

### Classifying Edits to Variability in Source Code

Paul Bittner, Christof Tinnes, Alexander Schultheiß, Sören Viegener, Timo Kehrer, and Thomas Thüm | Feb 24, 2023

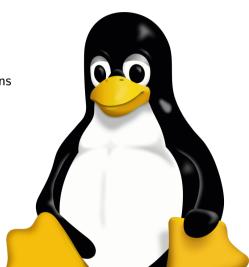


# **Software Comprises Massive Evolving Variability**

≥ **4,000** configuration options

 $\geq 10^{725}$  different variants

[2007]

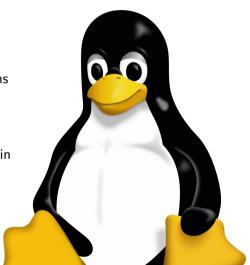


# **Software Comprises Massive Evolving Variability**

 $\geq$  **4,000** configuration options

 $\geq 10^{725} \\ \text{different variants} \\ \text{Only } \sim \! 10^{80} \text{ atoms in} \\ \text{observable universe!}$ 

[2007]

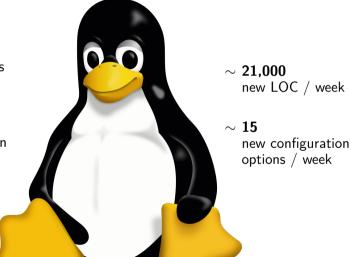


# **Software Comprises Massive Evolving Variability**

 $\geq$  **4,000** configuration options

 $\geq 10^{725} \\ \text{different variants} \\ \text{Only } \sim \!\! 10^{80} \text{ atoms in} \\ \text{observable universe!}$ 

[2007]



```
static void
f_foreground(/* params */)
{
#ifdef FEAT_GUI
    if (gui.in_use)
        gui_mch_set_foreground();
#else
# ifdef MSWIN
    win32_set_foreground();
# endif
#endif
}
```

```
static void
                                                              f_foreground(/* params */)
                                      FEAT\_GUI
                                                                 if (gui.in_use)
   static void
                                                                 gui_mch_set_foreground();
f_foreground(/* params */)
#ifdef FEAT GUI
   if (gui.in_use)
   gui_mch_set_foreground();
#else
# ifdef MSWIN
   win32 set foreground():
# endif
#endif
```

Vim Commit ab4cece.

```
static void
                                                             f_foreground(/* params */)
                                     FEAT\_GUI
                                                                if (gui.in use)
   static void
                                                                gui mch set foreground();
f_foreground(/* params */)
#ifdef FEAT GUI
                                  ¬FEAT_GUI, MSWIN
                                                                static void
   if (gui.in_use)
                                                             f_foreground(/* params */)
   gui_mch_set_foreground();
#6166
                                                                win32 set foreground():
# ifdef MSWIN
   win32 set foreground():
# endif
#endif
```

```
static void
                                                              f_foreground(/* params */)
                                      FEAT\_GUI
                                                                 if (gui.in use)
   static void
                                                                 gui mch set foreground();
f foreground(/* params */)
#ifdef FEAT GUI
                                   ¬FEAT GUI, MSWIN
                                                                 static void
   if (gui.in_use)
                                                              f_foreground(/* params */)
   gui_mch_set_foreground();
#6766
                                 \neg^{FEA}T\_{GUI}, \neg^{MSWIN}
                                                                 win32 set foreground():
# ifdef MSWIN
   win32 set foreground():
# endif
#endif
                                                                 static void
                                                               foreground(/* params */)
```

Vim Commit ab4cece.

## Edits to Variability via C Preprocessor

```
#ifdef A
foo();
#else
#ifdef B
baz();
#endif
#endif
```

## Edits to Variability via C Preprocessor

### Edits to Variability via C Preprocessor

```
#ifdef A
                       foo();
                     -#else
#ifdef A
                                                      #ifdef A
                     - #ifdef B
 foo();
                                                        foo();
                     + bar();
#else
                                                        bar();
                     +#endif
 #ifdef B
                                                      #endif
                     +#if B && (!A || C)
 baz();
                                                      #if B && (!A | | C)
                       baz():
 #endif
                                                        baz():
                     - #endif
#endif
                                                      #endif
                      #endif
```

### Related Work on Edit Classification is ...

### incomplete



[Stănciulescu et al., 2016] [Borba et al., 2012] [Al-Hajjaji et al., 2016] [Passos et al., 2016]

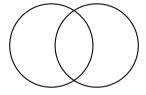
### Related Work on Edit Classification is ...

incomplete

or

ambiguous





[Stănciulescu et al., 2016] [Borba et al., 2012] [Al-Hajjaji et al., 2016] [Passos et al., 2016]

[Ji et al., 2015] [Stănciulescu et al., 2016]

### Related Work on Edit Classification is ...

incomplete

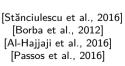
or

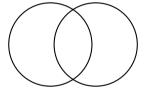
ambiguous

or

not automatable







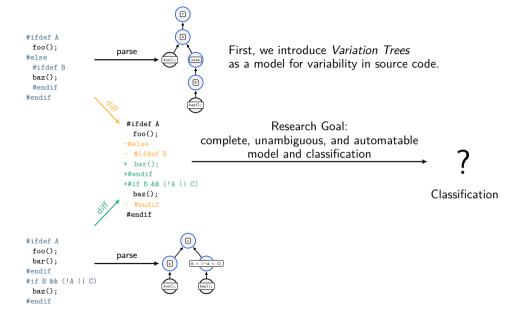
[Ji et al., 2015] [Stănciulescu et al., 2016]

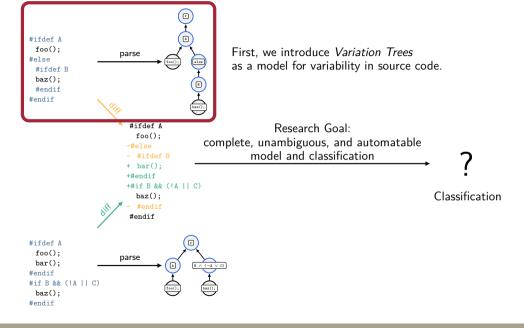


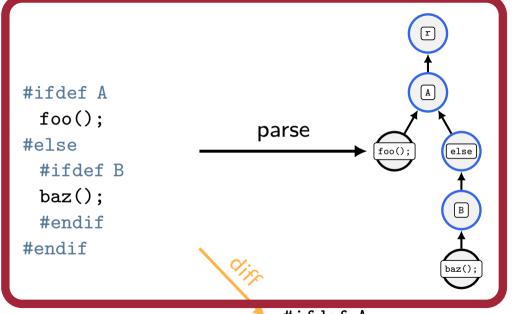
[Ji et al., 2015]

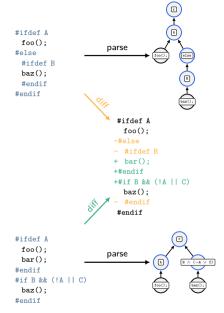
[Borba et al., 2012]

```
#ifdef A
 foo():
#else
 #ifdef B
 baz():
 #endif
#endif
                      #ifdef A
                                                       Research Goal:
                        foo():
                                       complete, unambiguous, and automatable
                     -#else
                                                  model and classification
                      - #ifdef B
                      + bar();
                     +#endif
                     +#if B && (!A || C)
                                                                                           Classification
                       baz();
                        #endif
                      #endif
#ifdef A
 foo():
 bar();
#endif
#if B && (IA | | C)
 baz();
#endif
```

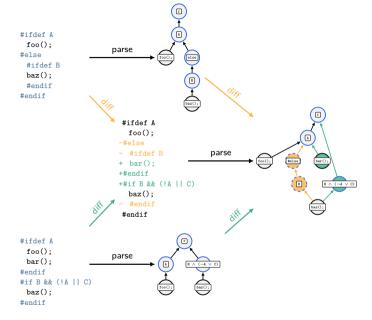




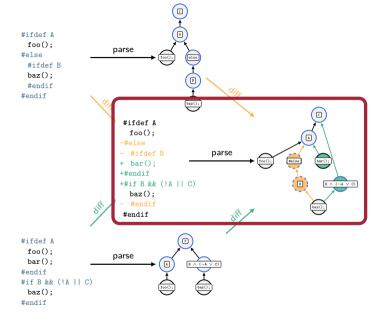




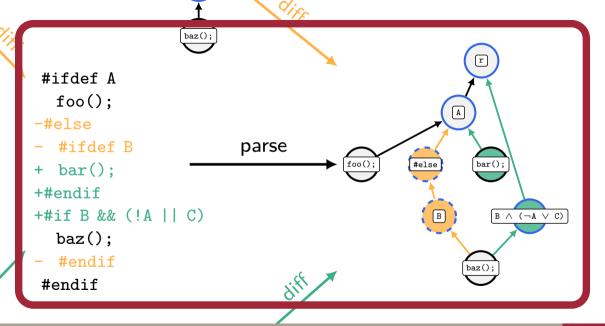
Edits to variability become edits to Variation Trees

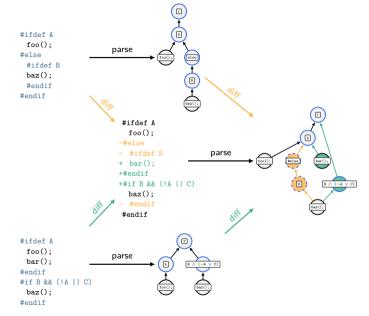


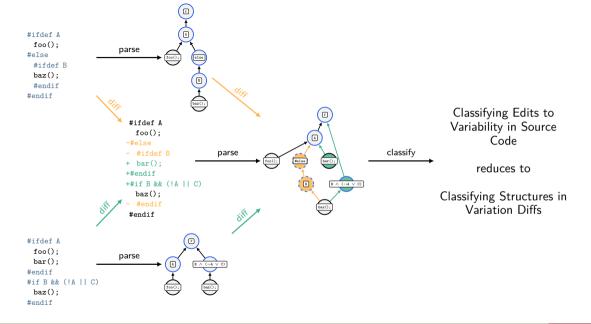
Edits to variability become edits to Variation Trees, for which we introduce Variation Diffs.

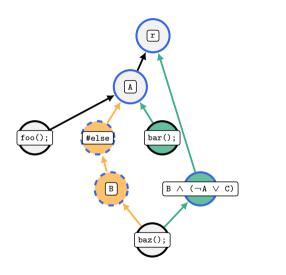


Edits to variability become edits to Variation Trees, for which we introduce Variation Diffs.

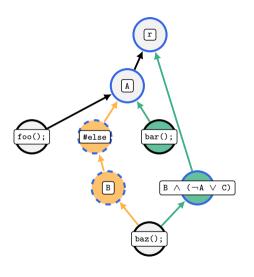


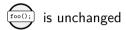




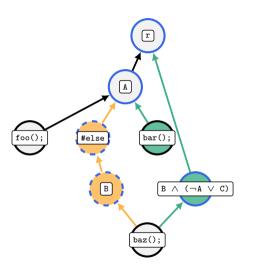


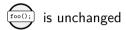




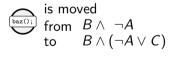


bar(); is added to feature A









#### **Classification** := **Set of Classes**

#### **Classification** := **Set of Classes**

$$\textit{class}: \bigcirc \hspace{-0.1cm} \bigcirc \hspace{-0.1cm} \rightarrow \{\textit{true}, \textit{false}\}$$

#### **Classification** := **Set of Classes**

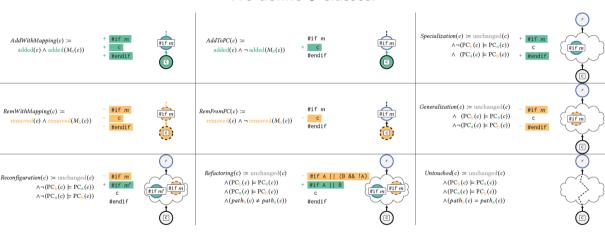
$$class: \bigcirc \longrightarrow \{true, false\}$$

#### Example:

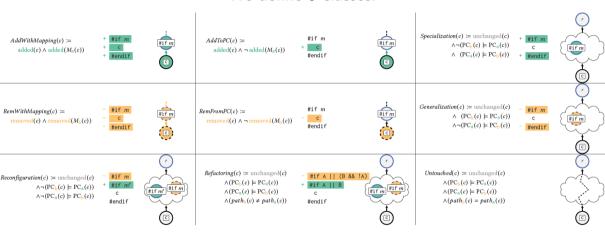
$$AddToPC(\bigcirc) := added(\bigcirc) \land \neg added(p_a(\bigcirc))$$



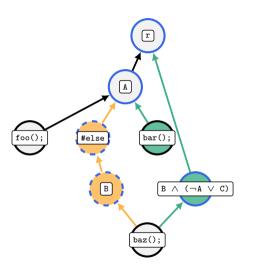
#### We define 9 classes.

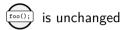


#### We define 9 classes.

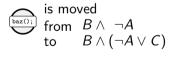


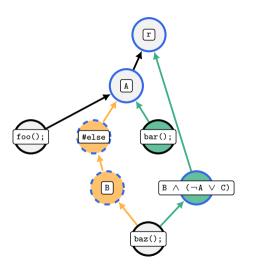
### Custom classifications possible.

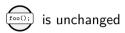






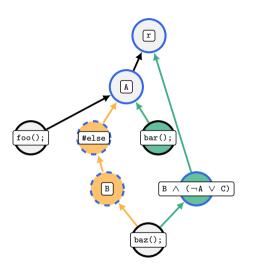






$$AddToPC(\underbrace{bar();}_{bar();}) = true$$

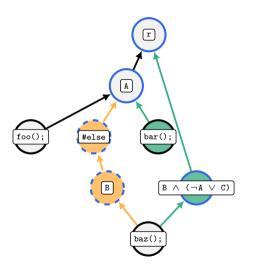
is moved from 
$$B \land \neg A$$
 to  $B \land (\neg A \lor C)$ 



$$Untouched(\widehat{\mathbb{G}_{0}}) = true$$

$$AddToPC(\underbrace{\mathtt{bar();}}) = true$$

is moved from 
$$B \land \neg A$$
 to  $B \land (\neg A \lor C)$ 

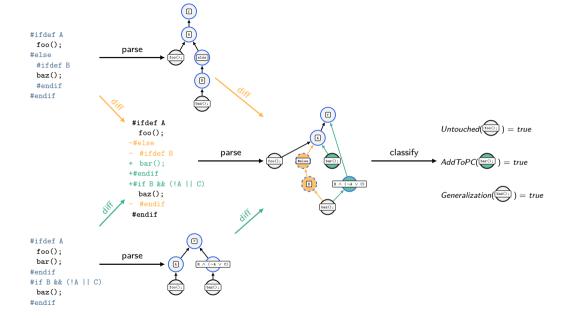


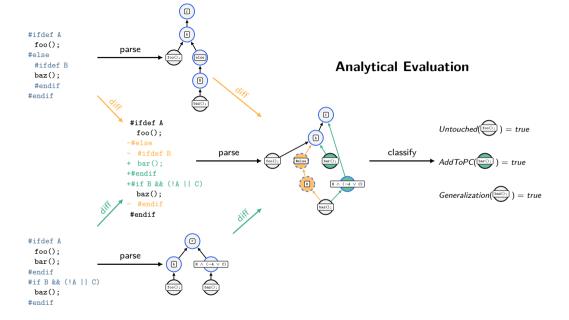
$$Untouched(\underbrace{foo()}_{:}) = true$$

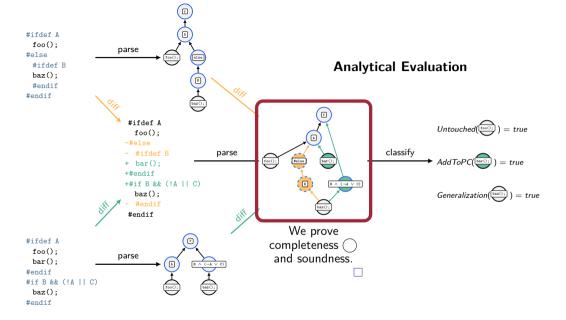
$$AddToPC(\underbrace{bar();}_{bar();}) = true$$

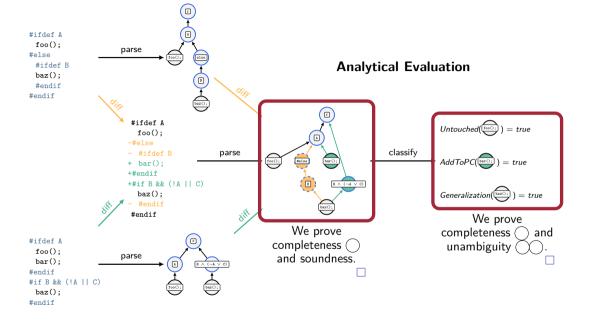
$$Generalization(\widehat{\mathbb{b}^{az()}}) = true$$

because 
$$B \land \neg A \models B \land (\neg A \lor C)$$





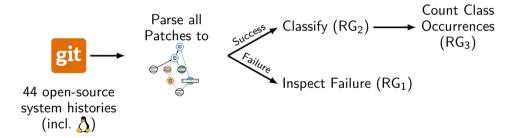








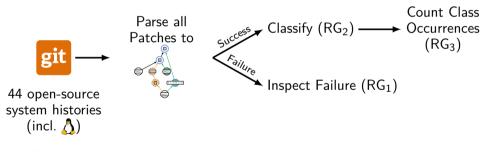






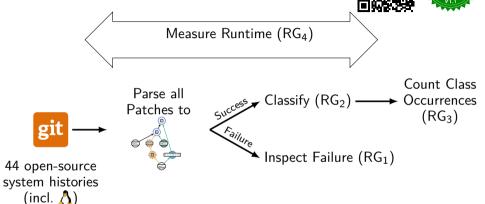






1.7 million commits 45 million edits





1.7 million commits 45 million edits

**RG**<sub>1</sub> Variation Diffs Validate completeness of variation diffs.

#### RG<sub>1</sub> Variation Diffs Result

Validate completeness of variation diffs. All patches with syntactically correct variability annotations can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs.

All patches with syntactically correct variability annotations

can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification

Validate completeness and unambiguity of classification.

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs.

All patches with syntactically correct variability annotations can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification Result Validate completeness and unambiguity of classification.

All edits were assigned exactly one class.  $\checkmark \Rightarrow \bigcirc \land \bigcirc$ 

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs.

All patches with syntactically correct variability annotations

can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification Result Validate completeness and unambiguity of classification.

All edits were assigned exactly one class.  $\checkmark$   $\Rightarrow$   $\bigcirc$   $\land$ 

RG<sub>3</sub> Relevancy

Validate that our edit classes are relevant (i.e., all classes occur in practice).

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs.
All patches with syntactically correct variability annotations

can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification Result

Validate completeness and unambiguity of classification. All edits were assigned exactly one class.  $\checkmark \Rightarrow \bigcirc \land \bigcirc \bigcirc$ 

RG<sub>3</sub> Relevancy

Validate that our edit classes are relevant (i.e., all classes occur in practice).

Result

All classes occur in practice (91,000 to 22 million

occurrences). √

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs.

All patches with syntactically correct variability annotations

can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification Result

Validate completeness and unambiguity of classification.

All edits were assigned exactly one class.  $\checkmark$   $\Rightarrow$   $\bigcirc$   $\land$ 

**RG**<sub>3</sub> Relevancy

Validate that our edit classes are relevant (i.e., all classes

occur in practice).

Result

All classes occur in practice (91,000 to 22 million

occurrences). √

RG<sub>4</sub> Scalability

Validate that edit classification can be automated and scales.

RG<sub>1</sub> Variation Diffs
Result

Validate completeness of variation diffs. All patches with syntactically correct variability annotations can be parsed (99.82%).  $\checkmark \Rightarrow \bigcirc$ 

RG<sub>2</sub> Classification Result Validate completeness and unambiguity of classification. All edits were assigned exactly one class.  $\checkmark \Rightarrow \bigcirc \land \bigcirc \bigcirc$ 

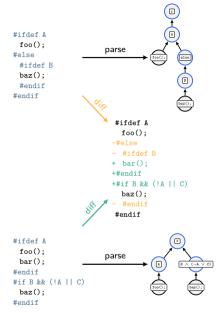
**RG**<sub>3</sub> Relevancy

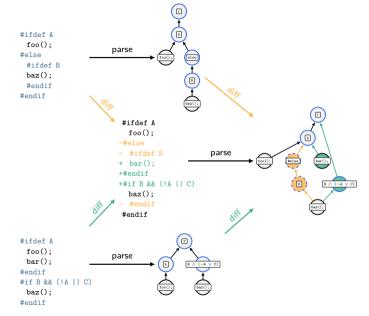
Validate that our edit classes are relevant (i.e., all classes occur in practice). All classes occur in practice (91,000 to 22 million occurrences). ✓

