



Integration of heterogeneous terrain data into Discrete Global Grid Systems

Mingke Li ^a, Heather McGrath ^b and Emmanuel Stefanakis ^a

^aDepartment of Geomatics Engineering, Schulich School of Engineering, University of Calgary, Canada; ^bCanada Centre of Mapping and Earth Observation, Natural Resources Canada, Ottawa, Canada

ABSTRACT

The Canadian Digital Elevation Model (CDEM) and the High-Resolution Digital Elevation Model (HRDEM) released by Natural Resources Canada are primary terrain data sources in Canada. Due to their different coverage, datums, resolutions, and accuracies, a standardized framework for national elevation data across various scales is required. This study provides new insights into the adoption of Discrete Global Grid Systems (DGGs) to facilitate the integration of multi-source terrain data at various granularities. In particular, the Icosahedral Snyder Equal Area Aperture 3 Hexagonal Grid (ISEA3H) was employed, and quantization, integration, and aggregation were conducted on this framework. To demonstrate the modeling process, an experiment was undertaken for two areas in Ontario, taking advantage of parallel computing which was beneficial from the discreteness of DGGs cells. The accuracy of the modeled elevations was estimated by referring to the ground-surveyed values and was included in the spatially referenced metadata as an indicator of data quality. This research can serve as a guide for future development of a national elevation service, providing consistent, multi-resolution elevations and avoiding complex, duplicated pre-processing at the user's end. Future investigation into an operational integration platform to support real-world decision-making, as well as the DGGs-powered geospatial datacube, is recommended.

ARTICLE HISTORY

Received 26 April 2021
Accepted 7 August 2021

KEYWORDS

Discrete global grid systems; terrain data; data integration; national elevation service; multi-resolution elevation; parallel computing

Introduction

Terrain data can be acquired by various technologies with different data formats, spatial resolutions, datum, projections, and update cycles. To make use of heterogeneous topographical data, an integration solution is needed to merge and quality-control the data in a standardized framework (Schumann & Bates, 2018). This type of solution will allow more efficient collaboration between data-providing communities with various acquisition and processing technologies. A standardized platform can provide end-users with unified, analysis ready terrain datasets, eliminating the need for users to apply time-consuming, multi-source data pre-processing steps.

In Canada, terrain datasets released by Natural Resources Canada (NRCan) primarily include the Canadian Digital Elevation Model (CDEM) and the High-Resolution Digital Elevation Model (HRDEM). The CDEM covers the full area across the country with lower accuracy. The HRDEM provides better quality but includes data gaps between tile footprints. Because of their different coverage, vertical and horizontal datums, and spatial resolutions, users working with the CDEM and HRDEM suffer from time-consuming pre-processing and inconsistent results due to different pre-processing methods. This can lead to

duplicated efforts across research groups and limit the resolution options individuals can access. An integration platform for national elevation data across various scales and reference systems is in need.

Discrete Global Grid Systems (DGGs) provide an opportunity to standardize and integrate heterogeneous terrain data (Schumann & Bates, 2018). DGGs are spatial informational reference systems tessellating the entire Earth's surface by nearly equal-area cells without overlaps at hierarchical resolutions (Open Geospatial Consortium, 2017). Each of the discrete cells is indexed by a unique address at each resolution. When modeling on a DGGs, values of geographic phenomena at a location are registered to the corresponding DGGs cell which has a constant location and index at a certain resolution (Alderson et al., 2020; Mahdavi-Amiri et al., 2015). The cell centroid of each DGGs cell is used as the representing point, and the value of a geographic phenomenon at the centroid location is viewed as the value of the certain DGGs cell accordingly (Open Geospatial Consortium, 2017). In this way, point-based spatial information is transformed into useful areal information in a DGGs. From the perspective of heterogeneous data management, DGGs can serve as a uniform data

model to integrate geospatial information from various sources and independent of the original data format, spatial reference system, spatial scale, and acquisition time (Peterson, 2016). In a DGGS, data are attached to spherical cells without projection; this is superior to planar rasters in projected coordinate systems where data are equally spaced on a grid because it is hard to define a single projected system that minimizes distortions over a large landscape such as Canada (Safanelli et al., 2020). Moreover, when handling big geospatial data, arranging computation across multiple memory allocations or processors can efficiently improve the usage of available computational resources (Yao et al., 2020). DGGS have shown great potential for parallel computing as DGGS cells are discrete and independent of each other, which facilitates distributed storage and parallel computing mechanisms for spatial operations (Peterson, 2016; Robertson et al., 2020; Yao et al., 2020). This potential is valuable and promising in the big data era.

Research on the standardized elevation framework has been undertaken by the United States Geological Survey (USGS), which carried out the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) and 3D Elevation Program (3DEP) projects. These activities explored integration of heterogeneous terrain data at global and national scales. The GMTED2010 project integrated raster terrain data from 11 sources, where weight-based mosaic functions were used to combine heterogeneous data to create smooth transitions over the overlapping zones of the neighboring grids (Danielson & Gesch, 2011). As one of the 3DEP products, the national seamless Digital Elevation Models (DEM) were generated by feathering and edge-matching seven sources of data into a successive terrain surface, where high-quality Light Detection and Ranging (LiDAR) and Interferometric Synthetic Aperture Radar (IfSAR) data had the highest priority (Arundel et al., 2017). In addition, researchers have proposed several DGGS implementations for storing or rendering terrain data. The Quaternary Triangular Mesh (QTM) was developed to assemble and manage global terrain data (Dutton, 1988). The QTM started from an octahedron circumscribed by a datum with two vertices at the poles and four vertices at the equator, dividing each triangular parent cell into four child cells (Dutton, 1988). Elevation values were assigned to consecutive triangular facets. This work simplified the Geodesic Elevation Model (GEM) which began as a cuboctahedron connected into a rhombic dodecahedron, recursively refining each face into nine partially nested equilateral triangles (Dutton, 1984). The Ellipsoidal Cube Map (ECM) and Crusta were built

upon quadtrees and designed for optimizing the rendering of global-scale terrain data. Crusta was based on a 30-sided polyhedron and ECM was developed on the circumscribing cube of a reference ellipsoid (Bernardin et al., 2010; Lambers & Kolb, 2012). Nonetheless, terrain data management on DGGS is still in its infancy, and a DGGS-driven integration solution has not been implemented specifically for Canada.

This study explored the adoption of DGGS as an integration platform for Canadian terrain datasets. The study developed algorithms to integrate the CDEM and HRDEM by direct quantization at various granularities and aggregate the modeled elevations by mean, maximum, and minimum statistics across the resolution levels. Different aggregation products are expected for different application purposes. For example, the minimum elevation helps determine stream channel areas while maximum elevation is useful for calculating the height of vertical obstructions (Danielson & Gesch, 2011). The remainder of this paper is organized as follows: Section 2 provides background information about the CDEM and HRDEM. Section 3 describes the DGGS configuration employed in this research. Section 4 demonstrates the modeling process of terrain data using DGGS. In Section 5, we experiment on two study areas using parallel computing mechanisms and present the results. Section 6 discusses the work, points out the limitations, and provides directions for future work. Section 7 concludes the paper.

Terrain data sources

Canadian Digital Elevation Model (CDEM)

The CDEM is a part of the Canadian altimetry system released by NRCAN (Natural Resources Canada, 2015). It provides nation-wide Digital Surface Model (DSM) data and Digital Terrain Model (DTM) data on the south of the tree line around 60° N. In this context, the DTM represents the bare-earth elevation without objects on the surface such as vegetation and buildings, and the DSM means the Earth's surface including all the objects on it. Nonetheless, because of the low density of vegetation and human infrastructures situated north of the tree line, both the DSM and the DTM can be viewed as representing the bare-earth elevation. The CDEM data were derived from the Canadian Digital Elevation Data, which were extracted from the 1:50,000 National Topographic Database (https://ftp.maps.canada.ca/pub/nrcan_rncan/vector/ntdb_bndt), the Geospatial Database (<http://www.geobase.ca>; <https://open.canada.ca/data/en/dataset?q=geobase&organization=nrcan-rncan>), various scaled positional data of provinces or

territories, and remote sensing imagery (Natural Resources Canada, 2013). The CDEM data are referenced to the North American Datum of 1983 (NAD83) Canadian Spatial Reference System (CSRS) datum with decimal degrees as the horizontal coordinate unit, and can be projected to a plane coordinate system at the time of extraction (Natural Resources Canada, 2013). The CDEM data are available at five resolutions. The base resolution is 0.75 arcsec (about 20 m) in the south-north direction and varies from 0.75 to 3 arcsec in the west-east direction depending on the geographic location, where higher latitudes correspond to coarser resolution in the west-east direction. The coarsest resolution available is 12 arcsec (about 400 m) in the south-north direction and 12 to 48 arcsec in the west-east direction. The CDEM records elevations in integer meters on the Canadian Geodetic Vertical Datum of 1928 (CGVD28), where waterbodies are assigned their known or estimated elevations (Natural Resources Canada, 2013). The CDEM includes measured altimetric accuracy per the applicable National Topographic System map sheet, where the absolute altimetric accuracy ranges from 0 to 70 m across the country (Natural Resources Canada, 2013). Additional CDEM information is provided in Table 1.

High-Resolution Digital Elevation Model (HRDEM)

The HRDEM is included in the CanElevation Series released by NRCan to support the National Elevation Data Strategy (Natural Resources Canada, 2020). The HRDEM greatly improves the absolute vertical accuracy and spatial resolution of terrain data in Canada. The HRDEM consists of the high-resolution DTM derived from the high-quality remote sensing imagery via the ArcticDEM project (Porter et al., 2018), and the LiDAR data collected by the Government of Canada in collaboration with the provinces and territories (Natural Resources Canada, 2019).

North of the tree line, the HRDEM provides DSM and covers the full area within Canada's borders (Natural Resources Canada, 2019). The data were originally generated from the optical stereo imagery in the ArcticDEM project and were corrected and improved by NRCan according to Canadian standards (Porter et al., 2018). The data are at 2 m horizontal resolution based on the World Geodetic System 1984 (WGS84) Polar Stereographic North projection. South of the tree line, the original data acquisition method was airborne LiDAR, where the Government of Canada worked in conjunction with the provinces and territories to release existing LiDAR data and to undertake new acquisitions (Natural Resources Canada, 2019). Both the DTM and

DSM are available at 1 m or 2 m resolution with reference to the NAD83 CSRS Universal Transverse Mercator (UTM) projections corresponding to specific zones (Natural Resources Canada, 2019). The DTM was generated at 1 m or 2 m resolution depending on the density of the LiDAR point cloud (Natural Resources Canada, 2019). The HRDEM is only available within the project footprints' extent south of the tree line at present. The HRDEM expresses elevation in decimal meters on the Canadian Geodetic Vertical Datum of 2013 (CGVD2013), the current standard vertical datum in Canada. Lakes on the north of the tree line are assigned their known constant elevation values, with water bodies in other locations denoted as void areas (Natural Resources Canada, 2019). This is due to the properties of the original acquisition method, LiDAR, used in the area to the south of 60° N, where LiDAR pulses are absorbed by water so that the point densities over waterbody areas are greatly reduced (Natural Resources Canada, 2019). Vertical accuracy varies depending on the source projects and data acquisition methods. The absolute accuracy is about 1.6 m and less than 1 m on the north and south of the forest line, respectively (Natural Resources Canada, 2019). Specific vertical accuracy information is available in the spatially referenced metadata for individual products. Additional information about the HRDEM is provided in Table 1.

