Approximation of Pareto frontier in complicated non-linear multi-objective problems

A.V. Lotov

Dorodnicyn Computing Centre

Federal Research Center “Computer Sci. and Control” of Russian Academy Sciences

Lisbon, October 24, 2019

This talk has been supported by the project
POCI-01-0145-FEDER-030391
(PTDC/ASP-SIL/30391/2017),
BIOECOSYS, Portugal

FCT Fundação para a Ciência e a Tecnologia

Cofinanciado por:

COMPETE 2020 PROGRAMA OPERACIONAL COMPETITIVIDADE E INTERNACIONALIZAÇÃO

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

REPÚBLICA PORTUGUESA

Plan of the talk

1. Pareto frontier in multi-objective optimization (MOO) problems
2. A few words on Pareto frontier approximation methods in non-linear case
3. Interactive Decision Maps (IDM) approach
4. The Baikal Lake level regulation problem as an example of a complicated real-life MOO problem
5. Application of IDM in the framework of Baikal Lake level regulation problem
6. Optima Injection Method for Edgeworth-Pareto Hull approximation in complicated non-linear MOO problems
7. Launch Pad Method (LPM) for Edgeworth-Pareto Hull approximation

A few words concerning Pareto frontier and Edgeworth-Pareto Hull in MOO problems

Usual (single-objective, scalar) optimization

Set of feasible decisions (Feasible set)

$$X \subset R^n$$

Objective function

$$y = f(x)$$

Optimization problem

$$y \to \min \text{ (or } y \to \max)$$

Solution of optimization problem

$$x^*: f(x^*) \le f(x) \text{ for all } x \in X$$

Multi-objective optimization

Feasible set $X \subset R^n$

Objective vector $y = f(x), y \in R^m$

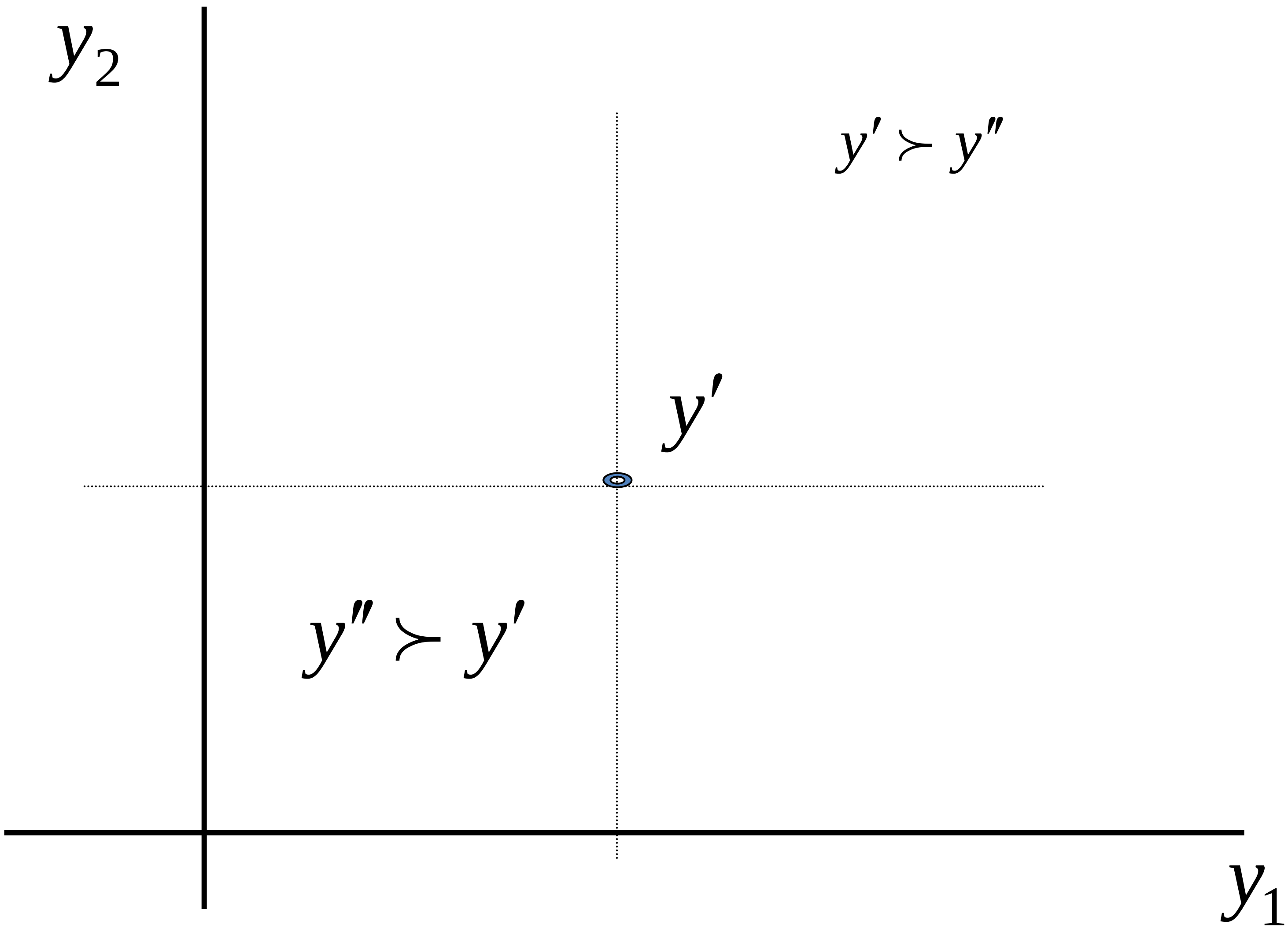
(If $m > 3$ one speaks about many-objective problems)

$y \to \min$ Means that minimization of all objective values is preferable

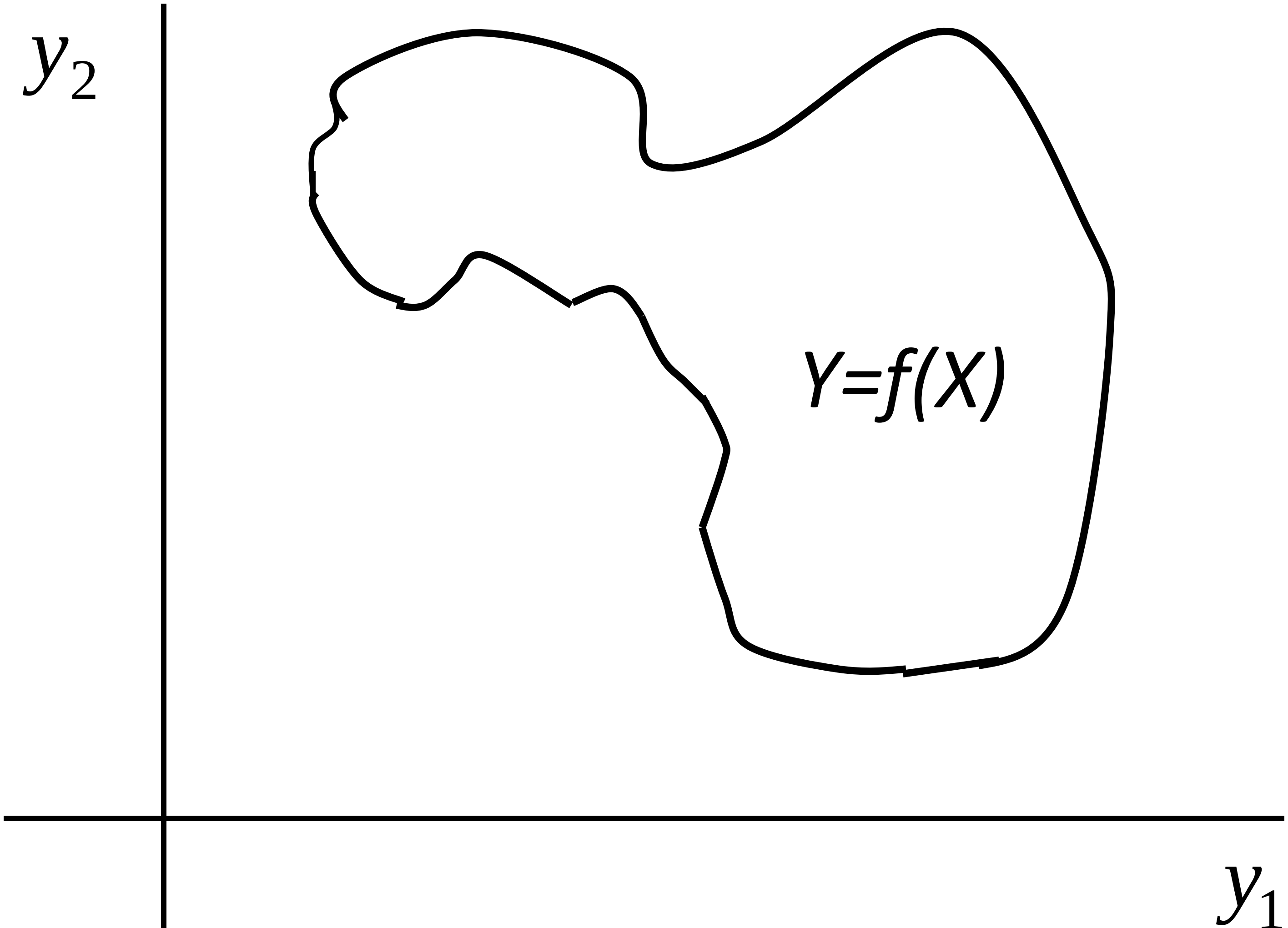
Pareto domination

$$y' \succ y'' \Leftrightarrow \{y'_i \le y''_i, \quad i = 1,2,\ldots,m;\ y' \ne y''\}$$

Pareto domination (minimization case)

