

A Sweep-Plane Algorithm for Calculating the Isolation of Mountains

Daniel Funke ✉

Karlsruhe Institute of Technology, Germany

Nicolai Hüning ✉

Karlsruhe Institute of Technology, Germany

Peter Sanders ✉

Karlsruhe Institute of Technology, Germany

Abstract

One established metric to classify the significance of a mountain peak is its isolation. It specifies the distance between a peak and the closest point of higher elevation. Peaks with high isolation dominate their surroundings and provide a nice view from the top. With the availability of worldwide Digital Elevation Models (DEMs), the isolation of all mountain peaks can be computed automatically. Previous algorithms run in worst case time that is quadratic in the input size. We present a novel sweep-plane algorithm that runs in time $\mathcal{O}(n \log n + pT_{NN})$ where n is the input size, p the number of considered peaks and T_{NN} the time for a 2D nearest-neighbor query in an appropriate geometric search tree. We refine this to a two-level approach that has high locality and good parallel scalability. Our implementation reduces the time for calculating the isolation of every peak on Earth from hours to minutes while improving precision.

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains

Keywords and phrases computational geometry, Geo-information systems, sweepline algorithms

Digital Object Identifier 10.4230/LIPIcs.ESA.2023.51

Related Version *Technical Report*: <https://arxiv.org/abs/2305.08470> [20]

Supplementary Material *Software*: <https://github.com/dfunke/mountains>
archived at `swh:1:dir:9081f9961a7408668c156e998a1facb21b93e7dc`

Funding This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 882500).



1 Introduction

High-resolution digital elevation models (DEMs) are an interesting example of large datasets with a big potential for applications but equally big challenges due to their enormous size. For example, WorldDEM provided by the TanDEM-X mission covers the entire globe with a resolution of 0.4 arcsecond and is currently the highest-resolution worldwide DEM available [39]. It consists of $6 \cdot 10^{12}$ individual sample points and amounts to approximately 25 TB of data. Modern LIDAR technology allows $< 1 \text{ m}^2$ samples, resulting in more than 300 TB of data for the land surface of the Earth. The algorithm engineering community has greatly contributed to unlocking the potential of DEMs by developing scalable algorithms for features such as contour lines, watersheds, and flooding risks [1, 11, 31]. This paper continues this line of research by studying the isolation of mountain peaks which is a highly nonlocal feature and requires us to deal with the Earth’s complicated (not-quite spherical) shape.

A mountain’s significance is typically characterized using three properties – elevation, isolation, and prominence [23]. Whereas elevation is a fundamental property, isolation and prominence are derived measures. Isolation – also referred to as dominance – measures the

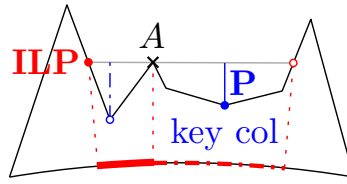


© Daniel Funke, Nicolai Hüning, and Peter Sanders;
licensed under Creative Commons License CC-BY 4.0
31st Annual European Symposium on Algorithms (ESA 2023).

Editors: Inge Li Gørtz, Martin Farach-Colton, Simon J. Puglisi, and Grzegorz Herman; Article No. 51;
pp. 51:1–51:17



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** Illustration of a mountain A 's isolation (red) and prominence (blue). The isolation is measured along the sea-level surface of the Earth.

distance along the sea-level surface of the Earth between the peak and the closest point of higher elevation, known as the *isolation limit point* (ILP). Prominence measures the minimum difference in elevation from a peak and the lowest point on a path to reach higher ground – called the *key col*. Refer to Figure 1 for an illustration.

Whereas previously both measures had to be determined manually by laboriously studying topographic maps, they can now be computed algorithmically using DEMs. Kirmse and de Ferranti [28] present the current state-of-the-art to compute both measures. They use an algorithm to determine a peak's isolation by searching in concentric rectangles of increasing size around the peak for higher ground. To determine the isolation of every peak on Earth, their algorithm has a worst case running time of $\mathcal{O}(n^2)$, with n being the number of sample points in the DEM. For high-resolution DEMs of the Earth and other celestial bodies, e. g. the Moon [6] or Mars [24], algorithms with better scalability are needed. Also, Continuously Updated Digital Elevation Models (CUDEMs), such as NOAA's DEM of North America's coastal regions [2], require efficient algorithms for frequent reprocessing. In particular, parallel and external algorithms are required.

Contribution and Outline. After presenting basic concepts in Section 2 and related work in Section 3 we describe the main algorithmic result in Section 4. We begin with a sequential algorithm that sweeps the surface of the Earth top-down with a sweep-plane storing the points having surfaces at that elevation. Processing a peak then amounts to a nearest neighbor query in the sweep-plane data structure. This results in an algorithm with running time $\mathcal{O}(n \log n + pT_{\text{NN}})$ where n is the input size, p the number of considered peaks and T_{NN} the time for a 2D nearest-neighbor query in an appropriate geometric search tree. To make this more scalable, we then develop an algorithm working on the natural hierarchy of the data which is specified in *tiles*. This algorithm performs most of its work in two scans of the data which can work independently and in parallel tile-by-tile. Only the highest points in each tile need to be processed in an intermediate global phase. Each of these three phases has a structure similar to the simple sequential algorithm.

In Section 5 we then explain how 2D-geometric search trees can be adapted to work on the surface of the Earth by deriving the required geometric predicates. After outlining implementation details in Section 6, in Section 7 we evaluate our approach using the largest publicly available DEM data.

2 Preliminaries

Spherical Geometry. Planets can generally be approximated by spheres. In the geographic coordinate system, a point $p = (\phi, \lambda)$ on the surface of a sphere is identified by its latitude ϕ and longitude λ . Latitude describes p 's north-south location, measured from the equator, $\phi \in [-90^\circ, 90^\circ]$ with negative values south of the equator. Longitude describes p 's east-west location, measured from the prime meridian through Greenwich, $\lambda \in (-180^\circ, 180^\circ]$ with negative values west of the prime meridian.

