Subjective Consistency

Ing. Pedro E. Colla

Motorola Argentina Center for Software (MACS) – Hipolito Yrigoyen 146 – P 9
X5000JHO – Córdoba – Argentina
E-mail: colla@motorola.com

Abstract. *An alternate effort es timation methodology is proposed to be used at mature software engineering organizations where situations of insufficient historical data or activities resilient to prediction by statistical models exist, in this scenario experts become the most used and sometime only way to perform estimations As experts enter the scene subjectiveness come into play and consistency couldn't longer be assessed when performing estimations across similar situations or technical contexts. This paper proposes the usage of pairwise comparison methodologies not only by its own merits but also to allow the subjective factors integrated in the experts' effort estimation to be measured, an therefore assessed in terms of its consistency The benefits of the pairwise comparison methodologies in general, and the Saaty's Analytic Hierarchical Process (AHP) methodologies often associated with complex data manipulation are briefly revisited and adaptations are made to it in order to get most benefits from their usage with simple requirements in terms of the hierarchical modeling of the problem and data manipulation. Criteria are discussed to be used as quality figures and validations of the proposed methodology, in particular the usage of the Consistency Ratio (CR)A preliminary experiment is discussed and evaluated in order to validate the scope of the work and to help identify areas that are candidate to receive further attention*

Overview

A significative effort is usually expended at software development organizations on predicting the overall effort a project will take.

For those organizations following the maturity path given by the *SEI Capability Maturity Model (CMM)* a whole *Key Process Area* (*Software Project Planning KPA, Goal 1*) is devoted to the subject[16].

The classical approach to estimations is to collect metrics as projects are deployed and to feedback those metrics into the estimation process of new projects, metrics are often used to derive models of different nature, from pure statistical correlation all the way up to sophisticated dynamic effort allocation projections.

The approach is certainly sound and has received substantial attention in the Software Engineering bibliography[13 et al]

In essence, metrics try to identify and capture predictors of difficulty after the belief that effort highly correlates with the overall complexity of the development.

However, not all metrics as predictors could be extrapolated well across different project teams, frameworks and technologies as it's often difficult to identify metrics that capture the significant and relevant effort drivers in a meaningful way, specially when only limited history project data is available.

In other cases the predictors capture reasonably well the essence of the effort drivers but the number of available historical data is small and doesn't represent a solid sample from the statistical meaning standpoint.

At those cases where other methods have a high likelihood to fail, the estimation made by an expert or a small team of experts is the preferred one, and quite often the only mechanism to perform estimations.

Human experts tend to capture very well the objective or subjective factors that correlate with the overall effort and are also usually able to bridge similarities and differences across different technologies, frameworks and projects [1,2,4,12].

Many methodologies to guide an estimation process made by human experts have been proposed; only few have survived the acid test of real usage and even fewer deliver reasonably good results in a consistent way.

The estimation process led by experts usually yield an error when compared with actuals, meassured for example using the *Mean*

Magnitude of Relative Errors (MMRE), in the neighborhood of $\pm 30\%$ [1].

Magnitudes like this might seem to be too high, but in fact are quite acceptable as initial estimations for most modern processes executed at high maturity(') organizations [13,14,15].

However, all process improvement methodologies will require the main drivers for the estimation to be understood and experts often have a hard time when asked to explicitly outline the chain of factors they considered for a given estimation as the drivers for it; it's not a surprise that it's very difficult to work on their estimation rationale to identify ways to improve it and as a result to reduce the error level [4].

The experts will express their necessarily *subjective* evaluation usually without any concious regard to whether or not it is made with any degree of *consistency* in terms of the same criteria and factors being used repetitively to make the evaluation or assessment on similar conditions or contexts.

This paper explores how the usage of pairwise comparison methodologies might provide the tools for *consistent* evaluations being made out of subjective criteria, or back to the title of the paper how to evaluate the consistency of the inherently subjective estimation made by the experts.

A methodology like the Analytic Hierarchy Process (AHP) [4,5] proposed by Dr.Saaty relies on a mathematical framework that ultimately allows not only the outcome of an estimation to be identified but also what the overall consistency of it is in terms of a *Consistency Index (CI)* and a *Consistency Ratio (CR)*. A quick overview is referred to at the sections *Pairwise Comparisons Model* and *Analytic Hierarchy Process (AHP)*.