DGGS configurations

A typical DGGS configuration includes the specifications of the base polyhedron and polyhedral projection method, cell shape and refinement ratio, and orientation of the grid relative to the Earth's surface (Open Geospatial Consortium, 2017; Sahr et al., 2003). The DGGS configuration used in this research is the Icosahedral Snyder Equal Area Aperture 3 Hexagonal Grid (ISEA3H; Figure 1(a)). The ISEA3H scheme has been suggested by Sahr et al. (2003) and has been adopted in several DGGS implementations, such as Global Grid Systems (Global Grid Systems, 2019) and geogrid (Mocnik, 2019).

The ISEA3H contains several advantageous characteristics. Compared to the other four common Platonic solids, namely tetrahedron, cube, octahedron, and dodecahedron, an icosahedron has the smallest face area and the smallest interior angles, resulting in the least angular distortion when being projected to a datum (Mahdavi-Amiri et al., 2015; Sahr et al., 2003). White et al. (1998) compared the distortion characteristics among five approaches to mapping from an octahedron or an icosahedron to the sphere surface, including the Gnomonic projection, Fuller's Dymaxion projection, Snyder's

Table 1. Detailed information on CDEM and HRDEM (Natural Resources Canada, 2013, 2019).

	CDEM		HRDEM	
	North of 60° N ^a	South of 60° N	North of 60° N ^a	South of 60° N
Terrain data type	DSM	DTM or DSM	DTM or DSM	DTM or DSM
Coverage	DSM is available national wide; DTM is available only on the south of 60° N	Entire Canadian landmass	Project footprints	Project footprints
Original data acquisition method	Stem from CDED which was extracted from NTDB, GDB, various scaled positional data, and satellite images		Optical satellite images, radar interferometry, and airborne LIDAR	Airborne LIDAR
Horizontal datum	NAD 1983 CSRS	NAD 1983 CSRS	WGS 1984	NAD 1983 CSRS
Plane coordinate projections	Pseudo-Mercator (EPSG: 3857) or Canada Atlas Lambert (EPSG: 3979) ^b		WGS84 Polar Stereographic North (EPSG: 3413)	NAD 1983 CSRS UTM based on the corresponding zones
Horizontal coordinate unit	Decimal degrees, while meters will be used when a projection is selected when downloading	Meter	Meter	Meter
Horizontal resolution	0.75, 1.5, 3, 6, or 12 arcsec ^c	2 m	1 m (most) or 2 m	1 m (most) or 2 m
Vertical datum	CGVD1928	CGVD2013	CGVD2013	CGVD2013
Vertical unit	Meter	Meter	Meter	Meter
Vertical resolution	Integer	Decimal	Decimal	Decimal
Vertical accuracy	Available as the altimetric accuracy per NTS map sheet, ranged from 0 to 70 m across the country	About 1.6 m	About 1.6 m	Less than 1 m, vary among source projects
Extracting extent	Predefined clipping area or customized clipping area ^d		Defined 50 × 50 km tiles	Defined 10 × 10 km or 20 × 20 km tiles
Waterbody value	Known or estimated elevation ^e		Constant value for lakes; void data for rivers	Void data
Derived topographical products	Slope, aspect, hillshade, color relief, color shaded relief maps, and point-based elevation query		Hillshade, color relief, and color shaded relief maps	Slope, aspect, hillshade, color relief, and color shaded relief maps
Metadata availability	Spatially referenced metadata		Spatially referenced metadata	
Data delivery format	GeoTIFF, ESRI ASCII Grid, ASCII Gridded XYZ, ASCII Gridded CSV		GeoTIFF	
Data update progress	No longer updated		Will gradually update once new LIDAR data are acquired	
License	Open Government License – Canada		Open Government License – Canada	
Link to data source	https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333		https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995	

^a60° N approximated the tree line used to separate the northern and southern parts of Canada.

^bRaw data included geographic coordinates; data with plane coordinate projections were available for download.

^cListed resolutions were those in the south-north direction, and the resolutions in the west-east direction vary depending on the geographic location, where higher latitudes correspond to coarser resolution in the west-east direction.

^dPredefined clipping area options included the National Topographic System (NTS) of Canada map sheets at the 1:250000 scale, Landsat footprints, and drainage areas.

^eFor CDEM data produced using ground elevations, waterbodies were considered as naturally occurring areas of constant elevation (lakes) or having a small slope (rivers). Waterbodies were assigned their known or estimated elevation.

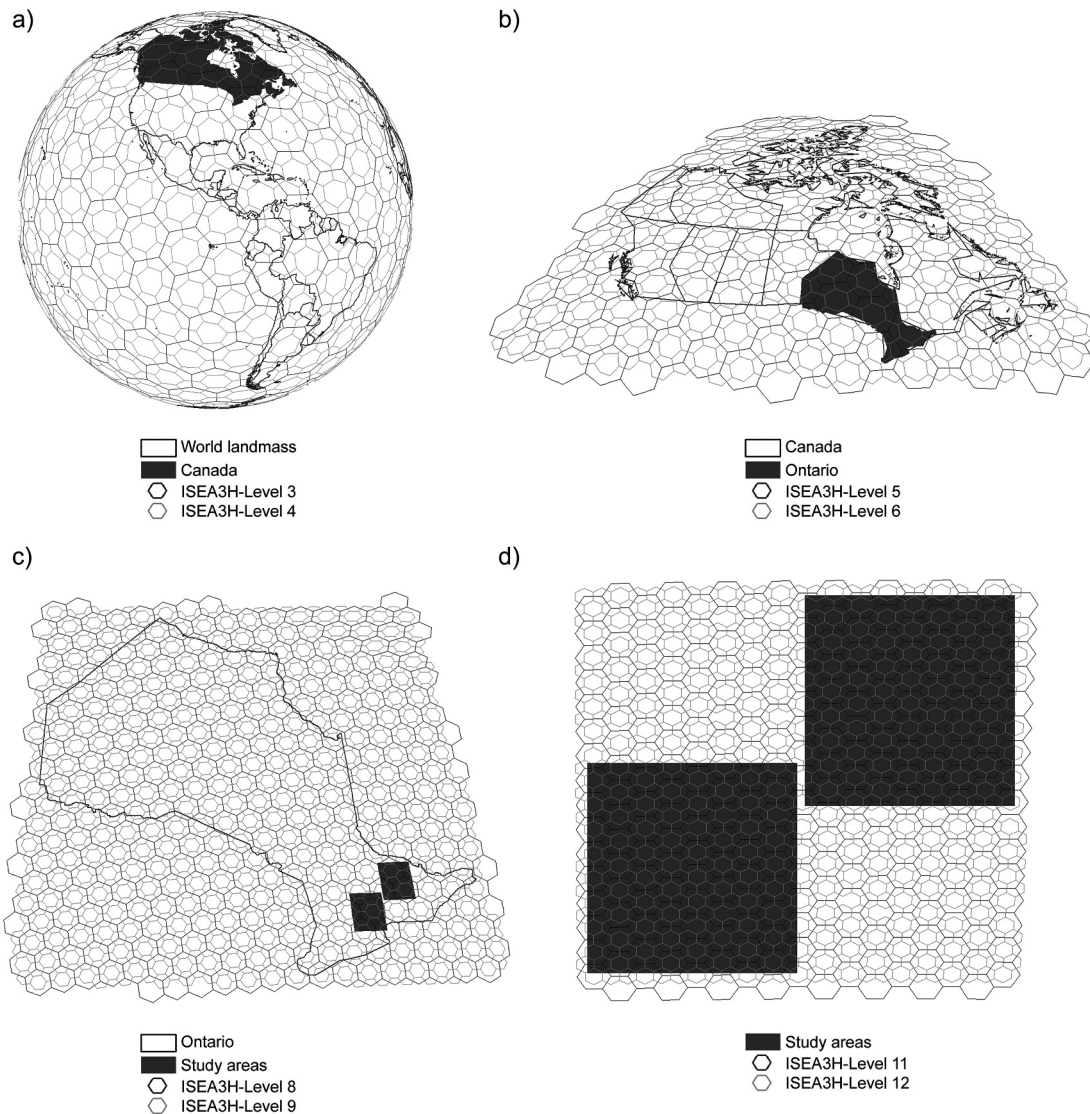


Figure 1. Representation of a) world landmass, b) Canada, c) Ontario, and d) study areas in the ISEA3H DGGs adopted in this research, with the orientation parameters: latitude of the pole (λ) = 37.6895°, longitude of the pole (φ) = -51.6218°, and azimuth (α) = -72.6482°.

equal area polyhedral projection, direct spherical subdivision by great circle arcs, and a hybrid method of the direct spherical subdivision and recursive Gnomonic projection. The researchers concluded that the icosahedral Snyder equal area projection contributed to the least area distortion (White et al., 1998).

Hexagonal grids have been increasingly accepted by researchers because of their greater angular resolution, uniform adjacency, lower average error when tiling the plane, and good approximation to Cartesian distance when generating the discrete distance metric (Conway & Sloane, 1998; Golay, 1969; Luczak & Rosenfeld, 1976; Sahr, 2011). Hexagonal pixels were also found to be

superior to square pixels in structuring rasters (Mersereau, 1979; Petersen & Middleton, 1962; Staunton, 1989). The “refinement ratio of three” is the smallest possible hexagon aperture. The hexagonal tessellation with aperture three contributes to a smoother transition between DGGs resolution levels compared to apertures of four or seven (Mahdavi-Amiri et al., 2015). The centroid-aligned, aperture-three hexagonal tessellation has the property that a parent cell’s six vertices are its child cells’ centroids (Sahr et al., 2003). This naturally leads to the important monotonical convergence characteristic, with which the representing centroid is infinitely closer to the point to be modeled at finer

resolutions (PYXIS, 2020). In other words, a finer resolution provides higher accuracy when modeling points on such a DGGS.

In the early stages of DGGS development, the orientation of a polyhedron was determined by achieving simple objectives such as aligning the poles and the prime meridian (Fekete & Treinish, 1990), maximizing the number of within-ocean polyhedral vertices (Fuller, 1982), or equatorial symmetry (Sahr & White, 1998). Barnes (2019) extended these criteria and developed algorithms to determine optimized orientations of DGGS under various global scenarios by solving the nonconvex optimization problems. To further satisfy local application scenarios, Zhou et al. (2020) managed to avoid the area of interest being split across multiple polyhedral faces and place the area of interest as central as possible within one face, so that the projection distortion was further reduced. As reported in their study, Canada's borders are centered on a single icosahedral face with the following orientation parameters: latitude of the pole (λ) = 37.6895°, longitude of the pole (φ) = -51.6218°, and azimuth (α) = -72.6482° (Zhou et al., 2020). This orientation of icosahedron was accepted in this study (Figure 1(a,b)).

Terrain data modeling process in DGGS

The modeling process of terrain data in the ISEA3H DGGS had the following main phases: data acquisition, pre-processing, quantization, aggregation, and quality control. The open-sourced library dggridR was used to complete conversion between geographic locations and ISEA3H DGGS cell indices (Barnes & Sahr, 2017). In the end, the library datashader, together with its native support for the library Matplotlib, was used to create the representations of modeled elevations on DGGS (Bednar et al., 2016). Figure 2 illustrates the workflow of the data acquisition, pre-processing, direct quantization, statistical aggregation, quality control, and visualization process. The modeling process was developed using a hybrid of Python 3.7.7 and R3.6.2 environments.

