Improving comparative analysis for the evaluation of AOSE methodologies

Luca Cernuzzi*

Departamento de Ingeniería Electrónica e Informática,
Universidad Católica 'Nuestra Señora de la Asunción'
Campus Universitario, C.C. 1683, Asunción, Paraguay
E-mail: lcernuzz@uca.edu.py
*Corresponding author

Franco Zambonelli

Dipartimento di Scienze e Metodi dell' Ingegneria,
Università di Modena e Reggio Emilia,
Via Amendola 2, 42100 Reggio Emilia, Italy
E-mail: franco.zambonelli@unimore.it

Comment [t1]: Author: Please confirm if L. Cernuzzi is the corresponding author.

Abstract: Agent-oriented methodologies, as they have been proposed so far, mainly try to suggest a clean and disciplined approach to analyse, design and develop MASs by using specific methods and techniques. Moreover, different studies have been proposed for the evaluation of agent-oriented methodologies adopting specific types of evaluation and criteria. However, little effort has been devoted to the comparison among such different evaluations. Comparison techniques may help in finding out new information from the existing studies, consolidating the results from the available evaluations, and consequently, in obtaining greater reliability from the evaluation results and their acceptance. With the aim of improving the acceptability of agent-oriented methodologies evaluation in the agent community, among the existing comparative techniques, the paper proposes the profile analysis technique for comparing evaluations carried out by different authors (perhaps using different evaluation frameworks). To exemplify the proposal, we present the application of the profile analysis technique on a case study.

Keywords: agent oriented methodologies; evaluation; profile analysis.

Reference to this paper should be made as follows: Cernuzzi, L. and Zambonelli, F. (xxxx) 'Improving comparative analysis for the evaluation of AOSE methodologies', *Int. Agent-Oriented Software Engineering*, Vol. X, No. Y, pp.000–000.

Biographical notes: Luca Cernuzzi is a Full Professor in Software Engineering Discipline in the Department of Electronics and Computer Science Engineering (DEI in Spanish) at the Universidad Católica 'Nuestra Señora de la Asunción' – UC (Asunción, Paraguay). Actually, he is the Dean of the School of Sciences and Technology at UC. He obtained his Laurea in Computer Science in the University of Milan – Italy, 1990, and PhD in Engineering at the University of Modena e Reggio Emilia – Italy, 2007. He has lectured as a Visiting Professor at several European and Latin American universities. His current research interests include: software engineering, multi-agent systems, web engineering, ICT for good, educational computer science (with emphasis on children with special needs). In these areas, he has published over 70 papers in journals, book chapters, and international conferences and workshops.

Franco Zambonelli is a Professor in Computer Science at the University of Modena and Reggio Emilia since 2001. He obtained his Laurea in Electronic Engineering in 1992, and PhD in Computer Science in 1997, both from the University of Bologna. His current research interests include: distributed and pervasive computing, agent-oriented software engineering, self-organisation in distributed systems engineering. In these areas, he has published over 130 papers in international fora, co-edited seven books, received several best paper awards, and has been invited Speaker and Tutorialist in several international conferences and workshops. He is a member of IEEE, ACM, AIIA, and TABOO.

1 Introduction and motivation

The need for reliable methodologies for designing and developing agents and multi-agents systems (MASs) is recognised as a fundamental issue in the agent oriented software engineering (AOSE) discipline (Zambonelli and Omicini, 2004). Nowadays, a wide range of agent-oriented methodologies are available for agent-based system designers (Iglesias et al., 1997; Ciancarini and Wooldridge, 2001). However, one of the open problems for agent-oriented software engineering to become 'mainstream' is the lack of consensus between the different methodologies that have been proposed. Moreover, in most cases, an agreement cannot be reached even on the kinds of concepts the methodology should support. Given this state of affairs, it may be very interesting for agent-based system designers to carry out an analysis or evaluation of the existing methodologies that would be most appropriate to use in each case or, at least, the interesting ones for their projects. In fact, one of the first topics that agent engineers have to deal with is the choice of a specific methodology for a particular project. Different factors may influence this choice and it would be interesting to consider them in the analysis of the advantages and drawbacks of a particular methodology.

In traditional software engineering, different efforts have been focused on the evaluation of processes and methodologies, especially in object-orientation and also in agent-orientation in the last few years. In AOSE, the specialised literature also presents different studies on the evaluation of AOSE methodologies (Shehory and Sturm, 2001; Cernuzzi and Rossi, 2002; Cuesta et al., 2003; Dam, 2003; Dam and Winikoff, 2003; Sturm et al., 2004; Tran and Low, 2005). Interesting results have been presented in such studies; however, some of them may be considered of low relevance since the authors themselves evaluate their own methodology. Despite the good intentions of such researchers, in general, it is quite hard to be objective in rating one's own work.

Still, one of the most relevant issues is related to the comparison among different evaluations, which is the main focus of this study. In fact, without a comparison (or carrying on just informal comparisons) between evaluations of the same methodologies from different experts, each individual evaluation offers a relative (reduced) value that is difficult to be accepted in the community of multi-agent researchers and developers. Attempting a comparison among the available studies we found some difficulties to deal with. Sometimes, the AOSE methodology evaluations presented so far adopt different criteria; which introduces some complications in comparing them. In addition, some proposals are based on different types of evaluations that are not totally congruent among them. It is worth noting that the vast majority of the studies are based on feature-based evaluation which is usually a qualitative technique and qualitative evaluations are difficult to compare. Qualitative evaluations can also be 'quantified' by assigning scores, assessing scales, weights and aggregating these results (Cernuzzi and Rossi, 2002). Though this approach still deals with qualitative aspects of the methodology, applying quantification can facilitate the use of comparison techniques between different evaluations. Comparison techniques may then help to find out new information from existing studies, to consolidate the results from available evaluations and consequently to obtain greater reliability from evaluation results and their acceptance. Among the existing comparative techniques, we pay special attention to the profile analysis method (Morrison, 1976; Nunnally, 1978; Reis, 1997), which is a multivariate statistical method for the quantitative evaluation of profiles. Profile analysis allows comparing information from different sources (i.e., the previous evaluations of AOSE methodologies) by contrasting two or more profiles and therefore, obtaining more reliable and relevant information on the previous evaluations.

This paper focuses on the application of the profile analysis method to assess the evaluations of some AOSE methodologies carried out by different authors. Our main goal is to present the potential of profile analysis in the evaluation comparison process.

As a second goal, whenever it is possible, we will look for evaluation similarities to confirm results and thus make them more acceptable in the agent community. For this second goal, it is also possible to carry out different comparisons. For example, a first comparison may focus on different profiles (two or more) produced by two or many other different evaluators on a particular methodology. If it is possible to find high similarity between the profiles, this can too confirm the evaluations done and consequently to increase the reliability of the evaluation on that particular methodology. A second comparison could consider the evaluations of different (two or more) methodologies for each evaluator. If the comparison shows high degrees of similarity between them it is possible to reinforce the evidence that a given methodology is better (or worse) than the others for covering some specific aspects. More generally, considering the average of the scores on each dimension, it would be possible to establish that a methodology performs 'better' than others.

Confirming the evaluation results for one or different AOSE methodologies carries benefits to both designers and practitioners, by allowing them to assess more reliably the advantages and shortcomings of all the methodologies under evaluation.

