
Analyzing a Software Process Model Repository for Understanding Model Evolution

Studying Follow-Up Changes to a Large Process Model

Martín Soto, Alexis Ocampo and Jürgen Münch

Fraunhofer Institute for Experimental Software Engineering
Kaiserslautern, Germany

Overview

Motivation

Execution: Study 1

Observations from Study 1

Execution: Study 2

Results

Conclusions

Motivation

Execution (1)

Observations

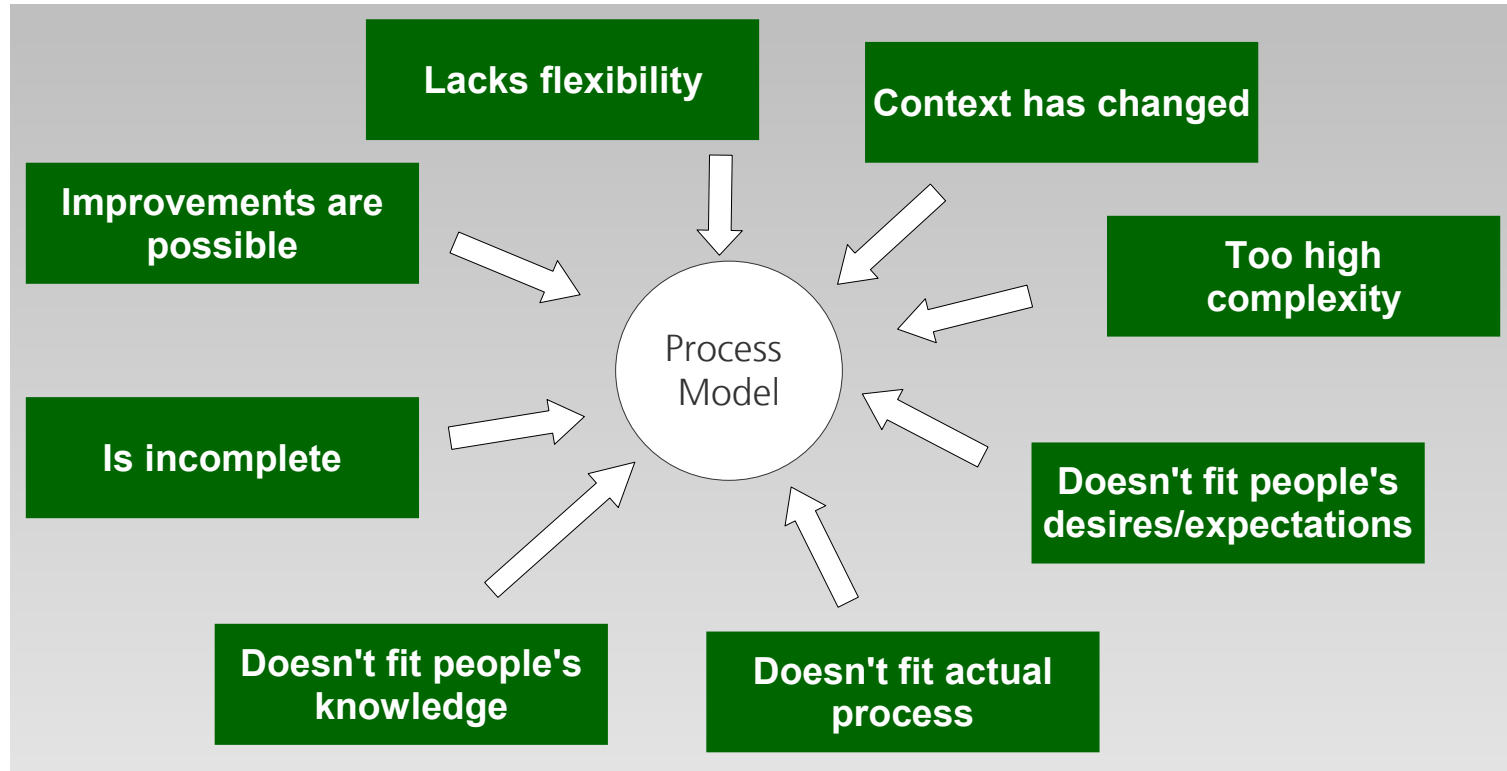
Execution (2)

Results

Conclusions

Motivation & Background

Reasons for Process Model Change



Central Question

Motivation

**Is process
description
evolution similar to
software evolution?**

Slide 5

Overview of Our Work

The Secret Life of a Process Model

ICSP 2008, Leipzig

Object: Development history of a large process description: the German V-Modell XT.

Exploratory study:

- No formal hypotheses
- Observation and visualization in order to recognize interesting patterns.

Slide 6

Overview of Our Work

The Secret Life of a Process Model

ICSP 2008, Leipzig

Analyzing a Software Process Model Repository for Understanding Model Evolution,

ICSP 2009, Vancouver

Object: Development history of a large process description: the German V-Modell XT.

Exploratory study:

- No formal hypotheses
- Observation and visualization in order to recognize interesting patterns.

Empirical study:

- Two formal hypotheses
- Analysis through statistical methods.

Slide 7

Motivation

Execution (1)

Observations

Execution (2)

Results

Conclusions

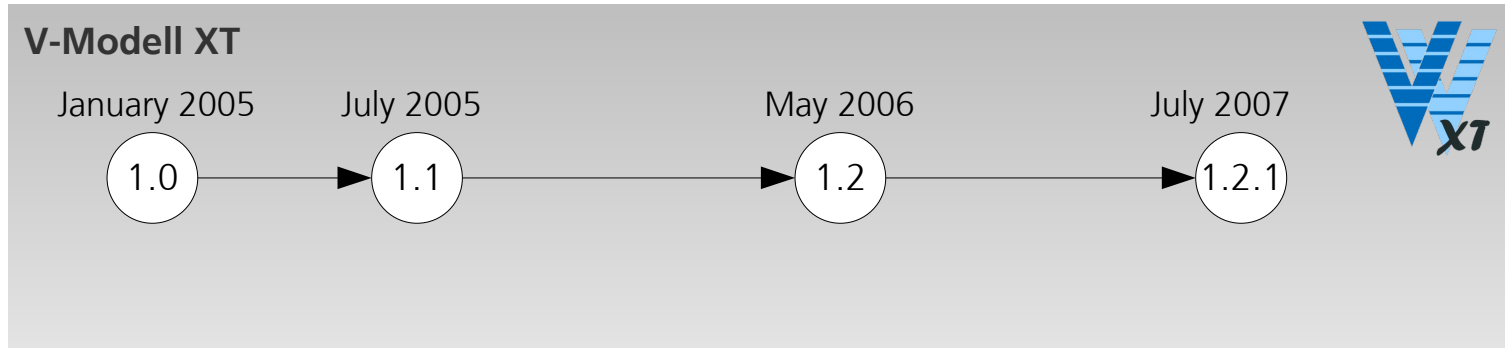
Execution of the Exploratory Study

Goals & Object of Study

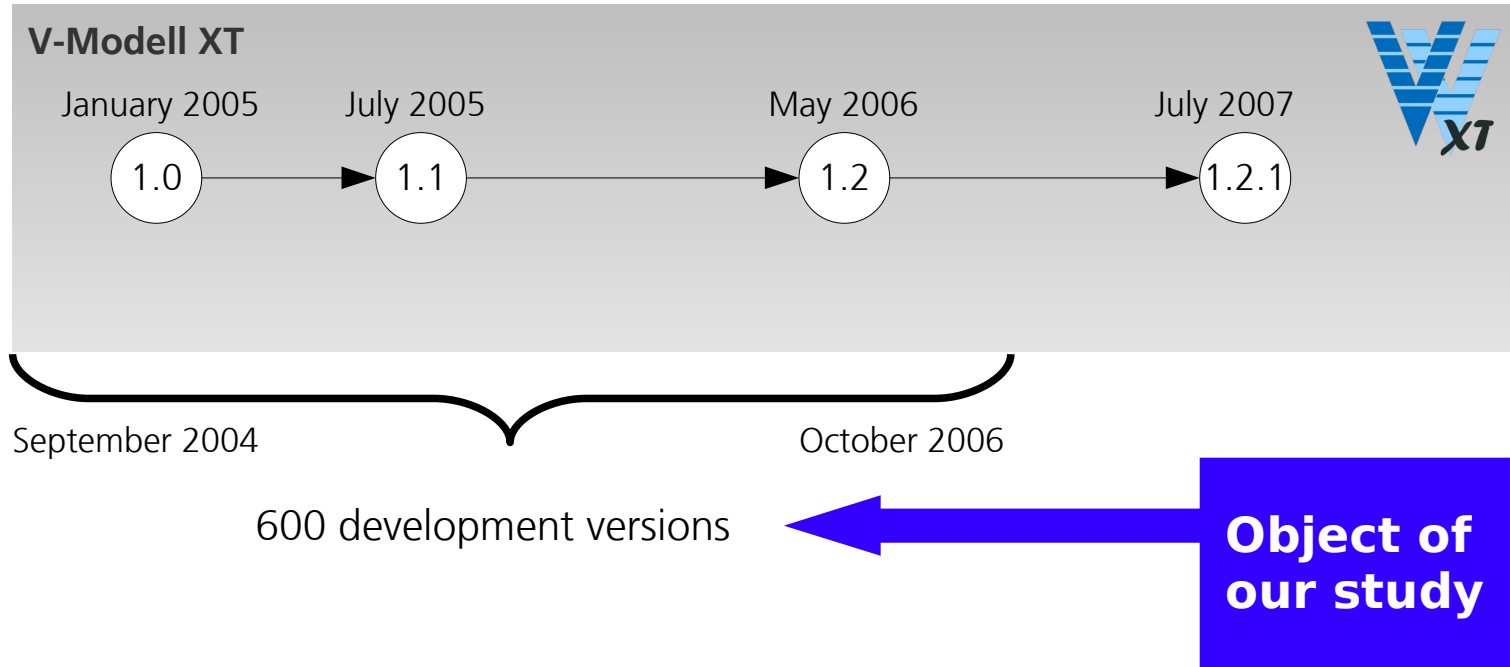


- German public standard software process description.
- Being increasingly adopted by the private sector.
- A prescriptive process model, not just requirements.
- Very large and complex (over 1.000 entities, 600 documentation pages)