Terrain data acquisition

The CDEM data were obtained in the GeoTIFF format through the Geospatial Data Extraction tool at the finest available resolution, namely 0.75 arcsec (about 20 m) along the latitudes, in the NAD83 CSRS reference system over the area of interest (Natural Resources Canada, 2017). Available HRDEM data within the area of interest were found by spatially overlaying the HRDEM data footprints, downloaded as 10 by 10 km GeoTIFF tiles for

data at 1 m resolution or 20 by 20 km tiles for data at 2 m resolution, with the NAD83 CSRS UTM projections corresponding to specific zones (Natural Resources Canada, 2020).

Pre-processing

The purpose of the pre-processing was to standardize the horizontal and vertical datums and the coordinate system of the CDEM and the HRDEM datasets before being modeled in DGGS. To standardize the vertical datum, the elevations of the CDEM were converted from CGVD28 to CGVD2013, the current standard datum in Canada. To standardize the horizontal datum and the coordinate system, the HRDEM tiles were converted from the projected coordinate system to the geographic coordinate system (i.e. the NAD83 CSRS) via inverse projections. The resulting resolution of the HRDEM rasters in the geographic space was about 0.04 arcsec or 0.08 arcsec where the original HRDEM tiles were at 1 m or 2 m resolution under the UTM projections, respectively. Transforming the CDEM and the HRDEM datasets into the same geographic coordinate reference system rather than a projected coordinate reference system accounts for the specificity of the Earth's curvature and avoids potential computational errors over a large landscape (Florinsky, 2017; Safanelli et al., 2020). This is in line with the essence of a DGGS which is a tessellation of the Earth's surface instead of a planar grid. Having the source DEMs in the same geographic coordinate system allows the following quantization process to be achievable, because the DGGS cell centroids are in a geographic coordinate system as well. It also contributes to greater workflow flexibility given that the specific UTM projections adopted by the HRDEM data vary among UTM zones. The HRDEM tiles were mosaicked after the coordinate reference system conversion.

Quantization

The OGC abstract Specification released in 2017 demonstrated that quantization was one of the basic operations required by a compliant DGGS implementation, supporting assigning data to cells and retrieving data from cells (Open Geospatial Consortium, 2017). The quantization strategy accepted by a DGGS can lead to uncertainties when applying a DGGS, such as data quality and topology validity (Li & Stefanakis, 2020a). For geospatial data originally stored in a raster format, such as remote sensing imagery and DEMs, a DGGS quantization strategy needs to consider the sample rate at the native pixel resolution, the algorithm of sampling

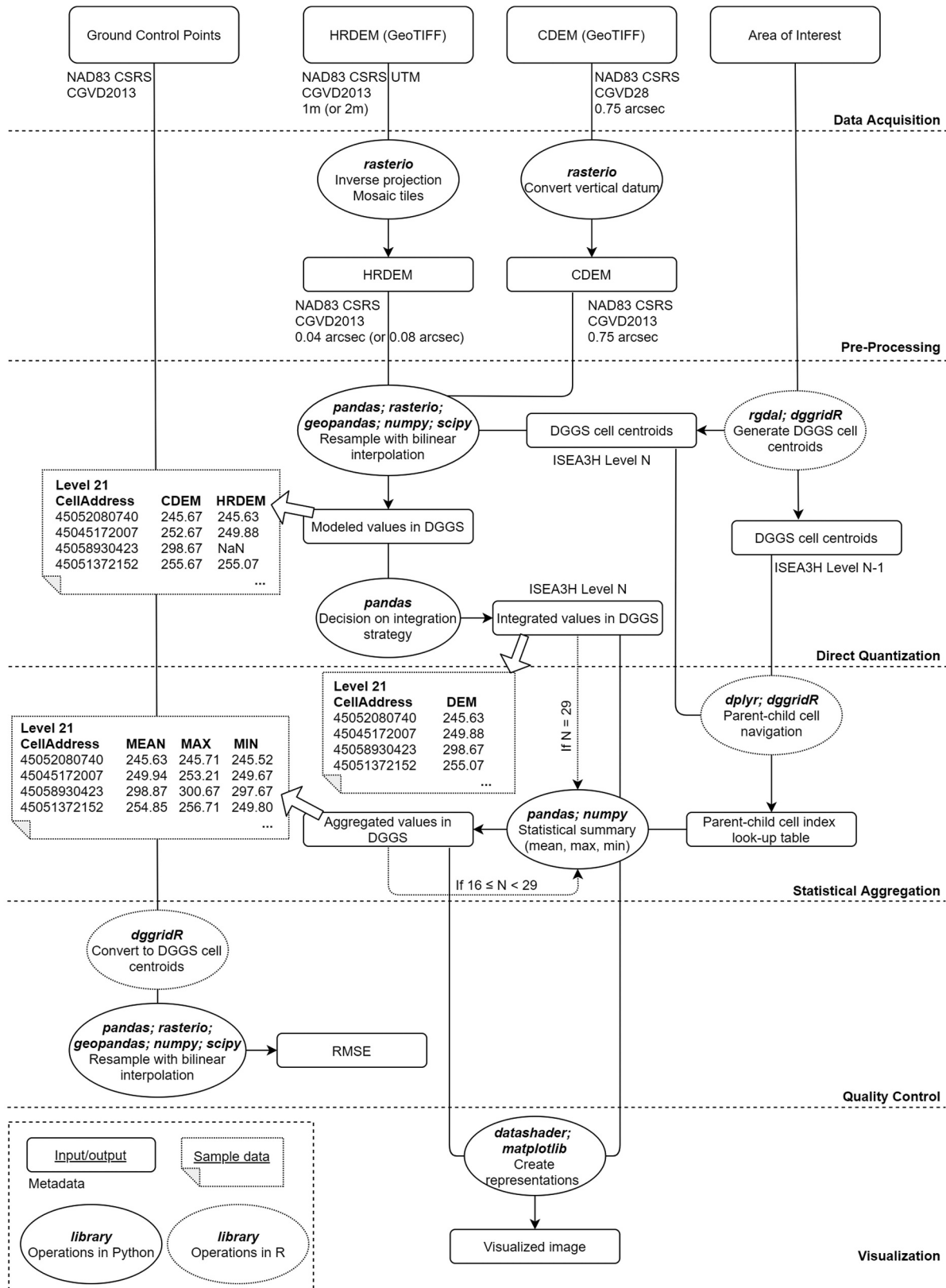


Figure 2. Workflow of data acquisition, pre-processing, direct quantization, statistical aggregation, quality control, and visualization process.

at the finest DGGs resolution, and the approach to aggregate geospatial information when moving from finer to coarser resolutions in a DGGs (Li & Stefanakis, 2020b).

In our case, resolution levels 16 to 29 in the ISEA3H DGGs were selected as the modeling granularities. The finest available resolution of the CDEM data was 0.75 arcsec (about 20 m) along the south-north direction, and the original resolution of HRDEM data was 1 m or 2 m in the projected space. Thus, the resolution level 29 with 0.74 m^2 hexagonal cell area was considered as the suitable finest DGGs level because of its higher yet closer sample rate compared to the original sample rate of CDEM and HRDEM data (Table 2). The hexagonal cell area is about 1.18 km^2 at resolution level 16, which was chosen as the coarsest resolution in this study (Table 2). The quantization can be carried out at each of the levels from 16 to 29, and aggregation can be implemented at levels 16 to 28. In this context, the quantization process aimed to resample elevation values with bilinear interpolation in DGGs at the cell centroid locations at a certain resolution level. Like a planar DEM raster, where the value at the central point of a rectangular grid is used as the elevation of the certain grid cell, the interpolated value at the cell centroid location is assigned to the certain DGGs cell to represent the elevation of the areal cell. Bilinear interpolation was recommended by (Ma et al., 2021) who resampled the Landsat 9 image over the Yellow River Estuary area, China and the GF-1 image over the province of Hainan, China, on aperture four hexagonal grids. The researchers demonstrated that the hexagonal grid image resampled by bilinear interpolation remains the most information and structural features because its quantitative evaluation indices (information entropy, mutual information index, deviation index, etc.) are closest to one, compared to

nearest-neighbor interpolation and cubic convolution (Ma et al., 2021). Other evidence included the smallest root mean square errors (RMSE) resulting from bilinear interpolation among the randomly selected checkpoints within the testing imagery (Ma et al., 2021).

Geographic locations of ISEA3H cell centroids along with their indices within the area of interest were generated by the library dggridR (Barnes & Sahr, 2017). Cell centroid locations were used to determine the elevations by bilinear interpolation based on the source rasters, and to determine the parent-child cell relationship among resolution levels. The latter usage is necessary in the aggregation process and explained in the next section. Within the area of interest, the CDEM data were interpolated for all available cell centroids, and the HRDEM data were interpolated for those centroids falling in the HRDEM extent. Specifically, when interpolating the CDEM data, the first-order neighbors of the target pixel in the CDEM raster were found and scanned to determine if there were enough valid interpolation inputs. In other words, if the corresponding raster pixel of the target DGGs cell centroid did not have enough interpolation neighbors, or if any of the necessary neighbors contained the void pixel value, the target DGGs cell would be assigned the void value. This scenario only occurred where the DGGs cells were close to the boundary of the area of interest. When modeling the HRDEM, only the cell centroids falling in the HRDEM extent were considered, and the same criteria were applied, namely, to run bilinear interpolation only when there were enough valid interpolation neighbors. This could happen when the target cell centroid fell in a waterbody or where the DGGs cells were close to the boundary of the HRDEM project footprint.

In addition, previous studies suggested that the ratio of horizontal to vertical resolution has impacts on terrain data applications (Garbrecht & Martz, 2000). For example, if the elevations are stored as integer meters and the horizontal resolution is as fine as 1 m, then the calculated slope can only take a limited number of discrete values which are not accurate enough over the flat terrain (Garbrecht & Martz, 2000). On the other hand, if the cell size is too small compared to the vertical resolution, local artifacts can be introduced and the computation of land-surface parameters will require more time (Hengl & Evans, 2009). Hence, during quantization at coarse DGGs resolutions, a rough ratio of horizontal to vertical resolution was maintained. Elevation values were stored with six decimal places in the original HRDEM data, and this was used as the vertical resolution at the finest DGGs level. The decimal places of modeled elevations decreased at coarse DGGs

Table 2. Quantitative information of the ISEA3H DGGs and vertical resolution at levels 16 to 29.

Resolution Level	Cell Area (m^2)	Spacing (m)	CLS ^a (m)	Decimal Places of Elevation
16	1,184,911.67	1,075.10	1228.28	0
17	394,970.56	620.70	709.15	0
18	131,656.85	358.36	409.43	0
19	43,885.62	206.90	236.38	1
20	14,628.54	119.45	136.48	1
21	4,876.18	68.97	78.79	2
22	1,625.39	39.82	45.49	2
23	541.80	22.99	26.26	3
24	180.60	13.27	15.16	3
25	60.20	7.66	8.75	4
26	20.07	4.42	5.05	4
27	6.67	2.55	2.92	5
28	2.23	1.47	1.68	5
29	0.74	0.85	0.97	6

^aThe diameter of a spherical cap of the same area as a cell of the specified resolution.

resolution levels. Table 2 lists the quantitative information of the ISEA3H DGGS and decimal places of modeled elevations at resolution levels 16 to 29.

After the quantization, each DGGS cell at a certain level had two elevation values along with its index, each of which corresponded to either the resampled CDEM data or HRDEM data, although some of the values could be void as explained before. A decision on the integration strategy needed to be made at this point to provide each DGGS cell with a unique elevation. In this research, the HRDEM values were preserved wherever the interpolated HRDEM values were available because of their better quality than the CDEM.

