

**lero**

THE IRISH SOFTWARE  
ENGINEERING RESEARCH CENTRE

# Systematic Literature Reviews: Theory and Practice

M. Ali Babar



## *What are Systematic Reviews?*

- A SR is a concise summary of the best available evidence that uses
  - explicit and rigorous methods to
  - identify, critically appraise, and synthesize relevant studies on a particular topic.
- The process and procedures are defined in advance and are documented so that others can critically appraise and replicate the review.



## *Traditional vs. systematic reviews\**

<b>Feature</b>	<b>Traditional review</b>	<b>Systematic review</b>
Question	Often broad in scope	Often a focused research question
Identification of research	Not usually specified, potentially biased	Comprehensive sources and explicit search strategy
Selection	Not usually specified, potentially biased	Criterion-based selection, uniformly applied
Appraisal	Variable	Rigorous critical appraisal
Synthesis	Often a qualitative summary	Qualitative and/or quantitative synthesis
Inference	Sometimes evidence-based	Usually evidence-based



# *Systematic Review Process*

- Planning the review
  - Identification of the need for a review
  - Development of a review protocol
- Conducting the review
  - Identification of research
  - Selection of primary studies
  - Study quality assessment
  - Data extraction
  - Data synthesis
- Reporting the review



## *Advantages*

- SRs create a solid foundation for future research
  - Close areas where no further research is necessary
  - Uncover areas where research is necessary
- They can contradict “common knowledge”
- They help the development of new theories

## *Disadvantages*

- SRs require more effort than informal reviews
- Difficult for lone researchers
  - Standards require two researchers to minimize individual bias
- Incompatible with requirements for short papers



# *SLR on Variability Management in Software Product Line Engineering*



## *Objective And Research Questions*

- The objective of this study is to review the status of evaluation of VM approaches in Software Product Line Engineering.
- The specific research questions are:
  - How have the variability management approaches in SPLE been evaluated?
  - What is the quality of the reported evaluations of the variability management approaches?



# Search Strategy

- Search strings were constructed using the following strategy:
  - Derive main terms based on the research question and the researched topics
  - Determine and include synonyms, related terms, and alternative spelling for major terms
  - Check the keywords in all relevant papers researchers already knew, and initial searches on the relevant databases
  - Incorporate alternative spellings and synonyms using Boolean “or” and
  - Link main terms using Boolean “and”
  - Pilot different combinations of the search terms
- Search string
  - software AND (product line OR product lines OR product family OR product families) AND (variability OR variation OR variant)





## *Data Sources*

- Databases
  - IEEExplore;
  - ACM Digital library;
  - Citeseer library (Google);
  - ScienceDirect;
  - EI Compendex / Inspec;
  - SpringerLink; and
  - Web of Science
- Proceedings of the SPLC conference series
- Non-reviewed literature
  - SEI's technical reports on SPL



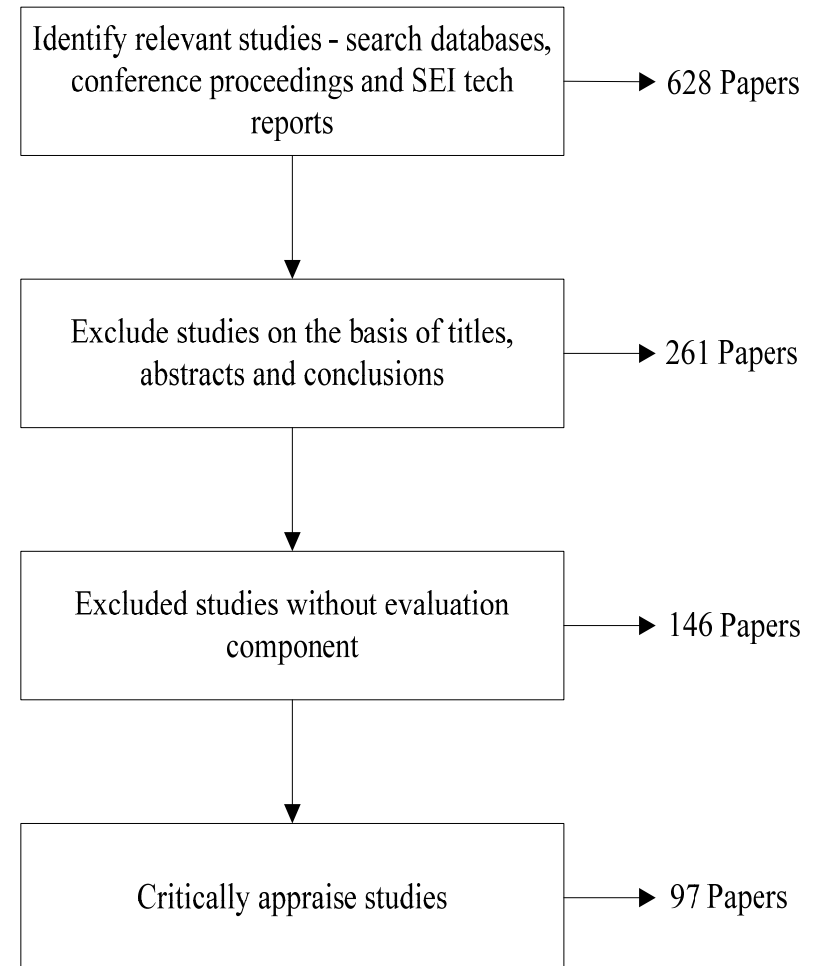
## *Time Span*

- We did not restrict our search based on publication year. We performed the search in September 2007. That means the papers published after that date were not included in this study



