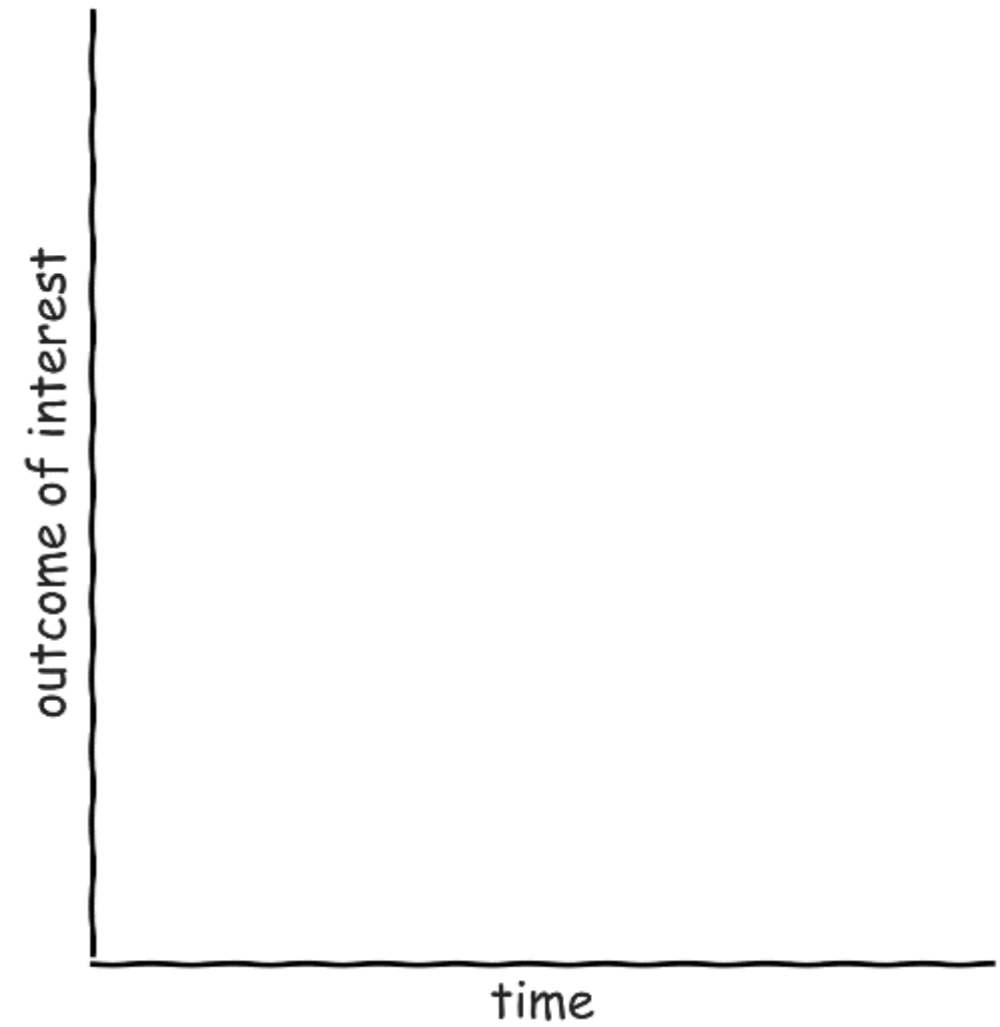
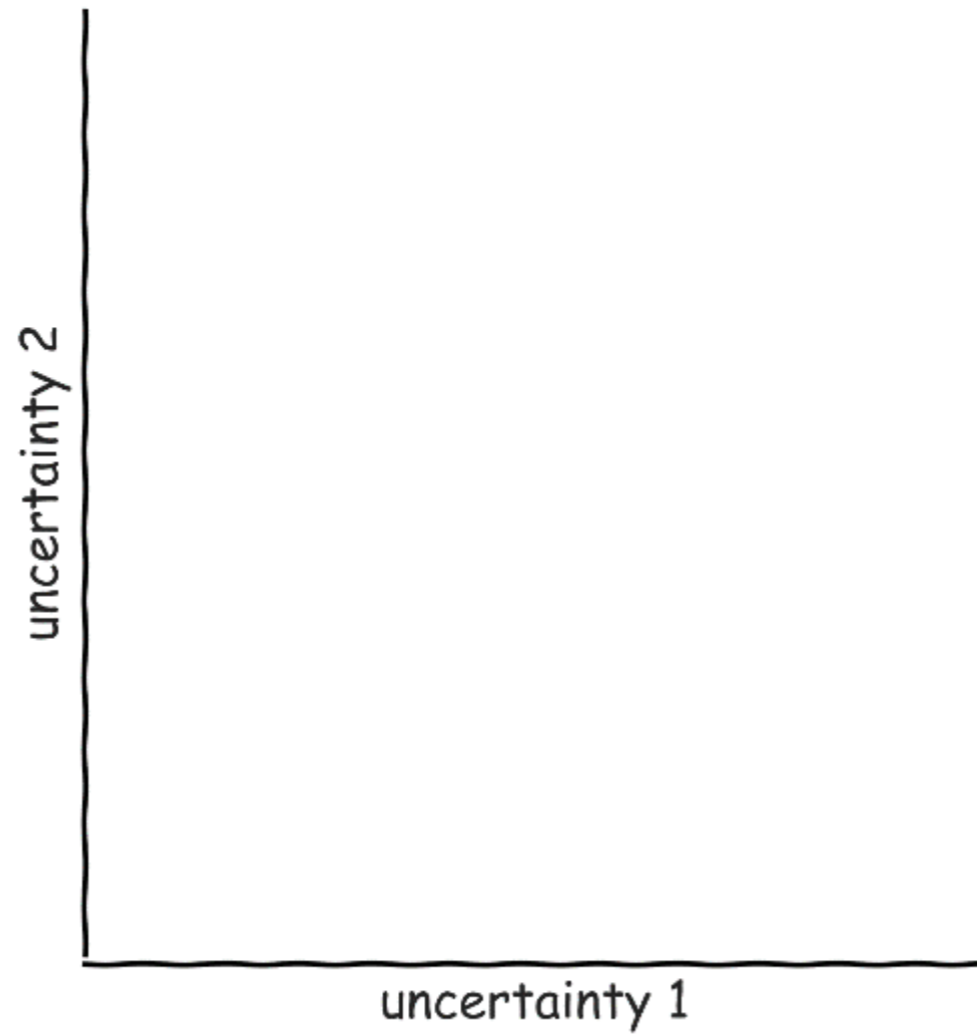


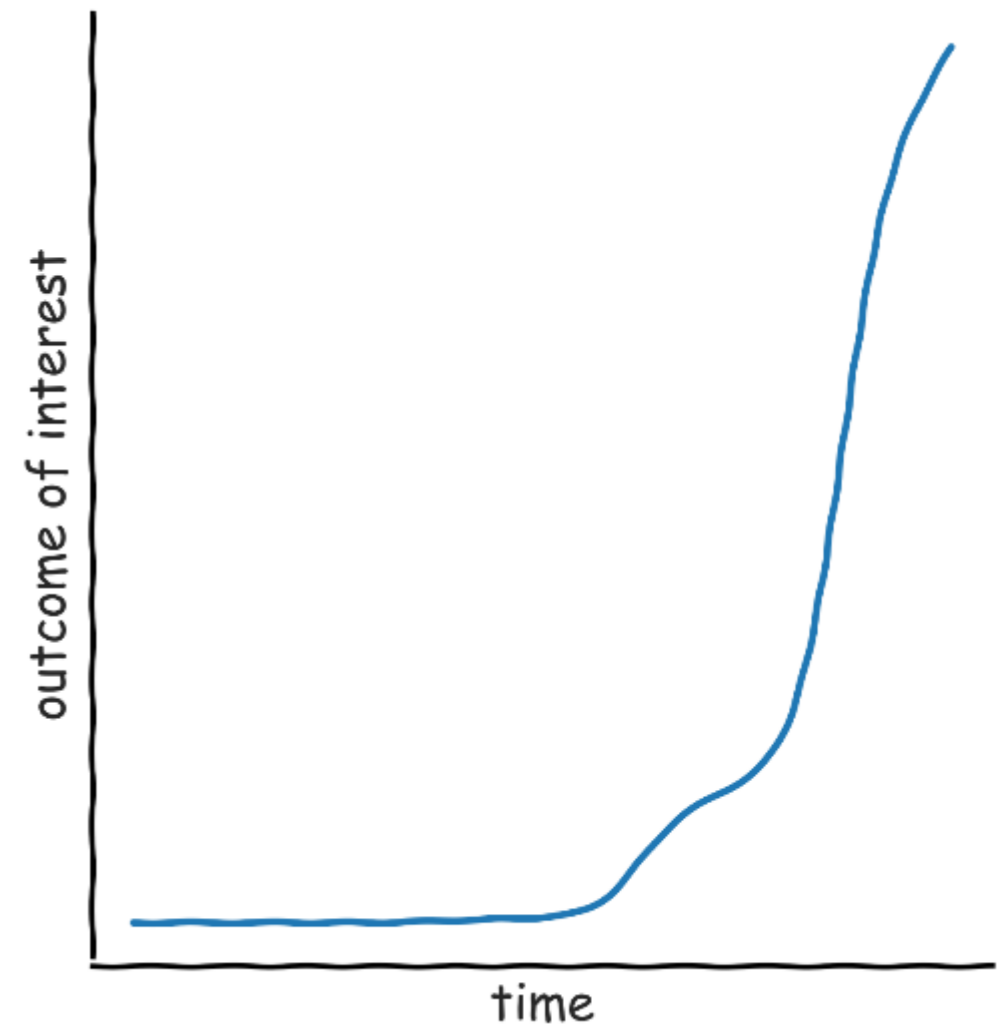
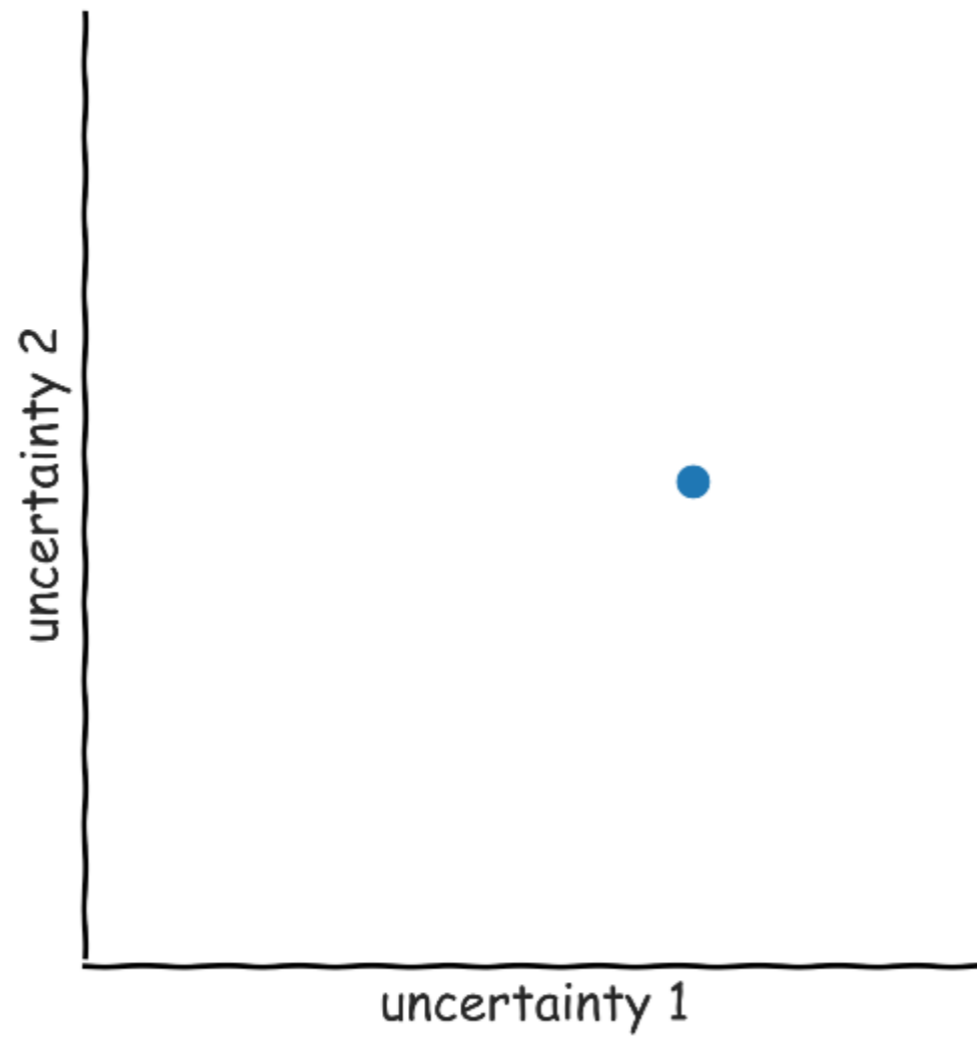
# A generalized many-objective optimization approach for scenario discovery

