A Performance Comparison of General Purpose Multi-Dimensional In-Memory Indexes – All Results

Revision 1.2 – 26th February 2018

Tilmann Zäschke
zoodb@gmx.de
zaeschke@inf.ethz.ch

1. INTRODUCTION

This document contains all TinSpin[1] test results from the test runs between November 2016 and January 2017.

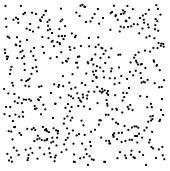
- Revision 1.0 2017-01-28 Initial version.
- Revision 1.1 2017-09-18 Added brief section on data.
- Revision 1.2 2018-02-26 Fixed labels in Fig. 16 and 17.

2. INDEXES

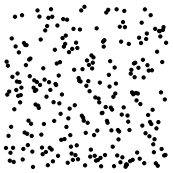
The following index implementations were tested:

- CBF CritBit tree by J. Fager[2]
- CBZ CritBit tree by T. Zäschke[3]
- KDL KD-Tree by Levy[4]
- KDS KD-Tree by Savarese[5]
- PH/PHM PH-Tree by T. Zäschke et al.[6]
- QTZ Quadtree by T. Zäschke[3]
- RSS R*Tree by N. Beckmann et al[7], optimized for in-memory use by T. Zäschke
- RSZ R*Tree by T. Zäschke[3]
- STRZ R*Tree by T. Zäschke[3]
- XTS X-Tree by S. Berchtold et al[7], optimized for in-memory use by T. Zäschke

[1] http://www.tinspin.org
[2] https://github.com/jfager/functional-critbit
[3] http://www.tinspin.org
[4] http://home.wlu.edu/~levys/software/kd/
[5] https://www.savarese.com/software/libssrckdtree-j/
[6] http://www.phtree.org
[7] http://chorochronos.datastories.org



(a) 2D CUBE



(b) 2D CLUSTER

(c) OSM Alps

Figure 1: The CUBE, CLUSTER and OSM Alps datasets

3. TEST DATA

The OSM-P (points) and OSM-R (rectangles) datasets are extracts from OpenStreetMap.org representing the European Alps[8], extracted on 2016-11-09 (Fig. 1c). The dataset consists of $\approx 2.1 \times 10^8$ points. The rectangles (OSM-R) are bounding boxes for all line segments in the dataset.

The synthetic CU-P/CU-R datasets (Fig. 1a), have the shape of a cube filled with up to 50,000,000 elements that are distributed uniformly at random between 0.0 and 1.0 in every dimension. Each element has unique coordinates.

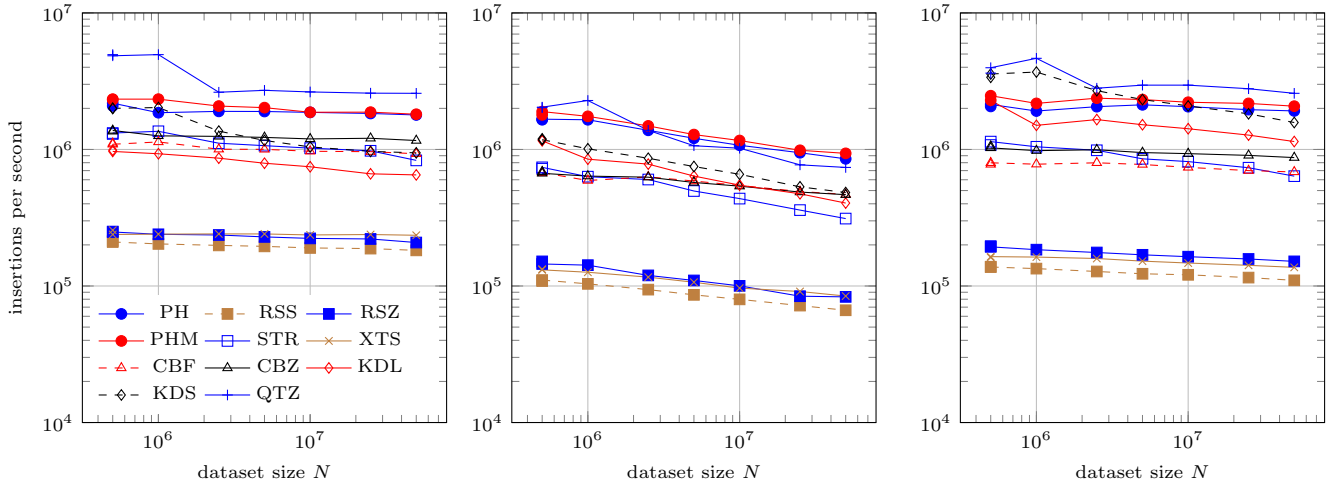
The synthetic CL-P/CL-R datasets (Fig. 1b) consists of 1000 clusters that are distributed uniformly at random between 0.0 and 1.0. In each cluster, elements follow a Gaussian distribution with standard deviation $\sigma = 0.001$. The CLUSTER dataset contains up to 50,000,000 elements.

4. RESULTS

Results are shown on in the following order:

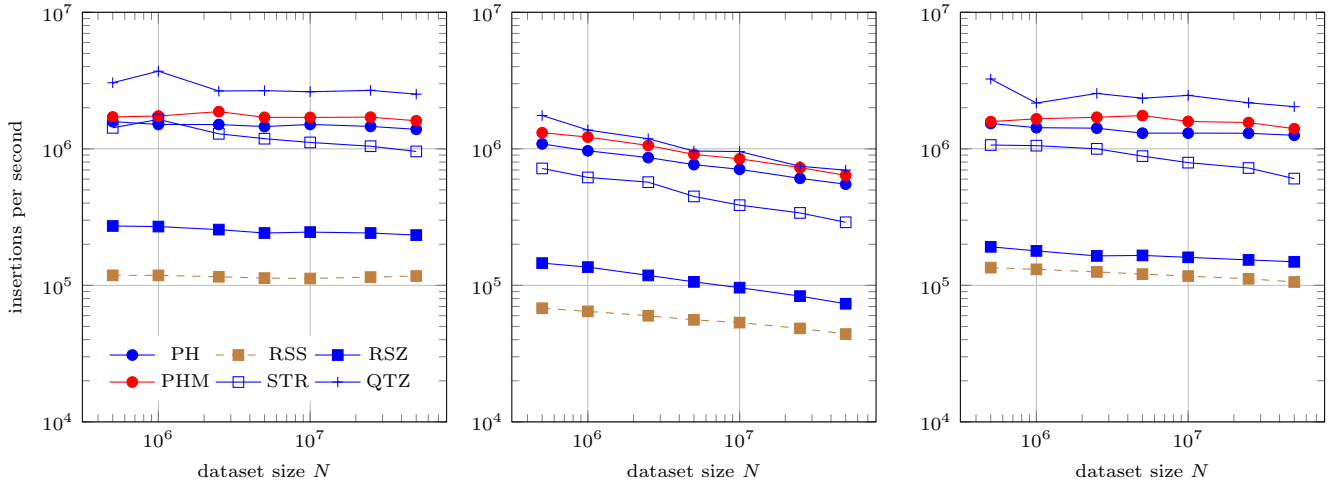
- Insertion: Figures 2 – 7
- Memory usage: Figures 8 – 13
- Window queries: Figures 14 – 21
- Exact match queries (point queries): Figures 22 – 27
- kNN queries: Figures 28 – 39
- Update: Figures 40 – 45
- Remove: Figures 46 – 51

[8] http://download.geofabrik.de/europe/alps.html



(a) Alps OSM-P (b) CUBE-P (c) CLUSTER-P

Figure 2: Insertion rates for the OSM, CUBE and CLUSTER datasets



(a) OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 3: Insertion rates for the OSM, CUBE and CLUSTER datasets

