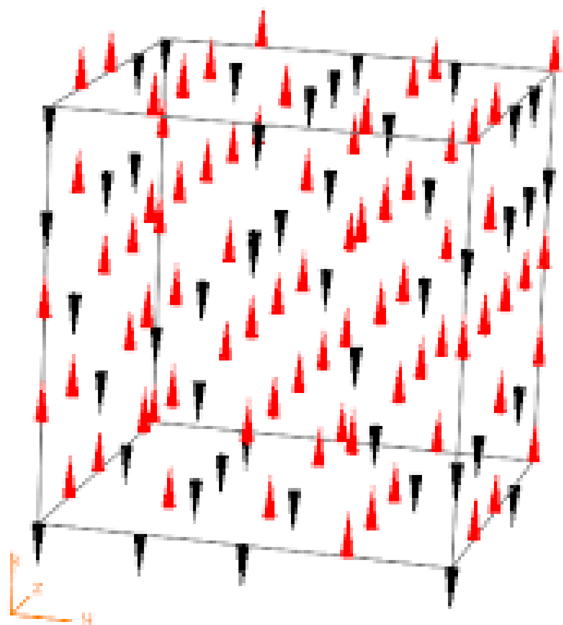
(Conservation laws and material constrains induce local "contradictions" between coupled parts of the system)

- Spin glasses
- Complex networks (for example Bianconi-Barabasi model)

- NB: very simple one-to-one relations between agents could give rise to non-local complex behaviour

Spin glasses

- Fluctuation of a spin glass system



"Internal state model"

- Describe a system by two sets of parameters

  - "Spatial" - provide "positions" of the agents

  - Internal - describe the state of an agent (random or deterministic)

- Very general setup, large class of systems are efficiently described

(Airports, spin glasses, complex networks, electric power transmission systems, arbitrage stock markets etc.)

- We call a system adaptive when the internal state of an agent depends on some interaction with other agents

Fibre bundles (or 2,5 min course in differential geometry)

- A real differentiable n-dimensional manifold M is a topological space which can be locally parametrized by real coordinate charts in $\mathbb{R}^N$
- A fibre bundle on a differentiable manifold is a differentiable manifold which is only locally isomorphic to the product of the base manifold M with some fibre space F
- Principal bundles – the fibre is a Lie group
- Associated vector (tensor) bundles – the fibre is a vector space
- Sections of vector bundles are analogs of vector valued functions on M
- The analogy is not complete! Smooth sections of certain bundles over certain manifolds could not exist
- A trivial bundle is globally isomorphic to the product MxF, it admits global smooth sections

Fibre bundles

- Fibre bundles are an appropriate tool for measuring the obstructions for certain point-wise geometric constructions to be extended to local or even global structures
- Plenty of examples for such phenomena coming from the theory of geometric structures on differentiable manifolds
- The integrability of Riemannian, complex, symplectic, spin structures measure the extent to which constructions defined on each tangent space can be defined on open subsets (existence of a set of holonomic i.e. compatible coordinates)

Local and global obstructions

- The Cartan's theory on integrability of geometric structures captures in coordinate independent way the local obstructions in terms of tensors on the variety i.e. sections of associated tensor bundles
- Torsion and curvature are typical examples for structure functions i.e. **local obstructions** represented by sections of tensor bundles
- Global properties of fibre bundles related to the topology of M
- Relations captured by **characteristic classes**
- Chern, Bott-Chern, Stiffel-Withney etc. cohomology classes etc. detect global obstructions for the extendibility of certain local phenomena (for example orientability of M, existence of a spin structure on the tangent bundle of M etc.)

Sections of vector bundles

- Generalization of the “internal sate model”: the set of possible states of a system is modelled over a vector (tensor) bundle over a smooth manifold of “spatial” parameters
- State of a system as a section of a vector bundle sampled in a finite number of points (its agents)
- The “geometrised” setup becomes interesting when the manifold of the deterministic parameters is not geometrically trivial
- Then non trivial bundles can be adopted for the description of the internal states by Introducing the stronger assumption that the state of the system is a sampling of a smooth section
- A strong “interaction” between the two types of parameters is introduced- “field of internal states”

Sections of vector bundles

- Interesting non local effects arise within this “geometrised” setup.
- The very identification of the internal states of the agents becomes a non local procedure.
- Parallel transport must be defined (a “signal” must be transmitted along the manifold).
- The “synchronization” of the internal states of the agents of a system - path-dependent problem - different paths with conflicting outputs
- The geometric obstructions (torsion, curvature, non trivial characteristic classes) for the extension of local phenomena introduce and model complexity of a system.

Manifold learning

- Manifold learning methods had origin from the idea that multivariate data are distributed on (or nearby) a differential submanifold M immersed in the space of parameters $\mathbb{R}^N$
- Manifold learning seems to be a suitable tool for treating big sets of multidimensional data:

1. Captures the geometric structure valid in the limit of continuous sampling of points on M

2. IF the underlying structure is well characterized, it is possible to regress and interpolate data in a efficient way i.e. reduce the quantity of data to handle

Manifold learning - redefined

- A more general reformulation of the manifold learning paradigm:

A feature map is a smooth section Y of a vector bundle E over a smooth manifold M sampled in a finite number of points X_i

- For example computing the dimension of M corresponds to a special case of feature extraction. i.e. extract a number of linearly independent tangent vectors in each point (local PCA)
- In view of Nash's C^1 isometric embedding theorem $\mathbf{X_i}$ can be considered as points immersed in $\mathbb{R}^N$ and M naturally inherits a Riemannian metric form the Euclidian space.

Standard methods

- Sections of vector bundles (fields) are “excitations” of non-local degrees of freedom on M
- Relevant sections can be for example detected by minimizing certain functionals on M
- The most popular manifold learning methods can be “tautologically” assimilated in this context. (In fact a global section of a vector bundle is an embedding of M in the fibre space)
- Given a finite sampling (X_i, Y_i) of a vector bundle section, functionals are defined by finite summations:

$$F(Y_i) = \sum_{i<j} L(X_i, X_j, Y_i, Y_j) Wij$$

Affinity graphs

- The role of the "affinity function" W_{ij} is to localize the contribution to the evaluation of the functional to some neighborhood of $\mathbf{X}_i$

$$W_{ij} = \begin{cases} 1 \text{ if a condition is fulfilled} \\ 0 \text{ if the complementary condition is fulfilled} \end{cases}$$

- Affinity graphs can be constructed by assigning edges from Xi to
  - the k nearest points
  - to the points closer than a fixed constant

L usually contains some relevant "characteristic" of the section which need to be preserved.

