Fast and user-friendly non-linear principal manifold learning by method of elastic maps

CONFERENCE PAPER · OCTOBER 2015

2 AUTHORS:

A. N. Gorban
University of Leicester

Andrei Zinovyev
Institut Curie

Fast and user-friendly non-linear principal manifold learning by method of elastic maps

Alexander N. Gorban
Department of Mathematics
University of Leicester
University Road, Leicester LE1 7RH, UK
Email: ag153@le.ac.uk

Andrei Zinovyev
Institut Curie, INSERM, U900
26 rue d'Ulm, 75248, Paris, FR
Email: andrei.zinovyev@curie.fr

***Abstract*—Method of elastic maps allows fast learning of non-linear principal manifolds for large datasets. We present user-friendly implementation of the method in ViDaExpert software. Equipped with several dialogs for configuring data point representations (size, shape, color) and fast 3D viewer, ViDaExpert is a handy tool allowing to construct an interactive 3D-scene representing a table of data in multidimensional space and perform its quick and insightfull statistical analysis, from basic to advanced methods. We list several recent application examples of manifold learning by method of elastic maps in various fields of life sciences.**

I. Introduction

Principal Component Analysis (PCA), introduced by Karl Pearson in 1901 [1] remains the most used and the most basic statistical tool used for dimension reduction of large and multi-dimensional datasets. The notion of principal manifolds (non-linear generalization of the linear principal manifold) was introduced into the practice of statistical data analysis in 1990s (starting from one-dimensional principal curves [2], [3], [4], [5]) but remained relatively underused because of costly algorithms for their computation [21]. In many practical examples from life sciences, it was shown that using non-linear manifolds allows getting more information from dimension reduction and formulate more hypotheses from exploratory data analysis compared to the linear ones [22], [26]. Due to necessity of dealing with big data in life sciences, increase of available computational power and due to appearance of new datasets with non-trivial internal structures, the interest to *manifold learning*, in general, rapidly increases. Despite this fact and many existing ideas for principal manifold learning [21], currently we are not aware of any open software allowing time-efficient construction of principal manifolds, usable by non-experts in scientific computation (i.e., within a user-friendly graphical user interface).

In mid-1990s the authors of this manuscript introduced one of the most efficient algorithm for approximating large and multidimensional datasets by non-linear principal manifolds, called the method of *elastic maps* [4], [6], [15], [7], [8], [11], [10], [9], [17], [20], [21], [24], [26]. Since then, the method found multiple applications in various fields of science. It was also generalized for non manifold-based data approximations such as *principal trees* and, more generally, *cubic complexes* [18], [20].

In this manuscript we briefly describe a user-friendly software, ViDaExpert, allowing computing and visualizing the elastic maps using a graphical user interface (GUI), equipped with many practical functions. ViDaExpert was already shown to be useful in many applications: some of them are listed in a section of this manuscript.

ViDaExpert software is freely available at http://bioinfo-out.curie.fr/projects/vidaexpert. The tool does not require installation. Currently, there is no implementation of ViDaExpert for platforms other than Windows (any version).

II. Method of elastic maps and its extensions

Elastic map provides a method for nonlinear dimensionality reduction. Elastic map is a system of elastic springs embedded in the data space and approximating a low-dimensional manifold. Tuning the elastic coefficients of springs allow switching from completely unstructured k-means clustering (zero elasticity) to the estimators located closely to linear PCA manifolds (for high bending and low stretching elasticities). With some intermediate values of the elasticity coefficients, this system effectively approximates non-linear principal manifolds. The approach is based on a mechanical analogy between principal manifolds, that are passing through "the middle" of the data distribution, and the elastic membranes and plates. The method was developed by the authors of this paper, starting from 1996-1998. The most exhaustive theoretical description of the elastic map methodology is provided in [24] together with a comprehensive review on the related methods. A more practice-oriented description is given in [26]. Here we just remind the basic principles of the method.

A. *Computing the principal manifold*

Construction of the elastic map is based on expectation-minimization-based energy optimisation algorithm, which utilises annealing methodology in order to achieve a deeper local minimum of the energy function. Therefore, the elastic map is trained in several epochs each characterized by certain elasticity coefficients. After optimising the position of grid nodes of the system of springs, the manifold is extended by linear extrapolation to its vicinity in the data space, in order to avoid projecting many data points onto the border of the manifold.

Let G be a simple undirected graph with set of vertices V and a set of edges E. *k-star* in a graph G is a subgraph with $k+1$ vertices $v_{0,1,...,k} \in V$ and k edges $\{(v_0, v_i)|i = 1, .., k\}$.

Suppose that for each $k \geq 2$, a family S_k of k-stars in G has been selected. Then we define an *elastic graph* as a graph with selected families of k-stars S_k and for which for all $E^{(i)} \in E$ and $S_k^{(j)} \in S_k$, the corresponding elasticity moduli $\lambda_i > 0$ and $\mu_{kj} > 0$ are defined.

Let $E^{(i)}(0)$, $E^{(i)}(1)$ denote two vertices of the graph edge $E^{(i)}$ and $S_k^{(j)}(0), ..., S_k^{(j)}(k)$ denote vertices of a k-star $S_k^{(j)}$(where $S_k^{(j)}(0)$ is the central vertex, to which all other vertices are connected). Let us consider a map $\varphi{:}V \rightarrow \mathbf{R}^m$ which describes an embedding of the graph into a multidimensional space. The *elastic energy of the graph embedding in the Euclidean space* is defined as

$$U^{\varphi}(G) := U_E^{\varphi}(G) + U_R^{\varphi}(G), \quad (1)$$

$$U_E^{\varphi}(G) := \sum_{E^{(i)}} \lambda_i \left\| \varphi(E^{(i)}(0)) - \varphi(E^{(i)}(1)) \right\|^2, \quad (2)$$

$$U_R^{\varphi}(G) := \sum_{S_k^{(j)}} \mu_{kj} \left\| \varphi(S_k^{(j)}(0)) - \frac{1}{k} \sum_{i=1}^{k} \varphi(S_k^{(j)}(i)) \right\|^2. \quad (3)$$

Elastic map is a continuous manifold $Y \in \mathbf{R}^m$ constructed from the elastic net as its grid approximation using some between-node interpolation procedure. This interpolation procedure constructs a continuous mapping $\varphi_c{:}\{\varphi_1, \ldots, \varphi_{dim(G)}\} \rightarrow \mathbf{R}^m$ from the discrete map $\varphi{:}V \rightarrow \mathbf{R}^m$, used to embed the graph in $\mathbf{R}^m$, and the discrete values of node indices $\{\lambda_1^i, ..., \lambda_{\dim(G)}^i\}$, $i = 1...|V|$. For example, the simplest *piecewise linear elastic map* is build by piecewise linear map φ_c.

