Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila

Mining metrics for understanding metamodel characteristics

Juri Di Rocco

{juri.dirocco|davide.diruscio|ludovico.iovino|alfonso.pierantonio}@univaq.it

Outline

- Introduction
- Motivation
- Measuring metamodels
- Calculation of metrics correlation
- Selection of metrics correlation
- Data Analisys
- Conclusions
- Future works

Introduction

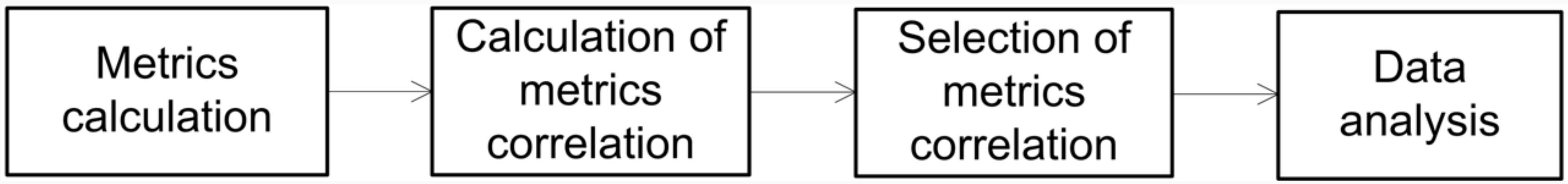
- **Metamodels** are a key concept in Model-Driven Engineering
- Metamodels formally define the modeling primitives used in modeling activities
- They represent the <<trait-d'union>> among all constituent components

Motivation

- Understanding common characteristics of metamodels
- How they evolve over time
- What is the impact of metamodel changes throughout the modeling ecosystem

Measuring Metamodels

- The applied process is able to identify linked structural characteristics
- Understand how they might change depending on the nature of metamodels



Metrics calculation

- Consists of the application of metrics on a data set of metamodels
- The applied metrics are borrowed in [1] and we added new ones by leading to a set of 28 metrics

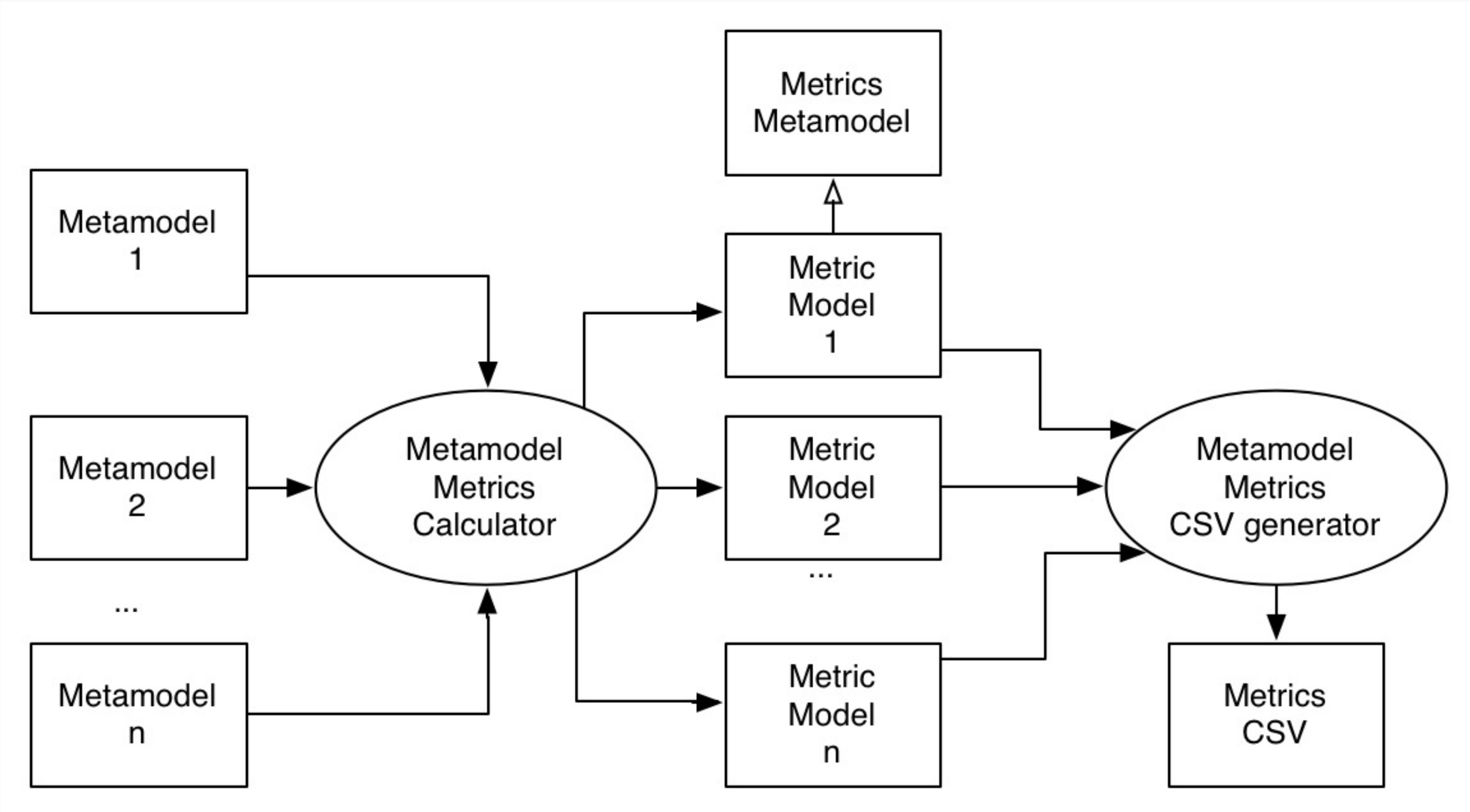
Acronym	Name	Description
AMC	Number of abstract MetaClass	Number of metaclasses that cannot be instantiated in models
ASF	Average Structural Features	Average number of attributes and references in a metaclass
CMC	Number of concrete MetaClass	Number of metaclasses that can be directly instantiated
IFLMC	Number of concrete Immediately Featureless MetaClass	The number of concrete metaclasses that have no attributes or references, but may inherit features from a superclass
LNS	Isolated metaclasses	It is the percentage of metaclasses that are not connected with any other one
MC	Number of total MetaClass	Number of metaclasses in the metamodel (MC = AMC + CMC)
MCWS	Number of class with a super type	Number of metaclasses having at least one super type
MGHL	Maximum generalization hierarchical level	Maximum hierarchical depth in the metamodel
MHS	Max Hierarchy Sibling	Maximun number of classes inheriting from a generic superclass
SF	Number of structural features	Number of attributes and references in the metamodel

