Processing and Modelling on Terrestrial Point Clouds

Igor Rychkov[1,**], James Brasington[*], Damià Vericat[2,*]

Institute of Geography and Earth Sciences, Aberystwyth University, UK

Abstract

A software library is presented for processing and analyzing high resolution massive point clouds obtained with a terrestrial laser scanner. Scans of gravel river beds on a reach scale were used as an example application. The number of points typical of these scans reaching billions, one of the main focuses was given to high performance and scalability of the software. Simplified, yet sufficient model choices underpinning the library are described. The library can be used to retrieve morphological and sedimentological models of gravel bed rivers, enabling study of the channel topology evolution and the movement and size distribution of the gravel. In particular, the library was successfully used on the consecutive scans of the rivers Feshie and Rees with more precise quantitative results and deeper insights obtained.

Keywords: TLS, point cloud, geomorphology, sedimentology, roughness

1. Introduction

TLS. Terrestrial laser scanning (TLS) is a recent advancement in light detection and ranging (LiDAR) technology exploiting a new generation of high performance lasers to study terrestrial objects in greater details over larger areas. It uses the same method of time-of-flight laser pulse distance ranging and a combination of mirrors to determine 3D positions on the surfaces being scanned. Compared to the more conventional air-borne LiDAR scans or GPS total-station measurements it produces much denser point clouds.

[*]Corresponding author

[**]Principal corresponding author

Email addresses: igor.rychkov@canterbury.ac.nz (Igor Rychkov), jtb@aber.ac.uk (James Brasington), damia.vericat@ctfc.cat (Damià Vericat)

[1]Present address: Department of Mathematics, University of Canterbury, New Zealand

[2]Present address: Forest Technology Centre of Catalonia, Spain

Surfaces can now be studied in sub-centimeter detail, as compared to the previous resolution of 1 point per square metre typical of GIS mapping studies. High sampling rate of 4-40 KHz enables rapid survey acquisition times for the range of hundreds of metres. The resolution, range and sampling rate of TLS permit data collection which transgresses system scales, enabling seamless integration of fluvial morphologies from grain-bar-reach within one dataset.

Geomorphology. Perhaps the most important model in geomorphology is the digital elevation model, DEM, which is the topology elevation map, or height field. DEMs from subsequent years are compared in order to reveal erosion and deposition areas, above a certain error threshold. TLS offers an improvement to this methodology because higher density means higher precision. This not only means determining the areas better but also reducing the threshold allowing more subtle changes can to be confidently determined.

Sedimentology. A novel contribution of TLS is in sedimentology studies. Its precision allows to study the surface roughness and grain size distribution. System scale patterns of bed roughness exert a fundamental control on flow resistance, sediment transport and river ecological processes. While field techniques for the measurement of roughness are well established they are invasive, labour Intensive and difficult to generalize over space. Patch-scale ($10^0 \sim 10^1$ m) experiments with hand-held and laboratory laser scanners, have indicated that an explicitly topographical approach to roughness, involving statistical analysis of elevations, may hold more potential to capture these complex effects (Aberle and Nikora, 2006). Upscaling this approach to acquire system-scale data however remains unresolved. In an attempt to address this, we outline a methodology to retrieve local topographical roughness and correlate this to grain-size using 3D point cloud data acquired from a terrestrial laser scanner. As a simple statistical measure of roughness of a point-based geometry is the standard deviation about the local ground level. This requires a model for the ground profile and its removal from the points coordinates, or “detrending”.

Software. TLS technology presents new challenges of effective and scalable point cloud processing software. While commercial products exist, such as Cyclone, that can handle massive point clouds, they usually do not go beyond their visualization. The point clouds of the same area but from different scanning positions are combined through the process called registration to produce a 3D photograph of the objects. Cyclone allows to export the registered point clouds in ASCII files. These files can be too

large to be used in the conventional GIS software such as ArcMap. It is worth mentioning the open source library GRASS which has modules to work with point clouds from air-borne or other conventional LiDAR (`http://grass.osgeo.org/wiki/LIDAR`). A library addressing the challenge of the size of the new generation high resolution laser scans and therefore bridging the chain gap between TLS instruments and GIS mapping software has been missing. In this communication we present an attempt to fill this gap with a library based on an expedient methodology and streamlined implementation.