Elastic principal manifold of dimension s for a dataset X is an elastic map, constructed from an elastic net Y of dimension s embedded in $\mathbf{R}^m$ using such a map $\varphi_{opt}{:}Y \rightarrow \mathbf{R}^m$, that corresponds to the minimal value of the functional

$$U^{\varphi}(X, Y) = MSD_W(X, Y) + U^{\varphi}(G), \quad (4)$$

where the weighted mean squared distance from the dataset X to the elastic net Y is calculated as the distance to the finite set of vertices $\{\mathbf{y}^1 = \varphi(v_1), ..., \mathbf{y}^k = \varphi(v_k)\}$.

In the Euclidean space one can apply an EM algorithm for estimating the elastic principal manifold for a finite dataset. It is based in turn on the general algorithm for estimating the locally optimal embedding map φ for an arbitrary elastic graph G (see [24]).

III. VIDAEXPERT SOFTWARE FOR FAST LEARNING OF NON-LINEAR MANIFOLDS

ViDaExpert (ViDa stands for Visualization of Data) is a software implementing a number of simple and advanced statistical methods and a user-friendly graphical user interface (GUI) for applying these methods to a table of data which can contain both numerical feature values and labels for objects and features. The primary objective of ViDaExpert is to implement a user interface to the method of *elastic maps* for non-linear data dimension reduction and visualization developed by the authors of this paper in [6], [7], [8], [11], [10], [9], [17], [20], [21], [24], [26]. It appeared advantageous to introduce standard methods of multivariate statistical analysis into the software to be able to visualize the result of their applications on the projections onto the elastic map (non-linear principal manifold). Currently, ViDaExpert contains quite a diverse set of statistical tools making it a handy self-consistent environment for performing fast exploratory statistical analysis and visualization of multivariate data.

A. Basic principles of ViDaExpert software

The basic scenario of working in VidaExperts consists of several simple steps.

(1) The user data should be prepared in a specific text-based format described below and loaded into ViDaExpert GUI. This will create an object called further DataTable.

(2) From DataTable, a user should create a DataSet: an object representing the cloud of points in multidimensional data.

(3) DataSet can be visualized using one or several Map objects representing a mapping of a multidimensional data into a low-dimensional space.

(4) A user can apply various statistical methods implemented in ViDaExpert, such as Linear Discriminant Analysis, and see visualization of the results of these methods in the constructed 3D-scene.

The steps of this scenario are explained and illustrated in the Tutorial in Supplementary Materials for this text.

B. Input format

The recommended format for using in ViDaExpert is a textual file having ".dat" extension. The file represents the table of numerical and string values with a simple header which contains, first, type of the column (FLOAT for numerical type and STRING for the labeling information) and, second, any labeling of columns (see example in the tutorial). Importantly, it is recommended to put in quotes any textual information and use tab symbol to delimitate the columns in the table and in the header.

C. DataTable object

When .dat file is loaded into ViDaExpert, a DataTable object is created. In one ViDaExpert session, a user can load several DataTable objects and switch between them. A user can save the content of the table into ViDaExpert format, which will have ".vet" extension. In this case, all attributes associated to objects (color, shape and size) will be saved as well and can be restored later.