The *Analytic Hiearchy Process (AHP)* methodology has received substantial attention about its applications in Software Engineering matters (see [1,2,3,4,5,6,7,8,9.10 et al at the Reference section]) but has received

---

1 For the purposes of this work this refers to organization operating at CMM Level 2 or above.

relatively sparse actual application, so this paper aims, in the first place, at revisiting it and at the body of published materials as a reference for mature organizations to inspect it closely when in search of alternate estimation methodologies.

However, a quick review of published resources shows that proposed usages assume or rely on the existence of a relevant, and sometimes large, body of historical evidence and the development of highly elaborated hierarchycal models of properly callibrated predictors and models [1,2].

An original contribution made by this paper is the exploration of simpler ways to use this methodologies without even the need of complex hierarchical models to be developed and therefore gaining in simplicity of use.

They will be just using straight evaluations made by experts instead. In this way the overall framework will still allow to extract very substantial conclusions in terms of the consistency of the estimations under evaluation.

In order to avoid relying excessively on analysis of speculative nature some evaluations of the proposed framework is exercised with real data in order to quickly validate the main concepts and to identify areas that must receive further attention, the preliminary good results could be seen at the section *Preliminary Verification*.

The experimental verification shows also a reasonably good correlation exists between the *Consistency Ratio (CR)* and the *Mean Magnitude of Relative Error (MMRE)* clearly showing the first being a good quality calibration factor for the methodology under study.

Having evaluated the framework, conclusions could be extracted (see *Conclusions*) and further work (see *Further Work*) could be briefly discussed.

Pairwise Comparisons Model

It's a belief, supported by a fair amount of empirical evidence, that experts got the result of their estimations by means of complex and very subjective distance algorithms between the case under estimation and the body of past experiences of the person acting as the estimator [1].

In other words, each expert models the problem in terms of relevant features which to different degrees act as the predictors of the final outcome of the estimation process, that is the prediction of a given value.

Unfortunately, the feature used vary from expert to expert and the actual usage of them, the overall consistency of their estimation, being unknown, might widely vary from estimation to estimation.

When comparing complex features, people are inconsistent in performing comparisons according to some subjective preference criteria [5], they are much better at comparing pairs of objects though [3,4,8].

Prof.Thomas Saaty [4 et.al.] developed a sound framework to apply pairwise comparisons across several objects or attributes of objects in a consistent way. This approach relies on objective mathematics to process the inescapably subjectiveness and personal preferences of an individual expert or a group of experts while they make a decision, the overall methodology is referred as the *Analytic Hierarchy Process (AHP)*.

The methodology has been in use for the past two decades since the original seminal papers were first published. Although it hasn't been adopted as mainstream on software engineering usages the bibliography where applications are documented is certainly very significant [1,3,4,8,11].

However, at the bibliography, most authors rely on the elaboration of a hierarchical network of relevant factors related or associated with the process to make the estimation as well as the relative weight those factors has between them.

The bibliography is extensive in terms of how factors should be identified, scaled and compared under different circumstances [6,7,9] but little references do exist on what happens when an hierarchy of factors are very difficult to identify in the first place, or even when reliable factors which could be used as solid predictors don't exist at all.

Fortunately, the foundations of the pairwise methodologies in general, an the AHP methodology in particular, don't really require a hierarchy of factors to be developed([2]); it works equally well when the final estimation is made directly by pairwise comparisons.

Judgements or measurements on pair of elements with respect to controlling elements to derive ratio scales that are then synthesized throughout an hierarchycal structure to select the best alternative to a given problem or the best solution to a given estimation.

At its essence the AHP works by identifying a set or a hierarchy of contributors (or drivers), the comparison criteria is at the mercy of the understanding of the decision maker. Factors with both tangible and intangible measurement scales could be compared and merged into a single decision result as long as proper scaling and normalization precautions are taken [6,7,10].

The *Analytic Hierarchycal Process (AHP)* methodology could be used equally well for processes where a given decision needs to be selected among several equivalents and at ocasions where an estimation has to be made relative to others [3,4]; both cases are of good use at Software Engineering problems but this paper will concentrate mostly on the second.

One of the main benefits of using the AHP approach developed by Saaty as the underlying methodology is that the mathematical foundations of it allows for the objective meassurements of the inconsistency of the decisions being made or estimations being

[2] It could certainly be argued that the bigger the hierarchy is the most diluted any individual source of subjective error would become.

formulated. This will be used to provide a very much needed objective quality factor that could be used to understand the overall quality of the expert estimation.

