

GENETIC ALGORITHMS AND REAL OPTIONS ON THE WILDCAT DRILLING OPTIMAL CHOICE

LUIGI DE MAGALHÃES DETOMI CALVETTE

MARCO AURÉLIO C. PACHECO

Abstract This paper discusses the problem of optimizing the drilling sequence of exploratory prospects considering its embedded real options and the potential learning effects. The portfolio value is the objective function to be maximized and it is modeled considering the rule-of-thumb that *the total is greater than the sum of its parts*. The main objective is to propose a general model to optimize portfolios with up to ten prospects using Genetic Algorithms.

Keywords Optimization; Genetic Algorithms; Learning Options; Chance Factor

1 Introduction An oil field exploration and development campaign is bounded with different kinds of uncertainty. The most basic one that an E&P portfolio manager deals with is the one related to the existence (or not) of oil in a given prospect. Typically, technical uncertainties are related to learning and tend to be reduced with investments to acquire additional information. From the correlation pattern on the prospects in a given exploratory portfolio, follows that the results from one initial wildcat drilling will, potentially, reveal, additional information about the oil existence (or not) in other prospects in the same geological play. This way, each prospect to be drilled might be understood as an learning option to be exercised (or not) depending on its respective success probability. In such case, one of the main factors on optimizing the exploratory campaign is choosing the ideal drilling sequence. Such choice is more complex, as the quantity and diversity of the prospects increases. Given such background, the present paper proposes a model that intends, using Genetic Algorithms, to optimize the drilling sequence and, as a consequence, the total portfolio value. The proposed model considers the interdependencies, each prospect's specific aspects and has as an objective function (to be maximizes) the portfolio net present value (NPV). Options and learning effects are the main aspects underlying the optimization model. The model was evaluated on

four different exploratory portfolios and, in every case, was able to deliver at least one sequence that could represent expressive NPV gains compared to the basic scenario.

This paper is organized as follows: In second section we will highlight some of the basic concepts related to the model such as bivariate Bernoulli Distribution; Learning Effects and Frechet-Hoeffding bound. Also, a two-prospect case is presented in order to introduce the basic feature of the potential learning embedded in the model. In section three we will present the basic model with up to four prospects. Besides the main equations we will present the optimal sequence, the portfolio value and the decision rule (to drill or not, given previous results). In section four we will present more complex and realistic portfolios with up to ten prospects. The chosen optimization method and the general results will be presented. Last, in section five will be the conclusions.

2 Basic Concepts and Motivation It was considered that the correlation between prospects plays a key-role in the optimization model. Modern finance theory¹ emphasizes that a negative correlation is a desirable feature among financial assets in order to properly diversificate the portfolio risk. In this paper, however, it was considered a different assumption. In real assets, a positive correlation means similar structures and consequently potential learning effects embedded. This subject has been researched both on the academic and on the practitioner's field in the past 20 years. DIAS (2005a), DIAS (2005b), SMITH and THOMPSON(2004), BROSCHE (2001) and BLAU (2001) were major contributors to the topic.

At the present model, the uncertainty related to the existence (or not) of oil in any prospect was considered and plays a major role in the portfolio's value discussion. The uncertainty was modelled using a Bernoulli distribution and, usually, as a starting point it was considered that all prospects have the same probability of success ("wet") represented by the Chance Factor (CF). Also, each prospect needs some investment to be drilled (INV) and, in case of success, the respective exploratory and development campaign will result in a certain Net Present Value (NPV). For a certain prospect (n), the Expected Monetary Value (EMV) is given by:

¹ See Markowitz (1951)

$$EMV_n = -INV_n + CF_n NPV_n \quad (1)$$

Where:

$$CF_n \sim Ber(p) \quad (2)$$

Let's consider initially a two-prospects portfolio with: (i) same probability of success of 30%; (ii) same wildcat investments USD 70 MM and (iii) same NPV of USD 200 MM. Representing the portfolio value when drilling first prospect 1 and then prospect 2 as Π_{1+2} then:

$$EMV_1 = -70 + 30\% 200 = -10 \quad (3)$$

$$EMV_2 = -70 + 30\% 200 = -10 \quad (4)$$

$$\Pi_{1+2} = -20 \quad (5)$$

The rational decision would be to not drill any prospect. Therefore, the previous analysis did not consider the portfolio learning effects. It wasn't considered that could be some degree of correlation among the prospects such that the drilling outcome of one prospect could bring some additional information to the possible outcome of the other prospect. The bivariate Bernoulli distribution was used to model the joint distribution of both prospects. As a constraint to the use of such distribution it was necessary to check if the Frechet-Hoeffding lower and upper limits are valid. For a given initial Chance Factor (CF_0) and probability of success (q) the Frechet-Hoeffding limits are given by²:

$$Max \left\{ -\sqrt{\frac{CF_0 q}{(1-CF_0)(1-q)}}, -\sqrt{\frac{(1-CF_0)(1-q)}{CF_0 q}} \right\} \leq \rho \leq \sqrt{\frac{Min\{CF_0, q\}(1-Max\{CF_0, q\})}{Max\{CF_0, q\}(1-Min\{CF_0, q\})}} \quad (6)$$

From the revelation distributions process literature the Chance Factor update formula is defined as³:

$$CF^+ = CF_0 + \sqrt{\frac{1-q}{q}} \sqrt{CF_0(1-CF_0)} \sqrt{\eta^2(CF|S)} \quad (7)$$

² See JOE (1997, p. 210) for the proof.

³ See DIAS (2005a) for the proof.

Where $\sqrt{\eta^2(CF|S)} = \rho$

$$CF^- = CF_0 - \sqrt{\frac{q}{1-q}} \sqrt{CF_0(1-CF_0)} \sqrt{\eta^2(CF|S)} \quad (8)$$

Assume, for example, that the prospects are correlated ($\rho = \eta^2 = 50\%$). In case of success on drilling the first prospect then the second prospect CF will be updated to $CF^+ = 65\%$ and, in case of failure it will be updated to $CF^- = 15\%$. As we can see, clearly the update process will chance portfolio value. Recalculating the EMV of each prospect, we have:

$$EMV_1 = -70 + 30\% 200 = -10 \quad (9)$$

$$EMV_2 \text{ (In case of success in 1)} = -70 + 65\% 200 = +60 \quad (10)$$

$$EMV_2 \text{ (In case of failure in 1)} = -70 + 15\% 200 = -40 \quad (11)$$

Considering that the decision to drill any prospect is an option, the portfolio value is given by:

$$\Pi_{A+B} = \max\{0, EMV_A + CF_A \text{Max}\{EMV_A^+, 0\} + (1 - CF_A) \text{Max}\{EMV_2^-, 0\}\} \quad (12)$$

$$\Pi_{1+2} = \max\{0, -10 + 30\% \text{Max}\{60, 0\} + (1 - 30\%) \text{Max}\{-40, 0\}\} = +8 \quad (13)$$

The general conclusion then would be *Drill prospect 1 to acquire information. If the prospect 1 is “wet”, then drill prospects 2, otherwise, do not drill the prospect 2.* As shown in this simple case, the correlation between prospects is, indeed, a key factor to correctly evaluate the portfolio and design the optimal sequence of the exploration campaign. The decision tree below illustrates this simple case which only has 4 end nodes.