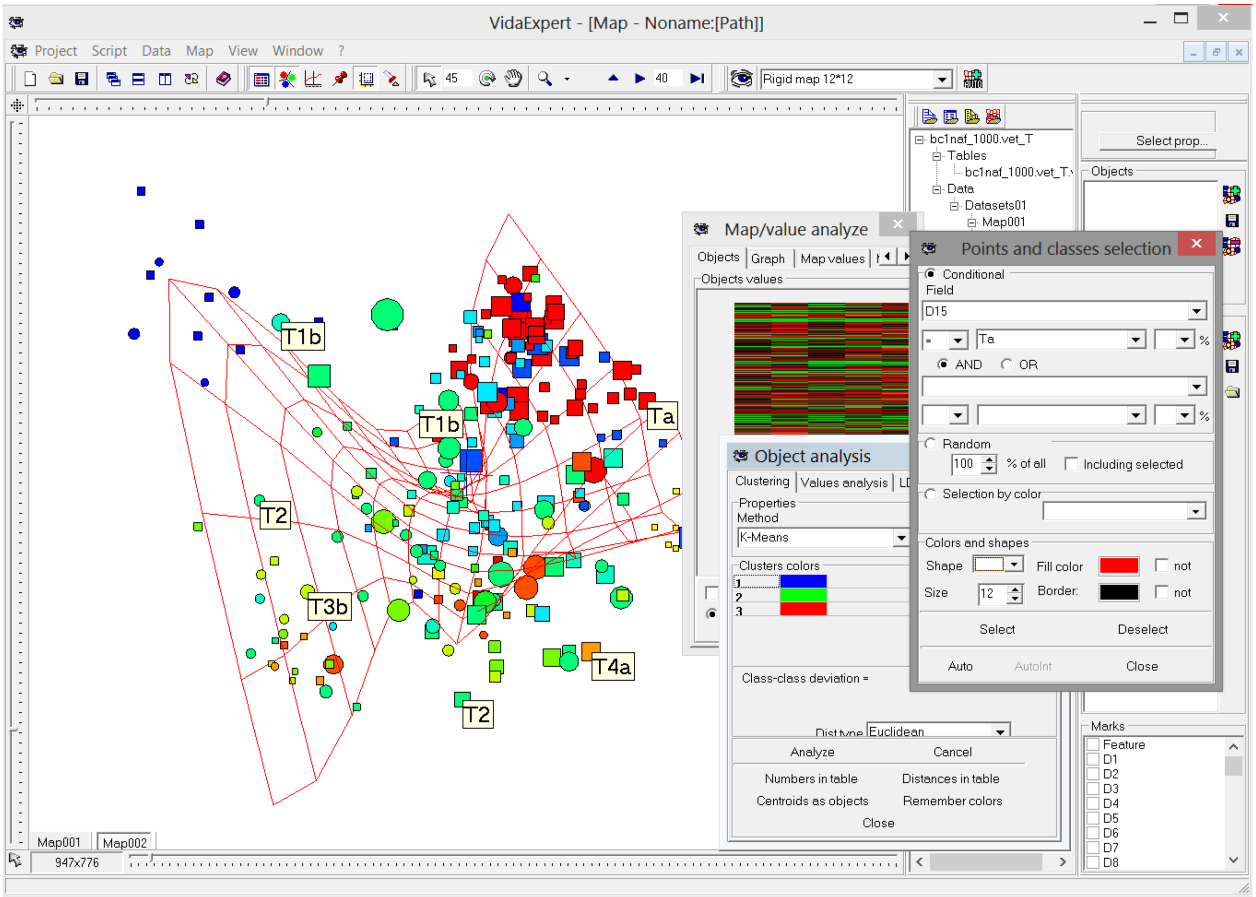

Fig. 1. ViDaExpert GUI screenshot. Most of the working space is occupied by interactive 3D-scene visualizing a cloud of data points and the elastic maps constructed for these data. Several specialized dialogs allows to launch various statistical procedures and visualize the results of their applications in the scene.

D. DataSet object

DataSet represents a numerical matrix in ViDaExpert. This matrix is formed by selecting certain columns having numerical values in DataTable and normalizing (scaling) them. The default normalization consist of subtracting the mean value and dividing by standard deviation of the values in the column. Other types of normalization are also available, including reducing the values to [-1;1] interval, using logarithm or hyperbolic tangent function. Several types of normalization can be combined for creating a DataSet. In addition, the user can decide not to use those rows in DataTable which contains undefined numerical values (gaps): in this case the number of rows in DataSet will become smaller than the number of rows in DataTable. However, many methods in ViDaExpert (including PCA and elastic maps) are capable of dealing with moderate number of gaps in a DataSet, even without prior imputation of missing values.

For each DataTable, a user can create several DataSets, which might be advantageous for using the same set of object attributes (color, size and shape) for showing the data points in different data spaces created by different column selections. For example, colors defined from application of K-Means clustering in one data space can be visualized in another data space, which was not used for clustering.

The DataSet can be saved into the ViDaExpert format having ".ved" extension and loaded later.

E. Map object

The Map object is a representation of an elastic map, low-dimensional manifold (screen), computed for a cloud of multidimensional vectors represented by a DataSet object. By default, the map represents a linear manifold spanned by the direction of the first two principal components. A user can change a number of predefined parameter configurations in order to compute a principal manifold by elastic map method, or construct the elastic map in a manual mode, by specifying the elastic parameters of the map.

After the map is created, a user can visualize both data points and the map together in several projections: (1) in the space of the three chosen by the user data space coordinate axes; (2) in the space defined by first three principal components and (3) in the internal space of the map after projecting the data points onto it. In all cases, the constructed 3D-scene can be rotated, zoomed and shifted by the user.

The map can be saved into the ViDaExpert format having ".vem" extension and loaded later.

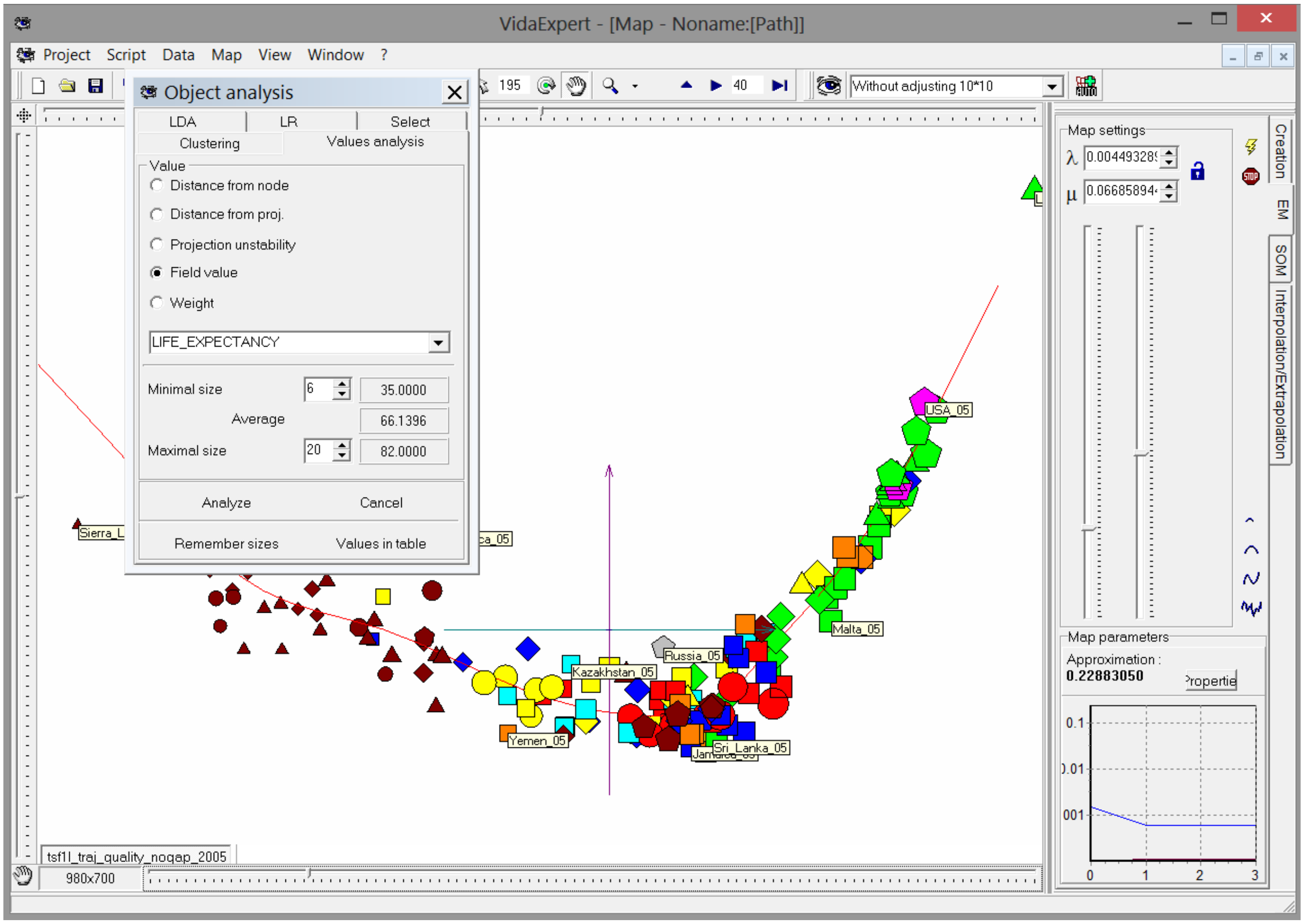

Fig. 2. Using VidaExpert for constructing a principal curve for a dataset of quality of life indices for all countries in 2005 [26], [27].