The Evolution of the V-Modell XT



The Evolution of the V-Modell XT



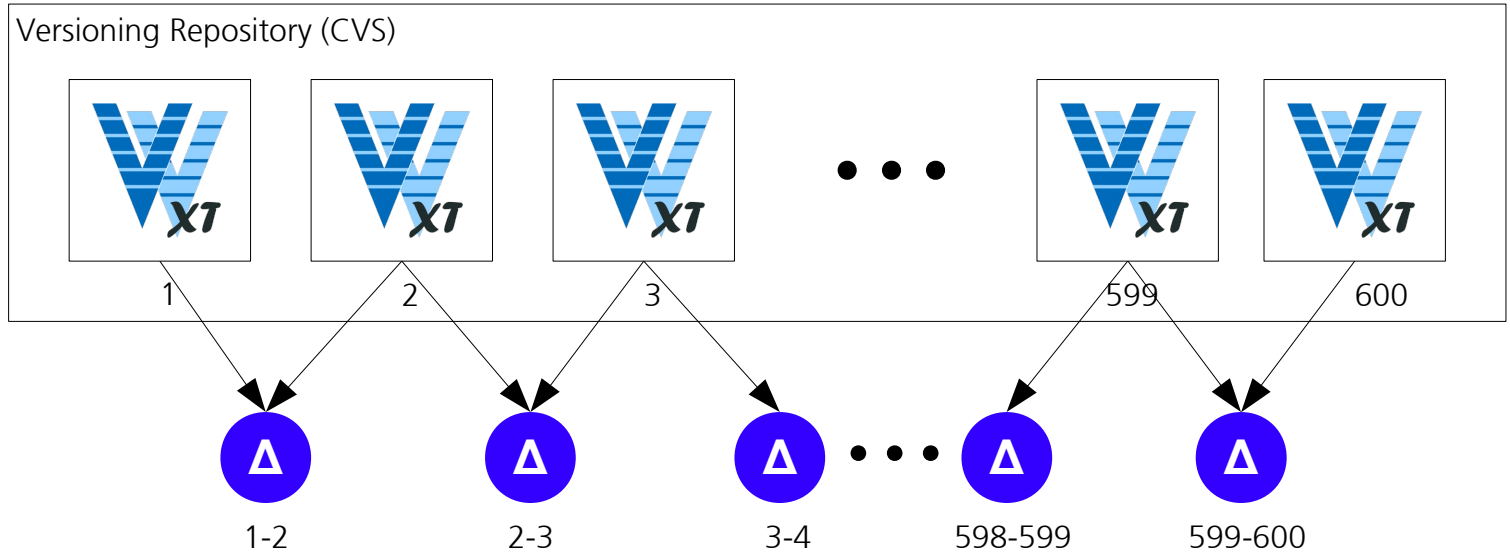
Version History



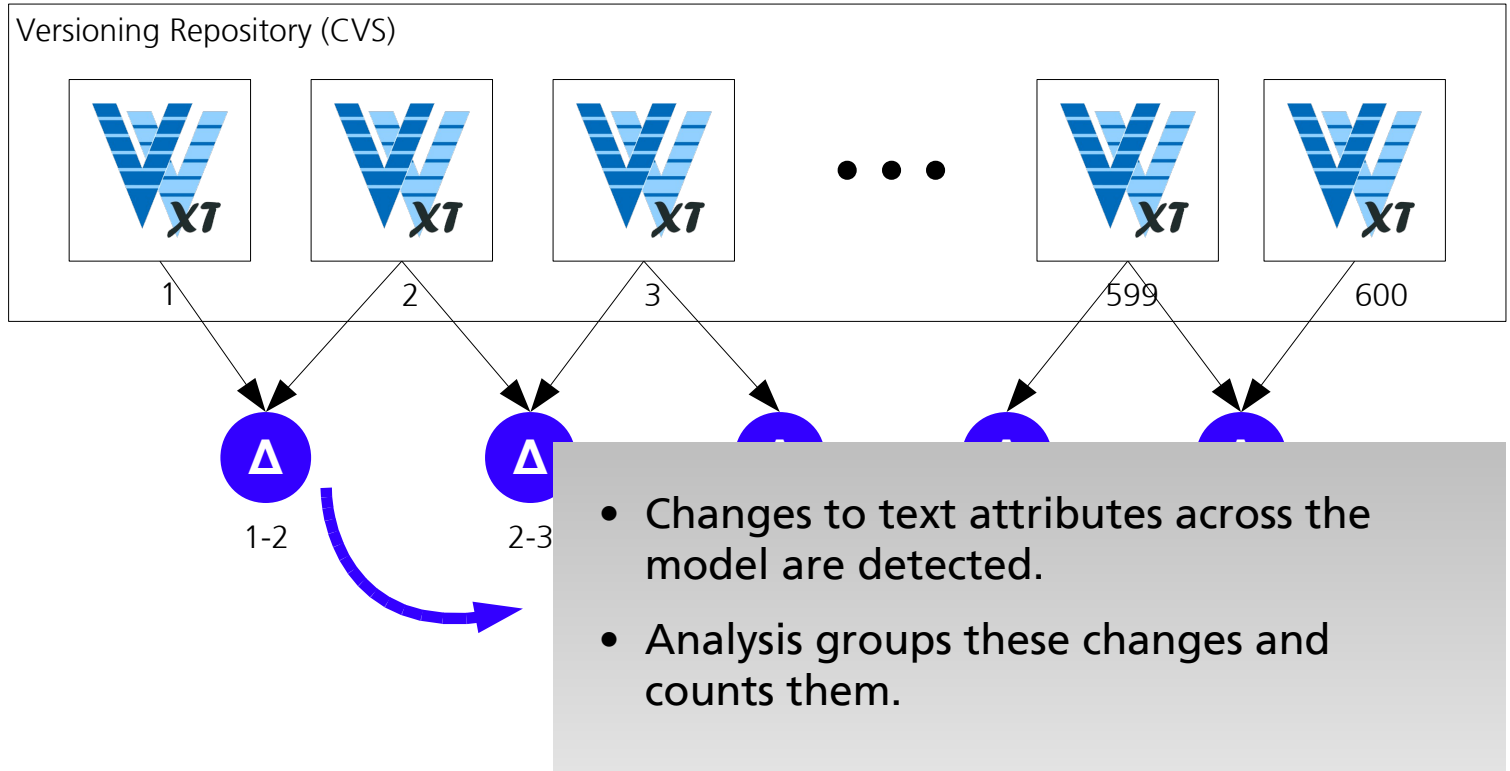
- Single XML files.
- Numbered in chronological order.
- Date, author and version log information available.

Slide 12

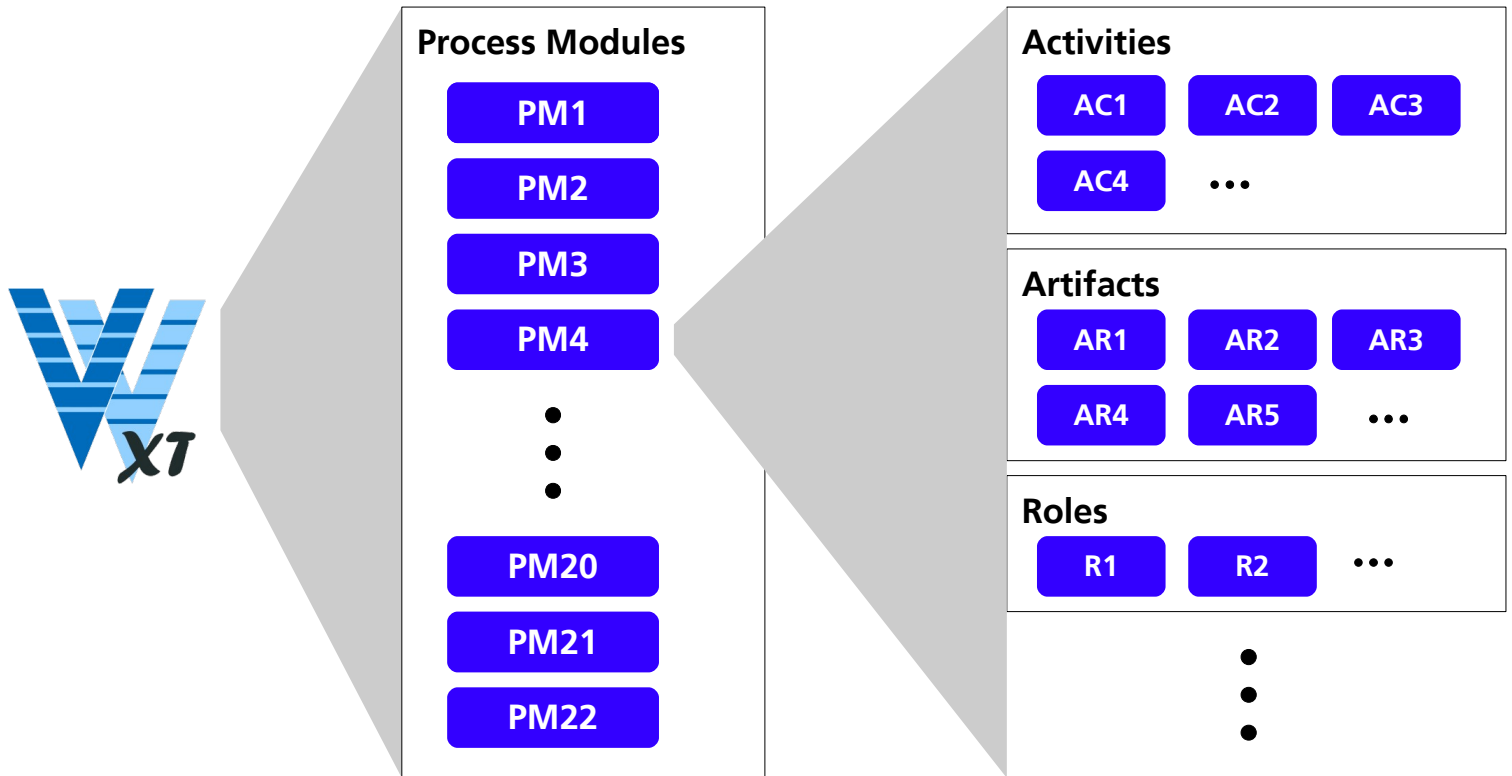
Analysis Approach



What's in a Version Comparison?