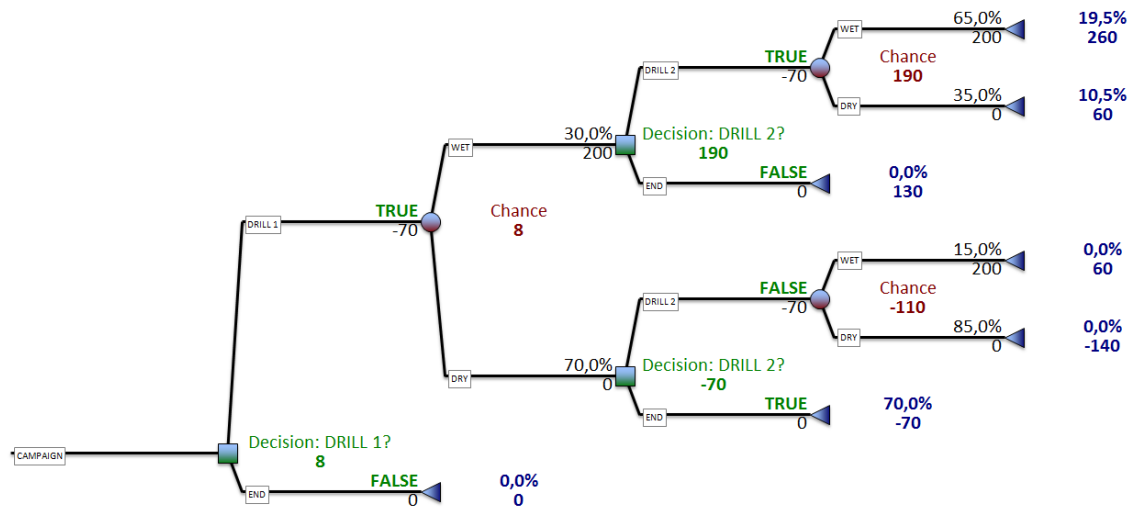


FIGURE I: Decision Tree – Two correlated prospects

It's interesting to highlight that even in the two-prospects case the drilling sequence might matter. If the prospects differ on any of the initial premises ($NPV_1 \neq NPV_2$, for example), the final portfolio value is different depending on the chosen drilling sequence⁴.

3 Modelling portfolios with three and four prospects It was shown that the correlation plays a key role on determining the portfolio value. In this model we assume that the correlation between two (or more) prospects is strictly related to the bi-dimensional distance between those prospects. It was considered that, for a given *Limit Distance* (D_{lim}), there might be similarities on the geological structure, such that the information acquired after drilling one certain prospect might be useful to reduce some of the uncertainties related to other (or others) correlated prospects. Evidently, the less distant the prospects are, the more correlated they tend to be. On the other hand, when determining a given prospect potential success, there are other factors that can't be directly related to distance. We assumed that, the maximum correlation strictly related to the distance is $\rho_{max} = 60\%$. Formally, given two prospects (a and b) bi-dimensional location (x,y) , the distance between them is defined by:

$$D_{a,b} = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \quad (14)$$

⁴ If, for example, $NPV_1 = \text{USD } 200 \text{ MM}$ and $NPV_2 = \text{USD } 300 \text{ MM}$, the EMV of the four possible drilling sequences are: $\pi_1 = -\text{USD } 10 \text{ MM}$; $\pi_2 = \text{USD } 20 \text{ MM}$; $\pi_{1+2} = \text{USD } 27,5 \text{ MM}$; $\pi_{2+1} = \text{USD } 38 \text{ MM}$.

It was assumed that the Limit Distance for which there might be some degree of correlation is $D_{lim} = 50$ km. The correlation between prospects a and b is defined by:

$$\rho_{a,b} = \frac{D_{lim} - D_{a,b}}{D_{lim}} \rho_{max} \text{ if } D_{a,b} < D_{lim} \quad (15)$$

$$\rho_{a,b} = 0 \text{ if } D_{a,b} \geq D_{lim} \quad (16)$$

The present model consists, basically of the following steps:

- (i). Given a certain portfolio configuration, find the distance matrix between the prospects and, as a consequence, the correlation matrix;
- (ii). Choose, randomly, one possible drilling sequence;
- (iii). For that sequence, calculate every updated Chance Factor in every possible outcome of the exploratory campaign (different *paths* on the decision tree);
- (iv). Calculate, *backwards*, the EMV of drilling each prospect;
- (v). Find the portfolio value for the sequence chosen on item (ii);
- (vi). Repeat steps (iii)-(v) for every other possible sequence for the portfolio defined in (i)
- (vii). Choose the sequence that accounts for the maximum EMV as the optimal one.

We will initially consider a portfolio composed of three prospects with the following configuration:

- (a) The distance between the prospects are given by: $D_{A-B} = 28,3$ km; $D_{B-C} = 26,2$ km and $D_{C-A} = 25,6$ km;
- (b) Every prospect has the same Chance Factors: $CF_A = CF_B = CF_C = 30\%$;
- (c) Every prospect needs the same amount of wildcat investments: $INV_A = INV_B = INV_C = USD 70$ MM;
- (d) Every prospect presents the same Net Present Value: $NPV_A = NPV_B = NPV_C = USD 200$ MM;
- (e) The decision to drill (or not) each prospect is optional (i.e. there's no obligation to proceed with any of the possible drilling procedures).