Analytic Hierarchycal Process

The *Analytic Hierarchycal Process (AHP)* first introduced by Saaty ([3] et.al) starts by having an n-dimensional Cardinal Ratio Scale ($\vec{W}$) from an nxn Matrix [A] obtained from

$$\binom{n}{2} = \frac{n(n+1)}{2} \quad \text{[Eq 1]}$$

pairwise comparisons. Given n objects, let

$$[A] = \lfloor a_{j,k} \rfloor \quad \text{[Eq 2]}$$

be an nxn reflexive matrix obtained by evaluating $\binom{n}{2}$ pairwise comparisons, such that

$$1 \le a_{ij} \le m$$

being m the *scale maximum* and with the reflexive condition being met

$$a_{jj} = 1 \text{ for all j=1,…,n}$$

Then considering the relation

$$a_{ij} = \frac{1}{a_{ji}} \quad \text{[Eq 3]}$$

We could express each element as the relation between the corresponding pair of elements been compared, what we are going to call *weights* (W) as in:

$$a_{ij} = \frac{w_i}{w_j} \qquad \text{[Eq 4]}$$

We could make at this point certain assumptions about the underlying structure of the matrix nxn [A] obtained in this way, one of them would be the *constructive assumption* which states that all entries are positive and expressed as:

$$a_{ij} > 0 \qquad \text{[Eq 5]}$$

Also, the consistency assumption will mean transitivity exists and could be expressed as:

$$a_{ik} \times a_{kj} = a_{ik} \qquad \text{[Eq 6]}$$

If transitivity exists then the following will hold true:

$$a_{ii} = 1 \qquad \text{[Eq 7]}$$

and also [Eq.3]will be verified, therefore the transpose of [A] is also its inverse.

If we define a *weight vector* as the relations between the parameters we want to determine:

$$\vec{W} = \lfloor w_1, \ldots, w_n \rfloor \qquad \text{[Eq 8]}$$

After obtaining the matrix [A] by performing pairwise comparisons using a suitable comparison scale [6] we would clearly obtain

$$[A] \times \vec{W} = n\vec{W} \qquad \text{[Eq 9]}$$

since

$$\begin{pmatrix} 1 & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \dots & \frac{w_2}{w_n} \\ & & \dots & \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix} = \begin{pmatrix} nw_1 \\ nw_2 \\ \dots \\ nw_n \end{pmatrix} \qquad \text{[Eq 10]}$$

Therefore, n is an *eigenvalue* of Matrix [A], since the entries of [A] are non-negatives all of A's eigenvalues are non-negatives as well, so the following relation is satisfied:

$$\forall i \rightarrow l_i \geq 0 \qquad \text{[Eq 11]}$$

As a result

$$l_{\max} = l_1 = n \qquad \text{[Eq 12]}$$

and for I=2,..n

$$l_i = 0 \qquad \text{[Eq 13]}$$

The [A] matrix will therefore be ideal in the sense it will be perfectly consistent an with rank 1 thus characterized by satisfying that only one row is independent, and using the consistency assumption we can recreate all other rows from the first one

$$a_{ik} \times a_{kj} = a_{ij} \qquad \text{[Eq 14]}$$

When the estimation process is carried a new real matrix [A'] will be elicited, the persons providing the pairwise comparisons may not be consistent and this will be reflected on the transitive nature of the new matrix not being true, so:

$$a_{ik} \times a_{kj} \neq a_{ij} \qquad \text{[Eq 15]}$$

Both [A] and [A'] should be fairly similar with very small changes between the entries of them, therefore we can manage the case as a small eigenvalue perturbation, in the new scenario we should expect then that the maximum eigenvalue of [A’] will be around n.

$$\lambda_{max} = \lambda_1 \neq n \qquad \text{[Eq 16]}$$

The other eigenvalues will be positive and around zero.

$$\lambda_i \geq 0 \qquad \text{[Eq 17]}$$

and

$$\lambda_i \leq \varepsilon \qquad \text{[Eq 18]}$$

With ε a small residual value close to zero, finally the solution will be

$$[A'] \times \vec{W} = \lambda_{max} \vec{W} \qquad \textbf{[Eq 19]}$$

The so called Weight vector ($\vec{W}$) is then the Eigenvector of [A’] corresponding to its maximum value, the methodology then introduces also a new factor that could be understood as the distance between the maximum eigenvalue and n that receives the name *Consistency Index (CI)* and it is expressed by:

$$CI = \frac{|\lambda_{max} - n|}{n-1} \qquad \text{[Eq 20]}$$

The *Consistency Index (CI)* could be used to measure how the real meassurement departs from the ideal one and therefore the degree of inconsistency of the expert while making the prediction using this methodology.