The Structure of the V-Modell XT



Motivation

Execution (1)

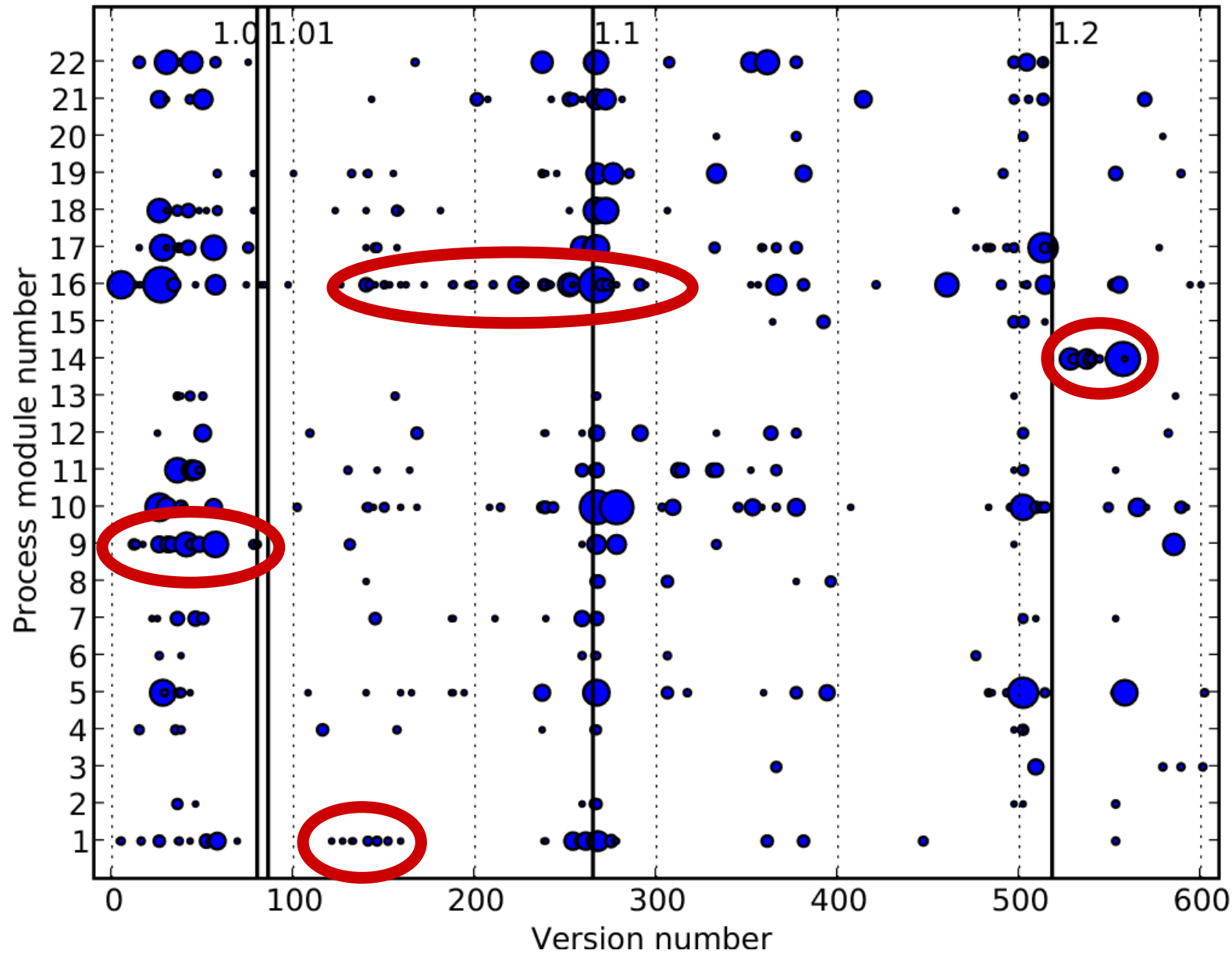
Observations

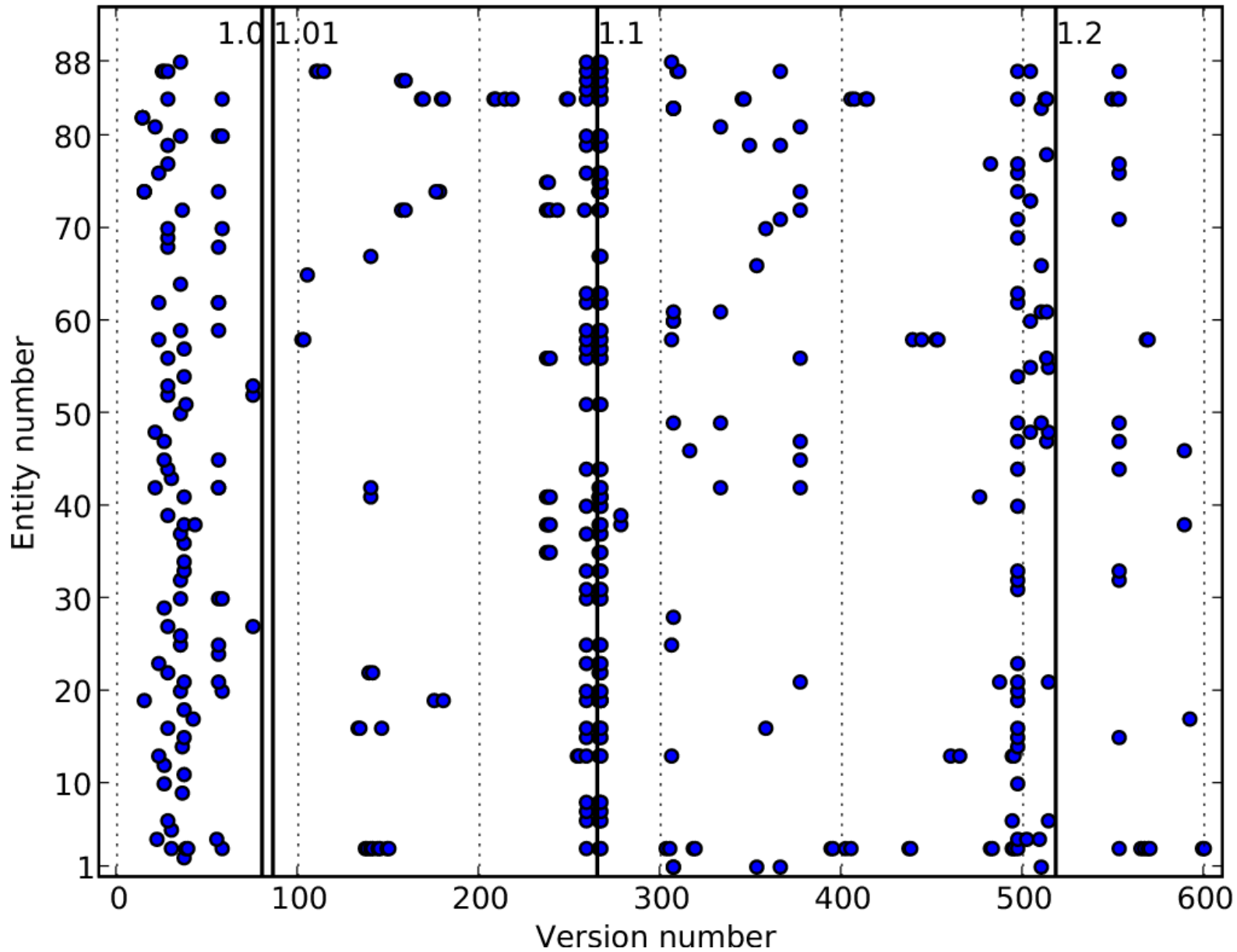
Execution (2)

