

## A research automation system for macroeconomic modeling\*

M. A. Bulyonkov, N. N. Filatkina

**Abstract.** The paper discusses several subsystems of the MIX system aimed to support various experimental researches in the field of economic modeling. The ultimate objection for each of these subsystems is to provide means for what-if analysis, impact estimation of strategic move, and decision making. Based on a comparative analysis of the design patterns and user scenario case studies, we propose directions for unified development.

**Keywords:** scientific research automation, situation room, development of the transport system, macroeconomic modeling, what-if simulation.

### 1. Introduction

In 2009, an initiative team consisting of the employees of the Siberian Institute of International Relations and Regional Studies, A.P. Ershov Institute of Informatics Systems, Institute of Economics and Industrial Engineering and Novosibirsk State University started the project MIX (model map-based information system) designed to create an information environment aimed at supporting the management of regions and their interaction. The main purpose of the MIX-PROSTOR system, which was presented in [1], is the automation of economic experiments related to the prediction of the Russian core transport network development. This system allows solving the problem of optimal transportation based on expert estimation of the changes in parameters that describe the freight services and handling in transport hubs. Thus we obtain a version of the basic transport network ensuring efficient interaction and competition of its constituent types of transport. In other words, we solve the problem of prediction by what-if simulation [2].

The main components of the system MIX-PROSTOR are as follows:

- a configuration manager, which allows to fix the structure of a particular transport network and global settings;
- a repository that stores experimental data and supports a series of experiments;

---

\*The work is a part of the Presidium of RAS project No. 31.3 “Integration role of large transportation project in special development”.

- an interaction interface, which provides connection to an external solver that implements the modeling method;
- various tools for experiment results visualization, with their mapping to geo-locations, when possible;
- algorithms and methods of comparative analysis; and
- a situation room, which realizes a decision support system (DSS).

Let us consider these components in more detail.

*Configuration manager.* The modeling method used in the MIX-PROSTOR system is quite general. It can be applied to the transportation network of various levels: international, federal, regional, etc. Moreover, even on the same level we can consider different versions of predicted network topology and different sets of the interacting types of transport (railroads, sea and river transport, pipelines, trucking, etc.) All information of that sort is aggregated in configuration files that can be edited by the system administrator (who may be the same analyst temporarily assuming another role).

*Data repository.* While the configuration of the system remains unchanged during what-if simulation analysis varying data of the experiment are related to the prediction of the network development. These data include the capacity of the main transportation lines (called shoulders) and the cost of transportation and handling in the network hubs for each type of transport. It turned out to be practically useful to specify series of experiments by a variation of individual parameters within a given range with a fixed increment. Such a specification is called a *variator* [1].

*Connection to an external solver.* The problem of optimal transportation may be reduced to a problem of integer linear programming. To solve it we exploit an external solver [3]. Therefore, the task of MIX-PROSTOR consists in the transformation of the given network configuration and varying parameters to the input data for the solver and interpretation of the solver results taking into account the geo-information and the structure of the network. In practice, the solver may take quite a long (and quite often unpredictable) time to obtain the results. Therefore, the caching of results is important for ensuring user interaction. In particular, it allows us to avoid recalculation before comparative analysis and/or visualization if the input has not been changed.

*Visualization tools.* Visualization is the most specific part of a particular task, and it requires creative design. The amount of data to be visualized makes it practically impossible to display them all at once on the limited size of the underlying map. The simulation may produce many different parameters even for a single geo-location, e.g. a transportation hub, and an attempt to show all numbers at this point would make a complete mess.

A good idea is to use different visualization means for different parameters. For instance, the capacity of a transport line may be reflected by its width; workload by the line color; the transport type by line type; etc. The visualization of some parameters, which are related to a region as a whole, can exploit choropleth maps.

*Algorithms and methods of comparative analysis.* The special kind of visualization is the one that enables the comparison of a series of pre-calculated variants. This task is of particular importance, because it contributes to the identification of the best prognosis by the expert. There are a number of useful “small” visual tricks: a display of relative deviations along with their absolute values, grouping and convolution data, such as “collapse” of transport corridors, etc., see Figure 1.

