A simplified geographical information systems (GIS)-based methodology for modeling the topography of bedrock: illustration using the Canadian Shield

Romain Chesnaux[1] · Mélanie Lambert[1] · Julien Walter[1] · Vianney Dugrain[1] ·
Alain Rouleau[1] · Réal Daigneault[1]

Abstract Many geology, mining, and geotechnical applications require or depend upon some form of modeling of bedrock topography. Optimizing the manner with which bedrock topography is modeled poses a significant challenge because of the unpredictable or erratic presentation of the surface shape of bedrock. Unlike surface topography, bedrock topography is more difficult to determine because direct observation points are often not readily or directly accessible, unless the bedrock outcrops at the surface and is exposed, a relatively rare occurrence. When bedrock is covered by granular deposits, the only methods that allow practitioners to objectively establish the location of the top of the bedrock are to drill boreholes or conduct geophysical surveys. This makes the determination of bedrock topography not only difficult but also expensive. This study proposes a new approach for optimizing the modeling of complex bedrock topography, whose originality is based on the addition of "virtual" data points derived from cross-sections located between known boreholes. The proposed methodology is thus composed of four steps: gathering the maximum amount of relevant surface and subsurface data (from observation points), selecting the most appropriate technique for interpolating the observed bedrock elevations that will be entered into the dataset to be modeled, enriching the quantity of modeling data by adding "virtual" data elements based on geological interpretations of cross-sections (inserted into the model alongside the original objective data), and finally the modeling itself. The proposed approach is illustrated using data from a study area located in the Canadian Shield. Thousands of borehole records and surficial geological data as well as geological cross-section records were integrated to construct a three-dimensional bedrock topography model. The new proposed methodology can be applied to other regions worldwide.

Keywords Interpolation methods · Bedrock topography · Spatial database · ArcGIS · Concept of virtual boreholes

Introduction

Modeling the geometry of geological formations is relevant for a variety of applications: mineral exploration (Paulen and McClenaghan 2003) groundwater resource assessment (Carmichael and Henry 1977) and geotechnical investigations (Elmahdy et al. 2010), among others. The bedrock forms the base of all geologic deposit environments. When the bedrock does not outcrop at the surface, knowledge of its topography is of crucial interest when assessing the thickness of the overlaying deposits. Modeling the topography of the bedrock when it does not outcrop at the surface is a challenge. Knowledge of the bedrock location is provided either by borehole observations that require drilling operations (invasive prospection) or by non-invasive geophysical measurements (Ugwu and Eze 2009; Farinotti et al. 2014; Tremblay Simard et al. 2015). Usually, only a limited number of observation points of the bedrock elevation are available at a regional scale. The observation points consist of boreholes, outcrops, and occasionally geophysical observations. The ability to determine an appropriate methodology for interpolating the bedrock elevation that exists between these observation points is the key factor for establishing a suitably representative layout of the bedrock topography. The observation data points available to the

✉ Romain Chesnaux
romain_chesnaux@uqac.ca

1 Centre d'Études sur les Ressources Minérales, Research group « Risk Resource Water », Université du Québec à Chicoutimi, 555, Bd de l'Université, Chicoutimi, Quebec G7H 2B1, Canada

modeler are obtained either from visual observations when the bedrock outcrops at the surface or from borehole observations when the bedrock is covered by granular deposits (Van Hoesen 2014). These borehole observations may come from groundwater well logs, exploration borehole logs, or geotechnical drilling logs. The critical question to be asked by the modeler is if the number and distribution of observation points are sufficient to allow a precise and accurate interpolation of the data. The key factors required for a successful (sufficiently representative and accurate) model of bedrock topography are the quantity (or density) of observation points and their distribution.

The root mean square error (RMS, Eq. 1) is the most frequently used indicator to assess modeling quality in terms of both accuracy and precision (MacCormack et al. 2011; Zimmerman et al. 1999; Jones et al. 2003; Wise 2000). As shown in Eq. 1, the RMS parameter corresponds to the mean of the differences between the observed and interpolated values (of the bedrock elevation in the case of our study). The RMS value consequently indicates the reliability of the model to represent reality. The lower the RMS, the higher the quality of the model; a lower value of RMS indicates a better reliability of the model in terms of accuracy and precision.

$$\mathrm{RMS} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{model}_i - \mathrm{observed}_i\right)^2\right)^{\frac{1}{2}} \tag{1}$$

Thus, the quality of an interpolation depends first and foremost upon the quality of the initial dataset in terms of density and spatial distribution of the observation points (Carter 1992; Chang and Tsai 1991). Figure 1 represents and allows us to visualize the resolution quality of the interpolation (accuracy) that can be obtained depending on the dataset properties: quantity of data ("low vs. high") and spatial distribution type (regular, random, or clustered). The modeling accuracy is represented by the size of the triangles in Fig. 1, which shows that a large dataset with a regular distribution of observation points makes it possible to reach a greater degree of modeling accuracy (larger size of the triangles in Fig. 1). Conversely, a smaller dataset with a clustered distribution yields a lesser degree of modeling accuracy (smaller size of the triangles in Fig. 1). Interpolation accuracy and consequently modeling quality also depend on variations in elevation of the bedrock that is being modeled. A flat area will be modeled more accurately than an area of more complex elevations (composed of marked valleys and peaks). Figure 1 also illustrates these differences and their impacts on the expected relative accuracy of the model, with Fig. 1a corresponding to a simple bedrock surface and Fig. 1b corresponding to a complex bedrock surface. For both types of surfaces, the more regular the spatial distribution of the data, the higher the accuracy of the interpolated model. As the relief increases in complexity, the quality

of the spatial distribution of data becomes more critical in order to obtain a good quality model. Figure 1b shows that a complex surface cannot be satisfyingly modeled from source data that are strongly clustered (note the absence of a triangle in the lower part of the figure), independently of the quantity of data contained in the dataset.

The quality of the interpolated result also depends on the chosen interpolation algorithm (Chaplot et al. 2006). When the dataset is of good quality, (regular distribution and high density), this choice is less critical and will have limited impact on the overall quality of the results (Carrara et al. 1997). Several authors have shown that the ordinary kriging (OK) interpolation method and the inverse distance weighting (IDW) interpolation method both yield the best results for modeling geological formations (Darsow et al. 2009; Slattery et al. 2011; Van Hoesen 2014). Other works have led to the same conclusion (Chaplot et al. 2006; Wise 2000) when comparing different interpolation methods for modeling surface topography and producing a digital elevation model (DEM). The same could be expected when modeling bedrock topography, considering that bedrock topography has a behavior similar to that of the DEM. The DEM is normally controlled by the topography of the bedrock when the thickness of the deposits is slight. In the case of complex surfaces, modeling the topography of bedrock is more challenging and accuracy strongly depends on the interpolation method.

This study proposes to develop a methodology for modeling the topography of bedrock. We first compared the accuracy of the results yielded by different interpolation methods, by means of a challenging case study in which the most critical conditions were presented—a complex bedrock topography combined with a low-quality dataset. The proposed methodology could then be used in other similar contexts worldwide.

We determined a sub-region (called test area) used to calibrate the chosen interpolation method, based on both density and distribution of data (observation points). Once the best interpolation had been selected, it was applied to the entire study region. In order to improve modeling accuracy, we propose a technique that makes it possible to produce more observation points. Using existing geological cross-sections, we generated what we call "virtual observation points" (issuing from virtual boreholes). Finally, the regional topography of bedrock was validated using a dataset of boreholes that do not reach the bedrock (but that provide known minimal depth values).

Study case

The bedrock to be modeled is composed of crystalline rocks in the Canadian Shield. Figure 1b shows the type of bedrock topography that was targeted. Because the surface is complex, it is interesting to develop a modeling approach that first

Fig. 1 Relationship between type of surface, type of data distribution and number of data points and their impact on the quality of interpolation results, i.e., the accuracy of the bedrock topography model (*RMS value* root mean square value)
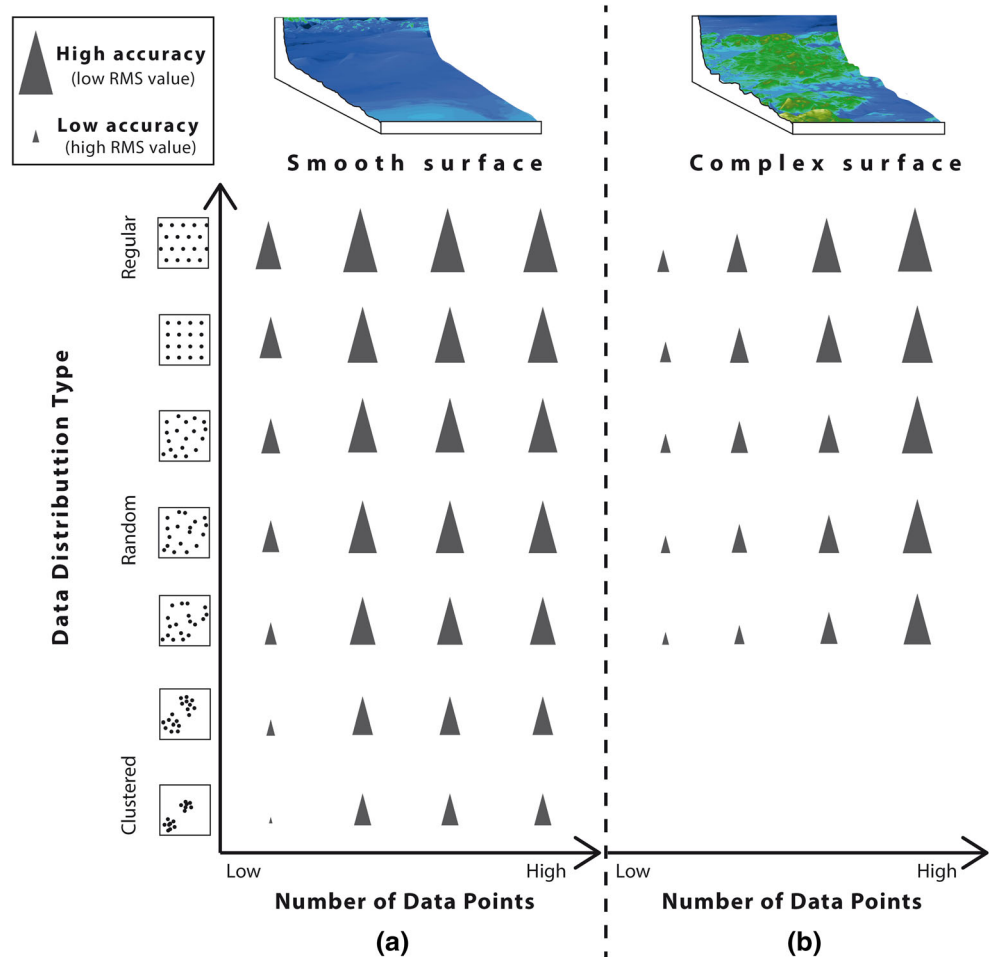
includes a trial of different methods of interpolation and an assessment of their accuracy, in order to recommend the most efficient of these methods before going forward. To perform this comparative test between interpolation methods, it was necessary to carefully select a study area which offered a large number of observation/controlling points of the bedrock topography. These points could be either a large density of boreholes or a widespread area where the bedrock outcrops at the surface and can be directly observed. Considering the cost of drilling, areas with a high density of boreholes are rare; areas of widespread outcrops are easier to find, relatively speaking. Access to a specific area with known or observable elevations of the bedrock makes it possible to better test and calibrate an interpolation method requiring a large number of observation points. It is then possible to better calculate the RMS values (Eq. 1) which are considered indicators of interpolation accuracy and ultimately, indicators of the quality of the proposed bedrock topography model.