The figure II illustrates the configuration for the three-prospects portfolio. Each letter indicates the prospect spot and the black circles around it defines de maximum distance for which there may be some degree of correlation

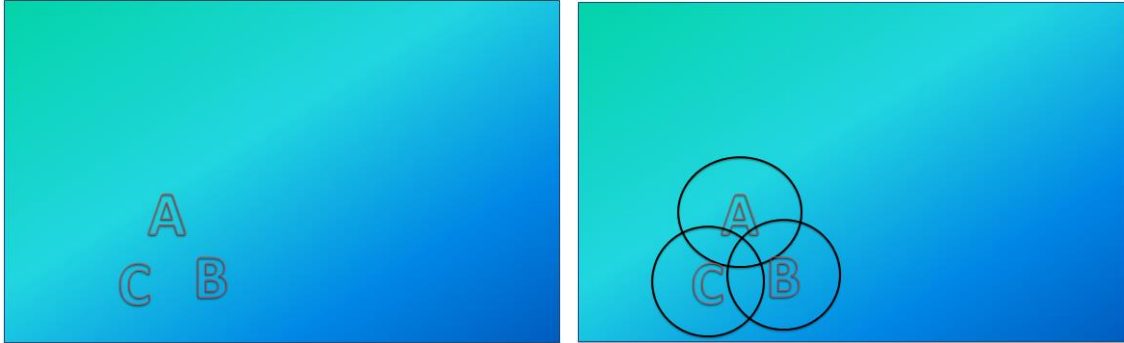


FIGURE II: Portfolio configuration – Three correlated prospects

Table I contains the correlation matrix for the three-prospects portfolio.

CORRELATION MATRIX	A	B	C
A	1,00	0,26	0,29
B	0,26	1,00	0,29
C	0,29	0,29	1,00

TABLE I: Three prospects portfolio correlation matrix

For one possible drilling sequence, *ABC* for example, it is necessary to calculate the portfolio EMV. To do that it must be followed six basic steps. (I) calculate the updated chance factors of both prospects B and C, given the information acquired when drilling the first prospect (A). From equations (7) and (8) it follows:

$$CF_B^+ = 30\% + \sqrt{\frac{1 - 30\%}{30\%}} \sqrt{30\%(1 - 30\%)26\%} = 48,2\% \quad (17)$$

$$CF_B^- = 30\% - \sqrt{\frac{30\%}{1 - 30\%}} \sqrt{30\%(1 - 30\%)26\%} = 22,2\% \quad (18)$$

$$CF_C^+ = 30\% + \sqrt{\frac{1 - 30\%}{30\%}} \sqrt{30\%(1 - 30\%)29\%} = 50,5\% \quad (19)$$

$$CF_C^- = 30\% - \sqrt{\frac{30\%}{1 - 30\%}} \sqrt{30\%(1 - 30\%)29\%} = 21,2\% \quad (20)$$

(II) Calculate every possible updated CF for the third prospect (C) given every possible sequence of failure or success so far obtained in the exploratory campaign (A and B). In this case, we will have four updated CFs for the prospect C as:

$$CF_C^{++} = 50,5\% + \sqrt{\frac{1 - 48,2\%}{48,2\%}} \sqrt{50,5\%(1 - 50,5\%)26\%} = 65,3\% \quad (21)$$

$$CF_C^{+-} = 50,5\% - \sqrt{\frac{48,2\%}{1 - 48,2\%}} \sqrt{50,5\%(1 - 50,5\%)26\%} = 36,7\% \quad (22)$$

$$CF_C^{-+} = 21,2\% + \sqrt{\frac{1 - 22,2\%}{22,2\%}} \sqrt{21,2\%(1 - 21,2\%)26\%} = 43,1\% \quad (23)$$

$$CF_C^{--} = 21,2\% - \sqrt{\frac{22,2\%}{1 - 22,2\%}} \sqrt{21,2\%(1 - 21,2\%)26\%} = 15,0\% \quad (24)$$

(III) Calculate the EMV of every possible outcome for the prospect C as:

$$EMV_C^{++} = -70 + 65,3\% * 200 = 60,6 \quad (25)$$

$$EMV_C^{+-} = -70 + 36,7\% * 200 = 3,37 \quad (26)$$

$$EMV_C^{-+} = -70 + 43,1\% * 200 = 16,23 \quad (27)$$

$$EMV_C^{--} = -70 + 15\% * 200 = -40,0 \quad (28)$$

(IV) Calculate the prospect C value backward to every possible outcome for prospect B as:

$$\begin{aligned} EMV_{C|B+} &= \text{Max}\{0; 48,2\% \text{Max}[60,6, 0] + (1 - 48,2\%) \text{Max}[3,37, 0]\} \\ &= \text{USD } 31,0 \text{ MM} \end{aligned} \quad (29)$$

$$\begin{aligned} EMV_{C|B-} &= \text{Max}\{0; 22,2\% \text{Max}[16,23, 0] + (1 - 22,2\%) \text{Max}[-40, 0]\} \\ &= \text{USD } 3,60 \text{ MM} \end{aligned} \quad (30)$$

(V) Calculate the EMV of every possible outcome for the prospect B as:

$$EMV_B^+ = -70 + 48,2\% * 200 = 26,4 \quad (31)$$

$$EMV_B^- = -70 + 22,2\% * 200 = -25,6 \quad (32)$$

At last, (VI) calculate prospects B and C value backward to every possible outcome for prospect A as:

$$\begin{aligned} EMV_{C,B|A^+} &= \text{Max}\{0; 30\% \text{Max}[26,4 + 31, 0] + (1 - 30\%) \text{Max}[-70 + 31, 0]\} \\ &= \text{USD } 17,2 \text{ MM} \quad (33) \end{aligned}$$

$$\begin{aligned} EMV_{C,B|A^-} &= \text{Max}\{0; 70\% \text{Max}[-25,6 + 3,60, 0] + (1 - 70\%) \text{Max}[-70 + 3,60, 0]\} \\ &= 0 \quad (34) \end{aligned}$$

Given that the EMV for prospect A is –USD 10 MM, the expected gains from drilling prospect B and C are more than enough to compensate that expected loss. According to that, a general conclusion related to the ABC sequence is that the exploratory campaign as a whole exhibits a positive expected monetary value of + USD 7,2 MM.

In order to find the optimal sequence for the three-prospects case, it is necessary to repeat the previous steps for every possible sequence. Table II shows the results.

SEQUENCE	PORTFOLIO VALUE
ABC	7,2
ACB	7,6
BAC	6,9
BCA	7,4
CAB	8,3
CBA	8,3

TABLE II: Portfolio value – Three-prospects possible sequences

The optimal choice, in this case, is to drill first the prospect C and then the others prospects. It is interesting to note that, for a planning perspective, the Expected Monetary Value for the whole campaign is positive. But, on the other hand, once the campaign started, the actual results (not the expected ones) should be considered to evaluate which next prospect should be (or not) drilled. The figure III shows the optimal

choice (or *path*) to be taken when executing the exploratory campaign. From it follows that the optimal rule is: *Drill prospect C. In case of success drill the others prospects, otherwise, end the exploratory campaign.*

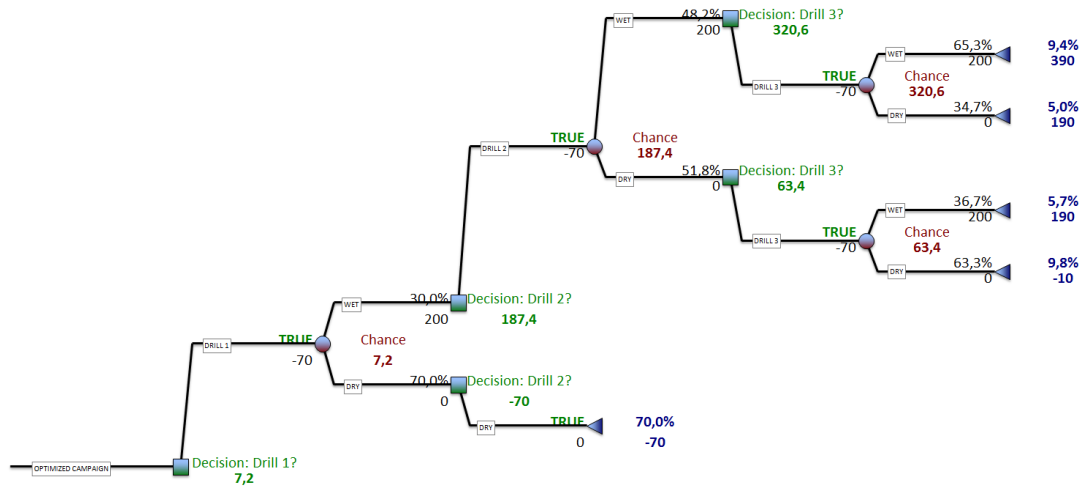


FIGURE III: Optimized Campaign – Sequence CAB

When extending the previous analysis to a four-prospects model evidently the amount of possible results and the calculation needed to find the optimal choice increases significantly. Besides that, the basic concepts and the methodology remains the same. Let's assume the same previous premises, except that now, the distance between the prospects are the ones shown on Table III:

DISTANCE MATRIX (KM)	A	B	C	D
A	0,0	28,3	25,6	49,0
B	28,3	0,0	26,2	28,8
C	25,6	26,2	0,0	28,6
D	49,0	28,8	28,6	0,0

TABLE III: Distance matrix – Four prospects portfolio

The figure IV illustrates the configuration for the four-prospects portfolio.

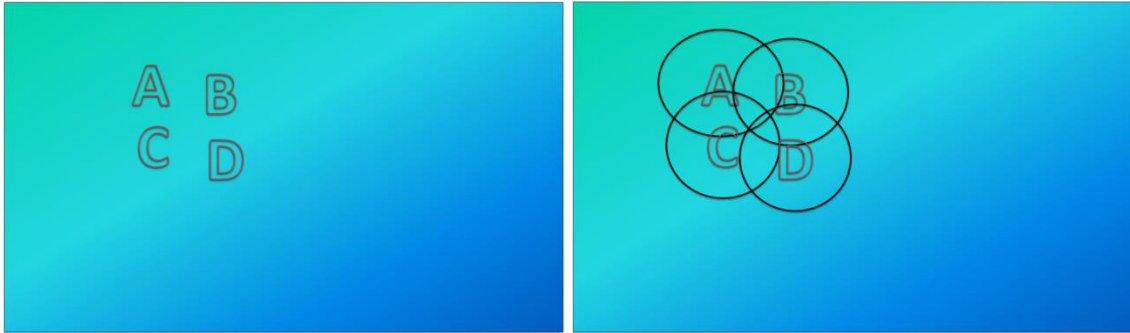


FIGURE IV: Portfolio configuration – Four correlated prospects

Table IV contains the correlation matrix for the four-prospects portfolio.

CORRELATION MATRIX	A	B	C	D
A	1,00	0,26	0,29	0,01
B	0,26	1,00	0,29	0,25
C	0,29	0,29	1,00	0,26
D	0,01	0,25	0,26	1,00

TABLE IV: Correlation matrix – Four prospects portfolio

After repeating the same steps used to calculate the three prospects portfolio EMV to the four-prospects model, table V shows the general results. It is worth noting at least two aspects: (i) Even though in every possible sequence the portfolio as whole presents a positive EMV, its value might increase significantly due to the optimal sequence chosen and (ii) in general, exploratory campaign that are initiated drilling the prospect C tend to present more expressive EMV.

SEQUENCE	PORTFOLIO VALUE
ABCD	10,47
ABDC	8,99
ACBD	10,85
ACDB	9,86
ADBC	5,83
ADCB	6,42
BACD	15,05
BADC	16,39
BCAD	15,06
BCDA	15,06
BDAC	16,4
BDCA	15,2

SEQUENCE	PORTFOLIO VALUE
CABD	16,44
CADB	17,63
CBAD	16,13
CBDA	16,1
CDAB	17,63
CDBA	16,21
DABC	4,97
DACB	4,72
DBAC	8,1
DBCA	9,22
DCAB	8,62
DCBA	9,36

$$\text{Amount of alternatives} = \sum_{p=0}^n \frac{n!}{(n-p)!} \quad (35)$$

where:

$$n \geq p$$

Considering, for example, a ten prospect portfolio, it follows that there will be 3.628.800 possible decision trees; each one composed of 1024 end nodes and up to 9.864.100 different alternatives to develop the exploratory campaign. Instead of calculating every possible alternative the model proposed uses Genetic Algorithm (GA) to find the optimal choice.

The GA considered is defined as a problem of order choosing whose representation is given by a list of prospects to be drilled. The GA basic premises considered were:

- (a) Cross-Over rate:80%
- (b) Mutation rate: 10%
- (c) Initial population: 100
- (d) Iteration limit: 3.000

Considerer the same premises about every prospect's CF, NPV and INV as before, such that, every prospect considered isolated from the portfolio doesn't present a positive EMV. The distances between the prospects are shown on Table VI.

CORRELATION MATRIX	A	B	C	D	E	F	G	H	I	J
A	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17	0,17	0,00
B	0,00	1,00	0,21	0,00	0,19	0,00	0,00	0,00	0,00	0,22
C	0,00	0,21	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,12
D	0,00	0,00	0,00	1,00	0,00	0,19	0,00	0,00	0,00	0,00
E	0,00	0,19	0,00	0,00	1,00	0,00	0,11	0,00	0,21	0,24
F	0,00	0,00	0,00	0,19	0,00	1,00	0,00	0,19	0,00	0,00
G	0,00	0,00	0,00	0,00	0,11	0,00	1,00	0,00	0,08	0,25
H	0,17	0,00	0,00	0,00	0,00	0,19	0,00	1,00	0,00	0,00
I	0,17	0,00	0,00	0,00	0,21	0,00	0,08	0,00	1,00	0,00
J	0,00	0,22	0,12	0,00	0,24	0,00	0,25	0,000	0,00	1,00

TABLE VI: Correlation matrix – Ten-prospects portfolio

The figure VI illustrates the configuration for the ten-prospects portfolio.

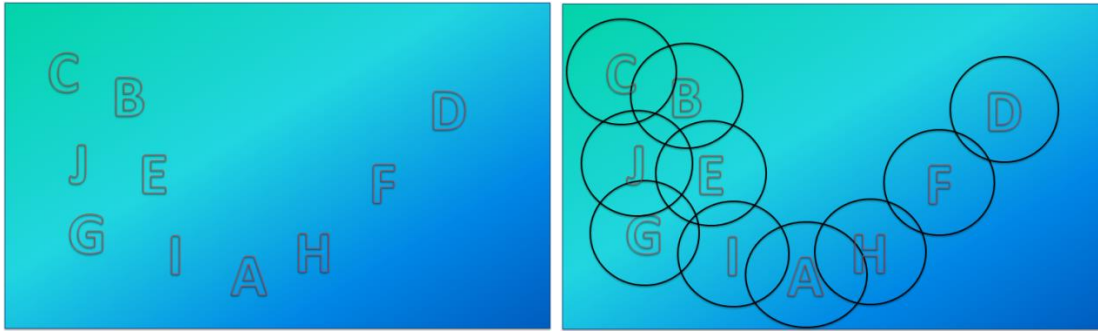


FIGURE VI: Portfolio configuration – Ten correlated prospects

Table VII contains the distance matrix for the ten-prospects portfolio.

DISTANCE MATRIX (KM)	A	B	C	D	E	F	G	H	I	J
A	0,0	94,7	122,9	93,6	61,67	63,39	77,98	35,46	35,51	87,45
B	94,7	0,0	32,5	115,6	34,28	111,63	58,80	112,49	67,02	31,42
C	122,9	32,5	0,0	147,0	61,27	144,17	67,33	143,87	91,24	40,40
D	93,6	115,6	147,0	0,0	105,39	34,14	144,22	67,98	109,57	133,67
E	61,67	34,28	61,27	105,39	1,00	91,85	40,85	84,44	32,87	30,25
F	63,39	111,63	144,17	34,14	91,85	1,00	126,04	34,07	86,41	122,00
G	77,98	58,80	67,33	144,22	40,85	126,04	1,00	110,03	42,92	28,80
H	35,46	112,49	143,87	67,98	84,44	34,07	110,03	1,00	67,26	113,76
I	35,51	67,02	91,24	109,57	32,87	86,41	42,92	67,26	1,00	53,27
J	87,45	31,42	40,40	133,67	30,25	122,00	28,80	113,761	53,27	1,00

TABLE VII: Distance matrix – Ten-prospects portfolio

The maximized EMV for the portfolio was reached after nearly 880 iteration on the GA as shown on Chart I.

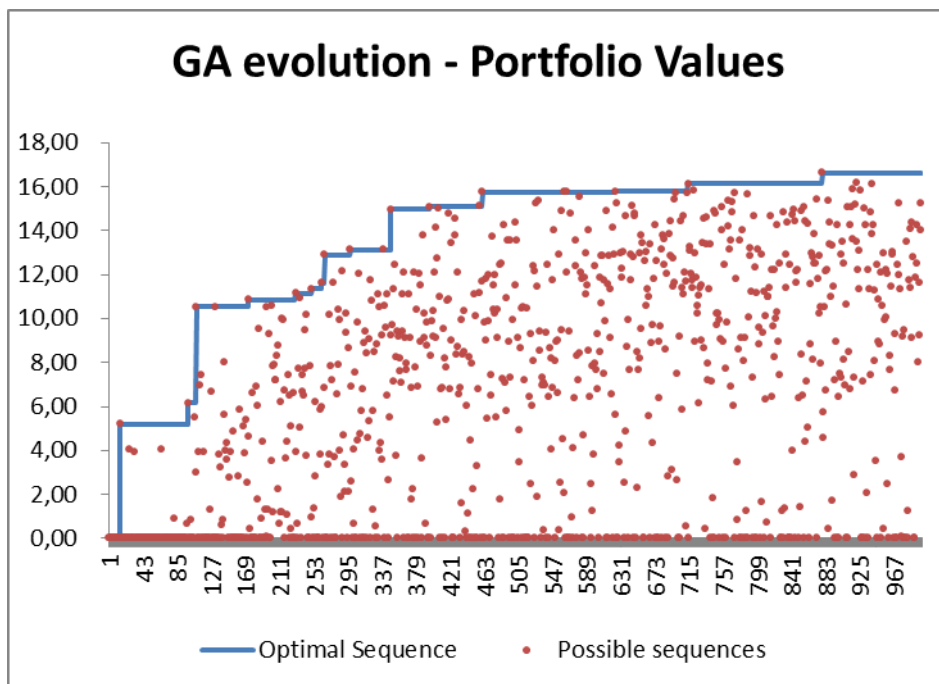


CHART I: GA Evolution

Table VII contains the optimal sequence and the value added by each prospect to the portfolio when drilled at that specific order. It is interesting to note at last two aspects of the results: (i) when drilling the first prospect (J) even though *directly* it doesn't add value to the portfolio, due to its potential to reveal information about others prospects (such as B,C,E and G), it is optimal to drill it first and (ii) when planning the optimized exploratory campaign, prospects D and F do not add any value to the portfolio at all. In such case the GA recommends to consider on the planned exploratory campaign only eight prospects. Evidently, once the campaign starts, the general results and the optimal decision choice might differ from the initially planned one.

Order	Prospect	EMV
1	J	0,00
2	B	0,00
3	G	4,05
4	E	7,32
5	C	3,88
6	I	0,81
7	A	0,42
8	H	0,17
9	D	0,00
10	F	0,00
TOTAL		16,65

TABLE VIII: Optimized portfolio – Ten-prospects portfolio

5 Conclusions Learning and optionality are two features that, when considered into the context of evaluating an exploratory campaign, might bring the problem to a more complex and realistic scenario. Also, on this background, the general conclusions might differ from the traditional analysis based only on the NPV with no optionality and potential learning.

The correlation among prospects is a desirable feature even in the case of “bad news”. As demonstrated, the appropriate portfolio valuation (with its embedded real options and learning effects) might avoid the investment in prospects that do not show economic viability.

The preliminary results shows that, prospects that reveal more information tend to be drilled at first, while, other prospects (uncorrelated to any other) should not be drilled at all. Additionally, the Genetic Algorithms demonstrated to be an effective way to approach the problem when considering portfolios with more than four prospects.

REFERENCE

BLAU, G. E. et al. Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. **Journal of Product Innovation Management**, [S.l.], v. 21, n. 4, p. 227-245, 2004.

BROSCH, R. **Portfolio-aspects in real options management**. Frankfurt, Main: Goethe-Universität, Feb. 2001. Working paper.

DIAS, M. A. G. Opções reais híbridas com aplicações em petróleo. Rio de Janeiro: PUC, Departamento de Engenharia Industrial, 2005a.

DIAS, M. A. G. Real options, learning measures, and Bernoulli revelation processes. In: ANNUAL INTERNATIONAL CONFERENCE ON REAL OPTIONS, 9th., 2005, Paris. **Proceedings...** Paris: [s.n.], 2005b.

JOE, H. **Multivariate models and dependence concepts**. London: Chapman & Hall: CRC, 1997. 399p.

MARKOWITZ, H. M. **Portfolio selection: efficient diversification of investments**. New Haven (USA): Cowles Foundation, Yale University Press, 1959.

SMITH, J. L.; THOMPSON, R. **Managing a portfolio of real options: sequential exploration of dependent prospects**. Dallas: Southern Methodist University, January 31, 2004. 26 p. Working paper.