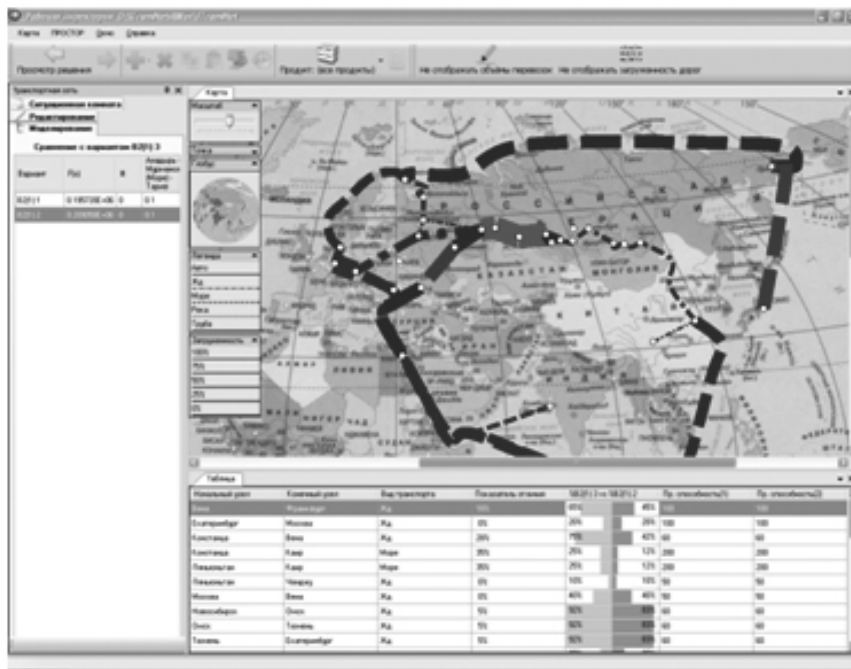


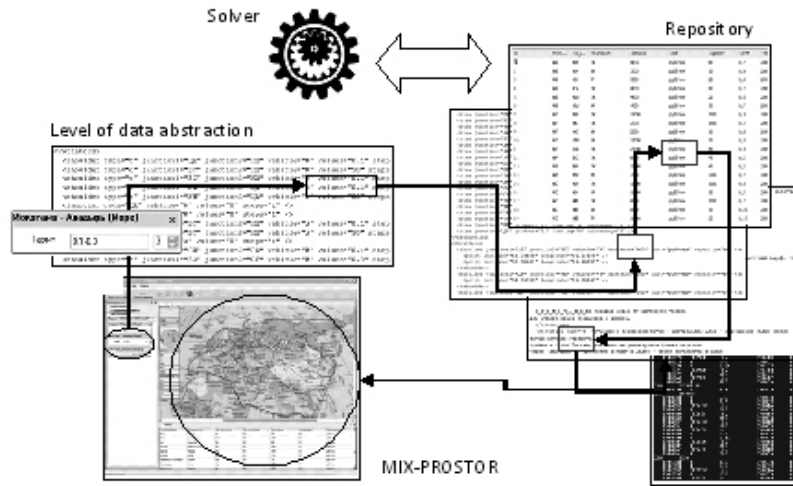
Figure 1. Methods of comparative analysis

To compare a large number of parameters produced by modeling, animation has turned out to be convenient: the compared variants (e.g., a series) smoothly “transform” one into others, thus allowing for tracking dynamics. In a sense, each variant is a frame in an animated film.

*Situation Room.* The practical usage of the MIX-PROSTOR leads to the idea of a *situation room* – a higher-level user interface, designed for decision makers rather than for experts in economics. The task of the experts in the framework is limited to the identification and formulation of basic situations.

Each basic situation characterizes and gives values to a group of related parameters. Here are some samples of such situations: “the elimination of bottlenecks on the Trans-Siberian Railway”, “the completion and rate reduction for the Trans-Caucasian arterial road”, “increase of the North Sea Route capacity”, “blocking of the Persian Gulf”, etc. Provided with a set of basic situations, a decision maker may use the situation room to combine them into *complex situation*, which describes a global perspective

The general design of the MIX-PROSTOR architecture is shown in Figure 2. As one can see, it complies with the principle of the separation of the levels of storage, computing and data visualization.



