Software Evolution Strategy Evaluation: Industrial Case Study Applying Value-Based Decision Model

Tero Tilus[1], Jussi Koskinen[1], Jarmo J. Ahonen[2],
Heikki Lintinen, Henna Sivula[1], Irja Kankaanpää[1]
[1] Information Technology Research Institute, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
`firstname.surname@titu.jyu.fi`
[2] Department of Computer Science, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland
`jarmo.ahonen@uku.fi`

Abstract

This study aims at supplementing the empirical basis of using software evolution evaluation methods in industrial settings. An industrial case study applying Value-Based Decision Model (VDM) for software evolution strategy evaluations is presented. VDM considers the value of a legacy software system and proposes convenient evolution strategies. Use of VDM is characterized, and the case study described. VDM has been applied in case of a production-use, large-scale industrial legacy system. Results produced by VDM, VDM's evaluation, and system evolution strategy recommendations are presented.

1. Introduction

Software maintenance and *evolution* are economically important issues [15]. The proportion of maintenance costs of the total lifecycle costs has traditionally been 50-75% in case of successful legacy systems with long lifetime [11]. *Legacy system* is an old, typically poorly documented software system, which has been implemented with old technology. It is at least partly outdated, but may contain invaluable business logic and information. According to *Lehman's first law* [10] a software system used in "real-world" environment must evolve or it will become progressively less satisfactory to its users. This is due to the typical changes of user requirements and characteristics of the technical environment. Therefore, during the maintenance phase of the legacy system, decisions have to be made concerning different system evolution strategies.

Evolution strategy selections often have large-scale, long-range economical effects including changes in system maintenance costs and in possibilities to preserve the initial legacy system investment. Viable strategic options might include system modernization or system replacement. It is in principle a desirable goal to preserve or reuse the knowledge that a successful legacy system contains. However, replacements are large investments and large-scale modernizations are often both technically demanding and economically relatively critical tasks. Earlier studies conducted within our research project include theoretical analysis and comparison of many of the important methods for software modernization estimation [8], including [3,5,15-17,20-21], empirical studies, including [9], and method development. Especially, the earlier analysis [8] reveals that more empirical studies should be conducted concerning the use and validation of the methods in this area. There is a general need to study and deal with industrial large-scale legacy systems which have been considered important or problematic to organizations. We have focused in one of our case studies to one such system. VDM (Value-Based Decision Model) [20] has been considered to be the best suited support method for the case. Therefore, it has been applied and feedback gathered.

The paper is organized as follows. Section 2 characterizes the main principles of VDM. Section 3 describes the structure and phases of the conducted case study applying VDM in industrial settings. Section 4 presents the results received. These include results produced by VDM, VDM's evaluation and suggestions of viable system evolution strategies. Finally, Section 5 summarizes the conclusions.

2. VDM: Value-Based Decision Model

VDM [20] is a model developed by Visaggio. It supports the selection of the most suitable software evolution strategy to be applied. Likewise as [1,13,17-18] it considers both economical and technical issues. It is one of the relatively well validated methods. VDM is based on a real-world renewal project on a banking system consisting of 653 programs and 1.5 million instructions.

First step in applying VDM is to divide the system into *parts*. The division does not necessarily correspond to system components. Second step is to determine and weight metrics and to set baselines (objective-levels). Both technical and economical metrics are used. Visaggio [20] has used anomalous files, dead data, semantic redundant data etc. The metrics used within individual evaluations may vary based on their availability and suitability for the evaluation purpose. Weights are set by the model users. Baseline represents the metric threshold value which is considered to be satisfactory.

The model produces two output variables: economical value (*economic score*) and technical value (*technical score*) of the evaluated system. The received value pair is presented as a point in two-dimensional (technical, economical) space. The relative potential of the strategies can then be deduced based on the relative locations of the received points. This enables evaluation of individual changes and priorization of system replacement, modernization and other available options. Main principles related to interpretation of the results produced by VDM are: 1) All maintenance should strive towards high economical and technical value. 2) Modernization is suitable in case of high economical and low technical value. 3) Replacement is suitable in case of low economical and technical value.

3. Case study

The general purpose of the case study was to systematically evaluate one industrial legacy system regarding its possible evolution strategies based on VDM. There were three involved organizations: University of Jyväskylä; the project group was mainly implementing the study, a large Finnish software house having supplied the target system (*supplier*), and a large public-sector organization using the target system (*customer*). About 100 work days of the project group resources were used to the planning, implementation and reporting of the case study (5/2003-3/2005).