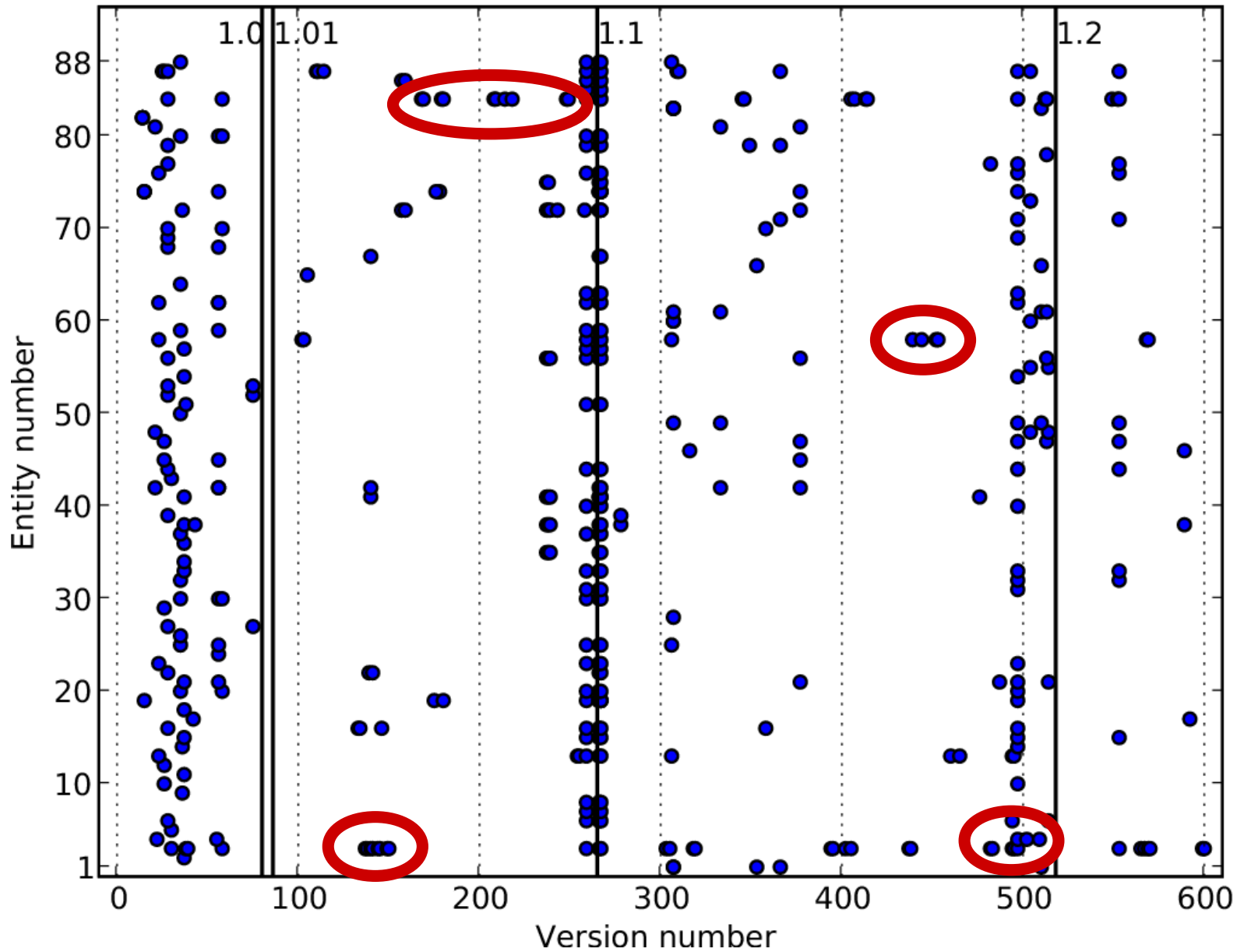
Results

Conclusions

Observations of the Exploratory Study







Specific Question

Motivation

**Do changes in the
model often lead to
follow-up changes?**

Slide 21

Motivation

Execution (1)

Observations

Execution (2)

Results

Conclusions

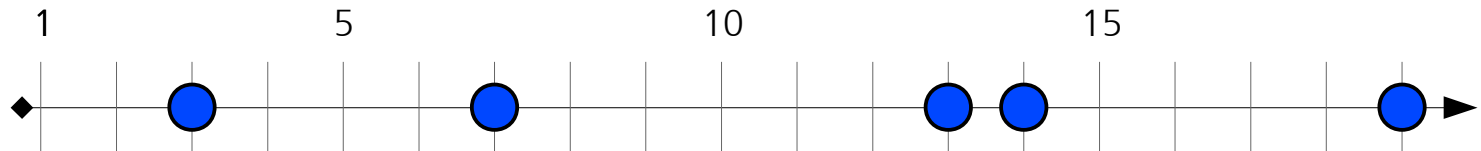
Execution of the Current Study

Hypotheses

- H1: Changing a process model entity in a particular version increases the probability of changing it again in subsequent versions.
- H2: Changing a process model entity at a given date increases the probability of changing it again in the following days.

Null Hypotheses (Versions)

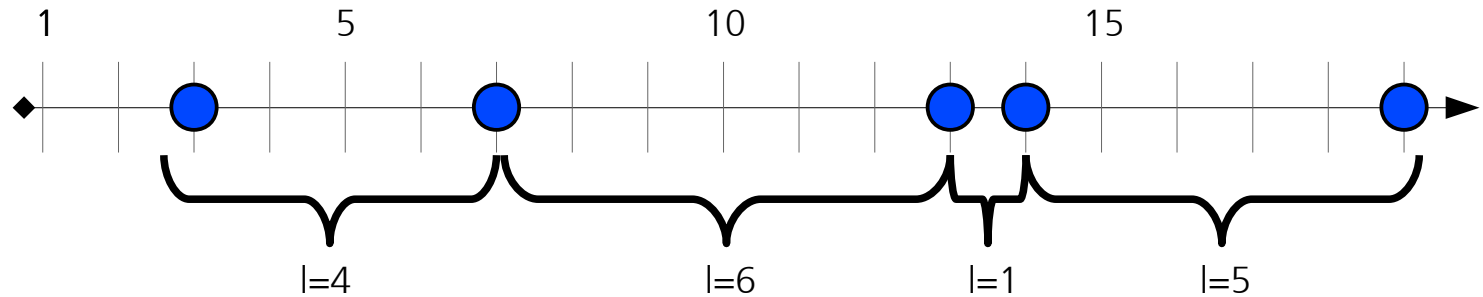
Entity History (Versions)



Changes to entity models are aleatory, i.e., the probability of an entity being changed in a particular version is fixed.

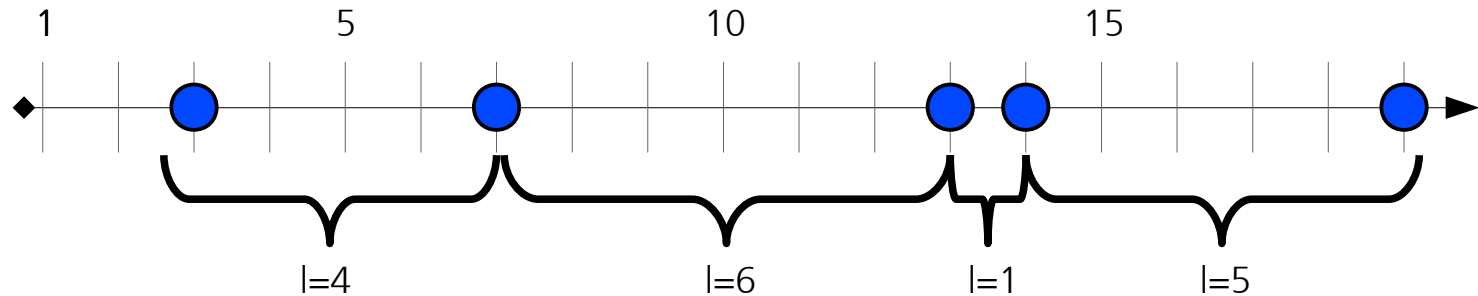
Pairs of Consecutive Changes

Entity History (Versions)



Operationalization of H10

Entity History (Versions)



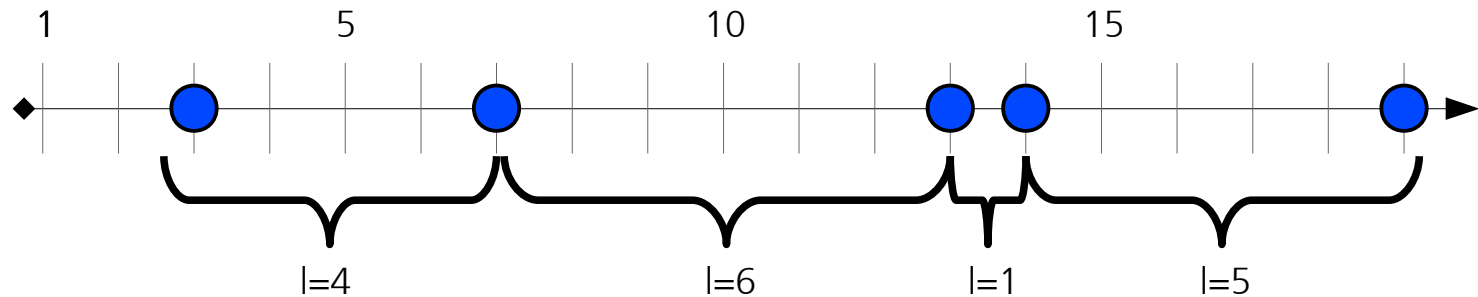
If H10 holds, the probability of a pair of changes having a length l is given by:

$$P(l) = (1 - p_v)^{l-1} \cdot p_v$$

where p_v is the probability of a change occurring at a particular version (geometric distribution).

Data Collection & Analysis

Entity History (Versions)



- For all entities with at least two changes in their whole history, collect the lengths of all pairs.
- Analyze the distribution (histogram) of this lengths.
- Test goodness of fit with respect to geometric distribution.

