Jul 1st, 12:00 AM

Response Surface approximation and Interactive Decision Maps for water quality planning

A. Castelletti

D. Limosani

Alexander V. Lotov

X. Quach

R. Soncini-Sessa

Response Surface approximation and Interactive Decision Maps for water quality planning

A. Castelletti[a], **D. Limosani**[a], **A.V. Lotov**[b], **X. Quach**[a], **and R. Soncini-Sessa**[a]

[a]*Dipartimento di Elettronica e Informazione, Politecnico di Milano, P.zza Leonardo da Vinci, 32, 20133, Milano, Italy (castelle@elet.polimi.it, dlimos@hotmail.com, quach@elet.polimi.it, soncini@elet.polimi.it)*

[b]*Dorodnicyn Computing Centre, Russian Academy of Sciences Vavilova str., 40, 119333, Moscow, Russia (lotov10@ccas.ru)*

Abstract: This study presents a new interactive procedure for supporting Decision Makers (DMs) in environmental planning problems involving large, process-based, dynamic models and many (more than two) conflicting objectives. Because of such features of the model, computationally-onerous simulations are the only feasible way of analysis, while the multi-objective nature of the problem entails the combined use of optimization techniques and appropriate tools for the visualization of the associated Pareto frontier. The procedure proposed is based on the iterative improvement of the current best compromise alternative based on interactions with the DM. At each iteration, the DM is informed about the Pareto frontier of a local multi-objective optimization problem, which is generated by linearizing the response surfaces that describe the objectives and constraints of the original planning problem. Interactive visualization of the multidimensional Pareto frontier is used to support the DM in choosing the new best compromise alternative. The procedure terminates when the DM is fully satisfied with the current best compromise alternative. The approach is demonstrated in Googong Reservoir (Australia), which is periodically affected by high concentrations of Manganese and Cyanobacteria. Results indicate that substantial improvements could be observed by simply changing the location of the two mixers installed in 2007 and adding another pair of mixers.

Keywords: Multi-criteria decision methods; response surface; Pareto frontier visualization; environmental planning; water quality.

1 INTRODUCTION

In this paper we propose a novel, interactive approach to environmental decision-making, in which the procedural approach proposed by Castelletti et al. (2010) is revisited as an iterative, local improvement of the current best compromise alternative (BCA), and combined with interactive decision map (IDM) (Lotov et al., 2004) techniques for supporting decision-making involving many ($q \geq 3$) objectives and requiring a large, process-based simulation model to compute the anticipated effects of the feasible alternatives. More precisely, given the current BCA, a small set of alternatives, opportunely selected in a neighbourhood of the BCA, is simulated through the process-based model and the corresponding values of the objectives computed. Using these simulation data, a local linear approximation of the RS around the BCA is identified. The Pareto frontier for the MO optimization problem formulated with the local linear RS approximation over the neighbourhood of the BCA is displayed by using IDMs. The DM selects the best tradeoff point on this Pareto frontier and the corresponding Pareto-efficient alternative is obtained by inverting the linear approximation of the RS. This alternative is then simulated with the process-based model and, if the DM is satisfied with the simulation results, it is assumed to be the new current

BCA. The procedure is repeated until a given termination test is satisfied. The proposed approach is demonstrated through a case study in Australia (Section 4).

