

# Visualization and 3D Printing of Multivariate Data of Biomarkers

Michael C. Thrun	Florian Lerch	Jörn Lötsch <sup>1</sup>	Alfred Ultsch
DataBionics, University of Marburg, Hans-Meerwein Str., 35032 Marburg, Germany	DataBionics University of Marburg, Hans-Meerwein Str., 35032 Marburg, Germany	Institute of Clinical Pharmacology, Goethe - University, Theodor Stern Kai 7, 60590 Frankfurt am Main, Germany	DataBionics University of Marburg, Hans-Meerwein Str., 35032 Marburg, Germany
mthrun@informatik.uni- marburg.de	lerchf@students.uni- marburg.de	j.loetsch@em.uni- marburg.de	ultsch@mathematik.uni- marburg.

(1) Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

## ABSTRACT

Dimensionality reduction by feature extraction is commonly used to project high-dimensional data into a low-dimensional space. With the aim to create a visualization of data, only projections onto two dimensions are considered here. Self-organizing maps were chosen as the projection method, which enabled the use of the U\*-Matrix as an established method to visualize data as landscapes. Owing to the availability of the 3D printing technique, this allows presenting the structure of data in an intuitive way. For this purpose, information about the height of the landscapes is used to produce a three dimensional landscape with a 3D color printer. Similarities between high-dimensional data are observed as valleys and dissimilarities as mountains or ridges. These 3D prints provide topical experts a haptic grasp of high-dimensional structures. The method will be exemplarily demonstrated on multivariate data comprising pain-related bio responses. In addition, a new R package “Umatrix” is introduced that allows the user to generate landscapes with hypsometric tints.

## Keywords

Self-Organizing Map (SOM), Multivariate Data Visualization, Dimensionality Reduction, High Dimensional Data, 3D Printing, U-Matrix.

## 1. Introduction

Some large data sets possess a high number of variables with a low number of observations. Projection methods reduce the dimension of the data and try to represent structures present in the high dimensional space. If the projected data is two dimensional, the positions of projected points do not represent high-dimensional distances. Therefore, low dimensional similarities could lead to incorrect interpretations of the underlying structures.

A certain solution for this problem is the self-organizing map (SOM) [Kohonen, 1982] with high number of neurons used as a projection method

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

[Ultsch, 1999]. SOM is an unsupervised neural learning algorithm. If used as a projection method, the picture of high-dimensional data is uniformly distributed on the neural grid. This distribution makes a direct interpretation demanding. The standard approach for this problem lies in generating a 2D visualization for SOM, because, for high-dimensional data, the SOM remains a reference tool for 2D visualizations [Lee/Verleysen, 2007, p. 227]. In literature, there are many approaches, which require experienced interpretations (e.g.[Kadim Tasdemir/Merényi, 2012; Vesanto/Alhoniemi, 2000]). Here, we focus on the method of U\*matrix, which is able to visualize distance and density based structures. The U\*matrix leads to a topographic map with hypsometric tints (for details see section 5), which seems like a 3D landscape for the human eye. But every 3D visualization still has to be viewed from multiple viewpoints and is often subject to serious occlusion, distortion and navigation issues [Jansen et al., 2013] cites [Shneiderman, 2003]. But [Jansen et al., 2013] showed that physical

visualizations can improve the user's efficiency at information retrieval tasks, because physical touch seems to be an essential cognitive aid. Three-dimensional printing addresses this important point through generating a haptic form. To facilitate this visualization of high-dimensional data for experts in the data's field, we propose the usage of colored three-dimensional printing.

3D printing is currently a quickly evolving technique. It represents a technical change from spraying toner on paper to adding up layers of materials to a 3D object [Sachs et al., 1993]. By enabling a machine to produce objects of any shape it has the potential to impact many production areas [D'aveni, 2013]. Main biomedical applications were so far 3D printing vascular implants, aerosol delivery technologies, cellular transplantation, endo-prosthetics, tissue engineering, biomedical device development and pharmacology including techniques such as individualized drug delivery formulations [Pillay/Choonara, 2015]. 3D printing is also employed for the visualization of biomedical data, for example to produce graspable three-dimensional objects for surgical planning [Rengier et al., 2010].

This work proposes the application of 3D printing to the enhancement of knowledge discovery in high-dimensional data transferring them into 3D haptic physical models with the goal of physical grasping a visualization of projections.

The results are shown using the example of pain data. Blue and green valleys indicate clusters of pain types and the brown or white watersheds of the U\*matrix point to borderlines of clusters (Fig. 4). Other SOM visualizations fail to display the information in an easily understandable form and do not allow the usage of 3D printing (see section 3).

We enable the user to achieve every step until the 3D printing using software: The tasks of SOM generation, visualization and supervised clustering can be performed interactively by the R package Umatrix [Version 2.0.0; Thrun et al., 2016]. The package also enables the usage of other SOM algorithms or comparing classifications with the U\*matrix visualization.

## 2. Emergent SOM

The first step for structure visualization is to project high-dimensional data in a two dimensional space. One approach is using self-organizing maps (SOM), which project to a fixed grid of neurons. Originally, the SOM algorithm was introduced by [Kohonen, 1982]. However, to exploit emergent phenomena in SOMs [Utsch, 1999] argued to use a large number of neurons (at least  $n = 4000$ ). The self-organization of many neurons allows emergent structures to occur in data. By gaining the property of emergence through

self-organization this enhancement of SOM is called Emergent SOM (ESOM).