dr.ir. Jan Kwakkel

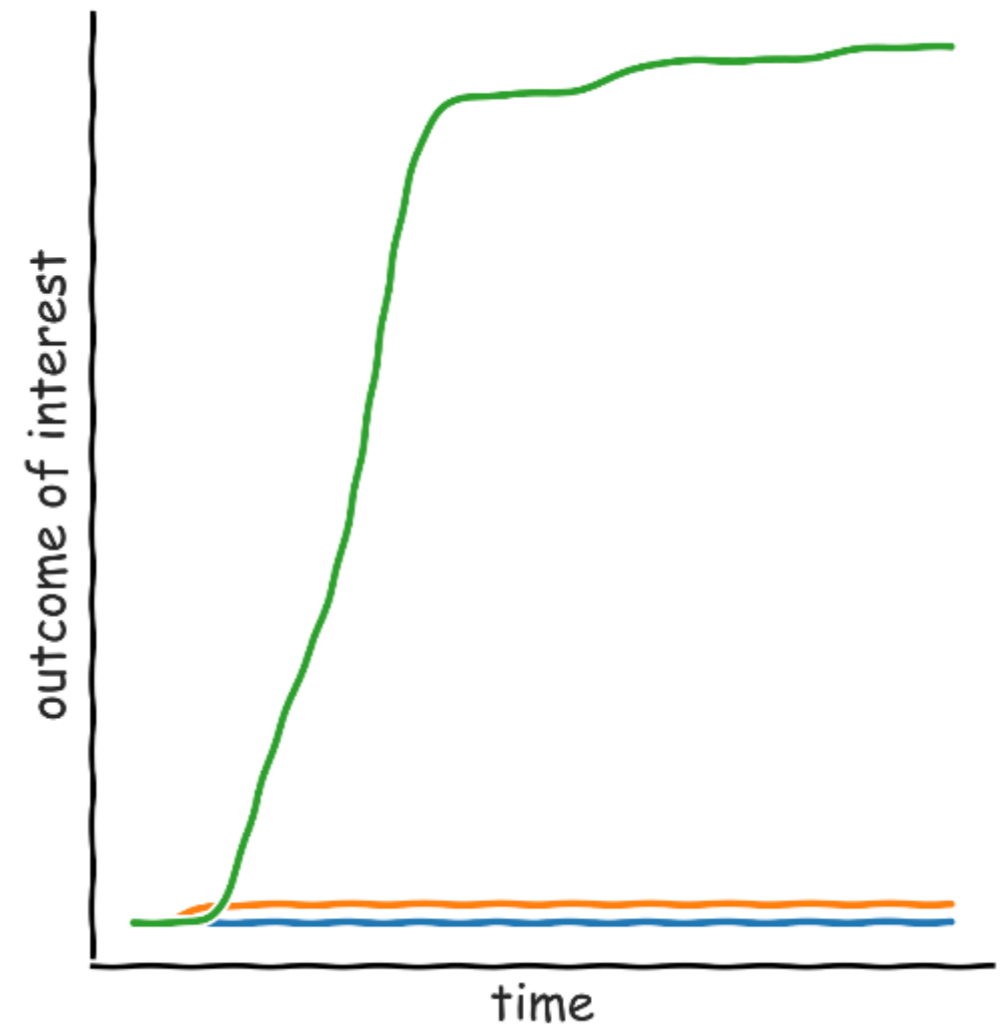
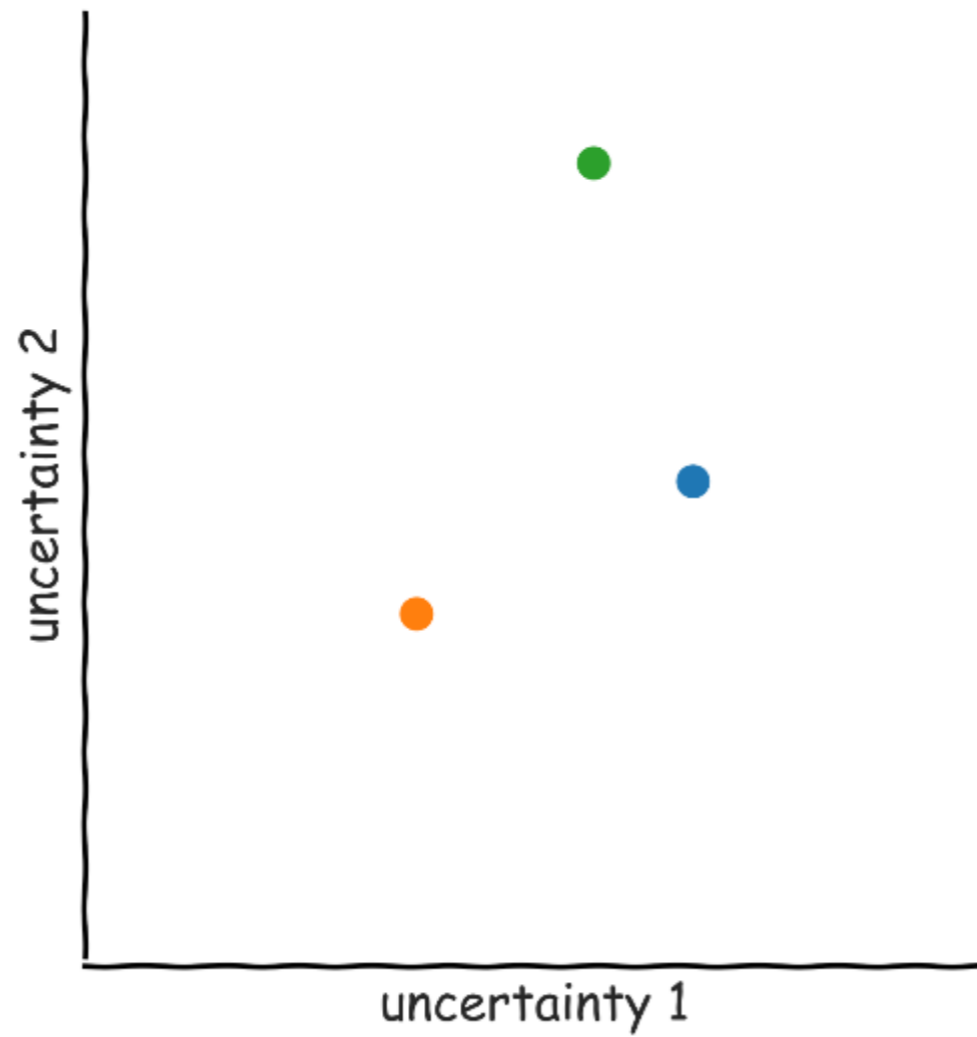
# Scenario Discovery



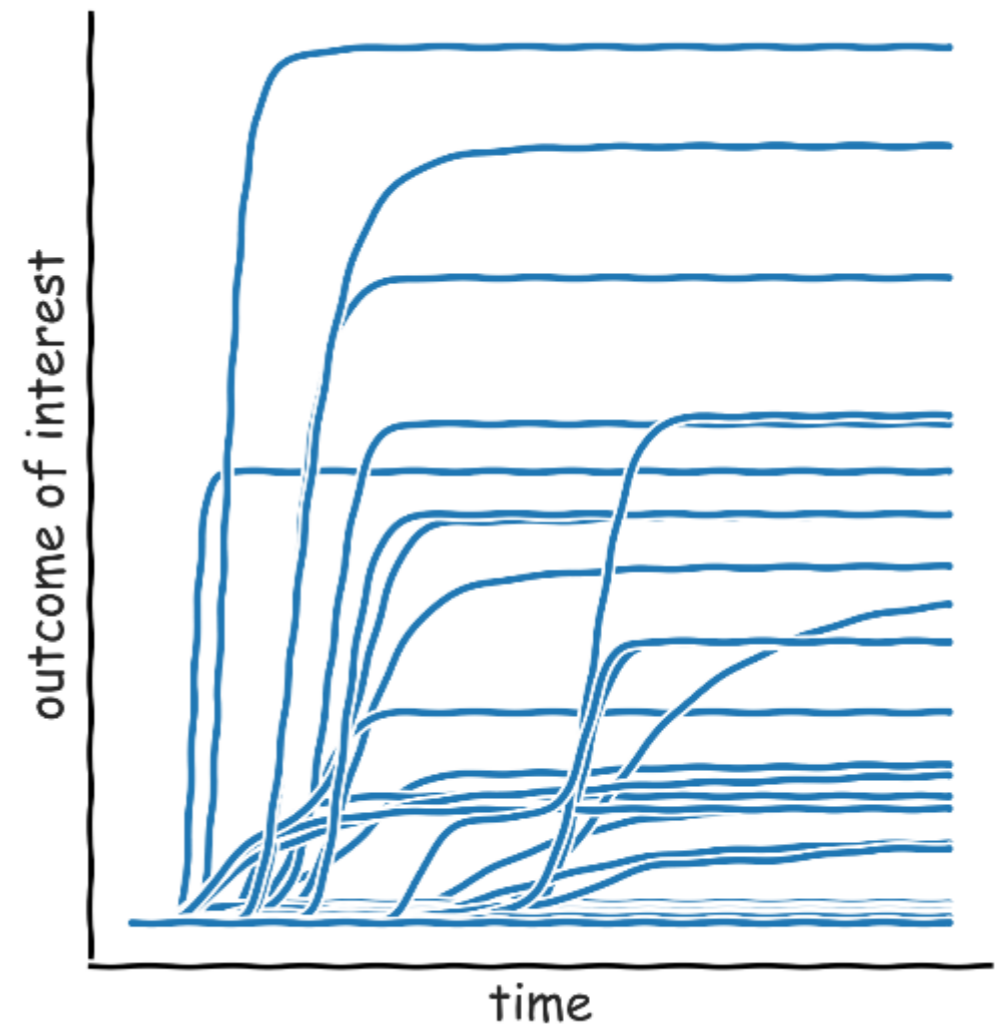
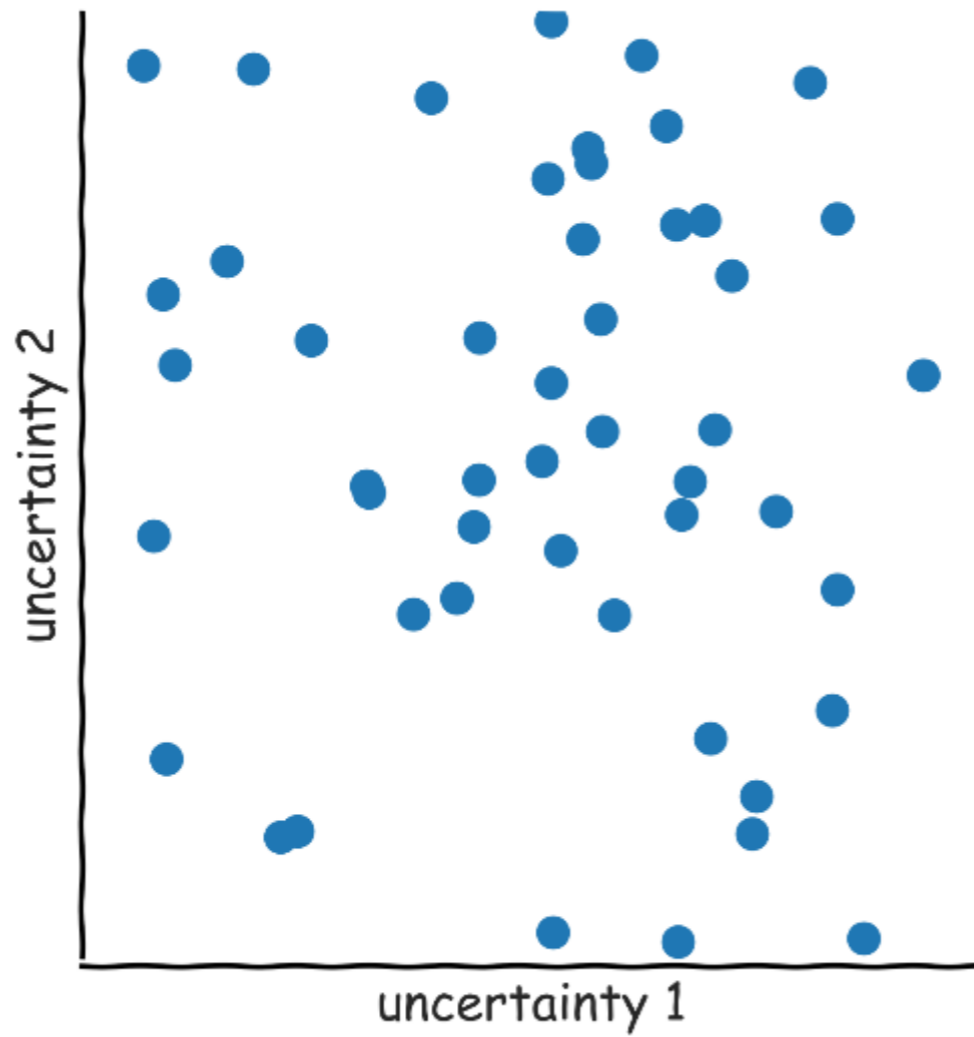
# Scenario Discovery



