Assessing a Requirements Evolution Approach: Empirical Studies in the Air Traffic Management Domain

Fabio Massacci, Deepa Nagaraj, Federica Paci, Le Minh Sang Tran
DISI
University of Trento
Povo, Trento, Italy
name.lastname@unitn.it

Alessandra Tedeschi
Deep Blue
Rome, Italy
name.lastname@dblue.it

***Abstract*—Requirements evolution is still a challenging problem in engineering practices. This paper presents a family of empirical studies about the applicability and usefulness of an approach for modeling evolving requirements. The empirical studies involved different categories of users (researchers, master students and domain experts) who have applied the approach to a real industrial evolutionary scenario drawn from the Air Traffic Management (ATM) domain. The results from the studies demonstrated the usefulness of the approach for requirements evolution in complex industrial settings such as the ones in the ATM domain. Furthermore, the validation provided us useful insights about the problem of requirements evolution faced in different industrial contexts.**

***Keywords*-requirements engineering; evolution; change management; user study; air traffic management domain.**

I. Introduction

In this paper we present the results of an empirical validation conducted on an RE approach to model and reason on requirements evolution (previously proposed in [1]). The evolution of mission-critical requirements at enterprise level is known to be possible, but it is unknown whether it would happen: the *known unknowns*. The objective of the approach is to capture what Loucopoulous and Kavakli [2] identified as the knowledge shared by multiple stakeholders about *"where the enterprise is currently"*, *"where the enterprise wishes to be in the future"*, and *"what alternative designs"* are needed for the desired future state. Unfortunately, large organizations cannot wait that the unknowns become known. The process of tendering and organizational restructuring requires a significant amount of time and planning. Decision makers at high-level must essentially bet on the final organizational solution and possibly minimize the risks that the solutions turn out to be wrong. The approach for evolving requirements should help the decision makers to select an optimal system design alternative that is resilient to requirements evolution.

The empirical validation we describe here aimed at assessing the applicability of the approach in an industrial setting.

For our validation we considered the air traffic management (ATM) domain for three main reasons. First, the ATM systems are complex and critical systems that are going through significant architectural, organizational, and operational changes as planned by the EU Single European Sky ATM Research (SESAR) Initiative [3]. Second, change management is a critical issue. The need of system engineering techniques to support change management is well recognized [4]. Last but not least there is a significant body of research about empirical evaluations of requirements engineering approaches in the ATM domain [5], [6], [7]. For example, in [6], DMAN (Departure MANager), a system for managing departure of aircrafts, is used as case study. This makes it easier to benchmark our study.

In our empirical validation, we have focused on changes associated with the introduction of a new decision supporting tool (the AMAN – Arrival MANager) and SWIM (System Wide Information Management) in the ATM domain.

Fig. 1 summarizes how our empirical studies developed along a two years horizon. The aim of the studies is to evaluate how easy to apply and to adopt is the modeling approach when used by subjects that have a different level of knowledge of the method and of the air traffic management domain. First we have conducted a study within the research team. Then, we have pushed the envelope further by carrying out a a study with MSc students which can yield additional benefits (see e.g. [8]), and a number of design workshops with ATM experts and industry practitioners as in [7] to avoid the pitfalls of students only studies [9].

The empirical studies allowed us to collect useful insights about weaknesses and advantages of our approach for modeling requirements evolution. On the positive side, the validation with the ATM experts highlighted that the approach could be a useful instrument in particular during the brainstorming phase to assess the impact of change on critical systems and operational procedures (this confirms the findings of [5]). We have further showed that it is reasonably possible for people different than the method's own inventor (such as students or domain experts) to build significantly large (and correct) models. On the negative side, it emerged during the discussion with ATM experts that efforts in larger or more accurate modelling (as opposed to brainstorming) would only be interesting if they would allow to save time

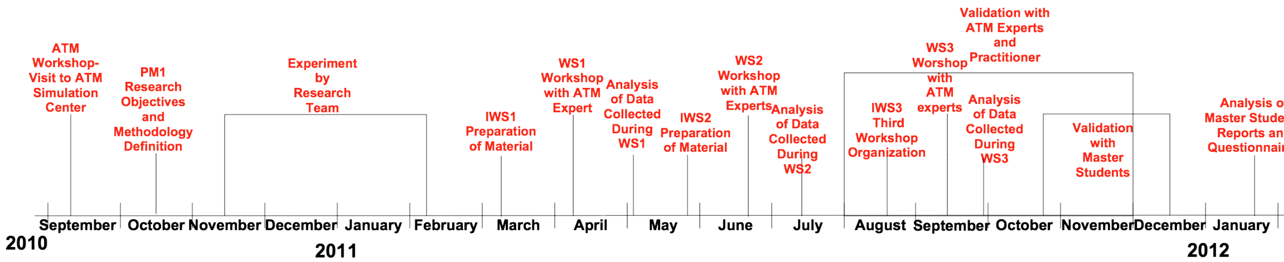

Fig. 1. Chronology of the family of empirical studies

in later phases and in particular costly what-if simulations with controllers. These would be large scale experiments and subject of a research project on its own.