F. Using Principal Component Analysis in ViDaExpert

Principal Component Analysis (PCA) is used in ViDaExpert for several purposes: (1) for constructing one of the projections of the data points and the elastic maps; (2) for initializing the elastic map; (3) for analysing contribution of a feature into the nth principal component and estimating the amount of variance explained by the nth principal component. Implementation of PCA in ViDaExpert differs from most of the standard implementations in the following aspects: (1) it is able to work with weighted data vectors (there is a dialog in ViDaExpert allowing to associate a weight to each data point) and (2) it is able to compute principal components for a matrix which can contain missing values without imputing them. The iterative algorithm for computing Singular Value Decomposition (SVD) for PCA is described in details in [24].

For user convenience, ViDaExpert allows to make standard DataTable transformations, frequently utilized in PCA such as transposing the DataTable or subtracting the first principal component (which can help to eliminate, for example, a major bias affecting the average feature values of an object, through all columns). Another useful implemented feature is using biplots for presenting the results of linear PCA [30] (see Figure 3).

G. Elastic maps method implementation in ViDaExpert

In ViDaExpert, the sequence of steps for constructing the elastic maps is either predefined in a number of standard scenarios which can be applied by a single click, or can be set manually by a user.

Elastic map algorithm allows constructing principal manifolds of various topology (rectangular, hexagonal, spherical) and dimension (1D manifolds or principal curves, 2D or 3D manifolds). Several such possibilities are provided to the ViDaExpert users. Among these build-in scenarios there are:

”Without adjusting” computes simple 2D linear principal manifold.

”Rigid map” computes relatively rigid and smooth elastic map, closer to the linear one.

”Soft map” computes much more non-linear than ”Rigid map” elastic map, which better approximates the data, but can be trapped in some too complicated locally optimal configurations.

”Soft spherical map” constructs a 2D principal manifold having topology of a sphere.

”Detailed map” constructs a systems of springs as a grid

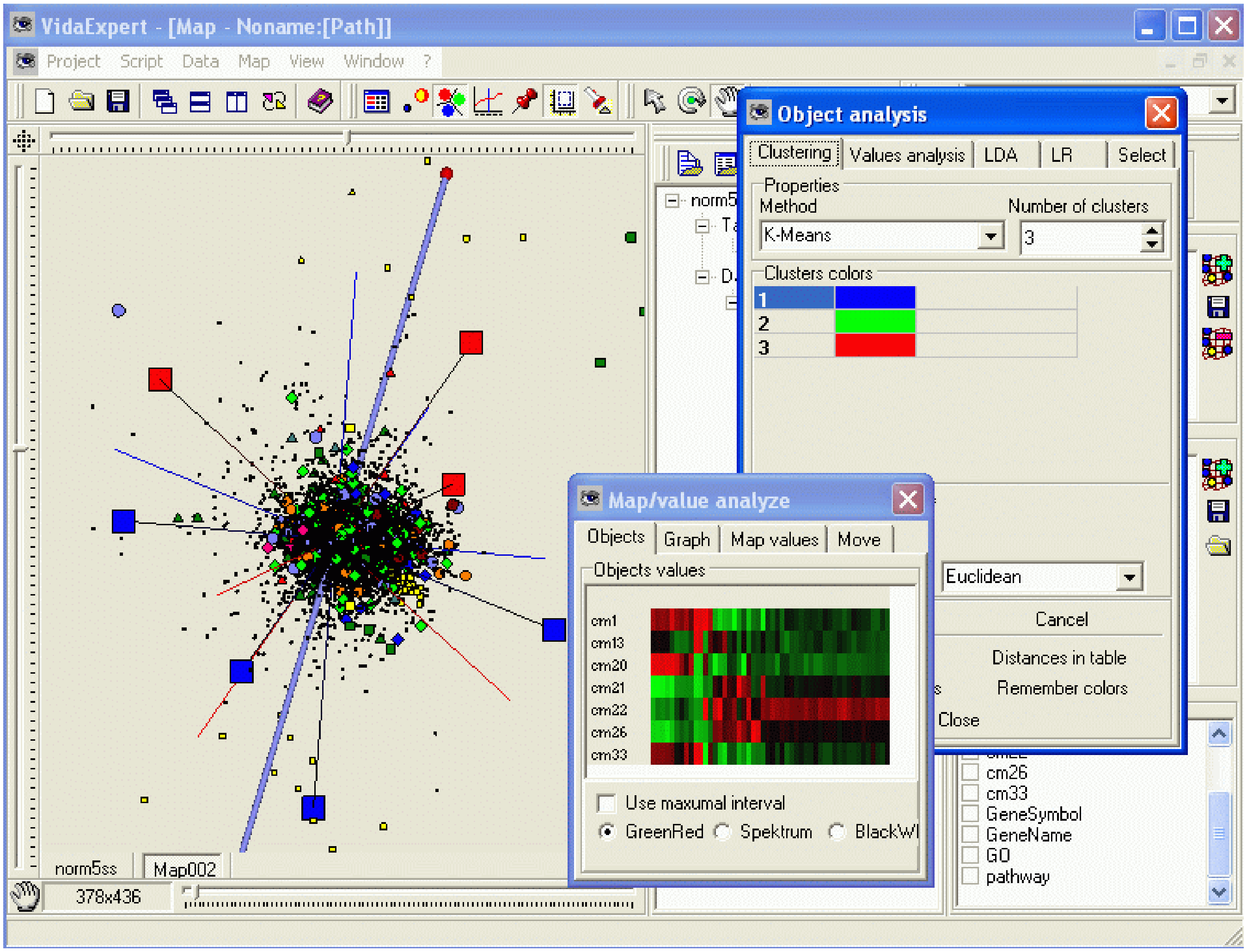

Fig. 3. Constructing biplot representation of principal components [30] in VidaExpert for a distribution of gene expression values.

with many nodes ($25 \times 25 = 625$ nodes) which is bigger than typically used in other scenarios ($10 \times 10 = 100$ nodes).