Motivation

Execution (1)

Observations

Execution (2)

Results

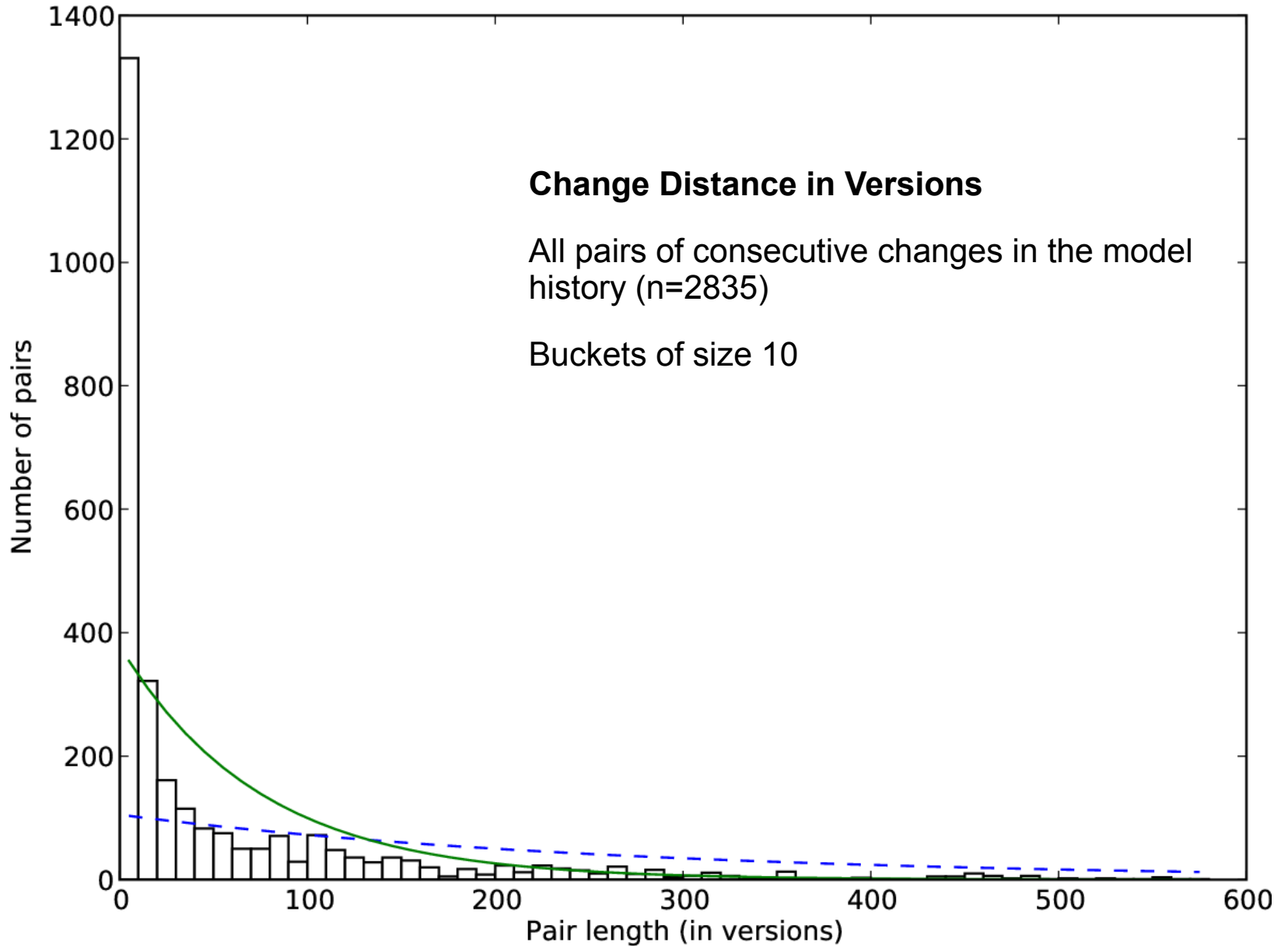
Conclusions

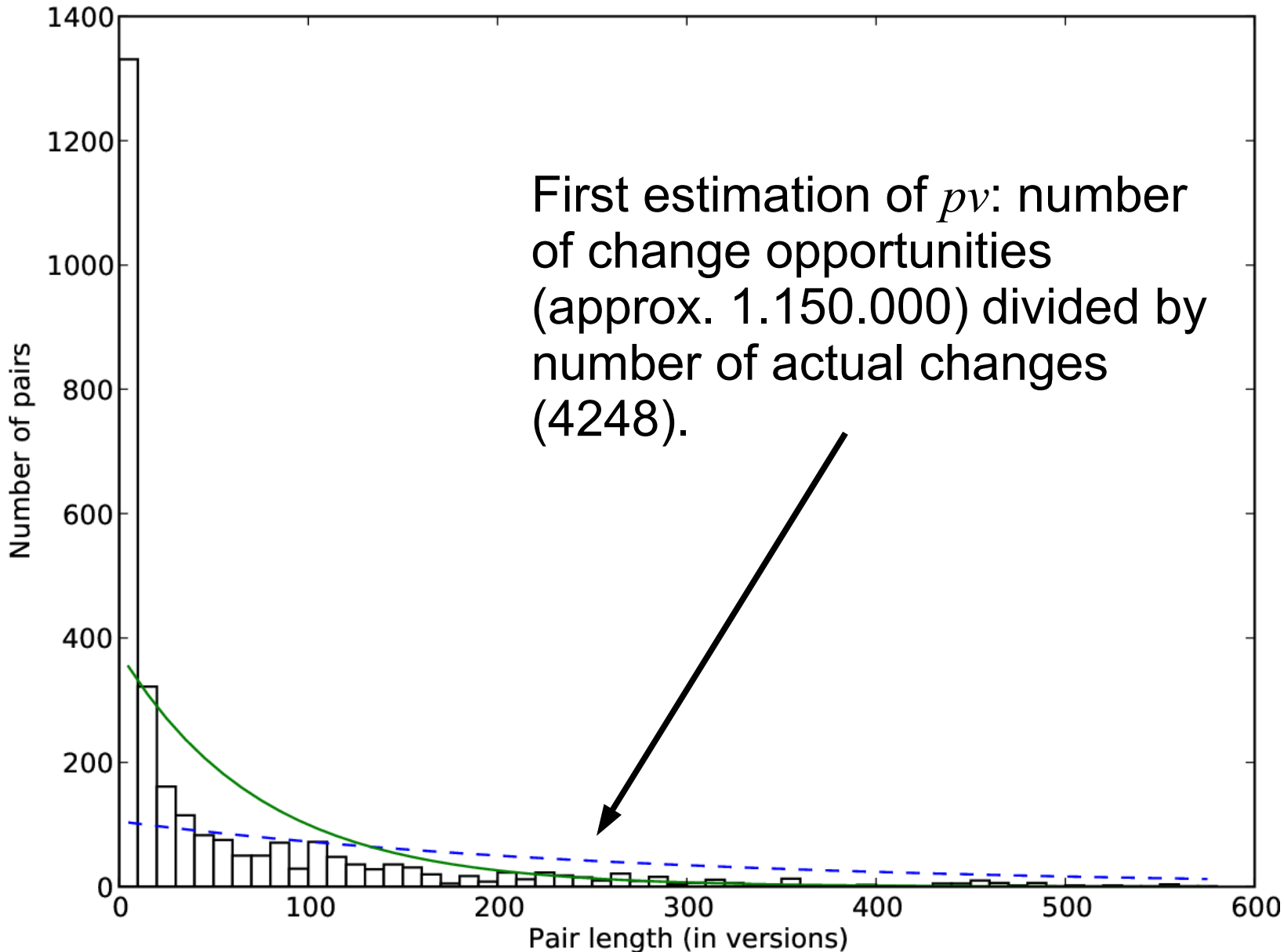
Results & Analysis

Change Distance in Versions

All pairs of consecutive changes in the model history (n=2835)