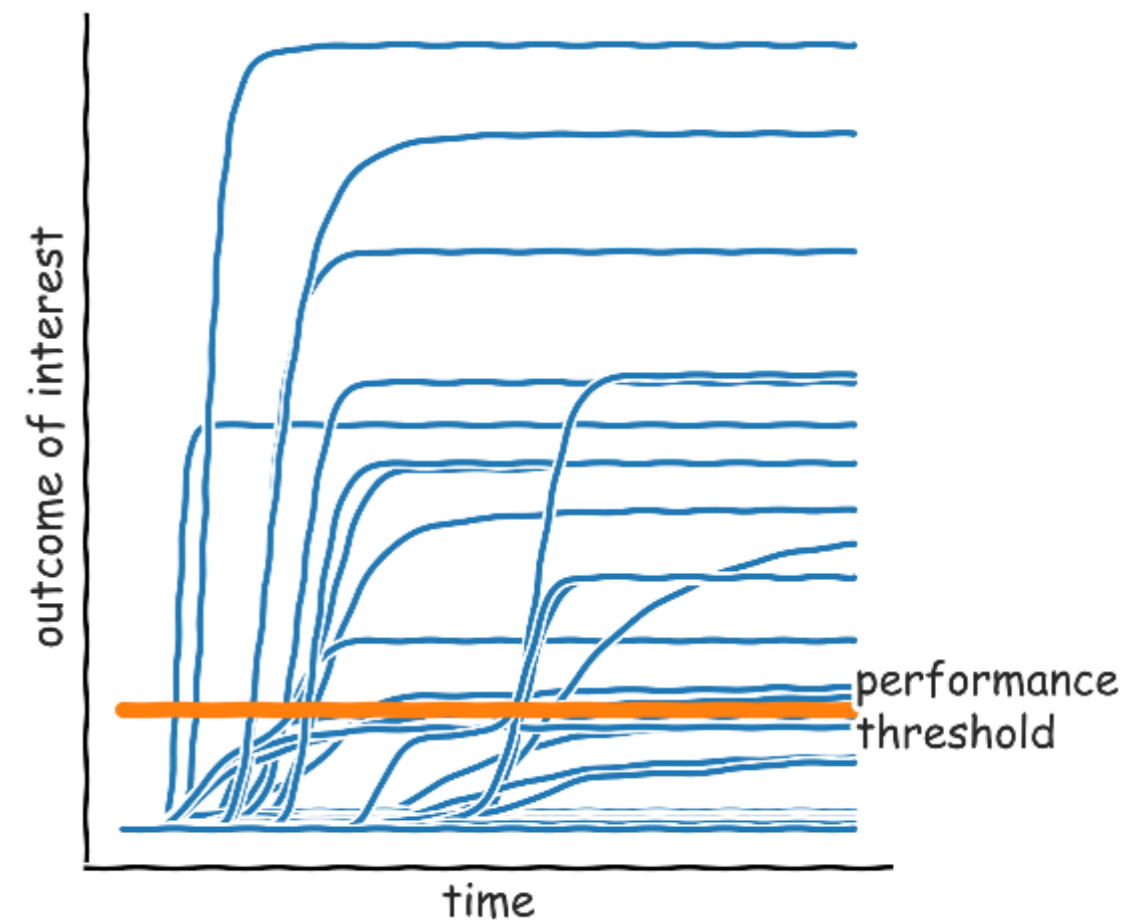
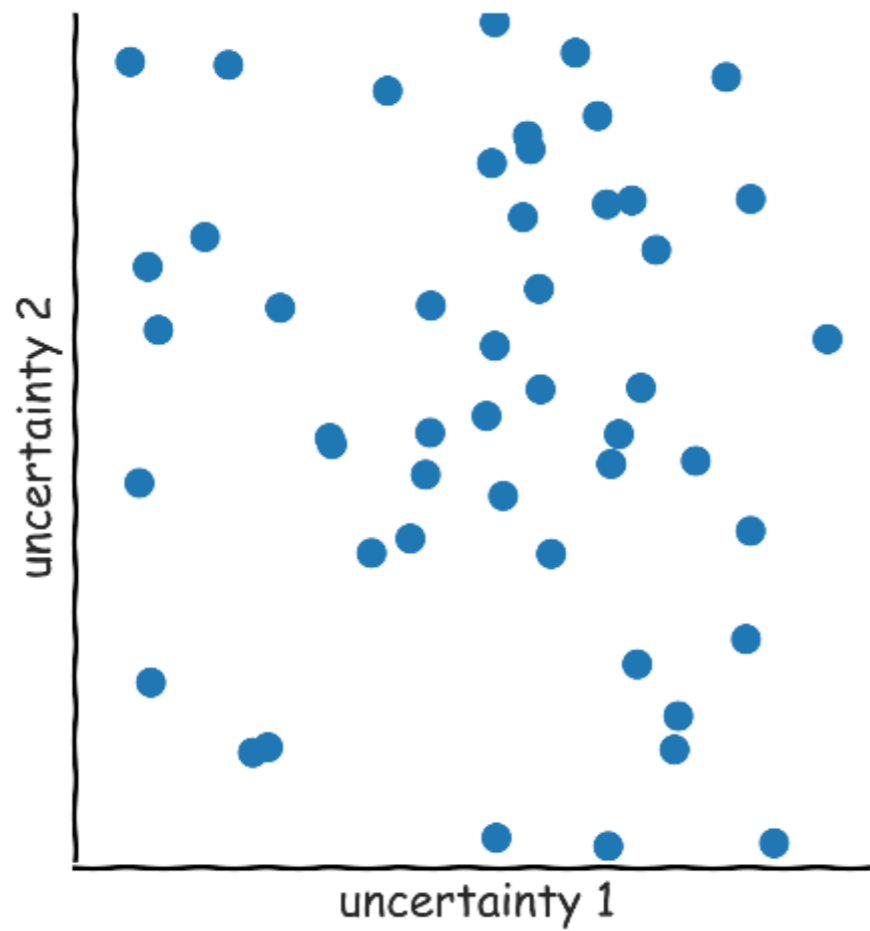
# Scenario Discovery



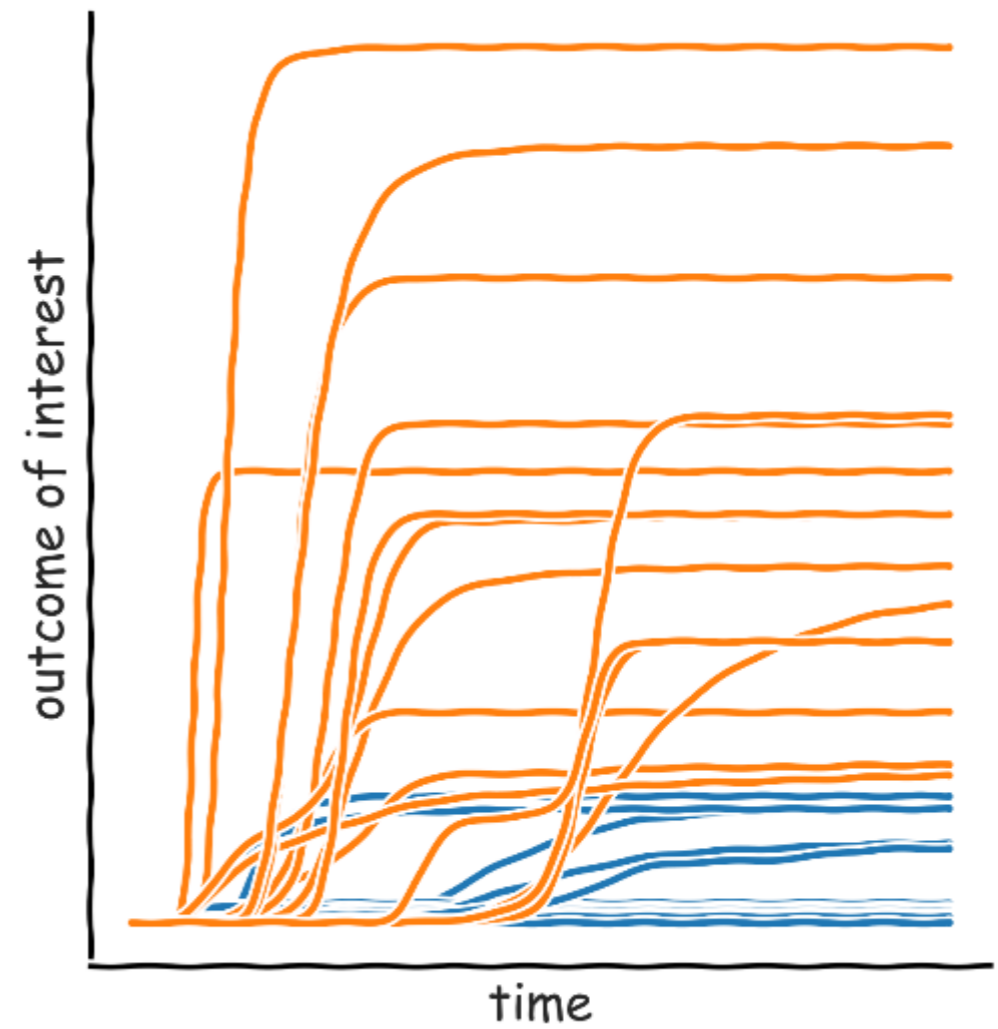
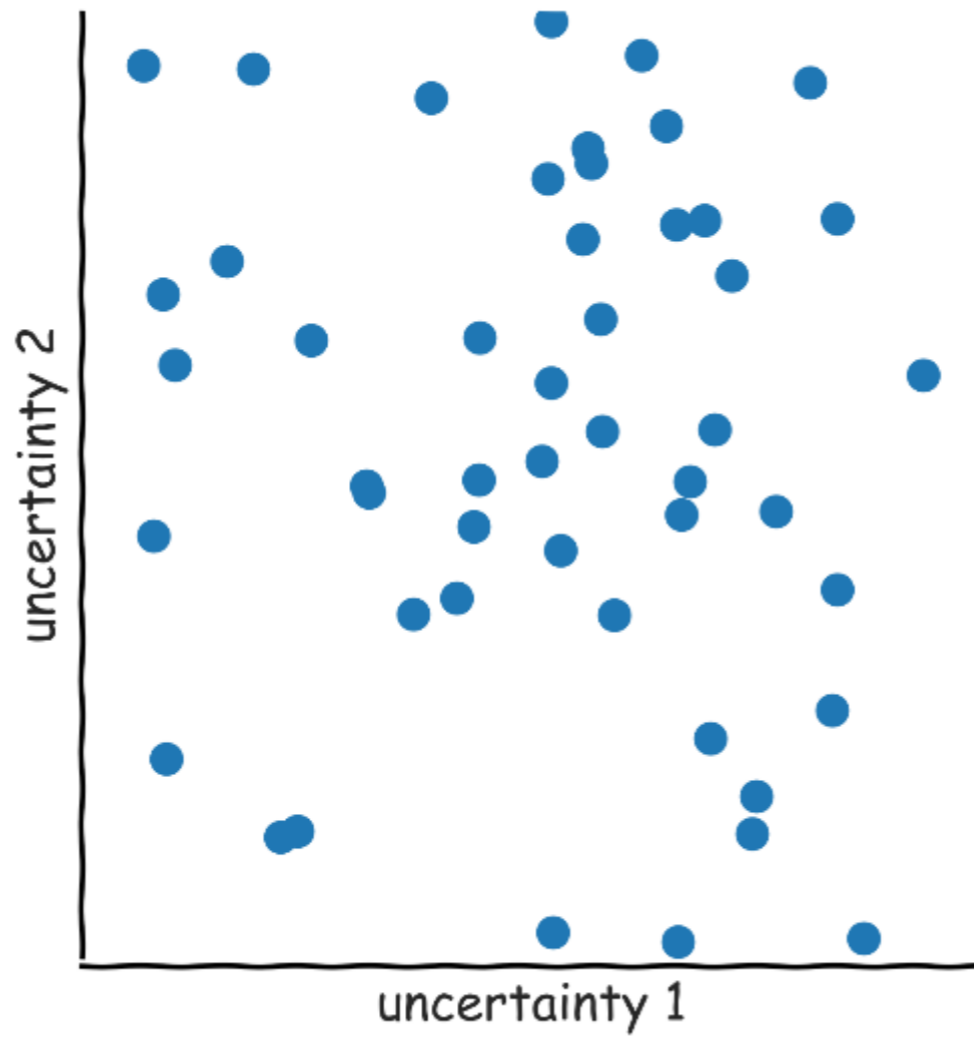
# Scenario Discovery