Let  $M = \{m_1, \dots, m_n\}$  be the positions of neurons on a two dimensional grid (map) and  $W = \{w(m_i) = w_i | i = 1, \dots, n\}$  the corresponding set of weights or prototypes of neurons, then the SOM learning algorithm constructs a nonlinear and topology preserving projection of the input space  $I$  by finding the bestmatching unit (BMU):

$$BMU(l) = \underset{m_i \in M}{\operatorname{argmin}} \{D(l, w_i)\}, \quad i \in \{1, \dots, n\} \quad (1)$$

$\forall l \in I$ , if  $D$  denotes a distance between input space  $I$ . Hence, the location of a given data point on the resulting map is depicted by the corresponding BMU. The topology of the map is toroid if the borders are cyclically connected [Utsch, 1999]. If the map was planar, the neighborhood of neurons at the edges would contain much less neurons compared to the middle of the map space. This would lead to undesired seam effects in the SOM algorithm [Utsch, 2003a].

In each step the SOM learning is achieved by modifying the weights in a neighborhood with

$$\Delta w(R) = \eta(R) * h(BMU(l), m_i, R) * (l - w(m_i)) \quad (2).$$

The cooling scheme is defined by the neighborhood function  $h: M \times M \times \mathbb{R}^+ \rightarrow [-1, 1]$  and the learning rate  $\eta: \mathbb{R}^+ \rightarrow [0, 1]$ , where the radius  $R$  declines until  $R = 1$  through the definition of the maximum number of epochs.

## 3. Other visualizations of SOMs

The result of Kohonen SOM algorithm are neurons, which are located on a map with a set  $W$  of prototypes corresponding to a set  $M$  of positions. In general, the positions on  $M$  are restricted to a grid, but a few approaches exist which change the positions in  $M$ , like Adaptive Coordinates [Merkl/Rauber, 1997]. Because these approaches are not based on a grid, they are not considered further.

BMUs define locations of input points on the map. However, they exhibit no structure of the input space for a SOM [Utsch, 1999]. But the goal is to grasp the structure of the high dimensional data and maybe even visualize cluster boundaries. Therefore, post-processing of the neurons is required for an informative representation of high dimensional data. Three standard approaches are found in literature:

The first approach projects the prototypes of the set  $W$  with Multidimensional Scaling (MDS) [Torgerson, 1952] or some of its variants to a two dimensional space [Kaski et al., 2000; Sarlin/Rönnqvist, 2013]. The result is mapped into the CIE Lab color space [Colorimetry, 2004]. This uniform color space is defined so that perceptual differences in colors

correspond to Euclidean distances in the map space as well as possible [Kaski et al., 2000]. The next two approaches visualize either distances or density of the prototypes.

The second approach defines receptive fields around each position in  $M$ . The unified distance matrix (Umatrix) [Ultsch/Siemon, 1990] or variants [Kraaijveld et al., 1995] [Häkkinen/Koikkalainen, 1997] [Hamel/Brown, 2011] represent distances of prototypes (see section 4 for details) by using proportional intensities of gray shades, color hues, shape or size. In [Kraaijveld et al., 1995] every neuron corresponds to a pixel. The gray value of each pixel is determined by the maximum unit distance from the neuron to its four neighbors (up, down, left, right). The larger the distance, the lighter the gray value. In [Häkkinen/Koikkalainen, 1997] additional visualization approaches for unit distances are explained. The shape and size of the receptive fields describe the dissimilarity of the corresponding neurons. Apart from the U-matrix, visualizations of receptive fields in three dimensions or specific components of prototypes with receptive fields in two dimensions were tried [Vesanto, 1999]. Also, SOM quality measures can be added to the receptive fields in a third dimension, e.g. [Vesanto et al., 1998].

The third approach connects the positions  $M$  by way of a specific scheme. In [Hamel/Brown, 2011] additional to a U-matrix neurons are connected with lines along the maximum gradient. The authors claim that clusters are the always connected components of the graph defined by the Umatrix .

[Merkl/Rauber, 1997] omitted the receptive fields approach by only connecting map positions with lines, where the intensity of the connections reflects the similarity of the underlying prototypes. [K. Tasdemir/Merenyi, 2009] proposed the CONNvis technique, which visualizes the grid by connecting the neurons, whose corresponding prototypes are adjacent in the space of input dimensionality, which is equal to the high dimensional data. The width of the connection line is proportional to the strength of the connection [K. Tasdemir/Merenyi, 2009].

In sum, all visualizations of large SOMs described above require an expert in the field for interpretation. In addition, a 3D print may not give a desirable result: in most cases the 2D visualization would have to be enhanced to 3D. But research indicates that 3D does not improve 2D visualizations [Cockburn, 2004; Cockburn/McKenzie, 2002; Sebrechts et al., 1999], and, to our knowledge, there are no 3D visualizations of ESOMs based on a 2D grid currently in use, besides the approach proposed in section 5.

#### 4. U\*matrix based on data distances and density

The Umatrix displays a folding of high-dimensional space, where each receptive field is called a U-height. Let  $N(j)$  be the eight immediate neighbors of  $m_j \in M$ , let  $w_j \in W$  be the corresponding prototype to  $m_j$ , then the average of all distances between prototypes  $w_i$  is called U-height regarding the position  $m_j$ :

$$u(j) = \frac{1}{n} \sum_{i \in N(j)} D(w_i, w_j), \quad n = |N(j)| \quad (3).$$

The Umatrix is a display of proportional intensities of grey shades of all receptive fields [Ultsch, 2003a]. By formalizing the displayed structures [Löttsch/Ultsch, 2014] showed that the Umatrix is an approximation of Voronoi borders of the high-dimensional points in the output space:

Let  $bmu(l)$  and  $bmu(j)$  be BMUs of data points  $l$  and  $j$ , where  $bmu(j)$  and  $bmu(l)$  have bordering Voronoi cells. On the borderline there is a vertical plane (AU-height), which is the distance  $D(l, j) > 0$  between the data points in the input space. In sum, the abstract Umatrix, (AU-matrix) is the Delaunay graph of the BMU's weighted by corresponding Euclidean distances in the input space.

