PhD research on nD point cloud data management

Haicheng Liu, Peter van Oosterom,
Martijn Meijers, Rod Thompsen,
Xuefeng Guan,
Edward Verbree, Mike Horhammer
H.Liu-6@tudelft.nl

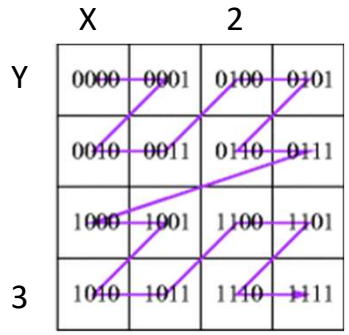
TU Delft

Outline

- nD-histogram for querying non-uniformly distributed data
- Executing convex polytope queries
- Benchmarks and applications

Background – PlainSFC

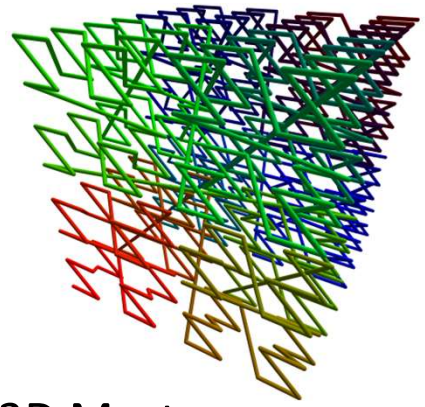
Data structure: Space filling curve (SFC) + Oracle index-organized tables (IOT)



