Comparing Variations of the Greedy Strategy to Find the Optimal Sugar Beet Processing Schedule

Albert I. Egamov* and *Elena A.* Ryabova

Lobachevsky University of Nizhny Novgorod, Nizhny Novgorod, Russia

Abstract. The article presents a mathematical model of sugar beet processing and finding the optimal schedule for this processing. Such task, with fully known data, reduces to a well-known assignment problem. However, in many applied problems of discrete optimization, it is not possible to use classical methods to find a solution in practice. This happens as a result of the fact that when setting the task at the initial moment of time, there is no complete information about some data. For example, in the assignment problem, information about the matrix on the basis of which the objective function is constructed may be incomplete. As a result, in practice it is necessary to search for heuristic algorithms for various possible situations. In particular, the greedy algorithm, as well as its various variations, claims to be such an algorithm. Computational experiment is being conducted that allows us to understand which of the variations of the greedy algorithm give an acceptable result when building an optimal schedule strategy for processing sugar beet batches.

1 Introduction

Finding optimal solutions for the urgent economic problems of the agro-industrial complex is a priority task for many researchers. Among them, the tasks related to sugar production are of particular importance [1–4]. Sugar production is strategic and energy-intensive, so presented optimization problem is highly significant. Often, mathematical models of applied tasks of the agro-industrial complex do not take into account certain auxiliary actions that have to be taken into account in real life. For example, auxiliary measurements of the share of useful ingredient, the cost of moving raw materials to the processing site, temporary unplanned production downtime, etc. Therefore, in order to take into account the real state of affairs, mathematical models have to be complicated, increasingly deviating from the standard basic models of classical applied mathematics. These features play important role in optimization problems. Here and further, one measurement is understood as the recognition of sugar content in one batch of sugar beet.

In continuous optimization, an additional condition for a solution or management almost always forces us to find a new optimal solution, or even a new technique for solving the task. In discrete optimization problems, the same thing happens in most cases. One or more vitally

* Corresponding author: albert.egamov@itmm.unn.ru

understandable and uncomplicated conditions and a standard task is transformed into a new one, the solution of which may be similar to the previous one, however, has its own subtleties and nuances. In [5–9] the problem of finding a heuristic strategy [10] for processing a finite number of batches of sugar beet of different varieties was considered. A greedy strategy, based on greedy algorithm [11, 12], was studied as a strategy claiming to be quasi-optimal with a fairly large range of acceptable parameters of sugar beet batches. When using it, it was necessary to measure the sugar content of all remaining (unprocessed) batches of sugar beet at the beginning of each of the processing stages. However, the correct measurement of sugar content of beetroot in the field is far from trivial and very time-consuming. Despite this, the cost of measuring sugar content and consumables was not taken into account. Simultaneously with the greedy strategy, the so-called strategy A was considered [6, 8], in the application of which only the initial sugar content of the batches was taken into account. That is, the measurements were carried out only once, at the beginning of the processing season. When beet degradation depends only on one parameter, namely the stage number (weather conditions), both strategies show better results than other strategies applicable in practice. The greedy strategy, of course, gave a better result than strategy A, since it "had" more information about the dynamics of degradation of sugar beet batches. In this regard, the following question arises: "How many intermediate measurements of the sugar content of batches should be made so that the relative losses used in a strategy with several intermediate measurements would be small compared to a greedy one?" Such relative losses can be calculated numerically using a virtual computational experiment, using for comparing their average values obtained as a result of a large number (50–100) experiments for various sugar levels generated for acceptable batch parameters. The results and conclusions of such computational experiment are presented in this article.

2 Mathematical model

Assume that at the beginning of the processing season of raw materials, namely sugar beet, the sugar plant has n batches of sugar beet, located for storage in the kagat fields (temporary storage for sugar beet). When storing beetroot, its mass is usually reduced due to the sugar consumed for respiration and moisture evaporation during drying. Suppose that all batches before processing have mass of M identically which coincides with the processing capabilities raw materials of the plant's on one stage.

