



UNIVERSITY OF GOTHENBURG

Using Semantic Model Comparison to Support Evaluation of Environmental Impact on Big Production Setups

VERA KAPAYEVA

Thesis work for Master of Science in Software Engineering and Management

Report No. 2010:104

ISSN: 1651-4769

Table of Contents

ABSTRACT	2
DEFINITION OF KEY TERMS	2
1. INTRODUCTION	2
1.1 Disposition	3
2. PROBLEM IDENTIFICATION	3
2.1 Initial VTEC's Requirements for Project Outcome	3
2.2 Research Questions	3
3. RESEARCH METHOD	3
3.1 Methodology	3
3.2 Data Collection Methods	4
3.3 Data Analysis Method	4
3.4 Collaboration with Other Master Students	5
4. RELATED WORK	5
4.1 Existing Software	5
4.2 Existing Software: Production Engineering Perspective	6
4.3 Summary of Related Work	6
5. INTERVIEWS' AND FIELD VISIT'S RESULTS	7
6. RESULTS	7
6.1 Prototype Description	7
6.1.1 Tools and Technologies	7
6.1.2 Prototype Functionality	7
6.2 Using Prototype Features for VTEC Project	12
7. ANALYSIS AND DISCUSSION	13
8. FUTURE CHALLENGES	13
9. VALIDITY ANALYSIS	14
10. CONCLUSIONS	14
11. ACKNOWLEDGEMENTS	14
REFERENCES	14
APPENDIX A	16
APPENDIX B	16

Using Semantic Model Comparison to Support Evaluation of Environmental Impact on Big Production Setups

Vera Kapayeva
IT University of Gothenburg
Gothenburg Sweden
+46 739 943 688
kapayeva@student.chalmers.se

ABSTRACT

Context:

Working with process modeling is widely spread in production systems engineering and software development. Thus models have essential part in representing, visualizing and analyzing data in the field of car's production for Value Stream Mapping (VSM) and Life Cycle Assessment (LCA) purposes.

Objectives:

This thesis aims to develop a software tool for comparison and version controlling of models by combining VSM and LCA for analyzing environmental impact. From the software engineering perspective the thesis solves the problem of model versioning and semantic comparison of models.

Method:

The research method used in the thesis is case study of a production set-up at one of Volvo Construction Equipment plants. The qualitative semi-structured interviews, literature review, one field visit to Volvo plant and validation meetings with primary users of thesis outcome were the methods for data collection.

Results:

The result of this thesis is an algorithm for semantic model comparison and its prototype implementation in Microsoft Visio.

Conclusions:

It is concluded that models can be compared both syntactically and semantically using simple algorithms assuming that the problem of model comparison can be simplified to a limited number of comparison parameters. In our case the semantics behind model items is defined by one main parameter - carbon dioxide (CO_2) emissions.

Keywords

Semantic model comparison, model version control, Value Stream Mapping, Life Cycle Assessment.

DEFINITION OF KEY TERMS

Value Stream Mapping (VSM). It is known as “Material and Information Flow Mapping”, is used to depict current and future, or ideal states in the process of developing implementation plans to install lean systems [23].

Life Cycle Assessment (LCA). It assesses the environmental impacts related to a product or a system during its whole life cycle; i.e. energy, other resource use and emissions from material production, use, maintenance, as well as disposal of the product are included in the analysis [22].

Model Semantics, literally Meaningful Modeling. “Semantics” originates from Greek word “semantikos” and is translated as “the study of meaning”. Semantics is the data in context - where the meaning is concentrated. [9]

1. INTRODUCTION

Modeling, as a tool for visualizing, analyzing information, making prognosis and predictions, appears to be useful in both software engineering and production engineering spheres. In case of software engineering one of the main fields where models are widely applicable is Model-Driven Development (MDD). As for production engineering, there is a need to model processes on the plants and assess environmental impact. Environmental concerns appear increasingly often in the manufacturing industry as a cause of the environmental awareness of our society.

These concerns pose strict demands for Volvo Corporation since the old production lines, which were built in 1970's, were not optimized from environmental point of view. Nowadays there are methods for minimization of energy consumption during the manufacturing phase using principles of LCA to establish sustainable manufacturing. But there is still a lack of generalized model with parameters to evaluate production systems, i.e. generalized template for data collection and analysis from environmental point of view.

This need of Volvo Technology Corporation (VTEC), i.e. modeling of environmental aspects could be solved by using semantic model comparison. In the thesis we set the goal to create new algorithms for semantic model comparison and employ it on a specific case of VTEC.

New algorithm intends to solve software engendering problem in the field of models comparison by, first of all, carefully splitting model's semantics from model's syntax and, secondly, not relying on unique identifiers. These two motioned aspects are main difficulties in already existing algorithms and are emphasized in related work description.

The solution is important for model-driven engineering of software systems, but the problem in general is too complex to tackle and therefore we limit ourselves to a smaller one: comparing VSM models. We claim that if the problem can be solved for a well-defined modeling language - VSM, it can be solved for other well-defined modeling languages, like for example Unified Modeling Language (UML).

The intended audience of this work is software engineers who are working with modeling as well as middle engineers and managers from production departments who will be actually using the tool for analyzing environmental impact and as a way of communication for making decisions.

1.1 Disposition

Section 2 defines the problem addressed in this research project and defines research questions. Section 3 presents research methodology used to address the research problem and research questions. Section 4 is description of related to current research work from both software engineering perspective and big production setups' perspective. Section 5 provides interviews' and field visit's results. Section 6 describes thesis results. Section 7 evaluates thesis results by analyzing developed prototype functionality. Section 8 provides future work ideas. Section 9 and 10 finalize the paper with conclusions and thesis results validity analysis. Appendixes, providing the list of attached documents, use cases for developed prototype and a reference list are included at the end of the report.

2. PROBLEM IDENTIFICATION

2.1 Initial VTEC's Requirements for Project Outcome

The following expected outcomes were formulated initially for the whole thesis project at VTEC:

1. Creation of a systematic method to handle environmental parameters in the design and upgrade of a production setup. This method intends to combine VSM and LCA in a way to visualize environmental bottlenecks.
2. Designing of a computer-based tool for assessment and optimization of environmental performance of a production setup. The requirements for this tool are to be derived from the created method mentioned in point 1. A prototype for version controlling models, visualization of differences in models, migration of production line data from existing LCA tools is also to be developed.

These problems are important to be resolved since there is a need to have simple user friendly computer based tool for assessing environmental impact quickly without the need for effort-intensive collection of detailed information about the plant under investigation. Thus, a general method should be created for most of production lines. Since one of the most vivid ways for visualising information is modelling - the idea of analyzing environmental impact through model comparison of current plant state and possible optimised future state naturally arose.

To lower the cost of the tool it should use existing LCA databases for estimating environmental parameters and is to be implemented as a plug-in for MS Visio which was already used for similar purposes at VTEC.

2.2 Research Questions

The current thesis work addresses the following research questions (RQ):

RQ 1: How can environmental parameters' analysis on big production setups be visualized?

RQ 2: How to version-control models and visualize comparison results?

These research questions arise from VTEC's needs to model, compare and visualize the environmental impact of various production set-ups. For example, the most common task is to find out which VSM model gives better environmental performance, in our case we asses CO_2 emissions. Especially it is important when comparing current and future state VSM models of the plant.

Since the environmental information is a semantic domain of the model, the problem of semantic model comparison and differentiation is also important. The semantic model comparison and differentiation is the main software engineering problem in this thesis [1, 2, 3].

The answers to research questions will also contribute to the solution of problems described in the literature review.

3. RESEARCH METHOD

The purpose of this section is to detail methods which we used to address our research questions. Specifically, we will explain which methodology, data sources and data collection methods were used and why.

3.1 Methodology

In this work the case study methodology as advocated by Yin [14] was used as a method. The method was chosen as we had the opportunity to work with a real-world case at VTEC, namely the case of comparing VSM models. VTEC provided concrete case where we had the possibility to evaluate whether version-controlling and comparison of models could be used efficiently. The chosen research method is effective since it enables to conduct a qualitative in-depth study to understand the semantics of VSM models and make generalizations by conducting a field observation

at VTEC, analyzing the existing solutions and proposing and evaluating a new one.

3.2 Data Collections

The following list summarizes the data sources we used throughout this study:

- Literature review was carried out. We aimed to find out which work in the model comparison field was already done and which computer-based tools exist. The detailed design of literature study is provided in Section 4.
- Semi-structured interviews with the people who are working with environmental analysis and are potentially affected by the results of our project. In total 5 interviews were conducted. Interviews were recorded and the interview transcriptions are attached in the document together with the questionnaire (see Appendix A “Interview Questions and Records” document, the roles and positions of interviewees are also listed there). The results of the interviews have significant value in establishing requirements for the computer-based tool.
- The field visit to Volvo Construction Equipment plant in Braås [17]. The data necessary for constructing VSM was collected during 2 visits to the plant. The list of necessary data was mainly created by another master student who worked on this project (see Section 3.4). The data collection was conducted by all members of the thesis project at VTEC.
- Regular meetings with VTEC’s supervisors and academic supervisor from IT University of Gothenburg were conducted to evaluate computer-based tool’s prototype functionality, propose new ideas and assure the implementation was on the right track.

3.3 Data Analysis

As data analysis method we used the qualitative data analysis (QDA) process developed by Seidel in 1998 [26] where objectives, requirements and results are continuously reflected upon and transformed to more and more specific and concrete goals using newly collected data. There are three parts “Noticing”, “Collecting” and “Thinking about the things” in the method and relationships among these parts are reflected in Fig. 1.