2. Methodology and Implementation

Digital Elevation Modelling. A compressed representation of 3D TLS point clouds suitable for analysis are 2D raster maps, where each "pixel" is coloured corresponding to the the value of an averaged attribute of interest in the vicinity of the point. Elevation and roughness maps are two typical examples. We should note that the DEM paradigm is not limited to horizontal surfaces and can be applied for example, to steep mountain sloped or cliff surfaces. Our library determines the principal least squares planes which can be used as the grid plane.

Sorting to regular 2d grid. The raster output query may suggest that we should work from the beginning with a regular 2D grid by sorting the points into the corresponding cells (Fig. 1). This operation along with the initial parsing of the multi-gigabyte ASCII input files is the most computationally expensive step which for a billion-order point cloud on a modern desktop computer can take an hour. A rather speedy step follows that finds the points corresponding to the mean, minimum and maximum elevations in each cell. The thinned point clouds have a much smaller size and are easily handled by mapping software, e.g. ArcMap.

Problem with high-resolution DEMs. One way to benefit from high resolution of the point clouds in studying terrestrial surfaces would be to increase the resolution of the DEMs from meters to the size of individual grains. Indeed, some results comparing DEMs of a resolution as high as 0.25 cm are presented below. However, the raster images representing the DEMs grow quickly to several gigabyte yet barely touching on the richness of the high density point clouds and still far from the scales of individual grains. Due to the sparsity of the point cloud most of the pixels or cells in these huge rasters are underpopulated with points and only waste space in the representation.

Subgrid statistics. A more preferred way could therefore be to keep the base grid at a reasonable size and the high resolution detail as the subgrid per-cell data. The subgrid data can be further analyzed to derive the maps with the same dimensions as those of the base grid of model parameters of interest. In this work we focused on statistical, point-based measures without any surface reconstruction or other models imposed on the subgrid points.

Standard deviation. One of the most important for us characteristics besides the already mentioned representative points, is the standard deviation as a measure of surface roughness caused by surface vertical irregularity. Other types or measures of roughness such as the porosity or frontal area would require some surface modelling which is beyond the scope of this work. We show empirically that there is linear dependence between the standard deviation and the grain size, one of the causes of the surface roughness. Before we describe this approach we should point out its relation to the multi-scale character of the problem and more particularly to the choice of deviations to use in the statistics.

Detrending. Intrinsic in this method is the locality of elevations relative to the local ground level. The resolution of the career grid determines the resolution of the ground model. The surface roughness then can be separated from the local ground slopes, the process called detrending. The quality and significance of the determination is generally much better for the detrended dataset.

Ground model: a triangular mesh of local slopes. DEMs not only allow to monitor the evolution of the ground level profile, they are also used to calculate the local slopes at the grid resolution. We use a local triangular mesh laid over the centroids of each cell and its 8 neighbour as vertices as continuous ground elevation model, (Fig. 2). This is perhaps the simplest model, assuming the least surface model structure. It corresponds to the piecewise linear interpolation in the approximation theory. The use of the centroid-thinned point cloud for the ground model acts as a low-pass filter separating the two scales, the grid and sub-grid.

When performing sub-grid analysis of the sorted point cloud the local slope represent an undesirable trend. In the case of small angled-slopes a simple detrending operation can be performed to exclude the large scale, low frequency ground profile. For each point the "absolute" z coordinate is replaced by its vertical displacement from the corresponding foot-point on the ground profile.