The row-antitonic matrix P [13] of dimension $n \times n$ consists of positive elements p_{ij}, which determine the sugar content of the i batch before the j stage, $i = \overline{1,n}$, $j = \overline{1,n}$. The property of sugar beet degrade during storage is taken into account, so matrix P is the row-antitonic matrix, that is, for each row, its elements decrease with increasing index j. We define the parameters b_{ij} as the degradation coefficients of the i batch after the jth stage, $b_{ij} = p_{ij+1}/p_{ij}$, $i = \overline{1,n}$, $j = \overline{1,n-1}$. The degradation coefficients are limited to $0 < \beta_1 \leq b_{ij} \leq \beta_2 < 1$, where β_1, β_2 are some constants. The batches are processed according to a certain strategy (schedule), which is determined by the permutation σ of numbers from 1 to n. The task is to find a permutation σ for which the value of objective function S would be the maximum. The function S has the form

$$S = \sum_{j=1}^{n} p_{\sigma(j)j}. \quad (1)$$

In this formulation, the problem is actually reduced to the well–known discrete optimization problem, and namely, to the assignment problem, under the additional condition: the matrix P is the row-antitonic matrix. We will call such permutation a strategy, in which, in turn, some algorithm of the same name is used. A similar formulation of the problem under study has been proposed in various works, for example, [6–9].

In some works, the presented mathematical model became more complicated. For example, we took into account the effect of artificial ripening [14], that is, a temporary increase in the sugar content of raw materials in storage areas after harvest. As a result, according mathematical model for each batch, the artificial ripening process lasted for the first $(\nu - 1)$ th stages [8]. In each row of the P matrix, the elements first increase and then decrease. In work [9] took into account the non-organic materials, which sugar binding substances that remained in sugar beet after soil enrichment [15]. Mathematically, this means that the matrix of states for the assignment problem is represented as the difference between the matrix P and another matrix Q of the dimension $n \times n$. The elements of q_{ij}, the matrix Q correspond to the proportion of sugar associated with non-organics for the i batch before the j stage. In the work of [7], the matrix P was filled in based on data from the Sergach sugar plant located in the Nizhny Novgorod region. In the article [16] mathematical model is considered in which plant shutdowns are possible due to force majeure situations.

There are algorithms for solving the assignment problem in polynomial time for an arbitrary matrix P, if it is fully known when setting the problem. These are, for example, the Hungarian algorithm, the Mack's algorithm, the Little method and others [17–19]. In practice, these methods are not applicable, since initially the matrix P is unknown. At the beginning of the recycling season, only its first column is known. The rest become known during the processing process. Before the jth stage, only the first j columns of the P matrix are known, and only the elements of p_{ij} that correspond to the unprocessed batches.

Thus, in order to obtain a quasi-optimal schedule for processing batches of raw materials, other strategies are needed using algorithms based only on the batch parameters known at the moment. Such strategies will be heuristic. Due to the lack of complete information about the Pmatrix, the strategy used in practice can be quasi-optimal only for some parameter values (see, for example, [20]).

3 Greedy strategy

One of the popular and intuitive strategies for processing batches of raw materials is the greedy strategy, which uses the greedy algorithm. In this case, it consists in the fact that before the start of the j stage, the batch (which has not yet been processed) with the biggest p_{ij} is selected for processing. Under quite much a variety of acceptable conditions for the parameters of the parties, a greedy strategy is not only quasi-optimal, but also optimal [6]. However, there are cases when it shows worse results compared to other well-known heuristic strategies.

Naturally, in order to apply a greedy strategy, it is necessary to take sugar content measurements for all non-processed batches before each stage. This is a rather time-consuming procedure if you do them according to the existing rules. This requires various resources, both human and fuel, as well as time resources. For n stages, $0.5(n^2 + n) - 1$ will have to be made per processing season sugar content measurements. We try to reduce the number of measurements.

Let measurements are performed at the beginning of the stages with the numbers $kl + 1$, $l = \overline{0, [n/k] - 1}$. The k batches with the biggest $p_{i\ kl+1}, i \in I_0$, are processed in the following k stages in descending order $p_{i\ kl+1}$, where I_0 is the set of unprocessed batch numbers. If k is not a divisor of n, then the last measurement (at the beginning of the stage with the number $[n/k] - 1$) is performed for the remaining batches. These batches are processed in the following steps in descending order of sugar content according to the data obtained during the last measurement too. If there is only one batch left for the last measurement is obviously not required. We will call the such strategy as G^k.