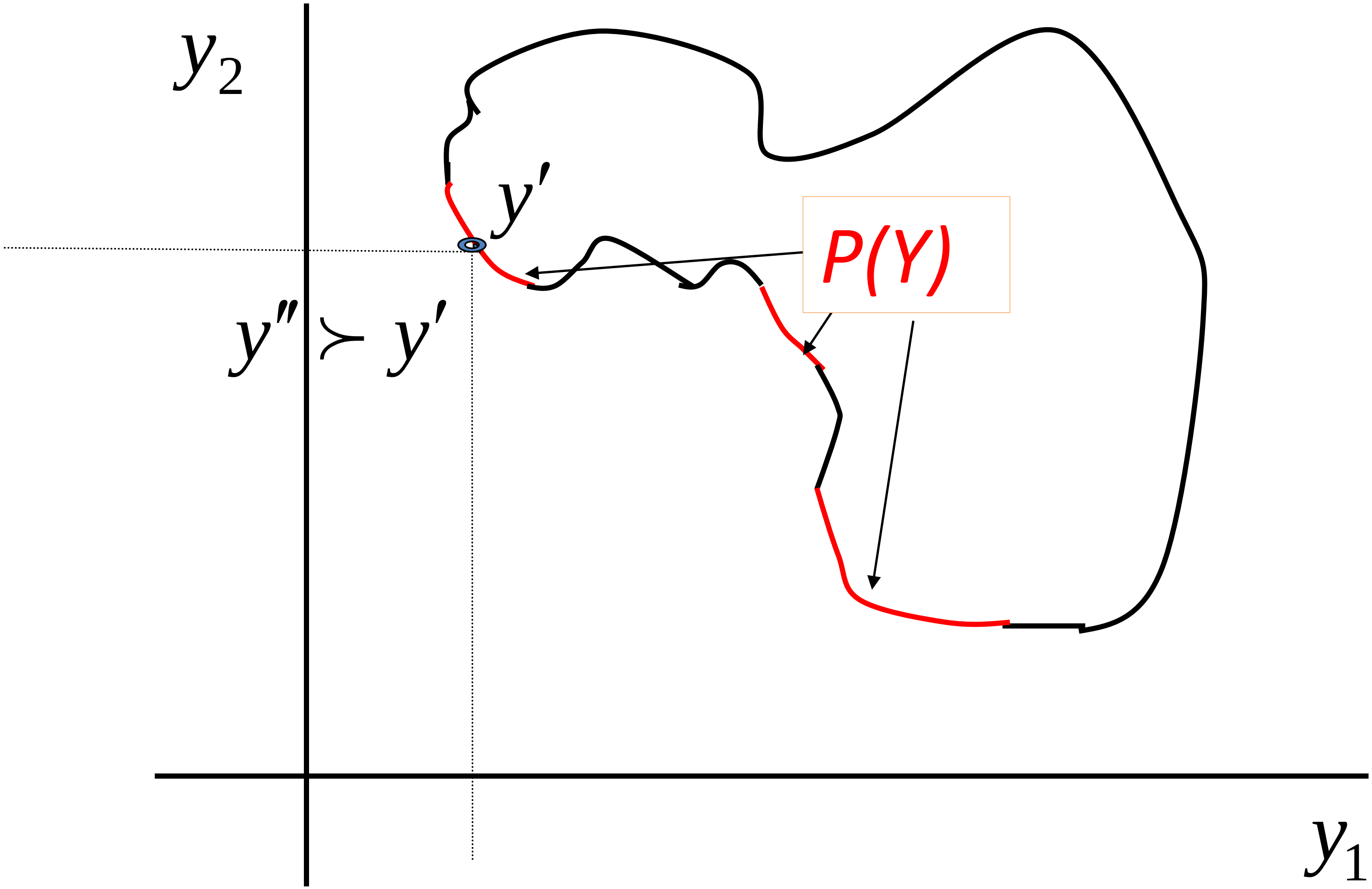


Example of *Y=f(X)* in a non-convex MOO problem



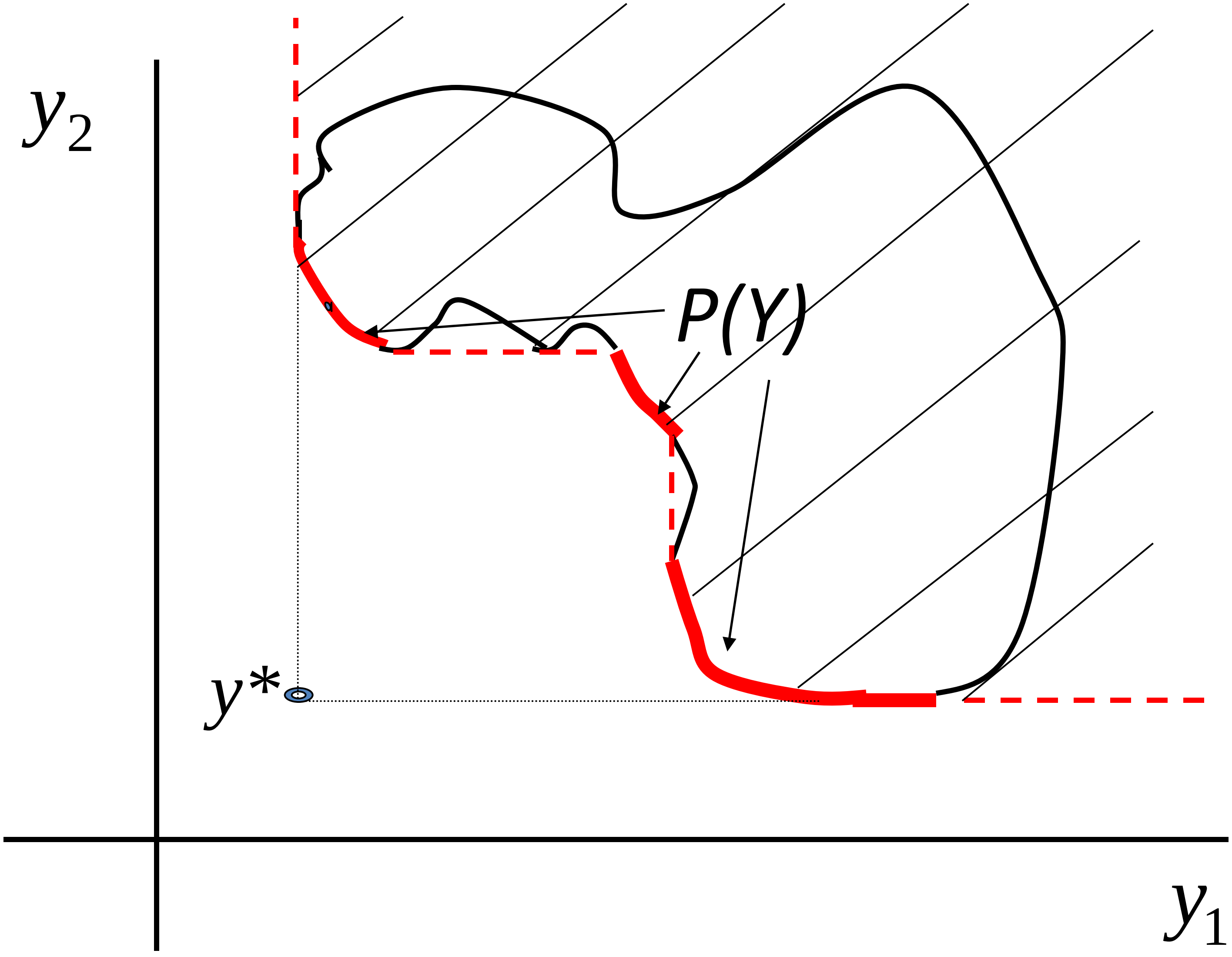
Feasible objective set $Y = f(X)$

Solution of the MOO problem: non-dominated (Pareto-optimal) frontier of *Y*



$$P(Y)=\{y\in Y:\{y'\in Y: y'\succ y\}=\varnothing\}$$

Edgeworth-Pareto Hull of *Y*

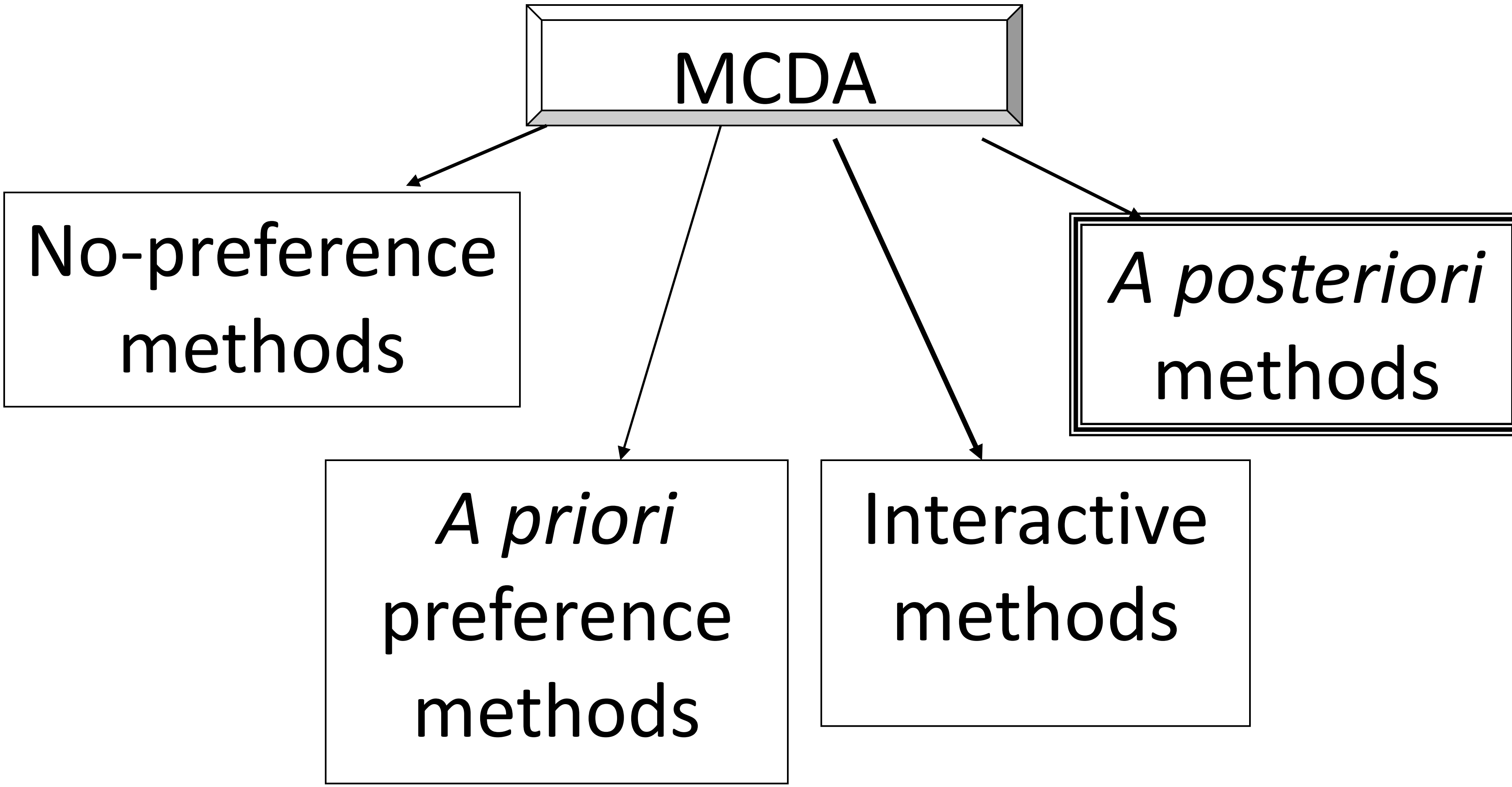


EPH has the same Pareto frontier as *Y*

How to select a unique decision from the set of Pareto-optimal decisions?

Decision makers are needed to select the most preferred solution among the Pareto-optimal solutions. Different MOO methods are based on different ideas concerning the role of decision maker.

Classification of MOO methods according to the role of DM



Pareto frontier approximation (*a posteriori*) methods

Two main problems that must be solved in the framework of Pareto frontier methods

- How to approximate the Pareto frontier
- How to inform the decision maker about the Pareto frontier

Two ways for informing the decision maker about the Pareto frontier

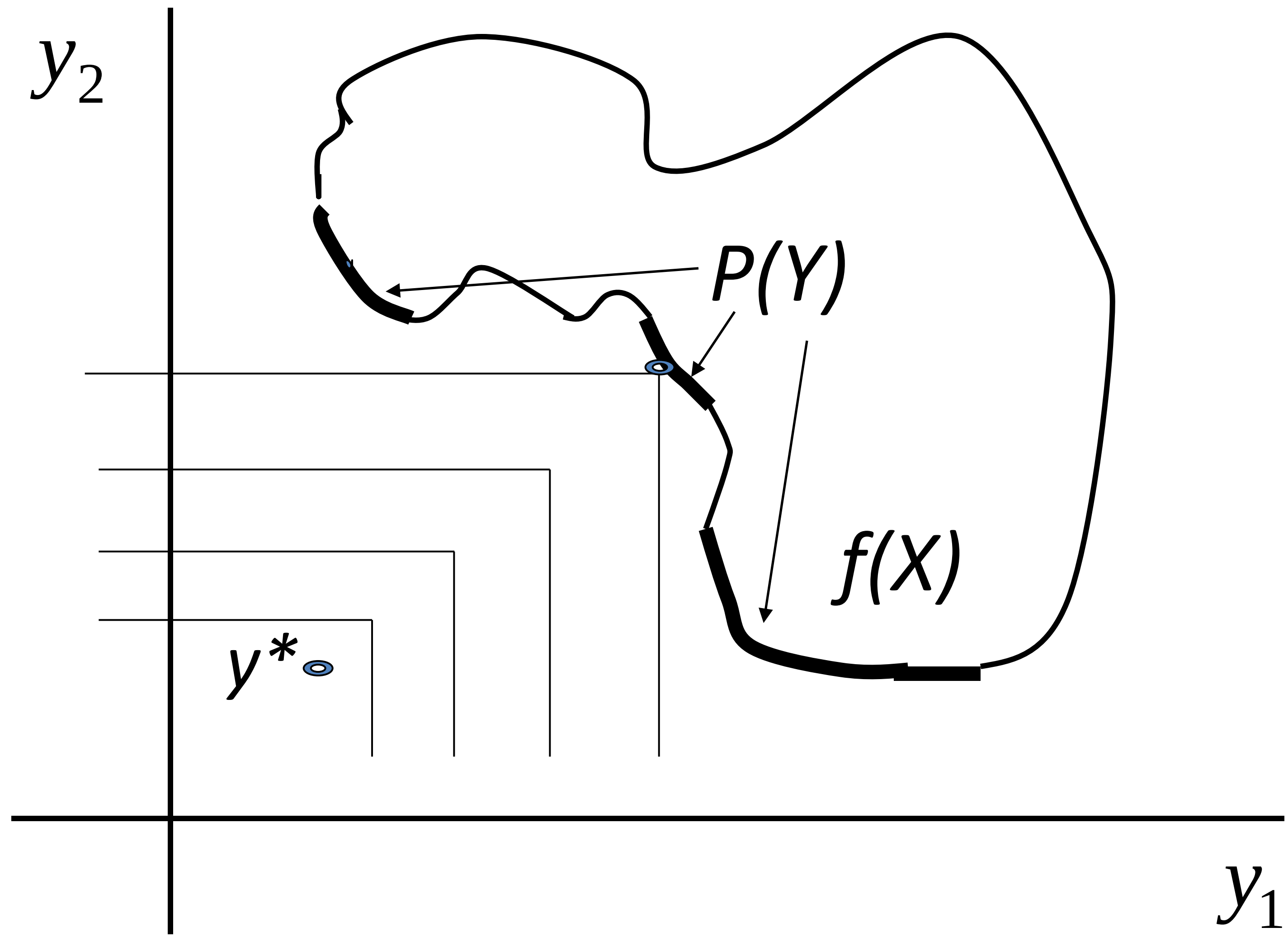
- By providing a large list of the points that belong to the Pareto frontier
- By visualization of the Pareto frontier

Approximating the Pareto frontier

Approximating the Pareto frontier is the *main computational problem* associated with the MOO techniques based on Pareto frontier analysis.

Multiple methods of Pareto frontier approximation were developed for the case of relatively simple non-convex non-linear MOO problems, for which solving of a sufficiently large number of global optimization of scalar functions of objectives (scalarizing functions) is possible.

Tchebycheff distance as scalarizing function



Application of the Tchebycheff distance

The problems

$$h(y) = \max\left[\lambda_i (y_i - y_i^*) : i = 1,2,\ldots,m\right] \to \min$$

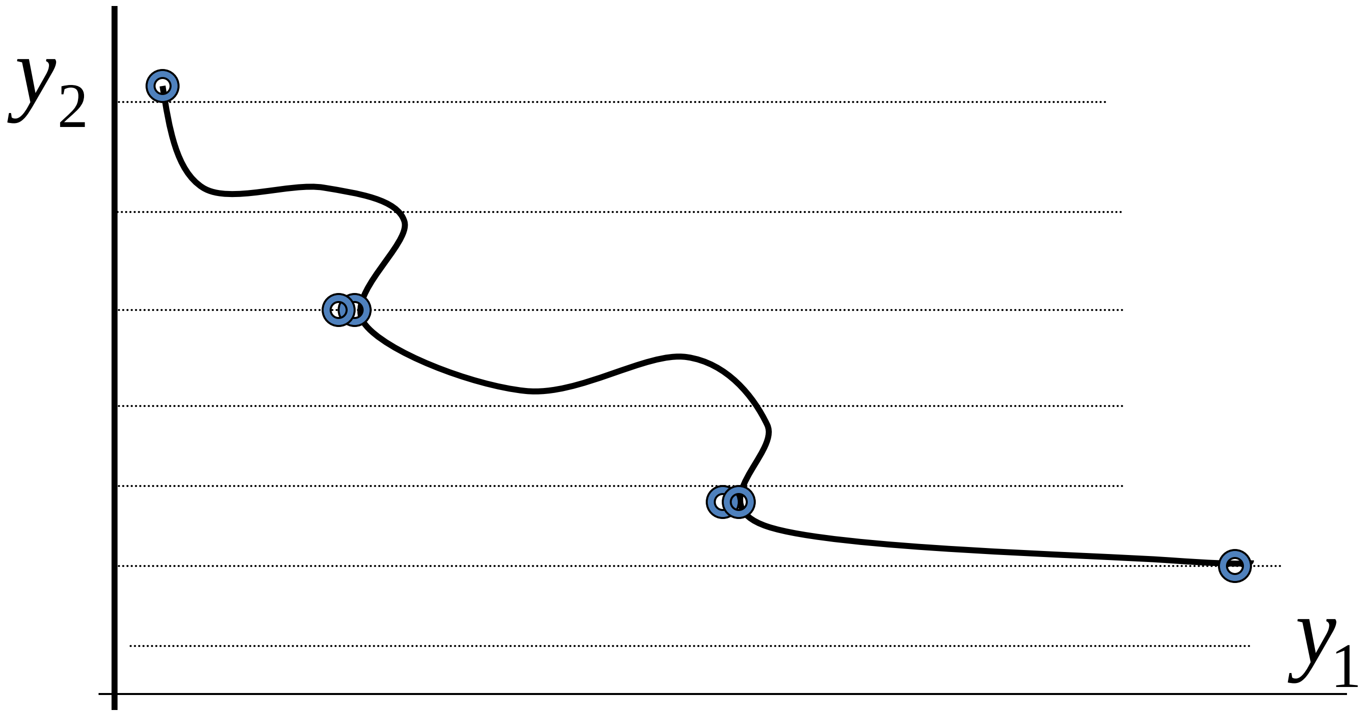
$$y = f(x), \; x \in X$$

are solved for a large number of parameter vectors

$\lambda = (\lambda_1,\ldots,\lambda_m)$ where $\lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1$.