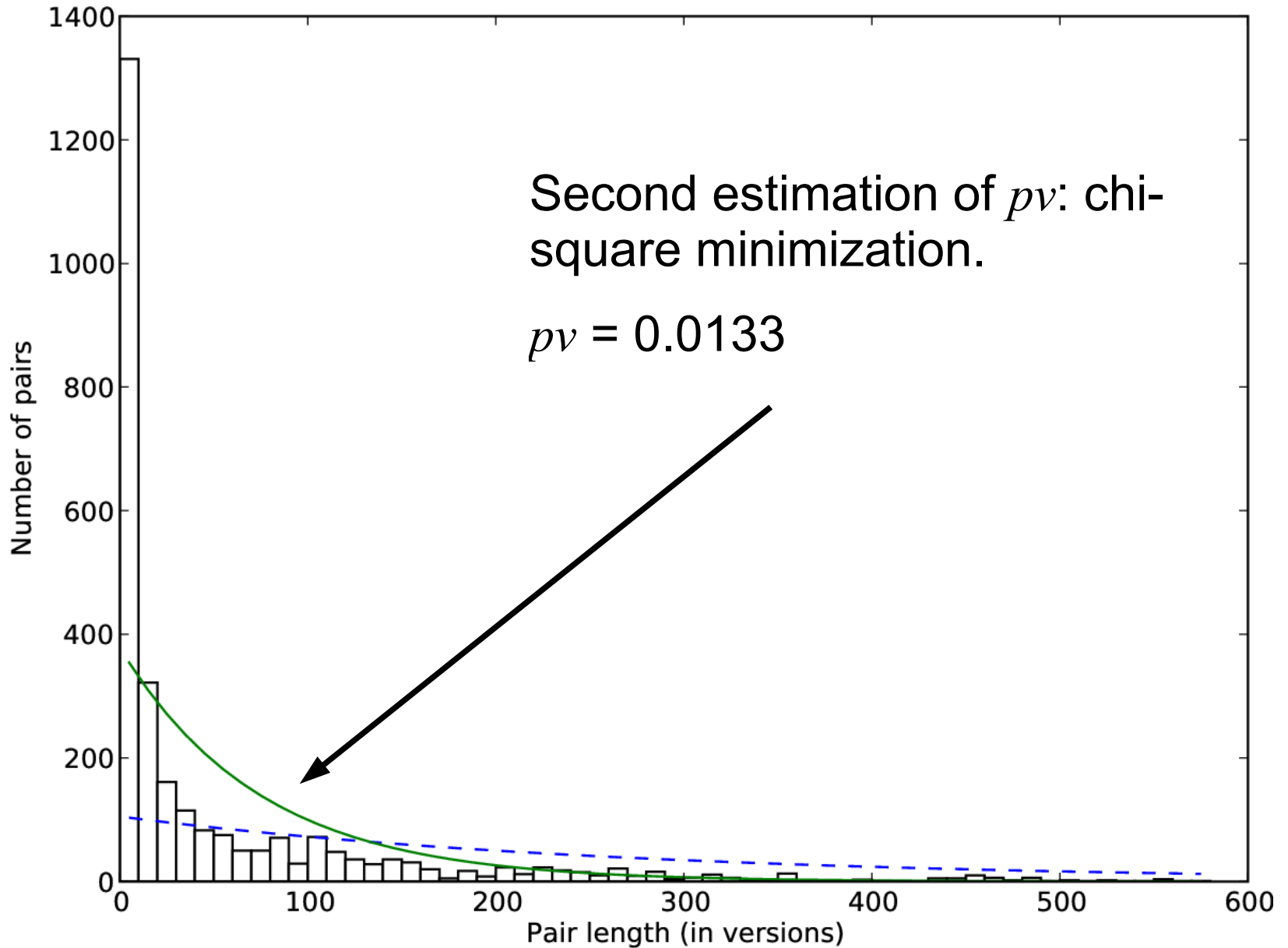
Buckets of size 10





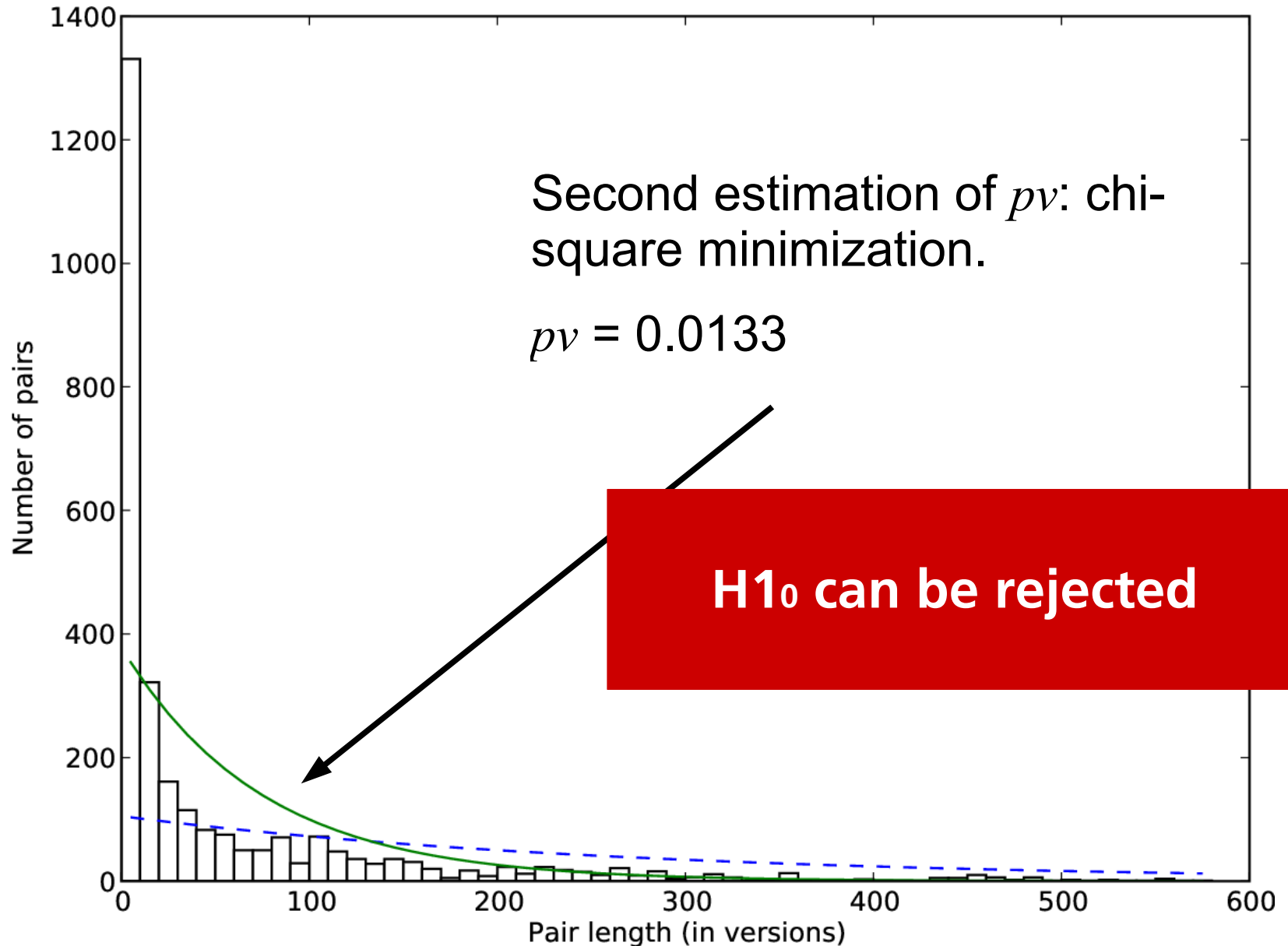
Second estimation of p_v : chi-square minimization.

$$p_v = 0.0133$$



Second estimation of p_v : chi-square minimization.