For close distances, Earth's surface is sufficiently flat to use planar Euclidean distances between points. For longer distances, spherical distance calculations are necessary, which are computationally more expensive. On the surface of the sphere, two points are always connected by at least two great circle segments. The shortest such segment, the *geodesic*, can be computed for points $p_1 = (\lambda_1, \phi_1)$ and $p_2 = (\lambda_2, \phi_2)$ with sphere radius R according to

$$\alpha = \sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right) \cos(\lambda_1) \cos(\lambda_2)$$

$$d(p_1, p_2) = R \tan^{-1}\left(\frac{\sqrt{\alpha}}{\sqrt{1-\alpha}}\right) \quad (1)$$

As the Earth and most other planets are not perfect spheres, ellipsoids are a more accurate approximation of their shape. The World Geodetic System (WGS) [29] defines such an ellipsoidal approximation of the Earth, requiring the following – even more computationally expensive – distance formula:

$$\begin{aligned} \Delta\lambda &= (\lambda_2 - \lambda_1)/2 & \Delta\phi &= (\phi_2 - \phi_1)/2 & \Lambda &= (\lambda_2 + \lambda_1)/2 \\ s &= \sin^2 \Delta\phi \cdot \cos^2 \Delta\lambda + \cos^2 \Lambda \cdot \sin^2 \Delta\lambda \\ c &= \cos^2 \Delta\phi \cdot \cos^2 \Delta\lambda + \sin^2 \Lambda \cdot \sin^2 \Delta\lambda \\ w &= \tan^{-1}(\sqrt{s}/\sqrt{c}) & r &= \frac{\sqrt{sc}}{w} \\ d(p_1, p_2) &= 2aw \left(1 + f \frac{3r-1}{2c} \sin^2 \Lambda \cdot \cos^2 \Delta\phi - f \frac{3r+1}{2s} \cdot \sin^2 \Delta\phi \cdot \cos^2 \Lambda\right) \end{aligned} \quad (2)$$

with a and f denoting the equatorial radius and flattening of the WGS84 ellipsoid [29].

Digital Elevation Models. Digital Elevation Models (DEMs) have become one of the most important tools to analyze the Earth's surface in geographic information systems. They represent the surface of the Earth by providing elevation measurements on a grid of sample points. The data is mostly stored in files of 1 square degree of coverage each, called tiles. A tile is addressed by its smallest latitude and longitude. In order to enable seamless processing of several tiles, each tile stores one sample row/column overlap with its neighbors. The resolution of DEMs is given by the length of one sample at the equator in arcseconds (").

State-of-the-art worldwide DEMs, such as WorldDEM, provide a resolution of 0.4" spacing between sample points – about 12 m at the equator [39]. Local DEMs, such as the swissALTI^{3D}, even have a resolution of only 0.5 m sample point spacing [21]. Freely available DEMs, such as The Shuttle Radar Topography Mission (SRTM) [25], provide a resolution of 3" – about 90 m at the equator – and an absolute vertical height error of no more than 16 m for 90% of the data [32]. Unfortunately, it does not provide global coverage¹ and contains large void areas, especially in mountainous regions, which are of particular interest for us. In a laborious process, de Ferranti [12] fused raw SRTM data with other publicly available datasets [14, 36] and digitized topographic maps to create a worldwide, void-free DEM available at www.viewfinderpanoramas.org.

¹ Only areas between 60° North and 56° South are covered.

3 Related work

Graff et al. [22] are the first to use DEM data to classify terrain into mounts, plains, basins or flats. As there is no definitive definition of these terrain features, several methods for terrain classification are studied in the literature [38], including fuzzy logic [16, 17] or, more recently, deep learning [37].

Prominence has received most of the attention when it comes to algorithmic computation of mountain metrics [26, 28]. Kirmse and de Ferranti [28] present the current state-of-the-art regarding isolation and prominence computation. As their main focus lies on the prominence calculation, they present a rather simple $\mathcal{O}(n^2)$ time isolation calculation algorithm.

In their algorithm they first calculate potential peaks which are samples that are at least as high as their eight neighboring samples. Afterwards, a search for the closest higher ground (ILP) is conducted, where, centered on each peak, concentric rectangles of increasing size are checked to find a sample with higher elevation. Since the closest higher ground of a peak could also be on a neighboring tile, these might be checked as well. If a higher ground was found, the distance to this sample is used to constrain the search in neighboring tiles. If not, tiles in increasing rectangles around the peak-tile are checked until an ILP is found or the complete world has been checked. Before searching a neighboring tile, the maximum elevation of the tile is checked. When it is smaller than the peak elevation, the tile can be ignored. Tiles and their maximum elevation are cached, because they need to be loaded rather frequently. Inside the tile that contains the peak, planar Euclidean distance approximations are used to find an ILP, for neighboring tiles, distances are computed according to the spherical distance function. Because a majority of peaks have a small isolation, the ILP is often within the same tile as the peak, so mostly planar Euclidean distance approximations are used to determine a peak's ILP. This reduces computational costs but also the accuracy for peaks with small isolation. Only in a final step before output is the distance between a peak and its determined ILP calculated using the precise, but expensive, ellipsoid distance function. Kirmse and de Ferranti's algorithm is unnamed, we refer to it as CONCIISO for brevity.

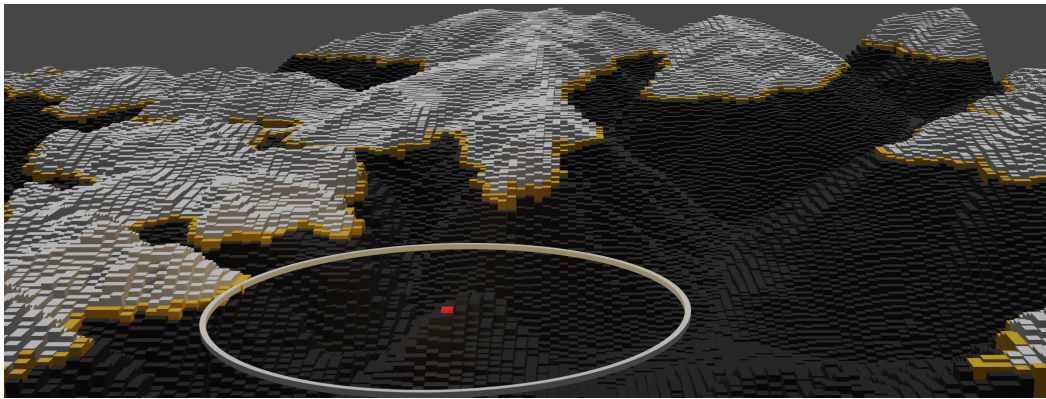
Sweep-line algorithms are introduced by Bentley and Ottmann [7] to compute line segment intersections. The technique is generalized by Anagnostou et al. to three-dimensional space [3]. Sweep-line algorithms have been applied to various geometric problems in two- and three-dimensional space, such as computing Voronoi diagrams [18] or route mining [40]. R-Trees are often used in spatial databases and have been adapted to geodetic distance computations by Schubert et al. [33].

4 Algorithm

In this section we present our novel sweep-plane algorithm to calculate the isolation of mountain peaks. The algorithm takes as input the search area – a quadrilateral A defined by its north-west and south-east corners – as well as the DEM data. We will first present a single-sweep algorithm that processes the entire search area in one sweep-plane pass. Subsequently, we present a scalable three-pass algorithm that reads the DEM-tiles of A twice and can process tiles in parallel.