Result: a large list of Pareto (and weak Pareto) points.

ε-constraints method



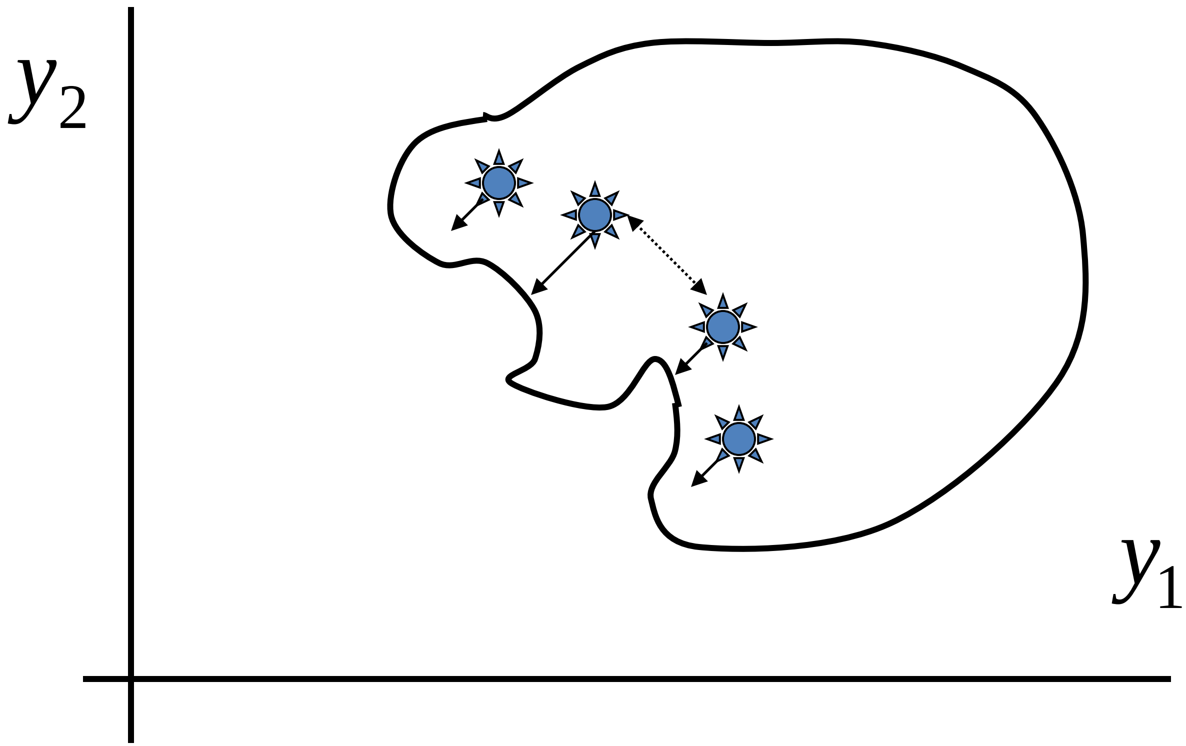
$$y_{i^*} \to \min \quad y = f(x), x \in X,$$

$$y_i \le l_i^p, i \ne i^*, \ p = 0,1,\ldots,P_i$$

Evolutionary (meta-heuristic) methods for Pareto frontier approximation

During two last decades, new kind of methods for Pareto frontier approximation were developed – the evolutionary (meta-heuristic) methods. They can be applied in various cases including complicated non-linear problems, for which solving of sufficiently large number of global optimization problems for scalarizing functions is impossible. The most important type of the evolutionary methods is provided by the genetic methods.

Genetic methods for Pareto frontier approximation



Genetic methods use mutation of decisions, cross-over of them and selection of ‘good’ decisions.

Result: a large list of points that are hopefully close to Pareto frontier. However, convergence to Pareto may be slow.

Interactive Decision Maps (IDM) technique

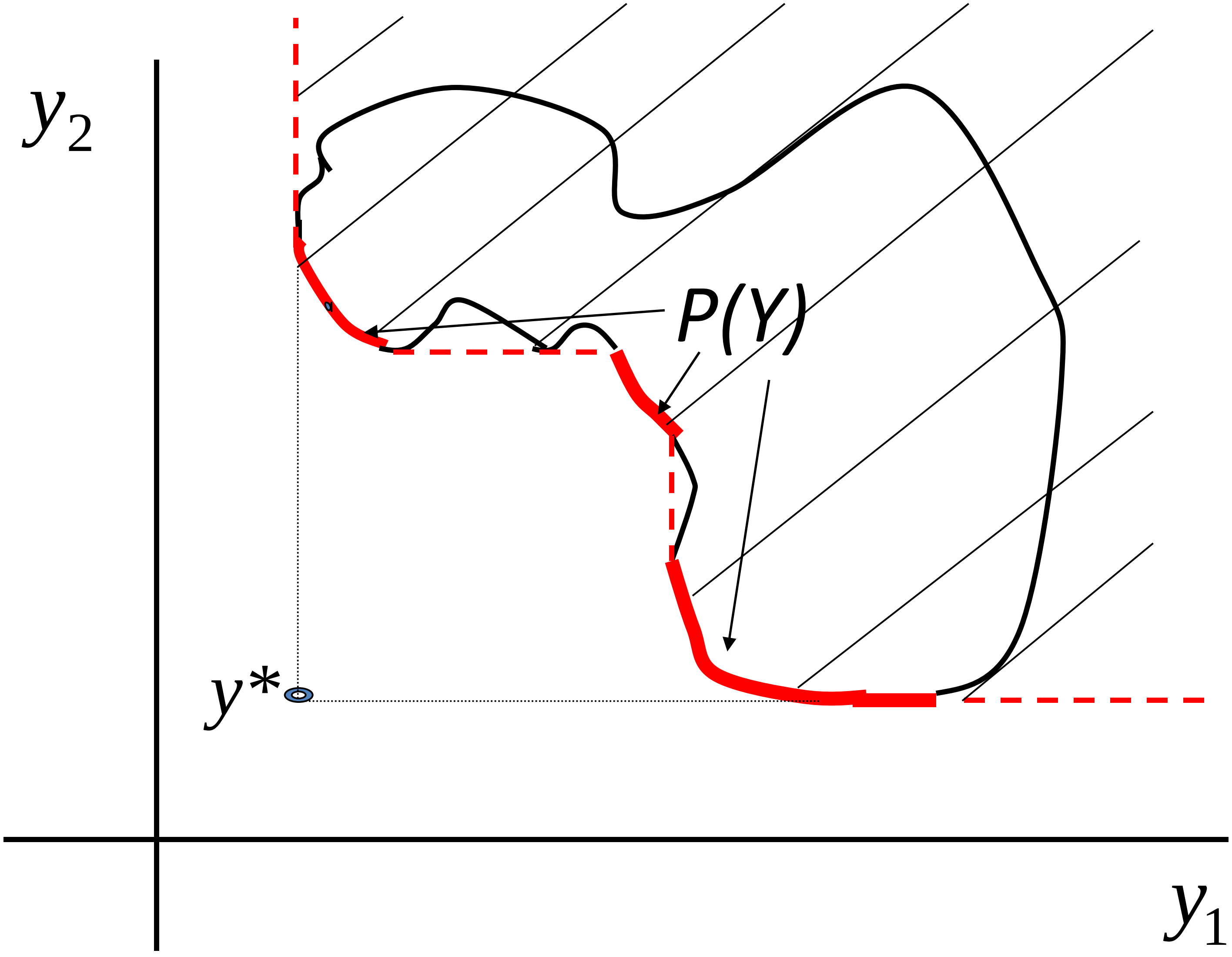
IDM technique

The IDM technique is based on ***approximating the Edgeworth-Pareto Hull (EPH), which has to be completed in advance***.

In the framework of IDM technique, decision makers study on-line various collections of bi-objective slices of EPH (decision maps).

In the IDM technique, the decision maps can be easily animated, zoomed, etc.

Edgeworth-Pareto Hull



Why one has to approximate the EPH instead of Pareto frontier?

Theoretical studies prove that the Pareto frontier is not stable with regard to disturbances of parameters of the MOO problem. See

Sawaragi Y., Nakayama H., Tanino T. Theory of multi-objective optimization. Orlando: Acad. Press, 1985.

Thus, the problem of Pareto frontier approximation is not correctly posed. As a rule, EPH is stable with regard to disturbances of parameters of the MOO problem.

IDM for linear and convex non-linear problems

In the case of linear and convex non-linear problems, multiple examples of EPH approximation and visualization in real-life decision/negotiation problems are given in

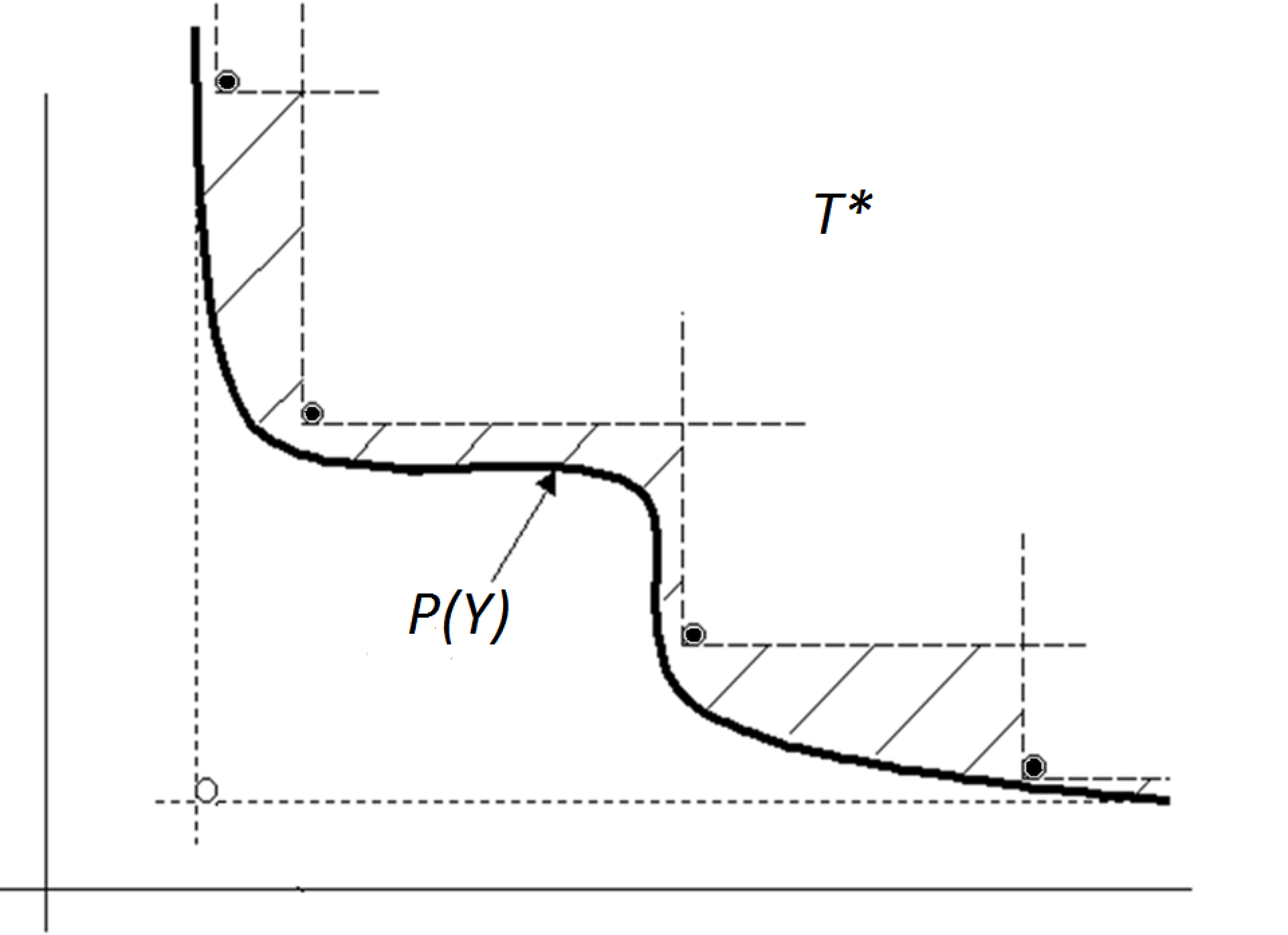
Lotov A.V., Bushenkov V.A., Kamenev G.K. Interactive decision maps. Approximation and Visualization of Pareto Frontier. Kluwer Academic Publishers. Boston / Dordrecht / New York / London. 2004. - 310 P.

EPH approximation and visualization in non-linear non-convex case

The EPH is approximated by the set T^* that is the union of the non-negative cones

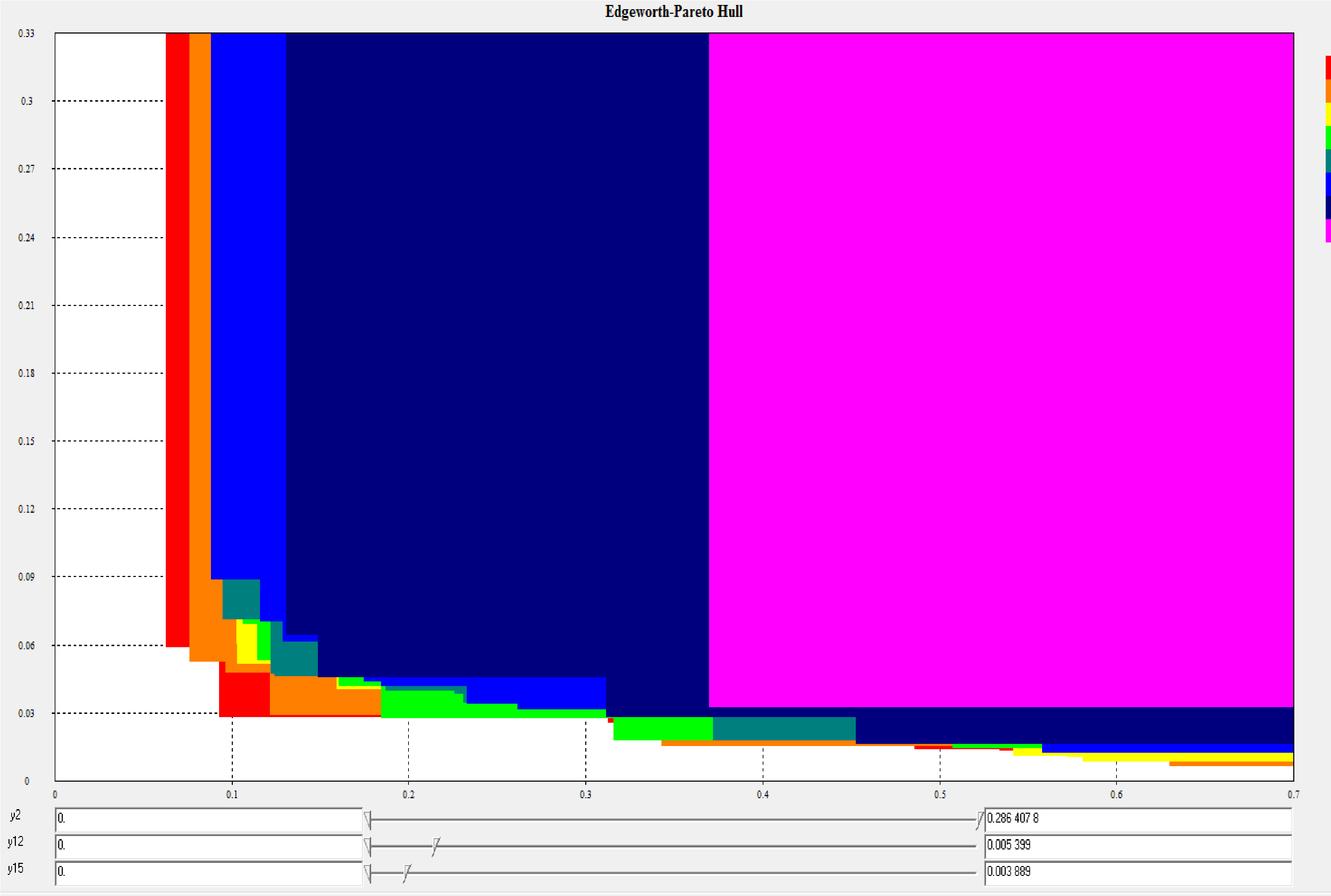
$$y + R_+^m$$

with apexes in a finite number of points y of the set $Y=f(X)$, which are close to $P(Y)$. The set of such points y is called the approximation base and is denoted by T.