”3D” map constructs three-dimensional non-linear principal manifolds by using a grid with $6 \times 6 \times 6 = 216$ nodes.

H. Coloring the manifold

After the manifold is constructed, it can be used in a number of ways. First of all, it can be used to visualize the distribution of data points after projecting them into the closest point of the manifold, and visually estimate if the points form a cluster structure (which in practice can be fuzzy and can not be easily determined by clustering algorithms). Second, the manifold itself can be used to visualize some functions defined in the multidimensional space or in the space of internal manifold coordinates.

In this way, we implemented a possibility to visualize (1) local density of data points both in multidimensional and in the projected space; (2) the feature values in each point of the manifold that represents a smoothed trend of the feature along the manifold; (3) some other functions such as those resulting from application of linear regression (LR) or linear discriminant analysis (LDA).

This technique is called ”Map coloring” in ViDaExpert. For example, coloring by density allows to visually perform clustering of data points (Figure 4). Coloring by a particular feature value produces a smoothed trend image of the values of this feature, along the manifold. Coloring by the result of LDA allows to visualize the distribution of misclassified data points. The user can use discrete or gradient coloring, choose tints of several pre-defined colors, use spectral-type coloring in which blue color denotes smaller and red color denotes bigger values, and use green-red coloring which is typically used in bioinformatics applications.

I. Interacting with the manifold

In ViDaExpert, a user can interact with constructed elastic manifold, using specialized dialogs. For example, it is possible to shift (translate) or rotate the manifold along one of the linear axes defining the three-dimensional subspace used for visualization. It is also possible to compute interactively the projection onto the manifold of a point in the dataspace defined

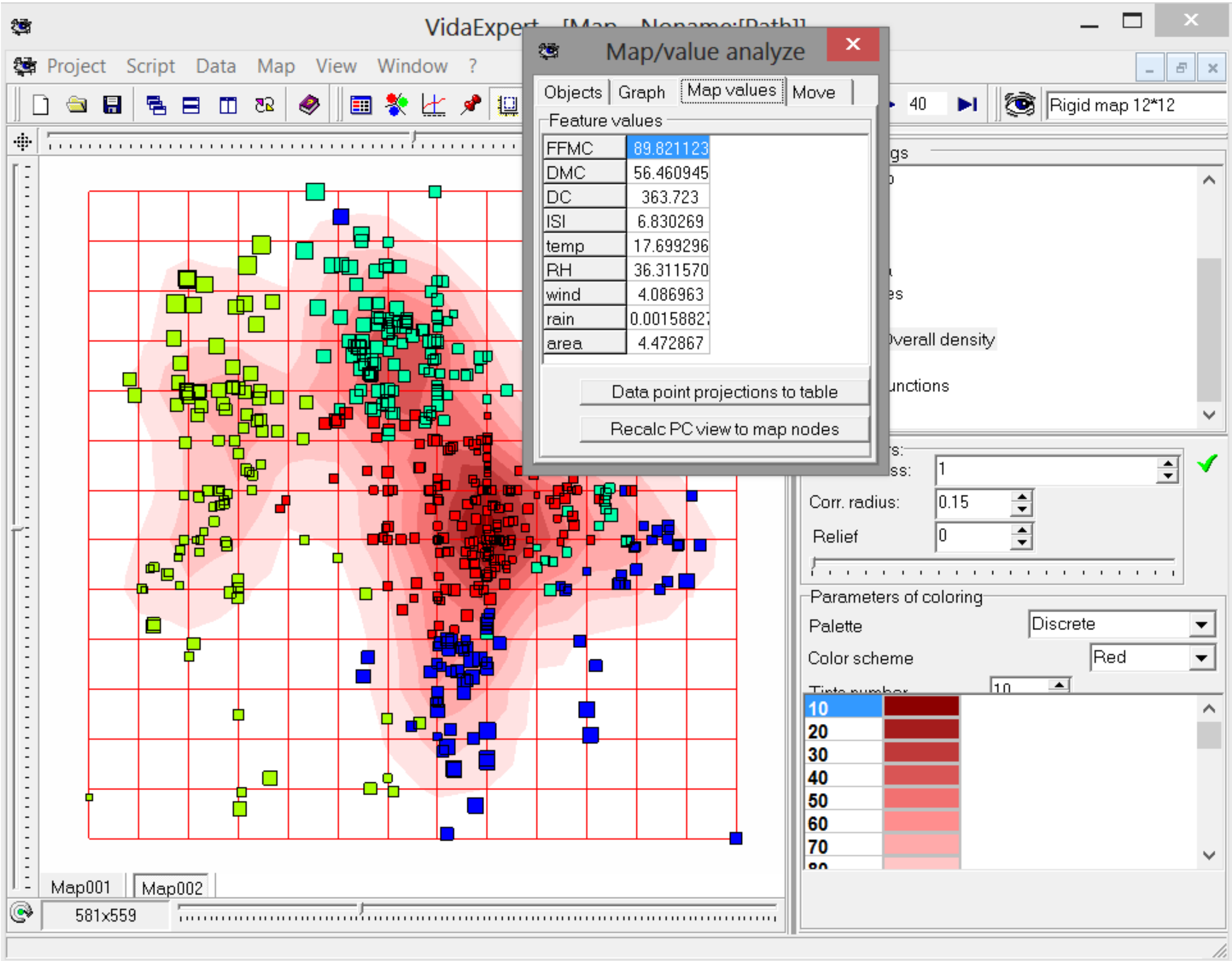

Fig. 4. Visulializing the density of data point projections onto the principal manifold (gradient in the background) in order to allow the visual estimation of existence of subgroups in the dataset describing forest fires [31] and comparing it with the results of K-Means clustering algorithm (shown by coloring the data points).

by mouse pointer (this answers the question "what would be projection of a data point if it appears in the position of the mouse pointer").

IV. Usefull ViDaExpert features

A. Configuring appearance of objects

ViDaExpert contains advanced dialogs containing multiple possibilities to change appearance of the objects in order to reflect their labels or numerical values by using shape, color (both background and the border colors of the points) and size. Many of these possibilities are automated: for example, it is possible to assign, in one click, different colors to data points for each distinct text they have in a certain label (point class information). Two features of the DataTable can be combined by OR or AND logics in order to define a certain visual appearance of data points.