4.1 Single-Sweep Algorithm

To determine the isolation of a peak, its closest point with higher elevation – the Isolation Limit Point (ILP) – needs to be found. Given the search area A , each sample point p of the DEM within A has two associated events: an *insert* event at p 's elevation and a *remove* event at the elevation of its lowest immediate (NESW) neighbor. Additionally, if a point is

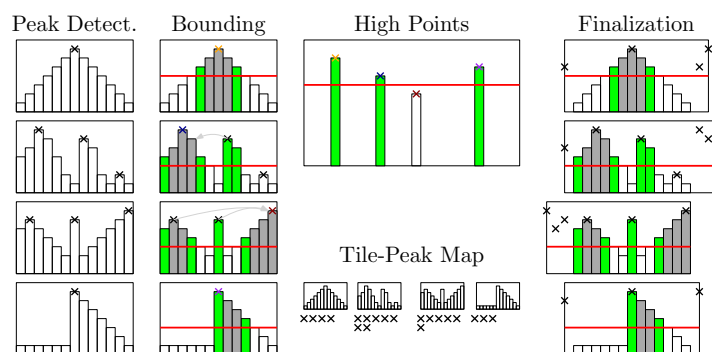


■ **Figure 2** Illustration of DEM grid and sweep-plane algorithm by using a voxel representation. Red represents the current peak. Orange DEM points are active and contained in the sweep-plane data structure. White points are inactive and already removed from the data structure. Black points have not yet been processed.

the high-point of its 3×3 neighborhood it is associated with a *peak* event at its elevation. We describe the peak detection routine in more detail in Section 6. All events are created at the beginning of the algorithm and can therefore be added to a static sequence sorted by descending elevation. Peak events are processed before other events at the same elevation.

The sweep-plane then moves downward from the highest sample point to the lowest and traces the contour lines of the terrain, refer to Figure 2. The sweep-plane data structure SL is a two-dimensional geometric search tree that maintains a set of currently *active* points. A sample point p becomes active at its insert event, when it is swept by the sweep-plane and inserted into SL . Point p becomes *inactive* and is removed from SL at its remove event, at which point its lowest neighboring point is activated and thus all of p 's neighboring points are either active or have already been deactivated. When a peak event for sample point p is processed, a nearest neighbor query for the closest active sample point to p in SL is performed. The returned point w is the ILP for the peak: Since w is active and peaks are processed before other events, w must be higher than p . There cannot be any closer ILP as points activated later are not higher than p and since higher points v that are already deactivated are surrounded by points that are all higher than p . At least one of them must be closer to p than v .

Analysis. Given a DEM with n points, we can detect its p peaks in time $\mathcal{O}(n)$. The resulting $2n + p$ events can be sorted in $\mathcal{O}(n \log n)$ time by elevation. Insertion and removal operations take $\mathcal{O}(\log n)$ worst case time in several geometric search trees [8, 15, 35]. Nearest neighbor search complexity is more complicated and still an open problem in many respects. Therefore, we describe it abstractly as T_{NN} . Simple trees used in practice such as k -D-Trees [8] or Quadtrees [15] achieve logarithmic time “on average”. Cover trees [9] achieve logarithmic time when an *expansion parameter* of the input set is bounded by a constant. Approximate queries within a factor of $1 + \epsilon$ are possible in time $\mathcal{O}(\log(n)/\epsilon^2)$ [4]. Overall, we get the claimed time $\mathcal{O}(n \log n + pT_{\text{NN}})$.



■ **Figure 3** Overview of our scalable multi-pass algorithm for four tiles in two dimensions. Peaks are detected for each tile with the peak finding algorithm of CONCISSO. The bounding pass determines an upper bound on the isolation for each peak and assigns it to tiles that could contain closer ILPs in the Tile-Peak map. The high-points of each tile without a local upper bound are processed by a separate pass considering only the high-points. In the finalization pass, all peaks assigned to a tile are checked for an ILP within the tile. The final determination of the closest ILP out of the found candidates for each peak is not depicted. For the sweep-plane data structure, green points are active in the sweep-plane passes, gray points are inactive and hollow points are not yet processed. Peaks are marked by crosses. The high-point of each tile is marked by a different color cross for each tile.

4.2 Scalable Multi-Pass Algorithm

High-resolution DEMs are massive data sets of currently up to 25 TB that require scalable algorithms. The algorithm described in the previous section can be parallelized to some extent but its sweeping character limits parallelism. Moreover, a geometric search tree covering the entire Earth could get quite large. We therefore develop a two-level algorithm that allows for more coarse-grained parallelism and better locality. We adopt the natural hierarchy of the input data using tiles of a fixed area but note that reformatting into smaller or finer tiles would be possible in principle. Furthermore, a more general multi-level algorithm could be developed using a similar approach.

Multi-level algorithms start at the finest level to extract information for global² processing at coarser level. The global results are then passed down to compute the final solution. In our two-level algorithm this implies that we have two passes reading the DEM tiles from external memory while a single global pass works with simple per-tile information. The first (*bounding*) and last (*finalization*) pass can work in parallel on each tile. The global (*high-point*) pass works in internal memory and is also parallelized. The first two passes establish a global Tile-Peak map that stores for each tile which peaks *can* have an ILP in it. The third pass processes these assigned peaks for a tile and determines their ILPs. All passes follow a very similar structure to our single-sweep algorithm from Section 4.1. They are described in detail in the following. Additionally, Figure 3 provides an overview of our algorithm.

Bounding Pass. The purpose of the first pass is to establish an upper bound on the isolation of a peak and therefore limit the the number of tiles that need to be searched for its ILP in the finalization pass. To establish this upper bound, we find a tile-local ILP for each peak using our single-sweep algorithm. Only the highest point in each tile will not have a tile-local ILP and is treated separately in the high-point pass. Given the upper bound on the isolation

² Pun intended.

of a peak p , we can assign p to all tiles within radius of the upper bound that could contain a closer ILP for p . In the third pass, these neighboring tiles then need to process p in order to find ILP candidates within them. To link peaks to tiles for processing in the third pass, we build a Tile-Peak map that stores for each tile a list of peaks that could have an ILP within it.

High-Point Pass. As there is no local ILP for the high-point h of each tile, we address high-points separately in the second pass. This pass uses only one type of event – a combination of insert and peak event from the other passes – and processes only one point per tile, namely their high-points. For each high-point h , we perform a nearest neighbor query in the sweep-plane data structure for h 's closest higher point p . The distance between h and p is an upper bound on h 's isolation. The sweep-plane data structure is then traversed again to find all tiles whose closest point to h is at least as close to h as p and h is linked to these tiles in the Tile-Peak map. Finally, h is inserted into the sweep-plane data structure and the next lower high-point is processed.

Finalization Pass. In the third pass, each tile is processed again by a sweep-plane algorithm similar to the single-sweep variant. The peak events of this pass are all peaks that have been assigned to the tile in the Tile-Peak map. For each assigned peak, the ILP candidate within the processed tile is determined. After all tiles have been processed by the third pass, each peak has as many ILP candidates as tiles it has been linked to. Thus, in a final step, for each peak the closest found ILP candidate is set as the true ILP of the peak.