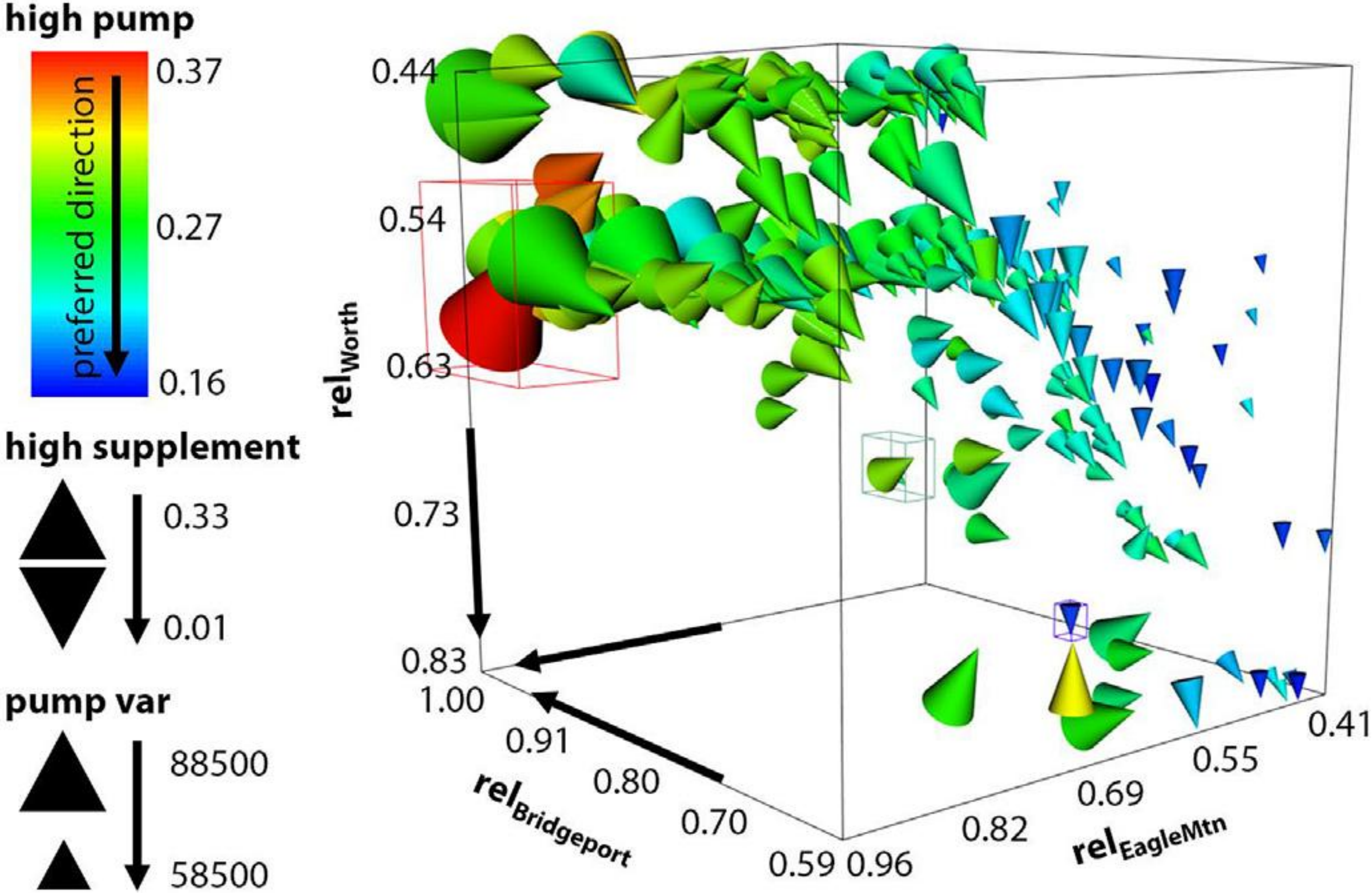


T*
P(Y)

Decision map with scroll-bars



Glyph plot



high pump
0.37
0.27
0.16
preferred direction
high supplement
0.33
0.01
pump var
88500
58500
0.44
0.54
0.63
0.73
0.83
rel$_{Worth}$
1.00
0.91
0.80
0.70
0.59
rel$_{Bridgeport}$
0.96
0.82
0.69
0.55
0.41
rel$_{EagleMtn}$

EPH approximation methods based on multi-start local optimization

These methods are used in complicated problems, in the framework of which a sufficient number of problems of global optimization of scalarizing function cannot be solved. The most practical EPH approximation method of this type, the two-phase method, is based on generation of random points x from X and application of local optimization of a scalarizing function for 'improving' the random point x. Statistical test of approximation quality are used to stop the iterations of the method.

V. E. Berezkin, G. K. Kamenev, A. V. Lotov. Hybrid adaptive methods for approximating a non-convex multi-dimensional Pareto frontier. Comput. Math. and Math. Phys. (2006) 46: 1918.

Examples of successful application of the two-phase methods in non-linear MOO problems with hundreds of decision variables

Lotov A., Berezkin V., Kamenev G., Miettinen K. Optimal control of cooling process in continuous casting of steel using a visualization-based multi-criteria approach // Appl. Math. Modelling. 2005. V. 29. No. 7. C. 653-672.

V. E. Berezkin, A. V. Lotov, and E. A. Lotova. Study of Hybrid Methods for Approximating the Edgeworth-Pareto Hull in Nonlinear Multicriteria Optimization Problems // Computational Mathematics and Mathematical Physics, 2014, Vol. 54, No. 6, pp. 919-930.

An example of a complicated real-life MOO problem, in which existing methods are not effective (Baikal Lake level regulation problem)

Lake Baikal

Lake Baikal located in Russia, in southern Siberia, is the largest freshwater lake by volume in the world. Baikal contains more water than the all North American Great Lakes together.

With a maximum depth of 1,642 m, Baikal is the world's deepest lake. It is considered among the world's clearest lakes, too.

Baikal is home to thousands of species of plants and animals, many of which exist nowhere else in the world. Lake Baikal was declared as UNESCO World Heritage Site in 1996.

Lake Baikal

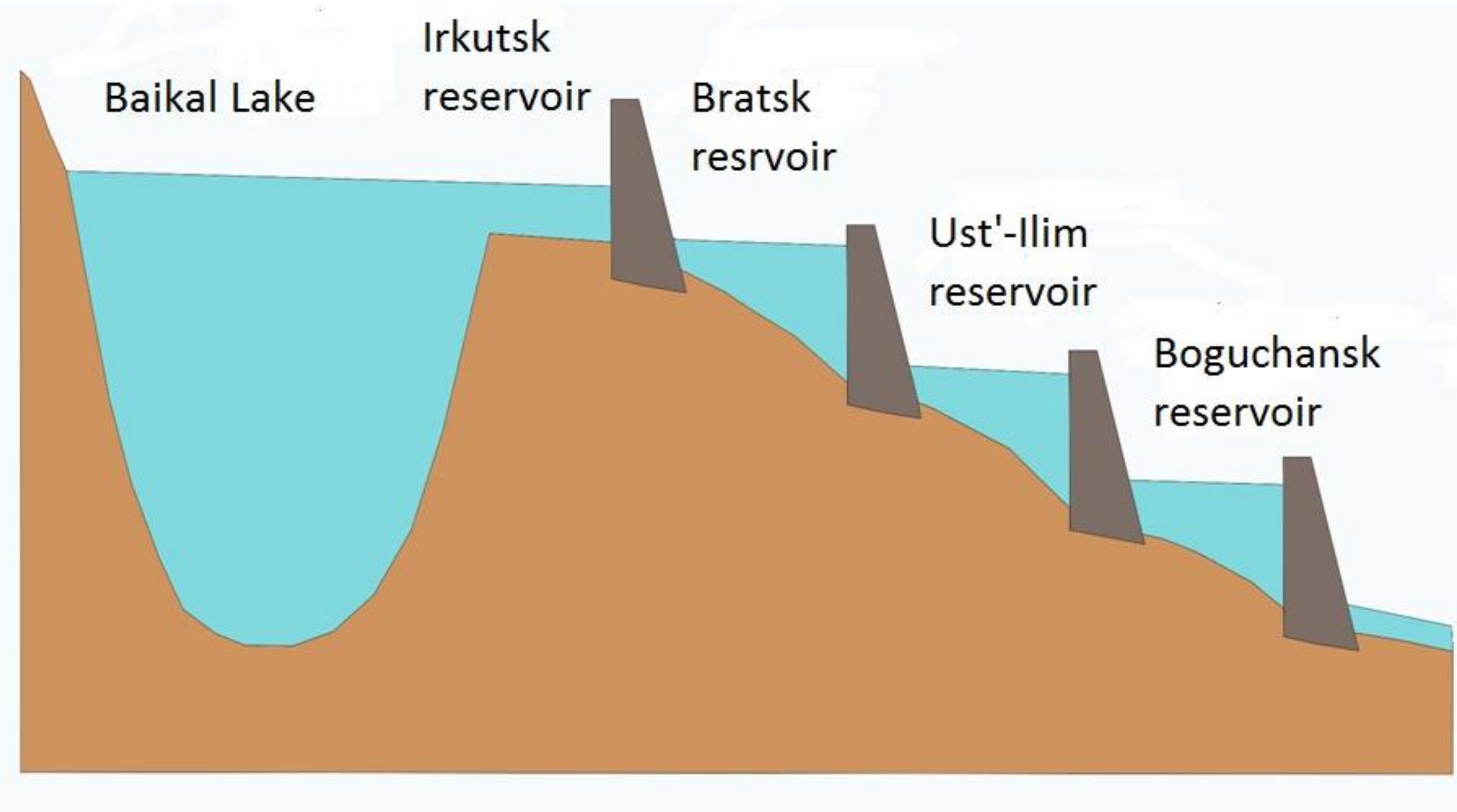

Россия
Russia
Kazakhstan
Mongolia
中国
China
S Korea
Japan
Afghanistan
Pakistan
India
Thailand
Usol'e-Sibirskoye
Angarsk
Irkutsk
Ulan-Ude

Lake Baikal and River Angara control problems

In the second part of the 20^{th} century, a cascade of dams and reservoirs was constructed at the Angara River, the only river that starts from the Baikal Lake.

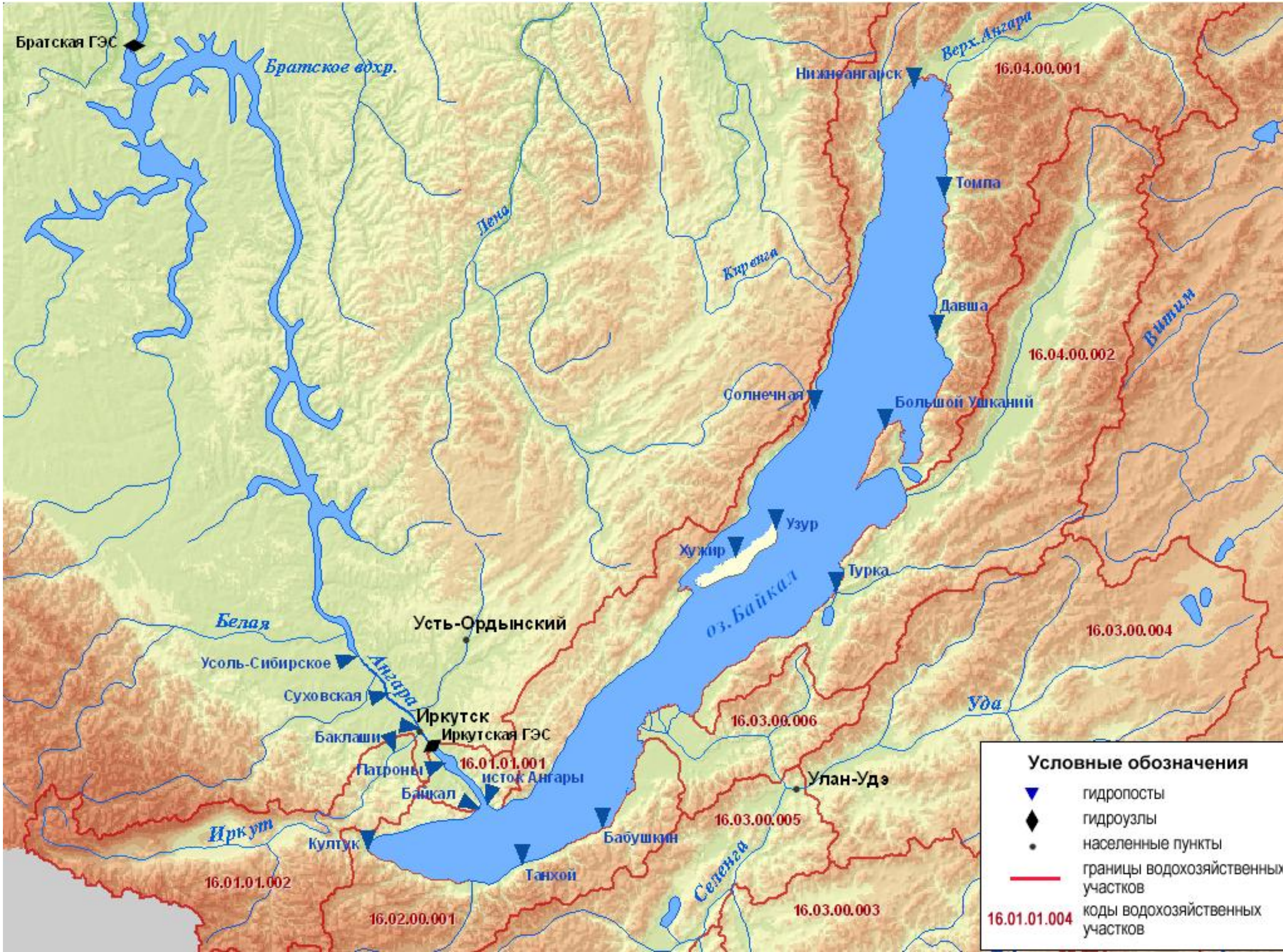
The cascade of reservoirs is intended for production of renewable hydroelectric energy, for providing transportation, for avoiding floods as well as for solving many other problems including regulation of the level of Baikal, for which the dam of the upper reservoir (Irkutsk dam) is used.