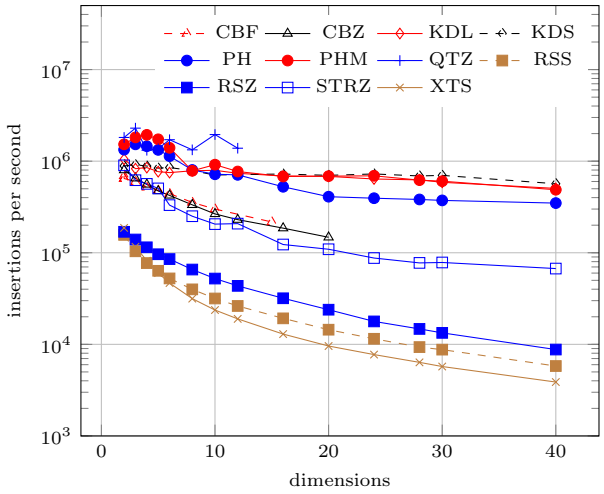


Figure 4: DIM: Insertion rates for CU-P

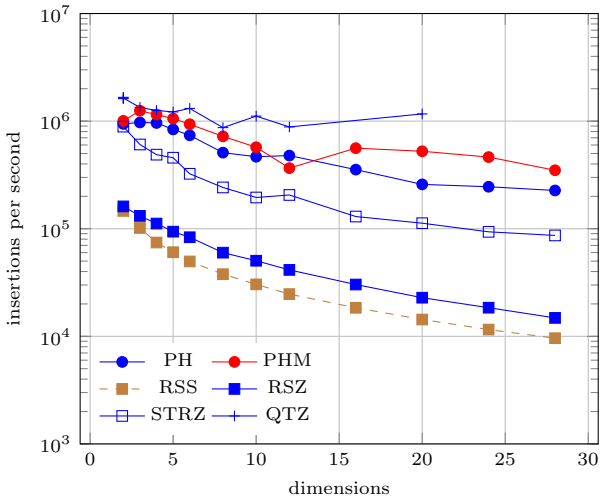


Figure 5: DIM: Insertion rates for CU-R

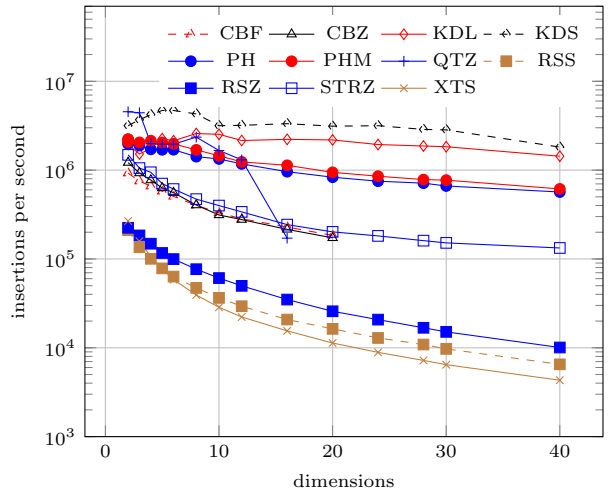


Figure 6: DIM: Insertion rates for CL-P

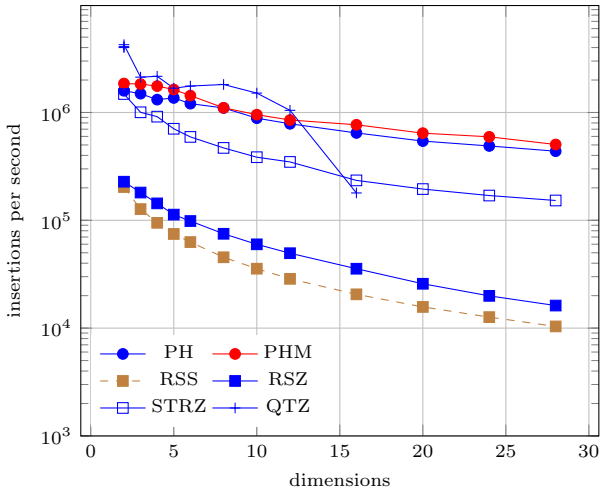
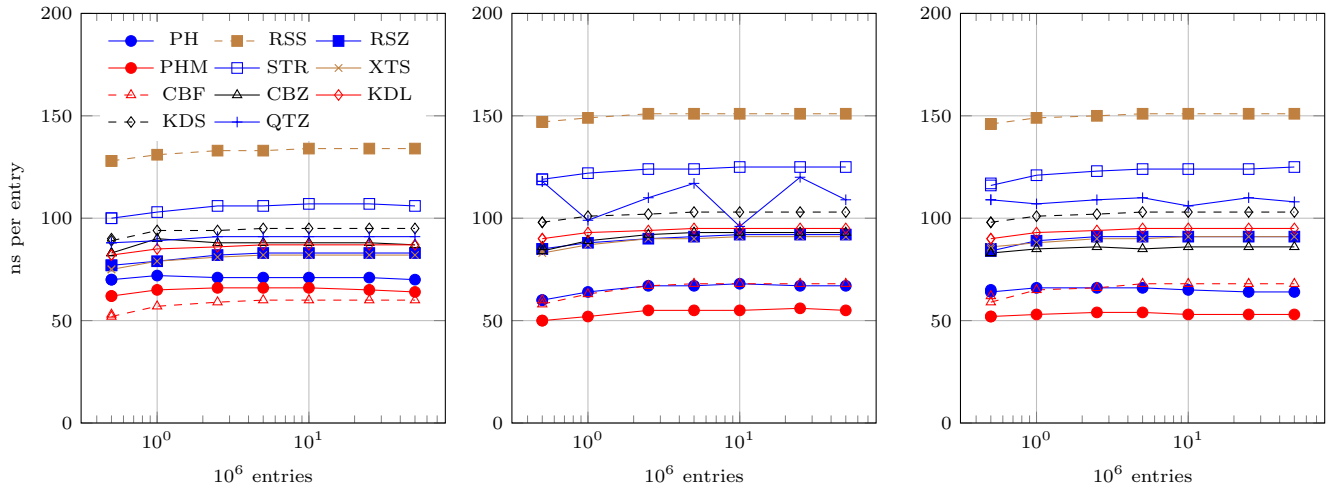
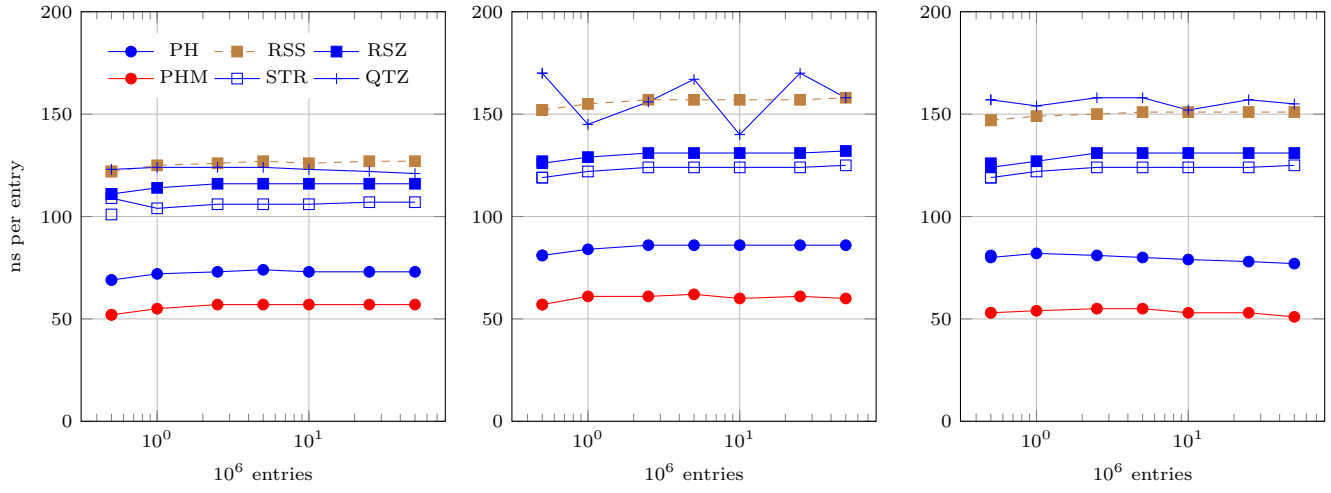


Figure 7: DIM: Insertion rates for CL-R



(a) Alps OSM-P (b) CUBE-P (c) CLUSTER-P

Figure 8: Memory usage per point entry for the OSM Alps, CUBE and CLUSTER datasets



(a) Alps OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 9: Memory usage per rectangle entry for the OSM Alps, CUBE and CLUSTER datasets