It is clear, that the strategy G^{n-1} is equivalent to the strategy G^n. The strategy G^1 is an ordinary greedy strategy, and the strategy G^n is strategy A [6], for which the processing order occurs in descending order of the initial sugar content of the batches. It is known that both the G^1 strategy and the G^n strategy are optimal if the P matrix is such that the degradation coefficients depend only on the stage number and do not depend on the batch number. In real life, this fact may correspond to the fact that the degradation of beetroot during storage is more dependent on the weather and the duration of storage, and practically does not depend on the variety of sugar beet. Also, through a computational experiment, the results were obtained that these strategies are more profitable to use with a large variance of the coefficients b_{ij}, as well as with their uniform distribution over the segment $[\beta_1, \beta_2]$.

4 Computational Experiment

For the computational experiment, two values were taken $n = 15$ and $n = 100$, in this regard, one stage lasts one week and one day, respectively. According to the available information, in one day the above-mentioned Sergach sugar factory can process $M = 3$ thousand tons of sugar beet. We define that one big experiment consists of 50 small experiments. At the beginning of each big experiment, an acceptable segment of initial sugar content is set, as well as β_1 and β_2. Next, for each small experiment, the batches parameters p_{i1} and b_{ij} are generated, $i = \overline{1,n}, j = \overline{1,n-1}$. The parameters b_{ij} are generated either using a uniform distribution or concentrated on specified allowable intervals, the essence of which will be explained later. The parameters a_i are generated a uniform distribution. The matrix P is built using them. The most convenient way to build it is by the formula $p_{ij} = p_{i1}b_{i1\ldots}b_{ij-1}$, $j = \overline{2,n}$. Then a virtual experiment is conducted, strategies of the type G^k are applied to the constructed matrix for various k. The objective function is calculated for each of them. In addition, by means of the so-called maximal strategy, using the Hungarian algorithm, the maximum possible value of the objective function is calculated. Based on the values of the objective functions of the strategy for all small experiments, the average value (arithmetic mean) of the objective function for all strategies G^k and the average maximal value for a big experiment are calculated.

For small experiments we define S_0 is the maximal value of the objective function S, S_{G^k} is the value of the objective function S at strategy G_k. Also we will write average value of the objective function as $\langle\cdot\rangle$. The average relative losses of the strategy S_{G^k}relative to the maximum value are denoted by $\mu^0_{G^k}$ and they have the form

$$\mu^0_{G^k} = \frac{\langle S_0\rangle - \langle S_{G^k}\rangle}{\langle S_0\rangle}. \quad (2)$$

The average relative losses of the strategy S_{G^k} relative to the strategy S_{G^1} (greedy strategy) are denoted by $\mu^1_{G^k}$ and they have the form

$$\mu^1_{G^k} = \frac{\langle S_{G^1}\rangle - \langle S_{G^k}\rangle}{\langle S_{G^1}\rangle}. \quad (3)$$

Comparison of these averages for a given n, other specified parameters, and a selected distributions b_{ij} It is the essence of every big experiment.

The first and second big experiments were conducted at$n = 15$, one stage is equal to one week, respectively. These experiments differ in the distribution of the degradation coefficientsb_{ij} over the given segments. In the first experiment it is uniform, and in the second experiment it is concentrated, that is, for all i the inequality $|b_{ij} - b_{ik}| < \delta$ is

fulfilled, positive $\delta << \beta_2 - \beta_1$ (for each batch, its degradation coefficients are located in a small neighborhood).

The results of the first experiment are presented in Table 1, and the second in Table 2. In the first big experiment, the greedy strategy shows the best result (see Table 1) and the more k the average relative losses increase. In the second big experiment, the greedy strategy gives worse results, which is confirmed by Table 2.