This study will focus on a representative bedrock located in the province of Quebec in Canada in the Saguenay-Lac-St.-Jean (SLSJ) region, where the Precambrian crystalline rocks are part of the Grenville Province within the Canadian Shield

(Fig. 2). The crystalline rocks of the Canadian Shield are mainly composed of three families: anorthosite/granite/syenite-migmatite-gneiss. The Kenogami uplands is a fractured bedrock composed of anorthosite (Fig. 3), which is a phanerozoic-aged structure in meridional Quebec (south of the 52nd parallel). The Kenogami upland structure consists of a 600-km^2 rock plateau bounded by the Saguenay River to the North and the Kenogami Lake to the South. Its eastern and western borders are marked by a sharp topographical transition to the plain (Fig. 2). Its surface undulates at elevations ranging from 150 to 200 masl, whereas the surrounding lower plain varies between 50 and 150 masl. The area predominantly presents exposed bedrock at the surface with a complex topography type as targeted for this study (Fig. 2). Much of the Kenogami uplands is covered by a thin and minimal soil cover over the Precambrian crystalline anorthosite (Chesnaux 2013), with many outcrops distributed throughout the area. As explained in more detail below, the Kenogami uplands area of the SLSJ region will first be used as an experimental area to determine the best interpolation method, because the topography of this sub-area is well known. Once the most appropriate interpolation method has been selected, it
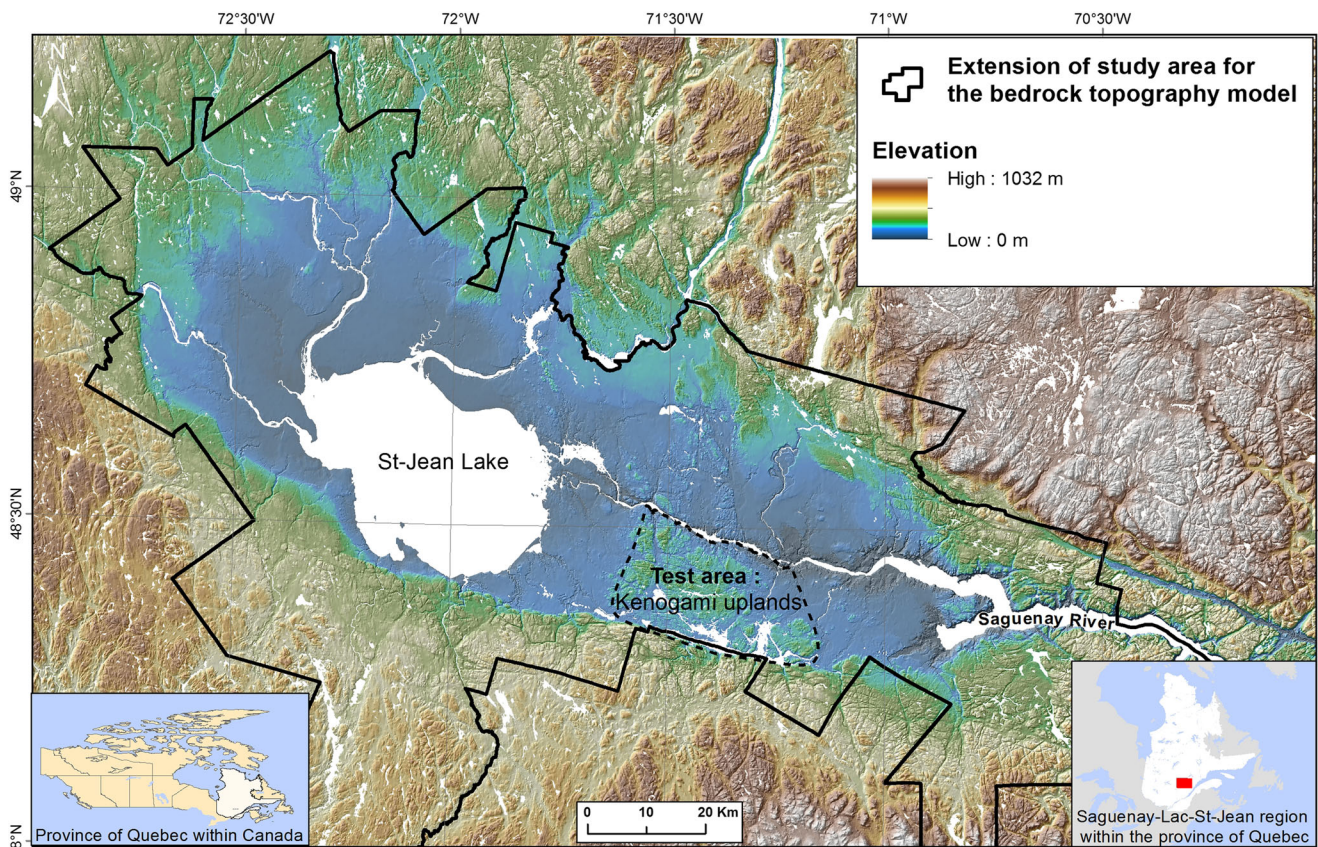
Fig. 2 Study area (Saguenay-Lac-St-Jean region) and test area (Kenogami uplands) locations and topography (digital elevation model, DEM). Crystalline rock aquifer of the Kenogami uplands in the province of Quebec, Canada (modified from Chesnaux 2013)

will be applied to the entire SLSJ region (13,210 km²) to model the topography of the bedrock on a larger, regional scale (Fig. 2).

During a previous hydrogeological characterization project of this region (Chesnaux et al. 2011), a spatial database for the Saguenay-Lac-St-Jean region (Fig. 2) had been generated

Fig. 3 Outcrop of fractured crystalline bedrock (anorthosite) in the Saguenay-Lac-St-Jean region located in Quebec, Canada (after Chesnaux 2013)

using ArcGIS. It was designed to provide relevant information on aquifers and groundwater properties and was generated using information from boreholes, among other sources of hydrogeological information (Chesnaux et al. 2011). This database was used for the present study. It contains stratigraphic information derived from 7627 boreholes and 40,473 outcrops providing surficial information (Fig. 4). Slightly more than 70% (5426) of the boreholes reach the top of the bedrock, providing direct observation of the bedrock elevation. The remaining boreholes are too shallow to reach the bedrock (2201), but nevertheless provide a minimal depth value of bedrock elevation (forcibly located below the bottom tip of the borehole). The total amount of boreholes represents a density of less than 0.26 observation points per square kilometer over the entire study area (SLSJ region), a territory where deposits cover most of the total surface of the bedrock. Moreover, approximately half the boreholes were drilled for purposes of installing residential groundwater supply wells. This introduces a bias in the distribution of the data, since most of these residential wells are located along roads or waterways (Fig. 4). The dataset is consequently of poor quality since it offers a limited quantity of observation points that are spatially distributed according to a random-clustered pattern. Furthermore, considering that the bedrock topography is of a
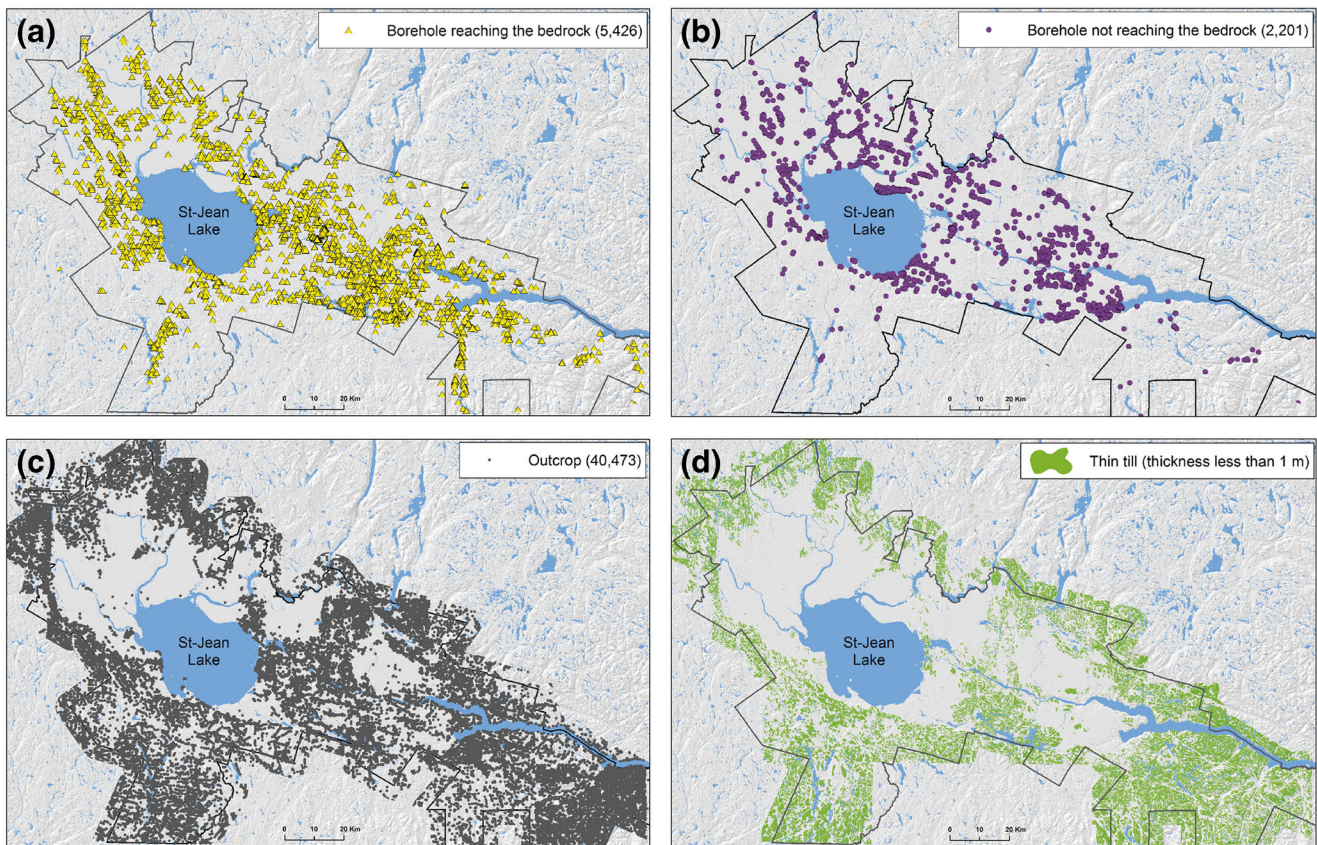
Fig. 4 Data points of bedrock elevation. **a** Boreholes reaching the bedrock (5426 observation points). **b** Boreholes not reaching the bedrock (2201 observation points). **c** Outcrops (40,473 observation points). **d** Thin till

complex type corresponding to the case represented in Fig. 1b, it is expected to produce a model of poor quality (of high RMS value). This is precisely the type of situation where the choice of interpolation algorithm becomes a critical decision that will exert significant impact on the quality of the modeling (Chaplot et al. 2006).

Methodology

The proposed methodology for generating a model of the bedrock topography in a study area (SLSJ region in the illustrated case) is composed of four main steps: (1) gathering the maximum amount of available and relevant data; (2) determining which of three different algorithms (ordinary kriging (OK), inverse distance weighting (IDW), and triangulated irregular network (TIN)) is the most appropriate interpolation method by comparing their performance on a test area (Kenogami uplands) contained within the wider region (study area) where the topography of the bedrock is to be modeled (Fig. 2); (3) using outcrop data and geological cross-sections to enlarge the pool of observed data in the study area (by creating "virtual" borehole data points) and to increase modeling quality; and (4) applying the selected interpolating method

to model the bedrock topography of the study area (SLSJ region) and validating the model.

