Answering the Call of the Wild?: Thoughts on the Elusive Quest for Ecological Validity in Variability Modelling

Seiede Reyhane Kamali, Shirin Kasaei, Roberto E. Lopez-Herrejon

École de technologie supérieure, Université du Québec, Canada

Ecological Validity – Form of External Validity

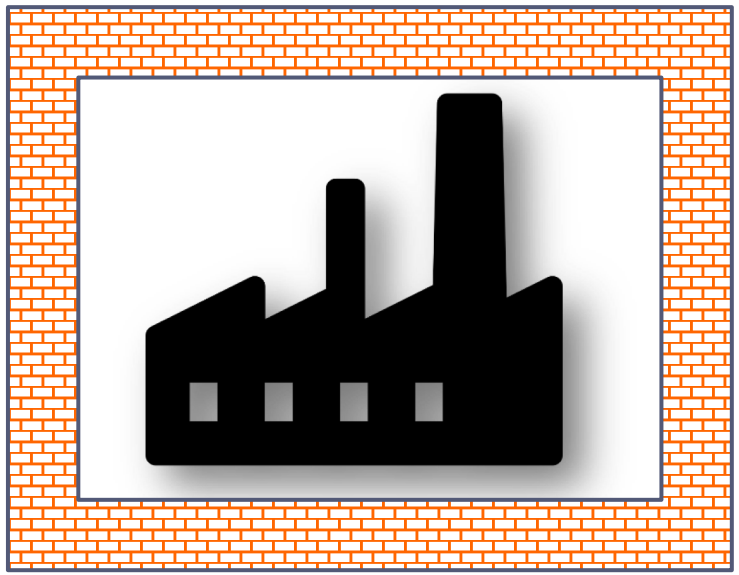
- Poses the question:
  - Is a study, method, artifacts, setting an approximation of what happens in real-life?

In the wild …

In the lab …

In SPL, what is “real-life”?

Industry



- Limited access
- Unfeasible reproducibility

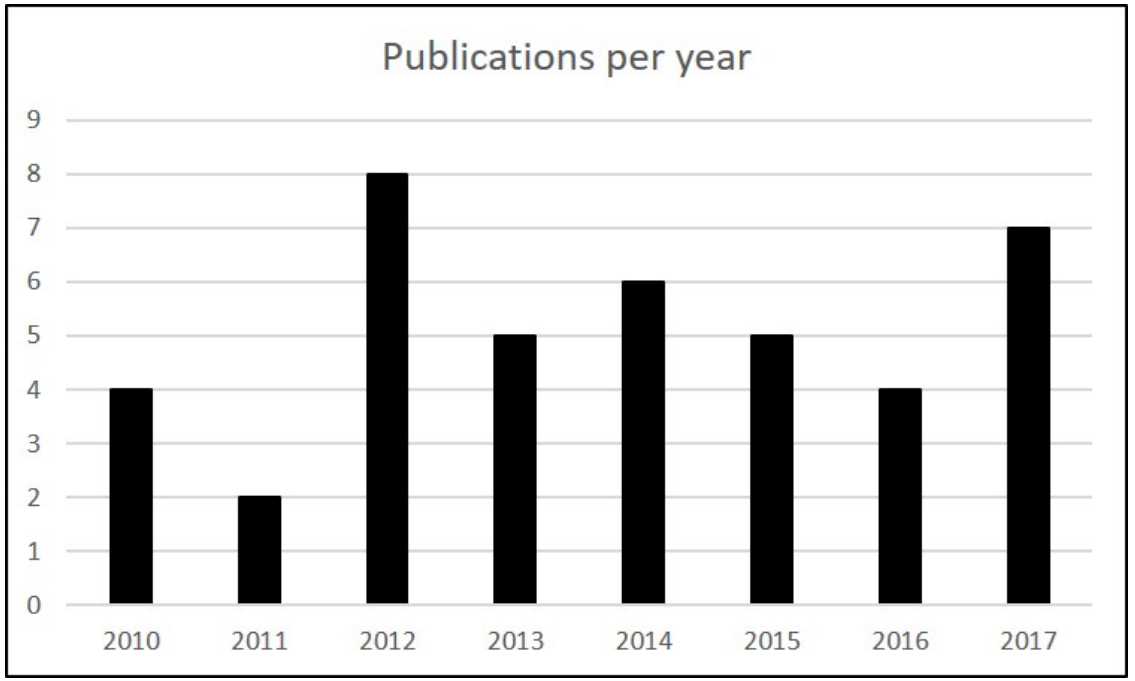
- Transparent access
- Enables reproducibility