The reminder of this paper is organised as follows. Section 2 analyses related work and points out the contribution of the present proposal. Section 3 describes the profile analysis method with its related coefficients. Section 4 introduces the methodological steps needed for the application of the profile analysis method and illustrates them through a case study. Finally, Section 5 concludes and highlights some future research directions.

2 Related works

As introduced in the previous section, important contributions in agent-orientation propose evaluation frameworks and apply them to the evaluation of different AOSE methodologies. According to Rossi and Siau (1998) the evaluations may be carried out with different techniques like the following: *feature-based analysis*, where the evaluation is done by referring to the available resources; *survey*, where the evaluation is done by examining the results of a survey distributed among practitioners; *case studies*, where the evaluation is done by examining the results of case study; and *field experiment*, where the evaluation is done by examining the results of field experiments. Other authors (Cuesta et al., 2003; Dam, 2003) introduce yet another technique called *structured analysis*, which tries to understand the common and different aspects of the AOSE methodologies exploring the models and processes the methodologies share and their distinguishing aspects. An additional relevant contribution comes from *action research* (Avison et al., 1999) which is an iterative process involving researchers and practitioners acting together on a particular cycle of activities, including problem diagnosis, action intervention, and reflective learning. Case study research frequently reports what researchers and practitioners say they do, while in action research the emphasis is more on what researchers and practitioners *actually do*. In the AOSE evaluations and assessments there are currently no studies that adopt the action research approach.

Initial contributions took a qualitative approach in which the main properties of the methodologies under study were outlined (Ciancarini and Wooldridge, 2001; Iglesias et al., 1997; Tveit, 2001). A further step on that direction was presented in Cernuzzi and Giret (2000) and Shehory and Sturm (2001). Both of these studies adopt a feature-based qualitative technique. In the first study, the authors address the criteria related to the agent characteristics and the process life cycle coverage. Their comparison covers six methodologies: agent oriented analysis and design (Burmeister, 1996); agent modelling technique for systems of BDI agents (Kinny et al., 1996); MASB (Moulin and Brassard, 1996); agent oriented methodology for enterprise modelling (Kendall et al., 1996); CoMoMAS (Glaser, 1996); and MAS-CommonKADS (Iglesias et al., 1997). In the second study, Shehory and Sturm (2001) introduce more specific criteria from agent-based systems and software engineering to analyse Gaia (Wooldridge et al., 2000), ADEPT (Jennings et al., 2000), and DESIRE (Brazier et al., 1997). Nevertheless, all these studies adopt qualitative evaluations, which are difficult to compare.

The structural analysis technique was adopted for the framework presented in Cuesta et al. (2003), which examines five categories that represent parts of a methodology:

1 the development process

2 the models

3 the supported agent concepts

4 additional modelling features

5 the methodology documentation.

The authors apply the framework to different relevant AOSE methodologies: Gaia (Wooldridge et al., 2000), MaSE (DeLoach et al., 2001), Tropos (Giunchiglia et al., 2002), MESSAGE/Ingenias (Caire et al., 2001; Gómez-Sanz and Pavón, 2003), and MAS-Common-KADS (Iglesias et al., 1997). Yet, this is another qualitative evaluation that is hard to compare with other evaluations.

An interesting proposal based on a survey technique used to compare five methodologies were presented in Dam (2003), and Dam and Winikoff (2003). The survey's subjects were the methodologies' authors and some students; thus, the evaluation results may be questionable for the possible biases of the authors and the low expertise of the students. However, two interesting points made in these works are:

1 the evaluation framework covers some relevant dimensions of agent concepts and the process associated to the development of agent-based systems

2 the proposed framework considers quantitative evaluation in at least some of the dimensions.

A more comprehensive study was presented in Sturm et al. (2004). In this work, the authors apply the feature analysis, the survey, and the structural analysis [according to the framework of Cuesta et al. (2003)] approaches for evaluating and comparing four agent-oriented methodologies (namely Gaia, Tropos, MaSE, and OPM/MAS (Sturm et al., 2003). One of the more interesting aspects of this study is the framework they adopt, which offers a well-defined and structured set of aspects that an agent-oriented methodology should include. Moreover, the provided framework is a qualitative one; however, it can be transformed into a quantitative one by borrowing the concepts from (Cernuzzi and Rossi, 2002) (as the same author's state).

To the best of our knowledge, none of the previous related studies in the agent-oriented area are directly focused on the issue of quantitative evaluation. Furthermore, as we have seen, usually the evaluation proposals presented so far are based on different types of evaluations that are not totally congruent and, even worse; they adopt different criteria introducing some difficulties in comparing the different evaluations. In addition, some works present the evaluation of a specific AOSE methodology done by the same authors of the methodology. Despite the good intentions of such researchers, in general it is quite hard to objectively evaluate one's own work. Moreover, another relevant problem of the previous studies is that the results are not compared (or are only informally compared) with those obtained by others authors. Consequently, each study offers a relative value that is harder to accept within the community of multi-agent researchers and developers.

Comparison techniques give the opportunity to improve the acceptance of the evaluation results depending of their accordance. In addition, since it is possible to systematically convert some qualitative evaluation into quantitative ones, useful contributions may be obtained by applying the profile analysis method to previous evaluation results. Therefore, profile analysis may be seen as a complement to the existing evaluation approaches. In effect, facilitating the comparison, it may be used by researchers, methodology designers, and practitioners to assess more reliably the advantages and shortcomings of all the methodologies under evaluation.

3 The profile analysis method

Profile analysis (Morrison, 1976; Nunnally, 1978; Reis, 1997) is a multivariate statistical method that allows the processing and the evaluation of profiles in specific areas in order to obtain quantitative results (Serafini, 1988). It allows comparing two or more profiles in the same space. It is possible to define a profile as "the results of the real or ideal description of a specific entity (human, project, institution, etc.) in terms of different dimensions, simultaneously considered, functionally or theoretically related, and which adopt the same scale of evaluation" (Serafini, 1988). In this definition, 'ideal' set consists of all variables having a value of 100% and represents the expected situation while 'real' represents the actual values of the dimensions. For example, choosing a methodology, and considering that we want to evaluate four specific criteria (e.g., autonomy, expressiveness, process phases, and scalability), each criterion is a dimension and consequently each profile has four dimensions.

In each evaluation relationship exists between the plan (what is expected) and the result (what is implemented) that is conceptually expressed in terms of distance/proximity. A greater distance means that we are far from the expected or ideal value. Considering that a profile may have n dimensions, to measure the distance we need to use an n-dimensional space; where the set of values of the n-dimensions determines a position of a point in this space. This way, it is possible to determine the Euclidean distance between the obtained point and the ideal one.

By having different profiles, it is possible to obtain a set of values, as estimation of each profile with regards to some ideal situation, to calculate the distance between two or more profiles and to analyse the similarities in the high and low points between profiles establishing the configurational similarity (CS). The coefficient of congruency may be used to quantitatively evaluate the Euclidean distance between profiles. It is defined as the proximity between the expected or ideal situation and the obtained or real one. The values of congruency range from 0 to 1.