The third, fourth and fifth big experiments were conducted at $n = 100$, one stage is equal to one day. In the third and fourth big experiments the distribution of degradation coefficients b_{ij} uniformly, and in the fifth experiment it is concentrated. The results of these experiments are presented in Table 3–5, respectively. In the third and fourth big experiments, the greedy strategy shows the best result (see Tables 3–4) and the more k the average relative losses increase. The third and fourth big experiments differ from each other by the value of β_1. In the fifth big experiment the conclusions are similar to the conclusions in the second case. For greater clarity, the results of the tables are shown in Figures 1–5.

1 big experiment

Here $n = 15$, $a_i = [0.16,0.2]$, $\beta_1 = 0.86$, $\beta_2 = 0.9999$, 100 little experiments. Coefficients b_{ij} are distributed uniformly over $[\beta_1, \beta_2]$. The results are presented in Table 1.

2 big experiment

Here $n = 15$, $a_i = [0.16,0.2]$, $\beta_1 = 0.86$, $\beta_2 = 0.9999$, 100 little experiments. Coefficients b_{ij} are distributed concentrated over $[\beta_1, \beta_2]$. The results are presented in Table 2.

3 big experiment

Here $n = 100$, $a_i = [0.16,0.2]$, $\beta_1 = 0.98$, $\beta_2 = 0.9999$, 100 little experiments. Coefficients b_{ij} are distributed uniformly over $[\beta_1, \beta_2]$. The results are presented in Table 3.

4 big experiment

Here $n = 100$, $a_i = [0.16,0.2]$, $\beta_1 = 0.95$, $\beta_2 = 0.9999$, 100 little experiments. Coefficients b_{ij} are distributed uniformly over $[\beta_1, \beta_2]$, $\delta = 0.01$. The results are presented in Table 4.

5 big experiment

Here $n = 100$, $a_i = [0.16,0.2]$, $\beta_1 = 0.95$, $\beta_2 = 0.9999$, 100 little experiments. Coefficients b_{ij} are distributed concentrated over $[\beta_1, \beta_2]$, $\delta = 0.01$. The results are presented in Table 5.

At concentrated distribution the unsatisfactory results of a greedy strategy and the impressively good results of the thrifty strategy (the opposite of greedy one in effect) were discussed in [21].

5 Discussion

As can be seen from the tables and figures, a greedy strategy and its variations should be applied if the degradation coefficients are evenly distributed over an admissible interval. If the degradation coefficients are distributed in a concentrated manner, then the greedy strategy and its variations should not be used. Further recommendations are given only for the uniform distribution of degradation coefficients. Let $n = 15$, for practical implementation, it is appropriate to use the strategies G^7 or G^8 instead of the greedy strategy G^1. In this case, the average relative losses compared to the greedy strategy are approximately 1%. For a greedy strategy, you need to spend 15+14+...+2=119 measurements. For strategy G^8, 15+7=22 measurements should be performed and for strategy G^7, 15+8=23 measurements should be performed. Let $n = 100$, for practical implementation, it is appropriate to use the strategy G^{20} instead of the greedy strategy G^1. In this case, the average relative losses compared to the greedy strategy are 1%. For a greedy strategy, you need to spend 100+99+...+2=5049

measurements. For strategy G^{20} 100+80+60+40+20=300 measurements should be performed.

It must be recognized that the permissible the lower limit is of degradation of sugar beet per day is greater than the main value in the experiments $\beta_1 = 0.98$, (for a week at the same time it is equal to $0.98^7 \approx 0.868$).

Table 1. 1 big experiment. Comparison of strategies.

Strategy	Value of objective function, S_0	Average losses relative to Max. strategy, $\mu^0_{G^k}$, %	Average losses relative to greedy strategy, $\mu^1_{G^k}$, %
Max.	1.852	0.00	–
G^1	1.763	4.81	0.00
G^2	1.758	5.08	0.28
G^3	1.755	5.25	0.47
G^4	1.753	5.37	0.60
G^5	1.752	5.42	0.64
G^6	1.748	5.63	0.87
G^7	1.747	5.71	0.95
G^8	1.745	5.80	1.05
G^9	1.742	5.95	1.20
G^{10}	1.741	6.00	1.26
G^{11}	1.740	6.07	1.33
G^{12}	1.738	6.15	1.42
G^{13}	1.738	6.19	1.46
G^{14}	1.738	6.18	1.45
G^{15}	1.738	6.18	1.45

