Chapter 1

Discussion

How effective is a group sampling algorithm compared to different sampling strategies? To learn the influence of configuration options on non-functional property, the tests on real world applications show, that fitting a linear regression on randomly sampled data performs quite well . As seen in **??** for most systems, the model describes the system very well with a relatively small sample size. Group sampling, as implemented here, does not provide a model with the same accuracy and precision as random sampling with linear regression. Even with increased sample size, the group sampling algorithm fails to produce better results. But the group sampling algorithm learns the key characteristics of the system faster, resulting in a more accurate model for smaller sample sizes. The group sampling algorithm implemented here works well for systems, with few constrains and one very influential feature. Due to the rules applied during the creation of the groupings, all features are represented in as few samples as possible, allowing for a quick identification of influential features. There are more optimization techniques which further reduce imposters during the influence calculations described by **?** but can not be properly applied in this scenario, which makes the here implemented group sampling algorithm more vulnerable to imposters. Also, the algorithm implemented here struggles to produce a good influence model, this can partially be accounted to the stepwise influence calculation.

hier gabs glaub n passendes paper

How can we mitigate the effect of constraints on the configuration options? In this work two approaches are presented to handle constrains on the configuration options.

How scalable is group sampling? TODO

Can group sampling be used to effectively identify feature interactions and include them in the resulting model? Maybe ...