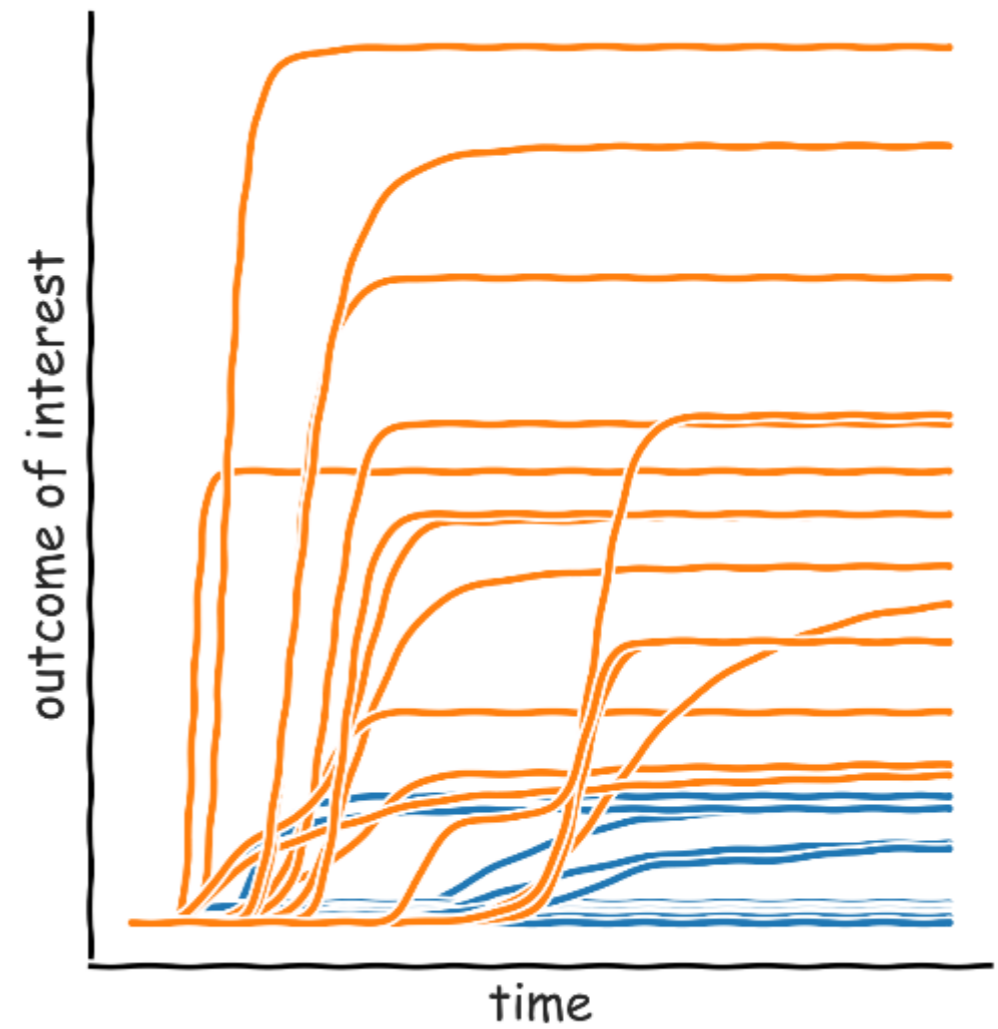
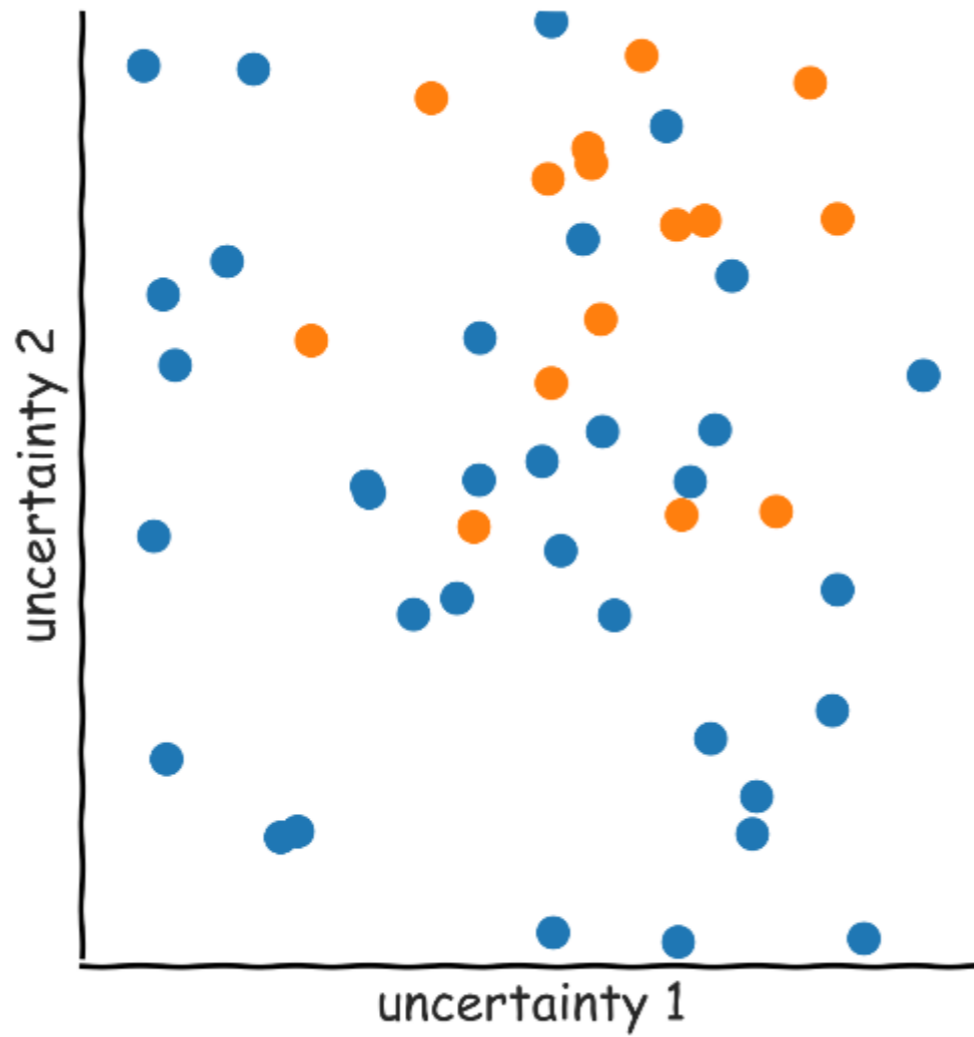
# Scenario Discovery



# Scenario Discovery

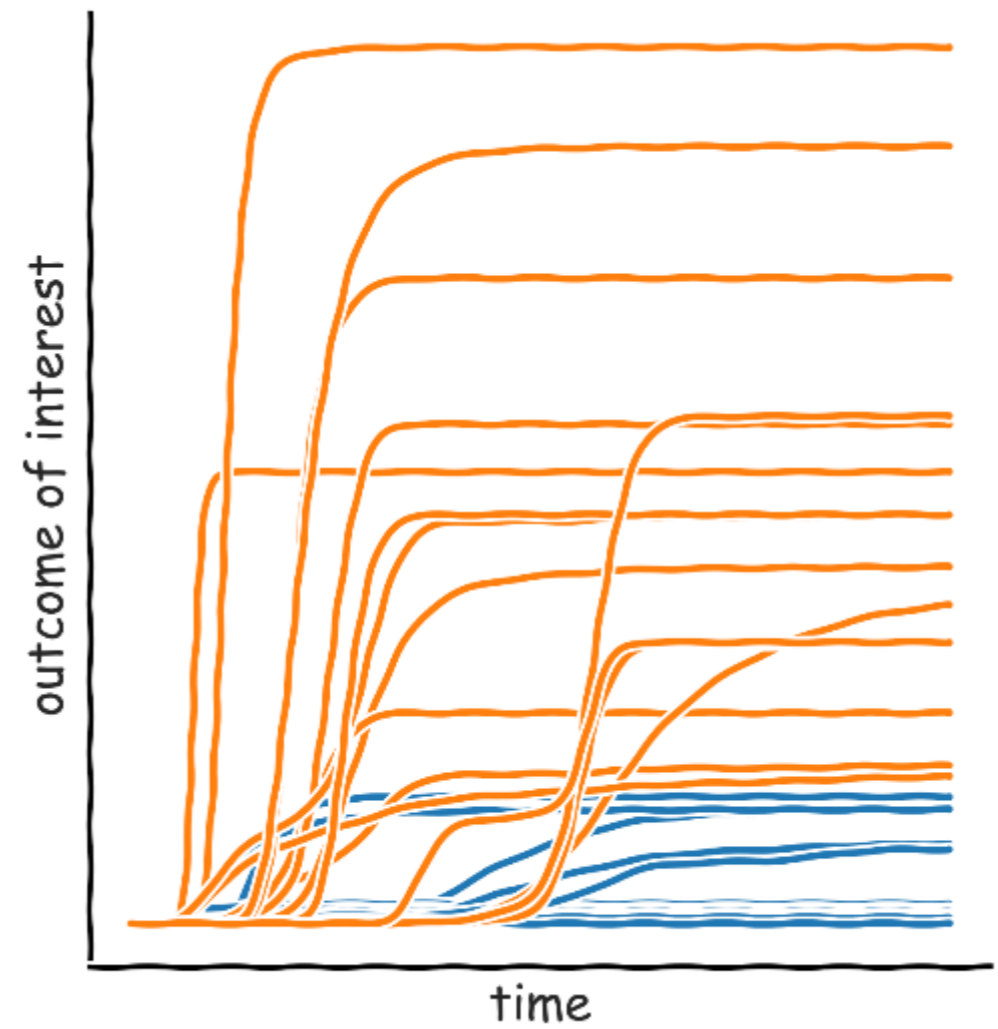
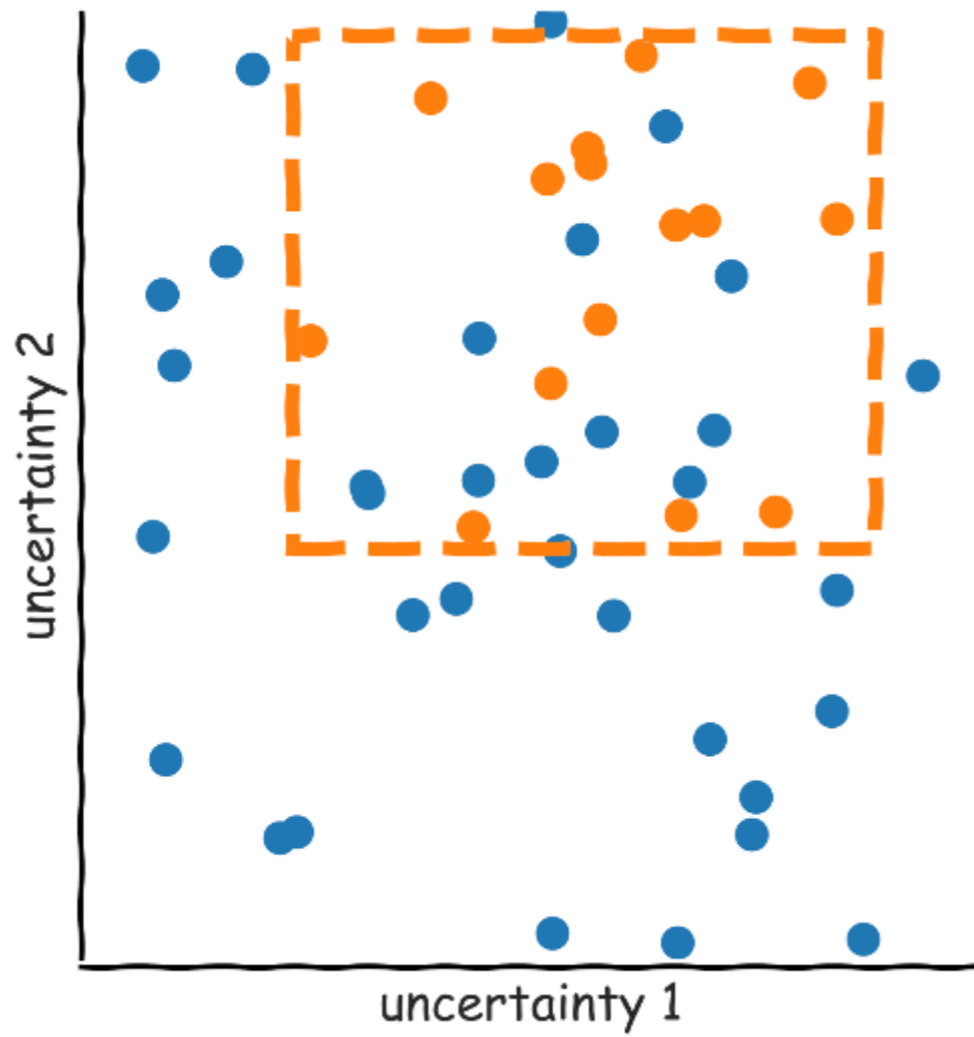