Since profile analysis needs quantitative evaluations as input, in the case of information sources based on qualitative scores, a first step implies the conversion of the scores to a quantitative common scale. In doing so, we are aware that the resulting profiles are usually based not on an interval scale but on an ordinal scale. Arithmetic's operations (e.g., mean or the formulas defined below for congruency and CS) may also be applied on such scales with the awareness that the resulting values have to be interpreted not as exact values but they may be useful to compare different evaluations. Table 1 shows the absolute values of congruency and CS and their interpretations. Since the resulting values may be not exact and perhaps depend on the order of attributes, it is worth to note that each class of the interpretation considers a consistent interval which may be representative of a class of values.

Table 1 Interpretation of the coefficients of congruency – C and configurational similarity – CS

Absolute value of C/CS	*Interpretation*
0.90–1.00	Virtually perfect congruency/configurational similarity
0.70–0.89	High congruency/configurational similarity
0.40–0.69	Moderate congruency/configurational similarity
0.20–0.39	Low congruency/configurational similarity
0.00–0.19	Virtually no congruency/configurational similarity

The coefficient of simple congruency is used when all the variables or dimensions have the same weight in the evaluation. The formula is (Serafini, 1981):

$$C_{ij} = 1 - \frac{D_{ij}}{D_{\max}} \tag{1}$$

where

C_{ij} simple congruency of the values

D_{ij} Euclidean distance between profiles in a p-dimensional space

$D_{\max}$ maximum allowed distance for the common scale of the p-dimensions.

$$D_{ij} = \left[\sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^2 \right]^{1/2} \tag{2}$$

$$D_{\max} = T.p^{1/2} \tag{3}$$

where

k 1, 2, ……, p

p number of dimensions of the profile

X_{ik} value of the profile i in the dimension k

X_{jk} value of the profile j in the dimension k

T difference between the maximum and the minimum values of the scale. Usually, T should be constant according to the adopted scale for the profile values (e.g., ranging from 0 to 10; in this case T assumes the value 10).

The coefficient of weighted congruency may be used when the variables or dimensions have different weights in the evaluation. Since all the evaluation studies presented so far for AOSE methodologies give the same weight to each criterion, we adopt the simple congruency formula.

A complementary coefficient is the CS, which measures the similarity of the shape of the profiles by measuring the grade of correspondence of the high and low values between different profiles in each dimension. When graphically represented, such values are easy to compare. It is worth mentioning that different profiles may have great distances among them but still have great similarities in their configuration. The formula for the CS coefficient (Serafini, 1988) is:

$$CS_{ij} = \frac{\sum_{k=1}^{p} \left| d_{ik} + d_{jk} \right|}{\sum_{k=1}^{p} \left| d_{ik} + d_{jk} \right| + \sum_{k=1}^{p} \left| d_{ik} - d_{jk} \right|} \tag{4}$$

where

k 1, 2, ……, p

p Number of dimensions of the profile

If $k = 1$ $d_{ik} = d_{jk} = 0$

If $2 \le k \le p$:

d_{ik} $X_i(k-1) - X_{ik}$

d_{jk} $X_j(k-1) - X_{jk}$

d_{ik} difference between successive values of the profile X_i

d_{jk} difference between successive values of the profile X_j.

Similarly to the congruency coefficient, the values of CS range from 0 to 1, 1 being the perfect matching in the evaluation of profiles. Also, the interpretations of CS coefficient values are the same of the congruency (see Table 1). As already stated above, we are aware that the resulting values may be not exact (i.e., may vary according to the order of the dimensions and attributes). Nevertheless, independently of the order, the values of all possible CS and the mean belong to the same class of interpretation.

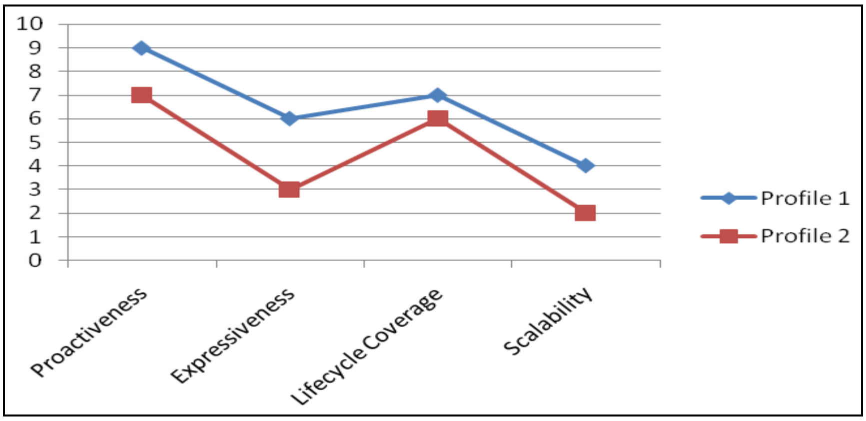
In Table 2, we show an example of the application of the CS coefficient.

A graphical representation of the values of Table 2 is shown in Figure 1. In the graph, it is possible to observe that there is a distance between the two profiles but both are similar in the high and low points. Specifically, they have a high CS.

Table 2 Values obtained considering two profiles for a methodology with four dimensions

Criteria or dimensions	*Profile 1 – values*	*Profile 2 – values*
Proactiveness	9	7
Expressiveness	6	3
Lifecycle coverage	7	6
Scalability	4	2

Figure 1 Graphical representation of the values of Table 2 (see online version for colours)



Let us illustrate this simple example calculating the value of congruency and CS.

From Table 3, we can obtain the congruency:

$$D_{12} = \left[2^2 + 3^2 + 1^2 + 2^2\right]^{1/2} = [4 + 9 + 1 + 4]^{1/2} = [18]^{1/2} \approx 4.24$$

$$D_{\max} = 10 * 4^{1/2} = 10 * 2 = 20$$

$$C_{12} = 1 - 4.24264 / 20 \approx 0.79$$

From Table 4, we can obtain the CS:

$$CS_{12} = 18 / (18 + 4) = 18 / 22 \approx 0.82$$

According to Table 1, it is possible to observe that the two profiles have high congruency and CS.

Table 3 Values for calculating the C coefficient among the two profiles of Table 2

k	*Profile 1*	*Profile 2*	*Profile1 – profile 2*
1	9	7	2
2	6	3	3
3	7	6	1
4	4	2	2

Table 4 Values for calculating the CS coefficient among the two profiles of Table 2

k	*Profile 1*	d_{1k}	*Profile 2*	d_{2k}	$d_{1k} - d_{2k}$	$d_{1k} + d_{2k}$
1	9	0	7	0	0	0
2	6	3	3	4	–1	7
3	7	–1	6	–3	2	–4
4	4	3	2	4	–1	7

4 Applying the profile analysis method

This section focuses on the application of the profile analysis method, by means of the congruency (C) and the CS coefficients, as a relevant approach to allow agent-oriented engineers to easily compare different evaluations of AOSE methodologies. It is worth recalling that the main objective of the present proposal focuses more on the potential of profile analysis for the evaluation comparison process than in the results of the comparison and the corresponding analysis.

To carry out a rigorous comparison several steps are needed. First, agent-oriented engineers have to choose the evaluation of methodologies to be compared; second, they have to define a common comparative framework; and finally, they may present the obtained results carrying out the corresponding analysis.

Therefore, the rest of the section is organised as follows. Section 4.1 analyses the issue of choosing the evaluations and the methodologies to be compared; evidently the different studies have to coincide in the methodologies and (probably) in the type of evaluation adopted. Section 4.2 illustrates the definition of the common comparative framework which includes the selection of common criteria, the definition of a common scale, and the transformation of qualitative attributes into quantitative values according the common scale. Section 4.3 applies the profile analysis method to a case study and presents an analysis of the results.