2 PROBLEM FORMULATION

Consider an environmental planning project, in which we need to fix the values of n decision variables u_i, with $i = 1, \ldots, n$. A planning alternative is a decision vector $\mathbf{u} = |u_1, \ldots, u_n|$ defined over a feasibility set $U \subseteq \mathbb{R}^n$ reflecting any existing technical or economic constraints on the decision values. The dimension n of the decision vector is assumed to be relatively small, but the set U might contain an infinite number of alternatives. A dynamic model $\mathcal{M}$ is available to simulate the system behaviour under each alternative. $\mathcal{M}$ transforms an alternative $\mathbf{u} \in U$ into a collection of output values $\mathcal{B}$, i.e. $\mathcal{B} = \mathcal{M}(\mathbf{u})$. Generally, $\mathcal{B}$ is composed of the time trajectories of spatially distributed outputs. Model $\mathcal{M}$ is given in the form of a computational code, which may require several days, to compute the output $\mathcal{B}$ for each single alternative $\mathbf{u}$. The information in $\mathcal{B}$ can be used to evaluate the alternative $\mathbf{u}$ and, by comparison with the other feasible alternatives, to choose the most preferable one. In general, two different approaches can be adopted to take the decision: 'What-if' analysis and Multi-Criteria approach. The later is better, but is hardly applicable when model $\mathcal{M}$ is computationally too onerous and/or the visualization of the Pareto frontier is made difficult by the high ($q \geq 3$) number of objectives.

To overcome these limitations, in this paper we propose an a-posteriori and interactive decision-making approach that iteratively improves the current BCA by combining the concept of Response Surface methodology (Box and Wilson, 1951), and the Interactive Decision Map technique (Lotov et al., 2004) , to compute the approximated Pareto frontier and support the DM in selecting, via visualization, the point she prefer on this frontier, and compute the associated BCA. Conceptually, at any given iteration (given the current BCA), the approach consists of two phases:

Learning: a sample of input (the alternative $\mathbf{u}$) and output (the associated indicator vector $\mathbf{y}$) points is generated through a set of suitably (in a sense that will be clarified later) designed simulation experiments, conducted on the process-based model $\mathcal{M}$. The sample is then used to identify an approximation $\hat{\mathbf{f}}(\cdot)$ of the RS $\mathbf{f}(\cdot)$.

Planning: an approximate solution to the above problem is obtained by substituting $\mathbf{f}(\cdot)$ for the current approximation $\hat{\mathbf{f}}(\cdot)$ and shrinking the set of alternatives considered. The resulting Pareto frontier is analyzed by the DM, who chooses a new point on it, which is used to compute the new associated BCA.

This iterative process can be formalized in a Learning and Planning procedure based on Local RS-approximation (LP-LRS) and on the use of Interactive Decision Maps (IDM). We will describe the LP-LRS procedure in the next section, while the IDM technique is refered to the book Interactive Decision Maps.

3 LEARNING AND PLANNING PROCEDURE BASED ON LOCAL RS-APPROXIMATION

The Learning and Planning procedure based on Local RS-approximation (LP-LRS) consists of four iterative steps, plus an initialization step and a termination test.

Initialization step The performance $\mathbf{y}(\mathcal{B}, \mathbf{u}_0^1)$ of the current BCA $\mathbf{u}_0^1$ is evaluated via simulation. This alternative is the best available before starting the procedure, e.g. the business-as-usual (BAU) alternative.

The four iterative steps at any iteration of the procedure, say the k-th, are as follows:

Step 1 *Simulation of the model $\mathcal{M}$ and computation of the indicators*: n simulations of the model $\mathcal{M}$ are run using n different decision vectors $\mathbf{u}_1^k, \ldots, \mathbf{u}_n^k$ as input, where n is the dimension of the vector $\mathbf{u}$. More precisely, the n vectors are obtained by perturbing - one by one - each component of the current BCA $\mathbf{u}_0^k$ by an amount Δ_i, that is

$$\mathbf{u}_i^k = \mathbf{u}_0^k + \boldsymbol{\delta}_i, i = 1, \ldots, n \tag{1}$$

