

Supplementary Report

Contents

1	Initialization	1
1.1	Load data	1
1.2	Clean data	1
2	Overview	2
2.1	Summary	2
2.2	Problems	2
2.3	Models	2
2.4	Configurations	4
2.5	Output metrics	5
3	Success rates (RQ1)	7
4	Variance of the output metrics (RQ2)	8
4.1	Individual distributions	8
4.2	Individual deviations	15
4.3	Summarized deviations	16

This is a supplementary report for the paper “**A Preliminary Analysis on the Effect of Randomness in a CEGAR Framework**” by **Akos Hajdu and Zoltan Micskei**, presented at the **25th PhD Mini-Symposium** (2018), organized by the *Department of Measurement and Information Systems* at the *Budapest University of Technology and Economics*.

1 Initialization

1.1 Load data

1.2 Clean data

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 10800 obs. of 13 variables:
## $ Model : Factor w/ 30 levels "eca/prob1_lab15",...: 14 14 14 14 14 14 14 14 14 14 ...
## $ Domain : Factor w/ 2 levels "EXPL","PRED": 1 1 1 1 1 1 1 1 1 1 ...
## $ Refinement : Factor w/ 2 levels "BIN","SEQ": 2 2 2 2 2 2 2 2 2 2 ...
## $ Result : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ TimeMs : int 28290 27241 27859 27581 27146 27307 27271 27633 27208 27318 ...
## $ Iterations : int 2 2 2 2 2 2 2 2 2 2 ...
## $ ArgSize : int 2 2 2 2 2 2 2 2 2 2 ...
## $ ArgDepth : int 1 1 1 1 1 1 1 1 1 1 ...
## $ ArgMeanBranchFactor: num 1 1 1 1 1 1 1 1 1 1 ...
## $ CexLen : int NA NA NA NA NA NA NA NA NA NA ...
## $ Randomized : Factor w/ 3 levels "DET","SEARCH",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Config : Factor w/ 12 levels "EB-D","EB-S",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Category : chr "hw" "hw" "hw" "hw" ...
```