4.1 Choosing the evaluations and the methodologies

The more interesting evaluations studies in agent-oriented methodologies evaluation are Dam (2003), Cernuzzi and Rossi (2002), Dam and Winikoff (2003), Sturm et al. (2004), Tran and Low (2005), Elamy and Far (2008). Among these, we select the works of Dam and Winikoff (2003), and Sturm et al. (2004) for diverse reasons. They are quite often referenced in the specialised literature. Both works evaluate common methodologies applying sometimes the same evaluation methods (feature analysis, survey, etc.). They also present a partial similarity with regards to the evaluation criteria adopted. Finally, none of the authors of the present study have been involved in such evaluations.

The evaluated methodologies in the selected works are: MaSE (DeLoach et al., 2001), Tropos (Giunchiglia et al., 2002), Gaia (the original version) (Wooldridge et al., 2000), Prometheus (Padgham and Winikoff, 2002), MESSAGE (Caire et al., 2001), and OPM-MAS (Sturm et al., 2003) (currently, more recent references are available for many of such methodologies; nevertheless, we refer to those specified in the selected works). Among them, three are common to both works: MaSE, Tropos, and Gaia. However, both works have considered Gaia in its first version even though there is an official extension that brings substantial improvements (Zambonelli et al., 2003). It is worth to observe that we are more interested in presenting the advantages of the profile analysis method than the results of its application to a specific methodology. For these reasons, we decided to take into account just MaSE and Tropos for our purposes. It is relevant to mention that none of the selected methodologies has any of the evaluators between its authors.

Dam and Winikoff (2003) apply the feature-based analysis, survey, and case study techniques of evaluation while Sturm et al. apply feature-based analysis, survey, and structured analysis. Thus, in a first attempt we focused our study on feature-based analysis and survey types of evaluation which are common for both works. However, the evaluations based on the Survey type present some limitations. The authors of Sturm et al., 2004) based their work on the experience of a 15 students' course in which each methodology was used by a group of two or three students. This experience, despite the interesting information it provides, is not relevant from a statistical point of view. Moreover, the results suffer the lack of expertise of the students engaged in the experiment. In the work of Dam and Winikoff (2003), the survey is based on the evaluation done by the authors of the methodologies complemented by an experiment with five students (one per each methodology to be evaluated). For the student experiment, it is possible to assert the same considerations of the work by Sturm et al. (2004). On the other hand, the form completed by the authors of these methodologies may offer interesting information considering the high expertise of the evaluators, but the results may be strongly influenced by a bias on their judgment. For all these reasons, we decided to just focus on the feature-based analysis evaluation of both works.

4.2 Defining a common comparative framework

In the definition of a comparative unified evaluation framework, it is necessary to do some steps:

1 select the common criteria

2 transform the qualitative attributes into quantitative values

3 convert the values of the attributes to a common scale previously defined.

Let us analyse each step in depth.

4.2.1 Select the common criteria

For the first step, both evaluation works classify the attributes in four major aspects or categories which are: concepts and properties, modelling and notations, process, and pragmatics. However, in the attributes definition we found many discrepancies. There exist criteria with the same definition but different names and, vice versa, criteria with the same name but different semantics. Moreover, there are some attributes that have been directly defined as individual in a work; while in the other comprise a set of criteria. In these cases we grouped the comprised criteria in a unique (more general) criterion assigning it the mean of the values of the comprised criteria [e.g., lifecycle coverage in Dam and Winikoff (2003)]. Table 3 shows the final set of common attributes which are the basis for the application of the comparative framework. Please observe that * means a change in the original name.

Table 5 Selected attributes for comparing the evaluations

Dam and Winikoff, 2003	*Sturm et al., 2004*
	Concepts and properties
Autonomy	Autonomy
Reactive	Reactiveness
Proactive	Proactiveness
Mental attitudes	Mental notions (*)
Teamwork	Organisation (*)
Protocols	Protocol
	Modelling and notations
Consistency check	Analysability
Refinement	Complexity management
Language adequate and expressive	Expressiveness
Easy to use + easy to learn + clear notation	Accessibility
	Process
Lifecycle coverage (*)	Lifecycle coverage
	Pragmatics
Domain specific	Domain applicability
Scalable	Scalability
Maturity (quality) (*)	Resources

4.2.2 Transform the qualitative attributes into quantitative values

In the second step, we need to transform every qualitative evaluation into quantitative values. Moreover, we need to unify the scale of evaluation. In effect, in Sturm et al. (2004), the authors adopt a metric ranging from 1 to 7 for each attribute, being 1 the minimum and 7 the maximum value; while in Dam and Winikoff (2003) the authors adopt different metrics according to the attribute. These metrics, clarified according the work of Dam (2003), are:

- 'High – medium – low – none' for the attributes in the category concepts and properties
- 'Strongly disagree – disagree – neutral – agree – strongly agree' for the attributes in the category of modelling and notations; and pragmatics (scalable and maturity)
- 'Yes – no' for the attribute domain specific in the pragmatics category
- '0 – 1 – 2 – 3 – 3' – 4' for the attribute in the process category.

It is possible to observe that some of the metrics adopted by Dam and Winikoff (2003) are qualitative; therefore, we propose a scale for each metric to be converted to a quantitative scale. Being such scales independent, the same value for different attributes may have different meanings. The proposed conversions are:

- None – low – medium – high – > 0 – 1 – 2 – 3
- Strongly disagree – disagree – neutral – agree – strongly agree – > 0 – 1 – 2 – 3 – 4
- Yes – no – > 0 – 1 (considering it is a reverse scale)
- 0 – 1 – 2 – 3 – 3' – 4 – > 0 – 1 – 2 – 3 – 4 – 5.

Table 6 Conversion for the scale of the work of Dam and Winikoff

Scales Dam – Winikoff 2003	*Common scale*
Concepts and properties	
None (0)	0
Low (1)	3
Medium (2)	7
High (3)	10
Modelling and notations	
Strongly disagree (0)	0
Disagree (1)	3
Neutral (2)	5
Agree (3)	8
Strongly agree (4)	10
Process	
0	0
1	2
2	4
3	6
3'	8
4	10
Pragmatics	
Yes (0)	0
No (1)	10

4.2.3 Convert the values of the attributes to a common scale

Finally, for the third step, it is necessary to convert all the values to a comparable metric to obtain meaningful inferences from them. In doing such conversion, it is possible to slightly modify the results that may influence the congruency and CS coefficients.

The definition of a common scale has been widely treated in the specialised literature and there are no strong reasons to prefer a metric over the others as, except in particular cases, none of them apply to our situation. Therefore, we decided to adopt a common scale varying from 0 to 10. For the conversion, we calculate the percentage of the value into its original scale and then convert it to a percentage over the scale of 10. In case of a fractional result, we round it to the closest integer. Table 6 presents the conversion for the work of Dam and Winikoff (2003), while Table 7 presents the conversion for the scales presented in Sturm et al.

Table 7 Conversion for the scale of the work of Sturm et al.