[1] W. James, Z. Athansios, M. Nicholas, R. Louis, K. Dimitios, P. Richard, and P. Fiona. What do metamodels really look like? Frontiers of Computer Science, 2013.

Metrics calculation

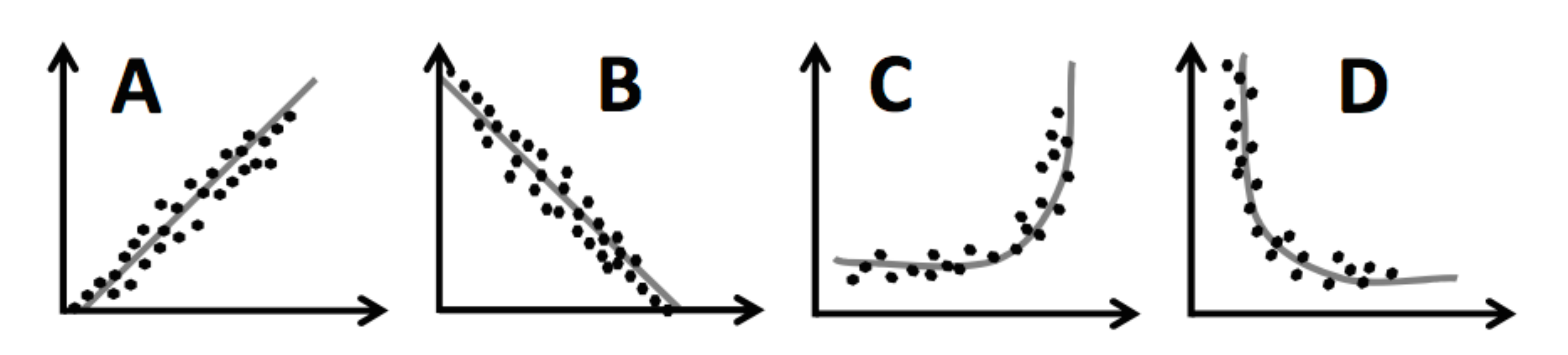
- For a total number of **466 metamodels** belonging to different technical spaces and domains
- The corpus of the analyzed metamodels has been obtained by retrieving metamodels from different repositories, i.e., EMFText Zoo, ATLZoo, Github, GoogleCode

Metrics Calculation



Calculation of metrics correlations

- Correlation is probably the most widely used statistical method to detect cross-links and assess relationships among observed data
- We have considered the Pearson's and Spearman's coefficients to measure the correlations among calculated metamamodel metrics