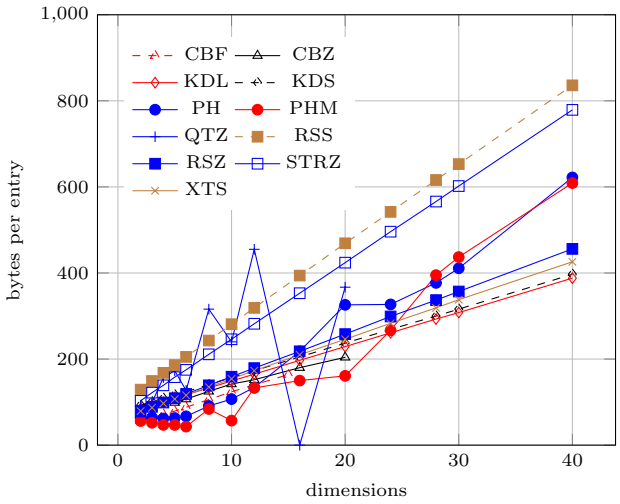


Figure 10: DIM: Memory usage per point for CU-P

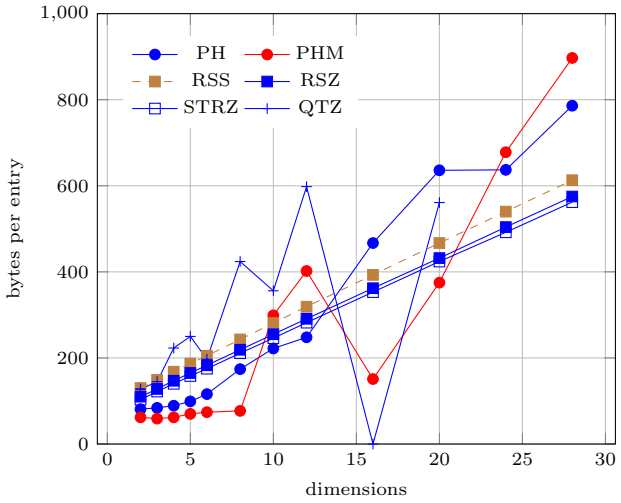


Figure 11: DIM: Memory usage per rectangle for CU-R

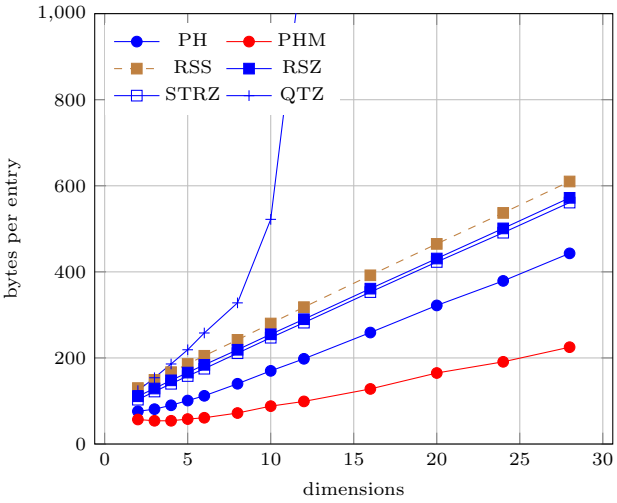


Figure 13: DIM: Memory usage per rectangle for CL-R

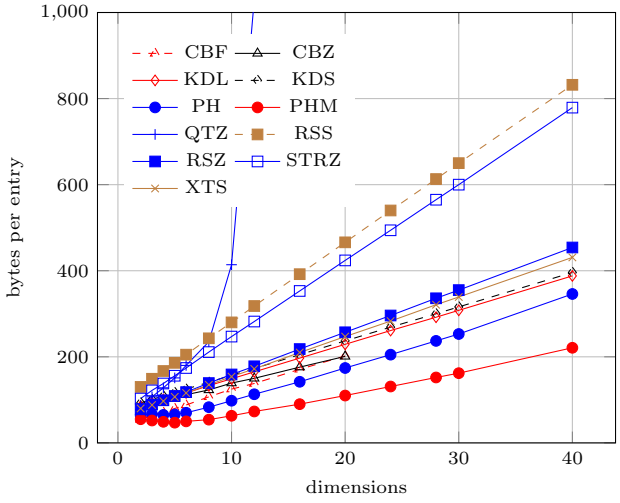
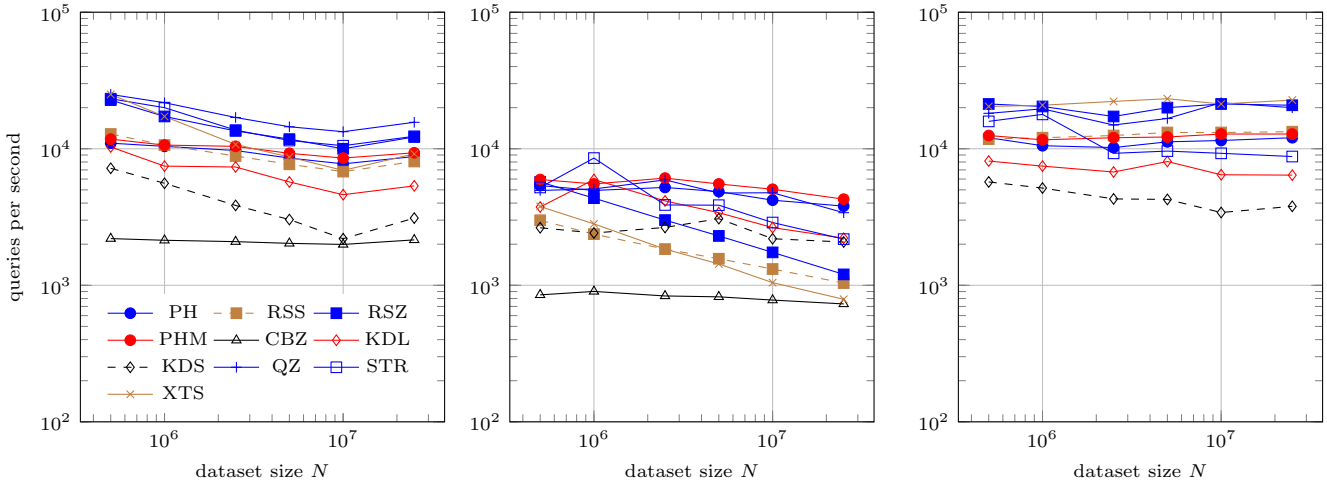
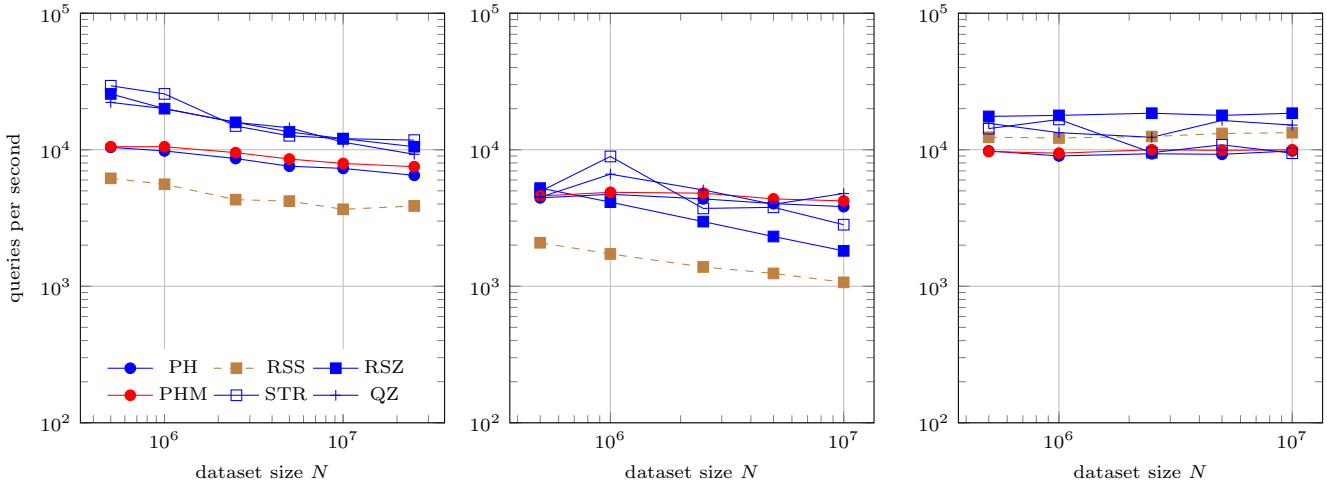


Figure 12: DIM: Memory usage per point for CL-P



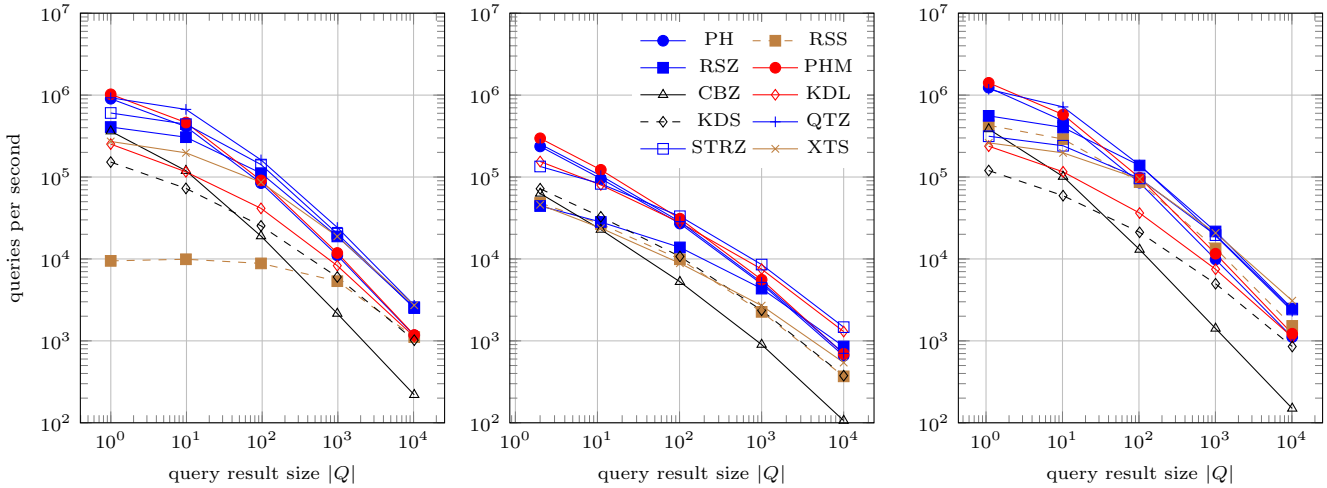
(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 14: Window query rates for the OSM, CUBE and CLUSTER datasets



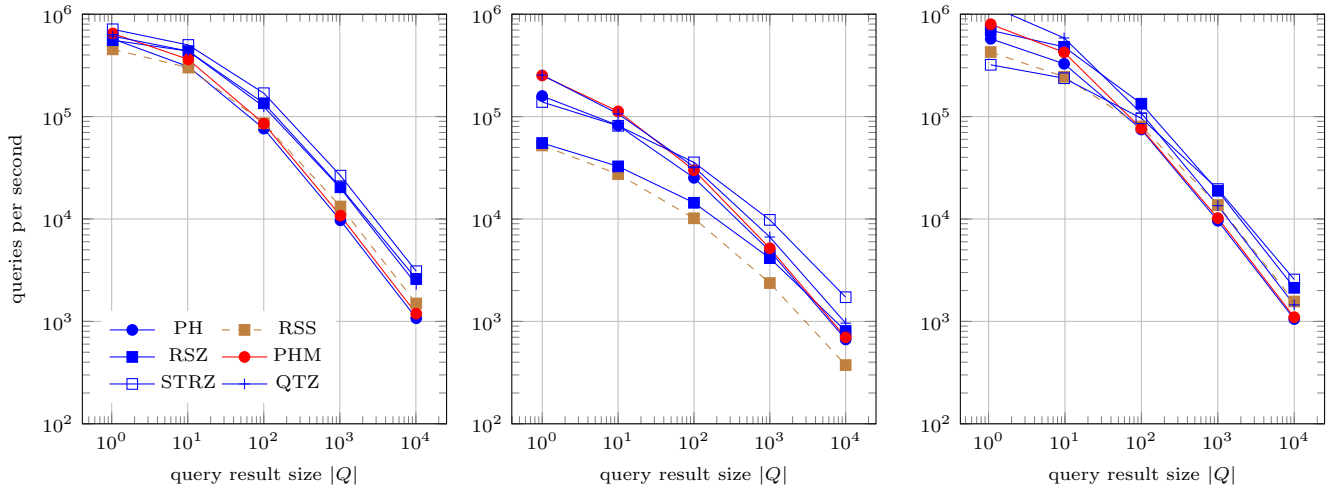
(a) Alps OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 15: Window query rates for the OSM, CUBE and CLUSTER datasets