In addition to the Umatrix, [Ultsch, 2003a] introduced the high-dimensional density visualization technique called P-Matrix, where P-heights on top of the receptive fields are displayed. The P-height  $p(m_i)$  for a position  $m_i$  is a measure of the density of data points in the vicinity of  $w(m_j)$ :

$$p(m_j) = |\{i \in I | D(i, w(m_j)) < r > 0, r \in \mathbb{R}\}| \quad (4).$$

The P-height is the number of data points within a hypersphere of radius  $r$ . Here, we choose the interval  $\varrho$  of the radius with

$$\varrho \in [\text{median}(C(D)), \text{median}(A(D))], \quad (5)$$

where  $D$  are all input space distances and  $A(D)$  is the group  $A$  of distances calculated by the ABCanalysis [Ultsch/Löttsch, 2015]. ABCanalysis tries to identify the optimum information that can be validly retrieved by using concepts developed in economical sciences. In particular, concepts are used in the search for a minimum possible effort that gives the maximum yield [Ultsch/Löttsch, 2015]. The distances are divided into three disjoint subsets  $A$ ,  $B$  and  $C$ , with subset  $A$  comprising largest values ("outer cluster distances"), subset  $B$  comprising values where the yield equals the effort required to obtain it, and the subset  $C$  comprising of the smallest values ("inner cluster distances"). We suggest the choice for the specific radius  $r$  through the proportion  $v$  of inter-versus intra-cluster distances estimated by

$$v = \frac{\max(C(D))}{\min(A(D))} \quad (6).$$

The radius  $r$  is estimated by  $r = v * p20(D)$ , where  $p20(D)$  is 20-th percentile of input distances [Ultsch, 2003b]. From this starting point the user may search interactively for the empirical Pareto percentile, which defines the radius  $r$  (see R package Umatrix).

The combination of a Umatrix and a Pmatrix is called U\*matrix [Ultsch et al., 2016]: It can be formalized as pointwise matrix multiplication:  $U^* = U * F(P)$ , where  $F(P)$  is a matrix of factors  $f(p)$  that are determined through a linear function  $f$  on the  $P$  heights  $p$  of the Pmatrix. The function  $f$  is calculated so that  $f(p) = 1$  if  $p$  is equal to the median and  $f(p) = 0$  if  $p$  is equal to the 95-percentile ( $p95$ ) of the heights in the Pmatrix. For  $p(j) > p95$ :  $f(p) = 0$ , which indicates that  $j$  is well within a cluster and results in zero heights in the U\*matrix.

## 5. Visualization as a 3D landscape

We concur with [Koikkalainen, 1997] that the content of information should be displayed in an understandable way. Hence, in the following section we formalize the idea of [Ultsch, 2003a] to visualize the U\*matrix as a landscape. We define a topographic map with hypsometric tints [Patterson/Kelso, 2004]. Hypsometric tints are surface colors which depict ranges of elevation. Here, a specific color scale is combined with contour lines.

The color scale is chosen to display various valleys, ridges and basins: blue colors indicate small distances (sea level), green and brown colors indicate middle distances (small hilly country) and white colors indicate high distances (snow and ice of high mountains). The valleys and basins indicate clusters and the watersheds of hills and mountains indicate borderlines of clusters (Fig. 1 and Fig. 4).

The landscape consists of receptive fields, which correspond to intervals of U\*heights edged by contours. This paper proposes the following approach: First, the range of U\*heights is assigned uniformly and continuously to the specific color scale above by robust normalization [Milligan/Cooper, 1988] and by splitting it up into intervals. In the next step, the color scale is interpolated by the corresponding CIE Lab colors space [Colorimetry, 2004]. The largest possible contiguous areas of receptive fields, which are in the same U\*height interval, are summarized and outlined in black as a contour. In sum, a receptive field is the display of one color in one particular place of the U\*matrix visualization within a height dependent contour. Let  $u(j)$  be the U\*height,  $q01$  the one-percentile and  $q99$  the 99-percentile of U\*heights, then the robust normalization of U\*heights  $u(j)$  is defined by

$$u(j) = \frac{u(j)-q01}{q99-q01} \quad (7).$$

The number of intervals  $in$  is defined by

$$\frac{1}{in} = \frac{q01}{q99}. \quad (8).$$

The resulting visualization consists of a hierarchy of areas of different height levels with corresponding colors (see Fig. 4). The visualization of SOMs using the tool Umatrix is consistent with a 3D landscape for the human eye, therefore one sees data structures intuitively. Contrary to other SOM visualizations, e.g. [K. Tasdemir/Merenyi, 2009], the 3D landscape enables layman to interpret the results of a SOM.