There were six project group members and nine external experts involved to the case study. Supplier experts were: technology manager, chief of department, two communication chiefs, and two technical experts. Customer experts were two system managers. These persons participated to the case study in various workshops regarding their own areas of expertise.

3.1. Initiation

Discussions between the involved parties concerning the possible goals and target systems of evaluations (5-7/2003) were followed by the selection of the target system by the customer and the project group. Then, the actual evaluation project was initiated (9/2003). VDM [20], SABA [3] and Renaissance [21] were considered as possible evaluation methods. VDM was presented to the involved experts in two occasions (9/2003, 2/2004) and via a literary summary of its contents (11/2003). Objectives of the evaluations were determined (10/2003). VDM was selected by the involved customer experts as the most suitable evaluation method for the organizational needs.

3.2. Target system characterization

Target system is a large industrial legacy system which is in constant production use. The system has been considered as a relevant target for the evaluations by the involved experts. It is a typical legacy system: implemented in relatively old technology but valuable or invaluable to the continued important business operations. The system is vital due to the large amount of specialized knowledge it contains. It also has a core-role in user organization’s business. The application area is tax payment monitoring. Payment information is received and calculations regarding the accuracy of the amounts and timings of the payments are performed at monthly and annual basis. The system is implemented mainly in Oracle Forms, COBOL and SQL. The size of the evaluated parts was about 67 KSLOC, which represented most of the system.

3.3. System division

System division is not supported by any of the methods cited earlier. Therefore, the division has been a subproblem to be solved. Our project has developed a method called MODEST [19] which is aimed at early system modernization pressure estimation. It also provides one solution to the division problem and has been applied here in that regard. The guidelines it follows regarding the division can be summarized into one central principle: *sufficient homogeneity of the values of each applied metric as such within the evaluated part*. If this criterion is not met, then it should be considered whether to divide the part into smaller sub-parts. Practical limitations of the resources allocated to the evaluations affect the possibilities to apply fine-grained system divisions. Division was selected in workshop meetings (3-7/2004). There were two selected main parts or segments (later denoted by *A* and *B*) divided to three subparts each (*A1*…*A3* and *B1*…*B3*).

3.4. Calculations

Calculations were conducted at general level in three different ways. The first variant is based on VDM [20]. The other two variants are mainly based on [17], and partly on [21], and will be used in evaluating the reliability and sensitivity of the main results produced by VDM.

1. VDM - *Baselined evaluation* (B). Technical and economical metrics were scaled based on per-metric baselines (i.e. preset objective-levels). Strength of this variant is that it covers organization's quality policy. Weakness is that the evaluations are relatively sensitive to the expert judgments.
2. *Relative evaluation* (R). Technical metrics were scaled per size of the units (intuitively: “size-relative density”).
3. *Simplified evaluation* (S). Technical metrics were weighted based on part sizes only (intuitively: “mass”).

Additionally metrics were weighted according to their importance reflecting organization’s quality policy. Weights were determined and preset by the involved experts based on their judgment during the measurements phase. Relative and simplified variants do not include quality objective considerations. Details of the calculations are presented in Appendix 2.

3.5. Selection and definition of the applied metrics

GQM [2] was used in the metrics selection process. Some adaptations and specifications were made to the metrics to improve their practicality. Feasibility studies were conducted to ensure the availability and relevance of the metrics. Based on these subphases, the applied metrics were finally selected.

3.5.1. GQM

The required selection of the applied metrics and their weights was achieved by applying Goal-Question-Metric (GQM) approach as presented in [2]. GQM has been applied also, e.g. in [1]. First, a goal(s) is set. Secondly, the goal is operationalized into a set of questions. Thirdly, questions are quantified in terms of metrics. The main question was: "What is the economic score and quality score of the target system?". Quantification into metrics was achieved during workshop meetings. These meetings were participated by experts from the involved organizations.

3.5.2. Adaptations and definitions

Some of the applied metrics (*portfolio expert judgments* and *data volume*) were such that they were not directly suitable to be evaluated based on the calculation suggested by Visaggio. Therefore, these were evaluated according to Sneed instead. Economical value did not include operation and maintenance costs. This needs to be taken into account while making conclusions. Metrics represented in the literature were somewhat adapted to increase their applicability, validity, and intuitive consistency in VDM evaluations.

