FUNCTIONAL CORRELATION OF FP AND DC METHODS

Marin Kaluža
Polytechnic of Rijeka
Vukovarska 58, 51000 Rijeka, Croatia
E-mail: mkaluza@veleri.hr

Mirjana Rakamarić - Šegić
Polytechnic of Rijeka
Vukovarska 58, 51000 Rijeka, Croatia
E-mail: mirjana.rakamaric-segic@veleri.hr

Mile Pavlić
University of Rijeka, Department of Informatics
Omladinska 14, 51000 Rijeka, Croatia
E-mail: mile.pavlic@ris.hr

Abstract

Most of organizations today use information-communication technologies (ICT) for building an information system (IS). IS is assembled of hardware, software, network resources, organizational and human resources. In IS development process, complexity is crucial for evaluating quantities of resources needed (time, people, money, equipment). Complexity of an IS can be evaluated and/or measured in different phases of development. There are many methods for measuring complexity, but mostly used and thoroughly described method is Function Point Analysis (FP). The opposite method, Database Complexity (DC), does not measure all the aspects of IS, but it could evaluate system complexity depending on the database complexity. DC method is intended to be used for measuring semantic complexity of the IS database, and can be shown by counting attributes A and foreign keys F. This paper describes a very high correlation between FP and DC methods, and defines a function which can in 95% of accuracy express FP values from measured DC values.

Key words: *Function point analysis, Database complexity, Measure, Evaluation, Complexity, IS, Information system, Software*

1. INTRODUCTION

Most of organizations today use information communication technologies (ICT) to perform everyday obligations. ICT can be implemented on a basic level, like the global file repository. On the other side, ICT can be implemented as an Information system (IS) which is based on process organization of the objective reality. IS is composed of hardware, software, network, organizational and human resources. IS cannot exist without even one of this resources. Building an IS is a complex process which demands methodological approach – the use of methods and methodology.

The amount of time used on design and development all the elements of an IS affects the whole IS. IS with a bigger number and more diverse business activities are more complex than those which carry out a lesser amount of activities. Different types of IS have different types of complexity. In the IS development process, complexity is important for evaluating needed resources (time, people, money, equipment).

Complexity of an IS can be evaluated and/or measured in different stages of development. It is better to evaluate in initial stages because based on the evaluation it is possible to forecast the amount of resources needed for building the IS.

Evaluation/measurement needs to satisfy a couple of criteria: speed – measurement process needs to be carried out as fast as possible; accuracy – a safe algorithm must be formed; simplicity – the measuring process must be simple.

There are number of methods for measuring some parts of an IS. Some methods measure specific elements of a system which partially evaluates complexity, whilst some methods evaluate a greater number of crucial elements of the system and are capable for more accurate complexity evaluation.

This paper compares two methods for complexity evaluation: Function Point Analyses (FP) and Database Complexity (DC). The FP method is used as a referenced method. The DC method will be tested as a method for complexity evaluation of an IS. A correlation analysis will be carried out between the complexity measures of tested systems expressed with both methods. Statistical analysis will ascertain the mutual connections between these two methods.

If the results of performed FP method on some IS can also be calculated by performing a mathematical function on the DC results, there is no need to perform FP method for complexity estimation.

An application area of DC method can be increased. The DC method can be used for measuring a whole IS complexity.

The main goal of this research is to prove that DC method can be used for measuring a whole IS complexity, besides the comprehensive and difficult performing a FP method.

2. COMPLEXITY MEASURING METHODS

There is great number of methods available which in some ways measure system complexity: FP (Abran and Robillard, 1996), NVC (Ahm and Baker, 1998), Genetic taxonomy (Brumec and Dušak, 1999) and its implementation in modeling IS (Brumec and Vrček, 2002), COCOMO[1], Use Case Points[2], Delphi[3], PND (Poščić, 2007), DC (Pavlić, Kaluža and Vrček, 2008). Each method measure some element/s or part/s of an IS. FP method is the most comprehensive method, and it is used as a standard of complexity measurement.