**Figure 2.** The general design of the MIX-PROSTOR architecture

## 2. Generalization of the method

The three-year user experience has shown that the system significantly increases the effectiveness of economic experimental research, especially in terms of perception and expert evaluation of the results. This motivated the expansion of the system application area to other related tasks. However, the specifics of the tasks themselves and their implementation required a significant generalization of the approach, which we will discuss below.

### 2.1. MIX-ATPK

The second application is related to forecasting the development of regional systems of Siberia on the example of the Taimyr-Yakutia aqua territorial industrial complex (ATPK as transliterated from Russian) in the impact

zone of the Northern Sea Route [4]. The concept of regional development depends on the interests of different players, such as large corporations, federal and local government agencies, foreign investors, etc. On the other hand, institutional factors play an important role; the specifics of both vertical integration and interregional cooperation must be taken into account. This problem shows that the interests of private investors and their desire to get a quick return on investment are in conflict with the strategic interests of the state. As companies participating in the development of natural resources have high economic, social, and political risks, their involvement requires a serious support from the public and state institutions. At the same time, in the current economic situation alternative solutions involving foreign investments may become more interesting.

The solver for this task is implemented as a MS Excel document: the user can “instantly” get output parameters by changing the values of certain cells that store the input parameters. Obviously, this document has a large number of auxiliary cells that organize calculations and are of no interest to the user. For the purpose of the integration of the solver in MIX, we introduced an intermediate level, abstracting the access to solver “variables” and events that occur when the variables are changed, as shown in Figure 3.

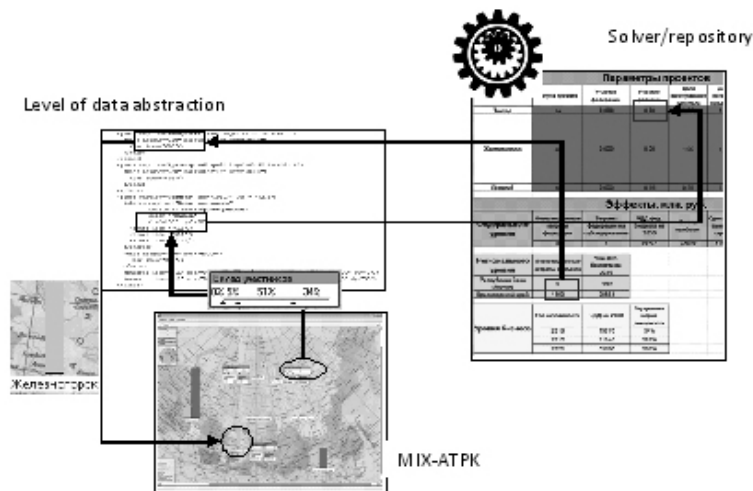
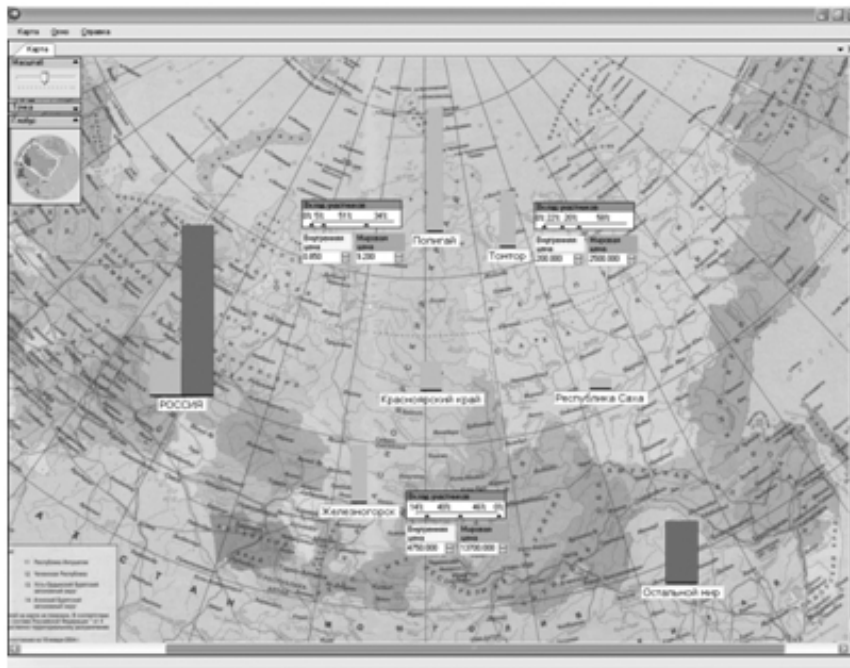


