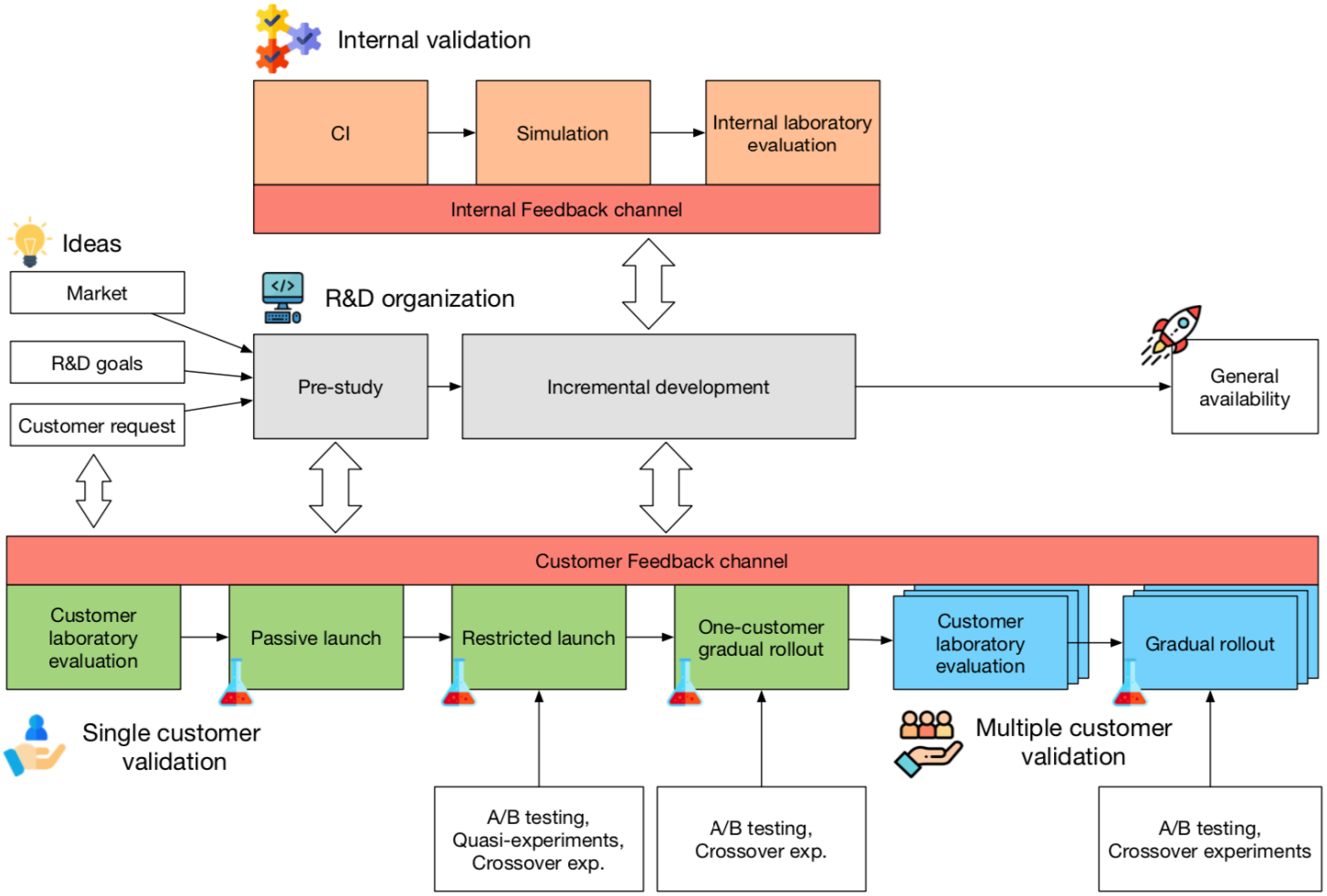
Data-Driven Continuous Evolution of Autonomous Systems of Systems (#19)

David Issa Mattos – davidis@chalmers.se
Jan Bosch – jan.bosch@chalmers.se
Helena Holmström Olsson – helena.holmstrom.olsson@mau.se

Experiments in B2B and mission critical features?



Experiments in B2B and mission critical features?

- Validation of a mission-critical feature in field
  - Faster than in-house validation
- Some customers see the value of faster field experiments that and are willing to cooperate!
  - Even if it breaks (because it will be fixed faster)
- Experimentation does not exclude internal validation, CI, laboratory experiments etc…
  - Those can happen in parallel or only with features that have evidence to work/deliver value

Experiments in B2B and mission critical features?

