

Visual analytics and its use in the NRNU MEPhI “Scientific Visualization” laboratory activities

V.V. Pilyugin^{1,A}, I. Milman^{2,B}

^A National Research Nuclear University MEPhI, Moscow, Russian Federation

^B Individual researcher, USA

¹ ORCID: 0000-0001-8648-1690, VVPilyugin@mephi.ru

² ORCID: 0000-0001-9705-9401, Igal.Milman@gmail.com

Abstract

The article discusses visual analytics, which, according to J. Thomas, is understood as the solution of problems of data analysis using a facilitating interactive visual interface. Today, visual analytics is widely used in various fields — in research, design, management and others, due to the well-known predisposition of people to spatial thinking. Despite the wide usage, according to the authors, the theoretical aspects of visual analytics are not well developed at present, which is certainly a limiting factor in the development of visual analytics tools and its effective use in practice. It is declared in the article, that one of the most common of the forms of visual analytics is solution of data analysis problems using the visualization method. The visualization method is described in details and a number of works done by the “Scientific Visualization” laboratory NRNU MEPhI in this area are presented as an example of usage of the visualization method. The works were performed in different subject areas from physics to finance and monitoring.

Keywords: Visual analytics, data analysis, visualization method, modeling.

1. Visual analytics

The term "Visual Analytics" was introduced by Jim Thomas [1] [2]. He defined visual analytics as solving data analysis problems using a facilitating interactive visual interface. The same term is used to name the scientific discipline about this activity as well.

Currently, visual analytics is widely used in various areas of human activity — in scientific research, design, management, financial monitoring [3], information security [4] and other areas. The key for wide and successful use of visual analytics in various fields and circumstances — is the well-known predisposition of people to spatial thinking and image recognition. [5]

Before reviewing Visual Analytics, a few words about the problem of data analysis in general. We define the data analysis problem as follows:

- There is an object of consideration Ob .
- The analyst knows some initial statements about the object Ob . We call it the “initial data” $G = \{G_i\}$ and it may be presented in one way or another.
- It is required to obtain new statements of interest for the analytic or check the given statements.

In the definition above, the object of consideration (Ob) can be one of the following:

- Either one or several real material objects $Ob = \{r_i\}$.
- Either one or more imaginary material objects $Ob = \{m_i\}$.
- Either one or more abstract objects $Ob = \{s_i\}$.

As follows from the above definition, visual analytics is a very broad conception and today we can talk about various forms of visual analytics. From our point of view, one of the most common forms of visual analytics is the solution of problems of analysis of various data using the visualization method.

2. The Visualization method

The solution of the problem of the initial data analysis consists of the sequential solution of the following two problems (Fig. 1): [6]

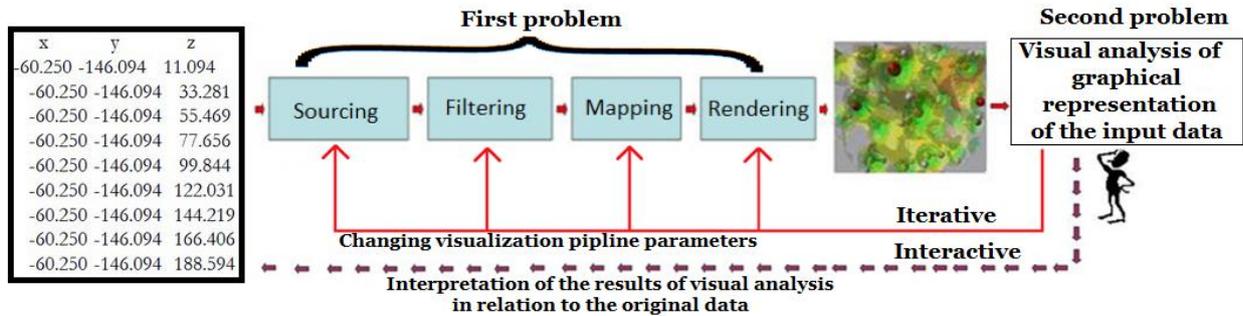


Fig. 1. Visualization method of data analysis

The first problem is the problem of obtaining a representation of the analyzed data as a graphic image (The problem of original data visualization). The problem is solved using the computational power of a computer. The second problem (the problem of obtaining information of interest to analytics), which is nonetheless important, is the visual analysis of the obtained image. The analyst has to work on this problem directly, it cannot be automatized.