Scale (Sturm et al., 2004)	*Common scale*
1	0
2	2
3	3
4	5
5	7
6	8
7	10

4.3 Results of the profile analysis

Hereinafter, we present the congruency and CS coefficients, the graphs with the comparisons, and a brief analysis of the results.

In Figure 2 and Figure 3, it is possible to compare the two evaluations (and the distance from the ideal profile) corresponding to MaSE and Tropos respectively.

Figure 2 Comparison for the MaSE methodology (see online version for colours)

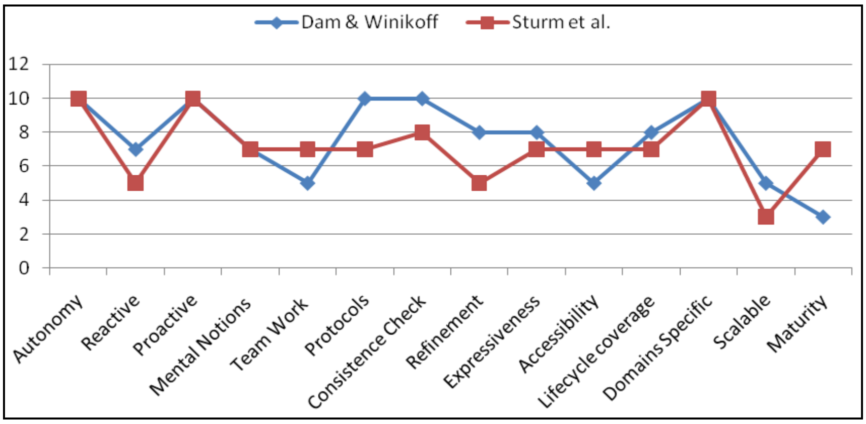
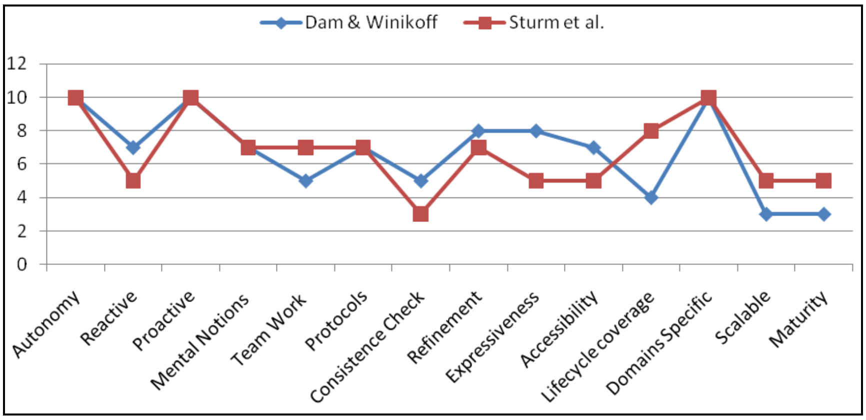


Figure 3 Comparison for the Tropos methodology (see online version for colours)



In Table 8, we present the obtained results considering the entire set of criteria. In all the tables below, DW means Dam and Winikoff, S means Sturm et al., and T is the constant defined in equation (3).

Table 8 Results considering the entire set of criteria

MaSE	C_{DWS} =	0.80	CS =	0.67
Tropos	C_{DWS} =	0.81	CS =	0.70

where

C_{DWS} represents the congruency between the profiles from the evaluations of Dam and Winikoff (2003) and Sturm et al. (2004)

CS represents the CS between both profiles.

It is possible to observe that in both cases the value of the congruency coefficient (C_{DWS}) between the two profiles reaches a 'high' level. This reflects the fact that the distance between the methodologies (MaSE and Tropos) is quite small. In other words, different evaluators coincide that MaSE and Tropos cover in a similar way the set of relevant attributes characterising AOSE methodologies. In addition, the CS for both methodologies varies from 'moderate' to 'high', that is, the profiles have a similar trend. Therefore, profile analysis shows that independent evaluators reach similar conclusions and this fact enhances the reliability of the results.

Another comparison could consider the evaluations of the two methodologies for each evaluator. Indeed, we are interested in assessing how the methodologies perform and in comparing them. If both evaluators confirm similar scores for a methodology, this reinforces the congruency of the methodology with regard to the better possible scores (i.e., the ideal profile). In addition, if both evaluators coincide in assessing a methodology as better than the other, this fact increases the reliability of such evaluation. Therefore, for this case it is not relevant to analyse the CS coefficient. The results of the comparison are shown in Figure 4 and Figure 5; and the congruency coefficients are presented in Table 9.

Figure 4 Comparison between MaSE and Tropos by Shehory et al. (see online version for colours)

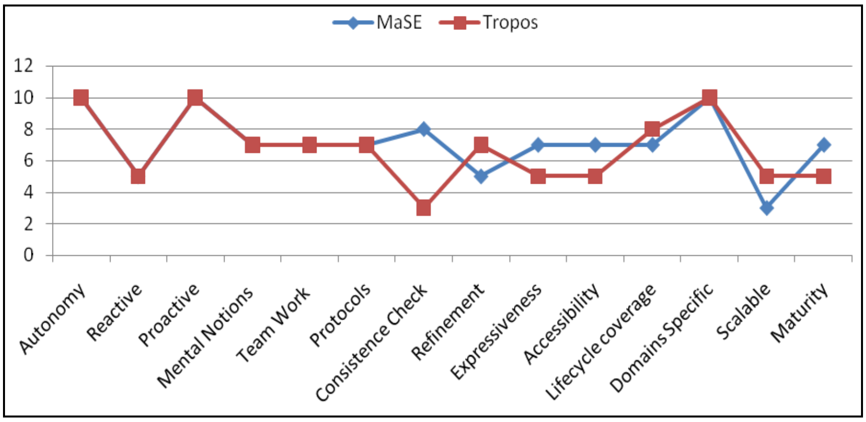


Figure 5 Comparison between MaSE and Tropos by Dam and Winikoff (see online version for colours)

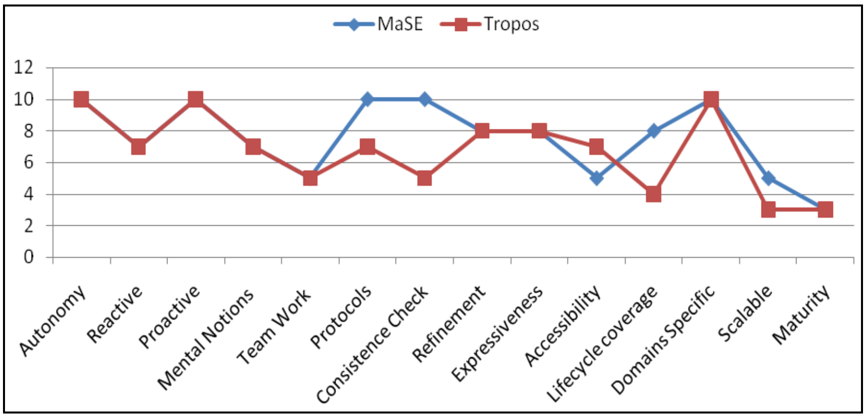


Table 9 Results considering the entire set of criteria

MaSE	C_{DWT} =	0.67	C_{ST} =	0.66
Tropos	C_{DWT} =	0.60	C_{ST} =	0.61

where

C_{DWT} represents the congruency between the ideal profile and the profile obtained from the evaluation of Dam and Winikoff (2003)

C_{ST} represents the congruency between the ideal profile and the profile obtained from the evaluation of Sturm et al. (2004).