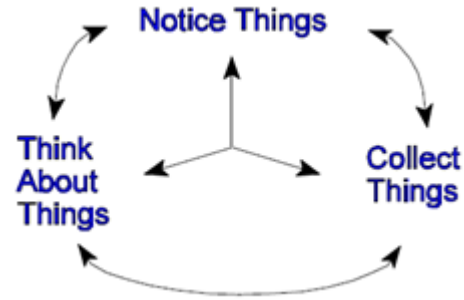


Fig. 1. Qualitative Data Analysis by Seidel [26]

In our case “things” are requirements to the model comparison algorithm and the prototype. As it is visible from the Fig. 1., the QDA process is not linear and the following characteristics should be taken into the account during analysis:

Iterative and Progressive: When we are thinking about things we also start noticing new things in the data. We then collect and think about these new things. In principle the process is an infinite spiral.

Recursive: While collecting the things, we might simultaneously start noticing new things to collect.

Holographic: Each step in the process contains the entire process. For example, when we first notice things, we are already mentally collecting and thinking about those things.” [26]

Thus, the following three parts were done iteratively through the case study using described above characteristics:

1. **Noticing:** during this phase initial purposes are set and hypotheses are formulated. It is decided what additional information is needed to confirm initial hypotheses.
2. **Collecting:** this phase aims to collect information. This was done using data collection sources mentioned in section 3.2.
3. **Thinking:** during this phase initial hypotheses are reexamined on the basis of analysis of data collected in the second phase.

“Noticing” and “Thinking” phases were done using regular meetings (see Appendix A “Meeting Conclusions”) with VTEC supervisors (usually once a week) where we evaluated all gained information and were coming up with ideas, plans and requirements for the computer-based tool which is the result of the current thesis work. During these meetings VTEC supervisors reflected upon how they are going to use the tool, and provided us with their feedback and comments. Rapid prototyping was used to present tool and model comparison algorithm to supervisors to get fast feedback and assure the development is on the right track.

We continued this iterative process until final validation of the thesis results by the end users.

3.4 Collaboration with Other Master Students

The whole project with the initial requirements from VTEC stated in Section 1.3 was carried out by two students: Vera Kapayeva and Yan Zhang who have different specialization backgrounds. Vera is a student of Software Engineering and Management master program at IT University of Gothenburg. Yan Zhang is a student of Production Engineering at Chalmers Institute of Technology. The purpose of such cooperation was to be able to develop a complete solution to VTEC for the stated tasks in the section 2.1 Initial VTEC's Requirements for Project Outcome. Thus, Yan was responsible for creation of systematic method to handle environmental parameters (1st requirement). Vera was responsible for implementing a computer-based tool for assessment and optimization of environmental performance of a production setup based on Yan's method (2nd requirement).

Yan's and Vera's common tasks were:

- Creating questionnaire for interviews,
- Managing interviews' appointments and conducting interviews themselves,
- Analyzing interview results,
- Visiting Braås Volvo plant and collecting data according to data sheets.

Specifically, Yan's tasks were:

- Preparing list of data to be collected for VSM construction during field visit to Braås,
- Construction of VSM model on the basis of collected at Braås information,
- Generalization of constructed VSM model which intended to fulfill the 1st requirement of VTEC project.

Specifically, Vera's tasks were:

- Making interview transcripts,
- Including LCA databases calculations into VSM model,
- Developing computer-based tool prototype with mechanisms for models' comparison and version control on the basis of VSM constructed by Yan. Thus, fulfilling the 2nd requirement of VTEC project.

Within the whole project at VTEC both students contributed equally to project's results although in this thesis only Vera's work results are described.

4. RELATED WORK

In this section we will give an overview of the work done in the area of model comparison, merging and versioning from software engineering perspective. The overview will also present the work done in analyzing environmental impact on big production setups and which computer-based tools already exist in this sphere.

We have a structured literature review approach: first the problem of analyzing environmental impact using combination of VSM and LCA was studied, then existing computer support was investigated in the field of big production setups and finally software engineering perspective was applied. The following sources of literature were used:

- Institute of Electrical and Electronic Engineers portal (IEEE),
- Association for Computing Machinery portal (ACM),
- Science Direct portal,
- literature recommended by academic supervisors and supervisors from VTEC.

As a key words for searching relevant articles we used "Semantic model comparison", "model version control", "model merging algorithm", "Value Stream Mapping" and "Life Cycle Assessment".

In total 35 relevant articles were read although only the most important articles listed in the references were used directly for making conclusions.

4.1 Existing Software

In this section we aim to give an overview of existing algorithms for models' comparison and list existing software tools implementing these algorithms.

Almanninger [1] explains the existing difficulty in model comparison: in order to identify the conflicts it is necessary not only to consider syntactical differences but also to "understand" the model's semantics, i.e. relationships between models' elements. Thus, the comparison should rely on logical structure of the model and visualize the conflicts which arise during merging compared models. Lin et al. [8] states that the commonly proposed technique for visualization the model differences is coloring. Almanninger [1] proposes "Semantically enhanced Model Version Control System" (SMoVer) which is based on unique identifiers (IDs) designed in the metamodel.

Alalen et al. [7], as well as Almanninger [1], rely on unique identifiers when proposing the algorithm for models comparison which is general for any modeling language based on the Meta Object Facility (MOF), i.e. it works for the Unified Modeling Language (UML).

Thus, in our algorithm we also need to establish uniqueness of the model items for comparison, i.e. algorithm should

identify which elements from models should be compared between each other. We can not use generated automatically by MS Visio model items' IDs since these IDs are regenerated, for example, if the shape is deleted and then added again to the model. In section 6 we explain how to define unique identifier for our case.

A technique for model comparison which does not rely on unique identifiers is presented by Treude et al. [12] where models are compared using transformations to an internal representation from an XML-based file format and similarity of elements does not depend on their compositional structure, but mainly on their neighborhood [6] We adopted this idea in our comparison mechanism. In VSM models order of the shapes has meaning, i.e. every shape meaning depends on its neighborhood. Thus, before comparing models we first sort them and then compare sorted arrays of models' items.

The work done by Oliveira et al. [2], [3] resulted in implemented tool Odyssey Version Control System (VCS) which is currently improving. Modeling elements may be split into two categories: syntactic and semantic elements. According to this principle during comparing we distinguish model items that are compared only syntactically and semantic items which contain environmental data and are compared against environmental parameters. In our case semantics is represented by the data from LCA database. Oliveira et al. [2], [3] reveal their model merge algorithm and acknowledge that they don't have visualization for model comparison and merging, they indicate it as a future work for Odyssey-VCS tool. That gives the room for current master thesis work.

Selonen [13] provides a review of five different approaches to model comparison. The author sets desirable qualities that any practical model comparison technique or tool should possess: identifier independence, reliability, usability, composability and non-intrusiveness. Thereby, several authors admitted that comparison mechanisms involving IDs are not enough to be able to find out the differences in models' semantics.

Harel et al. [9] provide an explanation what truly models semantic notion stands for as well as wrong ways to view semantics. Concisely, "when carefully devised, the semantics assigns an unambiguous meaning to each syntactically allowed phrase." [9] It is a challenge to distinguish syntax from semantics. We have faced this challenge when implementing our prototype and resolved it with the following main principles:

- all environmental information behind the Visio shapes which we store in the database or calculate using LCA databases defines items' semantics,
- items' order, as was already mentioned before, also has meaning in VSM models, thus contributes

to semantics aspect of the model comparison mechanism.

4.2 Existing Software: Production Engineering Perspective

This section aims to review work done in the field of combining VSM and LCA techniques and finding out whether there are any computed-based tools supporting this combination. Our case study needs industrial and software engineering perspectives for related work in order to be sure that our results are innovative in both spheres.

Steps to combining VSM and LCA techniques are done in "Lean and Clean VSM" article [10]. Authors propose to use VSM for analyzing production setups, as it is well known and mature approach, but include the list of input and output environmental parameters per each VSM process. The list of these parameters is clearly defined. This article gives understanding that VSM is a flexible tool which definitely gives the room for adding new data to be analyzed within VSM models. In our case this data is LCA data.

A considerable work is done in applying computer-based simulations for production setups: such computer programs as SIMTER and eQuest which are 3D simulation software tools for creating a "virtual environment in order to determine the best solution to achieve the goals, for example, for reducing energy consumption" [4], [5]. These kinds of tools are mature, expensive and require a lot of detailed initial data to be able to perform simulation. Our goal is to yield a simplified tool for environmental analysis based on 2D modeling.

There is eVSM [15] tool available on the market. The tool is a plug-in for Microsoft Visio and proposes the following useful functionality for constructing VSM models: analyzing and improving Value Streams, possibility to use Spaghetti Diagrams and constructing reach reports with diagrams.

We provide a prototype of plug-in for MS Visio which serves the needs for constructing VSM models including LCA data, possibilities for comparing, merging and version controlling these models and analyzing models' semantics. It should be emphasized that this functionality is not present in eVSM tool.

A very mature LCA tool is GaBi [19]. This tool provides modeling functionality and rich LCA databases for analyzing the whole life cycle of the product. In our work we are supposed to use only LCA thinking when constructing VSM.

4.3 Summary of Related Work

The main conclusions to be made from literature study are the following:

- Model comparison mechanisms should rely on more than just unique identifiers of model elements.
- There are already several model comparison algorithms developed, but none of them is implemented in a mature computer based tool which visualizes models comparison, merges, version-controls models and at the same time cover VTEC needs.
- When conflicts arise while merging models user has to resolve them manually. Software solution should help user to do it by visualizing model differences with coloring, graphical and textual representation of comparison results, i.e. proposing the possible ways of merging but only user can decide how the final model will look like.
- Although mature software tools exist for VSM and LCA they are not combined and don't cover VTEC needs, in particular the need for visualizing environmental information.

We took these conclusions as requirements when implementing tool prototype.

5. INTERVIEWS' AND FIELD VISIT'S RESULTS

This section intends to give main conclusions made from the interviews and visits to Volvo plant in Braås.

The following interviews results were obtained:

1. During semi-structured interviews all interviewees admitted the feasibility of combining VSM and LCA tools. This fact confirmed our original idea and let us proceed without changing project goals.
2. Interviews/meetings with VTEC supervisors resulted in establishing basic requirements for the computer-based tool: MS Visio integration and LCA databases integration. Also continuous evaluation of the tool prototype was held during these meetings. Thus, the following main requirements were added during evaluations:
 - create Visio stencil with custom shapes to be able to reuse them,
 - be able to save shapes' environmental information into the database to exchange data between different models,
 - dynamically generate semantic data for specific types of model items (station data sheets in VSM model) on the basis of user's answers to general questions regarding plant and stations. For example, estimate CO_2 emissions for ventilation energy consumption on the basis of approximation formulas which

use such parameters as station and plant sizes, production units per year, whole energy consumption for the plant per year etc.

As a stand along source of collected data are 2 visits to Volvo plant in Braås. There we took into the consideration only welding stations of the plant. The primary result of these visits was the VSM model constructed in MS Visio and using data collected on the plant. Then we analyzed how we can generalize obtained model and extend it to any plant. Thus, the following main steps were done to tune the model and make our computer-based tool prototype:

- All data sheet information collected in Braås was embedded into the model as Excel files per each plant station. These Excel sheets were generalized and implemented as custom MS Visio shapes.
- LCA data was connected to Visio model to calculate environmental parameters.
- Functionality for models version-control and comparison was added.

When final prototype was created, it was validated by users from VTEC.

6. RESULTS

This section provides full description of the thesis results:

- Description of the main prototype functionality,
- Model comparison algorithm with its pseudo-code,
- Model version-control algorithm,
- Impact of the prototype on VTEC.

6.1 Prototype Description

6.1.1 Tools and Technologies

The following software tools were used to develop the prototype:

- MS Visio 2007 – as a modeling tool,
- MS Access Database – as a database to store LCA data and models' configuration versions,
- MS Excel – as a tool for storing data collected during case study. We also used Excel documents within Visio models to store models' data in a structural way,
- Visual Basic for Applications (VBA) – as a programming language to implement prototype.

6.1.2 Prototype Functionality

First of all we want to stress that the evaluation of the prototype work is reflected in Interim Report document which is attached to this report (see Appendix A). The purpose of Interim Report artifact was to document in iterative manner on weekly basis all the changes in the

ideas regarding software solution for posted research questions during carrying out thesis. The current section gives the description only of the final prototype functionality which is divided into 3 main features:

Feature 1. Semantic Model Version Control, i.e. version control of the configurations for one model (RQ 2).

Feature 2. Models Comparison (RQ 1).

- 2.1. Within one model, i.e. comparison of the configurations of the same model. Configurations are represented by environmental data behind model items.
- 2.2. Comparison of 2 models.

Feature 3. Models Merging.

In order to describe prototype functionality in details use cases are presented in the Appendix B.

Feature 1: Semantic Model Version Control.

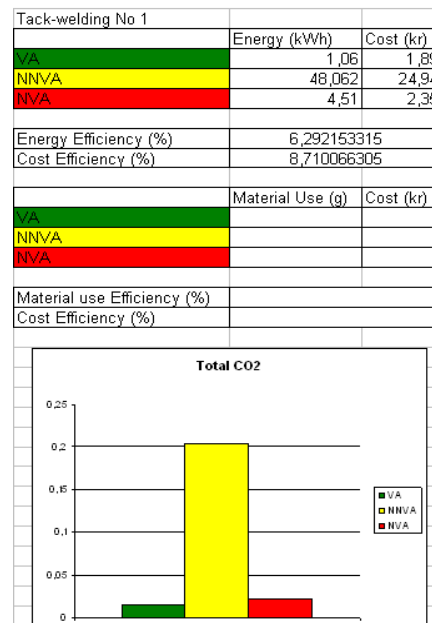
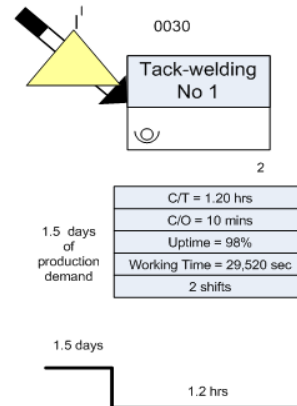
First, created VSM model (which was constructed on the basis of data collected at Braås) was static and didn't contain any macros and automatic calculations. Then we added custom shapes with embedded calculations of environmental parameters using LCA data.

Model's semantics is represented by the data contributing to environmental impact (see Fig. 2). This data is saved in the database and version-controlled per each model state. The **Use Case 1** describing this functionality in details is presented in Appendix B.

Parameters which represent operations and processes on the plant are split into 3 groups according to Lean development:

- Value Adding (VA) – represented in green color. These types of activities are operations which add value to a product through the processing of raw material or semi-finished products. For electricity consumption while welding it involves mainly electricity consumed when welding equipment is on. [16]
- Necessary Non-Value Adding (NNVA) – represented in yellow color. These operations may not add value to a product and be wasteful but are necessary for current operation procedures. Example for this category is electricity consumption of supporting systems such as ventilation, heating and lighting. [16]
- Non-Value Adding (NVA) – represented in red color. These activities are pure waste that should be removed entirely from the process. For example, these processes can include electricity consumption when equipment is in

stand-by mode, i.e. not used but is switched on. [16]



Excel Sheet

Fig¹. 1. Data Relative per Station.

Feature 2.1: Semantic Comparison within one model, i.e. comparison of model configurations.

All the comparison is carried out on the basis of the saved model configuration data using **Feature 1**. We have chosen the most important parameters to be saved and compared. The result of comparison is presented in the textual and graphical format (see Fig. 3).

¹ Red arrows and text comment pictures and are not present in the Visio document with the models.

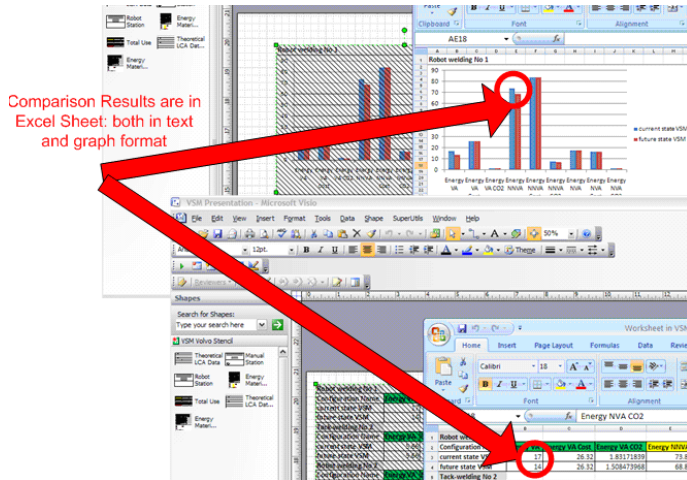


Fig. 3. Semantic Comparison Results.

The Use Case 2 from Appendix B describes this feature in details.

Feature 2.2: Comparison of 2 models.

The comparison of 2 different models is held against its syntactical and semantic differences at the same time. The algorithm for this feature is the following (in brackets see the references to the pseudo-code lines, the pseudo-code is presented after algorithm description):

1. All shapes are ordered on both models by the coordinates they are placed on the model's sheet (from the left to the right, from the top to the bottom). This action deals with semantics of the model since the order of shapes in VSM has meaning. [lines 16-18]
2. Ordered arrays of shapes are compared syntactically and semantically:
 - 2.1. In case compared shapes have type "Data Sheet" (see Fig. 2., "Data Sheet" is Excel file containing specific information for the station) they are compared semantically against the last saved configurations in the same way as in the **Feature 2.1** with Use Case 2 from Appendix B. As a result semantic comparison report is shown to the user (see Fig. 3.). [lines 55-59]
 - 2.2. In case compared shapes are of any other type except of "Data Sheet" they are compared only syntactically against shape type and shape text. See Fig. 4. for demonstration of syntactical differences highlighting [lines 20-48]:

- black color indicates that no differences were found,
- red text indicates that there is the shape in another model with the same shape type but the text is different,

- blue color indicates that the shape is present in one model but is absent in another,
- yellow color indicates that the shape types are different although the text is the same,
- when clicking on any shape after comparison the correspondent shape in another model is highlighted in red – thus user can easily see which shapes in the first model were compared to which shapes in the second model.

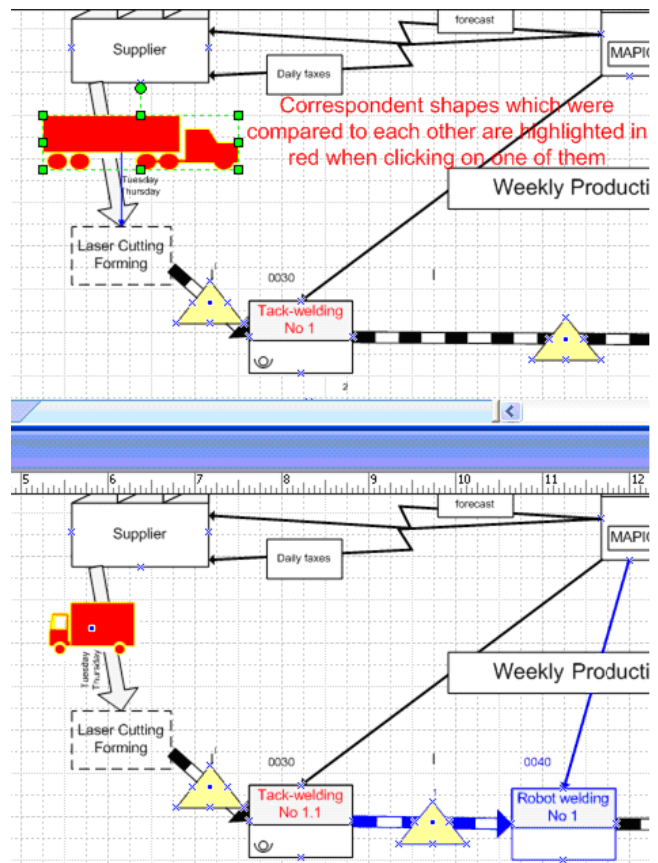


Fig. 4. Syntactical Comparison Results.

The Use Case 3 from Appendix B describes this feature from the point of view user-Model interaction.

Main advantage of the algorithm is that it clearly distinguishes semantic and syntactical differences and reports about the differences as well in a different manner. For syntactical differences we present shapes' highlighting (see Fig. 4), for semantic differences we present comparison report (see Fig. 3). The comparison algorithm doesn't rely on unique identifiers since in VSM models the order of the shapes has meaning and we managed to take into the consideration shapes surrounding. It should be pointed out that we distinguish the shapes by their types: there are types which are compared only syntactically and types which are

compared semantically (“Data Sheet” type, see **line 28** in pseudo-code).

We also managed to reduce big amount of semantic data in the shapes which are compared semantically (see Fig. 2, Excel Sheet, which is referred as shape of “Data Sheet” type in pseudo-code) to just one main parameter (see **line 59**) against which these types of shapes are compared. In the case of VSM models this parameter is CO_2 emissions produced by station (see Fig. 2, graph called “Total CO_2 ”). However in the comparison report (see Fig. 3) we present several parameters against which model configurations are compared. This is because there is a possibility to extend semantic comparison to as many parameters as it is needed by models’ semantics.

The pseudo-code for comparison algorithm is presented below:

```

1 ' Compare 2 opened models and highlight the differences
2 Public Sub CompareModels()
3
4     Dim oDoc1 As Document ' 1st model for comparison
5     Dim oDoc2 As Document ' 2nd model for comparison
6     Dim oDoc1Page1 As Page ' 1st page of the 1st document/model
7     Dim oDoc2Page1 As Page ' 1st page of the 2nd document/model
8     Dim oShapeDoc1 As Shape ' any shape on the 1st document/model
9     Dim oShapeDoc2 As Shape ' any shape on the 2nd document/model
10
11     Set oDoc1 = Visio.Application.Documents(1)
12     Set oDoc2 = Visio.Application.Documents(2)
13     Set oDoc1Page1 = oDoc1.Pages(1)
14     Set oDoc2Page1 = oDoc2.Pages(1)
15
16     ' order shapes before comparison
17     oDoc1Page1.Shapes = OrderShapes(oDoc1Page1.Shapes)
18     oDoc2Page1.Shapes = OrderShapes(oDoc2Page1.Shapes)
19
20     ' SYNTAX COMPARISON
21     ' loop through ordered arrays of shapes on both models
22     For Each oShapeDoc1 In oDoc1Page1.Shapes
23         For Each oShapeDoc2 In oDoc2Page1.Shapes
24             sShapeType1 = oShapeDoc1.NameU
25             sShapeType2 = oShapeDoc2.NameU
26             ' syntax comparison is not performed for shapes with the type "Data Sheet" - they are
27             ' compared semantically
28             If StrComp(sShapeType1, "Data Sheet") <> 0 And StrComp(sShapeType2, "Data Sheet") <>
0 Then
29                 If StrComp(sShapeType1, sShapeType2) = 0 And StrComp(oShapeDoc1.Text,
oShapeDoc2.Text) = 0 Then
30                     ' the shapes are the same, highlight shape border and shape text in green
31                     ' exit from loop for oDoc2Page1.Shapes and take next shape from
32                     ' oDoc1Page1.Shapes
33
34                     ElseIf StrComp(sShapeType1, sShapeType2) = 0 And StrComp(oShapeDoc1.Text,
oShapeDoc2.Text) <> 0 Then
35                         ' shapes are of the same type but with different text
36                         ' highlight text in red
37                         ' exit from loop for oDoc2Page1.Shapes and take next shape from
38                         ' oDoc1Page1.Shapes
39
40                     ElseIf StrComp(sShapeType1, sShapeType2) <> 0 And StrComp(oShapeDoc1.Text,
oShapeDoc2.Text) = 0 Then
41                         ' shapes are of different types but with the same text
42                         ' highlight shape border in yellow
43                         ' exit from loop for oDoc2Page1.Shapes and take next shape from
44                         ' oDoc1Page1.Shapes
45                     End If
46                 End If
47             Next
48         Next
49
50         ' if shape stays in black color (black text and border) after the previous loop it means the
51         ' shape doesn't present on one of the compared models->
52         ' change black color to blue
53         ' change green color back to black color, so that shapes which are the same are not
highlighted
54
55         ' SEMANTICS COMPARISON
56         ' shapes with the type "Data Sheet" were skipped during the syntax comparison
57         ' the function takes the latest configurations from the Models' databases for both models and
58         ' configuration contains several paramters per each station: the main paramter is only one
59         ' shows comparison report grouped by stations
60         ShowComparisonReport
61
62     End Sub

```

Model Comparison Algorithm: Pseudo-code.

Feature 3: Models Merging.

This feature is not fully implemented in the prototype although we provide the ideas for further development.

First, the models are compared as it is described in the **Feature 2.2**: syntactical differences are highlighted (see Fig. 4.) and semantics report is shown (see Fig. 3.).

1. Then user selects “Merge Models” command and all the shapes which have the same syntax are copied to a new model document. This functionality is implemented and copies all shapes in black color from the first model to a new document.
2. Now user has to decide how to actually merge the conflicts. When user is clicking on a shape in one model – the correspondent shape is highlighted in red in another model. This is also implemented and described in **Feature 2.2**. We propose to enrich this step with the following functionality:
 - 2.1. the form is show to user proposing to choose one of the 4 different options: (1) copy shape from the 1st model with the semantic data from the 1st model, (2) copy shape from the 2nd model with the semantic data from the 2nd model, (3) copy shape from the 1st model with the semantic data from the 2nd model, (4) copy shape from the 2nd model with the semantic data from the 1st model. These options should be visualized, i.e. all the data necessary for making the decision should be presented to user.

6.2 Using Prototype Features for VTEC Project

Described features fulfill VTEC’s needs in analyzing current VSM for plant, creating future VSM for plant and comparing VSM models against environmental parameters. It is done by performing **Use Case 1** and then **Use Case 2** from Appendix B. Thus, the following summarizes the opportunities which our tool provides for VTEC:

- User can quickly construct VSM models using custom Volvo Visio shapes with predefined structure for keeping data necessary for calculating environmental impact (Fig. 5.). User now can keep all the data necessary for VSM construction within the model through embedded Excel data sheets in predefined templates.

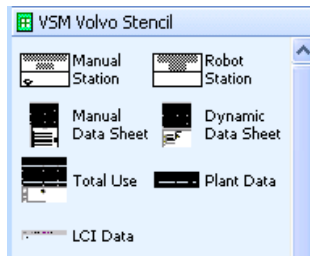


Fig. 5. Custom Volvo Stencil.

- CO_2 emissions are calculated automatically by the tool using LCA database and after user inputs data required by predefined Excel data sheets from Volvo custom stencil.
- User can save any states of the model performing **Use Case 1** for further comparison with other model states.
- User can analyze environmental impact of each station in VSM model as well as environmental impact of the whole plant using graphical information representation (see Fig. 6.).

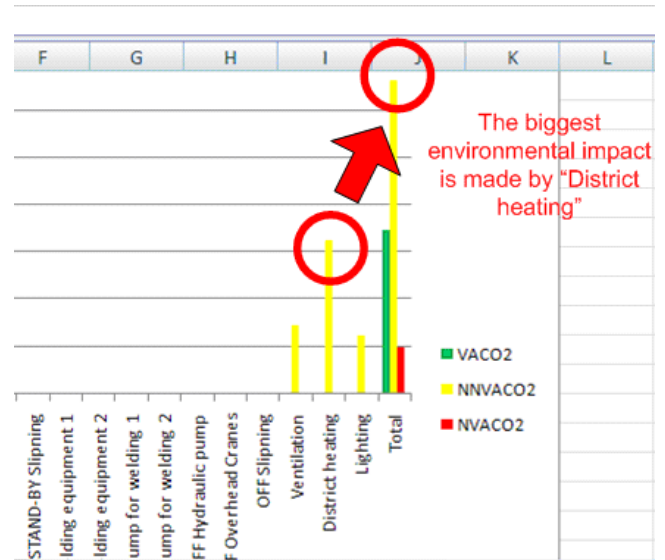


Fig. 6. Graph for Analyzing Environmental Information.

- User can compare different models configurations performing **Use Case 2** for finding the most environmental friendly one, i.e. there is always a possibility to compare current and future state VSM and conclude whether improvement actions are worth to be performed on the basis of visualized report (see Fig. 3).

It is important to emphasize that for VTEC’s project more functionality was implemented than described in Appendix B and User Guide document was created for VTEC users.

7. ANALYSIS AND DISCUSSION

Analysis of the developed prototype was held during regular meetings with supervisors from VTEC and IT University of Gothenburg. The main decisions regarding the prototype functionality were made after Braås plant visit when static VSM model was constructed. The main challenge was to enhance the model with dynamic behaviour replacing static shapes with custom shapes where environmental parameters calculations were embedded and then to come up how to split syntax and semantics of the model for further comparison.

On the final tool demonstration the following advantages of the prototype were admitted by the end users (production engineers) from VTEC:

- Usage of embedded Excel sheets for storing necessary information. This solution turned to be user friendly since Excel is widely used for similar purposes by production engineers.
- Creation of custom stencil with reusable shapes. It gives possibility to construct VSM models quickly using predefined formats for data sheets.
- The possibility to compare models and their configurations as the leading idea of the tool was unanimously accepted by all users.

The following disadvantages of the prototype were pointed out by thesis author and software engineer from VTEC:

- The calculations are slow because of usage of Excel files in the Model.
- VBA script as a programming language for the prototype makes restrictions in programming, for example, in dynamic event handling.

Thus, the end users were satisfied with the proposed solution although further improvements in the tool would be good advantage and are stated as future work. Developed prototype resolves the problem of combining VSM and LCA.

Analyzing contribution of the current thesis to existing software problems described in related work (see section “4.3 Summary of Related Work”) we emphasize that:

- Developed algorithm does not rely on unique identifiers when comparing and version-controlling models. It identifies model’s items by their types, location in the model and meaningful names assigned to the items.
- The ideas described in **Feature 3** resolve the problem of manual model merging visualizing model differences with coloring, graphical and textual representation of comparison results.
- The main contribution to the problem of semantic model comparison is done by introducing new algorithm which splits model’s semantics from

syntax by storing semantics in the database assigned to the model.

Summarizing the whole work we present the answers to posted research questions:

RQ 1: Environmental parameters’ analysis on big production setups may be visualized by modeling the material and information flow using VSM combined with LCA data for assessing environmental impact. We provided the prototype of the tool which includes functionality for analyzing this type of VSM models with LCA data by means of model comparison and version-control.

RQ 2: When comparing models’ semantics it is possible to find out one main aspect/parameter against which the model items may be compared. The idea is similar to comparing products by their cost. In the case of VSM models with LCA data this parameter is carbonate dioxide (CO_2) emissions. Syntax comparison of the models is implemented according to the differences in types and text of shapes. The comparison is not based on unique identifiers. It also takes to the consideration that in VSM all shapes are ordered, i.e. the order has the meaning, and models must be compared in the correspondent order of shapes. To version control the models’ semantics it is enough to version-control the minimum set of parameters representing this semantics.

Thus, answering the second research question we proposed the generalization of our approach when comparing models. It means that the solution is not specific for VSM models and could be used for other types of models.

8. FUTURE CHALLENGES

In this section we list some ideas for possible future work.

Firstly, semantically enhanced merging, i.e. conflicts resolution, should be implemented as it is described in Section 6 in **Feature 3**. The idea reminds the functionality provided by Araxis Merge tool [21] for comparison of textual files. However, this is a task that requires more advanced modeling tools and therefore was not addressed in this thesis.

From industrial perspective, the main future work is to extend our prototype to be able to connect to different LCA databases in different formats, for example to European Reference Life Cycle Database (ELCD) which contains more data and can be used without licensing issues.

We believe that our ideas will help to define comparison rules for further researches.

9. VALIDITY ANALYSIS

We evaluate validity of the presented thesis following schema described in [25] against four validity threats to empirical studies.

There is an *external validity* threat in this work that no interviews within other companies except of Volvo regarding comparing, version-controlling and merging models were conducted. However the fact that our supervisors at VTEC and IT Gothenburg University have solid experience within software modeling field makes obtained results reliable.

The main *internal validity* threat to the study is that there is a risk the results are specific for VSM model for Braås plant. To minimize this threat, generalizations were made to be able to apply results to other plants and to other types of models.

Construct validity threat was also identified. Not all model shapes with semantic data may be characterized by CO_2 emissions or this characteristic is not enough for understanding the whole picture of environmental impact. Nevertheless, we found out that we can use other environmental parameters to characterize shapes in VSM models.

As a *conclusion validity* threat we see the lack of testing the tool prototype ideas on other plants and other types of models which are different from VSM modelling.

10. CONCLUSIONS

This thesis addresses the problem: there is no mature computer based tool which supports specific VTEC's needs (MS Visio integration, LCA databases integration), visualizes models comparison, merging and version-control taking into account not only models' items syntax but also models' semantics. We have resolved this problem and answered posted research questions in the context of VSM modeling on big production setups by providing the prototype implemented in MS Visio. We also introduced critical vision of proposed solution indicating future work and possible problems in generalizing and extending our approach to be applicable to other types of models (for example, UML diagrams).

We emphasize that developed prototype covers two initial thesis requirements from VTEC since it uses LCA data to calculate environmental impact from VSM models. The first requirement coverage is mostly the result of thesis author partner's work at VTEC's project. The second requirement is covered by described prototype functionality and by the list of future ideas in the "Future Challenges" section. It was extremely difficult to resolve at the same time both software engineering and production engineering problems. We achieved this with project partner by posting different research questions and writing different thesis reports, although we were working in a close cooperation

on a daily basis exchanging ideas and providing each other with necessary materials.

Empirical data obtained from the interviews and field visit together with literature review results confirmed the need in our work. Our solution is validated both by VTEC supervisors and IT University of Gothenburg supervisor via regular supervision meetings.

11. ACKNOWLEDGEMENTS

The author would like to thank Dr. Mirosław Staron as an academic supervisor of the current thesis project for critical approach, valuable ideas and advices regarding thesis content, research questions, structure and focus.

The author is also grateful to VTEC's supervisors Thomas Lezama, Per Hanarp, Lisbeth Dahlöf and Maria Walenius Henriksson for initializing the project, providing requirements for it and constant support through the whole work with necessary information, ideas and reviews of our progress.

The author is particularly grateful to VTEC's project partner Yan Zhang and his supervisor Prof. Johan Stahre for professional collaboration.

We also want to admit huge contribution to the thesis results of all our interviewees and especially of Niklas Sjöden, the head of security at Volvo Braås plant, who managed our 2 visits to Braås.

Special thanks we say to MyCar and FUTURA EU projects for creating the need in the current thesis work.

And finally the author thanks Maria Scherbakova, Negar Fazeli and Zinaida Ratko for friendly support during the whole period of the thesis work.

REFERENCES

- [1] Kerstin Altmanninger. Models in Conflict – Towards a Semantically Enhanced Version Control System for Models. Department of Telecooperation, Johannes Kepler University Linz, Austria. H. Giese (Ed.): MoDELS 2007 Workshops, LNCS 5002, pp. 293-304, 2008. © Springer-Verlag Berlin Heidelberg 2008.
- [2] Hamilton Oliveira, Leonardo Murta, Claudia Werner. Odyssey-VCS: a Flexible Version Control System for UML Models Elements. COPPE/UFRJ – Systems Engineering and Computer Science Program Federal University of Rio de Janeiro. SCM 2005, September 5-6, 2005, Lisbon, Portugal.
- [3] Leonardo Murta, Hamilton Oliveira, Cristine Dantas, Luiz Gustavo Lopes. Odyssey-SCM: An integrated software configuration management infrastructure for UML models. COPPE/UFRJ – Systems Engineering and Computer Science Program Federal University of Rio de Janeiro. © 2006 Elsevier B.V.

- [4] Yimin Zhu. Applying computer-based simulation to energy auditing: A case study. Department of Construction Management, College of Engineering and Computing, Engineering Centre, Florida International University, Miami, 2005.
- [5] Juhani Heilala, Saija Vatanen, Hannele Tonteri, Jari Montonen, Salla Lind, Björn Johansson, Johan Stahre. Simulation-based Sustainable Manufacturing System Design. © 2008 IEEE.
- [6] Lars Bendix, Department of Computer Science Lund Institute of Technology S-221 00 Lund, Sweden. Pär Emanuelsson, Ericsson AB S-583 30 Linköping Sweden. Diff and Merge Support for Model Based Development.
- [7] Marcus Alanen, Ivan Porres. Difference and Union of Models. TUCS Turku Centre for Computer Science Department of Computer Science, Åbo Akademi University Lemminkäisenkatu 14, FIN-20520 Turku, Finland, 2003.
- [8] Yuehua Lin, Jing Zhang, Jeff Gray. Model Comparison: A Key Challenge for Transformation Testing and Version Control in Model Driven Software Development. Dept. of Computer and Information Sciences, University of Alabama at Birmingham Birmingham AL 35294-1170
- [9] David Harel, Weizmann Institute of Science. Bernhard Rumpe, Technische Universität Braunschweig. Meaningful Modeling: What's the Semantics of "Semantics". © 2004 IEEE.
- [10]. Green Suppliers Network, (2010). Lean and Clean Value Stream Mapping. Available: <http://www.greensuppliers.gov>
<http://www.epa.gov/>
- [11] Yolanda Gil, Varun Ratnakar. A Comparison of (Semantic) Markup Languages. USC Information Sciences Institute and Computer Science Department 4676 Admiralty Way Marina del Rey, CA 90292. Copyright © 2002, American Association for Artificial Intelligence. Available: www.aaai.org.
- [12] Treude, C., Berlik, S., Wenzel, S., Kelter, U.: Difference Computation of Large Models, in Proceedings of ESEC/FSE 2007, Dubrovnik, Croatia, September 3-7, 2007.
- [13] Selonen, P: A Review of UML Model Comparison Approaches, in Proceedings of Nordic Workshop on Model Driven Engineering, Ronneby, Sweden, August 27-29, 2007.
- [14] Robert K. Yin. Case Study Research: Design and Methods. Fourth Edition. © 2009 by SAGE Publications, inc.
- [15] GumshoeKI, Inc., (2010). eVSM software tool. Available: <http://evsm.com/>
- [16] Xiaojin Qiu, Xiaoxia Chen. Evaluation of the environmental impacts from implementing lean in production process of manufacturing industry. Chalmers university of Technology, Master Thesis 2009.
- [17] Volvo CE Braås, (2010). Available: <http://www.verkstadsklubben-braas.se/>
- [18] IEEE computer society, (2010). Available: <http://www.computer.org/portal/web/csdl/home>, ACM portal, (2010). Available: <http://portal.acm.org/portal.cfm>, Science Direct, (2010). Available: <http://www.sciencedirect.com/>
- [19] GaBi software: product sustainability, (2010). Available: <http://www.gabi-software.com/>
- [20] ELCD database source, (2010). Available: <http://lca.jrc.ec.europa.eu/lcaifohub/databaseList.vm>
- [21] Araxis Merge tool, (2010). Available: <http://www.araxis.com/merge/>
- [22] Baumann, Tillman. The HitchHiker's Guide to LCA, An orientation in life cycle assessment methodology and application. 2004.
- [23] Rother, Shook. Learning to see - value stream mapping to create value and eliminate muda, Lean enterprise institute: Cambirdge, Massachusetts, USA. 2003.
- [24] Womack, Jones. Lean Thinking: Banish waste and create wealth in your corporation. Free Press Business. London, Great Britain. 2003.
- [25] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslèn, A. Experimentation in Sw. Eng.: An Introduction, Kluwer, Boston MA (2000).
- [26] Seidel, J. V. (1998). Qualitative Data Analysis, Qualis Research. Available: <ftp://ftp.qualisresearch.com/pub/qda.pdf>

APPENDIX A

This report goes together with the following list of artifacts which are attached in the folder “Appendix A”:

- “Meeting Conclusions” document represents the records from all meetings with supervisors from VTEC.
- “Interview Questions and Records” document summarizes collected raw data from interviews.
- “Interim Report” document reflects the evaluation of computer-based tool prototype development.

APPENDIX B

Use Case 1. Save Model configuration for further comparison with another version: semantics version control.

1. User opens VSM Model constructed for the plant with or without filed in data for the stations.
2. Model is displayed as shown on Fig. 7.
3. User fills in or edits all station data sheets (see Fig. 2) with the information collected on the plan (for example, energy consumption in kW per hour for welding equipment) and presses “Generate Stations’ Data Sheets” button (see Fig. 7, left upper corner of the Model).
4. Model calculates CO_2 emissions using LCA data. The results of this action are stored within station data sheets which are Excel files and contain several tabs (see Fig. 8)
5. User assumes that Model is completed and goes to “Version Control” page in the Visio document.
6. “Version Control” page is displayed as shown on Fig. 9.
7. User presses “Save Current Configuration” button and provides the name for configuration.
8. Model saves current configuration into the database. Only the most important parameters per station are saved, for example, total energy consumption for value adding operations and CO_2 emissions for it.

Use Case 2. Compare Model’s configurations: semantics comparison.

Precondition: **Use Case 1** is completed.

1. User analyzes what could be improved on the plant, i.e. how to reduce environmental impact. Thus, user opens detailed data per station (see Fig. 8) and finds the processes which should be improved. All the processes impact is shown in CO_2 terms and it is easy to find processes which contribute most of all into total environmental

impact by looking on the detailed charts (see Fig. 8.1.).

2. User changes the data which he can improve for every station data sheet and presses “Generate Stations’ Data Sheets” button.
3. Model recalculates CO_2 emissions for all the stations.
4. User goes to “Version control” tab, presses “Save Current Configuration” button and provides the name for new configurations.
5. Model saves new configurations into the database.
6. User chooses 2 configurations which he wants to compare and presses “Compare” button.
7. Model displays the result of comparison on “Comparison Results” page (see Fig. 10.) both in textual and graphical format.

Use Case 3. Compare 2 Models: syntactical and semantic comparison.

Precondition: **Use Case 1** is completed for both Models which are to be compared.

1. User opens 2 VSM Models in Visio (see Fig. 11).
2. User executes macros “Compare Models”.
3. 2 types of comparison results are presented to user:
 - 3.1. syntactical differences are highlighted (see Fig. 11.1)
 - 3.2. semantic differences of the latest models’ configurations are presented in the same manner as in **Use Case 2** (see Fig. 10), comparison report is displayed on the “Comparison Results” tab of both compared models.

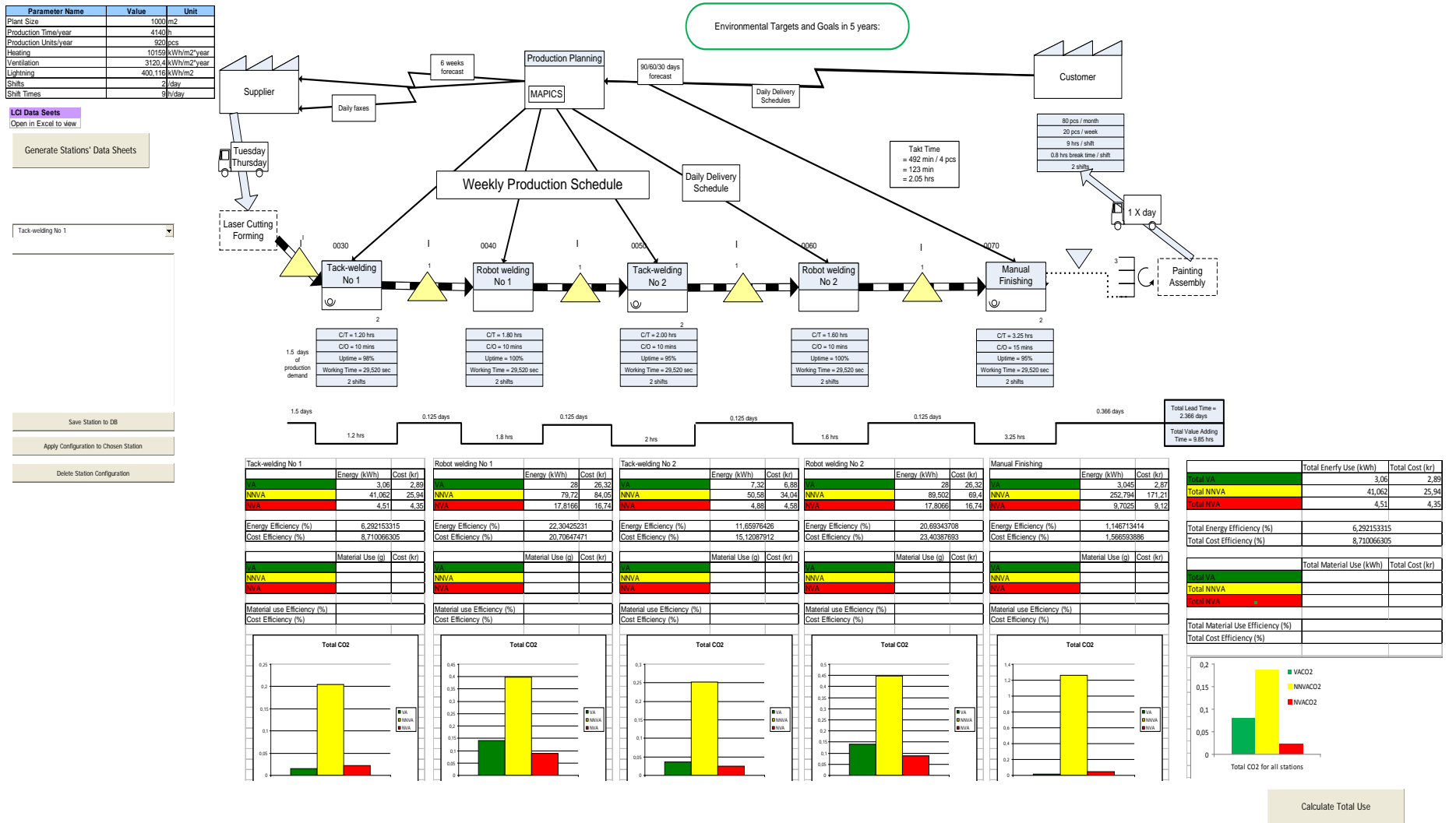


Fig. 7. Whole VSM model for Braås Volvo Plant.

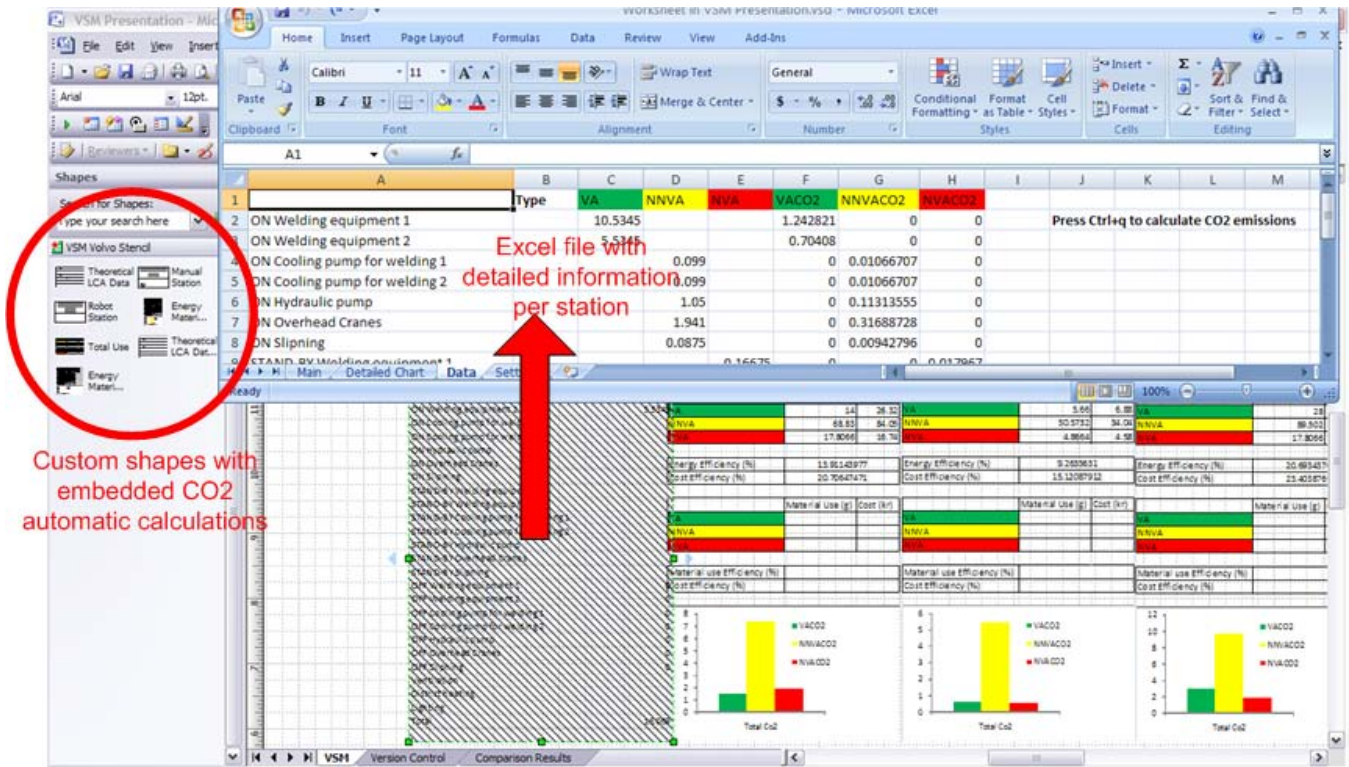


Fig. 8. Station Detailed Data in Excel File.

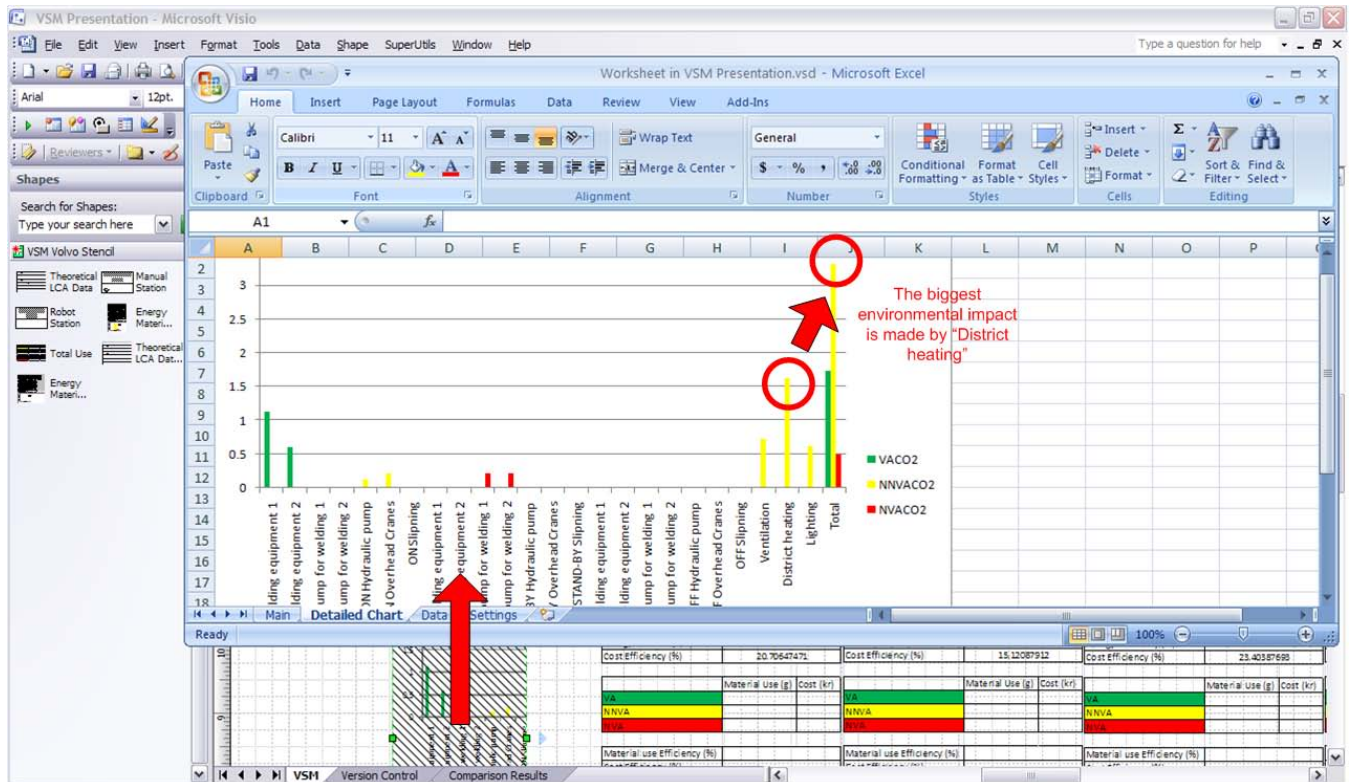


Fig. 8.1. Station Detailed Data in Excel File. Analyzing Environmental Impact.

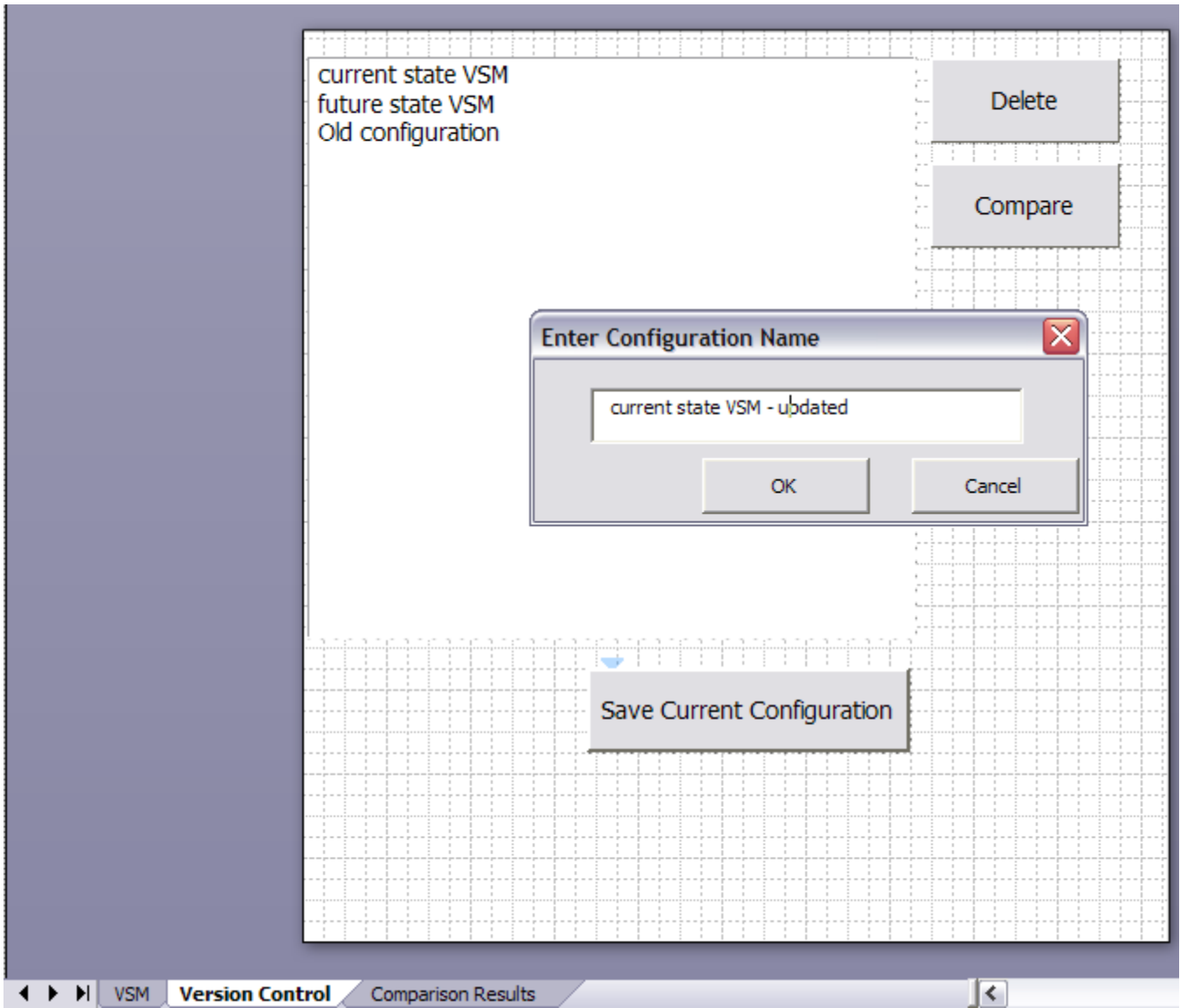


Fig. 9. Model Version Control – Saving Model Configuration.

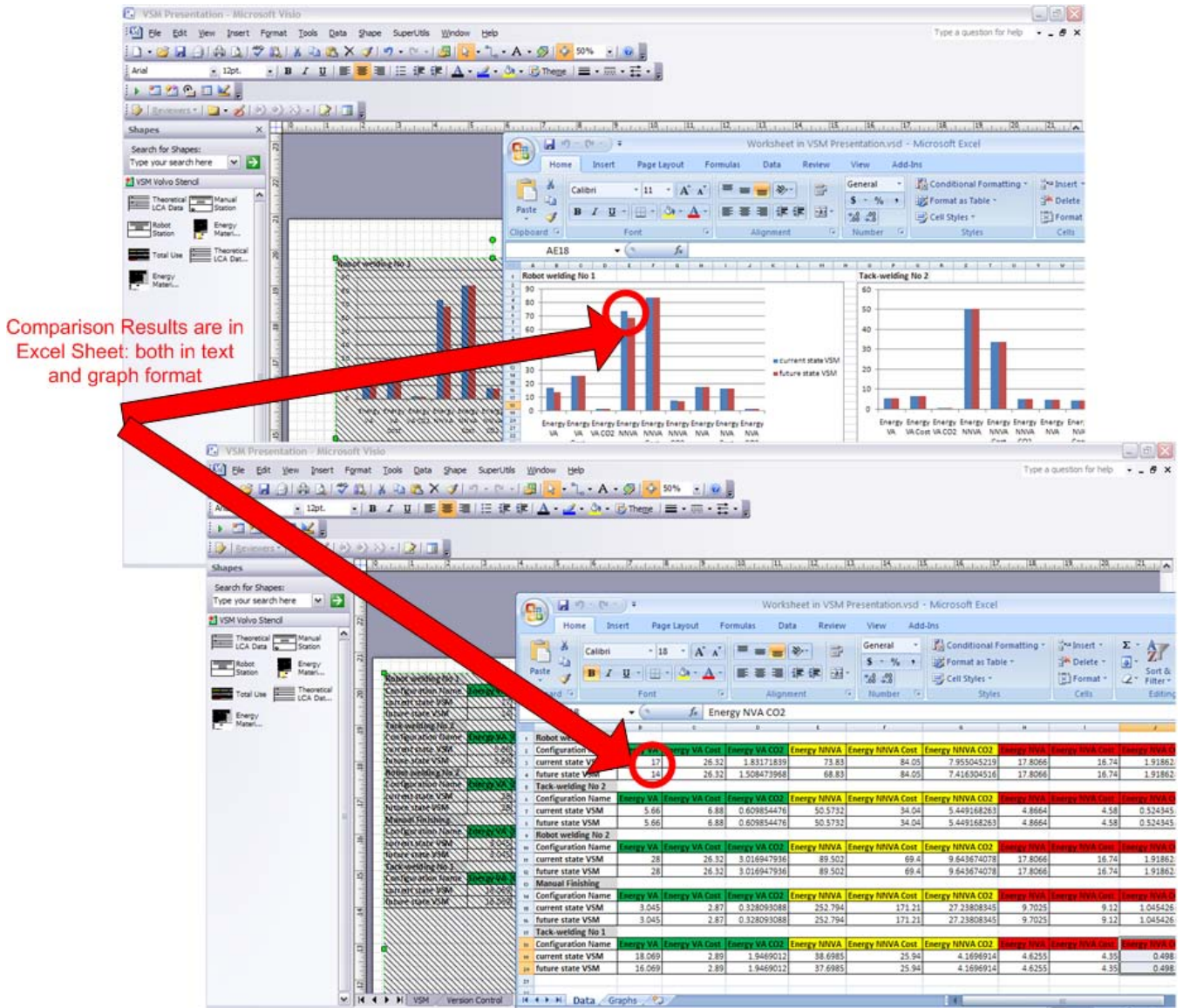


Fig. 10. Semantic Comparison Results.

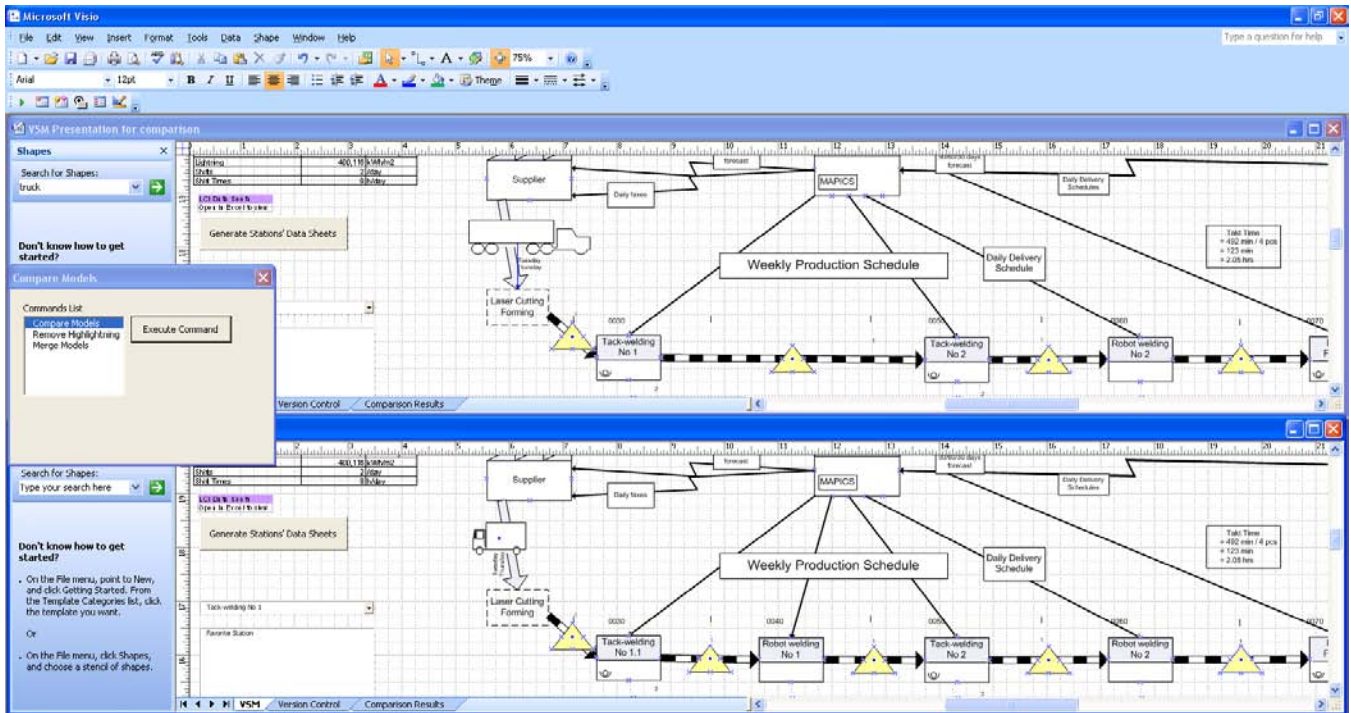


Fig. 11. Models to Be Compared.

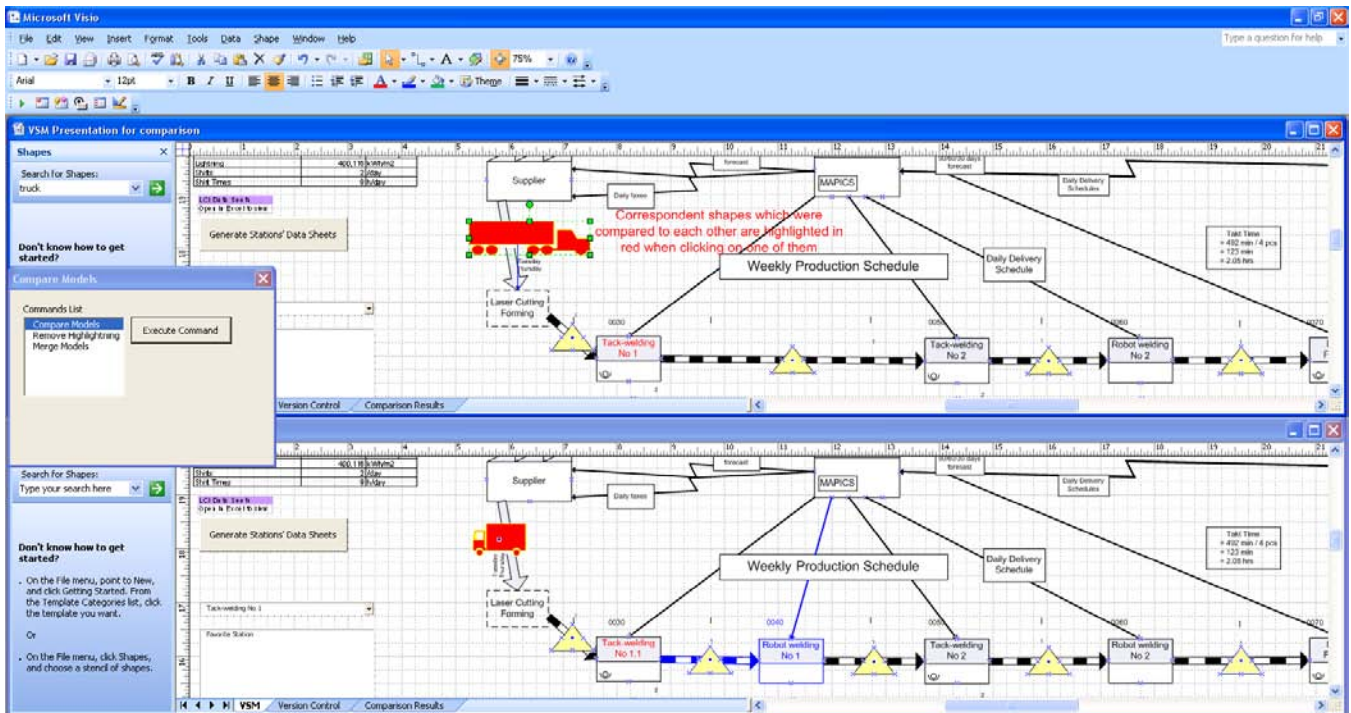


Fig. 11.1. Highlighting Models' Syntactical Differences.