$$p_v = 0.0133$$

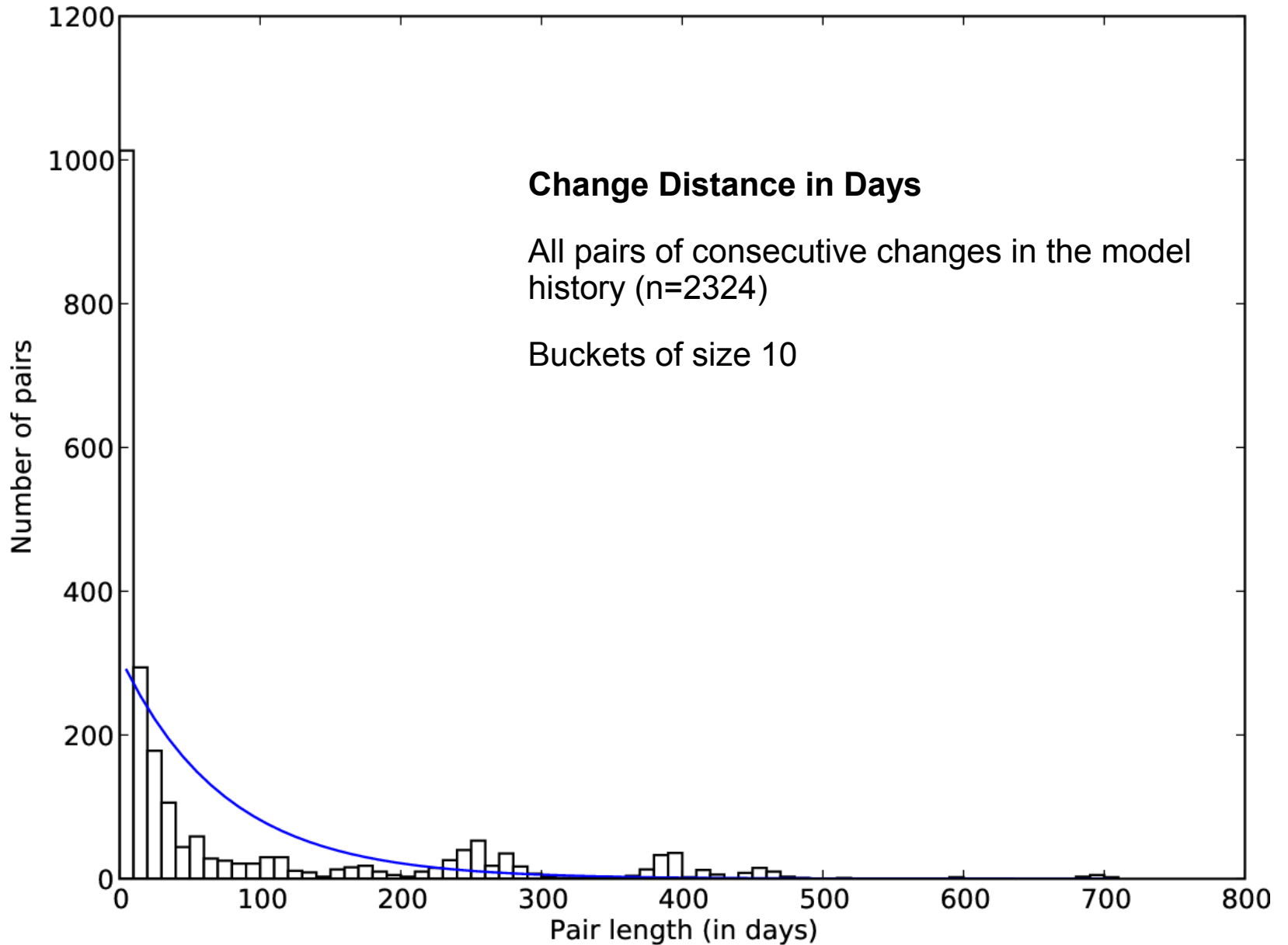


H1₀ can be rejected

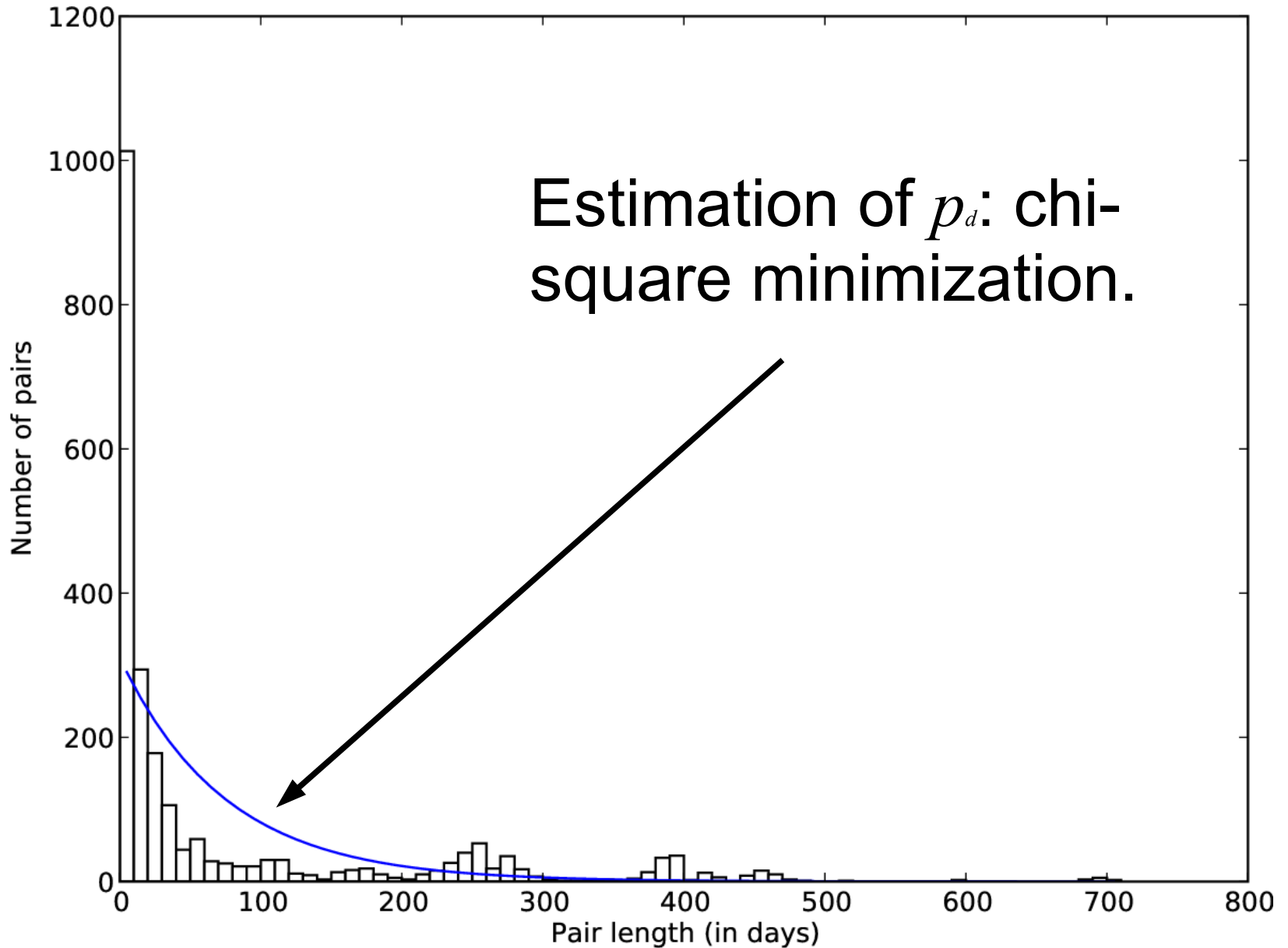
Change Distance in Days

All pairs of consecutive changes in the model history (n=2324)

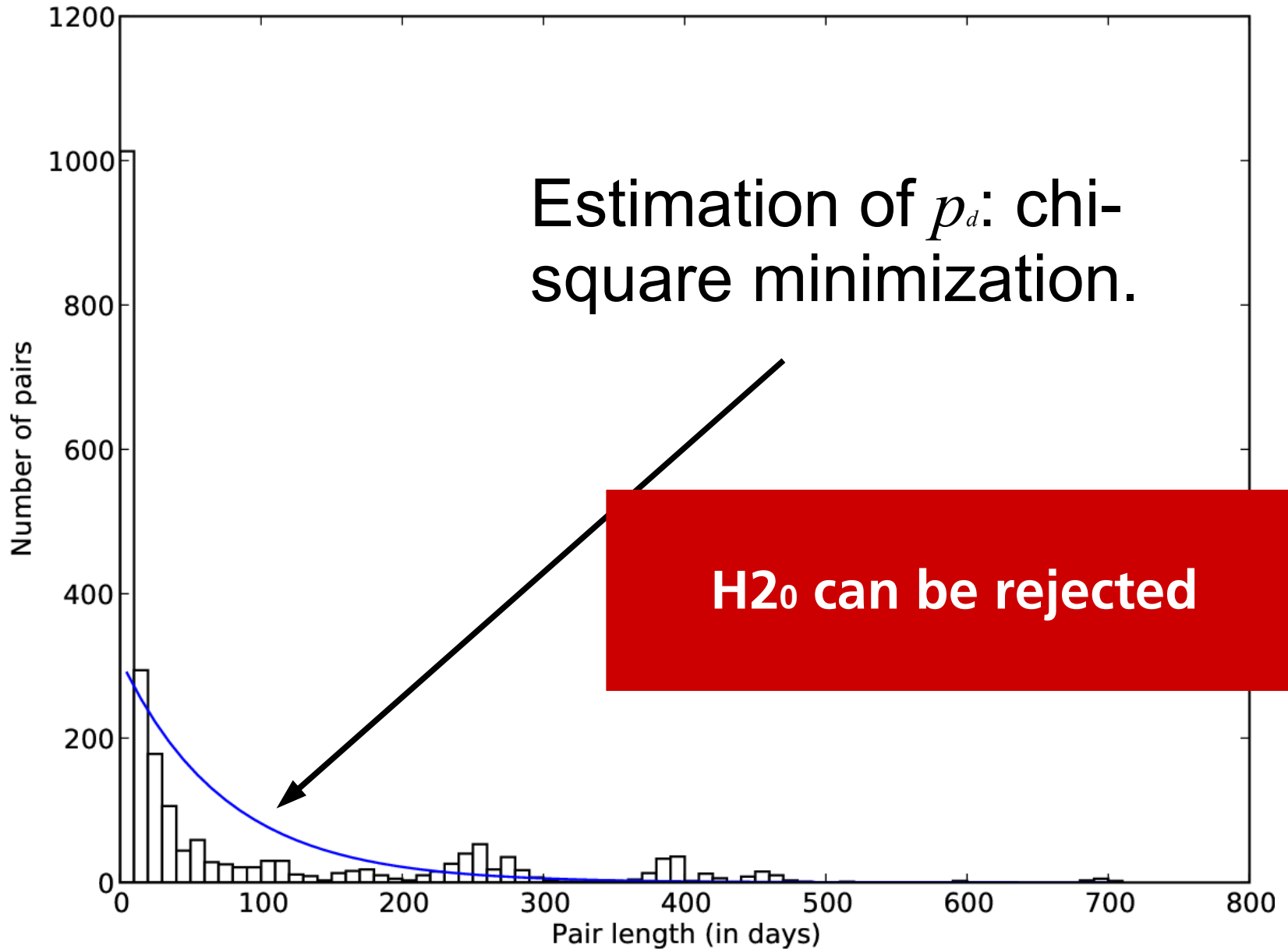
Buckets of size 10



Estimation of p_d : chi-square minimization.



Estimation of p_d : chi-square minimization.



H20 can be rejected

Motivation

Execution (1)

Observations

Execution (2)

Results

Conclusions

Conclusions

Conclusions

Conclusions

- Both null hypotheses can be rejected. Our original hypotheses can be accepted, that is, changes increase the probability of changes in the near future.
- Consequence: large changes probably “destabilize” the model and must be planned carefully.
- The previous effect can also be observed during the evolution of software systems.

Slide 37

Threats to Validity

Conclusions

- We see a dependency between changes, but have not discarded all alternative explanations.
- Consecutive changes may happen due to development strategy. This can only be discarded by manually looking at the change logs.
- Given our scarce knowledge regarding model evolution, we cannot yet generalize. We can expect however, that these results apply to some extent to the similar models (e.g., RUP).

Slide 38

Future Work

Conclusions

- Look into the specific causes of model changes.
- Investigate similar issues in other models, e.g., RUP.