2.1. Original data visualization

Solution of this problem involves a well-known sequence of steps. The algorithm of translating the initial data to a graphic image is known as visualization pipeline [7]. The steps are:

- **Sourcing** is the step of obtaining raw data G . This process may involve combining data from different sources or just formalizing the data into a computer representation.
- **Filtering** is the step of preliminary processing the data, cleaning it and performing different calculation tasks to obtain refined data G^f . This step is not mandatory and may be voided in some cases.
- **Mapping** is the step of obtaining of the mathematical model of spatial scene S from the filtered data G^f . Spatial scene is a set of spatial (usually two or three dimensional) geometric objects G^s with corresponding graphical attributes T . This step is often unique for every data analysis problem. It determines the effectiveness of the visualization method.
- **Rendering** — At this stage, based on the mathematical model of the spatial scene, its projection graphic image is built G^r . The effectiveness of rendering is usually defined by the quality of the previous step.

A simple pipeline is shown on fig. 2. Data from an spreadsheet file is converted to the memory as a table, being filtered and then mapped to equations using linear interpolation. The equation is being rendered as a regular Y(X) chart.

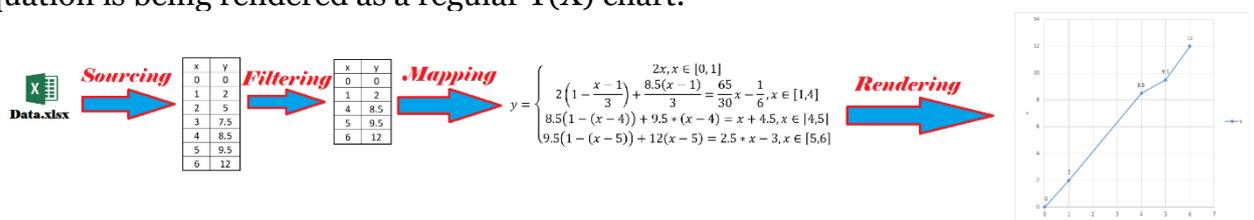


Fig. 2. Simple visualization pipeline example.

After the completion of the first problem, the initiative moves to the analyst, one needs to review the graphic images and do the statements.

2.2. Visual analysis of the graphic representation

As stated above, this problem is solved directly by the analyst. The results may and will vary depending on the analyst, their background and goals [8]. The solution of this problem consists of two generic steps:

- Visual analysis of the graphic representation. Fundamentally important to understand that the resulting graphics are only natural and convenient means of presenting spatial interpretation of the original data for the analyst. The spatial scene has to be visually analyzed. It allows the analyst to use the enormous potential possibilities of the spatial thinking in the analysis process. On this step, the analyst looks on the representation and notices some regularities or irregularities. This step cannot be strictly formalized, but we can define the main directions of analysis. The analyst may look on the following criteria's:
 - The shapes of the objects on the spatial scene. An example of this kind of statements is “The object number 1 is a cube, while object number 2 is a sphere”.
 - The relative positioning of the objects. There are two basic ways to do these kinds of statements: 1) “objects one to six are close to each other” or “Object seven is far from all the other objects”. 2) “objects one to eight are located on the sides of a cube”
 - The optical parameters (e.g. color, opacity etc.). An example is “Object one is red” or “Object two is opaque”.

Statements about the visual representation and the objects on the spatial scene are the results of this step. These statements have nothing to do with the initial data yet. It is important to highlight that the results of this step will differ depending on the analyst and their ability to spatial thinking, noticing details, focus etc.