The next section presents the related work. Section III presents the context of our studies. Section IV gives an overview of the RE approach which has been the object of the validation. We provide an overview of the family of empirical studies we have performed in Section V, while Section VI, VIII and VII describe in details the individual studies we conducted to evaluate the approach. Section IX presents the results of the analysis of the data collected during the validation. Section X discusses the threats to validity while Section XI concludes the paper with the lessons learned.

II. Related Work

We review here the proposals related to empirical studies on the visual modeling of requirements and on methods for capturing requirements evolution.

One of the first studies on the topic has been performed by Kamsties et al.[10]. They summarized the results of a case study on requirements interdependencies held with practitioners from ten small and medium enterprises. They showed that new requirements implementation can cause unpredictable interactions with requirements already implemented. Carlshamre et al.[11] surveyed five companies and showed that visualization of interdependencies allows for an efficient identification of key characteristics. With the aim of planning for changes beforehand, Villela et al.[12] proposed a software evolution model to address adaptation needs and potential changes in all levels of software abstraction. In [13], a quasi-experiment was conducted in the field of Ambient Assisted Living to characterize the adequacy and feasibility of this method. Maiden et al. [5], [6], [7] conducted several case studies in ATM domain to validate RESCUE, a scenario-driven requirements engineering process. The case studies included a series of workshops to which ATM experts with different expertise participated. The workshops were organized in three main phases: a *training* phase about RESCUE, a *brainstorming* phase, and then an *application* phase where the experts applied RESCUE to discover requirements for different ATM tools (e.g. DMAN, CORA-2, and MSP). To collect the results from each workshop, color-coded idea cards, post-it notes, A3 papers etc have been used.

Among the above proposals, the works by Maiden et al. [5] are the closest to ours. We have used similar case studies about ATM tools and also the organization of the workshops with ATM experts is mix of training and application sessions as in the works by Maiden et al. We have further based our approach on a broader set of guidelines proposed by Runeson et al. [14] about how to conduct qualitative research. Following those guidelines we have involved researchers, students, domain experts and practioners with different expertise and used semi-structured interviews, questionnaires and audio-video recordings to collect data in order to perform data triangulation.

The guidelines were adopted by McGee et al.[15] to conduct a case study on requirements change taxonomy. The work focused on how change classification helps designers to understand the impact of change, why and when it happens. This study investigated changes recorded during development cycles of an industrial project.

Another study by Herrmann et al.[16] on requirements evolution investigated the applicability of TORE, a requirements engineering approach to identify delta requirements for a plant engineering tool. The study measured improvements in the as-is-analysis, the to-be-analysis, and the prioritization of refinements. Here we focussed on the ability to capture complex evolution scenarios.

III. Application Scenario

The context for the validation of our modeling approach approach has been selected cooperatively by the University of Trento and Deep Blue Srl, an Italian consultancy company specialized in human factors, safety and validation of ATM concepts and systems. The collaboration was established within the SecureChange European project about secure software evolution. A workshop about ATM procedures and tools, and safety and security issues was organized by Deep Blue to introduce the researchers to the ATM domain. It also included a visit to an ATM real time simulation center. During the workshop, it was decided to use as context the introduction of AMAN, a new queue management tool, and SWIM, a new data transport network that will replace the current phone communication lines. These scenarios were selected because of their importance in the SESAR Initiative to which Deep Blue actively participates, and because they involved a number of changes at both organizational and technical level.

Before the introduction of AMAN, the Sector Team (composed by two air traffic controllers) has to manage and generate arrival sequence for arriving flights. Then, the AMAN will automatically generate the arrival sequence to support Sector Team. AMAN may also provide other functionalities, such as ad-hoc simulation, or generating advisories, or metering capabilities for runways. At organizational level, a new type of controller namely, *Sequence Manager* is needed to monitor and modify sequences generated by AMAN, and provide information updates to the Sector Team. At operational level, all ATM actors (including AMAN) will communicate via SWIM which should provide authenticity, integrity and availability guarantees comparable with the one provided by the dedicated communication lines (e.g phone) currently used by controllers.