Software Effort Estimation

The framework presented in the previous section could be employed as a concrete methodology to be used in the estimation of a software project effort.

A typical pairwise methodology allows for a fairly complex hierarchy of drivers and factors to be defined in order to enable the bottom-up computation of a final outcome.

As stated early we'll use the framework with just one parameter at the top level at a fraction of the overall complexity but still retaining most of the benefits of the methodology.

This will allow the methodology of being relatively free of complex validations of the correlation between the factors being used and the ones being calculated; it will focus also on a pure expert estimation while retaining at least partially the consistency measurement benefits already identified by means of the ability to compute the Consistency Index as a quality factor of it.

The initial step is to identify a body of historical data in terms of past projects (or activities) and the effort that each one took.

Given a set of *efforts* (E) from previous (known) projects:

$$E = \{E_1, ..., E_n\} \qquad \text{[Eq 21]}$$

The selected set has to be relevant in terms of the project and technology context of the development to be made; it could be arbitrarily long in size but for practical reasons we'll eventually try to keep the total number of cases at a reasonable level, the total number of pairwise comparisons will be

$$\frac{n(n-1)}{2} \cong \frac{n^2}{2} \qquad \text{[Eq 22]}$$

So it's in our best interest to keep the comparison process effort at manageable levels by not allowing an excessive number of comparisons.

The matrix [A] could therefore be built using the ratios between known efforts in a pairwise style, all ratios must satisfy:

$$a_{ij} \geq 0 \quad \text{[Eq 23]}$$

A pairwise comparison between efforts would allow the creation of a matrix with the pairwise relation between them:

$$[A] = \begin{pmatrix} a_{11} = \dfrac{E_1}{E_1} & a_{12} = \dfrac{E_1}{E_2} & \dots \\ \dots & \dots\dots & \dots\dots \\ a_{n1} = \dfrac{E_n}{E_1} & \dots\dots & \dots\dots \end{pmatrix} \quad \text{[Eq 24]}$$

In order to introduce the unknown factor to be computed (E_x) the dimension of the basic matrix [A] will be increased by one turning it into as **zxz** matrix with **z=n+1**

The added row and column will reflect the comparisons between the existing base of project data with the one under estimation (E_x).

Being z=n+1

$$[A^+] = \begin{bmatrix} a_{11} = \dfrac{E_1}{E_1} & \dots & a_{1z} = \dfrac{E_1}{E_x} \\ \dots & \dots & \dots \\ a_{z1} = \dfrac{E_x}{E_1} & \dots & a_{zz} = \dfrac{E_x}{E_x} \end{bmatrix} \quad \text{[Eq 25]}$$

Please note that only half the new matrix elements need to be computed since we could rely on the reflexiveness discussed in the previous section [Eq.3] to complete the other half.

However, we don't really know the magnitude of the effort that we're adding, that is exactly what we're trying to define, so we'll need to identify the values of the added matrix elements by indirect means.

In order to compare the magnitudes being evaluated against the reference set of data, a proper scale needs to be defined [3,6], several authors discuss extensively what the best criteria to define such scaling problem is, we'll propose to use:

Table 1 Calibration Factors

Numerical Values	Definition
1	Similar or Equal
2	Slightly bigger
3	Clearly Bigger
4	Much Bigger
5	Very much bigger

After we assign to each comparison a suitable value We do need to ensure all measurements are on the same scale so a normalization has to be made in a way that each node of the matrix is replaced by

$$\left[A^N\right]=\begin{pmatrix} \frac{a_{11}}{\sum_{i=1} a_{i1}} & \frac{a_{12}}{\sum_{i=1} a_{i2}} & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \frac{a_{n1}}{\sum_{i=1} a_{i1}} & \frac{a_{n2}}{\sum_{i=1} a_{i2}} & \ldots\ldots \end{pmatrix} \qquad \text{[Eq 26]}$$