ANGARA RIVER CASCADE OF RESERVOIRS



Baikal: the total volume - 23600 cubic km, depth - 1642 m, average annual inflow - 60 cubic км

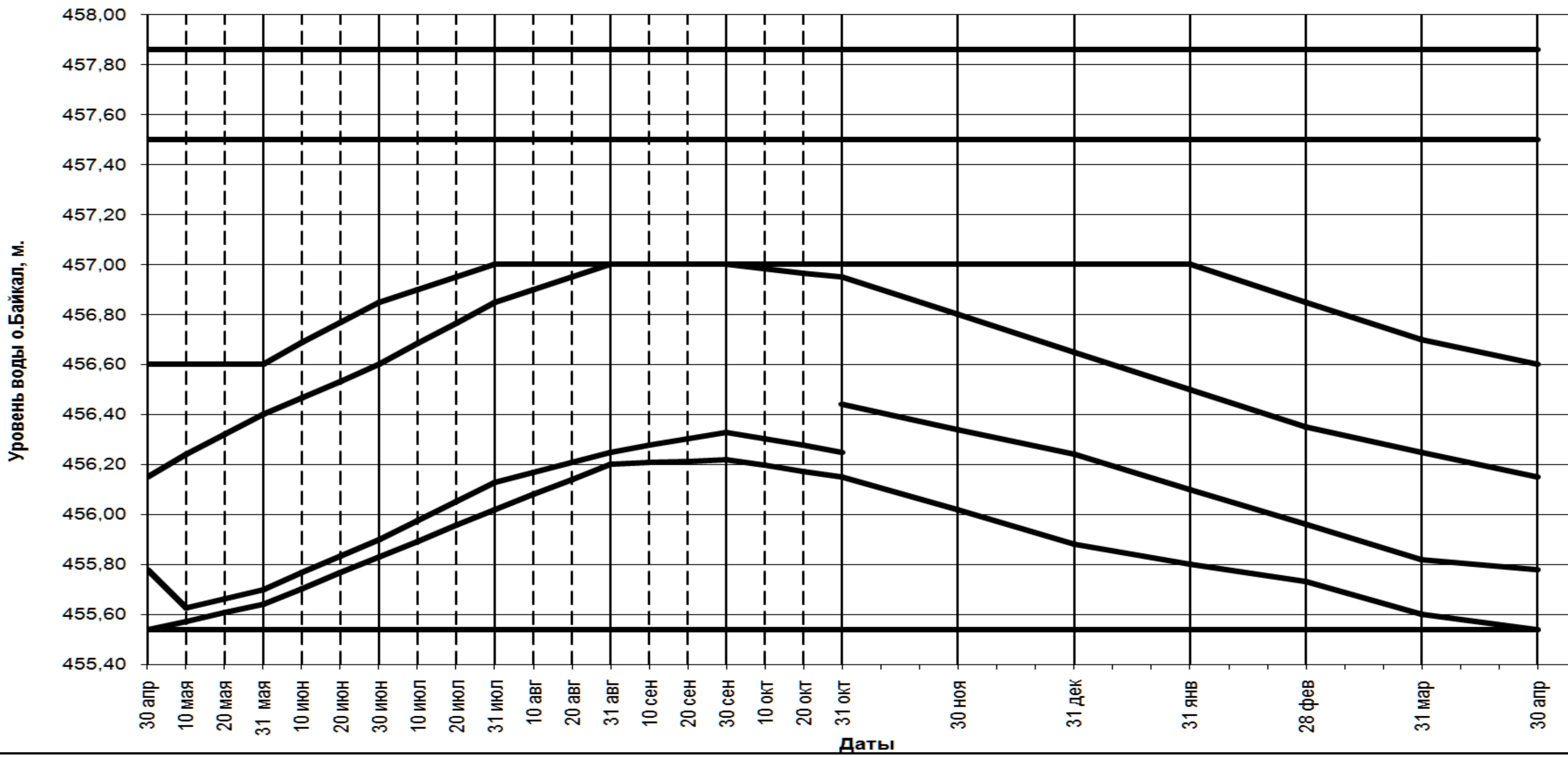
Baikal Lake, Irkutsk and Bratsk reservoirs



Problem of release rules design for a reservoir cascade

The level of the Baikal Lake and the flow of the Angara River is controlled by the release through the dams. Since one is not able to forecast the precipitation, water managers traditionally do not plan water release through the dam. Instead, they develop release rules, which relate the flow through a dam to the current level of the related reservoir. The rules may include the inflow forecast for the next time interval. The rules of different dams must be coordinated.

Dispatch schedules



Lake Baikal - Irkutsk reservoir

$$W^{t+1} = W^t + Q^t - R^t$$

balance equations

$$\nabla H^t = F(W^t)$$

Bathymetric function “Baikal lake level - volume of water”

$$\nabla \overline{H^t} = F(\nabla H^t, R^t)$$

Dependence of the level of upper pool on the Baikal level and dam release

$$\nabla \underline{H^t} = F(R^t)$$

Dependence of the level of lower pool on the Irkutsk dam release

$$N^t = F(\nabla \overline{H^t}, \nabla \underline{H^t}, R^t)$$

Electric power production

Here: *t* is time interval, W is volume of the Lake Baikal, Q is inflow, R is release

Reliability Criteria

$$y_i = \frac{1}{T}\sum_{t=1}^{T}\Theta(z_t^i)$$

where *i* =1,...,24, are the reliability criteria,
z is the violation value (*i* is requirement number, *t* is time interval),
and Θ(z) is Heaviside function, that is,

$$\Theta(z) = 0 \text{ for } z \le 0 \text{ and } \Theta(z) = 1 \text{ for } z > 0$$

The release rules

The release rule of a reservoir $R_t = f(W_t, p)$ is a piecewise constant function of the volume of water in the reservoir at the end of time-period, which depends on the release of the reservoir located upstream and the forecast of the side inflow during the current time-period.

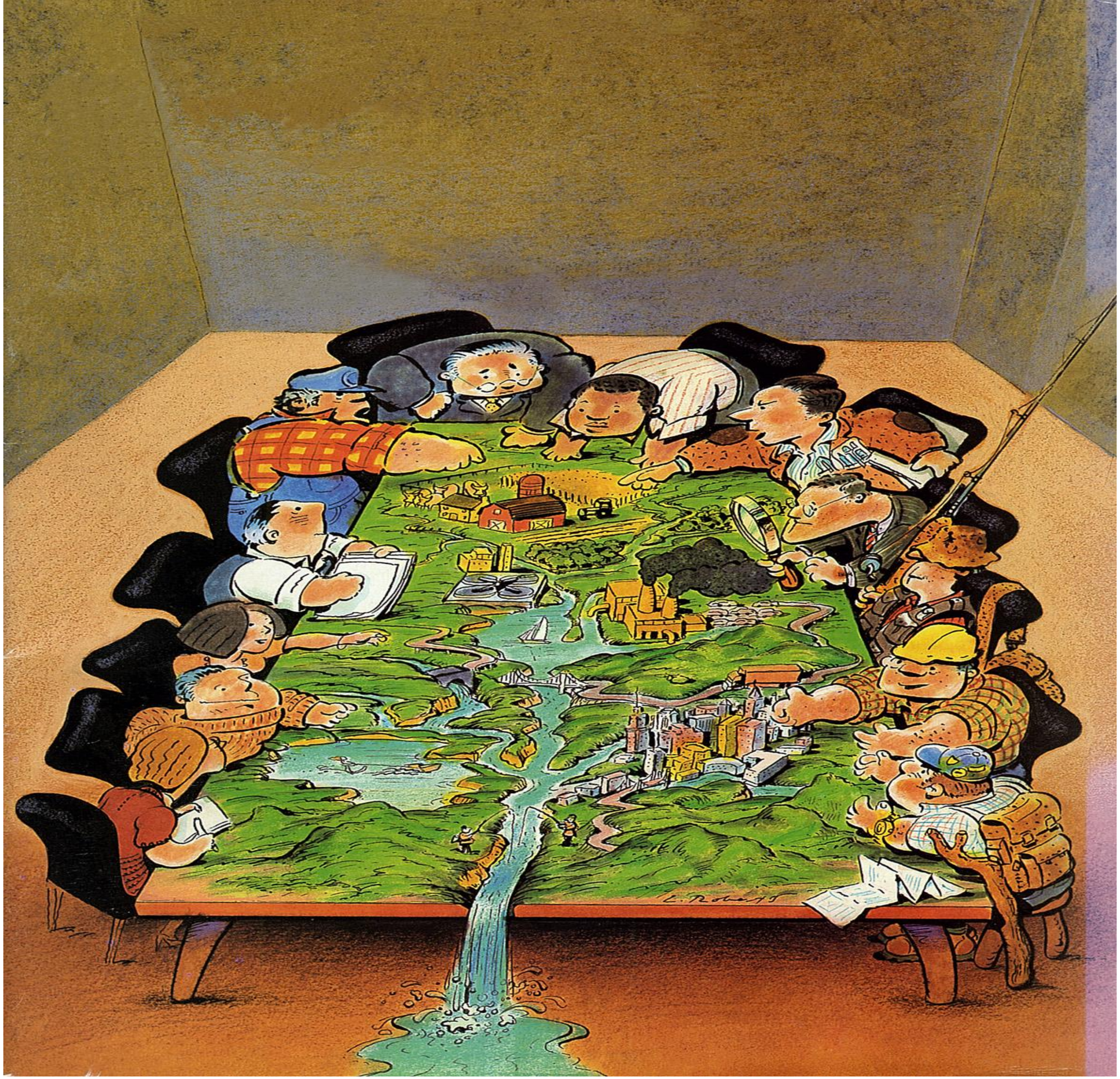
The rule includes parameters *p* that are subject of optimization. For any time-interval, 6 parameters of the rule are considered. Thus, we have got 132 parameters of the rule for 22 time-intervals for a reservoir. They are considered as the decision variables. Rules for three reservoirs contain about 300 decision variables.

Requirements

The parameters of the release rules must satisfy requirements of the industrial, agricultural and municipal users as well as safety and environmental requirements.

- Level of the Lake Baikal must be between 456 and 457 m over the ocean level
- Certain levels of electricity generation at the HPP must be provided;
- Transportation requirements must be satisfied;
- A set of environmental and municipal requirements must be satisfied;
- Floods at all reaches of the river must be avoided. Totally, 24 requirements are considered.

These requirements are in conflict. One has to find a balanced decision on release rules.
It is needed to inform participants of the decision/negotiation process on efficient decisions.
Who are these participants?



From Loucks D.P, van Beek E. Water resources systems planning and management. An introduction to methods, models and applications. Paris: UNESCO Publishing, 2005

The process of selecting a balanced decision in water control problems involves not only people who are formally delegated to make the decision.
People responsible for regional/municipal problems use opportunities available to them for influencing the resulting decision. Business is active, too. Finally, people who do not have senior positions but who live in regions which may be influenced by the decisions can also try to influence the decision.
The final decision is an informal compromise that is somehow reached by the various participants of the process.
We use the IDM technique to inform them on tradeoffs between the objectives.

Application of IDM:
Tradeoff visualization in the framework of Baikal Lake level regulation problem

Approximation of EPH may require a relatively long time (hours and even days).

After completion of this computing, fast on-line visualization of Pareto frontier can be carried out by displaying the bi-objective slices of EPH. The non-dominated frontiers of slices that are often called tradeoff curves are cross-sections of Pareto frontier.