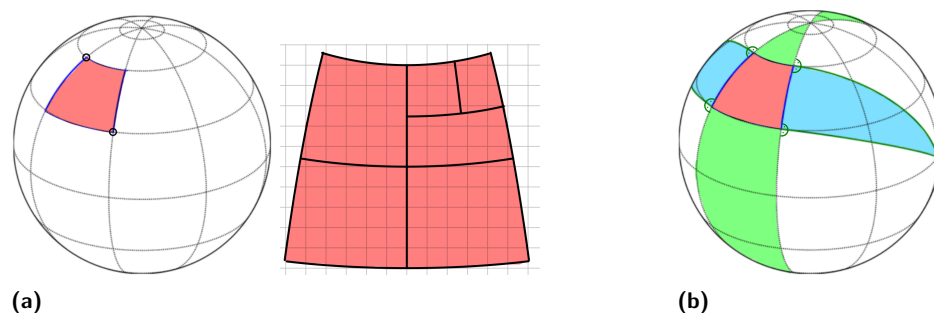
Algorithmic Details and Outline of Analysis. Let us first look at the total **work** performed to process A tiles containing n sample points overall. The bounding pass performs a similar amount of work as the global algorithm except that it defers nonlocal nearest neighbor searches to the subsequent passes. More precisely, the high-points of each tile are deferred to the high-point pass while other peaks are deferred to neighboring tiles. On average, there is a constant number of such neighbor tiles.³

The high-point pass potentially defers nearest-neighbor-search work for the highest point of each tile to a potentially large number of tiles. However, this is not so different from what a search-tree based nearest neighbor search in a global tree data structure would do. Our two-level algorithm can be viewed as a vertically split (quad-)tree algorithm where updates and nearest neighbor searches are reordered to improve locality. The overall amount of work done is quite similar though.

The finalization pass can be viewed as completing the deferred nearest neighbor searches. Thus, the main overhead of the two-pass algorithm compared to the global algorithm is that the data is swept through search trees twice. We mitigate this effect by building only a coarse search tree in the bounding pass. This has no negative effect on precision as the bounding pass is only needed to identify the tiles where an ILP can be. Only the finalization pass computes the actual ILPs.

Let us now look at **I/O-costs**. Assume one tile and the high-points fit in internal memory. This is similar to a “semi-external”-assumption used in previous algorithm engineering papers on DEM processing, e. g. [1]. The I/O volume of our two pass algorithm is dominated by

³ Near the poles, there are many candidate tiles that may be closer than a local ILP. However, averaged over all tiles, this effect is not large. It is interesting to note though that this is an artifact of a pseudo-high longitudinal resolution that is not reflected in the actual precision of the sensors but stems from artificially mapping the data using a Mercator projection. Meshes with more uniform cells are possible, for example the icosahedral grid used in some modern climate/weather models [27].



■ **Figure 4** Representation of a latitude-longitude aligned quadrilateral and areas for the different min-distance cases between a quadrilateral and a point.

reading all tiles twice.⁴ This is actually better than an external-memory implementation of the global algorithm as that algorithm would have to sort the data by altitude before being able to scan the input in the right order for the sweep. Even using pipelining [13], sorting would require at least two reading and one writing pass over the DEM data.

Finally let us consider **parallelization**. Bounding and finalization passes can work on A tiles in parallel. In order to also parallelize the high-point pass, we implement a variant that uses a static search tree with subtrees augmented by the maximum elevation occurring in a subtree. Then the A nearest-dominating-point searches can be done in parallel. See Section 6 for details.

5 Predicates for Search Trees on Spherical Surfaces

The sweep-plane data structure needs to be a dynamic data structures that supports efficient nearest neighbor queries. Space-partitioning trees such as k -D trees or Quadtrees are well-suited data structures for this application [8, 15]. These trees recursively divide the input space into smaller and smaller blocks. For a spherical surface, the space is divided into quadrilaterals which are aligned with latitude and longitude, refer to Figure 4a. A quadrilateral is defined by its north-west and south-east corners. Each quadrilateral can then be further subdivided into smaller quadrilaterals. The root of a space-partitioning tree covers the entire input area.

Our sweep-plane algorithm requires two geometric primitives for these trees: a) whether a given point p lies inside a quadrilateral Q and b) the shortest distance between p and Q . The former can be easily answered by comparing p 's latitude and longitude with Q 's north-east and south-west corner. For the latter, there are four configurations of p and Q that need to be considered, refer to Figure 4b:

1. $p \in Q$ (red area),
2. p is between the longitude lines of Q (green area),
3. p is between the four great circles through the corners of Q , which are perpendicular to the longitude edges of Q (blue area),
4. all other positions (white area).

⁴ In comparison, the Tile-Peak map has negligible data volume and can be handled in an I/O-efficient way by observing that the first two passes only insert to it and the third pass only reads it tile by tile. Thus, we can first buffer its data in an external log which is sorted by tile before the finalization pass.

For case 1 the distance is zero. In case 2, the closest point $s \in Q$ to p is on the intersection between the longitude line of p and one of the latitude-edges of the quadrilateral, as the shortest distance between two latitudes is along the longitude lines. The shortest distance of p to Q is thus the shorter distance between p and the north and south latitude of Q .

In the remaining cases, s must lie on one of the longitude lines of the quadrilateral. To determine the longitude edge closest to p , we calculate the center longitude of Q and rotate it to align with the meridian. We rotate p by the same amount. If the longitude of p is now positive, the west longitude-edge of Q is closer to p , otherwise the east one.⁵ Having determined the closest longitude edge, we can now calculate point s using linear algebra in Euclidean space as described in the accompanying technical report [20, Appendix A.2]. If the latitude of s is between the top and bottom latitude of Q , s is the point with the shortest distance to p in Q (case 3). Otherwise one of the corners is the point with the shortest distance to p (case 4).

Given these primitives, insert and query operations on space-partitioning trees for spherical surfaces are identical to the ones for Euclidean space.

6 Implementation

We implement our new sweep-plane algorithm in the MOUNTAINS C++ framework by Kirmse and de Ferranti [28].⁶ The framework provides essential functionalities for the work with tiled DEM data, such as data loading and conversion, peak discovery and distance computations. Our code is available on Github.⁷ In the following we provide details our implementation.

Data Structure. We implement a fully dynamic k -D-Tree, that supports insert and remove operations in $\mathcal{O}(\log n)$ worst case time and nearest neighbor searches in $\mathcal{O}(\log n)$ expected time [19] using the predicates described in the previous section.⁸ Points are stored in the leaves of the tree, which have a fixed capacity of C points. On exceeding capacity C , a leaf is split according to a center-split policy along the longer side of the quadrilateral. The first points inserted into the tree often belong to the highest peak in a tile and are thus close to each other. In order to prevent the tree from degenerating, we pre-build the first k levels in a Quadtree-like manner.⁹ To improve cache-efficiency, tree nodes are allocated in blocks and are re-used after deletion.