where $\boldsymbol{\delta}_i = |\delta_1, \ldots, \delta_n|$, with $\delta_h = 0$ for $h \neq i$ and $\delta_h = \Delta_i$ for $h = i$. By simulating model $\mathcal{M}$ for n times, the outputs $\mathcal{B}_1^k, \ldots, \mathcal{B}_n^k$ are obtained, from which the sample $\mathbf{y}_1 = \mathbf{y}(\mathcal{B}_1^k, \mathbf{u}_1^k), \ldots, \mathbf{y}_n = \mathbf{y}(\mathcal{B}_n^k, \mathbf{u}_n^k)$ is computed. If the set U contains an infinite number of points, the values of Δ_i are selected such that the n vectors $\mathbf{u}_1^k, \ldots, \mathbf{u}_n^k$ belong to a sufficiently small neighbourhood of $\mathbf{u}_0^k$, where the linear approximation can provide a good accuracy for non-linear RS. The definition of such values is not straightforward, and a trial-and-error approach is very likely to be used. If the set U is comprised of a finite number of points, neighboring feasible points can be used.

Step 2 *Identification of the local RS approximation*: the sample obtained in Step 1 and the performance $\mathbf{y}_0 = \mathbf{y}(\mathcal{B}, \mathbf{u}_0^k)$ of the current BCA $\mathbf{u}_0^k$ are used to identify a linear approximation in $\mathbf{u}_0^k$ of the unknown RS $\mathbf{f}(\cdot)$. Each component $f_l(\cdot)$ of $\mathbf{f}(\cdot)$ is of the form

$$\hat{f}_l(\mathbf{u}) = f_l(\mathbf{u}_0^k) + \mathbf{a}^T(\mathbf{u} - \mathbf{u}_0^k)$$

where the vector $\mathbf{a}$ is estimated as

$$\mathbf{a} = \left[\frac{f_l(\mathbf{u}_1^k) - f_l(\mathbf{u}_0^k)}{\Delta_1}, \ldots, \frac{f_l(\mathbf{u}_n^k) - f_l(\mathbf{u}_0^k)}{\Delta_n} \right] \tag{2}$$

Step 3 *Local MO optimization*: using a suitable algorithm, the Pareto frontier of the following MO optimization problem is determined:

$$\min_{\mathbf{u}} \left[\hat{f}_1(\mathbf{u}), \ldots, \hat{f}_q(\mathbf{u}) \right] \tag{3a}$$

subject to

$$\hat{f}_l(\mathbf{u}) \leq \bar{y}_l \qquad \text{with } l = q + 1, \ldots, L \tag{3b}$$
$$\mathbf{u} \in D \cap U \tag{3c}$$

where D is a neighbourhood of $\mathbf{u}_0^k$, within which the local linear approximation $\hat{\mathbf{f}}(\cdot)$ of RS can be assumed to be acceptable. Different methods are available for determining the Pareto frontier. The IDM technique (Lotov et al., 2004) is one of those that is used in this study.

Step 4 *Analysis of the Pareto frontier and choice of the BCA*: the DM analyses the Pareto frontier obtained at Step 3 with a suitable visualization tool and chooses the most preferable point according to her judgment. The associated candidate new BCA $\mathbf{u}^*$ is determined with a suitable algorithm.

Termination test If the performance $\mathbf{y}(\mathcal{B}, \mathbf{u}^*)$, obtained via simulation of $\mathbf{u}^*$, is preferred over the performance $\mathbf{y}(\mathcal{B}, \mathbf{u}_0^k)$ of the current BCA $\mathbf{u}_0^k$, then $\mathbf{u}^*$ becomes the current BCA $\mathbf{u}_0^{k+1}$ for the subsequent iteration $k + 1$. The procedure continues from Step 1. In the opposite case, the procedure terminates, i.e.either the DM is not able to find a point which improves the current BCA or is not satisfied with it.