Figure 3. Events handling

Similarly to the solution for the transport problem in MIX-PROSTOR, here the visual elements and their relationship with abstract variables are defined in a configuration file. Each visual element is associated with an abstract variable name, which is used to set or retrieve its value from the solver. In addition, the element should be given a geo-reference. A certain

problem arises here in the case when several elements are attached to the same geographic point that corresponds to a participant: an investor, a developer or another player. To avoid the overlapping of visual elements, not only the global coordinates - latitude and longitude – must be specified but also the offset relative to the corresponding screen point.

Some types of visual elements that display numerical parameters, for example, the price or volume of investments, require a specification of the range of acceptable values. Some elements may be grouped into a single business diagram, for example, columns of different color. Another very commonly used type of visual elements is a track bar with multiple handlers, which allows for manipulating both the range of values and their grouping. It is used, for instance, for the separation of investment shares between players. The system must allow for using third party or custom visual controls designed specially for the elements like the ones described above (see Figure 4).



**Figure 4.** Visual elements in MIX-ATPK

## 2.2. MIX-OMMM

The third application is related to the analysis and forecasting of inter-regional cooperation, taking into account economic and social conditions, as

well as the dynamics of individual regions development. The problem is formulated using an optimizational aggregated interregional model (OMMM, as transliterated from Russian). This model was proposed by A. G. Granberg [5]. From a mathematical point of view, it is arranged as a linear system, the variables of which are the volume of production for each of the studied industries in each region and the volume of deliveries of products from one region to another, as well as restrictions on production capacity, labor and natural resources utilization, and the growth of investment funds. The optimization problem is to maximize the total volume of the final consumption when a limited time interval is considered.

Thus, the analysis of the obtained solution makes it possible to compare different concepts of development, to perform global assessment of economic proportions, and to make recommendations for the development and use of the advantages of individual regions in order to identify preferred embodiments of interregional cooperation [6]. This model allows for a significant reduction of work complexity and empowering the complex calculation of the overall final effect.

On the one hand, a characteristic feature of this problem is its large dimension and, therefore, a long time required for calculations. From the user point of view it is much more convenient to perform all the calculations in advance, and then analyze the stored results. On the other hand, the amount of data generated is too large to be displayed at once. Hence, one needs to formulate and realize various visualization aspects, concentrating on the meaningful subsets of data. The system logical structure is similar to that of MIX-PROSTOR and is displayed in Figure 5.

In other words, when imaging data on a map, many information items are related to the same point or closely located points, making the whole picture overcrowded and incomprehensible. The usual solution for the problem is to increase zoom and locate the point of interest though in this case the user loses the overall view. The MIX-OMMM uses another approach:

- display information in a compact way, allowing the user to expand it only when and where needed, and
- layout the blocks of information without overlapping, connecting each block to the original point by a visible straight line.

The screenshot of the MIX-OMMM system realizing the above approach is shown in Figure 6.

The second requirement may be formulated as an interesting optimization problem: let  $B$  be a  $(h_1, w_1), \dots, (h_n, w_n)$  be a set of rectangles specified by height and width, and  $\{(ax_i, ay_i) | i=1..n\}$  be a set of “anchor points”, one for each rectangle. It is necessary to locate the rectangles, i.e. to find a set of points  $\{(px_i, py_i) | i=1..n\}$  such that rectangles do not overlap