Aggregation

The statistical aggregation in DGGS was conducted based on the aggregation values at the consecutive finer level, except that the aggregation at the second finest level (i.e. level 28) was based on the directly quantized elevations at the finest level (i.e. level 29). The elevations involved in the aggregation at level 28 were those integrated values, namely the interpolated HRDEM values wherever they were available at level 29. Ahead of aggregation, the parent-child cell index look-up tables were generated in order to locate seven specific child cells for each of the parent cells at a certain level. This was achieved by determining the geographic locations of six vertices as well as the centroid of a parent cell and converting these seven geographic locations to cell indices at the consecutive finer level. This was reasonable given the characteristic of the centroid-aligned, aperture-three hexagonal tessellation that a parent cell's six vertices locate at its child cells' centroids (Figure 3). Then, the elevations of seven child cells were determined, and the mean, maximum, and minimum elevations were calculated and assigned to their parent cell at the immediately coarser resolution

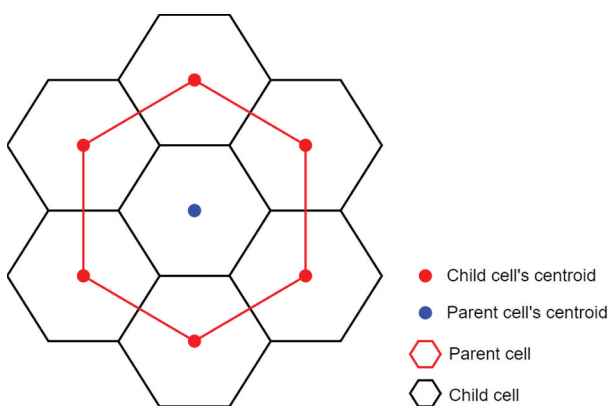


Figure 3. The relationship between a parent cell and its seven child cells at the immediately finer level.

level. In this way, the aggregated values were calculated hierarchically, and, for example, the maximum elevation of a cell at a certain level was always the maximum elevation of all its child cells at finer levels. During this process, the elevations were rounded to the corresponding decimal places at various resolution levels, as listed in Table 2.

Quality control

In this study, data quality evaluation contained three aspects: 1) comparison between the ground-surveyed elevations and the elevations on the CDEM or the HRDEM before being modeled in the ISEA3H DGGS (hereby referred to as “pre-DGGS elevation”); 2) comparison between the ground-surveyed elevations and the integrated elevations after being modeled in the ISEA3H DGGS (hereby referred to as “post-DGGS elevation”) among all resolution levels; and 3) comparison between the pre-DGGS elevations and the post-DGGS elevations among all resolution levels. The pre-DGGS elevation is essentially the elevation of the nearest central point of the grid in the original DEM raster, and the post-DGGS elevation is essentially the elevation of the nearest cell centroid in the DGGS. The RMSE were calculated to indicate the data quality evaluation at various modeling levels and reported as the accuracy indicator in the metadata. All the elevations were standardized on CGVD2013 before the computation of RMSE.

Metadata

The spatially referenced metadata of the modeled elevations in the ISEA3H DGGS at various levels were created in the GeoJSON format. Each GeoJSON file included two polygon geometries on the NAD83 CSRS datum, each of which represented the extent of the HRDEM or the CDEM (i.e. the full area of interest excluded the HRDEM extent). Attributes of the polygons were the metadata of the modeled elevations, containing the source data information, DGGS configuration, and the accuracy and precision. The specific items included in the metadata are listed in Table 3.

Experiment

The experiment was carried out to test the modeling process in the ISEA3H DGGS on the Advanced Research Computing cluster at the University of Calgary, Canada. Hardware partitions *cpu2019*, *razi-bf*, *apophis-bf* (40 cores, 2x Intel® Xeon® Gold 6148 CPU @ 2.40 GHz with 190 GB memory) and *bigmem* (80 cores, 4x Intel® Xeon® Gold 6148 CPU @ 2.40 GHz with 3022 GB memory) were used in this experiment whenever the requested number of

cores and memory were available. To improve the computational efficiency and take advantage of the discrete property of DGGS cells, the modeling process ran in a parallel fashion, which was a hybrid of the shared-memory parallelism and the job-level parallelism. The testing areas were sub-divided into fishnet grids to satisfy the job-level parallelism. Meanwhile, the Python multiprocessing module and the R doParallel module were used in the shared-memory parallelism when processing each of the fishnet grids (McKerns et al., 2011; Weston & Microsoft Corporation, 2020). The code used to conduct the experiment is available in the GitHub repository <https://github.com/Erin-1919/DGGS-Elevation-Integration>.

Study area and data acquisition

The modeling process in the ISEA3H DGGS was tested on two 1.5 by 1.5° regions in Ontario (Figure 1(c, Figure 1d)). The first region (hereby referred to as “vegetation region”) was around Algonquin Provincial Park, spanning 77.36 to 78.86° W and 44.51 to 46.01° N, where 78.31% of the area was forest, 10.85% open water, and 3.72% agriculture (Figure 4). The second region (hereby referred to as “urban region”) was around Toronto, spanning 78.92 to 80.42° W and 43.32 to 44.82° N, where 39.73% of the area was agriculture, 21.18% open water, and 13.41% community and infrastructure (Figure 4). The percentage of the land cover classification was calculated based on the Ontario Land Cover Compilation v.2.0 dataset (Ontario Ministry of Natural Resources and Forestry, 2016). The CDEM data over these two study areas were obtained at 0.75 arcsec resolution in the NAD83 CSRS reference system through the Geospatial Data Extraction portal in the GeoTIFF format (Natural Resources Canada, 2017). A total of 34 and 165 HRDEM GeoTIFF tiles with 10 by 10 km dimension, over the vegetation region and urban region, were obtained at 1 m resolution with the NAD83 CSRS UTM 17 N projection (Figure 4). Based on CGVD2013, elevations ranged from 113 to 576 m over the vegetation region and 70 to 542 m over the urban region. The south-east portion of the Toronto region around Lake Ontario represents territory of the United States and was assigned a void value for this study.

Direct quantization and statistical aggregation

Elevations over two regions were quantized in the ISEA3H DGGS at levels 16 to 29, and the statistical aggregation was conducted at levels 16 to 28, following the processing flow illustrated in Figure 2. During quantization and aggregation, each of the two regions were sub-divided into 100

Table 3. Categories and items contained in the metadata.

Category	Item
Source data information	Name
	Institution
	License
	Coordinate reference system
	Horizontal resolution
	Vertical resolution
	Source data accuracy
	Vertical datum
	Vertical unit
	Void data
	DGGS configuration
Cell shape	
Aperture	
Orientation (latitude)	
Orientation (longitude)	
Orientation (azimuth)	
Resolution level	
Cell area	
Centroid spacing	
Accuracy and precision	Precision of modeled elevations
	RMSE calculated against ground control points
	RMSE calculated against pre-DGGS values

fishnet grids (10 columns by 10 rows), which led to 100 individual jobs scheduled on the Advanced Research Computing cluster (Figure 5). Each of the jobs (i.e. each of 100 fishnet grids) involved 16 cores for the shared-memory parallelism on a requested hardware partition, allowing the 16 processes to run simultaneously. Outputs were merged after all processes were completed (Figure 5). A total of 100 individual tasks were submitted at a time as array jobs but began execution at different times depending on the availability of requested hardware partitions and the job scheduler.

The average processing time for a single fishnet grid at each modeling level is shown in Figure 6, where quantization time consisted of centroid generation time and bilinear resampling time, and aggregation time comprised of parent-child navigation time and statistical summary time. As shown in Figure 6, most quantization and aggregation time was consumed on the bilinear resampling and parent-child navigation, respectively. Specifically, for each fishnet grid, the centroid generation and bilinear resampling at the finest level (i.e. level 29; 261 M rows) took 1015 and 16,567 seconds in the vegetation region, and 1032 and 27,695 seconds in the urban region. The parent-child navigation and statistical summary process took 7850 and 1145 seconds in the vegetation region, and 8388 and 1221 seconds in the urban region, for each fishnet grid at the finest aggregation level (i.e. level 28; 87 M rows). The total wall-clock time consumed to conduct quantization or aggregation over the full regions depended on the available nodes and memory on the hardware partitions at the time of the request. Quantization results at the

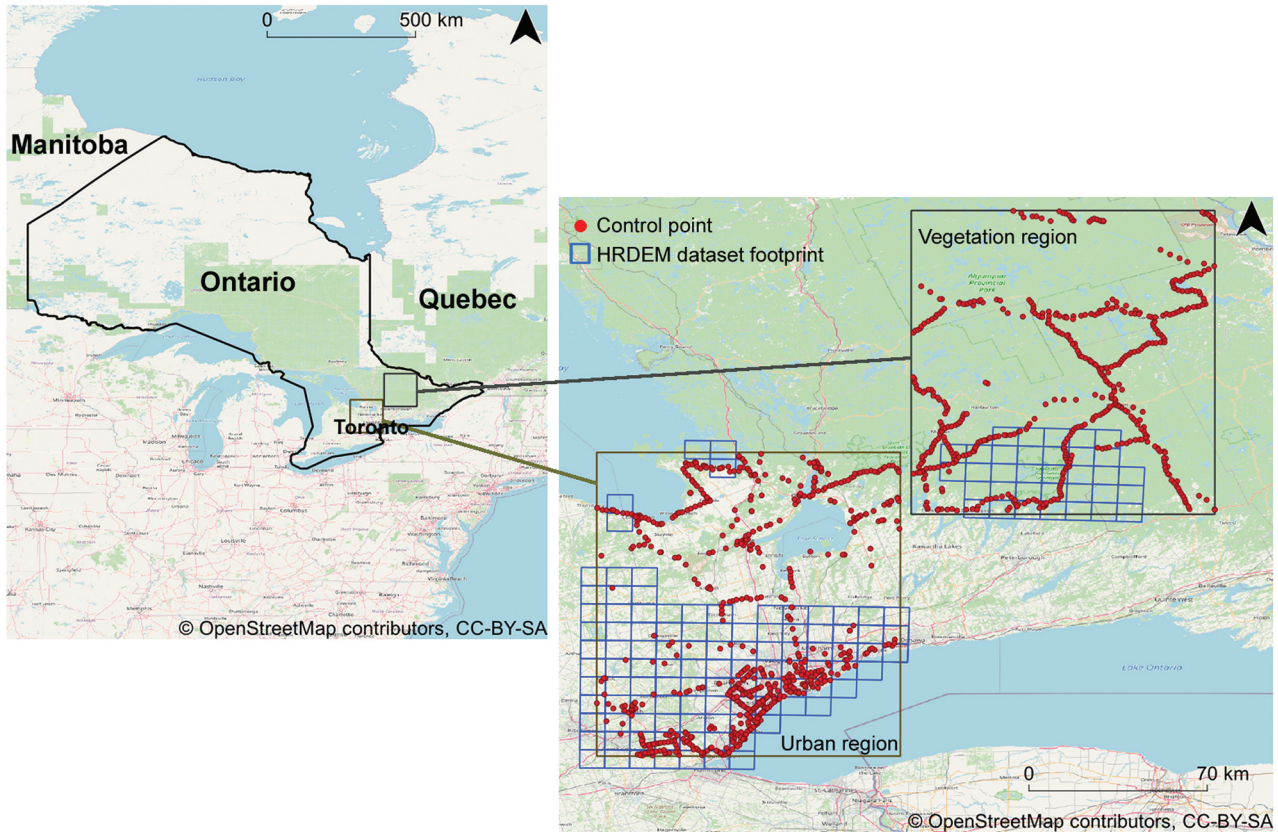


Figure 4. Study areas and the distribution of the HRDEM tiles and ground control points in Ontario, Canada.

finest resolution level are shown in Figure 7, with aggregation results at the coarsest resolution level illustrated in Figure 8.