- Interpretation of the results with respects to the initial data. On this step, the analyst converts the statements about the spatial scene to the object of consideration *Ob*. This step is one of the most crucial for the method, since it gives the overall results and it cannot be generally formalized. The analyst has to understand the connection and transformation of the initial data to the spatial scene and backtrack it to make the conclusions. The effectiveness of this step heavily depends on analyst's understanding of the domain and their experience. At the same time, this step gives the analyst much deeper understanding and, according to J. Thomas, insight into the data and domain as a whole.

An example of a simple visual analysis of the graphic representation is shown on fig. 3. The chart is being analyzed and a number of statements regarding the chart are given. Then the data is interpreted regarding the initial data (from fig. 2).

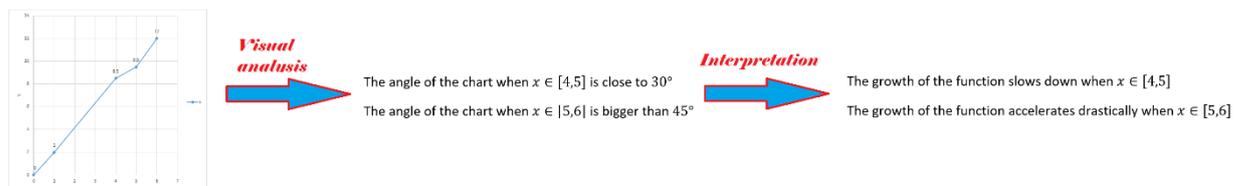


Fig. 3. Simple visual analysis explanation

Since the interaction between the human and the computer is the key in the solution of this problem, the solution makes the visualization method interactive.

After the analysis is complete, the results may satisfy the analyst, then the analysis of the initial data is complete or the analyst can decide to go back to one of the previous steps of the problems. The analyst can decide to repeat either step of either problem. This way, the method becomes iterative.

Defining the nature of the Visualization method, one can say that this method is a method of spatial modeling of the object of consideration. During the solution of the data analysis problem with the Visualization method, one models the object of consideration as a spatial scene, makes statements about the scene and then interprets the statements relating to the object of consideration. Summing up, we can say that the method is a modeling method with a spatial modeling of the object of consideration *Ob*. A spatial model is generated based on the object and the results of the spatial scene analysis are interpreted back to the object

3. Visual analytics in the “Scientific Visualization” laboratory of NRNU MEPhI

The purpose of the laboratory foundation was assistance to the physics departments in their studies using the visual analytics approach. This collaboration designed and implemented a number of passive and interactive visualization applications for various physics data. Some of them were developed with cooperation with British National Center of Computer Animation of Bournemouth University. Invisible objects, for example small as nanostructures or not physically visible as scalar or vector fields, got particularly high attention. The developed applications were static or animated, interactive or passive depending on the goals of the research and the needs of the researchers.

3.1. Applications in physics studies

The focus in physics studies is to illustrate the processes, so the researcher can see it. As Albert Einstein said, “If I can't picture it, I can't understand it”. The laboratory main focus was on making invisible things transparent and easy to observe.

Creating visualizations of nanostructure models is an example of such transformation. Replacing the pure numbers and formulas into an animated picture is one of the best ways to help the scientists understand the underlying processes. Giving the ability to change the initial data (like the conditions or the amount of initial objects) gives the analyst the ability to solve their tasks without showing the calculation results in numbers and wasting time on understanding what went wrong. All the needed information and processes are transparent and the calculations are done in the background.

The concatenation of two fullerenes is a good example of such problem. The created software had switches to tackle the fullerenes from different angles and with different speed. The analyst could change the parameters to get the picture they want to see. Two videos are shown on fig. 4. one is successful concatenation and the second one is not. [9]