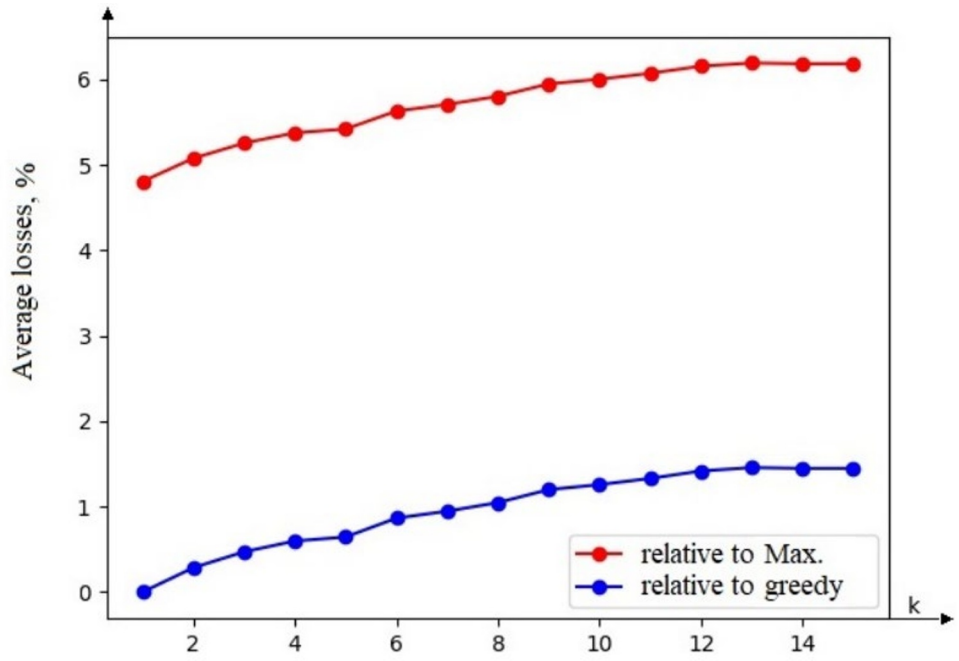

Fig. 1. 1 big experiment. The average relative losses, $\boldsymbol{n = 15}$

Table 2. 2 big experiment. Comparison of strategies.

Strategy	Value of objective function, S_0	Average losses relative to Max. strategy, $\mu^0_{G^k}$, %	Average losses relative to greedy strategy, $\mu^1_{G^k}$, %
Max.	1.982	0.00	–
G^1	1.695	14.98	- 0.00
G^2	1.710	13.74	-1.46
G^3	1.734	12.52	-2.89
G^4	1.754	11.51	-4.08
G^5	1.771	10.64	-5.10
G^6	1.786	9.00	-5.97
G^7	1.795	9.42	-6.53
G^8	1.803	9.06	-6.96
G^9	1.808	8.79	-7.28
G^{10}	1.812	8.59	-7.51
G^{11}	1.814	8.47	-7.65
G^{12}	1.816	8.38	-7.75
G^{13}	1.817	8.34	-7.80
G^{14}	1.817	8.33	-7.82
G^{15}	1.817	8.33	-7.82

