UNIVERSITEIT GENT

FACULTY OF ENGINEERING AND ARCHITECTURE

Search space optimisation for procedural content generators

Harald De Bondt

Problem statement

- Content creation problem
  - Unscalable
  - Expensive

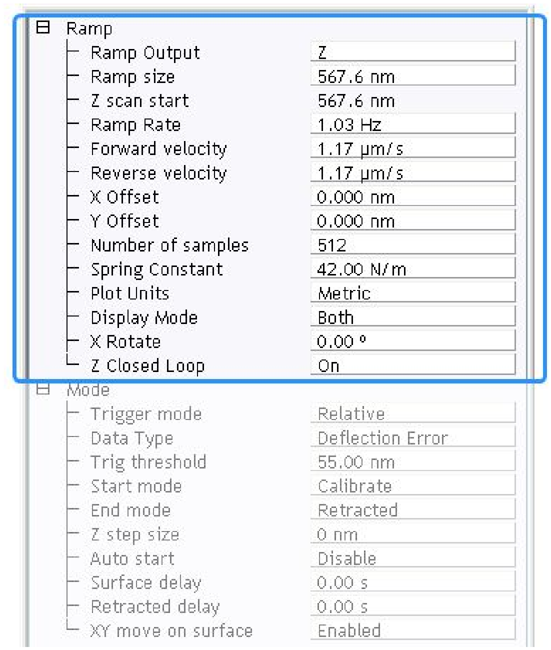
- Procedural Content Generation (PCG)
  - Automation
  - Variety
  - Replayability

- Research problem in PCG
  - Controllability
    - Intuitive configuration
  - Expressiveness
    - Avoid bias due to implementation

Related work
Controllability in PCG

- Expressiveness ↔ Controllability

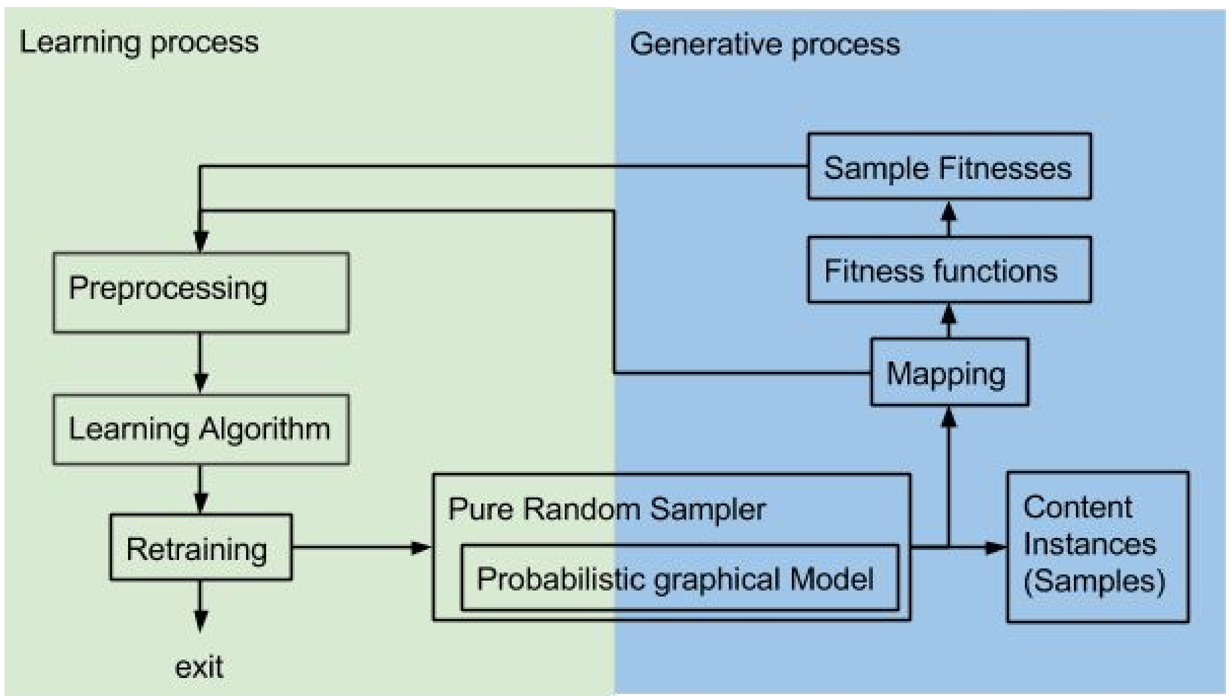
- Control types: Bottom-up vs top-down
  - How vs what is generated
  - Feed-forward vs iterative
  - Local control vs global
- Examples
  - Constructive
  - Declarative
  - Search-based
  - Probabilistic
  - Learning