- “The field deployment allowed us to identify a problem that none of our existing internal testing could identify”.
- “We are not more efficient when it comes to building time, but we know in advance that it works in the real world. That is the benefit of prototyping and experimenting”

Continuous experimentation

- A broader view of CE beyond A/B testing
- Different types of experiments
  - Business experiments
    - A/B testing, quasi experiments
  - Regression experiments
    - Dark launches, canary releases
  - Optimization experiments
    - Bayesian optimization, RL
  - Customer support experiments
    - Quasi-experiments, crossover
- Over 25 different techniques…

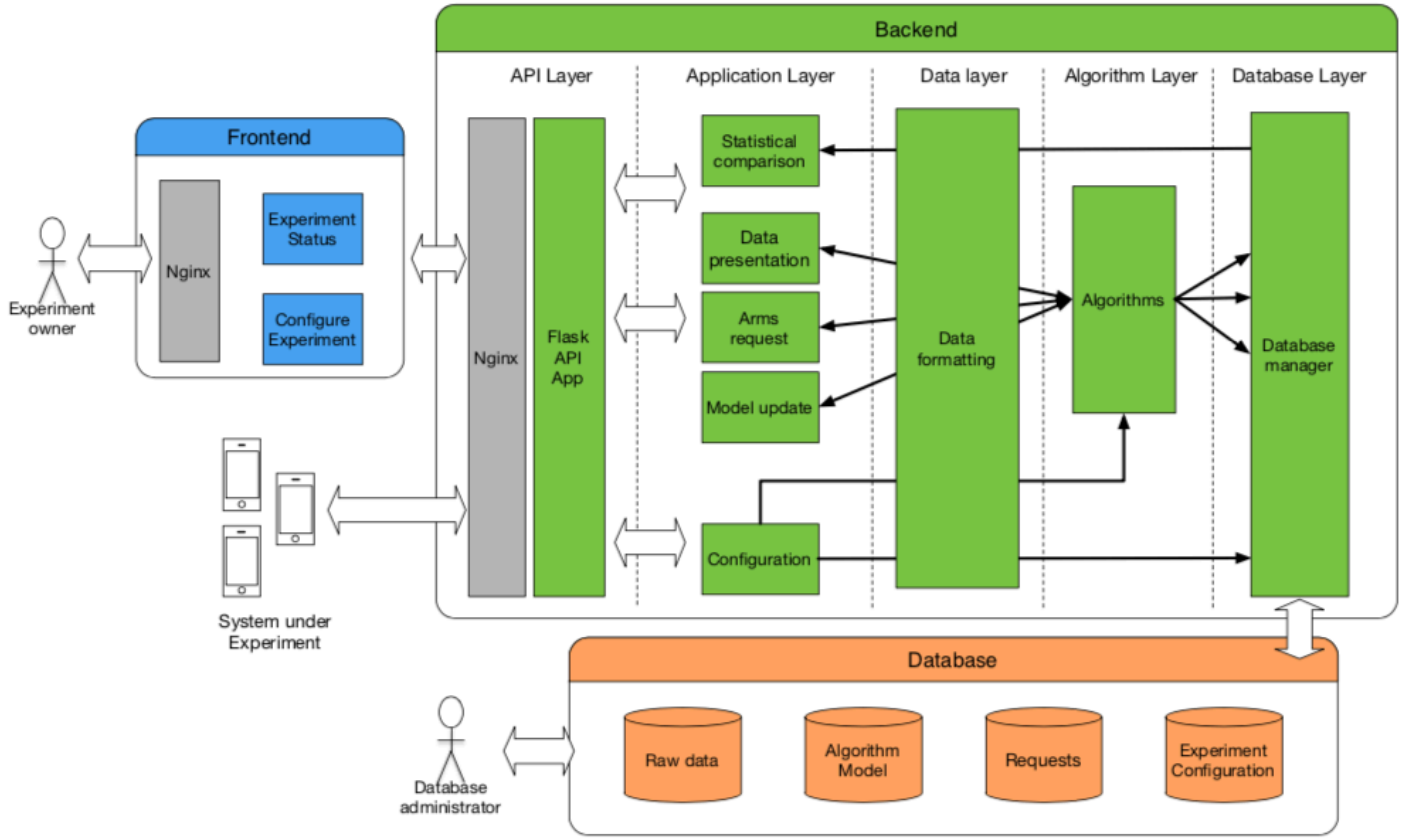
How to select a CE technique?

- Multiple case-study with 4 companies and 39 practitioners (with Ericsson from Software Center) + analysis of 108 papers