3.5.3. Feasibility studies

Relevant earlier maintenance and evolution metric surveyes, classifications, and implementations include [4-7,12,16]. The feasibility of the following technical metrics was initially carefully considered during the workshop meetings: *correctness, usability, maintainability, change impact area size, use of constants in code, used technology, direct SQL retrieves, operating system calls, source code comprehensibility, data base characteristics, redundandy, corrections/changes, change impacts*, and *number of relevant received email-messages from system users during 2003*. Similarly, many economical metrics candidates were initially considered.

3.5.4. Final selection of the applied metrics

Criteria for selecting the metrics to be applied were determined (3/2004): sufficiently small additional work to the organization responsible of the necessary data gathering, coverage, comparability, and repeatability of the measurements. Metrics included expert judgments and objective metrics. Expert judgments were supported by additional specifying instructions. Metrics were selected and final definitions specified based on the workshop results (3-

10/2004). The selected metrics (Appendix 1) are used in evaluation of the system parts.

3.6. Measurements

Measurements were planned and performed (7-12/2004) in cooperation with the project group and experts from the supplier and customer. Feasibility of tool support for automatically determining the needed metrics values was first considered. The very versatile tool Softcalc/COBANAL developed by Sneed [16] for COBOL code analysis was considered and tested. However, due to the used not-supported dialect of COBOL in the system it could not be applied sufficiently conveniently. Instead, objective technical metrics were gathered by a technical manager and supporting personnel from the supplier and objective economical metrics by the two system managers from the customer. *Technical expert judgment* was made by the developer team of the supplier. *Comprehensibility* and *adaptability* were evaluated also by a COBOL user support group of the supplier. *Economical expert judgment* was made by two system managers from the customer.

3.7. Gathered metrics

Table 1 gathers the selected metrics, received actual and derived metrics values for the evaluated six system parts, baselines, and set weights of the metrics. "ExpertJudgment-X-Y" refers to the value of the judgment of expert X in absolute or relative terms (Y). Absolute values (A) were asked, relative values (R) were derived. Best values are written in boldface. Noteworthy are the good technical values of parts *A1* and *B1* and good economical values of parts *A1* and *A3*. As noted earlier baselines and the weights were set by the experts.

Table 1. Summary of the gathered metrics.

Metric/Part	*A1*	*A2*	*A3*	*B1*	*B2*	*B3*	Baseline	Weight
Size (KSLOC)	17.8	7.2	5.8	8.3	13.0	15.8	-	-
Technical value								
ExpertJudgment-1-A	70	78	48	**97**	65	42	50	1.5
ExpertJudgment-1-R	**113**	51	25	74	77	60	50	1.5
ExpertJudgment-2-A	48	30	32	**60**	14	16	25	1.5
ExpertJudgment-2-R	**78**	20	17	46	17	23	25	1.5
Constants/KSLOC	1.7	13.2	0.7	5.4	3.1	**0.1**	0.2	1.0
DB-IB/KSLOC	**0.0**	1.0	2.7	**0.0**	2.6	3.0	0.1	1.0
OS-calls/KSLOC	5.6	1.1	1.5	**0.0**	0.1	0.1	0.1	1.0
Economical value								
ExpertJudgment-3-A	**70**	30	30	30	20	20	25	1.5
ExpertJudgment-3-Rel.	40	42	**53**	37	16	13	25	1.5
ExpertJudgment-4-Abs.	**60**	26	30	34	26	26	25	1.5
ExpertJudgment-4-Rel.	34	36	**52**	41	20	16	25	1.5
Automation (%)	90	66	**100**	86	92	87	97	2.0
Input vol (rec./MSLOC)	579	**915**	5	30	345	132	100	1.0

4. Results

Results regarding VDM included those produced by VDM itself (4.1), VDM's evaluation (4.2), and software evolution strategy recommendations for the target system (4.3). VDM produced technical and economical scores of the evaluated system parts. Evaluation was based on sensitivity analysis and gathered feedback. Final evolution strategy recommendations were achieved based on these analyses.

