

Emergent behaviour in a system of industrial plants detected via manifold learning

Gueorgui Mihaylov¹, Matteo Spallanzani²

¹ *Politecnico di Torino, Department of Mathematical Sciences, Turin, 10129 TO, Italy*
gueorgui.mihaylov@polito.it

² *Tetra Pak Packaging Solutions, Modena, 41123 MO, Italy*
matteo.spallanzani@tetrapak.com

ABSTRACT

The efficiency behaviour of an industrial plant, part of a huge international structure of plants, is modelled as an emergent phenomenon in a complex adaptive system. The study is based on real in-service data obtained from an industrial production line monitoring system. Models of complex adaptive systems and some modern manifold learning methods are introduced in a unified formalism. The emergent behaviour is efficiently described in this setup.

1. INTRODUCTION

This article should be considered as part of a larger research programme, which aims to investigate the applications of modern manifold learning as an efficient tool for modelling emergent phenomena in complex (adaptive) systems. Some ideas in this direction have been presented during the last years by the first author in a series of seminars, conference talks and tutorials (ISI Foundation in Turin, Winter School of Geometry and Physics in Srni, PHME 2016 Conference in Bilbao, ICTD-CMMNO 2016 Congress in Gliwice). A strong boost for this programme has been determined by a research project developed jointly by the Polytechnic University of Turin and Tetra Pak, in which a relevant motivating example based on large amount of in-service industrial data appeared.

Emergent phenomena in complex systems are a relevant and broad topic with important implications in many spheres of the modern science. Apart from the more quantitative problems related to storage and processing of huge amounts of information, in our view one of the main big data related challenges is the theoretical explanation of those cases in which qualitatively new and often unexpected features emerge. The

leading paradigm for the comprehension of emergent phenomena ascribes them to the existence of opposite tendencies in the local dynamics of subsystems or single agents (frustration). Very simple one-to-one relations between agents in a system could give rise to complex global behaviour when the “synchronisation” of the states of the agents on large scale is impossible. In other words a complex behaviour in a system appears when local phenomena cannot be extended to a global scale.

Differential geometry, in particular the theory of geometric structures and gauge theories, offer specific tools able to detect local and global obstructions for certain point-wise constructions to be extended to local and global ones.

Manifold learning methods are constructed in a way to be consistent in the continuous limit. For this reason they naturally appear as suitable tools for treating huge data sets. The more general research programme, of which this work takes part, aims to develop modern manifold learning methods able to introduce the obstruction detection techniques of the theory of geometric structures in a more concrete computational context. In our view this would be a remarkable tool for analysing, modelling and quantifying complexity of systems. Some of these ideas are explained in Section 3 and Section 4 by using analogous formalism. Examples of standard constructions in differential geometry are introduced in order to make the paper more intuitive also for non experts.

2. PLMS DATA AND THE “FINGERPRINT” OF A PLANT

Packaging machines and other packaging line equipment produced by Tetra Pak are purchased and installed in several thousands of industrial plants all over the world. The productive contexts in which this equipment operates are extremely heterogeneous depending on the specific country, the company which owns the plant, the market, climatic, seasonal and many other factors.

Gueorgui Mihaylov et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

For many issues as maintenance services and contracts, productive efficiency, safety, equipment and material supply, the industrial plant is the relevant level of organization in this complex system of customers and producers (see for example Choo (2016) and O'Donovan (2016)). An accurate characterization (feature extraction and classification) of the specific “technical behaviour” of a plant is a very important task also from the viewpoint of large scale anomaly detection and prognostic applications (see for example Xiao (2016)).

An advanced Packaging Line Monitoring System (PLMS) is the Tetra Pak standard data management system for collecting, analysing and monitoring the operational performance of Tetra Pak filling machines and packaging lines. It is installed on huge part of the machines produced by Tetra Pak. It provides detailed information on both the health and the mechanical efficiency of the productive lines. Many technical parameters (pressures, gear and bearing state indicators etc.) are measured and collected continuously, so the total amount of in-service data collected worldwide from all the monitored machines is very big.

Part of this huge is condensed into a set of indicators defined and computed on different time scales in order to monitor the mechanical performances of each machine. In our study we mainly considered a set of ten mechanical performance indicators computed monthly for each packaging line. These indicators have been reputed by the Tetra Pak Technical Service as relevant and representative features for the overall mechanical efficiency. The first two mechanical performance indicators are simply the quantity of produced packages and the production time of the packaging line. Roughly speaking (the precise formulas cannot be publicly disclosed) the rest of the selected indicators take into account the stops, the emergency stops of the equipment in different failure modes, the restoring procedures of the equipment after an emergency or a maintenance stop etc. In order to make the different magnitudes of the indicators comparable from multivariate viewpoint, the data have been standardised. Several hundreds of thousands of in service vectors collected worldwide during 2013, 2014 and 2015 were available.

The set of PLMS vectors collected from a single plant represents all the packaging lines installed in that plant. This set appears spread in huge portions of the parameter space (see below on Fig. 1). For this reason standard hierarchic or k-means clustering procedures lead to rather poor results in the definition of efficiency behaviour classes or categories.

The PLMS record appears to be influenced by a series of characteristics of the plant and its production. The PLMS record of each plant depends (obviously) on the specific set of filling machine systems installed in that plant, the distribution within the plants production of several characteristics of the packages like volumes, types, shapes etc. Less obviously, detectable influence of geographic, climatic, “corporative” and

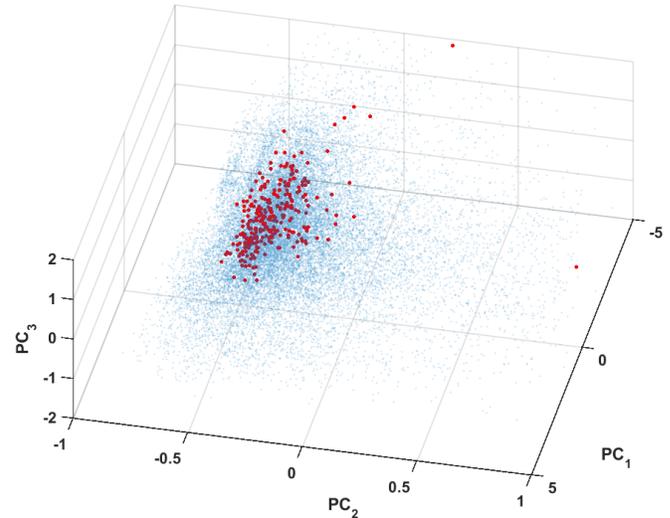


Figure 1. PCA plot of the PLMS record of an industrial plant collected in 2013 compared to the PLMS record of the whole calibration set.

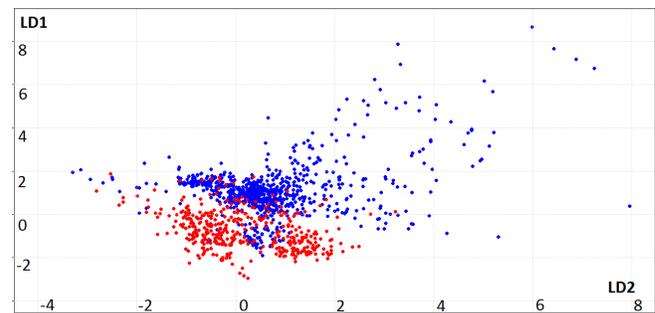


Figure 2. LDA on PLMS data collected in 18 different countries with equal levels of the other qualitative factors. As an example India is displayed in red and Japan in blue.

economic factors has been observed. The individual impact of all these qualitative criteria is easy to observe in homogeneous conditions i.e. by extracting subsets of data in which only one of the qualitative factors varies. Standard linear and quadratic multivariate statistical methods (Principal Component Analysis, Linear and Quadratic Discriminant etc.) highlight this fact (see Fig. 2 and Fig. 3).

An elementary “filtering” procedure allows to subdivide the PLMS record of a plant (or a set of plants) into subsets which are measured for equal levels of all the selected qualitative factors. We call these subsets the *homogeneous groups* of a plant (or a bigger calibration set of PLMS vectors). The number of homogeneous groups of a single plant is quite variable.

Specific multivariate normality tests (Mardia’s test, Henze - Zirkler’s test, Royston’s test implemented in R) show that the PLMS measurement vectors in a single homogeneous group is close to be normal with exception of less than 20% of outliers. The interpretation of this fact is that the splitting

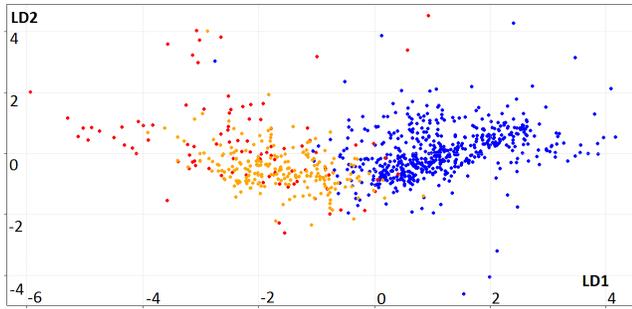


Figure 3. LDA on PLMS data with equal levels of all qualitative factors except the package volume. Red and orange dots represent two different single portion packages, blue dots - a family portion package.