(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 16: Varying query result size with the OSM Alps, CUBE and CLUSTER datasets



(a) Alps OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 17: Varying query result size with the OSM Alps, CUBE and CLUSTER datasets

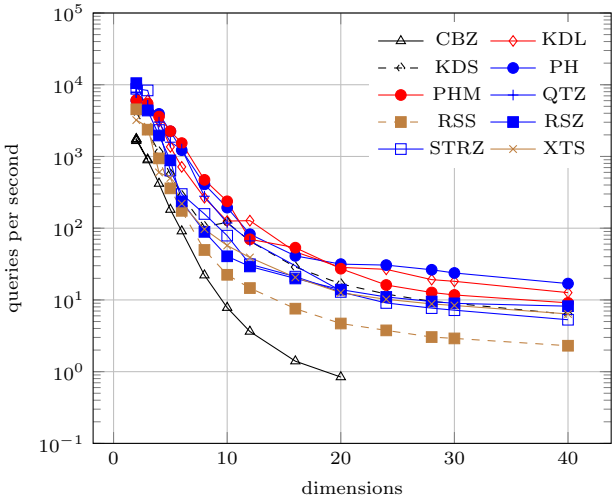


Figure 18: DIM: Window query rates for CU-P

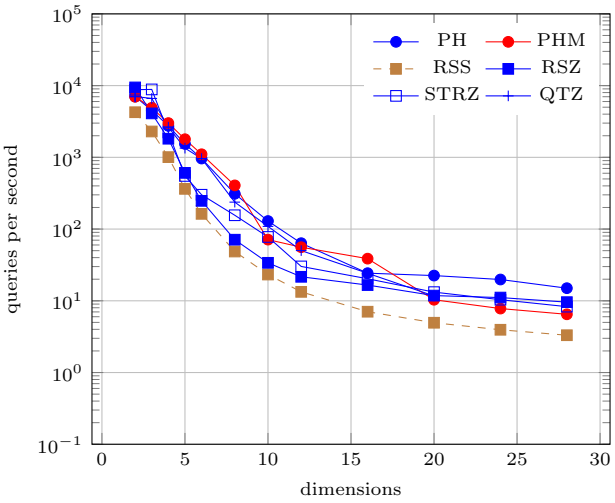


Figure 19: DIM: Window query rates for CU-R

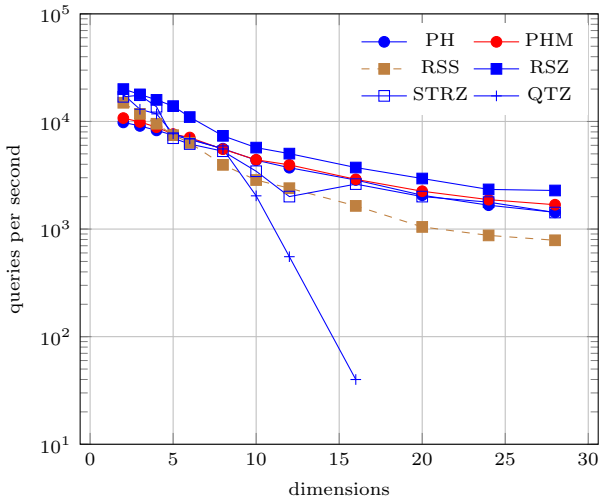


Figure 21: DIM: Window query rates for CL-R

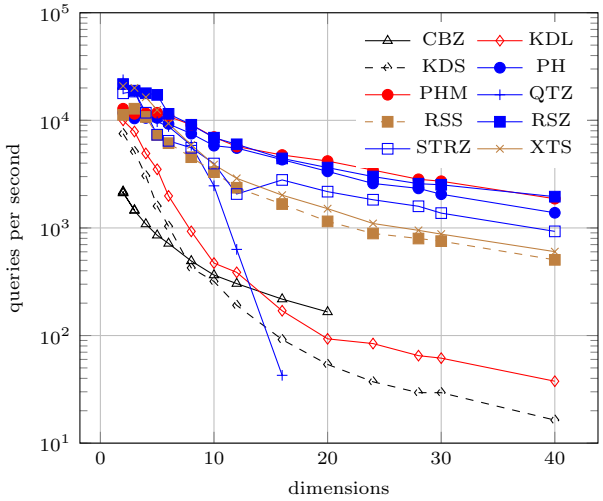
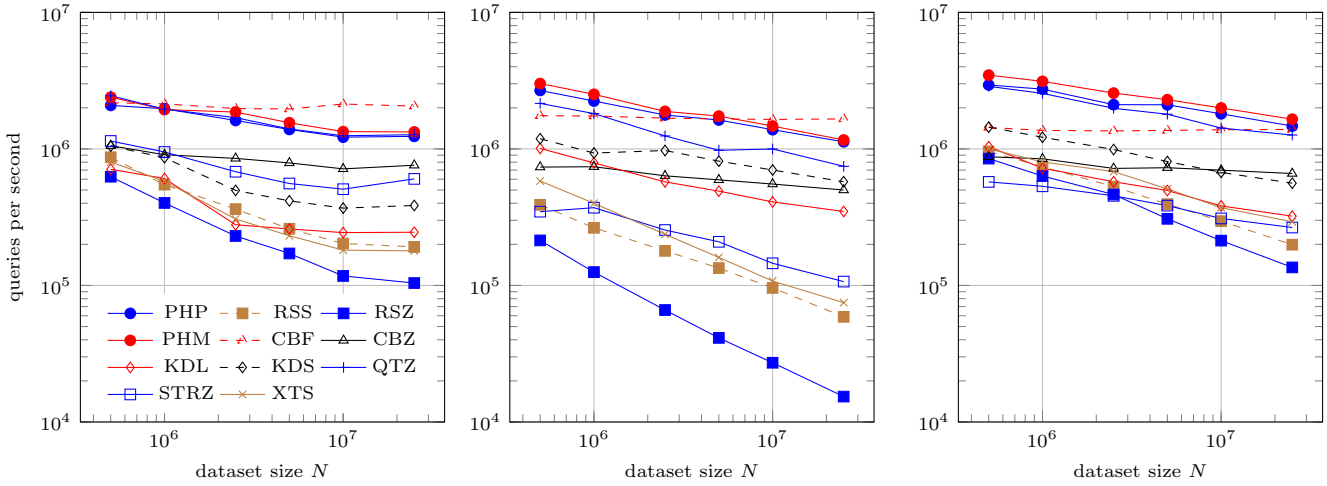
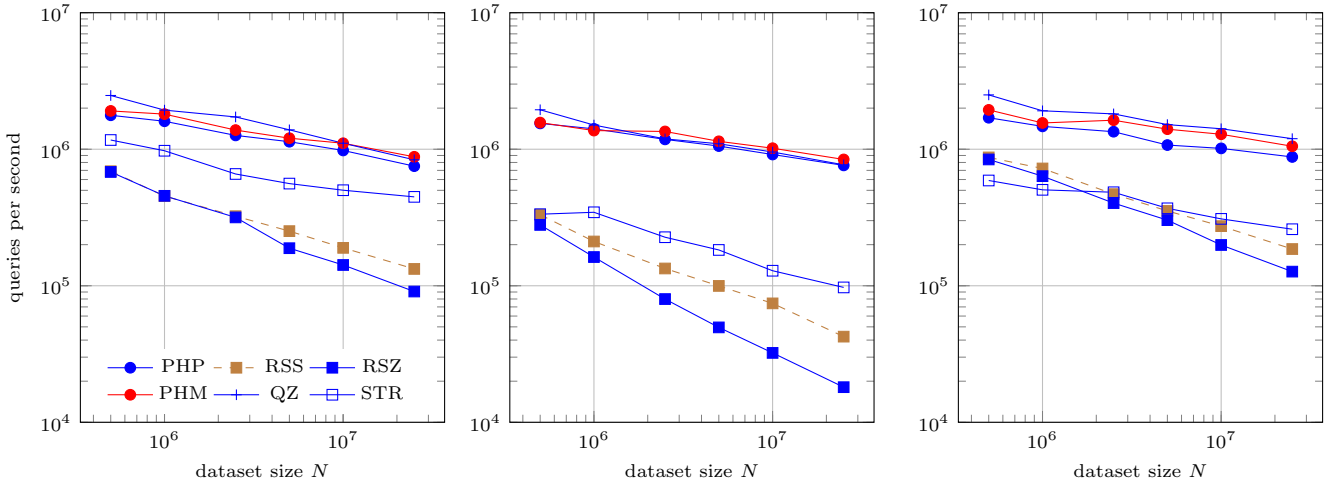


Figure 20: DIM: Window query rates for CL-P



(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 22: Exact match query rates for the OSM, CUBE and CLUSTER datasets