4.1. Results produced by VDM

As noted, VDM shows the results using the four-field-square notation. The four fields are determined based on the thresholds. Figure 1 shows the results received based on the baselined (B) evaluation (Appendix A2.1). This evaluation helps to identify the most valuable system parts. The figure also shows the recommended system evolution strategies for the four fields.

The results suggest that default evolution strategies of the parts would be the following. Status quo, *i.e.* continued conventional maintenance for *A1* and *B1*, modernization for *A2* and *A3*, and replacement for *B2* and *B3* (however, the replacement suggestion especially for *B2* is not particularly strong). The economical value of *A2* is here increased by *A2*'s very high level of automated operation. Similarly, the economical value of *A3* is increased by its small size and high input information volume. As noted earlier, the baselined variant is most sensitive to the expert judgments. In order to study the sensitivity of these results more closely, also relative (R) and simplified (S) evaluation values were mapped.

4.2. VDM's evaluation

The results produced by VDM (baselined evaluation) were compared to the results produced by relative and simplified evaluations. Table 2 characterizes the principles of using input metric values in these three variants of calculations. Only exception was that automation level is relative for all variants. Table 3 shows the final values for the six system parts calculated based on these variants. These values were received by applying the formulas presented in Appendix 2 based on the input values of Table 1 according to the main principles of Table 2. Threshold was preset by the involved experts and was used only in the baselined evaluation. It refers to a quality threshold separating good and poor systems.

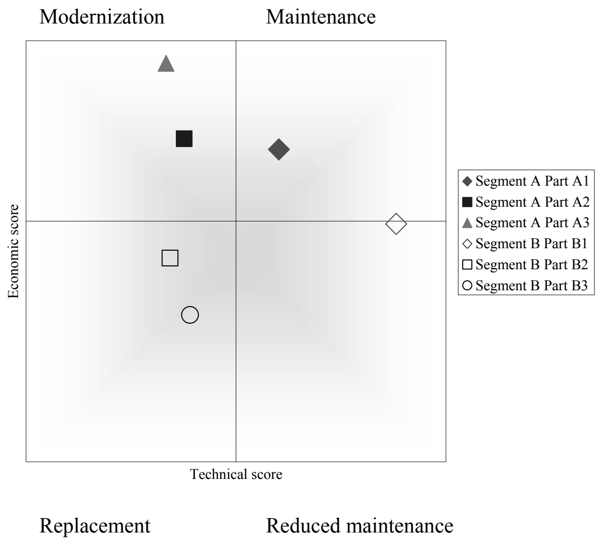


Figure 1. Results of VDM: baselined evaluation (B).

Table 2. Characterization of the principles of using input metrics in the three variants of calculations.

Metrics/Variant	Objective/ Technical	Objective/ Economical	Expert/ Technical	Expert/ Economical
Baselined (B)	Relative	Relative	Absolute	Relative
Relative (R)	Relative	Absolute	Absolute	Absolute
Simplified (S)	Absolute	Absolute	Relative	Absolute

Table 3. Summary of the technical and economical value calculation results: VDM - baselined variant (B), relative variant (R), simplified variant (S).

Value/Part	*A1*	*A2*	*A3*	*B1*	*B2*	*B3*	Threshold
Technical-B	7.2	4.5	4.0	**10.6**	4.1	4.7	6
Technical-R	4.2	3.4	3.3	**5.6**	3.2	3.0	-
Technical-S	**4.7**	2.8	3.2	4.4	3.2	3.2	-
Economical-B	7.8	8.1	**9.9**	5.9	5.1	3.7	6
Economical-R	**5.8**	3.3	3.4	3.3	3.4	3.0	-
Economical-S	**5.8**	3.3	3.4	3.3	3.4	3.0	-

Figure 2 shows the results received based on relative (R) evaluation. Sneed [17] does not use splitting of the field into squares, since there are no thresholds used. Interpretation is almost similar as in case of the previous figure. However, scores of the parts are purely relative to each others, and are here also interpreted as such. Differences of the evaluation results to the baselined evaluation are as follows. The economic value of *A2* and *A3* is relatively much smaller. Therefore, both modernization and replacement seem to be possible options for these parts.

