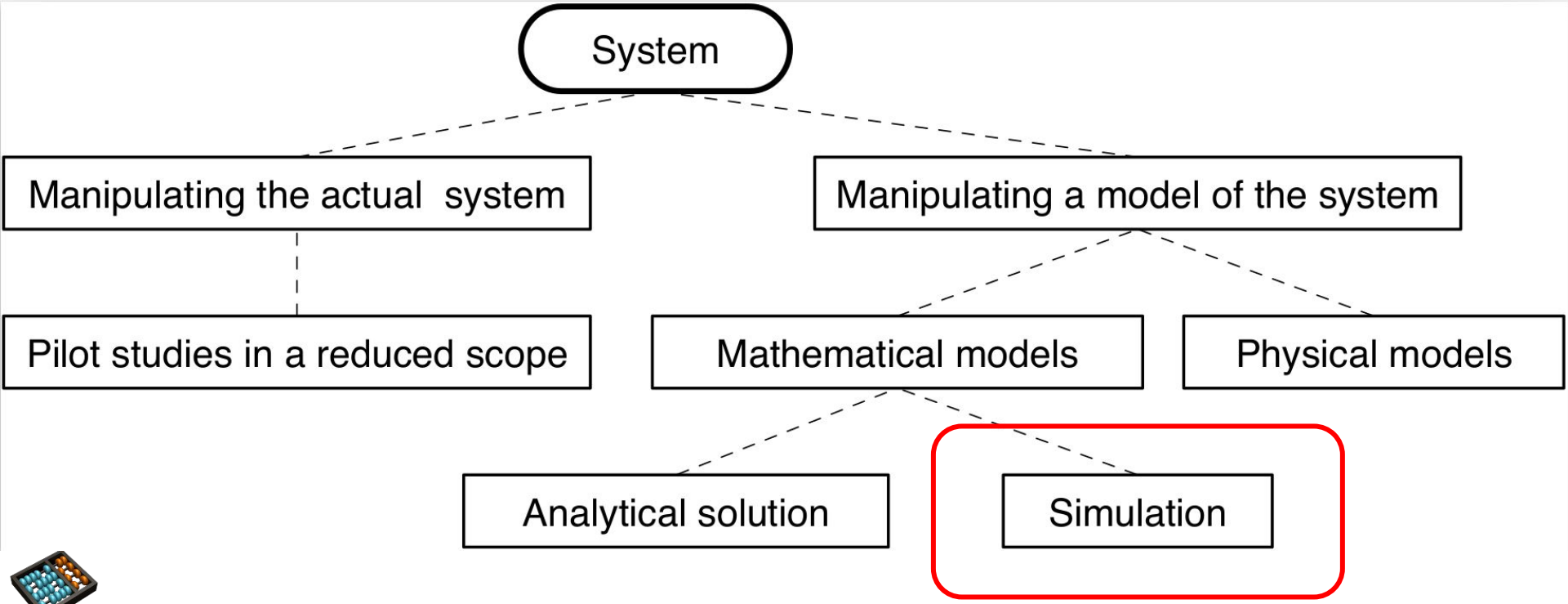
The Role of Simulation-based Studies in Software Engineering Research

Prof. Breno Bernard Nicolau de França
breno@ic.unicamp.br

Introduction

Ways of studying a system...



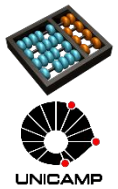
UNICAMP

Introduction

So, what is simulation?

"Simulation is the imitation of the operation of a real-world process or system over time. Simulation involves the generation of an artificial history of the system, and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented"

[Banks, 1999]

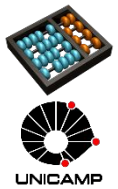


Introduction

Simulation supporting experimentation

- Biomedicine, Social Sciences, Physics and Engineering
  - Automobilistic Industry (crash tests)
  - Criminology (occurrence patterns, ethical and legal issues)

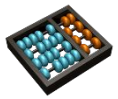
- Opportunity to investigate matters only observed at large scale



Introduction

Which systems can be simulated?

- The term “system” includes the notions of a process or phenomenon
- Physical existence is not an obligation
  - Concepts, ideas or proposals
- However, the system must have a “behavior over time” i.e., must be dynamic
  - Collection of interacting entities that produce an observable behavior in time interval
- The model is a representation or an abstraction of the real system
  - It is what we have been using science and engineering

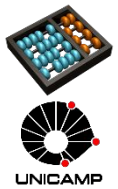


UNICAMP

Pros and Cons

Conducting experiments in labs or in the field may be:

- **Cost**: forecast the performance of a hardware update through all nodes in a network
- **Dangerous**: exploring alternative strategies to control a nuclear reactor
- **Take too long**: to evaluate the ecological impact of a long hunting period, for several years, in the population of involved species in an specific geographical region
- **Disturbing**: to evaluate the efficiency of one-way streets in the middle of a urban area
- **Morally/ethically unacceptable**: to evaluate the dispersion of radiation in a catastrophic failure
- **Irreversible**: to investigate the impact of changing the economical policy for a country



Claimed Advantages

We can fail!



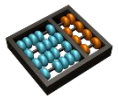




UNICAMP

Claimed Advantages

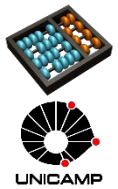
- To perform observations with high control
- To test hypothesis and assumptions
- To develop theories
- Low time and risk
- To replicate experiments
- Cheaper than other methods
- Proactive means to assess what will happen before actually committing resources for a change



UNICAMP

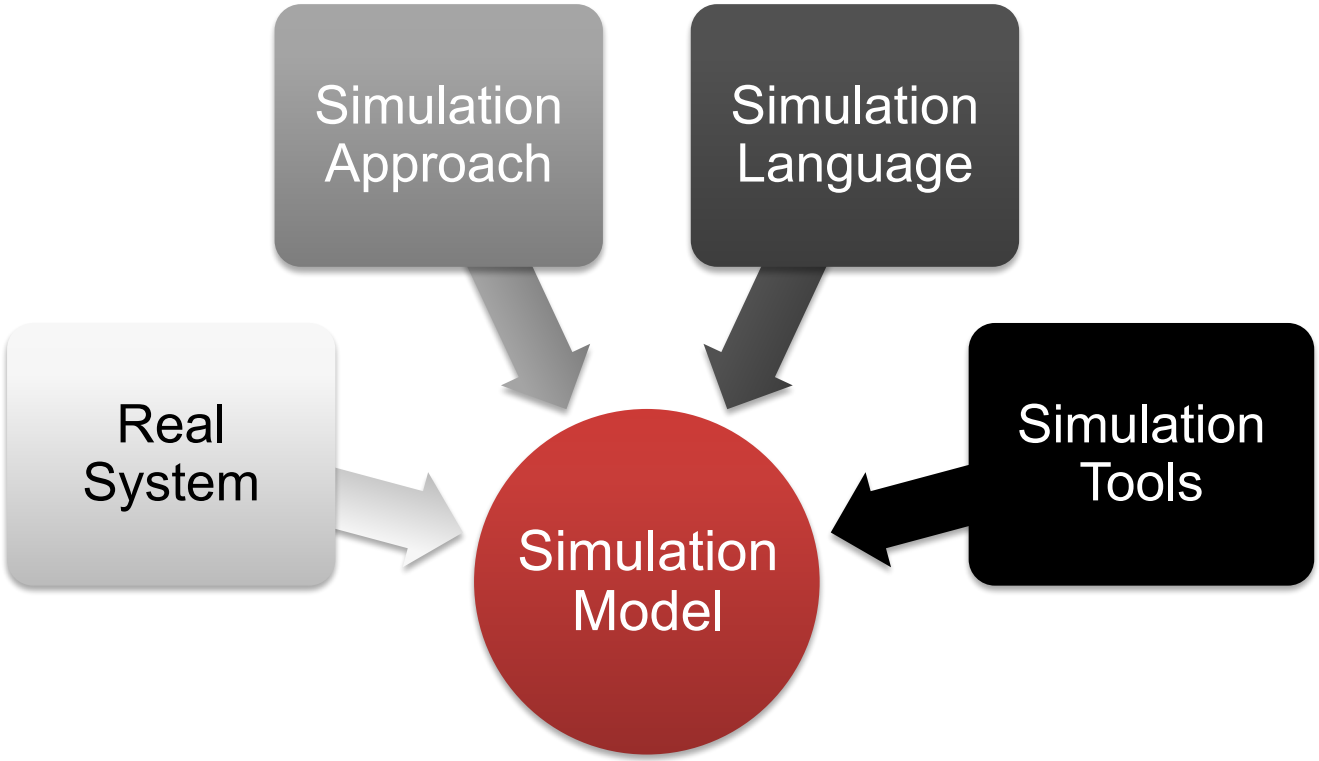
Disadvantages

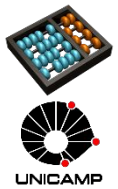
- High cost or effort to develop simulation models
- Models are abstractions or simplifications of the real world
  - Toy models
- How to evidence model validity?



Simulation Models

What do we need to develop a simulation model?