In addition, it is possible to calculate the mean of the evaluation for MaSE and Tropos for each evaluator.

Based on these results it is possible to observe that both evaluators coincide in that MaSE is more congruent than Tropos with regard to the 'ideal' profile (i.e., the highest score in all dimensions and attributes). In effect, in the congruency coefficient, MaSE reached higher scores; which implies that it better covers the required attributes for AOSE methodologies. Moreover, the differences of the coefficient are similar for both works (see Table 9).

The evaluations are consistent and these evidences are reinforced by the mean of the scores on each dimension presented in Table 10. In effect, since the mean score of MaSE is higher than the Tropos one for both evaluators, it is more reliable to affirm that MaSE performs 'better' than Tropos for covering the dimensions and attributes.

Table 10 Means of the scoring considering the entire set of criteria

$MaSE_{DW}$	7.57	$MaSE_S$	7.14
$Tropos_{DW}$	6.71	$Tropos_S$	6.71

Table 11 Values for the concepts and properties dimension to calculate the C and CS coefficients for MaSE

Concepts	$MaSE_{DW}$	d_{DW}	$MaSE_S$	d_S	$d_{DW} - d_S$	$D_{DW} + d_S$	$MaSE_{DW} - MaSE_S$
Autonomy	10	0	10	0	0	0	0
Reactive	7	3	5	5	–2	8	2
Proactive	10	–3	10	–5	2	–8	0
Mental notions	7	3	7	3	0	6	0
Team work	5	2	7	0	2	2	–2
Protocols	10	–5	7	0	–5	5	3

Table 12 Values for the concepts and properties dimension to calculate the C and CS coefficients for Tropos

Concepts	$Tropos_{DW}$	d_{DW}	$Tropos_S$	d_S	$d_{DW} - d_S$	$d_{DW} + d_S$	$Tropos_{DW} - Tropos_S$
Autonomy	10	0	10	0	0	0	0
Reactive	7	3	5	5	-2	8	2
Proactive	10	–3	10	–5	2	–8	0
Mental notions	7	3	7	3	0	6	0
Team work	5	2	7	0	2	2	–2
Protocols	7	–2	7	0	-2	–2	0

Table 13 C and CS coefficients for concepts and properties dimension

MaSE	C_{DWT} =	0.73	C_{ST} =	0.71	C_{DWS} =	0.83	CS =	0.73
Tropos	C_{DWT} =	0.71	C_{ST} =	0.71	C_{DWS} =	0.88	CS =	0.76

For a deeper analysis, it would be interesting to calculate the value of congruency and CS for each dimension. In the next figures, the subscripts refer to the evaluators and not to one of the characteristics of the methodology.

From the congruency and CS coefficients for the concepts and properties dimension, it is possible to observe that both the evaluation of Dam and Winikoff (2003), and Sturm et al. (2004) reach the high level. Moreover, for the Tropos methodology, the congruency is very close to the virtual perfection. In effect, in the concepts and properties dimension the coefficients are basically the same (see also the graphics in Figure 2 and Figure 3); this situation indicates that authors agree in that methodologies cover the set of criteria in a similar way. These results are confirmed by observing the graphics in Figure 4 and Figure 5, in which almost all the attributes of the concept and properties dimension obtain the same scores from both evaluators.

Table 14 Values for the modelling and notations dimension to calculate the C and CS coefficients for MaSE

Modelling and notations	$MaSE_{DW}$	D_{DW}	$MaSE_S$	d_S	$d_{DW} - d_S$	$d_{DW} + d_S$	$MaSE_{DW} - MaSE_S$
Consistency check	10	0	8	0	0	0	2
Refinement	8	–2	5	–3	1	–5	3
Expressiveness	8	0	7	2	–2	2	1
Accessibility	5	–3	7	0	–3	–3	–2

Table 15 Values for the modelling and notations dimension to calculate the C and CS coefficients for Tropos

Concepts	$Tropos_{DW}$	d_{DW}	$Tropos_S$	d_S	$d_{DW} - d_S$	$d_{DW} + d_S$	$Tropos_{DW} - Tropos_S$
Consistency check	5	0	3	0	0	0	2
Refinement	8	3	7	4	–1	7	1
Expressiveness	8	0	5	–2	2	–2	3
Accessibility	7	–1	5	0	–1	–1	2

Table 16 C and CS coefficients for modelling and notations dimension

MaSE	C_{DWT} =	0.71	C_{ST} =	0.66	C_{DWS} =	0.79	CS =	0.63
Tropos	C_{DWT} =	0.68	C_{ST} =	0.48	C_{DWS} =	0.79	CS =	0.71

In the modelling and notations dimension, the profile analysis shows a high congruency between the profiles. Nevertheless, the CS coefficient varies from MaSE to Tropos. For the Tropos case, being really on a borderline, CS reaches the high level. For MaSE, the CS is moderate. That means the two profiles have a reduced distance (high congruency) but they are not so similar in their high and low points. This evidence is perceivable by observing the graphics in Figure 2 and Figure 3, and confirmed from the differences in the score of the attributes of this dimension presented in Figure 4 and Figure 5. Thus, in this dimension the evaluators show some differences and consequently the evaluations results may be considered less reliable than in the concept and properties dimension.

Table 17 Values for the process dimension to calculate the C and CS coefficients for MaSE

Process	$MaSE_{DW}$	$MaSE_S$	$MaSE_{DW} - MaSE_S$
Lifecycle coverage	8	7	1

Table 18 Values for the process dimension to calculate the C and CS coefficients for Tropos

Process	$Tropos_{DW}$	$Tropos_S$	$Tropos_{DW} - Tropos_S$
Lifecycle coverage	4	8	–4

Table 19 C coefficients for process dimension

MaSE	C_{DW} =	0.80	CS =	0.70	C_{DWS} =	0.90	CS =	---------
Tropos	C_{DW} =	0.40	CS =	0.80	C_{DWS} =	0.60	CS =	---------

Since, in the process dimension, the Lifecycle coverage is the only criterion that determines the values of the coefficients, it is impossible to calculate the CS (according to formula 4 if we try to calculate it we find a division by zero). In addition, it is possible to observe the main differences in the congruency coefficients. It is worth to observe that for MaSE there is a virtual perfect congruency among the evaluations. Instead, for Tropos the congruency is moderate; that is, the difference makes the evaluations results partially reliable.