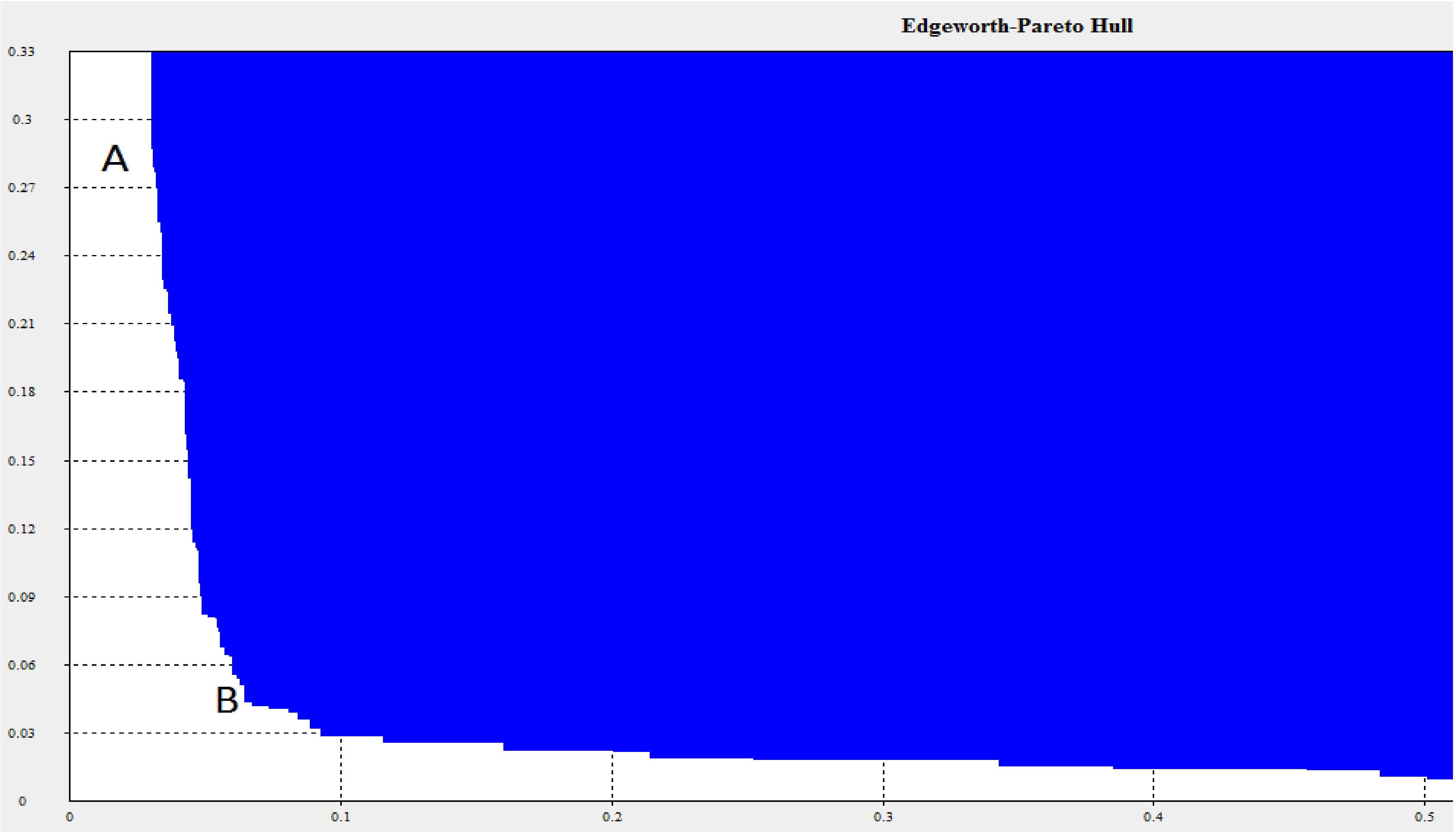
Tradeoff visualization

Let us first consider two objectives

y1 is the of the requirement that the level of the Baikal Lake must be between 456 and 457 m, that is, fraction of time intervals with violation of the requirement (horizontal axis)

y5 is the reliability of the requirement to electricity generation at the Irkutsk HPP, that is, fraction of time intervals with violation of the requirement (vertical axis)

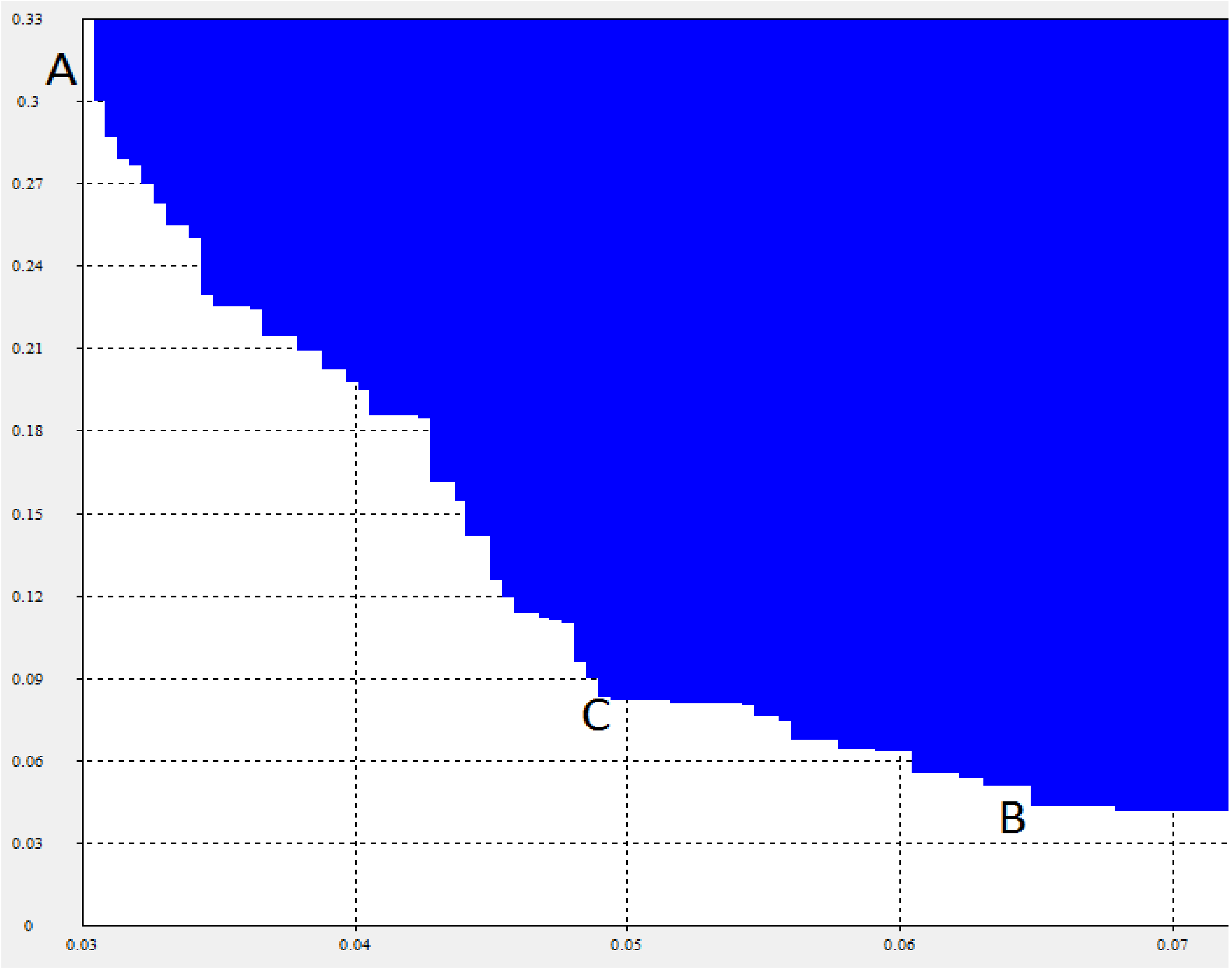
RELIABILITY OF LEVEL VERSUS ENERGY PRODUCTION



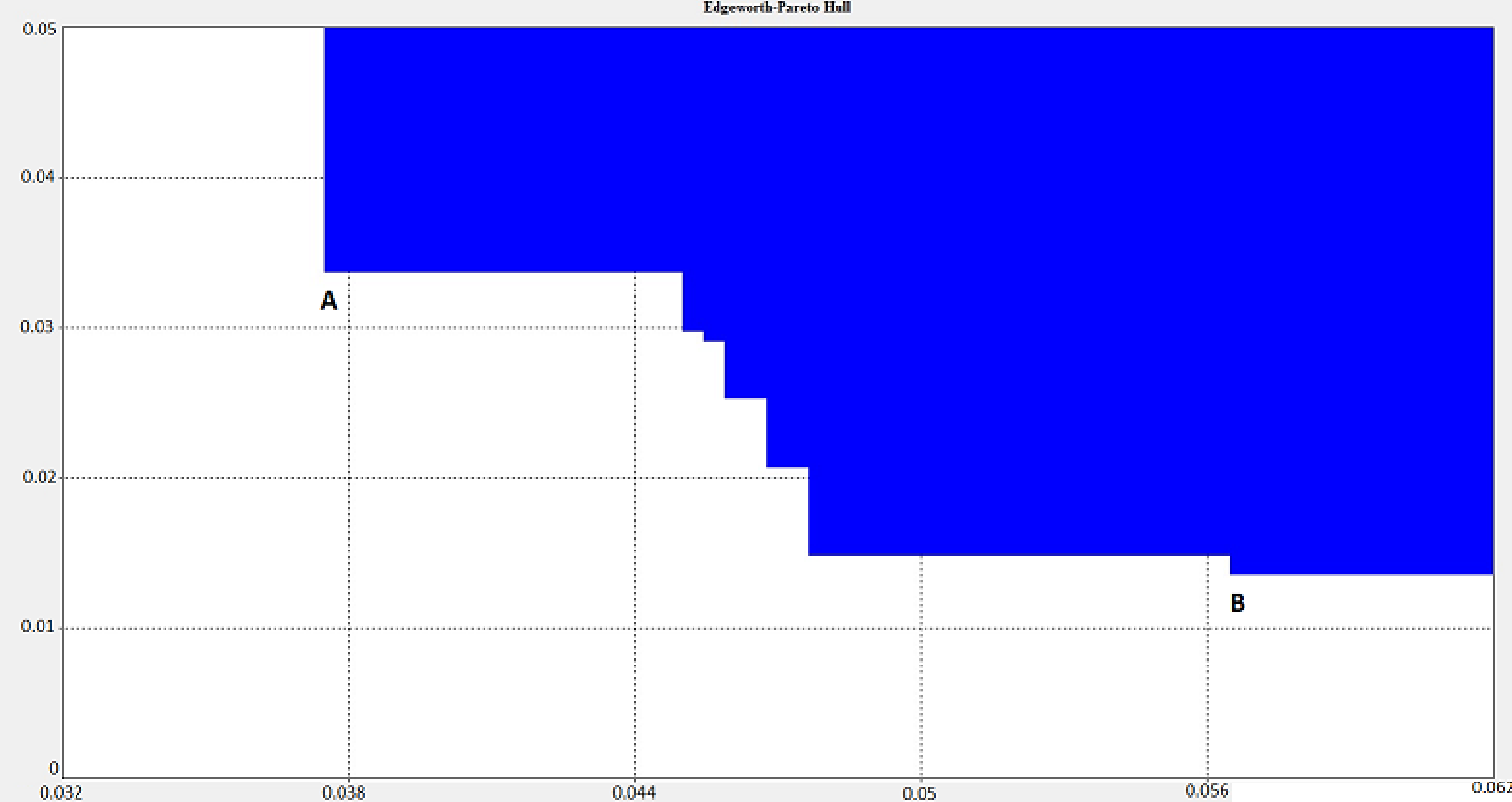
HORIZONTAL AXIS IS THE FRACTION OF FAILURES IN LEVEL OF THE LAKE, VERTICAL AXIS IS THE FRACTION OF FAILURES IN ENERGY PRODUCTION

A=(0.03, 0.29), B=(0.065, 0.04)

Zoomed image

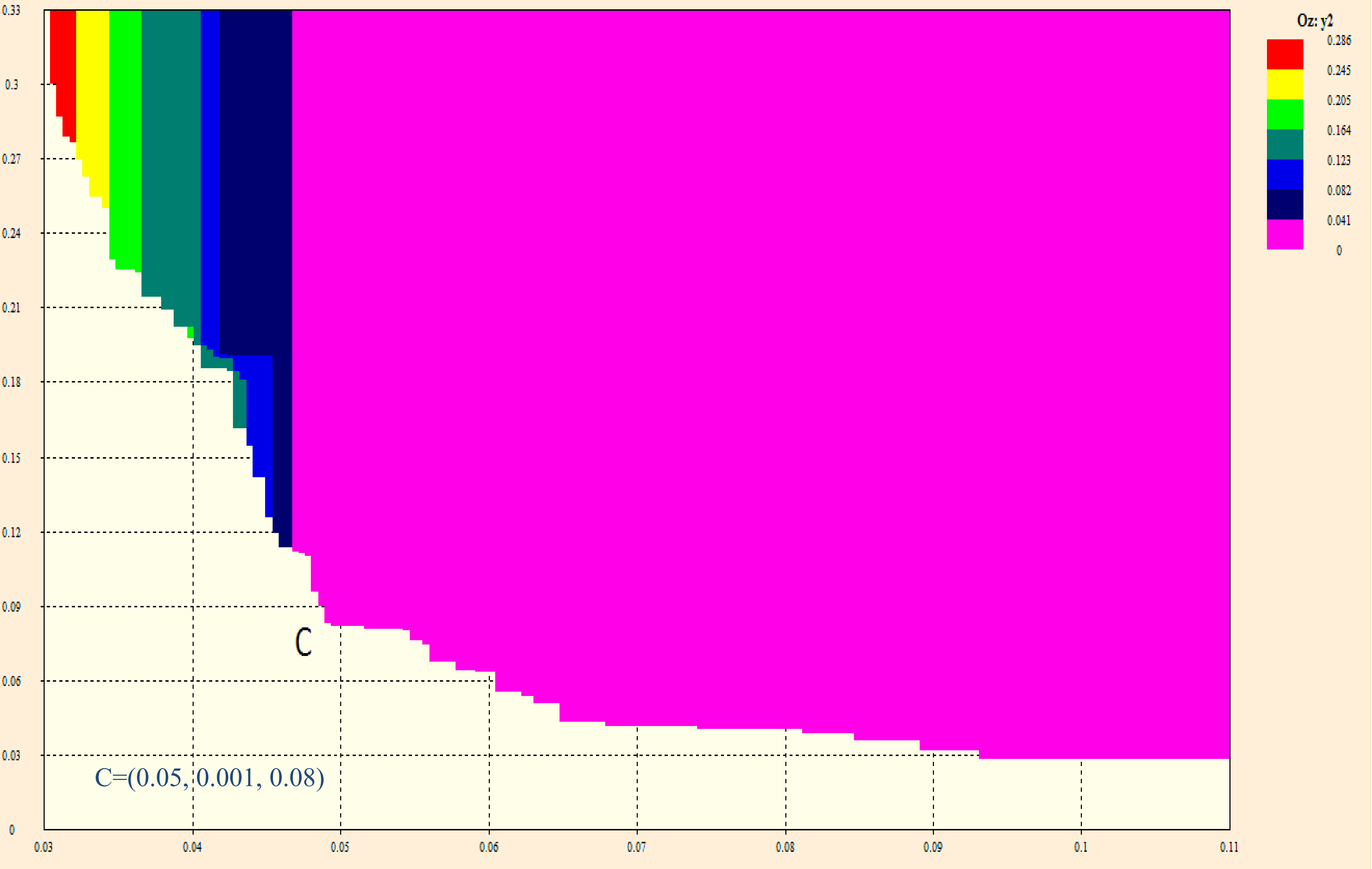


A=(0.03, 0.29), B=(0.065, 0.04) C=(0.05, 0.08)