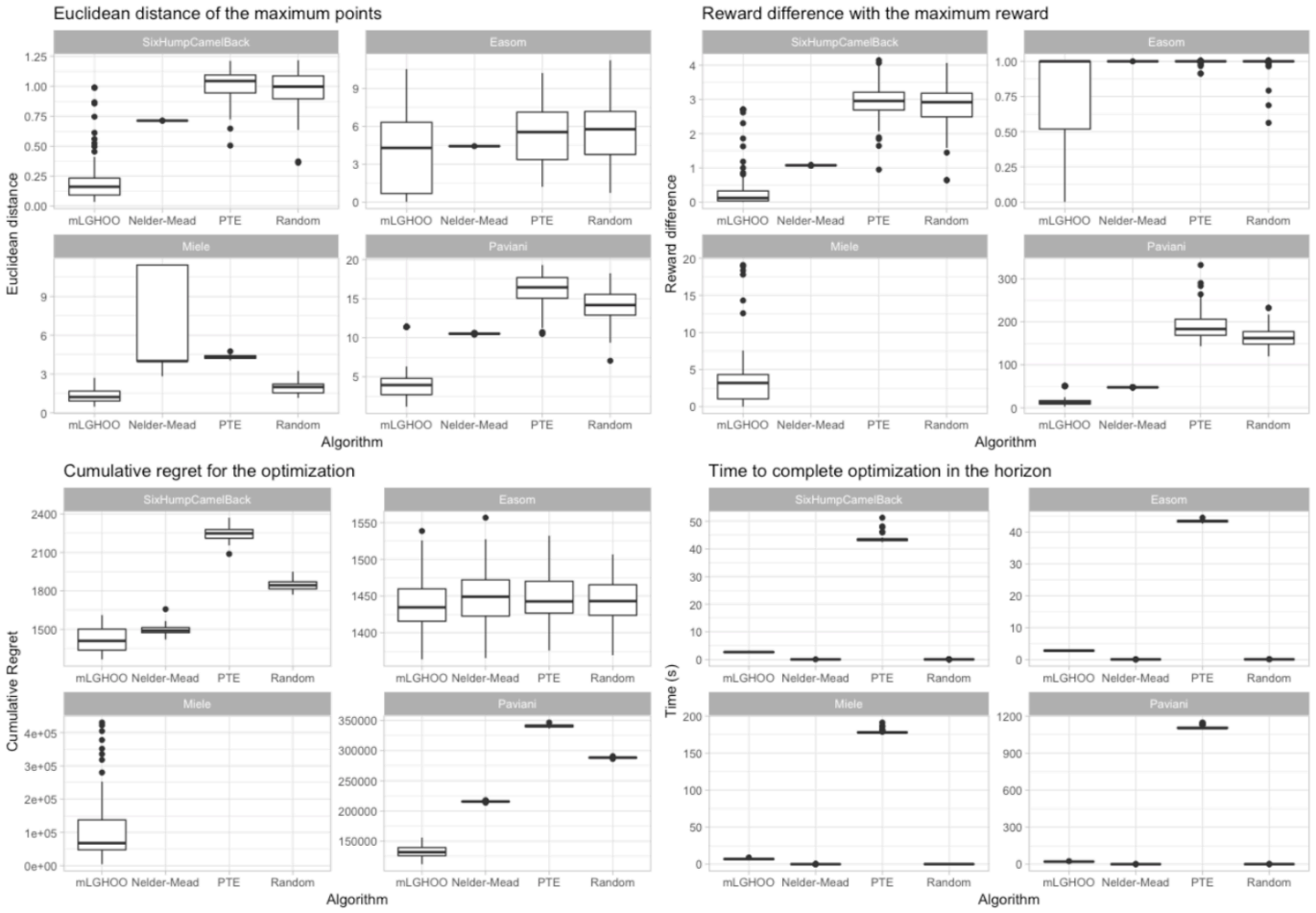
- When do I use each one of those?
- Why/when use one technique compared to other?
- Infrastructure to use

Optimization experiments

- Automatically identify best configuration parameters



How our system compares to other state-of-the-art systems?



Statistical comparison of black box algorithms for optimization of expensive functions

- Sometimes it is not clear which to select in which case!
  - There are hundreds of different algorithms
- We are doing a large scale comparison between algorithms
  - Over 25 algorithms
  - Over 170 benchmark functions
  - Different measurements metrics
  - Different conditions (noise, budget etc…)
- Applications
  - Field optimization
  - Hyperparameter tuning of ML models
  - AutoML

Customer aspects in CE

- Customers have different prioritization regarding experimentation
  - “I want to try the alpha 0.0.0.1a feature and help guide the development”
  - “I want the latest features in a stable release”
  - “I want the oldest supported software because I hate updates”
- What are the different profiles?
- How to plan experimentation activities based on customer profiles?
- Case study with Ericsson

It would be great to understand and see how this works with different companies

CE in Software Product Lines

- Case study with Ericsson
- How to plan the experiment? Where to start? How to analyze the data?
- Which product do you select to start experimenting?
- Transfer conclusions to other products?
- Incorporate knowledge and experience in the experimentation process
- E.g. How do you experiment in a feature available in different products?
- **Investigate/validate in other companies**

Sprint plans

- Understanding/adapting the HURRIER process in other companies
  - If you have interest contact me!
- Understand continuous experimentation from the software product line point of view.
  - Experiences from other companies
- Customer aspects in CE.
  - Experiences from other companies
- Comparison of algorithms for expensive black-box optimization