Table 20 Values for the pragmatics dimension to calculate the C and CS coefficients for MaSE

Pragmatics	$MaSE_{DW}$	d_{DW}	$MaSE_S$	d_S	$d_{DW} - d_S$	$d_{DW} + d_S$	$MaSE_{DW} - MaSE_S$
Domain specific	10	0	10	0	0	0	0
Scalable	5	–5	3	–7	2	–12	2
Maturity	3	–2	7	4	–6	2	–4

Table 21 Values for the pragmatics dimension to calculate the C and CS coefficients for Tropos

Pragmatics	$Tropos_{DW}$	d_{DWT}	$Tropos_S$	d_{ST}	$d_{DWT} - d_{ST}$	$d_{DWT} + d_{ST}$	$Tropos_{DW} - Tropos_S$
Domain specific	10	0	10	0	0	0	0
Scalable	3	–7	5	–5	–2	–12	–2
Maturity	3	0	5	0	0	0	–2

Table 22 C and CS coefficients for pragmatics dimension

MaSE	C_{DW} =	0.50	CS =	0.56	C_{DWS} =	0.74	CS =	0.64
Tropos	C_{DW} =	0.43	CS =	0.59	C_{DWS} =	0.84	CS =	0.86

For the pragmatics dimension, the profile analysis obtains a high congruency for both methodologies and a high (quite near to the virtual perfect) CS for the case of Tropos. These results increase the reliability of the independent evaluations about the strongest and weakness aspects of Tropos in this dimension. In the case of MaSE, the CS reaches a moderate level, thus, the conclusions have to be considered less reliable than those of Tropos. In effect, in Figure 2, it is possible to observe the differences between the two profiles for MaSE in this dimension (the last three attributes).

More generally, in all the dimensions the congruency coefficient values fall in the high level with the exception of the process dimension (which includes just an attribute) for which MaSE reaches a virtual perfect congruency and Tropos a moderate congruency. Moreover, CS varies from moderate to high reaching in every case more than 0.60. These results allow us to infer that both evaluations agree in the general evaluation of MaSE and Tropos and more specifically in the major part of high and low points of the methodologies, giving more reliability to the evaluations studies.

4.4 Discussion on the application of profile analysis and evaluation of AOSE methodologies

As presented in the Subsection 4.3., the application of profile analysis offers agent researchers and practitioners an interesting tool to improve the comparison among different evaluations. One of the advantages of the method is to allow comparing information from different sources (i.e., the previous evaluations of AOSE methodologies) by contrasting two or more profiles. By applying profile analysis, we look for evaluation similarities in terms of congruency and CS. The coefficient of congruency may be used to quantitatively evaluate the Euclidean distance between profiles. Clearly, a smaller distance between profiles of the same AOSE methodology evaluated by different authors, means that the evaluators assess the methodology in a similar way. Therefore, such similarity may reinforce the acceptability of the evaluation. Complementarily, the CS coefficient measures the grade of correspondence of the high and low values among different profiles in each dimension. Intuitively, different profiles may have great distances among them but still have great similarities in their configuration. This means that, independently from the distance, different evaluators may coincide in assessing the strengths and weaknesses of the methodologies in their coverage of relevant attributes or dimensions.

Using profile analysis it is possible to carry out different comparisons such as focusing on different profiles produced by different evaluators on a particular methodology, or focusing on the evaluations of different methodologies for each evaluator. Finding similarities may be used to increase the reliability of the evaluation on a particular methodology and, complementary, to reinforce the evidence that a methodology covers some specific dimensions or aspects better (or worse) than others. Consequently, profile analysis can improve the reliability of the advantages and shortcomings of all the methodologies under evaluation and thus, make them more acceptable in the community of multi-agent researchers and practitioners.

Still, the comparison among different evaluations of the same methodologies and from different experts have to deal with some difficulties like the adoption of different criteria, the use of different types of evaluations that are not totally congruent among them, and the use of qualitative evaluations which are difficult to compare.

Despite the research effort on the evaluation of methodologies (Shehory and Sturm, 2001; Cernuzzi and Rossi, 2002; Cuesta et al., 2003; Dam, 2003; Dam and Winikoff, 2003; Sturm et al., 2004; Tran and Low, 2005), we think that more effort is still needed in evaluating methodologies, specific methods and techniques in AOSE. In effect, the vast majority of the proposed works are currently focused on qualitative evaluation, while researchers and practitioners actually need for more quantitative results that may facilitate comparative analysis and the selection of specific methods. This is very relevant as, in the process of 'quantifying' qualitative evaluations, it is possible to introduce some approximations and therefore, to influence the final scores. In addition, a possible way to make the expertise accessible is to insist on evaluation frameworks that could highlight the advantages and drawbacks of AOSE methodologies in particular contexts. More specifically, instead of the general attributes used in this study, it may be useful the proposal of a catalogue of attributes refined for specific purposes. It is also possible to propose several ways to evaluate attributes, considering the type of agent-oriented system-to-be, and the opportunity to associate a priority or pondered weight to attributes to better cover significant aspects for the evaluation of methodologies.

5 Conclusions and future works

This study focuses on the evaluation of agent-oriented methodologies. Comparison techniques give the opportunity to improve (in the case of similar results) the acceptance of evaluation of different authors and thus, to improve the reliability of such results. Nevertheless, the existing evaluations do not focus on comparing their results with others ones carried out from different authors.

For carrying out this type of comparison, we proposed the adoption of the profile analysis method, which is a multivariate statistical method for the quantitative evaluation of profiles. Our main objective is to introduce the potential of profile analysis in the evaluation comparison process. In addition, we are also looking for evaluation similarities to confirm results and thus, make them more acceptable in the agent community. The application of such analysis method may help in the selection of a specific methodology with a specific level of congruency with regards to an expected profile, being it general or limited to a particular dimension.

To illustrate the application of the profile analysis method, we compared the evaluation done by different authors (Dam and Winikoff, 2003; Sturm et al., 2004) of MaSE (DeLoach et al., 2001) and Tropos (Giunchiglia et al., 2002). To compare the evaluations, we have selected the common evaluators' criteria, converted each of them to a common scale, and calculated the congruency and CS coefficients. Also, a brief analysis of the comparison results has been presented. Further interesting comparisons may be carried out to test the profile analysis method; e.g., comparing across methodologies.

Finally, it is possible to envision different future works to reach a greater reliability level with regards of the evaluation results. Among them, it is possible to carry out similar studies to compare other independent evaluations as well as to consider further AOSE methodologies. Moreover, it may be useful to include an analysis of the congruency and the CS for the evaluation of different methodologies carried out by the same evaluation group.

References

Avison, D., Lau, F., Myers, M. and Nielsen, P. (1999) 'Action research: making academic research relevant', in Lau, F., Myers, M. and Nielsen, P. (Eds.): *Communications of the ACM*, January, Vol. 42, No. 1, pp.94–97.

Brazier, F., Dunin-Keplicz, B., Jennings, N. and Treur, J. (1997) 'DESIRE: modelling multi-agent systems in a compositional formal framework', in Huhns, M. and Singh, M. (Eds.): *Intl. Journal of Cooperative Information Systems, Formal Methods in Cooperative Information Systems: Multi-Agent Systems*, Vol. 6, pp.67–94.

Burmeister, B. (1996) 'Models and methodology for agent-oriented analysis and design', in Fischer, K. (Ed.): *Working Notes of the KI'96 Workshop on Agent-Oriented Programming and Distributed Systems*, DFKI Document D-96-06.

Caire, G., Chainho, P., Evans, R., Garijo, F., Gómez Sanz, J., Kearney, P., Leal, F., Massonet, P. and Pavón, J. (2001) 'Agent oriented analysis using MESSAGE/UML', *Proceedings of Agent-Oriented Software Engineering – AOSE 01*, May, pp.101–107, Montreal, Canada.

Cernuzzi, L. and Giret, A. (2000) 'Methodological aspects in the design of a multi-agent system', *Proceedings of Workshop on Agent Oriented Information Systems. VII National Conference on Artificial Intelligence (AAAI-2000)*, July, pp.21–28, Austin, USA.