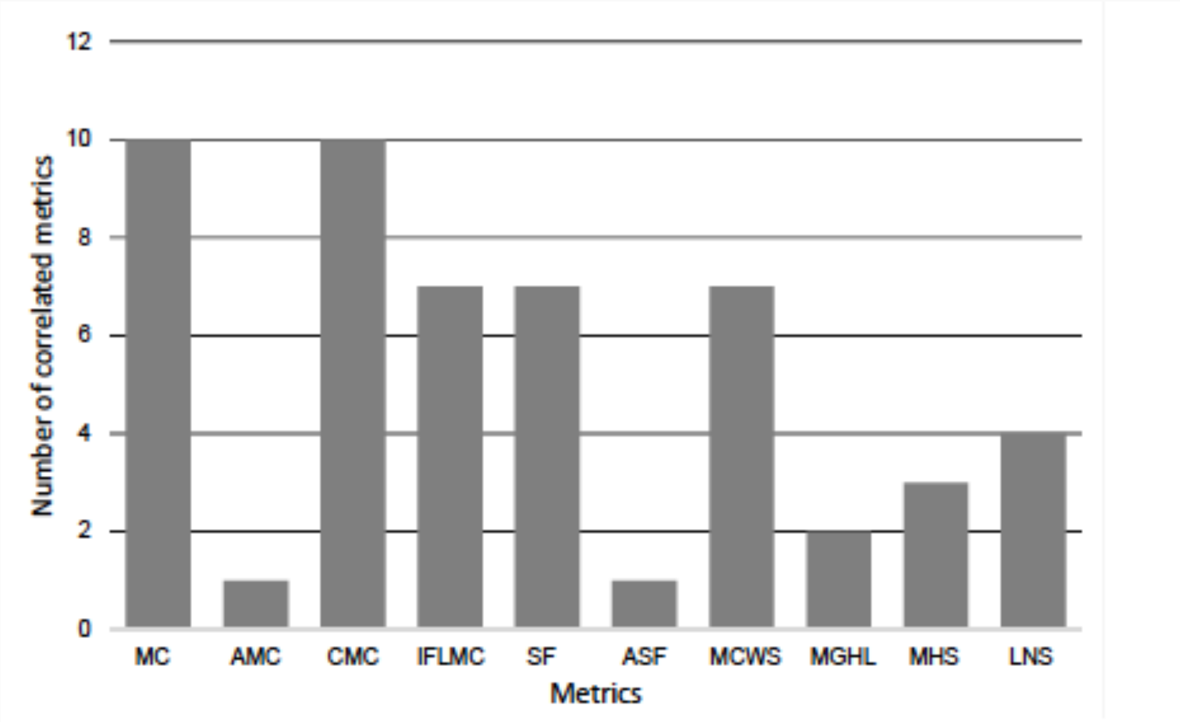


Selection of metrics correlations

- We have calculated the Pearson's and Spearman's correlation indexes for all the values of the considered metrics
- For each couple we have selected the coefficient index (between Pearson and Spearman) having the higher correlation value

Selection of metrics correlations

- The bar chart shows how much each considered metric is correlated with the others
- In particular, can be read as: the higher the bar, the more the metric is strongly correlated with others.