4   CASE STUDY

Googong Reservoir (Figure 1) is located in New South Wales, Australia. It is one of five sources supplying Canberra's water and is also used for recreational purposes. The reservoir has a history of low to medium levels of Cyanobacteria, namely Anabaena, which have the potential to produce taste and odour compounds and, at high concentrations, to produce neurotoxins. The reservoir

is also affected by a high concentration of Manganese, which causes stains on clothes or household fixtures. Artificial destratification was thought of as a suitable way to address the problem, as is common in many Australian reservoirs. The destratification system installed for Googong in March 2007 consists of two pairs of 5 m diameter downwards-pointing, surface-mounted impellers (the business-as-usual alternative, BAU in Figure 1 (b). The mixers are surrounded by a draft tube reaching a depth of approximately 10 metres and produce approximately 3 m^3/s per mixer. The addition of new mixers and/or the reallocation of the existing ones is evaluated with the purpose of improving the current solution. The search for Pareto-efficient decisions is performed according to the LP-LRS procedure proposed in.

4.1 The decision variables

The input vector **u** must univocally define the number and location of mixers, while the mixer thrust is assumed to be given and fixed in time. For this purpose, the reservoir surface corresponding to water column depth greater than 10 m in the driest observed conditions has been divided into three macroareas (Figure 1 (b)) corresponding to an equal share of subtended water volumes. The positioning of the mixers within each macroarea is univocally defined given the number of mixers assigned to that area in such a way that the volume of water influenced by each mixer is the same. Mixers are installed in pairs, counter-rotating for stability reasons. The maximum allowed number of mixer pairs compatible with both the recreational use of the lake and logistical considerations has been fixed at ten and thus there are ten possible configurations of mixer pairs for each macroarea. In conclusion, three decision variables u_i, with $i = 1, 2, 3$, have been defined to form the decision vector **u**, each one representing the number of mixer pairs in the i-th macroarea, and subjected to the constraint $\sum_{i=1}^{3} u_i \leq 10$.

4.2 The indicators

As explained above, the management concern relates to Anabaena and Manganese concentrations exceeding some reference value. As for the former, we consider as indicator the average annual number of days in which the concentration of chlorophyll-a in the epilimnion exceeds a given threshold, that is

$$y_A = \frac{365.25}{h} \sum_{t=1}^{h} \chi_t, \qquad \chi_t = \begin{cases} 1 & \text{if } A_t > \bar{A} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where A_t is the average concentration of chlorophyll-a in the epilimnion (temporal mean on day t), $\bar{A}$ is the critical threshold (2μg chlorophyll-a/L) and h is the total number of days in the simulation horizon. As for Manganese, any increase in its concentration in the hypolimnion can be due to poor (or null) oxygenation in the hypolimnion. We consider as indicator the average annual number of days in which the concentration of Manganese in the layer just above the sediments exceeds a given threshold, that is

$$y_M = \frac{365.25}{h} \sum_{t=1}^{h} \chi_t, \qquad \chi_t = \begin{cases} 1 & \text{if } Mn_t > \bar{M}n \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where Mn_t is the average concentration of Manganese in the benthic layer (temporal mean on day t) and $\bar{M}n$ is the critical threshold (0.065 mg/L).
Finally, the cost of installation should be taken into proper consideration. Assuming such cost to be linearly proportional to the number of mixers installed, we have the following indicator

$$y_C = \sum_{i=1}^{3} u_i \tag{6}$$

4.3 The process-based model

For the purpose of producing sample data for RS approximation, the coupled ELCOM-CAEDYM (The Estuary, Lake and Coastal Ocean Model - Computational Aquatic Ecosystem Dynamics Model) model was run on a 60 x 60 m grid bathymetry (Figure 1 (a)) with 1 m vertical grid resolution using a simulation step of 2 minutes[1]. The output required by the indicators was sampled with two different time steps, every 12 hours for the Manganese and every 3 hours for Cyanobacteria, which are much more sensitive to daylight variations. The ELCOM-CAEDYM model was calibrated and validated against available field data for temperature, dissolved oxygen, nutrients, metals and chlorophyll-a, with results reported in Hillmer et al. (2008).