Standard methods

- Curvilinear Component Analysis: (d is the Eucledian distance)

$$L = (d(X_i, X_j) - d(Y_i, Y_j))^2, \; W_{ij} = \begin{cases} 1, & d(X_i, X_j) \le \epsilon \\ 0, & d(X_i, X_j) > \epsilon \end{cases}$$

- Curvilinear Distance Analysis: same thing with distance along a graph

- Multidimensional Scaling Method

$$L = (f(d(X_i, X_j)) - d(Y_i, Y_j))^2, \; W_{ij} \equiv 1$$

- ISOMAP: same thing with graph (geodesic) distance
- Laplacian eigenmaps:

$$L = \sum_{i,j=1}^{N} (Y(X_i) - Y(X_j))^2 W_{ij}$$

Differential operators on Manifolds

- We are interested in "interacting sections" i.e. differential equations
- Differentiable operators on manifolds is a broad area in the modern (geometrical) mathematical analysis
- Differential operators act on sections of vector (tensor) bundles

$$\mathscr{D}\phi = \psi$$

- Particularly powerful approach produced far reaching implications (for example the celebrated Atiah-Singer Theorem, Calabi-Yau etc.)
- Fredholm's theory - very helpful for introducing differential operators in discrete context

Tensorial Green's functions

- Green's function G(X,Y) associated to a differential operator is a special integral kernel

$$\mathscr{D}\mathbf{G}(X,Y) = \mathbf{I}\delta(X-Y),$$

- In these terms

$$\phi(X) = \int_M \mathbf{G}(X,Y)\psi(Y)dY.$$

- If the operator has discrete spectrum

$$\mathbf{G}(X,Y) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i}\phi_i(X) \otimes \phi_i(Y).$$

Coarse discretization

- **DEFINITION** Consider a vector bundle E on a smooth manifold M and linear differential operator

$$\mathscr{D} : \Gamma E \longrightarrow \Gamma E$$

Denote by **G** the Green's function of $\mathscr{D}$, given a finite sampling $X_i \in M$ the coarse discretisation D is a linear mapping defined by:

$$D\mathbf{G}(X_i, X_j) = \mathbf{I}\delta_{ij}$$

The transformation

$$DY(X_i) = Z(X_i)$$

Is obtained by inverting

$$Y(X_i) = \sum_{j=1}^{N} \mathbf{G}(X_i, X_j) Z(X_j) \quad \text{or} \quad Y(X_i) = \sum_{j} \mathbf{G}(X_i, X_j) Z(X_j) W_{ij}$$

Finite dim. representations

- A feature can be represented by Nd-dimensional vector (a vector which contains N d-dimensional vectors)
- The action of D can be implemented by an Nd x Nd matrix A, a matrix with NN blocks of dimension dxd.
- Each dxd block represents a linear transformation in F which is “weighted” by a scalar factor.
- The affinity function is also constant on each block
- In particular analogous spectral decomposition:

$$\mathbf{G}(X_i, X_j) = \sum_k \frac{1}{\lambda_k} Y_k(X_i) \otimes Y_k(X_j)$$

Sample discretisation

- A manifold learning problem: How to build a sample estimate of a differential operator starting from a point cloud?
- Does a sample discretization coincide with a coarse discretization of the differential operator?
- A sample discretization is built by the same ingredients as D

  - An affinity graph

  - A set of linear transformations in each point

  - A kernel function $K(X_i, X_j)$ - certain regularity assumptions

$K(X_i, X_j)$ aims to reproduce the unknown Green's function of the manifold in terms of the coordinates of the immersed points Xi.

(natural guess- the Green's function on $\mathbb{R}^N$ corrected by curvature, variance and other terms)

Convergence

- Given a sample discretization of a differential operator, two types of convergence problems arise in the continuous limit of sampling.

  - point-wise convergence $DY(X_i) = Z(X_i) \longrightarrow \mathscr{D}\phi = \psi;$

  - spectral (stronger) convergence means that eigenvectors of D converge to eigenfunctions of $\mathscr{D}$ on M.

- The convergence of a sample discretisation of an operator is a highly non-trivial analytical problem
- Depends on the manifold M and its boundary, on the density of sampling, normalization procedures, opportune parameters choices (absorbing, re-defining overall constant factors, volumes) etc.
- Remarkable results in this direction have been achieved!

Kernel methods

- Kernel manifold learning methods can be included in our general setup
- Primary feature extraction remains implicit $Y(X_i)$. Sample covariance matrix (a new feature):

$$C = \frac{1}{N}\sum_i Y(X_i) \otimes Y(X_i)$$

- Given kernel function

$$\langle Y(X_i), Y(X_j) \rangle := k(X_i, X_j)$$

- compute the components of the feature vectors with respect to the basis of eigenvectors of C
- Global orthogonal basis change on the fibres of the bundle determine non-linear dimensional reduction
- Interpretation: passive gauge transformation (no interacting field)

The rough (connection) Laplacian

- Rough Laplacian operator acts on tensor fields on a Riemannian manifold M, defined as the trace of the second covariant derivative

$$\Delta f = -Tr \nabla_{LC} \nabla_{LC} f$$

- The rough Laplacian acting on real-valued functions is the Laplace-Beltrami operator, divergence of the gradient of a function (the trace of the Hessian.
- Spectral properties of Laplacian largely exploited in geometrical analysis.

The role of the Laplacian operator

- Laplacian operator related to two interesting contexts:

  1. Diffusion processes on manifolds

$$\begin{cases} \left(\frac{\partial}{\partial t} + \Delta\right) f = 0 \\ f(x,0) = u(x) \end{cases}$$

  2. Theory of harmonic functions, morphisms and mappings

$$E(f) = \int_M |df| d\omega$$

  (Harmonic morphisms preserve the Laplace's equation)

- Both interesting – "synchronization" and "global extremization" of fields
- Both have deep geometric meaning

Vector diffusion maps

- Vector diffusion maps represent discrete vector diffusion process of a sampled manifold
- Construction of a discrete Laplacian on the tangent bundle converges pointwise and in spectral sense
- Generalized construction includes Laplacian operators on other vector bundles

Principal bundle via manifold learning

- $\mathbb{R}^N$ is considered as the total space of a principal bundle
- Base manifold M and fibre is the set of nuisance parameters
- Structure group G acts simply and transitively on the vector fibre spaces

  (M parametrizes the set of orbits)
- Natural notion of parallel transport - products of the linear transformations associated to a sequences of vertices, acts on each associated bundle
- Special sections of the principal bundle obtained by optimal alignment procedure

Parallel transport

- The “synchronization” of the internal states of Xi and Xj by weighted sum of transformations along many paths.
- A path dependent “synchronization” which “contrasting outputs” – important for modelling complexity

Parallel transport procedure

$\updownarrow$

Affine connection on the principal bundle

(covariant derivative on the associated one)

$\updownarrow$

Gauge field via which sections of associated bundle interact

Assumptions

- The base manifold M is smooth and smoothly embedded in $\mathbb{R}^N$ a metric is induced by the canonical Euclidean metric, its boundary is either empty or smooth
- A principal bundle with a connection and an associated bundle on M are fixed, a metric is induced on the fibre space
- A probability density function $p \in C^3$ is uniformly bounded from above and from below
- A data cloud is sampled according to the probability density function, the elements in the fibre of the principal bundle are sampled uniformly over the structure group G
- The kernel function K is a strictly positive

Discrete Laplacian operators

- An undirected affinity graph W is constructed

$$\mathbf{P}_{ij} = \begin{cases} K(X_i, X_j) g_{ij} & (X_i, X_j) \in W \\ \mathbf{0}_{d\times d}, & (X_i, X_j) \notin W \end{cases}$$

- g is an orthogonal transformation on the vector state defined by summing the orthogonal transformations associated to all paths of certain length connecting i and j.
- K is normalized with respect to the sampling probability density then

$$\mathbf{D}_{i,i} = \sum_{j|(X_i, X_j)\in W} K(X_i, X_j)\mathbf{I}.$$

Discrete Laplacian operators

- Explicit form of the operator (and the true Green's function) involves sections (features) of several vector bundles, TM or better (the cotangent bundle), the Riemannian metric on M
- Great flexibility
- Interesting result (metric change):

Every symmetric local kernel with exponential decay corresponds to Laplacian operator in a Riemannian geometry and conversely any Riemannian geometry can be represented with an appropriate local kernel

Discrete Laplacian operators

- With these assumptions

$$\mathbf{D}^{-1}\mathbf{P}Y(X_i) - \mathbf{I}Y(X_i) \longrightarrow \Delta Y(X_i)$$

- Furthermore, spectral convergence Theorem:

Eigenvectors $Y(X_i)$ of $\mathbf{D}^{-1}\mathbf{P}-\mathbf{I}$ are discrete approximations of the eigenvector fields of the rough Laplacian on M with homogeneous Newman boundary conditions

$$\begin{cases} \Delta Y(X) = -\lambda Y(X), & X \in M \\ \nabla Y(X) = 0, & X \in \partial M \end{cases}$$

Eigenvectors of the discrete operator (linear maps) are special global degrees of freedom!

Discrete Laplace-Beltrami operator

- General construction is applied to real-valued functions, recover the diffusion maps and Laplacian eigenmaps models.
- Principal bundle with fibre G = e
- trivial total parallelism on the associated bundle.

- The discretisation of the operator is implemented by NxN matrices
- Normalised graph Laplacian:

$$L = \mathbf{D}^{-1}\mathbf{P} - \mathbf{I}$$

“Tension” functional

- Both harmonic and diffusion aspects collimate in this case:
- Minimize a discrete version of the tension functional

$$E(f) = \sum_{i,j=1}^{N} (Y(X_i) - Y(X_j))^2 W_{ij}, \;\; Y^T D Y = 1$$

$$E(f) = \int_M \|\nabla f\|^2 dx = \int_M \langle \nabla f(x), \nabla f(x) \rangle dx, \;\; \|f\| = 1$$

- On compact M $|f(x) - f(y)| \leq \|\nabla f(x)\| d_{\mathcal{M}}(x, y) + o(d_{\mathcal{M}}(x, y))$

- The minimization of the functional is done by the eigenvectors corresponding to smallest nonzero eigenvalues of the discrete Laplacian

Diffusion

- The solution of

$$\begin{cases} \left(\frac{\partial}{\partial t} + \Delta\right) f = 0 \\ f(x,0) = u(x) \end{cases}$$

- Is written as $f(x,t) = \int_{\mathcal{M}} H_t(x,y) u(y) dy$

- Where the heath kernel $H_t(x,y) = (4\pi t)^{-\frac{M}{2}} e^{-\frac{\|x-y\|^2}{4t}} (\phi(x,y) + o(t))$

$\|x-y\| \to 0$ and $t \to 0$ $\quad H_t(X_i, X_j) \approx (4\pi t)^{-\frac{M}{2}} e^{-\frac{|X_i - X_j|^2}{4t}}$

- The eigenvectors of the discrete Laplacian are relevant global degrees of freedom on M

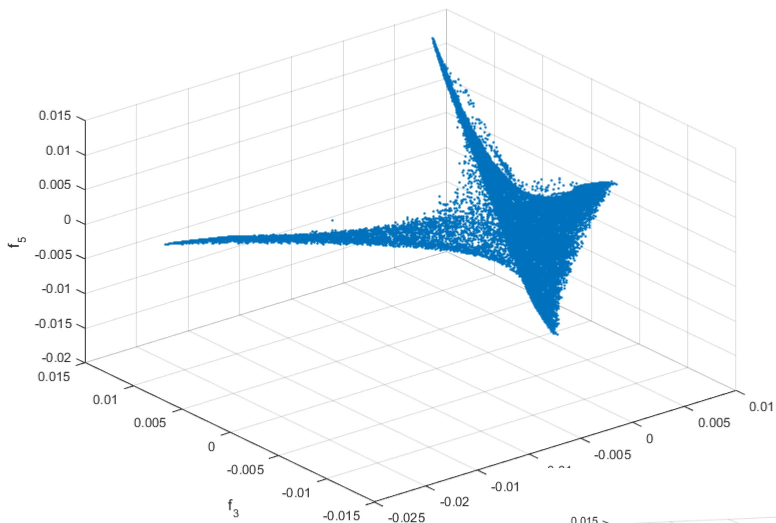
A model

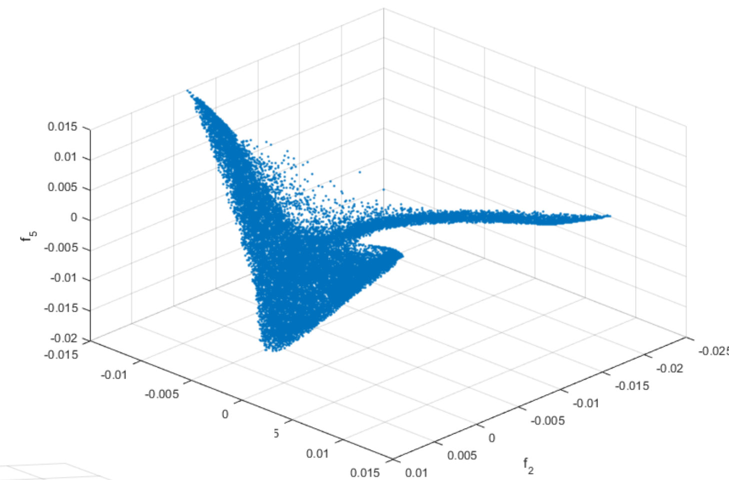
- The set of agents (single productive machines) is embedded in the physical space of the monitoring parameters
- efficiency behaviour at a plant level depends on the “interactions” between its agents (machines) in the monitoring space
- Effects captured by non local characteristics of the geometric structure of the fingerprint
- We model the internal state by real valued function
- Apply Laplacian eigenmaps with standard heath kernel (the distribution is overlapping of normally distributed homogeneous groups)

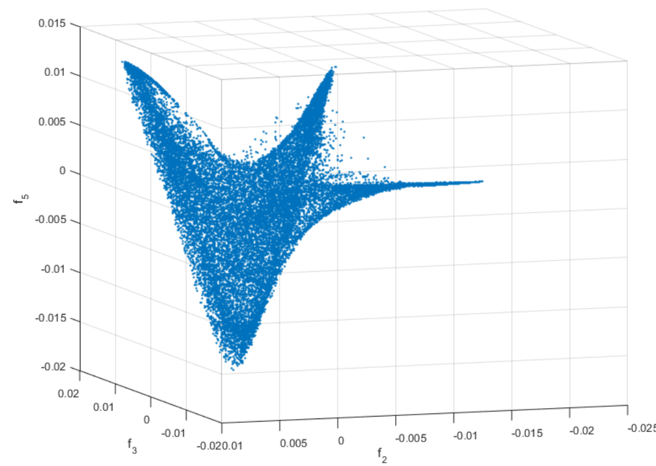
Remarks

- The discrete Laplace- Belrami operator transforms scalar “material” fields
- The analysis of its spectral structure provides deep geometric information on the manifold on which the monitoring data are sampled.
- The Laplace-Beltrami operator involves the Levi-Civita connection on the base manifold
- The extremisation of the tension functional involves local contrasting tendencies
- (Our interpretation goes beyond the well-known applications of the Laplace-Beltrami operator in spectral shape analysis)

Laplace scores of the monitoring record







Comparison with PCA and Kernel PCA

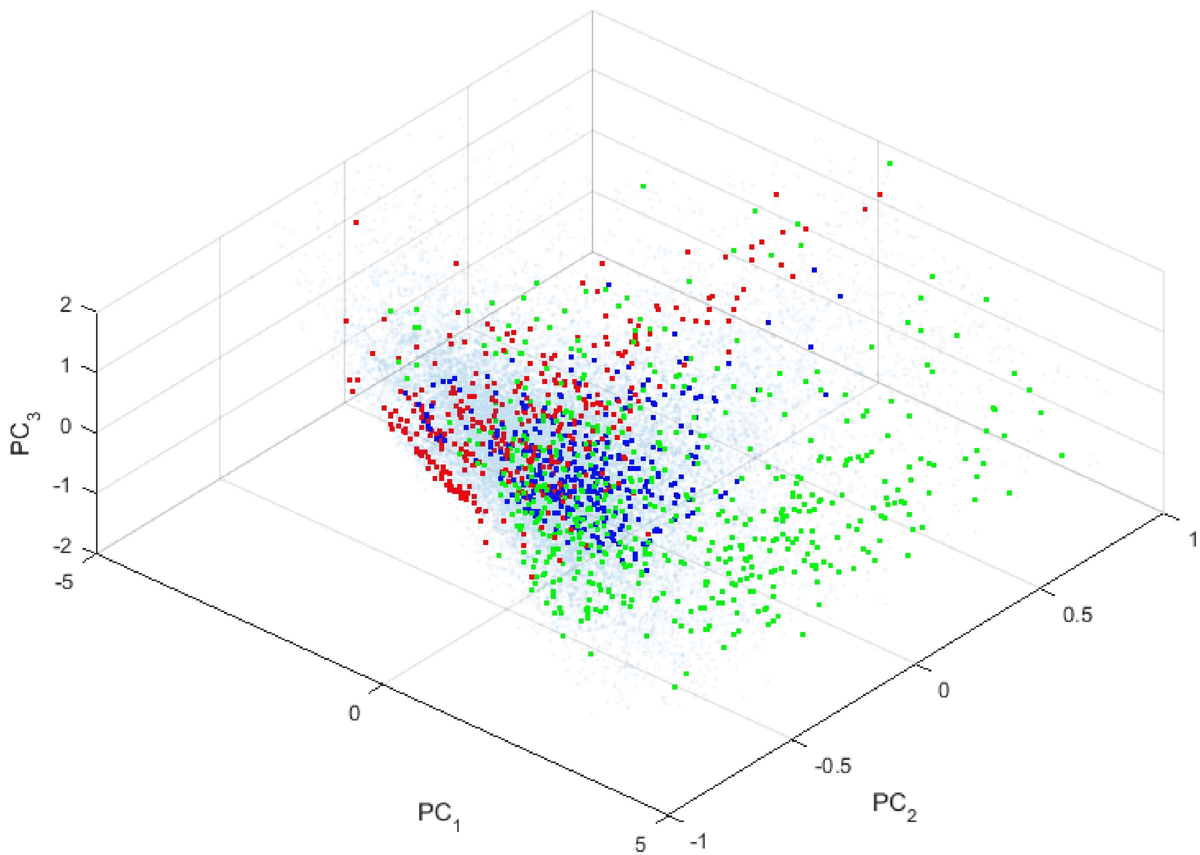
- The overall mechanical performance of a plant is modelled as an emergent phenomenon which depends on relations between its agents (lines)

RESULT: monitoring record of each single plant becomes compact and homogeneous in the space generated by the Laplacian eigenfunctions (global degrees of freedom)

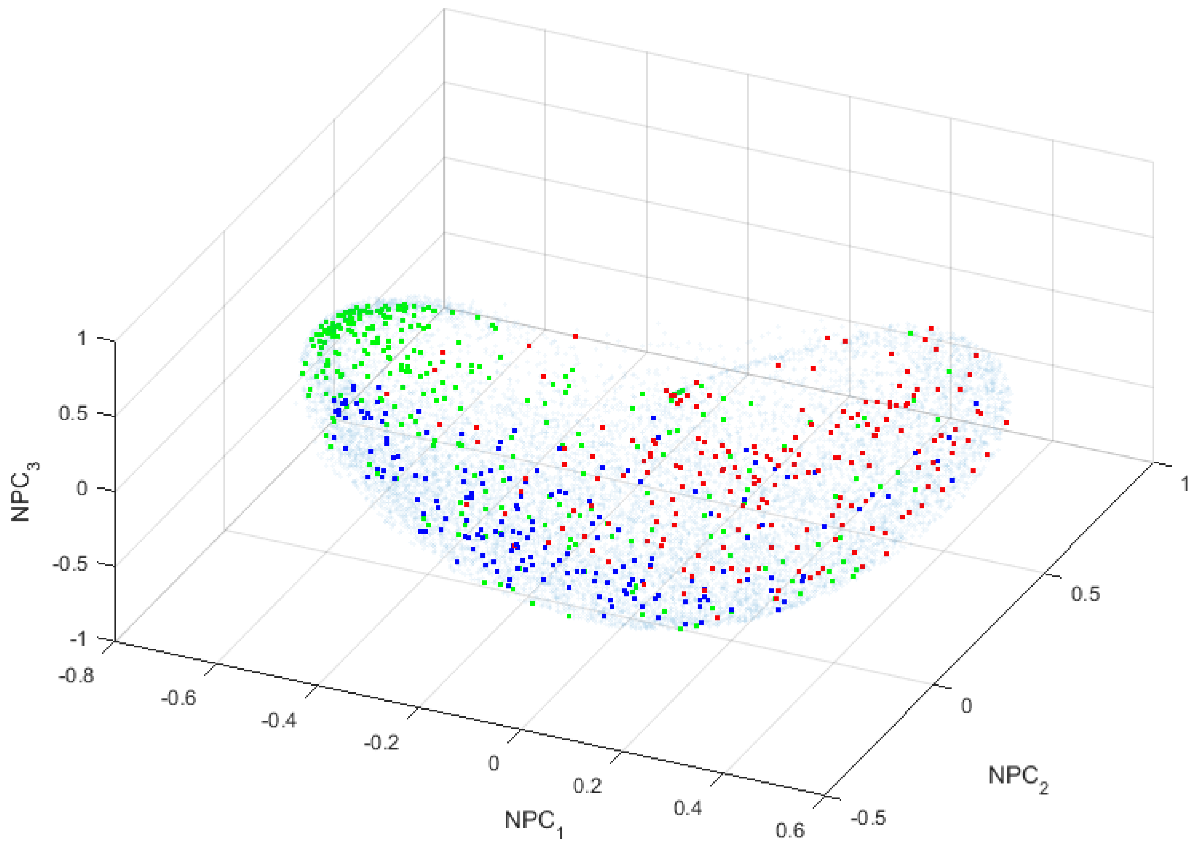
- As a check we appy kernel Principal Component with **the same heath kernel!** (mathematically different dynamic v.s. passive transformations)

$$\langle Y(X_i), Y(X_j)\rangle = k(X_i, X_j) = e^{-\frac{\|X_i - X_j\|^2}{\sigma^2}}$$