DATA ANALISYS

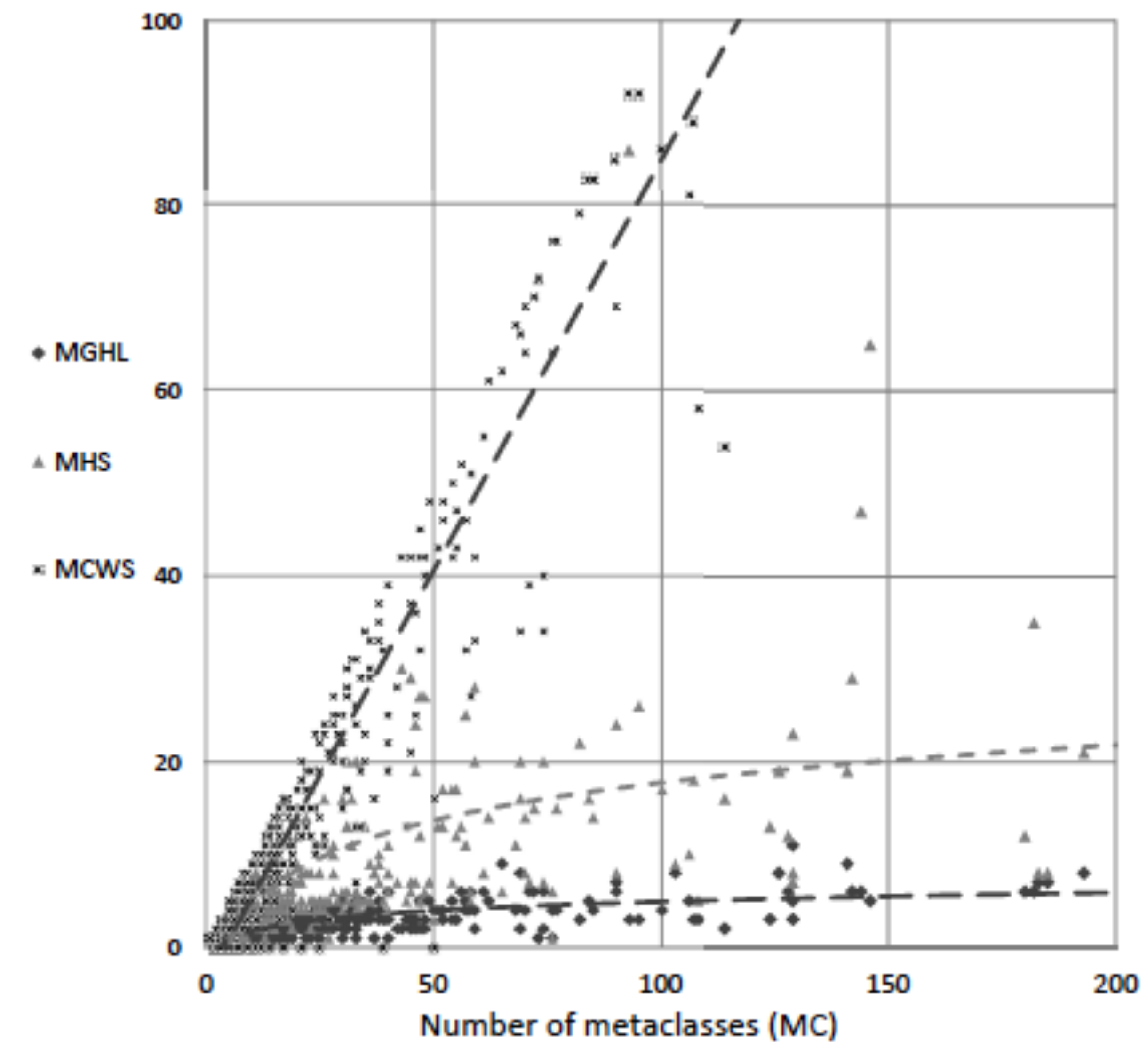
- We discuss some relevant correlations we have identified
- This permits to draw interesting conclusions about how some structural metamodel characteristics are coupled

DATA ANALISYS

- **How the size of metamodels expressed in terms of number of metaclasses is related to the adoption of abstraction constructs**
  - i.e., abstract metaclasses, and supertypes

DATA ANALISYS

How the size of metamodels expressed in terms of number of metaclasses is related to the adoption of abstraction constructs



Pearson correlation:

- MC – MCWS: 0.99
- MC – MHS: 0.70
- MC and MGHL: 0.66

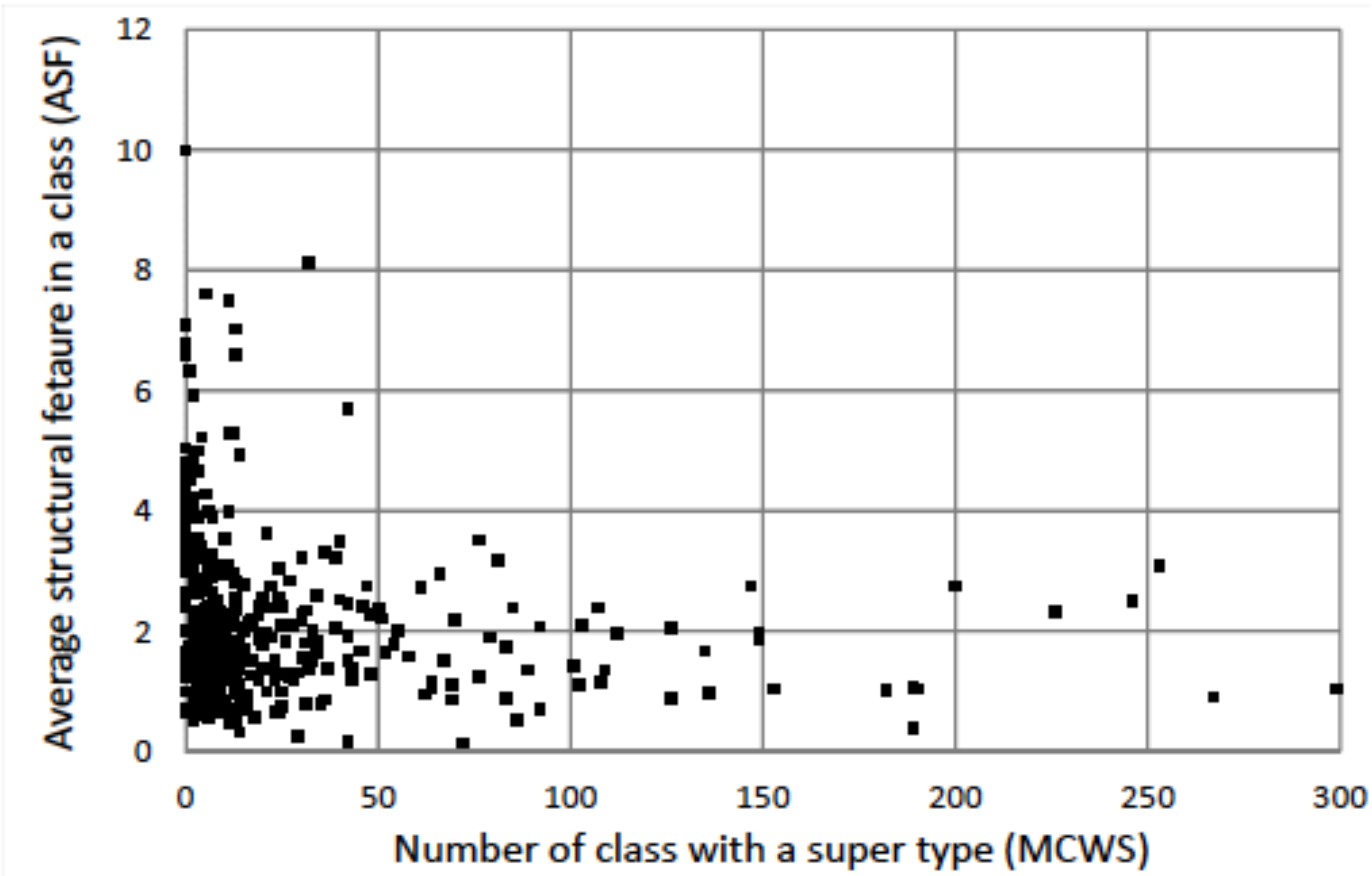
DATA ANALISYS

How structural features are used with hierarchies

- We can consider the average number of features (ASF) and the total number of metaclasses with supertypes (MCWS) metrics
- The Spearman approach permits to identify a greater correlation index