Note that this study was entirely conducted using ESRI's Geographical Information System ArcGIS version 10.2.

Selecting the interpolation algorithm

Selecting a test area

The bedrock of the Canadian Shield presents a highly irregular surface composed of successive peaks and valleys, qualifying it as a complex surface for modeling. Figure 1 shows that modeling such a complex surface would normally require a dataset of numerous and well-distributed observation points. In the case of the SLSJ region, the available hydrogeological database provides scarce data on bedrock elevations. Furthermore, the distribution of this information is clustered rather than regularly distributed. As a consequence, any interpolation results are expected to be of poor quality. To counter this, the choice of interpolation algorithm can be critical and will exert a significant impact on the modeling results.

We identified the area of the Kenogami uplands (Fig. 2) as an appropriate test area with a topography controlled by bedrock, i.e., an area where the bedrock is mainly outcropping or

covered by only a thin layer of deposits. In other words, the bedrock topography is reflected relatively accurately by the topsoil topography, as represented by the DEM. We may thus assume that the bedrock topography is well known in the selected test area. A series of datasets that represent different configurations of data density and distribution can be generated from the DEM (Fig. 5).

Following dataset generation, the different interpolation algorithms can be applied to each of these datasets. Each of the models thus produced can then be compared against the DEM to assess the RMS values and ultimately identify the interpolation approach that produced a model closest in accuracy compared to the test area (Fig. 5).
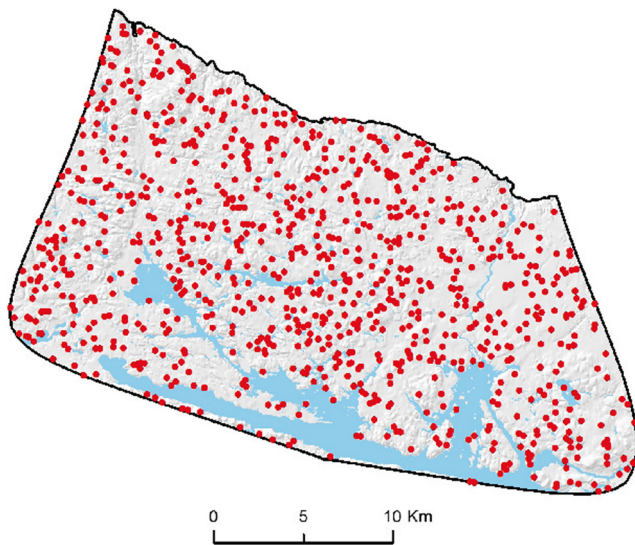
Creating datasets for the test area

Ten datasets with two different spatial distributions of data points of the bedrock's top elevation and with different quantities of data points were randomly extracted from the DEM (Fig. 5). Five datasets with a random distribution (Fig. 5a) were chosen and five additional datasets with a clustered distribution (Fig. 5b) were chosen. The datasets belonging to a same spatial distribution type differed according to the quantity of data points (observation points) and they contained the following: 50, 100, 500, 1245, and 3376 points in both random- and clustered-distribution dataset groups. These quantities correspond respectively to data densities of 0.08, 0.17, 0.8, 2.1, and 5.6 observation points per square kilometer. The five random-distribution datasets (Fig. 5a) were created by means of the Create Random Point tool in ArcGIS, whereas the five clustered-distribution datasets (Fig. 5b) were created by means of the Create Random Selection tool that randomly selects observation points in ArcGIS.

Testing process of the different interpolation methods

Thirty different test models were generated using the 10 datasets (5 random, 5 clustered) and the 3 interpolation methods (TIN, IDW, OK). In order to assess their accuracy, each one of the 30 models was compared against what we called the validation dataset. The validation dataset was composed of 1000 observation points that were randomly extracted from the DEM (which has a 10-m, grid-spaced resolution) of the study area (Fig. 6) using the Create Random Point tool available in ArcGIS. For the validation, we compared the same 1000 points of the DEM and of each of the 30 models. This validation process is presented in Fig. 7. For each of the 30 models, an RMS value was generated (representing the average of the differences between elevations of the model and of the DEM for each of the 1000 observation points). It should be remembered that the DEM of the test area is assumed to be accurately representative of the bedrock topography, which in this area is either outcropping or covered by a thin (of negligible thickness) layer of deposit. The lower the RMS value of a model (or the lower the differences between the modeled elevations and the observed elevations), the better the accuracy of the model in representing reality. Therefore by comparing RMS values, it is possible to select the best interpolation method (lowest RMS values) for both the random and the clustered distribution of data. Figure 7 summarizes the entire process described above, which was designed to select the best interpolation method for modeling our chosen bedrock topography. Note that the results of the interpolations are compared against the DEM by using the ArcGIS tool called Extract Values to Point available in the Spatial Analyst extension of ArcGIS.
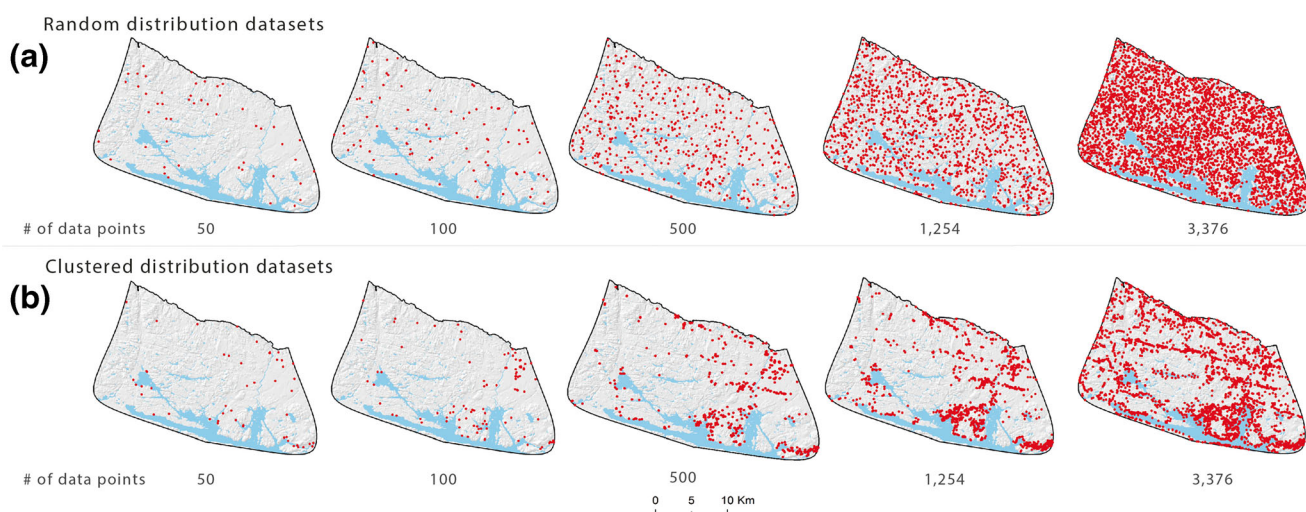
Fig. 5 Generating random (**a**) and clustered (**b**) datasets from the digital elevation model (DEM) for the test area (Kenogami uplands). The *red dots* represent the observation points of the top of the bedrock provided by the DEM

Fig. 6 Validation dataset created from the DEM for the test area (Kenogami uplands). A total of 1000 observation points were extracted from the DEM (*red dots* on the figure) and used to compare their real elevations against the modeled elevations
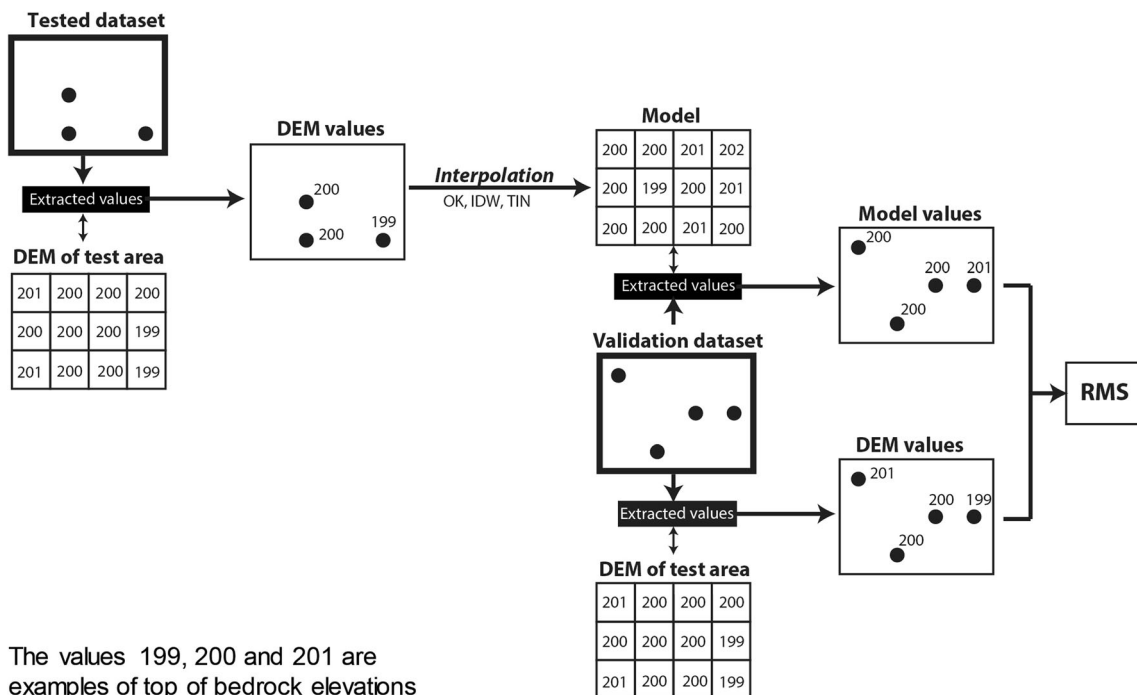
The tested interpolation methods

Three different methods of interpolation are tested using the datasets that have been previously described (Fig. 5). These methods are applied using the algorithms available in ArcGIS 10.2. The tested interpolation algorithms are ordinary kriging (OK), inverse distance weighting (IDW), and triangulated irregular network (TIN).

OK is a probabilistic method based on statistics, used for predicting and modeling surfaces. OK assigns a weight to the observation points based on the degree of similarity between these points, i.e., the covariance between the observation points as a function of the distance between the points. OK is executed using the analysis of a variogram and is used for interpolating a stationary variable of unknown mean, but it supposes that there exists a constant mean. In the framework of this study, the same kriging parameters were used for each dataset; these parameters were established using the 1000 points of the validation dataset. A spherical model of the semi-variogram was chosen. The Searching Neighborhood tool was applied from an ellipse with four sectors and the minimum number of points was set at 2, whereas the maximum number of points was set at 5.

The IDW method of interpolation is deterministic. It allows the interpolation of bedrock surface topography when considering the weighted average of neighboring observation points. The weight of the observation points is determined as a function of the distance between the points. The greater the distance between two points, the lower the influence of the observed values between these points. The IDW method can sometimes generate a "bull's eye" effect in the vicinity of the observation points. In the present study, the IDW interpolation was conducted by using a variable radius of influence which included a minimum of 12 observation points.

Fig. 7 Schematic diagram showing the interpolation testing algorithm used to calculate the RMS values, which are used to determine the accuracy of the generated models. Note that the DEM grid definition is 10 × 10 m

The TIN method is a vectorial structure used to represent a surface model in which the observation points are linked to each other by adjacent triangles which do not overlap (Delaunay triangles). These triangles form a continuous surface that in its entirety represents the surface topography of the bedrock in our case. The TIN method is particular in that it does not modify the observed input values (of the bedrock elevation) following the application of the interpolation. The TIN method is created from the 3D Analyst feature in ArcGIS and is then transformed into a matricial format.

Note that all three interpolation methods were performed using a 10-by-10 m resolution of the surface grids.

Increasing the quantity of "observed data points": observing outcrops and creating "virtual" boreholes

Once the best-suited interpolation method had been selected for the test site, an additional step was taken to improve the quality of the modeling, before applying the chosen interpolation method to the entire study area in view of modeling its bedrock topography. This intermediary step consists in increasing the number of "indirectly" observed data points. It should be remembered that different types of observation data are used to model the bedrock topography. The relevant observation data are as follows: (1) bedrock elevations observed in boreholes; (2) bedrock elevations observed at the surface when the bedrock outcrops or is located under a thin layer of deposits; (3) "virtual" bedrock elevations that are not directly observed but are deduced from geological cross-sections (note that such observations are artificially created); (4) non-observed bedrock elevations ("bedrock deeper than" in boreholes that are not deep enough to reach the bedrock). In this last case, such information is still relevant because it indicates that the elevation of the bedrock is *at least* below the end of the borehole. This pool of "deeper than" data was kept in reserve for a late-stage validation of the final model that we obtained for the entire SLSJ region, after application of the chosen (best) interpolation method. The following paragraphs describe in more detail the four abovementioned types of data that are used in our approach to model the topography of the bedrock.

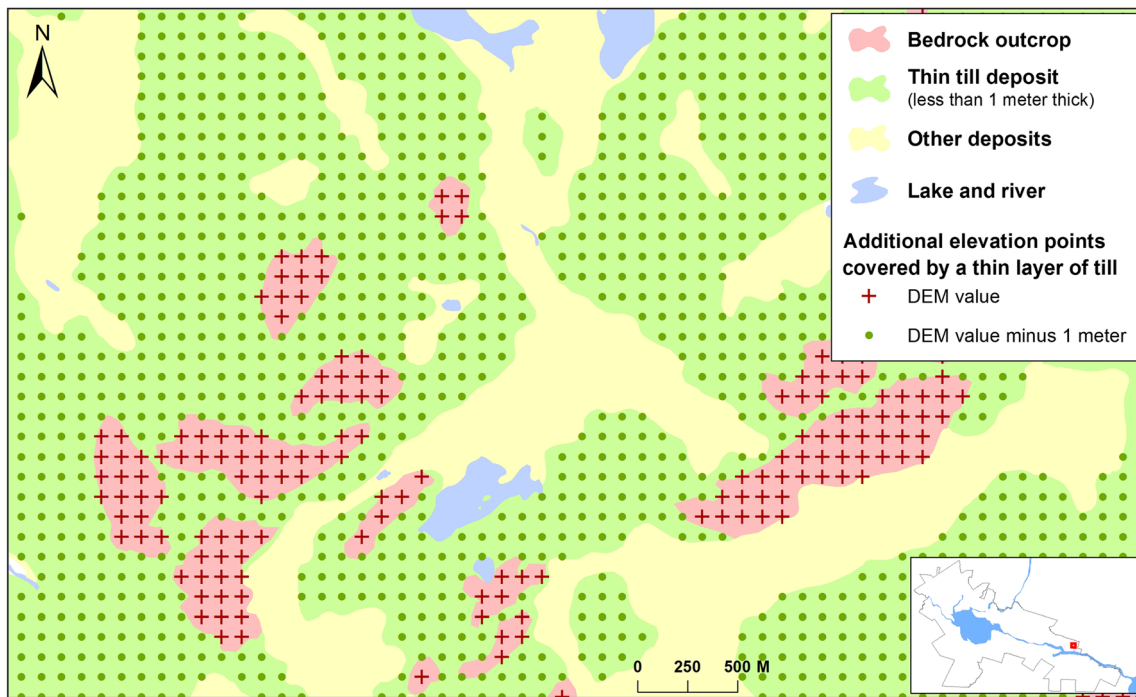#### Boreholes, outcrops, and thin deposits

The prime data for observing the locations of the top of the bedrock are the boreholes and the outcrops. For the SLSJ region, Fig. 4 shows the locations of these observation points (9272 boreholes and 40,473 outcrops). The mapping of surface deposits can also be useful to obtain more information about the location of the bedrock in a given region. Zones of minor deposit thickness are in fact zones of shallow bedrock. In this study, when deposits are less than 1 m in thickness and the bedrock was almost outcropping, it was considered to be

outcropping. This assumption made it possible to considerably increase the quantity of observation data on bedrock outcrops used in this study. In order to define the zones of thin deposit coverage (less than 1 m), the polygons on the map defined as these deposits (on the surficial deposits map of our database) are converted into a grid with a mesh of 100 m (Fig. 8). To each centroid of this grid is assigned a value of the bedrock elevation which corresponds to the DEM value at the center minus 1 m; that indicates that we designated the rock as outcropping at these locations (Fig. 8). Note that in the study region (SLSJ), the thin deposits (less than 1 m) are represented by till formations (Fig. 4). Figure 8 illustrates an example of the grid that is obtained with the two pools of data: real outcropping of the bedrock where the elevation of the bedrock is defined by the DEM and top of bedrock locations that are 1 m below the DEM in presence of a thin layer of till deposits.

"*Virtual*" boreholes generated from cross-sections

One original aspect of the proposed approach for modeling bedrock topography consists in the use of geological cross-sections (developed from the database) to generate new observation points that we called "virtual boreholes." One hundred thirty-four stratigraphic cross-sections were generated and drawn based on the geological data available in the database (Chesnaux et al. 2011). Some of this geological information is derived from the stratigraphy observed in the boreholes. Figure 9 shows the locations of these cross-sections. The cross-sections were distributed according to a regular spatial pattern to ensure a good coverage of the entire region. A simplified stratigraphic model was developed comprising only 5 categories: sand, gravel, clay, till, and rock. The cross-section lines (134 in total) each intercept several boreholes; the simplified stratigraphy of each borehole is reproduced along the lines. Based on this information, stratigraphic sections were produced based on a geological interpretation of a geologist of our research team who interpolated the stratigraphy between the projected boreholes. The cross-sections are georeferenced in a 3D environment and represented in what are called barrier diagrams (Chesnaux et al. 2011). Once the barrier diagrams are created, it becomes possible to extract the estimated elevation values of the bedrock at regular distances along the 134 cross-sections (in our study, every 500 m). These extraction points have been made to represent observation points that do not physically exist but have been artificially created; these have been called "virtual" boreholes for the purposes of this study. Each "virtual" borehole provides stratigraphic information (in actual fact, provided by the cross-section) including the elevation of the top of the bedrock. Figure 10 shows an example of one of the 134 stratigraphic cross-sections drawn in the SLSJ region. In this example, the cross-section intercepts 4 real boreholes and 6 outcrops. Of

Fig. 8 Grid showing the additional, virtual "top of bedrock" elevation points created from maps showing both the presence of thin deposits (less than 1 m in thickness) and the absence of deposits (outcrops)

these 4 real boreholes, it should be noted that only 2 intercept the bedrock; nevertheless, the remaining 2 boreholes are still of interest because they provide partial information on the stratigraphy.
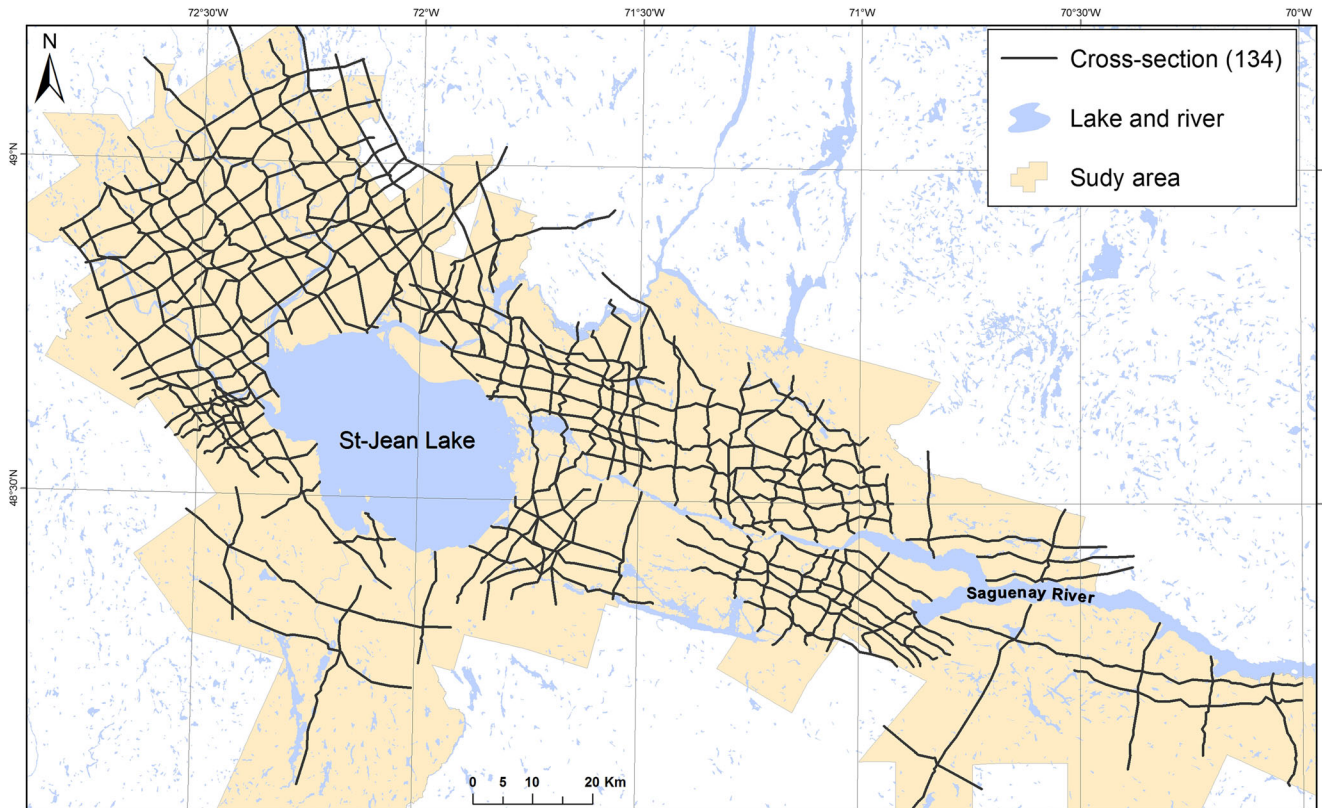
Fig. 9 Distribution of the 134 cross-sections throughout the study area (SLSJ region)
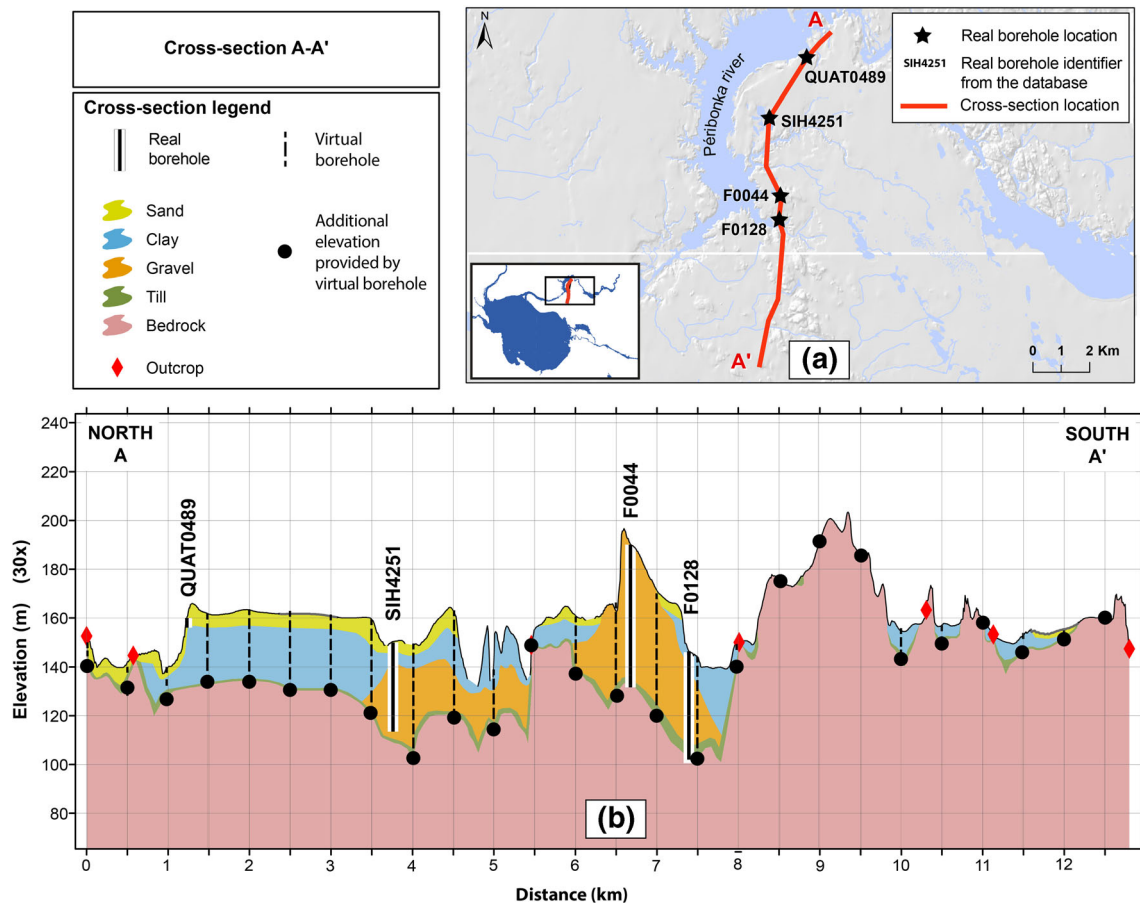
Fig. 10 View of real and virtual boreholes along an A-A′ cross-section. (*a*) Plan view. (*b*) Sectional view