# Scenario Discovery

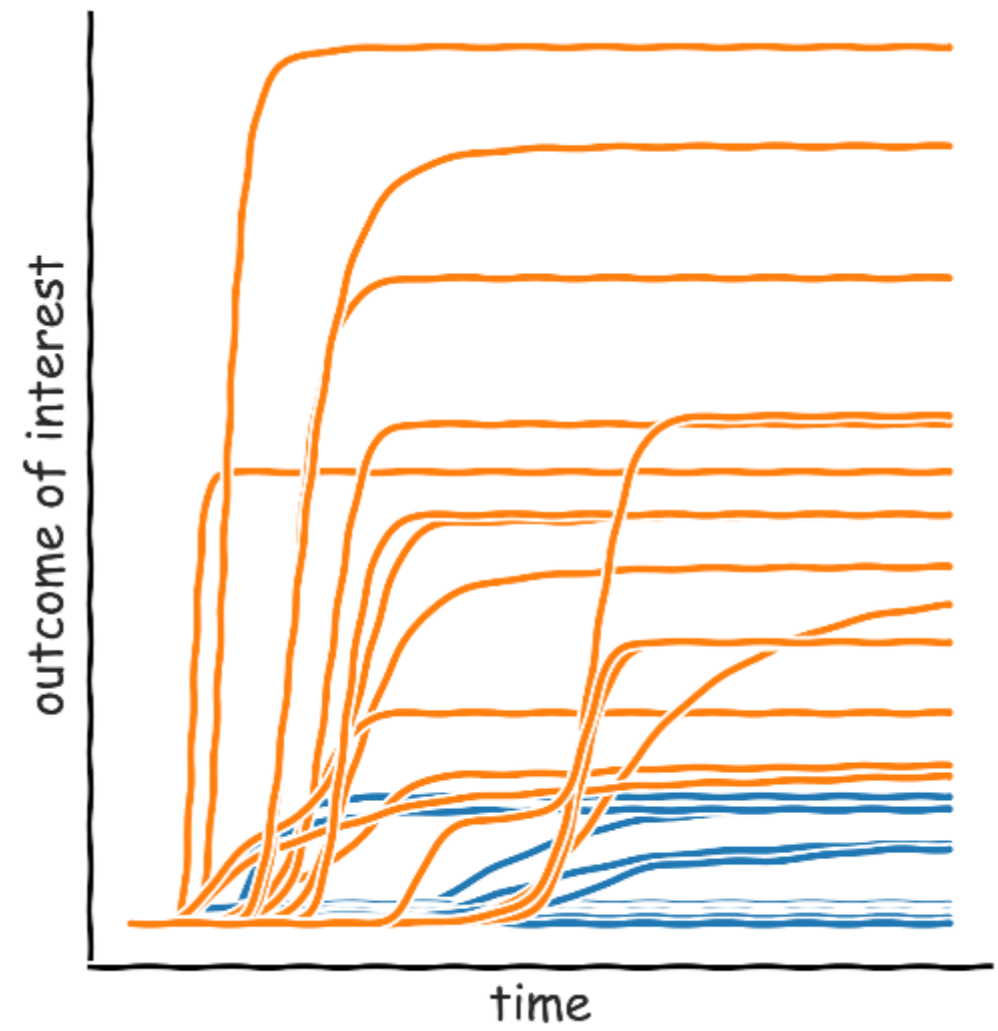
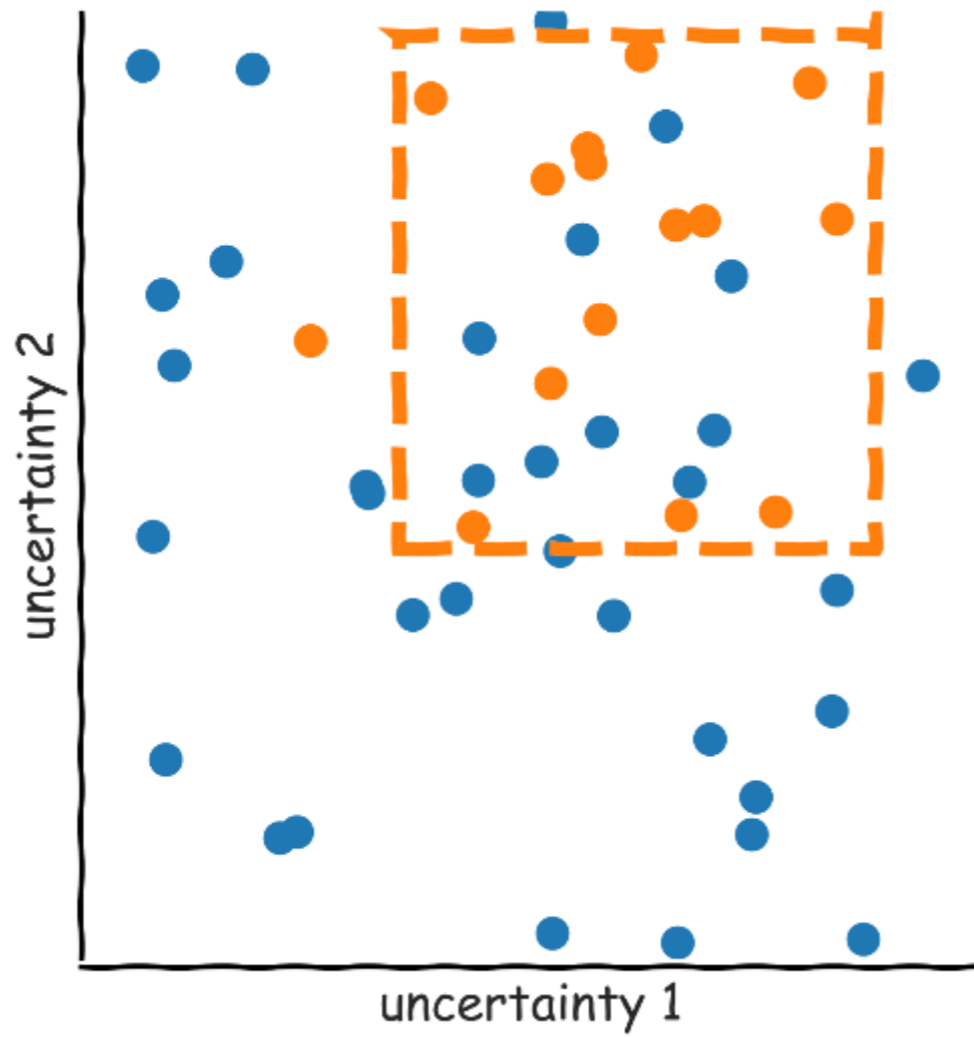


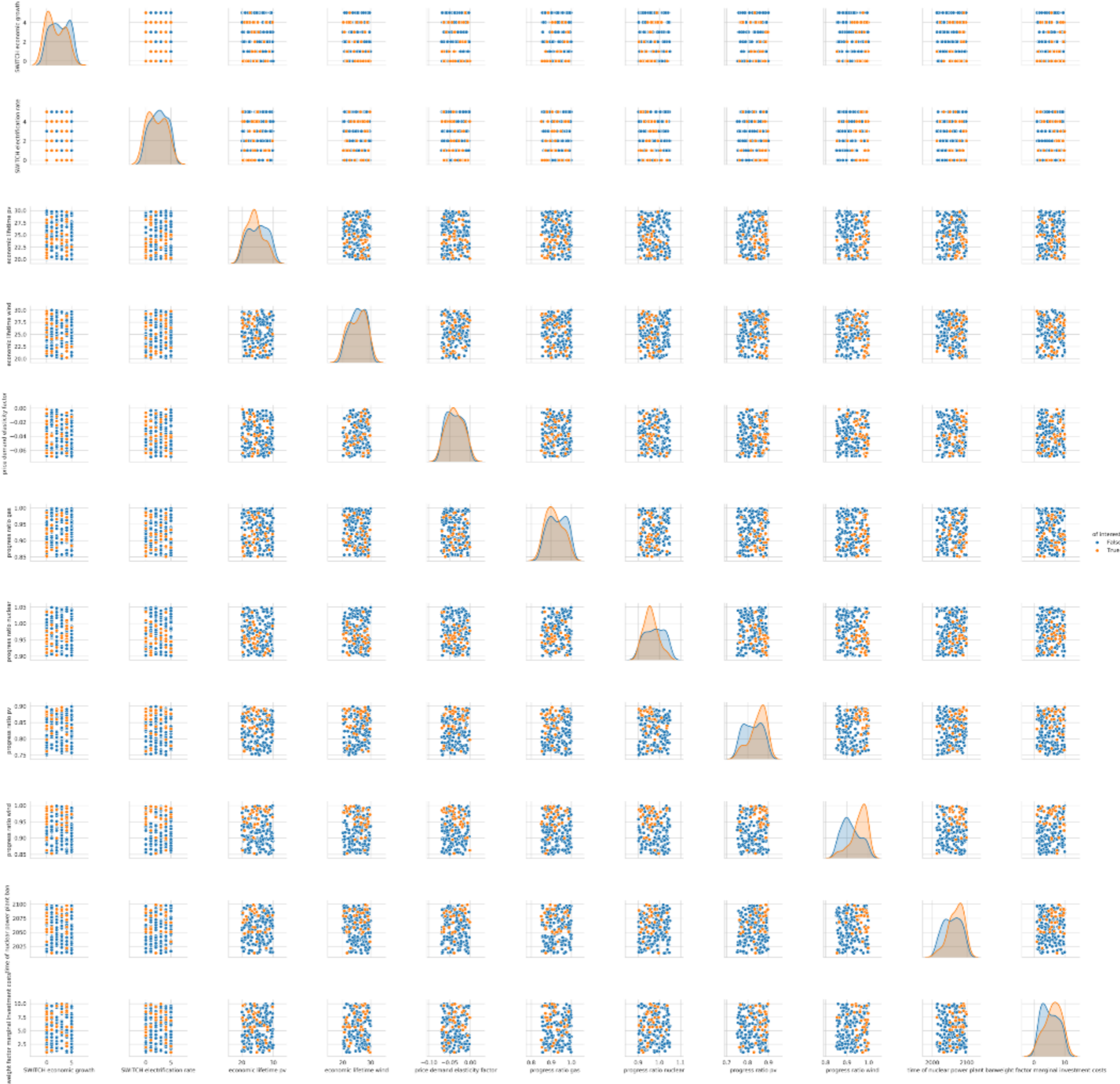


# Scenario Discovery



# Scenario Discovery





# Scenario discovery is a many objective optimization problem

*maximize*  $F(l) = (f_{coverage}, f_{density}, f_{interpretability})$

**Coverage** out of all the cases of interest, how many are in the the box?

**Density** of all the cases in the box, how many are of interest?

**Interpretability** number of restricted dimensions  
note: assumes orthogonal subspaces

# Yet we solve it using mono-objective rule induction algorithms

## Patient Rule Induction Method (**PRIM**)

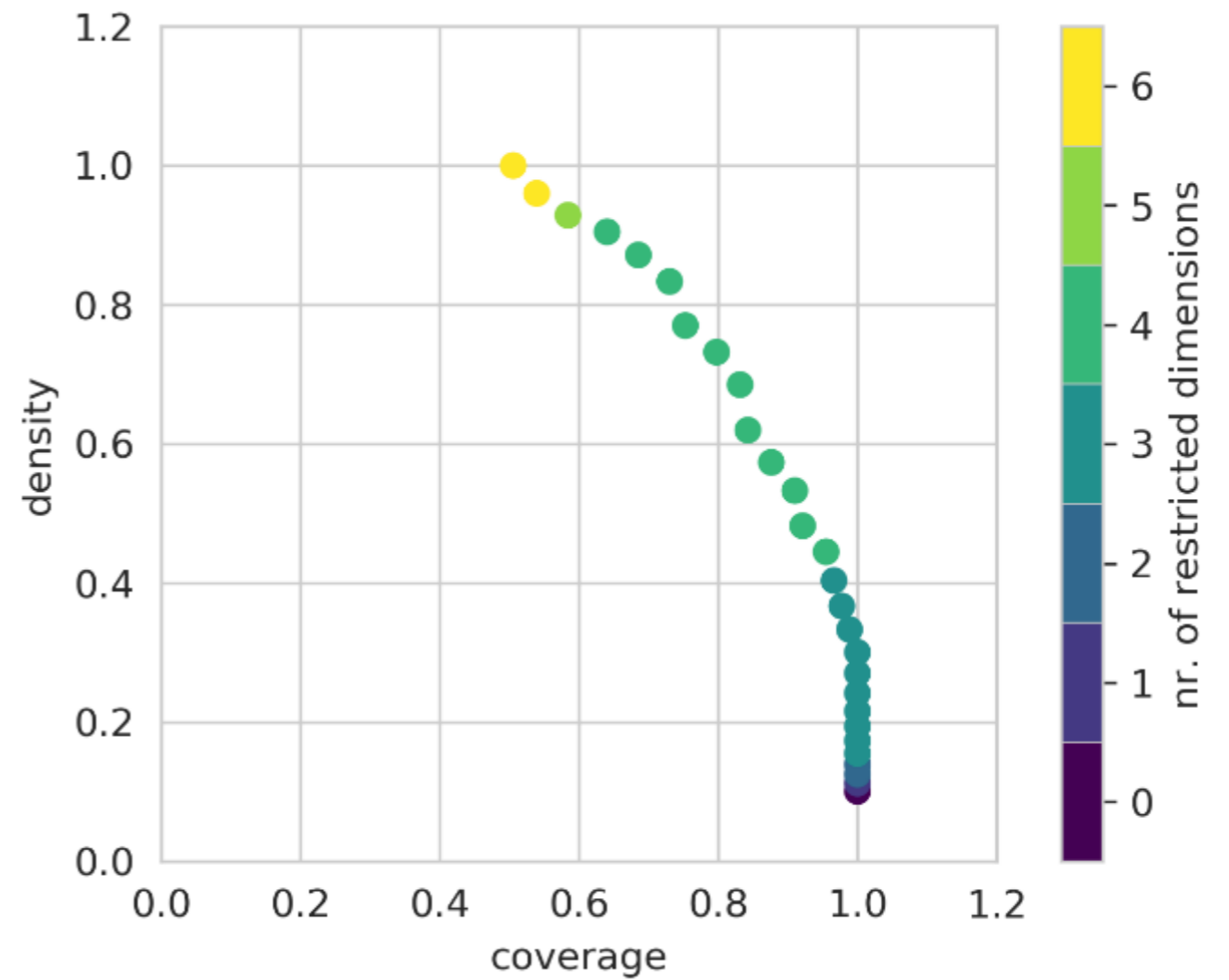
lenient hill climber for maximizing density coverage and interpretability through post processing  
a-posteriori selection of preferred trade-off