Accuracy estimation

The ground control points were used as referencing points to evaluate the quality of elevation data modeling results, which were offered by the Ministry of Natural Resources and Forestry, Ontario (Ontario Ministry of Natural Resources and Forestry, 2020). Among all the available ground-surveyed data, only those first-order control points with the highest accuracy were considered in the accuracy estimation process. The ground control points used in this study were typically located on roadways and sidewalks, and were originally surveyed by the Canadian Geodetic Survey at NRCan, Ontario Ministry of Transportation, and Ontario Ministry of Natural Resources and Forestry. There were 510 control points within the vegetation region, 72 of which fell in the HRDEM dataset footprints

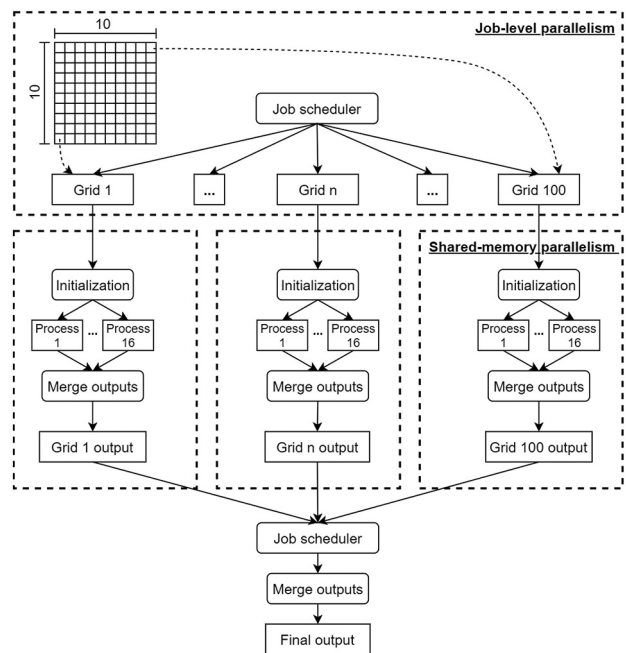


Figure 5. Job-level parallelism and shared-memory parallelism carried out for each of the study areas.

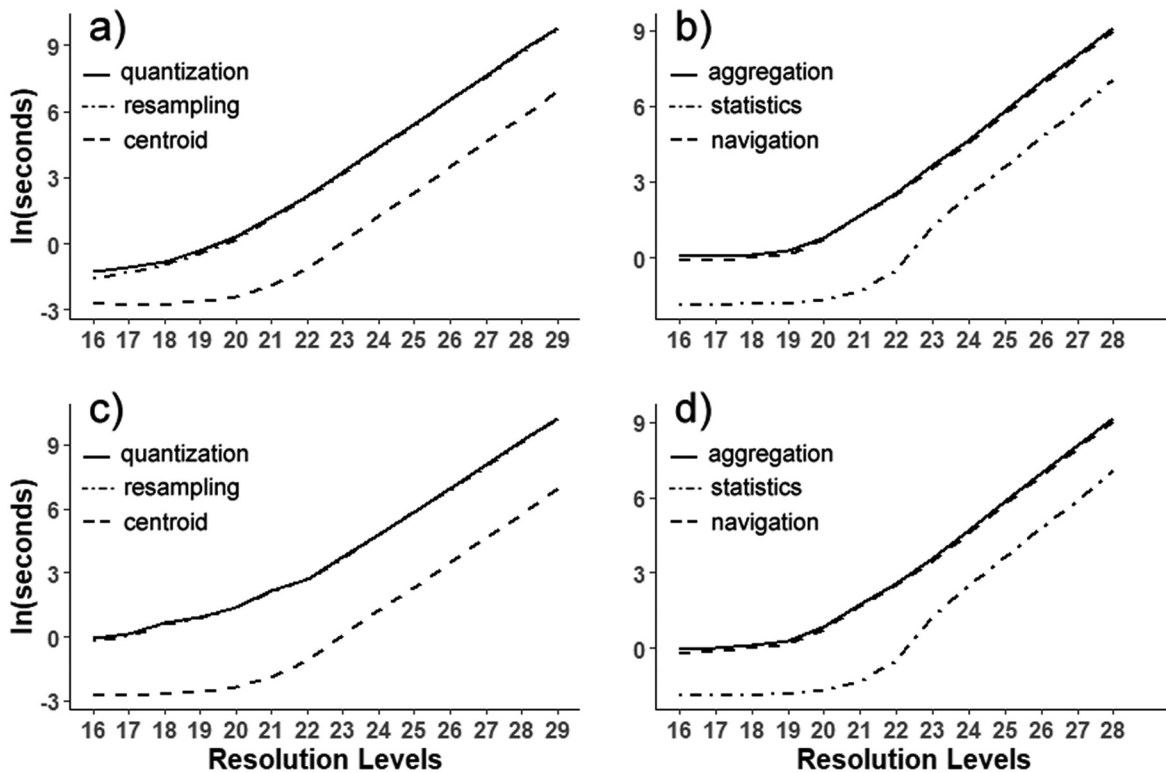


Figure 6. Average processing time of a) direct quantization in the vegetation region, b) statistical aggregation in the vegetation region, c) direct quantization in the urban region, and d) statistical aggregation in the urban region, for each fishnet grid at each modeling level.

(Figure 4). Within the urban region, 746 control points were available with 429 of them falling in the HRDEM extent (Figure 4).

Results showed that the calculated RMSE in the urban region were generally lower than those in the vegetation region (Table 4). Specifically, the RMSE of the ground-surveyed values versus the pre-DGGS values for those control points within and outside the HRDEM extent was 4.901 m and 6.410 m respectively over the vegetation region, and 3.679 m and 3.832 m respectively over the urban region. The original HRDEM and CDEM raster values standardized on CGVD2013 were involved in the calculation for points falling inside and outside the HRDEM extent. Comparing the ground-surveyed values with the post-DGGS values, the RMSE ranged from 6.233 m to 14.679 m in the vegetation region and 3.737 m to 9.625 m in the urban region among the modeling levels. A general trend of decreasing RMSE at finer levels was observed in both regions, although the trend was less evident at levels finer than 21 (Table 4). The RMSE of the pre-DGGS values versus

the post-DGGS values showed a decreasing tendency, despite not being strictly monotone, at finer resolution levels over both regions, where the RMSE at the finest level were 0.051 m and 0.039 m in the vegetation and urban regions, respectively (Table 4).

The spatially referenced metadata of the modeled elevations in DGGS was created in the GeoJSON format. Figure 9 shows the sample metadata of modeling results of the vegetation region at the finest level. The calculated RMSE were used as the accuracy indicators in the metadata, where the property “Source_Accuracy” represented the RMSE of the ground-surveyed values versus the pre-DGGS values, “Accuracy_GCP” was the RMSE of the ground-surveyed values versus the post-DGGS values, and “Accuracy_pre_DGGS” was the RMSE of the pre-DGGS values versus the post-DGGS values. Furthermore, the calculation of the “Accuracy_GCP” and “Accuracy_pre_DGGS” only used the ground control points inside or outside the HRDEM extent, depending on which of two polygon geometries they overlapped.

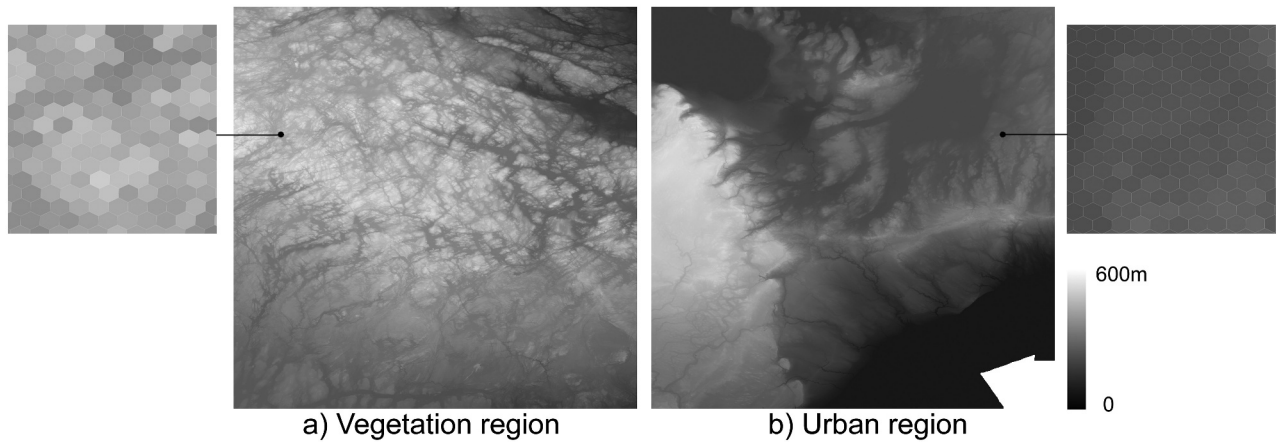


Figure 7. Quantization results in the a) vegetation region and b) urban region at the finest level in the ISEA3H DGGs.