In this manner 26 "virtual" boreholes (spaced 500 m apart) were created from the stratigraphy that had been drawn from real boreholes and outcrop observations. These new "virtual" boreholes make it possible to significantly increase the quantity of observation data indicating bedrock elevation. "Virtual" boreholes also make it possible to improve the density of information as well as its spatial distribution. Indeed, the resulting cross-section network covers certain areas of the SLSJ region where very little stratigraphic information is available (Fig. 4). Consequently, improved coverage of observation points on a regional scale may be expected to yield significantly better results in the interpolated bedrock topography model. Figure 11 shows the distribution of the 7304 "virtual" boreholes that were created from the 134 cross-sections (on average, approximately 54 "virtual" boreholes per cross-section).

Interpolation, modeling, and validation of the topography of bedrock on a regional scale

After selecting the interpolation method and integrating additional virtual data into the complete dataset, the next steps of the proposed methodology consist in executing the interpolation of the observation points throughout the entire SLSJ region, including all previously described observation points: real and virtual boreholes, outcrops and thin deposits. Figure 11 shows the locations of all of these types of data points. The interpolation is executed by means of the most appropriate interpolation method selected after the test interpolation done for the Kenogami uplands study area.

One possible type of validation, called cross-validation (not conducted for this study), would consist in keeping aside a portion of existing objective data points located in the study area, separate from the data actually used for the interpolation. After the model is generated, it can then be compared against the data that was kept aside. RMS values calculated for the data kept aside and for the model may be compared against each other to assess the model's accuracy.

Instead, we chose another type of validation process (using the set of boreholes not reaching the bedrock), less quantitative and more qualitative, which was deemed more appropriate for the characteristics of the SLSJ region. The justification for this decision lies with the clustered nature of data distribution in this case. Indeed, when data is clustered, the cross-validation method will usually yield good RMS values in any case, because the dataset used for verifying the model's accuracy is necessarily located in the same areas where the interpolation is accurate, since the modeled data points and the
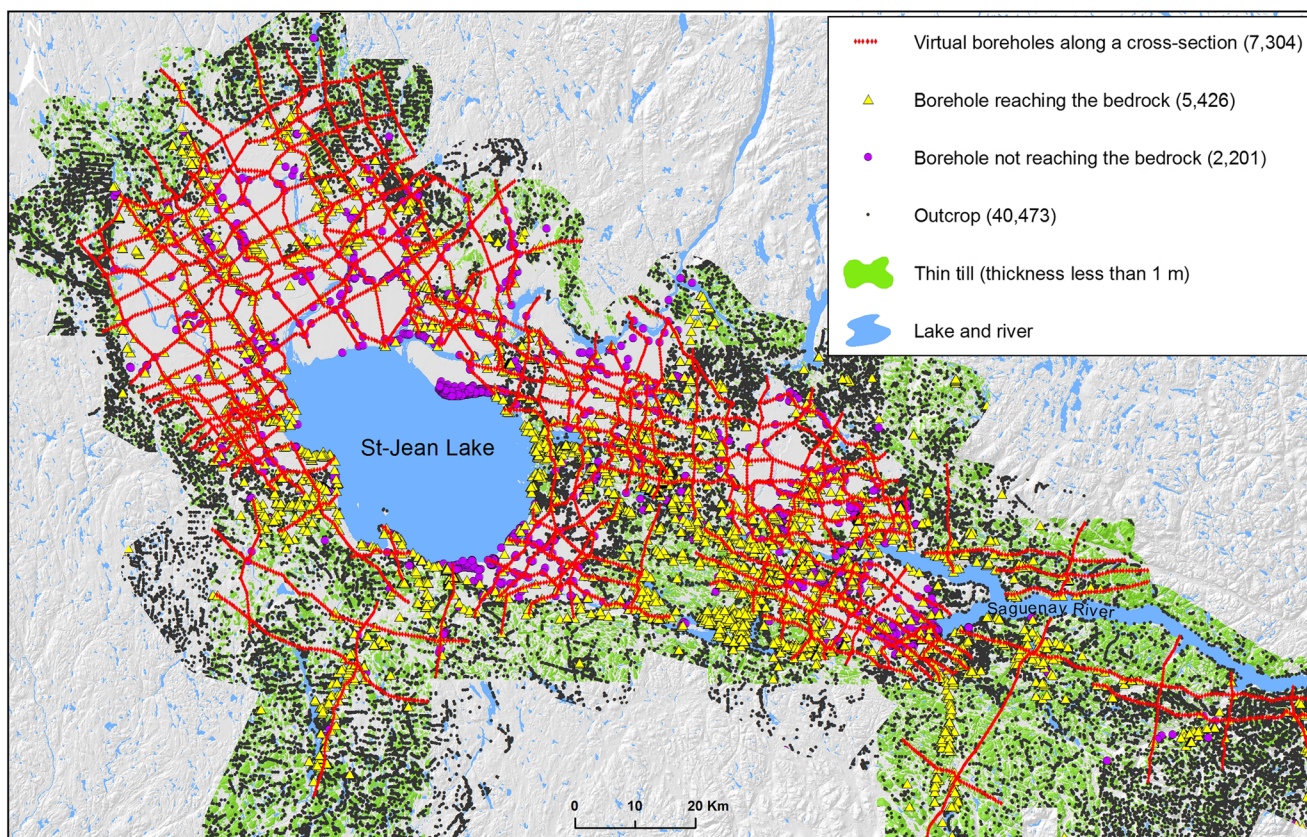
Fig. 11 Complete set of data points of bedrock elevation, including real and virtual data

verification data points are forcibly located close to each other. Clustered distribution thus introduces a bias in the cross-validation, by yielding RMS values which would have been too low and would not have provided a properly accurate assessment of the model's accuracy.

Instead, we used a new technique. A special set of data was kept in reserve for purposes of validating the topography model of bedrock elevation. These data were derived from the 2201 boreholes in the study region that do not reach the bedrock (shown in Figs. 4 and 11). At these locations, the top of the bedrock is deeper than the end of the borehole; in other words, the top bedrock elevation is known to be located below the borehole end elevation. This data set is considered to contain objectively certain information; as such, it will be used to assess by comparison the accuracy of the final model of bedrock topography of the SLSJ region. Figure 12 shows a conceptual cross-section that helps to understand the validation process of the topography model of the study area. A-type boreholes are the boreholes that are known to reach the bedrock (5426 in total) and that were used to generate the model. B and C types of boreholes are the "reserve pool of data" not used to generate the model, but only used to validate the model after it was generated. These B and C boreholes (quantity of 2201 in total) are known to not reach the bedrock; therefore, when the model showed a bedrock elevation higher than one

of these boreholes, then the model is known to be inaccurate at the specific location of that borehole. In this figure, borehole C appears deeper than the bedrock elevation as determined by the model; thus, the model is inaccurate at that location.

B-type boreholes validate the model since they are located not only above the observed bedrock but also above the modeled bedrock. In ArcGIS, the ratio of B to C borehole types will provide a qualitative assessment of the consistency of the model in representing the real topography of the bedrock in the study region. The higher the number of B-type boreholes relative to C-type boreholes (out of a total of 2201 boreholes not reaching the bedrock), the higher the degree of consistency of the model.

Results

Selecting the best interpolation method by means of an algorithm

Figures 13 and 14 present the bedrock topography modeling results in the test area (Kenogami uplands) using the 3 interpolation methods that were tested (IDW (b), OK (c), and TIN (d)), and according to the number and density of data points of the different datasets: 50, 100, 500, 1254, and 3376 data

Fig. 12 Conceptual cross-section showing how the "reserved" data points (*B* and *C*) are used after modeling to evaluate the model's accuracy. Borehole *B* (in *green*) is known to not reach bedrock; it shows that the model is accurate at that location because the borehole ends above the model (*dotted line*). Borehole *C* is known to not reach bedrock but shows that the model is inaccurate at that location because the borehole extends below the model
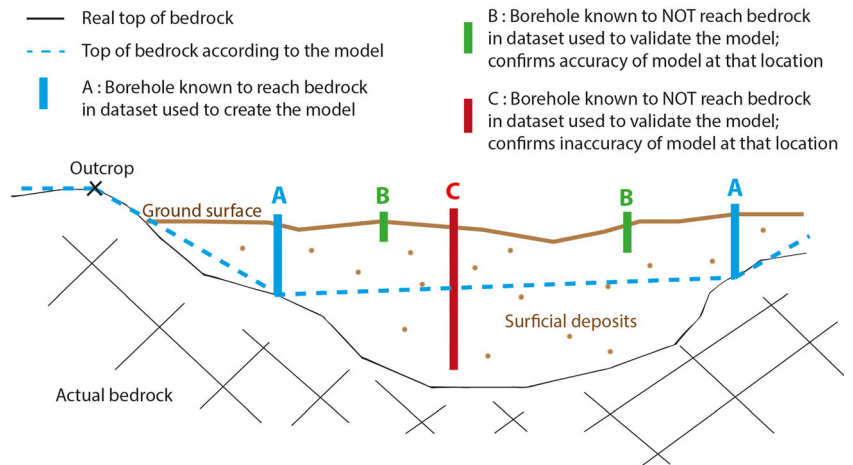

points. Figure 13 presents the results for the random-distribution cases whereas Fig. 14 presents the results for the clustered-distribution cases. All these interpolation results can be visually compared against the DEM of the test area, which is also shown in Figs. 13(a) and 14(a).

Figure 15 shows the calculated RMS values for the five different datasets (composed of 50, 100, 500, 1254, and 3376 data points) for the three different interpolation methods (IDW, OK, and TIN) and for both types of distributions, random (Fig. 15a) and clustered (Fig. 15b). In general for both random and clustered distribution, we observed the expected increase in accuracy when the number of data points increased (producing lower RMS values). However, it is interesting that for both random and clustered distribution, the relationship
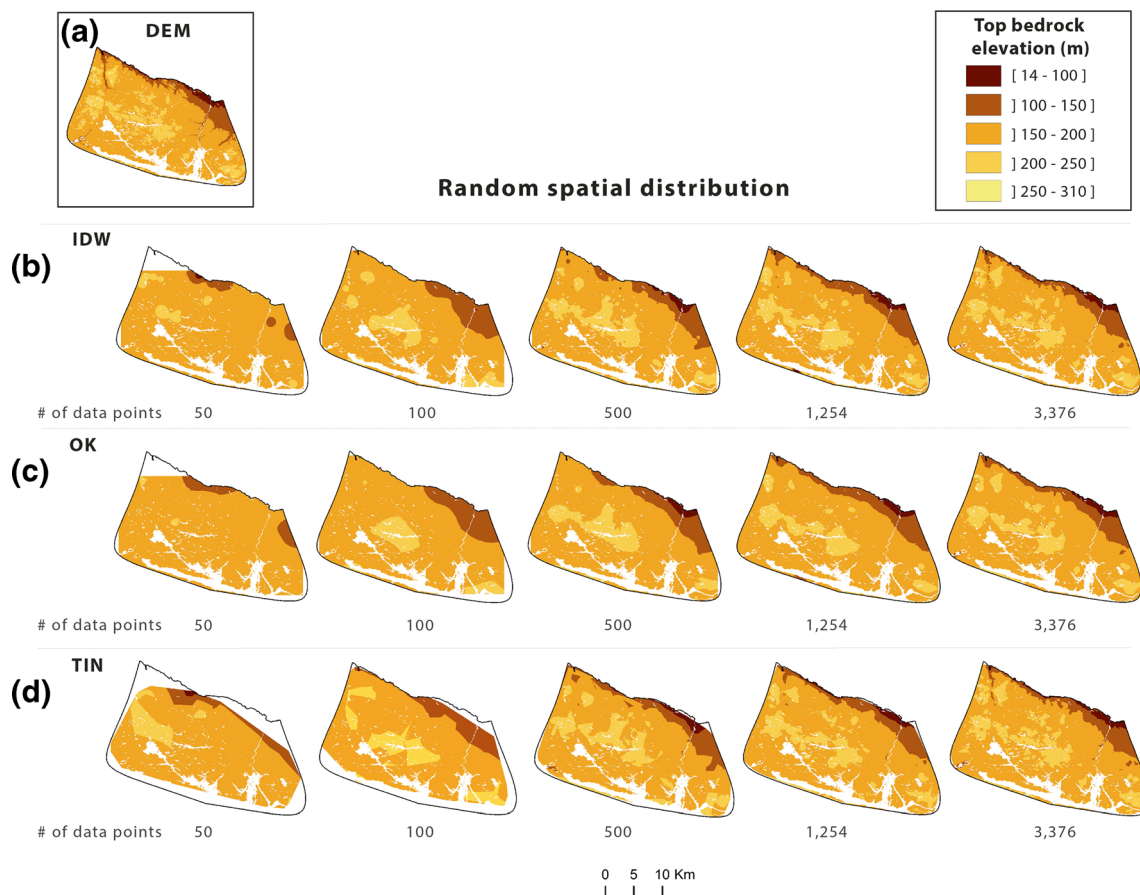
Fig. 13 (*a*) Reality as represented by the DEM (used as a reference for comparison) and interpolation results obtained from the random datasets: (*b*) using inverse distance weighting (*IDW*), (*c*) using ordinary kriging (*OK*), and (*d*) using triangulated irregular network (*TIN*)
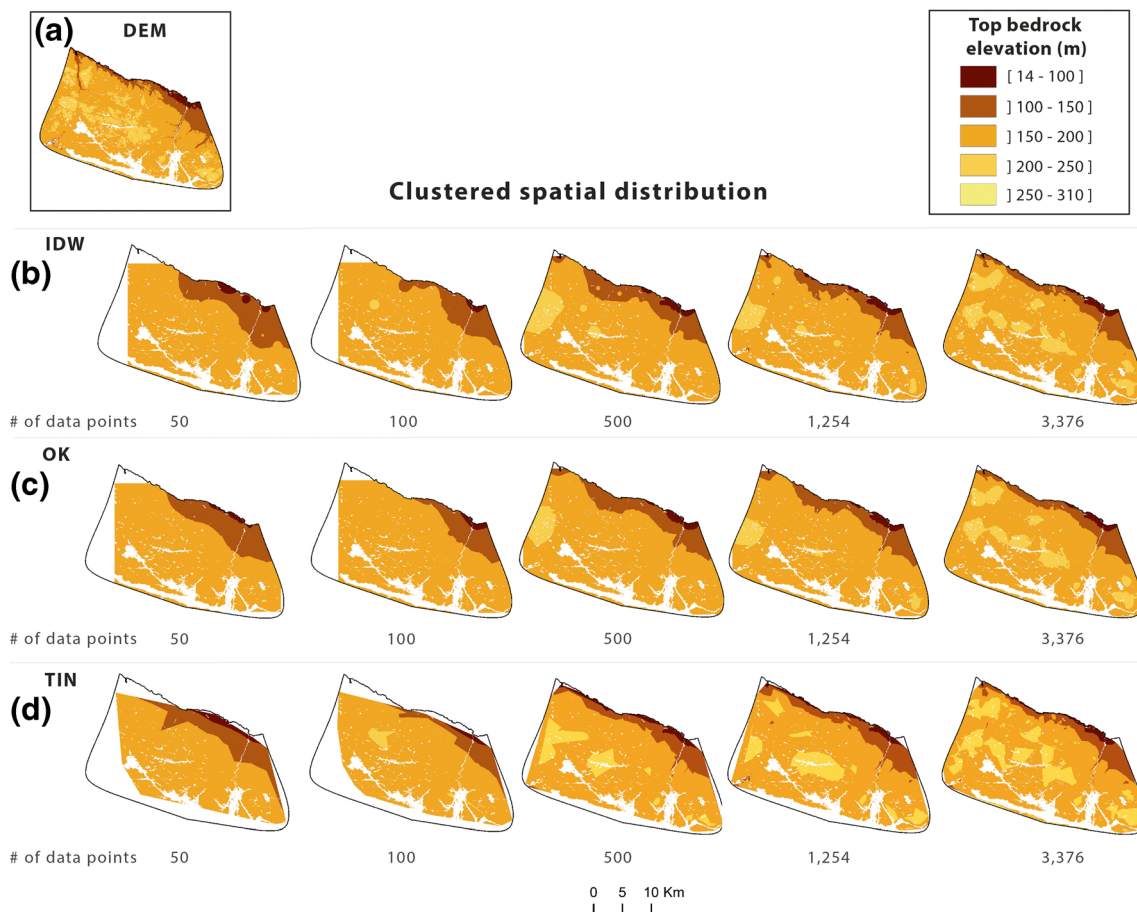
Fig. 14 (*a*) Reality as represented by the DEM (used as a reference for comparison) and interpolation results obtained from the clustered datasets: (*b*) using inverse distance weighting (*IDW*), (*c*) using ordinary kriging (*OK*), and (*d*) using triangulated irregular network (*TIN*)

between number of data points and accuracy is not linear. We observed a marked improvement of RMS values from 50 to 100 data points, but when the number of data points was higher than 100, the accuracy increased less markedly than when going from 50 to 100 data points. This observation is of interest when recommending a minimum density of information required to produce a model of acceptable representativity (of acceptable RMS value). As mentioned previously, 100 data points over the surface of our test area equates to a density of 0.17 data points per square km; this appears to be the minimal density required (considering that according to Fig. 15, the RMS values decrease more slowly above 100 data points) to obtain an acceptable degree of accuracy. From this minimal threshold, increasing the number of data points does not necessarily, significantly, or proportionally improve a model's RMS value.

When comparing random versus clustered distribution of data points, we observed better (lower) values of RMS using random distribution (ranging from 10 to 21 m, Fig. 15a) compared to clustered distribution (ranging from 13 to 26 m). This observation appears logical and was expected, considering that a more regular distribution of the data points (random)

ensures a better coverage of the area to be interpolated. When comparing the RMS values obtained from the three different interpolation methods, the TIN method yielded the best results for both clustered and random distribution and independently of data density. Based on the results obtained for our test area, the TIN method thus proved to be the most appropriate interpolation method for the SLSJ study area. The greater accuracy of results obtained using the TIN method is even more pronounced in cases of clustered data distribution. Indeed, Fig. 15b with clustered distribution showed greater differences between TIN and the other two methods (IDW and OK) than in the case of the random distribution in Fig. 15a. For the test area, the gain in terms of RMS when using the TIN method instead of IDW or OK varied between 0 and 3 m for random distribution and from 3 to 5 m for clustered distribution.

Based on all these observations, and because results obtained in our test area favored the TIN interpolation method, we concluded that when data distribution is clustered such as in the SLSJ region, it is justified to preferentially apply a TIN interpolation instead of IDW or OK for more accurate results.

(a)

(b)
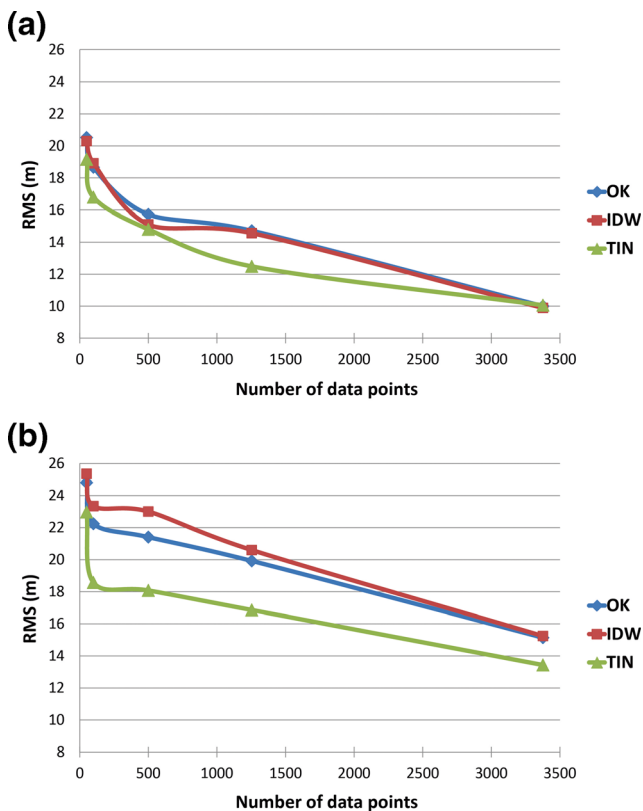
Fig. 15 RMS results for each interpolation method, for **a** random-distribution datasets and **b** clustered-distribution datasets

Generating and validating the bedrock topography model

After the TIN interpolation method was selected using the test area, TIN was applied to interpolate the entire, wider study area (the SLSJ region), using the complete set of all available observation points: 5426 real boreholes reaching the bedrock; 40,473 outcrops; 217,508 elevation points of the bedrock covered by a thin layer of till (DEM minus 1 m); and 7304 additional "virtual" boreholes. A total of 270,711 observation points were used for a surface area of 13,200 km^2, which equates to 20 observation points per square kilometer. This density is high compared to 0.17 data points per square kilometer that was used for the test area (previously established as the minimal density required to obtain an acceptable degree of accuracy). With 20 observation points per square kilometer instead of 0.17, we may expect to obtain a much more accurate model of bedrock topography, in particular because the virtual boreholes provide data points that are regularly distributed instead of clustered. The resulting bedrock topography model is presented in Fig. 16 (in 3D). The gain in accuracy is quantified in "Quantifying the impact of virtual boreholes.". It should be recalled that the resulting model still contains the uncertainties inherent to the data that were used to construct the DEM.

The next step was to validate the resulting model against the "reserved" pool of data comprised of 2201 boreholes that

are known to not reach the top of bedrock (see the locations of these boreholes in Fig. 4). Figure 17 shows the validation of the topography model of the SLSJ region using these 2201 boreholes. In a first instance, Fig. 17a shows that the 1568 boreholes that were correctly shown by the model to be located above the bedrock validated the model and, conversely, the 633 boreholes that were incorrectly shown by the model to be below the bedrock did not validate the model. This means that 71% of these boreholes agreed with the model, confirming the accuracy of the model at the locations of these boreholes. The remaining 29% of boreholes did not agree with the model, indicating the inaccuracy of the model at the locations of these boreholes.

An analysis of the 633 (29%) of boreholes not agreeing with the model revealed that almost half of these (quantity 284) were located 10 m or less below the modeled bedrock elevation (Fig. 18). Ten meters is also the accepted degree of error (the accuracy of the DEM is ±10 m) in the overall data points used to construct the complete model of the entire study area in ArcGIS. A further step was therefore conducted in ArcGIS in order to improve the accuracy of the model. A buffer of −10 m was applied to the simulation. The result was that the 344 boreholes that were previously not accurately represented by the model were now accurately represented by the model. Figure 17b shows the result after applying the buffer. The 1852 boreholes that are now located above the bedrock topography model validate the model and conversely, the 349 boreholes that are still located below the bedrock topography model do not validate the model. This means that 16% of boreholes still do not agree with the model, indicating the inaccuracy of the model at the locations of these boreholes, while the remaining 84% of these boreholes now agree with the model, confirming the accuracy of the model at the locations of these boreholes. Considering the difficulty in modeling bedrock elevations in regions where data is both scarce and clustered, this result may represent an optimized result compared to previous methods.

Quantifying the impact of virtual boreholes

In order to quantify the gain of accuracy that was achieved by adding cross-sections and virtual boreholes, a comparison was made between the results obtained when modeling with and without the virtual boreholes. Table 1 presents and compares the modeling results with and without the virtual borehole data and for both types of validations (without a buffer and with a 10-m buffer).

Table 1 shows that the model gains in accuracy when the virtual boreholes are included in the TIN interpolation that generates the bedrock topography model: the validation not using a buffer shows a difference of 15% whereas the validation using a ±10-m buffer shows a difference of 12%. It is interesting to note that the virtual boreholes represent only

Fig. 16 3D view of the regional bedrock topography model of the SLSJ region (view from the South-West) using the optimized methodology proposed in this study
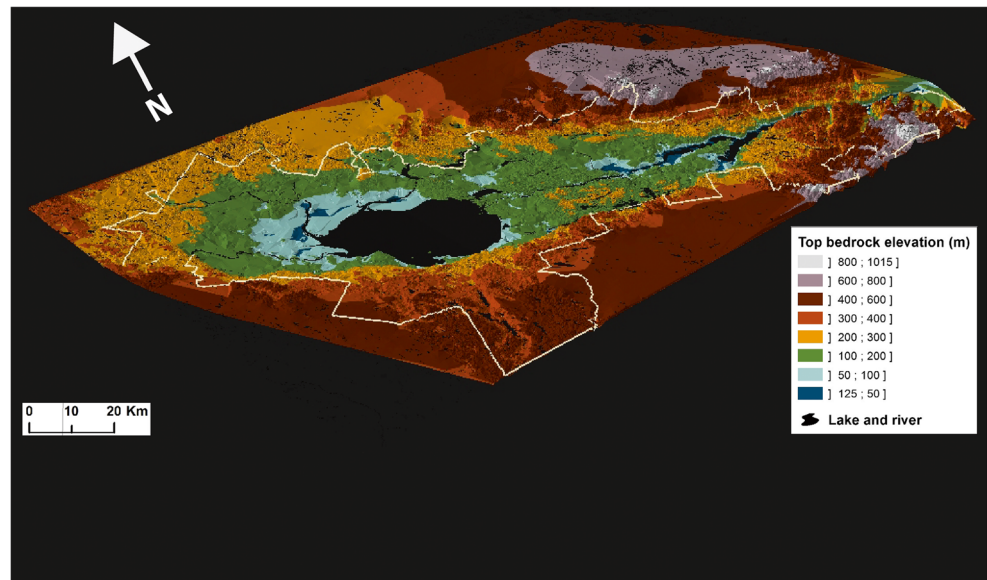
2.8% of all observation points (i.e., 7304 observation points over a total of 263,407). Despite this small number, they provide a significant improvement in the accuracy of the model. The gain in accuracy is due to a better distribution (regular rather than clustered).

Discussion and conclusion

The knowledge of the topography of the bedrock is a prerequisite when evaluating, for example, the quantity of available granular material that can be exploited (for example, for backfilling) or the geomechanical properties of the subsurface (for example, when installing building foundations). Determining the topography of the bedrock is also relevant when estimating the volume and potential productivity of groundwater reservoirs (aquifers). These aquifers are usually composed of sand and gravel at a regional scale, and they also delineate the top surface of the fractured-rock aquifers located underneath. In both granular and fractured-rock aquifers, the groundwater is contained in the porosity of the medium, respectively, either in the voids of the porous medium or in the fracture network of the fractured medium. Knowing the topography and thus the depth of the bedrock is critical when planning drilling operations with the intent of installing water wells or exploring for minerals.

In most cases, bedrock topography is complex and the data available on its location is of very low density and also unevenly distributed (Fig. 1). Data is habitually provided by boreholes drilled for municipal or individual water supply; these are usually located along roads, clustered at the outskirts of municipal centers. This reality, linked to patterns of social development, in fact controls the distribution of the data. It is difficult to create accurate bedrock topography models by interpolating these observation points, considering that the topography of bedrock is considered unpredictable (unlike the topography of a ground water table, for example, (Chesnaux 2013) whose location can be analytically determined based on equations that govern the groundwater flow).

For these reasons, it is a significant challenge to model the top of bedrock elevations with any degree of accuracy. The methodology that we have developed and presented in this study shows how better results may be obtained and how the accuracy of the models may be improved, by using techniques that optimize the density and distribution of data points. By drawing cross-sections between known boreholes and defining "virtual" boreholes that not only provide additional data, but more importantly a better distribution of data, better results can be derived when modeling bedrock topography. Figure 19 illustrates how the quality of bedrock topography modeling may be improved using our proposed approach. This approach can be used in other regions worldwide.

Our study, using a test area within the main study area, has also established TIN as the most appropriate interpolation method for the type of geological context under study: a complex bedrock surface of crystalline shield for which the available objective observation points are clustered and irregularly spaced over the territory being studied. This may not necessarily be the case for other regions. Even when a test area is not available, however, the TIN interpolation method may be recommended for modeling bedrock topography in similar geological contexts. The TIN method is commonly used to model surface topography, and our evidence also seems to support its use in modeling bedrock topography. This conclusion makes sense, considering the unpredictable nature of bedrock topography for which it is more logical to favor a linear type of interpolation between observation points (TIN) rather than adopting a probabilistic interpolation method such as OK.

Fig. 17 The two steps for validating the topography model of the SLSJ region. The top map (**a**) shows results before applying the 10-m buffer and the bottom map (**b**) shows the result after applying the 10-m buffer
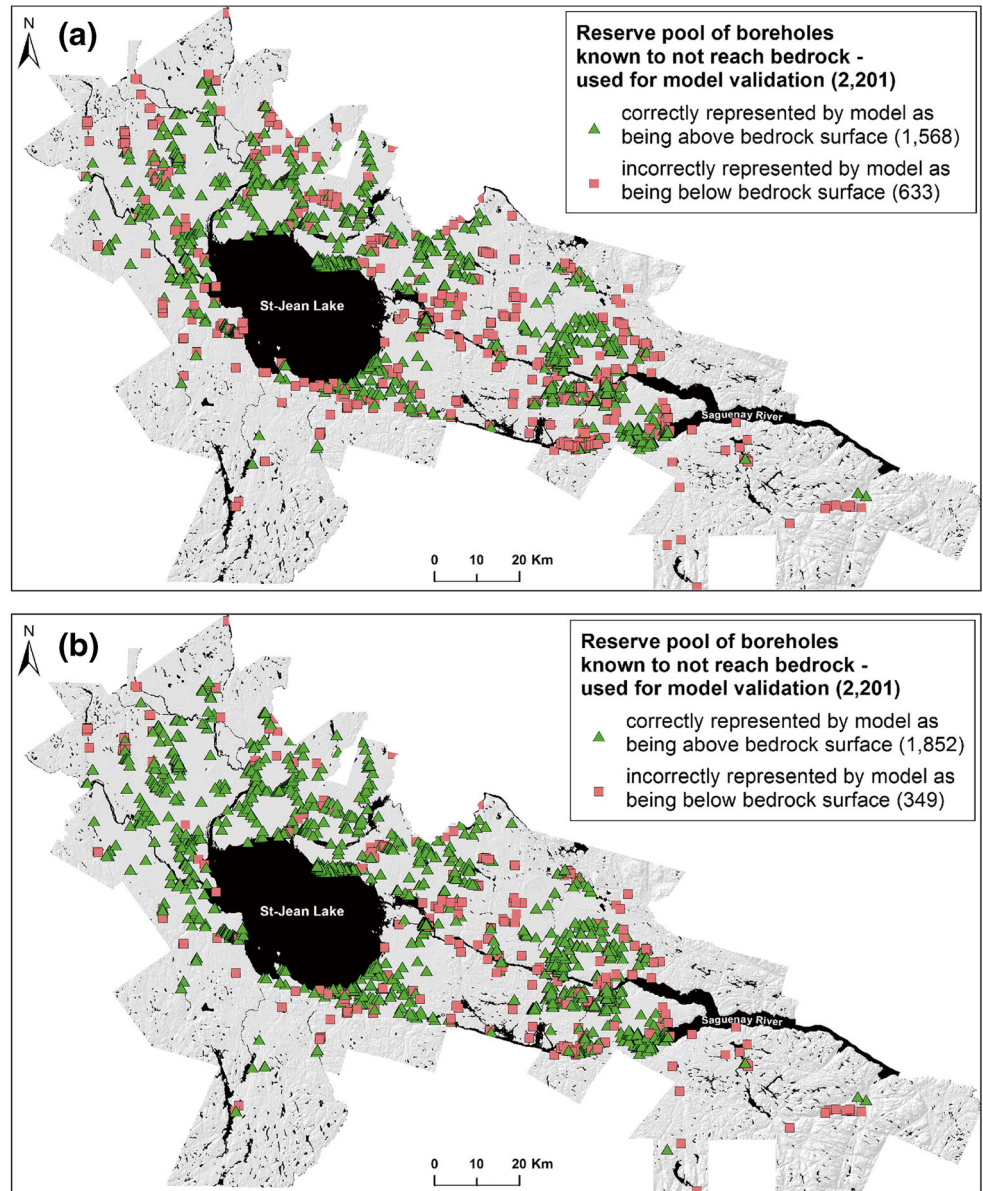
Fig. 18 Distribution of the 633 boreholes not reaching bedrock (validation dataset) that were inaccurately represented by the model as being deeper than the top of bedrock, according to the differences in elevation between the boreholes and the model itself
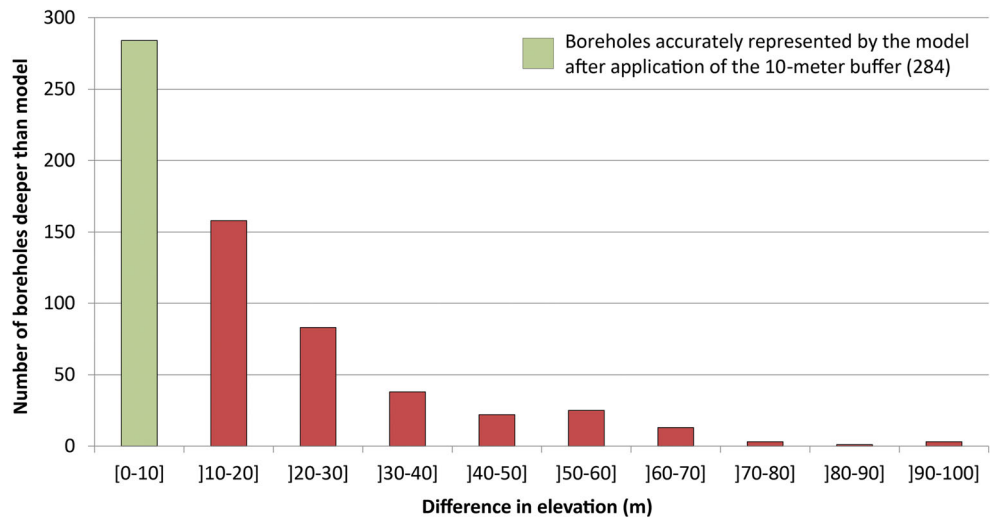
Table 1 Comparison of the validation results without virtual boreholes and with virtual boreholes