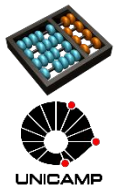


Simulation Models Life Cycle

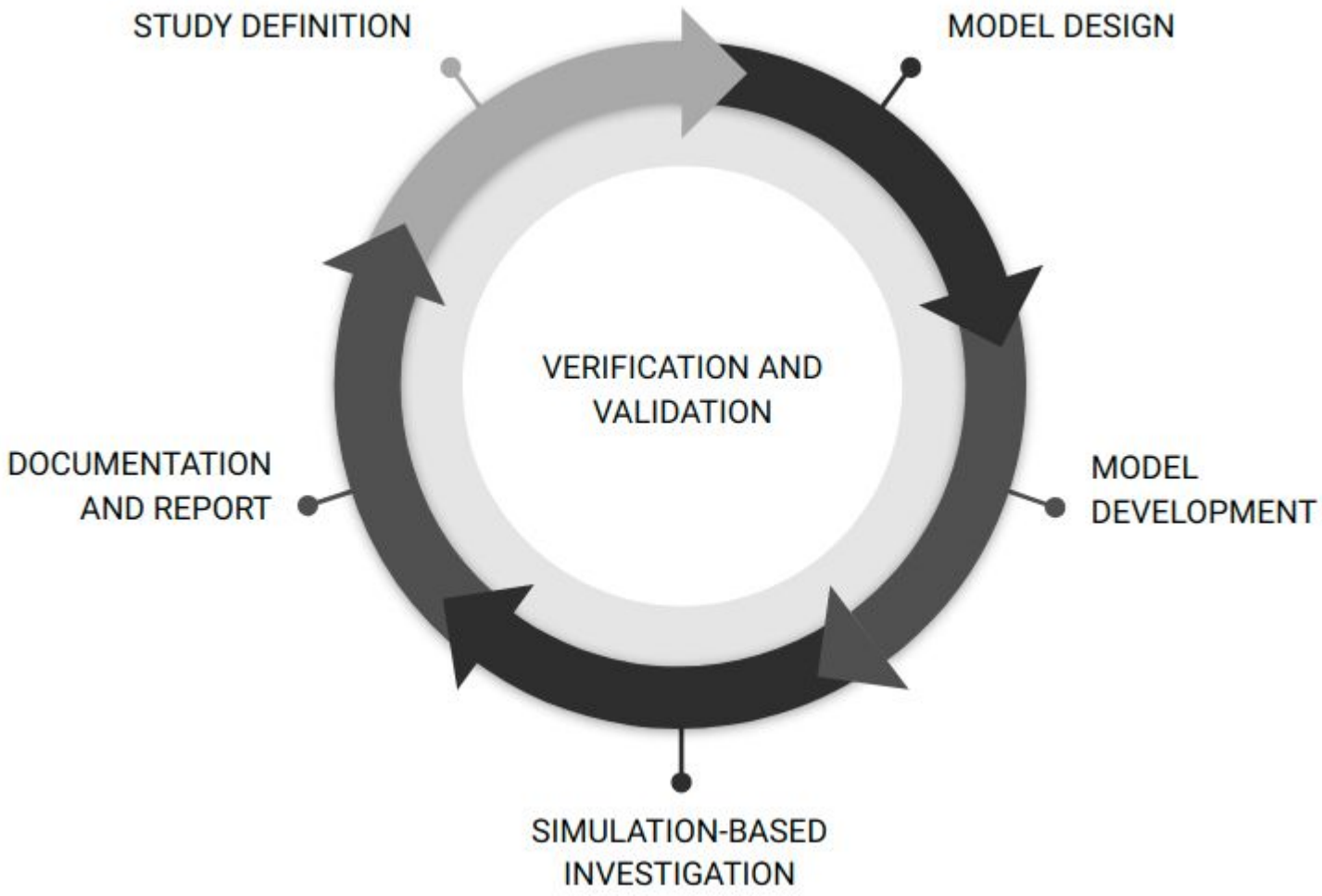
Simulation-based Studies: *"a series of steps, such as: data collection, coding and verification, model validation, experimental design, output analysis, and implementation"*

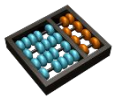
[Alexopoulos, 2007]

Models as instruments for observing the phenomenon under study



Simulation Models Life Cycle

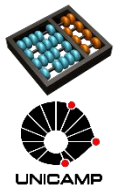




UNICAMP

Simulation Approaches

- **Systems Dynamics (Continuous Simulation)**
- **Discrete-Event Simulation**
- Agent-Based Simulation
- State-Based Simulation
- Hybrid simulation
- ...



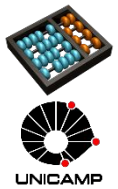
System Dynamics

- Mathematical models of continuous systems are defined by differential equations
- Time advances in constant intervals

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, \mathbf{p})$$

State variables (vector of stocks)

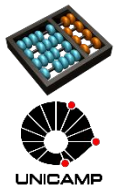
Input (list of parameters)



System Dynamics

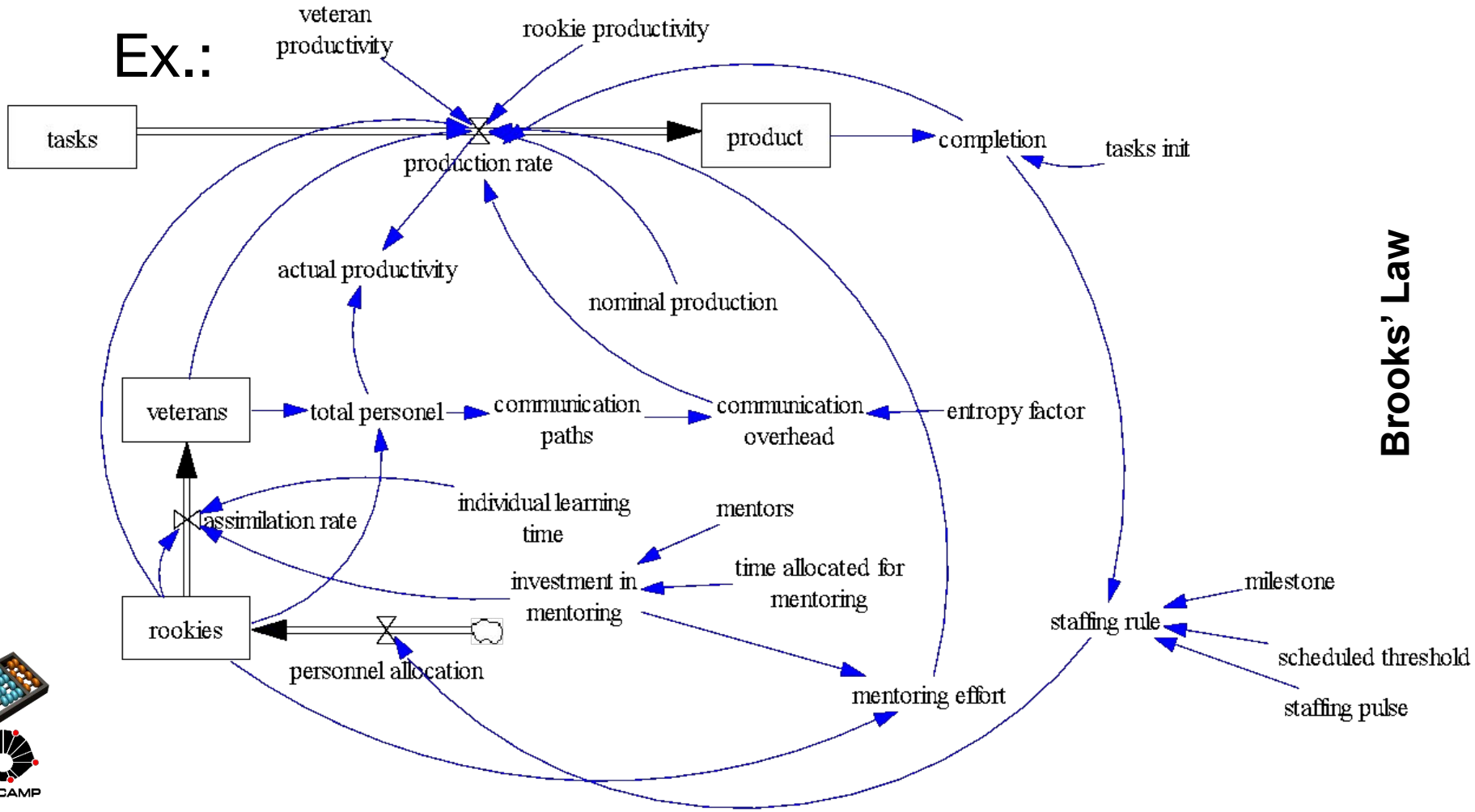
Modelling

- Causal loop diagram
  - Cause-effect relationships between variables
  - Positive (+) or negative (–) symbols: how dependent variables influences independent ones
  - Greater symbols (reinforcement and balance)
- Stock and flow models
  - **Stocks**: accumulation of values in the system that allows to measure the amount of a variable in a given instant
  - **Flows/Rates**: inputs and outputs of stocks that represents the rate of changes in stocks



System Dynamics

Ex.:

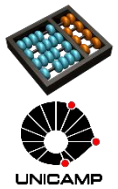


UNICAMP

David G Chernoguz. 2011. The system dynamics of Brooks' Law in team production. *Simulation,* 87, 11 (November 2011), 947-975.