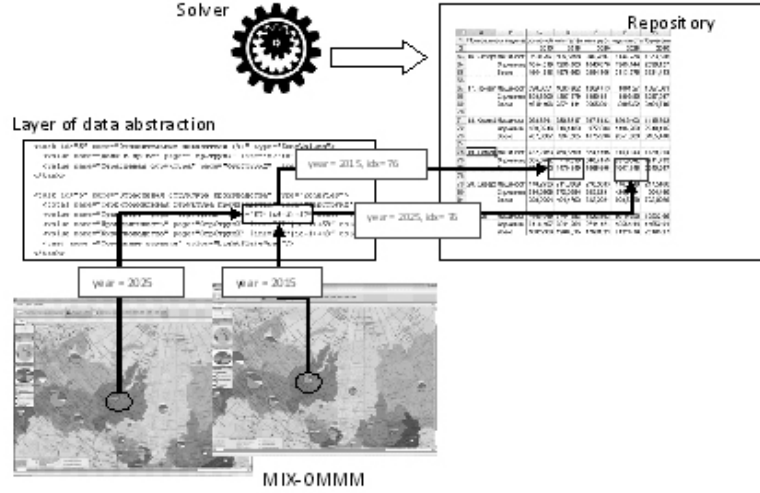


Figure 5. MIX-OMMM architecture

$$\forall ij = 1..n : ax_i > ax_j + w_j \text{ or } ax_j > ax_i + w_i \text{ or } ay_i > ay_j + h_j \text{ or } ay_j > ay_i + h_i$$

and minimize the distance from the anchor points

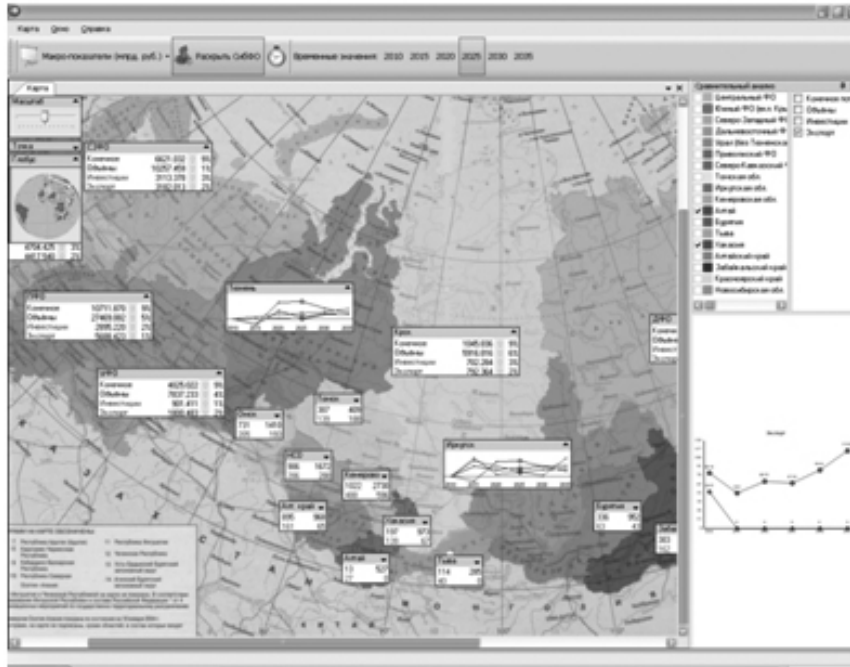
$$\sum_{i=1}^n \sqrt{(ax_i - px_i)^2 + (ay_i - py_i)^2} \rightarrow \min.$$

An additional requirement for the layout method is its stability: a small variation of rectangle sizes should cause small changes in the layout. For example, if the user expands a single information block, all other blocks must either remain in place or move only slightly. The problem is reduced to nonlinear function minimization in the restricted area. To handle it, we use a heuristic based on the iterative method of “simulated annealing”.

One may note that at a given moment an expert is interested not in the whole picture, but only in one of its aspects, which can be characterized by a comparatively small number of parameters. Examples of such aspects are the following:

- macro-indicators (final consumption, production volume, investments, exports),
- flows of supplies of goods (import/export of products from/to a region),
- indicators of the inter-regional impact (departure, arrival, transit, turnover),