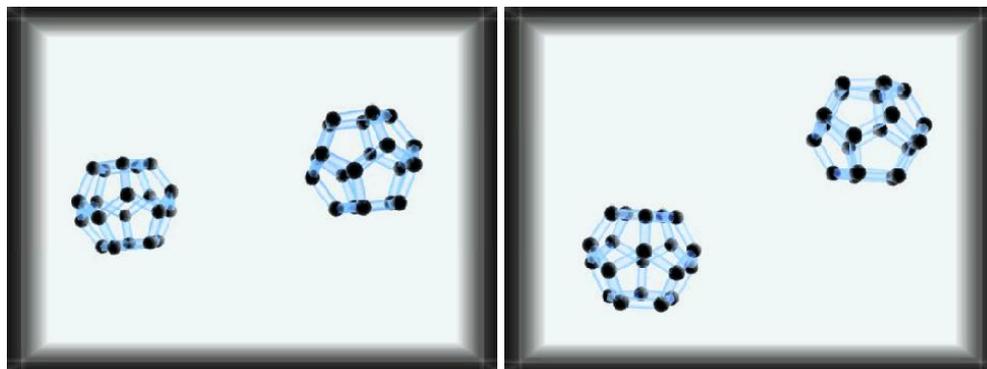


Fig. 4. Analysis of collision of two fullerenes

The other highly important object for visualization would be different fields. Our laboratory in collaboration with the physics departments worked with both scalar fields and vector (tensor) fields. In all cases, the main question was about the regions with low or high values, finding equivalent zones. In these researches, we used both shapes and graphical attributes to show the values and the directions of the fields.

An analysis of the superconductor parameter field. The superconductor is modeled according to Ginzburg-Landau theory. The designed software shows the current lines and the colors depends on the values of the field. This application allowed the physicists to actually see and understand the field, the directions and the values of the field.

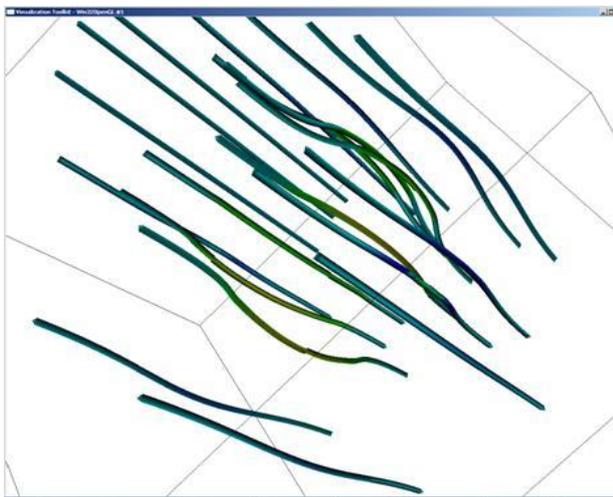


Fig. 5. Graphical representation of current lines of a vector field of the order parameter of a superconductor of the second kind

In some cases, there was a need of a combination of the two objects: a nanoobject and its field. In these cases the challenge was visualizations of both components in such a way, that they don't block each other and one can visually analyze both at the same time.

One of the projects of the laboratory was a visualization of a nanoobject and the electron density field. Our application first visualized the nanoobject (as spheres) and then added semitransparent isosurfaces to show the field. The color of the surfaces depends on the value of the field. This kind of visualization allows the researcher to see both the structure and the density of the field around it.

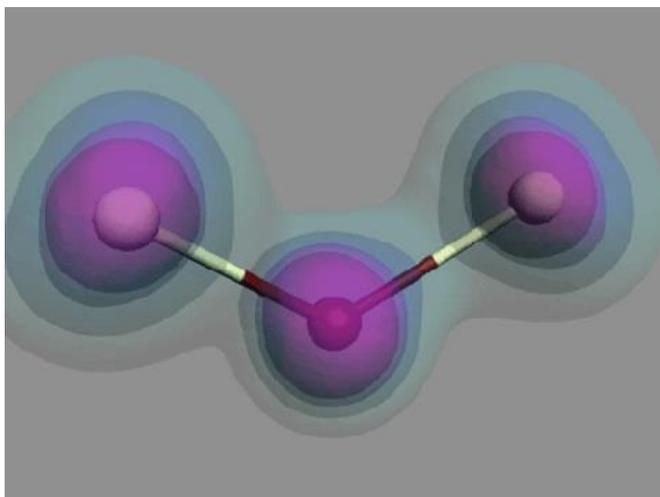


Fig. 6. Visualization of a nanostructure Cl_2O and its electron density.

3.2. Applications in finance

Our laboratory participated in credit institutes data analysis project managed by the finance department and the Federal Financial Monitoring Service of the Russian Federation. The main goal was finding the credit organizations that behave suspiciously, and passing the information for manual verification and a deeper check. Formalizing the goal, we can say that it was the anomaly detection problem. Out of all objects of consideration, the outstanding objects are objects of interest. [10] [11]

A multidimensional visual analysis tool was created to solve this problem. The main idea behind the tool is doing additional constructions in the multidimensional space. If the distance in between two objects in the multidimensional space is less than given by the analyst measure $d(\rho(x_i, x_j) \leq d)$, then we construct a section in the multidimensional space. The next step is projection to the 3-dimensional space and mapping the objects to spheres and the sections to cylinders. We use colors of the cylinders to show the distance between the multi-dimensional objects.

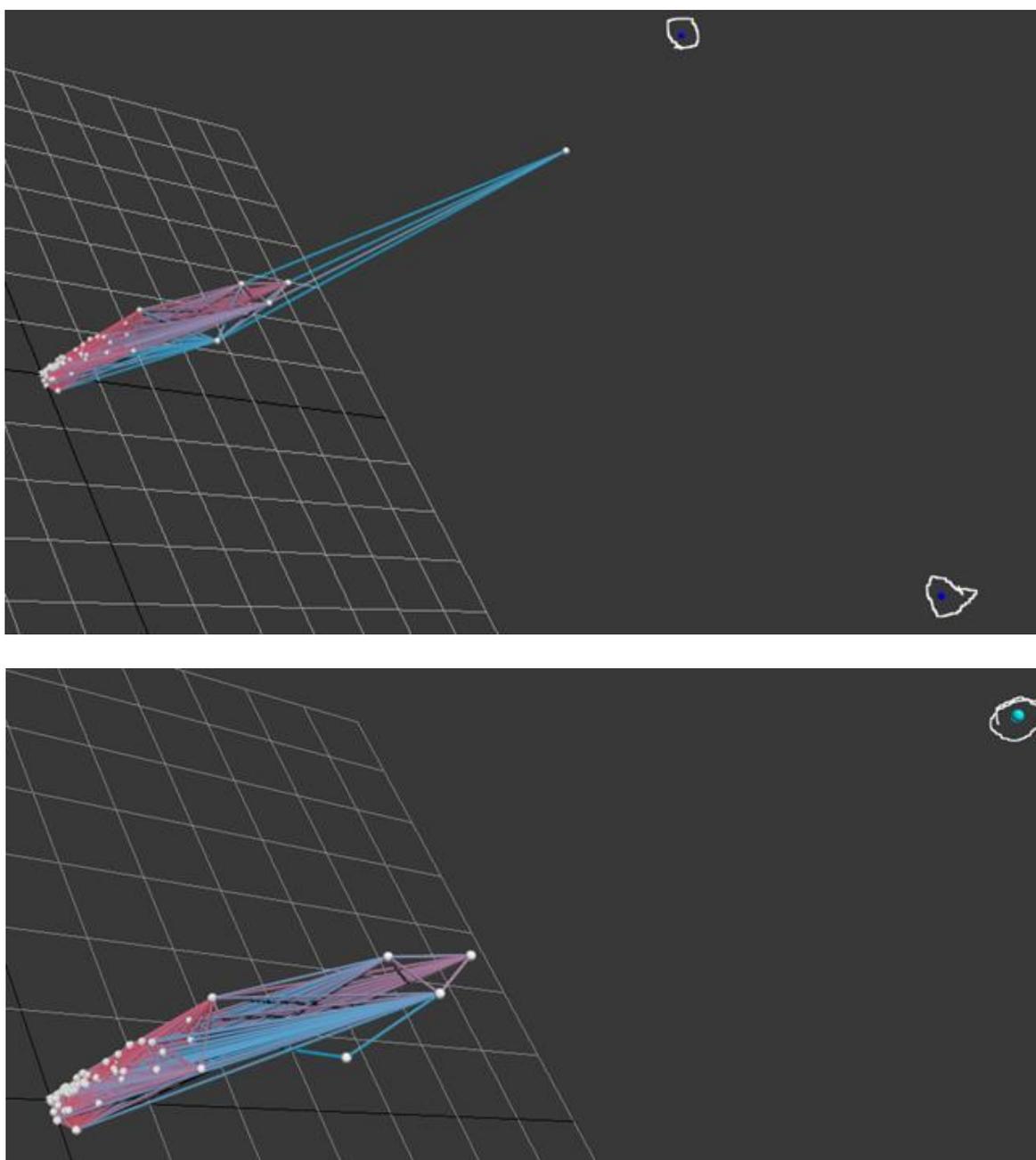


Fig. 7. Anomaly detection using the visualization method.

3.3. Application in Computational Centers Monitoring

Computational centers generate a lot of multidimensional metadata about its work. A huge issue is analyzing this data, since most of it reports normal results, but the interesting results are buried under huge amount of useless data. Our laboratory worked with a group from CERN on this project. The main goals were finding which computational centers have issues. [12]

In order to do this, two applications were developed – one took the datacenter’s metadata, clustered the data using some autonomous clustering algorithms and projected the data to a three dimensional space. The clusters were shown using colors.

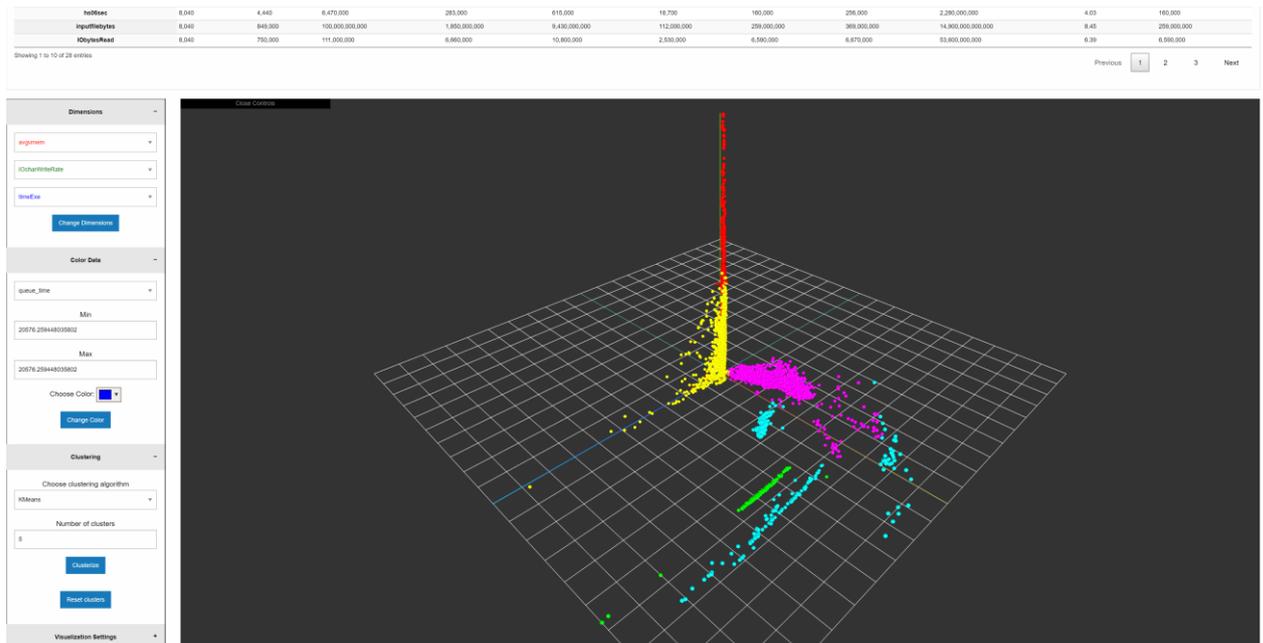


Fig. 8. Clustered data (more than 8000 objects with 28 parameters clustered to five clusters using K means algorithm)

The second application worked with the networking data and was designed to find the issues in communication between data centers. To solve this problem, the datacenters were put on two axes and the data was visualized as a data grid. With this kind of visualization, the analytic is looking for the spikes on the grid. An example is shown on fig. 9. The spikes are the red and yellow dots.

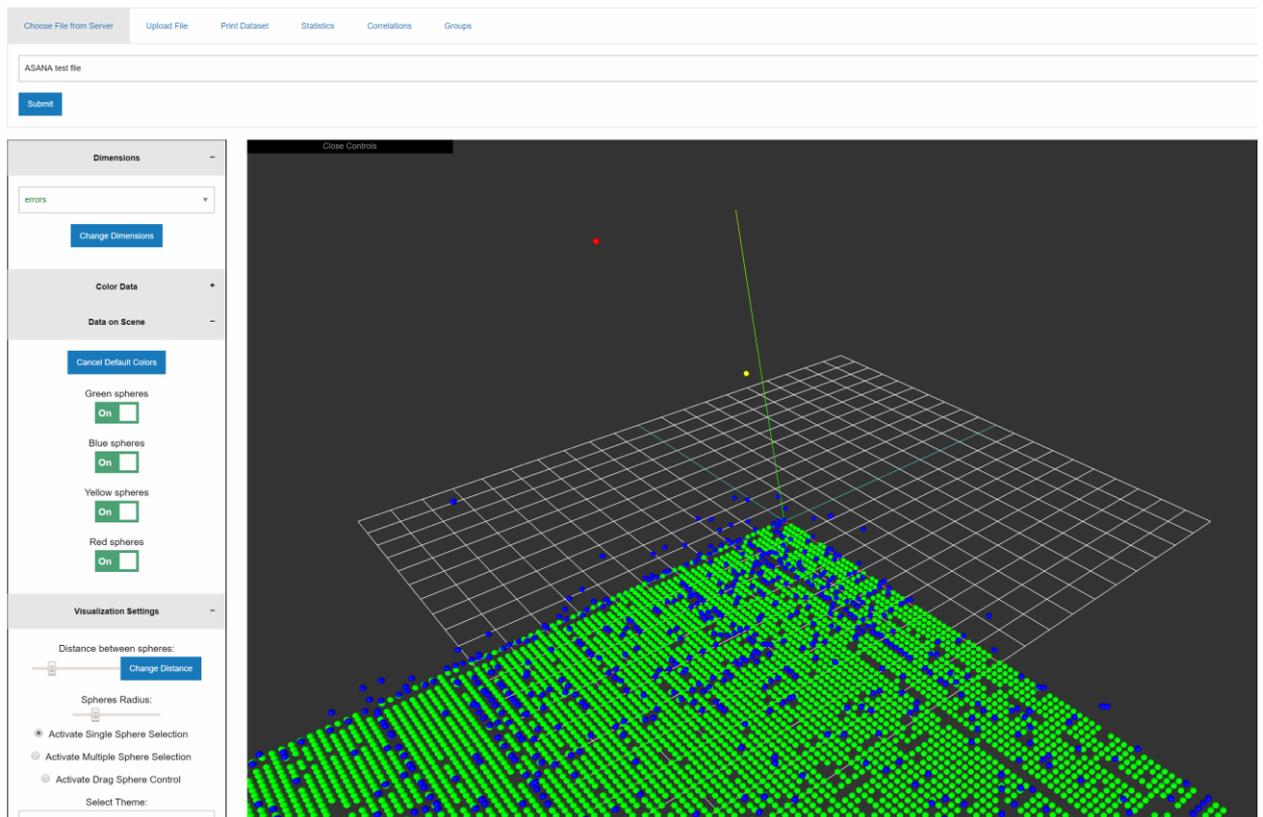


Fig. 9. Data grid with the spikes.