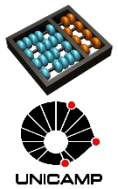
Discrete-Event Simulation

- Simplifications of continuous-time based systems
  - time instant *t* is a functional of the initial time t_0, a time interval *h*, and a discrete step *k*
  - $t = t_0 + hk$
- Intervals are not constant, it varies and it is not deterministic
  - Pseudo-random events controlled in a list, with the next instant
  - Dispatched in t0, t1, t2,... and enqueued!
  - Processing or service time is statistically distributed



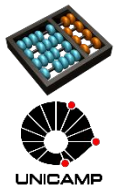
Discrete-Event Simulation





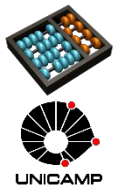
Experimentation in Software Engineering

- Low number of participantes
  - Reduced observation capacity (non-diverse population)
  - Reduced generalization capacity
- Hard to control or even unknown contextual variables
- Risk of failure during the execution of software products and projects when observing the phenomenon

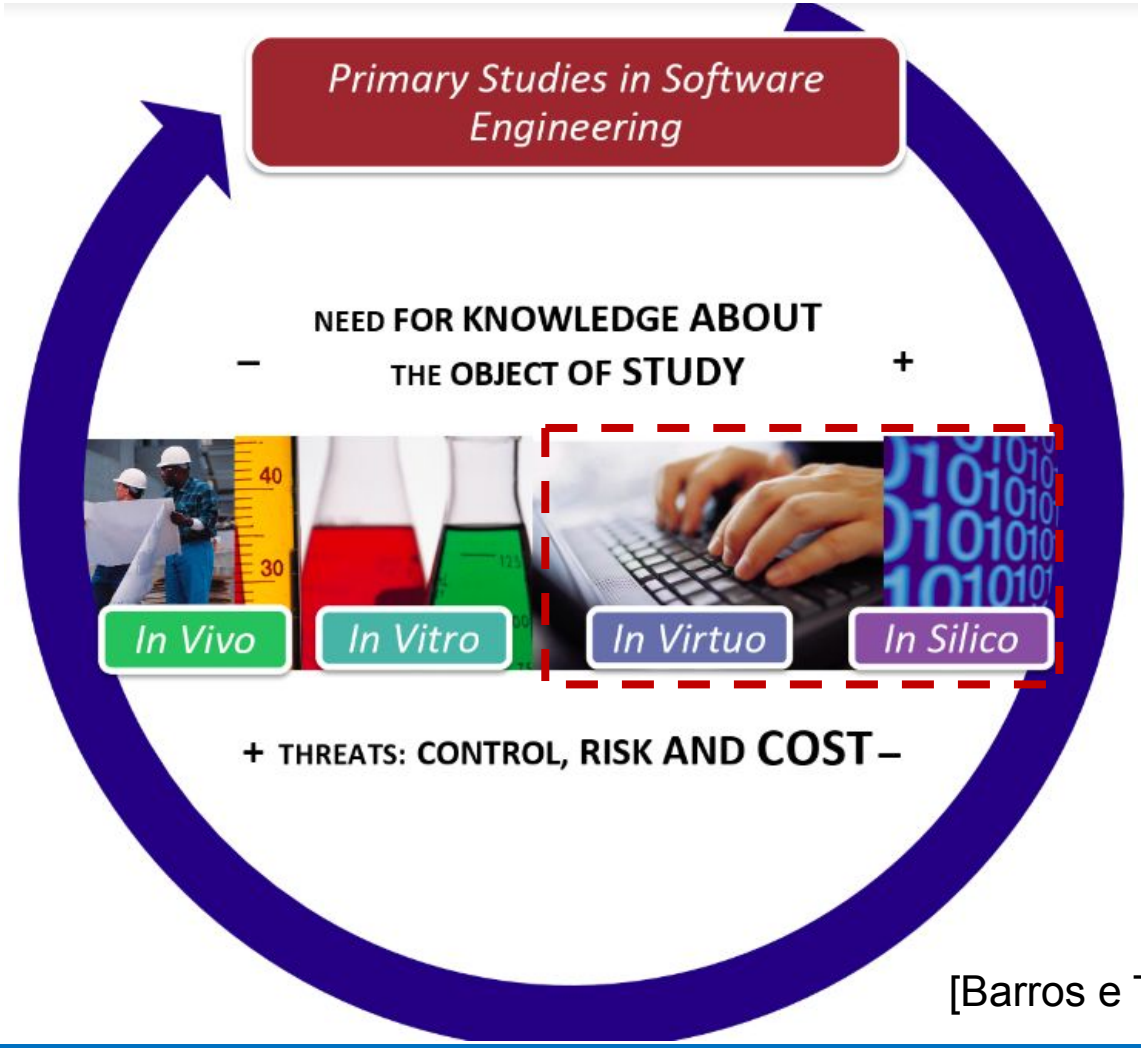


Experimentation in Software Engineering

- How to assess risks and feasibility in situations in which we do not have previous protocols?
- Difficulty on assuring control of variables in studies involving several variables or over a long period of execution



Experimental Cycle



[Barros e Travassos, 2003]

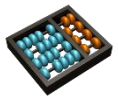
UNICAMP

Simulation in Software Engineering

Simulation appears in the SE literature as:

1. A problem-solving tool for decision support for SE practitioners
   a. the cost of conducting effective SBS is relevant!

2. A means to conduct controlled experiments and as an alternative to industrial case study research.
   a. the strength of evidence generated in SBS is relevant!.



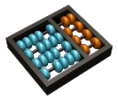
UNICAMP

Simulation in Software Engineering

1. A problem-solving tool for decision support for SE practitioners
   a. the cost of conducting effective SBS is relevant!

The cost of:

- developing and calibrating a simulation model
- design and analysis of simulation-based investigations
- the cost of necessary data collection
- or setting up a measurement program that can feed the simulation model with sufficiently reliable data.

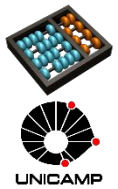


UNICAMP

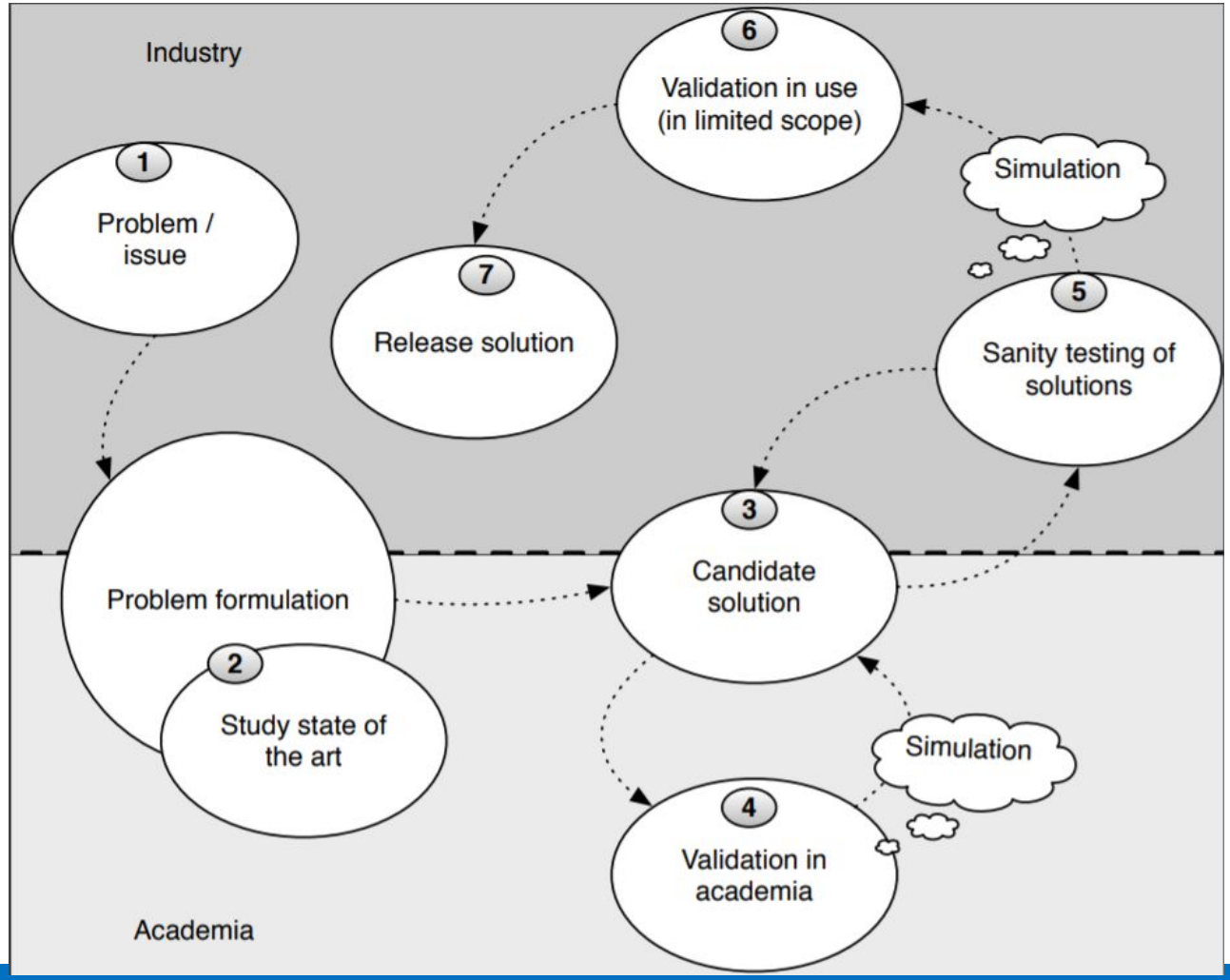
Simulation in Software Engineering

2. A means to conduct controlled experiments and as an alternative to industrial case study research.
   a. the strength of evidence generated in SBS is relevant!

- Models are simplifications
- Measurement challenges
- Lack of data



Simulation into SE Research and Practice



Adapted from: Gorschek, T., Wohlin, C., Garre, P., Larsson, S.: A model for technology transfer in practice. IEEE Software 23 (6), 88–95 (2006)

UNICAMP

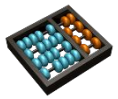
Simulation in Software Engineering

- Lack of …
  - Relevant information on reports
  - Methodological support for conducting simulation-based experiments
    - Model validity
    - Experimental design



Threats to validity

- Affecting
  - Understanding and reproducing studies
  - Credibility and confidence on results
  - Use of simulation as a supporting tool for SE research



UNICAMP

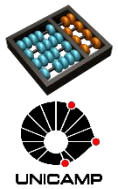
de França, B.B.N.; Travassos, G.H. Are We Prepared for Simulation Based Studies in Software Engineering yet? CLEI eletronic journal, CLEIej,v. 16, n. 1, paper 8, April, 2013. (CIbSE/ESELAW Special Issue)

Systematic Literature Reviews on Simulation in SE

Evolution of software process simulation from 1998 to 2008

- ~ 200 papers
- Results:
  - Categories to **classify** software process simulation models;
  - Research on improving the model performance
  - **Hybrid models** have attracted more attention from the community as a possibility to develop more realistic models.