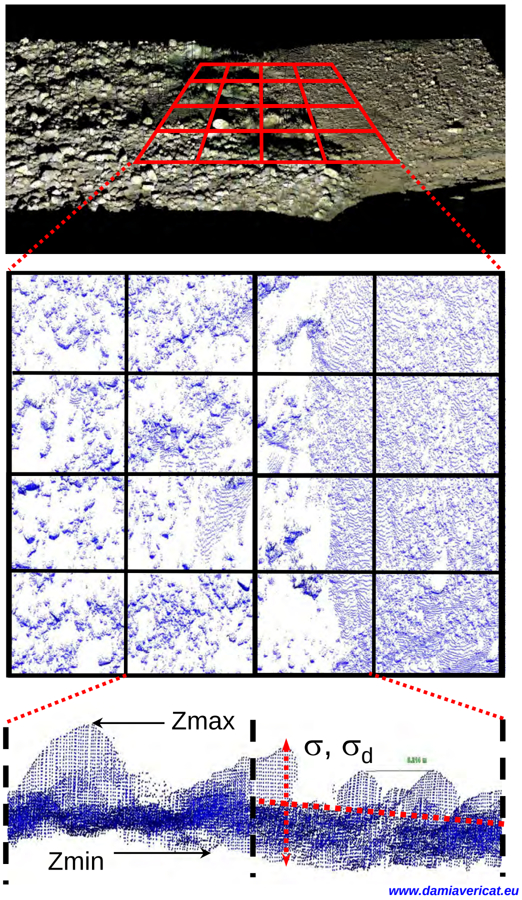

Figure 1: Differentiation of areas with different grain size by subgrid statistical analysis

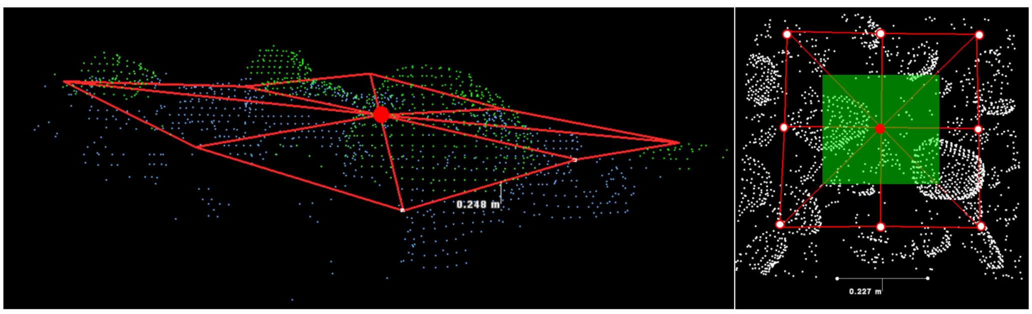

Figure 2: Triangular mesh is laid over thinned point cloud. Green mask represent current cell. Vertices in the current and nearest neighbour cells can be chosen to be either the centroids or minimum elevations. The facies of the mesh are the ground profile slopes.

Vegetation. Detrending allows us to apply this measure to study roughness at the subgrid resolution with more confidence. Mapping of the detrended standard deviation can reveal areas of high values that are usually identified with vegetation (Fig. 6) or other sharp vertical objects of subgrid scales. Compared to the those areas, scans of flat surfaces would yield a negligible standard deviation. The surfaces with vegetation or other vertical structures have the highest deviation values and can be masked out. One can then focus on the gravel patches with medium values of standard deviation.

STD~GSD. Empirical linear relation was derived using *in situ* measurements by means of pebble counts (Fig. 3) of grain size distribution representative parameter, D_{50} indicating a patch-scale Grain Size Distributions (GSD), for essentially different patches and plotting it against the detrended standard deviation from the laser scans (Fig. 4). The almost linear relation allows us to establish one-to-one correspondence between the standard deviation and GSD as gravel surface roughness.

3. Example application

Feshie. TLS was used to acquire detailed models of a braided reach of the River Feshie, Scotland, in summer 2006 and 2007 under low flow conditions when over 90% of the bed is exposed. The reach is dominated by a coarse gravel-cobble-sized armour layer ($d_{50} \sim 40 - 90$mm) of sub-rounded to platy schistose clasts. Scans were acquired from 19 locations, and registered to a common geodetic coordinate system using a network of 37 targets positioned by GPS and reflectorless total station. Registration of the scans to the targets was achieved using a global least-squares bundle adjustment, resulting in low RMS errors of <9 mm for over 250 coincident target measurements.