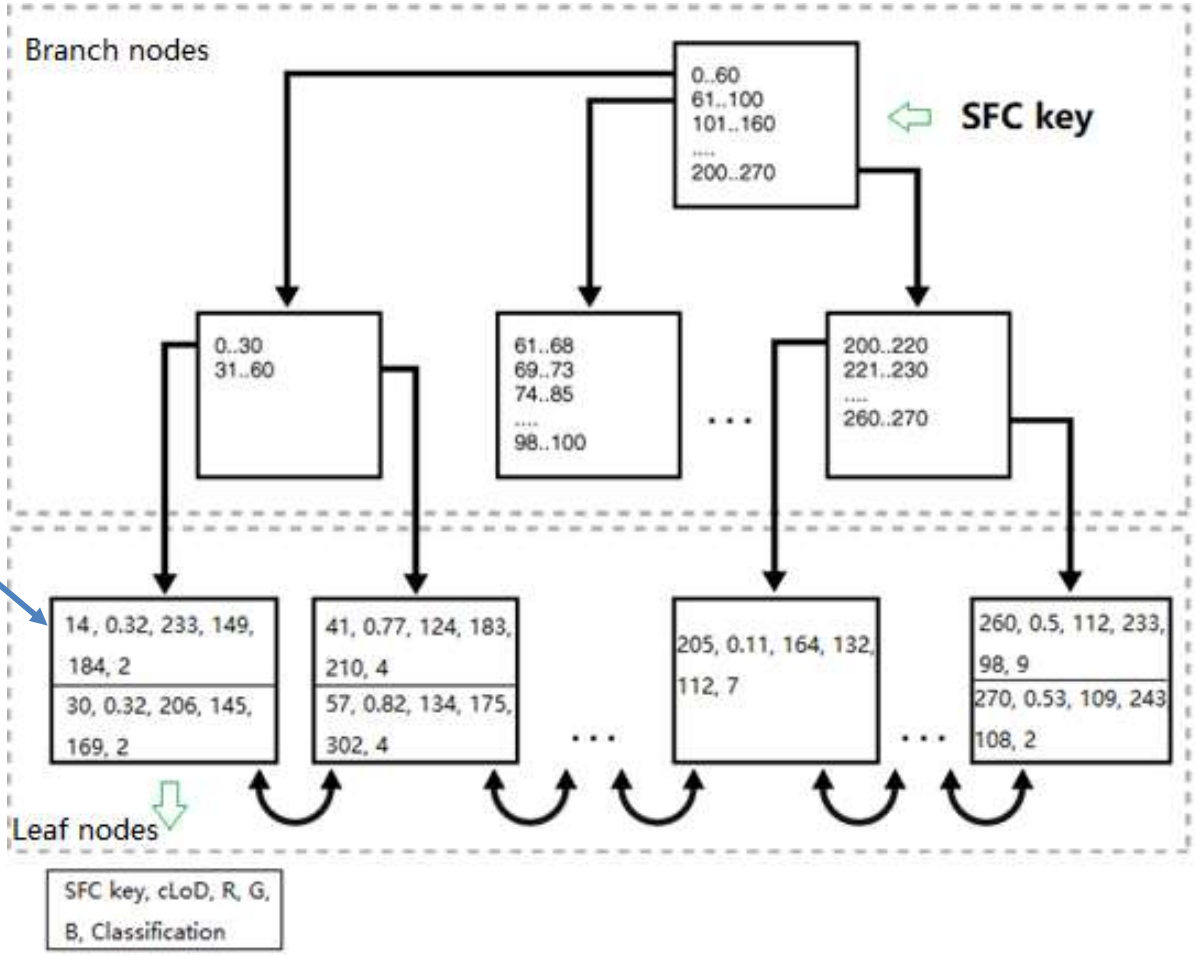
2D Morton curve

X: 2 -> 10 Y: 3-> 11

Morton key: 1110 -> 14



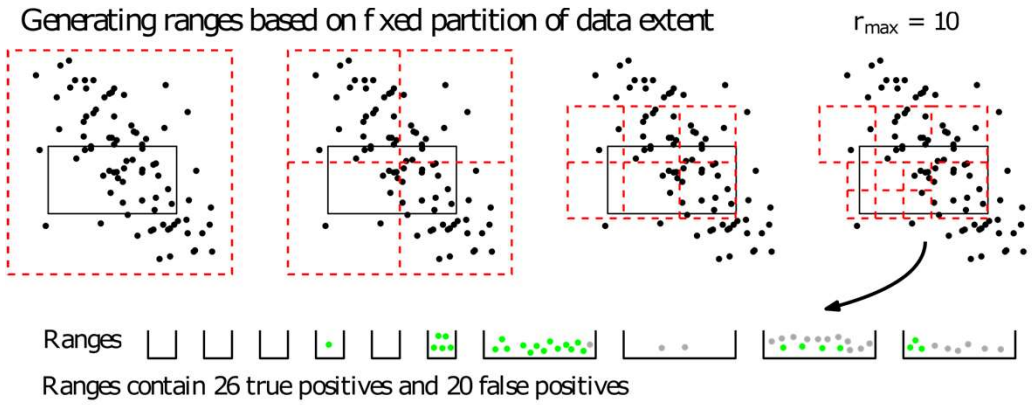
3D Morton curve -> nD



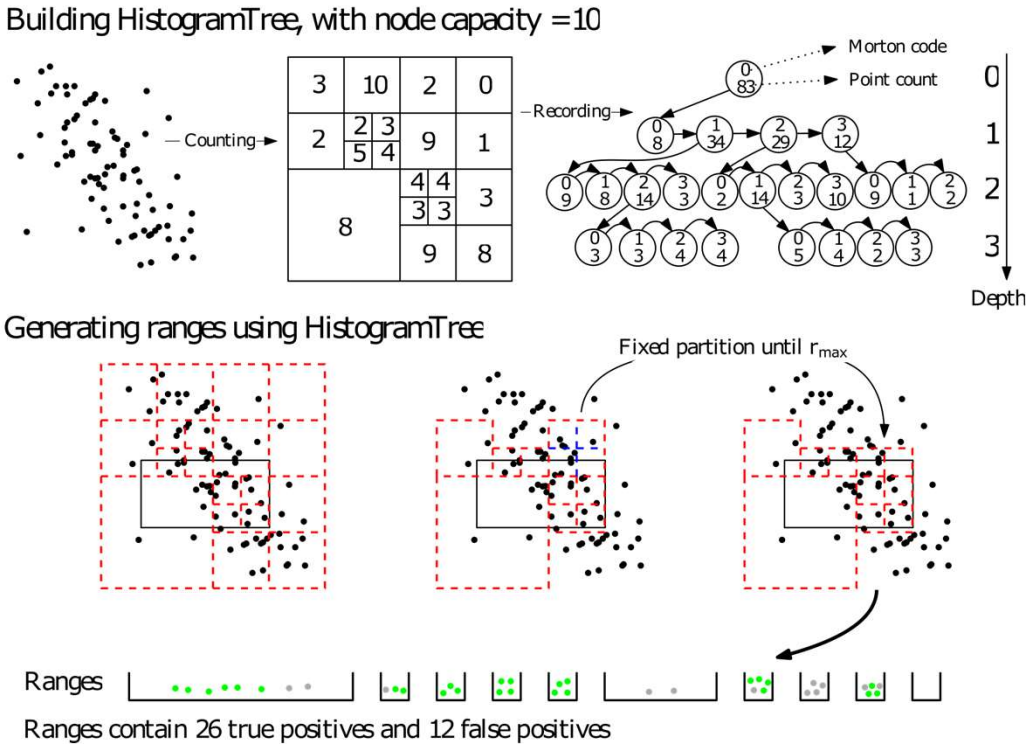
TU Delft

nD-histogram

PlainSFC querying

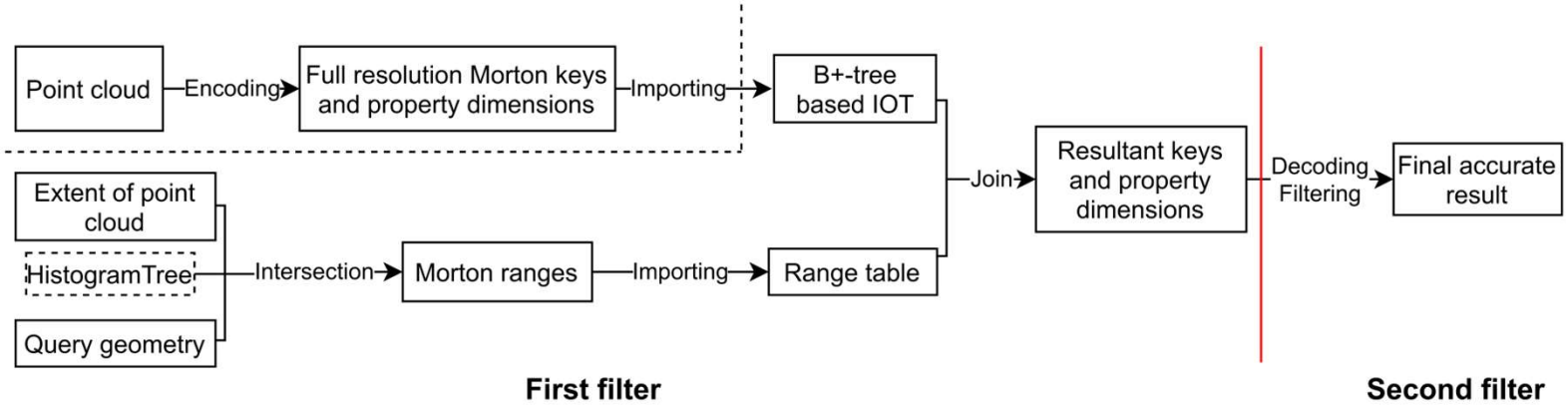


Improved by using an nD-histogram



nD-histogram

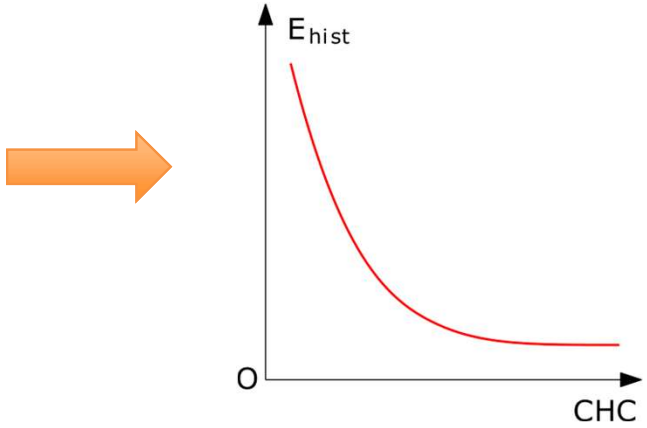
- Whole workflow



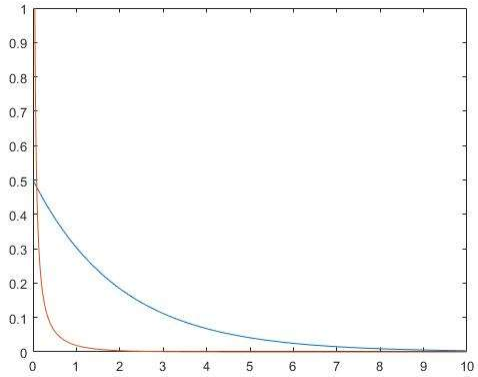
- Uniformity metric – cumulative hypercubic coverage (CHC)

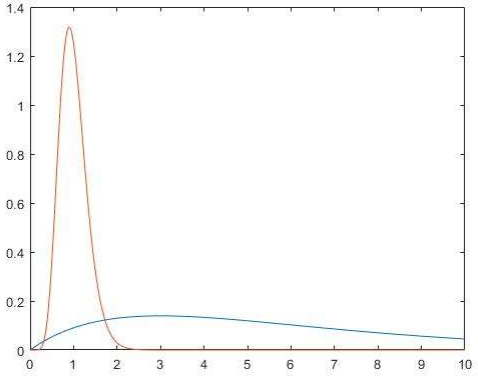
$$CHC = \int \cdots \int_{\Omega} \left(1 - e^{-f_n}\right) dv$$