(a) 2D OSM-R (b) 3D CUBE-R (c) 3D CLUSTER-R

Figure 23: Exact match query rates for the OSM, CUBE and CLUSTER datasets

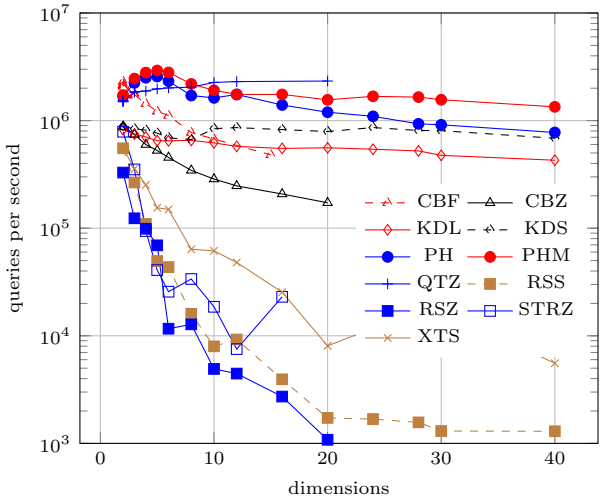


Figure 24: DIM: Exact match query rates for CU-P

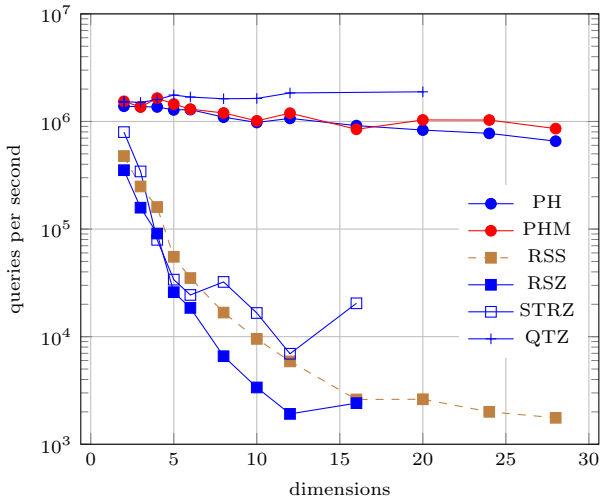


Figure 25: DIM: Exact match query rates for CU-R

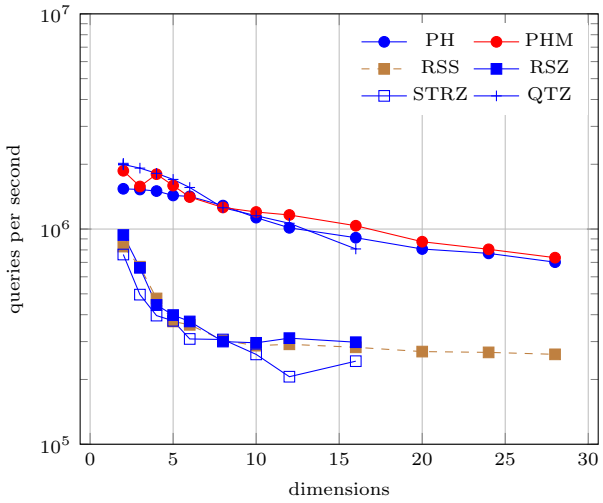


Figure 27: DIM: Exact match query rates for CL-R

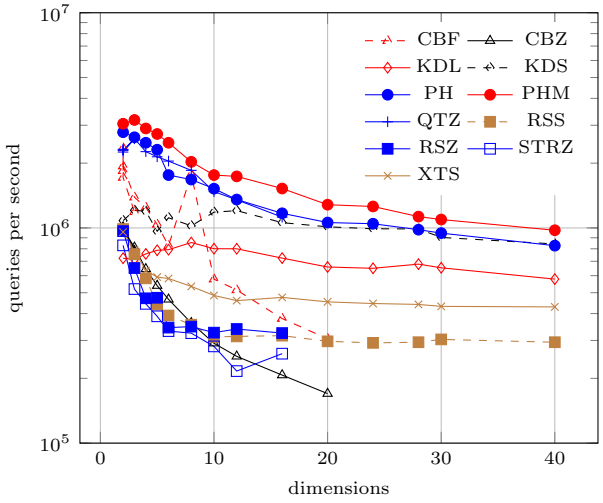
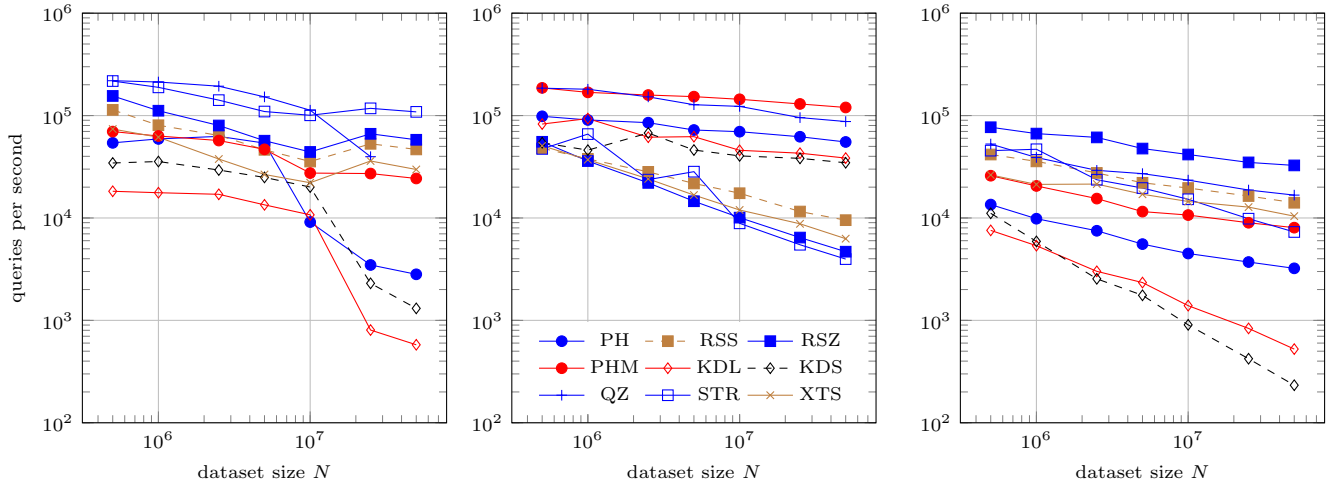
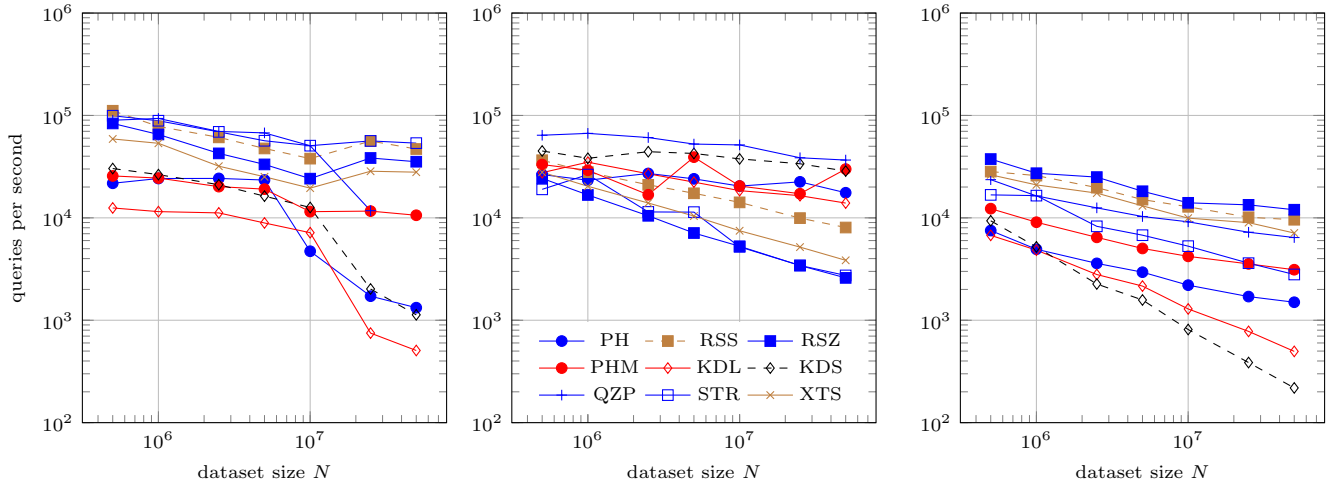


Figure 26: DIM: Exact match query rates for CL-P



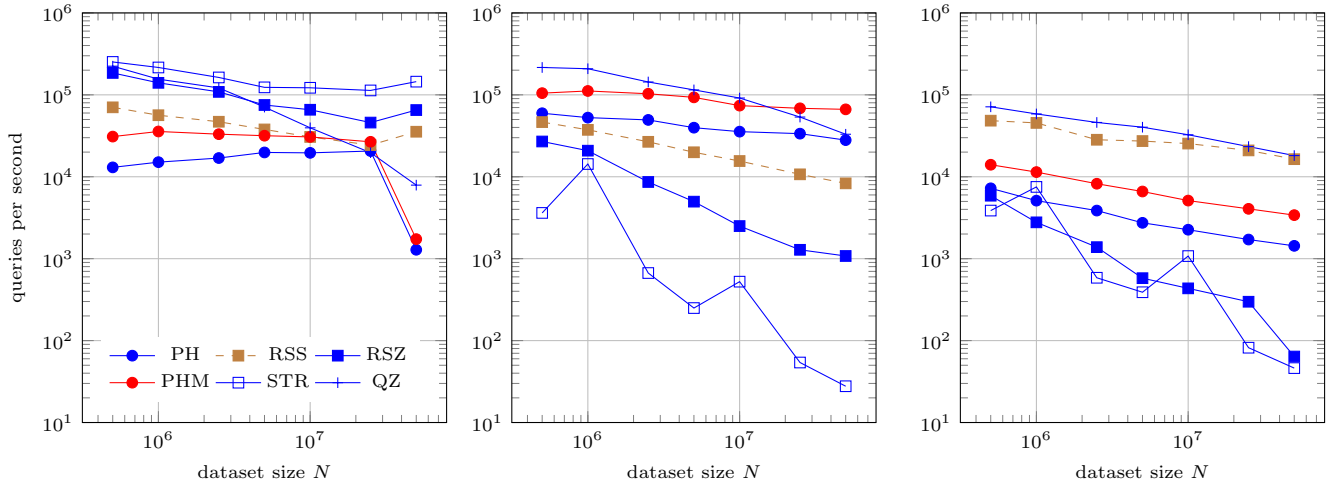
(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 28: 1NN query rates for the OSM, CUBE and CLUSTER datasets