Methodology

- Maximise expressiveness
  - Uniform sampling from optimised model ↔ optimised search algorithm

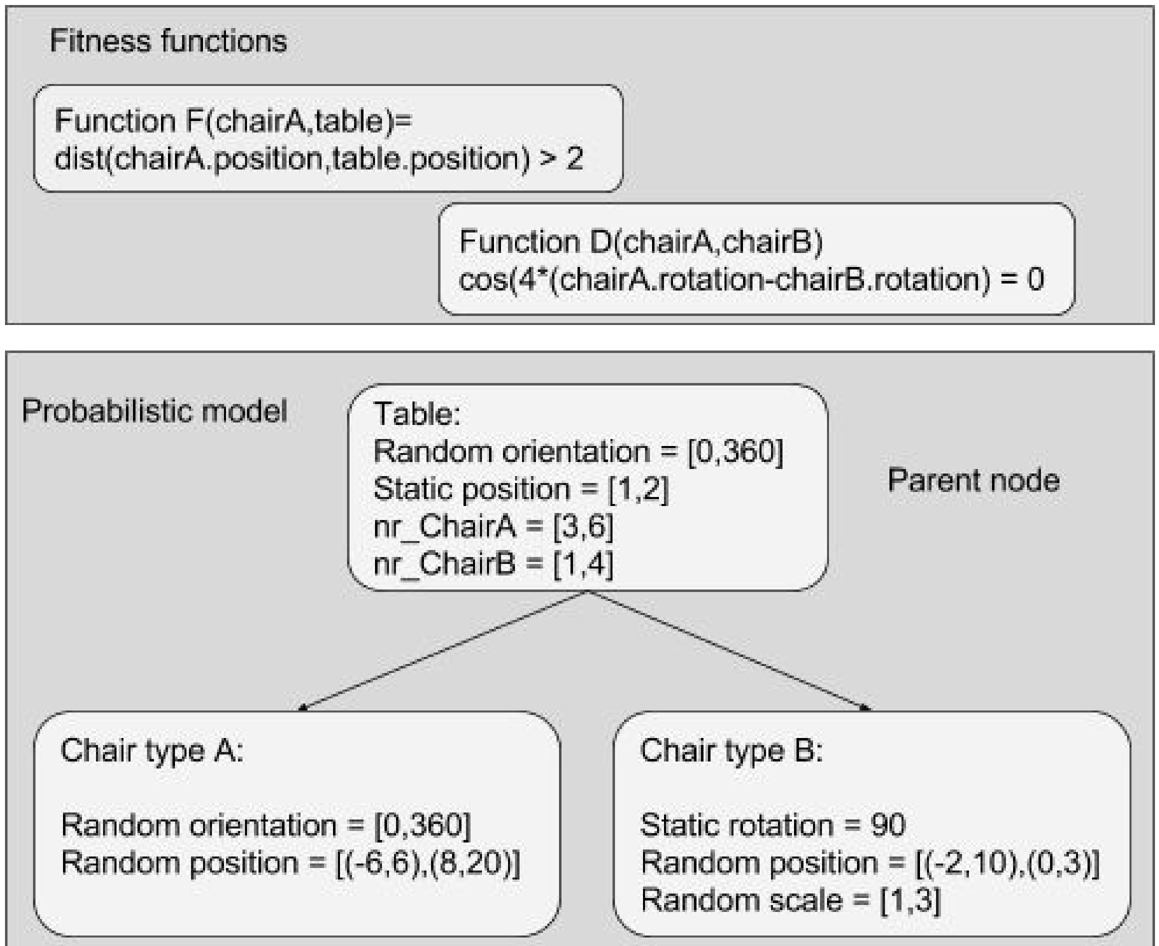
Two processes



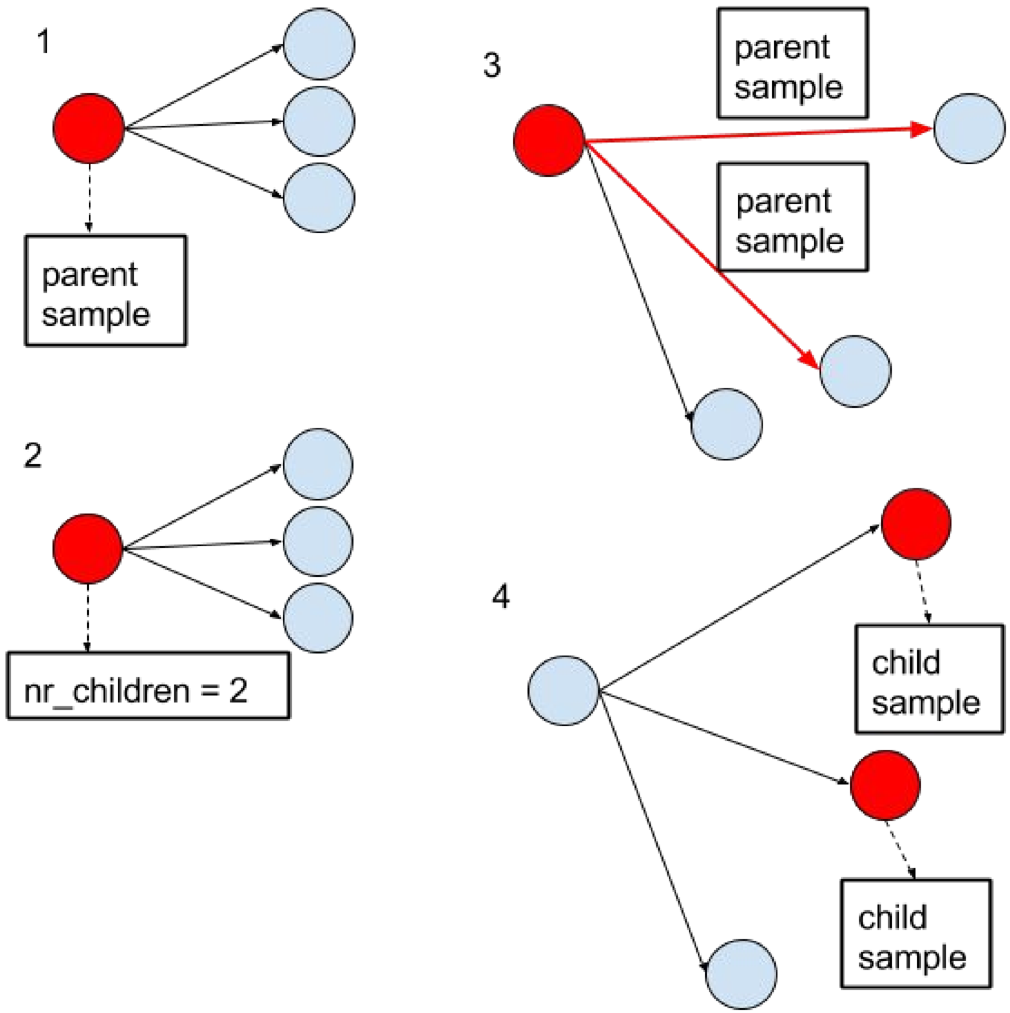