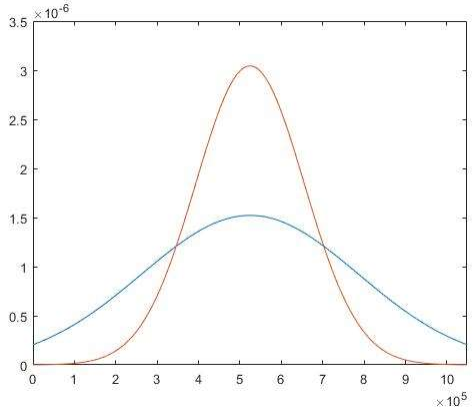
$$E_{hist} = \frac{Number\ of\ ranges\ exported\ by\ PlainSFC}{Number\ of\ non-vacant\ ranges}$$

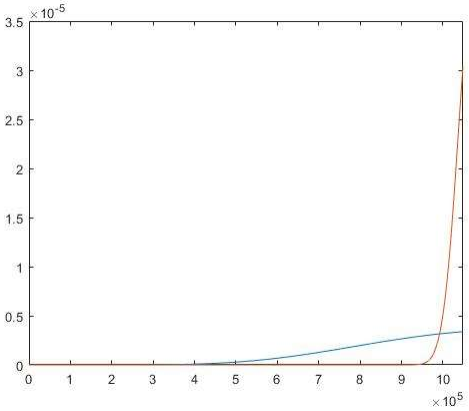


nD-histogram – simulations









Different distributions for simulating dimensions

CHC values of different data sets

	3D	4D	5D	6D
DG1	0.4790	0.3329	0.2210	0.1571
DG2	0.0704	0.0573	0.0484	0.0410
DG3	-	0.0102	0.0082	0.0090
DG4	-	-	0.0017	0.0019
DG5	-	-	-	0.0005

Average Effectiveness of nD-histogram

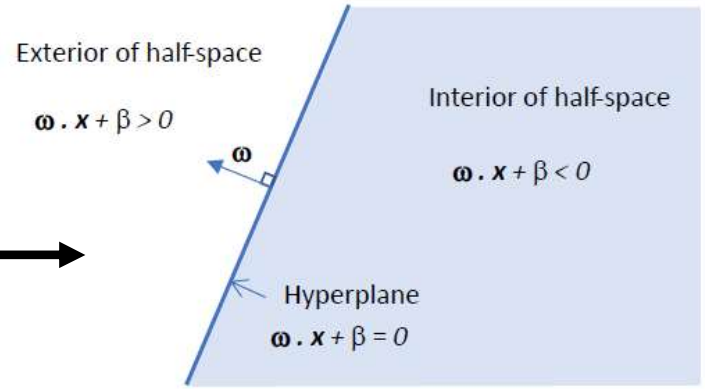
	3D	4D	5D	6D
DG1	1.01	1.05	1.06	1.35
DG2	3.70	5.08	2.92	4.24
DG3	-	6.44	6.68	8.00
DG4	-	-	40.83	10.07
DG5	-	-	-	62.17