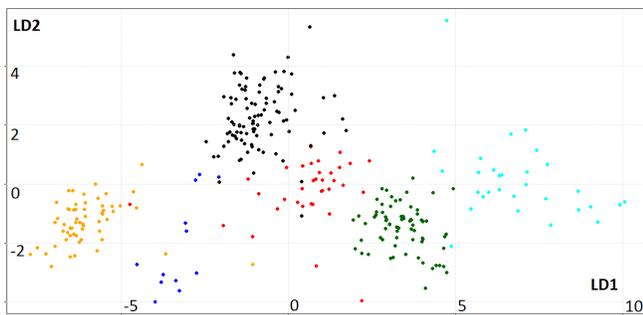


Figure 4. PLMS fingerprint of an industrial plant (the same as in Fig. 1) collected during 2013. Different colours represent different homogeneous groups in the PLMS record.

procedure eliminates part of the deterministic impact of the qualitative criteria and the remaining variations can be better modelled over random white noise. The number of remaining outliers could be further reduced by introducing additional qualitative criteria in the splitting procedure.

The set of the homogeneous groups in the PLMS data of a plant is called the “fingerprint of the plant”. On Fig. 4 is displayed a typical fingerprint of an industrial plant obtained from its PLMS record over 3 years. The distributions of each homogeneous group have been characterized by the estimates of their first and second momenta (the mean vector and the directions of the principle components of the group).

A practical classifier based on this information has been specifically implemented and validated. It has been first applied on (inspired by) a set of pre-existing segmentations of the plants based on commercial criteria and empirical observations of Tetra Pak Technical Service (plants owned by different companies, classes of plants with different commercial behaviour, plants with different types of service contracts etc.). An indexed calibration set is used for computing the fingerprints of each class. A plant is assigned to the class for which its fingerprint matches better with the calibrated benchmark, the class

which minimises (as first criterion) the following expression:

$$\frac{1}{N} \sum_j^N \sum_i^{n_j} \frac{1}{n_j} \frac{d_j - d_{ji}}{\sigma_j}$$

where N is the number of matching homogeneous groups in the calibration set of a class, d_j denotes the mean Mahalanobis distance (MD) from the mean vector of the j -th homogeneous group in the calibration set, d_{ji} denotes the MD between the mean vector of the j -th homogeneous group to the i -th element of the j -homogeneous group in the test set, σ_j is the standard deviation from the mean MD in the j -th calibration homogeneous group, and n_j is the number of elements in the j -th homogeneous group in the test set. Since there are 10 degrees of freedom, normal approximation and the corresponding sample estimates of the Fisher-Snedecor statistics can be exploited.

As a second subordinated criterion the plant is assigned to the class which maximises the cosines of the angles defined by the first four principal components of each homogeneous group computed in the calibration and in the test set.

A calibration-validation procedure was tested on a pre-existing internal empirical plant classifications by using 200 individuals as a calibration and 64 individuals as a test set and 73% of the test plants were correctly classified. For obvious reasons the “multiple fingerprint” classifier cannot be applied when none of the homogeneous groups of the plant is contained in the calibration set. The huge amount of available in-service data made possible avoiding such cases. Situations of partial covering of the homogeneous groups has been analysed and the multiple fingerprint classifier performs still very well.

The statistical aspects of the “multiple fingerprint” method, which was developed in this context, appear interesting on their own right. A separated paper will be devoted to their detailed description. Our aim in this paper is to point out that the objectivity and the consistency of our experimental results obtained via manifold learning techniques have been tested by an independently calibrated and validated method.

Any segmentation of the set of plants corresponds to a stratification with respect a number of a priori commercial or empirical criteria. Differences between the calibration fingerprints of a priori separated classes highlight some “interaction” between the qualitative factors which define the homogeneous groups and the new segmentation criteria. The application of the classifier to the existing empirical plant segmentations revealed some interesting and unexpected phenomena. The calibration fingerprints of the segments differ not only by slight deviations of the common homogeneous groups, but quite often more radical large scale modifications are observed. Our interpretation of these facts is that the large scale phenomena are provoked by some complex interaction of the qualitative factors and the segmentation criteria.

3. COMPLEX ADAPTIVE SYSTEMS

Non local or large scale phenomena which can be hardly modelled or explained by taking into account only the local degrees of freedom of a system or local interactions of its subsystems are called emergent. Large scale fluctuations, shock wave propagation, phase transitions, complex collective molecular motions etc. are examples for such somehow unexpected behaviour.

A system is a structure defined by the set of its elements (agents) together with the relations (interactions) among them. In the rest of the paper we will consider single filling machines as agents of bigger systems as plants, or even systems of plants. As we will see, the behaviour of a plant can be efficiently described if we take into account not just the values of the PLMS parameters of its productive lines, but if we also try to model an interaction between the machines.

The leading paradigm in the comprehension of emergent phenomena in a complex system asserts that they are generated by opposite tendencies in the local dynamical behaviour of its subsystems. This is usually called *frustration* and makes impossible the “synchronisation” of the states (or equivalently interpretation of the states) of subsystems on large scale in a complex system. Spin glasses, network fitness models (see Bianconi & Barabasi (2001)) etc. are important examples in which extremely simple one-to-one relations between the agents of a system give rise to a complex behaviour on large scale. In fact the interaction and the equilibrium configuration of a system of two magnetic dipoles is rather simple, but in a spin glass locally conflicting interactions make the definition of a global equilibrium state very difficult. Highly non trivial large scale fluctuations, phase transitions etc. arise.

In other words emergent phenomena appear when some local phenomena cannot be extended to global ones.

Neural networks provide excellent tools able to mimic complex behaviour, but less efficient for modelling and explaining its more specific features.

There are many classical examples of dynamical systems, whose normal modes are actually global degrees of freedom. The D’Alamberts equation, which rules the dynamics of a vibrating string, can be deduced by modelling the string as a system of coupled oscillators, but its dynamics is efficiently described in terms of global harmonic components i.e. the complete set of eigenfunctions of the space part of the equation. Another interesting example is the phonon i.e. the quantization of the vibrational degrees of freedom in quantum elastic systems. In these cases conservation laws and material constrains induce local “contradictions” between coupled parts of the system. More generally non-local extremisation procedures in “contradicting” conditions are expressed non-local degrees of freedom, a “symptom of complexity” (see below the discussion on harmonic mappings and diffusion).

An efficient approach (adopted for example by the network fitness model) consists of describing each agent in a system by two sets of variables, its position (usually a set of deterministic “spatial” parameters) and its internal state (often a random vector variable extracted from ideally continuous population). Extremely general class of systems can be modelled in this setup (systems of airports, arbitrage stock markets etc.). A state of a system can be identified in this case with a vector valued function (field) defined on the space of deterministic parameters which is sampled in a finite number of points.

We generalise this viewpoint to the case in which the “spatial” variables parameterize a differentiable manifolds.

A fibre bundle over a smooth manifold is itself a differentiable manifold E (total space) which is only locally isomorphic to the product of the base manifold M with some fibre space F . More precisely there exists a smooth surjection $\pi : E \rightarrow M$ which satisfies the following local triviality condition. Given a point $X \in E$, there exists an open neighborhood $U \subset M$ of $\pi(X)$ such that $\pi^{-1}(U)$ is diffeomorphic to $U \times F$.

Example 1. The tangent bundle over a smooth manifold is a vector bundle i.e. a fibre bundle in which F is a vector space (in this case $F = T_P M$, the tangent space in a point P of M). For the frame bundle LM over a differentiable manifold the fibre over $P \in M$ is the set of the linear bases of $T_P M$. One linear basis is transformed into another one by means of the action of the Lie group $Gl(\mathbb{R}, n)$ on $T_P M$. The frame bundle is an example of a principal bundle (the fibre space is isomorphic to a Lie group) and the tangent bundle is an example of a vector bundle associated to LM by means of the $Gl(\mathbb{R}, n)$ group action on the vector fibre space.

Sections of fibre bundles generalise the notion of scalar / vector / tensor-valued functions on manifolds and their transformations under basis changes. The analogy is not complete as global sections of a fibre bundle with specific characteristics could even not exist on a given manifold. In fact a principal bundle which is globally isomorphic to the product $M \times F$ is called trivial. Non trivial principal bundles do not admit global sections (as an example consider the “orientation” \mathbb{Z}^2 -principal bundle over the Möbius strip). Associated bundles in general admit global continuous sections, but other requirements (everywhere smooth, non-vanishing, non-singular) can be impossible to satisfy ¹.

Fibre bundles are an appropriate tool for measuring the obstructions to extend point-wise geometric constructions to local or even global structures.

There is plenty of examples for such phenomena coming from the theory of Geometric structures on differentiable manifolds. The integrability of Riemannian, complex, symplectic,

¹We invite the reader to construct a self parallel everywhere non-vanishing vector field on a 2-sphere...

spin etc. structures measures the extent to which constructions defined on each tangent space can be defined on open subsets, or at least to coordinate chart domains on the manifold (existence of a set of holonomic i.e. compatible coordinates). 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. obstructions represented by sections of tensor bundles (see Sternberg (1963)).

Example 2. A complex structure (an endomorphism J such that $J^2 = -Id$) can be always defined on the tangent space at a point of an even-dimensional manifold M . The Newlander-Nirenberg Theorems state that complex holomorphic coordinate charts on M exist if and only if a local obstruction on M , called the Nijenhuis tensor, vanishes (see for example Joyce (2000)). This obstruction is the intrinsic torsion (or first structure function) of what is called an almost complex structure over the manifold. The celebrated Darboux Theorem states an analogous result for almost symplectic structures.