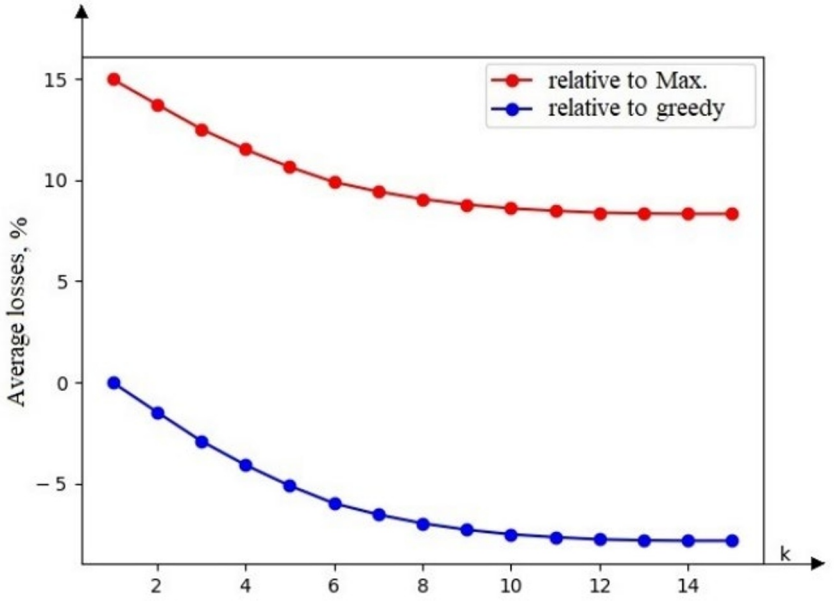

Fig. 2. 2 big experiment. The average relative losses, $\boldsymbol{n = 15}$.

Table 3. 3 big experiment. Comparison of strategies.

Strategy	Value of objective function, S_0	Average losses relative to Max. strategy, $\mu^0_{G^k}$, %	Average losses relative to greedy strategy, $\mu^1_{G^k}$, %
Max.	11.9192	0.00	–
G^1	11.6604	2.17	0.000
G^2	11.6600	2.17	<0.001

G^5	11.6570	2.17	0.004
G^{10}	11.6560	2.20	0.030
G^{20}	11.6533	2.21	0.038
G^{50}	11.6415	2.33	0.162
G^{100}	11.6308	2.42	0.254

Table 4. 4 big experiment. Comparison of strategies.

Strategy	Value of objective function, S_0	Average losses relative to Max. strategy, $\mu^0_{G^k}$, %	Average losses relative to greedy strategy, $\mu^1_{G^k}$, %
Max.	7.3444	0.00	–
G^1	7.0740	3.68	0.000
G^2	7.0701	3.74	0.055
G^5	7.0564	3.92	0.248
G^{10}	7.0385	4.17	0.501
G^{20}	7.0050	4.62	0.975
G^{50}	6.9454	5.43	1.816
G^{100}	6.9108	5.90	2.307

Table 5. 5 big experiment. Comparison of strategies.

Strategy	Value of objective function, S_0	Average losses relative to Max. strategy, $\mu^0_{G^k}$, %	Average losses relative to greedy strategy, $\mu^1_{G^k}$, %
Max.	8.591	0.00	–
G^1	7.238	15.74	0.00
G^2	7.286	15.19	-0.66
G^5	7.423	13.59	-2.55
G^{10}	7.599	11.54	-4.99
G^{20}	7.812	9.06	-7.93
G^{50}	7.934	7.64	-9.61
G^{100}	7.839	8.46	-8.30

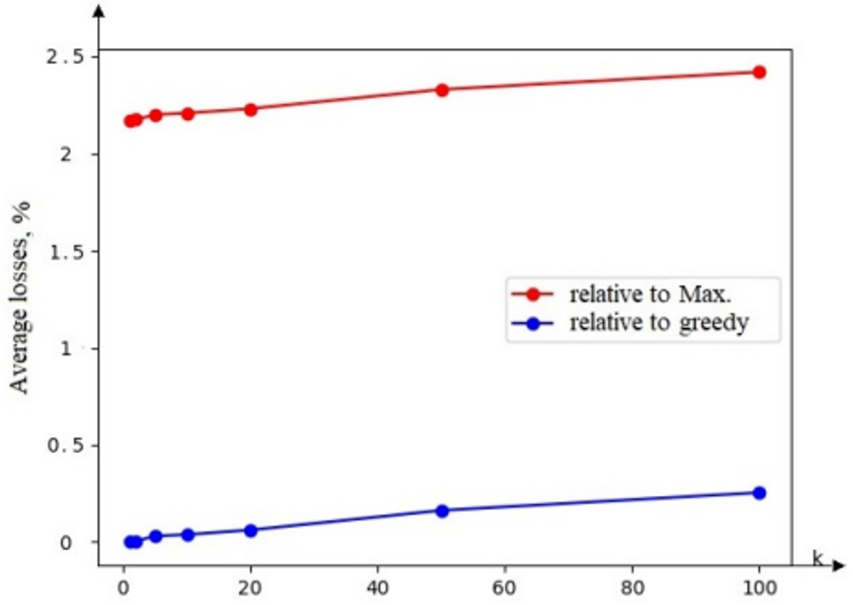

Fig. 3. 3 big experiment. The average relative losses, $\boldsymbol{n = 100}$.

