Pareto optimization via genetic algorithm for selecting a portfolio of oil and gas exploration projects.

Wellington C.R. Nascimento, Petróleo Brasileiro S.A.

Abstract

The oil and gas exploration and production industry deals with various uncertainties and risks, whether inherent in exploration projects or characterized by the high volatilities of commodity prices. Due to the high costs of exploration and production development, combined with the time required for oil fields to begin production, it is essential to use techniques that minimize uncertainties and risks throughout the production chain in order to maximize the value of a portfolio of projects, each project has its expected economic return for a given investment and a certain probability of success. Therefore, the decision-maker must select the combination of projects that maximize the portfolio value based on their limited investment capital. Thus, a careful portfolio analysis will allow for the identification of a combination of projects that result in the most efficient use of capital. In this work, we present an application of Genetic Algorithms to the problem of selecting projects that maximize the Net Present Value of the portfolio, while minimizing investments, maximizing the probability of success, and meeting other constraints deemed important by the decision-maker. For this, we will use multi-objective optimization, also known as Pareto Optimization, which presents a set of optimal solutions given a space of options with apparently conflicting objectives, such as high-return investments versus low-risk investments. Pareto Optimization allows for a comprehensive view of available options and for decision-making based on multiple criteria.

Introduction

Optimizing a portfolio of projects is a complex and challenging problem that involves selecting a set of projects to maximize return on investment, taking into consideration certain constraints such as budget, resources and risks. Geoscientists and engineers evaluate prospects in the subsurface and determine, with some degree of uncertainty, the location/depth, volume of oil/gas, the required investment to develop a reservoir and the probable returns calculated according to the price of a barrel of oil and the estimated production curve (Túpac et al. 2002). In addition to the probability of success or exploratory chance factor, recovery factors and the commercial viability of the exploration opportunity. Therefore, a company with multiple projects in its investment portfolio has the difficult task of selecting those that minimize risks and maximize profits, in addition to other strategic criteria and investment commitments with regulatory agencies. It is up to managers or even company directors to make decisions regarding the process of selecting exploratory projects that add value to the company. Traditional criteria use discounted cash flow to select projects with both the highest positive Net Present Value (NPV) and Expected Monetary Value (EMV). This traditional criteria is, however, limited in real-world situations (Lopes et al. 2013), because these parameters (NPV and EMV) miss to account for market strategies that the decision-maker must analyze in the project selection stage. Let us take, for example, the maximization of the volume of oil in place (VOIP) aiming to incorporate strategic reserves or even the restriction of available capital for investment. It is also worth noting that depending on the number of projects, the number of possible permutations becomes a very difficult task, and the traditional method of ranking projects by NPV, rate of return, or profitability index falls short in producing optimal results (Sarich, 2001).

In this work, we use Genetic Algorithms (GA) as a search and optimization method. As the problem is classified as a multi-objective optimization problem, as we will see later, the search space becomes very large and difficult to model. Therefore, we will use Pareto Optimization, which presents an optimized set of solutions in the space of feasible solutions, and within this set, it is possible to obtain an optimal solution that satisfies the problem constraints, having a decision model available.

Pareto Optimization is obtained through the NSGA-II algorithm, which is very popular in the literature for solving multi-objective optimization problems. NSGA-II obtains a set of optimal solutions for the problem known as non-dominated solutions. The image of this set in the space of objective functions forms the Pareto front.

Method

Genetic Algorithms

Genetic Algorithms are an optimization technique based on Charles Darwin's theory of evolution by natural selection. John Henry Holland was the pioneer in the introduction of genetic algorithms as an optimization technique inspired by natural evolution in his book *Adaptation in Natural and Artificial Systems*, published in 1975.

The basic concept of the natural selection process is related to the favoring of the hereditary transmission of beneficial characteristics to future generations of a population of reproducing organisms. Conversely, unfavorable characteristics become less common among descendants over time. Similarly, genetic algorithms operate to find the best solutions to a problem. The process begins with a random set of candidate solutions, evaluating their fitness in relation to the problem's objective function, which is the optimization target. In the case of a portfolio of exploratory projects, the value of each project, cost, risk and other constraints are considered. The candidate solutions are part of a generation. This generation is subjected to genetic crossover and mutation operators to produce a new generation of solutions. The new generation is used as input for subsequent algorithm iterations. The most favorable solutions are selected for reproduction and the next offspring is generated from them. The process is repeated for several generations until an acceptable solution is found.