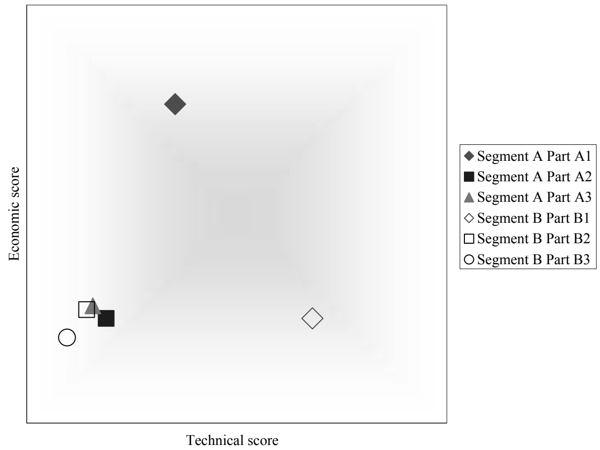


Figure 2. Results: relative evaluation (R).

Also, the economic value of *B2* and *B3* is somewhat smaller. Therefore, it appears that the replacement is the correct strategy afterall for these parts. Technical value of *A1* is somewhat smaller, suggesting that also modernization could be a viable evolution strategy for it.

Likewise, Figure 3 shows the results received based on simplified (S) evaluation. Differences to the relative evaluation are small, except that the technical value of parts *A2* and *B1* is smaller. This evaluation strongly suggests that some radical changes (modernization or replacement) are needed in case of part *A2*.

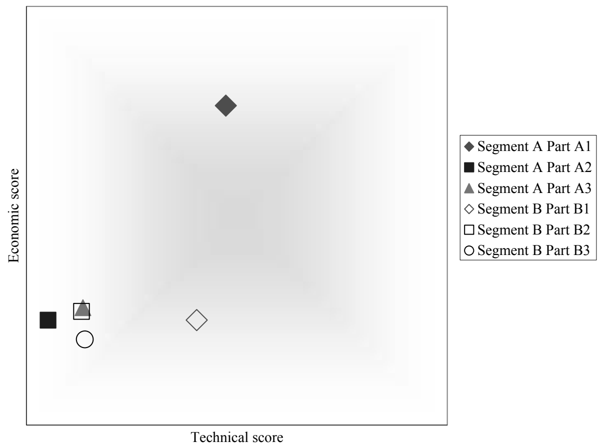
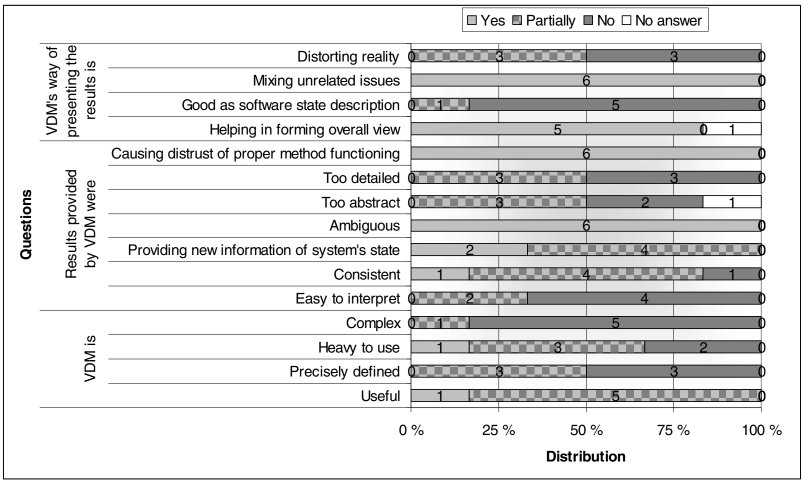


Figure 3. Results: simplified evaluation (S).

Additionally, feedback concerning the case study and VDM was gathered using a questionnaire. Experts from the industrial organization answered to the questions. The case study was considered successful: goals were well achieved and communication was fluent. Six persons out of the nine involved in VDM's application answered to the questions. Answers followed a three-option scale: agreement, partial agreement and disagreement. The questions and the distribution of the results are presented in Table 4. It can be concluded that VDM was generally considered useful, sufficiently precisely defined, and somewhat heavy to use. The results provided by it were easy to interpret, sufficiently consistent, somewhat ambiguous, possibly somewhat too abstract, definitively not too detailed, and highly reliable.

Table 4. Feedback concerning VDM.



The way that VDM presents its results (the four-field-square) was helping very well in forming overall view of the evaluated situation, sufficiently well in describing the state of the system and definitively not mixing or distorting issues. State of the system and overall view refer to the further maintenance and evolution of the system. Application experiences suggest that VDM evaluations could benefit from more specific instructions included into the method. These should specify how to consistently utilize the classifications and how the measurements could be handled across different evaluations in practice.

4.3. Evolution strategy recommendations

Table 5 summarizes the evolution strategy suggestions. At the most general level the analysis suggests that segments *A* and *B* are different. *A* should be modernized. *B* could be replaced. For *B2* and *B3*, replacement is a clear strategy. It should be started from *B3*. However, since replace is the most radical strategy, it is suggested that this analysis is first supplemented by detailed cost-benefit analysis. The now not included maintenance and operation costs should be taken into account while making the final decision. For *B1* it is not totally clear what the optimal strategy is. However, since conventional maintenance gives time to further considerations, it appears to be in that sense the currently right choice. For *A1*, *A2*, and *A3* modernization is the most consistent strategy. It should be started from *A3* and then, in case of positive results, continued to *A2* and *A1*. Especially for *A1* also continued maintenance is a quite viable solution if user satisfaction remains at sufficient level [14]. Next step for planning the modernizations is to gather more technical metrics for parts *A3*, *A2*, and *A1*. The need for change is clearly smallest in case of *A1* and *B1*. If there are insufficient applicable resources for radical system changes, and changes are not initiated, part of the resources should be reallocated from the maintenance of *B1* to the maintenance of *A1*.

Table 5. Suggested evolution strategies for the evaluated parts: VDM - baselined variant (B), relative variant (R), simplified variant (S), final suggestion (*).

Evolution strategy/Part	***A1***	***A2***	***A3***	***B1***	***B2***	***B3***
Replacement		RS	RS		BRS*	BRS*
Modernization		B*	B*			
Maintenance	BRS*			B*		
Reduced maintenance				RS		
Implementation priority		← Modernization				← Replacement

5. Conclusions

Software maintenance and evolution is an economically very important part of the system lifecycle. Especially, methods supporting the selection of proper system evolution strategies are needed. One of the central problems with most of the existing methods is that their empirical validation is quite limited. One of the important and relatively well empirically validated methods is Value-Based Decision Model (VDM). It considers both economical and technical characteristics of the evaluated legacy system.

This paper has presented a systematically performed industrial case study applying VDM for system evolution strategy evaluation. The application area of the evaluated target system was tax payment monitoring. The target system was divided into subsystems, metrics were selected, and expert judgments gathered. Multiple industrial system evolution decision making experts took part to the evaluations. Subsystems were characterized based on the analyses. VDM's baselined calculations produced economical and technical values for the evaluated parts. The analyses revealed those subsystems whose economical value or technical quality need to be improved. They also provided information for the selection of suitable evolution strategies for those parts.

Two other calculation variants were used in studying the sensitivity of the results produced by VDM. Feedback was gathered from the industrial partners. The case study was considered to have been successful. VDM was considered clearly useful for its intended purpose. Main confronted problems in applying it have been presented. The received results supplement the earlier empirical studies on the actual industrial usefulness of methods for system evolution strategy evaluations.