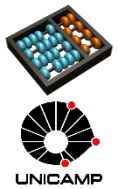
[Zhang et al, 2008]



Systematic Literature Reviews on Simulation in SE

Characterization of simulation-based studies in Software Engineering until 2011

- 108 papers
- Results:
  - Concentration on software process and project domains
  - SD and DEVS as the most adopted approaches
  - Lack of agenda on model purposes
  - Poor characterization
  - Lack of rigor
    - Verification & Validation
    - Planning
    - Output analysis procedure
  - Poor reporting quality
  - Need for supporting methods

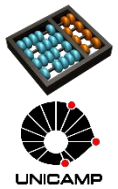


Systematic Literature Reviews on Simulation in SE

Characterization of software process simulation in industry until 2013

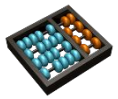
- 87 papers
- Results:
  - Trends on proof of concepts with different purposes (estimate, training, process improvement, etc.)
  - Studies with poor quality
  - Lack of model assessment w.r.t. its purpose
  - Lack of conclusive evidence regarding the usefulness of SPS
  - Among the reported costs, we cannot say it is cheaper
  - Need for supporting methods to conduct and report

[Ali, Petersen and Wohlin, 2014]



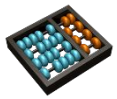
Verification & Validation

Procedure	Description
Face Validity	Consists of getting feedback from individuals knowledgeable about the phenomenon of interest through reviews, interviews, or surveys, to evaluate whether the (conceptual) simulation model and its results (input-output relationships) are reasonable.
Comparison with Reference Behaviors	Compare the simulation output results against trends or expected results often reported in the technical literature.
Comparison with Other Models	Compare the results (outputs) of the simulation model being validated to results of other valid (simulation or analytic) model. Controlled experiments can be used to arrange such comparisons.
Event Validity	Compare the "events" of occurrences of the simulation model to those of the real phenomenon to determine if they are similar. This technique is applicable for event-driven models.
Historical Data Validation	If historical data exists, part of the data is used to build the model and the remaining data are used to compare the model behavior and the actual phenomenon. Such testing is conducted by driving the simulation model with either sample from distributions or traces, and it is likely used for measuring model accuracy.
Rationalism	Use logic deductions from model assumptions to develop the correct (valid) model, by assuming that everyone knows whether the clearly stated underlying assumptions are true.



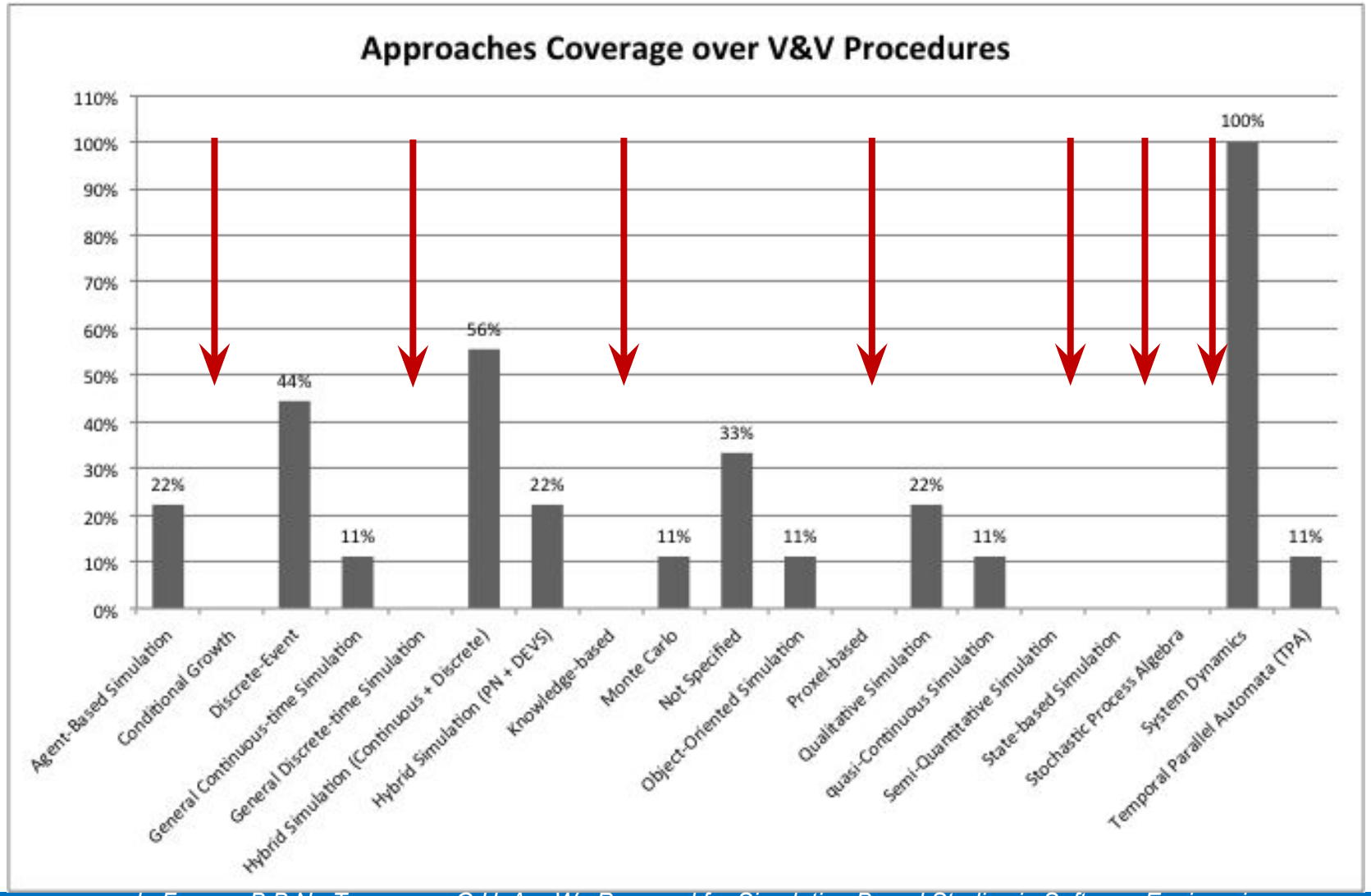
Verification & Validation

Predictive Validation	Use the model to forecast the phenomenon behavior, and then compares this behavior to the model forecast to determine if they are the same. The phenomenon data may come from the real phenomenon observation or be obtained by conducting experiments, e.g., field tests for provoking its occurrence. Also, data from the technical literature may be used, when there is no complete data in hands.
Internal Validity	It is likely used for measuring model accuracy. Several runs of a stochastic model are made to determine the amount of (internal) stochastic variability. A large amount of variability (lack of consistency) may cause the model results to be questionable, even if typical of the problem under investigation.
Sensitivity Analysis	Consists of systematically changing the values of the input and internal parameters of a model to determine the effect upon the model output. The same relationships should occur in the model as in the real phenomenon. This technique can be used qualitatively – trends only – and quantitatively – both directions and (precise) magnitudes of outputs.
Testing model structure and behavior	Submit the simulation model to tests cases, evaluating its responses and traces. Both model structure and outputs should be reasonable for any combination of values of model inputs, including extreme and unlikely ones. Besides, the degeneracy of the model behavior can be tested by appropriate selection of values of parameters.
Based on empirical evidence	Collect evidence from the technical literature (experimental studies reports) to develop the model causal relationships (mechanisms).
Turing Tests	Individuals knowledgeable about the phenomenon are asked if they can distinguish between real and model outputs.