Which also could be stated as

$$\left[A^N\right]=\begin{pmatrix} a_{11}^N & \ldots & a_{1z}^N \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ a_{z1}^N & \ldots & a_{zz}^N \end{pmatrix} \qquad \text{[Eq 27]}$$

The weight vector ($\vec{W}$) is then computed calculating the row averages of the normalized matrix:

$$\vec{W} = \begin{pmatrix} \frac{1}{z}\sum_{i=1}^{z} a_{1i}^{N} \\ \dots \\ \frac{1}{z}\sum_{i=1}^{z} a_{zi}^{N} \end{pmatrix} \qquad \text{[Eq 28]}$$

The consistency vector ($\vec{C}$) could be computed by multiplying the original (not normalized) [A] matrix by the weight vector to get the products and finally dividing it by the normalized weights to adjust for differences in scales.

$$\vec{I} = [A] \times \vec{W} \qquad \text{[Eq 29]}$$

And

$$\vec{C} = \begin{bmatrix} \frac{I_1}{w_1} \\ \cdots \\ \cdots \\ \frac{I_z}{w_z} \end{bmatrix} \qquad \text{[Eq 30]}$$

We're then able to compute the *Consistency Index (CI)* as defined by Saaty using the relation:

$$CI = \frac{l - z}{n} \qquad \text{[Eq 31]}$$

Where the *maximum principal eigenvalue of the comparisson matrix* (λ) *is the average value of the consistency vector*, and z is the number of potential factors being compared including the unknown one so z=n+1.

For each matrix of size z, random matrixes are generated and their mean *Consistency Index (CI)* value is computed, which is called the *Random Index (RI).*

Using these values, the *Consistency Ratio (CR)* is defined as the ratio of the *Consistency Index* and the *Random Index*

$$CR = \frac{CI}{RI} \qquad \text{[Eq 32]}$$

The meaning of the Consis tency Ratio is the measure of how a given matrix is as compared to a purely random matrix.

In other words, it is a measurement of the coherence of the information represented by that matrix, so it could be used as an indicator that provides a quality factor saying how consistent were the subjective decisions put together by the experts on their judgement.

The *Random Index* could be computed as a reference for n dimension matrices as follows [3]:

Table 2 Random Index Values (n=1,..,10)

n	RI
2	0
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.50

A value of the CR < 0.1 is typically considered acceptable; larger values require the decision-maker to reduce the inconsistencies by revising judgements.

Once the [A’] matrix computation is completed and it’s coherence could be objectively assessed using the Consistency Ratio, the weights obtained could be used to compute the unknown estimation under evaluation.

Being

$$p = w_z \qquad \text{[Eq 33]}$$

The zth weight corresponding to the unknown activity under estimation the associated *Effort* (E_x) could be computed relative to the other efforts associated with the known activities using:

$$E_x = \frac{p}{(1-p)} \times \sum_{i=1}^{n} E_i \qquad \text{[Eq 34]}$$

A way to meassure the distance between the forecasted and actual values is to hand in the concept of *Mean Magnitude of Relative Error (MMRE)*

$$MMRE = \frac{|E - E'|}{E} \times \frac{100}{n} \qquad \text{[Eq 35]}$$

Preliminary Verification

A preliminary verification is proposed by means of an experiment aiming at exercising the proposed methodology with real data and capturing the methodology that could be exercised on practical cases as a preliminary empirical validation.

The experiment will start by using process data collected during the execution of a development and maintenance project related to a representative application which is being run using a mature software engineering process with the highest ratings of compliance, quality metrics and customer satisfaction.

A sample of 18 requirements were collected from the project metrics database, all of them with known initial forecasted and actual metrics in order to allow the computation of error magnitudes between estimated and actual values.

The requirements were split using a random methodology into two groups, a *control group* and a *target group*.

The control group will be used during the experiment as the reference historical information in order to predict the target group metrics.