Study description

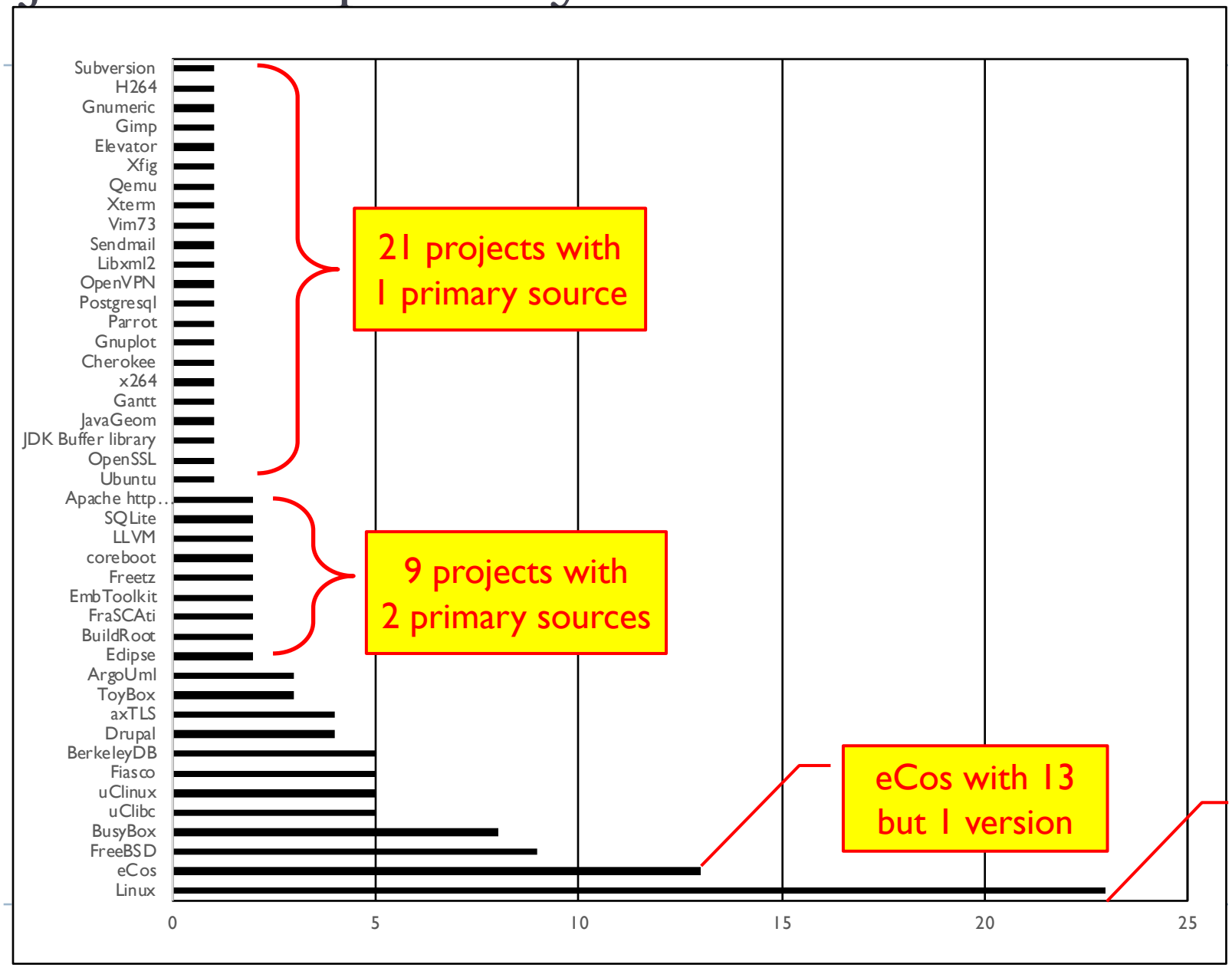
- General question:
  - What is the impact of open source projects in SPL research?
- Our paper focus:
  - Variability Modeling
- Starting point:
  - José A. Galindo, David Benavides, Pablo Trinidad, Antonio Manuel Gutiérrez Fernández, and Antonio Ruiz-Cortés. 2019. Automated analysis of feature models: Quo vadis? Computing 101, 5 (2019), 387–433.
  - Corpus of 242 articles from 2017 to 2017
- Open source project – less restrictive interpretation:
  - Publicly available code repository (URL)
  - More than one author
  - No academic projects