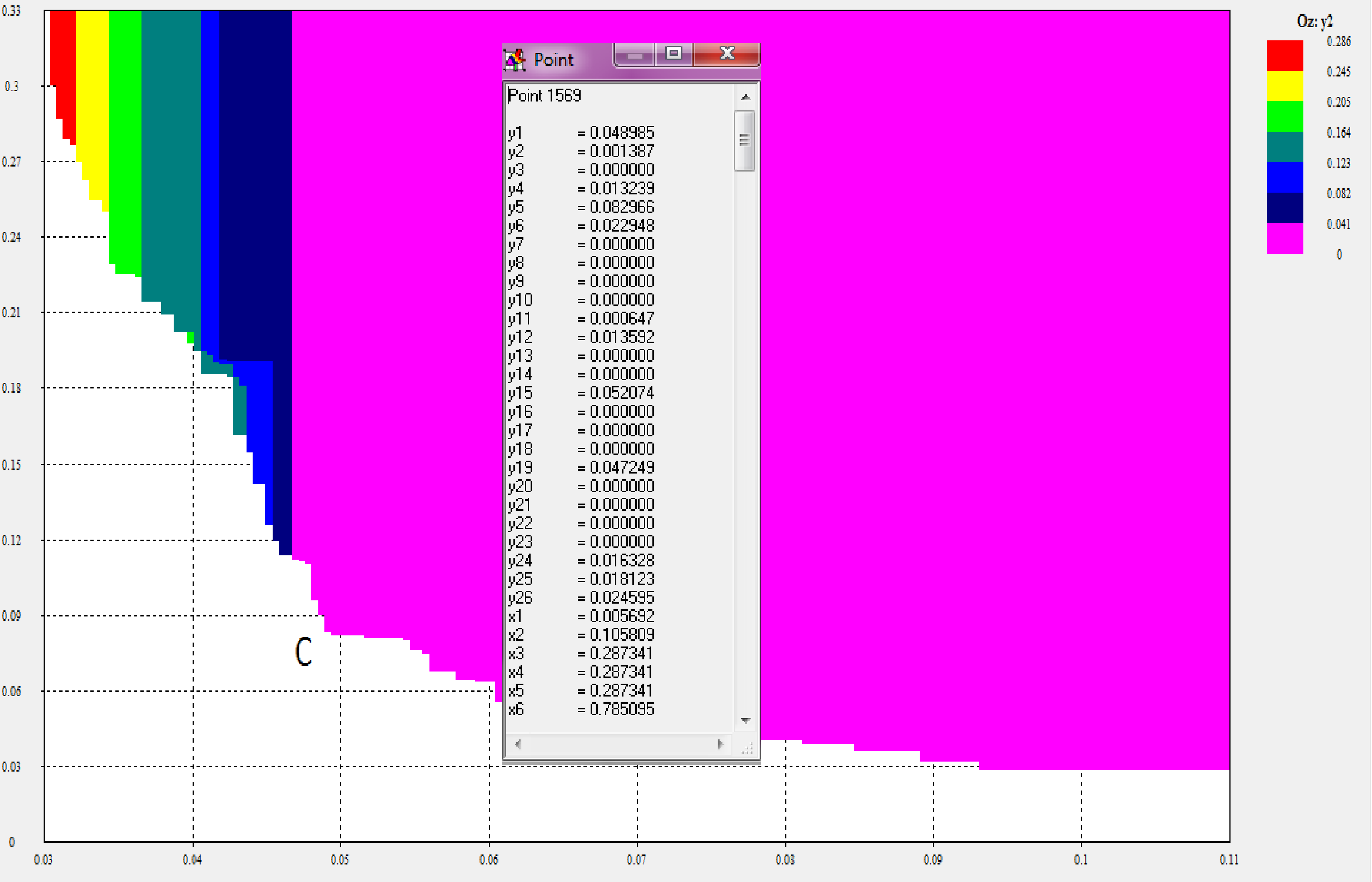
Example of a single bi-objective slice for the reliability of the Baikal level (horizontal axes) and of the navigation level of the upper pool of the Ust'-Ilim dam (vertical axes). Pareto frontier of the slice is the tradeoff curve between these two criteria.

Decision map: Colored scale is the fraction of violation intervals of minimum release through the Irkutsk dam



Feasible goals method, i.e. identification of the preferred feasible objective point directly at a decision map.

The same decision map plus the window with the decision which results in point C



EPH approximation methods

Black box models

We develop methods that do not use particular properties of the model, which is considered as a black box (computational module). The module computes outputs (objective values) for given inputs (decision variables). Thus, one can use the black box for generating objective vectors for random decisions, to carry out the simulation-based local optimization of scalarizing functions or for genetic optimization. A very broad scope of non-linear models can be considered and studied as a black box.

Optimization criteria

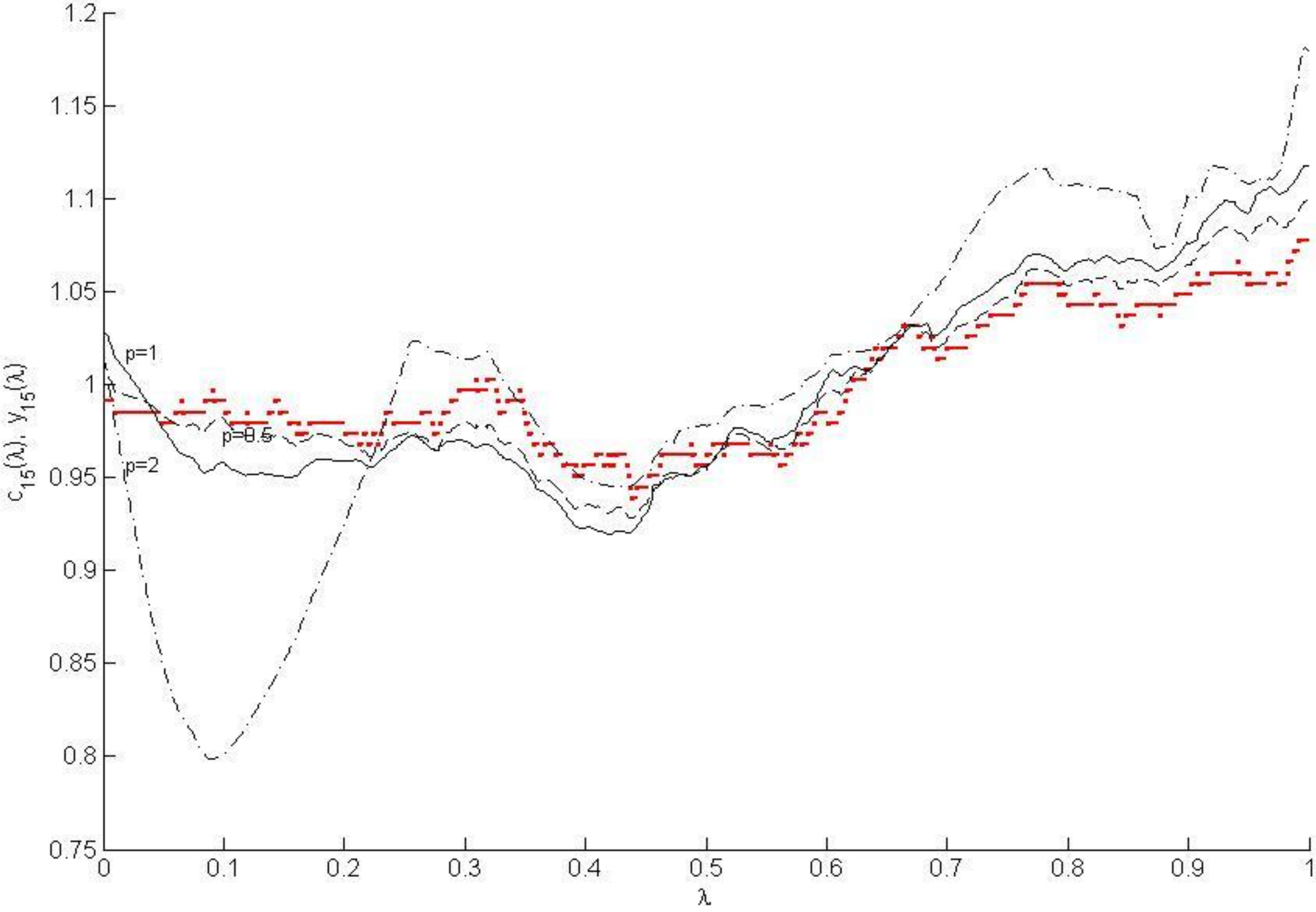
Optimization criteria (objectives) are equal to the fraction (or percentage) of time intervals with failure to satisfy the requirements, i.e.

$$y_i = \frac{1}{T}\sum_{t=1}^{T}\Theta(z_t^i)$$

where z_t^i is the value of violation of the *i*-th requirement during the time interval *t* and $\Theta(\cdot)$ is Heaviside function. Considering 100 years long time period, one gets the value of any objective as the sum of more than 2200 Heaviside functions. These objective functions must be minimized.

The number of objective functions is 24.

Objective function (red) and surrogate functions



General discussion of EPH approximation problem in Baikal problem

In the Baikal problem, existing EPH approximation methods, such as multi-objective multi-start methods, proved to be not able to construct a good approximation of the EPH. This experimental evidence is related to the traps provided by local minima of the scalarizing functions. The number of local minima in the Baikal problem is extremely large. Experiments show that the well-known genetic method NSGA-II is not sufficient as well.

This is why special methods for complicated problems, as the Baikal Lake problem, are needed.

Optima injection method

Optima Injection method

The gradient-based optimization method is used for searching for optima (optimal decisions) for individual objectives. Solving the problem of searching for such decisions takes a lot of time in the problems under study, but it turned out to be possible to solve it. Then, the optimal decisions are injected into population of a genetic algorithm. For this reason, the proposed method has got the name Optima Injection (OI) method.

OI method

Step 1. Preparation of the initial set for GA
The optimal decisions are constructed for all *m* by solving global minimization problems for the individual objectives. The initial population of a genetic method includes them as well as any convenient number of stochastic points from *X*.
Step 2. Genetic algorithm NSGA-II (proposed by K. Deb) complemented with the new stopping rule and the periodic injection of optima computes the EPH approximation. Optima are injected if the number of the NSGA-II iteration equals to a given number *K*. The stopping rule is based on comparison of the maximal distance ε_k of the new objective points obtained at the *k*-th iteration with a given value ε^* . Namely, if $\varepsilon_k < \varepsilon^*$, then stop Step 2.

Comparison of OI method with NSGA-II algorithm

The NSGA-II algorithm was compared with the simplified OI method (injection of optima into the initial population only). EPH approximations were constructed with the help of both methods while the number of computing the (vector) objective function was the same. In the both methods, the initial population consisted of 10 000 points. Computing of the optima required about 10 million calculations of the objective function, and 3269 iterations of step 2 ($\varepsilon^* = 0.005$) required about 42 million calculations. This number of calculations of the objective function was sufficient for 42 thousand iterations of the NSGA-II algorithm.

Comparison of approximations obtained by OI method and NSGA-II algorithm

The first comparison technique is based on deviation of approximations from a Pareto-optimal objective point known in advance(it was found by a different method). It turned out that the deviation δ_N of the approximation constructed by the NSGA-II algorithm was about 10 times greater than the deviation δ_o of the approximation constructed by the OI method.

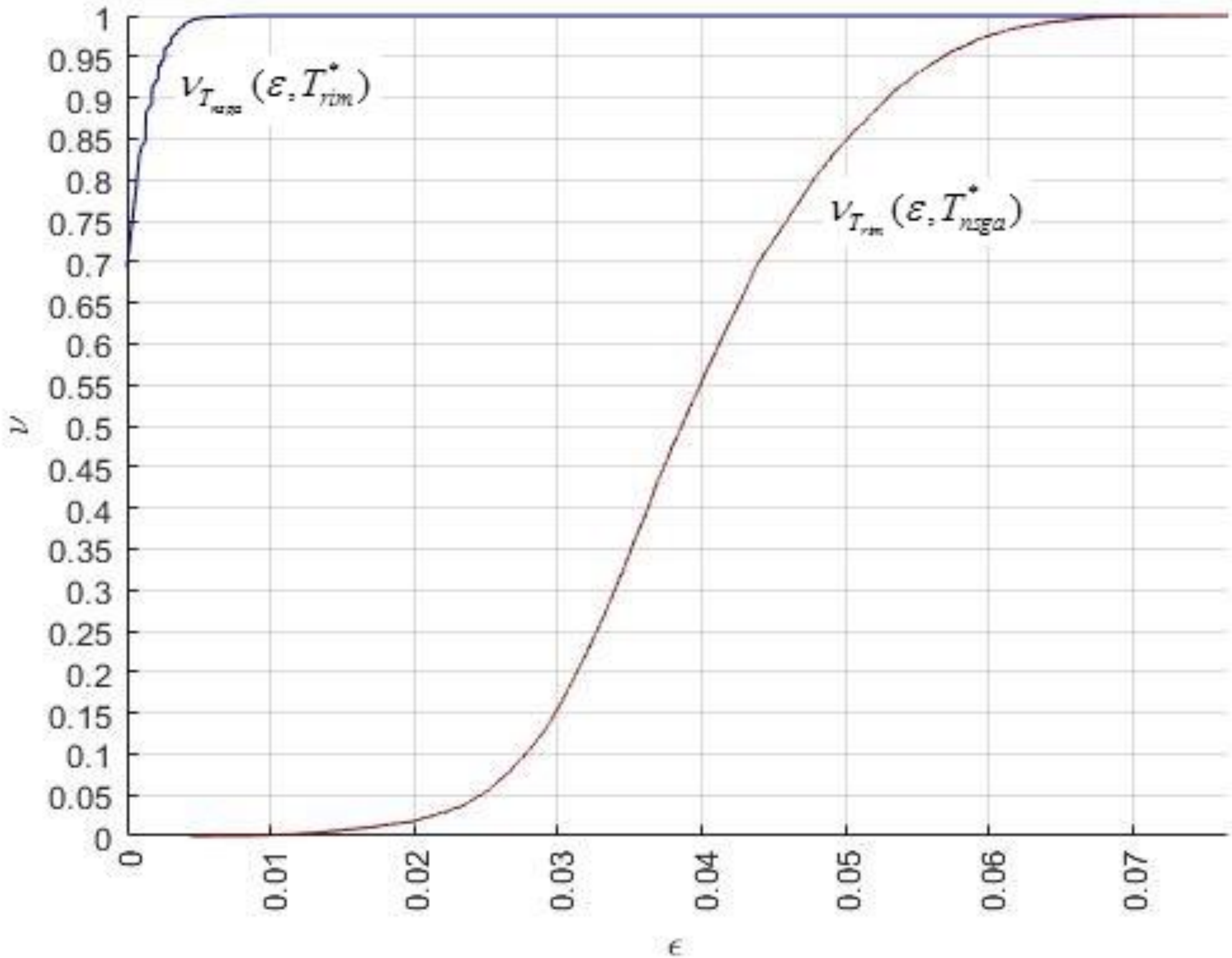
Approximations of lesser precision were compared, too. For 33 million calculations of the objective function the value of δ_o was 6 times less than the value of δ_N. For example, for 33 million calculations of the objective function the value of δ_o was 6 times less than the value of δ_N. For 5.5 million calculations the value of δ_o was only 1.5 times less than the value of δ_N. It is clear that the OI method is more efficient when high approximation precision is needed.