2 Overview

2.1 Summary

Terminology

- A *run* is a single execution of the algorithm *configuration* on an input *model*.
 - A run is *successful* if it terminates within the time limit.
- A *measurement* is the collection of all repeated runs of a configuration on a model.
 - A measurement is *successful* if it includes at least one successful run. In this case we also say that the configuration *verified* the model.

Results

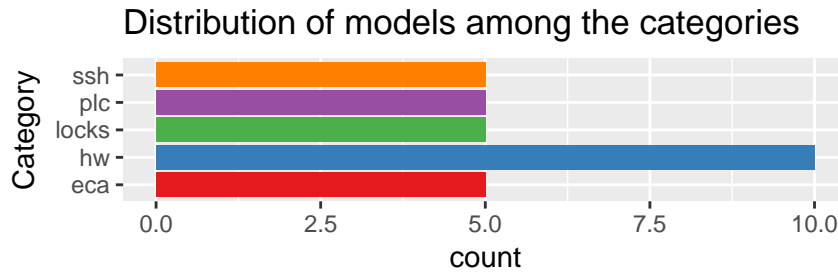
- There are **30 models** (from 5 categories) and **12 configurations**, giving **360 measurements**.
 - Each measurement was **repeated 30 times**, giving a total number of **10800 runs**.
- **7080 / 10800 runs (66%)** are **successful** with a timeout of **timeout of 180 s**.
 - **261 / 360 measurements (72%)** are **successful**.
 - There are **389 problems** (see details later).
- **28 / 30 models (93%)** were **verified** by at least one configuration.

2.2 Problems

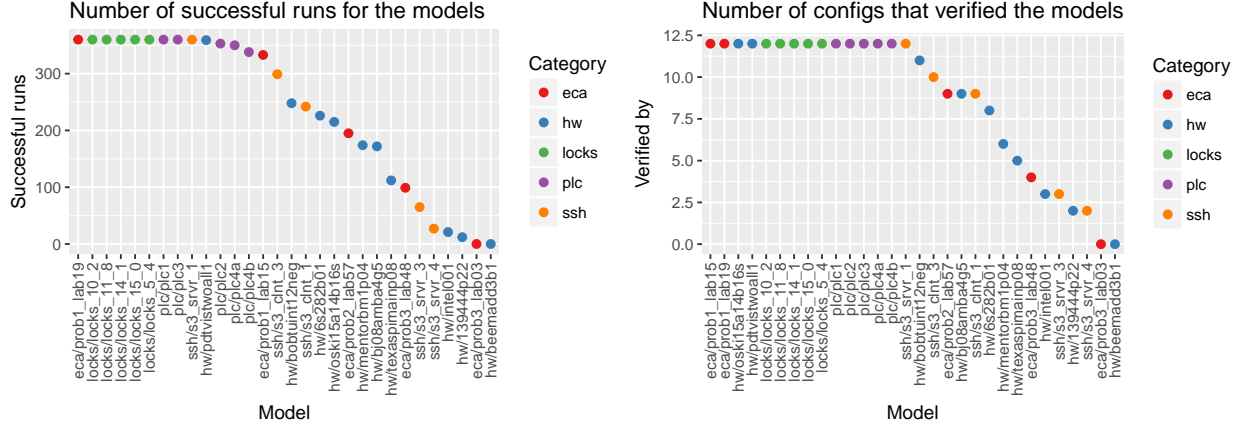
- There are **389 empty** results (no timeout, no exception). This usually happens when the SMT solver crashes.
- There are **0 exceptions**.

2.3 Models

The **30 models** are distributed in the **5 categories** in the following way.



The following plots show for each model the number of successful runs and the number of configurations that verified them.