3.4. Other Activities

Our laboratory does a number of activities outside of research, but related to visual analysis. The first one is education on visual analysis and visualization. Currently two courses are being taught based on the laboratory: a basic one (visual analytics) and an advanced one (scientific visualization). The first one is aimed at master's degree students, while the advanced one is for PhD students. [6]

The second remarkable activity is the production of the scientific visualization journal. It is an open access electronic journal with editorial boards from all over the world. Currently there are five issues a year, in cooperation with two conferences. Issuing around ten to twelve articles per issue. The journal is being indexed in a number of databases: SCOPUS, RSCI and Compendex. Visual analysis is the main theme of the journal: all the articles are related in one way or another either to visual analysis or to computer graphics. [13]

Conclusion

In recent years, the activities of our laboratory are mainly focused on theory and development of applications for visual analysis of multidimensional data. [14] The research we perform may or may not use the visualization method, but we always focus on combining calculations with the interactive visualization. Using this technique, the analytic gets the best of the both worlds and can use their formally logical thinking as well as spatial thinking to solve complex problems.

References

- [1] J. Thomas, K. Cook, V. Crow, B. Hetzler, R. May, D. McQuerry, R. McVeety, N. Miller, G. Nakamura, L. Nowell and P. Whitney, "Human—Computer Interaction with Global Information Spaces—Beyond Data Mining," *Digital Media: The Future*, pp. 32-46, 2000.
- [2] J. Thomas and K. Cook, *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE-Press, 2005.
- [3] M. L. Huang, J. Liang and Q. V. Nguyen, "A Visualization Approach for Frauds Detection in Financial Market," in *2009 13th International Conference Information Visualisation*, Barcelona, Spain, 2009.
- [4] A. A. Cárdenas, P. K. Manadhata and S. P. Rajan, "Big Data Analytics for Security," *IEEE Security & Privacy*, vol. 11, no. 6, pp. 74-76, 2013.
- [5] M. S. Khine, "Spatial Cognition: Key to STEM Success," *Visual-spatial Ability in STEM Education*, pp. 3-8, 2017.
- [6] V. Pilyugin, "Scientific Visualization Laboratory of NRNU MEPhI," NRNU MEPhI, [Online]. Available: <http://sv-journal.org/unl/>. [Accessed 25 11 2019].
- [7] V. Pilyugin, E. Malikova, A. Pasko and V. Adzhiev, "Scientific Visualization As Method Of Scientific Data Analysis," *Scientific Visualization*, vol. 4, no. 4, pp. 56-70, 2012.
- [8] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas and H. Ziegler, "Visual Analytics: Scope and Challenges," *Visual Data Mining. Lecture Notes in Computer Science*, vol. 4404, 2008.
- [9] M. Strikhanov, N. Degtyarenko, V. Pilyugin, E. Malikova, M. Matveeva, V. Adzhiev and A. Pasko, "Computer Visualization Of Nanostructures Experience At NRNU "MEPHI"," *Scientific Visualization*, vol. 1, no. 1, pp. 1-18, 2009.
- [10] I. Milman, A. Pakhomov, V. Pilyugin, E. Pisarchik, A. Stepanov, Y. Beketnova, A. Denisenko and Y. Fomin, "Data Analysis Of Credit Organizations By Means Of Interactive Visual Analysis Of Multidimensional Data," *Scientific Visualization*, vol. 7, no. 1, pp. 45-64, 2015.
- [11] I. Milman and V. V. Pilyugin, "Interactive Visual Analysis of Multidimensional Geometric Data.," in *24 th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG 2016*, Plzen, Czech, 2016.
- [12] T. Galkin, M. Grigoryeva, A. Klimentov, T. Korchuganova, I. Milman, S. Padolski, V. Pilyugin, D. Popov and M. Titov, "The new approach to monitor the workflow management system ProdSys2/PanDA of the ATLAS experiment at LHC by using methods and techniques of visual analytics," *Scientific Visualization*, vol. 10, no. 1, pp. 77-88, 2018.
- [13] "Scientific Visualization," [Online]. Available: <http://sv-journal.org/>. [Accessed 25 11 2019].
- [14] O. Maslennikov, I. Milman, A. Safiulin, A. Bondarev, S. Nizametdinov and V. Pilyugin, "Development Of A System For Analyzing Of Multidimensional Data," *Scientific Visualization*, vol. 6, no. 4, pp. 30-49, 2014.