The result of administering the FP method is dimensionless number which shows the number of functional elements (Input, Output, Inquiry, Master File). This method counts those functional elements and multiplies them with weighted factors (Abran and Robillard, 1991). The sum of all functional weights produces a number of unadjusted functional points (Abran and Robillard, 1996). Systems with a greater number of FP are more complex than those with a lesser number. The FP method can be applied in different stages of IS development and it is the most commonly used metric in software projects (Garmus, Herron, 2004). It can be applied in IS design process on process and data models, it can be applied on program solution, or even in IS re-engineering process. Applying the FP method is a difficult task which demands expert knowledge of technics and technologies for IS development.

The DC method shows semantic complexity of a database. Attributes (A) and foreign keys (F) are counted in a normalized database (Pavlić, Kaluža, and Vrček, 2008). The DC method does not evaluate process elements of the IS. The DC method is fast, accurate and objective because it is carried out by an automated procedure directly from the database.

3. MEASURED SYSTEMS

Measuring is carried out on 10 different systems. Table 1 shows results given through FP and DC methods. With FP method there is only one value per system. With DC method there are two values: attributes (A) and foreign keys (F).

---

[1] ftp://ftp.usc.edu/pub/soft_engineering/COCOMOII/cocomo99.0/modelman.pdf
[2] http://www.bfpug.com.br/artigos/ucp/damodaran-estimation_using_use_case_points.pdf
[3] http://millennium-project.org/FRMv3_0/04-Delphi.pdf

Table 1: Measured systems

PROJECT	ACRONYM	FP	DC	
			A	F
Asset management	OSA	301	427	48
Glass breakage	LSA	1798	1822	259
Earthquake	POT	1849	1840	259
Fire stocks	POZ	1886	1883	265
Machinery breakage	LST	1893	1887	266
Household	KU	2048	1939	274
Burglary and robbery	PKR	1933	1945	275
Fire summed/contracted	POS	2112	1984	280
Car all-risk insurance	KA	2261	2090	289
Car insurance	AO	2423	2179	296

Measuring is carried out on 10 different systems. Table 1 shows results given through FP and DC methods. With FP method there is only one value per system. With DC method there are two values: attributes (A) and foreign keys (F).

4. CORRELATION ANALYSIS

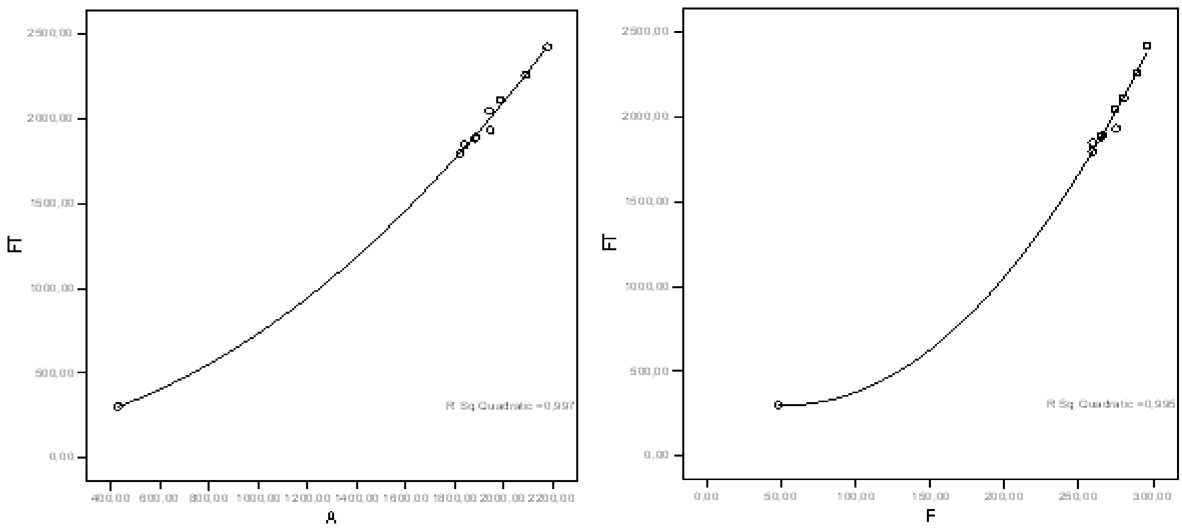

Figure 1: Scatter diagrams: A and FT, F and FT

Figure 1 shows scatter diagrams of paired values A and FT (FP) on the left side and F and FT (FP) on the right side.