IV. The RE Approach to Be Validated

This section gives an overview of our approach [1] to deal with requirements evolutions at design phase. The approach helps decision makers to select an optimal design solution to be implemented so that system is evolution-resilient.

Evolutions are classified into two types, namely, *observable* and *controllable* evolution. *Observable evolutions* are changes in a system due to external factors (such as rules, business agreement, new concepts or changes in standards). These changes can be foreseen with some level of certainty based on expert knowledge (for example the discussions and opinions held by different national representatives at EuroControl). An example can be *"Which types of aircraft advisories will AMAN support?"*: some advisories could be more likely to be mandated, others might be less likely to be supported and yet few others might not be supported at all. We capture these changes as observable evolutions, and the levels of certainty of ATM experts are captured by evolution probabilities.

Controllable evolutions correspond to design alternatives that can be chosen by the stakeholders. They originate from different options that can be proposed to fulfil requirements. For example, to *"generate advisories"*, the AMAN can either *"generate advisories based on time of battery"* or *"generate advisories based on speed of battery"*. These options are often proposed from designers who analyze requirements and identify design solutions for the system. Since different options can further lead to different designs for the same system, we consider them as intentional changes under the control of the designer.

These evolutions are incorporated into requirements model using *controllable rule* and *observable rule*, which

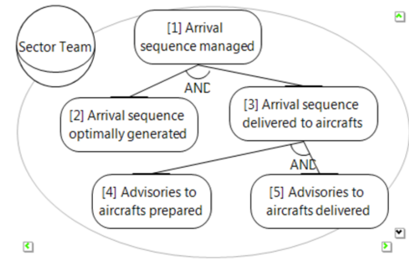

Fig. 2. Goal model of Sector Team.

are defined as follows.

$$\text{\textit{Observable rule:}} \qquad r_o = \bigcup_{i=1..n} \left\{ \textit{Before} \xrightarrow{p_i} \textit{After}_i \right\} \qquad (1)$$

$$\text{\textit{Controllable rule:}} \qquad r_c = \bigcup_{i=1..n} \left\{ \textit{Before} \rightarrow \textit{After}_i \right\} \qquad (2)$$

where: *Before* is the before-change requirements model, and $\textit{After}_i$ are after-change requirements models (a.k.a *evolution possibilities*) which *Before* might possibly evolve to, and p_i is *evolution probability*. The sum of p_i is always equal to one.

Example 1: Fig. 2 shows the goal model of Sector Team related to managing arrival sequence. The top goal is to have *"arrival sequence managed"* (g_1) which is decomposed in *"arrival sequence optimally generated"* (g_2), *"generate advisories"* (g_3). The latter goal is further AND-decomposed in *"advisories to aircrafts prepared"* (g_4) and *"advisories to aircrafts delivered"* (g_5).

According to the changes foreseen by SESAR initiative, the goal model can evolve as follows:

- *Arrival sequence creation*. Initially, the arrival sequence is created manually. In the future, this sequence should be automatically generated by the AMAN. Thus the goal g_2 is delegated to the AMAN. On the left side of Fig. 3 two evolution possibilities (AS1 and AS2) are represented. The probability that AS1 materializes is 60%, the goal g_2 delegated to the AMAN can be fulfilled either by *"simple arrival sequence generator"* (g_6) or by *"advance arrival sequence generator"* (g_7). Meanwhile, the probability that AS2 materializes is 25%. A new goal *"ad-hoc simulation"* (g_8) is introduced. And both g_2 and g_8 can be fulfilled by goal *"advance arrival sequence generator"* (g_7).
- *Aircraft advisories*. The decomposition of goal g_3 can change with 35% probability (AD1) so that the subgoal g_4 can be fulfilled either by the goal *"basic advisory generator"*(g_9), or the goal *"detail advisory generator"* (g_{10}). With a probability of the 40% (AD2), the goal *"detail advisories to aircrafts prepared"* (g_{11}) becomes mandatory and substitutes g_4. The observable rule for goal g_3 is illustrated on the right side of Fig. 3. ■

Since evolutions are known unknowns (we know they are possible but not sure whether they would happen), one design solution may be useful in a before-change situation