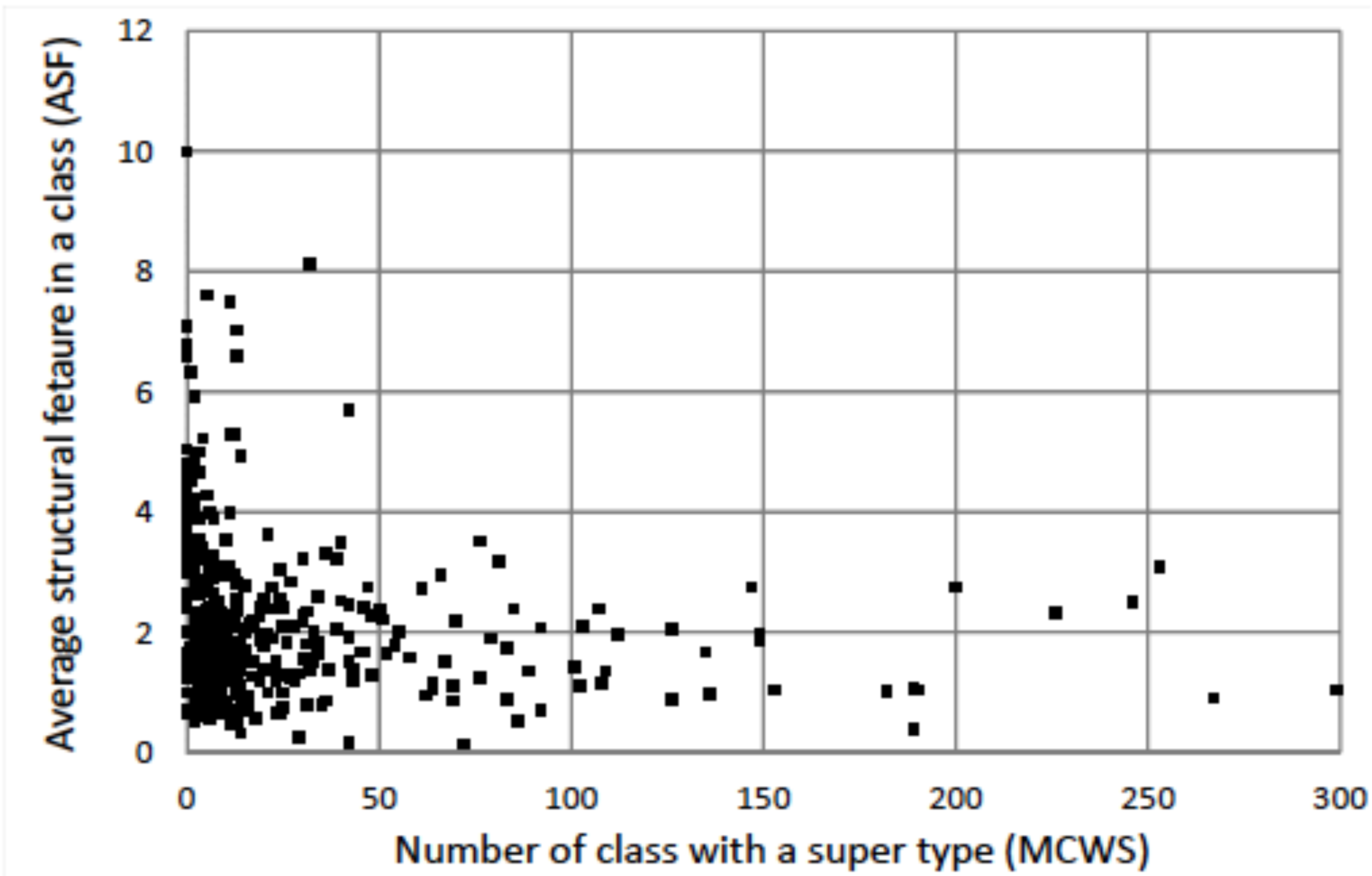
DATA ANALISYS

- Increasing the number of metaclasses with supertypes, the average number of structural features in a metaclass decreases



DATA ANALISYS

- By considering metamodels having in between 1 and 50 metaclasses with supertype, the average number of features (excluding the inherited ones) of a metaclass ranges between 1 and 5