Table 2: Correlations: A and FP, F and FP

Correlation		A	FP
A	Pearson Correlation	1	0,991
	Sig. (2-tailed)	.	0,000
	N	10	10
FP	Pearson Correlation	0,991	1
	Sig. (2-tailed)	0,000	.
	N	10	10

Correlation		F	FP
F	Pearson Correlation	1	0,983
	Sig. (2-tailed)	.	0,000
	N	10	10
FT	Pearson Correlation	0,983	1
	Sig. (2-tailed)	0,000	.
	N	10	10

Value of the Pearson coefficient is in both cases > 0,95 with a reliability level of Sig=0,000. These values show a strong positive connection between A and FP, and F and FP values pairs. These correlations are statistically significant because of the Sig value (Sig<0,05).

Table 3: FP forecast functions

Function		Unstandardized coefficients			R^2	t
		Constant	C1	C2		
FP=f(A)	Linear	-237,344	1,160		0,983	0,000
	Square	143,982	0,2017	0,0004	0,997	0,000
FP=f(F)	Linear	-122,92	7,8587		0,966	0,000
	Square	403,479	-3,8285	0,0355	0,995	0,000

Table 3 shows the forecast of FP values through linear and square regression for measured A and F values. It is visible in both cases (A and F) that higher forecast accuracy is gained through square regression (more than 99%). Square functions of FP forecast from A and F are:

- $FP = 143{,}982 + 0{,}2017A + 0{,}0004A^2$
- $FP = 403{,}479 - 3{,}8285F + 0{,}0355F^2$

4.1. Function definition

In the last paragraph a strong mutual connection is shown between values A and FP, and F and FP. It means that A can define FP with >99% certaincy. Also it is >99% certain that F can define the value of FP.

A and F are elements counted from the database and FP is a value gained by using function point analysis method. Function point analysis method can evaluate complexity of an IS. It is possible to evaluate the FP value with >99% certaincy based on singular values A and F. If it's possible to determine FP based on the values of A, and to determine FP based on the values of F, than it's obvious that both of these values together influence the complexity of an IS.

Because of excellent relationship between A and FP, and F and FP (especially square relationship), it's reasonable to assume that there is a square relationship between A and F together and FP. It is possible to set that kind of relation through general square function with two independent variables:

$$FP = aA^2 + bA + cAF + dF + eF^2 + f \tag{1}$$

Taking into consideration that *a*, …., *f* are ponders on A and F which are needed to forecast FP. Known measured values of 10 different systems are in A, F and FP. It's necessary to calculate the values of ponders *a*, …., *f*.

4.2. Calculation of ponders

The method of least squares has been used in the process of calculating the values of ponders. The method of least squares calculates functional dependency between experimental data by searching for function y=f(x) which has the least approximation error. In this case that is function FP= f(A,F). It is expected that the function applied on couples A and F would be somewhat different from evaluated values FP. That's why the searched function doesn't need to pass through points f(A,F) and FP. Searched function must enable the least possible scattering of given values around its graph - f(A,F) ≈ FT. If evaluated values A and F are set on the base of space, than there is no need to look for horizontal deviations e.g. deviations on the base of space. It's required to look only for vertical deviations e.g. the deviation of FP measured values and FP values gained from function: FP – f(A,F) or f(A,F) – FP. It's not necessary to evaluate if the error is positive or negative so the measurement for errors on any point can be observed as │f(A,F) - FT│. There will be a necessity for error derivability so it's wise to express the error gained from function approximation in a particular point with a formula:

$$E_i = \left(f(A,T) - FP\right)^2 \tag{2}$$

Total approximation error in that case is calculated as a sum of such singular, local errors:

$$E = \sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \left(f(A,T) - FP\right)^2 \tag{3}$$

(1) represents a function which demonstrates a relation between A, F and FP values. To calculate ponders *a*,…, *f* it is necessary to find the function's minimum in form of E(*a*,*b*,*c*,*d*,*e*,*f*).

