Massively parallel Landscape Evolution Modelling

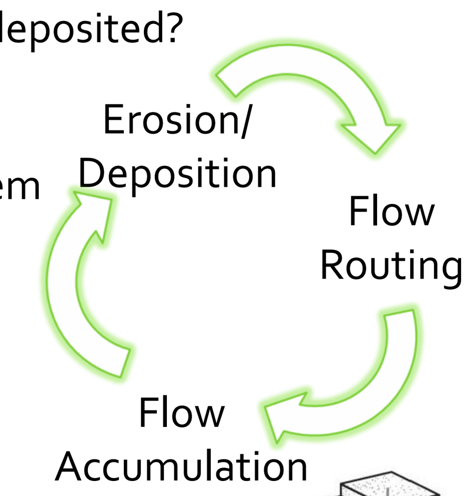
Alex Brown
Computing Science
100559242
email: a.j.r.brown@ncl.ac.uk

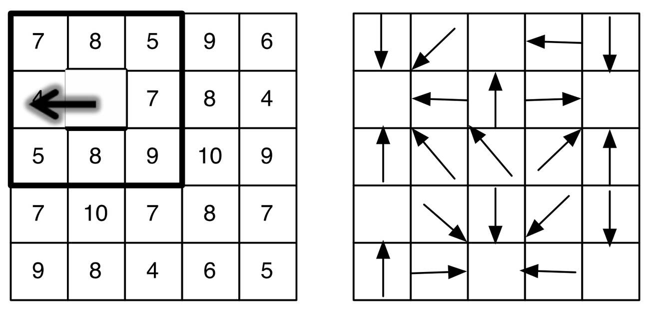
Landscape Evolution Modelling

Modelling how landscapes evolve over millennia.
For each iteration (normally a year):

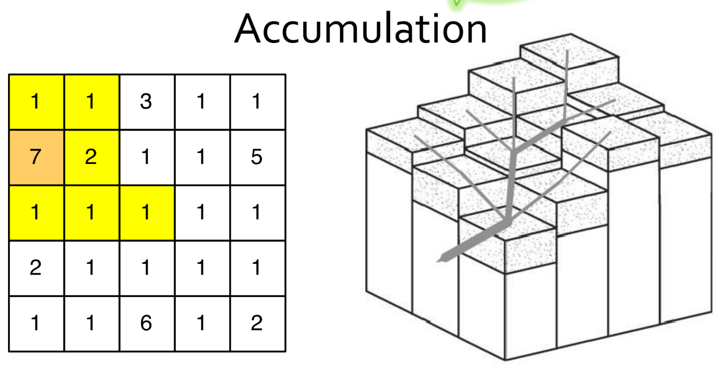
How much material will be removed?
How much material will be deposited?

Each step is 'fairly' fast…
But we want to do lots of them
800K to 1M years
If we could do 1 iteration in 30 seconds this would take ~ 277 – 347 days





Current sequential version is much slower than this:



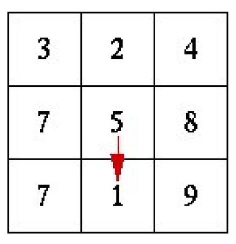
For a 51 x 100 grid this takes 72 hours
(optimized to 2.5 hours)

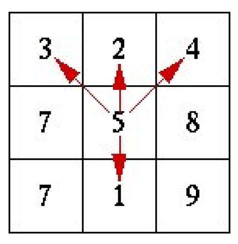
We want to do
> 11000 x 11000

What I did - Parallel Flow Routing

The direction water will flow out of each cell is based on the steepest decent to the surrounding cells. Each cell can be have this calculated independently, given the heights of the surrounding cells are known.

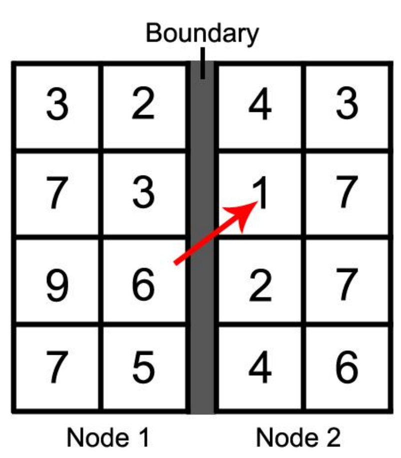
Single Flow Direction





Multiple Flow Direction

The landscape can be divided into smaller sections and these sections can be allocated to separate nodes on a network. Each can compute the flow direction of all cells independently with the exception of the cells around the boundary of the smaller section.

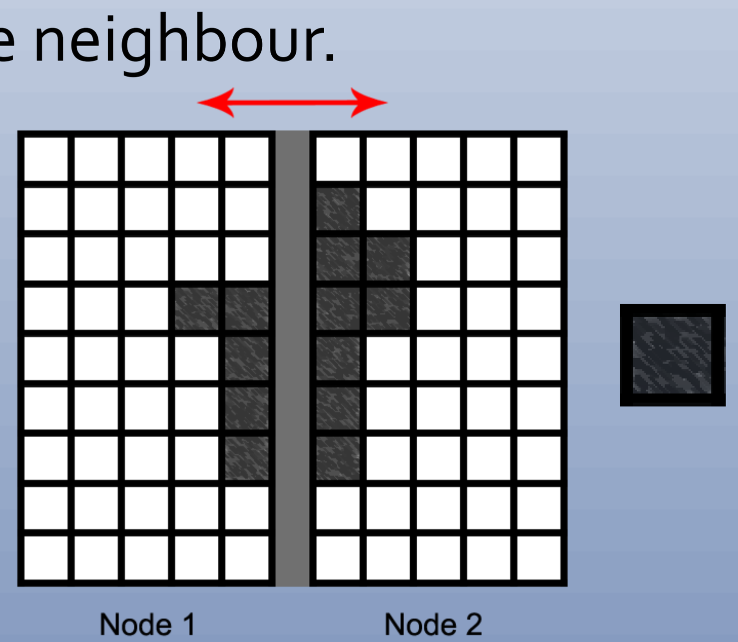


Flow Direction over boundary between nodes

This boundary data, known as a 'halo' can be exchanged between nodes through the use of MPI (Message Passing Interface).

MPI Solution to boundary Plateaus/Sinks

Each node must, for each boundary with a neighbour: locate all the plateaus/sinks on the boundary, calculate the size of each plateau/sink and exchange this data with the neighbour.



Each neighbour can then independently calculate the flow direction for all of its cells.

Aims

What we could do
from 5.1k cells, 100m resolution
(Wainwright, 2006) …
CPU-based model
800, 000 years , annual time-step
72 hrs (2006)
2.5 hrs (2011)

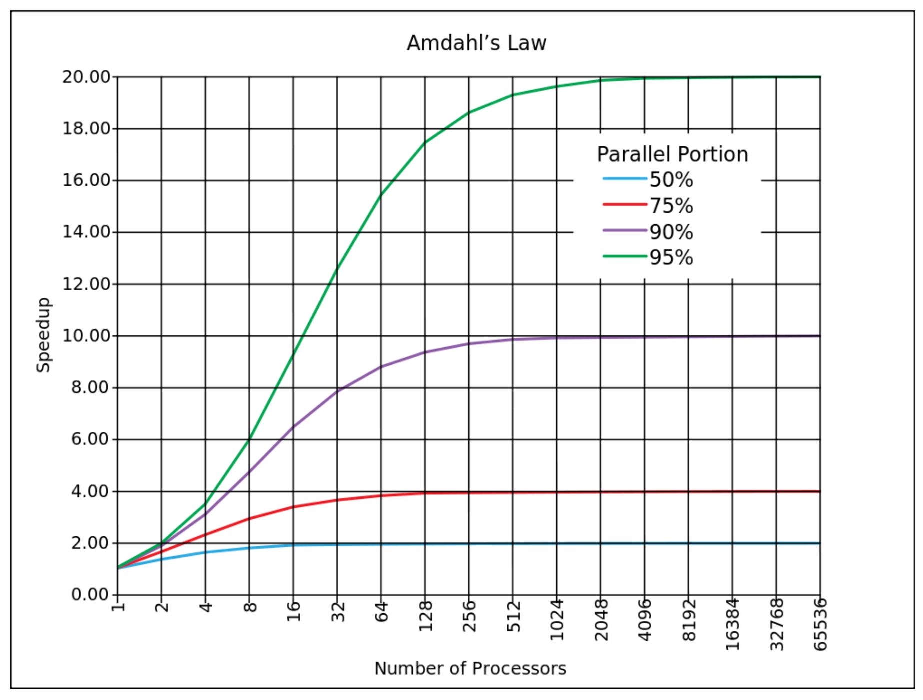
What we want to do
…. ~135M cells, 10m resolution, 1 million years
(Maddy et al, in prep)
Fully Parallelised based
???

Where we are
~46M cells, 10m resolution,
1million years, annual time-step
(McGough et al 2012, submitted)
Partial GPGPU based ~200 days

How will the problem be solved?

MPI (Message Passing interface) allows a complex computation problem to be divided across x nodes on a network, potentially increasing the performance by a factor of x.

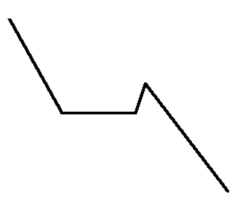
However, not all problems can be completely parallelised, so messages have to be passed between nodes.



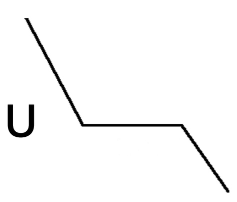
More Complex…

Sinks and Plateaus make it more difficult to perform flow routing. Both are regions where all cells have the same height. Need to compute how water can leave a plateau or sink.

Sink



Plateau



Routing from a plateau: Find the closest cell with lower height. Parallel breadth first search.
Routing from sink: Flood each sink until it becomes a plateau then use plateau routing. Can be complex as sinks can interact with each other. Sink identification using pointer jumping.

Boundary plateaus/sinks problem…

A problem arises when a plateau or sink crosses over the boundary between two neighbouring nodes. The entire plateau or sink has to added to the halo, and exchanged with the neighbouring node.

Conclusion

This research has proven that MPI can be applied to Landscape Evolution Modelling.
However, further study will be required to determine whether or not the overhead of passing messages outweighs the benefits of dividing computation between nodes.

Alex Brown, A. Stephen McGough, Darrel Maddy, John Wainwright[†]
Newcastle University, Newcastle upon Tyne, NE1 7RU, UK, [†]Durham University, Durham, DH1 3LE, UK