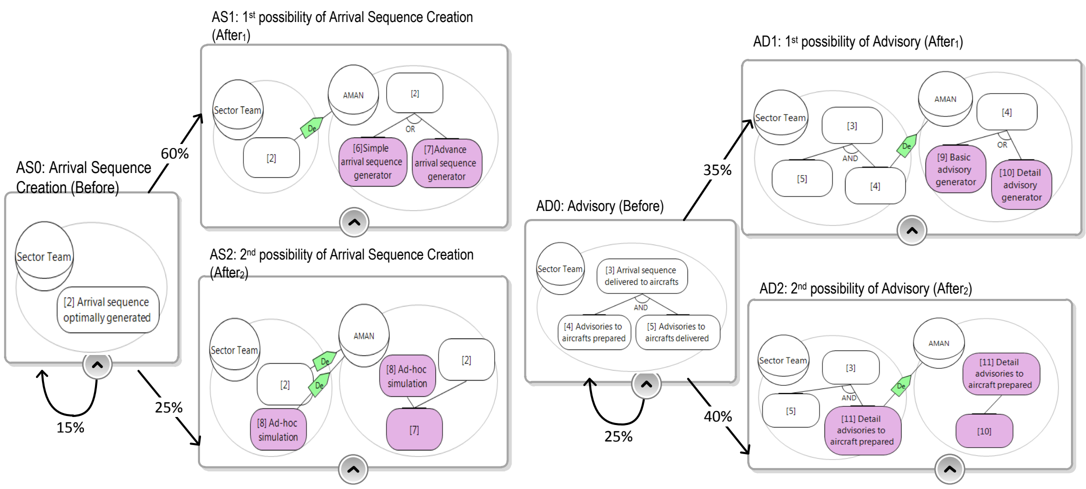

Fig. 3. Two observable rules of the arrival sequence creation (left) and aircraft advisories (right).

but might turn out to be useless in an after-change situation. Hence, in our approach [1], we introduce a reasoning to measure the level of usefulness of different solutions based on two metrics, namely, *MaxBelief* and *ResidualRisk*. *MaxBelief* measures the maximum evolution probability that a solution is still useful after evolution happens. *ResidualRisk* is the probability that a solution becomes useless after evolution. We do not introduce the reasoning here since the studies only focused on the validation of the modelling approach.

Table I
STUDIES SUMMARY

Participants	Method Knowledge	Domain Knowledge	Study	Easy to Apply	Easy to Adopt
Researchers	Good	Partial	Study 1		X
Master Students	Partial	Limited	Study 3	X	
ATM Experts and Professionals	Partial	Good	WS3	X	X

V. Overview of the Empirical Studies

The studies that we have conducted aim at investigating how easily the modeling approach can be applied and adopted by subjects who have different levels of expertise in the approach and in the ATM application scenario (see Table I for an overview). First, the researchers who have proposed the modeling approach (the authors of this paper) have investigated if the approach can be adopted in the air traffic management domain. In order to do that, the researchers have modeled the AMAN case study with possible observable and controllable evolution rules. Then, as in the studies by Maiden et al. three workshops were conducted having as participants air traffic controllers, managers, and system engineers that are the main actors in the work practice associated with the ATM application scenarios. The workshops aimed to test more thoroughly the applicability and potential for adoption of the modeling approach in an industrial setting. In order to collect data on the application of the modelling approach, the workshops have been audio-video recorded, questionnaires have been administered to the participants, and semi-structured interviews were conducted. The workshop activity was followed up by an off-line session which involved two ATM experts and a security engineer from industry.

Then, we have investigated how difficult is to apply the modeling approach for master students who are familiar to the requirement engineering domain and have no previous knowledge of the approach and the ATM domain. Data have been collected through questionnaires and a final report describing the application of the approach on the ATM scenario. After the course, students have been requested to fill in a questionnaire on the method application.

Each study has been organized in two phases: first a *Training* phase where the subjects attend training sessions either on the modeling approach or on the ATM appplication scenario or on both; and an *Application* phase where subjects work in groups and apply the approach on the ATM scenario.