TU Delft

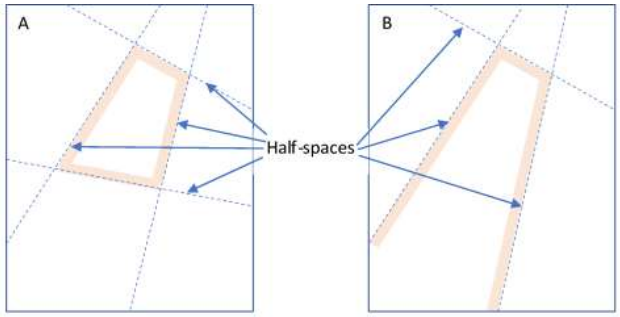
Convex polytope querying – definition

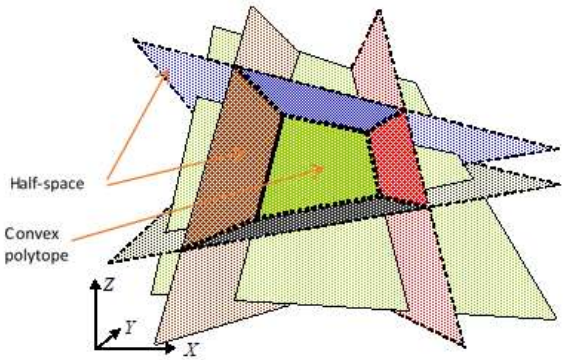
$$C = \cap_{i=1}^{m} H_i$$

where H_i is a set of m half-spaces

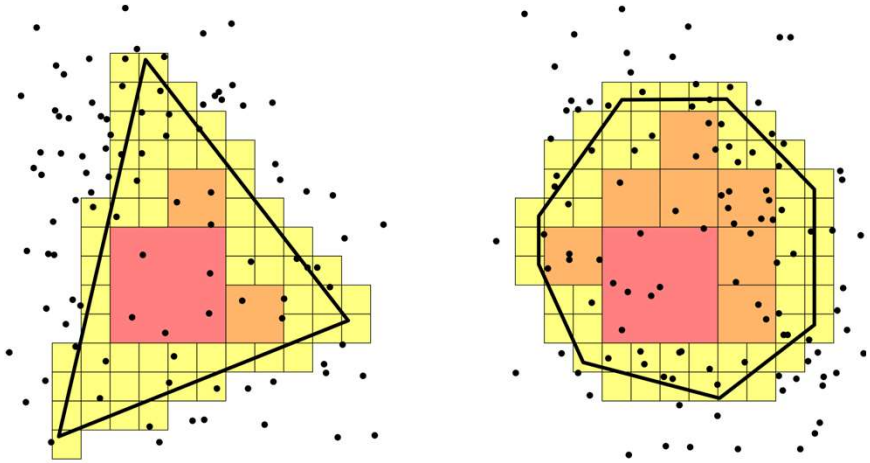


2D and 3D polytopes



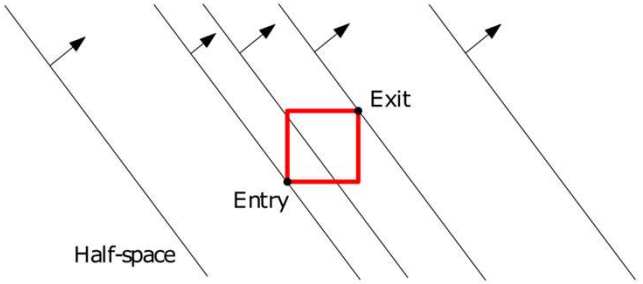


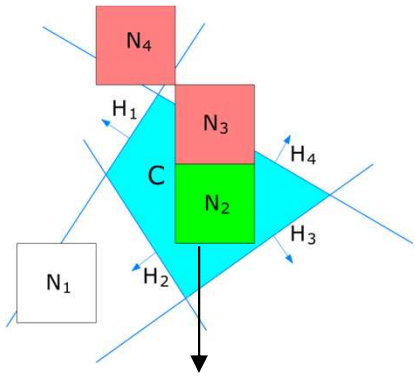
Range computation for a triangle and octagon query geometry