In most scenario, the background color of data points is associated to the class or cluster information. In contrast, the size of the points can be used to map the values of certain features, or some other values such as the distance to the constructed manifold (approximation residue).

B. Labeling objects

One of the user dialogs in ViDaExpert allows the user to attach a label to all or to a subselection of objects (data points). This label can reflect one of the row labels contained in the DataTable, a numerical value of a feature, or combine several labels together. In addition, more advanced labels can be assigned to data points, such as the name of the "least expected feature" (the feature that has the least typical value for this point).

C. Interaction with Microsoft Excel software

A usefull feature of ViDaExpert is using OLE technology to call Excel software and communicate some table information without saving it on the disk. Most of the dialogs for the implemented methods in ViDaExpert are equipped with "In

Excel" button, allowing to transfer the results of the method's application (i.e., the feature contributions into the first three principal components) into Excel, where the user can create different plots or manipulate the data at his discretion.

D. Introducing auxiliary objects into ViDaExpert 3D-scene

ViDaExpert allows describing a scene composed from spheres and cylinders positioned in the multidimensional space and then projected into a low-dimensional space. The corresponding file describing positions, colors and sizes of objects has ".veo" extension ("o" stands for "objects"). This possibility might be usefull in many situations, for example, for marking some distinguished points (centroid of a class) in the space and connections between them. Another application is visualizing the structures of complex molecules in ViDaExpert, using its 3D-viewer for rotating them (Figure 5). Another possibility consists in introducing line segments into the ViDaExpert scene, which, for example, can be used for visualization of a vector field.

V. Applications of the ViDaExpert software and the method of elastic maps

ViDaExpert software has been applied in many domains of science where there is a need to visually represent tables of numerical data.

Thus, it was used to visualize economic indicators of Russian economy [6], [15], [10], [7]. ViDaExpert was applied in political science for data visualization [34], [26]. In particular, this technology solves a classical problem of unsupervised ranking of objects. It allows to find the optimal and independent on expert's opinion way to map several numerical indicators from a multidimensional space onto the one-dimensional space of the "quality" or "index" [27], [32], [33] (see Figure 2).

The method is adapted as a support tool in the decision process underlying the selection, optimization, and management of financial portfolios [25].

Most of the applications of ViDaExpert software and elastic map method found in bioinfomatics. It was used to visualize the universal 7-cluster structure of bacterial genomes [12], [13] and the structure of codon usage in genomes of various organisms [16], [28], [29]. Elastic maps allow approximation of molecular surfaces of complex molecules and visualizing them in ViDaExpert [17] (see Figure 5). ViDaExpert is routinely used for analysis of microarray data in cancer biology [22], [26], [28] and in biology of microorganisms [19]. The method of elastic maps is applied in quantitative biology for reconstructing the curved surface of a tree leaf from a stack of light miscroscopy images [38].

The method of elastic maps was successfull in tracing skeletons of handwritten symbols [17]. The method of elastic maps has been systematically tested and compared with several machine learning methods on the applied problem of identification of the flow regime of a gas-liquid flow in a pipe [39]. Generalizations of elastic map method can be used to quantify and compare the complexity of large sets of data [35], [36], [28].

VidaExpert software is used in mathematical modeling of biological systems for visualizing multidimensional trajectories of the dynamical systems [40], distribution of stable states of discrete models of biological mechanisms [41]. Recently, ViDaExpert software was applied for visualizing the properties of genetic interactions predicted from mathematical modeling of regulatory networks [42]. Mapping biological networks with associated multi-variate data, applying elastic maps approach, was used for constructing the data-driven layouts of biological networks [43].

VI. Conclusion

Computing non-linear principal manifolds remains underused in life science, because of its reputation to be computationally demanding method, compared to the classical PCA. Method of elastic maps represents of the most efficient algorithm for approximating low-dimensional non-linear manifolds. Its implementation in ViDaExpert software makes this method available for final users, not familiar with programming. We believe that ViDaExpert will allow making more sense of large datasets constantly appearing in various domains of life science.

Acknowledgment

The authors would like to thank Alexander Pitenko for participating in ViDaExpert implementation, and all users of ViDaExpert for providing their feedback and use cases.

References

[1] K. Pearson. "On lines and planes of closest to systems of points in space." *Philos. Mag.*, vol.2, pp.559-572, 1901.

[2] T. Hastie, W. Stuetzle. "Principal curves." *Journal of the American Statistical Association*, vol.84, pp.502-516, 1989.

[3] B. Kegl, A. Krzyzak, T. Linder, K. Zeger. "A polygonal line algorithm for constructing principal curves". in *Neural Information Processing (NIPS 98), Denver, CO*, 1-3 December 1998, pp.501-507.

[4] A.N. Gorban', A.A. Rossiev. "Neural network iterative method of principal curves for data with gaps." *Journal of Computer and Systems Sciences International*, vol.38(5), pp.825-830, 1999.

[5] B. Kegl, A. Krzyzak, T. Linder, K. Zeger. "Learning and design of principal curves." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22(3), pp.281-297, 2000.

[6] A.N. Gorban, A.Yu. Zinovyev, A.A. Pitenko. "Visualization of data using method of elastic maps." (in Russian) *Informatsionnie technologii, 'Mashinostrornie' Publ.(Moscow)*, vol.6, pp.26-35, 2000.

[7] A.N. Gorban, A.A. Pitenko, A.Y. Zinov'ev, D.C. Wunsch. "Vizualization of any data using elastic map method." in *Proc. Artificial Neural Networks in Engineering, St. Louis, MO. Smart Engineering System Design*, 2001, vol.11, pp.363-368.

[8] A.N. Gorban, A.Yu. Zinovyev. "Method of Elastic Maps and its Applications in Data Visualization and Data Modeling." *International Journal of Computing Anticipatory Systems (CHAOS)*, vol.12, pp.353-369, 2001.

[9] A.N. Gorban, A.Yu. Zinovyev, A.A. Pitenko "Visualization of data. Method of elastic maps." (in Russian), *Neurocomputery: Razrabotka, Primenenie [Neurocomputers: Development and Application]*, vol.4, pp.19-30, 2002.

[10] A.Yu. Zinovyev, A.A. Pitenko, T.G. Popova. "Practical applications of the method of elastic maps." (in Russian), *Neurocomputery: Razrabotka, Primenenie [Neurocomputers: Development and Application]*, vol.4, pp.31-39, 2002.