# Study Selection

- A paper was included if it:
  - introduces an approach to dealing with some aspect of VM in SPLE or;
  - reports an evaluation of an existing VM approach.
- The paper was excluded if it:
  - does not deal with VM in SPLE.
  - does not include an evaluation of a VM approach.
  - is a short paper.





# Quality Assessment

- Type 1: intends to assess the quality of the studies with respect to their ability and suitability to answer our research questions, and with respect to the impact on the drawn conclusions.
  - Treats each paper equally assuming that each of them was of sufficient quality as they were published in peer reviewed journals, conferences, and workshops; or by highly credible institute (i.e. SEI)
- Type 2: serves as an instrument to answer one of our main research questions.

1. Is the paper based on research (or is it merely a “lessons learned” report based on expert opinion)?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to address the aims of the research?
5. Was the recruitment strategy appropriate to the aims of the research?
6. Was there a control group with which to compare treatments?
7. Was the data collected in a way that addressed the research issue?
8. Was the data analysis sufficiently rigorous?
9. Has the relationship between researcher and participants been considered to an adequate degree?
10. Is there a clear statement of findings?
11. Is the study of value for research or practice?



## *Data Extraction*

- Each of the 146 primary studies was fully read for data extraction purposes
- 49 papers were excluded in this stage
- Data extraction attributes: reviewer's name, extraction date, title, authors, publication venue, publication year, publication source, research method used, variability management approach, evaluator, evaluation method, and industrial evaluation)
- Doubts were clarified/checked



## *Data Synthesis*

- Most of the selected primary studies were grounded in qualitative research, so a meta-analytical approach was not suitable for synthesizing the data
- Manually review and link the extracted data in the Excel Spreadsheet
- Then used descriptive statistics (e.g. sum, average) for analysing the data



# Results – Demographic Data

- Publication venues
  - Literature on VM is scattered in different publication venues
  - The premier events of the SPL community (i.e. SPLC, PFE) do not have clear dominance
  - There are 39 venues with only one study published
  - No study appeared in the Empirical Software Engineering Journal
- Type of publication
  - Conference: 75
  - Journals: 21
  - Tech report: 1
- Publication year
  - No study prior to 1990
  - 4 studies before 2000
  - From 2000 onwards we found an increased number of studies with a peak in 2004

## Channels with at least 2 studies

ID	Channel	#	%
1	SPLC	11	11.34
2	SCP	6	6.19
3	PFE	5	5.15
4	RE	5	5.15
5	ICSR	4	4.12
6	APSEC	3	3.09
7	ICSE	3	3.09
8	SERA	3	3.09
9	ECBS	2	2.06
10	ESEC/FSE	2	2.06
11	GPCE	2	2.06
12	IEE Proceedings -Software	2	2.06
13	ISESE	2	2.06
14	JSS	2	2.06
15	NODe	2	2.06
16	REFSQ	2	2.06
17	SPE	2	2.06
<b>SUM</b>		58	59.76



## Results – Approaches Proposed

- Identified 91 different approaches
- A large majority of the approaches are based on feature modelling and/or UML based techniques
- There are very few approaches (4, 4.40%) based on mathematical techniques
- A large majority of the VM approaches are quite amenable to empirical evaluation

### The kinds of VM approaches proposed

Nature of Solution	# of approach
Feature model	33
Using UML and its extensibility	25
Express variability as part of a technique that models the architecture of the system	8
Using natural language	6
Expressed variability as part of a technique that models the components of the system	5
Formal techniques based on mathematics	4
X-frames organized into a layered hierarchy	4
Domain specific language	3
Ontology based techniques	3
Solution from the perspective of Aspect-Orientation	2
Orthogonal Variability Management	2
Configuration management based Modelling	1
Using information visualization techniques	1





## Results – Evaluation Methods Used

- Example application was most frequently used
- Most studies did not follow scientific approaches to rigorously evaluate a specific technology
- 87 (96%) of the presented approaches were evaluated in only one study
- Only one study can be considered as an independent evaluation
- We did not find a clear trend of improvement
- A large majority of the VM approaches awaits rigorous empirical evaluation

### Different evaluation approaches used

Type of Evaluation	# of Studies	%
Example Application	57	58.76
Experience Report	17	17.53
Case Study	13	13.4
Discussion	4	4.12
Laboratory Experiment with Human Subjects	1	1.03
Simulation	1	1.03
Laboratory Experiment with Software Subjects	2	2.06
Field Experiment	1	1.03
Rigorous Analysis	1	1.03
SUM	97	100