Slide 39

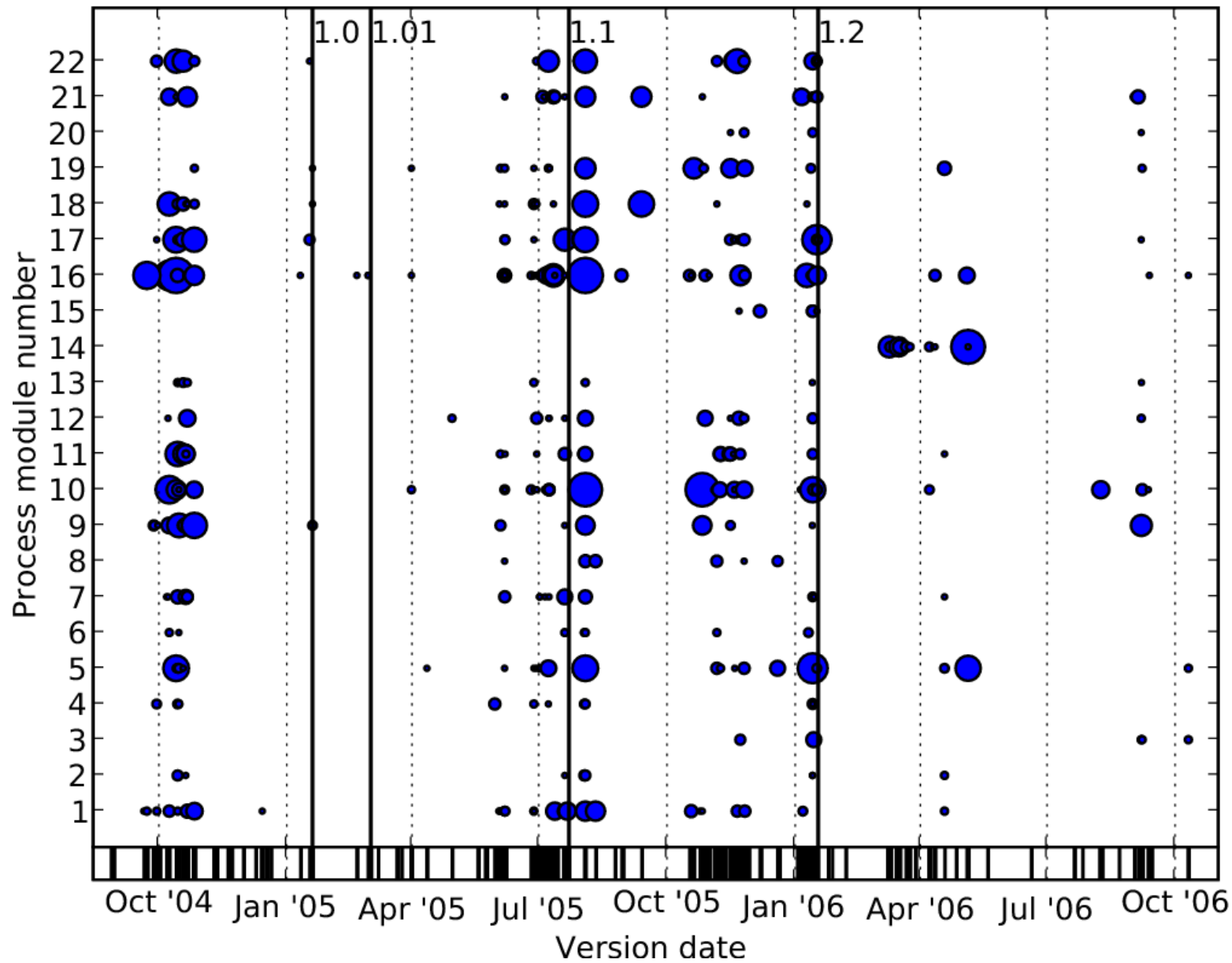
**Questions?
Comments?**

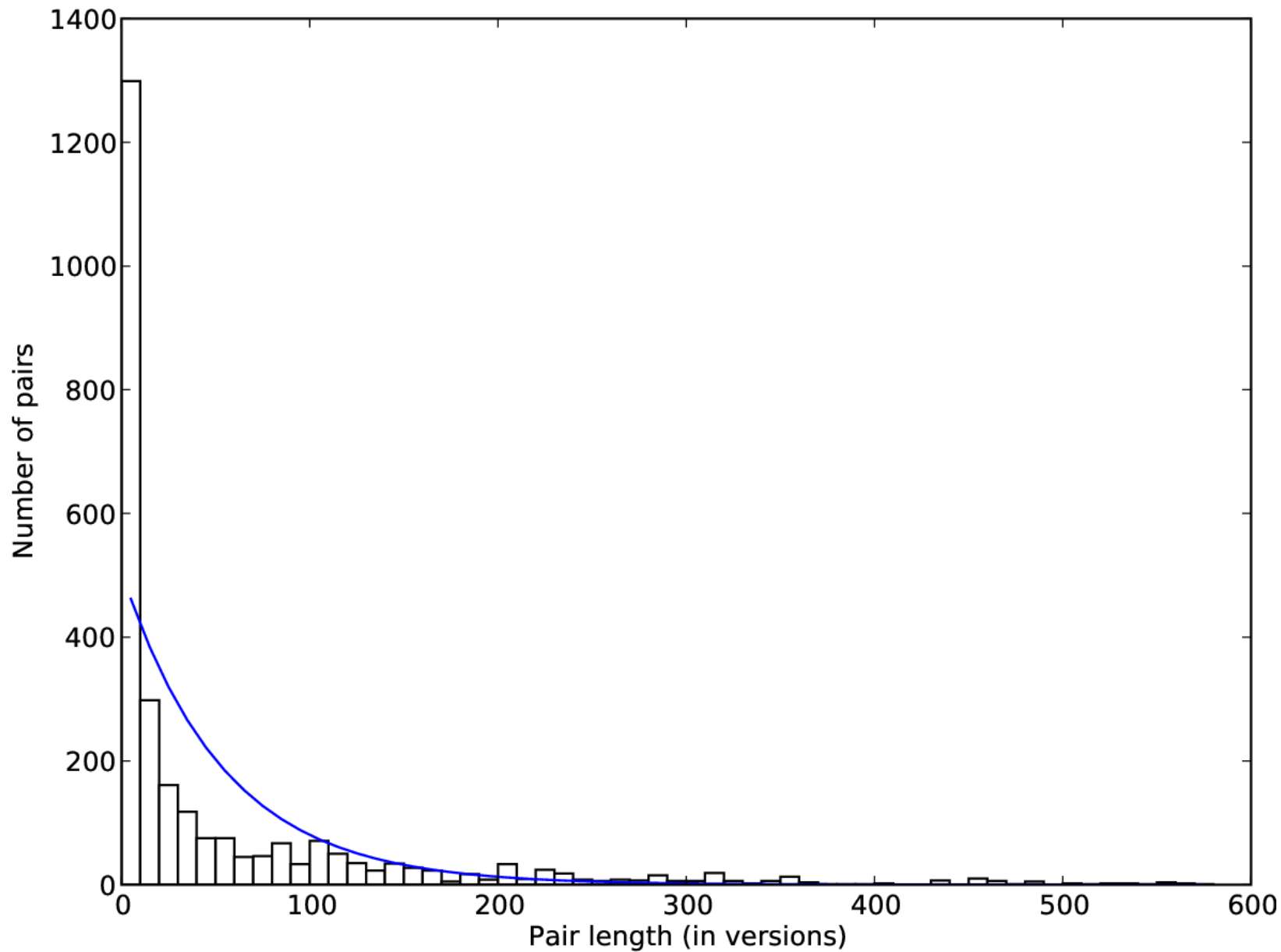
Backup Slides

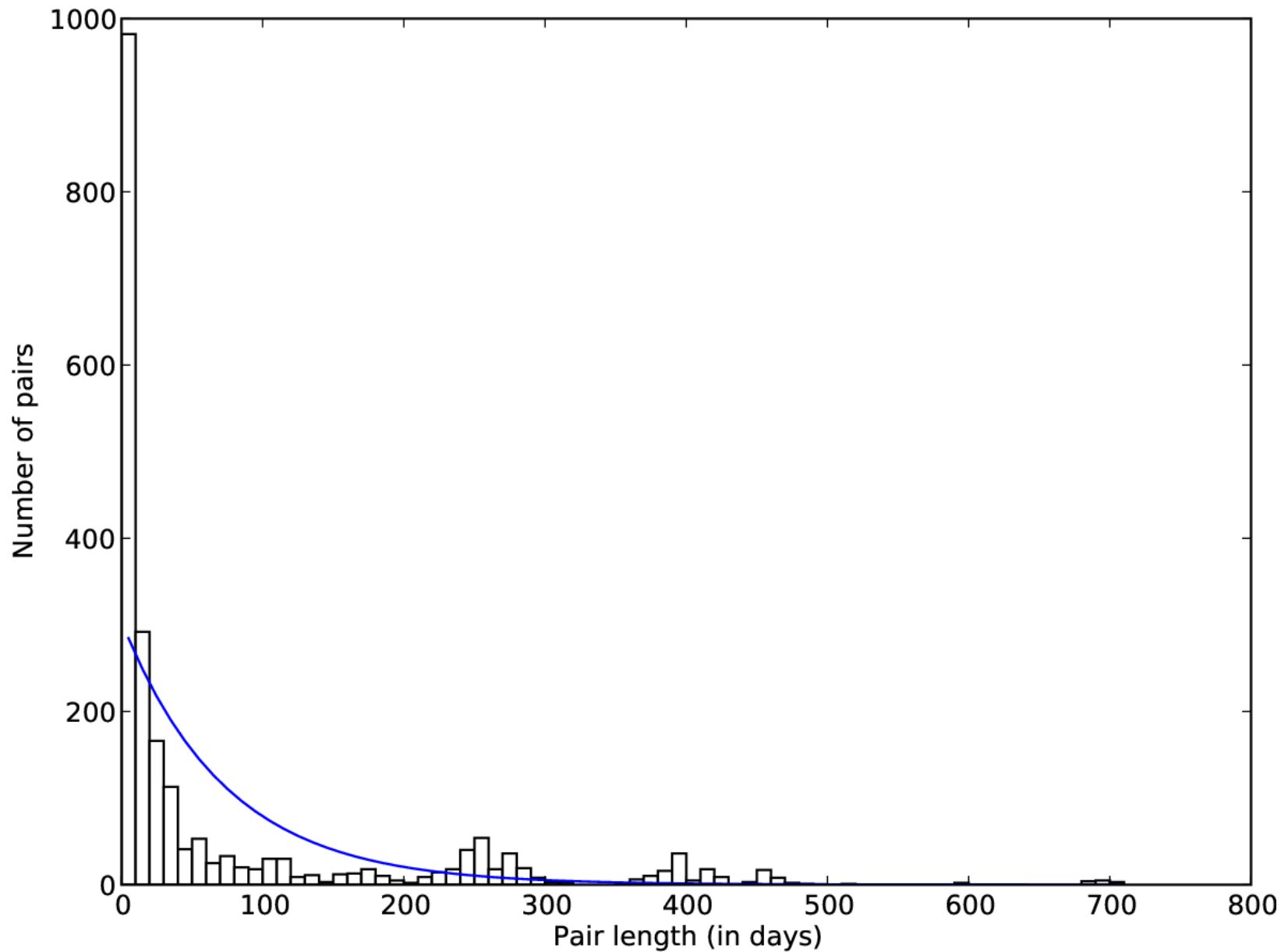
Are Software Processes Software Too?

- Short answer: not really.
 - We still strongly rely on natural-language text.
- Long(er) answer:
 - Process descriptions are often large and complex.
 - We use formalization in order to keep complexity under control: process entities connected by relations.
 - Many entity types, attributes, relations...

Slide 42







Null Hypotheses (Versions)

- H1₀: Changes to a process model entity in a particular version are independent from changes to the same entity in all other versions. Moreover, there is a fixed probability p_v of changes occurring to an entity in a particular version, for all versions and for all entities present in each version.

Null Hypotheses (Time)

Execution

- H2₀: Changes to a process model entity on a particular day are independent from other changes to the same entity. Moreover, there is a fixed probability p_d of changes occurring to an entity on a particular day, for all days in the studied period and for all entities present in the model during that period.