First findings

- Primary sources selected: 41
- Open source projects identified: 43
- Publications per year:
  - Median 5

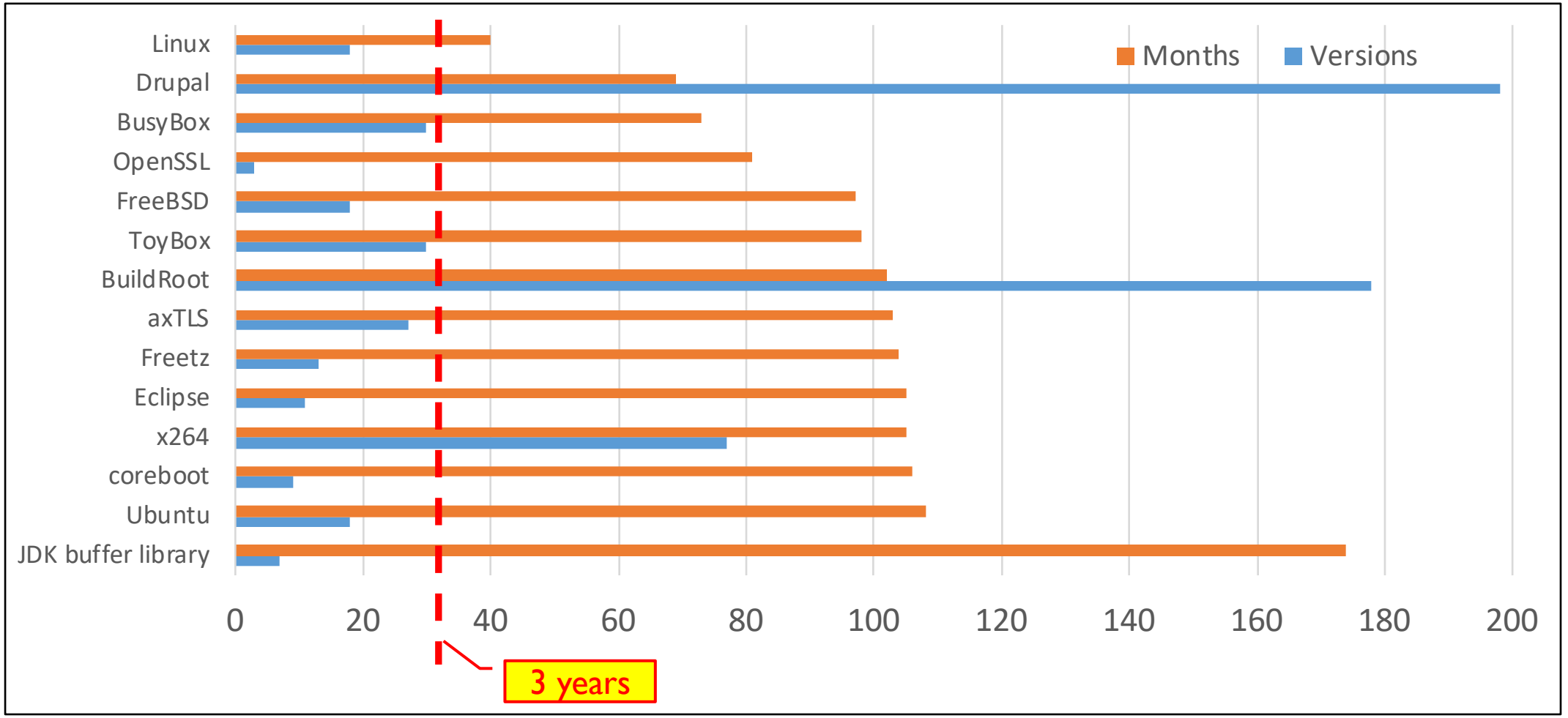


Projects and primary sources

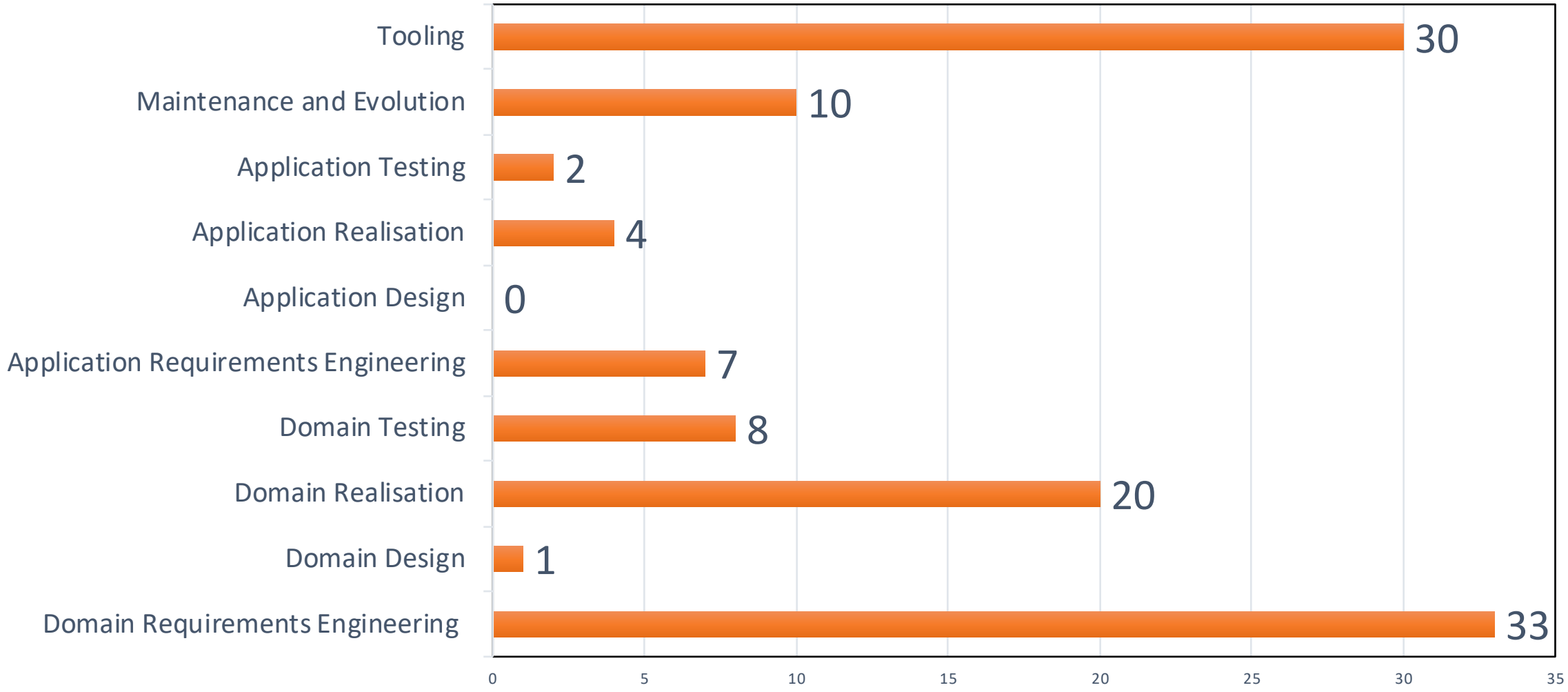


- Inconsistent or incomplete report of:
  - Version numbers
  - Number of features
- Projects with more than one version reported:
  - Eclipse (7)
  - BusyBox & FraSCati (3)
  - Drupal & uClibc (2)