We looked as sediment budgets by differencing a timeseries of DEMs. TLS helped addressing a typical challenge of the method which is to identify only statistically significant changes. This is complicated by low relief changes typical of fluvial systems: the changes are comparable to survey and interpolation errors. It requires intelligent, spatially variable thresholding taking into account the local slope and so on. TLS advanced the DEM differencing method to a new level: up to 160% more information as compared to GPS techniques could be recovered on bed changes due to much larger point density and sampling precision and more confident spatially sensitive error metric (Fig. 5).

The main benefit of the TLS and the new software applied to this study site was looking into more detailed changes that contributed to the lump annual

Figure 3: Patch-scale Grain Size Distributions (GSD, D50) by method of pebble counts

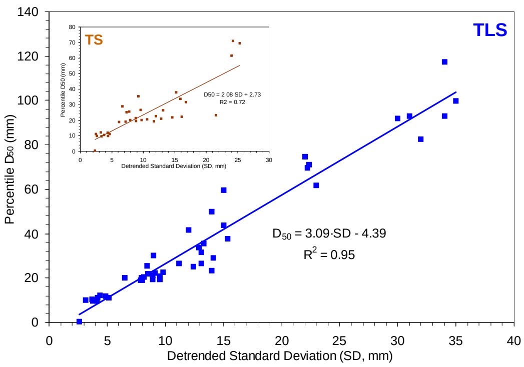

Figure 4: Model linear relationship between the STD of the targeted patches (n= 46) overall the 4 independent data sets. Relationship is more significant in the TLS data set than in the TS

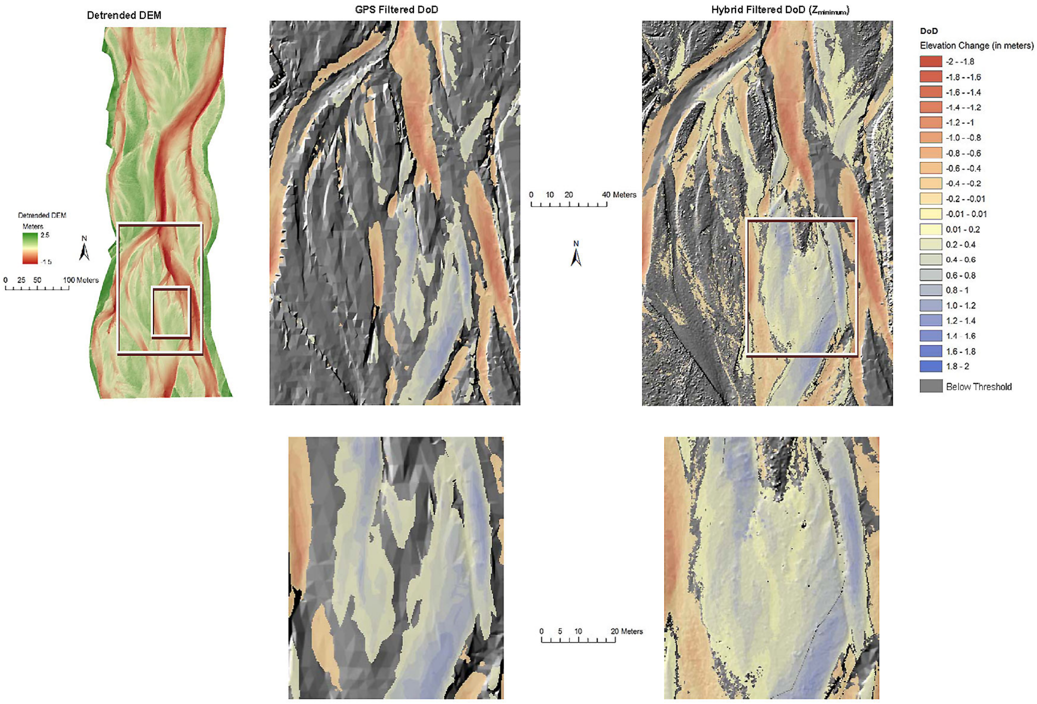

Figure 5: Feshie 2006-2007 DEM differencing. Comparison of TLS DEMs vs. GPS DEMs shows higher details on sediment budgeting and channel morphology.

changes in sediment budgeting. It allowed us to classify the facies including vegetation and different grain types and look at their annual evolution (Fig. 6).