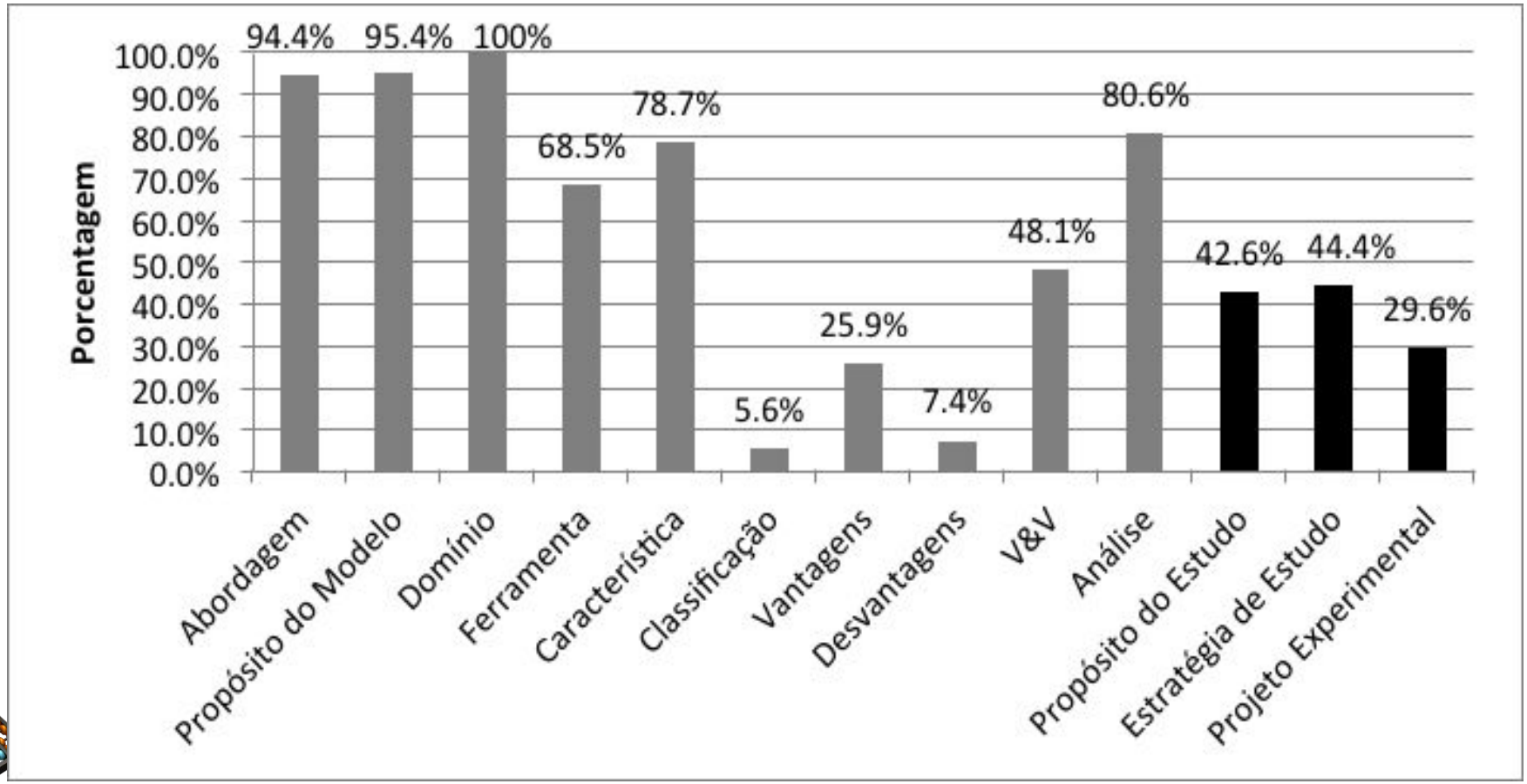
UNICAMP

Verification & Validation



UNICAMP

Quality of Reporting



UNICAMP

de França, B.B.N.; Travassos, G.H. Are We Prepared for Simulation Based Studies in Software Engineering yet? CLEI eletronic journal, CLEIej,v. 16, n. 1, paper 8, April, 2013.

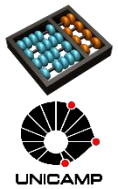
Simulation in Software Engineering

Validity threats are different from other methods!

From Context to Research Questions
Simulation Feasibility
Simulation Model
Model Validation
Subjects (*in virtuo* and *in silico*)
Experimental Design
Intermediate Experimental Trials
Supporting Data
Simulation Environment
Output Analysis
Study Report

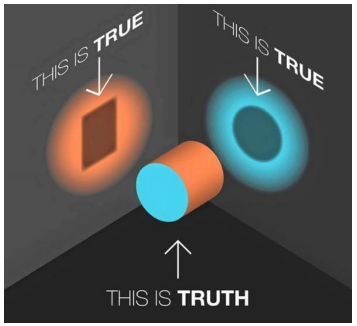


Prevent some risks
but new arise



Bernard Nicolau de França, B., & Horta Travassos, G. (2015). Simulation based studies in software engineering: A matter of validity. *CLEI electronic journal*, *18*(1), 5-5.

Worldviews (Paradigms)



Esta Foto de Autor Desconhecido está licenciado em CC BY-NC-SA



Esta Foto de Autor Desconhecido está licenciado em CC BY-NC-SA

Paradigm	Thruth	Research Method	Data Collection
Positivist/ Post positivist	Objective (independent from participants)	Controlled experiment, case study, and survey	Quantitative (Random sampling)
Constructivist/ Interpretivist	Subjective (constructed by participants)	Case study, survey, Interview	Quantitative and Qualitative (Random and purposeful)
Advocacy/ Participatory	Subjective (constructed by participants and the observer)	Action research (intervention driven)	Quantitative and Qualitative (Random and purposeful)
Pragmatist	Depends on what works at the time	Multi method research	Quantitative and Qualitative (Random and purposeful)



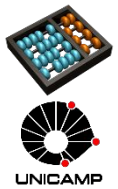
UNICAMP

Simulation in Software Engineering

Validity threats are different from other methods!

- Conclusion Validity
  - Ex: Considering only one observation when dealing with stochastic simulation, rather than central tendency and dispersion measures

- Internal Validity
  - Ex.: Simulation model simplifications (assumptions) forcing the desired outcomes
  - *“In order to introduce the Test-First Development practice into the FLOSS simulation model, we make the following assumptions: (1) The average time needed to write a line of production code increases; (2) The number of defects injected during coding activities decreases; (3) The debugging time to fix a single defect decreases”.*

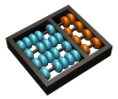


Bernard Nicolau de França, B., & Horta Travassos, G. (2015). Simulation based studies in software engineering: A matter of validity. *CLEI electronic journal*, *18*(1), 5-5.

Simulation in Software Engineering

Validity threats are different from other methods!

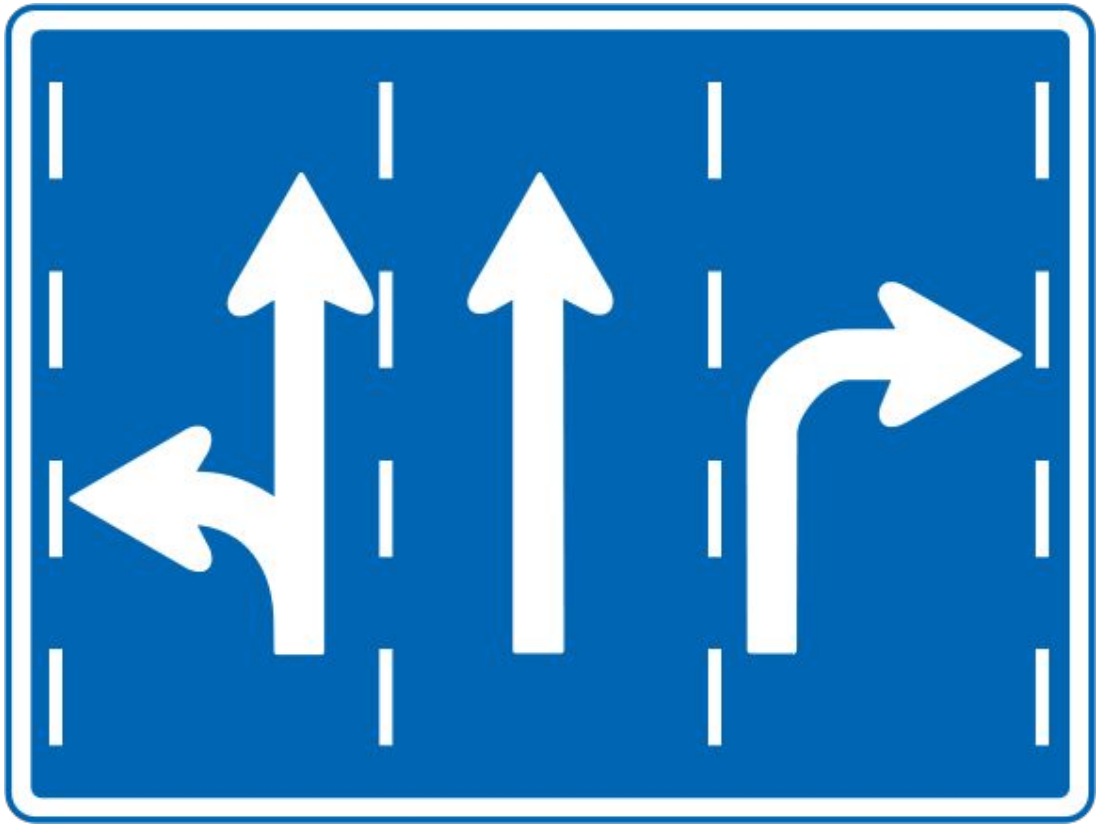
- Construct Validity
  - Ex: Inappropriate real-world representation by model parameters
  - The lack of evidence regarding model validity reduces the findings only to the simulation model
- External Validity
  - Ex.: Simulation results differ from the outcomes of empirical observations
  - Simulation model not based on empirical evidence



UNICAMP

Bernard Nicolau de França, B., & Horta Travassos, G. (2015). Simulation based studies in software engineering: A matter of validity. *CLEI electronic journal*, *18*(1), 5-5.

But, what can we do?

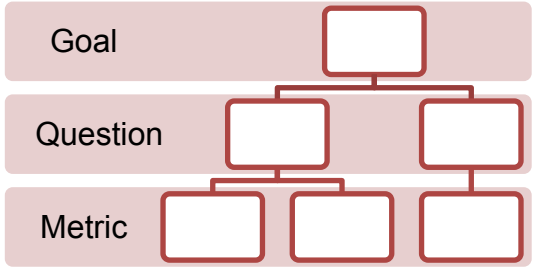


Guidelines for Reporting SBS

Identification
Study Definition
- **Context**
- **Problem**
- **Goals**
- **Questions**
- **Hypothese**

Simulation Feasibility
Background and Related Work
Simulation Model and Validity
Participants
Experimental Design
Intermediate Runs
Supporting Data
Simulation Environment
Output Analysis
Threats to Validity
Conclusions and Future Works





Extension of Balci (1999):
Custo, tempo e benefícios □ Fenômeno observável, Efeitos de alta ordem, Riscos na ocorrência do fenômeno, Disponibilidade de dados.