The following table summarizes the attributes of the models. Hardware models can be described by the number of hardware elements, while software and PLC models can be described by the number of control locations and statements.

Model	Inputs	Latches	AndGates
hw/139444p22	244	322	5549
hw/6s282b01	44	637	3185
hw/beemadd3b1	60	56	876
hw/bj08amba4g5	11	33	13585
hw/bobtuint12neg	212	207	1937
hw/intel001	31	23	240
hw/mentorbm1p04	100	2373	17508
hw/oski15a14b16s	1023	3451	33367
hw/pdtvistwoall1	6	31	725
hw/texaspimainp08	14	42	1955

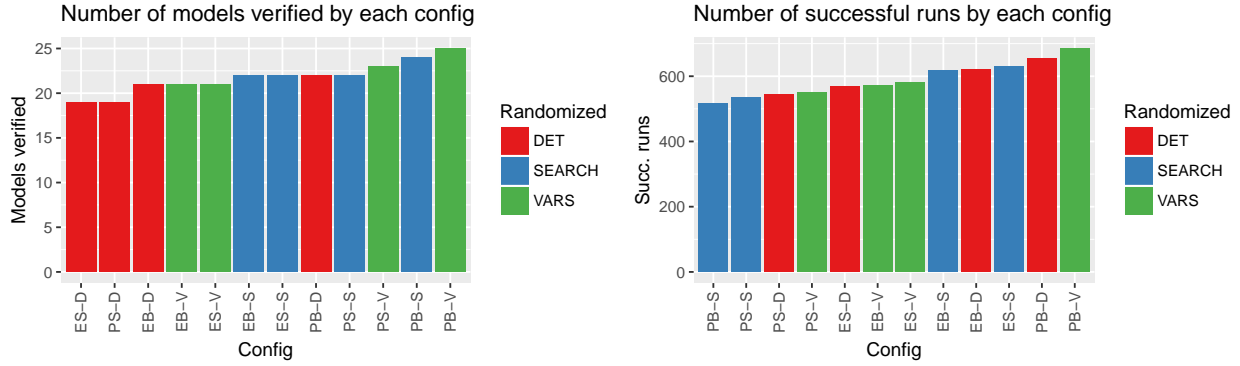
Model	Locations	Edges	Assigns	Assumes	Havocs
eca/prob1_lab15	317	393	236	156	1
eca/prob1_lab19	322	403	236	166	1
eca/prob2_lab57	312	408	211	196	1
eca/prob3_lab03	1261	1436	1081	354	1
eca/prob3_lab48	1280	1474	1081	392	1
locks/locks_10_2	26	36	10	24	2
locks/locks_11_8	16	21	5	14	2
locks/locks_14_1	9	10	2	6	2
locks/locks_15_0	9	10	2	6	2
locks/locks_5_4	13	16	4	10	2
plc/plc1	66	70	51	11	8
plc/plc2	175	196	135	40	21
plc/plc3	175	196	135	40	21
plc/plc4a	175	196	135	40	21
plc/plc4b	175	196	135	40	21
ssh/s3_clnt_1	187	262	79	154	29
ssh/s3_clnt_3	193	268	85	154	29
ssh/s3_srvr_1	233	323	102	184	37
ssh/s3_srvr_3	230	320	100	184	36
ssh/s3_srvr_4	230	320	100	184	36

2.4 Configurations

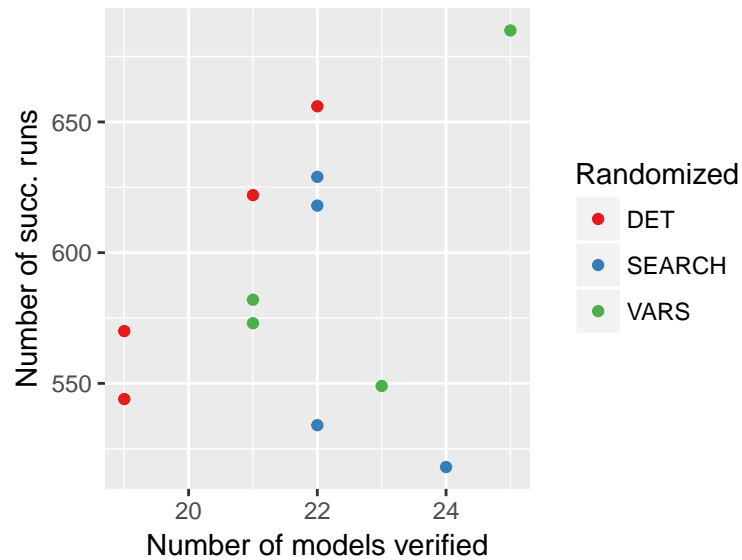
Configurations are listed in the following table with their abbreviation, parameters, successful runs and verified models.

Config	Randomized	Domain	Refinement	SuccRunCount	SuccCount
EB-D	DET	EXPL	BIN	622	21
EB-S	SEARCH	EXPL	BIN	618	22
EB-V	VARs	EXPL	BIN	573	21
ES-D	DET	EXPL	SEQ	570	19
ES-S	SEARCH	EXPL	SEQ	629	22