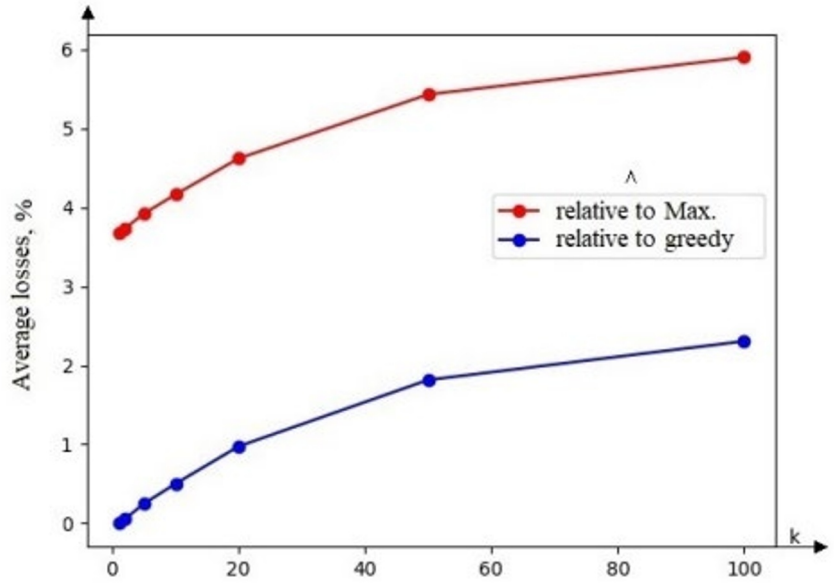

Fig. 4. 4 big experiment. The average relative losses, $\boldsymbol{n = 100}$.

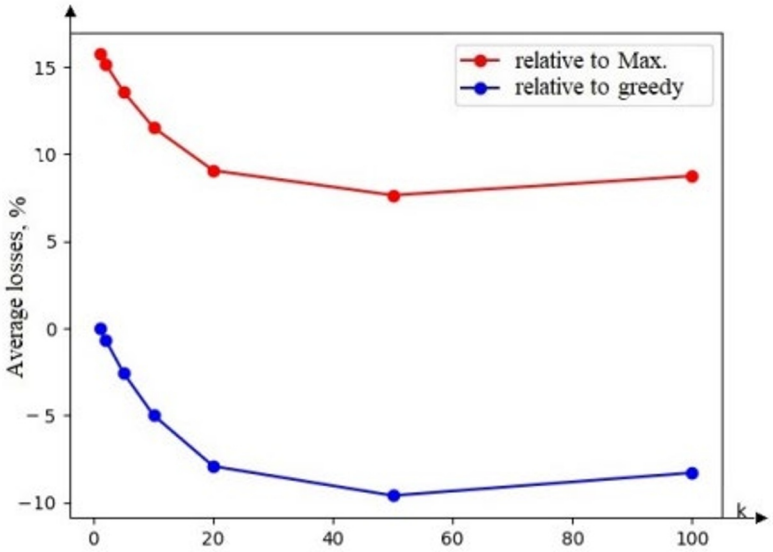

Fig. 5. 5 big experiment. The average relative losses, $\boldsymbol{n = 100}$.

However, sugar beet diseases, the presence of inorganic substances, possible low temperatures at the end of the processing season and other negative aspects in practice often make their own adjustments downward the lower limit.