## Classification and Regression Trees (**CART**)

greedy minimization of Gini impurity  
coverage and interpretability through post processing

# Drawbacks

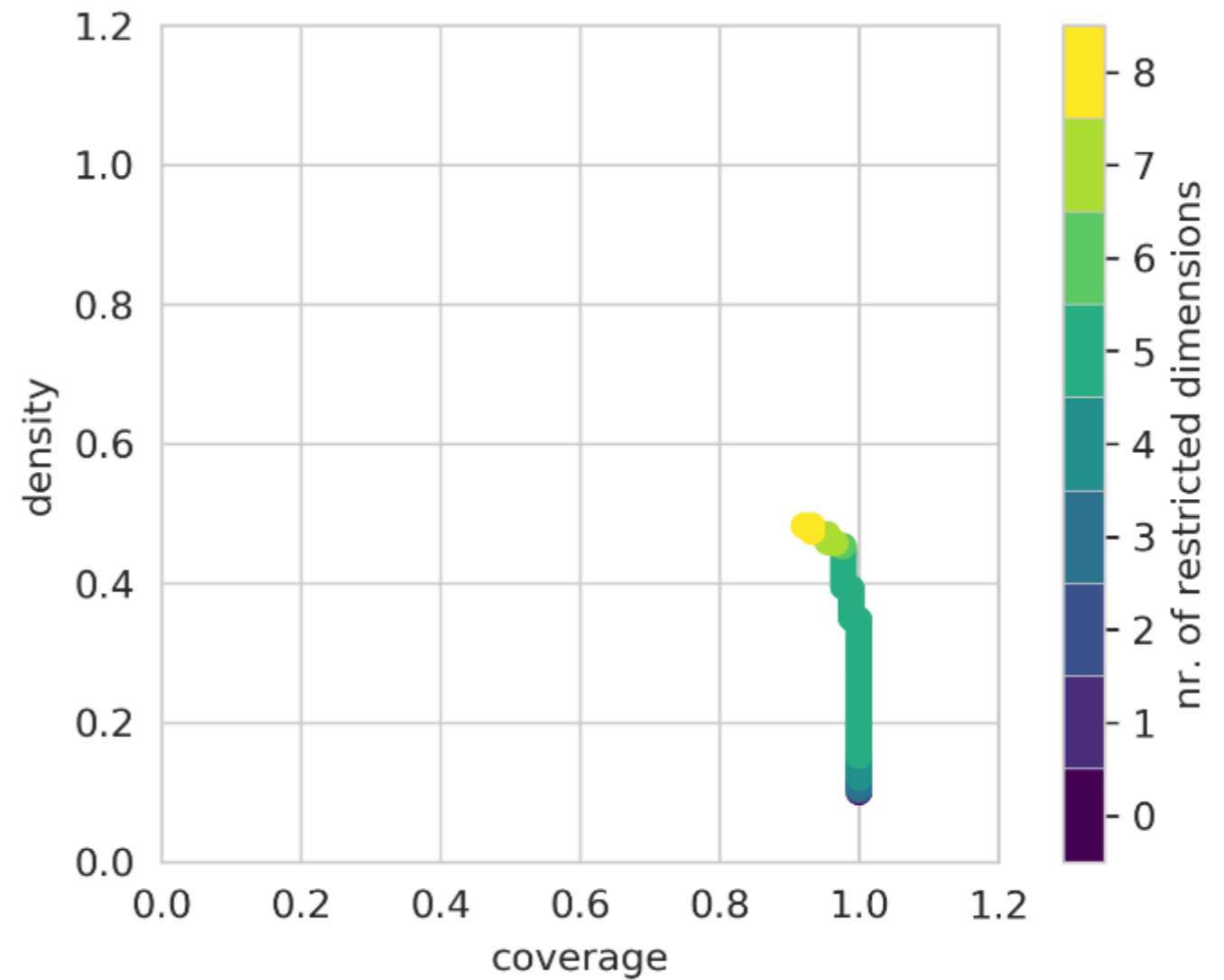
trade-off space not fully explored



# Drawbacks

trade-off space not fully explored

local versus global optimum

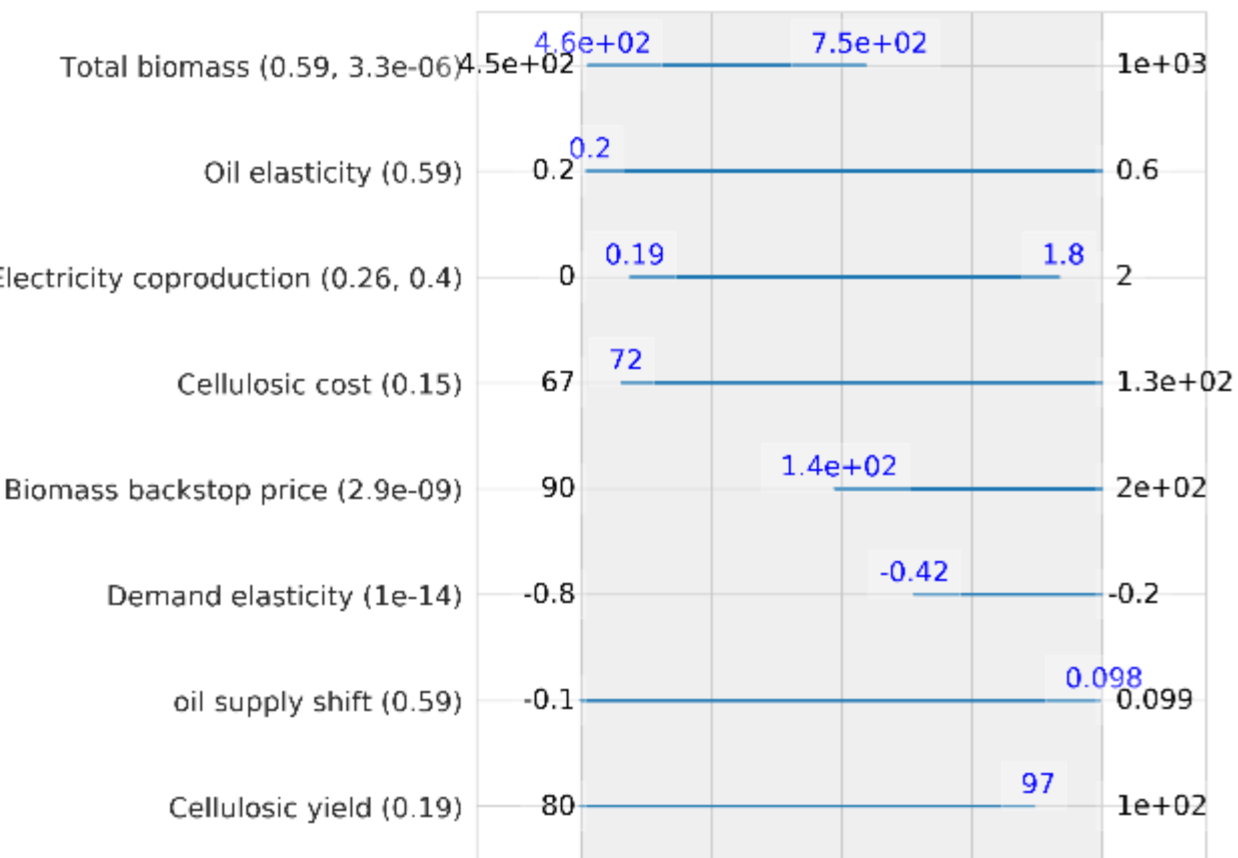


# Drawbacks

trade-off space not fully explored

local versus global optimum

PRIM has a tendency to over fitting



coverage	0.64
density	0.864

	Reproduce coverage	Reproduce density
Biomass backstop price	70	100
Total biomass	70	90
Cellulosic yield	60	90
Feedstock distribution	60	80
Electricity coproduction	60	80
Cellulosic cost	50	80
Oil elasticity	50	80
Demand elasticity	50	80
oil supply shift	50	80



# Improving usage of PRIM

## Map trade-off space in more detail

run PRIM using subsets of the uncertain factors →

$1, 2, 3, \dots, n$  combinations of significant uncertain factors

merge results runs into a single set of results using non dominated sort

## Address over fitting

quasi-p values (Bryant & Lempert 2010), one sided binomial test, for each restricted dimension

remove any candidate box that has non-significant quasi-p values

# Solving it using a Many-Objective Evolutionary Algorithm

## $\epsilon$ -NSGAII

### Initialization

real valued: truncated exponential distribution

integer and categorical valued: Boltzmann distribution

limited subset of uncertain factors

### Evolution

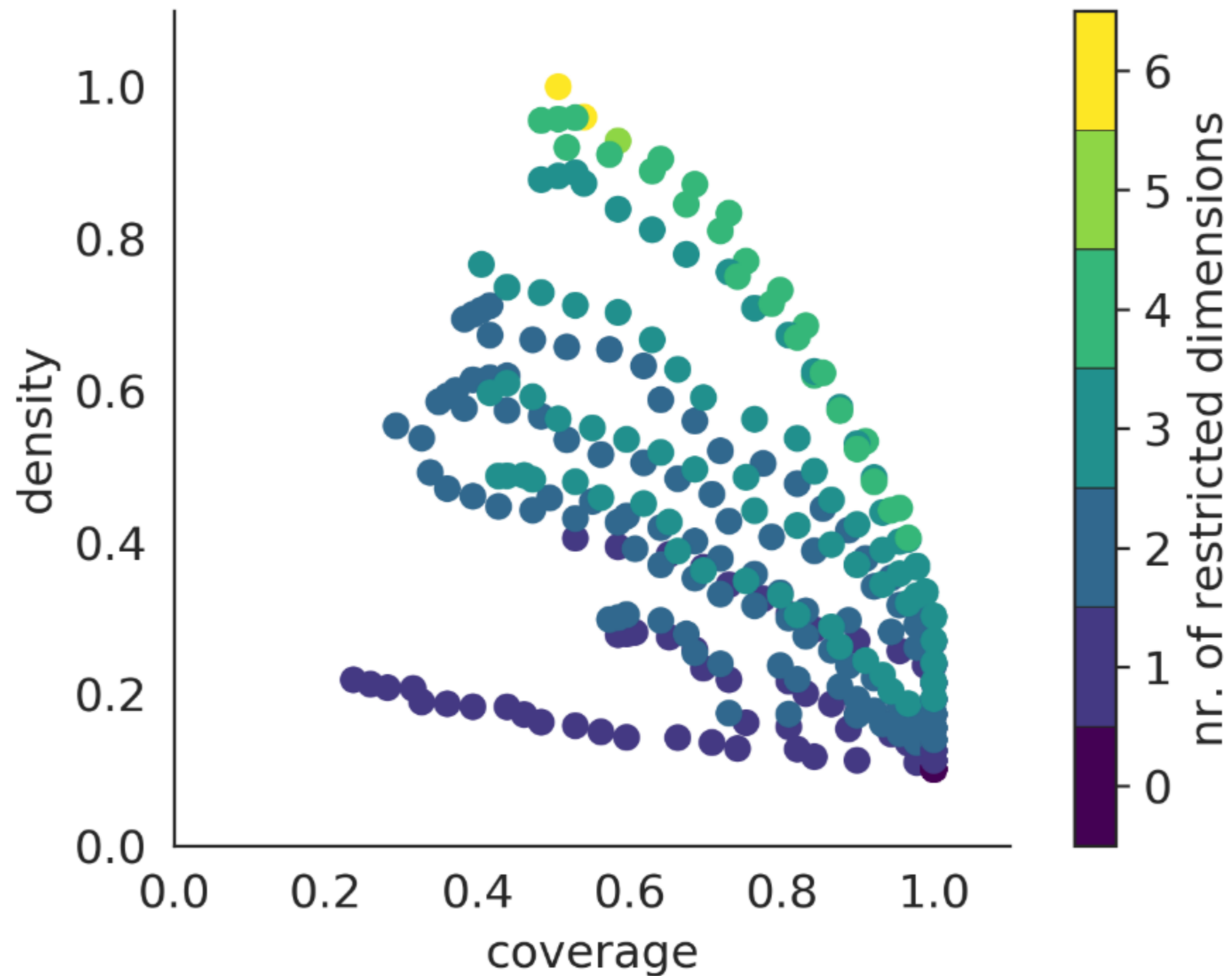
GeneAS (Deb, 1999) for handling heterogeneously typed decision variables

