Earth System
Science
Data

Open Access   Discussions

Catalogue of coastal-based instances with bathymetric and topographic data

Owein Thuillier[1,2], Nicolas Le Josse[1], Alexandru-Liviu Olteanu[2], Marc Sevaux[2], and Hervé Tanguy[1]

[1]Thales, Defence and Mission Systems, Brest, France
[2]Université Bretagne-Sud, Lab-STICC, UMR CNRS 6285, Lorient, France

Correspondence: Owein Thuillier (owein.thuillier@fr.thalesgroup.com)

Abstract. We provide a catalogue of 17 700 unique coastal-based instances distributed throughout the globe and derived from bathymetric and topographic data made publicly available by the GEneral Bathymetric Chart of the Oceans (GEBCO) as of 2022. These instances, or Digital Elevation Models (DEMs), are delivered in the form of raster grids with a 15 arc-seconds resolution and are divided equally into three libraries "A", "B" and "C". In a given library, the dimensions range from a

5   minimum of 10×10 cells to a maximum of 300×300 with an incremental step of 5, i.e. 59 unique dimensions with 100 instances per dimension. In addition, for each dimension, these instances are ordered by increasing number of maritime cells and have in common the presence of a unique maritime connected component with a ratio of maritime cells lying between 25 % and 95 %, so as to cover a broad spectrum of different coastlines geometries. In this paper, we will describe in detail the procedure used for their automated generation. The resulting catalogue can be downloaded from Zenodo, a general-purpose repository

10 operated by CERN (European Organisation for Nuclear Research) and developed under the European OpenAIRE program, at the following persistent address: https://doi.org/10.5281/zenodo.10530247 (last access: February 15th, 2024) (Thuillier et al. (2024b)). Additionally, a set of 18 colour palettes specifically designed for the visualisation of DEMs has been derived for this occasion and is available at the following address: https://doi.org/10.5281/zenodo.10530296 (last accessed: February 15th, 2024) (Thuillier et al. (2024c)). Both of these repositories come with a comprehensive documentation.

15 # 1   Introduction

Within the current state of the literature, and to the best of our knowledge, there is no catalogue of instances, or Digital Elevation Models (DEMs)[1], of identified coastal geographical areas of varying dimensions and geometries. It is therefore on the basis of this observation that we hereby propose the first catalogue made up of a collection of 17 700 coastal instances covering a wide range of dimensions and coastal layouts. More precisely, the organisation of this catalogue is as follows: **3 libraries** "A",

20 "B" and "C", each containing **59 unique dimensions** ranging from 10×10 to 300×300 with a 5-step increment, and for each of them we have **100 instances** sorted in ascending order of the number of maritime cells, which adds up to a total of **17 700 instances** (i.e. 5 900 per library).

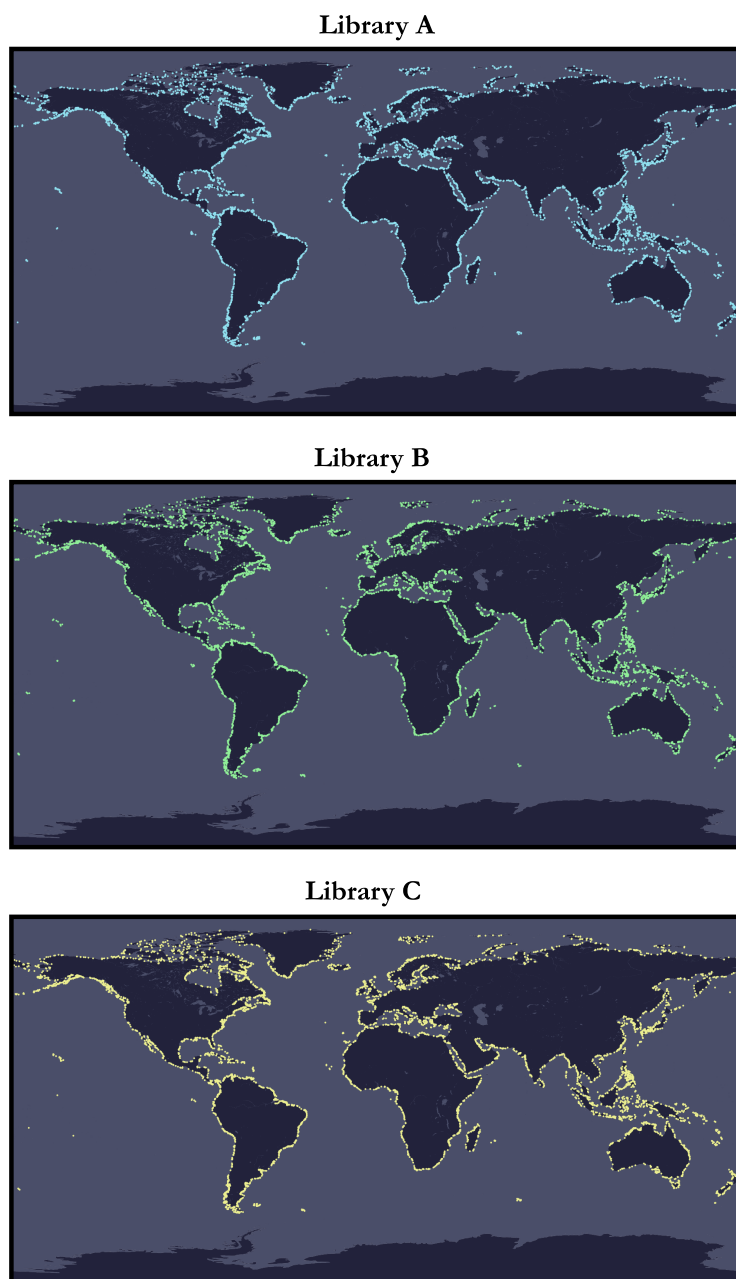[1]For more information on DEMs, the reader is referred to Section B.

These instances were automatically generated from bathymetric and topographic data[2] made available in the public domain
by the GEneral Bathymetric Chart of the Oceans (GEBCO) as of 2022 (GEBCO Bathymetric Compilation Group (June 22,
25 2022)), although an up-to-date 2023 version (GEBCO Bathymetric Compilation Group (April 18, 2023)) is available at the
time of writing[3]. Figure 1 shows an overview of the spatial distribution of the catalogue instances for the three libraries on
plate carrée projections (a special case of an equirectangular projection centred on the equator) with a base layer consisting of
contour lines derived from Natural Earth data (Kelso and Patterson (2010)) (public-domain). