ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Incorporating distributed Dijkstra's algorithm into variogram calculation with locally varying anisotropy

Oscar Peredo, Felipe Navarro, Mauricio Garrido and Julián M. Ortiz

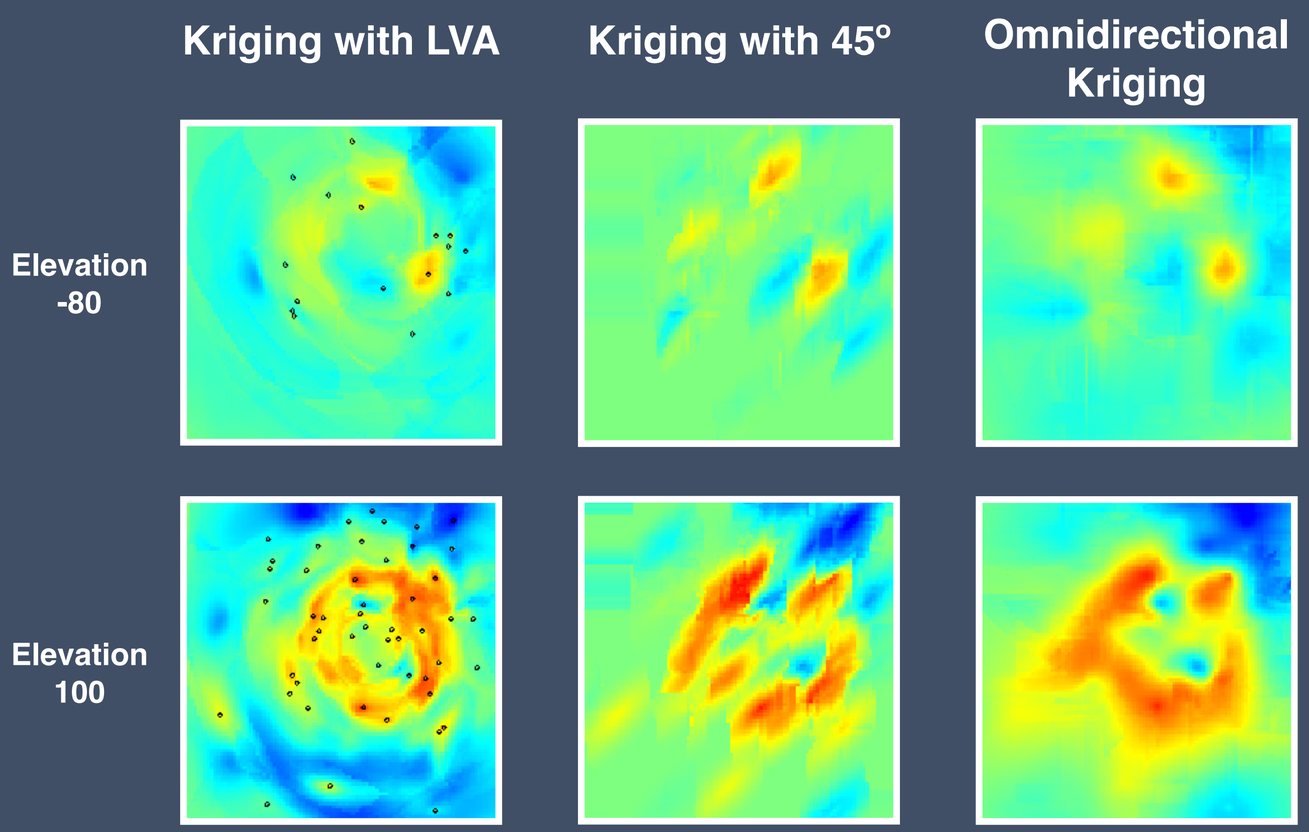
Department of Mining Engineering – FCFM – Universidad de Chile
Advanced Mining Technology Center – Universidad de Chile

37th APCOM Conference
Fairbanks, Alaska 2015

fcfm
Ingeniería de Minas
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

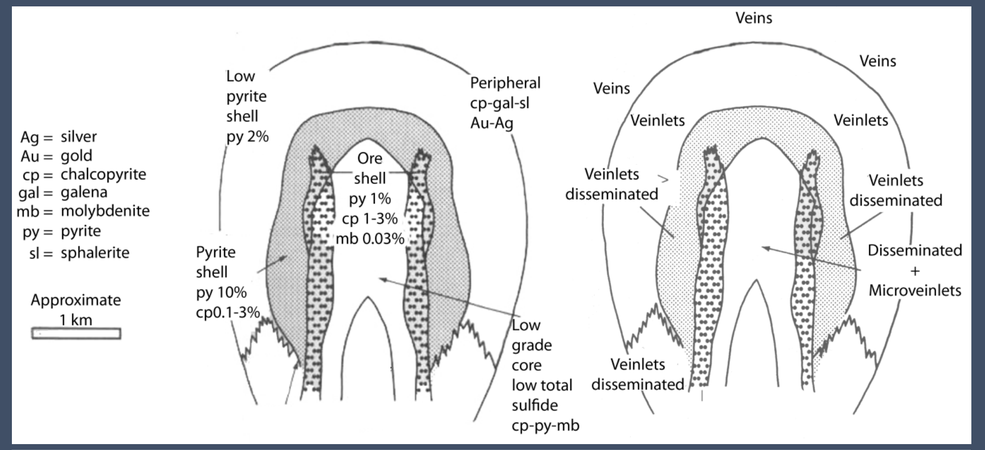
amtc
ADVANCED MINING TECHNOLOGY CENTER

Locally varying anisotropy (LVA)



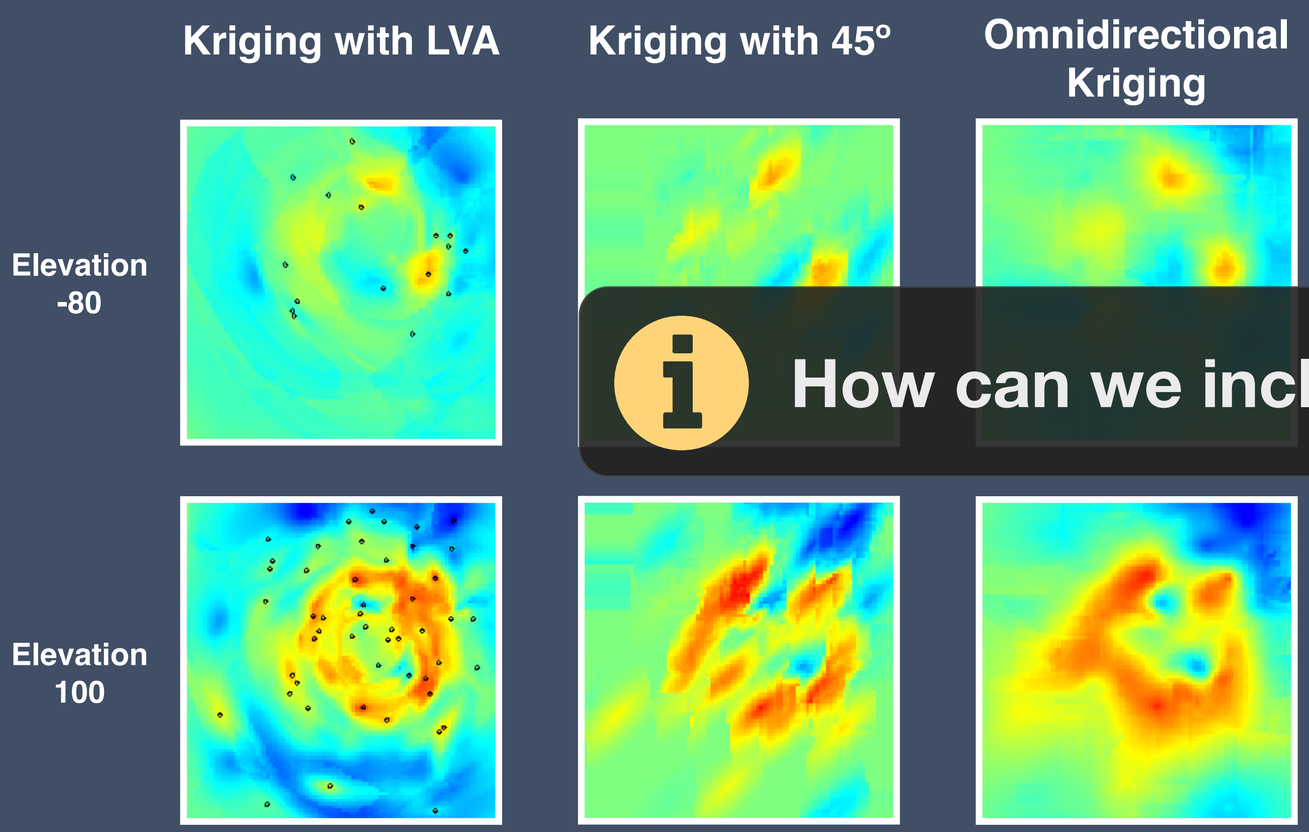
Source: *Boisvert* (2010)

Copper porphyry deposit



Source: *Guilbert and Park* (2007)

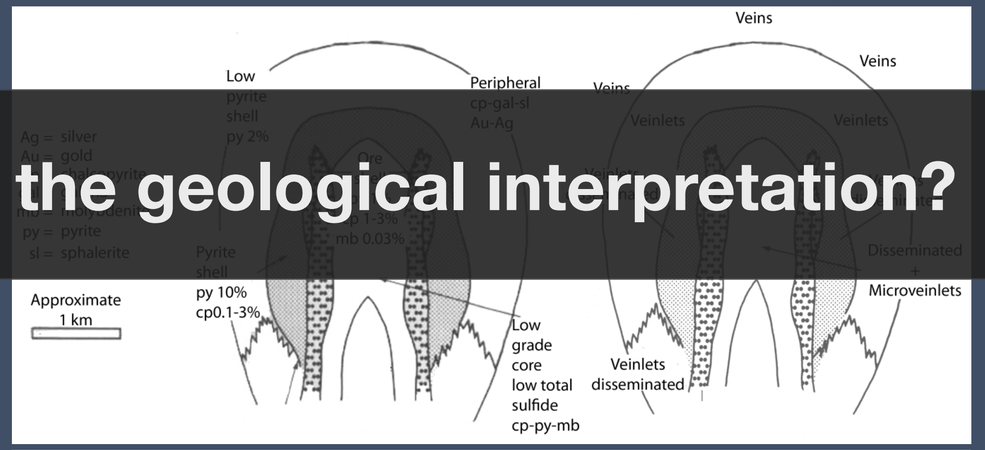
Locally varying anisotropy (LVA)



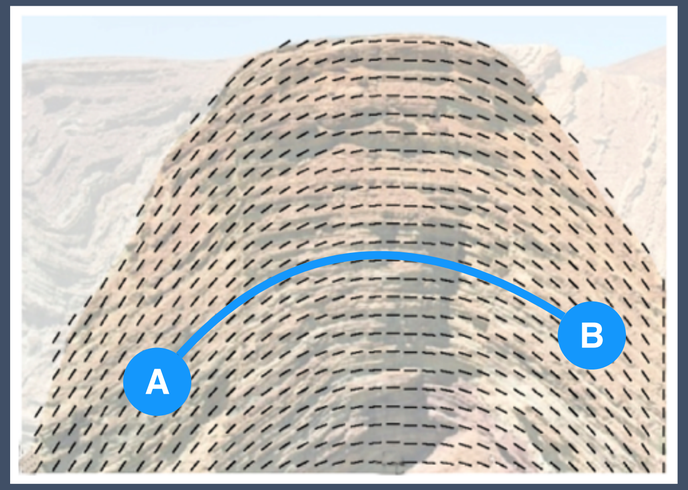
Source: *Boisvert* (2010)

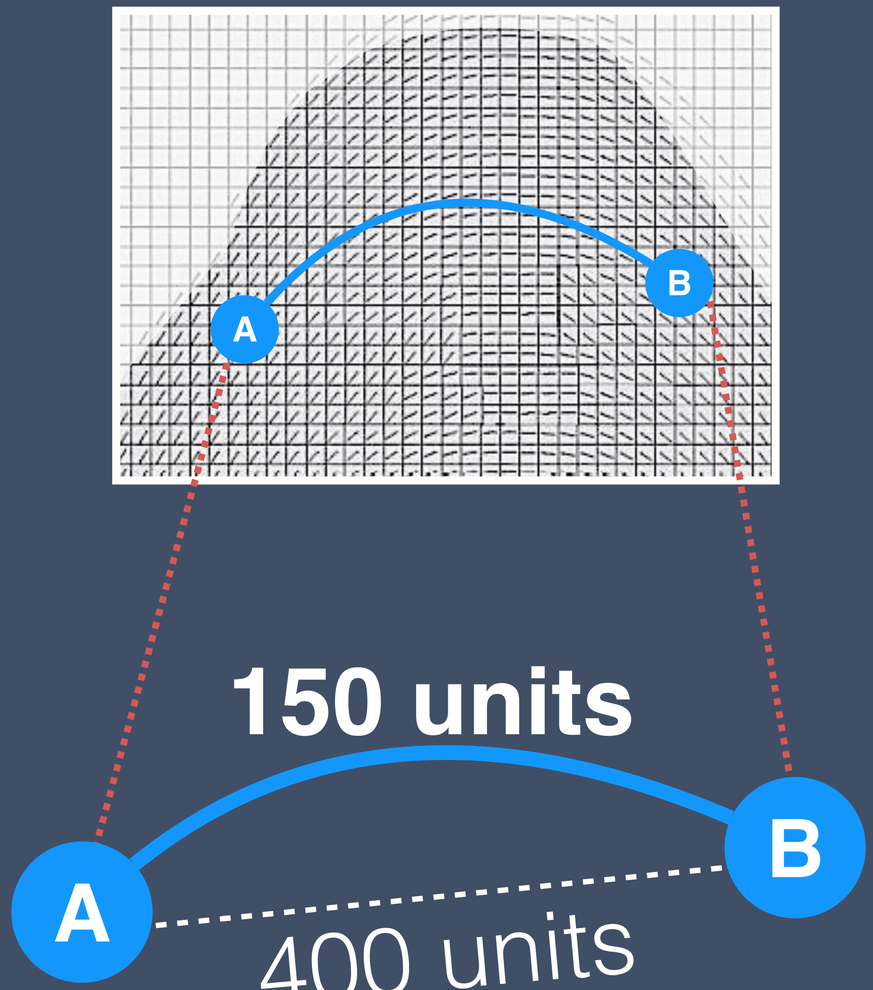
How can we include the geological interpretation?

Copper porphyry deposit



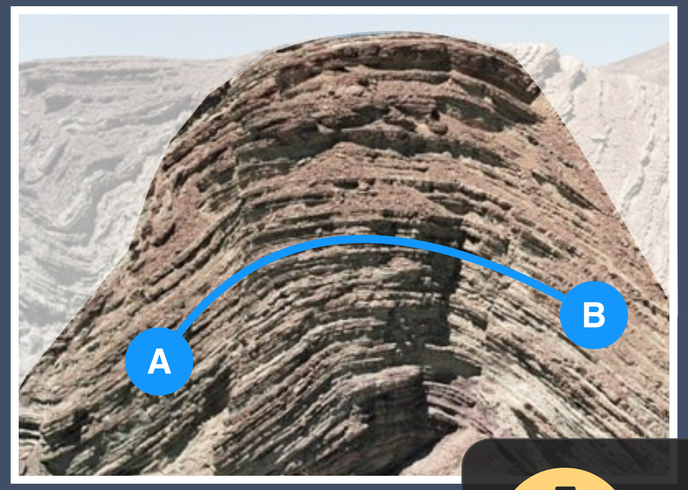
Source: *Guilbert and Park* (2007)

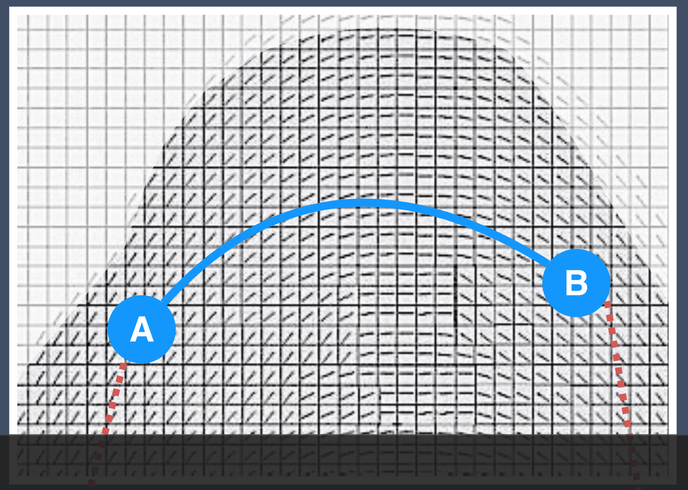




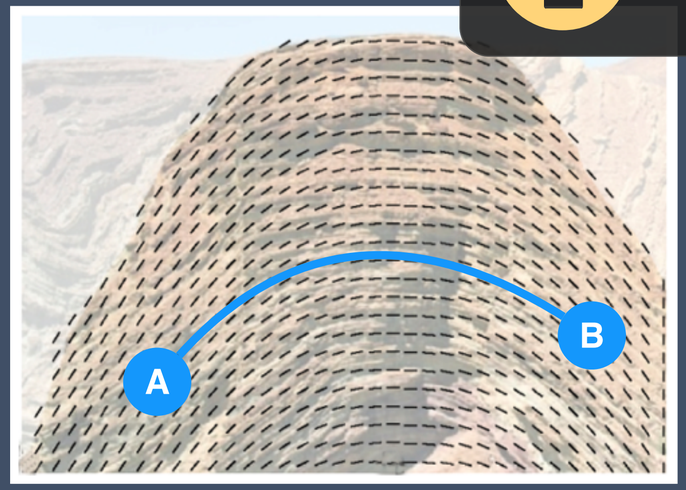
(*) Picture from an anticline with LVA (www.geology.about.com)

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING



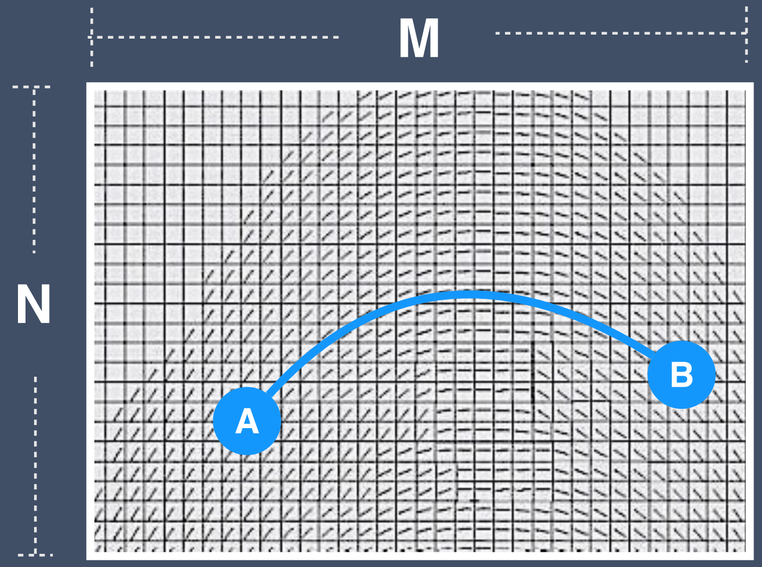


How can we calculate the anisotropic distance?



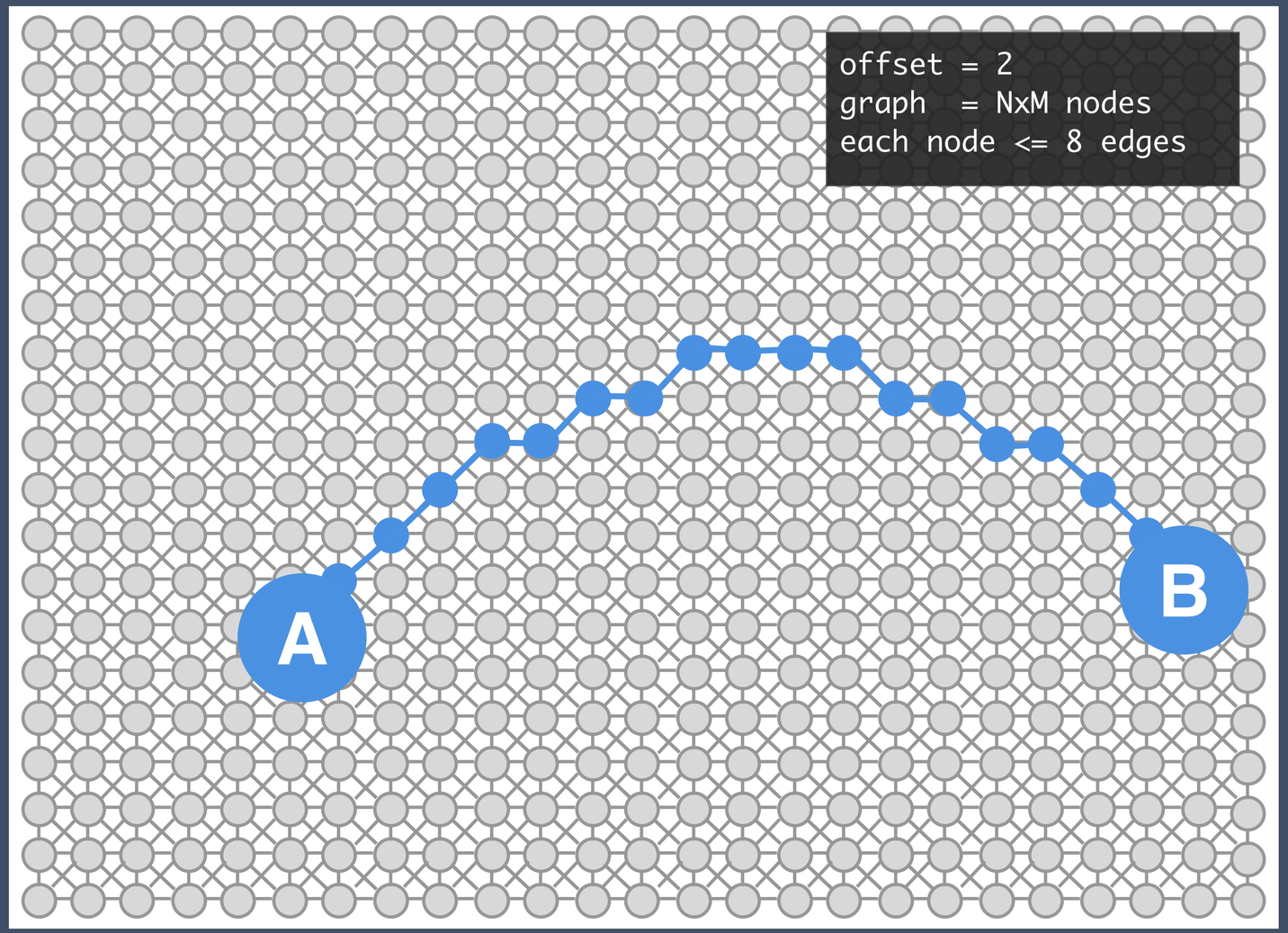
(*) Picture from an anticline with LVA (www.geology.about.com)

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

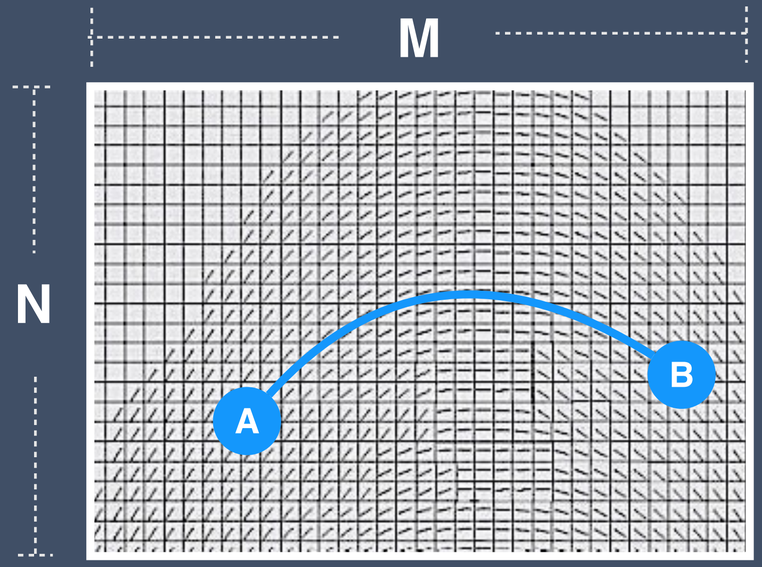


How to build the anisotropic distance matrix

1. Define Graph offset (local connectivity)
2. Build Graph
3. Find shortest path between A and B (using single source Dijkstra)
4. Repeat for each node (***NxM*** Dijkstra runs)

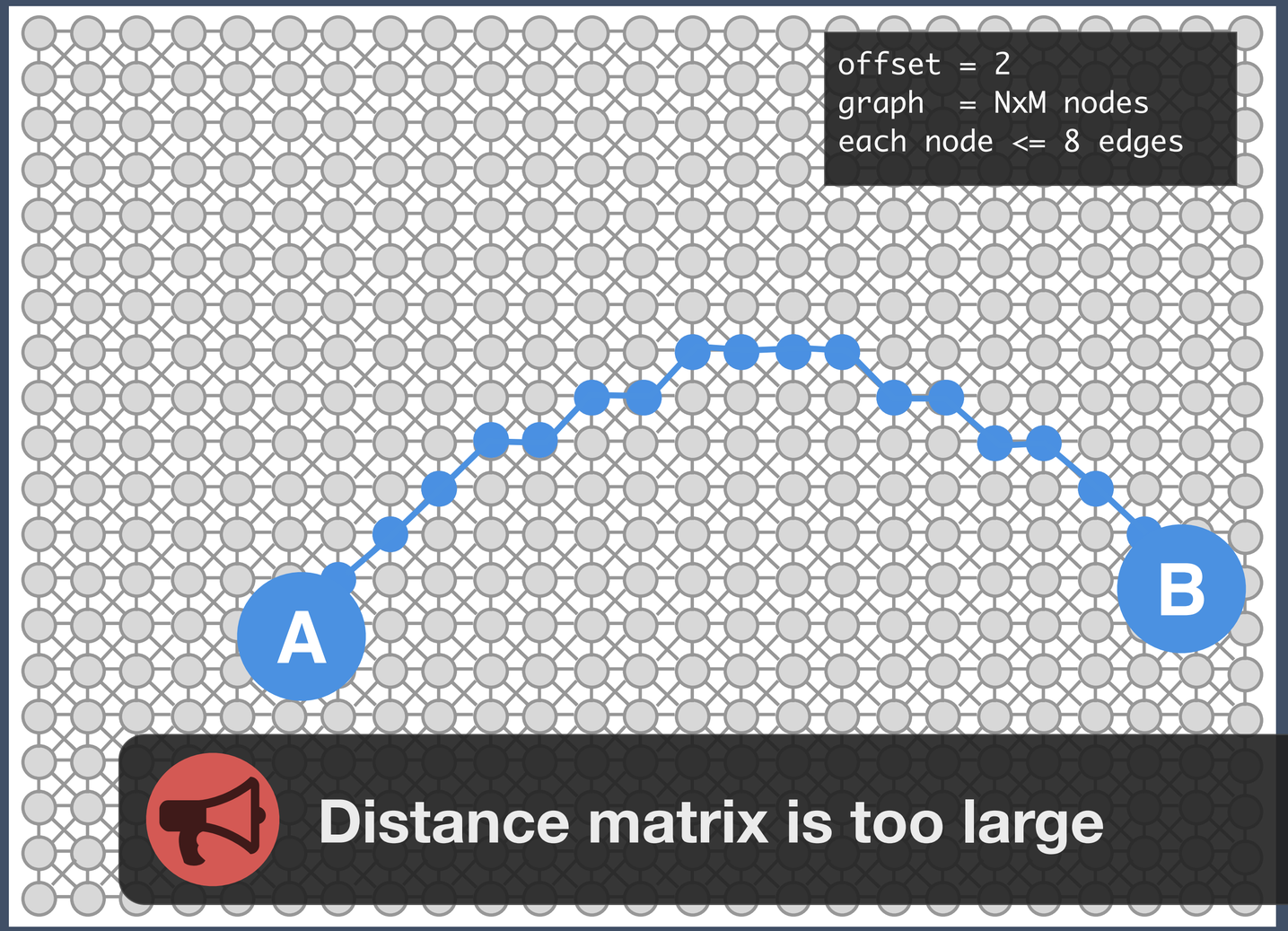


ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING



How to build the anisotropic distance matrix

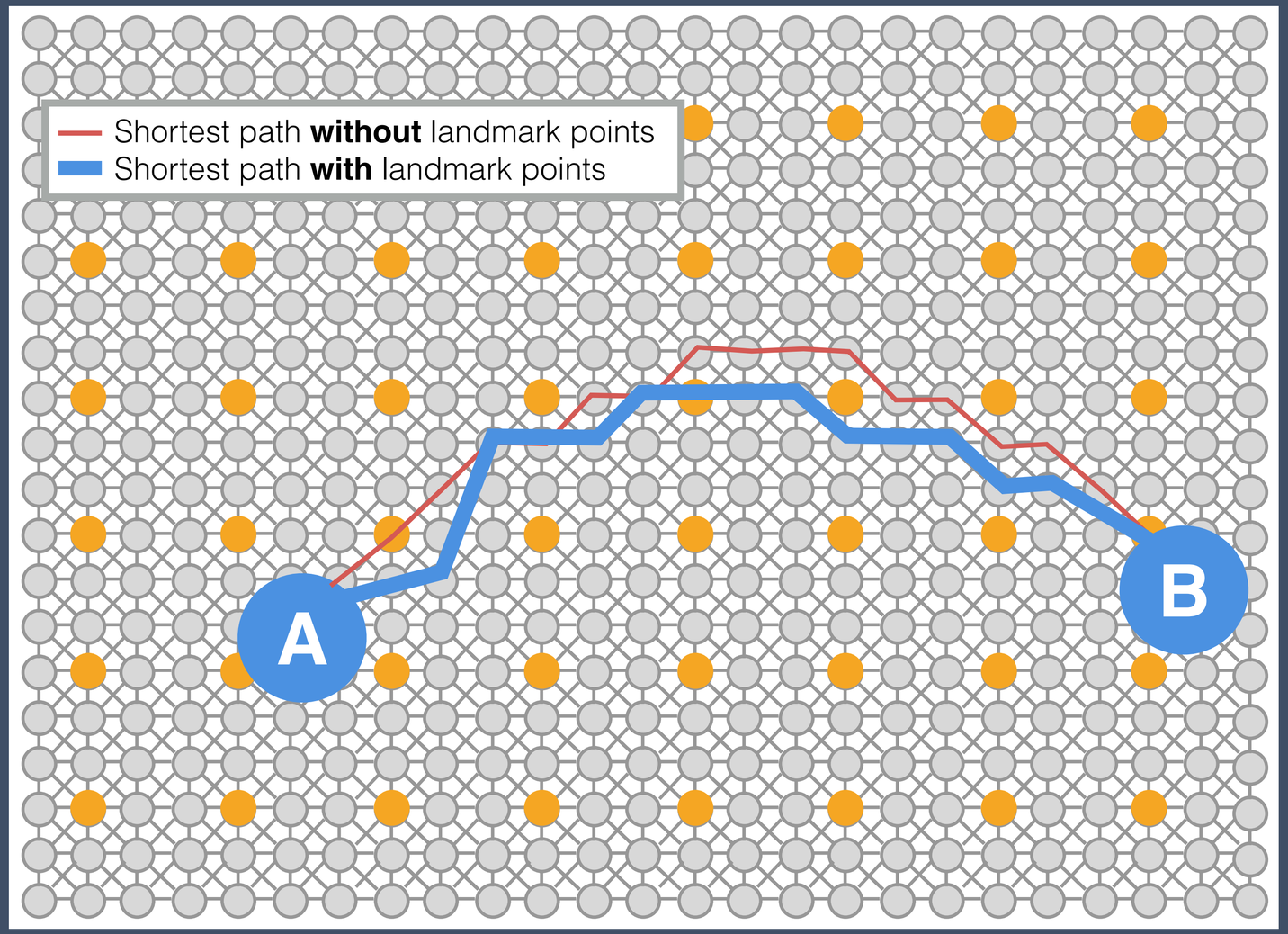
1. Define Graph offset (local connectivity)
2. Build Graph
3. Find shortest path between A and B (using single source Dijkstra)
4. Repeat for each node (***NxM*** Dijkstra runs)



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Dimensionality reduction using landmark points

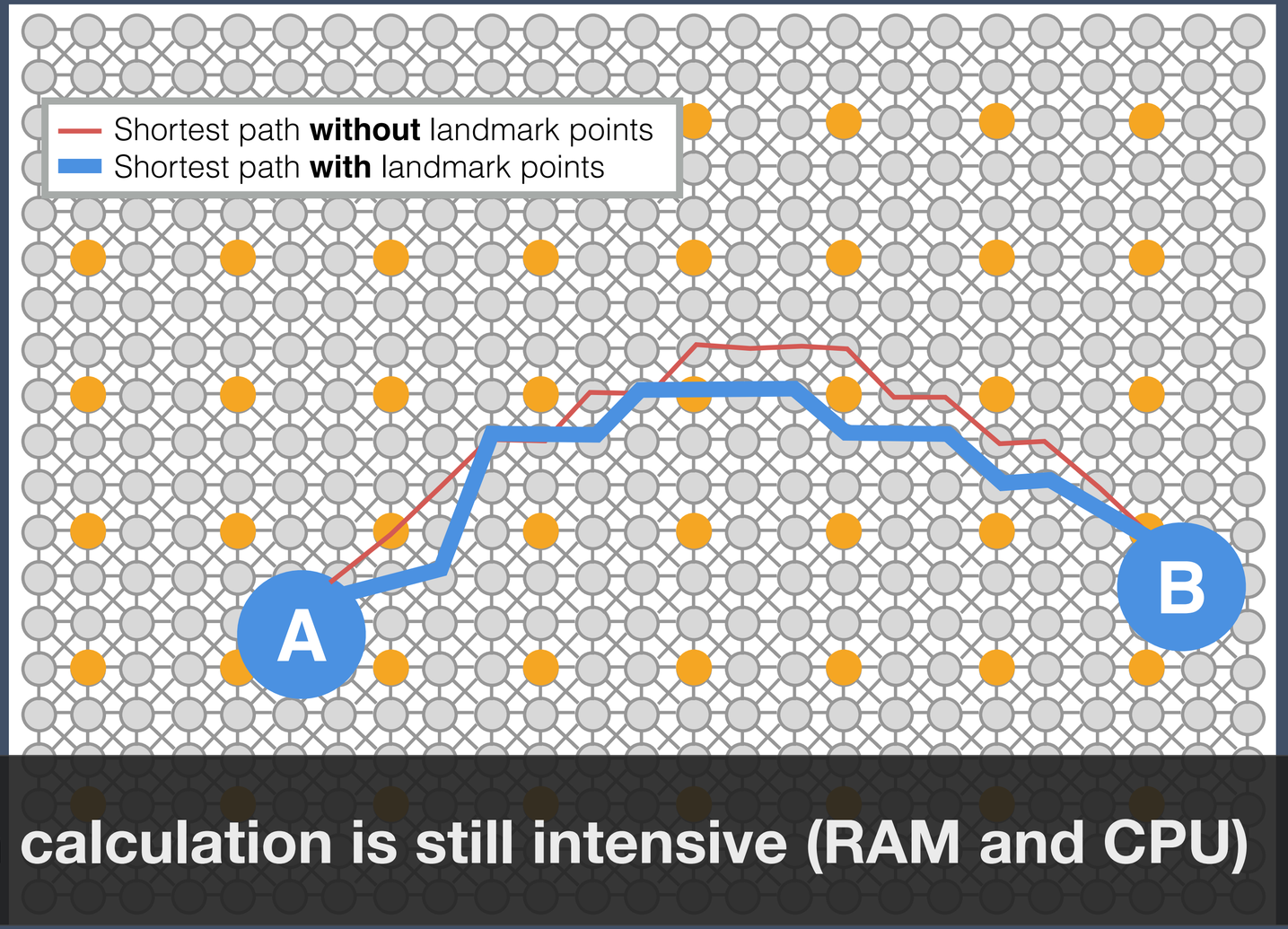
- L-ISOMAP is used
- Distance matrix size is reduced
  L *x (NxM)* **<<** *(NxM)*^{2}
- Trade-off: Numerical accuracy of anisotropic distance
- Shortest path calculation between landmark points and nodes (instead of between every node, ***L*** Dijkstra runs)



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Dimensionality reduction using landmark points

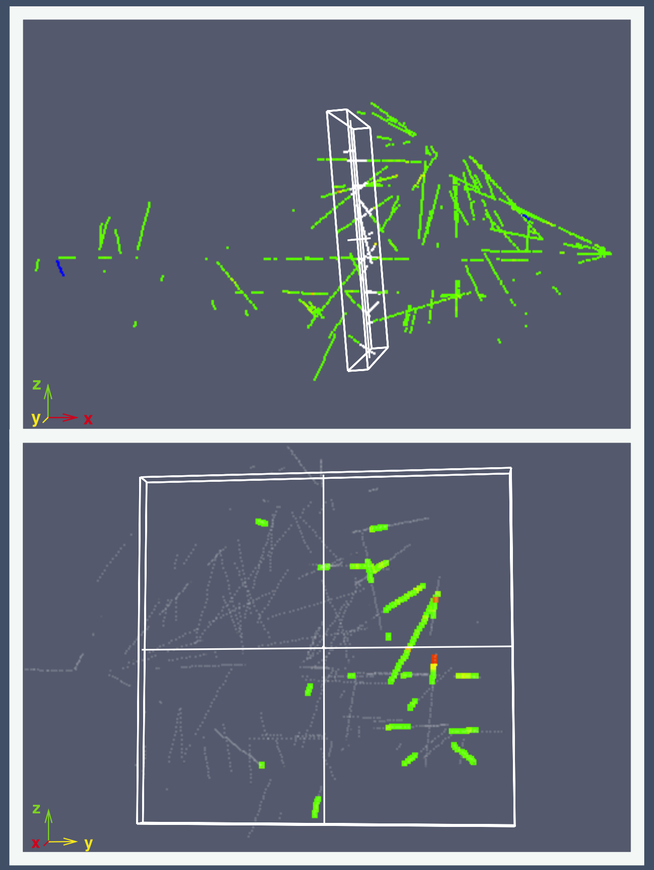
- L-ISOMAP is used
- Distance matrix size is reduced
  L *x (NxM)* **<<** *(NxM)*^{2}
- Trade-off: Numerical accuracy of anisotropic distance
- Shortest path calculation between landmark points and nodes (instead of between every node, ***L*** Dijkstra runs)



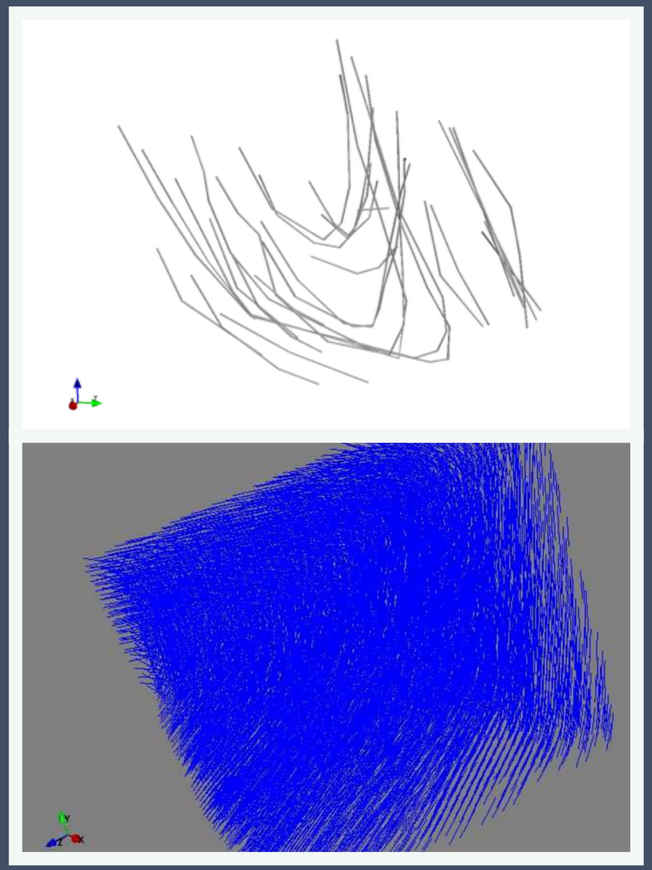
Shortest path calculation is still intensive (RAM and CPU)

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

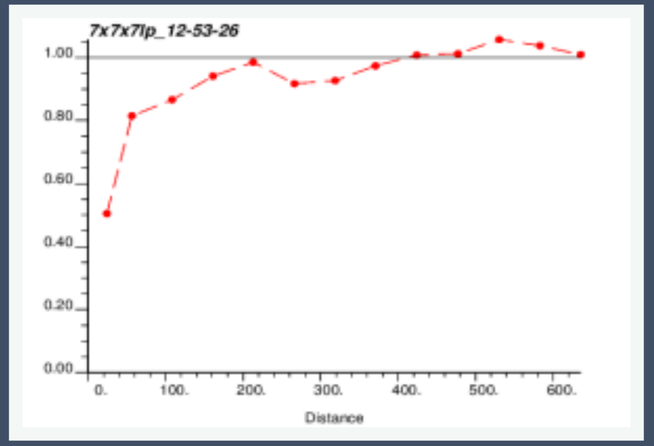
Case study



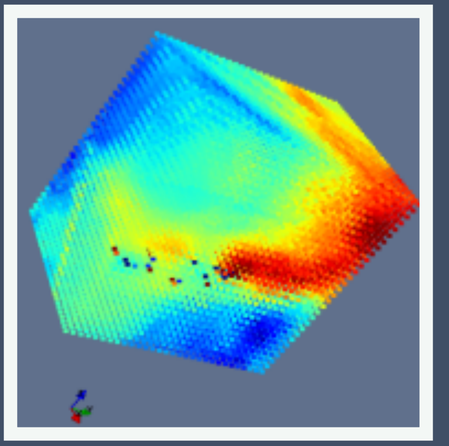
Drillholes
(800 gold grade samples)



DXF lines
LVA field (30 x 30 x 30)



Semi Variogram

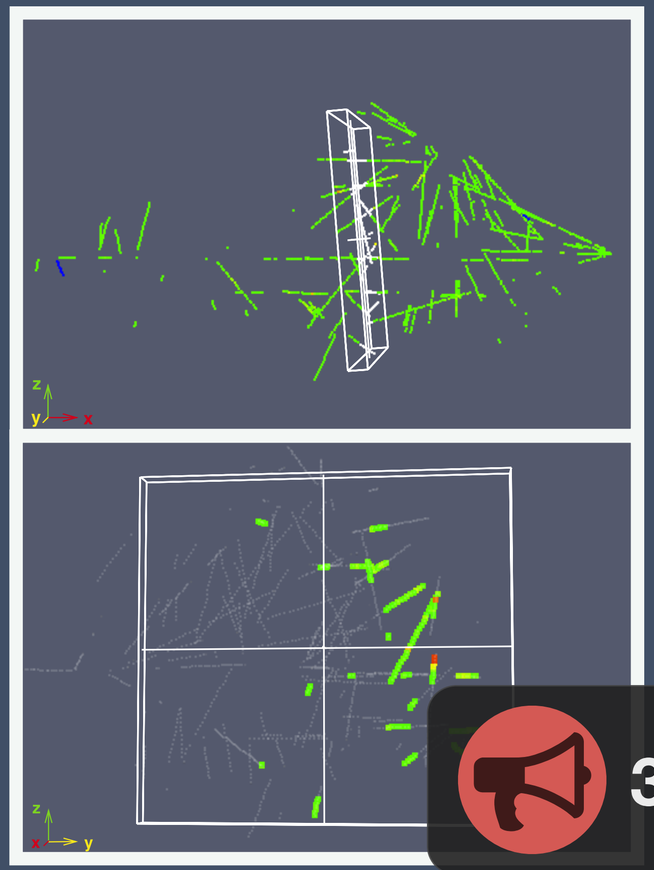


Simple Kriging

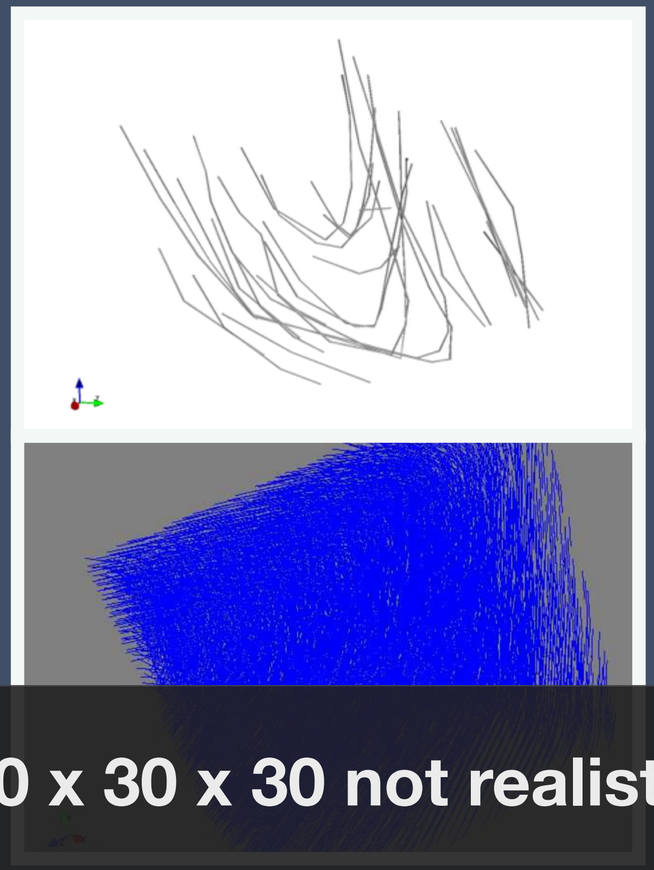
offset = 2
Landmark grid (7 × 7 × 7)
Estimation grid (30 x 30 x 30)

NOTE: *Carlin type deposit containing gold, copper, molybdenum, zinc, iron and lead associated with dolomitic skarns*

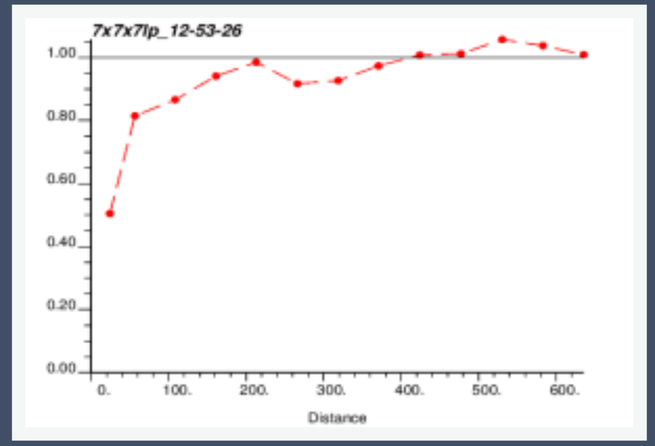
Case study



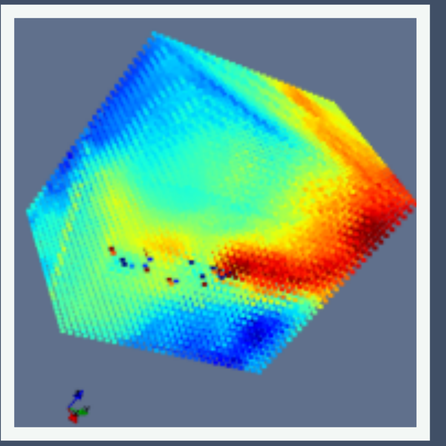
Drillholes
(800 gold grade samples)



DXF lines
LVA field (30 x 30 x 30)



Semi Variogram



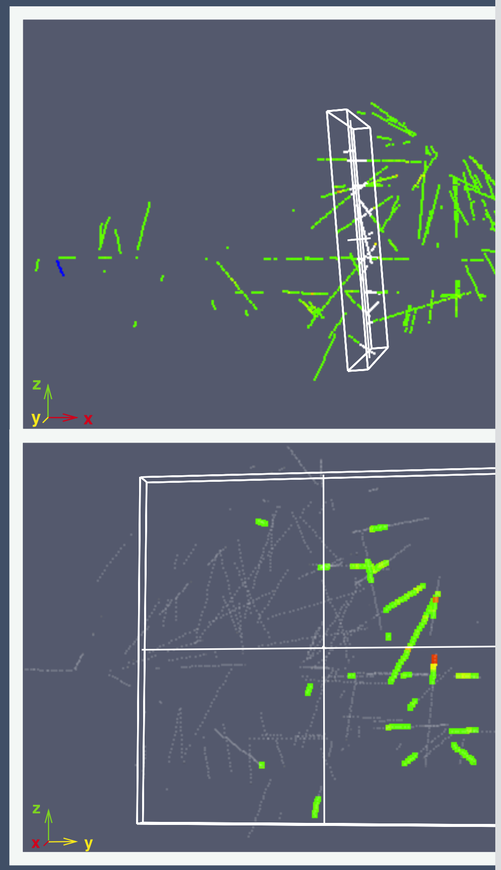
Simple Kriging

offset = 2
Landmark grid (7 x 7 x 7)
Estimation grid (30 x 30 x 30)

30 x 30 x 30 not realistic. Can we scale to larger grids?

NOTE: *Carlin type deposit containing gold, copper, molybdenum, zinc, iron and lead associated with dolomitic skarns*

Case study



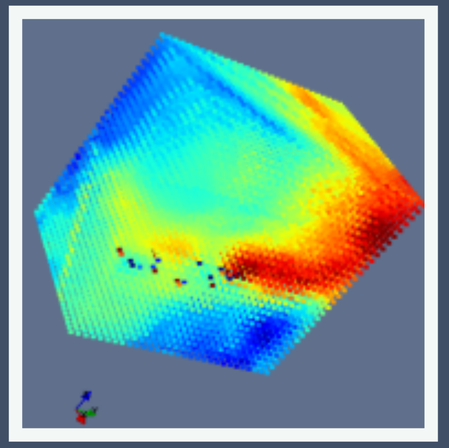
Drillholes
(800 gold grade samp

Shortest path calculations
(Dijkstra)

Sequential	Sequential
14 GB	34 GB
1900 s	9300 s
120^3 grid	160^3 grid

** time in seconds*

offset = 2
Landmark grid (7 × 7 × 7)

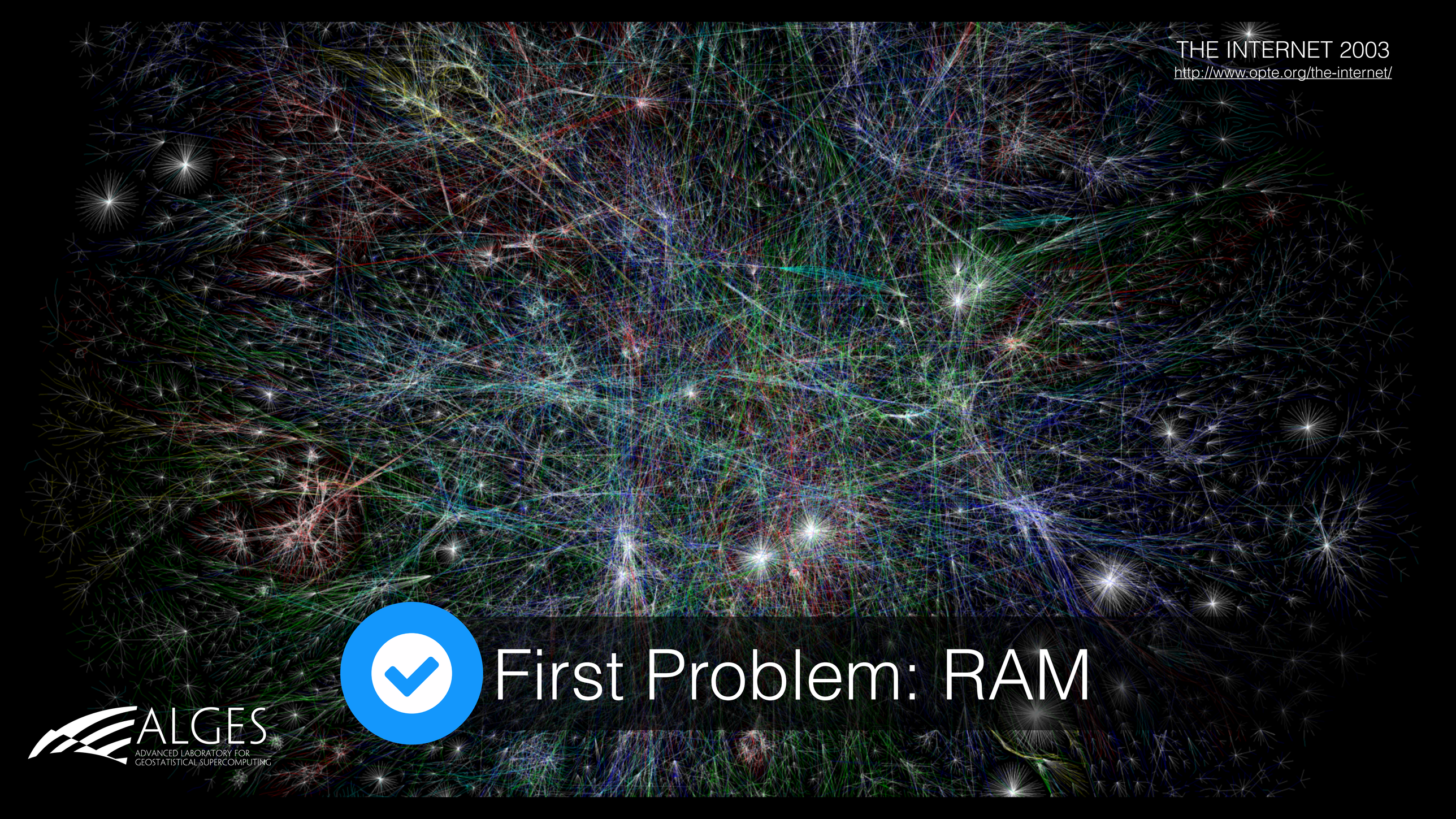


Simple Kriging

× 7 × 7)
0 x 30 x 30)

Dimensionality curse: larger grids require *too much* **RAM**

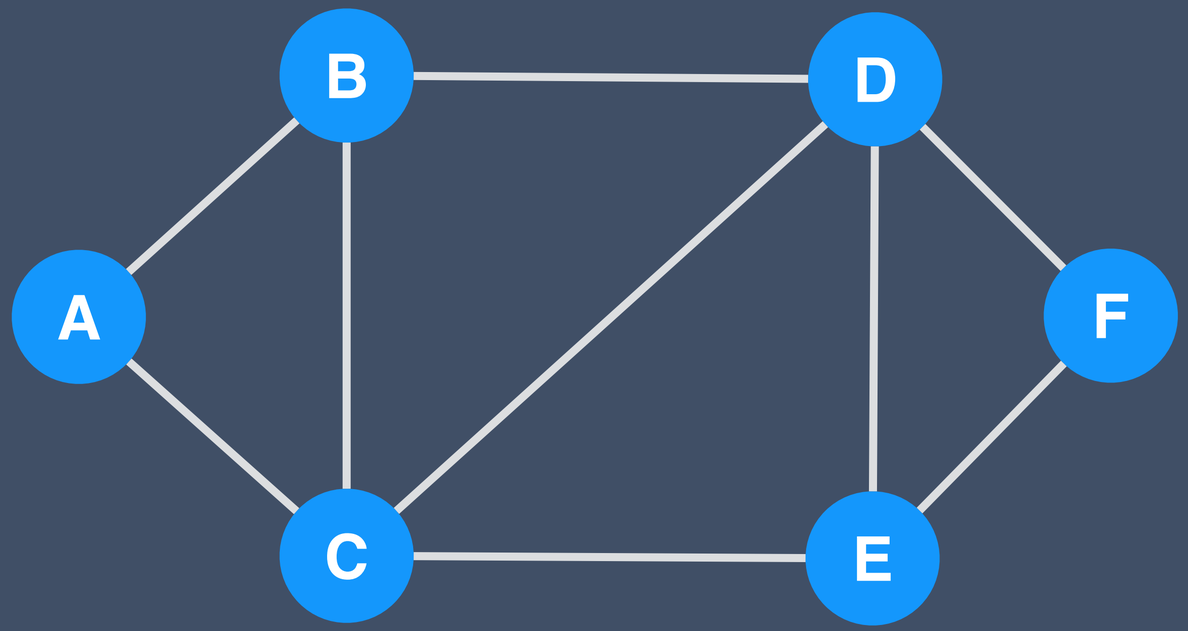
NOTE: *Carlin type deposit containing gold, copper, molybdenum, zinc, iron and lead associated with dolomitic skarns*



First Problem: RAM

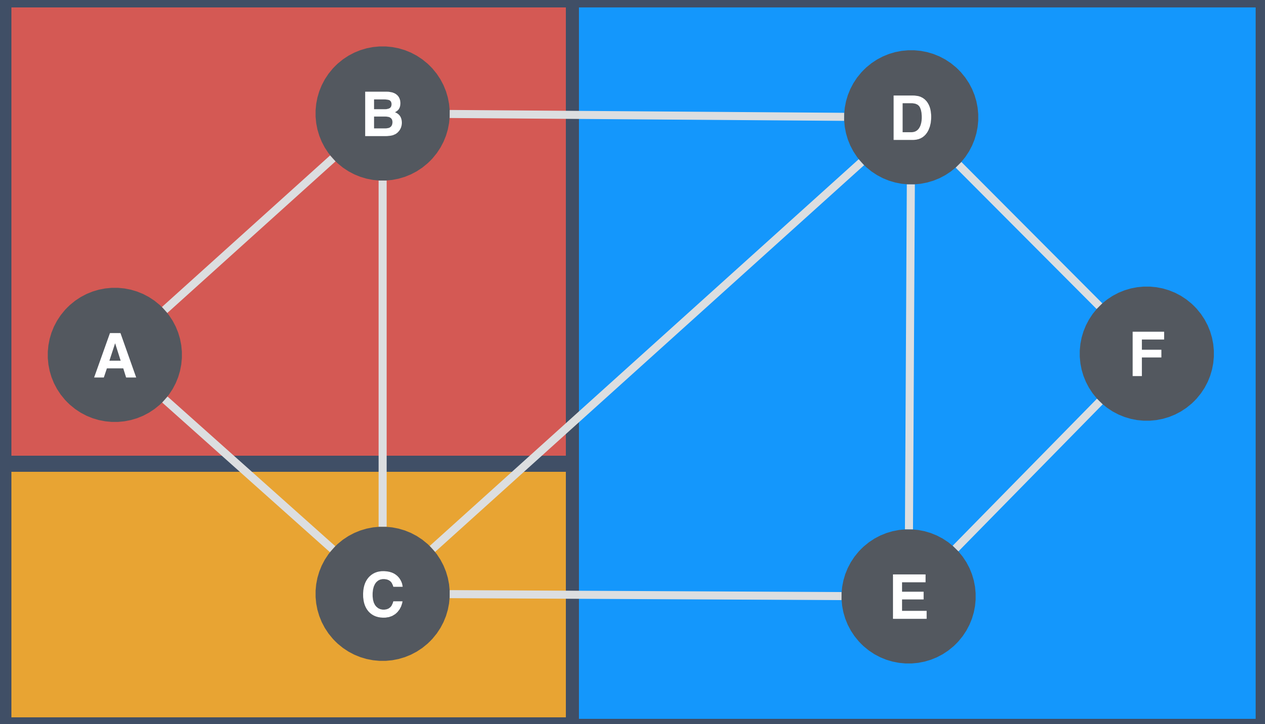
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Distributed graph



- When working with large estimation grids **huge** graphs must be stored in memory
- LVA methods generates *undirected graphs*
- A grid of *120**x**120**x**120 uses* more than 41GB (offset = 3)

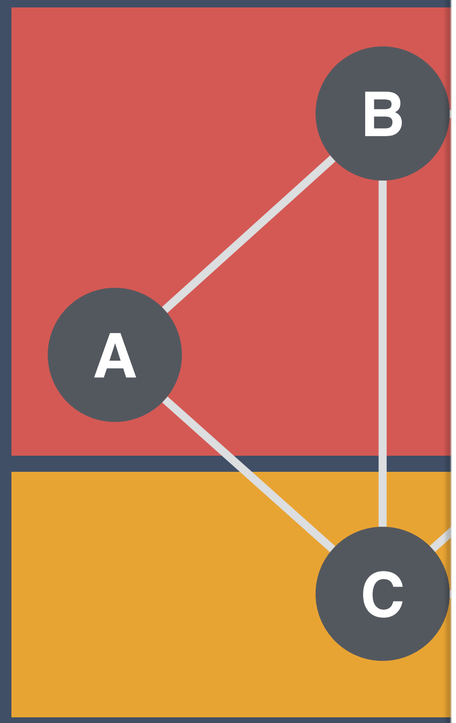
Distributed graph



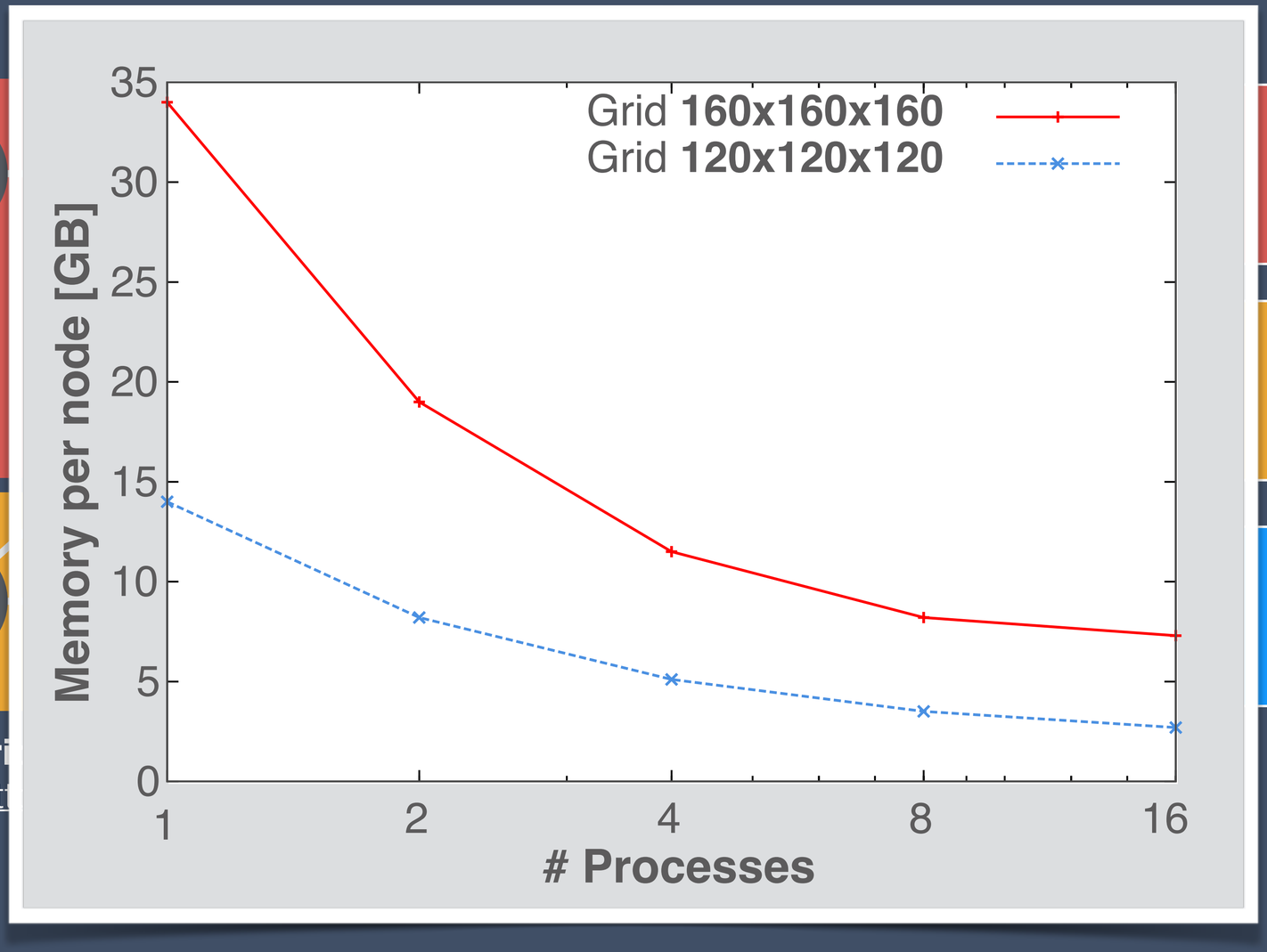
processor 0	vert: A, B edges: 4
processor 1	vert: C edges: 2
processor 2	vert: D,F,E edges: 3

Distribution using *Parallel Boost Graph Library*
http://www.boost.org/libs/graph/doc/index.html

Distributed graph



Distri
htt



vert: A, B
edges: 4

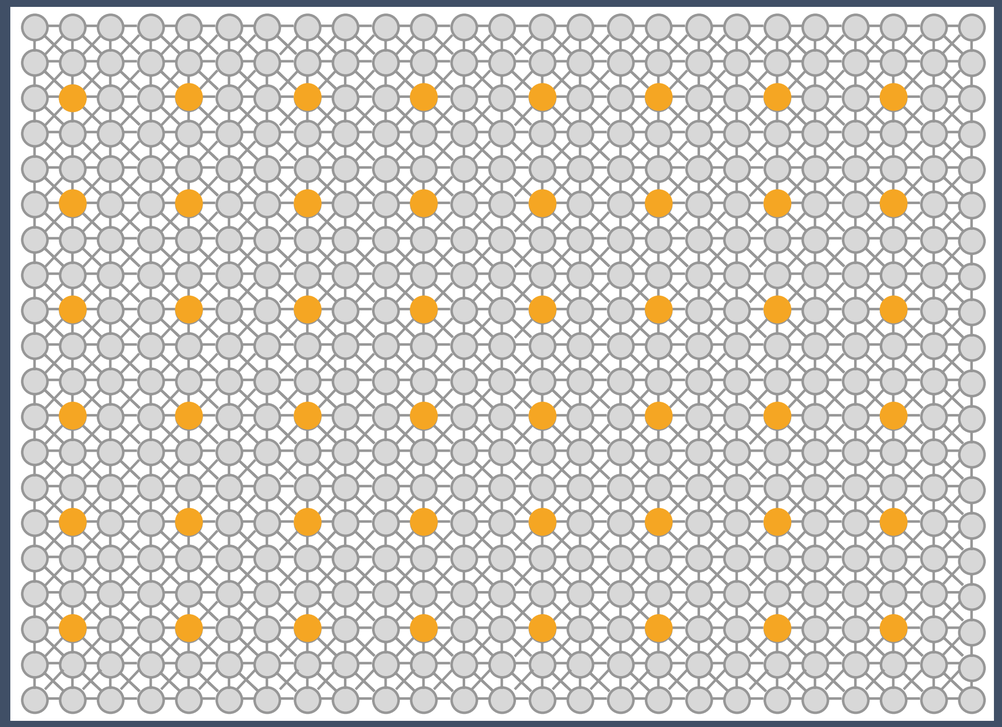
vert: C
edges: 2

vert: D,F,E
edges: 3

Second problem: CPU

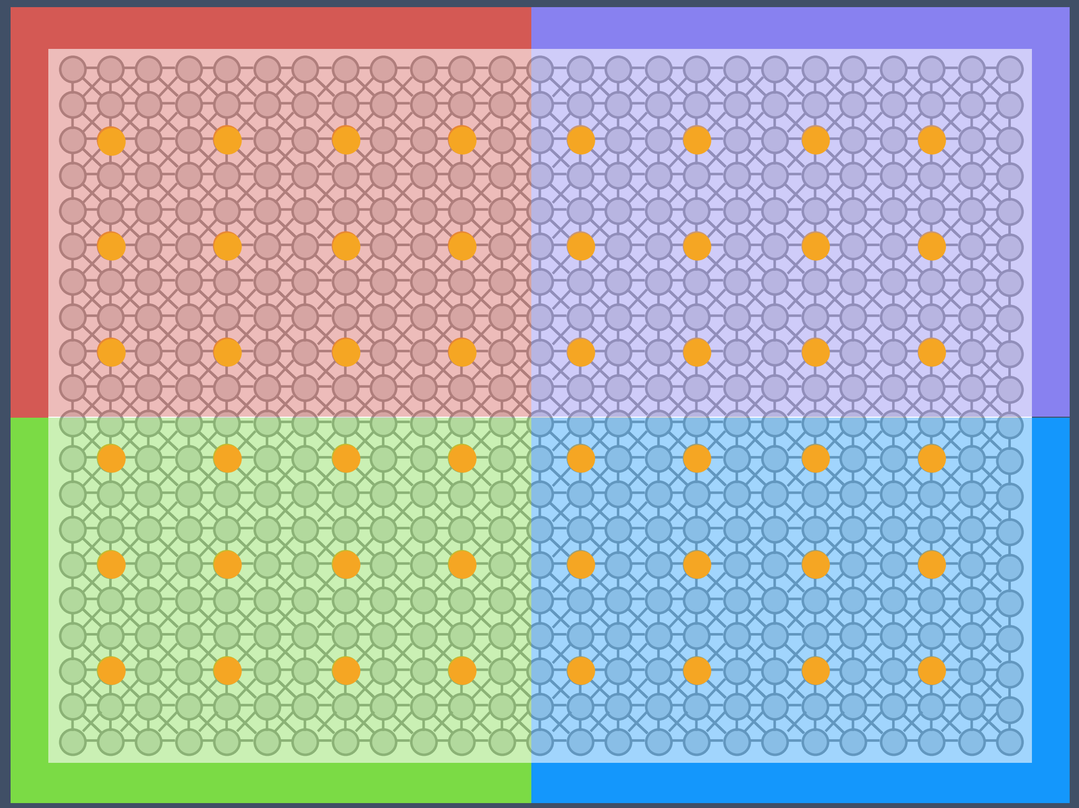
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Parallel Dijkstra runs



- For each landmark point, we must run a single source shortest path calculation
- Every calculation can be executed in parallel by different processes
- Each run can use the distributed graph explained before

Parallel Dijkstra runs



```
set 0: 12 Dijkstra runs
```

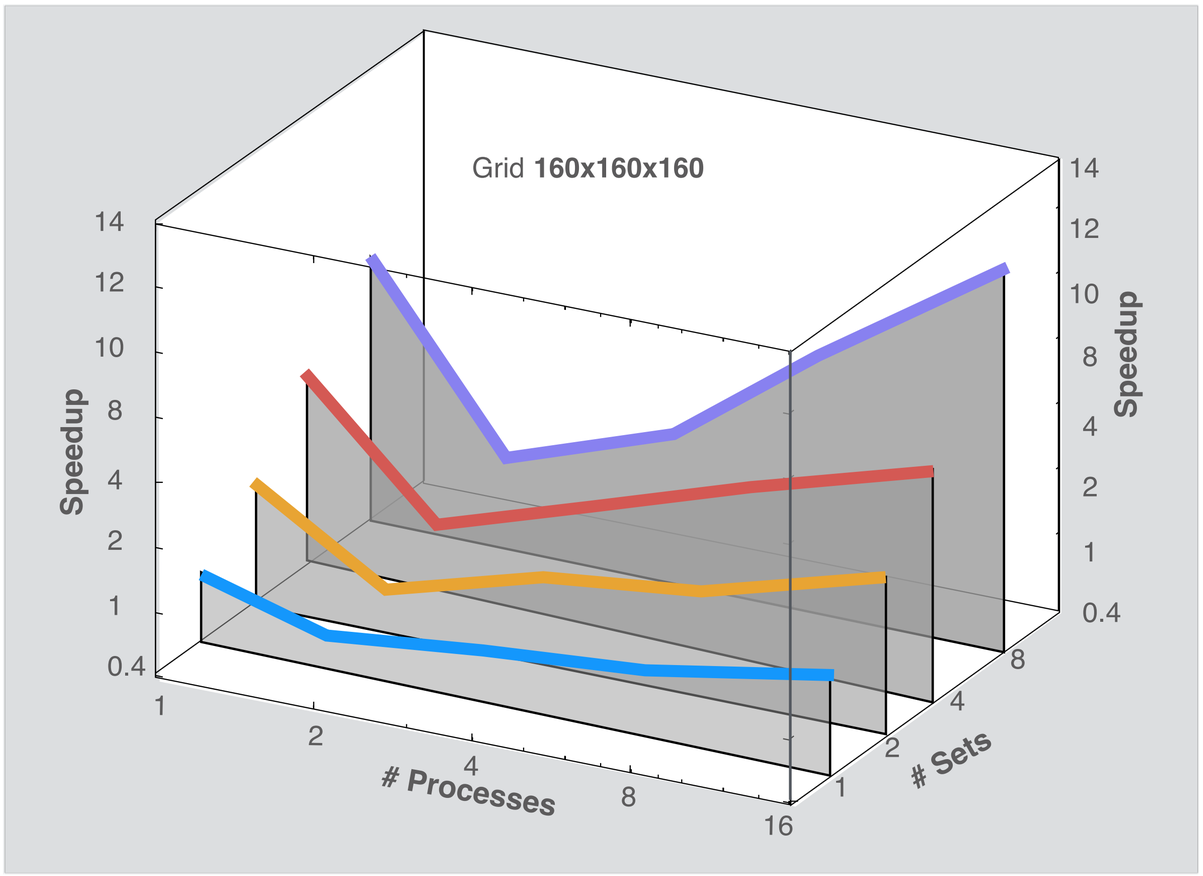
```
set 1: 12 Dijkstra runs
```

```
set 2: 12 Dijkstra runs
```

```
set 3: 12 Dijkstra runs
```

Distribution using *Parallel Boost Graph Library +* Parallel *MPI* Executions

Parallel Dijkstra runs



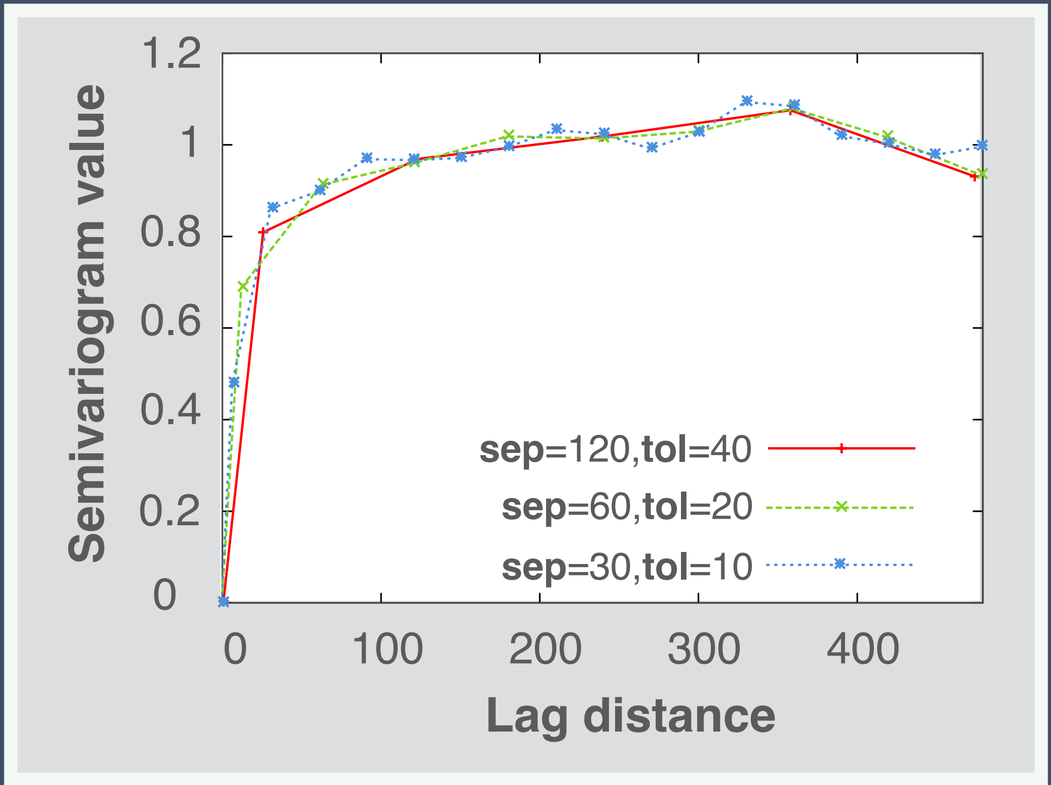
#Processes / # Sets	1	2	4	8	16
1	1x	0.44x	0.67x	1.07x	1.49x
2	2x	0.88x	1.34x	2.13x	2.97x
4	4x	1.73x	2.68x	4.27x	5.96x
8	7.5x	3.29x	5.03x	8.04x	11.19x

Distribution using *Parallel Boost Graph Library* + Parallel *MPI* Executions

Grid **160x160x160**

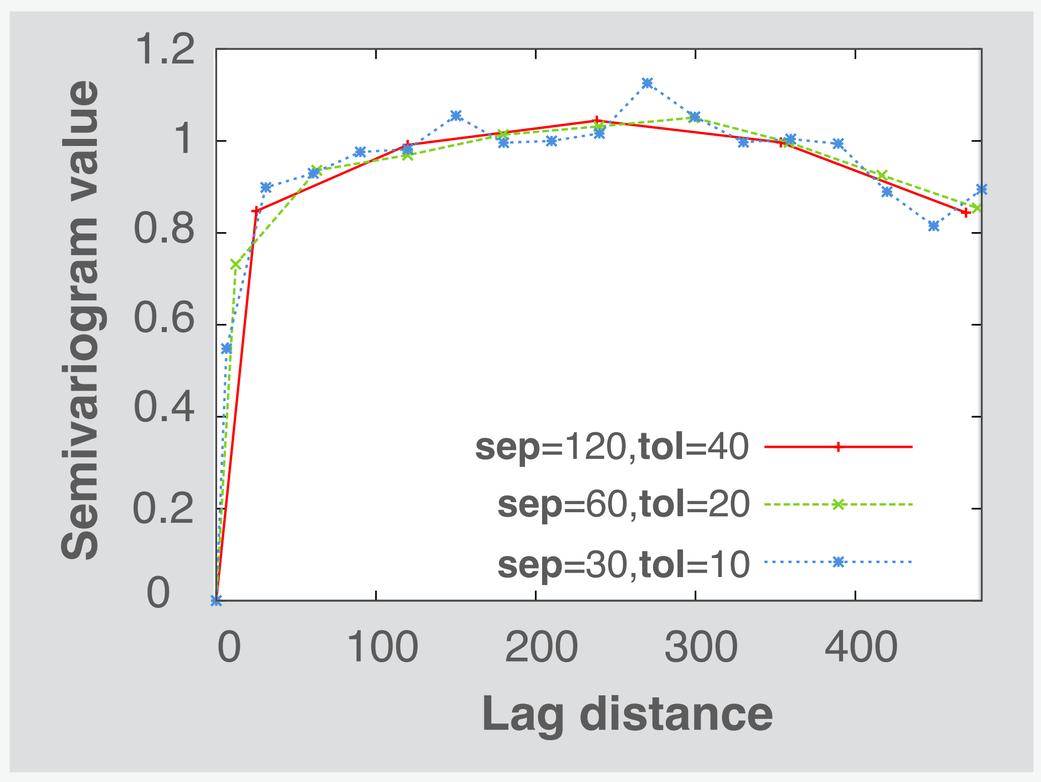
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Results



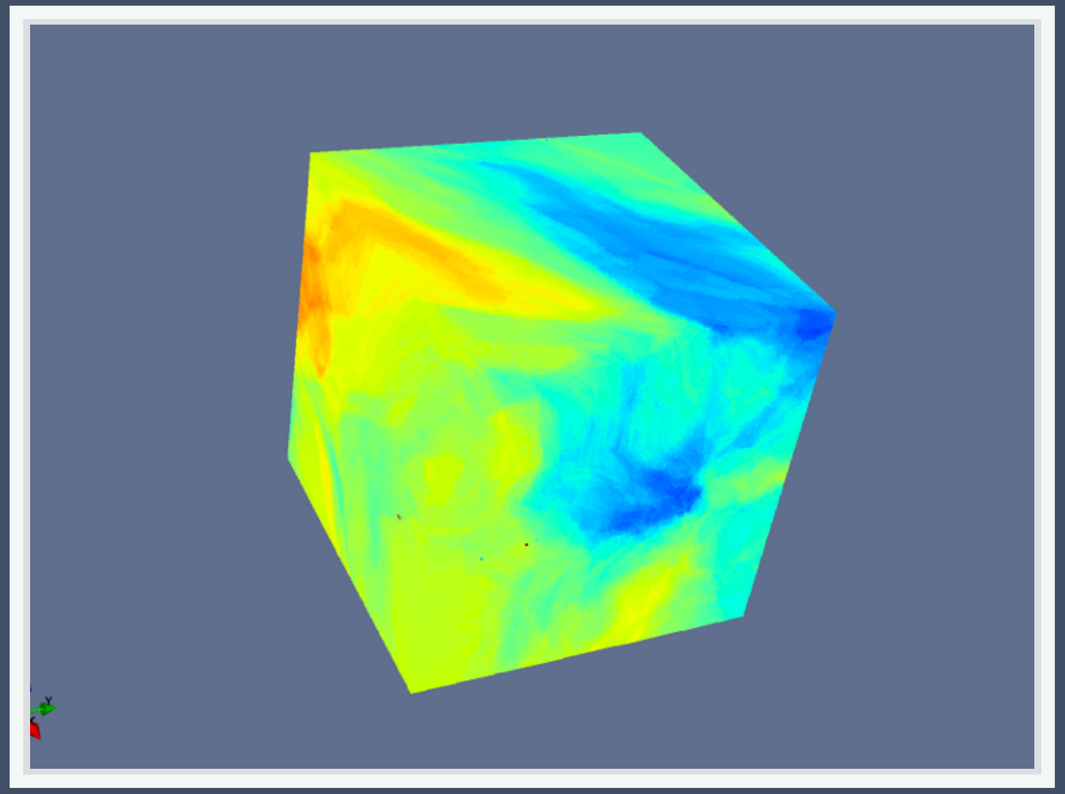
Estimation grid of
1,728,000 nodes
(120 × 120 × 120)

Semi Variogram
`offset = 2`
`landmark = 7x7x7`



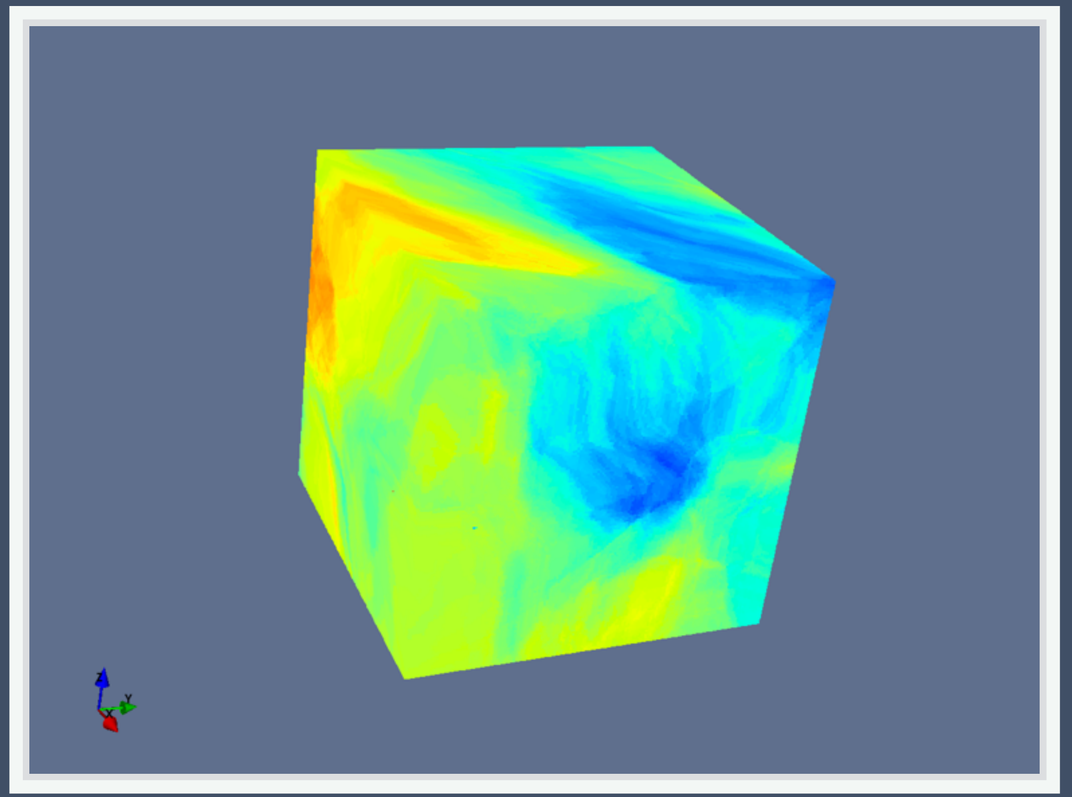
Estimation grid of
4,096,000 nodes
(160 × 160 × 160)

Results



Estimation grid of
1,728,000 nodes
(120 × 120 × 120)

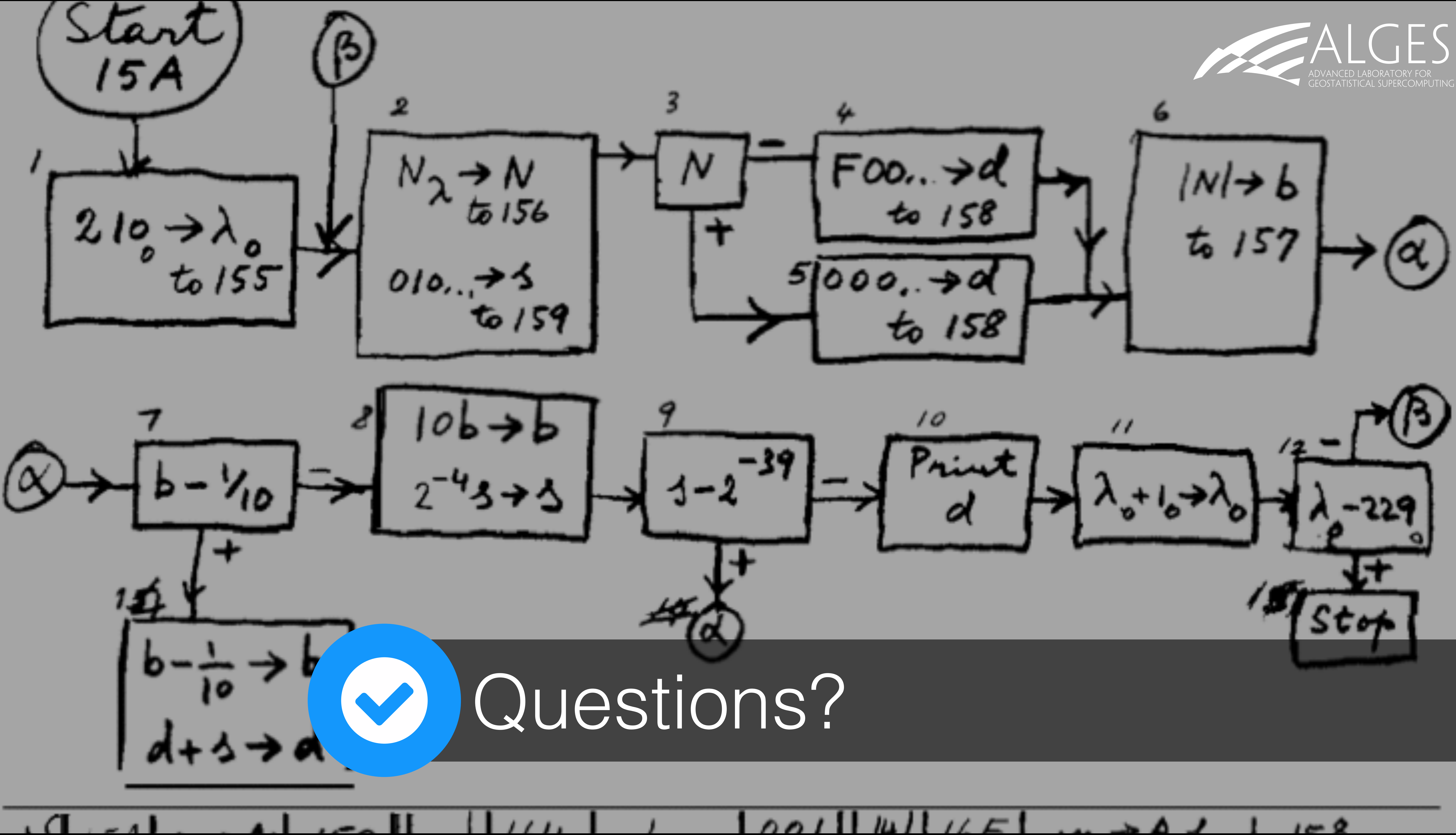
Simple Kriging
`offset = 2`
`landmark = 7x7x7`



Estimation grid of
4,096,000 nodes
(160 × 160 × 160)

Conclusions

- The scope and possibilities that the parallel approach can bring in the LVA context are related with the size of the problems to be solved
- Due to the possibility to distribute the memory usage among many computational nodes, very large graphs can be stored in memory in order to be processed by all workers
- As a future work, a new version of the original LVA codes will be developed, enabling the parallel processing capabilities of the current hardware


Start
15A
β
210₀ → λ₀
to 155
N_λ → N
to 156
010.. → s
to 159
N
F00.. → d
to 158
000.. → d
to 158
|N| → b
to 157
α
b − 1/10
10b → b
2^-4 s → s
s − 2^-39
Print
d
λ₀ + 1₀ → λ₀
λ₀ − 229₀
Stop
b − 1/10 → b
d + s → d

Questions?

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Incorporating distributed Dijkstra's algorithm into variogram calculation with locally varying anisotropy

Oscar Peredo, Felipe Navarro, Mauricio Garrido and Julián M. Ortiz

Department of Mining Engineering – FCFM – Universidad de Chile
Advanced Mining Technology Center – Universidad de Chile

jortiz@ing.uchile.cl

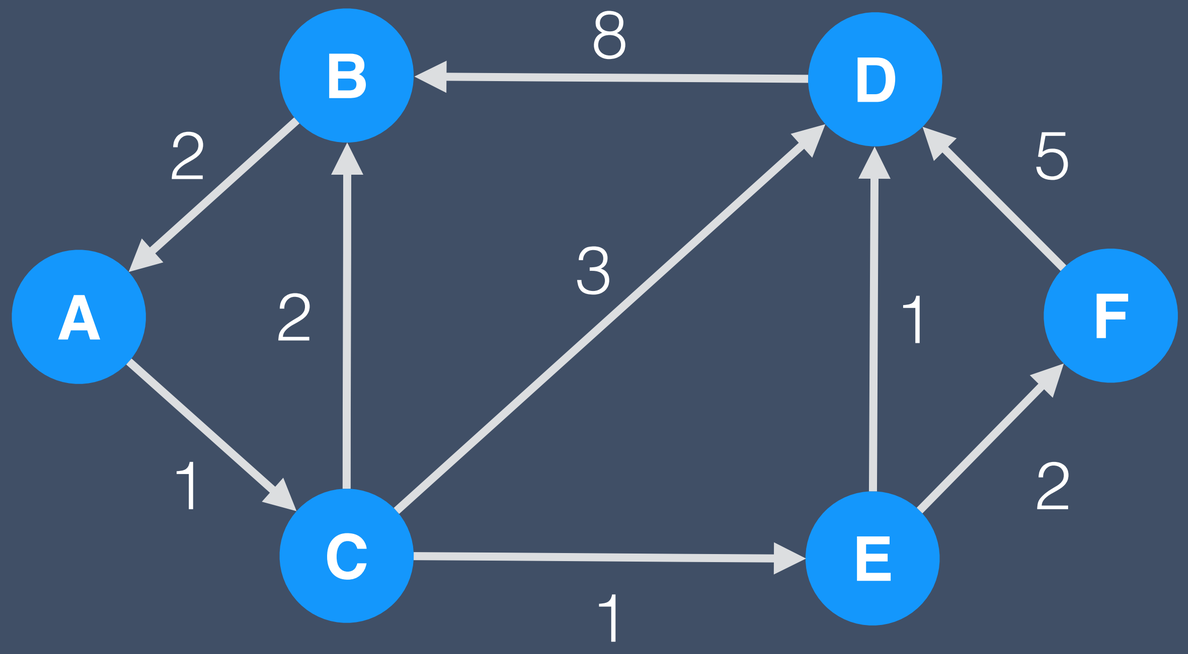
37th APCOM Conference
Fairbanks, Alaska 2015

Ingeniería de Minas
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

amtc
ADVANCED MINING TECHNOLOGY CENTER

Dijkstra calculation

ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING

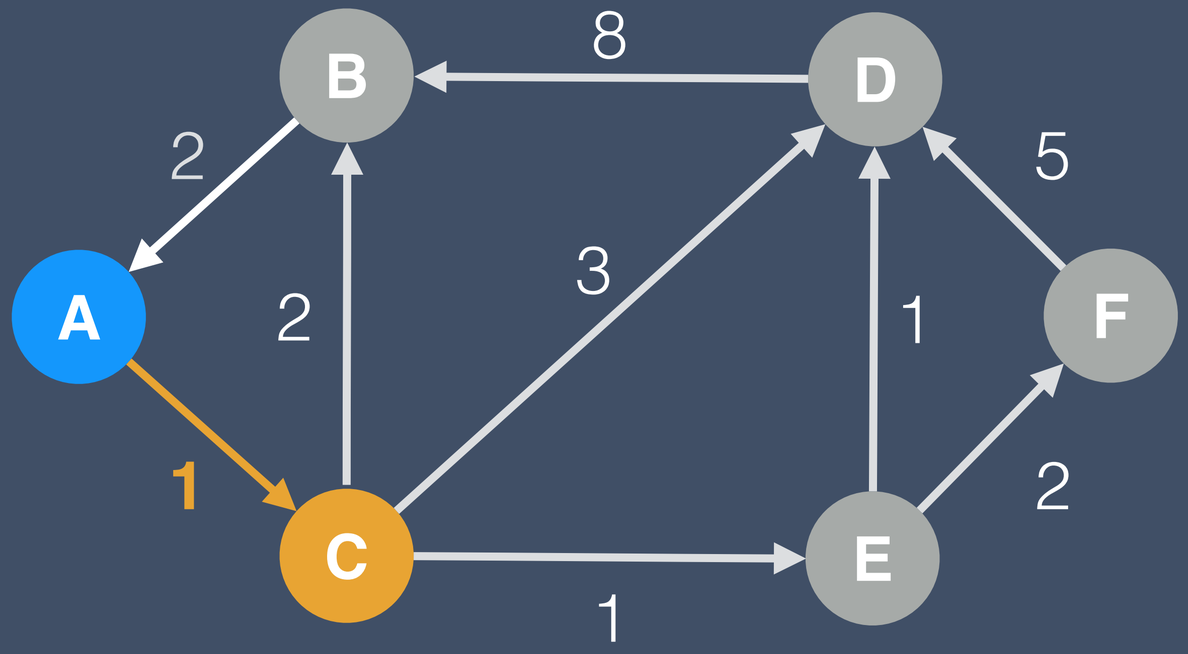


total cost

From A	**B**	**C**	**D**	**E**	**F**

destination

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

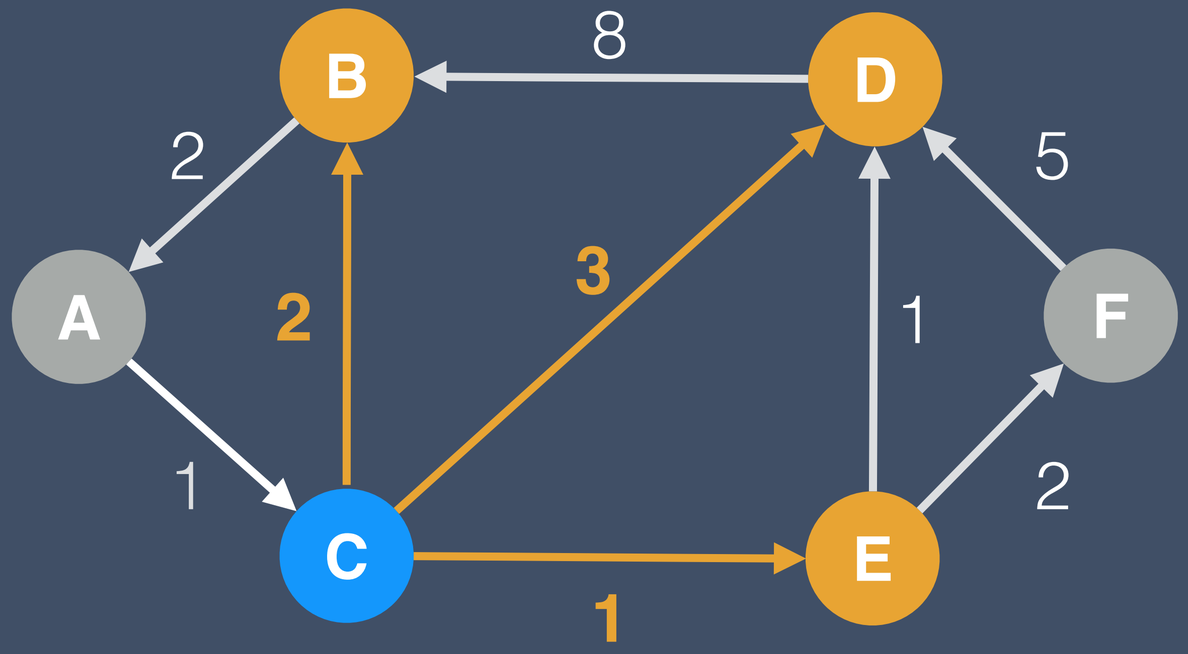


total cost

From A	**B**	**C**	**D**	**E**	**F**
A	∞	**1**	∞	∞	∞

destination

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

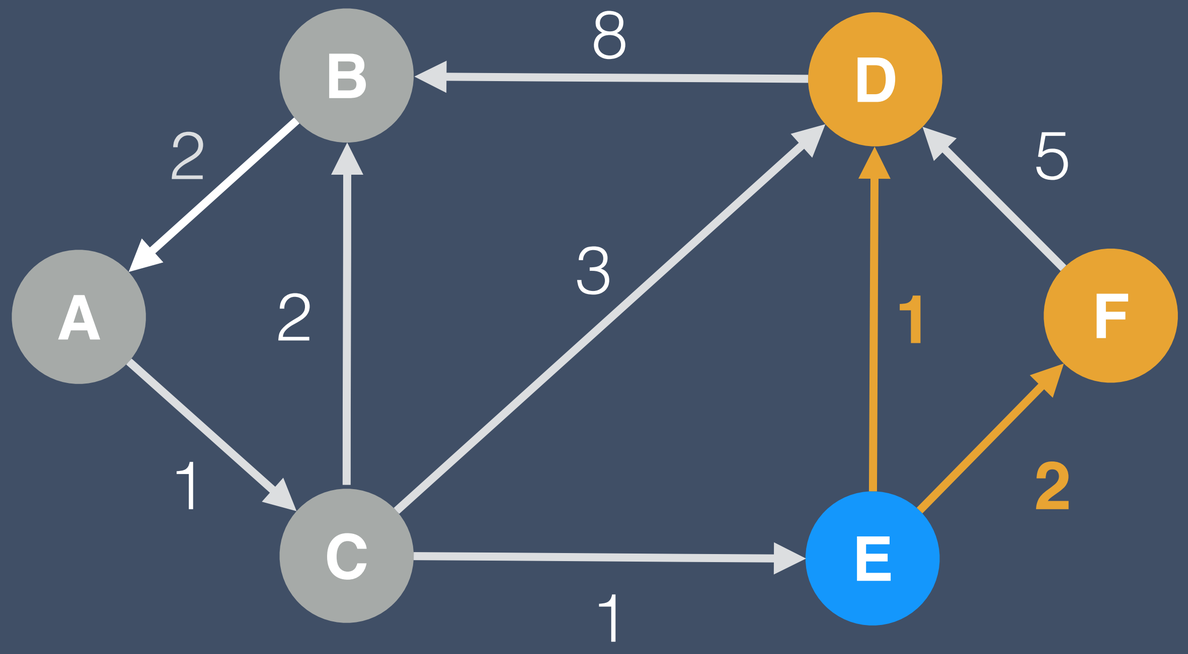


total cost

From A	**B**	**C**	**D**	**E**	**F**
A	∞	1	∞	∞	∞
C	3	1	4	2	∞

destination

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

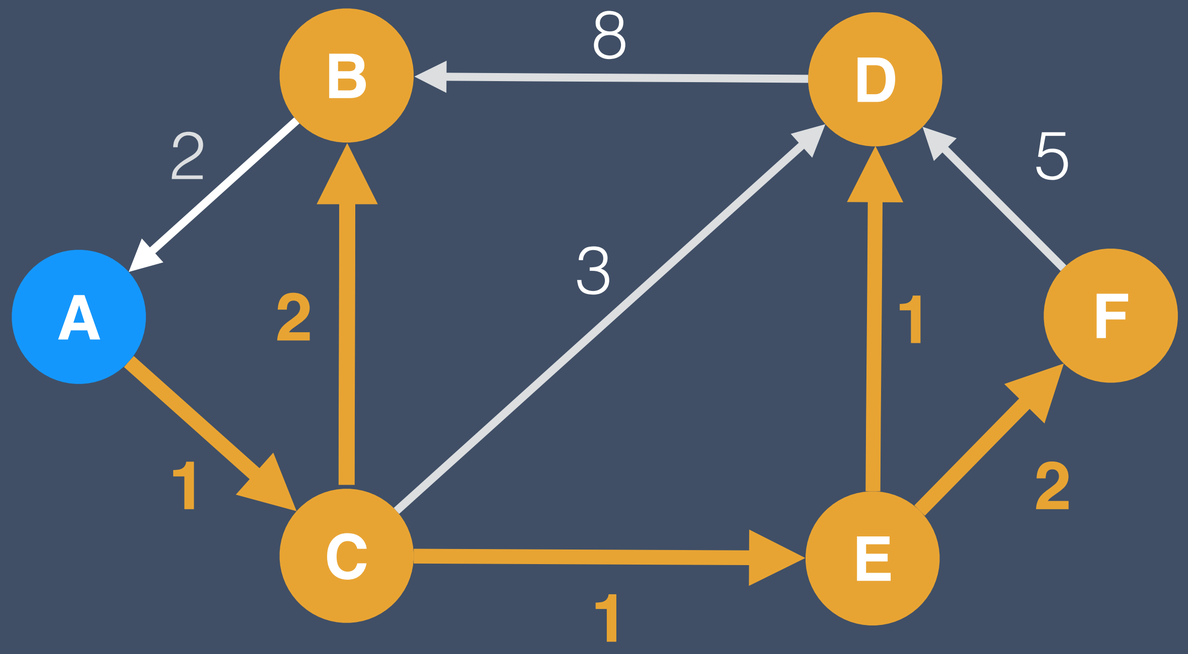


total cost

From A	**B**	**C**	**D**	**E**	**F**
A	∞	1	∞	∞	∞
C	3	1	4	2	∞
E	3	1	3	2	4

destination

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING



total cost

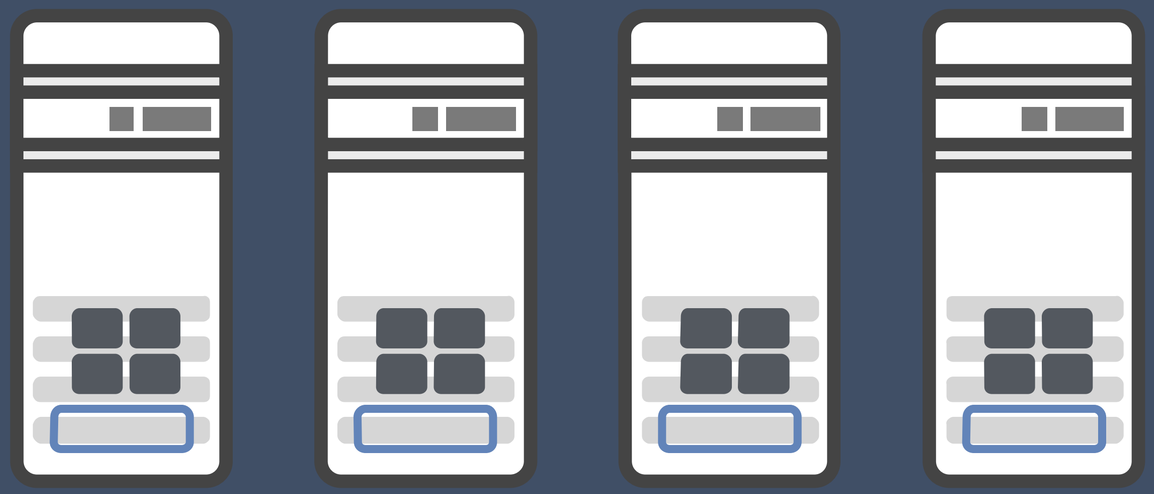
From A	**B**	**C**	**D**	**E**	**F**
destination					
A	∞	1	∞	∞	∞
C	3	1	4	2	∞
E	3	1	3	2	4

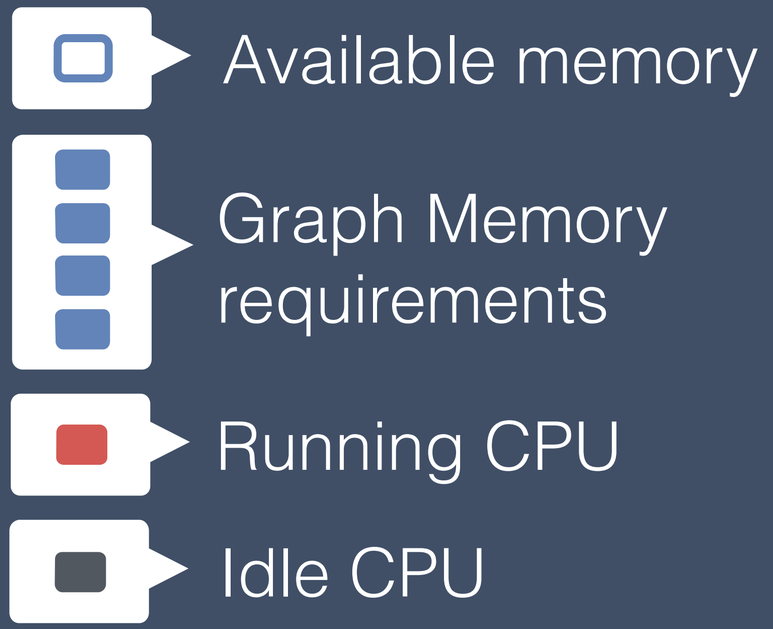
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

How does it work?

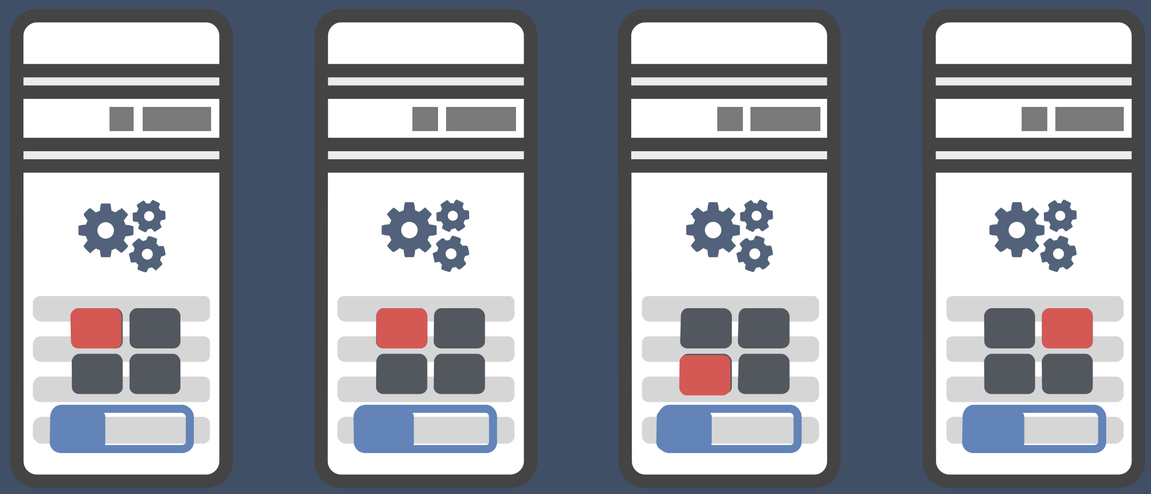
ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING

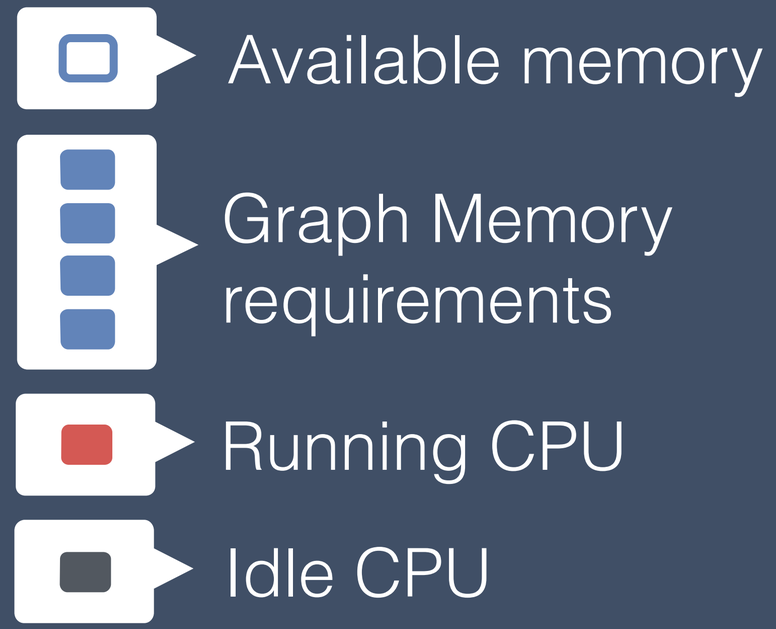
How does it work?





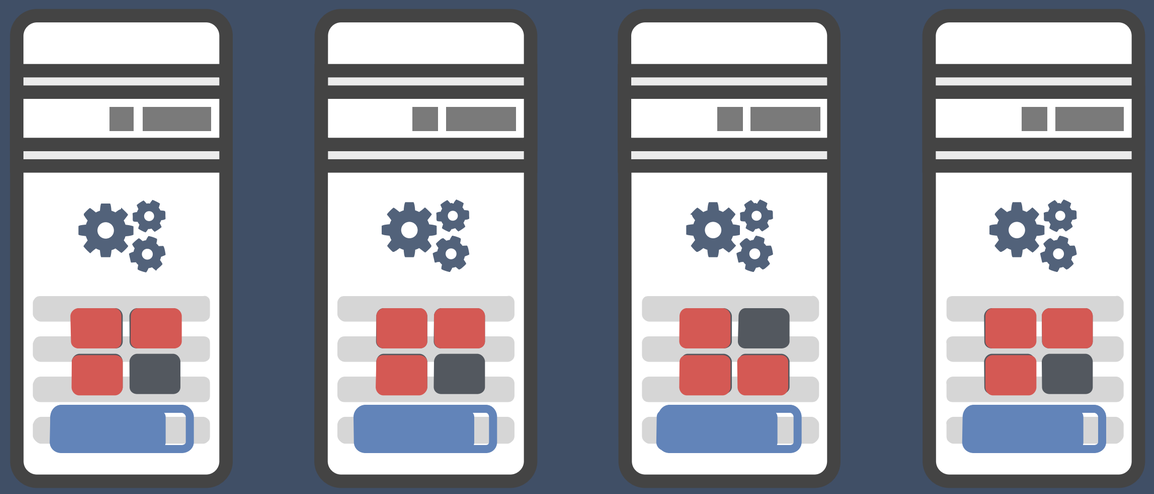
How does it work?

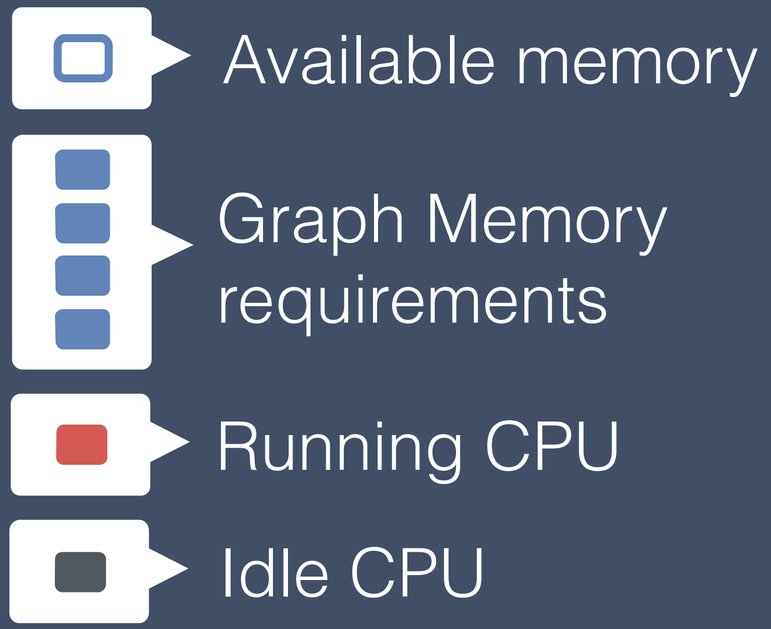




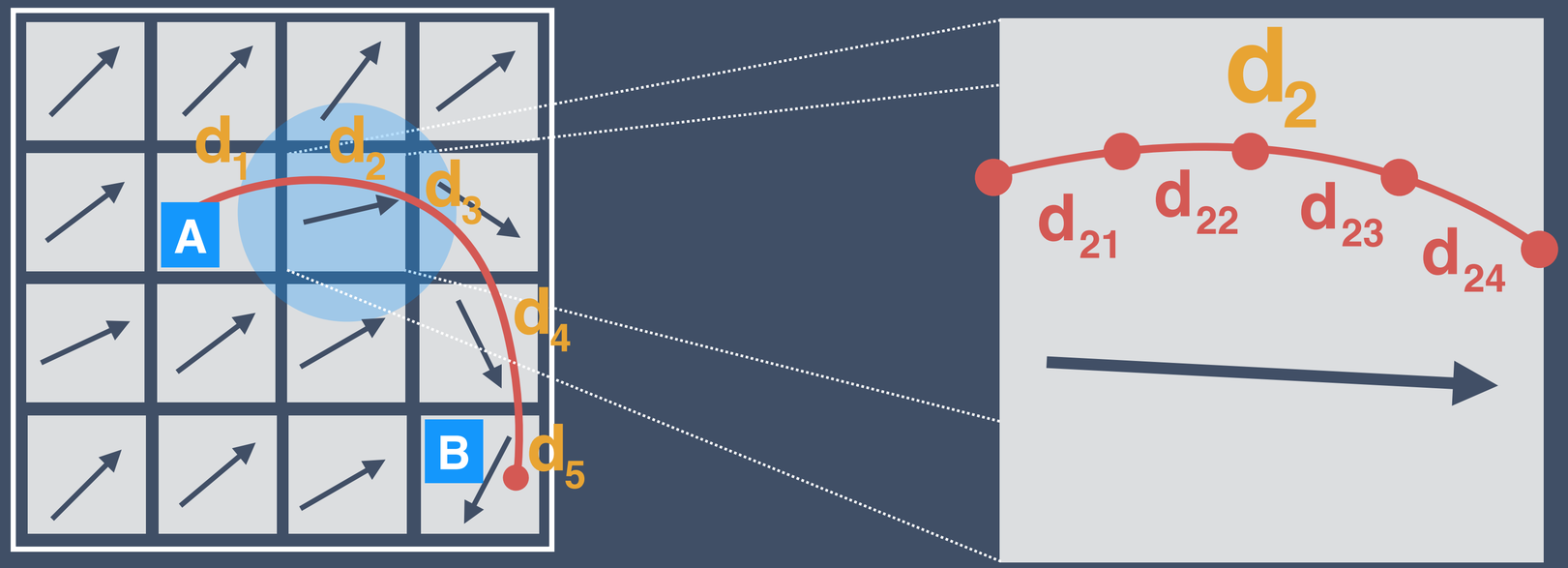
The graph is stored in 4 distributed machines

How does it work?





3 sets of landmark points are being processed in parallel



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING