Keynote

Modeling

A prime factor analysis

Harald Störrle
hsbrrle@acm.org, @stoerrle

HuFaMo, Munich, 2019-09-16
Abstract

Modeling is an intrinsically human activity, often embedded in social contexts. Consequently, social sciences theories and research methods apply to many questions related to modeling. Owing to the cultural mismatch between social sciences and computing sciences, though, they are but a fringe phenomenon.

In this keynote talk, I will demonstrate the power and scope of human-centered research on modeling. I will present research projects on human factors in modeling inspired by observations from my work in industry and show what impact they can have in industrial practice. With this, I want to encourage and support the audience to use such methods more often and with greater confidence in the future.
<table>
<thead>
<tr>
<th>Lead Observation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question</td>
<td></td>
</tr>
<tr>
<td>Research Method</td>
<td></td>
</tr>
<tr>
<td>Trick</td>
<td></td>
</tr>
<tr>
<td>Faults</td>
<td></td>
</tr>
<tr>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>Insights</td>
<td></td>
</tr>
<tr>
<td>Publication</td>
<td></td>
</tr>
<tr>
<td>Benefit</td>
<td></td>
</tr>
</tbody>
</table>
Querying, Constraints, and Transformations, visually
All of UML, and beyond
Implemented
Table 4.1 Main publications on VM* and its precursors. The column “Intent” is used to classify the primary purpose of each publication, with values ranging from Query (Q) to Transformation (T). The column “Type” distinguishes between Conference (C) and Journal (J) publications. The column “Ref” is a reference number for each publication. The “Method” column indicates whether the study was based on a query (QE), transformation (TA), or model manipulation (M). The “Participants” column shows the number of participants. The “Languages” column lists the languages used in the study. The “Mode” column indicates whether the study was conducted in a single mode (S) or multiple modes (M). The “Year” column lists the year of publication.

Table 4.4 Main empirical studies evaluating VM* and its precursors. The column “Method” is used to classify the study methodology, with values ranging from Experiment (E) to Quasi-Experiment (Q). The column “Participants” shows the number of participants in the study. The “Languages” column lists the languages used in the study. The “Mode” column indicates whether the study was conducted in a single mode (S) or multiple modes (M). The “Year” column lists the year of publication.

QAware | 8/13
<table>
<thead>
<tr>
<th><strong>Lead Observation</strong></th>
<th>Model querying is important, but difficult. Having to learn a second (complex) query language is a poor business proposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question</strong></td>
<td>My (new) model query language is cool. Isn't it?</td>
</tr>
<tr>
<td><strong>Research Method</strong></td>
<td>Experiments comparing OCL v. VMQL, then NLMQL, then OCL+</td>
</tr>
<tr>
<td><strong>Trick</strong></td>
<td>Exploit students</td>
</tr>
<tr>
<td><strong>Faults</strong></td>
<td>poorly planned &amp; executed experiments</td>
</tr>
</tbody>
</table>
| **Findings**          | VMQL is not the best approach.  
                        | OCL offers very poor usability.  
                        | Substantial differences between students and practitioners. |
| **Insights**          | The syntactic and the conceptual dimensions are independent, and the syntactic one is not decisive.  
                        | Motivation and perseverance make a massive difference.  
                        | Professional engineers are a lot better than the best students. |
| **Publication**       | VL/HCC 2009/2010, JVLC 2011, …, Book chapter 2019 |
| ** Benefit**          | Very uncomfortable for many colleagues |
Model Usage
Study Design

Sampling
- convenience
- snowballing

Questionnaire Flow
- senior practitioners
- other participants

Question Types
- five closed questions
- two open questions

Questionnaire
- Consent Instruction Feedback
- Cultural Background
- Industrial Experience
- Modeling Experience
- Modeling in Software Developm.
- Modeling Opinions

Online Survey (descriptive, cross sectional)
- 65 Questions
  - 38 senior, industry
  - 58 junior/academia

Answers
- Demographics
  - age, education, experience, cultural background

- Modes of Modeling
  - Modeling languages used ways and scenarios of modeling

- Benefits & Beneficiaries
  - Benefits gained from and beneficiaries of modeling

- Opinions on Modeling
  - MDE endorsement, meaning of modeling
Models are primarily used for communication and cognition, not so much for code generation.
### Three modes of modeling

<table>
<thead>
<tr>
<th></th>
<th><strong>Informal models</strong></th>
<th><strong>Partially formal models</strong></th>
<th><strong>Fully formal models</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>support thinking and communication, utilizing rich information implicit in the situational context. Capturing contextual information in a model is not just no improvement, but effectively damages the usefulness of a model in such settings.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>support design and documentation activities. Here, more detail must be included and greater precision must be exacted such that the model can stand for itself, outside a given situational context.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>are to be taken literal and binding, so as to allow the analysis of system properties, simulation, and generation of code and test cases. Fully formal models can also be used like legal documents such as contracts, or other formalized agreements.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Beneficiaries of modeling

Who benefits how much from modeling?

<table>
<thead>
<tr>
<th>Role</th>
<th>None</th>
<th>Little</th>
<th>Some</th>
<th>A lot</th>
<th>Crucial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Architects</td>
<td>7%</td>
<td>16%</td>
<td>28%</td>
<td>44%</td>
<td>91%</td>
</tr>
<tr>
<td>Developers &amp; Testers</td>
<td>9%</td>
<td>24%</td>
<td>37%</td>
<td>36%</td>
<td>63%</td>
</tr>
<tr>
<td>Domain Experts</td>
<td>14%</td>
<td>37%</td>
<td>35%</td>
<td>36%</td>
<td>39%</td>
</tr>
<tr>
<td>Requirements Analysts</td>
<td>24%</td>
<td>37%</td>
<td>35%</td>
<td>36%</td>
<td>13%</td>
</tr>
<tr>
<td>Project Managers</td>
<td>29%</td>
<td>37%</td>
<td>35%</td>
<td>36%</td>
<td>39%</td>
</tr>
<tr>
<td>Clients</td>
<td>60%</td>
<td>37%</td>
<td>35%</td>
<td>36%</td>
<td>39%</td>
</tr>
<tr>
<td>End Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Benefits of modeling

- Modeling helps me deliver software with higher quality: 93%
- Modeling was important when it was applied: 86%
- Modeling should be used in more projects: 80%
- Modeling helps me deliver software with less effort: 72%
- Modeling helps me deliver software faster: 66%
- Modeling helps me react faster to client or market demands or changed requirements: 61%
- I share the vision of generating complete applications from models: 49%
<table>
<thead>
<tr>
<th><strong>Lead Observation</strong></th>
<th>Academic perception is ludicrously warped and self-centered: Few people in industry generate code, but everybody draws sketches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question</strong></td>
<td>How are models used in industry?</td>
</tr>
<tr>
<td><strong>Research Method</strong></td>
<td>Survey among practitioners</td>
</tr>
<tr>
<td><strong>Trick</strong></td>
<td>a tour with talks in regional ACM chapters or similar venues, advertise survey at the end</td>
</tr>
<tr>
<td><strong>Faults</strong></td>
<td>small n, recruiting bias, regional/cultural bias</td>
</tr>
<tr>
<td><strong>Findings</strong></td>
<td>3 distinct modes, sharply separated</td>
</tr>
<tr>
<td><strong>Insight</strong></td>
<td>None - we knew this before. Fowler wrote it in &quot;UML distilled&quot; in 1998. Proof was dearly needed.</td>
</tr>
<tr>
<td><strong>Publication</strong></td>
<td>EASE 2017</td>
</tr>
<tr>
<td><strong>Benefit</strong></td>
<td>If models really are mostly used for communication, maybe they should be studied from a linguistic viewpoint?</td>
</tr>
</tbody>
</table>
Linguistic Analysis
Splitting a diagram into layers conveys a narrative with as much information as the model proper.
Several usage scenarios for layers are common

**Orthogonal Aspects**
- Compile viewpoints/opinions
- Define features/modes

**Alternative Parts**
- Juxtapose variants
- Isolate special cases

**Consecutive Stages**
- Explain domain stepwise
- Specify release plan
Qualitative Validation

1. Expert assessment: Presented idea to 22 experienced modelers from academia and industry
   • Ensure validity & generalizability of earlier findings.
   • Unanimously positive feedback, potential benefit of layers was obvious and directly applicable in their respective fields in previous and current projects.
   • Some proposed more (new) usage scenarios
2. Field testing: Modeling in the context of several courses at different academic levels.
   • ~10 teams of 4-6 students tasked with UI design.
   • Students picked up concept very quickly (no intro required) and invented new usage scenarios on the fly.
   • Unanticipated usage modes were invented on the fly (e.g., using layers topographically to split diagram into sectors by responsibility).
3. Recent (new) field application
   • Showed the paper to a colleague who went on to use layers (poor man's style) to great effect.
Context

One-Click™ credit approval

OnlineCredit

- provide client & credit details
- compute score
- compute credit rate

Client

One-Click credit approval on-line opens a new market segment for Mutual 1886. It helps stabilize profits for the next 10 years.

Legend

System
(application providing services)
Use Case
(service provided by application)
Actor
(person interacting with service)

Project
Credit-NxtGen

Status: QA-approved
Stage: Design
Author: Warren, D.H.
Date: 12/07/2016

Type and title of a diagram set the scene.

Vignettes add texture and meaning to the scenario.

A legend makes the notation more accessible and invites a more diverse set of stakeholders.

Traditional plan headings show administrative information helping to keep track of the multitude of documents in a project. Including client logo may increase buy-in.
Modalities and Moods

- Natural languages offer rich tools to express varying degrees of reality.
  - Linguistics collectively refers to such phenomena as (grammatical) **moods** signaling epistemic **modality**.
  - English has the moods indicative, imperative, and subjunctive; there are languages with up to 16 moods.
  - As the usual modeling languages don’t have moods, all model elements have the same epistemic status.

- Thus, in UML, BPMN etc., it is not possible to express statements like the following **as part of the model**.
  - “This use case exists now, and that one will exist after the next release.”
  - “I’m not sure about this class. Maybe it should be split up into two classes?”
  - “This DB column is decided upon, that one is still up in the air.”
  - “This message must not be sent.”

- In practical modeling, however, such situations are ubiquitous.
  - We typically add spoken texts (the “voice track”), possibly even or comments in the model
  - Such additional information easily gets lost, and cannot be exploited formally.
Scenario 1: Purposes "descriptive" vs. "prescriptive"

- Alex wants to improve an existing booking process:
  - documents existing process, complete with weaknesses and omissions,
  - describes extensions and changes to be implemented.
- There is a modality "Purpose" with two different levels "descriptive" and "prescriptive".
- Combining both moods in one model is common and effective.
  - Using two complete separate models that uniformly have just one mood is possible,
  - but combining two modalities in one model is cheaper and faster.
Scenario 2: Certainty: "certain" vs. "uncertain" knowledge

- During prescriptive as well as descriptive modeling, open questions arise.
  - Challenging assumptions and discovering gaps is the point of modeling, after all.
- Model elements may have different degrees of certainty.

- Theoretically, certainty is a degree (e.g., percentage), but practically, two or three levels are more than sufficient.
  - Both levels are needed in the model.
  - Restricting the model to one of the two levels omits important information.

- Moods like "purpose" and "certainty" are independent of each other, so it must be possible to model all their combinations.
Do modalities exist, and which are most common?

- All of the scenarios described before can be found in practical modeling.
- Judging by the models I have seen since 2000, mood is a very common phenomenon, but it is rarely expressed inside the model.
- Instead, mood is often conveyed as part of the "soundtrack", a jargon term for the oral narrative accompanying a model presentation.
- Inside a model, modelers would use comments, annotations, or graphical styling to indicate mood.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Levels</th>
<th>Prevalence</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td><strong>Descriptive</strong>, prescriptive</td>
<td>●●●●●</td>
<td>model portrays something in existence or something to be created</td>
</tr>
<tr>
<td>Certainty</td>
<td>Certain, uncertain</td>
<td>●●●</td>
<td>knowledge represented by model or model element is certain or not</td>
</tr>
<tr>
<td>Finality</td>
<td>Final, ongoing</td>
<td>●●</td>
<td>modeling of element is completed or not</td>
</tr>
<tr>
<td>Attitude</td>
<td>Positive, negative</td>
<td>●●</td>
<td>element is supposed to be there or not be there</td>
</tr>
<tr>
<td>Entanglement</td>
<td>None, rely, conflict</td>
<td>●</td>
<td>element’s status and/or existence depends on other elements</td>
</tr>
<tr>
<td>Fidelity</td>
<td>Low, ..., high</td>
<td>●</td>
<td>degree of detail of a model relative to the original</td>
</tr>
</tbody>
</table>
How to express modalities?

Legend: non-default modalities
- High Fidelity (element with details attached)
- Uncertain (no consensus about element existence)
- Dependent (depends on other parts of model)
Test it in reality

Diagram editor
(Diagram elements are independent of model elements)

Model element editor
(Each element has a form for all the details)

FIDELITY describes the level of detail of a model element

CERTAINTY captures the reliability of the information represented in the model

ID, Name, and basic properties

Textual use case description

Pre- and Postconditions

Other descriptive aspects
<table>
<thead>
<tr>
<th><strong>Lead Observation</strong></th>
<th>Linguistic categories might apply to UML et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question</strong></td>
<td>What linguistic phenomena are present in models, but ignored by research?</td>
</tr>
<tr>
<td><strong>Research Method</strong></td>
<td>Explain phenomenon, provide examples, elaborate concepts</td>
</tr>
<tr>
<td><strong>Trick</strong></td>
<td>Take a new standpoint outside your field, even if it feels weird at first.</td>
</tr>
<tr>
<td><strong>Faults</strong></td>
<td>There is no representative body of models (&quot;corpus&quot;) as there is in linguistics.</td>
</tr>
<tr>
<td><strong>Findings</strong></td>
<td>Several such phenomena exist in models (narrative structure, moods, context, implicature, metaphor)</td>
</tr>
<tr>
<td><strong>Insight</strong></td>
<td>As modelers are humans, they imprint their communication methods on any medium,</td>
</tr>
<tr>
<td><strong>Publication</strong></td>
<td>various small papers starting 2014, MiSE 2016, FlexMDE 2019, …</td>
</tr>
<tr>
<td><strong>Benefit</strong></td>
<td>establish notions, raise interest, pave ground for a theory of &quot;communication with models&quot; (cd. Petri's PhD-thesis).</td>
</tr>
</tbody>
</table>
Layout of Diagrams
Good and Bad Diagrams

- Here are two different diagrams of the same model.
- Obviously, the diagram on the left has a better layout than the diagram on the right.

- But exactly why is this the case?
- And just how good and bad are they? How much better is the good one?
- To answer these questions, we need an objective, repeatable, and practical metric for diagram (layout) quality.
- Also, our previous definition of diagram size was flawed in that it contained aspects of quality.
Study Design

**Intervention**
(Independent Variables)

- Diagram Type
  - [UCD, AD, SM, SQD, CD]
- Diagram Size
  - [#Elements]
- Diagram Quality
  - [#Flaws]

**Experiments A-F**
(within subject, randomized)

- 60 Diagrams

**Observation**
(Dependent Variables)

- Modeler Performance
  - Test Score [0..10]
  - Test Score Variation [0..10]

+ subjective assessment, follow-up questions
Diagram layout has a significant impact on diagram (and model) understanding.

Accuracy
- a: correct, b: wrong or missing

Response time
- e: per answer, f: per correct answer

Preference
- c: quality, d: clarity

Individual Score

Average Score per Sheet

Legend:
- + = good layout
- - = bad layout

D (BEng)
E (MSc)
F (Elite)
Figure 5: Distribution of diagram sizes per diagram type: the bottom/grey bars show numbers of elements, the top/red bars show number of layout flaws per diagram. The boxplots to the right show distribution of elements and flaws, respectively, in total and by diagram. The box with index G refers to [13].
Towards a theory

<table>
<thead>
<tr>
<th>Layout Level</th>
<th>Governing Principles</th>
<th>Variation Points</th>
<th>Layout Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - Pragmatics</td>
<td>Modeler Intent</td>
<td>Narrative</td>
<td>convey message to target diagram to audience, realize implicature</td>
</tr>
<tr>
<td>2 - Layout</td>
<td>Gestalt Laws</td>
<td>Flow, Grid, Symmetry</td>
<td>exhibit global structure through symmetric, regular, or ordered arrangement, visual flow</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Topology</td>
<td>avoid local mistakes of intersecting, overlapping, and touching elements, line bends</td>
</tr>
<tr>
<td>1 - Graphics</td>
<td>Psychophysics</td>
<td>Bertin-Variables</td>
<td>reduce noise from uniform visual style of color, texture, direction, size, ... of elements</td>
</tr>
</tbody>
</table>

... but does this show in the actual behavior?
Point of reference / Replication
Validation of previous studies by repeating the same experiment, with the same (subjective) measurements on a sub-sample of previously applied stimuli.
<table>
<thead>
<tr>
<th><strong>Lead Observation</strong></th>
<th>Client paid for it</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question</strong></td>
<td>Does diagram layout improve model understanding? How much? Which factors/demographic? What are suitable metrics? What are cognitive processes?</td>
</tr>
<tr>
<td><strong>Research Method</strong></td>
<td>series of large scale student experiments, major eye tracking study</td>
</tr>
<tr>
<td><strong>Trick</strong></td>
<td>be popular as a teacher and your students will volunteer to help you</td>
</tr>
<tr>
<td><strong>Faults</strong></td>
<td>need more input from cognitive psychologists</td>
</tr>
<tr>
<td><strong>Findings</strong></td>
<td>Size matters, diagram type doesn't. Expertise matters, experts have distinct behavior. Mechanical metrics for diagram size &amp; quality,</td>
</tr>
<tr>
<td><strong>Insight</strong></td>
<td>Layout is a massive factor. Findings are actionable - nudging is sufficient for improvement.</td>
</tr>
<tr>
<td><strong>Benefit</strong></td>
<td>Reference point, enough evidence to support theory</td>
</tr>
</tbody>
</table>
MDE Adoption
Interview workflow

**Activity** | **Effort** | **Caveat**
---|---|---
recruit, prepare | 1.2h | spreads over weeks
interview | 1.5h | plus travel time
proceed notes | 0.5h | do immediately
process recording | 0.25h | little bit of learning
transcribe | 9h | do soon after
translate | 6h | avoid if possible
encode | 15h | do soon
re-encode | 5h | one pass per interview
Total (per interview) | ~40h | per interview
Total (10 interviews) | ~450h | per study

**Admissible tricks:** spread work over consortium, use commercial transcription, group similar activities

**Contingent tricks:** predefined code tree, no translation, spread out over time (piggy back) => may deteriorate results

**Dirty tricks:** no transcription/encoding (just notes and quotes) => no pain, no gain
Study Summary

- We have conducted an extensive interview study regarding how and to what extent models are used in different industries.

Areas of Expertise
- Automotive
- Aerospace
- Medical Devices
- Telecom
- Banking & Finance
- Insurance & Pension
- Public administration
- Military, Consumer, Retail, Logistics

Interviews
- 25 Participants
- 20 Interviews
- 18 Organizations
- 7+ Industries

Observations
- Audio Recording (24h 41min.)
- Transcript (260k words)
- Field Notes (20 pages)

Encoding
- open coding
- selective coding
- conceptual saturation
- code memos
- recoding
- code tree refactoring
- axial coding

Theoretical Sampling

- It appears that different industries exhibit different patterns of MDSD adoption, governed by industry specific economic drivers.
- The official MDE claims (productivity, ...), are irrelevant, though.
<table>
<thead>
<tr>
<th>Lead Observation</th>
<th>Academic perception is ludicrously warped and self-centered: Few people in industry generate code, but everybody draws sketches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question</td>
<td>How are models used in industry?</td>
</tr>
<tr>
<td>Research Method</td>
<td>Interview campaign among practitioners</td>
</tr>
<tr>
<td>Trick</td>
<td>use your industry contacts, spread the word, use all kinds of journeys to piggy back another benefit</td>
</tr>
<tr>
<td>Faults</td>
<td>Grounded theory is (yet) too far out of SE mainstream</td>
</tr>
<tr>
<td>Findings</td>
<td>Decisive factor is not UX, technology, or scientific maturity, but business factors that differ by industry, region, culture.</td>
</tr>
<tr>
<td>Insight</td>
<td>It's the economy, stupid.</td>
</tr>
<tr>
<td>Publication</td>
<td>Not ICSE'18, EASE'19, or MODELS'19 :-(</td>
</tr>
<tr>
<td>Benefit</td>
<td>none yet. If published: tons of interesting new questions. A theory of technology adoption substantially more realistic than TAM.</td>
</tr>
</tbody>
</table>
Wrapping up
Learnings

- As far as modeling is concerned, I have learned that:
  - People use diagrams for communication, linguistic analysis applies.
  - For practical relevance, only the practitioners' voice counts.
  - Modeling is not an important topic, globally speaking.
  - In terms of maturity, this modeling community lag behind general SE, which lags behind Empirical SE.

- Creating a language, an algorithm, a tool is engineering at best, but not science.
- Science is a curious observation, followed by systematic application of suitable scientific methods.

- Different research methods offer different benefits:
  - Insight may be generated by qualitative methods;
  - Certainty may be generated by experimental methods.

- A (senior) researcher needs proficiency in multiple methods, and acquaintance with multiple viewpoints.
Tutorial

"Qualitative Research Methods in Modeling"

This afternoon, here at MODELS