Convex polytope querying – intersection algorithms

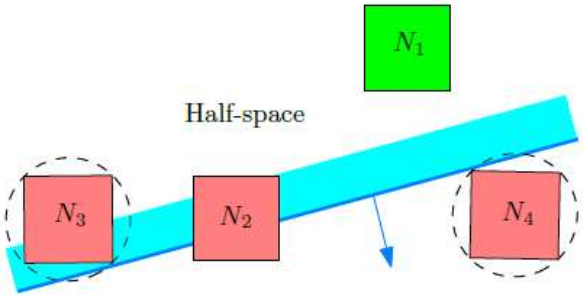
- SWEEP





HistogramTree node

- SPHERE



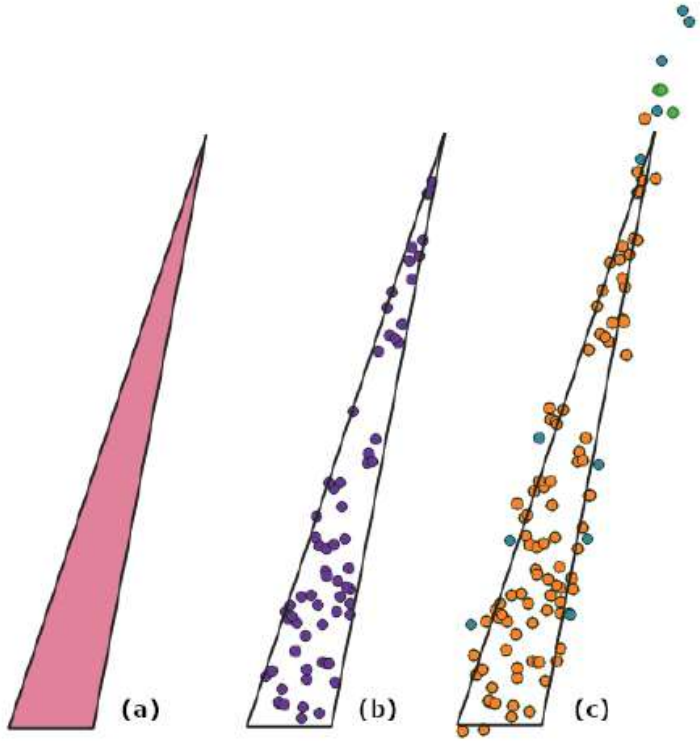
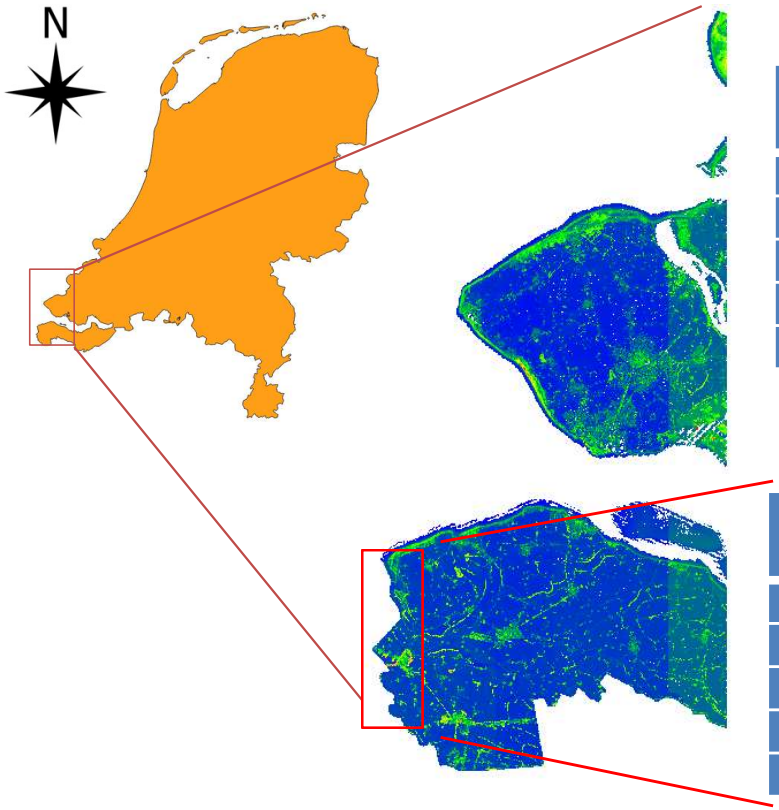


FIGURE 12 Querying results from different algorithms: (a) querying geometry, (b) accurate result, (c) Overlapping results of CPLEX (orange), SWEEP (green) and SPHERE (blue), without a second filter

- IBM CPLEX – rigorous linear programming method

Benchmarks and applications – AHN2

- Data: AHN2 sample (XYZ coordinates), 10 billion points, south-western part of the Netherlands; split into 5 data sets for scalability test
- cLoD: computed for each point and values range: (0, 12000)



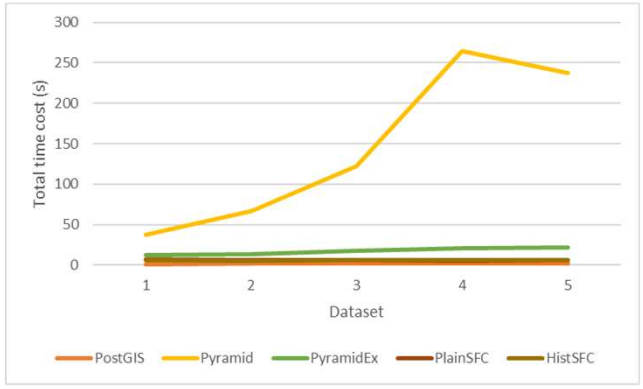
Storage size of datasets on the disk (in GB)

Dataset	Number of points	Raw TEXT files	SFC IOT	Pyramid	PyramidEx	PostGIS R-tree
1	5×10^8	16.49	10	18.24	18.52	7.212
2	10^9	32.98	19.95	36.39	36.95	14.17
3	2×10^9	64.42	38.97	71.06	72.01	28.1
4	6×10^9	193.9	118.3	213.6	216.2	82.32
5	10^{10}	323.4	199.7	356.9	360.7	138.0

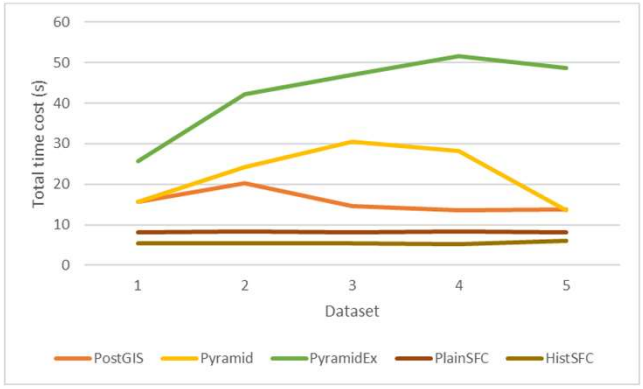
Query windows with different selectivity

	X	Y	Z	cLoD	Whole (output/total)
Small1	1.73%	4.37%	99.46%	94.31%	0.05%
Small2	20.5%	79.11%	1.01%	23.48%	0.05%
Small3	20.25%	17.58%	98.8%	1.03%	0.05%
Medium	20.29%	35.02%	98.29%	11.23%	0.67%
Large	20.39%	55.3%	98%	40.03%	4.53%