Example 3. There is no a priori assigned way for identifying elements in the fibres of a bundle. The identification of elements which belong to different fibres (for example two tangent vectors) is itself a non-local procedure. The parallel transport is a map from one fibre into another which is associated to what is called an affine connection on principal bundles (with a corresponding covariant derivative on associated ones). A connection field is by definition a gauge field by means of which sections of the associated bundles interact. Parallel transport on smooth manifolds is usually path-dependent. According to the Ambrose-Singer Theorem the local obstruction for the triviality of the parallel transport, associated to a nontrivial holonomy transformation, is a tensor called curvature (see for example Joyce (2000)).

Global properties of fibre bundles strongly involve topological features of the manifold. These relations are captured by the theory of characteristic classes which associates to each principal bundle on M a set of cohomological classes of M . Characteristic classes are global invariants which measure how far is the global structure of the bundle from being a trivial one. Well known characteristic classes as Chern, Bott-Chern, Stiefel-Whitney etc. detect global obstructions for the extendibility of specific local phenomena (orientability, existence of a spin structure on the tangent bundle etc.). Practically the vanishing of the obstructions is often related to the vanishing of the integral of a special differential form on M .

The "internal state model" can be generalised by saying that the configuration space of possible states of a system is modelled over a vector (tensor) bundle over a smooth manifold of "spatial" parameters. More precisely we interpret a state of a system as a section of a vector bundle sampled in a finite number of points (its agents). We call a system *adaptive*

when the internal state of each agent is not merely random, but is determined by some interaction with other agents (a spin glass is an example).

This "geometrised" setup becomes interesting when the manifold of the deterministic parameters is not geometrically trivial. Non trivial bundles can be adopted for the description of the internal states. Introducing the stronger assumption that the state of the system is a sampling of a locally smooth section over "spatial" manifold corresponds to introducing a strong "interaction" between the two types of parameters. Interesting non local effects arise within this "geometrised" setup. The very identification of the internal states of the agents must be done by means of a parallel transport. A "signal" must be transmitted along the manifold and this is a non-local procedure.

In our view a consistent construction of affine connections, curvature and characteristic classes are essential tools in the elaboration of a gauge theory on emergent phenomena in a complex adaptive system. The "synchronisation" of the internal states of the agents of a system becomes a path-dependent problem Ilinski (2001), Mack (2000) and different paths could give rise to conflicting outputs. The geometric obstructions (torsion, curvature, non trivial characteristic classes) for the extension of local phenomena introduce and model complexity in the system. This is the main focus of our larger research project mentioned in the introduction to this paper.

4. SOME IDEAS ON MODERN MANIFOLD LEARNING

Manifold learning methods had their origin from the idea that multivariate data are distributed on (or nearby) a differential submanifold M immersed in the space of measure variables \mathbb{R}^n . There are two reasons for which manifold learning seems to be a suitable tool for treating big sets of multidimensional data. It is constructed in a way to capture some geometric structure valid in the limit of continuous sampling of points on M . If one manages to characterise consistently the underlying geometric structure of a data set, it is possible to regress and interpolate missing or out of sample data in a particularly efficient way. One can actually reduce the quantity of data to handle. In Singer & Huettenlocher (2012) is described a rather general procedure of regressing and interpolating fields in out of sample points. A relevant point discussed in Shawe-Taylor et al. (2005) is that once captured the geometric structure of the data cloud, the error on out of sample points is rather small.

The study of the properties of differentiable operators on manifolds is a broad area in the modern (geometrical) mathematical analysis. Differential operators act on sections of fibre bundles. This approach revealed itself as particularly powerful and produced relevant and far reaching implications (for example the celebrated Atiyah-Singer Theorem). Another remarkable example for application of analytical techniques

for describing specific geometric properties is given by the proof of the Calabi conjecture given by Yau. The existence and uniqueness of a metric with certain reduced holonomy is proved by solving a partial differential equation of Monge-Ampere type (see for example Joyce (2000)).

The study of the spectral properties of operators concentrates the analysis onto a reduced set of “special” sections. Close relations between the spectral properties of differential operators and geometry have been analysed. There are many examples of such constructions as harmonic sections, spectral analysis of the Dirac operators on differentiable manifolds (just as examples we cite Atiyah (1973), Esposito (1998), I. Agricola & Friedrich (1999), Jardim & Leao (2008), , Friedrich (2012), F. T. Agricola I. & Kassuba (2008) etc.). Spectral analysis gives both local and non-local information on manifolds.

We re-interpret the manifold learning problem in view of the following more general “geometrization paradigm”:

Definition 1. A feature map is a locally smooth section Y of a vector bundle E over a smooth manifold M sampled in a finite number of points $X_i \in M$.

NB. The value of Y at each point of the sampling is intended expressed with respect to a local basis assigned on the fibre. The existence of a discrete sampling X_i in an ambient space \mathbb{R}^n and sampled sections do not imply the triviality of any bundle.

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. This is usually done by well known method of local PCA (see Singer & Hau-tieng (2012)). The tangent bundle associated to the frame bundle can be constructed Singer & Hau-Tieng (2013).

Furthermore, in view of Nash’s C^1 isometric embedding theorems (Nash (1954)) under quite general assumptions M inherits a Riemannian metric from the Euclidian space \mathbb{R}^n .

By definition sections of vector bundles represent “excitations” of non-local degrees of freedom on M . Relevant sections of vector bundles can be selected by minimising functionals $F(Y)$ on M or by analysing the spectral structure of differential operators \mathcal{L} (if the differential operator admits a formal adjoint the two conditions coincide).

Many of 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 $(X_i, Y(X_i)) := (X_i, Y_i)$ of a vector bundle section, functionals are defined by finite summations on that finite set.

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

The role of the “affinity function” W_{ij} is to localize i.e. to

restrict (or not) the contribution to the evaluation of the functional to some neighborhood of 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 an edge between each couple of points for which $d(X_i, X_j) \leq \epsilon$ (where d denotes the Euclidian distance) or by connecting to X_i the k nearest points etc. The first case is geometrically intuitive, but could produce graphs with more than one connected components, the second choice provides one connected component.

The function L usually contains some relevant characteristic of the section which need to be preserved. We list below some popular examples of this type:

- In Curvilinear Component Analysis (see for example Demartines & Hérault (1997)) we have:

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

where d denotes the Euclidian distance both in the space of measured variables and in the fibre. In similar setting Curvilinear Distance Analysis can be introduced by replacing the Euclidian distance in the space of variables by a distance along a neighbour graph.

- In Multidimensional Scaling Method (see for example Borg & Groenen (1997); Bronstein & Kimmel (2006)) we have:

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

where f is a function which defines the model.

- ISOMAP is multidimensional scaling method based on the graph geodesic distance matrix (see deSilva & Tanenbaum (2003); Tanenbaum & Langford (2000)).

- Laplacian eigenmaps are defined by minimising some energy functional (see below).

Whether or not the continuous limit of the above expressions represents meaningful functionals

$$F(Y) = \int_M \mathcal{L}(X, Y(X)) dX$$

on sections on M is a rather delicate analytic question. The research of a hypothetical general procedure which leads to the deduction of the above empiric functionals from continuous analogues (by replacing differences by derivations and other discretisation tricks) goes beyond the aims of this paper.

In general if the expression of “characteristic” on the fibres is preserved by the action of some group G , features can be naturally considered as sections of a vector bundle associated to a principal G -bundle.

Fredholm’s theory suggests a natural way of implementing

discrete forms of differential operators acting on finitely sampled sections. Consider a linear differential operator acting on sections of vector bundles $\mathcal{D} : \Gamma E \rightarrow \Gamma F$ via:

$$\mathcal{D}\phi = \psi \quad (1)$$

The Green's function $G(X, Y)$ associated to a differential operator is a special integral kernel related to the Dirac delta function. More precisely (written component-wise) we have:

$$\mathcal{D}G_k(X, Y) = v_k \delta(X - Y),$$

where v_k is an arbitrary constant unit vector. For example the Green's functions associated to the Maxwell's equations in relativistic and non relativistic formalism, Green's functions associated to Dirac operator on spin bundles have been studied. The above equation can be written in compact tensorial form by setting:

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

where \mathcal{D} acts on the columns of \mathbf{G} and \mathbf{I} denotes the unit dyad. In these terms Equation (1) can be re-written in integral form:

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

In particular for differential operators $\Gamma E \rightarrow \Gamma E$ with discrete spectrum λ_i the Green's function can be expressed in terms of its eigenfunctions ϕ_i :

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

Definition 2. Consider E and F , vector bundles on a smooth manifold M and a linear differential operator

$$\mathcal{D} : \Gamma E \rightarrow \Gamma F.$$

Denote by \mathbf{G} the Green's function of \mathcal{D} . Given a finite sampling $X_i \in M$ the coarse discretisation D of \mathcal{D} is a linear mapping defined by:

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

where δ_{ij} denotes the Kronecker's symbol.

Let $Y(X_i)$ and $Z(X_i)$ denote features sampled in N points of M . The transformation $DY(X_i) = Z(X_i)$ induced by the linear mapping D is obtained by inverting the relation:

$$Y(X_i) = \sum_{j=1}^N \mathbf{G}(X_i, X_j)Z(X_j). \quad (2)$$

The contribution to the value of $Y(X_i)$ can be localised by means of an affinity function W_{ij} . In these terms Equation (2)

becomes.

$$Y(X_i) = \sum_j \mathbf{G}(X_i, X_j)Z(X_j)W_{ij}.$$

The affinity function can be absorbed in the definition of \mathbf{G} .

If the fibres of E and F are isomorphic to \mathbb{R}^n and \mathbb{R}^m then D is element of $\mathbb{R}^N \otimes \mathbb{R}^N \otimes \mathbb{R}^m \otimes \mathbb{R}^n$. A feature can be represented by Nn -dimensional vector (a set of N n -dimensional vectors) and the action of D implemented by $Nm \times Nn$ matrix A with $N \times N$ blocks of dimension $m \times n$. Each $m \times n$ block represents a linear map $\mathbb{R}^n \rightarrow \mathbb{R}^m$ which is "weighted" by a scalar factor. The affinity function W_{ij} is constant on each block and puts some blocks equal to zero.

The spectrum of (a "square") D is the set of eigenvalues λ_i of a square matrix A . If A is symmetric, (this means that $\mathbf{G}(X_i, X_j) = \mathbf{G}(X_j, X_i)$ and $W_{ij} = W_{ji}$), it is easy to prove that the above definitions lead to:

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

The described discretisation of given differential operator is rather straightforward procedure. The definition of the coarse discretisation involves a finite summation, so analytical troubles are easily avoided. A typical manifold learning problem arises when one wishes to recover the expression of specific linear differential operator on (a priori unknown) manifold from a point cloud. We call such a linear map the *sample discretisation of a differential operator*. The problem can be formulated as: "Does a sample discretisation coincide with a coarse discretisation of the differential operator".

The ingredients for the construction of a sample discretisation are an affinity graph and a kernel function $K(X_i, X_j)$ which must satisfy certain symmetry and regularity assumptions (see Singer & Hau-Tieng (2013); Berry & Sauer (2015)). The kernel function $K(X_i, X_j)$ in the manifold learning setup aims to reproduce the expression of the (unknown) Green's function of the manifold in terms of the coordinates of the points X_i in the space of measured variables. A first approximation attempt regarding the expression of the kernel can be the (often known) Green's function of the same operator on \mathbb{R}^n computed with the embedding coordinates X_j . Such an estimate must be corrected by bias terms which involve the curvature on M , other potential-like terms, finally terms which take into account the variance of the sample etc.

Formally the explicit form of the operator (and the true corresponding Green's function) involves sections of several vector bundles, the tangent bundle TM , the Riemannian metric on M is itself a section of the vector bundle $S(TM^* \otimes TM^*)$ (where S denotes the symmetrisation of the tensor product). From our viewpoint this means that the very definition of the discrete Hessian, the connection Laplacian, the Laplace-

Beltrami operators is based on several implicit feature extractions. This point actually gives considerable flexibility to the construction.

In Berry & Sauer (2015) is provided an interesting result in this direction i.e. “tracing” the influence of the implicit features extraction procedures. The “intrinsic geometry” of so called local kernels is described both in case of uniform and nonuniform sampling. The expression of the Laplace-Beltrami operator changes if the Riemannian metric on M (one of the implicitly extracted features) changes in a family of metrics which includes the one induced by the embedding $M \hookrightarrow \mathbb{R}^n$. This changes are captured by modifications in the form of the Green’s function of the operator and subsequently in its discretization. Every symmetric local kernel with exponential decay corresponds to a Laplacian operator in a Riemannian geometry and conversely any Riemannian geometry can be represented with an appropriate local kernel. In the same paper are introduced other relevant geometric applications of that technique like data driven geometry regularization via local kernels, conformally invariant embeddings etc.

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

- point-wise convergence

$$DY(X_i) = Z(X_i) \longrightarrow \mathcal{D}\phi = \psi;$$

- spectral convergence means that eigenvectors of D converge to eigenfunctions (sections) \mathcal{D} on M .

Spectral convergence is a stronger condition than point-wise convergence. Establishing the convergence properties of a discretized operator is a highly non-trivial analytical problem. It depends on the manifold M and its boundary, on the density of sampling, normalization procedures, opportune parameters choices (absorbing, redefining overall constant factors, volumes) etc. Remarkable results in this direction have been achieved in Belkin & Niyogi (2005); Singer & Hau-Tieng (2013) (see below).

Our viewpoint on manifold learning can be easily applied to a class of techniques called Kernel methods. In these cases primary feature extraction remains implicit. An interesting example is the kernel Principle Component Analysis. Consider a feature extraction $Y(X_i)$. The sample covariance matrix (a new feature) matrix determined by the section is:

$$C = \frac{1}{N} \sum_i Y(X_i) \otimes Y(X_i) \quad (3)$$

A well-known fact is that the condition

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

for some given kernel function is sufficient in order to com-

pute the components of the feature vectors with respect to the basis of eigenvectors of the covariance matrix C . In other words this information is sufficient in order to construct a global basis change on the fibres of the bundle determined by the orthogonal matrix which diagonalises C . (Taking the first two or three components of $Y(X_i)$ with respect to the orthonormal eigenvector basis provides a non-linear dimensional reduction technique.)

In our view this type of approach appears as less descriptive with regards to emergent phenomena in complex systems. In fact no functional extremisation is involved, no contrasting tendencies appear, no action of a differential operator. In other words no “interaction” of sections appear, just a global passive, so called gauge transformation induced by 3.

Observe that many of the geometric obstructions mentioned in the previous section are represented by sections of tensor bundles i.e. features. Finite (integral) estimates of these features can be obtained in discrete context. For a manifold learning implementation of parallel transport see Zhang & H. Zha (2004) and Singer & Hau-tieng (2012). Also components of the Riemannian curvature tensor can be estimated in manifold learning context, the Ricci curvature captures the deviations in the volume of a geodesic ball (see for example Ache & Warren (2014)). The Weyl tensor is the obstruction for conformal flatness etc.

We call these type of features *high level features*. We expect that they could play the role of *complexity markers*.

5. GEOMETRY, HARMONIC MAPPINGS AND DIFFUSION

Sample discretisations of several relevant differential operators have been defined in a series of publications. The spectral analysis of these discrete realizations provides a set of eigenfunctions which play the role of global degrees of freedom in a non linear dimensional reduction in a space of non-local features (see for example Donoho & Grimes (2003)).

The rough Laplacian operator acts on tensor fields on a Riemannian (or pseudo-Riemannian) manifold M and is defined as the trace of the second covariant derivative of the field associated to the Levi-Civita connection:

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

The rough Laplacian acting on \mathbb{R} -valued functions on M is known as the Laplace-Beltrami operator. It can be interpreted as the divergence of the gradient of a function (the trace of the Hessian of a real valued function). The spectral properties of Δ have been largely exploited in geometrical analysis.

The Laplacian operator is of great interest in relation to the discussion developed in Section 3 for two main reasons. First of all the solutions of the Laplace’s equation involve non local extremisation. Then, it generates diffusion processes by

means of the heat equation. A diffusion process on a vector field involves non local “synchronization” process.

The connection Laplacian operator is very important in differential geometry (we refer the reader to Rosenberg (1997)). It is directly used in spectral shape analysis, in the theory of harmonic functions and harmonic vector fields on Riemannian manifolds, which provide specific geometric information. Harmonic morphisms are maps which preserve Laplace’s equation. More generally the theory of harmonic mappings between Riemannian manifolds, harmonic sections of fibre bundles leads to many far reaching implications (see for example Eells & Sampson (1964); Baird & Wood (2003)). It is interesting to recall that a smooth map $f : M \rightarrow N$ between Riemannian manifolds is called a harmonic mapping if it extremises the energy functional:

$$E(f) = \int_M |df|^2 d\omega$$

where $d\omega$ denotes the volume form. In particular every harmonic morphism between Riemannian manifolds is a harmonic map.

The deep geometric meaning of diffusion processes has been highlighted also by an alternative proof of the Atiyah-Singer index Theorem based on the analysis of the heat equation (see for example Gilkey (1995)).

The construction of the so called Vector Diffusion Maps leads to a consistent sample discretisation of the connection Laplacian on different vector bundles which converges in both pointwise and spectral sense. In Singer & Hau-tieng (2012) is shown how the rough Laplacian on the tangent bundle of the manifold can be approximated from random samples. In particular it generates a vector diffusion process and its eigenvectors generate an Euclidean space in which the data sample can be embedded. The Euclidean distance this space induces a metric on the sampled manifold which is called the vector diffusion distance. This distance approximates, via short time asymptotic expansion of the heat kernel, the geodesic distance between nearby points on the manifold.

In Singer & Hau-Tieng (2013) connection Laplacians on other vector bundles are discretised. The introduction of a principal bundle construction from a sampled set of data provides good theoretical setup in which the Laplacian eigenmaps method, the diffusion maps method, the vector diffusion maps method, and the orientable diffusion maps (see Singer & Hau-Tieng (2011)) can be successfully unified. These models are introduced in terms of sections of suitable different associated bundles.

In this construction the space \mathbb{R}^N in which the sample X_i is collected is considered as the total space of a fibre bundle in which M is the base manifold and the fibres represent the overabundant nuisance parameters. A group G of linear

transformations with simple transitive action on the space of some nuisance parameters can be considered as a structure group of a principal bundle in which M parameterises the set of orbits.

For example the principal bundle exploited in the construction of orientable diffusion maps is the \mathbb{Z}_2 orientation bundle on the tangent bundle of M .

A natural notion of parallel transport is introduced in discrete setup. Special local sections of the principal bundle can be defined by means of the optimal alignment procedure with fixed radius (we refer the reader to Zhang & H. Zha (2004) and Singer & Hau-tieng (2012)). The products of the linear transformations associated to a sequences of vertices (discrete paths) generates a translational group i.e. a consistent notion of parallel transport on each associated bundle. In this framework, the affinity between two agents is defined by considering the parallel transport along all the paths of certain length which connect them. The “synchronisation” of the internal states of X_i and X_j is obtained by weighting and summing the transformations computed along all these paths. A path dependent “synchronisation” which “contrasting outputs” is very important from the viewpoint of the complexity modelling.

Observe that the Euclidean metric in \mathbb{R}^N takes into account the nuisance parameters. A metric on M which defines distances between orbits in the total space must be invariant under the action of G on the total space. Invariant metric can be defined also by optimal alignment (see Singer & Hau-Tieng (2013)).

The above construction fits perfectly with the general discussion developed in Section 5. In Singer & Hau-Tieng (2013) a sample estimate of a differential operator is built under the following assumptions.

- The base manifold M is smooth and smoothly embedded in \mathbb{R}^N , a metric is induced by the canonical Euclidean metric, ∂M 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 $K(|X|) \in C^2(\mathbb{R}^+)$ is a strictly positive function etc.

Consider an undirected affinity graph defined by W_{ij} and set

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

where g_{ij} is an orthogonal transformation on the vector state defined by summing the orthogonal transformations associated to all paths of certain length connecting X_i to X_j .

The diffusion map method evaluates given kernel $K(X_i, X_j)$ on all pairs from the data set, but then different types of normalisation are applied.

The first normalization divides the columns of the matrix associated to the operator by the column sums, to some power. This normalization is equivalent to multiplying the matrix on the right by a diagonal matrix with entries equal to a power of the sampling probability p . The aim of this procedure is to remove the bias introduced by the fact that the differential operator is estimated from a data set evaluated according to some probability density.

The second normalization takes the right-normalized matrix and divides the rows by the row sums. In this way the matrix which implements the differential operator becomes a row-stochastic matrix (or Markovian) matrix. In the continuous limit this type of normalization eliminates the curvature dependent term in the expansion of the kernel. Consider the $Nd \times Nd$ block diagonal matrix

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

With these assumptions (up to some constant factors not quite relevant in this general discussion) have been proved the pointwise convergence in the continuous limit:

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

Furthermore a spectral convergence Theorem in Singer & Haultieng (2013) states that the eigenvectors $Y(X_i)$ of $\mathbf{D}^{-1} \mathbf{P} - \mathbf{I}$ are discrete approximations of the eigenvector fields of the rough Laplacian operator on M with homogeneous Newman boundary conditions which satisfy:

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

In other words the above sample discretisations are proved to be coarse discretisations of the connection Laplacian operator. Observe that a different convention on the adopted normalisation leads to a different embedding in the space of eigenfunctions.

If the above general construction is applied to \mathbb{R} -valued functions, we recover the diffusion map and the standard Laplacian eigenmaps models. It is based on a principal bundle with fibre $G = e$ and fibre \mathbb{R} . In other words we have a trivial total parallelism on the associated bundle. The discretisation of the operator is implemented by $N \times N$ matrices with scalar entries and the normalised graph Laplacian is again $L = \mathbf{D}^{-1} \mathbf{P} - \mathbf{I}$.

Both the harmonic and the diffusion aspects are immediate in this case. The sample discretisation of the Laplace-Beltrami operator on a Riemannian manifold is obtained by minimising a discrete graph version (see Belkin & Niyogi (2003))

$$E(f) = \sum_{i,j=1}^N (Y(X_i) - Y(X_j))^2 W_{ij}, \quad Y^T D Y = 1 \quad (4)$$

of the energy functional.

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

In fact (see Belkin & Niyogi (2003)) for a C^2 \mathbb{R} -valued function f on $M \subseteq \mathbb{R}^n$ smooth and compact Riemannian manifold occurs that

$$|f(x) - f(y)| \leq \|\nabla f(x)\| d_M(x, y) + o(d_M(x, y))$$

where d_M is the distance inherited from \mathbb{R}^n and $\|\cdot\|$ is the L^2 norm on M .

We remark again that the minimisation is a global ‘‘dynamical’’ process which combines contrasting local tendencies. Since $\mathcal{L} = -\Delta$ and ∇ are formally adjoint operators, according to the Courant-Fischer-Weyl minimax Theorem the energy expression is minimized by the eigenfunctions of \mathcal{L} corresponding to its lowest eigenvalues. In fact Stokes’ theorem allows to write

$$\int_M \langle \nabla f(x), \nabla f(x) \rangle dx = \int_M f(x) \mathcal{L}(f)(x) dx,$$

which corresponds to:

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

By construction L is thus a positive semidefinite matrix. Standard numerical analysis considerations (see Sameh & Wisniewski (1982)) imply that the minimization is achieved by the eigenvector of L corresponding to its lowest eigenvalue. The trivial solution $Y(X_i) \equiv 1$ corresponding to $\lambda = 0$ is ignored within the minimization procedure. The dimensional reduction is performed by the lowest accomplished by the L non-constant solutions of:

$$L f = \lambda D f$$

An immediate way to implement the ‘‘diffusion process’’ viewpoint in this case consists of sampling the Green’s function H_t of the Laplace operator on points $X_i \in \mathbb{R}^n$. In these terms the generic solution u to the Laplace equation with initial condition f ,

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

is written as:

$$f(x, t) = \int_M H_t(x, y)u(y)dy$$

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

where $\phi \in \mathcal{C}^\infty(M \times M)$ has the property $\phi(x, x) = 1$, so that when $\|x - y\| \rightarrow 0$ and $t \rightarrow 0$, H_t can be approximated by a Gaussian function

$$H_t(X_i, X_j) \approx (4\pi t)^{-\frac{\dim M}{2}} e^{-\frac{|X_i - X_j|^2}{4t}}$$

In Belkin & Niyogi (2003), Coifman & Lafon (2006) etc is proved that the sample discretisation constructed by means of the graph Laplacian with the above choice of kernel (wh oportune normalisation, constant definitions etc.) converges to the Laplace-Beltrami operator both point-wise and in spectral sense for an uniform and not uniform sampling.

6. EFFICIENCY BEHAVIOUR SEGMENTATION

In the context of the considerations presented in the previous sections, manifold learning methods have been applied to the available PLMS record of industrial plants. The calibration set contains the PLMS record for 2013, 2014 and 2015 of 771 plants installed worldwide which contain 3363 monitored packaging lines.

The set of agents (single filling machines) is embedded in the physical space of the PLMS parameters. In our “double vision” setup the efficiency behaviour at a plant level depends both on the positions and the “interactions” between its agents (machines) in the PMLS space. The effects of these relations are captured by the non local characteristics of the geometric structure of the plants fingerprint. Our aim is to characterise and categorise the geometric structure of the plants fingerprints. Assigning a plant to a class means finding the class to which the shape of the plant’s fingerprint adapts better.

We decide² to describe the internal states of the machines, responsible for their relations, by means of a scalar “material” field (section of a bundle with fibre \mathbb{R}). The discrete Laplace-Beltrami operator transforms such features and the analysis of its spectral structure provides deep geometric information on the smooth manifold M on (nearby) which the PLMS data are sampled.

²First of all for simplicity, then in order to keep the analogy with the standard fitness model, as in Ilinski (2001) we avoid (at least for the moment) the complications introduced by less trivial gauge transformations, but most of all because we had some purely technical computational difficulties in working with high order matrices, and implementing VDM procedure on the available PLMS data set enhances considerably the dimensions of the matrix which represents the discretized operator. Intention of the authors is to work on a separated case study of the application of VDM on the same PLMS data set.

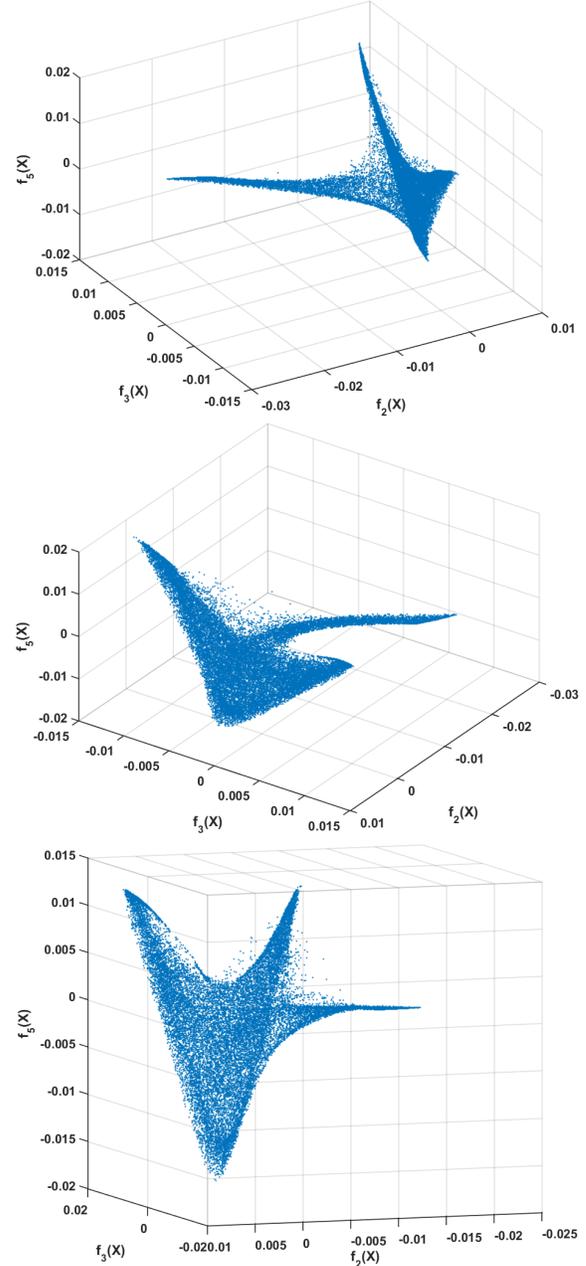


Figure 5. Different viewpoints on the Laplacian eigenmap scores of the whole available PLMS data set.

As we pointed out, the structure of the PLMS data can be considered approximately as overlapping of normally distributed homogeneous groups. So we can consider the hypotheses of Belkin & Niyogi (2005); Coifman & Lafon (2006) as satisfactorily fulfilled and compute the discrete eigenfunctions f_i of the Laplace-Beltrami operator from the complete experimental data set. On Fig. 5 are displayed different viewpoints on the plot of the whole available data set with respect to the second, the third and the fifth eigenfunctions of the Laplacian (denoted by f_2 , f_3 and f_5). These are the eigenfunc-

tions which appear as more “sensitive” to the phenomenon we are interested in. Recall that the coordinates of a point X of the sample in that feature space are given by $\mathbf{f}_i(X)$.

The eigenfunctions of the Laplacian on M provide a set of global degrees of freedom. Their level sets $\mathbf{f} = \text{const}$ define hyper-surfaces in M (as observed before, the discretisation permits to neglect many regularity issues). The set of points X_i of a single plant could be largely spread in M , but the set of values of $\mathbf{f}(X_i)$ could be much more condensed. This would mean that the shape of the cloud X_i resembles (adapts to) a level set of an eigenfunction \mathbf{f} .

Now we can introduce a similarity (affinity) relation based on the internal states of the agents of a plant. Interpreting a plant as an adaptive system, in this sense means hypothesising that its PLMS record assumes certain characteristic shape in the physical space because of the internal states of its machines. The eigenfunctions of the Laplace-Beltrami operator (which itself involves the Levi-Civita connection on the base manifold) are not arbitrary functions on M , but they are obtained by means of the extremisation of the tension functional involving local contrasting tendencies. The definition of the eigenfunctions is a dynamical geometric process. In this sense our interpretation goes beyond the well-known applications of the Laplace-Beltrami operator in spectral shape analysis (see Reuter & Peinecke (2005)).

To sum up, a plant is modelled as a complex adaptive system, its overall performance as an emergent phenomenon which depends on a deep interaction between the spatial positions and the internal states of its agents. This approach appears as very efficient. In fact the PLMS plot of each single plant becomes quite compact and homogeneous (see below) in the space generated by the Laplacian eigenfunctions.

Next we construct a spontaneous data driven clustering of plants in this feature space. This segmentation is based on sampling some special configurations of the internal state field (eigenfunctions of the Laplacian) in the complete experimental data set. It can be therefore interpreted as sampling and classifying the possible relations between the machines in a plant. In Shi & Malik (1997) and Ng A.Y. & Weiss (2002) it was first asserted that the spectral properties of graph Laplacians can allow to overcome classical non-linear constraints of data points clusters. Interpreting the observations as vertices of an undirected graph with weighted edges, the goal is in fact to minimize the “flow of similarity” amongst different clusters while maximizing intra-cluster affinity.

The first step is the creation of a weight matrix W with non-negative entries, which element W_{ij} represents the strength of the similarity between X_i and X_j ; then suppose the graph is partitioned into K clusters C_1, \dots, C_K , and define the “flow of similarity” between cluster C_i and the remaining part of the graph as $F(C_i, \bar{C}_i) = \sum_{X_{j_1} \in C_i, X_{j_2} \notin C_i} W_{j_1 j_2}$. Spectral

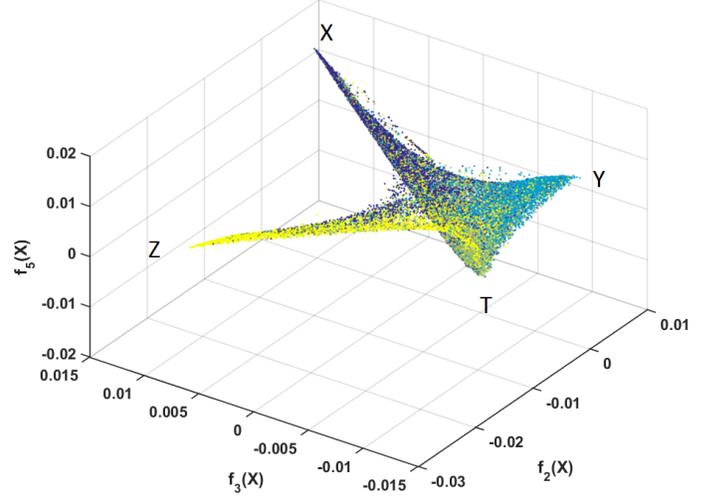


Figure 7. Plant affinity clustering.

clustering aims to minimize the so-called “Ratio Cut” of the partition

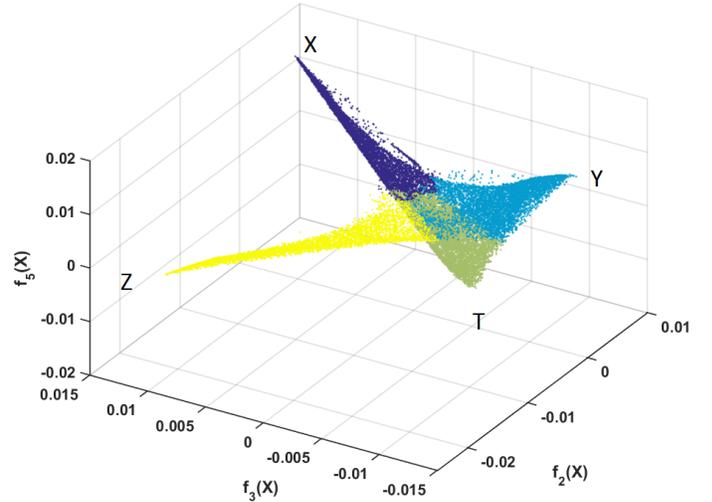


Figure 6. Spectral clustering on machines.

$$RC(C_1, \dots, C_K) = \frac{1}{2} \sum_{i=1}^K \frac{F(C_i, \bar{C}_i)}{|C_i|} \quad (5)$$

where the cardinality $|C_i|$ is taken into account to normalize the similarity with respect to the dimension of the cluster. Let $H \in M_{N \times K}(\mathbb{R})$ be an indicator matrix such that

$$H_{ij} = \begin{cases} \frac{1}{\sqrt{|C_j|}} & \text{if } X_i \in C_j \\ 0 & \text{if } X_i \notin C_j \end{cases}$$

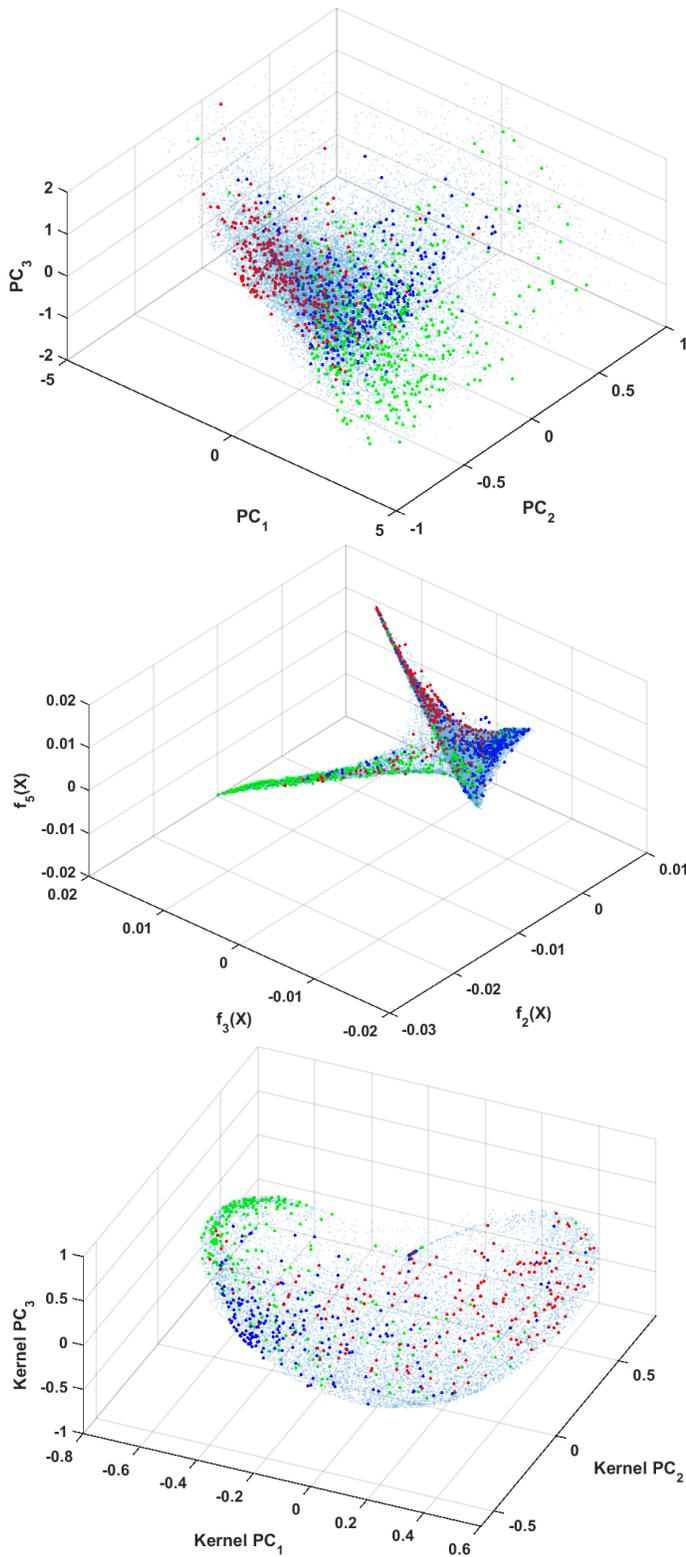


Figure 8. PCA scores, Laplacian eigenmap scores and kernel PCA scores of the PLMS records of 18 industrial plants (6 randomly extracted from each of X, Y and Z classes) compared to the whole PLMS calibration set in 2013.

