High Performance Computing and Information Theory with D-HYDRO 1D2D on Cloud Infrastructure

MSc Thesis Defense – Demi de Rijke

TU Delft

Why HPC & information theory in the cloud?

Hydrodynamic modelling is compute intensive and can often not be ran locally → **computing cluster**

Cloud platforms offer utility computing in a pay-as-you manner

How to efficiently make use of this?

The computation time decreases and the costs to perform computations become explicit.

How to combine these parameters with the quantification of the obtained information to make a trade-off of the model configurations?

TU Delft

What is information?

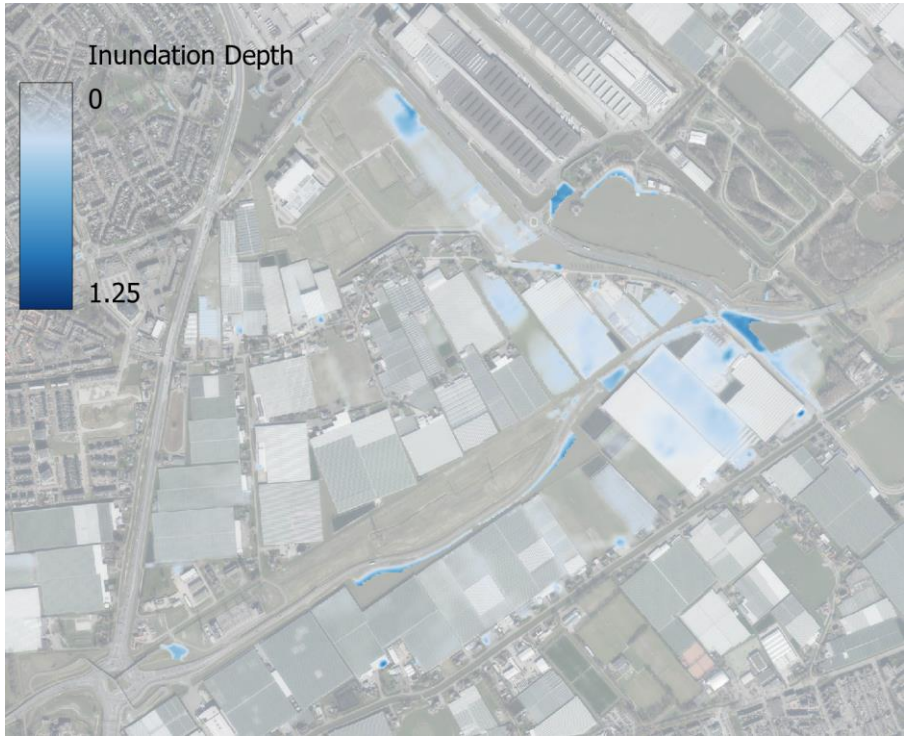
The weather forecast, *with Shannon's Entropy*

$$H(x) = -\sum_{i=1}^{n} p(x) \,.\log\big(p(x)\big)$$

TU Delft

How does this relate to inundation maps?

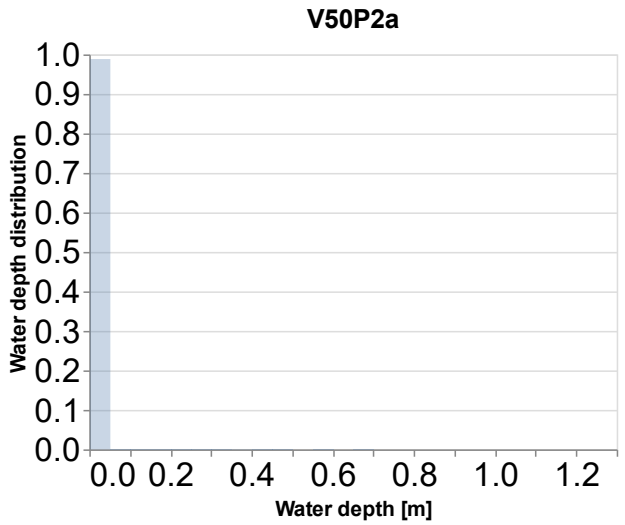
V50P0

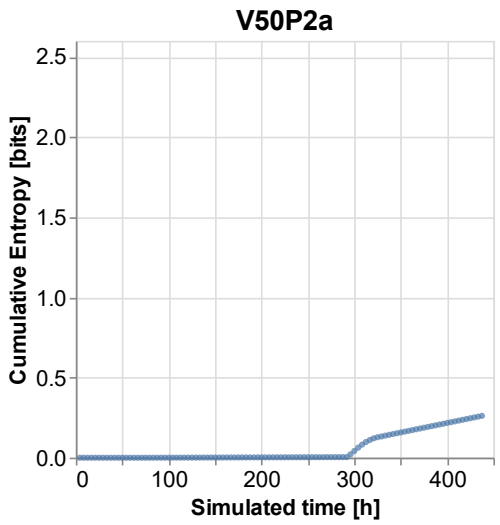


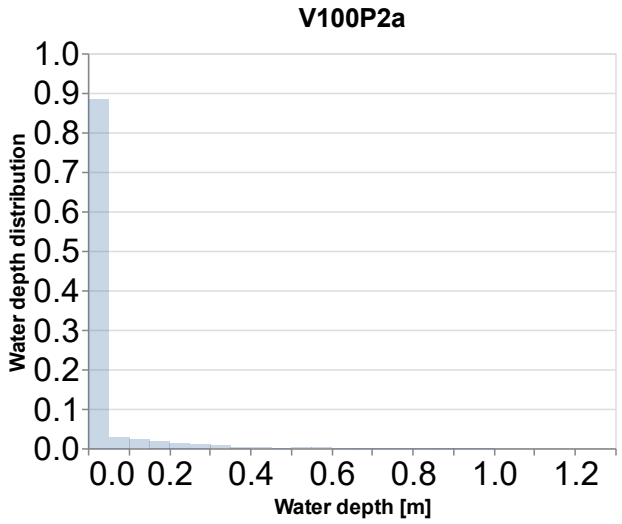
V100P2a

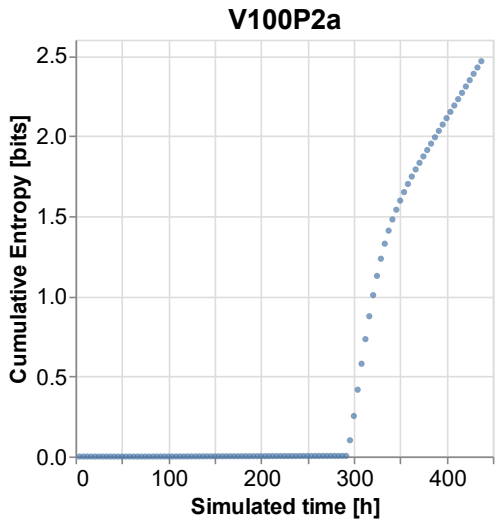
TU Delft

How has Shannon's Entropy been applied?









TU Delft

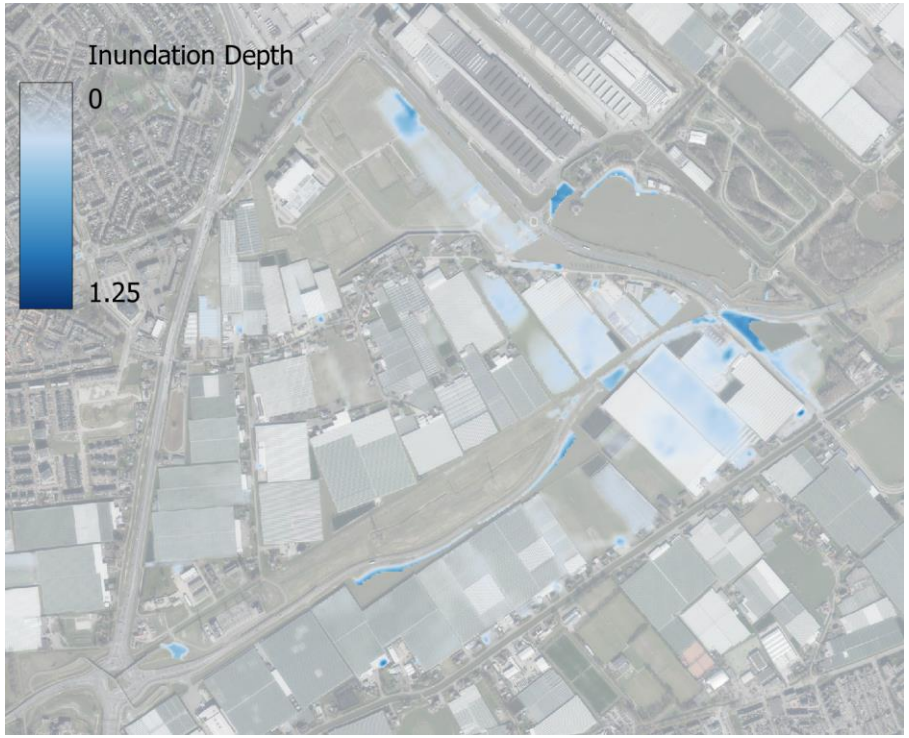
What does(n’t) Shannon’s Entropy tell us?

It expresses the quantity of information, not the quality!

TU Delft

Now we defined how to quantify information, let's produce it!

Water System Analysis for the Vlietpolder



- Norms for regional water systems (NBW)
- Rain showers come in different shapes and sizes
  - Volume → How much?
  - Pattern → How is it volume distributed in time?
  - 133 events for fine and course resolution
- Get the return period of the inundation levels

TU Delft

What is the cloud? What can(‘t) we do with it?

Possibility

- Elasticity

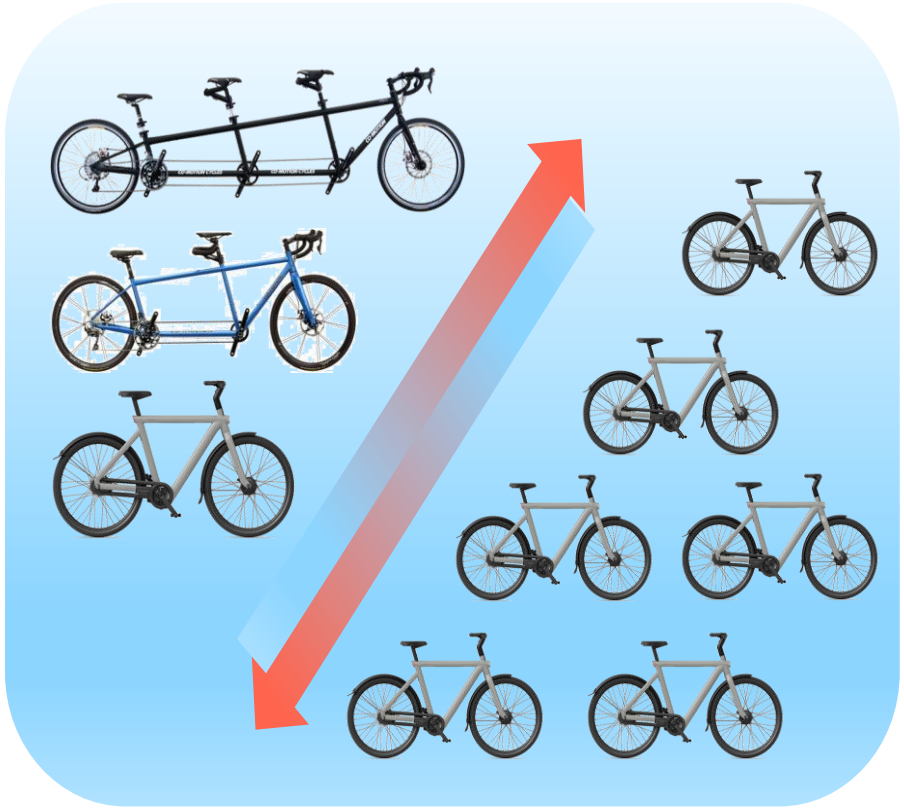
Limitations

- Communication
- Required expertise

What does the elasticity look like?

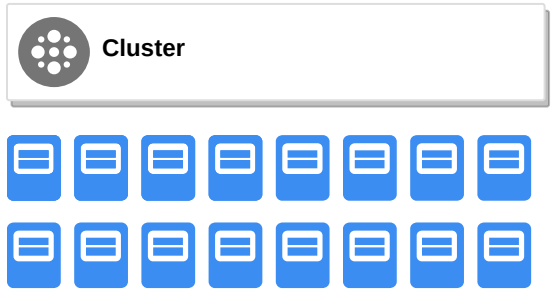
Vertical scaling
Adjust the capacity of an individual node: vCPU, RAM, etc.

Horizontal scaling
Increase or decrease the number of individual nodes



TU Delft

The Cluster Configuration

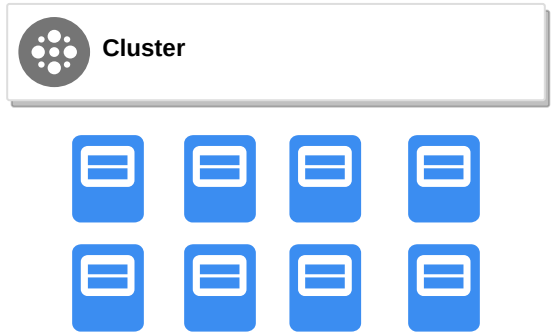


16x 4vCPU + 16 Gb M

Total =

64vCPU

256 Gb Memory

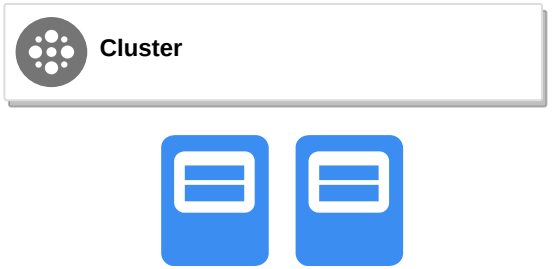


8x 8vCPU + 32 Gb M

Total =

64vCPU

256 Gb Memory

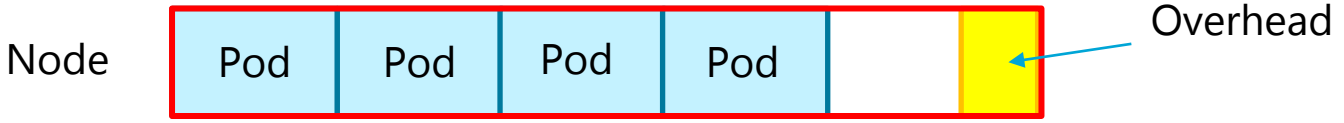


2x 32vCPU + 128 Gb M

Total =

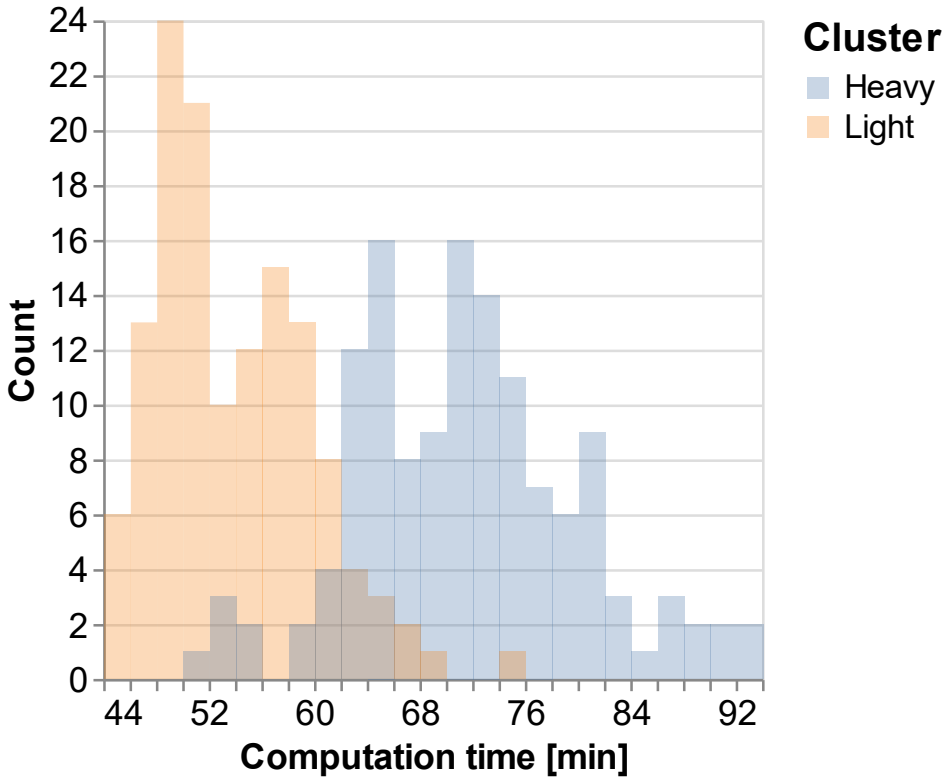
64vCPU

256 Gb Memory



TU Delft

Comparison of Computation Times



Cluster type	No. of nodes	vCPU	Memory	Simultaneous processes on nodes
Standard light	16	4	16	~2
Standard heavy	2	32	128	~20

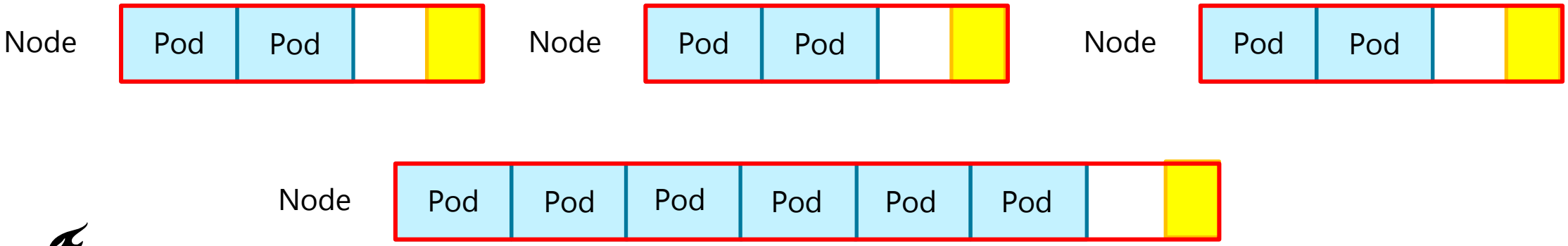
Why?

- Processes cannot take place simultaneously
- Too much network traffic
- Communication overhead

TU Delft

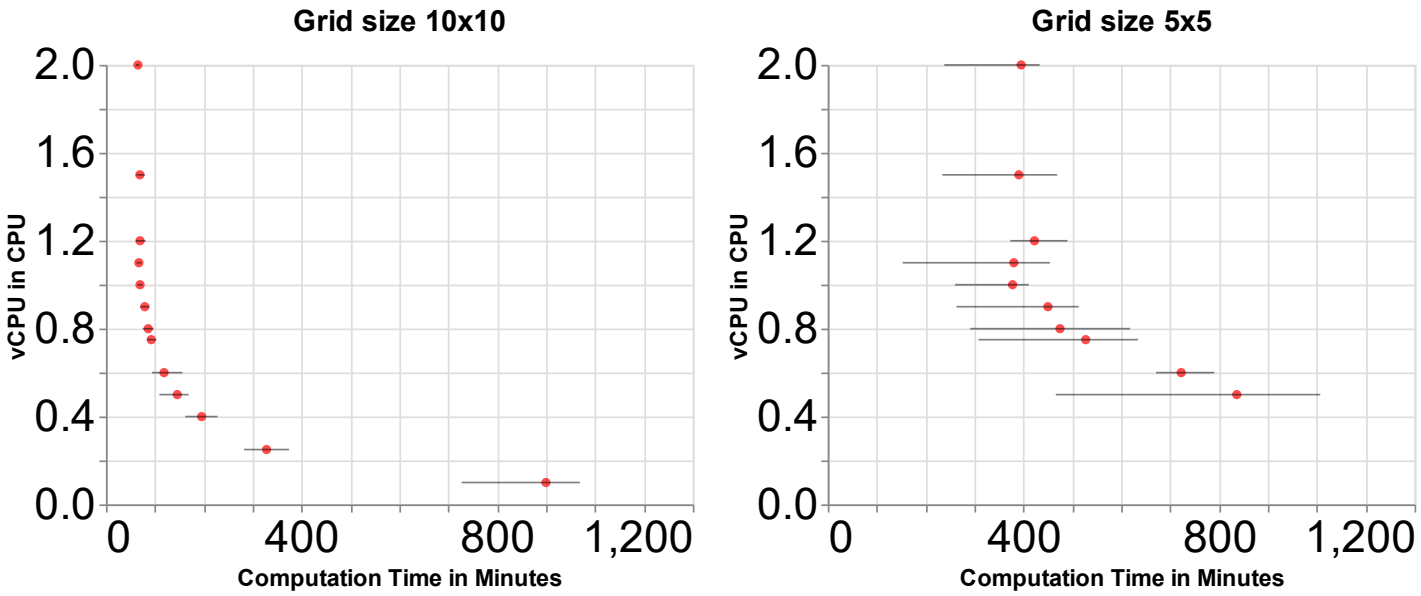
Influence on Computation Times

	Cluster light	Cluster medium	Cluster heavy
HPC time [hrs]	4:20	4:12	4:30
D-Hydro time [min]	7102	8247 (+16%)	9438 (+33%)



TU Delft

Resource Allocation of Pods



vCPU

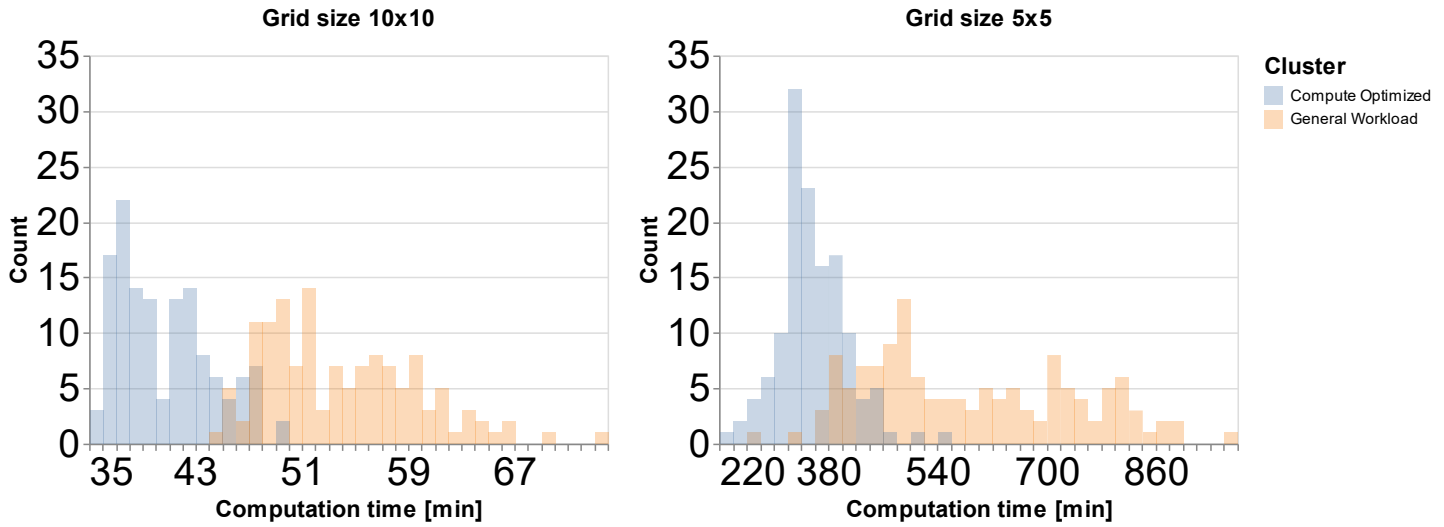
- More than 1 vCPU does not result in a speedup
- Interesting area between 1vCPU – 0.8vCPU to increase utilization

Memory

- Threshold value, with too little RAM computation is not performed

TU Delft

General workload vs. Compute optimized nodes

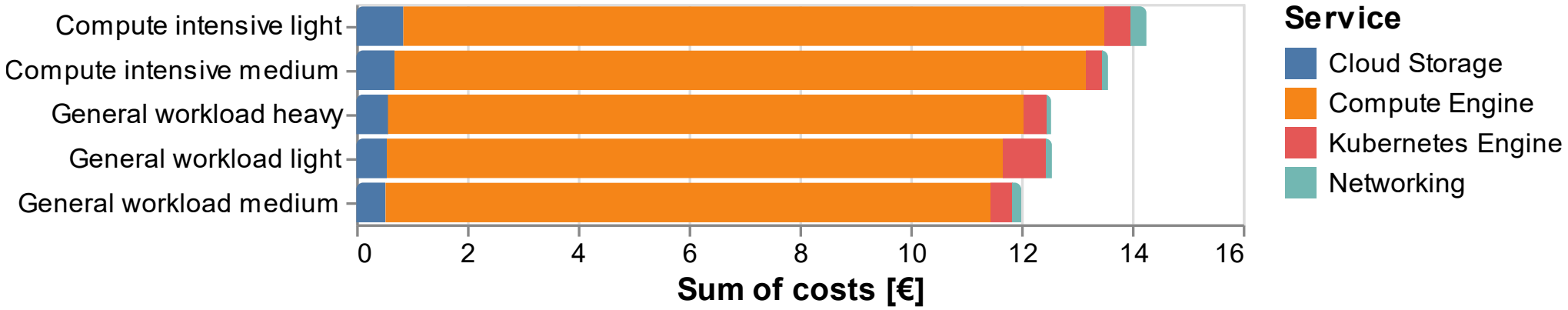


	10x10		5x5	
	Standard	Compute	Standard	Compute
HPC time [hrs]	4:20	3:16 (-24%)	30:25	25:20 (-16.7%)
D-Hydro time [min]	7102	5342 **(-24.8%)**	77410	47055 **(-39%)**

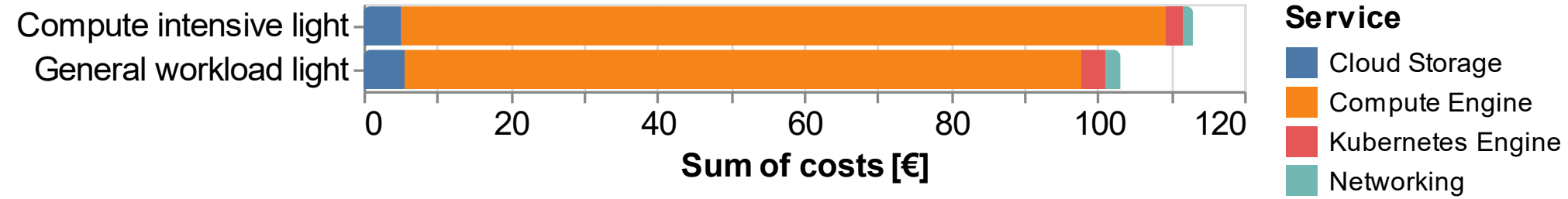
TU Delft

How about the costs?

10x10



5x5



TU Delft

What can we do with this?

- Cost associativity
  - One machine for 1000 hours costs the same as 1000 machines for one hour.
  - If possible, run everything in parallel on thin nodes.

- Compute Intensive nodes are faster, but you must pay for it.

- The pod size could be adjusted to increase the utilization and decrease wasted resources

✓ Quantify information

✓ Use the cloud infrastructure

→ Combine to CBA

Cost Benefit Analysis

- How to compare different model configurations with a cost-benefit analysis?
  - Model resolution
  - Node type
  - Cluster configuration

- Cloud infrastructure costs versus entropy?
- Cloud infrastructure costs versus damage assessment?
- What are the costs and benefits?

TU Delft

Multi Criteria Analysis

The cheapest

Entropy/€

The fastest

Entropy/HPC Time

Largest Entropy

The most amount of information

Entropy/D-HYDRO Time

The most efficient

TU Delft

Multi Criteria Analysis

The cheapest

Medium – general – 10x10

The fastest

Medium – compute – 10x10

5x5

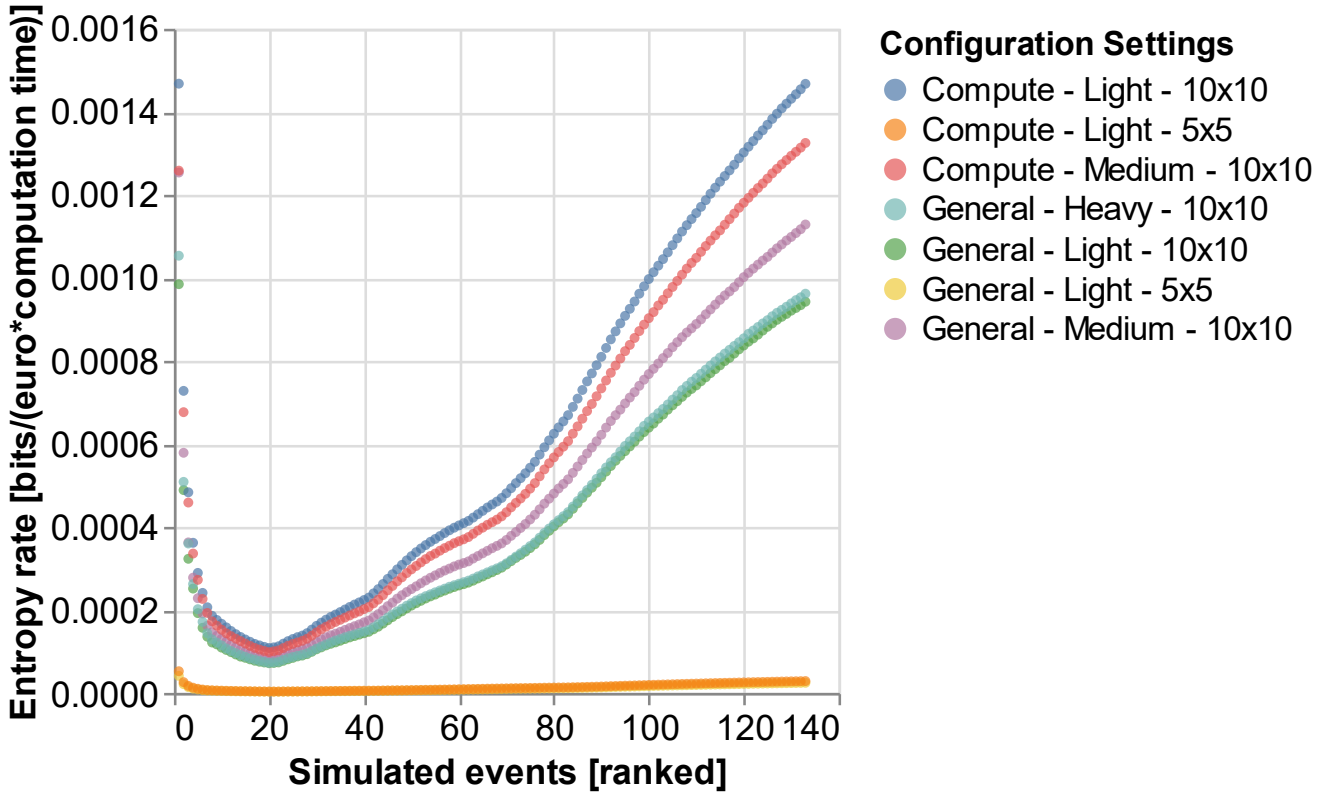
The most amount of information

Light – compute – 10x10

The most efficient

TU Delft

Entropy Rate



TU Delft

To wrap it up:

- The cloud has *'unlimited resources'*, but you pay for everything that you do.
- Where computation time is often a limiting factor, this does not have to be the case anymore with the use of the cloud. However, another important parameter is added: Computation Costs!
- Shannon's Entropy can be used to quantify the model results; however, further research is required to use it to select events that need to be computed upfront.

TU Delft

Thank you for your attention

TU Delft