4.4 The MO optimization problem

The MO optimization problem for Googong reservoir can be formulated as

$$\min_{\mathbf{u}} [y_A(\mathcal{B}), y_M(\mathcal{B}), y_C(\mathbf{u})] \tag{7a}$$

subject to

$$\mathcal{B} = \mathcal{M}(\mathbf{u}) \tag{7b}$$

$$\mathbf{u} = |u_1, u_2, u_3| \in U \tag{7c}$$

$$U \triangleq \{\mathbf{u} \in \mathbb{R}^3 : u_i \in \{0, 1, \ldots, 10\}, i = 1, 2, 3, \sum_{i=1}^{3} u_i \leq 10\} \tag{7d}$$

where the numerical subscripts denotes the three macroareas and $\mathcal{M}$ is the ELCOM-CAEDYM model described above.

4.5 Application of the LP-LRS procedure

The LP-LRS procedure was applied step-by-step to the above problem with the purpose of reproducing the decision-making process described in Castelletti et al. (2010) and comparing the two approaches in terms of accuracy and computational requirements (i.e. number of simulations of model $\mathcal{M}$ required). Given the experimental nature of this planning exercise, the role of DM was played by an expert from the Centre for Water Research at the University of Western Australia. Initially, the configuration of mixers installed in 2007 (i.e. the BAU alternative) was used as the first BCA. This is defined as $\mathbf{u}_0^1 = |2, 0, 0|$. The associated values of the objectives were computed via simulation of the ELCOM-CAEDYM model. In the step 1, new alternatives were generated in the neighbourhood of the current BCA using expression (1). The associated values of the objectives y_A and y_M were obtained via simulation of ELCOM-CAEDYM, and the value of y_C was directly computed. Then, in the step 2, a linear approximation $\hat{f}(\cdot)$ of the unknown RS $\mathbf{f}(\cdot)$ was obtained. Base on the RS approximation, and the specified neighbourhood set (D) of the current BCA, with assistance of IDM, non-dominated alternatives were found in the step 3. These alternatives assumed to be compared by DM in the step 4. The best one will be the new current BCA for the next iteration if it satisfies conditions of the termination test.

4.6 Discussion

The sequence of points chosen by the DM at each iteration of the LP-LRS procedure and the associated BCAs are summarized in Figures 2 and 3. The final BCA obtained (point 3 in the figures) improves the performance of the current mixer configuration by more than 50% on y_M and nearly 50% on y_A. Operationally, this new configuration can be easily implemented by adding 2 mixers in the third macroarea of the lake and moving the existing ones from the first to the second macroarea (see Figure 1). This is one of the solutions obtained in Castelletti et al. (2010)

---

[1] Approximatively 7 days for a one year long simulation on a AMD Athlon(tm) 64 3.5 MHz.

by applying a global RS approach. In the same work, a detailed physical interpretation of the processes that lead to the improved perfomances is proposed. Here, we are interested in comparing the two approaches from the point of view of the computational requirements and the interaction with the DM.

With the global RS approximation, the number of simulations of the ELCOM-CAEDYM model required to obtain the final BCA was equal to 42. With the local RS approach, this reduces to less than a half (18). In both cases, the time required to find the final BCA (nearly 240 and 90 days respectively on an Intel Xeon 3.16 Ghz QuadCore with 16GB Ram) is incomparably smaller than using an exhaustive 'what-if' analysis that would have required about 5.5 years of computation. Of course these figures have only a relative value, as they might significantly vary with both the complexity of the problem under study (e.g. number and shape of objectives and constraints) and the subjective judgment of the DM. In principle, with highly non-linear RSs, the global approach (based on non-linear RS approximation) should be preferable over the local one. However, the size of the data sample available to estimate the RS approximation can be a limiting factor in producing an accurate approximation and thus, with time-demanding models (e.g. 3D models), the local approach can turn out to be a more accurate and advisable option. A minor drawback of the local approach is that the non-dominated points of the local problem at first iterations could be far from the Pareto-efficient alternatives of the original problem, and, therefore, they might constitute a (biasing) reference for the DM, who should be made aware of this in order to avoid the so-called *anchoring bias* (Tversky and Kahneman, 1974).

5   CONCLUSIONS

In this paper we presented and demonstrated a novel methodology based on local Response Surface (RS) approximation and Interactive Decision Maps (IDMs) for supporting DMs in environmental planning projects involving (i) large, process-based, distributed models, for which simulation is the only feasible way of analysis and a simulation run takes several days (or weeks) even when a modern supercomputer is used, and (ii) many ($q \geq 3$) objectives. Because of such features of the model, only a small number of simulations are possible within a reasonable decision-making timeframe, while the many-objective nature of the problem makes it hard for the DM to explore the Pareto frontier without the support of appropriate visualization tools. The approach was demonstrated on a real world case study to derive approximated Pareto-optimal combinations of mixers number and location to improve water quality in a reservoir. Results indicate that significant improvements could be observed by simply changing the location of the current mixers and installing an additional pair of mixers. The same solution was obtained by Castelletti et al. (2010) using a global RS approach and the same visualization tool, but in more than double the time required by the local approach. The generalization of this figure is not easy, as a number of factors might affect the shape of the original unknown RS and thus the effectiveness of the different approaches available for its approximation, such as the kind and number of objectives considered, the mathematical formulation adopted for the decision variables. Future research will include the evaluation of different formulations of the decision vector (e.g. using spatial coordinates) and their effects on the final decision.

REFERENCES

Box, G.E.P., and K.B. Wilson (1951), On the experimental attainment of optimum conditions (with discussion), *Journal of the Royal Statistical Society Series B*, *13*(1), 1–45.

Castelletti, A., Pianosi, F., Soncini-Sessa, R., Antenucci, J.P., 2010. A multi-objective response surface approach for improved water quality planning in lakes and reservoirs. Water Resources Research, doi:10.1029/2009WR008389, in press.

Hillmer, I., Alexander, R., Yeates, P.S., 2008. Murrumbidgee to Googong raw water transfer project - Hydrodynamic modelling. Tech. Rep. WP2185, Centre for Water Research, University of Western Australia, Crawley, WA.
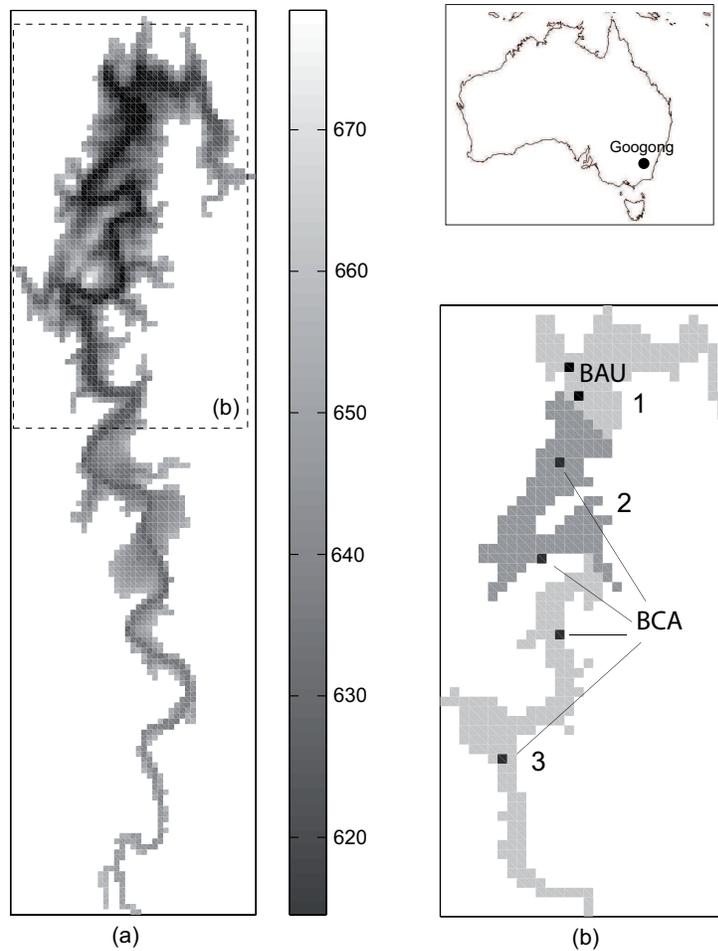
Figure 1: The Googong reservoir bathymetry (a) with elevation shown as metres above sea level. The partitioning of the reservoir in three macroareas (b) with the location of the two pairs of mixers installed in 2007 (the business-as-usual (BAU) alternative) and the improved configuration obtained in this study (the final best compromise alternative (BCA), see Section 4.5).

Lotov, A.V., Bushenkov, V.A., Kamenev, G.K., 2004. Interactive Decision Maps. Vol. 89 of Applied Optimization. Kluwer Academic Publishers, Norwell, MA.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heurisitcs and biases. Science 185, 1124–1130.
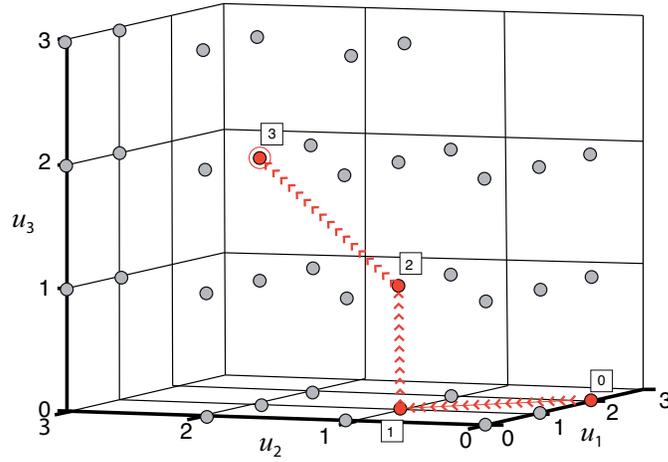
Figure 2: The sequence of BCAs iteratively chosen by the DM in the three iterations performed of the LP-LRS procedure starting from the current mixer configuration (Initialization).
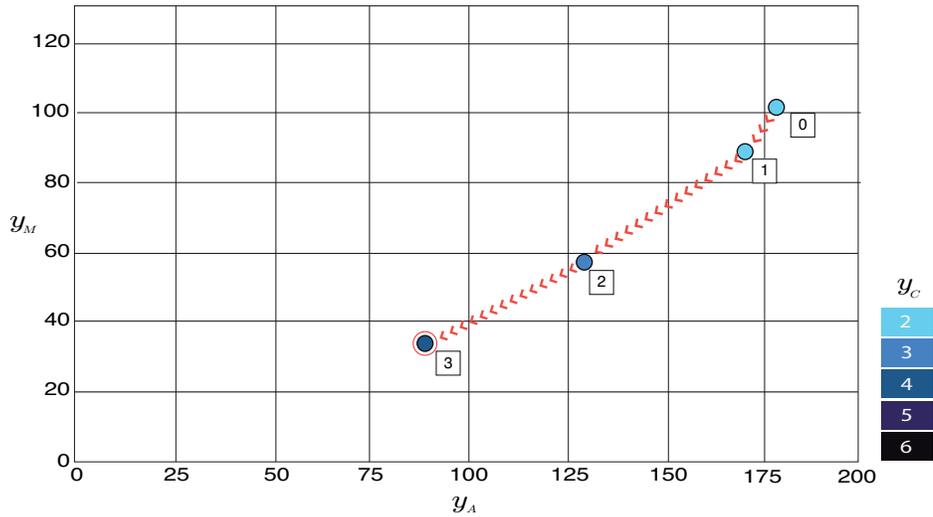

Figure 3: The improvement of the current mixer configuration performed in the three iterations of the LP-LRS procedure.