VI. Study 1: Study within The Research Group

The researchers (the authors of this paper) have first gained knowledge about the domain by attending half a day workshop organized by Deep Blue, related to ATM procedures and tools, safety and security issues in the ATM. Deep Blue also provided the research team with documentation about ATM process, AMAN and SWIM architecture and their functional and non functional requirements. After the training on the ATM domain, the researchers have modeled several evolutionary scenarios following the approach to model requirements evolution. The scenarios considered include the introduction of the AMAN; the introduction of the ADS-B, a new surveillance tool used to determine aircrafts' positions, the introduction of the SWIM, and the introduction of the AMAN and SWIM to connect AMAN with queue management tools in other airports. For each of the evolutionary scenarios, the researchers have drawn an original model *Before* and identified an evolution possibility $After_i$. The *Before* model and the *After*$_i$ models have been modeled in the Si* language. Eight SI* models have been drawn by the research team of medium complexity in terms of number of actors, goals, tasks, resources, trust and delegation dependencies between actors.

VII. Study 2: Workshops with ATM Experts

The validation was organized into three separated workshops held in April (WS1), June 2011 (WS2), and September 2011 (WS3). The workshops involved both researchers and ATM experts. The role of researchers was to facilitate the workshop and make observations. The role of ATM experts was to apply the modeling approach and provide feedback about its applicability to model requirements changes.

The objective of the first workshop (WS1) was to present the modeling approach to the ATM experts. Seven ATM experts have participated to WS1: four of them are Deep Blue consultants with various background (e.g., Computer Science, Human Factors, Safety and Security) who have worked in several projects related to the ATM domain. The other three ATM experts have been working for an European Air Navigation Service Provider with different roles and responsibilities: one is a system administrator, while the other two are air traffic controllers. The ATM experts have also extensive experience with the validation of new operational concepts [17] and are currently involved in various SESAR validations. The workshop started with a training session to introduce the experts to the requirement engineering domain and the modeling approach for evolving requirements. Then, the workshop involved role-playing scenario where ATM experts assessed the representation of changes (in terms of goals), the likelihood of particular change scenarios and the representation of such changes. Then, the requirements analysts held a brainstorming session with the participants to identify possible evolution rules.

The focus of second workshop (WS2) was the evaluation of the correctness of the before and after models drawn by the researchers. Most participants were the same from the first workshop: there was an additional participant who works as ATM manager. During the workshop, the requirements analysts have shown the original model and the possibility of evolution model *After*$_i$ they have drawn in Study 1. The models have been discussed and revised with the domain experts. The requirements analysts also conduct a semi-structure interview with the experts to understand what the experts think about the approach. At the end of the workshop, additional feedback was collected by an evaluation questionnaire.

During the third workshop (WS3), ATM experts were asked to apply the modeling approach. We had a total of twelve participants: a security engineer from industry and eleven ATM experts. The ATM experts were the same as the other workshop plus two other Deep Blue consultant who have expertise in Security and Safety for ATM systems. The workshop started with a brief presentation of the scenario to which the experts have to apply the modeling approach to requirements evolution and a summary of the steps they have to follow. The participants were divided in four different heterogeneous groups (in terms of expertise). Each group had to draw an original model and one possibility of evolution model *After*$_i$ using the SI* tool. A total of eight models were produced. The groups annotated on wild cards specific problems they encountered while applying the approach. During the validation session each group was observed by a researcher. At the end of the validation session the research team administered a questionnaire to be filled in by the participants. Then, the research team discussed with the ATM experts the feedback annotated on wild-cards and asked them to provide additional feedback about the modeling approach.

An additional off-line evaluation session was conducted, which involved one of the ATM experts who works as Deep Blue consultant and one security engineer from industry who participated in workshop WS3. The session had a duration of four months from August 2011 to November 2011. The two participants produced eight SI* models representing the ATM system as-is and after the introduction of the AMAN. The models represent the main ATM actors, the resources, the whole ATM system, and the arrival management procedure.

VIII. Study 3: Case Study with Master Students

Eleven MSc students, divided in four groups, participated to the validation. The students had a major in Computer Science and had basic knowledge about requirements engineering. First, students were trained about the approach for evolving requirements during the Security Engineering course and they were also introduced to the ATM application scenarios. As additional material, they received three documents describing AMAN and SWIM users requirements, SWIM content and information services, and AMAN and SWIM core architecture.

Each group chose a possible scenario associated with the introduction of AMAN and SWIM network, and had to apply the approach for evolving requirements. After examining the scenarios, they identified controllable and observable evolutions, and constructed $Before$ goal model as well as several $After_i$ models representing evolution possibilities.

At the end of the validation the students were asked to deliver a report describing in details the application of the approach and the generated models. They also had to fill out a questionnaire about the approach.

IX. Qualitative and Quantitative Data Analysis

This section reports the results of the analysis we have conducted on the data collected during the studies. First, we assess if the modeling can be adopted to capturing complex requirements changes that characterize the ATM domain. Then, we investigate how easy was the application of the modeling approach depending on the participants's background.