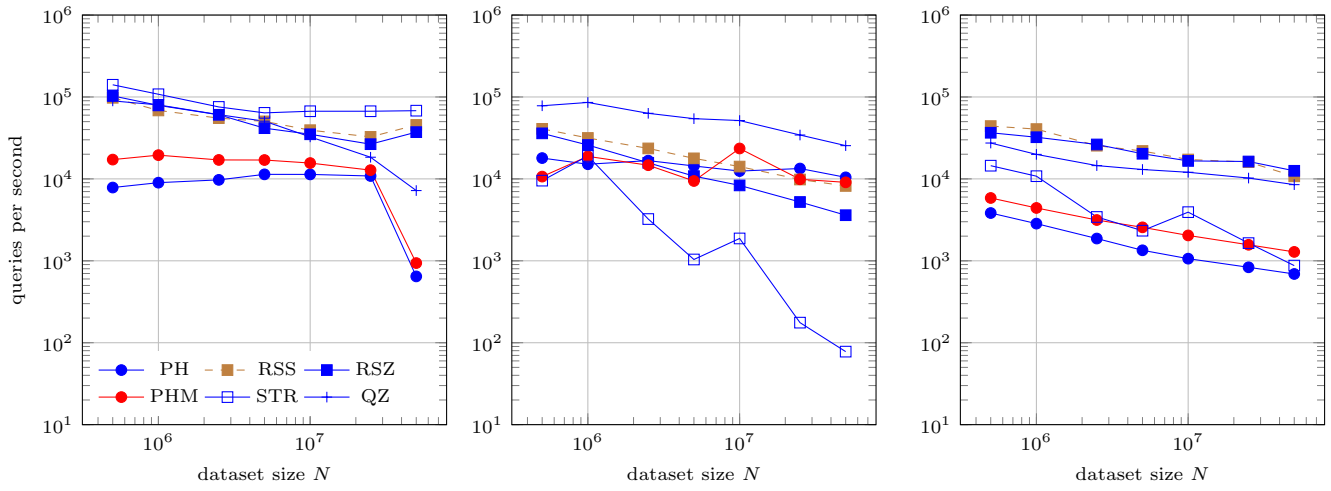
(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 29: 10NN query rates for the OSM, CUBE and CLUSTER datasets



(a) OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 30: 1NN query rates for the OSM, CUBE and CLUSTER datasets



(a) OSM-R (b) CUBE-R (c) CLUSTER-R

Figure 31: 10NN query rates for the OSM, CUBE and CLUSTER datasets

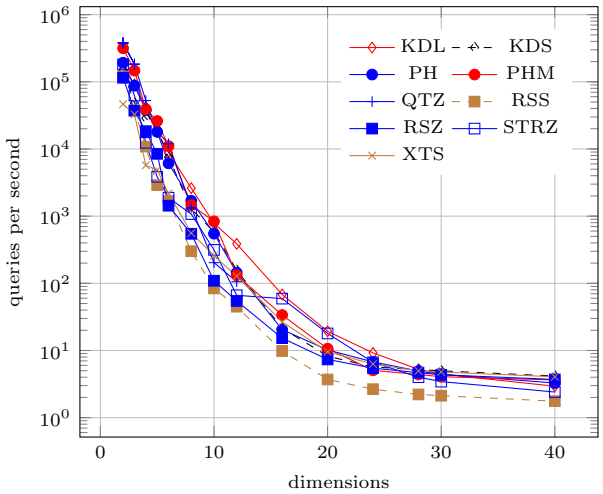


Figure 32: DIM: 1-NN query rates for CU-P

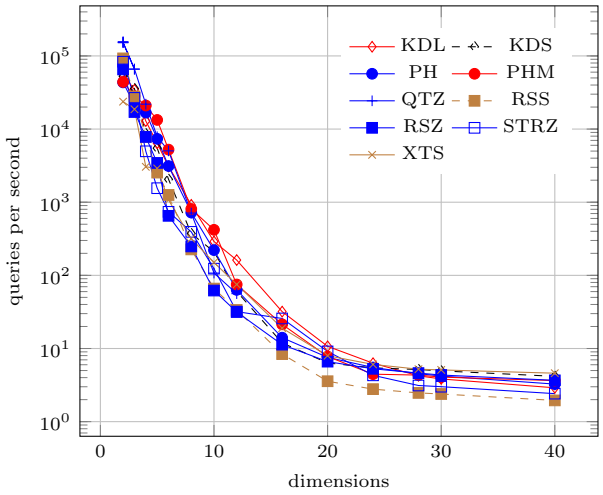


Figure 33: DIM: 10-NN query rates for CU-P

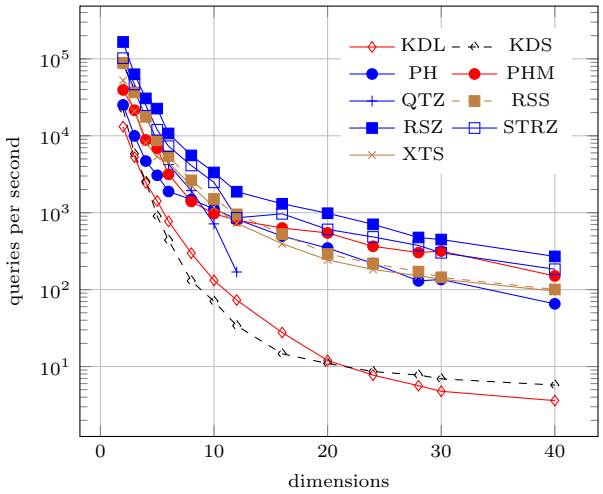


Figure 34: DIM: 1-NN query rates for CL-P

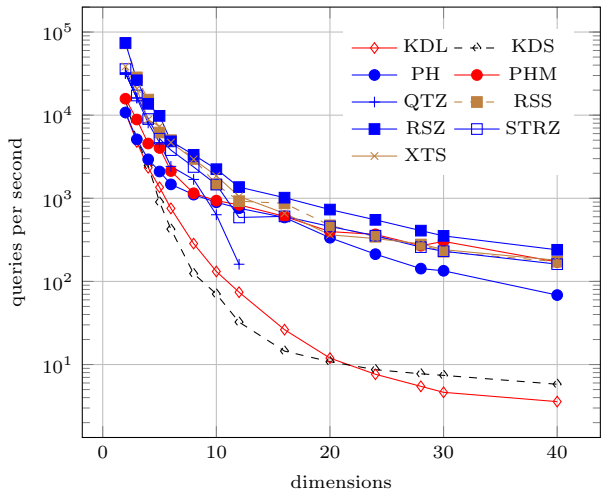


Figure 35: DIM: 10-NN query rates for CL-P

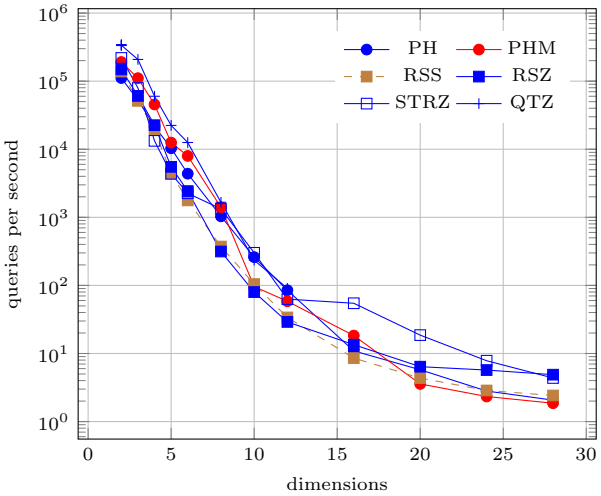


Figure 36: DIM: 1-NN query rates for CU-R

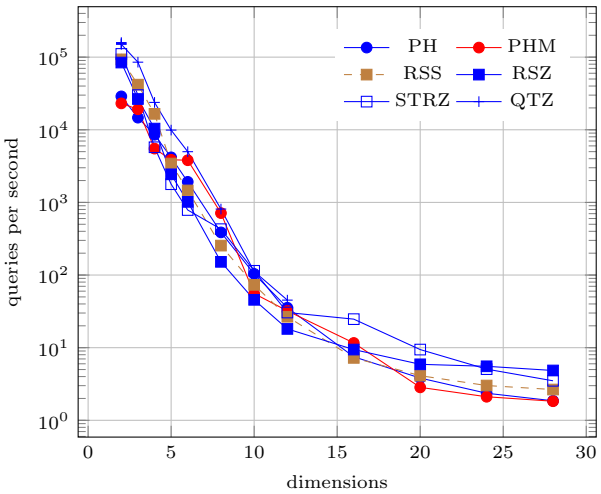


Figure 37: DIM: 10-NN query rates for CU-R

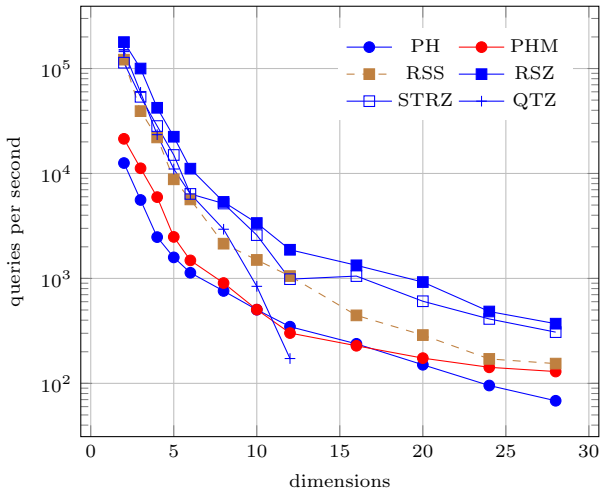


Figure 38: DIM: 1-NN query rates for CL-R

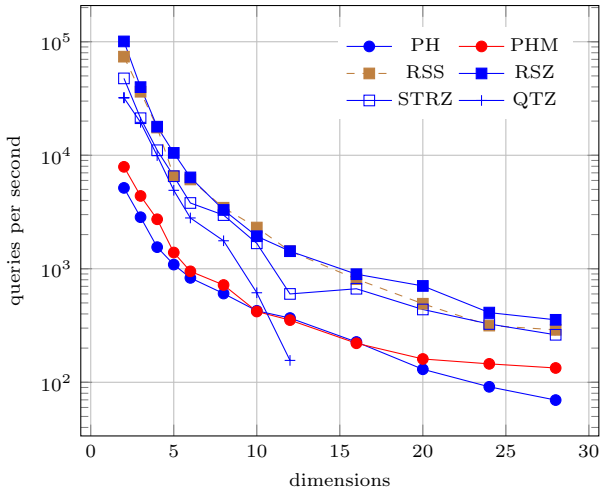
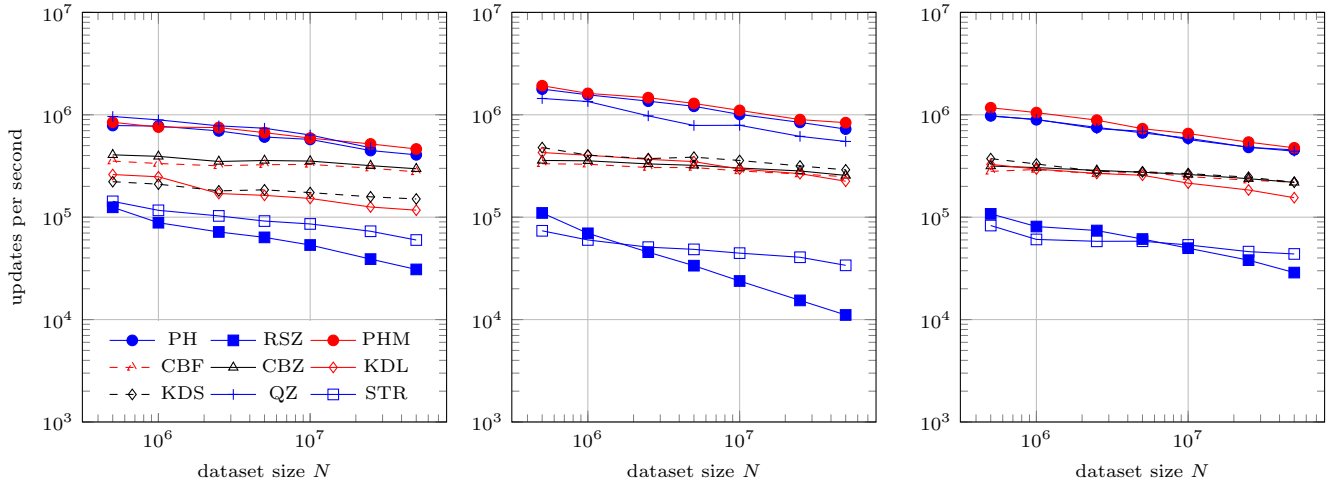
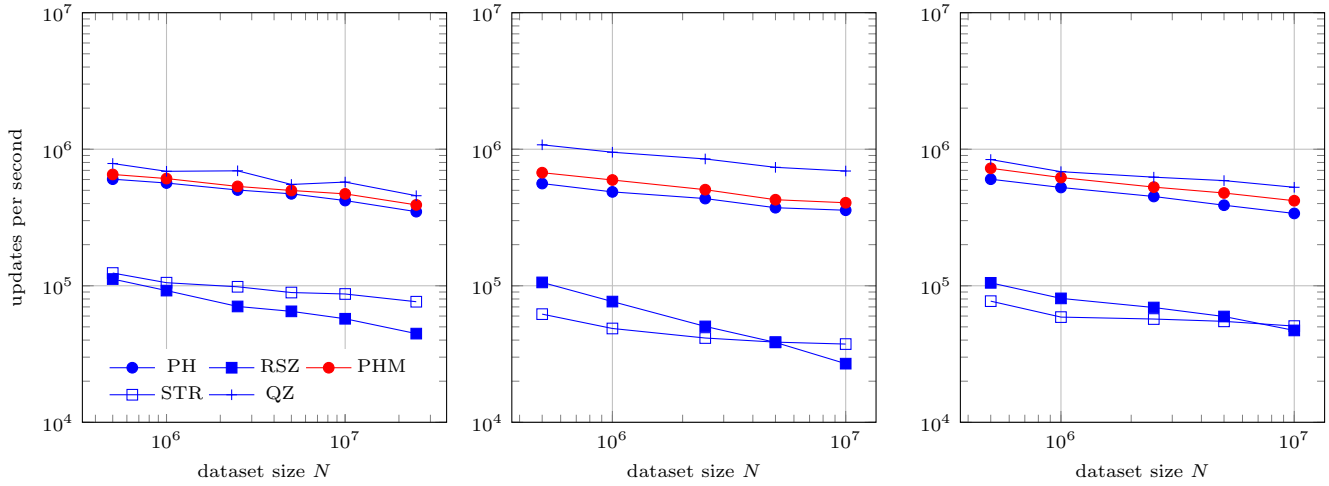


Figure 39: DIM: 10-NN query rates for CL-R



(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 40: Update rates for the OSM Alps, CUBE and CLUSTER datasets



(a) 2D OSM-R (b) 3D CUBE-R (c) 3D CLUSTER-R

Figure 41: Update rates for the OSM Alps, CUBE and CLUSTER datasets

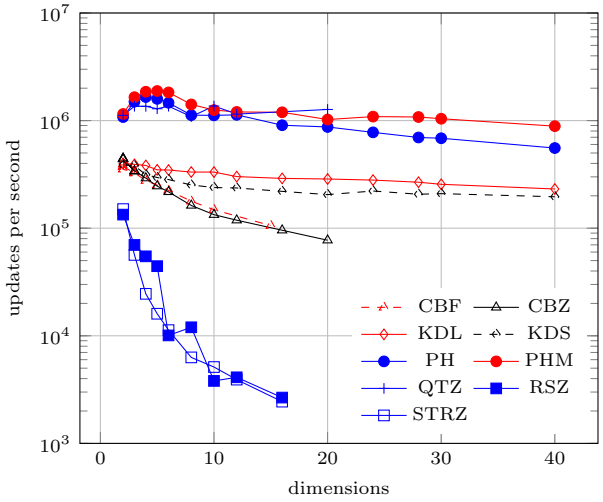


Figure 42: DIM: Update rates for CU-P

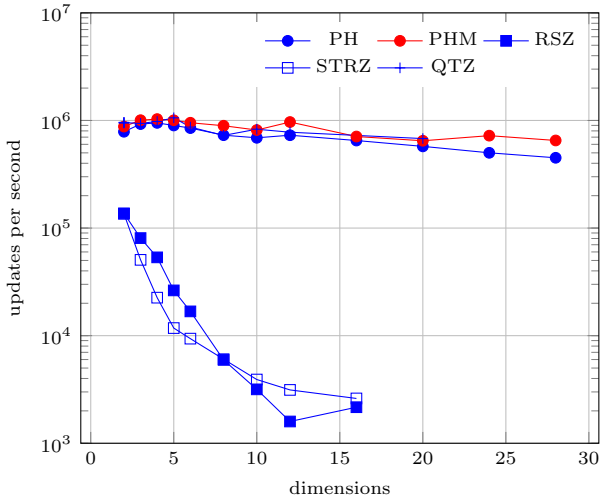


Figure 43: DIM: Update rates for CU-R

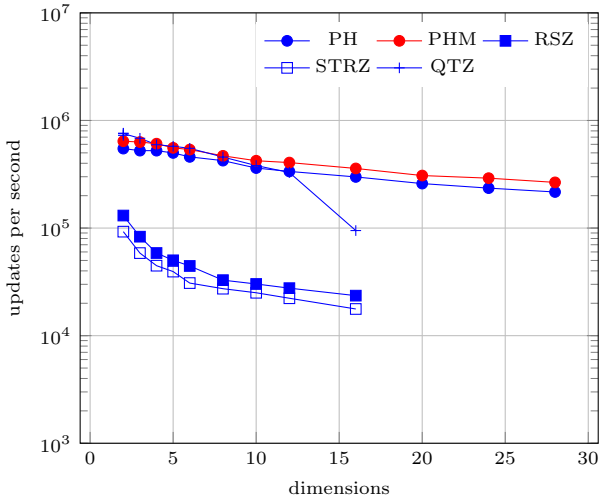


Figure 45: DIM: Update rates for CL-R

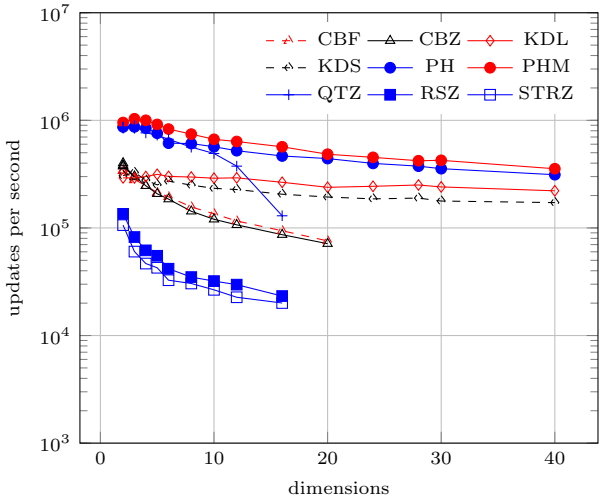
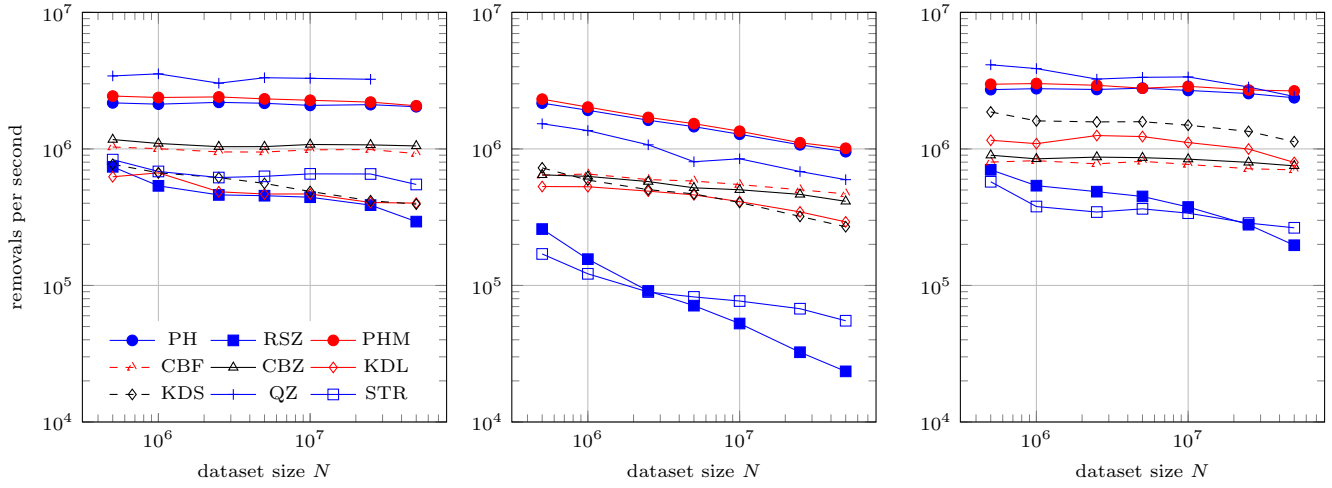
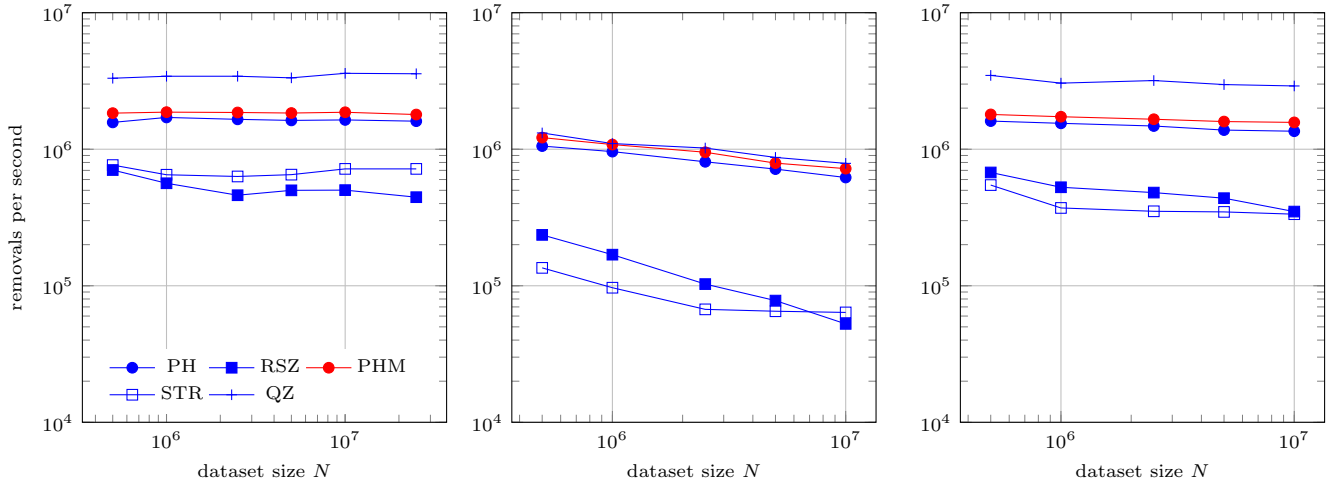


Figure 44: DIM: Update rates for CL-P



(a) 2D OSM-P (b) 3D CUBE-P (c) 3D CLUSTER-P

Figure 46: Removal rates for the OSM Alps, CUBE and CLUSTER datasets



(a) 2D OSM-R (b) 3D CUBE-R (c) 3D CLUSTER-R

Figure 47: Removal rates for the OSM Alps, CUBE and CLUSTER datasets

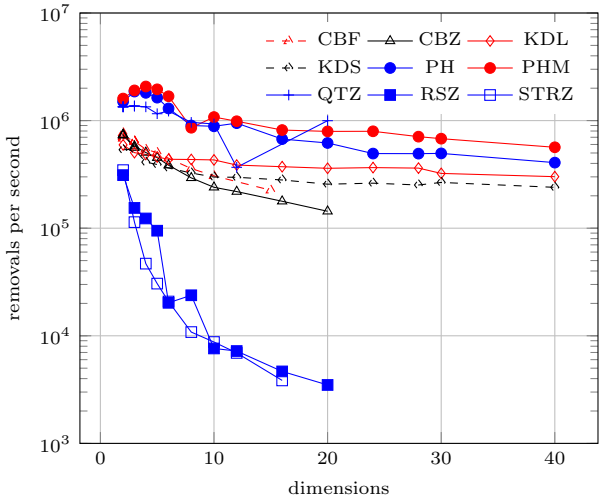


Figure 48: DIM: Removal rates for CU-P

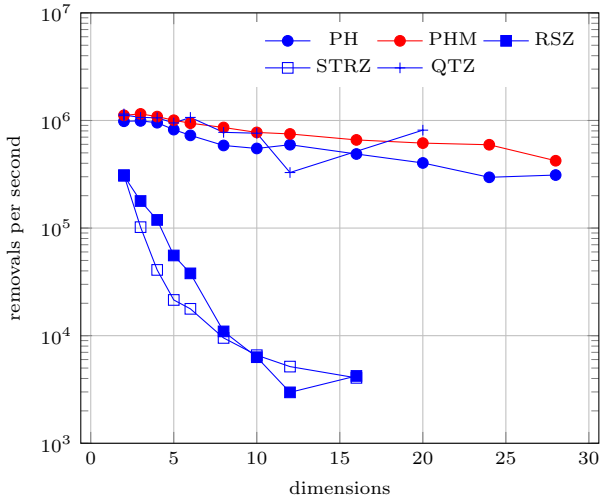


Figure 49: DIM: Removal rates for CU-R

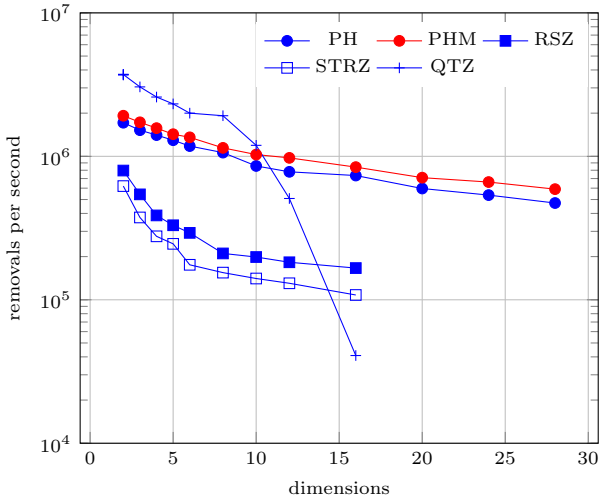


Figure 51: DIM: Removal rates for CL-R

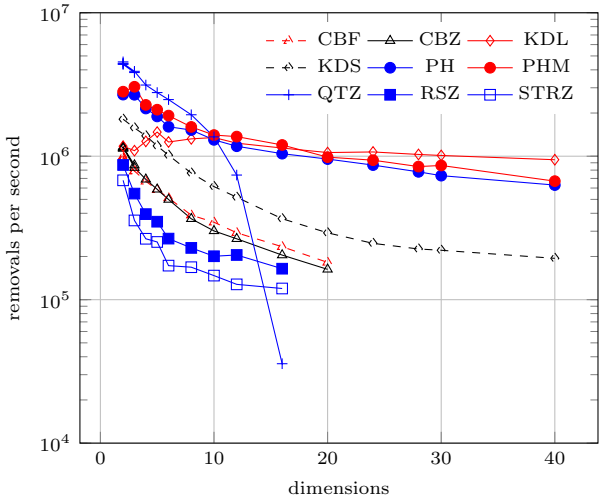


Figure 50: DIM: Removal rates for CL-P