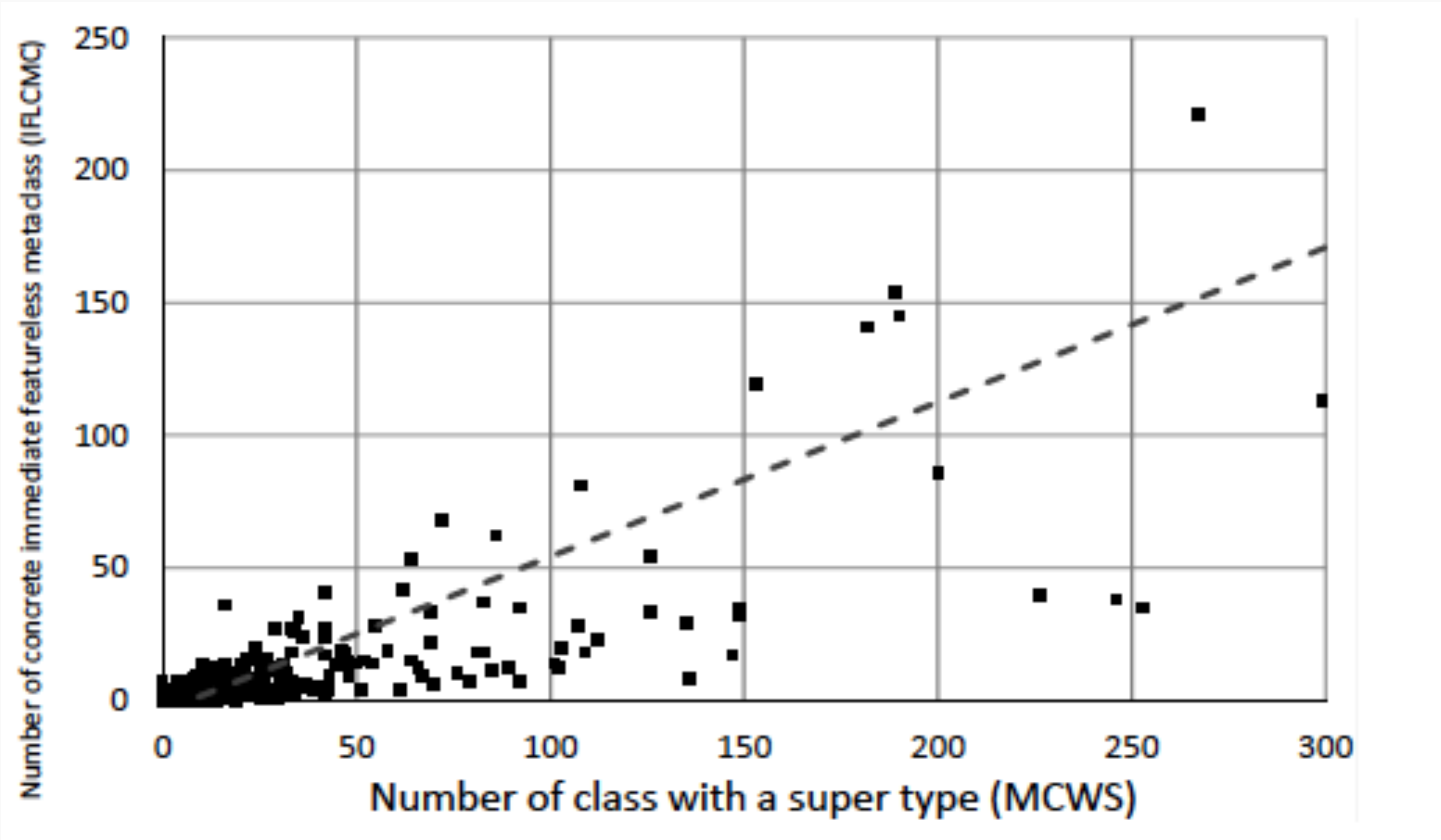
DATA ANALISYS

How the number of featureless metaclasses is related to hierarchies height

- This can indicate how specializations of metaclasses can introduce or reduce structural features in metamodels
- MCWS and IFLMC (immediate featureless metaclass) are strongly correlated as supported by the Pearson's index having value 0.890

DATA ANALISYS

How the number of featureless metaclasses is related to hierarchies height



DATA ANALISYS

- By increasing the number of metaclasses with super types, the number of metaclasses without attributes or references increases too
- This means that when hierarchies are introduced, usually existing features are subject to refactoring operations mainly to move them to super classes and to create leaves in the hierarchies inheriting features from the super types.

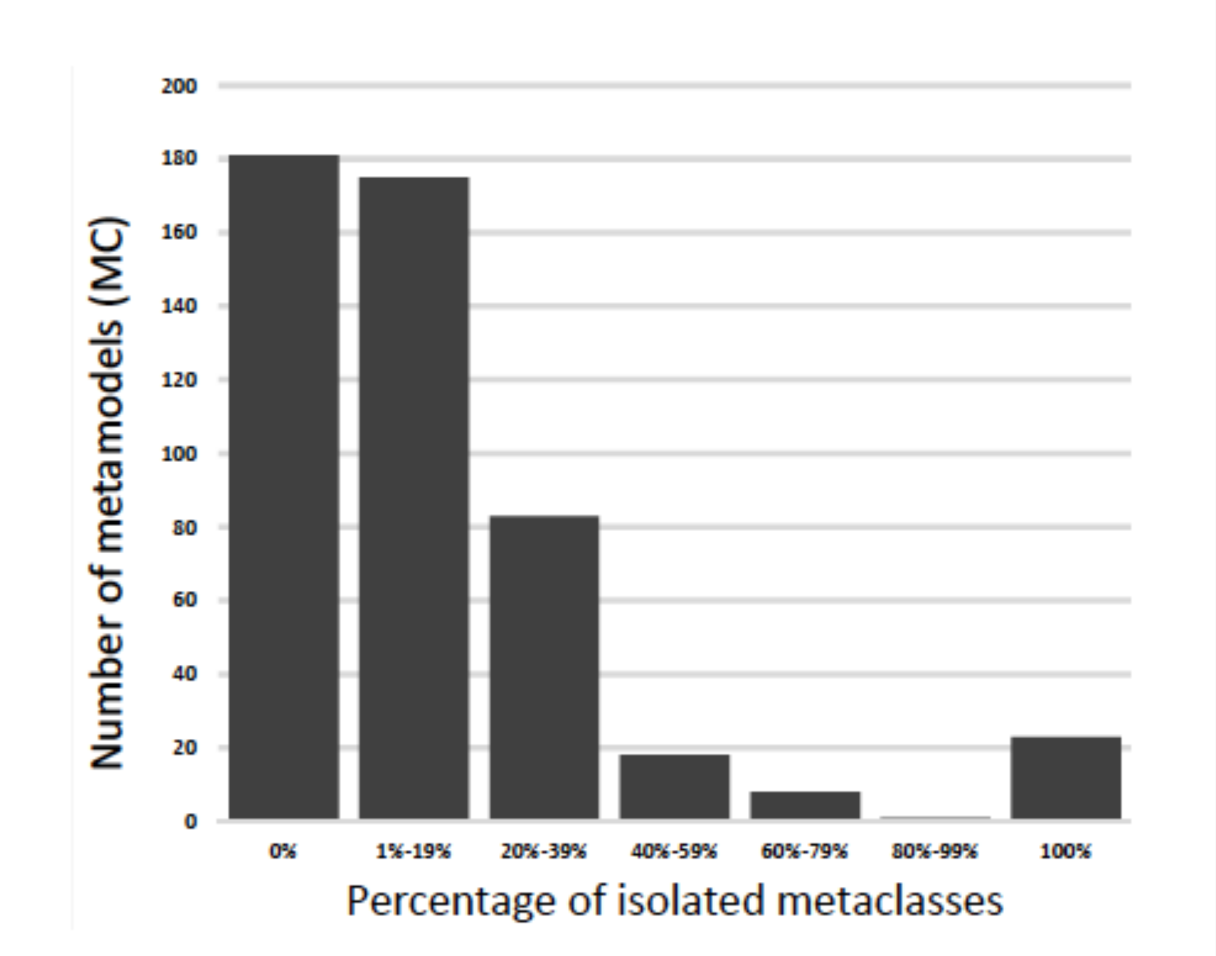
DATA ANALISYS

How isolated metaclasses are distributed

- In this case we considered another statistical instrument named **Pareto analysis**, (20% of the causes account for 80% of the defects)
- About 80% of the analyzed metamodels have a percentage of isolated metaclasses in the range 0-19%, by testifying the fact that isolated metaclasses are not commonly used

DATA ANALISYS

- We can claim that they are used only for testing or educational purposes



Conclusions

- We proposed a number of metrics which can be used to acquire objective, transparent, and reproducible measurements of metamodels
- The major goal is to better understand the main characteristic of metamodels, how they are coupled, and how they change depending on the metamodel structure
- A correlation analysis has been performed to identify the most cross-linked metrics, which have, in turn, been computed over 450 metamodels

Conclusions

These figures have been discussed in detail highlighting the followings

- the adoption of inheritance is proportional to the size of metamodels
- the number of metaclasses with supertypes are inversely proportional to the average number of structural features
- the number of metaclasses with supertypes is proportional to the number of metaclasses without attributes or references
- isolated metaclasses are not commonly used apart from testing or educational purposes.

Conclusions

These figures have been discussed in detail highlighting the following

Correlated Metrics	
↑ usage of inheritance	↑ metamodel size
↑ #metaclasses w/ supertypes	↓ avg. #features
↑ #metaclasses w/ supertypes	↑ #metaclasses w/o features

Future works

- One of the main goal in the next future is to extend the approach to analyse the characteristics of coupled modeling artifacts
- How structural characteristics of metamodels affect those of model transformations or any metamodel-based artifact
- The long term goal of this work is to define an approach able to estimate or even predict the cost of developing or refining modeling artifacts by considering the structural characteristics of the corresponding metamodels.

References (1)

[1] W. James, Z. Athansios, M. Nicholas, R. Louis, K. Dimitios, P. Richard, and P. Fiona. What do metamodels really look like? Frontiers of Computer Science, 2013.

[2] N. E. Fenton and S. L. Pfleeger. *Software Metrics: A Rigorous and Practical Approach*. PWS Publishing Co., Boston, MA, USA, 2nd edition, 1998.

[3] Di Rocco, J., Di Ruscio, D., Iovino, L., Pierantonio, A.: Mining Metrics for Understanding Metamodel Characteristics. In: Proceedings of the 6th International Workshop on Modeling in Software Engineering. MiSE 2014, New York, NY, USA, ACM

References (2)

[4] C. Spearman. The proof and measurement of association between two things. The American journal of psychology, 15(1):72–101, 1904.
[5] E. Vépa, J. Bézivin, H. Brunelière, and F. Jouault. Measuring model repositories. In *Proceedings of the 1st Workshop on Model Size Metrics (MSM'06) co-located with MoDELS'2006*, 2006.
[6] A. Vignaga. Metrics for measuring atl model transformations. Technical report, 2009.

[7] R. Harrison, S. Counsell, and R. Nithi. An evaluation of the mood set of object-oriented software metrics. *IEEE Transactions on Software Engineering*, 24:491–496, 1998.