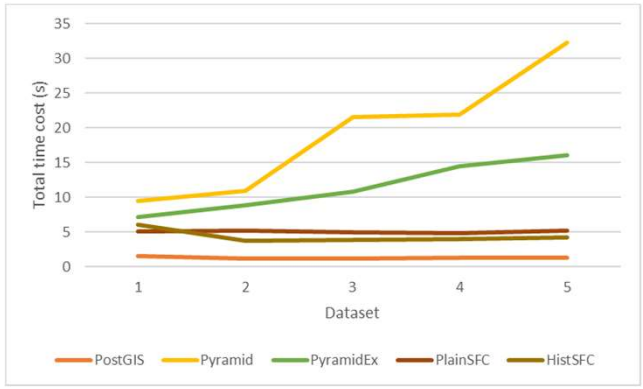
Benchmarks and applications – AHN2



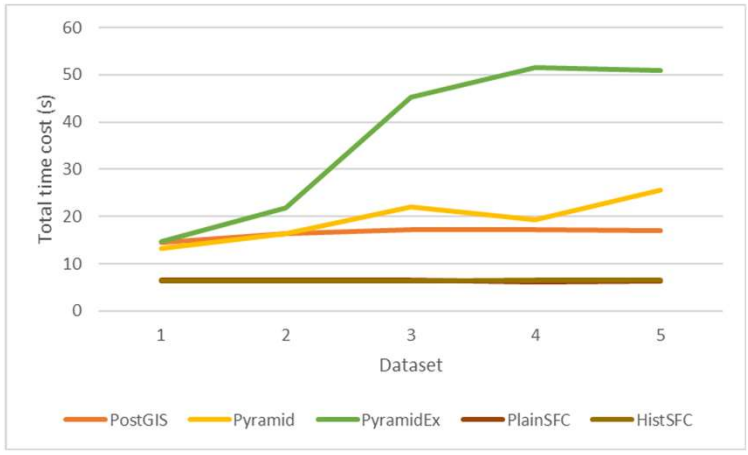
Small-1 query



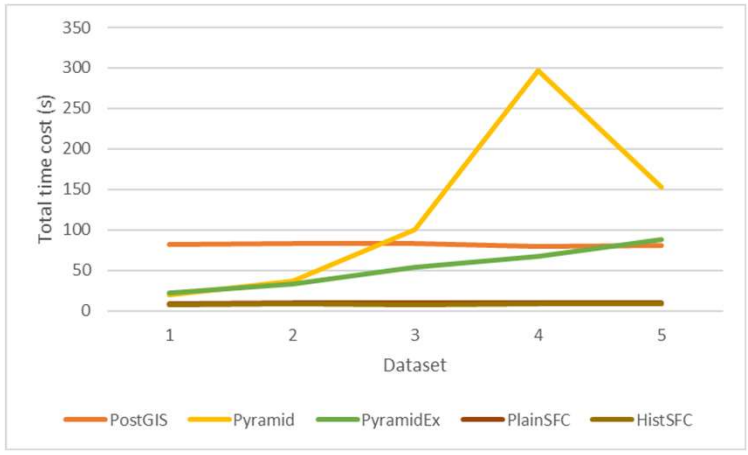
Small-2 query



Small-3 query



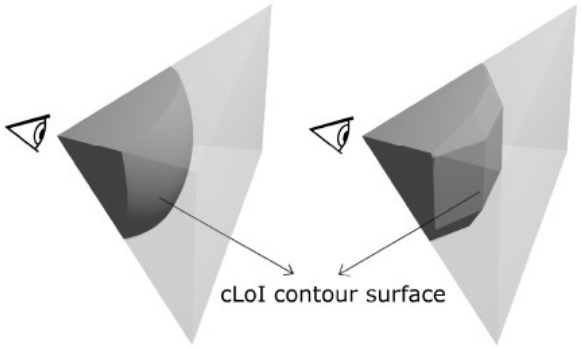
Medium query

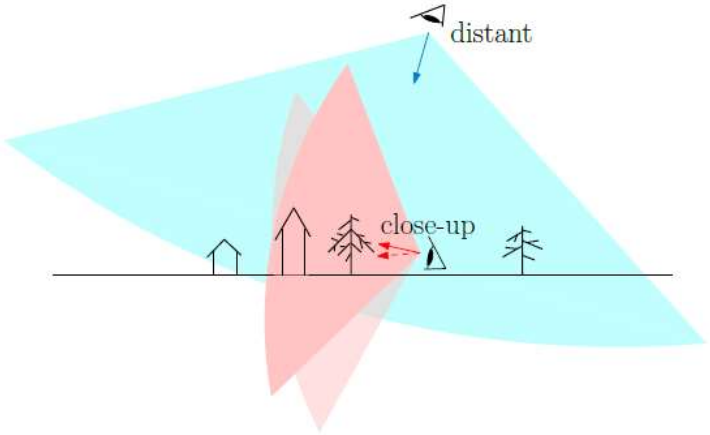


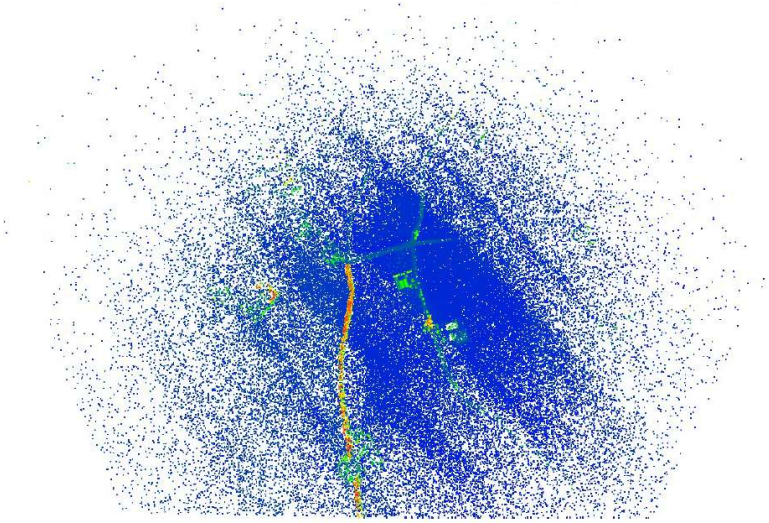
Large query

Benchmarks and applications – AHN2

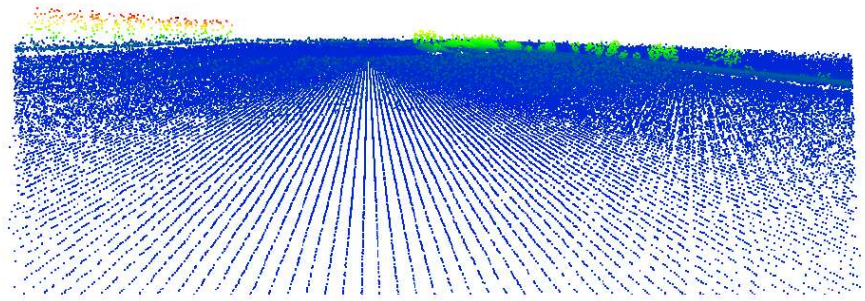
Perspective view selection using the polytope, response time < 1 s







Distant perspective view



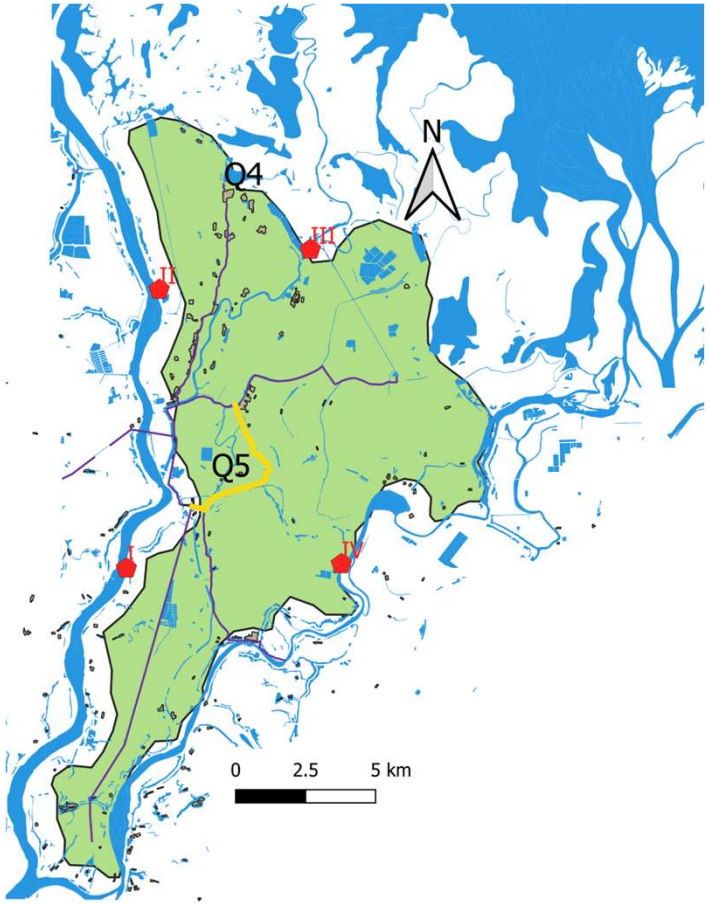
(a) Perspective view, colored by elevation