A. Qualitative Analysis

During the structured interviews conducted during the workshops, the ATM experts were asked whether the approach can be adopted to represent the evolution of ATM systems.

They all pointed out that it is not possible to predict all the possible changes in advance especially for complex systems such as ATM systems. The ATM Manager said "*Sometimes, when you apply you discover a third change that is better than the one you have predicted*", and that "*The model may be good but when you switch form theory to practice you realize that there are many situations that you did not consider*". The Senior Deep Blue consultant also remarked that "*We are talking about very complex systems. You don't know from the beginning all the actors involved in the process. There are always certain changes that you cannot predict due to the complexity of the system.*". The ATM experts went further and suggested that an incremental approach should be applied to identify all the possible evolution alternatives: "*It should be an iterative process* ", and that "*you need to have more iterations if you want to reach 100%.You cannot foreseen everything at the beginning*".

The ATM experts were also asked whether the graphical representation can be adopted to capture complex evolutions. The ATM Manager said that the representation "*No - it is not so immediate*". The Senior Deep Blue consultant was more confident but again he remarked for the need to adopt an incremental approach to elicit all the possible changes: "*yes, but you need more than one iteration*". Another Deep Blue consultant reported that it is necessary to simplify the graphical representation because: "*the representation becomes very complex expecially when SI* is used to model before and after models*".

B. Quantitative Analysis

We wanted to assess how difficult is to apply the modeling approach depending on the subject's level of knowledge in the method and in the ATM domain. As evaluation criteria, we have used the complexity of the before and after models drawn by the researchers, the master students and the group composed by the Deep Blue consultant and the security engineer. We did not include in the comparison, the models drawn by the ATM experts because they were produced in two hours of work during the third workshop while the models drawn by the other experimental subjects have been produced over a longer period of time (months) and their correctness have been assessed by Deep Blue consultants.

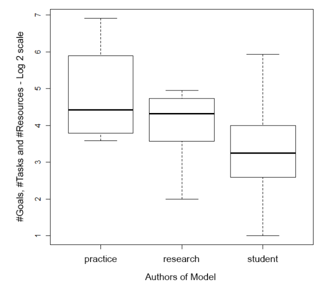

Fig. 4. Distribution of Number of Model Elements by Different Authors

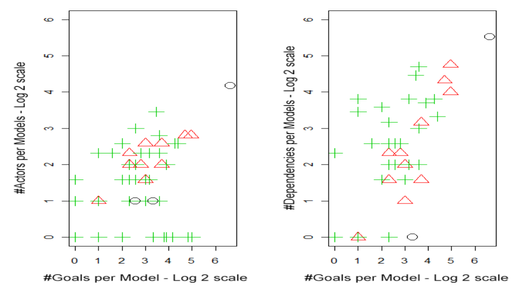

Fig. 5. Scatter-Plot of Measures of Model Complexity

Fig. 4 and Fig. 5 represent the complexity of the models (number of goals, actors and dependencies) drawn by the practioners, the researchers and the master students (practitioners by ovals, researchers are denoted by triangles, and students by crosses). Both figures hightlights that the team composed by the Deep Blue consultant and the security engineer has produced complex models that represent the main ATM systems and actors related to the arrival procedures. The researchers have drawn models of medium complexity while the master students have produced very simple models with a limited number of goals,tasks and resources. However students have identified more evolution possibilities than the team of practitioners and the researchers who just identified one evolution possibility. The plot in Fig. 6 shows that on average observable rules identified by the groups of MSc students have four alternative evolution possibilities, while controllable rules have around two possibilities.

The comparison of the complexity of the models hightlights that the modeling approach can easily be applied by subjects that are novice either in the method (ATM experts and practitioners), in the application domain (researchers), or both (master students). From the complexity of the models, it seems that in order to draw models that capture all the aspects of the case study a partial knowledge of the requirement engineering domain and a good knowledge of the case study are required.