Using a toroid map for the ESOM computation requires a tiled display of the landscape in the interactive tool Umatrix [Version 2.0.0; Thrun et al., 2016] which means that every receptive field is shown four times. So in the first step the visualization consists of four adjoining pictures of the same Umatrix [Ultsch, 2003a] (the same for the U\*matrix after loading of a SOM or computing one). To get the 3D landscape this paper proposes to cut the tiled U\*matrix visualization rectangular:

Let  $v_{Lines}$  be the vector of row sums,  $v_{Columns}$  be the vector of column sums of the U\*heights and let  $b_{Lines}$  be the number of BMU's of the corresponding row line of  $v_{Lines}$  (for  $b_{Columns}$ ,  $v_{Columns}$ ), then we define the upper border  $up = \max(v_{Lines}/f(b_{Lines}))$ , the left border by  $lb = \max(b_{Columns}/f(v_{Columns}))$  and the other two borders by the length and width of the U\*matrix, if the vector  $f(b)$  is the addition  $f(b) = \hat{b} + b + \check{b}$  with  $\hat{b} = (b_n, b_1, \dots, b_{n-1})$  and  $\check{b} = (b_2, \dots, b_{n+1})$ , where the grid is toroid. For better comprehensibility see the axes in Fig 1, which are defined from one to  $max(Lines)$  and from one to  $max(Columns)$ .

## 6. 3D Printing of pain phenotypes

3D landscapes can be better grasped when viewed from multiple perspectives. This can be easily achieved with a haptic form. As an example of a haptic 3D presentation of biomedical data, complex pain phenotypes composed of responses to four different types of nociceptive stimuli are used. Nociceptive stimuli activate nociceptors, which are sensory nerve cells responding to pressure (mechanic), electric, cold or heat. Data was acquired with the help of 206 healthy volunteers as described previously in detail [Flühr et al., 2009; Lötsch/Ultsch, 2013; Neddermeyer et al., 2008]. Data was projected using the ESOM algorithm and clusters were identified by interpreting its U\*matrix visualization (Fig 1). In a last step, pain sub-phenotypes were identified by interpreting the clusters using classification and regression tree classifiers (Cart) [Lötsch/Ultsch, 2013]. By way of extracting decision rules through the conditional information of the GINI impurity [Hill et al., 2006], the interpretation based on measured stimulus intensities evoking pain at threshold level. Eight

different pain phenotypes were observed, involving individuals who shared complex pain threshold patterns across five variables. Subsequently, the specific properties of each phenotype could be interpreted clinically. Three main pain sensitivity groups were identified: high-pain sensitivity (HPS), average pain sensitivity (APS) and low-pain sensitivity (LPS) [Lötsch/Ultsch, 2013]. HPS was divided into two clusters (1,2), APS into four (3-6) and LPS into two (7,8). All clusters were interpretable (further details see [Lötsch/Ultsch, 2013]). From this data set, a 3D Landscape could be generated (Fig. 1 top view and Fig. 4) and printed by means of a 3D color printer (Fig. 2). Due to technical limitations, printing is restricted to three colors blue, green and white, while the digital 3D landscape consists of many more different height dependent colors (Fig. 2. and Fig. 4). On the other hand, contrary to Fig 1, Fig. 4 had to be reworked manually by using a graphics editor program. Otherwise the structures on the borders of the island would be difficult to interpret. Note, that the 3D print of Fig. 2 was generated using Fig. 1 and not Fig 4.

Data processing was done using the interactive tool Umatrix [Version 2.0.0; Thrun et al., 2016] with the freely available R software [Version 3.2.5; R Development Core Team, 2008] for Windows 7 64bit, and the graphical interface by the open source web application framework shiny [Version 0.13.2; RStudio, 2014]. To our knowledge, the 3D print of an U\*matrix is the first application of 3D printing techniques used directly for data mining and knowledge discovery in high-dimensional data in a haptic form. In addition, the political map of the eight clusters is shown in Fig 3. The political map of an ESOM is the coloring of the Voronoi cells of the BMUs with different colors for each cluster [Lötsch/Ultsch, 2014].

## 7. Summary

Projection methods visualize the structures of high-dimensional data in a low-dimensional space. The unsupervised neural learning algorithm, which is called self-organizing map (SOM), may be used as a non-linear projection method. In that case SOM projects high-dimensional data onto a two dimensional grid, where the positions of projected points do not represent high-dimensional distances. The standard approach to this problem is the generation of a visualization for SOM. Because common SOM visualizations fail to display the information in an easily understandable form and do not allow the usage of 3D printing, we combined a large SOM with the U\*matrix visualization technique. The U\*matrix is able to visualize distance and density based structures. This 3D visualization is a topographic map with hypsometric tints and representable as a 3D landscape. The details of

creating the 3D landscape were introduced in the paper in section 5. The tasks of SOM generation, visualization and supervised clustering can be performed interactively by the published R package Umatrix [Version 2.0.0; Thrun et al., 2016]. We allow the user to choose a different SOM based projection method, on which our visualization techniques still can be used. The package also enables comparing of classifications to the U\*matrix visualization.

The main step forward presented in this paper is the color 3D printing of landscapes based on the visualization originating from the U\*matrix. Through its haptic form, the 3D print makes high-dimensional structures more understandable for experts in the data's field. Structural features of high-dimensional data were depicted with the use of 3D printing (Fig 2) and pain data. Blue and green valleys indicate clusters of pain types and the brown or white watersheds of the U\*matrix visualization point to borderlines of clusters. In our opinion, the task of height depending 3D color printing is still very trying. Automatically cutting a non-rectangular island defined by curved borders remains also an unsolved problem.

To our knowledge, this 3D print is the first application of 3D printing techniques used directly for data mining and knowledge discovery in high-dimensional data in a haptic form.

