BY **JUHA OKSANEN, VILLE MÄKINEN, TAPANI SARJAKOSKI**, NATIONAL LAND SURVEY OF FINLAND and **JAN WESTERHOLM**, ÅBO AKADEMI UNIVERSITY

# Hitting the limits of high-performance computing platforms

▶ With the increased size of data, our analysis and spatial modelling tasks have become more complex. A perfect case for advanced geocomputation.

Anyone operating with geospatial analysis has found themselves in a situation where a GIS application is unable to open the data due to memory limitations, or performing spatial analysis tasks takes far too long, or the application halts completely without explaining the real reason for the stop. In these cases, advanced geocomputation becomes of real value.

Geocomputation has been seen as a new paradigm for approaching spatial analysis. A characteristic of geocomputation is the use of large datasets and computationally intensive data processing. By 'advanced geocomputation' we mean the running of highly optimized code for spatial analysis that makes use of aggregated computing resources, such as multi-node clusters, accelerators like general-purpose graphics processing units (GPUs), and multicore processors. These computing resources may be available in high-performance computing (HPC) supercomputer clusters, but also in the hardware-accelerated workstations commonly used for running GIS applications.

## Too much data

The need for advanced geocomputation stems from factors related to data in terms of size and actuality, analysis and modelling, and the eligible level of interaction in user interfaces.

Development of sensor technology has led us to a situation where the amount of collected data extends far beyond the physical memory limits of any computing resource. Increased spatial coverage together with improved spatial and temporal resolution inevitably lead to a situation where the data will not fit in the memory of your device.

Simultaneously with the increased size of data, our analysis and spatial modelling tasks have become more complex. For example, we need to perform repeated computations due to input data based on real-time environmental sensors and the use of Monte Carlo method with 100–1,000 simulation runs that enables uncertainty awareness of the analysis. We also want to make large pre-renderings of extensive geospatial data for cached Web map services and advanced animated geovisualisation. On top of that, we expect our applications to support a high level of user interaction. Luckily, we seldom need all these requirements simultaneously; rather, factors vary from case to case and also their impact on the consumption of computing resources varies.

At the Finnish Geospatial Research Institute and Åbo Akademi University, we have worked on several research projects to develop solutions for advanced geocomputation. Recently, our special interest has been on quick-response geocomputation, which enables batch processing, as well as interactive use through graphical user interfaces. Examples of these are presented in the following two cases, giving different perspectives for advanced geocomputation.

## Uncertainty–aware drainage basin analysis based on GPUs

In uncertainty-aware geospatial analysis we compute not only the solution to a given problem, but also estimates of the uncertainty of the solution.

Although the foundation for uncertainty-aware geospatial analysis is rather well established, it has received relatively little practical usage. One of the reasons for this situation is the fact that uncertainty analysis is computationally very demanding. The implementations depend on Monte Carlo simulation in which the underlying analysis is typically repeated hundreds of times. It is evident that to carry out uncertainty-aware geospatial analysis with large datasets covering geographically extensive areas stretches computation facilities to their limits. We have designed and implemented an uncertainty-aware drainage basin delineation program that utilizes multiple GPUs to speed up calculations and to permit to efficient processing of large digital elevation models (DEMs).

Incorporating multiple GPUs and using them in parallel is implemented by dividing the DEM into rectangular partitions. In this way, large sections of the partitions can be processed independently of other partitions, and only the values at the local bounda-

ries must be communicated between the neighbouring partitions (Figure 1a). The complexity of this communication depends on the task at hand. While parallelization of, for example, D8-based flow direction calculation for non-flat areas is trivial, filling pits preventing outflow from the DEM may require several iterations to be completed appropriately.

## Country–wide solution using HPC cluster

For benchmarking the HPC solution, we used the country-wide DEM10 (10 m grid) of the whole of Finland from the National Land Survey of Finland (NLS), which fits inside a bounding box of 55,000 × 114,000 elevation points (Figure 1b).

In the first experiment, we used ten NVIDIA K40 GPUs in the HPC cluster operated by CSC – IT Center for Science Ltd. By this arrangement, each data partition with overheads filled the GPU memory, and the number of partitions was minimized. The results of each simulation cycle were ready in 54.6 s and the whole task, with 1,000 simulation cycles, was completed in a little over 15 hours. When the number of GPUs was raised to 30, allowing underuse of memory resources, one simulation run took only 20.6 s, meaning that the whole task would have been ready in five hours and 45 minutes. Still, when applying the new NLS DEM2 (2 m grid) based on airborne laser scanning to similar tasks, these figures need to be multiplied by at least a factor of 25.
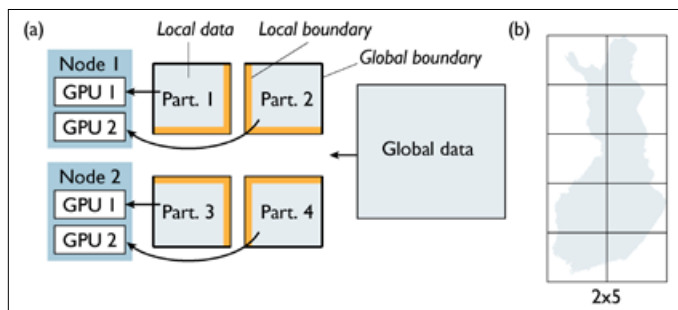


*Figure 1. (a) An example of dividing data into partitions and assigning them for processing in a multi-node multi-GPU environment. In advanced geocomputation, appropriate handling of local boundaries is essential. (b) An example of a regular data partitioning for the uncertainty-aware catchment delineation for the country-wide DEM10 of Finland.*

## Interactive solution for regional–level analysis as a QGIS plugin

To demonstrate the use of the same computational principles as part of an interactive session on generic GIS software, we created a plugin for QGIS (Figure 2).

The plugin serves as the front end for the catchment delineation program: it allows interactive selection of stream segments, setting of DEM error model parameters, and number of partitions and simulation runs, after which the system makes the uncer-

tainty-aware catchment delineation and displays the result in a map window. The example in Figure 2, where an area of 20 km × 20 km (2,000 × 2,000 elevation points on DEM10) was used, 100 simulation runs took 10 s on a computer with an NVIDIA GeForce GTX TITAN GPU. For larger DEMs, the tool also allows the study area to be divided into smaller partitions. The benefit is that less memory is required to perform the analysis; however, for such small areas, running out of memory is rarely a problem, and the analysis time often increases due to the communication needed to keep the partitions in synchronization.

These results demonstrate that we have reached a situation where uncertainty maps of drainage basins for user-selected pour points can be computed efficiently as part of interactive analysis sessions on a local workstation, and even for the whole country in a HPC environment. Potential users of the tool are in national and regional environmental administration, where it could be used for water management, conservation and research, as well as national and international reporting and information system work related to water resources.
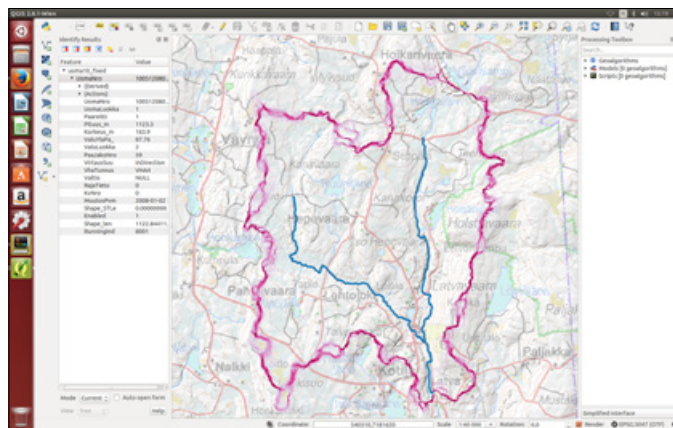


*Figure 2. Uncertainty-aware catchment delineation of Lake Latvajärvi (Puolanka, Finland) calculated using our GPU-enabled QGIS plugin.*

DR. **JUHA OKSANEN** IS A RESEARCH MANAGER AND A LEADER OF THE RESEARCH GROUP ON ANALYSIS AND VISUALISATION OF GEOSPATIAL DATA AT FINNISH GEOSPATIAL RESEARCH INSTITUTE FGI, NATIONAL LAND SURVEY OF FINLAND.
DR. **VILLE MÄKINEN** IS A RESEARCH SCIENTIST AT FINNISH GEOSPATIAL RESEARCH INSTITUTE FGI, NATIONAL LAND SURVEY OF FINLAND.
PROF. **TAPANI SARJAKOSKI** IS A HEAD OF DEPARTMENT IN FINNISH GEOSPATIAL RESEARCH INSTITUTE FGI, NATIONAL LAND SURVEY OF FINLAND.
PROF. **JAN WESTERHOLM** IS A LEADER OF THE HPC GROUP IN ÅBO AKADEMI UNIVERSITY, FACULTY OF NATURAL SCIENCES AND TECHNOLOGY.

FIRSTNAME.LASTNAME@NLS.FI
JAN.WESTERHOLM@ABO.FI