Some values of it would be:

Table 3 Control Group Metrics

CaseId	Size Estim (Locs)	Effort Estim (X)	Size Actual (Locs)	Effort (Actual) (Y)
ACECQ00000054	113.00	70.69	285.00	122.75
ACECQ00001678	40.00	38.45	64.00	45.47
ACECQ00001904	81.00	68.99	210.00	93.38
ACECQ00002028	60.00	43.31	45.00	38.24
ACECQ00002050	48.00	60.56	166.00	63.82
ACECQ00002076	45.00	60.12	83.00	98.86
ACECQ00002128	63.00	73.25	547.00	112.61
ACECQ00002249	70.00	48.00	70.00	26.38
ACECQ00002252	60.00	39.00	35.00	11.00

At the control group, a correlation between the *Estimated size (in LOCS)* and the *actual effort (E) in staff-days* shows up to be extremely poor ($r = 0.45$) which might indicate the lines of code indicator behave poorly[3] as an effort predictor.

At the same time the forecasted effort (E') also in staff-days and the actual effort (E) correlate very well ($r = 0.91$) leading to the interpretation that experts are better
at capturing the essence of the effort drivers than actual metrics estimators.

In our case, the MMRE of the estimations made by the regular process are still somewhat high, being the prediction consistently lower than the actual with an average error in the neighbor of 20%.

The experiment is started building the matrix [A] with the activities at the control group being used as a reference.

An estimating session with the participation of a group of experts is then planned and held comparing each requirement in the target group as the unknown effort to be evaluated with the entire control group.

A separate matrix is buiilt for each estimation being made, with a total of 9 pairwise comparisons estimated, as a result a total of 9 different z=10 matrixes are developed during the process.

Each estimation is computed separately to avoid coupling and inconsistency issues because of the similar size of the control group with the target group.

In order to dilute a likely skew in the comparison process only the upper half of the estimations made showing the better CR value are finally used to perform the experimental evaluation.

Each 10x10 matrix with the full control data corpus and one additional row and column for the requirement from the target group is being estimated.

---

[3] This is not intended to be yet another argument in the Lines of Code (LOC) goodness or poorness as a predictor, in this particular case it's likely that other metrics would fail as well.

Using the weights vector for the matrix associated which each target requirement allows the computation of the unknown effort variable using the relation shown at [Eq 35].

Since all actual efforts of the target group are known, the error between the estimated value and the actual one could be computed by means of the MMRE of it.

Each matrix will then have an associated CR, so the relation between this factor and the MMRE of each estimation performed with this methodology could be computed.

The results could be seen at the follow ing table:

Table 4 Relation CR vs. MMRE for the experimental data

ID	Effort Estim	CR	MMRE	Effort Actual
ACECQ00002133	25.77	0.011	0.062	27.47
ACECQ00002048	65.75	0.016	-0.010	65.12
ACECQ00002159	263.33	0.029	-0.069	246.33
ACECQ00002104	273.12	0.031	0.037	283.54

The absolute magnitude of the MMRE of the estimation increases as the CR of the set used rises, which should be expected and it is very consistent with the assumptions of the model.

On further analysis, a simple regression model linking the CR with the resulting MMRE of the estimation using the associated matrix could be developed, both factors result on showing a fairly high correlation ($r = 0.94$)

$$MMRE' = 19.19 \times CR - 0.21 \quad \text{[Eq 36]}$$

This simple regression model suggests that if we want to keep the MMRE at or below the levels obtained with other methodologies, we should accept pairwise comparison matrix satisfying that

$$CR \leq 0.02 \qquad [Eq\ 37]$$

Which could be used until further evidence is collected as acceptance criteria and quality for the methodology.

Matrixes that showns Consistency Ratio values higher than this threshold limit should guide the estimator or team of estimators to do additional estimations in order to increase its consistency.

Conclusions

The proposed approach exhibits the ability to allow the estimation activities to capture the good ingredients of expert subjective judgement and still allowing to increase the consistency.

The usefulness of the Consistency Ratio as the quality gate for the methodology has been strongly suggested by the preliminary verification of the methodology.

Mature organizations with limited history of metrics or executing projects whose nature makes them resilient to conventional estimation methodologies should pay a closer look to the pairwise comparison methodologies in general and the adaptations proposed by this paper in particular.

Further Work

Preliminary results clearly encourage to continue evaluating the methodology at situations where experts' opinions are required and consistently provide better results than metrics oriented or other methods of heuristic nature.

Additional experimental corroboration of the results and limits of the methodology is clearly indicated before it could be used as the main estimation method, in particular the role of the Consistency Ratio (CR) as a predictor must receive further exploration and analysis.