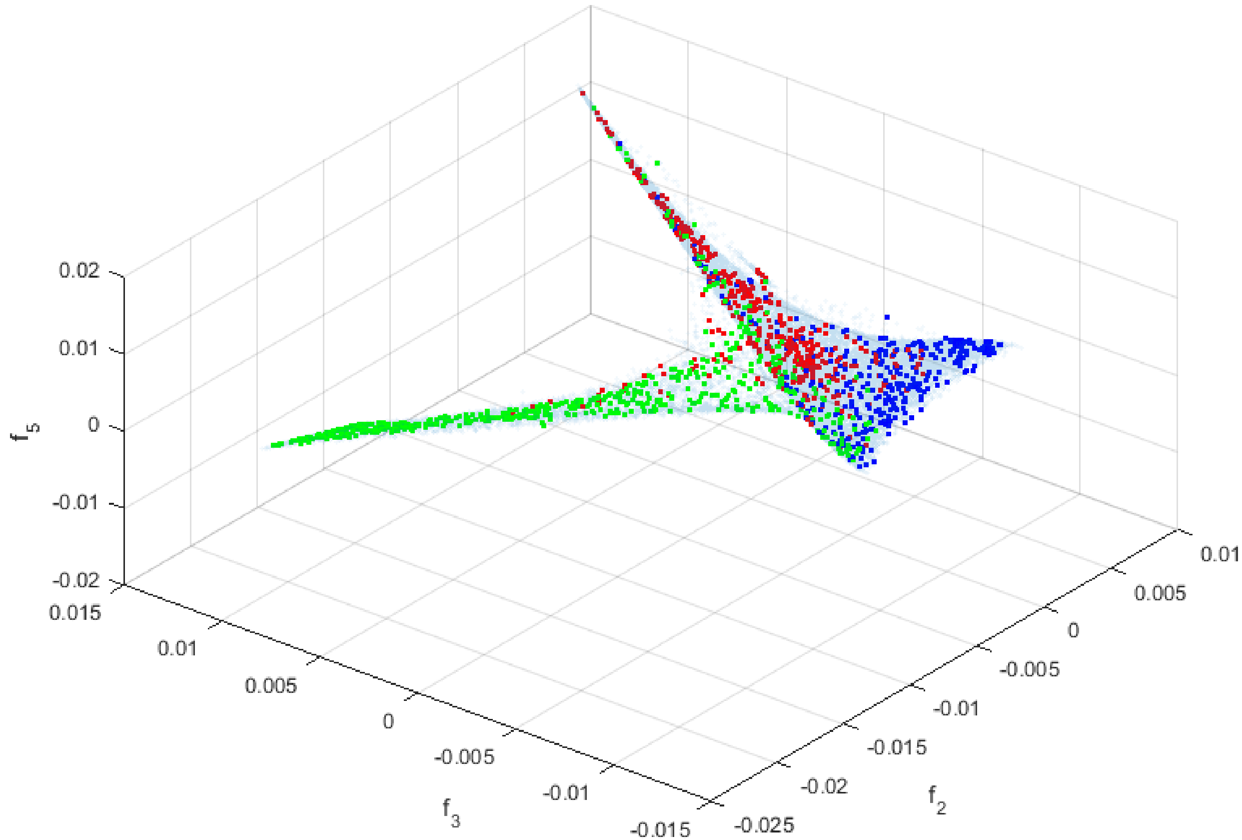
Example 1: Physical data



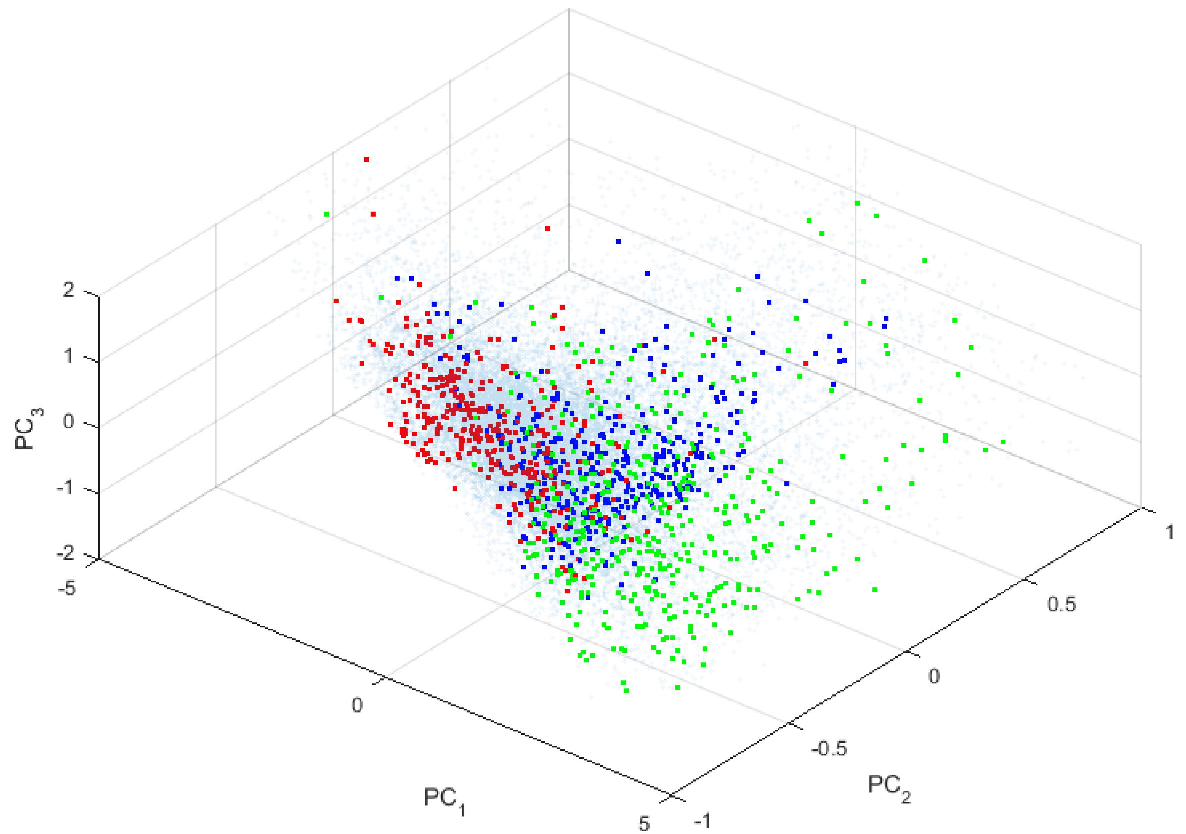
Example 1: Kernel PCA



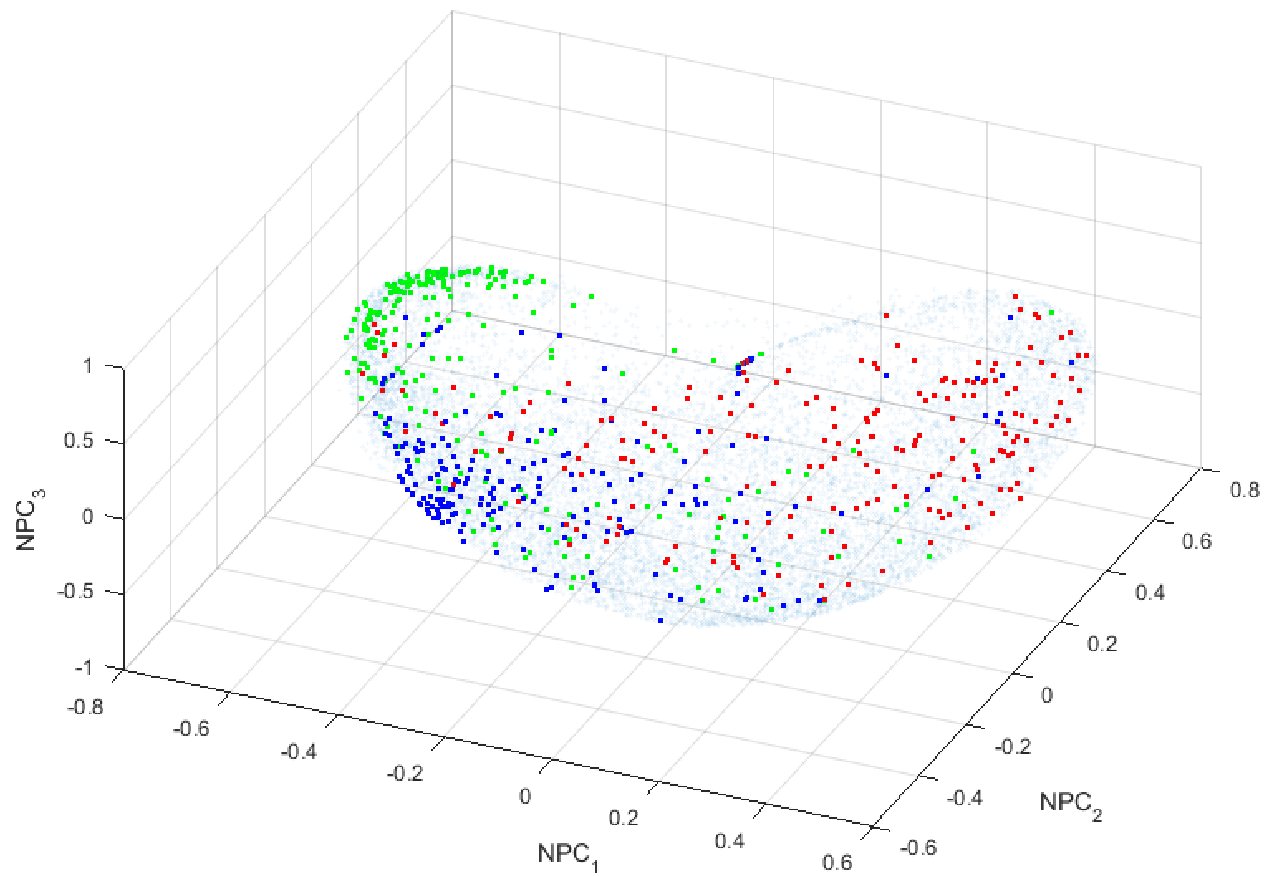
Example 1: “Laplacian” scores



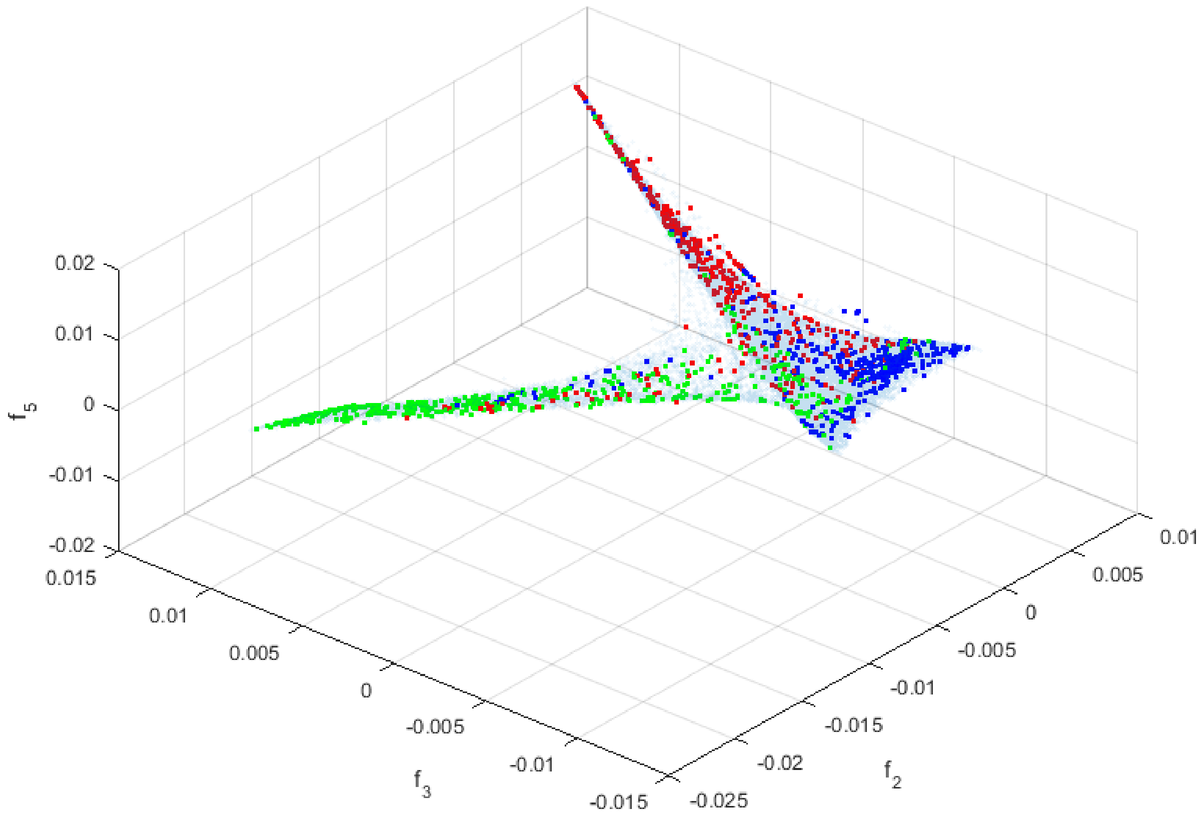
Example 2: Physical data



Example 2: Kernel PCA



Example 2: Laplacian scores



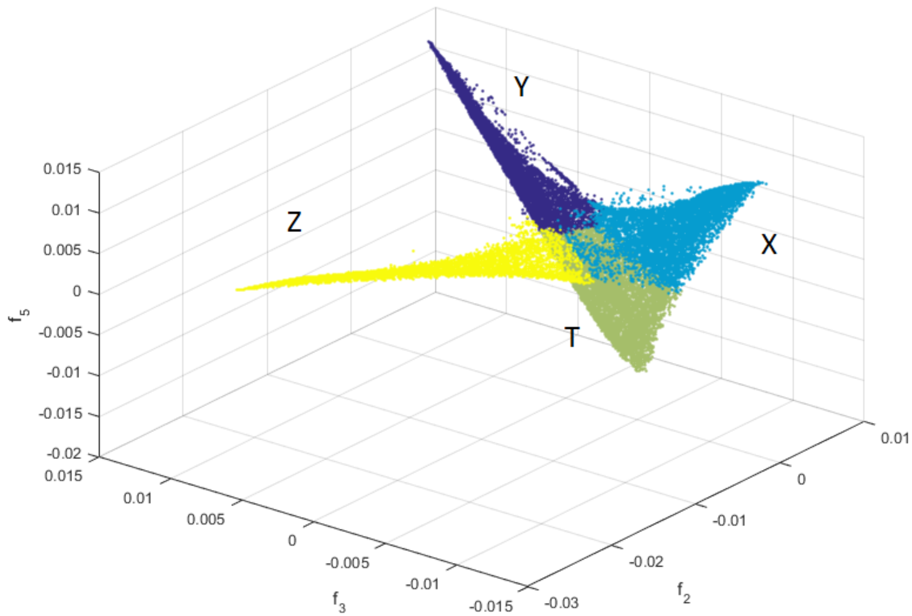
Classes of plants

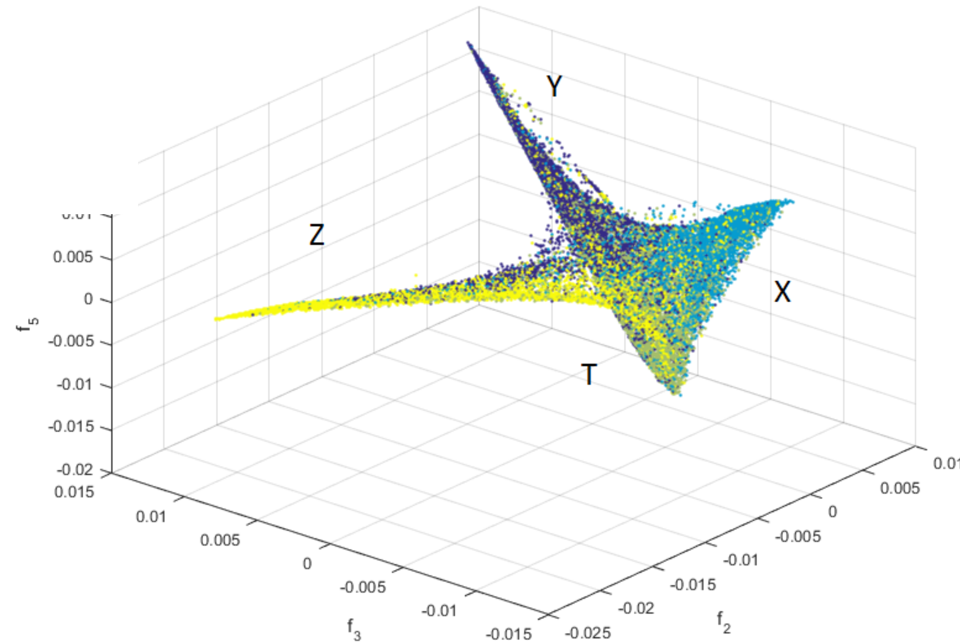
- Neat and compact images of the plants in the feature space
- So let us cluster in the feature space:

1. Spectral clustering - apply k-means clustering to Laplacian scores (regards machines!)
2. Create a define a pseudo-convolutional measure of similarity between plants (slightly imprecise in certain circumstances) and the apply spectral clustering on plants.

$$s(I_l, I_h) = \sum_{Y_i \in \phi_l} \sum_{Y_j \in \phi_h} e^{-\frac{\|Y_i - Y_j\|^2}{\sigma^2}}$$

Both methods collimate remarkably in the case of four groups!