Generative process: Initialization

Top-down

Bottom-up



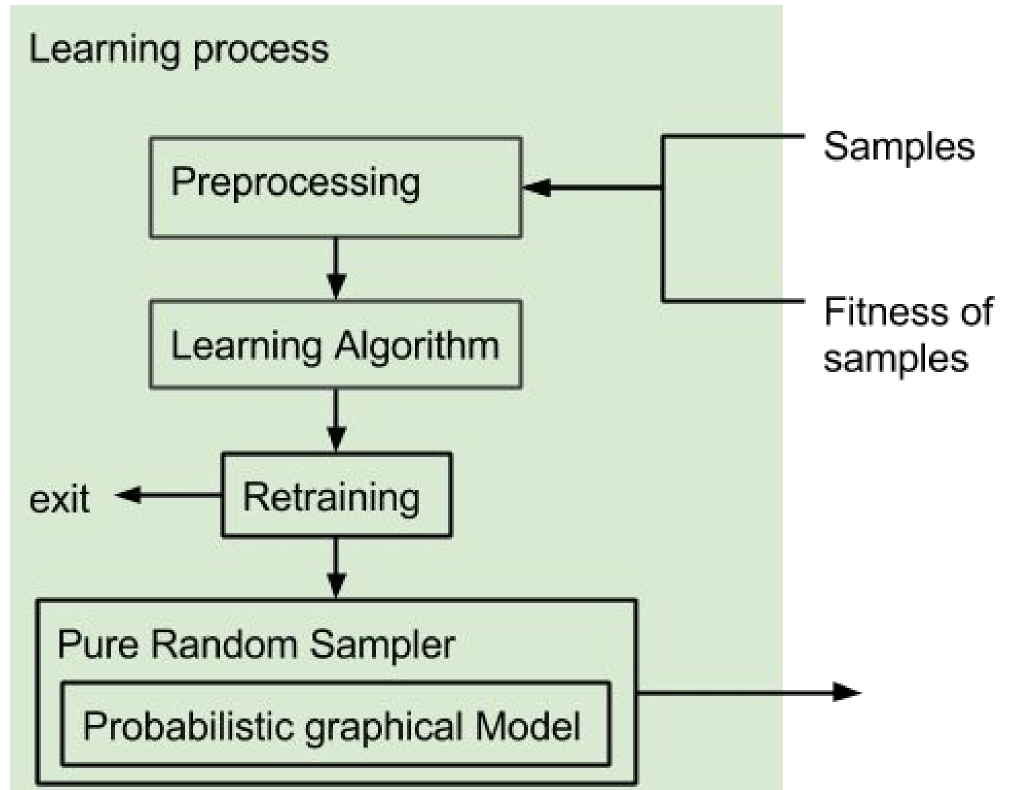
Generative process:
Sampling (uniform)



Learning process