6 Conclusion

This article provides some recommendations for applying the greedy strategy and its variations. The interest in the greedy strategy is explained by the fact that in many cases it is quasi-optimal. It is shown that for practical purposes it can be replaced by some variation with a much smaller number of sugar content measurements in sugar beet. Variations of the greedy strategy are proposed, using which the resulting sugar mass is approximately 99 percent of the sugar mass, that can be obtained using the greedy strategy. In turn, the sugar mass of this strategy is about 95 percent of the maximum possible sugar mass. In practice, such deviations supposed to be admissible.

References

1. O. Pirogova, R. Nuzhdin, B. Pivovar, E3S Web of Conferences, **244,** 10056 (2021)
2. D. Rudoy et al., *E3S Web of Conferences* **381** 01082 2023)
3. Z. Jiao, A.J. Higgins, D.B. Prestwidge, Computers and Electronics in Agriculture, **48,** 170-181 (2005)
4. R. Junqueira, R. Morabito, International Journal of Production Economics, **231(1),** 150-160 (2019)
5. D.V. Balandin, O.A. Kuzenkov, V.K. Vildanov, Modern information technologies and IT-education, **17 2,** 442-452 (2021) https://doi.org/10.25559/SITITO.17.202102 [in Russian]
6. D.V. Balandin et al., *Mathematical Modelling and Optimization of Scheduling for Processing Beet in Sugar Production* In book Balandin D., Barkalov K., Meyerov I.

(eds) Communications in Computer and Information Science **1750** 227–238 (2022) https://doi.org/10.1007/978-3-031-24145-1_19

7. D.V. Balandin, O.A. Kuzenkov, A.I. Egamov, IOP Conf. Ser.: Earth Environ. Sci., **1206,** 012046 (2023)
8. D.V. Balandin et al., Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies, **2,** 62–76 (2023) [in Russian]
9. D.V. Balandin, O.A. Kuzenkov, A.I. Egamov, *Development of components for monitoring and control intelligent information system* In book Voevodin V., Sobolev S., Yakobovsky M., Shagaliev R. (eds). Supercomputing. 9th Russian Supercomputing Days, RuSCDays. 2023. Moscow, Russia, September 25–26, 2023. Revised Selected Papers, Part II. Lecture Notes in Computer Science, book series, LNCS, 14389, 162–177 (2023).
10. V.B. Popov, Chapter 7. Metaheuristic algorithms for problems of economic optimization and forecasting *Information economics: development, management, models*: A collective monograph. (Under the scientific editorship of N.V. Apatova. Simferopol, 2017) 401–416. [in Russian]
11. T. Roughgarden, Algorithms Illuminated Part 3: Greedy Algorithms and Dynamic Programming *Sound-likeyourself Publishing* (LLC. New York, NY, 2019) 232.
12. T.H. Cormen et al., *Introduction to Algorithms. Third Edition* (The MIT Press Cambridge, Massachusetts. London, England, 2009) 1313.
13. A.A. Koroleva, Journal of the Belarusian State University. Economy, **1,** 26–36 (2021) [in Russian]
14. R.B. Harvey, Bulletin (University of Minnesota. Agricultural Experiment Station), **247**, 36 (1928)
15. V.N. Kukhar et al., Sugar, **1**, 18-36 (2019) [in Russian]
16. A.I. Egamov, E3S Web of Conferences, **395,** 03007 (2023) https://doi.org/10.1051/e3sconf/202339503007
17. B. Bunday, Basic linear programming (London, 1984) 163.
18. B Rainer, M Dell'Amico, S Martello, Assignment problems. Society for Industrial and Applied Mathematics (USA. Philadelphia, 2009) 382.
19. N.A. Shchukina, Problems of Economics and management, **5(57),** 169–174 (2016) [in Russian]
20. V.P. Khranilov, P.N. Burago, A.I. Egamov, E3S Web of Conferences, **537,** 09009 (2024) https://doi.org/10.1051/e3sconf/202453709009
21. D.V. Balandin, A.I. Egamov, O.A. Kuzenkov, Comparison of Heuristic Strategies for Sugar Beet Processing Schedules *Applied Mathematics, Computational Science and Mechanics: Current Problems* (Voronezh, Russian Federation, AMCSM, 2023) 1–6. https://doi.org/10.1109/AMCSM59829.2023.10525768