Another data structure used is the Tile-Peak map, which we implement as an internal memory hash map with the latitude and longitude of a tile as key and a list of peaks as value.

Algorithm. Given the search area A , all contained DEM tiles can be processed in parallel. We use a work queue and thread pool to distribute the computations among the available processing elements. The Tile-Peak map is initialized in advance with all tiles. To sort the events of a pass, we use the efficient sorting algorithm `ips4o` of Axtmann et al. [5].

⁵ This is possible because we split the Earth at the antimeridian. Therefore, the western longitude of Q is always smaller than the eastern one.

⁶ <https://github.com/akirmse/mountains>

⁷ <https://github.com/dfunke/mountains>

⁸ We also adapted a Quadtree using our predicates, which was however outperformed by the k -D-Tree in our experiments.

⁹ Our experiments show $k = 4$ to be a good choice.

Bounding Pass. We use the peak detection algorithm of Kirmse and de Ferranti [28, Sec. 2.2] to find all peaks within a tile. It considers all points to be peaks that have eight neighboring points of lower or equal elevation. If a peak consists of several equally high sample points, only a single one is added to the set of peaks.¹⁰ Since only an upper bound is calculated during the bounding pass, we down-sample the resolution of the DEM after peak detection. Since all peaks are within the tile which is processed, distances are rather small and can be approximated using planar Euclidean geometry during the nearest-neighbor search. The upper bound is computed using the spherical distance function according to Equation (1). If the determined upper bound on the isolation of a peak is below a threshold I_{\min} we discard the peak as insignificant. Peaks with an upper bound above I_{\min} are added to the Tile-Peak map for processing in the finalization pass as described in Section 4.2.

High-Point Pass. While processing the tiles in the bounding pass, we build a geometric search tree T that partitions the entire search area down to the tile level. Internal nodes of T save the highest elevation in its sub-tree. After all tiles have been processed, we can use this information to efficiently determine upper bounds on the isolation of the high-points of each tile. For each high-point h , we find the closest tile containing a point of higher elevation than h in search tree T . The maximum distance between h and any point within the found tile serves as an upper bound on h 's isolation. Given this upper bound we can add h to all tiles containing potential ILPs in the Tile-Peak map. This approach is trivially parallelizable over the number of high-points in the search area.

Finalization Pass. In this pass we use the full resolution of the input DEM. For each tile, the peaks processed in this pass are the ones that are assigned to it in the Tile-Peak map. We use ellipsoid distance computations according to Equation (2).

7 Evaluation

In this section we evaluate our novel sweep-plane algorithm to calculate the isolation of mountain peaks, which we named SWEEPISO. We evaluate it with regard to runtime behavior and solution quality and compare it against CONCIISO from Kirmse and de Ferranti [28].

¹⁰The algorithm by Kirmse and de Ferranti always chooses the north-west corner of a peak.

■ **Table 1** DEM models of the Earth, Mars and the Moon used in our experiments. The reported runtime is the single-threaded runtime for the entire data set.

Name	Resolution	Pixels	Size	Coverage	Runtime
Earth SRTM	3" (90 m)	20×10^9	71 GB	Global	2.07 h
Earth SRTM NA-EU	1" (30 m)	49×10^9	105 GB	Cont. US+CAN+EU	3.8 h
Moon SLDEM2015	7" (59 m)	11×10^9	22 GB	[60°N, 60°S]	2.5 h
Mars MGS MOLA - MEX HRSC Blended	12" (197.6 m)	5.7×10^9	11 GB	Global	1.3 h

Experimental Setup. All benchmarks are conducted on a machine with an AMD EPYC Rome 7702P with 64 cores and 1024 GB of main memory. The DEM data is stored on a Intel P4510 2 TB NVMe SSD. We use the 3'' data set from viewfinderpanoramas [12] with worldwide coverage. To build smaller test instances from the worldwide data set, we choose a random starting tile and add neighboring tiles in a spiraling manner around it until the desired number of tiles is reached. For each instance size, we generate several instances to cover a wide range of terrain. Additionally, we use the 3'' and 1'' DEM for the USA, Canada and most of Europe from viewfinderpanoramas [12] to study the scaling behavior for higher-resolution DEMs. This data set corresponds to 4294 tiles or roughly 16 % of the total test data set. We will call this data set NA-EU. We also use data sets from other celestial bodies: Mars [24] and the Moon [6]. Table 1 lists some properties of the studied data sets along with the runtime of our algorithm to determine the isolation of every peak contained in them.

For all benchmarks we use an isolation threshold I_{\min} of 1 km and report the mean runtime of 5 runs. I/O costs are not part of the reported figures as we use the MOUNTAINS framework [28] for them without an attempt at optimization. They are about the same time as computation for 3'' DEMs and about 10 % of computation time for 1'' ones and thus could be overlapped with the computation in an optimized framework.

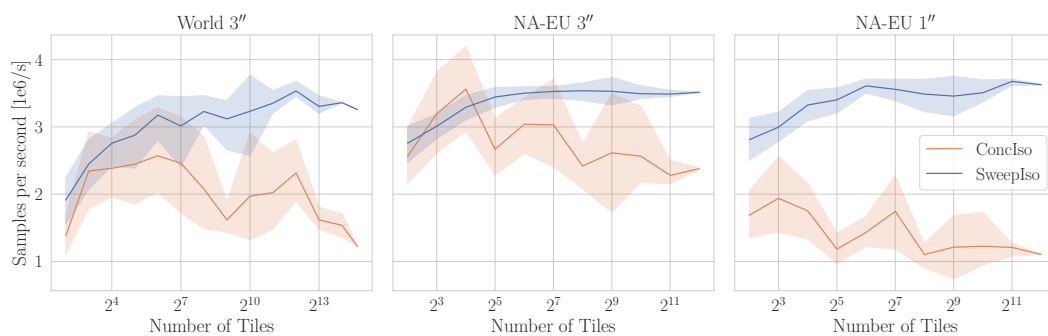
7.1 Runtime and Scaling Behavior

The runtime of the algorithms depends on the number of sample points in the DEM. These can either increase due to a larger search area or a higher-resolution DEM. Another factor is the number of processed peaks, since every peak starts a local search in CONCISO and a nearest neighbor query in SWEEPISO. A larger search area increases the number of tiles and the number of peaks, whereas higher-resolution data mostly increases the number of points per tile. High-resolution DEMs can contain more peaks than lower-resolution ones due to the more truthful representation of the terrain, however these are predominantly low-isolation peaks and are filtered out in the first pass. We study both effects in our experiments by using different resolution DEMs as well as increasing search areas.