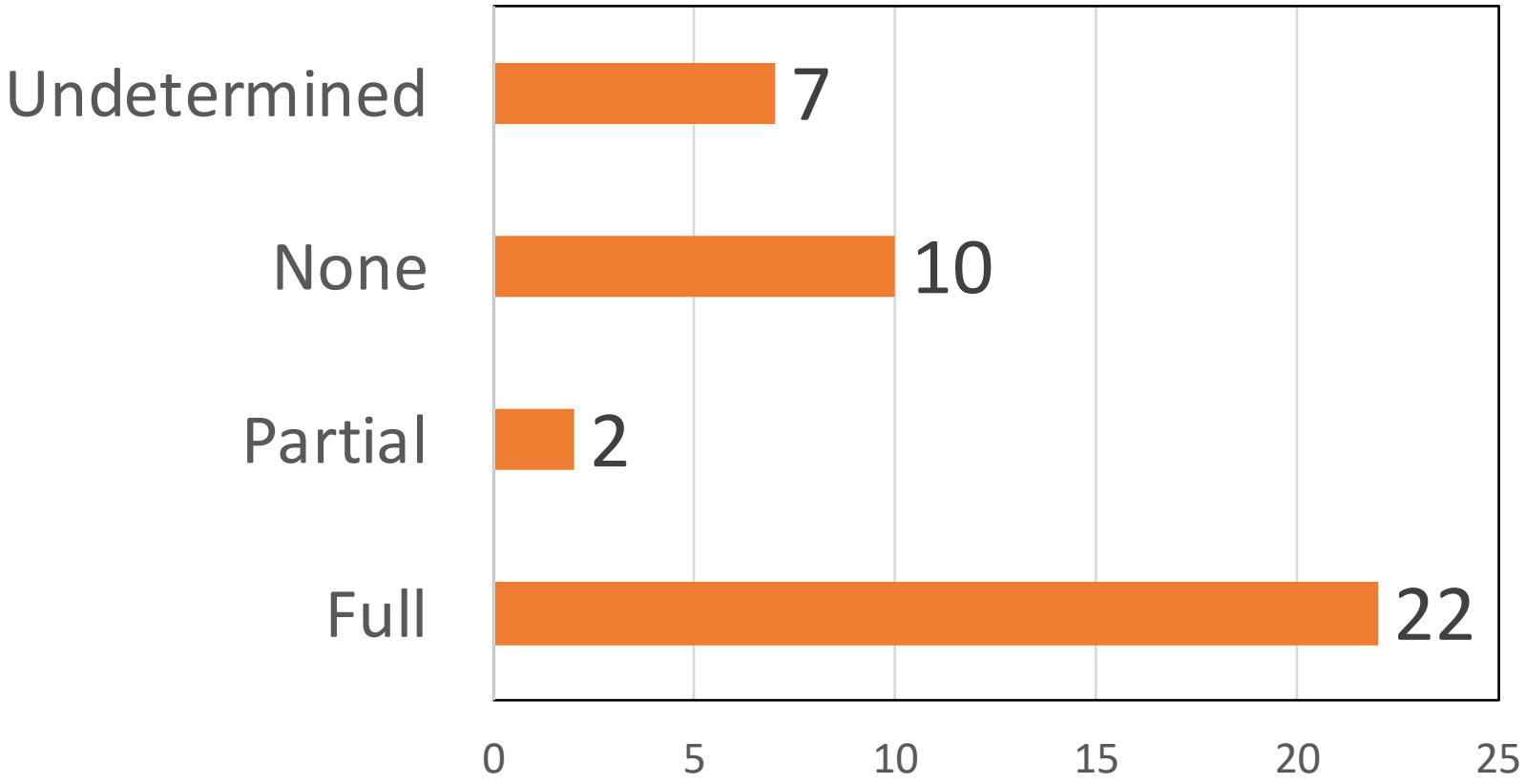
Recency – gap between versions analyzed and released