ES-V	VARs	EXPL	SEQ	582	21
PB-D	DET	PRED	BIN	656	22
PB-S	SEARCH	PRED	BIN	518	24
PB-V	VARs	PRED	BIN	685	25
PS-D	DET	PRED	SEQ	544	19
PS-S	SEARCH	PRED	SEQ	534	22
PS-V	VARs	PRED	SEQ	549	23

The following plots show the number of models verified and the successful runs by each config. The number of models verified ranges between 19 and 25, while the number of successful runs is between 518 and 685.



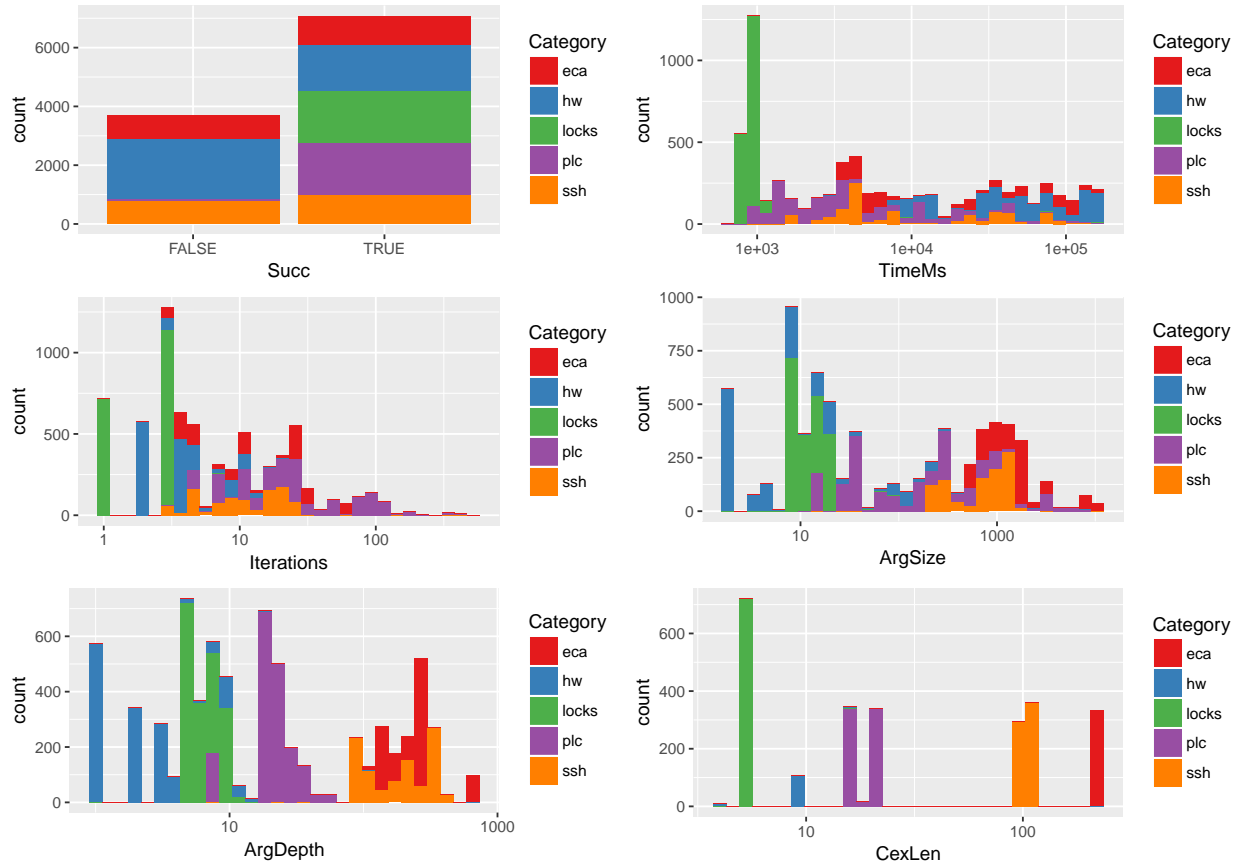
The following plot compares the number of verified models to the number of successful runs for each configuration.



2.5 Output metrics

The following histograms show an overview on the range of the output metrics colored by the category.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The following scatterplot matrix shows the numerical output metrics.

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 4546 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 4546 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 4546 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 4546 rows containing missing values

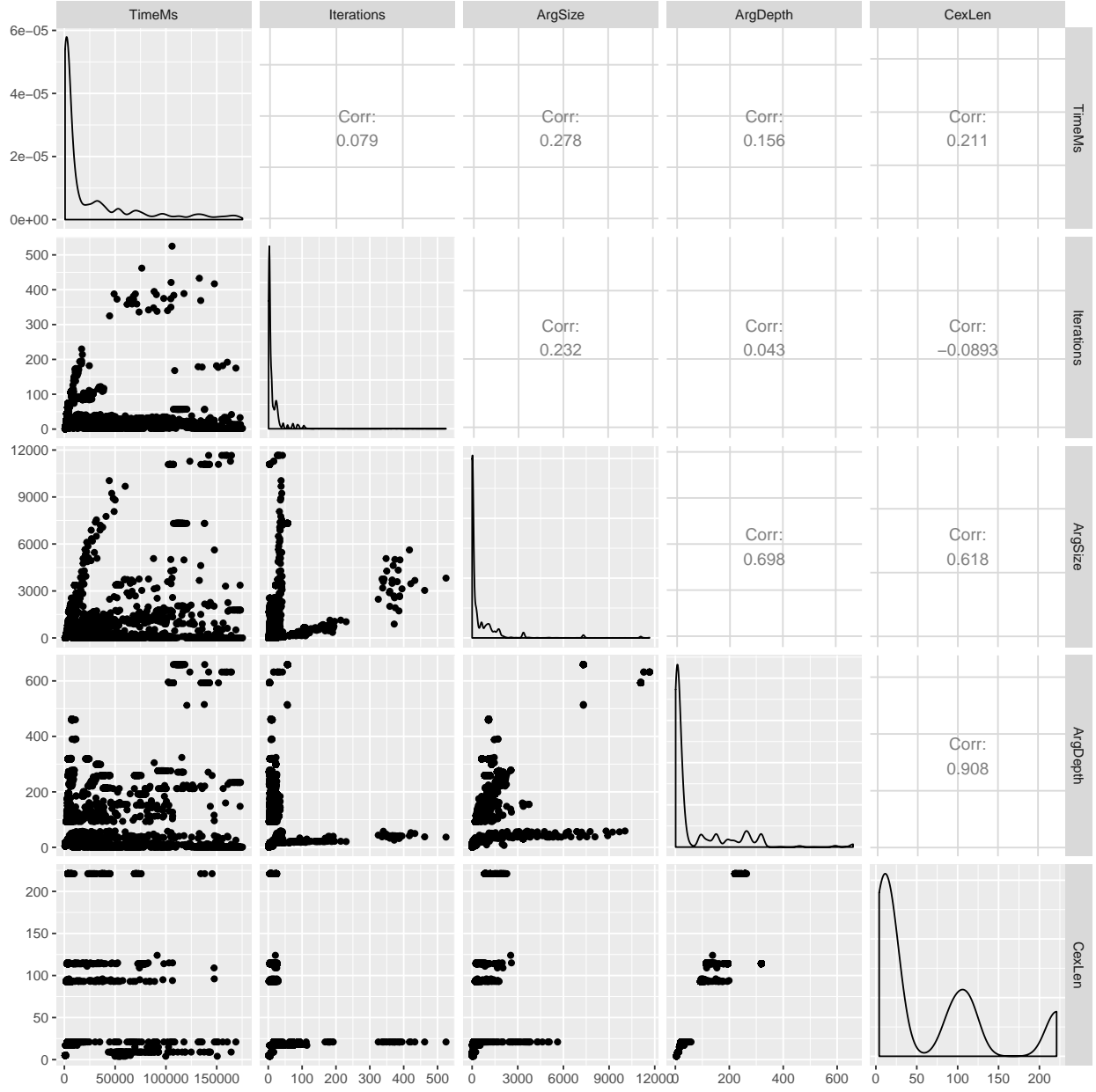
## Warning: Removed 4546 rows containing missing values (geom_point).

## Warning: Removed 4546 rows containing missing values (geom_point).

## Warning: Removed 4546 rows containing missing values (geom_point).

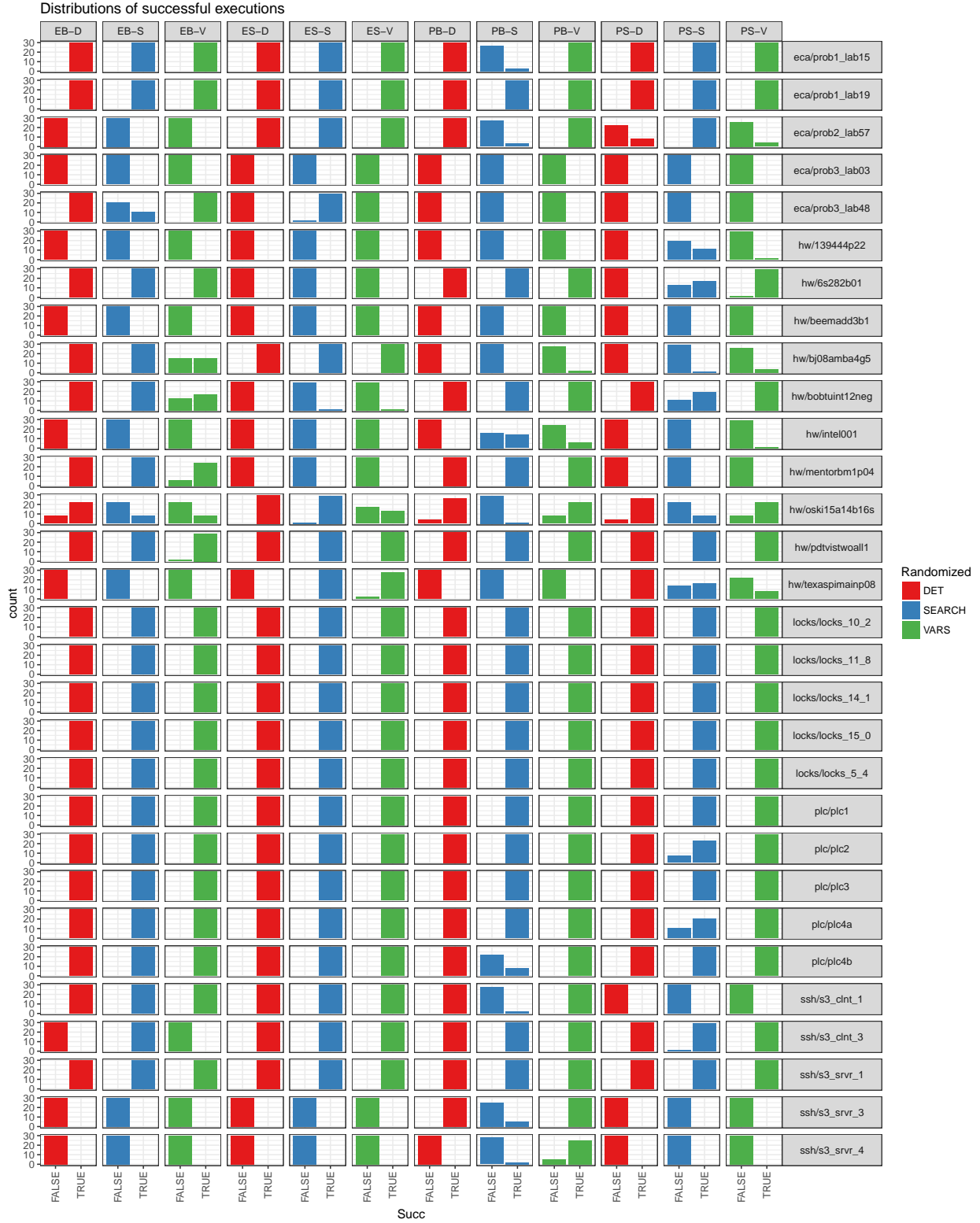
## Warning: Removed 4546 rows containing missing values (geom_point).

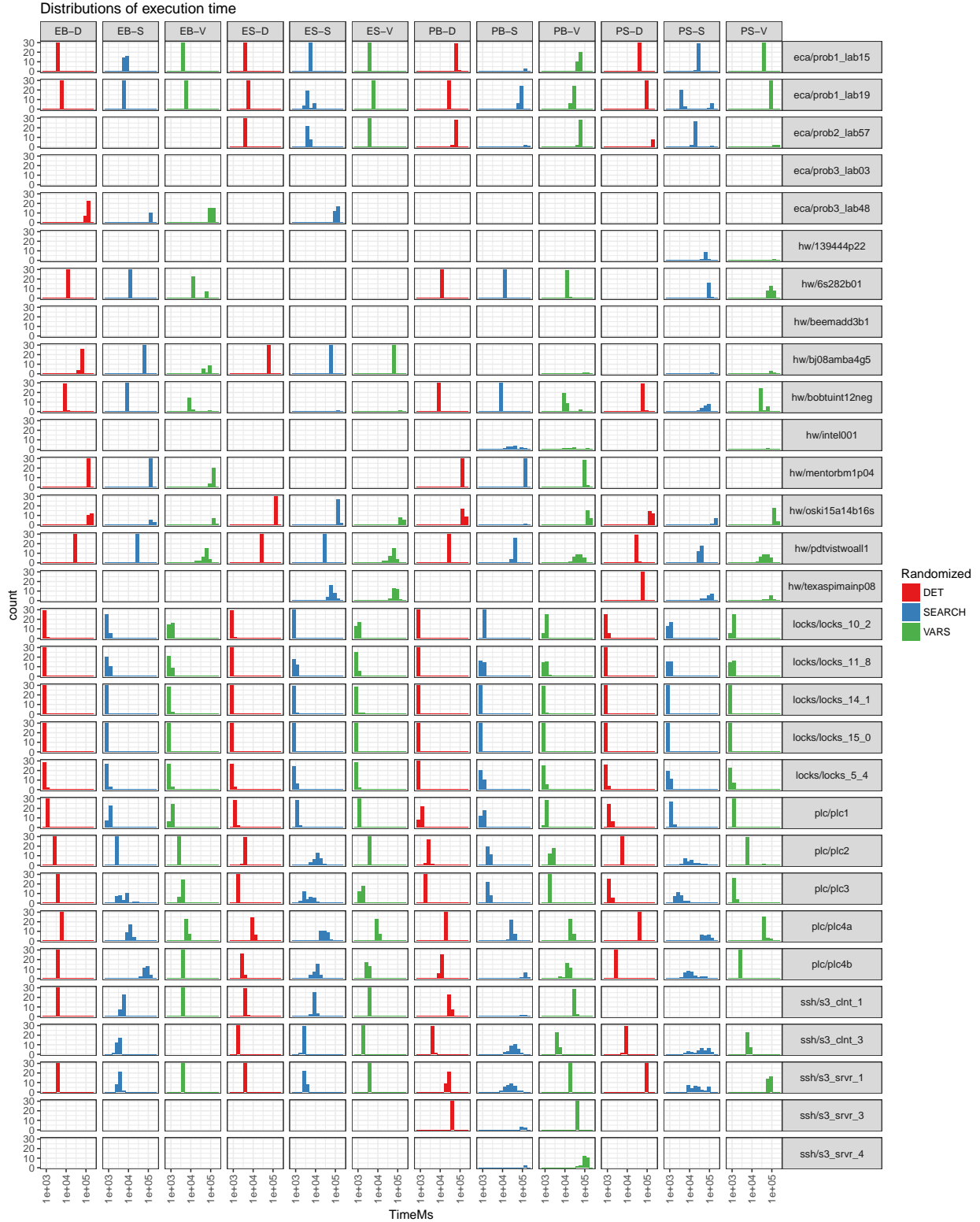
## Warning: Removed 4546 rows containing non-finite values (stat_density).
```