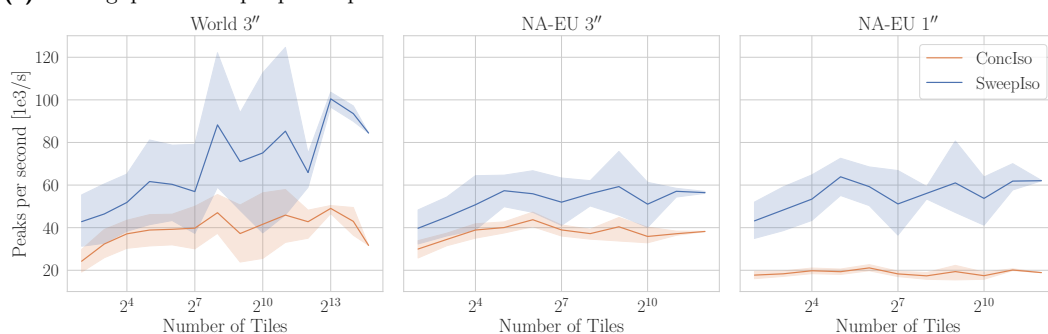
SWEEPISO exhibits a nearly constant throughput of sample points per second with increasing search area, while CONCISO's throughput degrades – refer to Figure 5a. Figure 8a shows that SWEEPISO outperforms CONCISO by a factor of 2 to 3 in terms of sample point throughput. For instance, we reduce the time required to compute the isolation of every peak on Earth from 9 h down to 3.5 h. However, CONCISO's runtime scales significantly better than its $\mathcal{O}(n^2)$ worst case runtime bound would suggest. This is because most peaks have a relatively low isolation. In fact 99.996 % of discovered peaks on the world data set have an isolation below 50 km and more than 99 % below 10 km. Since one tile covers on average an area of 70 km \times 111 km, the nearest higher point is most of the time within the same tile as the peak, where fast approximations for the distance calculation are used. Nevertheless, for high-resolution DEMs the computation cost per peak is significantly higher for CONCISO than for SWEEPISO, refer to Figure 5b. A more in-depth analysis of the runtime behavior of both algorithms can be found in the accompanying technical report [20, Appendix B].

Figure 6 shows the runtime composition of the bounding and the finalization pass. As the DEM resolution is reduced in the bounding pass, the finalization pass requires the majority of the runtime, especially for higher-resolution inputs. Geometric computations only make a small fraction of the runtime. The high-point pass requires just 0.001 % of the total execution time and is therefore omitted in the figure.

51:12 A Sweep-Plane Algorithm for Calculating the Isolation of Mountains

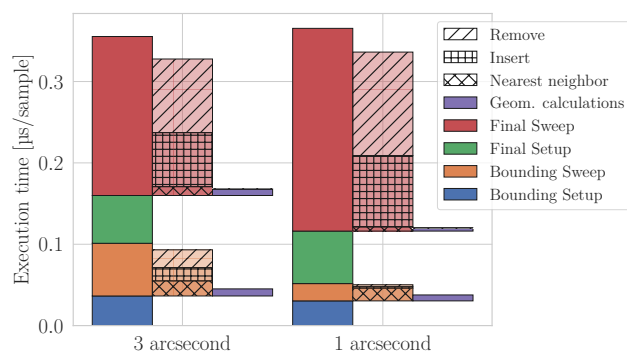


(a) Throughput in sample points per second over number of tiles.

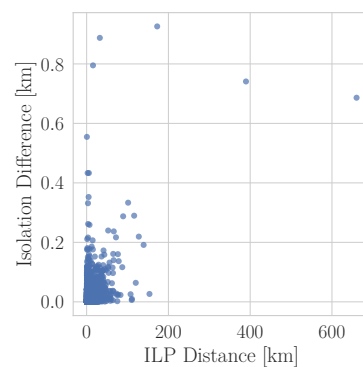


(b) Throughput in processed peaks per second over number of tiles.

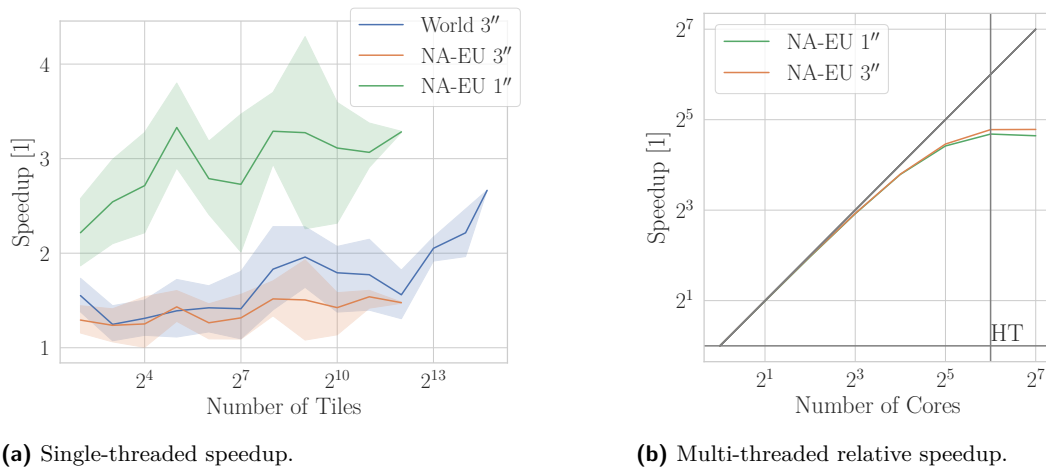
■ **Figure 5** Single-threaded runtime comparison of SWEEPISO and CONCLISO.



■ **Figure 6** Runtime composition of the bounding and finalization pass for 3'' and 1'' data. The sweep-plane operations of these passes are further subdivided into remove, insert, nearest neighbor search operations as well as the time for distance calculations.



■ **Figure 7** Comparison of difference in isolation between SWEEPISO and CONCLISO and distance between found ILPs.



(a) Single-threaded speedup.

(b) Multi-threaded relative speedup.

■ **Figure 8** Single-threaded speedup of SWEEPISO over CONCISO and relative speedup of SWEEPISO for multi-threaded execution.

Figure 8b shows the results of our multi-threaded runtime experiments with the NA-EU data set. SWEEPISO scales well with physical cores, but does not benefit from hyper-threading (HT). For 64 cores it reaches a speedup of 25. We verified the scaling behavior of the algorithm for the entire Earth and were able to confirm a speedup of 25 on 64 cores. This corresponds to a runtime of 8 min to calculate the isolation of every peak on Earth. We were not able to execute CONCISO with multiple threads due to issues in the implementation.

7.2 Solution Quality

In principle, at any particular time, the isolation and witnessing ILP of a peak are a well defined. However, the actually computed values depend on imprecisions in both the data and the used algorithms.