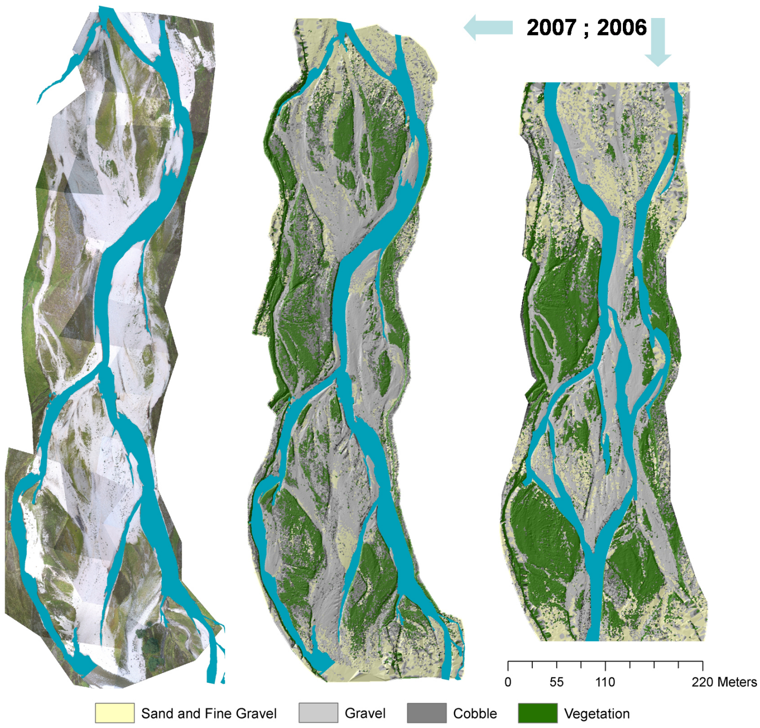

Figure 6: Feshie facies classification and movement from 2006 to 2007.

Rees. The methodology was applied to model the morphological and sedimentary dynamics of an actively braided reach of the Rees River at the head of Lake Wakatipu, New Zealand, during a consecutive sequence of floods. The Rees has a labile gravel bed, which braids extensively in response to high sediment delivery from the uplifting Southern Alps and the wide accomodation space in this recently deglaciated landscape.

Laser Scanning in the Rees was performed using a Leica ScanStation (with 1.2 mm minimum spacing, over 4000 points/second, spot size of 4 mm at 50 meters, 4-40 KHz sampling rate, range of 250-350 meters, horizontal and vertical field of view of 360° and 270° respectively and a digital camera built in).

Analyzing the scans before and after a flooding event we could register the ediment budgeting a good confidence using high resolution DEM differencing

(Fig. 7, D, H). The detrended standard deviation mapping provided insight into which types of gravel was moved on the selected patch by the flood (Fig. 7 E, F G).