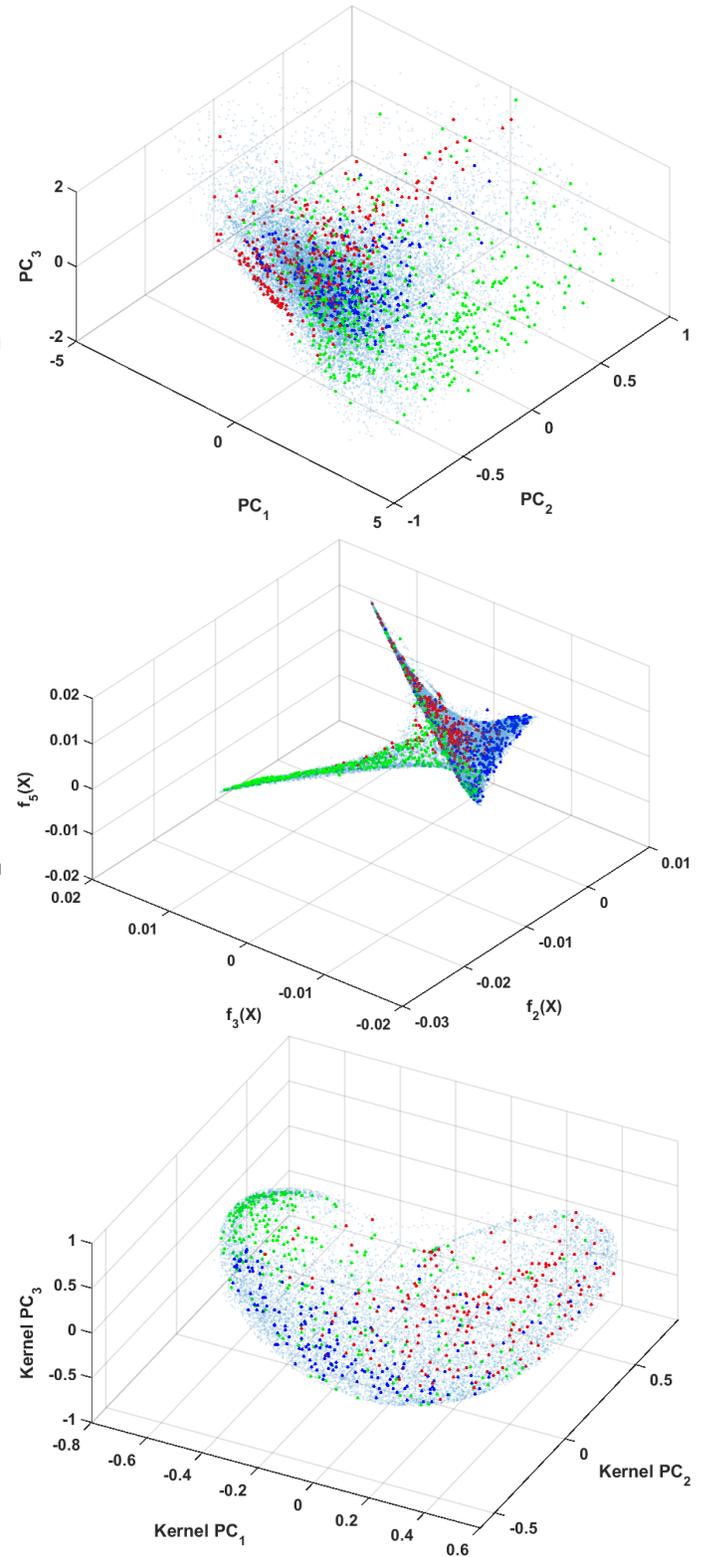


Figure 9. PCA scores, Laplacian eigenmap scores and kernel PCA scores of the PLMS records of 18 industrial plants (6 randomly extracted from each of X, Y and Z classes), compared to the whole PLMS calibration set in 2013.

In these terms the minimization of (5) is expressed as:

$$\min_{C_1 \dots C_K} \text{Tr}(H' LH)$$

where L denotes the graph Laplacian. Since the indicator matrix H takes values into a discrete set, this is a NP-hard problem; the relaxation of H to a matrix \mathcal{Y} which is allowed to take real values, with the only constraint $\mathcal{Y}'\mathcal{Y} = I_K$ (i.e., an arbitrary scaling constant is irrelevant), finally reveals the intimate relation with Laplacian eigenmaps. A simple k -means algorithm with $k = K$ is thus applied onto H 's rows to extract the clusters. The version proposed by Ng A.Y. & Weiss (2002) normalizes \mathcal{Y} 's rows before running k -means, in order to limit the influence of the degrees of the nodes on the value of the eigenvectors' components.

We remark that the connectivity of the affinity graph, whose vertices represent machines, is determined by their internal states i.e. clustering occurs in the feature space. The analogy with fitness network models is obvious, but in this case the state of each agent in the adaptive system is not random, but is obtained by a non-local extremisation process.

Spectral clustering is closely related to the geometric structure captured by the Laplacian eigenmaps, but the direct application of this method provides clusters of filling machines and not plants. The fact that plants, identified with the set of observations of their machines $I_l = \{X_{i_1} \dots X_{i_{N_l}}\}$ (where N_l denotes the number of observations on systems of plant I_l), are characterised by condensed images in the feature space, suggested an alternative clustering method. We define a pseudo-convolutional measure of similarity between plants I_l and I_h based on feature coordinates $\phi_l = \{Y_{i_1} \dots Y_{i_{N_l}}\}$:

$$s(I_l, I_h) = \sum_{Y_i \in \phi_l} \sum_{Y_j \in \phi_h} e^{-\frac{\|Y_i - Y_j\|^2}{\sigma^2}} \quad (6)$$

where σ^2 is an arbitrary parameter. Once the similarity matrix

$$S_{lh} = \begin{cases} s(I_l, I_h) & \text{if } l \neq h \\ 0 & \text{if } l = h \end{cases} \quad (7)$$

is evaluated, the spectral clustering algorithm is executed using (7) as a weight matrix and considering plants as nodes of a graph. Given the diagonal degree matrix

$$D_{ll} = \sum_{h=1}^{N_{plants}} S_{lh},$$

a graph Laplacian normalised by $LS = D^{-\frac{1}{2}}SD^{-\frac{1}{2}}$ is then computed, and the first \tilde{K} eigenvectors (excluding the first one), are grouped into a feature matrix Z , which is normalized by rows; finally, a k -means algorithm with $k = \tilde{K}$ is applied on Z 's rows, labeling plants by their features' convolutional affinity.

This clustering method is applied on plants, but care must be taken as the similarity measure (6) depends on a metric, so it is liable to errors when computed on discrete sets of points. In fact "pathological" examples of particular situations of wrong plant assignment can be easily constructed, but their impact in huge data samples appears to be marginal.

Both algorithms was applied for $k = 3, 4, 5, 6$. For $K = 4$ and $\tilde{K} = 4$ the two methods lead to remarkably collimating outputs, neat homogeneous clusters emerge both on level of plants and single machines (compare Fig. 6 to Fig. 7). A 4-group classification has been therefore adopted. From one side, the global features seem to model efficiently the relations between agents in the system (single machines). From the other, emergent clusters correspond to classes of geometrically similar plant fingerprints.

The emergent clusters have been analysed and interpreted as different production efficiency plant profiles i.e. efficiency segments. Segments Z and Y represent highly efficient plants with very different profiles from the viewpoint of quantity and complexity (number of different volumes, types shapes etc...) of their production. The segment X contains plants which are less efficient. Segment T represents sort of transition or "at risk area" between low and specific high performances.

On the first diagram of Fig. 8 are displayed the PCA scores of eighteen plants (six plants are randomly extracted from each of the segments X, Y and Z). Plants form the same segment are colored in the same way. The second diagram of Fig. 8 represents the Laplacian eigenmap scores of the same eighteen plants. The reader can compare the highly spread and overlapped images of the segments (the same regards the image of a single plant) in the physical space to the neat, compact and homogeneous images in this feature space. The same procedure of random extraction has been repeated many times with analogous results, Fig. 9 represents a different set of eighteen plants.

We have thus constructed a spontaneous classification of the efficiency categories of plants. Building completely data driven models typically runs into two types of risk. First of all structures could emerge accidentally from the data set, without being expression of some objective phenomenon. Furthermore objective but trivial structures might emerge.