		Resulting model excluding virtual boreholes	Resulting model including virtual boreholes	Difference %
	Reserve set of boreholes not reaching bedrock			
Validation not using a buffer	Boreholes correctly represented by model	1235 (56%)	1568 (71%)	15
	Boreholes incorrectly represented by model	966 (44%)	633 (29%)	−15
	Total	2201 (100%)	2201 (100%)	
Validation using a ±10-m buffer	Boreholes correctly represented by model	1579 (72%)	1852 (84%)	12
	Boreholes incorrectly represented by model	622 (28%)	349 (16%)	−12
	Total	2201 (100%)	2201 (100%)	

It should be noted that knowledge of bedrock topography makes it possible to model the thickness of overlaying deposits, by simple subtraction of surface elevations from bedrock elevations. Such applications are of particular interest in several fields of applied geological and geotechnical engineering and in general, in the earth sciences. Also, it should be mentioned that any model of bedrock topography may be improved by further considering the rock's structural features (lineaments, faults, fracture networks…) that introduce changes in the topography. Such changes can be characterized, and this information included in the datasets used to generate bedrock topography models. Considering such additional structural information (that can be contained in old survey reports for example) was beyond the scope of this study, but we speculate that such a refinement may possibly further improve the accuracy of bedrock topography modeling.

Fig. 19 Increased accuracy of the model. (*a*) Model quality obtained from original datasets. (*b*) Model quality obtained using additional "virtual" top bedrock elevations extracted from the cross-sections between real boreholes
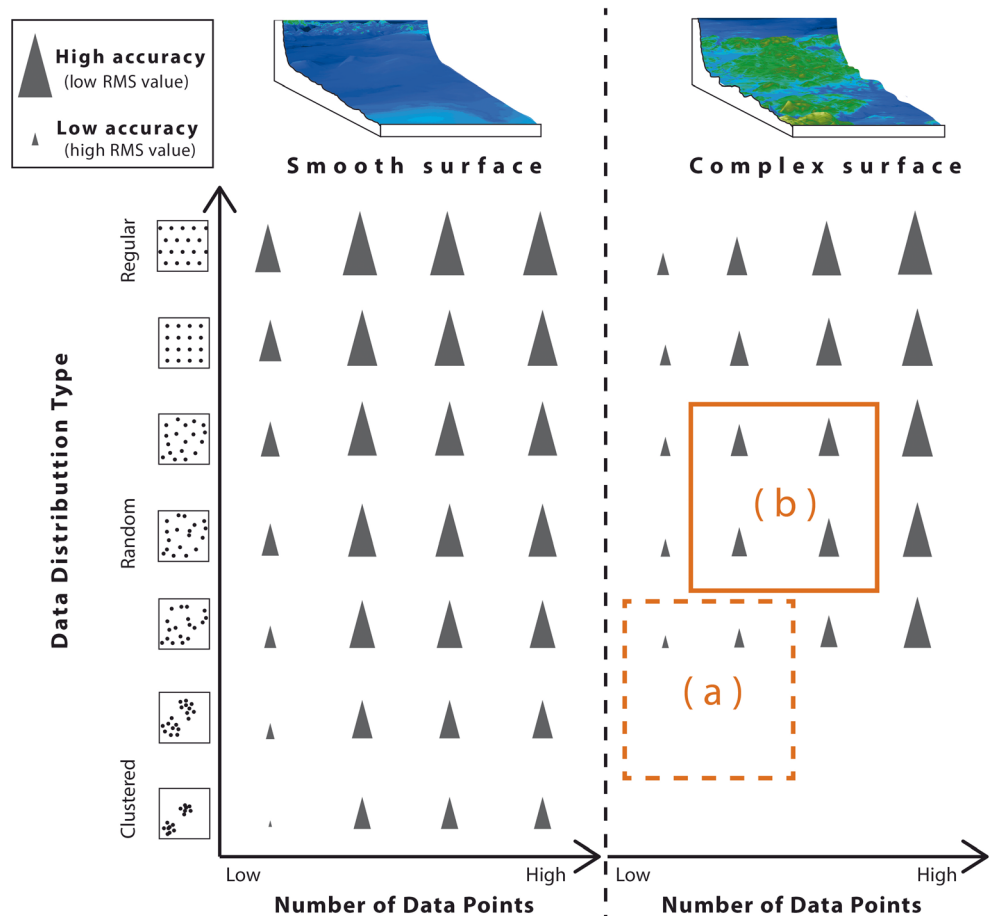
And finally, the approach that consists in using virtual boreholes could be extended to building geological layered models where stratigraphic information is required for interpolating between the interfaces of the different geological layers composing the model. Virtual boreholes could be generated to increase the number of stratigraphic observation points as well as to improve the distribution of these observation points in the same way as what has been presented in this study.

References

Carmichael RS, Henry G (1977) Gravity exploration for groundwater and bedrock topography in glaciated areas. Geophysics 42(4):850–859

Carrara A, Bitelli G, Carla R (1997) Comparison of techniques for generating digital terrain models from contour lines. Int J Geogr Inf Sci 11(5):451–473

Carter J (1992) The effect of data precision on the calculation of slope and aspect using grid DEMs. Cartographica 29(1):22–34

Chang K-T, Tsai B-W (1991) The effect of DEM resolution on slope and aspect mapping. Cartography and Geographic Information Systems 18(1):69–77

Chaplot V, Darboux F, Bourennane H, Leguedois S, Silvera N, Phachomphon K (2006) Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. Geomorphology 77:126–141

Chesnaux R (2013) Regional recharge assessment in the crystalline bedrock aquifer of the Kenogami uplands, Canada. Hydrol Sci J 58(2):421–436

Chesnaux R, Lambert M, Fillastre U, Walter J, Hay M, Rouleau A, Daigneault R, Germaneau D, Moisan A (2011) Building a geodatabase for mapping hydrogeological features and 3D modeling of groundwater systems: application to the Saguenay-Lac-St-Jean region, Canada. Comput Geosci 37:1870–1882

Darsow A, Schafmeister M-T, Hofmann T (2009) An ArcGIS approach to include tectonic structures in point data regionalization. Groundwater 47:591–597

Elmahdy SI, Mansor S, Rodsi A, Bujang BK (2010) Geotechnical modeling of fractures and cavities that are associated with geotechnical engineering problems in Kuala Lumpur limestone, Malaysia. Environmental Earth Sciences 62(1):61–68

Farinotti D, King EC, Albrecht A, Huss M, Gudmundsson Hilmar G (2014) The bedrock topography of Starbuck Glacier, Antarctic Peninsula, as determined by radio-echo soundings and flow modeling. Ann Glaciol 55(65):22–28

Jones NL, Davis RJ, Sabbah W (2003) A comparison of three-dimensional interpolation techniques for plume characterization. Groundwater 41:411–419

MacCormack KE, Brodeur JJ, Eyles CH (2011) Assessing the impact of data quantity, distribution and algorithm selection on the accuracy of 3D subsurface models, Proceeding of the GeoHydro 2011 conference, Quebec City, Quebec, Canada, 6pp

Paulen RC, McClenaghan MB (2003) Three-dimensional bedrock topography and drift thickness models from the Timmins Area, northeastern Ontario: An application of GIS to the Timmins Overburden Drillhole Database, On the Edge: Earth Science at North America's Western Margin. *Annual Meeting of the GAC, MAC and SEG*. Sheraton Wall Centre, Vancouver, British Columbia. May 25–28, 2003

Slattery SR, Barker AA, Andriashek LD, Jean G, Stewart SA, Moktan H, Lemay TG (2011) Bedrock topography and sediment thickness mapping in the Edmonton-Calgary Corridor, Central Alberta: an overview of protocols and methodologies, ercb/ags open file report 2010–12

Tremblay Simard P, Chesnaux R, Rouleau A, Daigneault R, Cousineau PA, Roy DW, Lambert M, Poirier B, Poignant-Molina L (2015) Imaging glacial quaternary deposits and basement topography using the transient electromagnetic method for modeling aquifer environments. J Appl Geophys 119:36–50

Ugwu SA, Eze CL (2009) Mapping bedrock topography using electro-magnetic profiling. J Appl Sci Environ Manag 13(4):43–46

Van Hoesen JG (2014) A geographic information systems (GIS)-based approach to derivative map production and visualizing bedrock topography within the town of Rutland, Vermont, USA. International Journal of Geo-Information 3:130–142

Wise S (2000) Assessing the quality for hydrological applications of digital elevation models derived from contours. Hydrol Process 14:1909–1929

Zimmerman D, Pavlik C, Ruggles A, Armstrong MP (1999) An experimental comparison of ordinary and universal kriging and inverse distance weighting. Math Geol 31:375–390