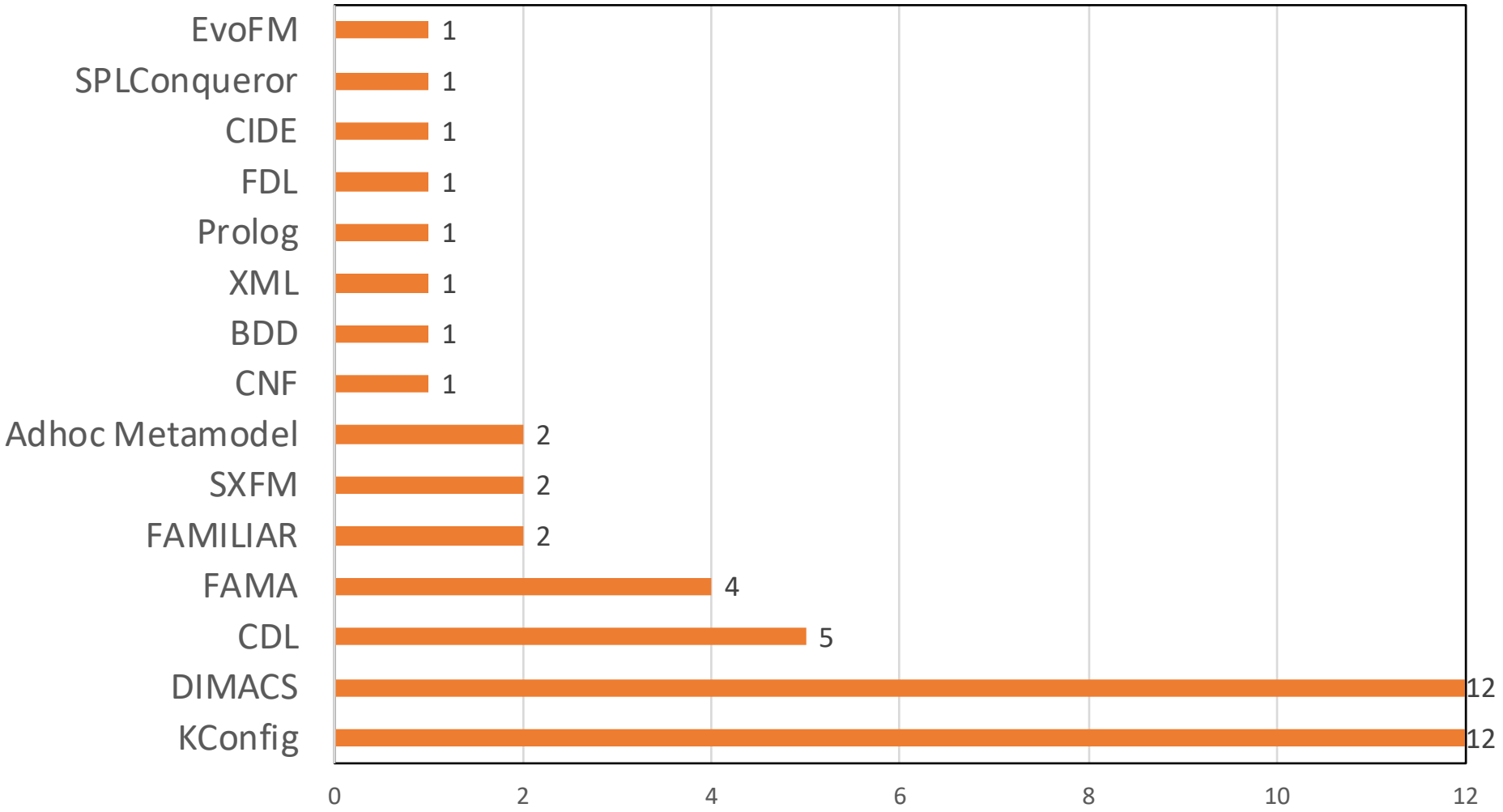
Feature Models – What for?



Availability for replication



Model Formats



Analysis (1)

- Predominance of Linux and KConfig
  - Number of primary sources, number of versions, most recent, …
  - Tool ecosystem built around it
- Aging and outdated datasets
  - Only Linux has been updated but still at a lower rate than releases
- Artifacts beyond feature models and source code
  - Open sources projects are used for multiple purposes hence the need to study other types of artifacts → e.g. faults, test cases, test code, …

Analysis (2)

- Replication as a pillar for empirical SPL research
  - Reproducibility should be the norm not the exception
- Building bridges to open source communities
  - Linux has had enormous impact on SPL research, but how about the other way around?

Future Work

- Expand the focus of our study on impact of open source projects in SPL research to:
  - Primary sources beyond Galindo et al.'s work
  - Other artifacts → related to testing activities

- Other open source projects
  - Based on KConfig
  - Mining Repositories community at large
  - Open source communities at large

The End

Any questions?

Roberto Erick Lopez Herrejon
École de technologie supérieure
Université du Québec
Montreal, Canada
Roberto.lopez@etsmtl.ca

Natural Sciences and Engineering Research Council of Canada