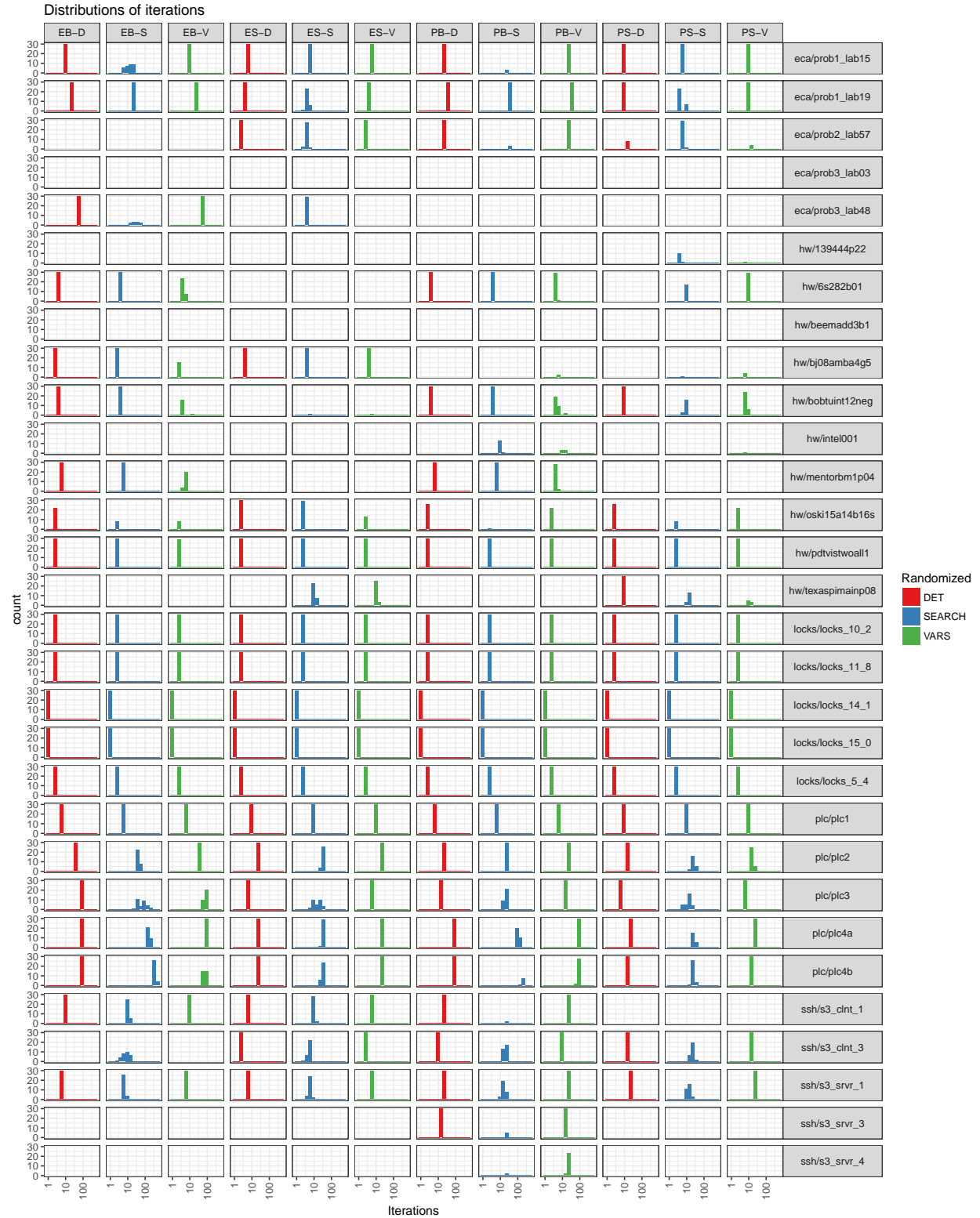


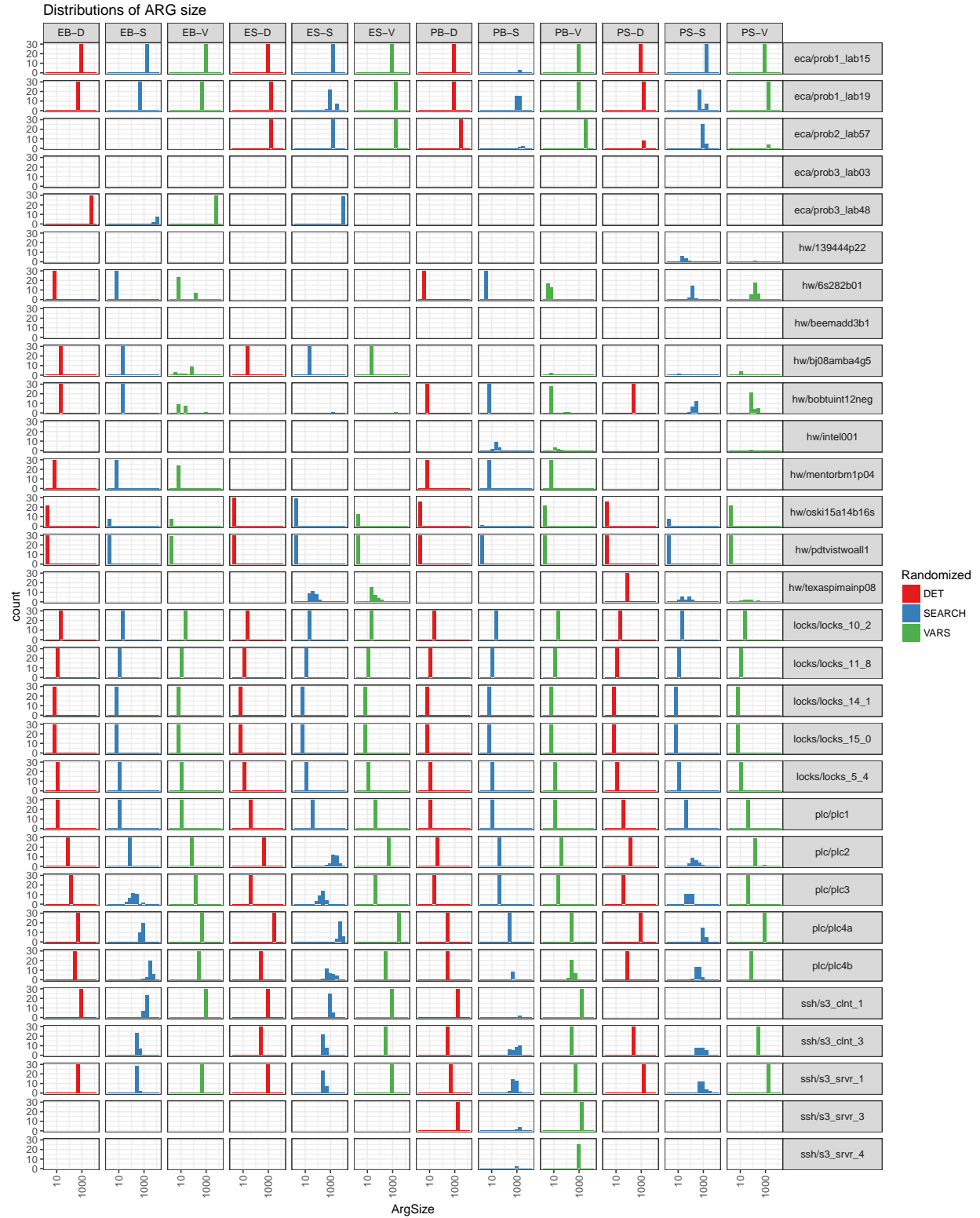
3 Success rates (RQ1)

The following heatmaps show for each measurement whether it was successful (at least one successful run) and the number of successful executions (if there were any). In the latter plot, those measurements are marked (with circles for randomized search and triangles for randomized variable names) where a randomized configuration verified a model that its deterministic counterpart could not.