(3) can be dissolved as a binomial square:

$$E = \sum_{i=1}^{n} f(A,T)^2 - 2\sum_{i=1}^{n} f(A,T)FP + \sum_{i=1}^{n} FP^2$$

It is necessary to calculate the extreme of the differentiable real function with six variables. Critical points of such a function are stationary points and possible ponders *a*,…, *f* are calculated by solving the system:

$$\nabla E = 0 \qquad \text{e.g.} \qquad \frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial c} = \frac{\partial E}{\partial d} = \frac{\partial E}{\partial e} = \frac{\partial E}{\partial f} = 0$$

Based on the definition above these expressions can be postulated:

1. $\frac{\partial E}{\partial a} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\cdot A^2\right) - 2\sum_{i=1}^{n}\left(A^2 \cdot FP\right) = 0$
2. $\frac{\partial E}{\partial b} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\cdot A\right) - 2\sum_{i=1}^{n}\left(A \cdot FP\right) = 0$
3. $\frac{\partial E}{\partial c} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\cdot A\cdot F\right) - 2\sum_{i=1}^{n}\left(A\cdot F \cdot FP\right) = 0$
4. $\frac{\partial E}{\partial d} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\cdot F\right) - 2\sum_{i=1}^{n}\left(F \cdot FP\right) = 0$
5. $\frac{\partial E}{\partial e} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\cdot F^2\right) - 2\sum_{i=1}^{n}\left(F^2 \cdot FP\right) = 0$
6. $\frac{\partial E}{\partial f} = 2\sum_{i=1}^{n}\left((aA^2 + bA + cA\cdot F + dF + eF^2 + f)\right) - 2\sum_{i=1}^{n}\left(FP\right) = 0$

After calculating the partial derivations from above:

1. $a\sum_{i=1}^{n} A^4 + b\sum_{i=1}^{n} A^3 + c\sum_{i=1}^{n} A^3F + d\sum_{i=1}^{n} A^2F + e\sum_{i=1}^{n} A^2F^2 + f\sum_{i=1}^{n} A^2 = \sum_{i=1}^{n} A^2FP$
2. $a\sum_{i=1}^{n} A^3 + b\sum_{i=1}^{n} A^2 + c\sum_{i=1}^{n} A^2F + d\sum_{i=1}^{n} A\cdot F + e\sum_{i=1}^{n} A\cdot F^2 + f\sum_{i=1}^{n} A = \sum_{i=1}^{n}\left(A\cdot FP\right)$
3. $a\sum_{i=1}^{n} A^3F + b\sum_{i=1}^{n} A^2F + c\sum_{i=1}^{n} A^2F^2 + d\sum_{i=1}^{n} A\cdot F^2 + e\sum_{i=1}^{n} A\cdot F^3 + f\sum_{i=1}^{n} A\cdot F = \sum_{i=1}^{n}\left(A\cdot F\cdot FP\right)$
4. $a\sum_{i=1}^{n} A^2F + b\sum_{i=1}^{n} A\cdot F + c\sum_{i=1}^{n} A\cdot F^2 + d\sum_{i=1}^{n} F^2 + e\sum_{i=1}^{n} F^3 + f\sum_{i=1}^{n} F = \sum_{i=1}^{n}\left(F\cdot FP\right)$
5. $a\sum_{i=1}^{n} A^2F^2 + b\sum_{i=1}^{n} A\cdot F^2 + c\sum_{i=1}^{n} A\cdot F^3 + d\sum_{i=1}^{n} F^3 + e\sum_{i=1}^{n} F^4 + f\sum_{i=1}^{n} F^2 = \sum_{i=1}^{n} F^2\cdot FP$
6. $a\sum_{i=1}^{n} A^2 + b\sum_{i=1}^{n} A + c\sum_{i=1}^{n} A\cdot F + d\sum_{i=1}^{n} F + e\sum_{i=1}^{n} F^2 + f = \sum_{i=1}^{n} FP$

Values A, F, and FP are gained by measuring and they are known. In singular sums of each equation there is a certain combination of A, F and FP. After substituting A, F and FP with measured values a system of six equations with six unknown parameters is gained. These unknown parameters are now ponders *a*,…, *f*.