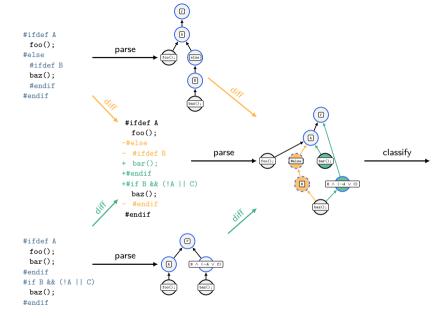
Result

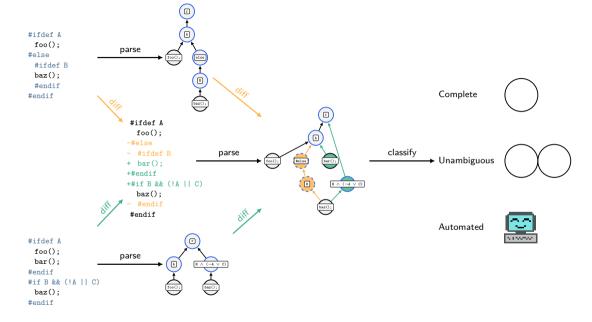
RG<sub>4</sub> Scalability Result Validate that edit classification can be automated and scales. 99.89% of commits processed in < 1s with 7ms/commit as median.  $\checkmark \Rightarrow \blacksquare$ 

```
#ifdef A
 foo();
#else
 #ifdef B
 baz();
 #endif
#endif
                        #ifdef A
                         foo();
                       -#else
                       - #ifdef B
                       + bar();
                       +#endif
                       +#if B && (!A || C)
                         baz();
                       - #endif
                        #endif
#ifdef A
 foo();
 bar();
#endif
#if B && (!A || C)
 baz():
#endif
```



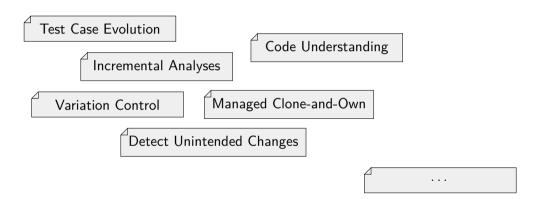






Backup Slides

#### **Use Cases**



#### Software projects are heterogeneous



All identifiable *artifacts* within a software project might be subject to variability.

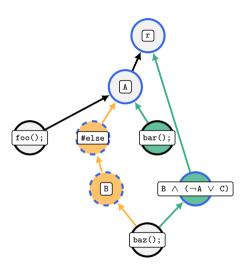
#### Variability comes in many forms

```
#ifdef A
    foo();
#else
#ifdef B
baz();
#endif
#endif

Variability Annotations
(in different dialects)
Frameworks

Language Extensions
```

Commonality: Each artifact is identified with its configuration options.



We prove that variation diffs are

**complete** (i.e., any edit to a variation tree can be described as a variation diff) and

**sound** (i.e., any variation diff describes an edit to a variation tree)

giving us a complete and sound model for edits to variability.



Al-Hajjaji, M., Benduhn, F., Thüm, T., Leich, T., and Saake, G. (2016).

Mutation Operators for Preprocessor-Based Variability.

In Proc. Int'l Workshop on Variability Modelling of Software-Intensive Systems (VaMoS), pages 81–88. ACM.



Borba, P., Teixeira, L., and Gheyi, R. (2012).

A Theory of Software Product Line Refinement.





Ji, W., Berger, T., Antkiewicz, M., and Czarnecki, K. (2015). Maintaining Feature Traceability with Embedded Annotations.

In Proc. Int'l Systems and Software Product Line Conf. (SPLC), pages 61-70. ACM.



Passos, L., Teixeira, L., Dintzner, N., Apel, S., Wąsowski, A., Czarnecki, K., Borba, P., and Guo, J. (2016).

Coevolution of Variability Models and Related Software Artifacts.

Empirical Software Engineering (EMSE), 21(4).



Stănciulescu, S., Berger, T., Walkingshaw, E., and Wąsowski, A. (2016).

Concepts, Operations, and Feasibility of a Projection-Based Variation Control System.

In Proc. Int'l Conf. on Software Maintenance and Evolution (ICSME), pages 323-333. IEEE.