The main steps of the algorithm are:

- **Fitness**: an analysis of solutions (population individuals) is performed to find candidate solutions based on the objective function.
- **Reproduction**: individuals are selected and copied to the next population according to their fitness.
- **Crossover**: recombination of the selected solutions, generating new individuals.
- **Mutation**: Random Exchange of characteristics in individuals, adding diversity to the population.
- **Update**: the new individuals are inserted into the current generation population.
- **Finalization**: analysis of the termination conditions of the evolution (acceptable solution found).

Figure 1 illustrates the execution steps of the Genetic Algorithm.

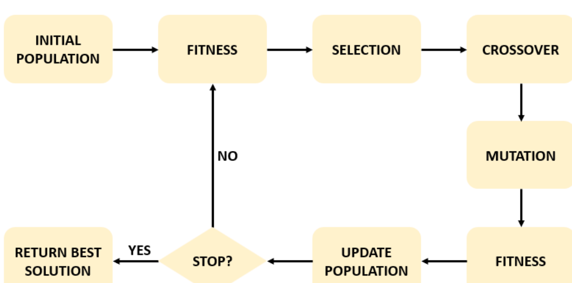

Figure 1: GA flowchart.

Genetic Algorithms can be customized to address a wide range of problems, such as optimization problems with complex constraints and multi-objective decision problems, which is the case analyzed in this work.

Pareto Optimization

Pareto Optimization, also known as multi-objective optimization, is common in many problems in engineering, economics and logistics and basically consists of obtaining a set of solutions that satisfy certain constraints and optimize a function composed of several objectives. These complex problems become interesting when their objectives are conflicting, meaning that decreasing the value of one objective necessarily increases the value of another. In these scenarios of conflicting objectives, a solution that maximizes all objectives is infeasible. However, it is possible to find a set of solutions that represents the best trade-off between objectives.

Conflicting objectives are common in oil exploration projects, where we constantly need to maximize economic return while minimizing exploration costs, minimizing geological uncertainties while also maximizing the discovery of new reserves. Therefore, the optimal solution of one objective function does not coincide with the optimal solutions of the other objective functions. The solution to this impasse is not unique, but rather a family of solutions known as Pareto-Optimal solutions. Pareto-Optimal solutions are optimal in a broad sense, meaning that no other solution in the search space will be superior to them when all objectives are considered simultaneously. This concept was introduced by the Italian economist Vilfredo Pareto in the late 19th century.

More formally, the multi-objective optimization problem can be defined as follows:

$$\begin{gathered}\textit{Given a vector of decision variables with dimension } n,\\ x=\{x_1,\ldots,x_n\} \textit{ in the search space } X, \textit{we want}\\ \textit{to find a vector } x^* \in X \textit{ that simultaneously}\\ \textit{minimizes (maximizes) the } r \textit{ objetive functions.}\end{gathered}$$

The essence of Pareto Optimization is to find the set of solutions P* that contains a family of Pareto-Optimal solutions. Solutions are compared through the Pareto dominance property.

Definition (Dominance): A point $x_1 \epsilon X$ is said to dominate $x_2 \epsilon X$ if $f(x_1) \leq f(x_2)$ and $f(x_1) \neq f(x_2)$.

Definition (Pareto-Optimal Solution): We say that $x^* \epsilon P^*$ is a Pareto-Optimal solution of the multi-objective problem if there is no other solution $x \epsilon P^*$ such that $f(x) \leq f(x^*)$, i.e., if x^* is not dominated by any other feasible point.

A Pareto-Optimal solution cannot be improved with respect to any objective function without worsening at least one other objective function.

In the case of two objectives, Figure 2 (Arroyo, 2002) illustrates the concept of dominance where points A and B dominate C, points E and F are dominated by C, and points D and G are indifferent to C.

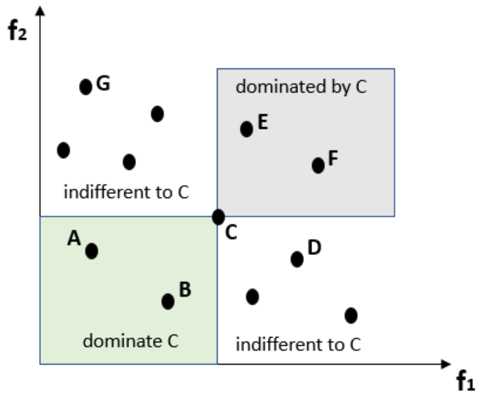

Figure 2: Pareto dominance.

Figure 3 shows two examples of solutions families for two objective functions in the case of minimization (left) and maximization (right) with their respective Pareto-Optimal solutions (Pareto front).

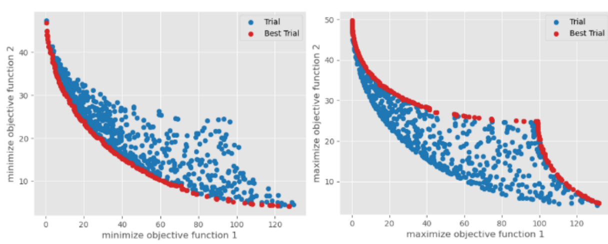

Figure 3: Pareto-Optimal Solutions (red).

The possible optimal solutions for the case of two objectives (bi-objective) functions are illustrated in Figure 4 (Bektas, 2020).

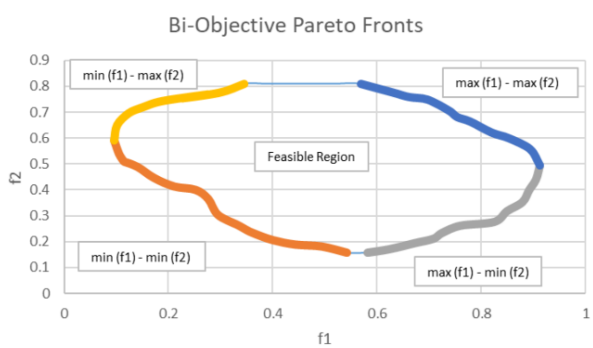

Figure 4: Possible Pareto fronts.

NSGA-II

The NSGA-II (*Non-dominated Sorting Genetic Algorithm*) algorithm is one of the pillars of multi-objective optimization based on genetic algorithms. It was developed as a response to the deficiencies of early evolutionary algorithms. The basic idea is to allow a population of candidate solutions to evolve towards the best solution for solving a multi-objective optimization problem. The NSGA-II was designed to search for the optimal solution in an exhaustive list of candidate solutions, resulting in a large search space.

Figure 5 illustrates the execution steps of the NSGA-II algorithm.

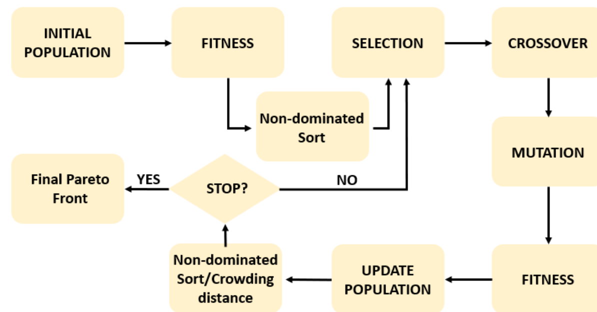

Figure 5: NSGA-II flowchart.

The main objective of the NSGA-II algorithm is to calculate the Pareto-Optimal solution, which corresponds to a set of optimal solutions, called non-dominated solutions. A non-dominated solution provides a suitable combination of all objective functions without degrading any of them. For more details, interested readers can refer to Deb et al. (2002).

Results

Due to issues of confidentiality, to apply multi-objective optimization, we used the data presented in the paper of Lopes et al. (2013), listed here in Table 1 located at the end of the paper. As mentioned by the authors, the data are realistic enough in the sense of a particular decision-making context and for the structure of the relationship between the variables and parameters considered.

The list of potential projects presented in Table 1 has the following attributes:

- **NPV**: Net Present Value in case of success in the exploration phase.
- **PoS**: Probability of Success.
- **RES**: Estimate of the size of the hydrocarbon volume (reserves).
- **SYN**: Synergy. Relates to the project's influence on others projects.
- **DHC**: Dry hole cost. The risk capital of the project.
- **EXT**: Qualitative criterion related to the influences of external factors.

The decision Variable in our problem is the selection or not of the exploratory project, which, according to Table 1, are named P1, P2, …, P30. These projects are converted in terms of genetic data by binary coding. The genetic algorithm starts with the creation of the initial population of chromosomes. The chromosome length consists of thirty genes (proposed projects). If the project is selected, the encoding will be 1, otherwise 0. Gene selection occurs randomly. Figure 5 gives us an idea of the concept of gene, chromosome and population.

Figure 5: Gene, Chromosome and Population.

Thus, the gene represents a project, the chromosome (individual) represents a solution (portfolio) and the population represents a family of solutions. A population consists of a certain number of individuals, each representing a solution to the problem. In our case, we consider an initial population of 1000 individuals as shown in Figure 6.

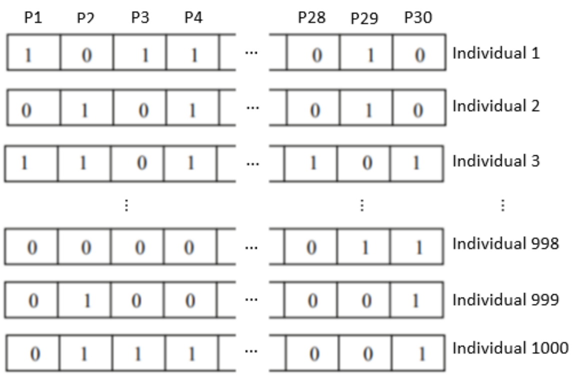

Figure 6: Initial population.

A simple scenario to test the algorithm is to consider the maximization of NPV and RES simultaneously. In this case, there is only one solution (red dot), which is the selection of all projects. This is because, in this particular case, there are no constraints. The result is presented in Figure 7 with NPV = 17066 US$MM and reserves RES = 17996 MMBOE.

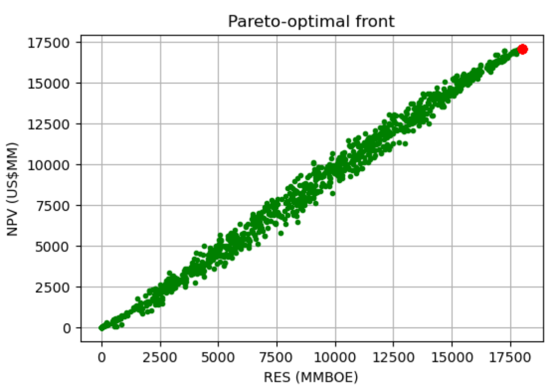

Figure 7: Maximization of NPV and RES.

Figure 8 presents the progress of the NSGA-II algorithm in generating the Pareto front. As the number of generations increases, solution families converge towards a single family which is the set of solution P* containing a family of Pareto-Optimal solutions.

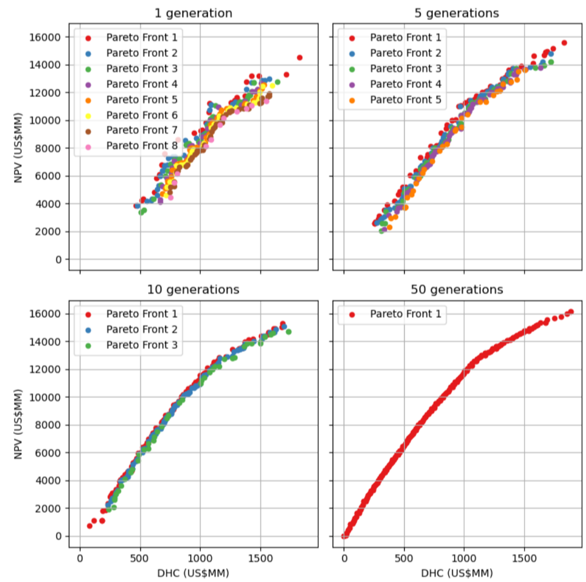

Figure 8: Pareto Fronts.

A first experiment was conducted to maximize NPV and minimize DHC. The result of simulation for 100 generations is shown in Figure 9. We can see all solution families (green dots) found for the problem's imposed constraints. In addition to the Pareto-Optimal solutions in red dots.

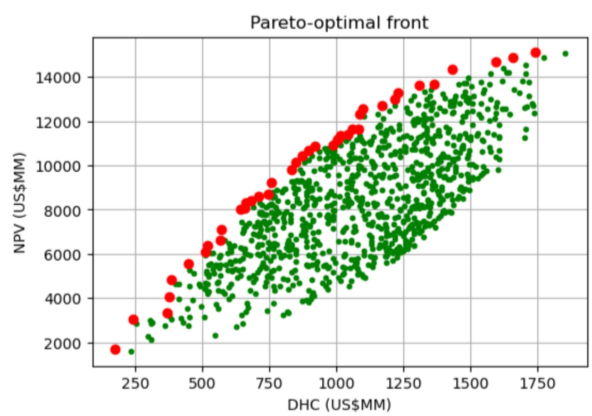

Figure 9: Pareto-Optimal solutions.

The evaluation of everyone in the population was done by calculating the NPV and DHC. The next step is the selection of chromosomes that will be transmitted to the next generation. This was done using the ranking method that classifies individuals according to their fitness.

To perform the crossover, we used the one-point recombination method with a 50% probability, as shown in Figure 10.

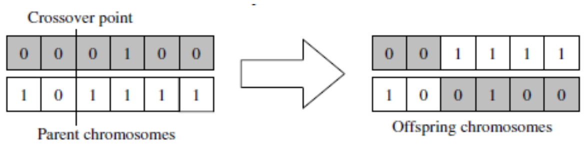

Figure 10: One point crossover.

In the case of mutation, which consists of a random change in one or more genes of the chromosome, we use a rate of 1%. The mutation is performed as shown in Figure 11.

Figure 11: Before and after the mutation.

With the solutions presented in Figure 9, the decision-maker enters the circuit with a strategic model or limitations regarding the capital to be invested. Let's say the decision-maker has a capital of 1000 US$MM at their disposal. In this case, we would need to present the solution on the Pareto frontier that maximizes the NPV of projects with this capital limitation. To do so, simply select the solutions that are limited to a DHC of 1000 US$MM, as shown in Figure 12.

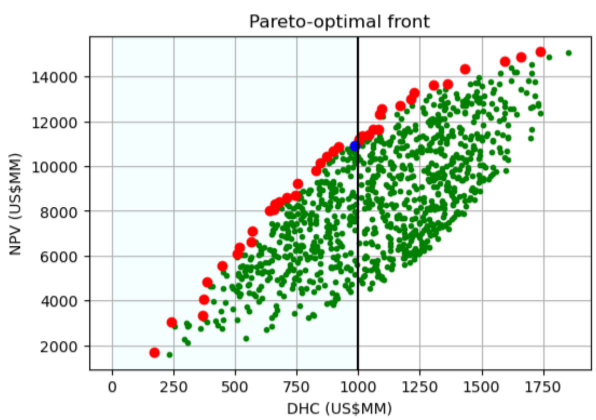

Figure 12: Pareto-Optimal solution (blue dot).

The Pareto-Optimal solution that maximizes NPV with the imposed capital limitation is NPV = 10911 US$MM and total DHC = 987 US$MM, represented in Figure 12 by the blue dot.

The portfolio that satisfies the problem constraints is: {P1, P3, P4, P5, P7, P8, P14, P15, P16, P18, P19, P29, P30}. A total of thirteen projects.

Another scenario is the maximization of reserves with the qualitative criterion EXT, which is related to the influence of external factors in the project's management, such as political situation and local infrastructure (Lopes et al. 2013). This attribute varies from 1 to 5 on a Likert scale, where 1 means very negative influence and 5 when they are very positive. The Pareto-Optimal solutions of the simulation are shown in Figure 13. Faced with several optimal solutions available, the decision-maker selects the one that fits the adopted strategy.

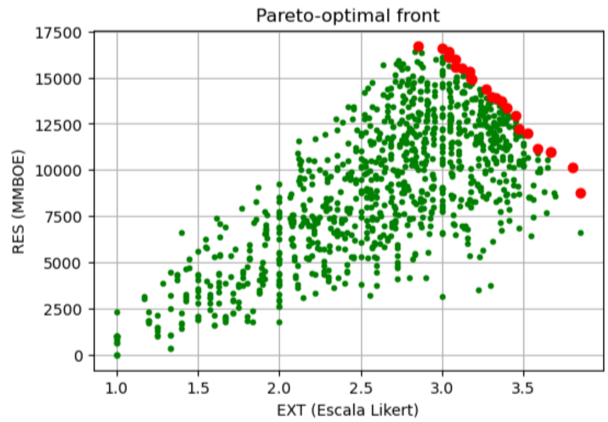

Figure 13: Pareto-Optimal solutions (RES x EXT).

The scenario that considers the probability of success (PoS) of the project is the one that maximizes the EMV and minimizes the DHC. To obtain the VME we use the following formula:

$$EMV = PoS * NPV - DHC$$

The simulation result is shown in Figure 14 where the blue dot represents the Pareto-Optimal solution that maximizes the EMV with the limitation of capital available for investment.

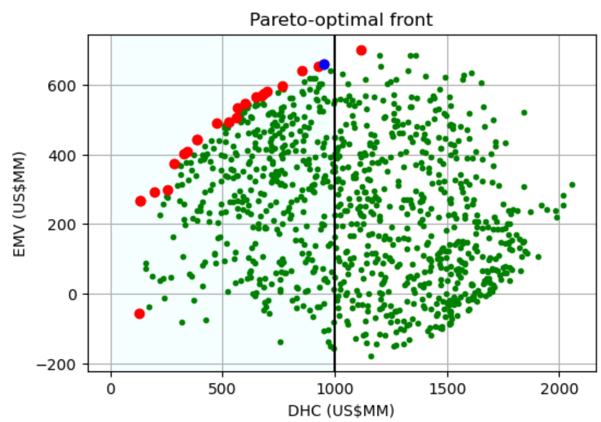

Figure 14: Pareto-Optimal solution (blue dot).

Conclusions

This work presents an application of Genetic Algorithms and the NSGA-II algorithm to obtain the Pareto front containing solutions for optimized selection of projects in an exploratory portfolio. Depending on the number of available projects, it becomes practically impossible to consider all possible alternatives for economic maximization of the portfolio. Thus, for decision-makers with capital constraints and a business rule, Pareto optimization proves to be an excellent tool for those who want to maximize the value of their portfolio of oil and gas exploration projects.

Acknowledgments

The author would like to thank Antonio Affonso Filho, Elizeu Santos-Neto and Sandinei Ugo for reading, critical review and valuable suggestions and also Petróleo Brasileiro S.A. represented by manager Suzanna Carreiro Calache for supporting the research.

References

Al-Mudhafer, W.J.; Shahed, M.A. 2011. Adopting Simple & Advanced Genetic Algorithms as Optimization Tools for Increasing Oil Recovery & NPV in an Iraq Oil Field. *SPE Middle East Oil and Gas Show and Conference*. SPE 140538.

Arroyo, J.E.C. 2002. Heurísticas e Metaheurísticas para Otimização Combinatória Multiobjetivo. PhD Thesis. Unicamp.

Bektas, H.O. 2020. Considerations on Multiobjective Optimization Regarding Municipal Water Supply Systems. PhD Thesis. Middle East Technical University.

Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, Vol.6, No.2.

LaCosta, W.C.P.; Milkov, A.V. 2022. Petroleum exploration portfolios generated with different optimization approaches: Lessons for decision-makers. *Journal of Petroleum Science and Engineering*, 214, 110459. https://doi. org/10.1016/j.petrol.2022.110459.

Lopes Y.G.; Almeida, A.T. 2013. A multicriteria decision model for selecting a portfolio of oil and gas exploration projects. *Pesquisa Operacional*, 33(3), 417-441.

Sarich, M.D. 2001. Using Genetic Algorithms to Improve Investment Decision Making. *SPE Asia Pacific Oil and Gas Conference and Exhibition*. SPE 68725.

Túpac, Y.J.; Vellasco, M.M.B.R.; Pacheco, M.A.C. 2002. Selection of Alternatives for oil field development by genetic algorithms. *Engenharia Térmica,* n°2, 51-54.

Table 1: Project list.

	NPV(US$MM)	PoS(%)	RES(MMBOE)	SYN	DHC(US$MM)	EXT
P1	1086	10.8	1311	5	85	2
P2	670	31.8	582	1	145	5
P3	2131	9.0	1710	4	180	2
P4	991	8.1	799	2	95	3
P5	1172	11.3	750	2	120	4
P6	385	31.4	512	4	80	2
P7	1164	9.8	850	5	120	4
P8	1639	12.3	1355	4	110	1
P9	451	26.8	678	2	150	2
P10	829	31.6	700	4	55	1
P11	752	14.0	708	5	60	5
P12	457	29.8	510	4	90	5
P13	463	29.0	480	1	90	4
P14	709	7.4	800	1	60	1
P15	557	9.4	850	5	50	1
P16	430	9.8	651	2	35	1
P17	383	12.6	575	3	35	3
P18	374	17.3	550	5	40	4
P19	320	17.4	423	1	40	2
P20	338	22.4	500	3	100	3
P21	455	30.3	450	3	105	3
P22	337	12.2	492	5	80	4
P23	56	17.7	101	1	41	3
P24	28	24.8	580	3	25	2
P25	155	5.6	304	4	30	5
P26	95	15.8	180	1	41	2
P27	266	28.2	204	5	40	3
P28	35	25.2	50	2	18	4
P29	185	11.1	176	5	22	1
P30	153	8.4	165	4	30	3