Future work will include the abstract U\*matrix [Ultsch et al., 2016] into the current visualization techniques and allow the height dependent 3D print of an U\*matrix in more than three colors.

## 8. Acknowledgments

This work has been funded by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (JL), by the German Research Foundation (DFG) under grant agreement (BE4234/3-1, UL159/10-1), and by the Else Kröner-Fresenius Foundation (EKFS), Research Training Group Translational Research Innovation – Pharma (TRIP, JL). Special acknowledgment goes to the 3D printing by Michael Weingart, Weingart Ingenieur-Büro + CNC Fräsen, Kirchheim-Teck, Germany for the practical 3D print and the consistent coloring of the print.

## 9. References

- Cockburn, A.: Revisiting 2D vs 3D implications on spatial memory, Proc. Proceedings of the fifth conference on Australasian user interface-Volume 28, pp. 25-31, Australian Computer Society, Inc., 2004.
- Cockburn, A., & McKenzie, B.: Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments, Proc.

- Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 203-210, ACM, 2002.
- Colorimetry.C.I.E., Vol. CIE Publication, Central Bureau of the CIE, Vienna, 2004.
- D'aveni, R. A.: 3-D printing will change the world, *Harvard business review*, Vol. 91(3), pp. 34-35. 2013.
- Flühr, K., Neddermeyer, T. J., & Lötsch, J.: Capsaicin or menthol sensitization induces quantitative but no qualitative changes to thermal and mechanical pain thresholds, *The Clinical journal of pain*, Vol. 25(2), pp. 128-131. 2009.
- Häkkinen, E., & Koikkalainen, P.: SOM based visualization in data analysis, *Artificial Neural Networks—ICANN'97*, (pp. 601-606), Springer, 1997.
- Hamel, L., & Brown, C. W.: Improved interpretability of the unified distance matrix with connected components, *Proc. 7th International Conference on Data Mining (DMIN'11)*, pp. 338-343, 2011.
- Hill, T., Lewicki, P., & Lewicki, P.: *Statistics: methods and applications: a comprehensive reference for science, industry, and data mining*, StatSoft, Inc., 2006.
- Jansen, Y., Dragicevic, P., & Fekete, J.-D.: Evaluating the efficiency of physical visualizations, *Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2593-2602, ACM, 2013.
- Kaski, S., Venna, J., & Kohonen, T.: Coloring that reveals cluster structures in multivariate data, *Australian Journal of Intelligent Information Processing Systems*, Vol. 6(2), pp. 82-88. 2000.
- Kohonen, T.: Self-organized formation of topologically correct feature maps, *Biological cybernetics*, Vol. 43(1), pp. 59-69. 1982.
- Koikkalainen, E. H. P.: The neural data analysis environment, *Proceedings of the Workshop on Self-Organizing Maps Map*, Vol., pp. 69-74. 1997.
- Kraaijveld, M., Mao, J., & Jain, A. K.: A nonlinear projection method based on Kohonen's topology preserving maps, *Neural Networks, IEEE Transactions on*, Vol. 6(3), pp. 548-559. 1995.
- Lee, J. A., & Verleysen, M.: *Nonlinear dimensionality reduction*, New York, USA, Springer, 2007.
- Lötsch, J., & Ultsch, A.: A machine-learned knowledge discovery method for associating complex phenotypes with complex genotypes. Application to pain, *Journal of biomedical informatics*, Vol. 46(5), pp. 921-928. 2013.
- Lötsch, J., & Ultsch, A.: Exploiting the Structures of the U-Matrix, in Villmann, T., Schleif, F.-M., Kaden, M. & Lange, M. (eds.), *Proc. Advances in Self-Organizing Maps and Learning Vector Quantization*, pp. 249-257, Springer International Publishing, Mittweida, Germany, 2014.
- Merkel, D., & Rauber, A.: Alternative ways for cluster visualization in self-organizing maps, *Proc. Proc. of the Workshop on Self-Organizing Maps (WSOM97)*, pp. 4-6, Citeseer, 1997.
- Milligan, G. W., & Cooper, M. C.: A study of standardization of variables in cluster analysis, *Journal of Classification*, Vol. 5(2), pp. 181-204. 1988.
- Neddermeyer, T. J., Flühr, K., & Lötsch, J.: Principle components analysis of pain thresholds to thermal, electrical, and mechanical stimuli suggests a predominant common source of variance, *Pain*, Vol. 138(2), pp. 286-291. 2008.
- Patterson, T., & Kelso, N. V.: Hal Shelton revisited: Designing and producing natural-color maps with satellite land cover data, *Cartographic Perspectives*, Vol. (47), pp. 28-55. 2004.
- Pillay, V., & Choonara, Y.: 3D Printing in Drug Delivery Formulation: You Can Dream it, Design it and Print it. How About Patent it?, *Recent patents on drug delivery & formulation*, Vol., pp., 2015.
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing (Version 3.2.5)*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Rengier, F., Mehndiratta, A., von Tengg-Kobligk, H., Zechmann, C. M., Unterhinninghofen, R., Kauczor, H.-U., & Giesel, F. L.: 3D printing based on imaging data: review of medical applications, *International journal of computer assisted radiology and surgery*, Vol. 5(4), pp. 335-341. 2010.
- RStudio, I. (2014). *shiny: Easy web applications in R (Version 0.13.2)*. Retrieved from <http://shiny.rstudio.com/>
- Sachs, E., Cima, M., Cornie, J., Brancazio, D., Bredt, J., Curodeau, A., . . . Lee, J.: Three-dimensional printing: the physics and implications of additive manufacturing, *CIRP Annals-Manufacturing Technology*, Vol. 42(1), pp. 257-260. 1993.
- Sarlin, P., & Rönnqvist, S.: Cluster coloring of the Self-Organizing Map: An information visualization perspective, *arXiv preprint arXiv:1306.3860*, Vol., pp., 2013.
- Sebrechts, M. M., Cugini, J. V., Laskowski, S. J., Vasilakis, J., & Miller, M. S.: Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces, *Proc. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 3-10, ACM, 1999.
- Shneiderman, B.: Why not make interfaces better than 3D reality?, *Computer Graphics and Applications*, IEEE, Vol. 23(6), pp. 12-15. 2003.
- Tasdemir, K., & Merenyi, E.: Exploiting Data Topology in Visualization and Clustering of Self-Organizing Maps, *IEEE Transactions on Neural Networks*, Vol. 20(4), pp. 549-562. doi 10.1109/tnn.2008.2005409, 2009.
- Tasdemir, K., & Merényi, E.: SOM-based topology visualisation for interactive analysis of high-dimensional large datasets, *Machine Learning Reports*, Vol. 1, pp. 13-15. 2012.
- Thrun, M. C., Lerch, F., & Ultsch, A. (2016). *Umatrix (Version 2.0.0)*. Marburg. R package, requires CRAN packages: Rcpp, ggplot2, shiny, ABCanalysis, shinyjs, reshape2, fields, plyr, abind, tcltk, png, tools, grid, rgl. Retrieved from