Due to the design of SWEEPISO, more expensive distance approximations can be used to find the ILP than in CONCISO. This results in closer ILPs being found, as shown in Figure 7, which displays the distribution of the difference in isolation and the distance between the found ILPs between SWEEPISO and CONCISO, using the same data. Even if the isolation values between the two algorithms do not vary greatly, the distances between the ILPs do. For example for the Cerro Gordo summit in Mexico both algorithms find ILPs that are more than 600 km apart while SWEEPISO's ILP is merely 0.8 km closer to the peak.

To further evaluate the results, we used the collection of peaks with more than 300 km isolation from the website peakbagger.com [34]. The comparison showed that for about 75% of peaks in this list the isolation deviation is below 2 km. In a DEM, a sample point corresponds to the average elevation of the area it represents. This often leads to an underestimation of a peak's elevation, sometimes significantly [28]. For instance, according to our DEM data, Galdhøpiggen (the highest point of Scandinavia) is 11 m lower than Glittertind. This causes a significant change in isolation of both mountains. Table 2 lists the five summits with the biggest differences in isolation between peakbagger and our calculation. The five most isolated peaks on Mars and the Moon as determined by our algorithm are presented in Table 3. To the best of our knowledge, this is the first such list.

51:14 A Sweep-Plane Algorithm for Calculating the Isolation of Mountains

■ **Table 2** Biggest isolation differences between discovered and peakbagger (PB) data.

Mountain	PB Rank	PB Isolation	SweepIso Isolation	Notes
Galdhøpiggen (Norway)	47	1568.3	12.9	<i>Galdhøpiggen</i> DEM elev.: 2455 m – real elev.: 2469 m [30] <i>Glittertind</i> DEM elev.: 2466 m – real elev.: 2457 m [30] Isolation 1563.5 km
Mt. Kirkpatrick (Antarctica)	45	1585.2	53.5	found ILP at -83.8967;168.3733 DEM elev.: 4416 m – real elev.: 4528 m [30] ILP elev.: 4416 m
Mt. Hope (Antarctica)	75	1113.5	203.4	No data found in other sources
Dome Charlie (Antarctica)	92	971.8	248.6	found ILP at -76.3783;116.3708 DEM elev.: 3265 m – real elev.: 3233 m [10] ILP elev. 3266 m
Mauga Silisili (Samoa)	23	2245.1	2502.8	DEM elev.: 1854 m – real elev.:1863 m [30] PB ILP elev.: 1879 m found ILP elev.: 1836 m
Cocos Islands High Point	94	961.4	947.4	Found ILP at Enggano Island

■ **Table 3** Most isolated peaks on Mars and the Moon according to the DEM data listed in Table 1.

Name	Coordinates		Elevation	ILP		Isolation
<i>Mars</i>						
Olympus Mons	17.33°N	133.42°W	21 226 m	-	-	∞
Cruls crater wall	42.27°S	163.78°E	4331 m	46.79°S	147.63°W	2037.43 km
Near Cyclopia	6.34°S	129.19°E	3806 m	18.59°N	149.99°E	1913.59 km
Huygens crater wall	10.06°S	54.62°E	4604 m	10.50°S	85.09°E	1776.76 km
Ascraeus Mons	11.76°N	104.53°W	18 321 m	17.89°N	131.64°W	1593.92 km
<i>The Moon</i>						
Engel'gardt crater wall	5.41°N	158.63°W	10 783.3 m	-	-	∞
Between Tacitus and Fermat craters	18.93°S	19.17°E	4832.33 m	60.0°S	72.65°W	2261.5 km
Near Lewis crater	21.04°S	112.33°W	9459.83 m	5.26°N	157.98°W	1574.56 km
Near Calippus crater	39.09°N	9.46°E	3625.68 m	9.63°S	2.06°E	1492.13 km
Dellinger crater wall	7.25°S	142.0°E	7561.73 m	12.91°S	170.22°W	1434.95 km

8 Conclusion and Future Work

We have presented SWEEPISO, a scalable and efficient algorithm to compute the isolation of peaks. SWEEPISO considerably outperforms the previous, more brute-force, state-of-the-art approach. The performance gains also enable more accurate distance calculations at decisive places resulting in higher accuracy. SWEEPISO is able to process the entire Earth for currently publicly available data within minutes. This is relevant as it indicates that SWEEPISO can also handle higher-resolution data that is available commercially or that will be available in the future. SWEEPISO’s two-level semi-external sweeping architecture may also be an interesting design pattern for other computations on massive DEM data. Furthermore, SWEEPISO could serve as a benchmark for dynamic nearest neighbor search data structures.

Future Work. From an application perspective, it would be interesting to compute not only isolations for a given, necessarily imprecise data set but to compute confidence bounds that take into account error margins in the input data. This would be possible at a moderate increase in cost. Peaks could be replaced by enclosing boxes/circles while vertical errors could be handled by having “may-be-there” and “must-be-there” insertion events and sweep-plane data structures. Geographically most interesting would be those isolations that change a lot depending on how high exactly particular pairs of peaks are. Those pairs could then be valuable targets for additional data cleaning or new measurements.

For algorithm engineering, it would be interesting to close the gap between theory and practice with respect of nearest-neighbor data structures. We have reasonable empirical performance of simple data structures like k -D trees but no well-fitting performance guarantees applicable to SWEEPISO. For example, one could look for a more general characterization of inputs where cover trees [9] work provably well.

References

- 1 Pankaj K. Agarwal, Lars Arge, Thomas Mølhave, and Bardia Sadri. I/O-efficient efficient algorithms for computing contours on a terrain. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*. ACM, June 2008. doi:10.1145/1377676.1377698.
- 2 Christopher J. Amante, Matthew Love, Kelly Carignan, Michael G. Sutherland, Michael MacFerrin, and Elliot Lim. Continuously Updated Digital Elevation Models (CUDEMs) to Support Coastal Inundation Modeling. *Remote Sensing*, 15(6), 2023. doi:10.3390/rs15061702.
- 3 Eftymios G. Anagnostou, Leonidas J. Guibas, and Vassilios G. Polimenis. Topological sweeping in three dimensions. In Tetsuo Asano, Toshihide Ibaraki, Hiroshi Imai, and Takao Nishizeki, editors, *Algorithms*, pages 310–317, Berlin, Heidelberg, 1990. Springer Berlin Heidelberg.
- 4 Sunil Arya, David M Mount, Nathan S Netanyahu, Ruth Silverman, and Angela Y Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, 45(6):891–923, 1998.
- 5 M. Axtmann, S. Witt, D. Ferizovic, and P. Sanders. In-Place Parallel Super Scalar Samplesort (IPSSSSo). In *25th Annual European Symposium on Algorithms*, volume 87, pages 9:1–9:14, Dagstuhl, Germany, 2017. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ESA.2017.9.
- 6 M.K. Barker, E. Mazarico, G.A. Neumann, M.T. Zuber, J. Haruyama, and D.E. Smith. A new lunar digital elevation model from the Lunar Orbiter Laser Altimeter and SELENE Terrain Camera. *Icarus*, 273:346–355, 2016. doi:10.1016/j.icarus.2015.07.039.
- 7 J. L. Bentley and T. A. Ottmann. Algorithms for reporting and counting geometric intersections. *IEEE Transactions on Computers*, pages 643–647, 1979.