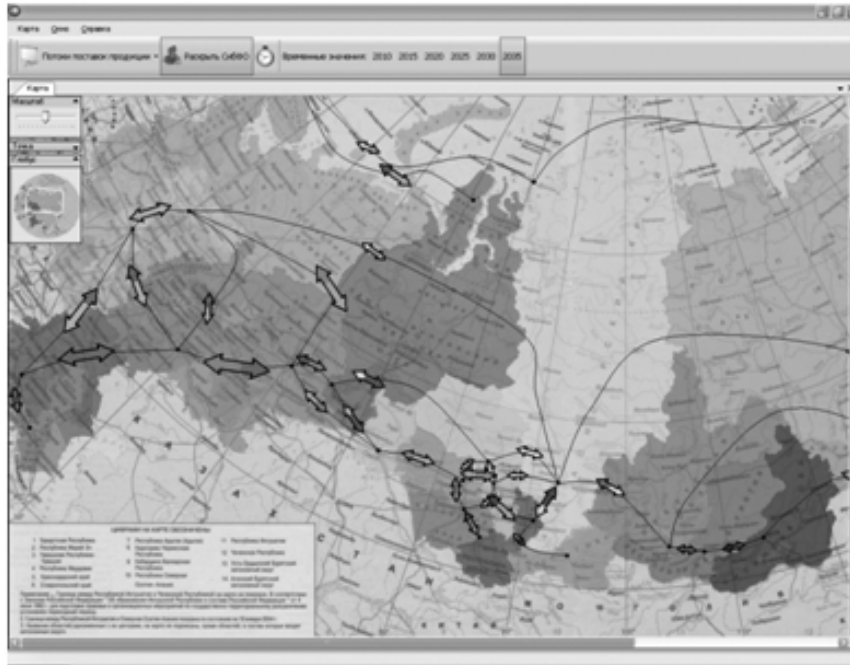


**Figure 6.** A screenshot of the MIX-OMMM system

- sectoral structure of production (transport, industry, livestock, etc.).

Each aspect may be displayed most appropriately in its own specific way. For example, the industrial structure is commonly represented in the form of a pie chart, while macro-indicators are usually shown in the form of a column chart or a table with specific values. For most parameters, it is interesting to know not only their absolute values, but also the dynamics of their increase/decrease. Therefore, it is necessary to switch not only between compact and expanded information blocks with absolute values, but also to display a graph of trends for selected indicators over a reviewed period of time.

A special visualization method is used to display solution parameters associated with communications between regions. We have tried a number of different methods, but perhaps the most demonstrative is the one that shows the flow between two regions as an arrow, with its size proportional to the volume of deliveries and the coloring reflecting the ratio of the incoming and outgoing volumes. The arrow itself is located along the path connecting specified points within the regions, see Figure 7. Other options, such as displaying numerical data above or below the path or at the entry and exit points do not provide sufficient clarity. Nor can the thickness of the line be used because of a large spread of values.



**Figure 7.** Traffic flow visualization

### 3. Comparative study and development perspectives

The following comparative analysis of the systems described above has two objectives. First, this is the development of common machinery and tools, which can be shared by all the systems. Second, the functionality of each system may be enhanced further through an analogy with another system. The three main areas of development are as follows:

- from interactive experiments to a series of experiments,
- from a linear series of experiments to a multidimensional experiment space, and
- from experimentation to situational analysis.

#### 3.1. Response time

System design depends heavily on computer performance and effective implementation of modeling algorithms. It is clear that there is a certain psychological limit of waiting time that the user accepts for obtaining computation results in an interactive session.

Among all the systems discussed above, the smallest response time is demonstrated by the MIX-ATPK system: any change to a parameter made

by the user leads to the “instant” recalculation and refresh of the visual representation of all output parameters. On the other hand, the calculations performed in the MIX-OMMM system require so much time that it makes an interactive session impossible. The performance of the MIX-PROSTOR is close to the waiting time threshold: modeling a variant usually takes no more than a few minutes, which allows for using it in the interactive mode. However, the user would be greatly disappointed if he had to wait again for getting the same results when he wished to return to the same variant. This happens in particular when several variants are to be compared. The problem is solved by caching the results and tracking the modification of input data.

A rather unexpected disadvantage of high interactivity of the MIX-ATPK system is that it becomes difficult to compare variants which differ slightly in several input parameters. While the user changes the value of parameters one by one, the display is changed in real time, and some output parameter may increase and/or decrease many times during the changes. So it is hard to compare the final state with the original one. Therefore, for comparative analysis and forecasting the user-scenario should start with taking a “snapshot” of the current state, then you can make all changes necessary for the transition to the new state, then take a snapshot of the second state, and after that you can switch between the steps as many times as needed. Essentially, this is similar to the saving of variants to the repository and the introduction of the concept of a series of experiments as is done in other systems.

### **3.2. Series of experiments**

In the MIX-OMMM system, a series of results to be compared is determined by a variation of the year. On the other hand, the system focuses on economic development trends, and in this sense the series of experiments should be treated as a whole. The nearest enhancement planned is designed to allow for varying the initial data in order to obtain different trends and compare them rather than a single trend dynamics.

In the MIX-PROSTOR system, it is possible to define explicitly a series of variants by specifying the range and fixed increment for each of the parameters of interest, so that the system can process all variants of the series in a batch mode. Since variations of parameters may be independent, the space of experiment results is multidimensional. However, at the moment the MIX-PROSTOR system simply enumerates the variants, thus losing its multi-dimensional nature. The solution of the problem cannot be straightforward because some parameters essentially depend on one another. For example, to simulate a change in a railway bandwidth, one must set the capacity of each of its constituent shoulders in a consistent manner.

There is no concept of a series of experiments in the MIX-ATPK system at the moment since, as we have already mentioned, the very concept of an experiment is not clearly formulated. Further development of the system includes visualization of economic games with the purpose to find the balance of interests of different economic subjects, such as regions, the federal government, and foreign investors. This implies that the experiment data space should reflect different behavior strategies separately for each subject, which again leads to a multidimensional space. Thus, the repository must maintain an effective access to various projections of the space of experiments. A promising basis for the implementation is the OLAP technology [7, 8].

### **3.3. Situation room**

The ultimate goal of any of the systems considered is decision making based on expert analysis. In practice, an expert is not necessarily a decision maker, and vice versa. A decision maker takes into account many factors in addition to economic modeling, which is one of the reasons why the process cannot be fully automated.

From this we can conclude that, firstly, different levels of abstraction require different user interfaces: at the highest level a decision-maker does not have to be aware not only of the scientific experiment technology, but also of the specific data, methods of calculation and in general of the solution of mathematical economic problems. The situation room of the MIX-PROSTOR system is an example of such high-level interface.

Secondly, we need more advanced tools to combine and generate variants for the situation room. The system should provide the basic space of situations composed into a model hierarchy by relationships such as “a situation is an integral part of another one”, “a situation is determined by a variety of alternative situations”, etc.

Economic games add an economic subject as a new dimension to the space of situations and allow iteration of the modeling process in order to achieve a reasonable compromise between the subjects: results of the experiment can be taken as the specification of new variants. In this case, it would be useful to trace the history of the emergence of variants.

### **3.4. Conclusions**

We can conclude that all the components of the MIX system have growth points and may adopt features from each other. Also, comparative analysis sketched a common platform of a system that supports experimentation in economic modeling; its logical structure is shown in Figure 8:

- An external solver that performs calculation for a specific economic

problem;

- The Repository that stores a multi-dimensional space of variants and provides a caching mechanism for the modeling results;
- The layer of data abstraction that allows for the mapping of subject domain objects to the repository data items and/or projections;
- The core MIX system that supports extension by domain specific visual layers and components, which may be bound to geo-location;
- A situation room that provides a high level user interface for decision making based on the possibility of variant composition, which may be domain specific.

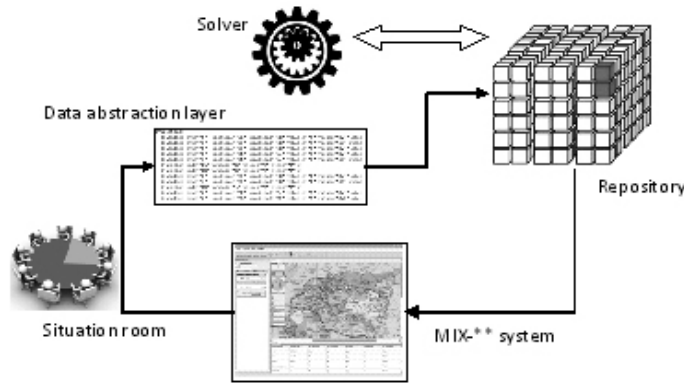


Figure 8. Generic MIX architecture

#### 4. Related works

The mutual influence of regions realized by the OMMM model can be used not only to solve economic problems. For example, the problem of resource consumption is considered in [11] for the purpose of predicting the region’s ecology. An even wider range of applications of this model, including tourism, health, energy, and sociology, was discussed at the conference [13].