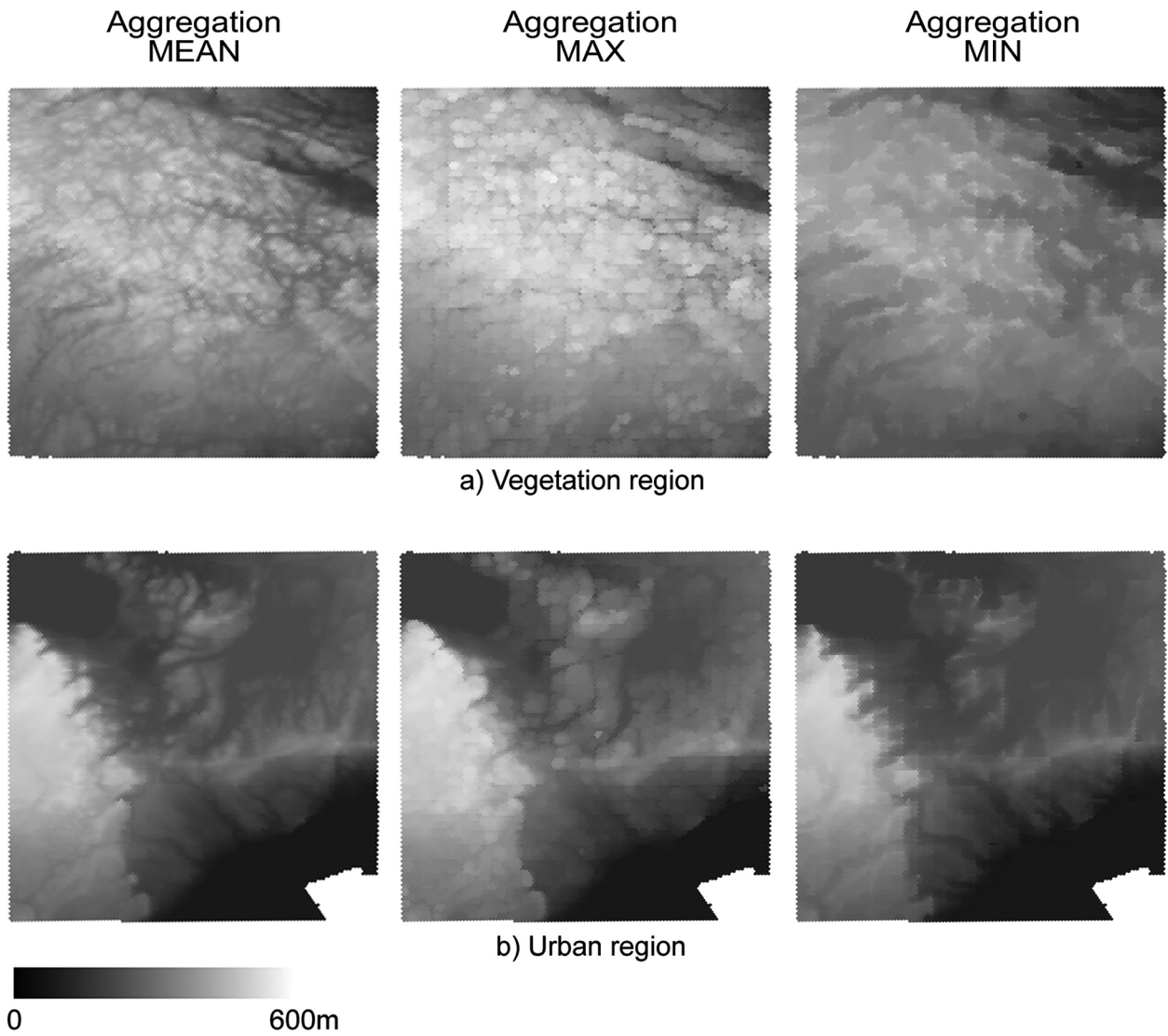


Figure 8. Aggregation results in the a) vegetation region and b) urban region at level 16 in the ISEA3H DGGs.

Table 4. The root mean square errors of the ground control points values versus post-DGGS values and pre-DGGS values versus post-DGGS values over the vegetation region and the urban region among the ISEA3H DGGS levels 16 to 29 (unit: m).

Resolution Level	GCP versus Post-DGGS		Pre-DGGS versus Post-DGGS	
	Vegetation	Urban	Vegetation	Urban
16	14.679	9.625	14.440	9.203
17	12.026	6.517	12.091	5.824
18	9.492	4.762	7.655	3.608
19	7.973	3.944	5.345	2.725
20	6.767	4.075	3.391	1.902
21	6.233	3.862	2.227	1.351
22	6.374	3.737	1.366	0.985
23	6.310	3.842	1.039	0.657
24	6.371	3.745	0.759	0.487
25	6.258	3.773	0.142	0.276
26	6.265	3.750	0.072	0.178
27	6.261	3.762	0.052	0.140
28	6.264	3.754	0.073	0.099
29	6.258	3.745	0.051	0.039

Discussion

Data quality

To set the stage for a DGGS-based elevation service in Canada, maintaining high data quality was one of the main considerations in this research. Data quality largely depends on the uncertainty of the spatial modeling process, where the uncertainty derives from inaccuracy and imprecision, meaning the lack of correlation between observations and reality, and the lack of detail in an observation (Worboys & Duckham, 2004).

Terrain data accuracies vary based on differences in primary data acquisition technologies, processing methodology, terrain roughness, and land cover types (Erdoğan, 2010; Gesch et al., 2014). In our case, the HRDEM had higher accuracy than CDEM because of its advanced data acquisition and processing technologies. The forested area with higher terrain roughness led to lower accuracies compared to a bare, flat landscape, which was in line with previous studies (Hobi & Ginzler, 2012; Mukherjee et al., 2013). The inaccuracy existing in the original data can be propagated through interpolation when the data were resampled and modeled in DGGS. With NRCan-released HRDEM and CDEM data as the baseline, this research managed to maintain their original fidelity and minimize the error propagation by adopting a closer, yet higher sample rate compared to the possible highest sample rate of CDEM and HRDEM data when modeling at the finest level in DGGS. Moreover, the ISEA3H tessellation used in this research guaranteed the monotone convergence in terms of the point displacement, i.e. a finer level resulted in less displacement between the original point location and the corresponding cell centroid location. This attribute makes the ISEA3H DGGS superior to other lattice

structures not converging monotonically with each refinement, for example, the square quadtree. However, one should note that the corresponding modeled values of a non-monotone geographic phenomenon (e.g. elevations) are not necessarily closer to the true value at a finer ISEA3H DGGS level. Bilinear interpolation was applied in this research when resampling DEM rasters onto hexagonal DGGS. Although it was suggested by the prior study (Ma et al., 2021), the most appropriate interpolation method should depend on the landforms of interest and specific scenarios. It is meaningful to extend the approach to a wider geographic area and test among different terrain types.

In terms of the terrain data precision, the ratio of horizontal resolution to vertical resolution of DEM has an impact on the real-world applications such as the topographical parameter computation (Garbrecht & Martz, 2000; Hengl & Evans, 2009). As mentioned earlier, at the finest level, this research provided elevations with six decimal places, which was the elevation precision of the original HRDEM data. The modeled elevations were gradually rounded to fewer decimal places at the coarser resolutions to reduce the unnecessary precision and to preserve a rough ratio of the horizontal to the vertical resolution among the DGGS levels. This was meaningful for the application of terrain data, where the generation of topographical features can avoid the potential local artifacts over a low relief landscape.

Computation efficiency

As conducted in our experiment, the job-level parallelism was executed in combination with shared-memory parallelism. Compared to the traditional spatial algorithms and serial computation, the discrete nature of a DGGS ensures that the shared-memory parallelism or even multi-processing parallelism can be achieved to make use of multiple processors accessing the same or different nodes to divide and run the full computation task (Peterson, 2016; Robertson et al., 2020; Yao et al., 2020). This is important for enabling a national elevation service in Canada considering the tremendous size of the existing datasets and more LiDAR-derived terrain data accessible in future. Although some parallel-database frameworks have been developed for big geospatial data (e.g. Eldawy et al., 2017; Jing et al., 2017), DGGS are promising because DGGS cells can be processed independently of each other. Consequently, the mechanism of parallel computing and level of parallelism are flexible and scalable to fit computation tasks. By integrating with cloud computing technologies, DGGS can adequately fit in the distributed storage and distributed computing in the context of big geospatial data (Yao et al., 2020).

```

1 {
2   "type": "FeatureCollection",
3   "crs": {
4     "type": "name",
5     "properties": {
6       "name": "EPSG:4617"
7     }
8   },
9   "features": [
10    {
11     "type": "Feature",
12     "id": 0,
13     "properties": {
14       "Source_Data": "Canadian Digital Elevation Model",
15       "Source_Institution": "Government of Canada; Natural Resources Canada;
16       Earth Sciences Sector; Canada Centre for Mapping and Earth Observation",
17       "Source_Licence": "Open Government Licence - Canada",
18       "Source_EPSG": 4617,
19       "Source_Resolution": "0.75 arcsec",
20       "Source_Accuracy": 6.410
21       "Vertical_Datum": "CGVD2013",
22       "Vertical_Unit": "Meter",
23       "No_Data": -32767,
24       "Polyhedral_Projection": "Icosahedral Snyder Equal Area Projection",
25       "Cell_Shape": "Hexagon",
26       "Aperture": 3,
27       "Orientation_Latitude": 37.6895
28       "Orientation_Longitude": -51.6218
29       "Orientation_Azimuth": -72.6482
30       "DGGS_Resolution": 29
31       "Cell_Area": 0.74
32       "Centroid_Spacing": 0.85
33       "Decimal_Places": 6
34       "Accuracy_GCP": 6.455
35       "Accuracy_pre_DGGS": 0.048
36     },
37     "geometry": {
38       "type": "Polygon",
39       "coordinates": [
40         [
41           [
42             -77.369352042562696,
43             46.0147577505698
44           ],
45           ...
46         ]
47       ]
48     },
49     {
50     "type": "Feature",
51     "id": 1,
52     "properties": {
53       "Source_Data": "High Resolution Digital Elevation Model",
54       "Source_Institution": "Government of Canada; Natural Resources Canada;
55       Earth Sciences Sector; Canada Centre for Mapping and Earth Observation",
56       "Source_Licence": "Open Government Licence - Canada",
57       "Source_EPSG": 26917,
58       "Source_Resolution": "1 m",
59       "Source_Accuracy": 4.901
60       "Vertical_Datum": "CGVD2013",
61       "Vertical_Unit": "Meter",
62       "No_Data": -32767,
63       "Polyhedral_Projection": "Icosahedral Snyder Equal Area Projection",
64       "Cell_Shape": "Hexagon",
65       "Aperture": 3,
66       "Orientation_Latitude": 37.6895
67       "Orientation_Longitude": -51.6218
68       "Orientation_Azimuth": -72.6482
69       "DGGS_Resolution": 29
70       "Cell_Area": 0.74
71       "Centroid_Spacing": 0.85
72       "Decimal_Places": 6
73       "Accuracy_GCP": 4.757
74       "Accuracy_pre_DGGS": 0.066
75     },
76     "geometry": {
77       "type": "Polygon",
78       "coordinates": [
79         [
80           [
81             -78.136770200384376,
82             44.915028999988408
83           ],
84           ...
85         ]
86       ]
87     }
88   ]
89 }

```

Figure 9. Sample metadata of modeling results over the tested vegetation area at the finest level in the ISEA3H DGGS.

With regard to the quantization method, bilinear interpolation was used to resample DGGS cell centroids in this research. Among nearest-neighbor, bilinear, and bicubic interpolation, bilinear interpolation requires less complex computation than bicubic interpolation while providing more accurate, continuous results than the nearest-neighbor approach. In other words, a trade-off has to be achieved between the computation complexity and the resulting accuracy. Furthermore, edge effects were noticed during quantization, where DGGS cell centroids close to the boundary of an area of interest suffer from not having enough valid inputs for interpolation and end up with invalid values. Typically, this issue can be encountered when applying bilinear interpolation and cubic convolution. This study managed to avoid this issue by assigning a void value to

the target centroid unless enough valid neighbors were found. As a result, the border region of the area of interest can remain as void after quantization especially at a coarse level.

One main limitation of the computation efficiency in this research is the navigation among parent and child cells when conducting statistical aggregation between levels. The computation complexity of hierarchical navigation is largely based on the cell indexing mechanism of a DGGS implementation. The R library *dggridR* used in this study does not support functions for direct navigation across hierarchies on the ISEA3H tessellation, and the linear addresses of cells were not organized in a way where a parent cell and its child cells share the common pattern such as a prefix (Barnes & Sahr, 2017; Bondaruk et al., 2020). As a result, the parent-child cell index look-up tables were computed separately before the aggregation, requiring additional time. In contrast, a DGGS library based on a completely hierarchical index scheme can facilitate parent-child navigation. For instance, the H3 library developed by Uber Technologies Inc. adopts a nested one-to-seven hexagonal tessellation and assigns additional digits zero to six at each finer resolution based on its parent cell's address via a Central Place Indexing arrangement (Sahr, 2019; Uber, 2017). This allows an advanced hierarchical query across resolutions, especially given the fact that the H3 library has developed a binding to ClickHouse, which is a column-oriented database management system supporting online analytical processing of queries (Uber, 2017; Yandex, 2020).

Impact and future work

This study demonstrated the potential of DGGS as an integration platform for national elevation data across various scales and reference systems in Canada. With the DGGS-driven platform, the HRDEM and CDEM data are integrated and standardized to provide complete coverage over the country and improve data quality relative to original CDEM data. The proposed processing flow can serve as a reference for the development of a future national elevation service, which aims to offer consistent and multi-resolution elevations and save end-user time on pre-processing tasks. Over the rivers and lakes, voids on the original HRDEM data are filled by their known or estimated values from the CDEM data, which ensures the availability of uninterrupted terrain data over the country. Furthermore, multiple aggregation methods were proposed, where the minimum and maximum elevation products at different resolution levels can meet various application purposes.

An important direction of future work is to develop in-database analytical functions in the ISEA3H DGGS. Currently, applying analysis algorithms of square grids to the spheroidal equal angular DEM can result in computational errors. In-database analytical functions would allow geospatial analysis to be undertaken directly in the hexagonal DGGS, eliminating the need to convert the DGGS data back to the traditional raster model and applying the algorithms developed for square grids (Hojati & Robertson, 2020). In this way, one can avoid working with the projected terrain rasters and take the Earth's curvature into account. Tomlin's model, namely the local, focal, and zonal operations, needs to be extended in the context of DGGS (Tomlin, 1990). Specifically, focal operations in a hexagonal DGGS need to consider the special cell connectivity characteristics of hexagons (Li & Stefanakis, 2020b). The moving window can be defined as the number of rings or a distance (Li & Stefanakis, 2020b; Robertson et al., 2020). As well, new algorithms and equations need to be developed for topographical and hydrological analysis. For example, to compute slope or aspect in a hexagonal DGGS, equal weights should be given to six neighbors for each center cell (Hojati & Robertson, 2020; Li & Stefanakis, 2020b). Hydrological functions such as flow direction and accumulation analysis require new algorithms due to the feature of uniform cell adjacency on a hexagonal tessellation (Liao et al., 2020; Wang & Ai, 2018). Topographical and hydrological features can also be extracted in the hexagonal DGGS and compared against those generated from traditional rasters, which can serve as additional evaluation indicators.

Another future research direction is the extension of the elevation-integration platform to a DGGS-powered geospatial datacube. A traditional geospatial datacube contains multiple dimensional arrays where spatio-temporal values of observations are stored along each dimension (Baumann et al., 2018). As congruent geography, a geospatial datacube can be built upon a DGGS around a data repository by adopting the underlying spatial architecture of DGGS, providing access to analysis-ready data and offering more comprehensive datacube solutions (Goodchild, 2018; Purss et al., 2019). In Canada, a DGGS-powered geospatial datacube can contain terrain data as well as Earth observation data along time series. The quality of modeling remote sensing data on hexagonal grids can be evaluated by basic indicators such as information entropy, image features like edges, and other geometric evaluation indices (Ma et al., 2021). Several on-going projects are making attempts to apply such an infrastructure such as EO4wildlife (EO4wildlife, 2020). Future research may focus on standardizing the implementation details, improving the quality of

spatio-temporal data, accelerating data accesses or queries, enabling advanced spatio-temporal analysis, developing appropriate service interfaces, and reaching larger adoption and acceptance (Baumann et al., 2018; Purss et al., 2019; Salehi et al., 2007).

Conclusions

Adopting the ISEA3H DGGS, we proposed methods to integrate multi-source terrain data in Canada. The ISEA3H tessellation was preferred mainly because of its reduced projection distortion as well as the monotonical convergence property. The DGGS was orientated to locate Canadian territory at the center of one single icosahedral face. The CDEM and HRDEM were converted to the uniform horizontal and vertical datums and quantized in the ISEA3H DGGS at various resolution levels. With quantization results at the finest level as the baseline, this research also executed statistical aggregations at coarser levels to fulfill the potential needs of local minimum or maximum elevation usage. Applying the experiment over two study areas, we demonstrated the full modeling process and carried out accuracy estimation by comparing the ground-surveyed elevations, pre-DGGS elevations, and post-DGGS elevations. The results showed a general improvement of accuracy with higher DGGS resolutions where areas of variable elevation (e.g. vegetation) resulted in lower accuracies in comparison with flat landscapes. Due to the discreteness of DGGS cells, computation tasks can be accelerated by parallelism. Nonetheless, the cell indexing mechanism implemented by a DGGS has a great impact on the computation efficiency, especially the aggregation across resolution levels. In summary, this study contributed to multi-resolution elevations by integrating multi-source data in DGGS and improved the coverage and data quality of single-source data. It set the stage for a national elevation service across various scales for Canada. Future work directions include the development of in-database DGGS analytics, exploration of supporting real-world decision-making, and extending outcomes to build a DGGS-powered geospatial datacube.

Acknowledgments

We acknowledge the technical advice from the Research Computing Services at the University of Calgary. We appreciate Natural Resources Canada with the contribution number: 20210033. We also thank Dr. Titus Tienaah for his valuable insights.

Data availability statement

The data that support the findings of the experiment in this study are available with the identifier 10.5281/zenodo.4718019. The original Canadian Digital Elevation Model data can be obtained via the Geospatial-Data Extraction tool in Canada's Open Government Portal: <https://maps.canada.ca/czs/index-en.html>.

The High Resolution Digital Elevation Model data (Digital Terrain Model) are available at: https://ftp.maps.canada.ca/pub/elevation/dem_mne/highresolution_hauteresolution/dtm_mnt. Ground control points are accessible using COSINE, Ontario's geodetic control database, on the website of Ontario Ministry of Natural Resources and Forestry: <https://www.liaapplications.lrc.gov.on.ca/COSINE/index.html?viewer=COSINE.OntarioViewer&locale=en-CA>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Canadian Natural Sciences and Engineering Research Council (NSERC) Discovery Grant program. under Grant RGPIN/03977-2019.

ORCID

Mingke Li  <http://orcid.org/0000-0001-6310-4964>
 Heather McGrath  <http://orcid.org/0000-0002-6439-3339>
 Emmanuel Stefanakis  <http://orcid.org/0000-0002-0062-5647>

References

- Alderson, T., Purss, M., Du, X., Mahdavi-Amiri, A., & Samavati, F. (2020). Digital earth platforms. In H. Guo, M. Goodchild, & A. Annoni (Eds.), *Manual of digital earth* (pp. 25–54). Springer. https://doi.org/10.1007/978-981-32-9915-3_2
- Arundel, S. T., Bulen, A. N., Adkins, K. F., Brown, R. E., Lowe, A. J., Mantey, K. S., & Phillips, L. A. (2017). Assimilation of the national elevation dataset and launch of the 3D elevation program through the USGS spatial data infrastructure. *International Journal of Cartography*, 4(2), 129–150. <https://doi.org/10.1080/23729333.2017.1288533>
- Barnes, R. (2019). Optimal orientations of discrete global grids and the poles of inaccessibility. *International Journal of Digital Earth*, 13(7), 803–816. <https://doi.org/10.1080/17538947.2019.1576786>
- Barnes, R., & Sahr, K. (2017). dggridR: Discrete Global Grids for R. R package version 2.0.4. <https://github.com/r-barnes/dggridR>
- Baumann, P., Misev, D., Merticariu, V., & Huu, B. P. (2018). Datacubes: Towards space/time analysis-ready data. In J. Doellner, M. Jobst, & P. Schmitz (Eds.), *Service oriented mapping - changing paradigm in map production and geoinformation management* (pp. 269–299). Springer. https://doi.org/10.1007/978-3-319-72434-8_14
- Bednar, J. A., Crist, J., Cottam, J., & Wang, P. (2016). *Datashader: Revealing the structure of genuinely big data* [Paper presentation]. 15th Python in Science Conference, Austin, Texas.
- Bernardin, T., Cowgill, E., Kreylos, O., Bowles, C., Gold, P., Hamann, B., & Kellogg, L. H. (2010). Crusta: A new virtual globe for real-time visualization of sub-meter digital topography at planetary scales. *Computers & Geosciences*, 37(1), 75–85. <https://doi.org/10.1016/j.cageo.2010.02.006>
- Bondaruk, B., Roberts, S. A., & Robertson, C. (2020). Assessing the state of the art in discrete global grid systems: OGC criteria and present functionality. *Geomatica*, 74(1), 9–30. <https://doi.org/10.1139/geomat-2019-0015>
- Conway, J. H., & Sloane, N. J. A. (1998). *Sphere packings, lattices and groups*. Springer Science & Business Media.
- Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (Open-File Report 2011-1073). Earth Resources Observation and Science (EROS) Center, U.S. Geological Survey. <https://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf>
- Dutton, G. (1984). Part 4: Mathematical, algorithmic and data structure issues: Geodesic modelling of planetary relief. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 21(2–3), 188–207. <https://doi.org/10.3138/R613-191U-7255-082N>
- Dutton, G. (1988). Computational aspects of a quaternary triangular mesh. In G. H. Dutton (Ed.), *A hierarchical coordinate system for geoprocessing and cartography* (pp. 41–70). Springer-Verlag. <https://doi.org/10.1007/BFb0011620>
- Eldawy, A., Sabek, I., Elganainy, M., Bakeer, A., Abdelmotaleb, A., & Mokbel, M. F. (2017). Sphinx: Empowering impala for efficient execution of SQL queries on big spatial data. In M. Gertz, M. Renz, X. Zhou, E. Hoel, W.-S. Ku, A. Voisard, C. Zhang, H. Chen, L. Tang, Y. Huang, C.-T. Lu, & S. Ravada (Eds.), *Advances in spatial and temporal databases* (pp. 65–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-64367-0_4
- EO4wildlife. (2020). *EO4wildlife project*. Retrieved January 15, from <http://www.eo4wildlife.eu/>
- Erdogan, S. (2010). Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study. *Computers & Geosciences*, 36(1), 34–43. <https://doi.org/10.1016/j.cageo.2009.06.005>
- Fekete, G., & Treinish, L. A. (1990). Sphere quadtrees: A new data structure to support the visualization of spherically distributed data. In *Proceedings of SPIE 1259, extracting meaning from complex data: Processing, display, interaction* (pp. 242–253). International Society for Optics and Photonics. <https://doi.org/10.1117/12.19991>
- Florinsky, I. V. (2017). Spheroidal equal angular DEMs: The specificity of morphometric treatment. *Transactions in GIS*, 21(6), 1115–1129. <https://doi.org/10.1111/tgis.12269>
- Fuller, R. B. (1982). *Synergetics: Explorations in the geometry of thinking*. MacMillan Publishing Co. Inc.
- Garbrecht, J., & Martz, L. W. (2000). *Digital elevation model issues in water resources modeling*. In Proceedings of 1999 Esri User Conference. Esri. <https://proceedings.esri.com/library/userconf/proc99/proceed/papers/pap866/p866.htm>
- Gesch, D. B., Oimoen, M. J., & Evans, G. A. (2014). *Accuracy assessment of the U.S. Geological Survey national elevation dataset, and comparison with other large-area elevation datasets: SRTM and ASTER* (Open-File Report 2014-1008). Earth Resources Observation and Science (EROS) Center, U.S. Geological Survey. <https://pubs.usgs.gov/of/2014/1008/pdf/ofr2014-1008.pdf>