EQ 1: $(1{,}33*10^{14})a+(6{,}77*10^{10})b+(1{,}86*10^{13})c+(9{,}47*10^{9})d+(2{,}61*10^{12})e+(3{,}46*10^{7})f=7{,}04*10^{10}$

EQ 2: $(6{,}77*10^{10})a+(3{,}46*10^{7})b+(9{,}47*10^{9})c+(4{,}84*10^{6})d+(1{,}33*10^{9})e+(1{,}80*10^{4})f=3{,}59*10^{7}$

EQ 3: $(1{,}86*10^{13})a+(9{,}47*10^{9})b+(2{,}61*10^{12})c+(1{,}33*10^{9})d+(3{,}65*10^{11})e+(4{,}84*10^{6})f=9{,}85*10^{9}$

EQ 4: $(9{,}47*10^{9})a+(4{,}84*10^{6})b+(1{,}33*10^{9})c+(6{,}78*10^{5})d+(1{,}86*10^{8})e+(2{,}51*10^{3})f=5{,}02*10^{6}$

EQ 5: $(2{,}61*10^{12})a+(1{,}33*10^{9})b+(3{,}65*10^{11})c+(1{,}86*10^{8})d+(5{,}11*10^{10})e+(6{,}78*10^{5})f=1{,}38*10^{9}$

EQ 6: $(3{,}46*10^{7})a+(1{,}80*10^{4})b+(4{,}84*10^{6})c+(2{,}51*10^{3})d+(6{,}78*10^{5})e+(1{,}00*10^{0})f=1{,}85*10^{4}$

Table 4 shows calculated values of ponders *a*,…., *f* – for original function (1).

Table 4: Calculated ponder values

Ponder	Value	Truncated Value
a	$-1{,}32340466211827*10^{-03}$	-0,001
b	$1{,}91303075442960*10^{+00}$	1,913
c	$1{,}82965620016386*10^{-02}$	0,018
d	$-1{,}14014197931288*10^{+01}$	-11,401
e	$-4{,}43924850928387*10^{-02}$	-0,044
f	$-3{,}97144575043959*10^{-03}$	-0,004

Based on calculated values of ponder *a*,…, *f* shown in Table 4, the original function (1) can be expressed:

$$FP=-0{,}001A^2+1{,}913A+0{,}018AF-11{,}401F-0{,}044F^2-0{,}004 \quad (4)$$

(4) shows that some of the values of ponders are very small, almost insignificant. It's presumed that the lesser ponder value, especially those >1, are insignificant. Although that is not always correct. Ignorance of the ponder *f* leaves the function without one of the parameters, so it becomes more simple but the forecast error it necessary enlarged. Ponders *a,c* and *e* have very low values but in combination with A^2, A*F and F^2 these values gain significance. The only ponder which always has a low absolute value is ponder *f*. If ponder *f* is removed from the function it would trigger a small

correction of the value f(A,F). Ponder *f* = -0,004 is completely insignificant when taking into consideration the measurement unit of the whole function. This function calculates values which are usually of great absolute values so a low ponder value can be ignored. Taking that into consideration a function can be written as such:

$$FP = -0{,}001A^2 + 1{,}913A + 0{,}018AF - 11{,}401F - 0{,}044F^2 \quad (5)$$

Ponder *f* has a very small absolute value so it is insignificant for the forecast of the FP value; it can be very well forecasted without ponder *f*.

It would be logical to approximate gained ponders *a*,…, *f* to values which are easily remembered. Table 5 shows these ponders and their approximated values:

Table 5: Ponders approximation

Ponder	Calculated Value	Approximated Value
a	-0,001	-0,001
b	1,913	1,2
c	0,018	0,02
d	-11,401	-11,5
e	-0,044	-0,05

Final formula can now be written in such form:

$$FP = -0{,}001A^2 + 1{,}2A + 0{,}02AF - 11{,}5F - 0{,}05F^2 \quad (6)$$

A formula above depicts a way to forecast FP value based on A and F values from the DC method.

4.3. Discussion

Table 5 shows measured A, F and FP values per each system observed. Also the calculated FP value is displayed by (4) in the fifth column – f(A,F). Table also shows absolute deviation of f(A,F) from real measured FP value. Results depict absolute deviations which vary from 0,02 to 75,17. The range of absolute deviation is great, but it is necessary to show the value of each of these relative deviations e.g. the value of relative forecast error. Relative deviations of measured value FP, calculated through (4), are also shown. Relative error varies from 0,01% to 3,89%. Relative error is less than 5% so it can be established that these results gained through (4) are statistically significant for the chosen sample of measured systems and (4) can, with 95% accuracy, express the FP value.