- 8 Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, September 1975. doi:10.1145/361002.361007.
- 9 Alina Beygelzimer, Sham Kakade, and John Langford. Cover trees for nearest neighbor. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 97–104, New York, NY, USA, 2006. Association for Computing Machinery. doi:10.1145/1143844.1143857.
- 10 COMNAP: Council of Managers of National Antarctic Programs. Antarctic Station Catalogue, 2017. Online; accessed 22-Apr-2023. URL: https://www.comnap.aq/s/COMNAP_Antarctic_Station_Catalogue.pdf.
- 11 Mark de Berg and Constantinos Tsiriogiannis. Exact and approximate computations of watersheds on triangulated terrains. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, November 2011. doi:10.1145/2093973.2093985.
- 12 Jonathan de Ferranti. Digital elevation data, 2011. Online; accessed 05-Juli-2022. URL: <http://viewfinderpanoramas.org/dem3.html>.
- 13 R. Dementiev, L. Kettner, and P. Sanders. STXXL: Standard Template Library for XXL data sets. *Software Practice & Experience*, 38(6):589–637, 2008.
- 14 Alaska Satellite Facility. Radarsat antarctic mapping, 2008. URL: <https://asf.alaska.edu/data-sets/derived-data-sets/radarsat-antarctic-mapping-project-ramp/>.
- 15 R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974. doi:10.1007/bf00288933.
- 16 Peter Fisher and Jo Wood. What is a mountain? or the englishman who went up a boolean geographical concept but realised it was fuzzy. *Geography*, pages 247–256, 1998.
- 17 Peter Fisher, Jo Wood, and Tao Cheng. Where is helvellyn? fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29(1):106–128, 2004.
- 18 S. Fortune. A sweepline algorithm for Voronoi diagrams. *Algorithmica*, 2(1):153–174, 1987.
- 19 Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.
- 20 Daniel Funke, Nicolai Hüning, and Peter Sanders. A Sweep-plane Algorithm for Calculating the Isolation of Mountains. Technical report, Karlsruhe Institute of Technology, May 2023. doi:10.48550/arXiv.2305.08470.
- 21 Bundesamt für Landestopografie swisstopo. *swissALTI 3D Das hoch aufgelöste Terrainmodell der Schweiz*, March 2022.
- 22 L. H. Graff and E. L. Usery. Automated classification of generic terrain features in digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 59(9):1409–1417, 1993.
- 23 Peter Grimm. *Die Gebirgsgruppen der Alpen: Ansichten, Systematiken und Methoden zur Einteilung der Alpen*. Deutscher Alpenverein, 2004.
- 24 K. Gwinner, F. Scholten, F. Preusker, S. Elgner, T. Roatsch, M. Spiegel, R. Schmidt, J. Oberst, R. Jaumann, and C. Heipke. Topography of Mars from global mapping by HRSC high-resolution digital terrain models and orthoimages: Characteristics and performance. *Earth and Planetary Science Letters*, 294(3-4):506–519, 2010.
- 25 Heather Hanson. Shuttle radar topography mission (srtm). Technical report, NASA, 2019. Online; access 21-september-2022. URL: <https://eospsso.gsfc.nasa.gov/missions/shuttle-radar-topography-mission>.
- 26 Adam Helman. *The Finest Peaks-Prominence and Other Mountain Measures*. Trafford Publishing, 2005.
- 27 Johann H Jungclauss, Stephan J Lorenz, Hauke Schmidt, Victor Brovkin, Nils Brüggemann, Fatemeh Chegini, Traute Crüger, Philipp De-Vrese, Veronika Gayler, Marco A Giorgetta, et al. The ICON earth system model version 1.0. *Journal of Advances in Modeling Earth Systems*, 14(4):e2021MS002813, 2022.
- 28 Andrew Kirmse and Jonathan de Ferranti. Calculating the prominence and isolation of every mountain in the world. *Progress in Physical Geography*, 41(6):788–802, 2017.

- 29 National Center for Geospatial Intelligence Standards. World geodetic system 1984: Its definition and relationships with local geodetic systems. Technical Report NGA.STND.0036_1.0.0_WGS84, Department of Defense, 2014.
- 30 PeakVisor. Peakvisor. Online; accessed 22-Apr-2023. URL: <https://peakvisor.com/panorama.html>.
- 31 Mathias Rav, Aaron Lowe, and Pankaj K. Agarwal. Flood risk analysis on terrains. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, November 2017. doi:10.1145/3139958.3139985.
- 32 E Rodriguez, CS Morris, JE Belz, EC Chapin, JM Martin, W Daffer, and S Hensley. An assessment of the SRTM topographic products. Technical Report JPL D-31639, Jet Propulsion Laboratory, 2005.
- 33 Erich Schubert, Arthur Zimek, and Hans-Peter Kriegel. Geodetic distance queries on r-trees for indexing geographic data. In *Advances in Spatial and Temporal Databases*, pages 146–164. Springer, Berlin Heidelberg, 2013. doi:10.1007/978-3-642-40235-7_9.
- 34 Greg Slayden. Peakbagger. Online; accessed 22-Apr-2023. URL: <https://peakbagger.com>.
- 35 Michiel Smid. Closest-point problems in computational geometry. In *Handbook of computational geometry*, pages 877–935. Elsevier, 2000.
- 36 Tetsushi Tachikawa, Masami Hato, Manabu Kaku, and Akira Iwasaki. Characteristics of ASTER GDEM version 2. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 3657–3660, 2011. doi:10.1109/IGARSS.2011.6050017.
- 37 Rocio Nahime Torres, Piero Fraternali, Federico Milani, and Darian Frajberg. A deep learning model for identifying mountain summits in digital elevation model data. In *IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 212–217. IEEE, 2018.
- 38 Rocio Nahime Torres, Federico Milani, and Piero Fraternali. Algorithms for mountain peaks discovery: a comparison. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pages 667–674, 2019.
- 39 Dimitra I. Vassilaki and Athanassios A. Stamos. TanDEM-X DEM: Comparative performance review employing LIDAR data and DSMs. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160:33–50, 2020. doi:10.1016/j.isprsjprs.2019.11.015.
- 40 Wang Yitao, Yang Lei, and Song Xin. Route mining from satellite-AIS data using density-based clustering algorithm. *Journal of Physics: Conference Series*, 1616(1):012017, August 2020. doi:10.1088/1742-6596/1616/1/012017.