The classification of the efficiency behaviour which emerged in our case is highly non-trivial with respect to the pre-assigned qualitative criteria in the sense that the efficiency segment are completely transversal with respect to their levels. The segmentation does not reproduce the distribution of the machine systems, the volumes, types or shapes of the produced packages or any geographical grouping. Actually the marginal distributions of all these qualitative criteria change in the segments and this gives the possibility to identify a typical plant in the segment. More importantly Z and Y represent plants

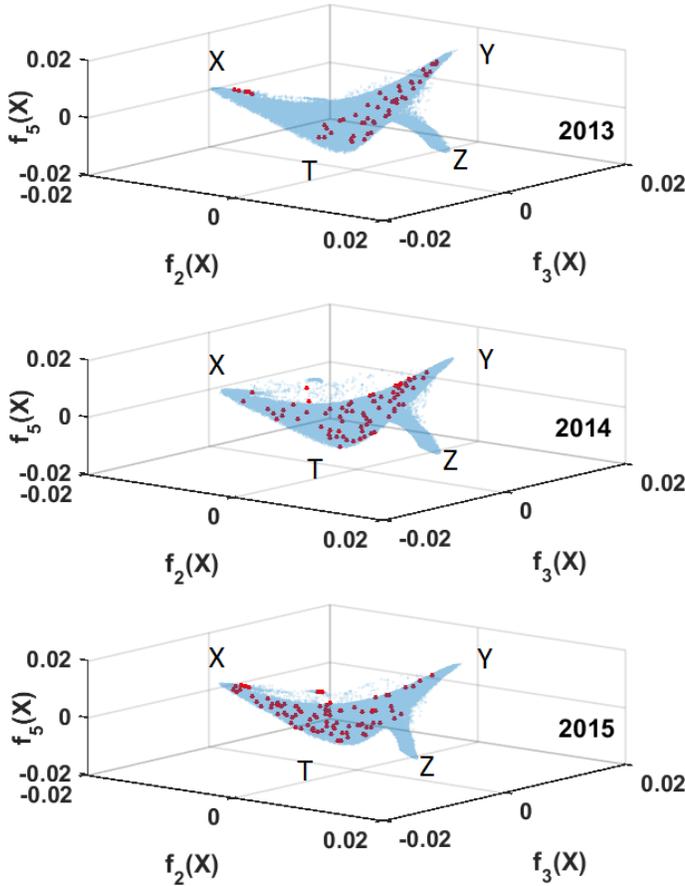


Figure 10. A plant “migration” in the period 2013-2015.

which are both efficient compatibly with quite different productive profiles.

A series of consistency checks have been done in order to confirm the objectivity of the classification:

- the application of the multiple fingerprint classifier to this segmentation on the same calibration and validation set mentioned before led to a correct re-classification of 93% of the plants. This result is much better if compared to the direct k-means (21%) and hierarchical clustering procedures (36%) applied directly on the physical PLMS data.

- more that 81% of the plants kept stable their segments in the period 2013-2015, this is a strong indication for the low level of “noise” in the model.

- the history of many cases of “migrations” of plants from one segment to another has been analysed by the experts of Tetra Pak. Remarkably the majority of these migrations can be attributed to objective causes. Our model detected concrete events as system upgrades, radical changes in the commercial profile of the plant or even changes in the plant management. We illustrate this last case on Fig. 10, where a clear transition of a plant from Y to X through T is visible.

Kernel PCA model is calibrated on the same PLMS set. We use the same functional form of the heat kernel:

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

where σ is a real parameter. The plot of the kernel PCA scores of the same set of 18 plants is displayed at the bottom of Fig. 8 and Fig. 9. We observe that the Kernel PCA plot is rather similar to the PCA plot. A slight “improvement” in fingerprints separation occurs with respect to classical linear PCA, but in our view kernel PCA does not provide particular advantages because of its “passive” nature. On the contrary the Laplacian method takes into account a dynamically interacting field and for this reason models better a complex adaptive system.

7. CONCLUSIONS

The overall mechanical performance behaviour of an industrial plant has been efficiently described as an emergent phenomenon in a complex adaptive system. The behaviour was highlighted by means of modern manifold learning methods. Many consistency tests have confirmed the objectivity of the results obtained from large amount of in-service data. This study provides an interesting motivating example for the further exploration of the use of modern manifold learning methods in modelling complex emergent phenomena.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Gabriele Molari and Roberto Fontana who made possible the realization of this joint research project. We thank professor Keith Worden for the interesting discussions and Nikolaos Dervilis for some useful technical suggestions.

REFERENCES

- Ache, A. G., & Warren, M. W. (2014). Coarse Ricci curvature with applications to manifold learning. *preprint available at arXiv:1410.3351*.
- Agricola, F. T., I., & Kassuba, M. (2008). Eigenvalue estimates for Dirac operators with parallel characteristic torsion. *Differential Geometry and its Applications*, 26, 613–624.
- Agricola, I., & Friedrich, T. (1999). Upper bounds for the first eigenvalue of the Dirac operator on surfaces. *Journal of Geometry and Physics*, 30, 1–22.
- Atiyah, M. (1973). Eigenvalues and Riemannian geometry. In *Proceedings of the international conference on manifolds and related topics in topology* (pp. 5–9).
- Baird, P., & Wood, J. (Eds.). (2003). *Harmonic morphisms between Riemannian manifolds*. Oxford University Press.
- Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural*

- Computation*, 15, 1373–1396.
- Belkin, M., & Niyogi, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *Proceedings of the 18th annual conference on learning theory* (pp. 486–500).
- Berry, T., & Sauer, T. (2015). Local kernels and the geometric structure of data. *in press, Applied and Computational Harmonic Analysis*.
- Bianconi, G., & Barabasi, A. (2001). Competition and multiscaling in evolving networks. *Europhysics Letters*, 54, 436–442.
- Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: theory and applications*. Springer.
- Bronstein, B. M., A.M., & Kimmel, R. (2006). Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 1168–1172.
- Choo, A. S. W. B. M. J. B. P., B.Y. (2016). Adaptive multi-scale prognostics and health management for smart manufacturing systems. *IJPHM ISSN2153-2648*, 014.
- Coifman, R. R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21, 5–30.
- Demartines, P., & Herault, J. (1997). Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of datasets. *IEEE Transactions on Neural Networks*, 8, 148–154.
- deSilva, V., & Tanenbaum, J. B. (2003). Global versus local methods in nonlinear dimensionality reduction. In *Advances in neural information processing systems* (pp. 721–728).
- Donoho, D. L., & Grimes, C. (2003). Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 5591–5596.
- Eells, J. J., & Sampson, J. H. (1964). Harmonic mappings of Riemannian manifolds. *American Journal of Mathematics*, 86, 109–160.
- Esposito, G. (1998). *Dirac operators and spectral geometry*. Cambridge University Press.
- Friedrich, T. (2012). The second Dirac eigenvalue of a nearly parallel G_2 -manifold. *Advances in Applied Clifford Algebras*, 22, 301–311.
- Gilkey, G. (1995). *Invariance theory, the heat equation, and the Atiyah-Singer index theorem*. CRC press.
- Ilinski, K. (2001). *Physics of finance: gauge modelling in non-equilibrium pricing*. Wiley.
- Jardim, M., & Leao, R. (2008). Survey on eigenvalues of the Dirac operator and geometric structures. *International Mathematical Forum*, 3, 49–67.
- Joyce, D. (2000). *Compact manifolds with special holonomy*. Oxford Mathematical Monographs.
- Mack, G. (2000). Universal dynamics, a unified theory of complex systems. emergence, life and death. *preprint available at arXiv:hep-th/0011074*.
- Nash, J. (1954). C^1 isometric imbeddings. *Annals of Mathematics*.
- Ng A.Y., M. I., Jordan, & Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. In *Advances in neural information processing systems 14* (pp. 849–856). MIT Press.
- O'Donovan, B. K. O. D., P. (2016). Adaptive multi-scale prognostics and health management for smart manufacturing systems. *IJPHM ISSN2153-2648*, 026.
- Reuter, W. F., M., & Peinecke, N. (2005). Laplace-spectra as fingerprints for shape matching. In *Proceedings of the ACM symposium on solid and physical modeling* (pp. 101–106).
- Rosenberg, S. (1997). *The laplacian on a Riemannian manifold: an introduction to analysis on manifolds*. Cambridge University Press.
- Sameh, A., & Wisniewski, J. (1982). A trace minimization algorithm for the generalized eigenvalue problem. *SIAM Journal on Numerical Analysis*, 19, 1243–1259.
- Shawe-Taylor, W. C., J. S., Cristianini, N., & Kandola, J. (2005). On the eigenspectrum of the Gram matrix and the generalization error of kernel pca. *IEEE Transactions on Information Theory*, 51, 2510–2522.
- Shi, J., & Malik, J. (1997). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 888–905.
- Singer, A., & Hau-Tieng, W. (2011). Orientability and diffusion map. *Applied and Computational Harmonic Analysis*, 31(1), 44–58.
- Singer, A., & Hau-tieng, W. (2012). Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65, 1067–1144.
- Singer, A., & Hau-Tieng, W. (2013). Spectral convergence of the connection Laplacian from random samples. *preprint available at arXiv:1306.1587*.
- Sternberg, S. (1963). *Lectures on Differential Geometry*. New York Chelsea Publishing Company.
- Tanenbaum, d. V., J. B., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319–2323.
- Xiao, W. (2016). A probabilistic machine learning approach to detect industrial plant faults. *IJPHM ISSN2153-2648*, 007.
- Zhang, Z., & H. Zha, H. (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.*, 26, 313–338.