Significance of a formula, that is able to forecast the FP value with a 95% of accuracy, based on values A and F is shown previously. FP values are measured by function point analysis method which is already comprehensive so a great amount of time is necessary for its implementation. A and F values are measured by DC method.

Table 6: Measured and calculated FP values

Acronym	Measured Values			f(A,F)	Abs. Error f(A,F) – FT	Relative Error	
	FP	A	F			Under FP	Under f(A,F)
OSA	301	427	48	301,02	0,02	0,01%	0,01%
LSA	1798	1822	259	1795,50	2,50	0,14%	0,14%
POT	1849	1840	259	1828,00	21,00	1,14%	1,15%
POZ	1886	1883	265	1900,91	14,91	0,79%	0,78%
LST	1893	1887	266	1907,55	14,55	0,77%	0,76%
PKR	1933	1945	275	2008,17	75,17	3,89%	3,74%
KU	2048	1939	274	1997,64	50,36	2,46%	2,52%
POS	2112	1984	280	2077,53	34,47	1,63%	1,66%
KA	2261	2090	289	2266,06	5,06	0,22%	0,22%
AO	2423	2179	296	2421,59	1,41	0,06%	0,06%

5. CONCLUSION

Two methods are analyzed in this paper: Function Point Analysis and Database Complexity. Operational (in use) systems are measured with these methods.

FP method was used as a source of referent values. FP method is a generally accepted method for measuring an IS complexity. The procedure of implementing the FP method and calculating the unadjusted number of functional points is difficult.

DC method measures complexity of a database. A database is an element of software. DC does not measure process elements of the system.

A correlation analysis of system complexity gained by both methods is implemented. There is a high level (>99%) of connectivity between singular elements (A and F) from DC with the FP. The least square method is used to calculate ponders $a,\dots,f$ implemented on general square function with two variables (A and F). A function is re-controlled and a relative error <5% is gained.

This analysis proved that FP values can be expressed using mathematical function on DC values.

Taking into consideration that DC method can express FP values with over than 95% accuracy, it's possible to ascertain that DC method can also be used in purpose of measuring complexity of whole IS. Furthermore, there is no need for empirical performing a FP method. The FP method can be replaced with the DC method in common environment.

Practical advantages which DC has over FP are: user friendliness and measuring velocity.

REFERENCES

Abran, A. and Robillard, P. N. (1991), *Identification of the structural weaknesses of Function Point metrics*, 3rd Annual Oregon Workshop of Software Metrics, Portland, Oregon.

Abran, A. and Robillard, P. N. (1996), Function Point Analysis: An Empirical Study of Its Measurement Processes, *IEEE Transactions on Software Engineering*, Vol. 22, No. 12.

Ahm, S. I. and Baker, S. (1998), *The NVC Method of Software Project Estimation*, Jacksone Open Kilobyte Enterprises, http://www.tribalsmile.com/nvc [Accessed 27/04/2008].

Brumec, J. and Dušak, V. (1999), *The Assessment of IS Complexity Based on Genetic Taxonomy*, Evolution of Challenges in System Development, Kluwer Academic/Plenum Publishers, New York.

Brumec, J. and Vrček, N. (2002), *Genetic Taxonomy: the Theoretical Source for IS Modelling Methods*, Proceedings of the ISRM2002 Conference, Las Vegas, NV, USA.

Garmus, D. and Herron, D. (2004), *Function point analysis – measurement practices for successful software projects*, Addison-Wesley.

Pavlić, M., Kaluža, M. and Vrček, N. (2008), *Database complexity measuring method*, CECIIS, Varaždin.

Poščić, P., Pavlić, M. and Vrček, N. (2008), "Method for Estimating the Complexity of Designing Business Information System", *Journal of Information and Organizational Science (JIOS),* Vol. 32, No. 2, pp. 123-136.

ftp://ftp.usc.edu/pub/soft_engineering/COCOMOII/cocomo99.0/modelman.pdf, [Accessed 09/12/12].

http://www.bfpug.com.br/artigos/ucp/damodaran-estimation_using_use_case_points.pdf, [Accessed 09/12/12].

http://millennium-project.org/FRMv3_0/04-Delphi.pdf, [Accessed 09/12/12].