[11] A. Zinovyev. *Visualization of Multidimensional Data.* (in Russian), KGTU Publ., Krasnoyasrk, 2000, p.180.

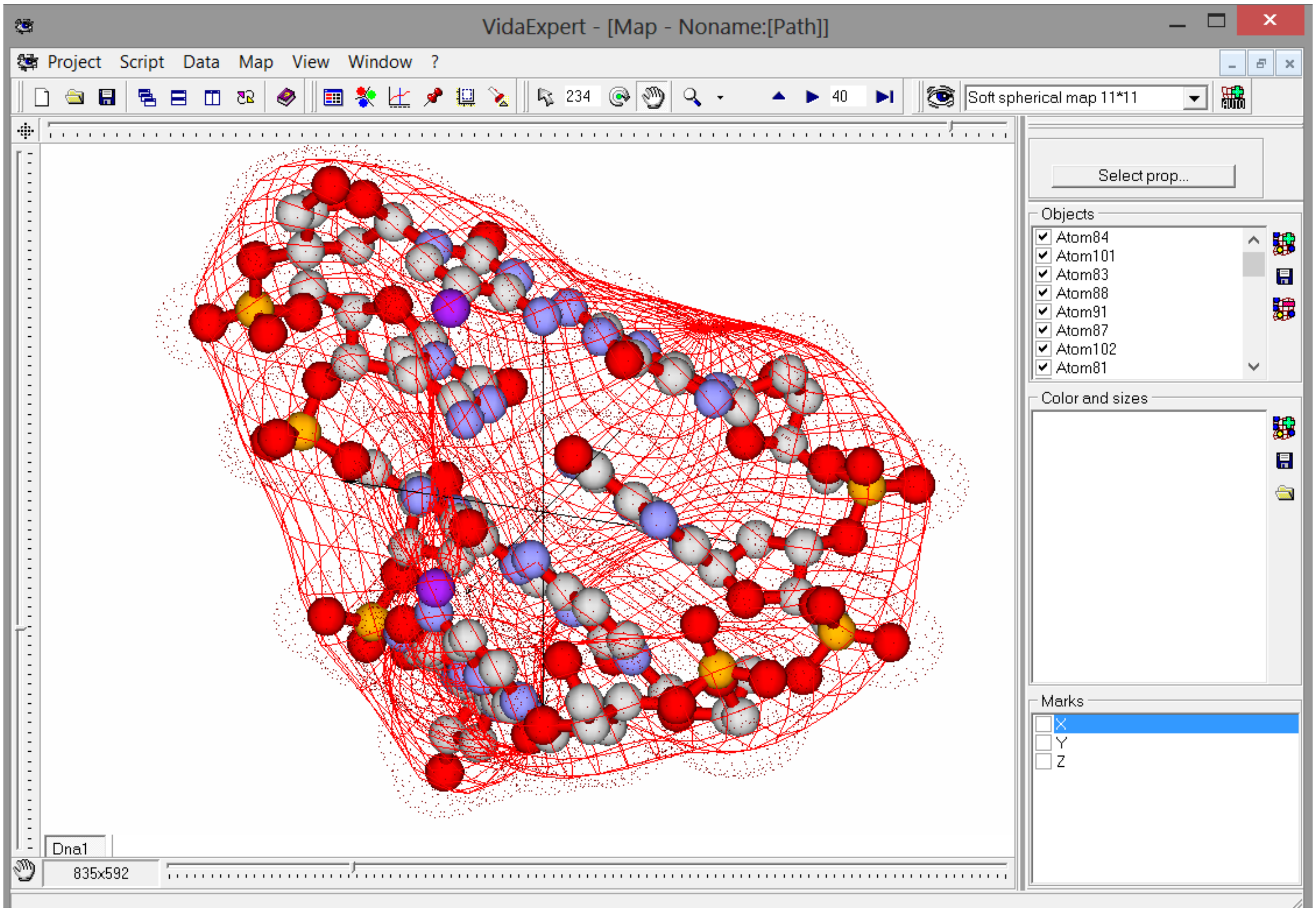

Fig. 5. Using VidaExpert for visualizing the backbone of a molecule (several DNA nucleotides in this case) in 3D, and approximating the continuous molecular surface by computing the principal manifold with spherical topology [17].

[12] A.N. Gorban, A.Y. Zinovyev, D.C. Wunsch "Application of the method of elastic maps in analysis of genetic texts.", in *Proc. International Joint Conference on Neural Networks (IJCNN2003), Portland, Oregon*, 2003, vol.3, pp.1826-1831. Available: http://cogprints.org/3961/01/IJCNN2003.pdf.

[13] A.N. Gorban, A.Y. Zinovyev, T.G. Popova. "Seven clusters in genomic triplet distributions." *In Silico Biology*, vol.3, 0039, 2003. Available: http://www.bioinfo.de/isb/2003030039/.

[14] A.N. Gorban, A.Y. Zinovyev, T.G. Popova. "Four basic symmetry types in the universal 7-cluster structure of microbial genomic sequences." *In Silico Biology* vol.5(3), 265-282, 2004. Available: http://www.bioinfo.de/isb/2005/05/0025/.

[15] A.N. Gorban, A. Zinovyev. "Visualization of data by method of elastic maps and its applications in genomics, economics and sociology." *IHES Preprints*, IHES/M/01/36, 2001. Available: http://preprints.ihes.fr/M01/Resu/resu-M01-36.html.

[16] A. Carbone, A. Zinovyev, F. Kepes. "Codon Adaptation Index as a measure of dominating codon bias." *Bioinformatics* vol.19(13), pp.2005-2015, 2003.

[17] A. Gorban, A. Zinovyev. "Elastic Principal Graphs and Manifolds and their Practical Applications." *Computing* vol.75, pp.359-379, 2005. Available: http://arxiv.org/abs/cond-mat/0405648.

[18] A.N. Gorban, N.R. Sumner, A.Y. Zinovyev. "Branching principal components: elastic graphs, topological grammars and metro maps," in *Proceedings of International Joint Conference on Neural Networks Orlando, USA*, 2007.

[19] M. Chacòn, M. Lévano, H. Allende, H. Nowak. "Detection of gene expressions in microarrays by applying iteratively elastic neural net" in *Lecture Notes in Computer Sciences*, vol.4432, B. Beliczynski et al., Eds., Berlin Heidelberg: Springer, 2007, pp.355-363.