Comparison of approximations obtained by OI method and NSGA-II

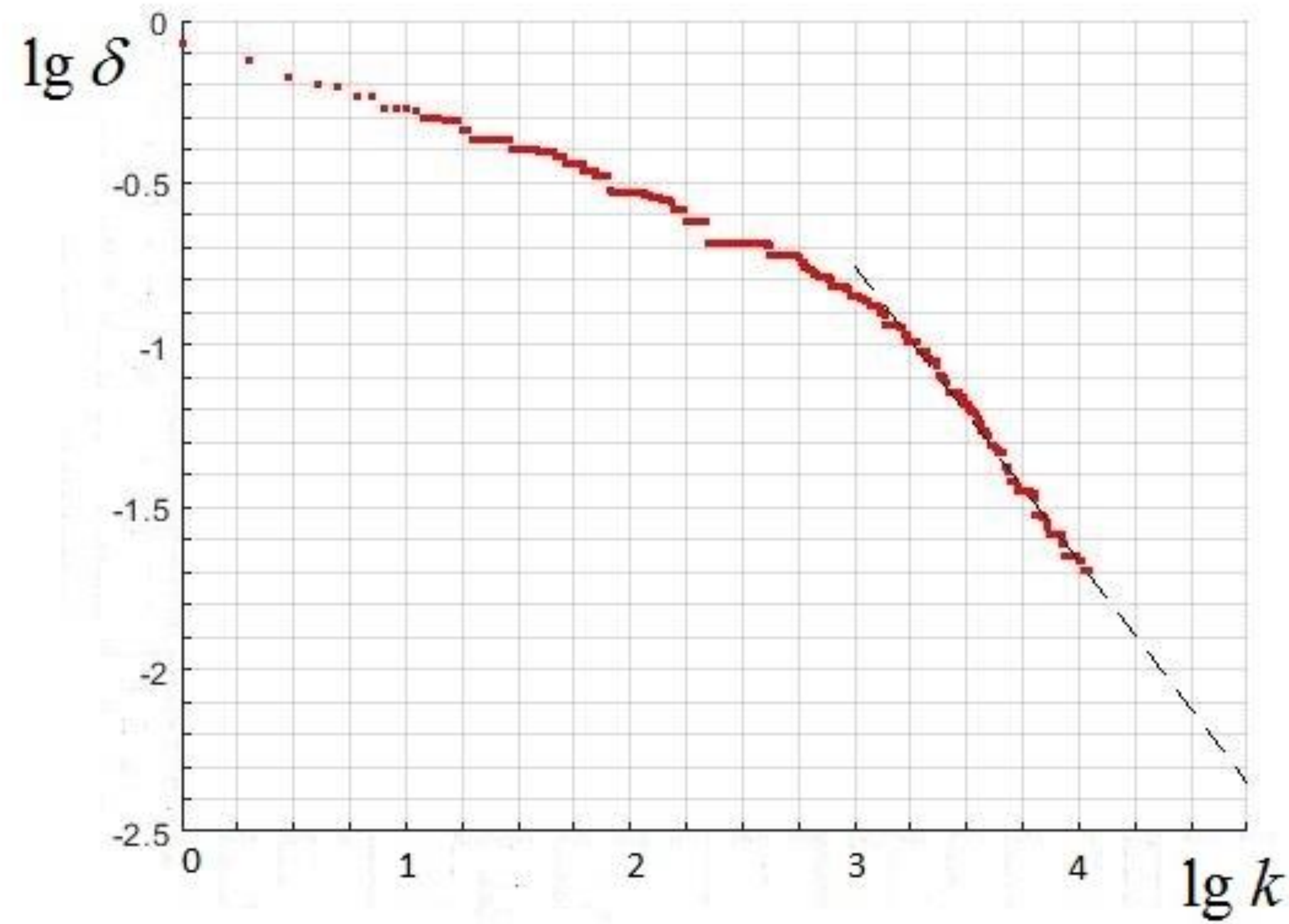
The second idea is based on the comparison of two EPH approximations with the help of the inclusion functions, which graph shows what fraction of points of the base of the first approximation belongs to the ε - vicinity of the second EPH approximation.

Locking at the graphs, one can easily find that the approximation constructed by the OI method is much better, than the other one.

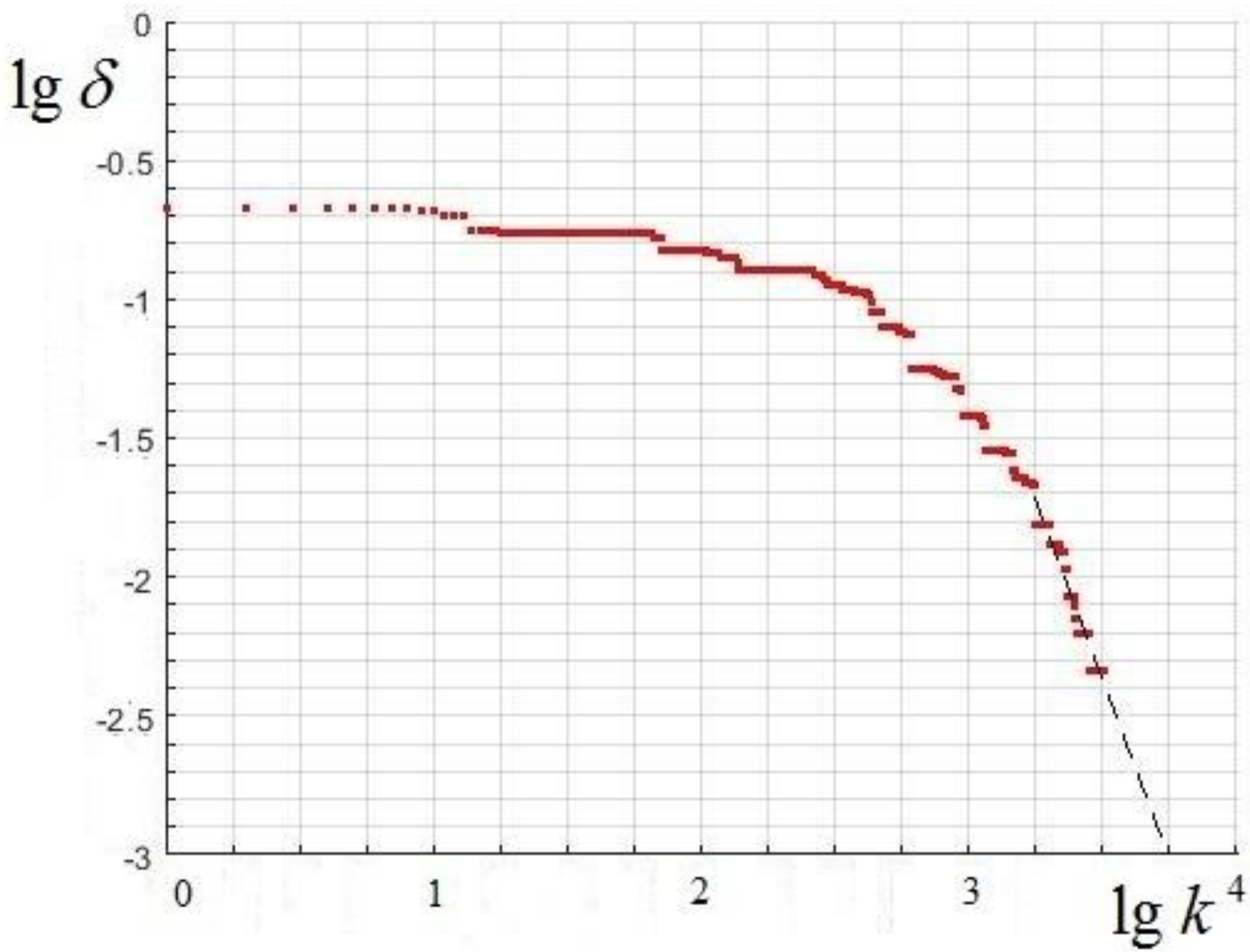
Inclusion functions



Comparison of convergence rates.
For NSGA-II we have $\delta_N \sim 10^2/k$ where k is the number of NSGA-II iterations



For OI method we have $\delta_N \sim 10^5/k^2$ where k is the number of OI iterations



Launch Pad Method for EPH approximation in complicated non-linear non-convex MOO problems

Launch Pad Method (LPM)

The main feature of the LPM consists in constructing the Launch Pad, i.e. subset X_0 of X, which is used as the set of random starting points in multi-start local optimization-based EPH approximation method. Several approaches for constructing the Launch Pad were proposed. One of them is based on approximation of the EPH by using the OI method. The set of obtained decisions is used as Launch Pad. It turned out that multiple local minima do not hinder to find a good approximation of $P(Y)$.

Steps of LPM

Step 1. The extremes of individual objectives are constructed by using global optimization methods and used as a part of initial population of genetic algorithm .

Step 2. A genetic algorithm NSGA-II is used for constructing an approximation of the Pareto set in decision space, i.e. such subset X_0 of X that points of $Y_0=f(X_0)$ approximate the Pareto frontier. The obtained set X_0 is used as the Launch Pad, i.e. the set, from which local optimization is started.

Step 3. Iterations of the local optimization are carried out from random points of X_0 . Results of local optimization (points in objective space) are used for constructing the approximation base of the EPH approximation.

An iteration of Step 3 of LPM

The number of points N in random sample H_N, the desired precision ε^* as well as a local optimization algorithm, which result is denoted by $\Phi(x)$, must be given in advance.

An iteration. A current approximation base T must be constructed at the previous iteration. By this, the current EPH approximation T^* is given, too.

1. A random sample H_N of N points of X_0 is generated.
2. *Testing* T^*. If $\varepsilon_{\max}=\delta(\Phi(f(H_N)),T^*) <\varepsilon^*$, then stop.
3. *Forming the new approximation base*. The new approximation base consists of non-dominated points of unification of T and $\Phi(f(H_N))$.

Start the next iteration.

Comparison of LPM with NSGA-II algorithm

The LPM method was compared with the NSGA-II algorithm. Two EPH approximations were constructed with the help of both methods while the number of the objective function calculation was the same. In the both methods, the initial population consisted of 10 000 points. Computing the extreme decisions for all $m=24$ objectives at Step 1 of LPM required about 10 million calculations of the objective function. Step 2 of LPM (NSGA-II) required about 5 million calculations. Finally, Step 3 required 3 million calculations. In total, we spent 18 million calculations of the objective function. This number of calculations of the objective function was sufficient for about 1800 iterations of the NSGA-II algorithm.

Comparison of LPM with NSGA-II

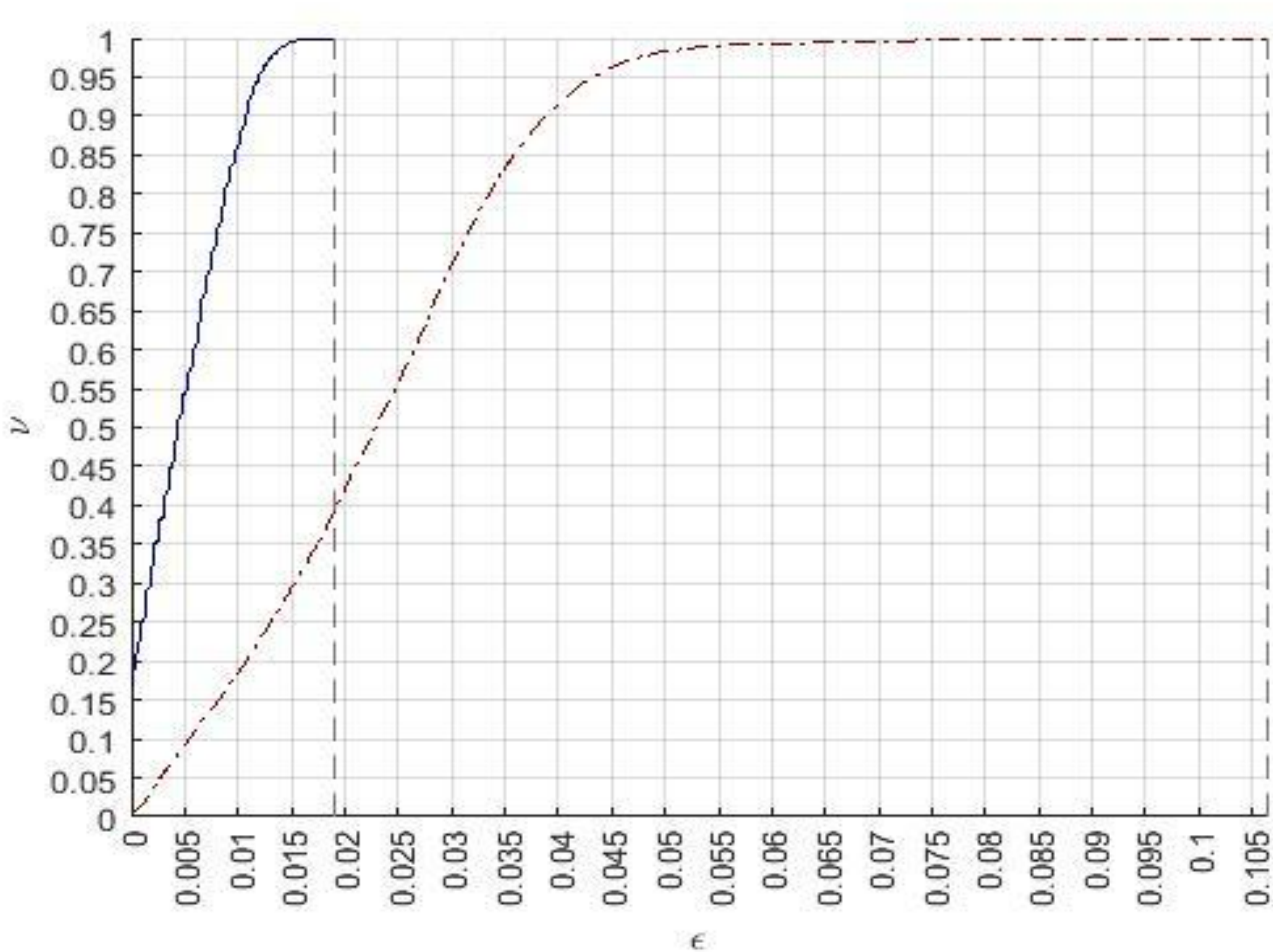
The first technique is based on the comparison of deviation of a known feasible objective point from approximations. It has turned out that 18 million calculations of the objective function were sufficient for LPM to decrease the deviation denoted by δ_L till 0.01. This number of calculations of the objective function were sufficient for 18 thousand iterations of NSGA-II that resulted in the deviation $\delta_N = 0.1$. Thus, LPM turned out to be about 10 times more effective.

Comparison of LPM with NSGA-II algorithm (cont-n 2)

The second technique is based on a direct comparison of two EPH approximations. The comparison is carried out with the help of the inclusion functions. A graph shows what fraction of points of the approximation base of one approximation belongs to the ε- vicinity of the second EPH approximation.

Locking at the graphs, one can easily find that the approximation constructed by the LPM is much better, than the other one.

Inclusion functions



Thanks for your attention!