6. References

1. L. Aversano, R. Esposito, T. Mallardo, M. Tortorella, "Supporting Decisions on the Adoption of Re-engineering Technologies", *Proc. 8th European Conference on Software Maintenance and Reengineering* (CSMR 2004), Tampere, Finland, 2004, pp. 95–104 (IEEE Computer Soc.).
2. V.R. Basili, G. Caldiera, H.D. Rombach, "The Goal Question Metric Approach", *Encyclopedia of Software Engineering*, 1994 (Wiley).
3. K. Bennett, M. Ramage, M. Munro, "Decision Model for Legacy Systems", *IEE Proceedings - Software* **146** (3), 1999, pp. 153–159.
4. D. Coleman, D. Ash, B. Lowther, P. Oman, "Using Metrics to Evaluate Software System Maintainability", *Computer* **27** (8), 1994, pp. 44–49.
5. A. De Lucia, M. Di Penta, S. Stefanucci, G. Venturi, "Early effort estimation of massive maintenance processes", *Proc. Int. Conf. on Software Maintenance - 2002* (ICSM 2002), 2002, pp. 234–237 (IEEE Computer Soc.).
6. G.A. Di Lucca, A.R. Fasolino, P. Tramontana, C.A. Visaggio, "Towards the Definition of a Maintainability Model for Web Applications", *Proc. 8th European Conf. on Software Maintenance and Reengineering* (CSMR 2004), Tampere, Finland, 2004, pp. 279–287 (IEEE Computer Soc.).
7. C.F. Kemerer, "Software Complexity and Software Maintenance: A Survey of Empirical Research", *Annals of Software Engineering* **1**, 1995, pp. 1–22.
8. J. Koskinen, H. Lintinen, H. Sivula, T. Tilus, "Evaluation of Software Modernization Estimation Methods Using NIMSAD Meta Framework", *Publications of the Information Technology Research Institute* **15**, University of Jyväskylä, Jyväskylä, Finland, 2004, Technical report, 62 p.
9. J. Koskinen, J.J. Ahonen, H. Sivula, T. Tilus, H. Lintinen, I. Kankaanpää, "Software Modernization Decision Criteria: An Empirical Study", *Proc. Ninth European Conf. on Software Maintenance and Reengineering* (CSMR 2005), Manchester, UK, 2005, pp. 324–331 (IEEE Computer Soc.).
10. M. Lehman, D. Perry, J. Ramil, "Implications of Evolution Metrics on Software Maintenance", *Proc. Int. Conf. on Software Maintenance - 1998* (ICSM 1998), 1998, pp. 208–217 (IEEE Computer Soc.).
11. B. Lientz, E. Swanson, "Problems in Application Software Maintenance", *Communications of the ACM* **24** (11), 1981, pp. 763–769.
12. P. Oman, J. Hagemeister, "Metrics for Assessing a Software System's Maintainability", *Proc. 1992 IEEE Software Maintenance Conf.* (ICSM 1992), 1992, pp. 337-344 (IEEE Computer Soc.).
13. J. Ransom, I. Sommerville, I. Warren, "A Method for Assessing Legacy Systems for Evolution", *Proc. IEEE Conf. Software Re-engineering* (CSMR 1998), Florence, Italy, 1998, pp. 84–92 (IEEE Computer Soc.).
14. I. Sahin, F. Zahedi, "Policy Analysis for Warranty, Maintenance, and Upgrade of Software Systems", *Journal of Software Maintenance: Research and Practice* **13** (6), 2001, pp. 469-493.

15. R. Seacord, D. Plakosh, G. Lewis, *Modernizing Legacy Systems: Software Technologies, Engineering Processes, and Business Practices* (SEI Series in Software Engineering), 2003, 332 p. (Addison-Wesley).
16. H. Sneed, "Estimating the Costs of Software Maintenance Tasks", *Proc. Int. Conf. on Software Maintenance - 1995* (ICSM 1995), 1995, pp. 168–181 (IEEE Computer Soc.).
17. H. Sneed, "Planning the Reengineering of Legacy Systems", *IEEE Software* **12** (1), 1995, pp. 24–34.
18. P. Stevens, R. Pooley, "Systems Reengineering Patterns," *ACM Software Engineering Notes* **23** (6), 1998, pp. 17–23.
19. T. Tilus, J. Koskinen, J.J. Ahonen, H. Lintinen, H. Sivula, I. Kankaanpää, "MODEST: A Method for Early System Modernization Pressure Estimation", submitted, 2005.
20. G. Visaggio, "Value-Based Decision Model for Renewal Processes in Software Maintenance", *Annals of Software Engineering* **9**, 2000, pp. 215–233.
21. I. Warren, J. Ransom, "Renaissance: A Method to Support Software System Evolution", *Proc. 26th Annual Int. Computer Software and Applications Conference* (COMPSAC 2002), 2002, pp. 415–420 (IEEE Computer Soc.).

Appendix 1: Selected metrics

The finally selected metrics are as listed. Those which are used in the final calculations are written in *italics*:

***Technical value*:**

- ***Technical expert judgment*:**
  - ***Correctness*:** Low amount and low severity of errors
  - ***Comprehensibility*:** Proper component size, low complexity, high structuredness, high quality of the related documentation and comments, high code quality, high conformance to the established conventions, non-redundancy of data
  - ***Adaptability*:** Low difficulty, effort and impact of changes, high reusability
  - ***Efficiency*:** High actual efficiency as compared to the required
- **Objective technical metrics:**
  - ***Constants*:** Amount of the numeric and string constants out of the declaration and definition contexts in code/size unit
  - ***DB-IB*:** Number of SQL retrieve (select, insert, update) statements bypassing the specified database interface/size unit
  - ***OS-calls*:** Relevant operating system calls within the area