Finally, suitable tools must be evaluated and eventually adopted in order to cope with the sheer numerical intensity of the model on most real world situations.

Acknowledgements

The *Test Management System (TMS)* project team at *Motorola Argentina Centro de Software (GSG)*, Ramiro Perez Brisco, Leonardo Frittelli, Natacha Lascano, Julio Staude and Carlos Lucero, provided the metrics and valuable time to run the experiments presented in this paper. This very paper won't be possible without their process execution excellence. Finally, my thanks to Adriana Buglio who translated this paper into English.

References

[1] Shepperd M. and Cartwright M.: Predicting with Sparse Data, -IEEE Transactions of Software Engineering Vol 27 No 11 Nov 2001

[2] Shepperd M. and Schofield M.: Estimating Software Project Effort Using Analogies, IEEE Transactions of Software Engineering Vol 23 No 12 Nov 1997

[3] Bonissone P. : Determination of Weight Vectors from Pairwise Comparisons, Lectures, © 1998.

[4] Saaty T.: The Seven Pilars of the Analytic Hierarchy Process.

[5] Saaty T.: Fundamentals of the Analytic Network Process, ISAHP 1999 Kobe Japan 12-14,1999

[6] Zhang Q., Nishimura T. : A Method of Evaluation for Scaling in the Analytic Hiearchy Process, Section of Transportation Planning, Dept of Civil Engineering, Osaka City University. IEEE 1996.

[7] Perego A. : On integrating Tangible and Intangible Measures in AHP Applications. A Reference Framework. Dept of Mechanics, Politecnico di Milano (Italy) and Andrea Rangone, Dept of Economics and Production, Politecnico di Milano (Italy). IEEE 1996.

[8] Yannou B. , Limayem F. : Les Methodes de Comparaison par paires, Interet foundamental, Methodes practiques, Avancees scientifiques et logiciel. Ecole Central Paris. Reserche Scientifique.

[9] Sugihara K., Ishii H. , Tanaka H. : Fuzzy AHP with incomplete Information - 2001 IEEE.

[10] Shrabonti G. Alam S. : Ranking by AHP: A Rough Approach - Indian Institute of Technology Kharagpur. ISIF 2002.

[11] Mustajoki J, Hamalainen R. : Web-HIPRE, a Java Applet for AHP and Value Tree Analysis. Systems Analysis Laboratory - Helsinki University of Technology. ISAHP 1999. Kobe Japan August 12-14 1999

[12] Barker S., Shepperd M. , Aylett M. : The analytic hierarchy process and almost dataless prediction- Project Control for Software Quality extract. Editors Rob Kusters, Adrian Cowderoy, Fred Heemstra and Erik van Veenendaal. Shaker Publishing 1999 ISBN 90-423-0075-2

[13] Cerrillo D. : Planificacion y Gestion de Sistemas de Informacion, Estimacion de Software. Universidad de Castilla-La Mancha. Escuela Superior de Informatica. Mayo de 1999.

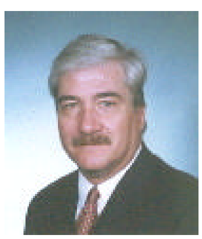
[14] Ferens D. : Software Size Estimation, Quo Vadis? Air force Research Laboratory (AFRL/IFSD). National Estimator, Winter 1999.

[15] Briand L.C. , El Emam K. , Surmann D. , Wieczorek I. : An Assessment and Comparison of Common Software Cost Estimation Modeling Techniques. Fraunhofer Institute for Experimental Software Engineering. Germany. International Software Engineering Research Network Technical Report ISERN-98-27

[16] Paulk H. , Weber H. , Weber C. , Garcia S. , Chrissis M.B. , Bush M. : Key Practices of the Capability Maturity Model (CMM) Version 1.1- SEI Carnegie Mellon University Technical Report – CMU/SEI-93-TR-025.

About the Author

Pedro E. Colla *Graduated as Electronic Engineer at the University of Buenos Aires in 1981, pursued BCS postgraduate studies in 1991 and earned a MBA degree in 1997. He has extensive background in the IT industry at both technical and management responsibilities.. Now he works at Motorola GSG – Argentina Center as Development Operations Manager.*



Address:
Motorola GSG - Argentina
Hipólito Yrigoyen 146 – piso 9
X5000JHO – Córdoba
ARGENTINA
E-mail: colla@motorola.com