- www.uni-marburg.de/fb12/datenbionik/software-en
- Torgerson, W. S.: Multidimensional scaling: I. Theory and method, *Psychometrika*, Vol. 17(4), pp. 401-419. 1952.
- Ultsch, A.: Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series, In Oja, E. & Kaski, S. (Eds.), *Kohonen maps*, (1 ed., pp. 33-46), Elsevier, 1999.
- Ultsch, A.: Maps for the visualization of high-dimensional data spaces, *Proc. Workshop on Self organizing Maps (WSOM)*, pp. 225-230, Kyushu, Japan, 2003a.
- Ultsch, A. Optimal density estimation in data containing clusters of unknown structure, technical report, Vol. 34, University of Marburg, Department of Mathematics and Computer Science, 2003b.
- Ultsch, A., Behnisch, M., & Lötsch, J.: ESOM Visualizations for Quality Assessment in Clustering, In Merényi, E., Mendenhall, J. M. & O'Driscoll, P. (Eds.), *Advances in Self-Organizing Maps and Learning Vector Quantization: Proceedings of the 11th International Workshop WSOM 2016*, Houston, Texas, USA, January 6-8, 2016, (10.1007/978-3-319-28518-4\_3pp. 39-48), Cham, Springer International Publishing, 2016.
- Ultsch, A., & Lötsch, J.: Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data, *PloS one*, Vol. 10(6), pp. e0129767. doi 10.1371/journal.pone.0129767, 2015.
- Ultsch, A., & Siemon, H. P.: Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis, *International Neural Network Conference*, pp. 305-308, Kluwer Academic Press, Paris, France, 1990.
- Vesanto, J.: SOM-based data visualization methods, *Intelligent data analysis*, Vol. 3(2), pp. 111-126. 1999.
- Vesanto, J., & Alhoniemi, E.: Clustering of the self-organizing map, *Neural Networks, IEEE Transactions on*, Vol. 11(3), pp. 586-600. 2000.
- Vesanto, J., Himberg, J., Siponen, M., & Simula, O.: Enhancing SOM based data visualization, *Proc. Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems. Methodologies for the Conception, Design and Application of Soft Computing*, Vol. 1, pp. 64-67, 1998.