Z-efficient + certain production profile
Y-efficient + different profile
X-inefficient
T- transition i.e. “at risk zone”

Back to reality

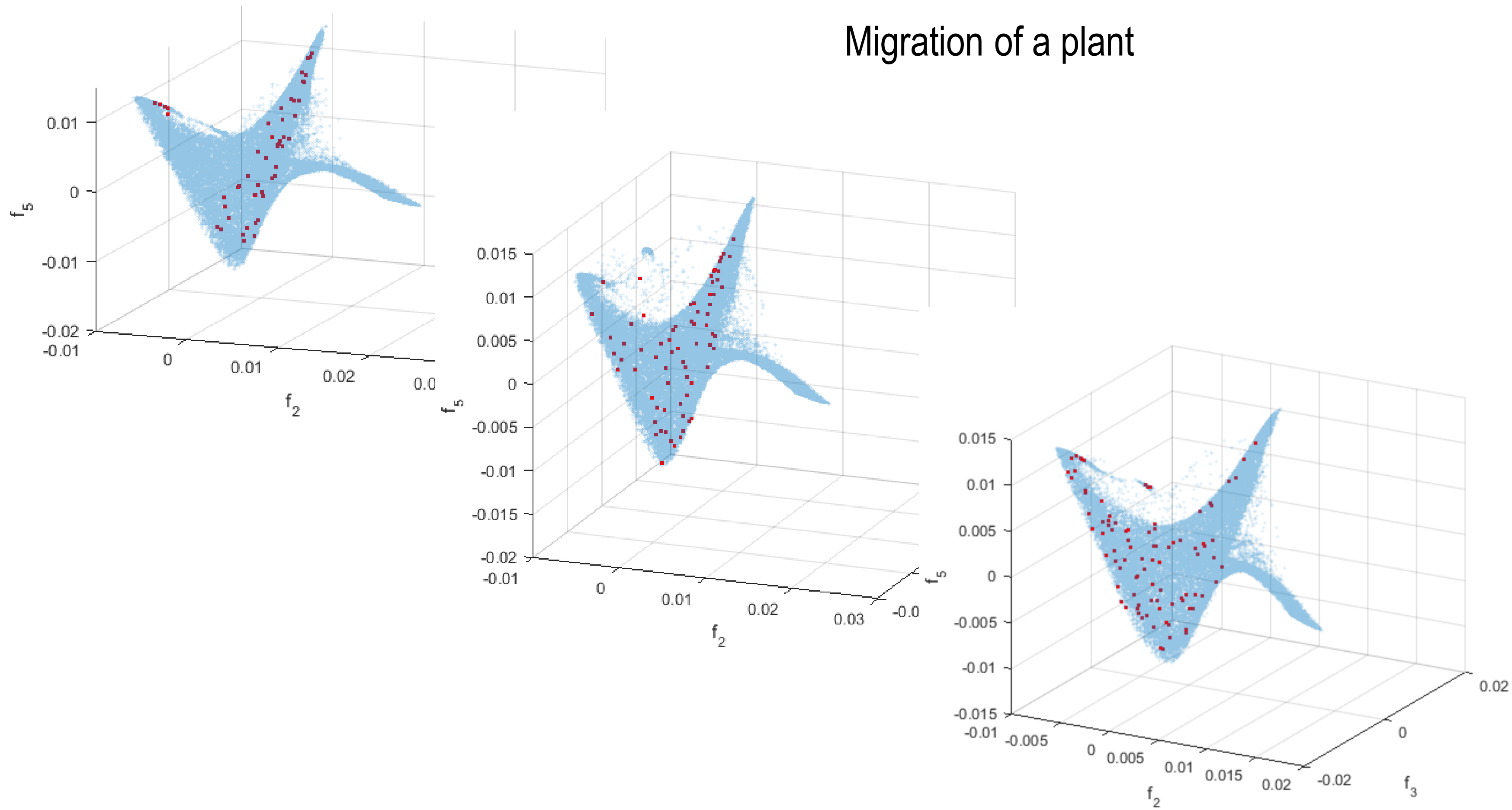
- Constructed spontaneous data-driven segmentation of the overall efficiency behaviour of industrial plants
- Highly non trivial i.e. transversal with respect to the original qualitative factors (interesting marginal distributions – indications for a typical plant)
- Plants are efficient in doing SPECIFIC things!
- Objective i.e. not accidental
- Series of objectivity tests on the segmentation done by Company's experts

 - Statistical checks + relations wit other data (commercial, customer sat. …)

 - Low "noice" 81% of plants remain stable in their segments over time

 - Remarkably! migrations can be attributed to objective causes. Our model detected concrete events as system upgrades, changes in the commercial profile, changes of the plant management (!!!)

Migration of a plant



Conclusions

- The overall efficiency behaviour of an industrial plant has **been modelled as an emergent phenomenon in a complex adaptive system**
- Modern manifold learning appears as good way to model emergent phenomena in complex adaptive systems
- Ambition 1: build a theory able to implement local and global “geometric” obstructions as “**complexity markers**”
- Ambition 2: segment plants which are inefficient in a specific way

Reserences

- A. Bellazzi, G. Jacazio, B. Maino, G. Mihaylov, F. Pellerey, M. Sorli, Integrated multivariate health monitoring system for helicopters main rotor drives: development and validation with in-service data. Proceedings of the Annual Conference of the Prognostics and Health Management Society ISBN: 978-1-936263-17-2, ISSN: 2325-0178 – 2014

- G. Jacazio, G. Mihaylov, F. Pellerey, Multivariate processing of accelerometric condition indicators, ScienceDirect IFAC-PapersOnLine, Volume 48, Issue 21, 2015, Pages 571-576, ISSN: 2405-8963, DOI: 10.1016/j.ifacol.2015.09.587

- G.Mihaylov, M. Spallanzani, Emergent behaviour in a system of industrial plants detected via manifold learning, submitted to ijPHM special issue on big data 2016

Gueorgui Mihaylov
Department of Mathematical Science
Politecnico di Torino