Mutation stays quite close to original value

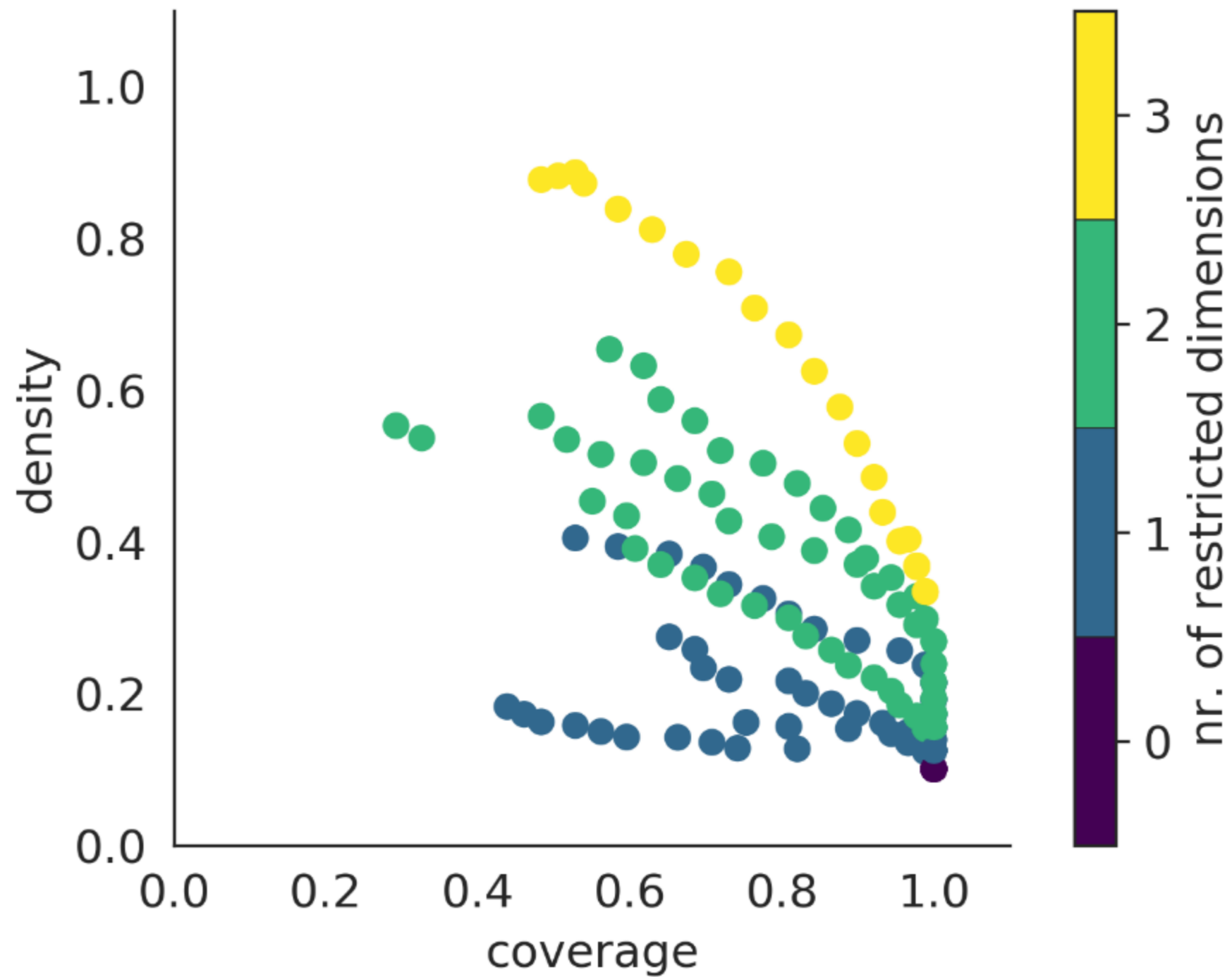
### Constraints

quasi p-values used as constraints to avoid over fitting

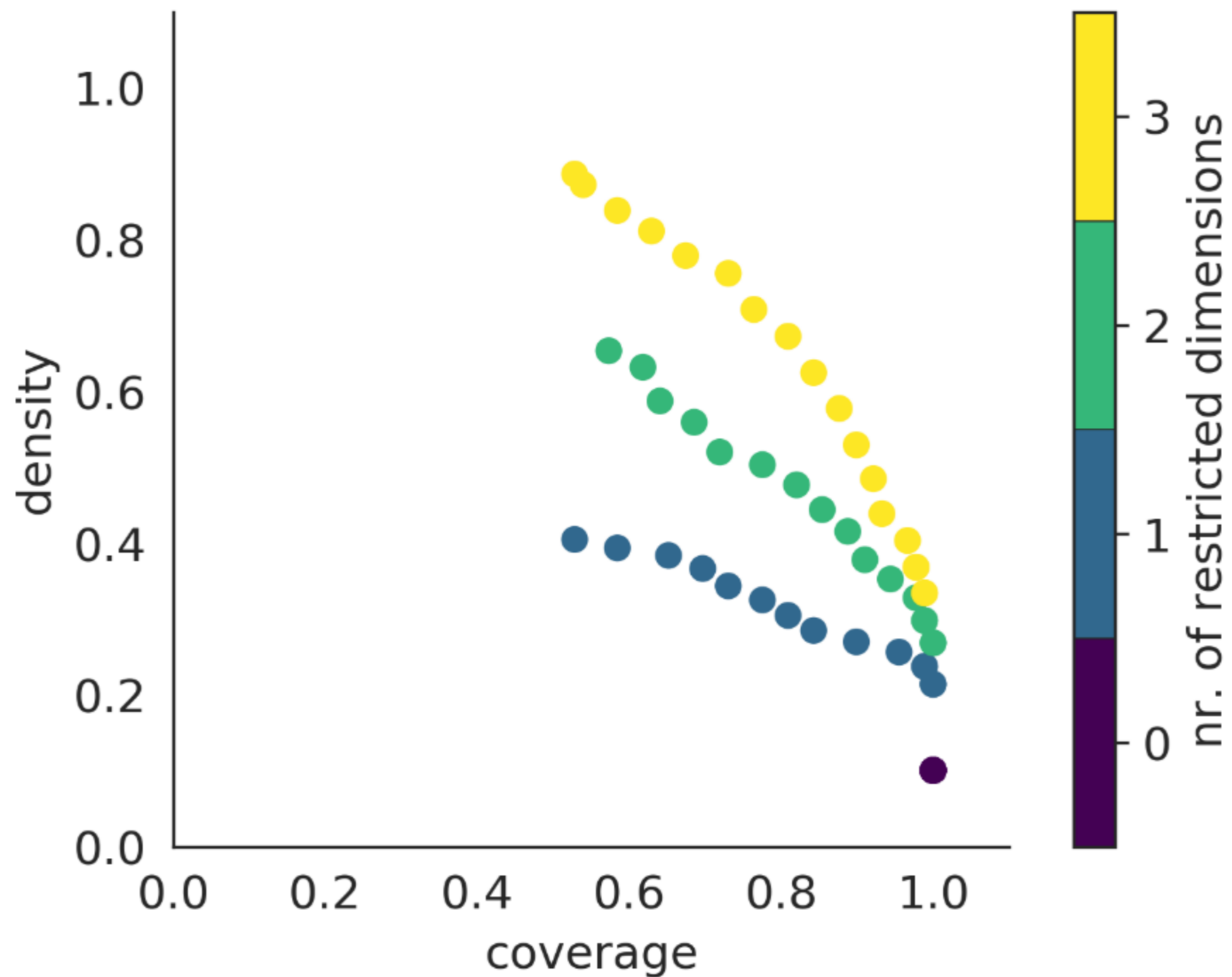
# Case 1: Bryant & Lempert (2010) dataset



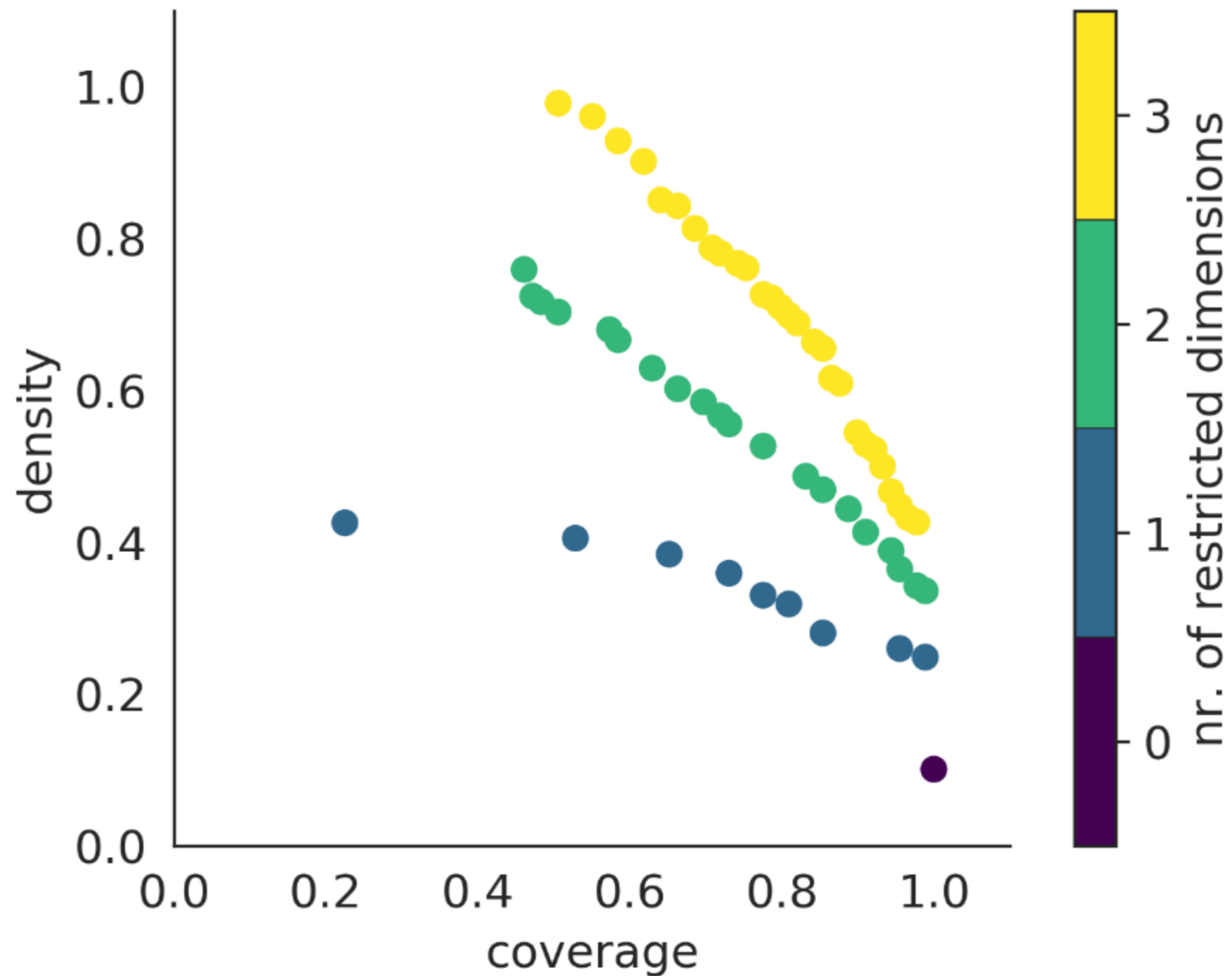
# Case 1: Bryant & Lempert (2010) dataset



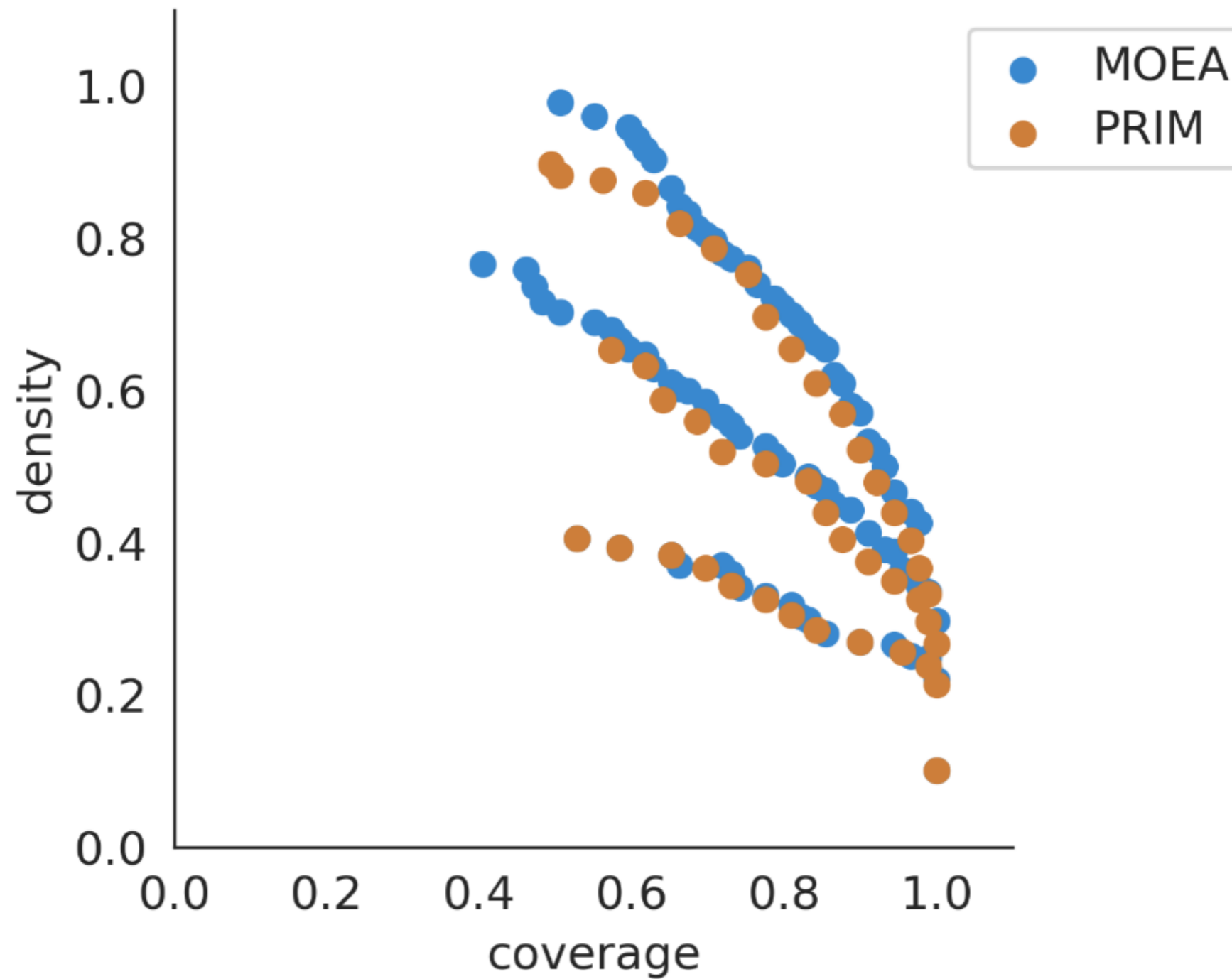
# Case 1: Bryant & Lempert (2010) dataset



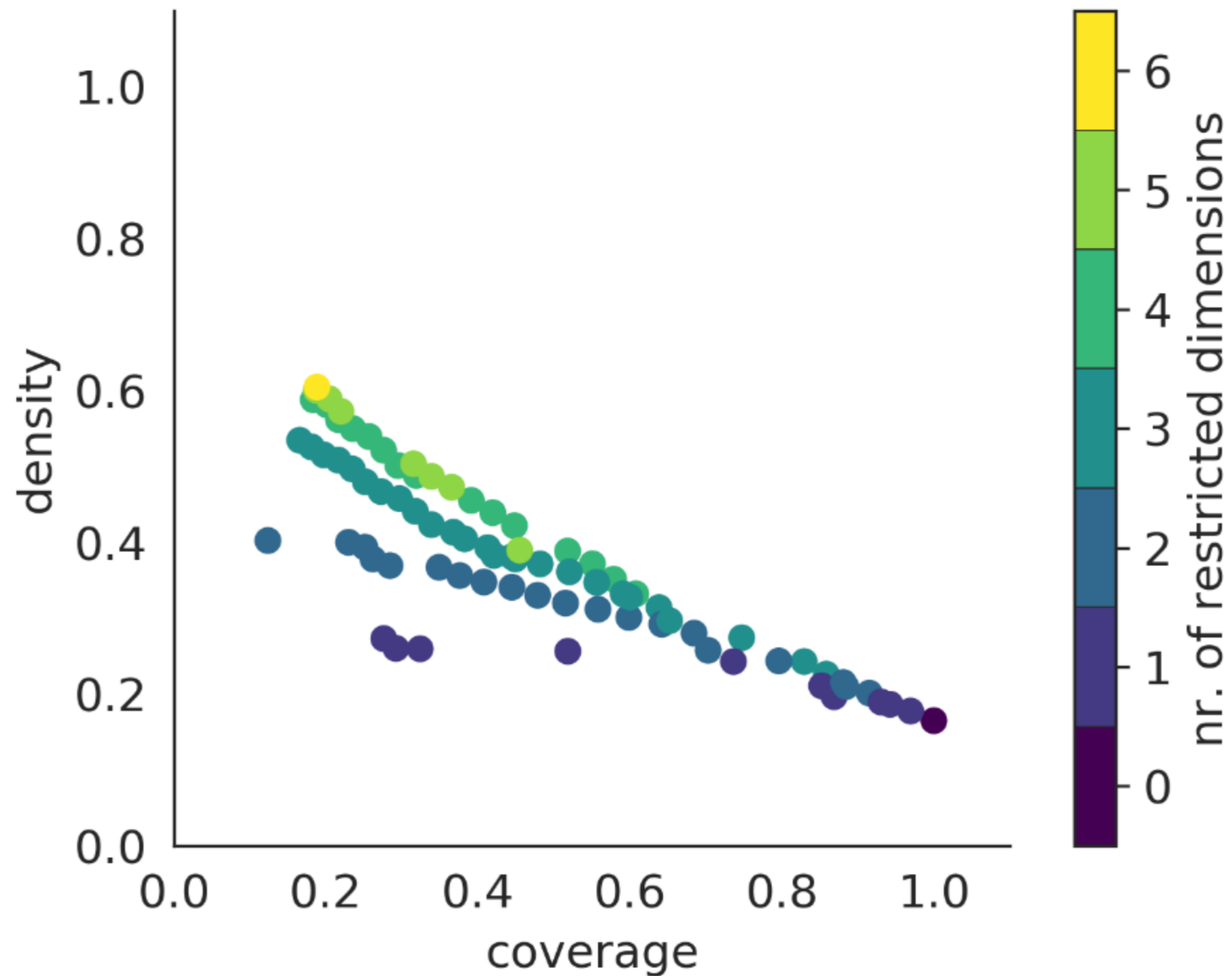
# Case 1: Bryant & Lempert (2010) dataset



# Case 1: Bryant & Lempert (2010) dataset

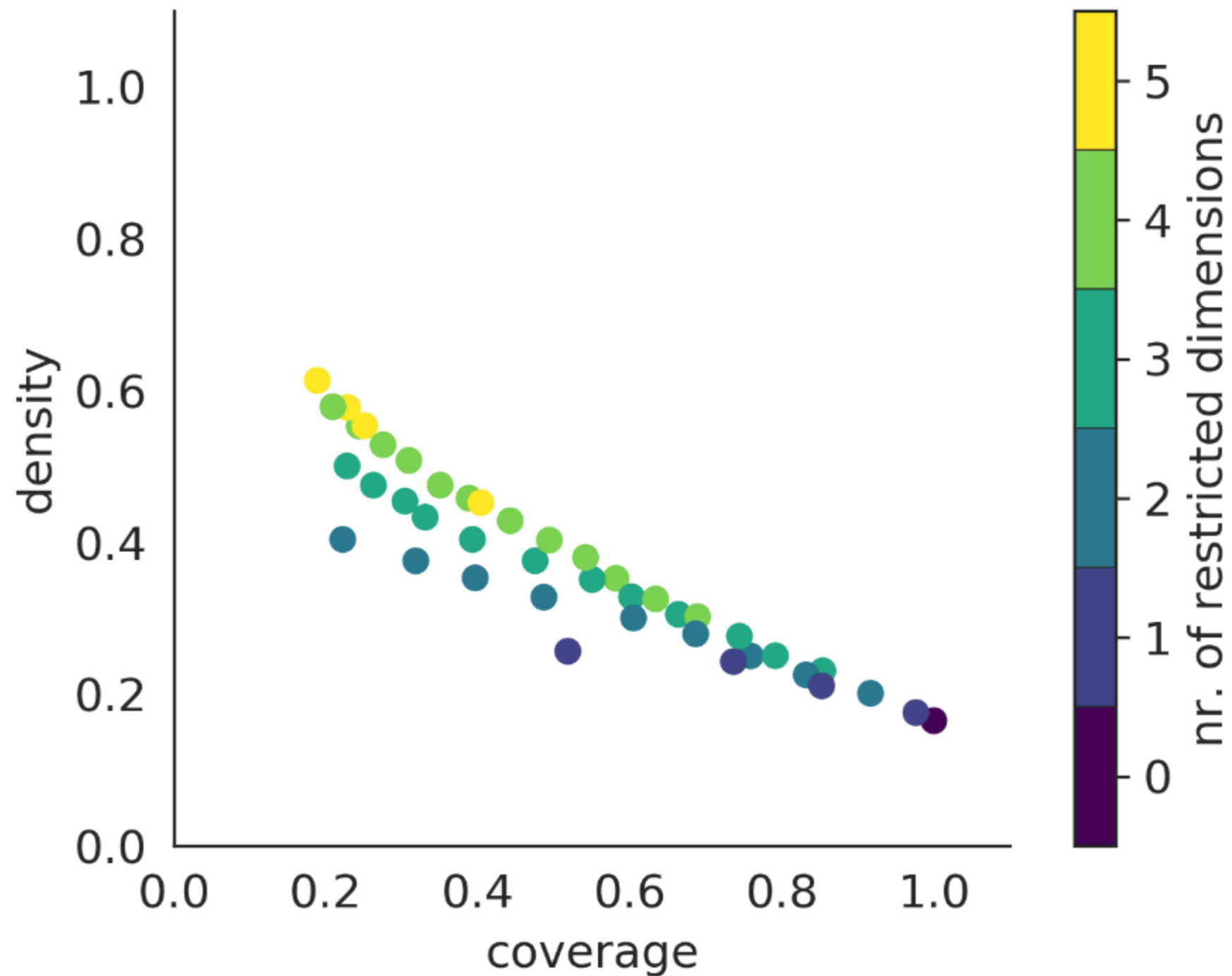


## Case 2: Hamarat et al (2013) dataset

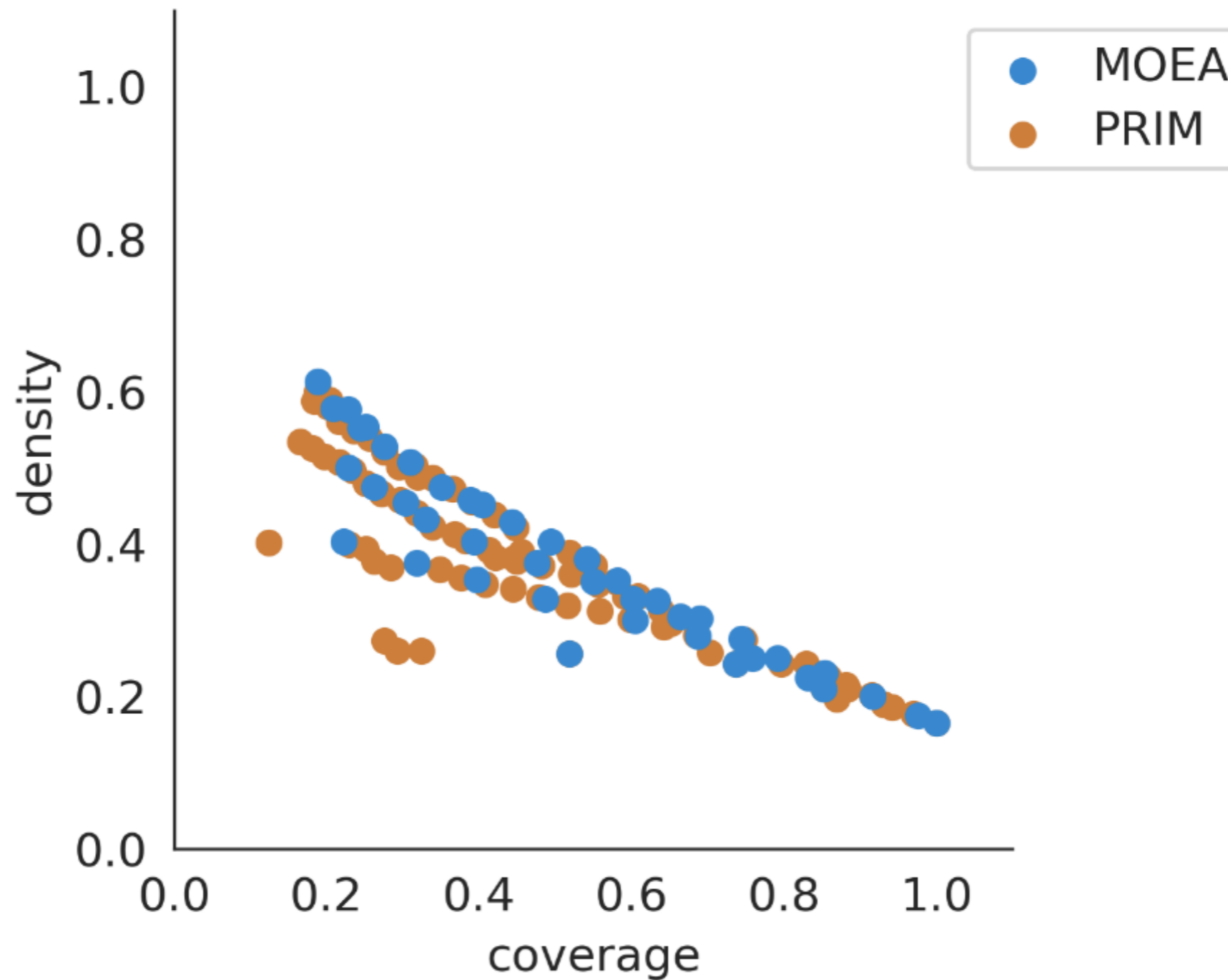




## Case 2: Hamarat et al (2013) dataset



## Case 2: Hamarat et al (2013) dataset



# Conclusions

## **Scenario discovery is a many-objective optimization problem**

expandable with additional objectives (e.g. consistency)

## **Solving the problem using a MOEA**

viable

dominates (slightly) results from improved PRIM based approach

computationally expensive

## **Future work: genetic programming approach**

generating non orthogonal subspaces

Finding multiple subspaces in one go