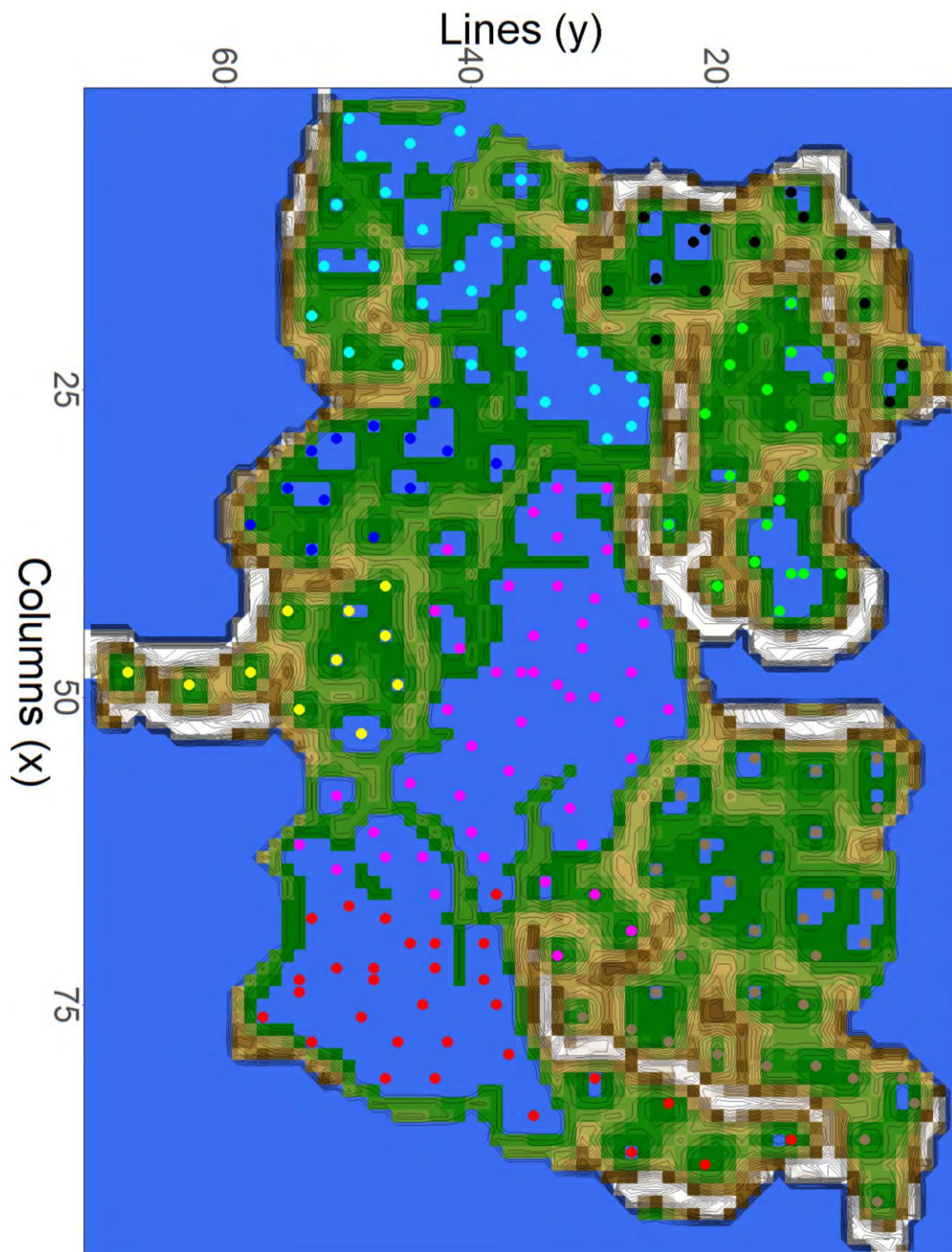


Figure 1: Top view of the 3D landscape of the pain data generated with the Umatrix tool: After the rectangular cut (section 5), the cutting lines of visualization of the  $U^*$ matrix were improved interactively. The points are the BMU's with different colors as cluster labels. The top view was used for 3D printing.



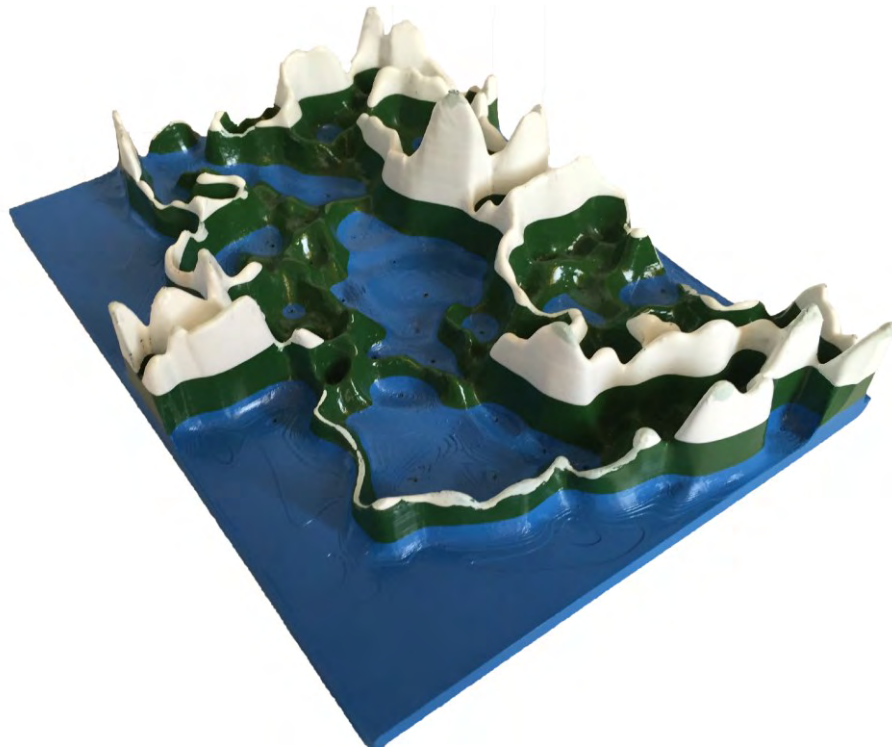


Figure 2: The 3D print of Fig 1 in three height dependent colors white, green and blue. The valleys indicate clusters of pain types and the watersheds of the  $U^*$ -matrix borderlines of clusters.