***Economical value*:**

- ***Economical expert judgment*:**
  - **Criticality:** Importance of the system functionality to the operation of the organization in worst-case scenario of operation failure
  - **Appropriateness:** Conformance to the current business process, fitness to the intended purpose
- **Objective economical metrics:**

Automation: Ratio of the input data which can be processed without any (cost-producing) manual interventions
Input volume: Volume of the events/records received to be processed

Appendix 2: Calculations

This appendix details the three variants: VDM – baselined (B), relative (R), and simplified (S), used to calculate the final technical and economical scores for the evaluated system. Definitions:

- Mt_j technical quality metric j (1: DB-IB (number of SQL-retrieves), 2: OS-calls, 3: Constants, 4: Expert judgment #1 (application-team), 5: Expert judgment #2 (COBOL-team))
- Mb_j economical quality metric j (1: Expert judgment #3, 2: Expert judgment #4, 3: Input volume, 4: Automation (relative level, %))
- C_i part (component) to be evaluated
- S_i size of the part i (kSLOC)
- mt_{ji} technical quality metric j for part i
- mb_{ji} economical (business) quality metric j for part i
- T_j total amount of expert judgment based evaluation points of metric j
- n number of the components to be evaluated
- Vb_i economical value of part i
- Vt_i technical value of part i
- B_j baseline (objective) of the metric j
- w_j weight of the metric j

A2-1: VDM - baselined evaluation (B)

Technical metrics were normalized based on the preset baselines of objectives and part sizes (for the sake of brevity the internals of normalizations are not detailed here). Most of the economical metrics were normalized based on part sizes. Automation level was reversed into "manual level" and was normalized likewise as code metrics. Input volume was scaled linearly likewise as the expert judgments of technical quality. The initial proportionality of the evaluation scale to the baseline was found misrepresentative. This problem was corrected for code and other metrics by performing an unlinear transformation $m(x) = x^{0,5}$. The result was that the scale behaved approximately likewise as in case of other metrics.

$$Vt_i = \sum_{j=1}^{3} \frac{2B_j w_j}{mt_{ji} + B_j} + \sum_{j=4}^{5} \frac{w_j mt_{ji}}{B_j}$$

$$Vb_i = \sum_{j=1}^{2} \frac{w_j T_j mb_{ji}}{B_j S_i \sum_{k=1}^{n} \left(\frac{mb_{jk}}{S_k} \right)} + w_3 \sqrt{\frac{m_{3i}}{B_3}} + \frac{2(1 - B_4) w_4}{(1 - m_{4i}) + (1 - B_4)}$$

A2-2: Relative evaluation (R)

Used code metrics were normalized such that the effect of part sizes was eliminated (size relativity). The focus of interest is here on densities instead of amounts. All metrics were normalized linearly to [0,1]. The result is a sum of the metrics weighted based on the preset weights.

$$Vt_i = \sum_{j=1}^{3} w_j \left(1 - \frac{mt_{ji}}{S_i \cdot \max\{ mt_{jk} / S_k \mid 1 \le k \le n \}} \right) + \sum_{j=4}^{5} \frac{w_j mt_{ji}}{\max\{ mt_{jk} \mid 1 \le k \le n \}}$$

$$Vb_i = \sum_{j=1}^{4} \frac{w_j mb_{ji}}{\max\{ mb_{jk} \mid 1 \le k \le n \}}$$

A2-3: Simplified evaluation (S)

The evaluation was performed likewise as in case of size relative evaluation, except that some metrics were treated differently before the [0,1] normalization. Technical metrics were proportioned to the sizes of the parts. Code metrics were not normalized. Technical expert judgments were part size weighted.

$$Vt_i = \sum_{j=1}^{3} w_j \left(1 - \frac{mt_{ji}}{\max\{ mt_{jk} \mid 1 \le k \le n \}} \right) + \sum_{j=4}^{5} \frac{S_i w_j mt_{ji}}{\max\{ S_k mt_{jk} \mid 1 \le k \le n \}}$$

$$Vb_i = \sum_{j=1}^{4} \frac{w_j mb_{ji}}{\max\{ mb_{jk} \mid 1 \le k \le n \}}$$