[20] A.N. Gorban, N.R. Sumner, A.Y. Zinovyev. "Topological grammars for data approximation." *Applied Mathematics Letters* vol.20(4), pp.382-386, 2007. Available: http://arxiv.org/abs/cs/0603090.

[21] A.N. Gorban, B. Kegl, D.C. Wunch, A. Zinovyev (eds.) *Principal Manifolds for Data Visualisation and Dimension Reduction*, Berlin Heidelberg: Springer, 2008.

[22] A.N. Gorban and A.Y. Zinovyev. "Elastic Maps and Nets for Approximating Principal Manifolds and Their Application to Microarray Data Visualization" in *Principal Manifolds for Data Visualisation and Dimension Reduction. Lecture Notes in Computational Science and Engineering*, vol.58, A.N. Gorban et al, Eds., Berlin Heidelberg: Springer, 2008, pp.97-128. Available: http://arxiv.org/abs/0801.0168.

[23] A.N. Gorban, N.R. Sumner, A.Y. Zinovyev. "Beyond The Concept of Manifolds: Principal Trees, Metro Maps, and Elastic Cubic Complexes" in *Principal Manifolds for Data Visualisation and Dimension Reduction. Lecture Notes in Computational Science and Engineering*, vol.58, A.N. Gorban et al, Eds., Berlin Heidelberg: Springer, 2008, pp.223-240. Available:http://arxiv.org/abs/0801.0176.

[24] A.N. Gorban and A.Y. Zinovyev. "Principal Graphs and Manifolds" in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques* E.S. Olivas et al, Eds., Hershey, PA: IGI Global, Information Science Reference, 2009, pp. 28-59. Availalble: http://arxiv.org/abs/0809.0490.

[25] M. Resta "Portfolio optimization through elastic maps: Some evidence from the Italian stock exchange" in *Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Sci-*

ence, vol.4693, B. Apolloni et al, Eds. BerlinHeidelberg:Springer, 2010, pp.635-641.

[26] A.N. Gorban, A. Zinovyev "Principal manifolds and graphs in practice: from molecular biology to dynamical systems." *Int J Neural Syst* vol.20(3), pp.219-32, 2010. Available: http://arxiv.org/abs/1001.1122.

[27] A. Zinovyev, A.N. Gorban. "Nonlinear quality of life index." arXiv preprint, arXiv:1008.4063, 2010. Available: http://arxiv.org/abs/1008.4063.

[28] A. Zinovyev. "Dealing with complexity of biological systems: from data to models." HDR thesis, ENS, Paris, France, 2014. Available: http://arxiv.org/abs/1404.1626.

[29] A. Zinovyev. "Overcoming Complexity of Biological Systems: from Data Analysis to Mathematical Modeling." *Mathematical Modelling of Natural Phenomena*, vol.10(3), pp.186-205, 2015. Available: http://www.mmnp-journal.org/articles/mmnp/abs/2015/03/mmnp201510p186/mmnp201510p186.html.

[30] K.R. Gabriel. "The biplot graphic display of matrices with application to principal component analysis." *Biometrika* vol.58(3), pp.453-467, 1971.

[31] P. Cortez and A. Morais. "A Data Mining Approach to Predict Forest Fires using Meteorological Data," in *New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal*, 2007, pp.512-523.

[32] Ch.-G. Li, X. Mei, B.-G Hu. "Unsupervised ranking of multi-attribute objects based on principal curves." arXiv preprint, arXiv:1402.4542, 2014. Available: http://arxiv.org/abs/1402.4542.

[33] Ch.-G.Li., X. Mei, B.-G. Hu. "Two-phase attribute ordering of unsupervised ranking of multi-attribute objects," in *IEEE International conference on Data Mining Workshop proceedings*, 2014.

[34] A. Zinovyev. "Data visualization in political and social sciences," in *International Encyclopedia of Political Science* Badie et al, Eds., Thousand Oaks, California: SAGE Publications, 2011. Availalble: http://arxiv.org/abs/1008.1188.

[35] A. Zinovyev and E. Mirkes. "Data complexity measured by principal graphs." *Computers and Mathematics with Applications* vol.65, pp.1471-1482, 2013. Availalbe: http://arxiv.org/abs/1212.5841.

[36] E.M. Mirkes, A. Zinovyev, A.N. Gorban. "Geometrical complexity of data approximators," in *Proceedings of IWANN-2013, Puerto de la Cruz, Advances in Computation Intelligence, Lecture Notes in Computer Science*, 2013, vol.7902, pp.500-509. Available: http://arxiv.org/abs/1302.2645.

[37] A. Zinovyev, U. Kairov, T. Karpenyuk, E. Ramanculov. "Blind Source Separation Methods For Deconvolution Of Complex Signals In Cancer Biology." *Biochemical and Biophysical Research Communications* vol.430(3), pp.1182-1187, 2013. Available: http://arxiv.org/abs/1301.2634.

[38] H. Failmezger, B. Jaegle, A. Schrader, M. Hlskamp, A. Tresch. "Semi-automated 3D leaf reconstruction and analysis of trichome patterning from light microscopic images." *PLoS Computational Biology* vol.9(4),e1003029.

[39] H. Shaban, S. Tavoulari "Identification of flow regime in vertical upward airwater pipe flow using differential pressure signals and elastic maps." *International Journal of Multiphase Flow* vol.61, pp.62-72, 2014.

[40] O. Radulescu, A. Zinovyev, A. Lilienbaum. "Model reduction and model comparison for NFkB signalling," in *In Proceedings of Foundations of Systems Biology in Engineering, Stuttgart, Germany*, 2007.

[41] L. Calzone, L. Tournier, S. Fourquet, D. Thieffry, B. Zhivotovsky, E. Barillot, A. Zinovyev. "Mathematical modelling of cell-fate decision in response to death receptor engagement." *PLoS Computational Biology*, vol.6(3),e1000702, 2010.

[42] L. Calzone, E. Barillot, A. Zinovyev. "Predicting genetic interactions from Boolean models of biological networks." *Integrative Biology*, vol.7(8), pp.921-929, 2015.

[43] U. Czerwinska, L. Calzone, E. Barillot, A. Zinovyev. "DeDaL: Cytoscape 3 app for producing and morphing data-driven and structure-driven network layouts." *BMC Systems Biology* vol.9, 46, 2015.