## *Results – Evaluation in an Industrial Setting*

- The large majority of the reviewed approaches have never been evaluated in an industrial setting
- Among those approaches that have been tried in industrial settings, more than half of them were reported as experience reports

Industrial evaluation of the VM approaches reported in the papers included in this SR

Industrial evaluation/trial	# of Approaches	Percent
Not evaluated in industrial settings	65	71.43
Tried in industrial settings	26	28.57
<b>SUM</b>	91	100



## Results – Quality of Evaluation

- We decided not to apply the whole quality assessment criteria to the studies using Example Application, Experience Reports, Discussion and Rigorous Analysis
- This decision left us with only 18 studies to be assessed against the quality criteria
- Except the first, last criterion and the criterion regarding description of the context, there is a huge gap (scored around 0.60) that needs to be filled in order to improve the quality of the evaluation of the VM approaches.

### Quality assessment of 18 empirical studies

	1	2	3	4	5	6	7	8	9	10	11
	Research	Aim	Cxt	R.design	Sampling	Ctrl Grp	Data coll	Data anal	Reflex.	Findings	Value
OK (1)	18	7	17	3	6	3	7	5	0	5	18
Partial (0.5)	0	11	1	15	11	0	8	13	0	13	0
No (0)	0	0	0	0	1	15	3	0	18	0	0
AVG	1	0.6	0.9	0.58	0.64	0.17	0.61	0.64	0	0.64	1



## *Key Points*

*1/2*

- No rigorous evaluation of 80.41% of the approaches indicates a general lack of robust assessment of a large majority of the VM approaches
- Huge quality deficiencies in the majority of the reported studies on most (8 of 11, 72.73%) of the elements of the quality assessment criteria used
- A lack of number of studies carried out to evaluate a particular approach as 95.60% of the proposed approaches were evaluated by only one study
- Lack of industrial evaluation of the proposed approaches



## *Key Points*

*2/2*

- The dominance of some well-known publication channels is not that clear; thus, limiting the scope of search of primary studies to well-known publication channels may miss a large number of primary studies
- Some of the papers without an evaluation component were excluded, but they were highly cited, which may indicate a lack of appreciation of empirical studies in this community
- The papers in the area of VM in SPLE rarely reference literature on empirical research methodologies
- ESE community can be encouraged to work with SPLE community to improve the state of the practice of rigorously evaluating research outcomes



## *Limitations*

- Bias in selection of publications
  - We may have not found those studies whose authors might have used other terms for VM
  - We may have missed those approaches that have been commercialized but have not been reported in literature with an evaluation component
  - Quality of search engines could have influenced the completeness of the identified primary studies
- Inaccuracy in data extraction
  - Sometimes we had to infer certain pieces of the required information, because many articles lacked sufficient details about the design and execution of the reported studies
- Reliability of classifying the evaluation approaches reported and quality assessment
  - A third researcher checked the papers about which the primary researchers were in doubt or felt uncertain. All discrepancies were resolved through reviews of the papers and discussions.



## *Conclusion*

- The status of evaluation of VM approaches in SPLE is unsatisfactory
- Most of the approaches are amenable to empirical evaluation, but the available evidence is sparse and the quality of the presented evidence is very low
- The findings highlight the areas of improvement for rigorous evaluation of VM approaches
- Contributions
  - Help practitioners to correct misunderstanding and over interpretation of the reported evidence in the literature
  - The revelation of paucity of empirical evidence and the scattered distribution of papers over a large number of publication venues can be useful information for researchers who are going to do literature review in this area



## *A Few Relevant Papers*

- J.P.T. Higgins and S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.0* (updated Feb. 2008), The Cochrane Collaboration, 2008.
- B.A. Kitchenham, *Guidelines for performing Systematic Literature Reviews in Software Engineering, Version 2.3*, Keele University and University of Durham, EBSE Technical Report, 2007.
- G.W. Noblit and R.D. Hare, *Meta-Ethnography: Synthesizing Qualitative Studies*, Thousand Oaks: Sage, 1988.
- M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*, Oxford, UK: Blackwell, 2006.
- L. Chen, M. A. Babar, and N. Ali, "Variability Management in Software Product Lines: A Systematic Review," in the 13th International Software Product Line Conference, San Francisco, CA, USA, 2009.
- L. Chen, M. A. Babar, and C. Cawley, "A Status Report on the Evaluation of Variability Management Approaches," in 13th Int'l Conf. on Evaluation and Assessment in Softw. Engineering, UK, 2009.





## *Acknowledgements*

- Some of the slides are adapted or taken from the following tutorial
  - M. Ali Babar & Tore Dyba, Empirical Research in Software Engineering: concepts, Methods, and Challenges, XP 2008, Limerick, Ireland
- Researchers involved in the SLR were
  - Lianping Chen, Ciaren Cawley, & M. Ali Babar



**lero**

THE IRISH SOFTWARE  
ENGINEERING RESEARCH CENTRE

*Thank you!*