- Optimisation of the generative process → learning problem (synthetic data)

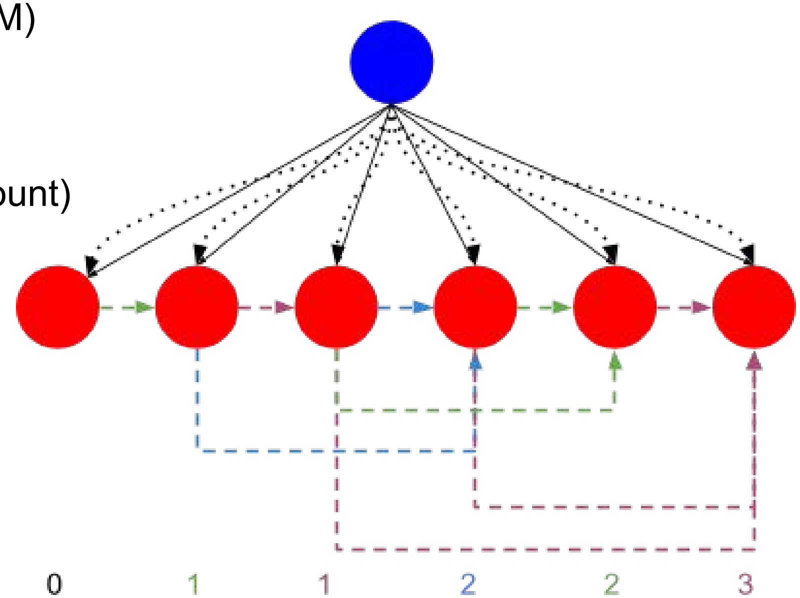
- Components
    - Preprocessing
    - Retraining
    - Probabilistic model
    - Learning Algorithm



