A Systematic Review of the Impact of Model-Based Testing on Code Coverage

August 11, 2020

Protocol

Team: Sten F. Andler (University of Skövde, Skövde, Sweden); Fabiano C. Ferrari (Federal University of São Carlos, São Carlos, Brazil); Jeff Offutt (George Mason University, Fairfax, USA); Mehrdad Saadatmand (RISE SICS, Västerås, Sweden); Nils Muellner (German Aerospace Center, Germany); Vinicius H. S. Durelli (Federal University of São João del Rei, São João del Rei, Brazil).

Primary study selection: Fabiano C. Ferrari, Sten F. Andler, Vinicius H. S. Durelli.

1 Description

This document consists in a protocol to be followed over the course of the conduction of our systematic literature review. To establish this protocol, we followed the guidelines for conducting secondary studies proposed by Kitchenham et al. (2015) and Wohlin (2014). The next sections describe how we followed these guidelines to answer the research questions posed by this study.

2 Question Formulation

The research questions (RQs) emphasize the creation of an overview of the literature in a way that is interesting to researchers and practitioners and also gives them insights into several aspects of model-based testing, *e.g. (i)* test case generation at model level and *(ii)* the mapping of model-level test suites to code-level test suites. The scope and goal of our study can be better formulated through the Goal-Question-Metric approach (Basili et al., 1994) as follows.

> **Analyze** the state of the art in model based testing
> **for the purpose of** exploration and analysis
> **with respect to** the intensity of the research in the area, how source code coverage can be gauged from test suites generated using model based testing approaches, hindrances to turning model-level test suites into code-level test suites, trends, to what extent test generation at model level has been empirically evaluated, and most widely used model representations.
> **from the point of view of** researchers and practitioners
> **in the context of** software testing.

As pointed out by Kitchenham et al. (2015), research questions (RQs) must embody the goal of secondary studies. Thus, the goal of our study can be broken down into three main RQs and a sub-question:

- RQ1: *How are test suites that are developed at the model level mapped to the code level; code which may or may not be created by automatic transformation?*
  - RQ1.1: *What is required of the model-to-code transformation to allow the mapping to the tests?*
- RQ2: *How are tests generated from the model specifications (e.g. UML or Simulink)?*
- RQ3: *How does coverage of the model by abstract tests relate to coverage of the code for the corresponding concrete tests?*

Keywords (to be expanded):

- Group 1: “model-based testing”; MBT; “model-to-text”; “model-to-code”; “model-to-model”; “executable model”;
- Group 2: “code coverage”; “structural coverage”; “statement coverage”; “branch coverage”; “decision coverage”; “data flow coverage”

Base search string: not applicable; the study selection is based on the snowballing method (Wohlin, 2014).

Intervention: Proposition or application of model-based testing to executable models.

Control: Not applicable (a start set of studies will be defined; more details in Section 3 of this protocol).

Effect (Results): A characterization of the current state-of-the-art including how source code coverage can be gauged from test suites generated using model based testing approaches, hindrances to turning model-level test suites into code-level test suites, trends, what extent test generation at model level has been empirically evaluated, and most widely used model representations.

Outcome Measure: Not applicable.

Population: Software testing literature on model-based testing.

Application: Researchers and practitioners interested in model-based testing.

3 Selection of Sources

3.1 Criteria for Selection of Sources

The set of sources for primary studies naturally grows as long as the snowballing-based search evolves. Traditional digital libraries that publish reputable international journals and a number of relevant software engineering related conferences, symposia, and workshops proceeding are expected to compose our set of sources. Examples are the ACM Digital Library,[1] IEEE Xplore,[2] Elsevier ScienceDirect,[3] Springer SpringerLink,[4] and Wiley Online Library.[5]

3.2 Identification of Sources

The snowballing technique will be performed manually, by analyzing lists of references of selected studies, as well citations to selected studies. The set of study sources will evolve in parallel with the set of selected studies.

4 Selection of Studies

This section defines the inclusion and exclusion criteria that will be used during the conduction of this secondary study.

[1] http://dl.acm.org – accessed on February, 2020.
[2] http://ieeexplore.ieee.org/Xplore/home.jsp – accessed on February, 2020.
[3] http://www.sciencedirect.com – accessed on February, 2020.
[4] http://link.springer.com – accessed on February, 2020.
[5] https://onlinelibrary.wiley.com – accessed on February, 2020.

4.1 Inclusion (I) Criteria

I1: The study proposes/applies model-based testing for/to models.

I2: The study addresses automatic model-to-code (or model-to-text) transformation.

I3: The study addresses the mapping from test suites developed at model level to source code level.

I4: Our selection relies on the quality filtering performed by the peer-review process, so all studies must have undergone peer-review to be eligible to be selected. We take into account only studies published in scholarly venues such as conference, symposium, and workshop proceedings and scientific journals.

A study is selected if it passes ($I1$ AND $I2$ AND $I4$) or ($I3$ AND $I4$). More details about how these criteria are applied are provided in Section 4.4.

4.2 Exclusion (E) Criteria

Studies that fall into at least one of the following categories will not be selected.

E1: Studies that emphasize hardware testing.

E2: Secondary studies.

E3: Peer-reviewed studies that have not been published in journals, conferences, symposia, or workshop proceedings (*e.g.* Ph.D. theses and technical reports).

E4: Studies that are not written in English.

4.3 Types of Studies

In the context of this literature review, we are interested in studies that fall into one of the following categories: methodological, experimental, characterization study, and position papers.

4.4 Study Selection Process

During study selection, we apply the inclusion and exclusion criteria. In hopes of avoiding spending too much time analyzing papers that are clearly irrelevant to our literature review, we initially filter papers based on title and abstract. However, it may be the case the selection requires a more thorough analysis of the studies in order, for instance, to determine whether a paper to some extent describes model to code (or model to lower level model transformations). Therefore the criteria $I1$, $I2$, and $I3$ are not applied during the first selection step. Additionally, $E1$ is not applied during the first selection step because applying such a criterion requires a more in-depth examination of the studies: often, title and abstract are not enough to pin down the content of a paper. During the second step of the search process, one reviewer read the candidate papers in their entirety. Indecision on whether a paper should be selected or not is resolved by discussion with, at least, two other reviewers.

This study selection process follows the guidelines for performing SLR using the snowballing search method (Wohlin, 2014; Kitchenham et al., 2015). Snowballing, which is also referred to as citation analysis, is a literature search method that can take one of two forms: backward snowballing and forward snowballing (Kitchenham et al., 2015). Backward snowballing consists in starting the search from a set of papers that are known to be relevant (*i.e.* an initial set of selected studies): basically, it involves searching the references sections of the studies. Forward snowballing entails finding all studies that cite a study from the set of selected studies. Both search methods update the set of selected studies in an iterative fashion; only the studies included in the previous step are taken into account in each search iteration, and both backward and forward snowballing end when no new papers are found in the search iterations.

In our SLR, we planned the following snowballing steps:

Step-1: Define the ***start set*** (S_S): the initial set of studies, upon which snowballing will develop, must be defined to conform with the following properties:

- It must include studies from different communities (to cover independent clusters).
- It must cover different publishers, years and authors.
- It should be of relatively medium size (depending on the breadth of the surveyed area).
- It must match the predefined set of keywords and their synonyms (to allow varied terminology).

At the end of this step, the S_S becomes the ***current set*** (S_C) of selected studies.

- For this SLR, S_C is composed by the studies of Briand et al. (2016), Camus et al. (2016), and Eriksson and Lindström (2016).

Step-2: Perform *backward snowballing* over the current set (S_C):

(a) Perform the pre-selection of studies of interest:

i. For each study s_i in S_C, scan its reference list to identify studies of interest; this includes analyzing the place studies are referenced in s_i (in the background section, related work section, and experimental setup section) and the type of the referenced studies (*i.e.* whether they are peer-reviewed or not);

ii. Apply the inclusion and exclusion criteria to predefined parts of the identified studies (*i.e.* title, abstract and keywords).

At the end of this step, the pre-selected studies constitute the ***pre-selected backward delta set*** ($S_{PB\Delta}$), that is, a set that will potentially append S_C.

(b) Perform the final selection, based on the full reading of studies in $S_{PB\Delta}$.
At the the of this the step, the selected studies constitute the ***backward delta set*** ($S_{B\Delta}$), that is, a set that will append S_C.

Step-3: Perform *forward snowballing* over the current set (S_C):

(a) Perform the pre-selection of studies of interest:

i. For each study s_i in S_C, identify studies of interest that cite s_i; this includes analyzing the place the citation occur (in the related work section and/or experimental setup sections) and the type of the studies (*i.e.* whether they are peer-reviewed or not). The Google Scholar[6] automated search mechanisms is used to support this step.

ii. Apply the inclusion and exclusion criteria based on the predefined parts of the identified studies (*i.e.* title, abstract and keywords).

At the end of this step, the pre-selected studies constitute the ***pre-selected forward delta set*** ($S_{PF\Delta}$), that is, a set that will potentially append S_C.

(b) Perform the final selection, based on the full reading of studies in $S_{PF\Delta}$.
At the the of this the step, the selected studies constitute the ***forward delta set*** ($S_{F\Delta}$), that is, a set that will append S_C.

Step-4: Update S_C:

(a) Create the ***delta set*** (S_Δ) as the union of $S_{B\Delta}$ and $S_{F\Delta}$.

(b) Append S_C with S_Δ.

(c) Start a new iteration from Step 2, only considering studies included in S_Δ.

General Procedures:

- In any selection step, duplicated entries must be discarded.
- Up-to-date studies must replace prior versions of the same study in S_C. Examples are studies that update a technique previously published, or studies that extend a prior publication. In these cases, the replacing study must become part of S_Δ for a new iteration. Studies removed in the replacement process as said to be *subsumed* by more recent studies.

[6] http://scholar.google.com/ – accessed on February, 2020.

5 Data Extraction

To answer the RQs described in Subsection 2, we will extract from primary studies the information outlined in the data extraction form. Subsection 6.1 lists the extraction fields. Beyond the fields that we will use to answer the RQs, our data extraction form includes fields designed to gather general information about the primary studies, *e.g.* title and year of publication. Before starting the review, the data extraction forms will be revised by all involved reviewers.

After extracting all data from all selected studies, the three reviewers in charge of primary study selection will check all extracted data to make sure the data is accurate and ready for further analysis.

6 Additional Information

6.1 Main Data Extraction Fields

- The general goal of the study
- A description of the study from the perspective of each research question
- The main results of the study
- The conclusion of the study, *cf.* the original authors
- The conclusion of the study, *cf.* the reviewers
- The target specification language (at model level)
- The target programming language (at code level)
- The tool used for model-to-text transformation (for the main software artifacts)
- The tool used for automatic test case generation (at the model level)
- The tool used for test suite transformation (from model to source code)
- The obtained code coverage obtained with model-based test suites
- Level of automation for model-based test generation
- Level of automation for test re-execution (model → code)
- Level of traceability of model elements → code elements
- Level of automation for traceability of model elements → code elements:

References

Basili, V. R., Caldiera, G., and Rombach, H. D. (1994). The Goal Question Metric Approach. In *Encyclopedia of Software Engineering*. Wiley.

Briand, L., Nejati, S., Sabetzadeh, M., and Bianculli, D. (2016). Testing the Untestable: Model Testing of Complex Software-intensive Systems. In *Proceedings of the 38th International Conference on Software Engineering (ICSE) - Visions of 2025 and Beyond Track*, pages 789–792, Austin, TX, USA. ACM.

Camus, J.-L., Haudebourg, C., Schlickling, M., and Barrho, J. (2016). Data Flow Model Coverage Analysis: Principles and Practice. In *Proceedings of the 8th European Congress on Embedded Real Time Software and Systems (ERTS)*, pages 1–10, Toulouse, France. Centre pour la Communication Scientifique Directe.

Eriksson, A. and Lindström, B. (2016). UML Associations: Reducing the Gap in Test Coverage Between Model and Code. In *Proceedings of the 4th International Conference on Model-Driven Engineering and Software Development (MODELSWARD)*, pages 589–599, Rome, Italy. IEEE.

Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). *Evidence-Based Software Engineering and Systematic Reviews*. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series.

Wohlin, C. (2014). Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, pages 1–10, London, UK. ACM.