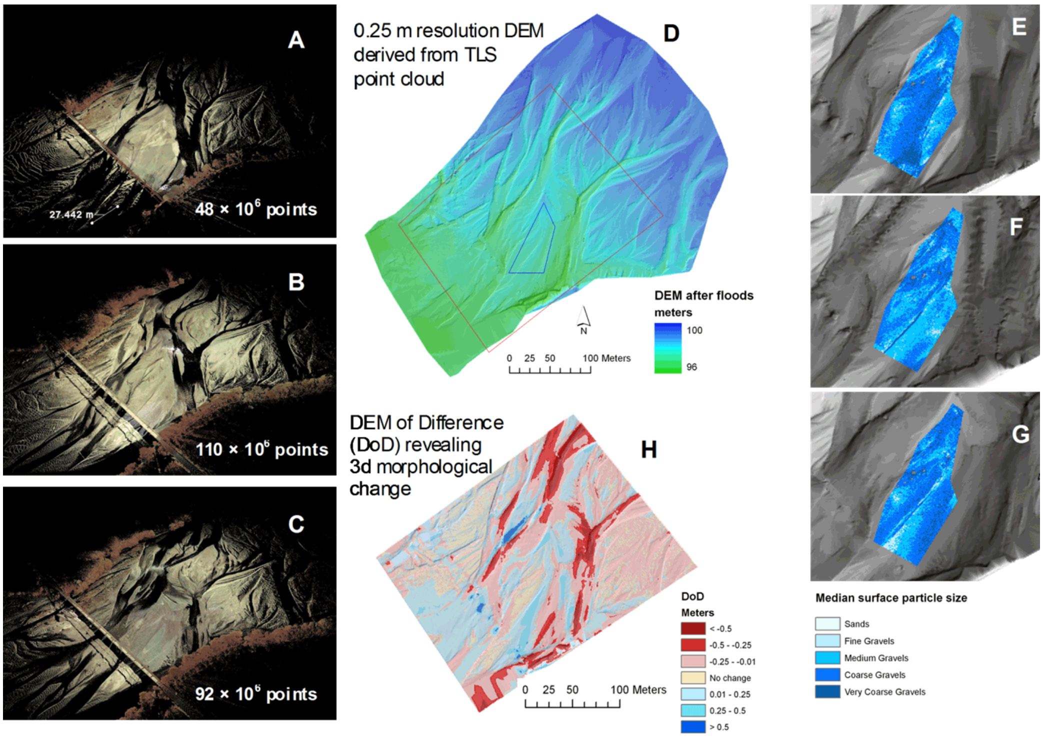

Figure 7: Rees flood (A,B,C) sediment movement study by detailed elevation (D) differencing (H) and grain size, as derived from the detrended standard deviation, mapping (E, F, G), where the darker and lighter blue coloring corresponds to the areas of different gravel sizes.

4. Computational details

We further describe technical computation details. Although serious benchmarking of the computational decisions is beyond the scope of this work, they definitely made the library more effective and scalable.

Space partitioning and sparsity. The "grid and sub-grid" paradigm can be seen as a first step towards a multiresolution analysis which will be the subject of subsequent study. In this simplified approach we did not have to use a hierarchical space partitioning, such as a quad-tree, but used a regular grid instead of the same shape as our output maps. One important issue should be taken into account in when using a regular grid. TLS scans may

have a highly inhomogeneous distribution of points. Cells in a m^2 regular plaid grid, for example, may contain anywhere between 0 and 10^5 points. Adaptive binary trees data structures would be one way to deal with the inhomogeneity of scattered data. The regular grid approach provides a simple method to describe the sparsity too. We used an index 2D array with zero values corresponding to empty cells and non-zero values being the offsets into a densely packed list of non-empty cells.

Sorting the raw point cloud.. Every cell, in turn, holds a pointer to the points that project into it. In that way we preserve the full point cloud as subgrid information that can still be used to model the surface with a subgrid resolution possibly with 3D features such as cliff overhangs, vertical objects and underrock areas.

Sorting of multi-gigabyte raw scanner datasets into a plaid grid is a time consuming operation that fortunately only need to be done once. Although there is some order even in the unified point cloud, the locality is not guaranteed. Therefore at least two iterations are required to read the whole dataset: firstly to make lists of all points belonging to each cells, then to reorder the points into a new dataset. Often the original data comes in ASCII format which involves additional parsing, such are PTS and PTX file formats used by Cyclone software to export the point clouds. Once parsed, the data can be written as binary stream reducing the file size by half and speeding up further file I/O. While sorting our program also produces a few rasters, most importantly the entry indices and number of points per cell, but also some statistical measures such as the mean elevation. Decimated point clouds containing per-cell minimum and maximum 3D points can also be produced at this stage.

Effective disk I/O. Binary stream files have another advantage of being suitable for memory mapping, or virtual memory. Memory mapping is used in disk paging, a fundamental and, presumably, highly optimized function of modern operating system, whereby a disk file is used to mirror a memory block which can be released for use by other programs or can be loaded back to memory. By supplying a binary storage of points coordinates as for memory mapping we could work with it as if it was completely residing in the main memory, while letting the OS to decide which parts of it are actually loaded. The only apparent shortcoming is that the maximum size of mapped file can only be the size of addressable memory. Thus, for 32-bit OS, the maximum addressed memory and the maximum file size would be about 4GB. The restriction is practically removed for 64-bit OS, the OS of choice

anyway for processing of massive point clouds on the computers with RAM exceeding 4GB.

Minimizing page faults. While using the OS service of memory mapping greatly simplified the programming of disk access and therefore avoided possible ineffectiveness of hand-coded implementations, an appropriate ordering of grid cells in memory is desirable in order to further minimize memory page faults. A locality preserving space-filling curve can be used to efficiently map the 2D grid onto 1D memory in such a way that the cells that are close are neighbours in space are close in the memory as well. The added complexity of the mapping procedures should be well compensated by the gain of less paging operation. We have implemented a recent algorithm (Chen et al., 2007) of ordering with the Hilbert curve, the space-filling curve with the best locality-preserving properties (Moon et al., 2001).

Parallel processing and I/O. Parallel I/O is possible with memory-mapped files where different CPU threads access consecutive blocks in the paged memory. The disk or memory access for memory mapped files is transparently serialized by OS between working threads without the need for any special programming. This complements the parallel processing already trivial on rasterized data structures where cells can be processed independently.

5. Software availability

The library is an open source project freely available online (Rychkov, 2009). It is implemented in platform independent C++ code using standard runtime and open source libraries such as OpenMP (`http://openmp.org`), Boost (`http://openmp.org`), Geometric Tools (`http://www.geometrictools.com`). Memory-mapped files are available on both Windows and POSIX systems.

The use of memory-mapped files effectively limits the size of the files to that of the address space which for a 32-bit code is 4GB. Therefore for large multi-gigabyte files the 64-bit compilation, 64-bit Python installation, 64-bit operating systems, and 64-bit capable hardware is required. A binary of the library pre-built for 64-bit Windows is provided, other compilations will be made on demand.

A highly customizable user interface is achieved through Python scripting which allows the user to specify the logic workflow of function calls directly into the compiled library and to specify the full set of arguments. Thus, a need for a middle-tier user interface is eliminated, be it a GUI or command

line parsing. The sample Python script and a sample data file is provided demonstrating the capabilities of the toolset. ASCII files are used as as both input and output file format for compatibility with other software. The library can export thinned point clouds, DEMs, detrended standard deviation and and statistical parameters of local point distribution such as kurtosis and skewness.

6. Conclusions and future work

We have outlined an effective work-flow for taming the data unwieldiness of massive terrestrial point clouds and performing point-based modelling and statistical analysis. We have demonstrating that advances in geomatics together with a suitable processing and modelling of the terrestrial point clouds offer the opportunity to create the first hyperscale models of braided systems, from the scale of individual grains upwards. Moreover, the method is sensitive enough to study the morphodynamics caused by events over shorter periods of time. High-performance implementation allows to work with multi-billion point clouds in a couple of hours on an ordinary modern desktop computer. The methodology could be used to study other terrestrial surfaces such as mountain slopes or cliffs. The software library has laid a foundation for our current and planned point cloud based research. We will soon include into the library a smooth ground level model to be used as the boundary condition in hydraulic simulations. Development of the full multiresolution analysis aiming to build a hierarchy of scale dependent surface models is also in progress.

Aberle, J., Nikora, V., 2006. Statistical properties of armored gravel bed surfaces. Water Resources Research 42, W11414.

Chen, N., Wang, N., Shi, B., 2007. A new algorithm for encoding and decoding the hilbert order. Software: Practice and Experience 37, 897–908.

Moon, B., v. Jagadish, H., Faloutsos, C., Saltz, J. H., 2001. Analysis of the clustering properties of the hilbert space-filling curve. IEEE Transactions on knowledge and data engineering 13, 124–141.

Rychkov, I., 2009. Point-cloud-tools: A library for sorting, processing, and analysis of massive point clouds generated with terrestrial laser scanning. Google Project Hosting.
URL `http://code.google.com/p/point-cloud-tools/`