(b) Orthogonal projection, colored by maximum cLoI value

Close-up perspective view

Benchmarks and applications – flood

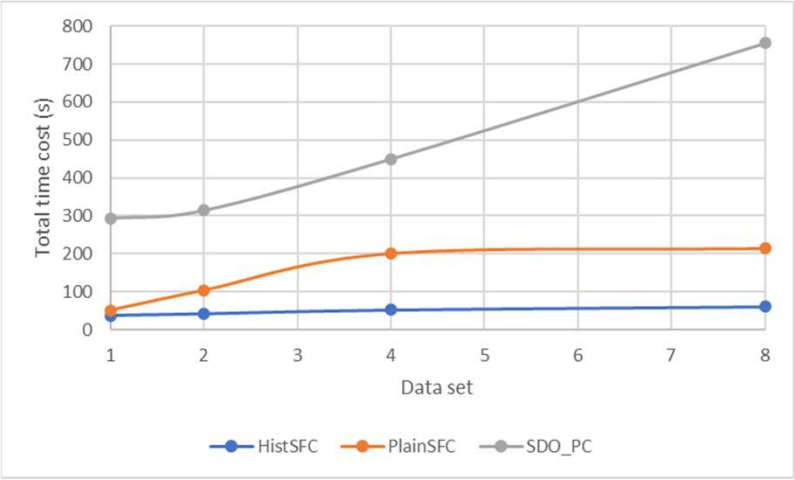
Niansi Levee, Jiangxi Province, China



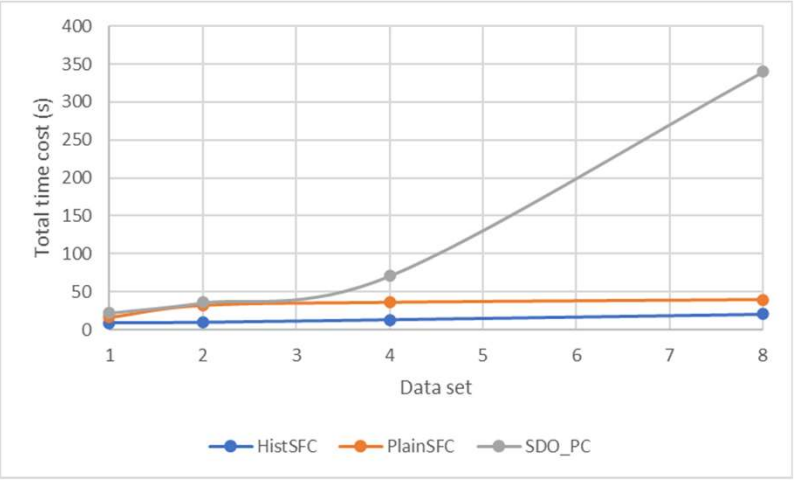
- For flood risk analysis, a large database storing modelling results is needed, e.g., generate flood maps
- Computation based on mesh, storage using raster is cumbersome:
  - Triangular cells have to be averaged, accuracy damaged;
  - The converted rasters also contain lots of empty cells
- Extract Centroids of mesh to form a point cloud
  - 8 dimensions, caseID, XYZT, flow velocity, depth and direction, 332,775,680 points

Benchmarks and applications – flood

Qid	Description
Q1	Select the area that is flooded with depth greater than 3 m, in case 1
Q2	Select the area that is flooded in 24 hours, in case 1
Q3	Select the maximum inundation area, in case 1
Q4	Select the area that is flooded around several houses (a rectangle area), in case 1
Q5	Select the dangerous points along a country road (velocity ≥ 0.5), in case 1
Q6	Select dangerous locations evaluated by human instability (depth × velocity ≥ 2), in case 1

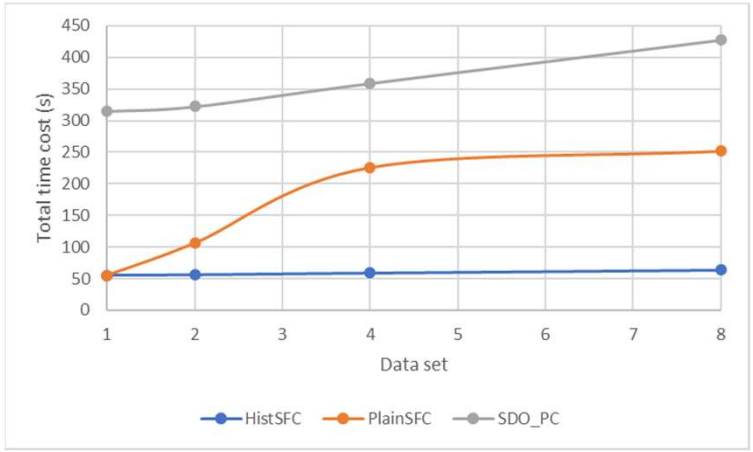


Q1 time cost (s)

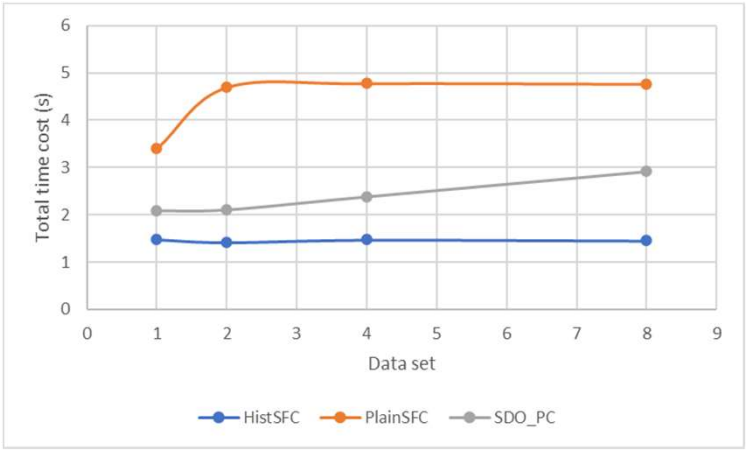


Q2 time cost (s)

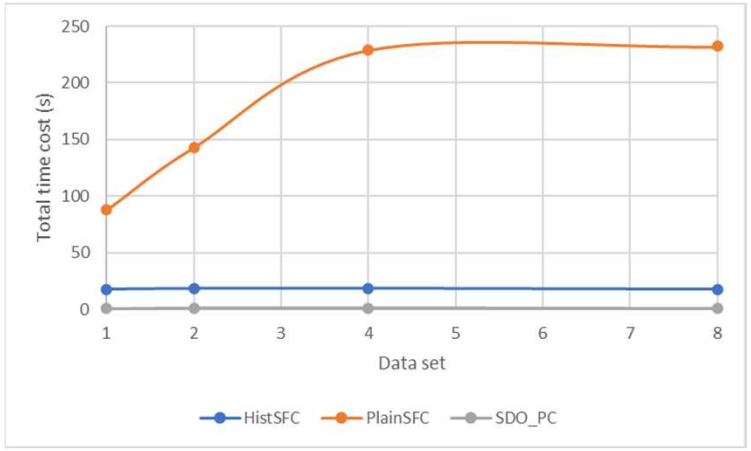
Benchmarks and applications – flood



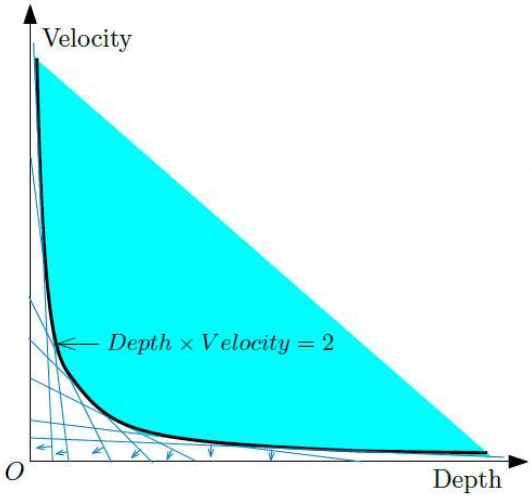
Q3



Q4



Q5



Q6 polytope modelling

	GEOM	SWEEP	CPLEX
Case 1			
First filter	1.273	1.297	87.86
Second filter	14.99	15.08	15.08
Total	16.25	16.37	102.9
Case 5			
First filter	1.336	1.474	98.46
Second filter	15.29	15.01	15.01
Total	16.62	16.48	113.5

Q6 time cost (s)

Thanks for your attention