- Global Grid Systems. (2019). *Global Grid Systems*. Retrieved November 20, from <https://www.globalgridsystems.com>
- Golay, M. J. E. (1969). Hexagonal parallel pattern transformations. *IEEE Transactions on Computers*, C-18(8), 733–740. <https://doi.org/10.1109/T-C.1969.222756>
- Goodchild, M. F. (2018). Reimagining the history of GIS. *Annals of GIS*, 24(1), 1–8. <https://doi.org/10.1080/19475683.2018.1424737>
- Hengl, T., & Evans, I. S. (2009). *Mathematical and digital models of the land surface (Geomorphometry-concepts, software, applications)* (pp. 31–63). Elsevier. [https://doi.org/10.1016/s0166-2481\(08\)00002-0](https://doi.org/10.1016/s0166-2481(08)00002-0)
- Hobi, M. L., & Ginzler, C. (2012). Accuracy assessment of digital surface models based on WorldView-2 and ADS80 stereo remote sensing data. *Sensors (Basel)*, 12(5), 6347–6368. <https://doi.org/10.3390/s120506347>
- Hojati, M., & Robertson, C. (2020). Integrating cellular automata and discrete global grid systems: A case study into wildfire modelling. In *Proceedings of the 23rd AGILE conference on geographic information science*. Copernicus Publications. <https://doi.org/10.5194/agile-giss-1-6-2020>
- Jing, W., Huo, S., Miao, Q., & Chen, X. (2017). A model of parallel mosaicking for massive remote sensing images based on spark. *IEEE Access*, 5, 18229–18237. <https://doi.org/10.1109/ACCESS.2017.2746098>
- Lambers, M., & Kolb, A. (2012). Ellipsoidal cube maps for accurate rendering of planetary-scale terrain data. In C. Bregler, P. Sander, & M. Wimmer (Eds.). *Proceedings of the Pacific conference on computer graphics and applications* (pp. 5–10). The Eurographics Association. <https://doi.org/10.2312/PE/PG/PG2012short/005-010>
- Li, M., & Stefanakis, E. (2020a). Geo-feature modeling uncertainties in discrete global grids: A case study of downtown Calgary, Canada. *Geomatica*, 74(4), 175–195. <https://doi.org/10.1139/geomat-2020-0011>
- Li, M., & Stefanakis, E. (2020b). Geospatial operations of discrete global grid systems—A comparison with traditional GIS. *Journal of Geovisualization and Spatial Analysis*, 4(2), 26. <https://doi.org/10.1007/s41651-020-00066-3>
- Liao, C., Tesfa, T., Duan, Z., & Leung, L. R. (2020). Watershed delineation on a hexagonal mesh grid. *Environmental Modelling & Software*, 128, 104702. <https://doi.org/10.1016/j.envsoft.2020.104702>
- Luczak, E., & Rosenfeld, A. (1976). Distance on a hexagonal grid. *IEEE Transactions on Computers*, 25(5), 532–533. <https://doi.org/10.1109/TC.1976.1674642>
- Ma, Y., Li, G., Yao, X., Cao, Q., Zhao, L., Wang, S., & Zhang, L. (2021). A precision evaluation index system for remote sensing data sampling based on hexagonal discrete grids. *ISPRS International Journal of Geo-Information*, 10(3), 3. <https://doi.org/10.3390/ijgi10030194>
- Mahdavi-Amiri, A., Alderson, T., & Samavati, F. (2015). A survey of digital earth. *Computers and Graphics*, 53 (Part B), 95–117. <https://doi.org/10.1016/j.cag.2015.08.005>
- McKerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. G. (2011). Building a framework for predictive science. In S. V. D. Walt & J. Millman (Eds.). *Proceedings of the 10th Python in science conference (SciPy 2011)* (pp. 76–86). SciPy. <https://doi.org/10.25080/Majora-ebaa42b7-00d>
- Mersereau, R. M. (1979). The processing of hexagonally sampled two-dimensional signals. In *Proceedings of the IEEE* (pp. 930–949). IEEE.
- Mocnik, F.-B. (2019). *Geogrid*. Retrieved March 5, from <https://github.com/giscience/geogrid>
- Mukherjee, S., Joshi, P. K., Mukherjee, S., Ghosh, A., Garg, R. D., & Mukhopadhyay, A. (2013). Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *International Journal of Applied Earth Observation and Geoinformation*, 21, 205–217. <https://doi.org/10.1016/j.jag.2012.09.004>
- Natural Resources Canada. (2013). *Canadian digital elevation model product specifications*. Retrieved March 1, from https://ftp.maps.canada.ca/pub/nrcan_rncan/elevation/cdem_mnec/doc/CDEM_product_specs.pdf
- Natural Resources Canada. (2015). *Canadian digital elevation model, 1945-2011 [Grid]*. https://ftp.maps.canada.ca/pub/nrcan_rncan/elevation/cdem_mnec/
- Natural Resources Canada. (2017). *Geospatial data extraction*. Retrieved October 15, from <https://maps.canada.ca/czs/index-en.html>
- Natural Resources Canada. (2019). *High Resolution Digital Elevation Model (HRDEM) – CanElevation Series – Product specifications*. Retrieved March 1, from https://ftp.maps.canada.ca/pub/elevation/dem_mne/highresolution_hauteresolution/HRDEM_Product_Specification.pdf
- Natural Resources Canada. (2020). *High Resolution Digital Elevation Model (HRDEM)-CanElevation Series [Grid]*. Canada Centre for Mapping and Earth Observation, Strategic Policy and Results Sector, Natural Resources Canada, Government of Canada. https://ftp.maps.canada.ca/pub/elevation/dem_mne/highresolution_hauteresolution/dtm_mnt/
- Ontario Ministry of Natural Resources and Forestry. (2016). *Ontario land cover compilation v.2.0 [Grid]*. Ministry of Natural Resources and Forestry, Science and Research Branch, Natural Resource Information Section, Forest Resource Information Unit. <https://ws.gisett.lrc.gov.on.ca/fme/datadownload/Packages/OntarioLandCoverComp-v2.zip>
- Ontario Ministry of Natural Resources and Forestry. (2020). *COSINE online service*. Ministry of Natural Resources and Forestry, Science and Research Branch, Natural Resource Information Section, Forest Resource Information Unit. <https://www.lioapplications.lrc.gov.on.ca/COSINE/index.html?viewer=COSINE.OntarioViewer&locale=en-CA>
- Open Geospatial Consortium. (2017). *Topic 21: Discrete global grid system abstract specification*. Retrieved November 15, from <http://www.opengis.net/doc/AS/dggs/1.0>
- Petersen, D. P., & Middleton, D. (1962). Sampling and reconstruction of wave-number-limited functions in n-dimensional Euclidean spaces. *Information and Control*, 5(4), 279–323. [https://doi.org/10.1016/S0019-9958\(62\)90633-2](https://doi.org/10.1016/S0019-9958(62)90633-2)
- Peterson, P. (2016). Discrete global grid systems. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds.), *The international encyclopedia of geography* (pp. 1–10). John Wiley & Sons Ltd. <https://doi.org/10.1002/9781118786352.wbieg1050>
- Porter, C., Morin, P., Howat, I., Noh, M.-J., Bates, B., Peterman, K., Keeseey, S., Schlenk, M., Gardiner, J., Tomko, K., Willis, M., Kelleher, C., Cloutier, M., Husby, E., Foga, S., Nakamura, H., Platson, M., Wethington, M., Jr., Williamson, C., Bauer, G., ... Bojesen, M. (2018). *ArcticDEM [Grid]*. Harvard Dataverse. <https://doi.org/10.7910/DVN/OHHUKH>
- Purss, M., Peterson, P., Strobl, P., Dow, C., Sabeur, Z., Gibb, R., & Ben, J. (2019). Datacubes: A discrete global grid systems perspective. *Cartographica the International Journal for Geographic Information and Geovisualization*, 54(1), 63–71. <https://doi.org/10.3138/cart.54.1.2018-0017>

- PYXIS. (2020). *Quantization*. Retrieved April 9, from <http://sdk.pyxisinnovation.com/quantization>
- Robertson, C., Chaudhuri, C., Hojati, M., & Roberts, S. A. (2020). An integrated environmental analytics system (IDEAS) based on a DGGS. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 214–228. <https://doi.org/10.1016/j.isprsjprs.2020.02.009>
- Safanelli, J. L., Poppiel, R. R., Ruiz, L. F. C., Bonfatti, B. R., Mello, F. A. D. O., Rizzo, R., & Demattê, J. A. M. (2020). Terrain analysis in Google Earth engine: A method adapted for high-performance global-scale analysis. *ISPRS International Journal of Geo-Information*, 9, 400. <https://doi.org/10.3390/ijgi9060400>
- Sahr, K., & White, D. (1998). Discrete global grid systems. In S. Weisberg (Ed.), *Computing science and statistics (Volume 30): Proceedings of the 30th symposium on the interface, computing science and statistics* (pp. 269–278). Interface Foundation of North America.
- Sahr, K. (2011). Hexagonal discrete global grid systems for geospatial computing. *Archives of Photogrammetry, Cartography and Remote Sensing*, 22, 363–376. <https://www.infona.pl/resource/bwmeta1.element.baztech-8fbd4fa0-6092-44bb-83ff-52f2ce89f7e6/tab/summary#1>
- Sahr, K. (2019). Central place indexing: Hierarchical linear indexing systems for mixed-aperture hexagonal discrete global grid systems. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 54(1), 16–29. <https://doi.org/10.3138/cart.54.1.2018-0022>
- Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 30(2), 121–134. <https://doi.org/10.1559/152304003100011090>
- Salehi, M., Bédard, Y., Mostafavi, M., & Brodeur, J. (2007). From transactional spatial databases integrity constraints to spatial data cubes integrity constraints. In A. Stein (Ed.), *Proceedings of the 5th international symposium on spatial data quality*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Schumann, G. J. P., & Bates, P. D. (2018). The need for a high-accuracy, open-access global DEM. *Frontiers in Earth Science*, 6, 225. <https://doi.org/10.3389/feart.2018.00225>
- Staunton, R. C. (1989). The design of hexagonal sampling structures for image digitization and their use with local operators. *Image and Vision Computing*, 7(3), 162–166. [https://doi.org/10.1016/0262-8856\(89\)90040-1](https://doi.org/10.1016/0262-8856(89)90040-1)
- Tomlin, C. D. (1990). *Geographic Information Systems and cartographic modeling*. Prentice Hall.
- Uber. (2017). *H3: A hexagonal hierarchical geospatial indexing system*. Retrieved November 25, from <https://github.com/uber/h3>
- Wang, L., & Ai, T. (2018). The comparison of drainage network extraction between square and hexagonal grid-based DEM. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(4), 687–692. <https://doi.org/10.5194/isprs-archives-XLII-4-687-2018>
- Microsoft Corporation & Weston, S. (2020). *doParallel: Foreach parallel adaptor for the 'parallel' package*. R package version 1.0.16. Retrieved March 3, from <https://CRAN.R-project.org/package=doParallel>
- White, D., Kimerling, A. J., Sahr, K., & Song, L. (1998). Comparing area and shape distortion on polyhedral-based recursive partitions of the sphere. *International Journal of Geographical Information Science*, 12(8), 805–827. <https://doi.org/10.1080/136588198241518>
- Worboys, M., & Duckham, M. (2004). *GIS: A computing perspective*. CRC Press.
- Yandex. (2020). *ClickHouse*. Retrieved March 12, from <https://clickhouse.tech/>
- Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L., & Zhu, D. (2020). Enabling the big earth observation data via cloud computing and DGGS: Opportunities and challenges. *Remote Sensing*, 12(1), 62. <https://doi.org/10.3390/rs12010062>
- Zhou, J., Ben, J., Wang, R., Zheng, M., Yao, X., & Du, L. (2020). A novel method of determining the optimal polyhedral orientation for discrete global grid systems applicable to regional-scale areas of interest. *International Journal of Digital Earth*, 13(12), 1553–1569. <https://doi.org/10.1080/17538947.2020.1748127>