UNICAMP

Guidelines for Reporting SBS

Identification
Study Definition
- **Context**
- **Problem**
- **Goals**
- **Questions**
- **Hypothese**

Simulation Feasibility
Background and Related Work
Simulation Model and Validity
Participants
Experimental Design
Intermediate Runs
Supporting Data
Simulation Environment
Output Analysis
Threats to Validity
Conclusions and Future Works

Procedure
Face Validity
Comparison to Reference Behaviors
Comparison to Other Models
Event Validity
Historical Data Validation
Rationalism
Predictive Validation
Internal Validity
Sensitivity Analysis
Testing model structure and behavior
Based on empirical evidence
Turing Tests

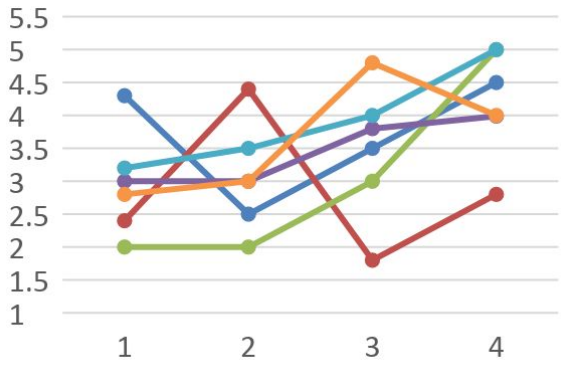


UNICAMP

Guidelines for Reporting SBS

- **Identification**
- **Study Definition**
  - **Context**
  - **Problem**
  - **Goals**
  - **Questions**
  - **Hypothese**
- **Simulation Feasibility**
- **Background and Related Work**
- **Simulation Model and Validity**
- **Participants**
- **Experimental Design**
- **Intermediate Runs**
- **Supporting Data**
- **Simulation Environment**
- **Output Analysis**
- **Threats to Validity**
- **Conclusions and Future Works**

Scenario	Factor 1	Factor 2	…	Factor K
1	0	0	…	0
2	1	0	…	0
…	…	…	…	…
N	1	1	…	1





UNICAMP

Guidelines for Reporting SBS

Identification
Study Definition
- **Context**
- **Problem**
- **Goals**
- **Questions**
- **Hypothese**

Simulation Feasibility
Background and Related Work
Simulation Model and Validity
Participants
Experimental Design
Intermediate Runs
Supporting Data
Simulation Environment
Output Analysis
Threats to Validity
Conclusions and Future Works

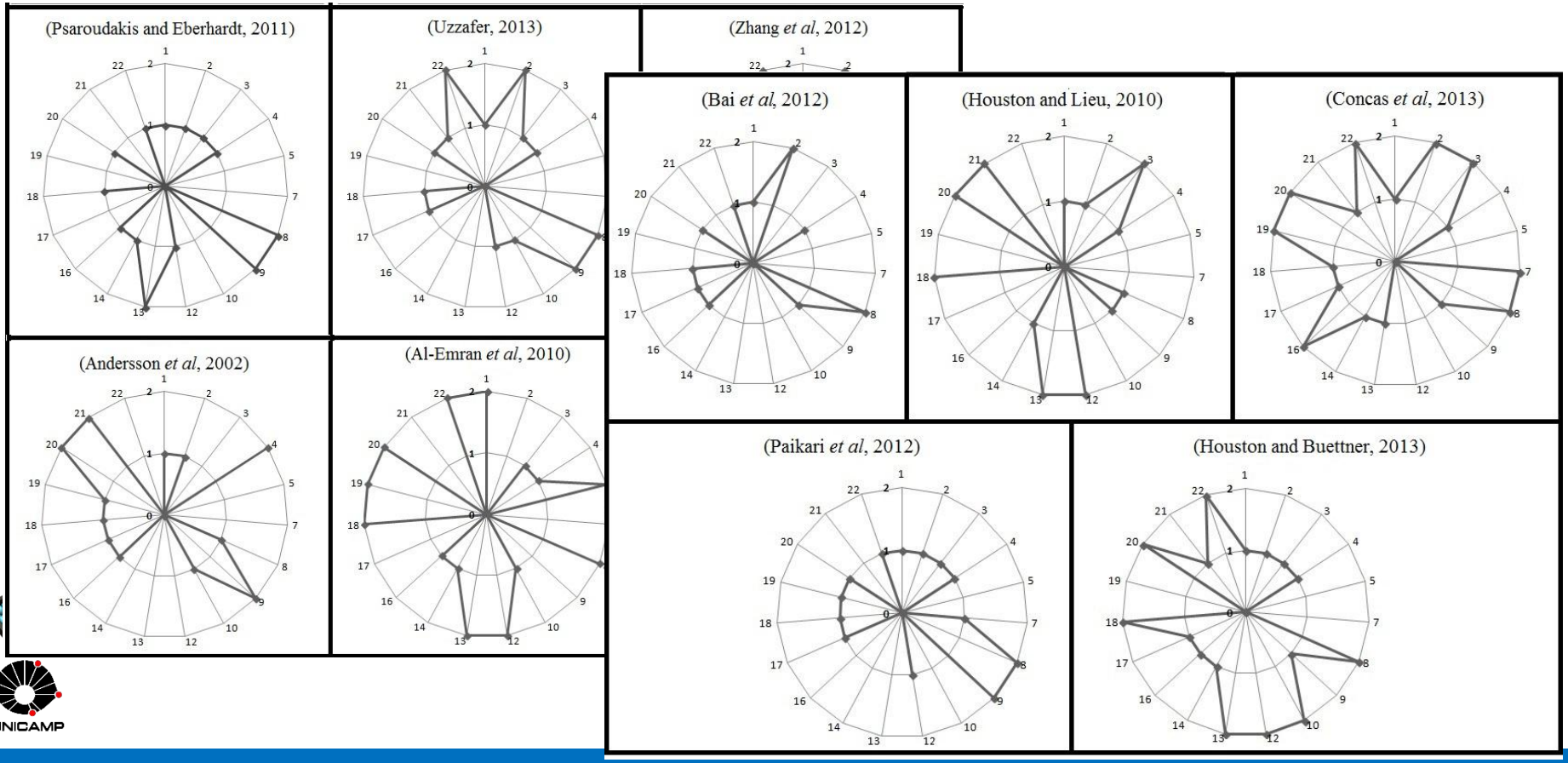
Dados reais ou artificiais?



UNICAMP

Guidelines for Reporting SBS

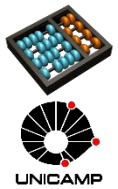
Analysis against SE literature



Guidelines for Planning SBS

Context (*in silico*)

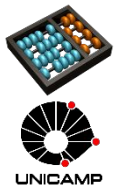
- Planning of release frequency
- Web IS for controlling business process
  - Research Support Organization
- Iterative and Incremental Development Process
  - Emphasis on V&V activities
  - Two sites geographically distributed
  - 12 software developers using Java as platform



Guidelines for Planning SBS

Problem

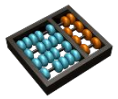
- Maintenance cycles (corrective, evolutive or perfective)
- Quality, time and budget constraints release planning
  - **However**, several factors impact on these variables that may over or underestimate
  - Causing, **thus**, delays, increase of cost, and quality decay



Guidelines for Planning SBS

Goal (GQM)

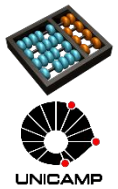
Purpose	
Analyze	the evolution of a Web information system
For the purpose of	characterization
Perspective	
With respect to	the duration of maintenance cycles as well as its effect on product quality
From the point of view	the SE Researcher in the context of simulating quality decay
Environment	
In the context of	a large-scale information system with the use of a SD model as instrument.



UNICAMP

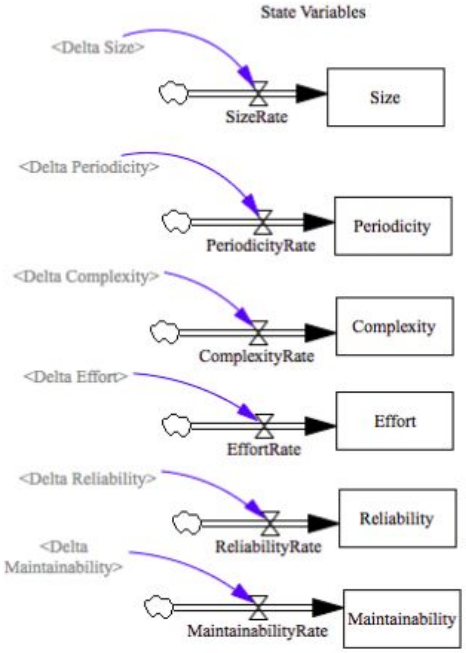
Guidelines for Planning SBS

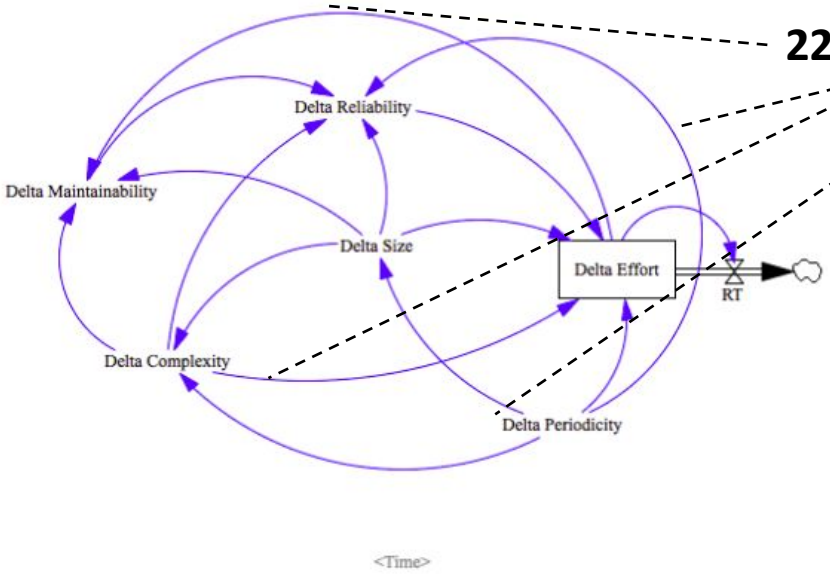
- Research Questions
  - *Q_1: Which periodicity (shorter or longer cycles) performs better for the next 6 months after the last release?*
  - *Q_2: Which strategy (fixed or variable duration cycles) performs better regarding product quality?*
- Why simulation?
  - Análise em longo prazo com diversas variáveis de interesse a serem controladas **sem impor risco** ao projeto
  - Observação dessas variáveis e suas interações, considerando **efeitos de alta ordem**



Guidelines for Planning SBS

Descrição, Validade e Evidências: SG9, 10 e 12





22 Evidências (RSL)

Validação Histórica: *dataset* dividido em calibração e validação

(ARAÚJO, MONTEIRO and TRAVASSOS, 2012)

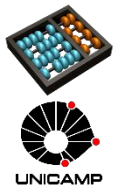
$$Delta\ Reliability = (0.199239 \times Delta\ Periodicity) + (2.23944 \times Delta\ Complexity) + (0.167732 \times Delta\ Maintainability) - (0.92313 \times Delta\ Size)$$

Guidelines for Planning SBS

Verifying Assumptions

- Logical formulation: expected behavior for each characteristic
  - Increase - ↑, decrease - ↓ or maintained - ↔
- Tested as hypotheses when observed any Law of Software Evolution
  - Behavior can be observed or not, depending on the context

ARAÚJO, MONTEIRO and TRAVASSOS (2012)



Guidelines for Planning SBS

Participants

- *In silico*, focused on product characteristics
- Tem behavior embedded in data

Experimental Design and Sensibility Analysis

Scenario	Strategy	*Periodicity*
1	Fixed-duration	2
2	Fixed-duration	10
3	Fixed-duration	40
4	Variable-duration	Low mean (2) and variation (1)
5	Variable-duration	Medium mean (10) and variation (5)
6	Variable-duration	High mean (20) and variation (10)



UNICAMP

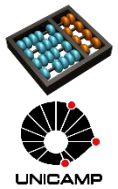
Guidelines for Planning SBS

Number of runs

- Duração fixa (determinístico): 3 rodadas
- Duração variável (estocástico): 300 rodadas

Supporting Data

- Medições para as seis variáveis, em 13 releases
- Fontes: Controle de versão, *Bug tracking,* Planilhas de esforço



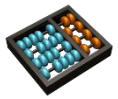
Guidelines for Planning SBS

Contextual Data

- Documentation and project members
- Releases: corrective, adaptive, and perfective maintenance
  - Perfective: security, performance, maintainability and usability
  - No new feature is added
  - Just bugs reported are considered

Data for Calibration x Experimentation

- Input data determined from collected data
  - Observed values and distributions
  - Time Sensitiveness
  - Two years of real project data

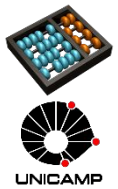


UNICAMP

Guidelines for Planning SBS

Simulation Environment

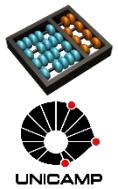
- Vensim PLE
  - System Dynamics
  - Academic version with limited support to experimental design, but free of costs
  - Analysis: causal trees, traces and charts plotting variables over time
- Spreadsheets to support output analysis



Guidelines for Planning SBS

Output Analysis

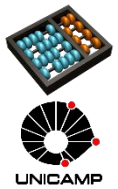
- Histogramas e Gráficos de Sequência para as variáveis de resposta
  - Histograma: distribuições
  - Sequência
    - Comportamento em função do tempo
    - Comparação entre cenários
    - Analise de múltiplas rodadas, com medidas de centralidade e dispersão



Guidelines for Planning SBS

Threats to validity

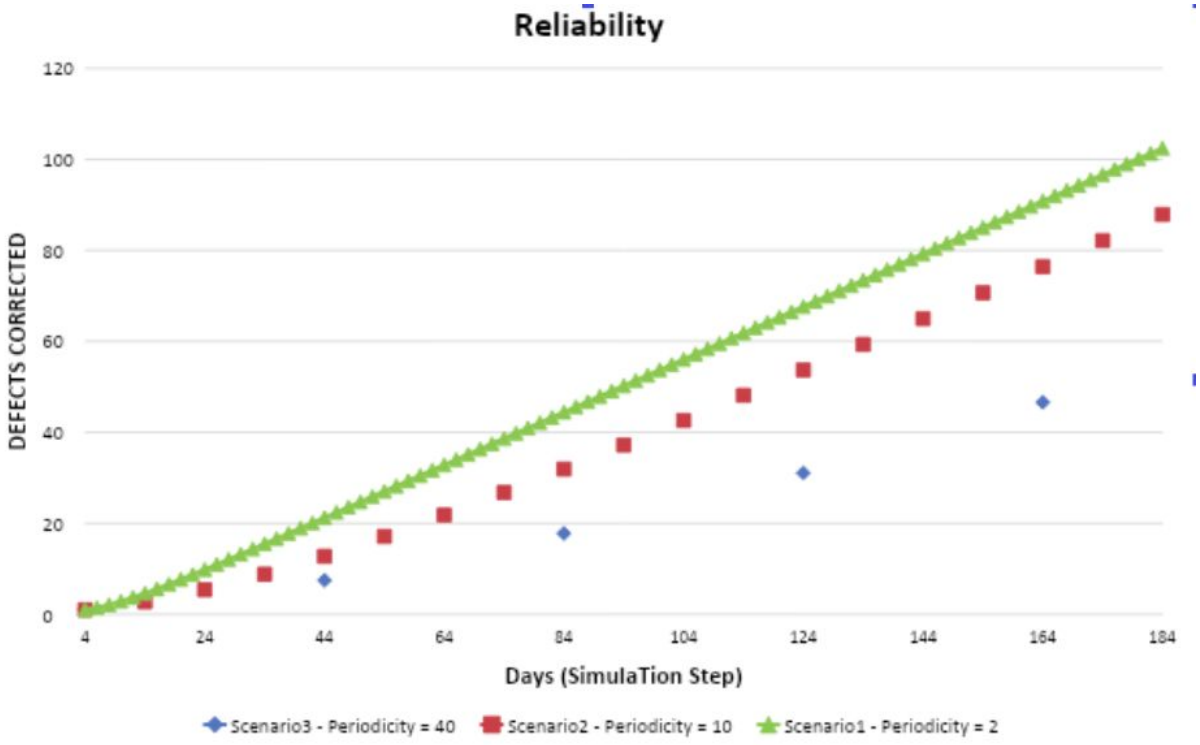
- Modelo abstrai detalhes de processo
  - Limita as possíveis explicações
  - Informações de contexto
- Horas para correções é problemático
  - Não considera manutenção perfectiva, que pode melhorar manutenibilidade e remete a ciclos longos
- Somente tendências são observadas, explicando apenas comportamentos gerais
  - Ciclos curtos de melhoria e longos de correção nos dados



Guidelines for Planning SBS

Results and Applicability

- Ciclos de manutenção curtos levam à maior confiabilidade

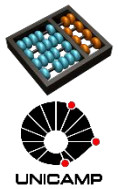


UNICAMP

Guidelines for Planning SBS

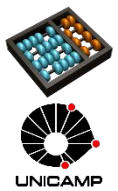
Results and Applicability

- Ciclos de duração fixa promovem mais confiabilidade para ciclos curtos e médios
- Ciclos curtos tendem a reduzir a manutenibilidade
  - Manutenção perfectiva não é realizada
- Estabilização da confiabilidade e manutenibilidade em longo prazo
  - Maior controle da qualidade em ciclos de duração fixa



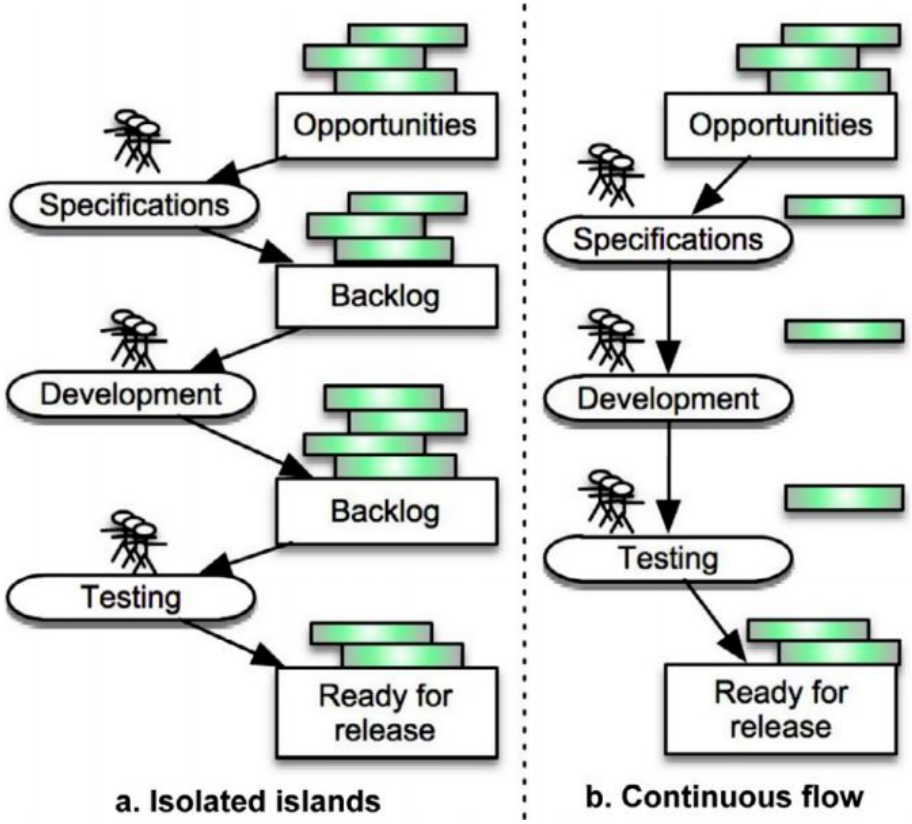
What more to simulate in SE?

- DevOps/Continuous Deployment
  - Data from repositories, CI/CD tools, and production environment
- Large-scale Systems
  - Systems of Systems, Smart Cities
- Cyber-physical Systems
  - Simulation abstracting hardware issues: Robotics, IoT, …



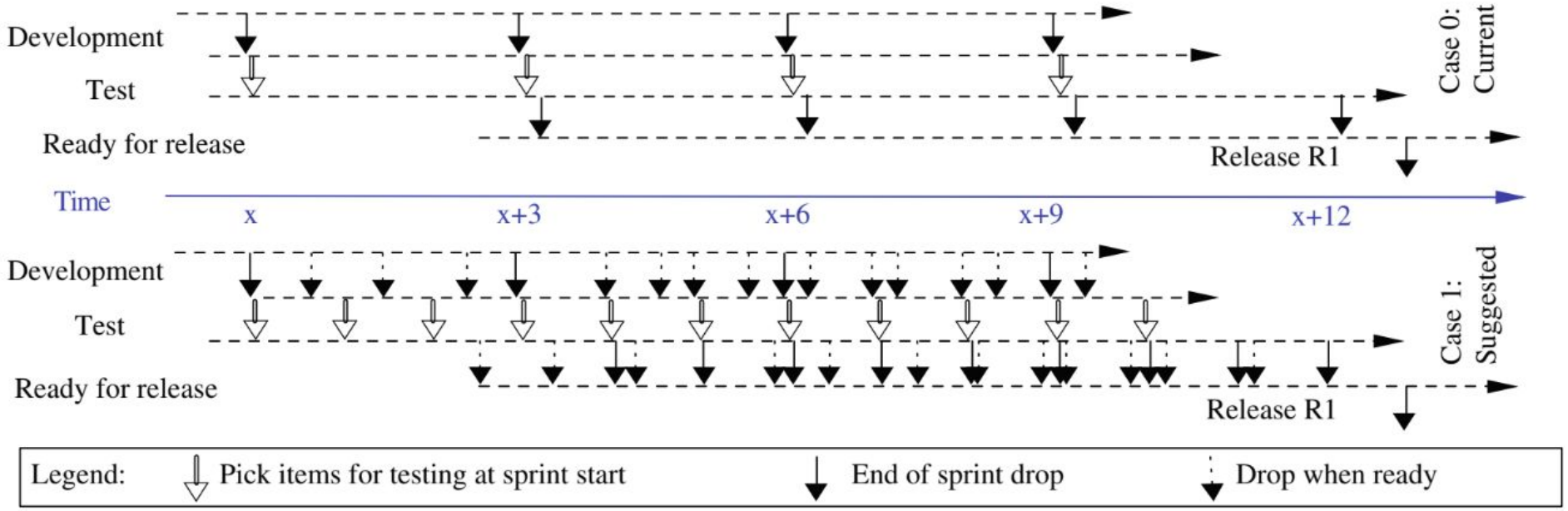
What more to simulate in SE?

Continuous Development Flow



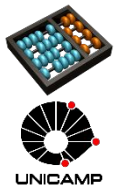
What more to simulate in SE?

Continuous Development Flow

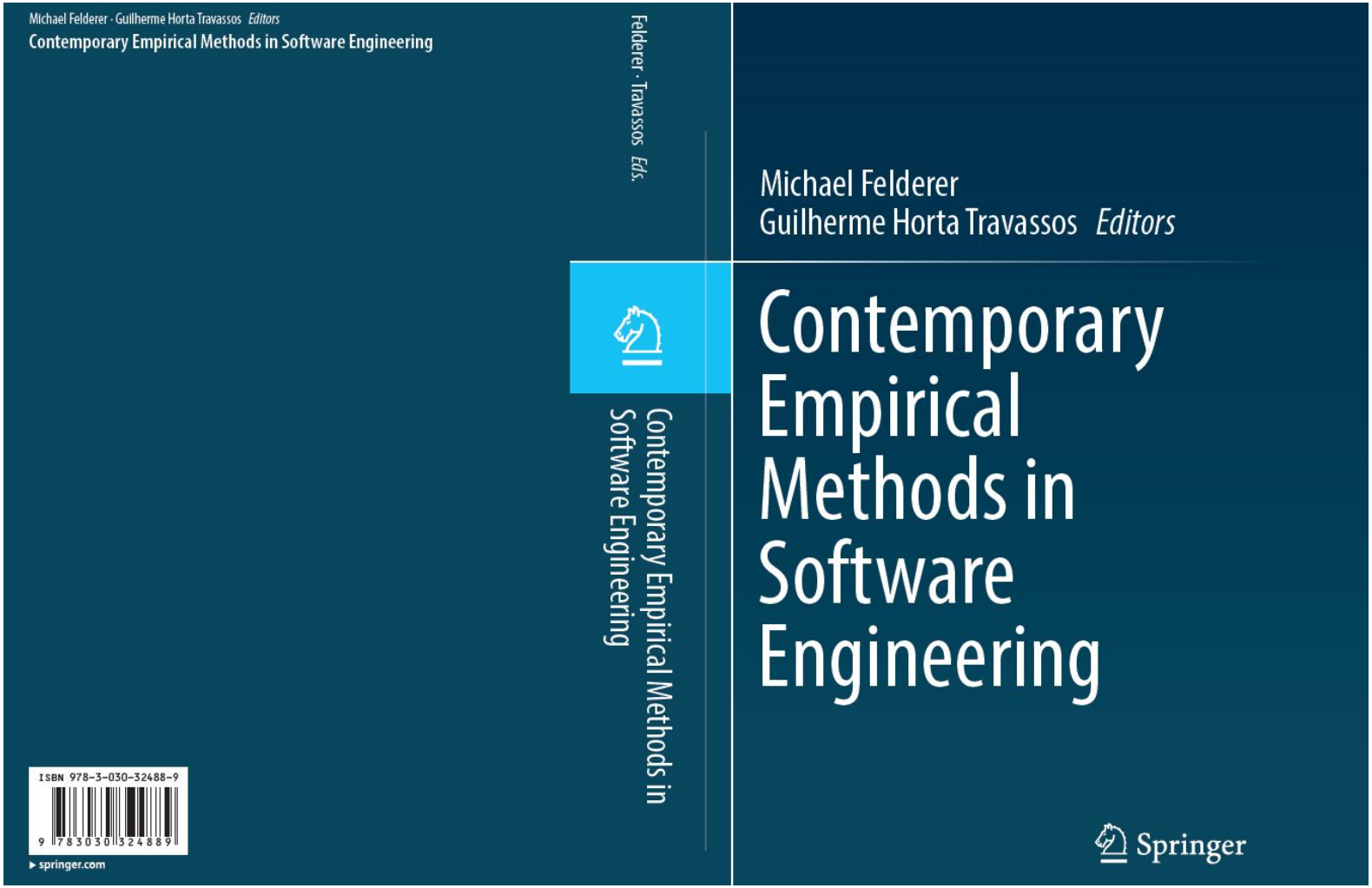


What more to simulate in SE?

- Critical Systems
  - Safety-critical, business-critical
- Focused interventions on software processes
  - Wide-scope simulation of software projects may involve several variables and unavailable data



ESE Book coming...



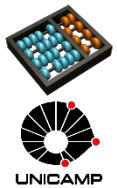
The Role of Simulation-based Studies in Software Engineering Research

Thanks!

Prof. Breno Bernard Nicolau de França
breno@ic.unicamp.br

Referências

- Banks, J. Introduction to Simulation. In: Winter Simulation Conference (WSC'99). Phoenix, AZ, USA (1999)
- Alexopoulos, "Statistical analysis of simulation output: State of the art". Simulation Conference, 2007 Winter, 2007.
- Barros, M. O.,Travassos, G. H. Contributions of In Virtuo and In Silico Experiments for the Future of Empirical Studies in Software Engineering. In: WSESE03. Fraunhofer IRB Verlag, Rome (2003)
- O. Balci, "Guidelines for successful simulation studies," Proc. Winter Simulation Conference, pp. 25-32, 1990.
- B. B. N. França, G. H. Travassos, "Are We Prepared for Simulation Based Studies in Software Engineering yet?," Proc. IX Experimental Software Engineering Latin American Workshop, Buenos Aires, Argentina, 2012.
- França, B.B.N.; Travassos, G.H. Reporting Guidelines for Simulation-Based Studies in Software Engineering. In: International Conference on Evaluation and Assessment in Software Engineering, EASE, Ciudad Real, Spain. 2012.
- Höst, M., Regnell, B., Tingström, C. A framework for simulation of requirements engineering processes EUROMICRO 2008 - Proceedings of the 34th EUROMICRO Conference on Software Engineering and Advanced Applications, SEAA 2008, 2008, 183-190
- Müller, M.; Pfahl, D. Simulation Methods. Guide to Advanced Empirical Software Engineering. 2008, Section I, 117-152, DOI: 10.1007/978-1-84800-044-5_5



Referências

- Ahmed, R.; Hall, T.; Wernick, P.; Robinson, S. & Shah, M. Software process simulation modelling: A survey of practice. Journal of Simulation, 2008, 2, 91 – 102.
- Zhang, H., Kitchenham, B., Pfahl, D.: Reflections on 10 years of software process simulation modeling: A systematic review. LNCS, vol. 5007, pp. 345-356 (2008)
- Birta, L.G.; Arbez, G. Modelling and Simulation: Exploring Dynamic System Behaviour. Springer. 2007.
- John A. Sokolowski; Catherine M. Banks. Principles of Modeling and Simulation: A Multidisciplinary Approach. John Wiley & Sons, Inc. 2009.
- Severance, F. L. System Modeling and Simulation: An Introduction. John Wiley & Sons, Inc. 2001.