X. Threats To Validity

- **Construct Validity**. A main threat to construct validity in our study was represented by a communication gap

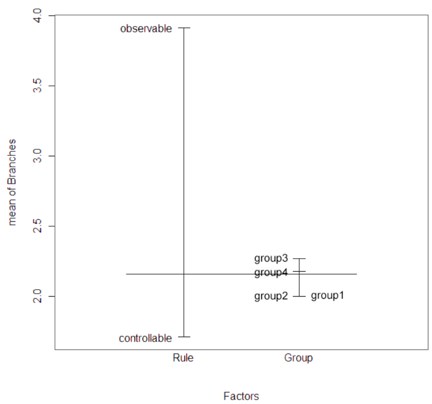

Fig. 6. Plot-design Mean for Different Factors

between the research team and the domain experts. Research team and domain experts might use same terms with different meanings and this can lead to misunderstandings and to provide wrong or unrelated feedback. For example, the distinction between *goal* and *resource* was difficult to understand for the experts because in ATM domain, a goal is something that is produced while in the requirement domain is something that an actors wants to be achieved. To mitigate this threat we have included a "*mediator*" who occasionally reformulated questions of the research team for the domain experts and reformulated domain experts' feedbacks for the researchers. The mediator role in this experiment was played by a member of Deep Blue who was familiar with our approach.

- **Internal Validity**. The feedbacks provided by the ATM experts on the possible adoption of the graphical representation to model evolution of requirements in the ATM domain can be biased by the fact that the before and after model were drawn in the SI* requirements language. SI* graphical notation tends to get very complex even for simple models and this aspects may have influenced the feedbacks of the ATM experts. We should organize another study using a different requirements language to evaluate whether the feedbacks depend on the use of SI*.
- **External Validity**. Since we have evaluated the applicability of the approach with different kind of subjects - researchers, practitioners, domain experts, and students - our results can have general validity. However, to show that our findings are true for other cases we should test the applicability of the approach in other industrial contexts.
- **Conclusion Validity**. An important threat to the conclusion validity of our studies is the relatively small number of subjects that participated to the studies. The number of participants was small expecially for the workshops with ATM experts because of constraints on costs and time availability. We should organize other studies to have a bigger data sample to draw our conclusions.

XI. Conclusions and Lessons Learnt

We have described the case studies that we have conducted in the ATM domain to evaluate the applicability and the potential for adoption of an approach for modeling evolving requirements proposed by us.

The validation has highlighted that the modeling supported by the approach can be a useful decision support tool for decision makers during brainstorming and change assessment. Moreover, the ATM experts reported a problem related to the graphical representation of evolution rules which tend to be complex even for simple evolution scenarios. The effectiveness of visual notations for requirements is a problem that has received little attention in the requirements engineering community. A relevant work on the subject is by Moody et al. [18] who evaluated the cognitive effectiveness of i* visual notation. However, the evaluation done in the paper is not based on case studies research like we did in this paper. Effectiveness of a method can only be evaluated by conducting case studies research with users that are novice to the approch.

The validation also highlighted a number of aspects that should be taken into account to conduct case study research.

The selection of domain experts strongly influence the relevance of feedback collected and the satisfaction of the success criteria chosen for the case studies. In the case studies, the domain experts selected had a different background and so we were able to collect feedback about the approach to requirements evolution from different perspectives.

However, an issue of the domain is the separation between ATM organizations and IT suppliers. They have different and often competing stakes. In future validation studies, we think that one should validate the approach separately with two groups of ATM organizations and IT supplier and identify methods to *firewall feedback* by different groups. This might highlight competitive advantages that one group might gain over the other by adopting the method.

Another interesting lesson concerns the choice of the communication medium. The level of engagement of the domain experts depends on two main factors: the means to provide feedback, and the language in which such feedback needs to be provided. Our workshop sessions included Hungarians, Indians, Italians, Norwegians, and Vietnameses; juggling between languages made our meetings lively. Albeit obvious in hindsight, the *foreign language gap* was not mentioned in the previous work by N. Maiden and others [5], [6], [7] because their studies were clearly English-to-English. A possible solution is that the domain experts can discuss in their mother tongue language and then provide summary feedback in English, but this hampers the immediacy of the feedback, and "minority opinions" might not be reported (we noticed this phenomenon during the workshops). The mediator was a useful tool to mitigate the internal validity threats also in this setting.

However, a major factor in the level of engagement of domain experts is the *perceived compliance* with the practice in industry where requirements management tools such as IBM Rational DOORS are adopted. A show-stopper in the discussion with a practitioner was simply "We use DOORS" (and therefore cannot use and should not waste time evaluating requirements models in format different than DOORS). This was purely a syntactical limitation, not a semantical or methodological one: we could have perfectly used DOORS to link requirements expressed by goal models, but our tool simply did not do it, as we thought this was just "Engineering". This is indeed true if we considered limiting our validation to an experiment (as noted in [19] this is what the vast majority of RE papers report). Being able to syntactically interface with these tools (even for just gathering requirements IDs to label goals), is essential to obtain better perceived compliance and thus a better engagement and case-study based validation. Since a full fledged integration does not make sense for research purposes, in our future work we will investigate how to increase the perceived compliance with a light integration of our methodology as a reasoning plug-in into industry-level requirements management tools.

ACKNOWLEDGEMENT

This work is partly supported by EU-FP7-IST-NoE-NESSOS, EU-EIT-ICTLabs, EU-FP7-SEC-SECONOMICS, PAT-TRISE.

REFERENCES

[1] L. M. S. Tran and F. Massacci, "Dealing with known unknowns: Towards a game-theoretic foundation for software requirement evolution," in *Proc. of CAiSE'11*, 2011, pp. 62–76.

[2] P. Loucopoulos and E. V. Kavakli, "Enterprise knowledge management and conceptual modelling," in *Proc. of ER'99*. Springer Verlag, 1999, pp. 123–143.

[3] "EUROCONTROL, ATM Strategy for the Years 2000+," 2003.

[4] H. K. Rober Graham, Nadine Pilon and P. Ravenhill, "Performance framework and influence model in atm," in *Digital Avionics Systems Conference(DASC)*, 2009.

[5] N. Maiden and S. Robertson, "Integrating creativity into requirements processes: Experiences with an air traffic management system," in *Proc. of RE'05*. IEEE Press, 2005, pp. 105–116.

[6] N. Maiden, S. Jones, S. Manning, J. Greenwood, and L. Renou, "Model-driven requirements engineering: Synchronising models in an air traffic management case study," in *Proc. of CAiSE'04*, ser. LNCS. Springer Verlag, 2004, vol. 3084, pp. 3–21.

[7] C. Ncube, J. Lockerbie, and N. Maiden, "Automatically generating requirements from <i>i*</i> models: Experiences with a complex airport operations system," in *Proc. of REFSQ'07*, ser. LNCS. Springer Verlag, 2007, vol. 4542, pp. 33–47.

[8] J. Carver, L. Jaccheri, S. Morasca, and F. Shull, "Issues in using students in empirical studies in software engineering education," in *Proc. of METRICS'03*. IEEE Press, 2003, pp. –.

[9] M. Host, B. Regnell, and C. Wohlin, "Using students as subjects. a comparative study of students and professionals in lead-time impact assessment," *Empirical Softw. Engg.*, vol. 5, no. 3, pp. 201–214, Nov. 2000.

[10] E. Kamsties, K. Hörnmann, and M. Schlich, "Requirements engineering in small and medium enterprises," in *Proc. Conf. on European Industrial Requirements Engineering*, 1998.

[11] P. Carlshamre, K. Sandahl, M. Lindvall, B. Regnell, and J. N. och Dag, "An industrial survey of requirements interdependencies in software product release planning," 2001.

[12] K. Villela, J. Doerr, and A. Gross, "Proactively managing the evolution of embedded system requirements," in *Proceedings of the 2008 16th IEEE International Requirements Engineering Conference*, ser. RE '08. IEEE Computer Society, 2008, pp. 13–22.

[13] K. Villela, J. Drr, and I. John, "Evaluation of a method for proactively managing the evolving scope of a software product line." in *REFSQ*, ser. Lecture Notes in Computer Science, R. Wieringa and A. Persson, Eds., vol. 6182. Springer, 2010, pp. 113–127.

[14] P. Runeson and M. Host, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Softw. Engg.*, vol. 14, no. 2, pp. 131–164, Apr. 2009.

[15] S. McGee and D. Greer, "Software requirements change taxonomy: Evaluation by case study," in *Requirements Engineering Conference (RE), 2011 19th IEEE International*, 29 2011-sept. 2 2011, pp. 25 –34.

[16] A. Herrmann, A. Wallnöfer, and B. Paech, "Specifying changes only — a case study on delta requirements," in *Proc. of REFSQ '09*. Springer-Verlag, pp. 45–58.

[17] "EUROCONTROL, European Operational Concept Validation Methodology, E-OCVM Version 3.0," 2010.

[18] D. Moody, P. Heymans, and R. Matulevicius, "Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation," *Requirements Engineering*, vol. 15, pp. 141–175, 2010.

[19] N. Condori-Fernández, M. Daneva, K. Sikkel, R. Wieringa, O. Dieste, and O. Pastor, "Research findings on empirical evaluation of requirements specifications approaches," in *WER*, 2009.