Learning process: components
Probabilistic model

- Probabilistic model compatible with generation process
  - Hierarchical Gaussian Mixture Models (GMM)
    - Conditional relation
      - Parent and child
      - Child and it’s siblings
  - Multiple-instance learning (varying object count)

- Optimisations
  - Variable sibling order
  - Marginalisation

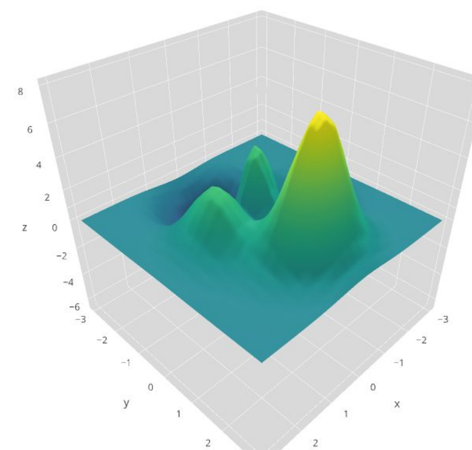


Learning process: components

Learning algorithm

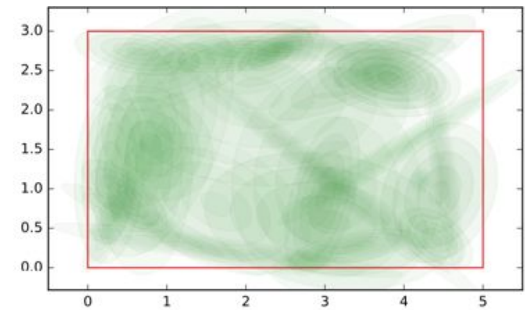
- Supervised learning
  - Fit GMM based on sample and fitness data
    - EM algorithm (unsupervised)

- Incorporate fitness in EM
  - Weighted density estimation (WDE)
    - Likelihood ~ fitness
  - Regression → P(Variables|Fitness)



Evaluation and discussion

- Measuring expressiveness is open problem [1]
- Elaborate evaluation of simplified case (2 dimensional)
  - Expressiveness relative to fitness
  - Sampling Performance using heatmaps
- Simplified evaluation of multiple cases (>3 dimensional)
  - Visualized expressiveness
  - Sampling Performance using Average Fitness Gain
  - Runtime complexity/computational time



[1] G. Smith and J. Whitehead, “Analyzing the expressive range of a level generator,” Proceedings of the 2010 Workshop on Procedural Content Generation in Games PCGames 10, 2010.

Simplified case:

Expressiveness results

		Predicted condition	
	Total population	Predicted condition positive	Predicted condition negative
True condition	Condition positive	True positive: 89 %	False negative: 11 %
	Condition negative	False positive: 15 %	True negative: 85 %