Cernuzzi, L. and Rossi, G. (2002) 'On the evaluation of agent oriented methodologies', *Proceedings of the OOPSLA 02 – Workshop on Agent-Oriented Methodologies*, November, pp.21–30, Seattle, USA.

Ciancarini, P. and Wooldridge, M. (2001) 'Agent-oriented software engineering', *Proceedings of the 1st International Workshop on Agent-Oriented Software Engineering*, Springer Verlag, LNCS, No. 1957, pp.1–24.

Cuesta, P., Gómez, A., González, J.C. and Rodríguez, F.J. (2003) 'A framework for evaluation of agent oriented methodologies', *Actas del Taller de Agentes Inteligentes en el tercer milenio (CAEPIA'2003)*, November, San Sebastián, Spain.

Dam, H.K. (2003) 'Evaluating agent-oriented software engineering methodologies', MSc thesis, School of Computer Science and Information Technology, RMIT University, Australia.

Dam, H.K. and Winikoff, M. (2003) 'Comparing agent-oriented methodologies', *Proceedings of Agent Oriented Information Systems-AOIS'03*, July, pp.78–93, Melbourne, Australia.

DeLoach, S., Wood, M. and Sparkman, C. (2001) 'Multiagent systems engineering', *International Journal of Software Engineering and Knowledge Engineering*, Vol. 11, No. 3, pp.231–258.

Elamy, A.H. and Far, B. (2008) 'On the evaluation of agent-oriented software engineering methodologies: a statistical approach', in Kolp, M., Henderson-Sellers, B., Mouratidis, H., Garcia, A., Ghose, A. and Bresciani, P. (Eds.): *Proceedings of the 8th international Bi Conference on Agent-Oriented information Systems IV*, Lecture Notes In Computer Science, pp.105–122, Hakodate, Japan, Springer-Verlag, Berlin, Heidelberg.

Giunchiglia, F., Mylopoulos, J. and Perini, A. (2002) 'The Tropos software development methodology processes, models and diagrams', *Proceedings of Agent-Oriented Software Engineering (AOSE-2002)*, Springer Verlag, LNCS, No. 2585, pp.63–74.

Glaser, N. (1996) 'Contribution to knowledge modelling in a multi-agent framework (the CoMoMAS approach)', PhD dissertation, L'Universtité Henri Poincaré, Nancy I, France.

Gómez-Sanz, J. and Pavón, J. (2003) 'Agent oriented software engineering with INGENIAS', *Proceedings of the 3rd Central and Eastern Europe Conference on Multiagent Systems*, Springer Verlag, LNCS 2691, pp.394–403.

Iglesias, C.A., Garijo, M., González, J.C. and Velasco, J.R. (1997) 'Analysis and design of multiagent systems using MAS-CommonKADS', in *AAAI'97 Workshop on Agent Theories, Architectures and Languages – ATAL, Providence, RI, (An extended version of this paper has been published in INTELLIGENT AGENTS IV: Agent Theories, Architectures, and Languages, Springer Verlag, 1998).*

Jennings, N., Faratin, P., Norman, T.J., O'Brien, P., Odgers, B. and Alty, J.L. (2000) 'Implementing a business process management system using ADEPT: a real-world case study', *Intl. Journal of Applied AI*, Vol. 14, No. 5, pp.421–465.

Kendall, E.A., Malkoun, M.T. and Jiang, C. (1996) 'A methodology for developing agent based systems for enterprise integration', in Luckose, D. and Zhang, C. (Eds.): *Proceedings of the First Australian Workshop on DAI, Lecture Notes on Artificial Intelligence*, Vol. 1087, Springer-Verlag, Heidelberg, Germany.

Kinny, D., Georgeff, M., and Rao, A. (1996) 'A methodology and modelling technique for systems of BDI agents', in van der Velde, W. and Perram, J. (Eds.): *Agents Breaking Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World MAAMAW'96*, LNAI, Vol. 1038, Springer-Verlag, Heidelberg, Germany.

Morrison, D.F. (1976) *Multivariate Statistical Methods*, McGraw Hill, NY, USA.

Moulin, B. and Brassard, M. (1996) 'A scenario-based design method and an environment for the development of multiagent systems', in Luckose, D. and Zhang, C. (Eds.):*Proceedings of the First Australian Workshop on DAI*, LNAI, Vol. 1087, pp.216–231, Springer-Verlag, Heidelberg, Germany.

Nunnally, J. (1978) *Psychometric Theory*, McGraw Hill, NY, USA.

Padgham, L. and Winikoff, M. (2002) 'Prometheus: a methodology for developing intelligent agents', *Proceedings of the Third International Workshop on Agent-Oriented Software Engineering AOSE-2002 at – AAMAS '02*, Springer-Verlag, LNCS, No. 2585, pp.135–146.

Reis, E. (1997) *Multivariada Aplicada*, Ediçoes Sílabo, Lisboa, Portugal.

Rossi, M. and Siau, K. (1988) 'Evaluation of information modeling methods – a review', in *Proceedings of 31 Annual Hawaii International Conference on System Science*, pp.314–322.

Serafini, O. (1981) *Indicadores Cuantitativos de la Distancia Evaluativa: Coeficientes y Congruencia Simple (C) y Ponderada (Cp)*, Brasilia, Brazil.

Serafini, O. (1988) 'Análisis de perfiles en ciencias de la educación: coeficientes ES1 y ES2 de similaridad configuracional entre perfiles cuantitativos', *Revista Paraguaya de Sociología*, No. 72, pp.193–200.

Shehory, O. and Sturm, A. (2001) 'Evaluation of modeling techniques for agent-based systems', *Proceedings of Autonomous Agents*, pp.624–631, Montreal, Canada.

Sturm, A., Dori, D. and Shehory, O. (2003) 'Single-model method for specifying multi-agent systems', *Proceedings of the Second International Conference on Autonomous Agents and Multi-Agent Systems – AAMAS'03*, July, pp.121–128, Melbourne, Australia.

Sturm, A., Dori, D. and Shehory, O. (2004) 'Comparative evaluation of agent-oriented methodologies', *Methodologies and Software Engineering for Agent Systems*, Kluwer, pp.127–149.

Tran, Q.N. and Low, G. (2005) 'Comparison of ten agent-oriented methodologies', in Henderson-Sellers, B. and Giorgini, P. (Eds.): *Agent-Oriented Methodologies*, Idea Group Publishing, Chapter 12, pp.341–367.

Tveit, A. (2001) 'A survey of agent-oriented software engineering', in *NTNU Computer Science Graduate Student Conference, Norwegian University of Science and Technology*.

Wooldridge, M., Jennings, N.R. and Kinny, D. (2000) 'The Gaia methodology for agent-oriented analysis and design', *Journal of Autonomous Agents and Multi Agent Systems*, Vol. 3, No. 3, pp.285–312.

Zambonelli, F. and Omicini, A. (2004) 'Challenges and research directions in agent-oriented software engineering', *Journal of Autonomous Agents and Multiagent Systems*, Kluwer Academic Publishers, Vol. 9, No. 3, pp.253–283.

Zambonelli, F., Wooldridge, M. and Jennings, N.R. (2003) 'Developing multiagent systems: the Gaia methodology', *ACM Transaction on Software Engineering and Methodology*, Vol. 12, No. 3, pp.417–470.