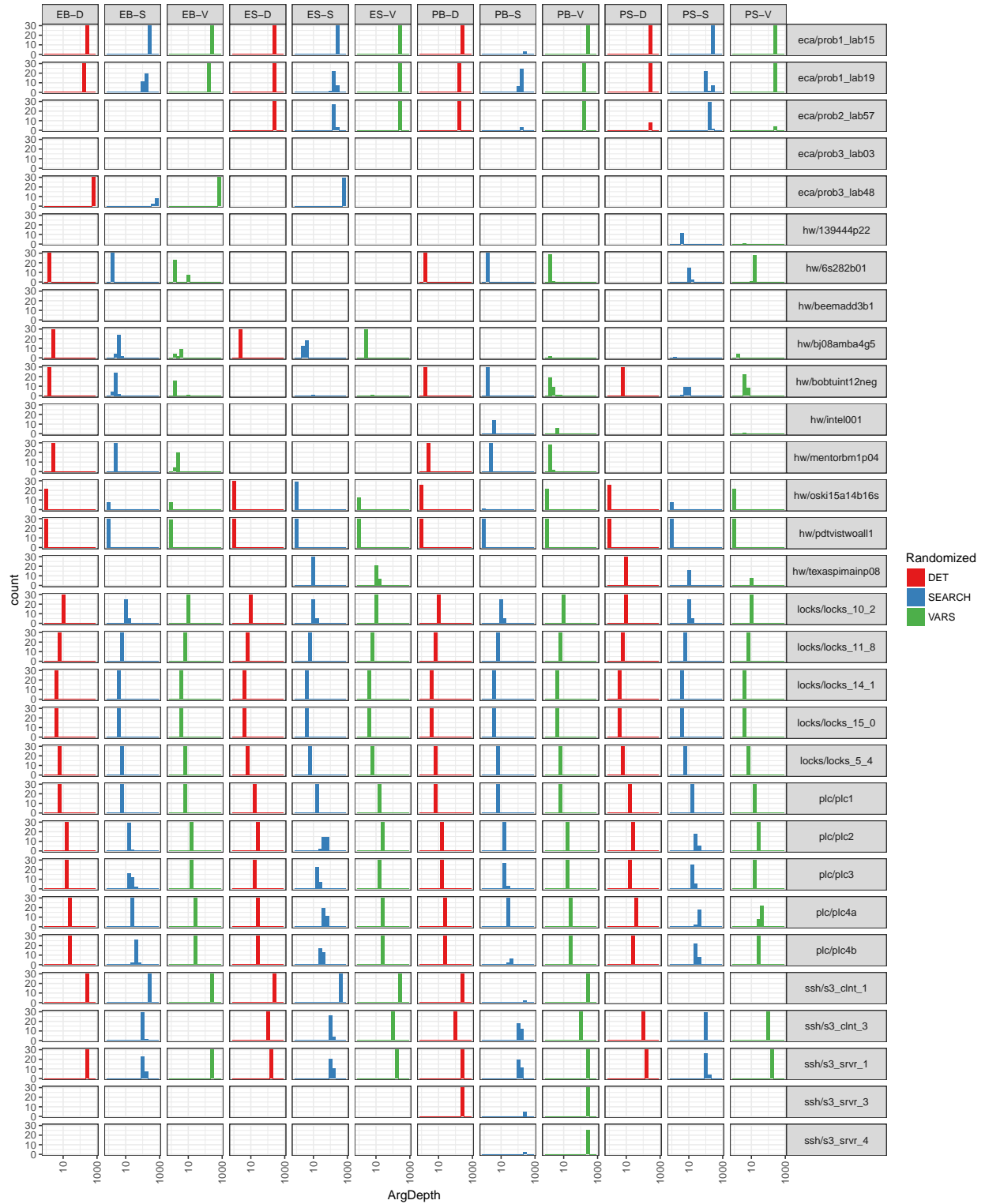




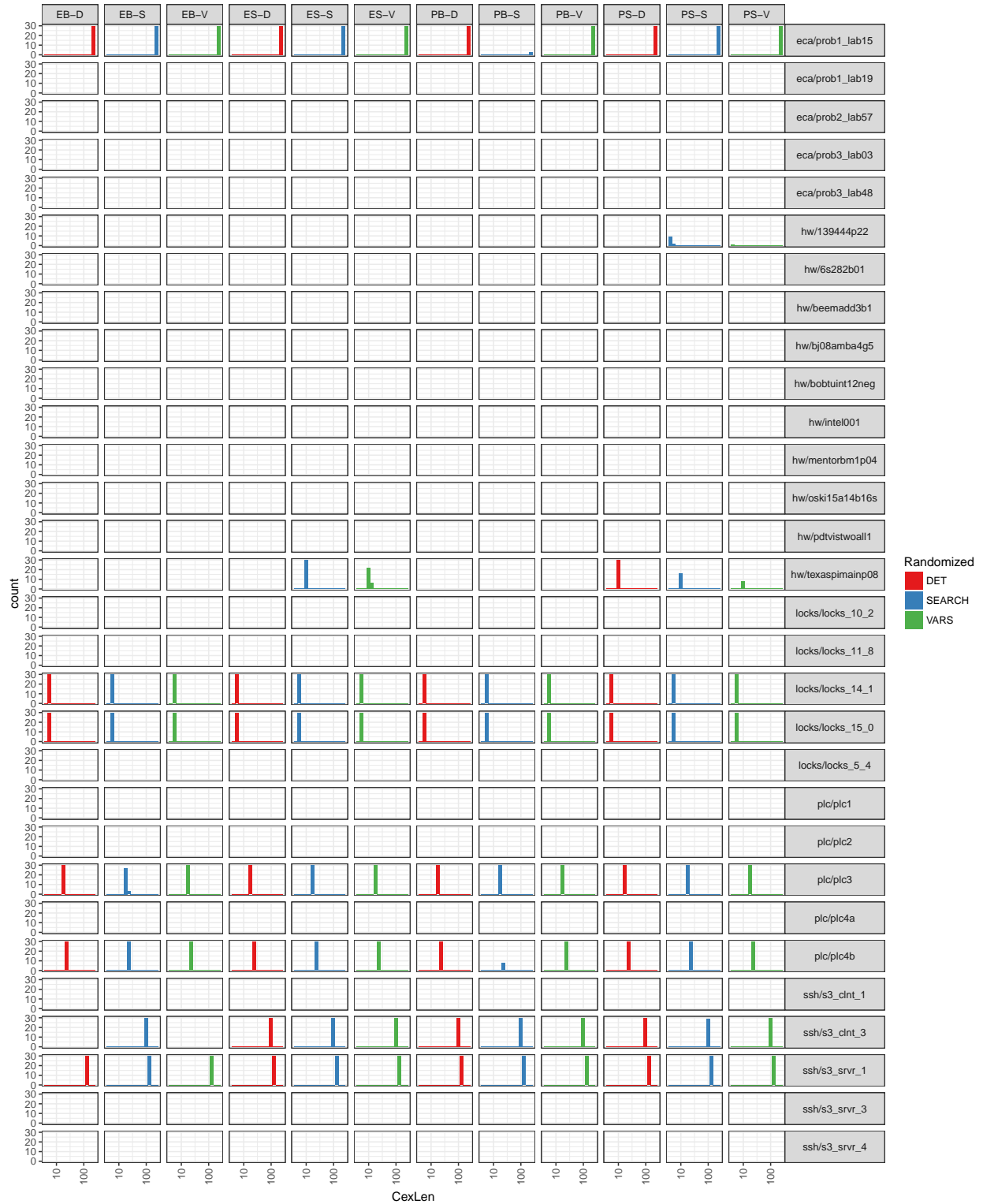




Distributions of ARG depth

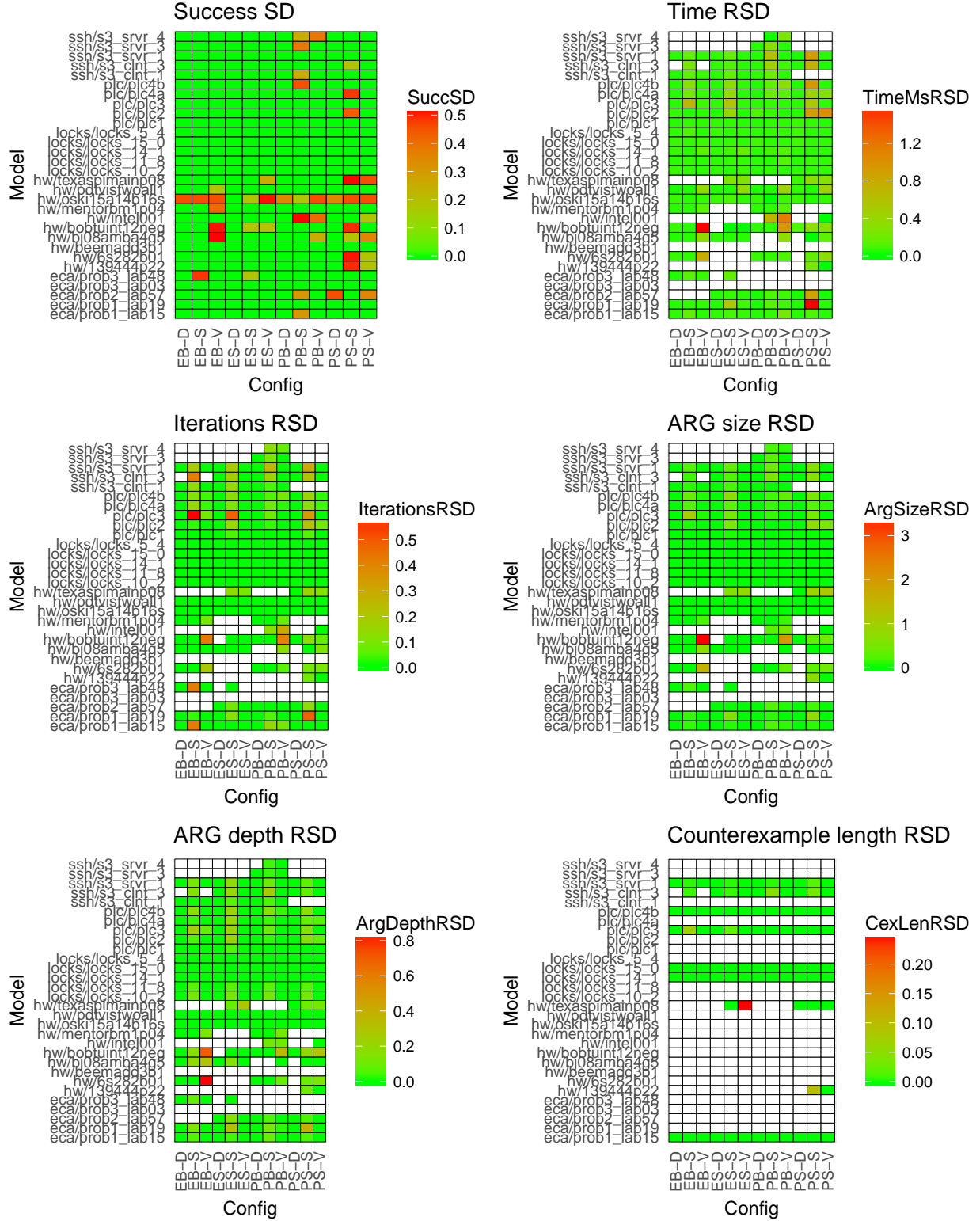


Distributions of cex. length



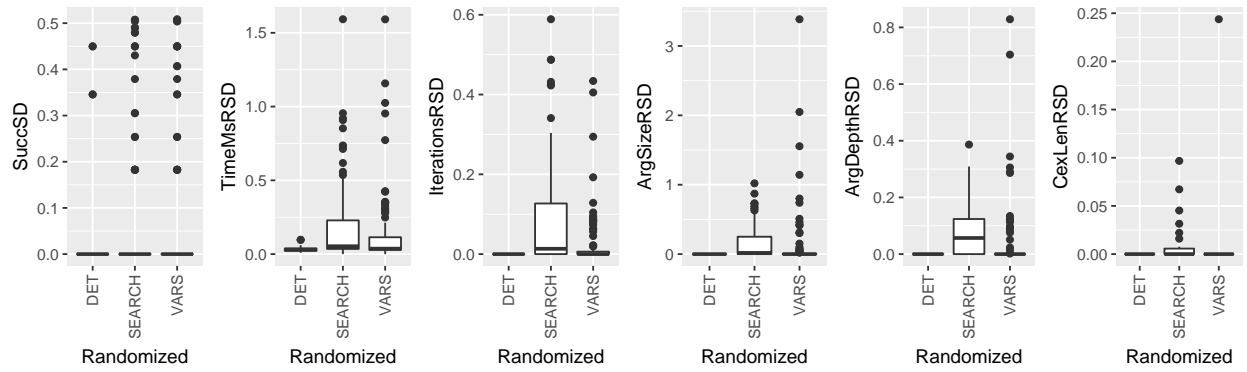
4.2 Individual deviations

In this research question we are mainly interested in the **deviations** of the distributions. The following heatmaps summarize the deviations for each measurement and each output metric. The *SuccSD* variable shows the *standard deviation* of the success of the verification (0: false, 1: true). Other metrics correspond to the *relative standard deviation* ($RSD = SD / \text{mean}$) of the output metrics.