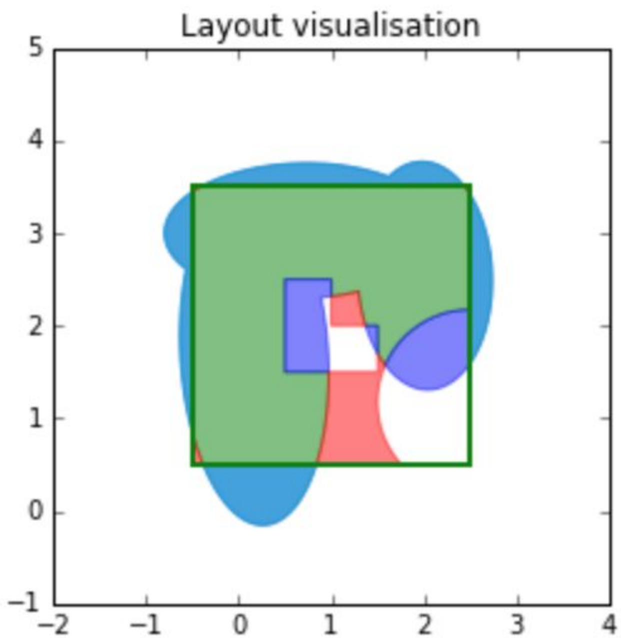


Fig. 5. GMM surface area coverage

Simplified case:

Sampling performance results

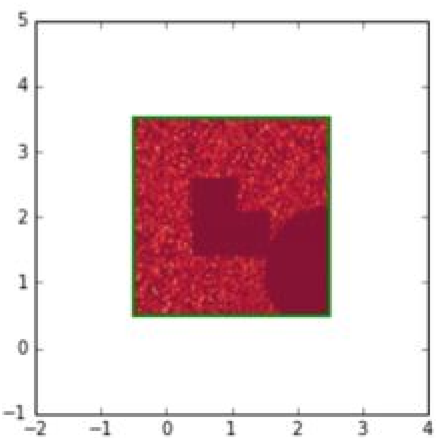


Fig. 3. Heat-map for naive sampling

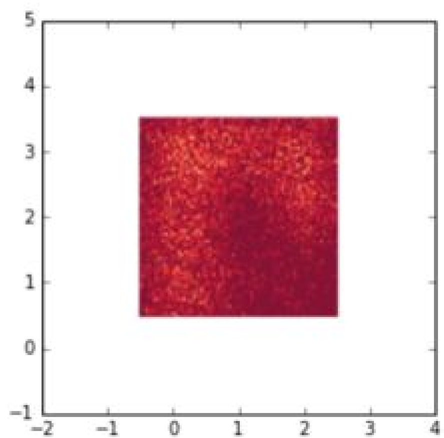


Fig. 4. Heat-map for sampling from trained model

Average acceptance rate:
Naive: 70%
Optimised: 86%

Multiple cases

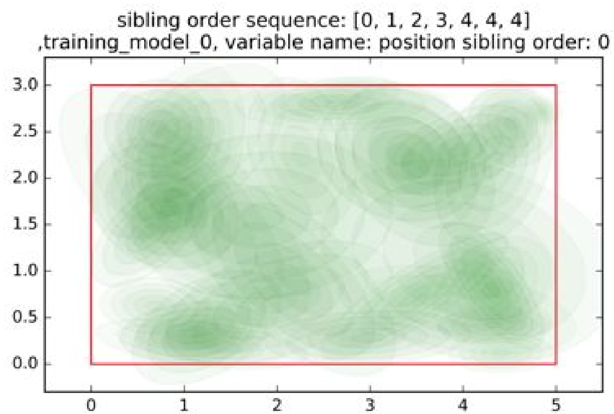
- Evaluation of 15 parameters in the methodology
- Generality: 5 separate fitness functions
    - Functions derived from constraints in architectural context [1]

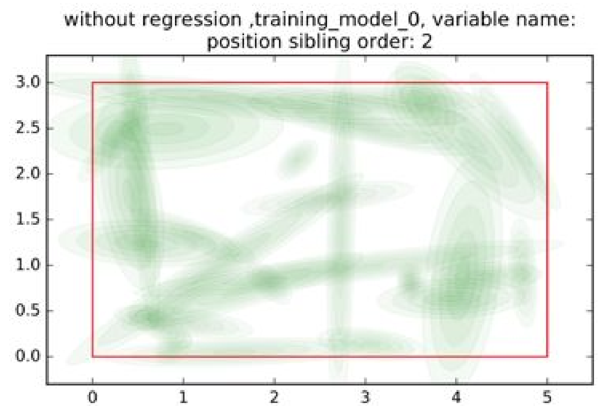
	Hierarchical relation	Analytic family
Minimum polygon overlap	pairwise	geometric
Target surface ratio	absolute	geometric
target distance	pairwise	distance based
Minimum centroid difference	absolute	geometric/distance based
Closest side alignment	pairwise	goniometric/geometric

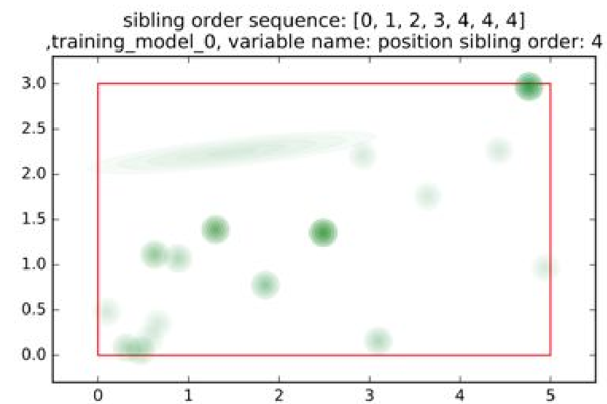
[1] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun, “Interactive furniture layout using interior design guidelines,” ACM Trans. Graph., vol. 30, no. 4, p. 1, 2011.

Multiple cases: Visualized expressiveness

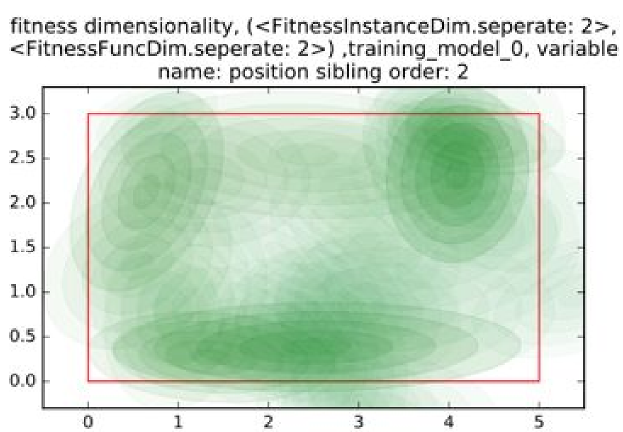
With weighted density estimation

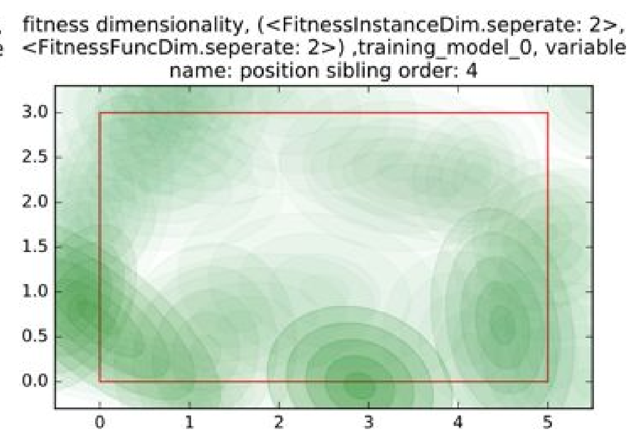






With regression





Multiple cases: Results Summary

- Most significant parameters
  - Number of training samples ↑
    - Increase expressiveness and sampling performance ↑
  - Marginalisation ↑
    - Reduce computation time ↓
    - Expressiveness
- Optimised fitness functions (⅗)
  - Minimum polygon overlay: 100%
  - Target distance: 22%
  - Minimum centroid difference: 45%

Conclusion and Future work

- Novel expressiveness metric
- New approach as alternative for current PCG solutions
  - Combines top-down and bottom-up control
  - Automatic search space optimise
    - Minimal decrease in expressiveness
- Optimisation not significant enough
- Possible paths for future research:
  - Evaluate more complex fitness functions
  - Train on data from optimised search algorithm
    - E.g Markov chain Monte Carlo
  - A more elaborate probabilistic model
  - Learn hierarchical structure from the data