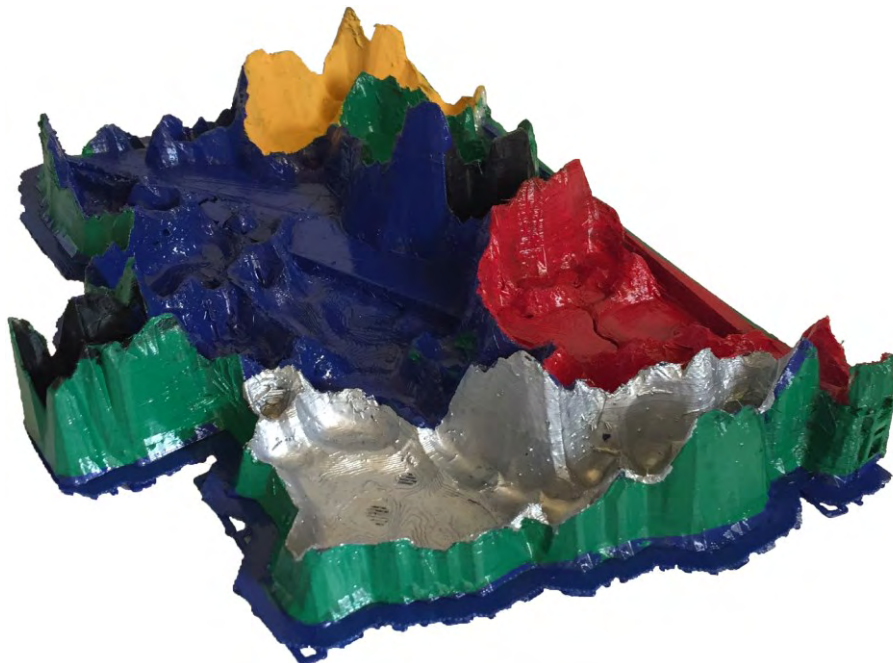


Figure 3:  $AU^*$ -clustering based on the Voronoi cells formalizes the distance and density based structures and leads from Fig 1 to a political map (further details in [Ultsch et al., 2016]). Above the 3D print of this political map is shown. Every color indicates one cluster as described in section 6.

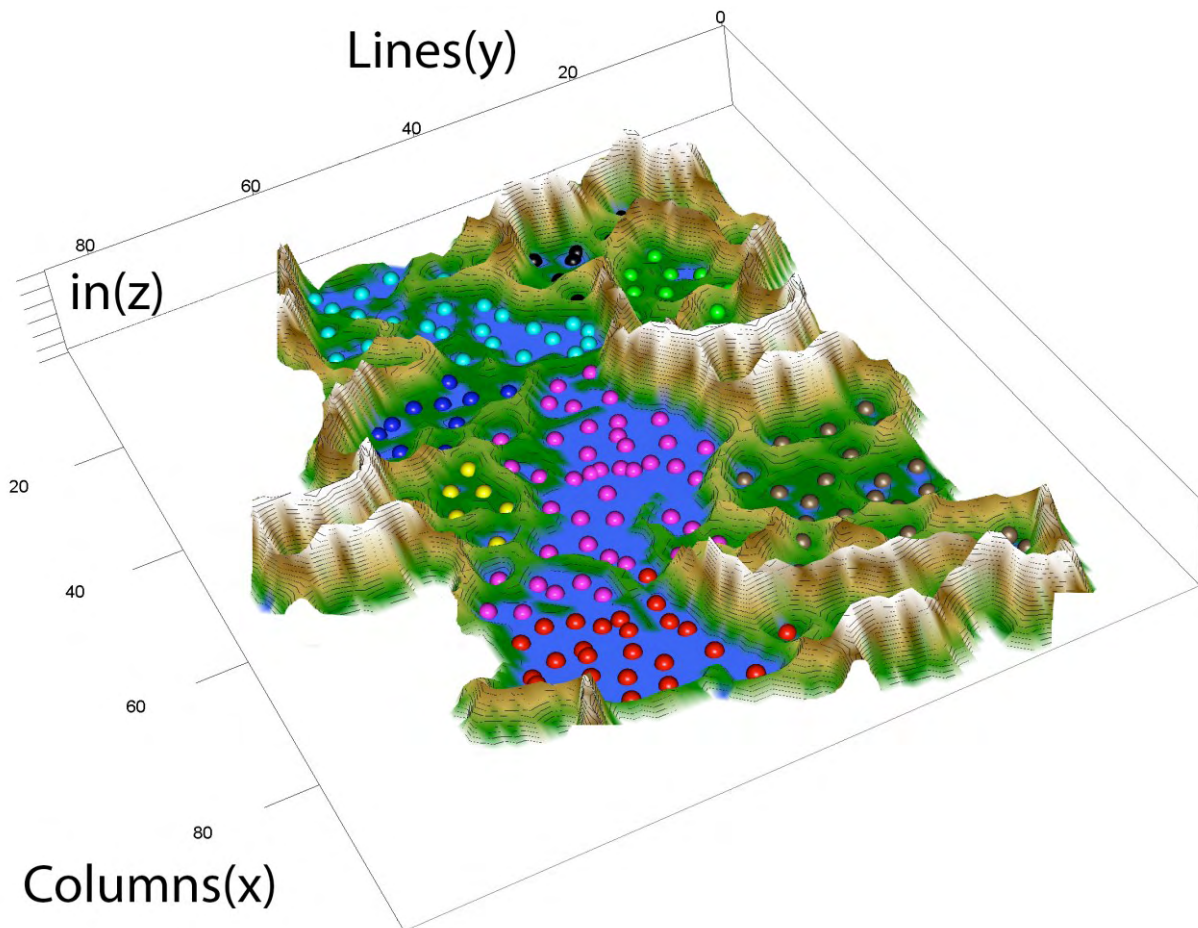


Figure 4: 3D landscape of the pain data generated with the Umatrix tool: After the rectangular cut (section 5), the cutting lines of visualization of the  $U^*$  matrix were improved interactively with shiny in R. The points are the BMU's with different colors as cluster labels. Contrary to Figure 1, the borders around the island had to be reworked manually using graphics editor program afterwards. Otherwise the borders of the island would be difficult to interpret.