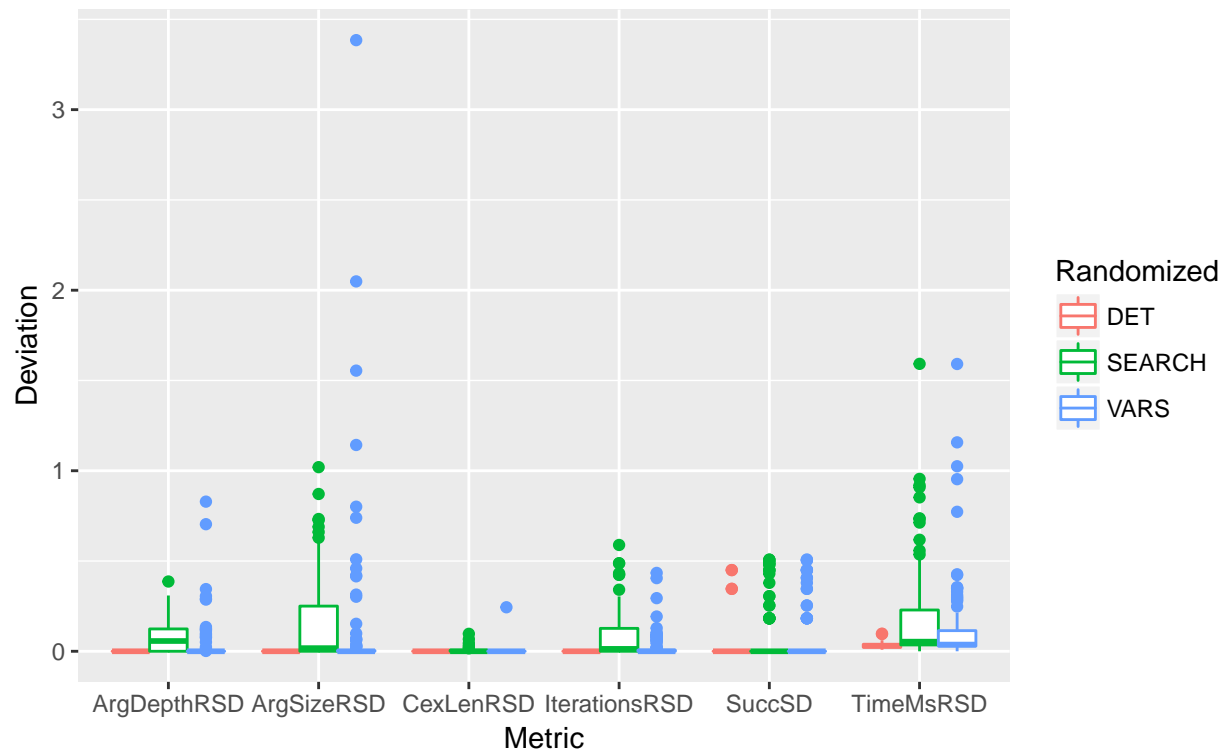


4.3 Summarized deviations

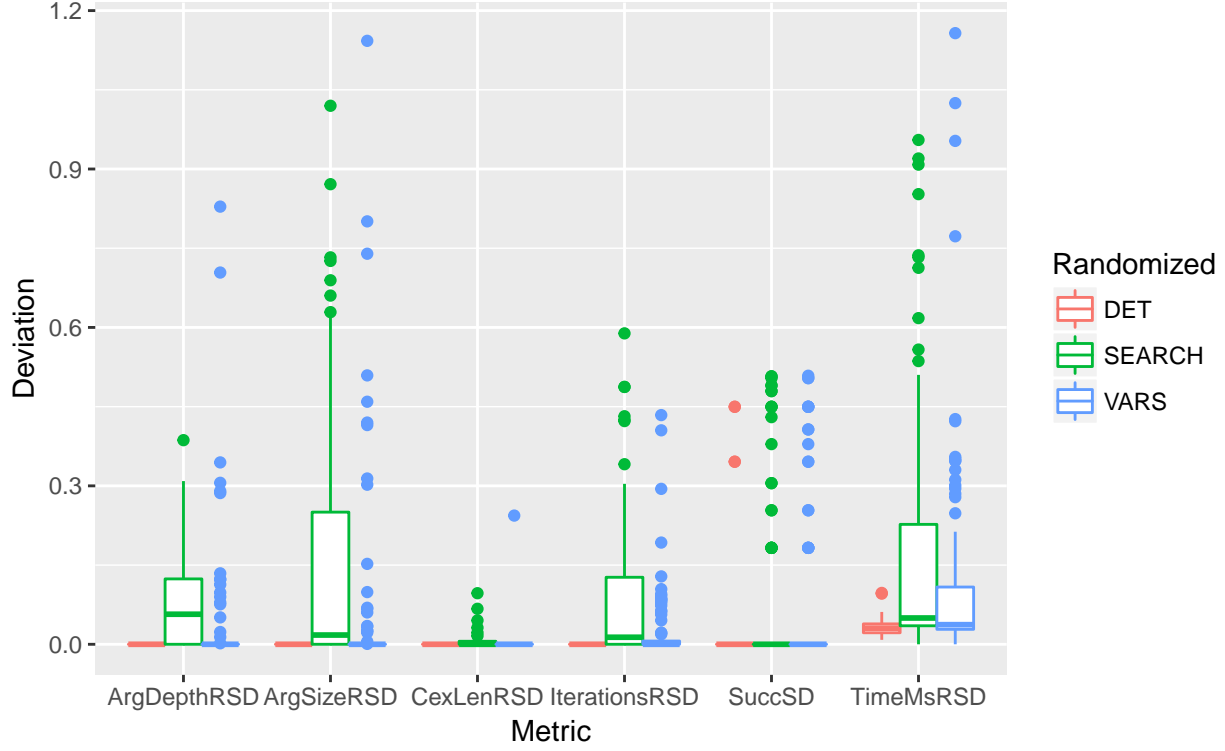
The following plots summarize the distribution of the deviations, grouped by the parameter that is randomized.



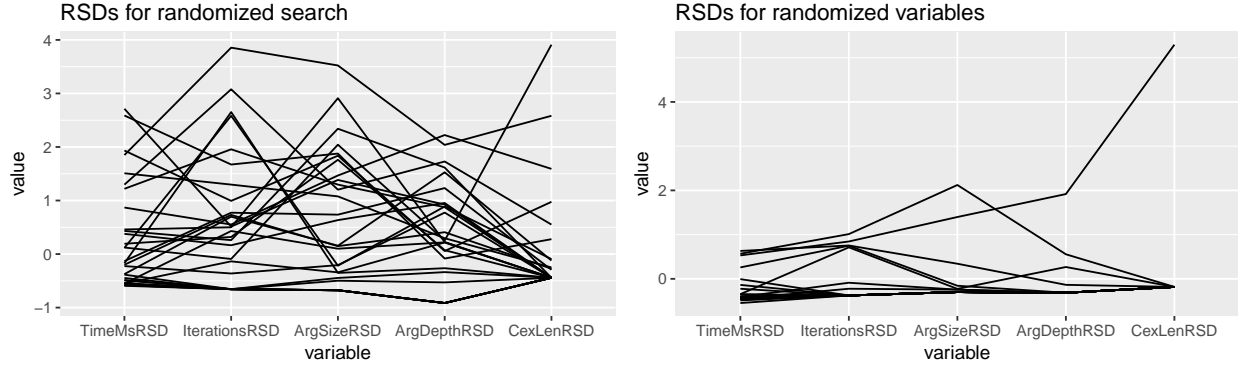
The following plot shows the same data, but using a uniform y axis scale.



The following plot shows the same data, but outliers above 1.2 are trimmed (there are 5 of them).



The following parallel coordinate plots show the deviations of each output metric, where each line corresponds to a measurement.



The following table shows the mean values of the deviations for each output metric grouped by the parameter that is randomized.

Table 4: Mean value of the deviations

Randomized	Succ	TimeMs	Iterations	ArgSize	ArgDepth	CexLen
DET	0.0132587	0.0326082	0.0000000	0.0000000	0.0000000	0.0000000
SEARCH	0.0652476	0.1948332	0.0856263	0.1641061	0.0757643	0.0098402
VARS	0.0551681	0.1447746	0.0282938	0.1409662	0.0409520	0.0081250

The following table shows the mean values of the deviations for each output metric for each configuration.

Table 5: Mean value of the deviations

Config	Succ	TimeMs	Iterations	ArgSize	ArgDepth	CexLen
EB-D	0.0149925	0.0318736	0.0000000	0.0000000	0.0000000	0.0000000
EB-S	0.0309747	0.1130352	0.1113018	0.1039979	0.0685456	0.0112486
EB-V	0.0683914	0.1843390	0.0394242	0.2805611	0.0975525	0.0000000
ES-D	0.0000000	0.0330502	0.0000000	0.0000000	0.0000000	0.0000000
ES-S	0.0182574	0.1560224	0.0769143	0.1800610	0.1006171	0.0040703
ES-V	0.0313430	0.0641824	0.0043733	0.0381368	0.0138147	0.0304686
PB-D	0.0115249	0.0318854	0.0000000	0.0000000	0.0000000	0.0000000
PB-S	0.0878830	0.1376548	0.0476057	0.0979967	0.0533533	0.0070693
PB-V	0.0496457	0.1849910	0.0470859	0.1368538	0.0277514	0.0000000
PS-D	0.0265174	0.0338148	0.0000000	0.0000000	0.0000000	0.0000000
PS-S	0.1238753	0.3778186	0.1101400	0.2803788	0.0825784	0.0160288
PS-V	0.0712922	0.1385212	0.0195455	0.1118677	0.0283993	0.0000000