The transportation problem in a more complex, non-linear formulation is considered in [10]. It takes into account not only economic factors, but also natural disasters (such as earthquakes) and their economic impact spread through the transport network and mutual influence of the regions.

The perspectives of the what-if analysis and its application are considered in detail in [9]. These and many other works are focused on the economic

issues per se. The aspects of data representation and visualization are either not considered at all or mentioned only briefly.

An approach to the visualization of global production networks using data clustering, which reduces the volume of the rendered information, is proposed in [12]. However, this study does not provide ways to compare different variants and visualize the dynamics of their changes.

It is also worth mentioning the Global Model Workstation system developed by Oxford Economics Company [14], which provides a user interface to visualize various economic parameters of selected regions using business charts, tables, and mapping. The most interesting is the huge amount of relevant, regularly updated data of various areas of the economy and regions of the world. The system, however, is not suitable for integration custom experimental modeling and visualization methods.

## References

- [1] Buluonkov M.A., Filatkina N.N. Situational analysis in the transport forecast system MIX-PROSTOR // Information Technologies. – 2013. – No. 8. – P. 43–52 (In Russian).
- [2] Vorob'eva V.V., Malov V.Yu., Radchenko V.V., Potter M.V., Serebryannokov I.E. The model of the basic transport network development forecast // Modelling of the Industrial and Regional Systems Based on GIS and IT / Eds: Ju. Sh. Blam, V.V.Radchenkov. – Novosibirsk: IEIE SB RAN, 2011. – P. 68–96 (In Russian).
- [3] Zabinyako G.I. A package for integral linear programming // Discrete analysis and operation research. Ser. 2. – 1999. – Vol. 6, No. 2. – P. 32–41 (In Russian).
- [4] Tarasova O.V. Configuration of economic interests and cooperation parameters for arctic ATIC participants // Institutional Transformation of Economics: Russian Vector of the New Industrialization. – Proc. 4th Intern. Conf. (Russia, Omsk, October 21–23, 2015). – Omsk, 2015. – Vol. 2. – P. 463–469 (In Russian).
- [5] Granberg A.G. Optimization of Territorial Proportions of the National Economy. – M.: Economics, 1973 (In Russian).
- [6] Melentiev B.V. The real possibilities of modern mathematical models to predict the economic development of the regions // Analysis, modeling, management, development of socio-economic systems (AMUR-2015). – Proc. Scientific. IX Intern. Shk.-Sympos., 12–21 September, 2015, Sevastopol / [ed. AV Segal]. – Sevastopol, 2015. – P. 240–245 (In Russian).
- [7] Chaudhuri S., Dayal U. An overview of data warehousing and OLAP technology // SIGMOD Record. – 1997. – Vol. 26. – P. 65–74.
- [8] Lawrence M., Rau-Chaplin A. Dynamic view selection for OLAP, Data Warehousing and Knowledge Discovery // DaWaK 2006. – Lect. Notes Comput. Sci. – 2006. – Vol. 4081. – P. 33–44.

- [9] Golfarelli M., Rizzi S., Proli A. Designing what-if analysis: towards a methodology // Proc. DOLAP-2006. – P. 51–58
- [10] Ham H., Kim T.J., Boyce D. Implementation and estimation of a combined model of interregional, multimodal commodity shipments and transportation network flows // Transportation Research Part B. –2005. – P. 65–79.
- [11] Wiedmann Th. A review of recent multi-region input-output models used for consumption-based emission and resource accounting // Ecological Economics. – 2009. – Vol. 69. – P. 211–222.
- [12] Ortega-Segura L., Gaytán-Rodríguez M. Visualization of information analysis as a tool for cluster identification of global production networks // Proc. 22nd Internat. Input-Output Conf. – <https://www.ioa.org/conferences/22nd/papers.html>.
- [13] 22nd International Input-Output Conference & 4th Edition of the International School of I-O Analysis. – <https://www.ioa.org/conferences/22nd/papers.html>.
- [14] Global Model Workstation by Oxford Economics. – <http://www.oxfordeconomics.com/techlabs/global-model-workstation-v2.0>.

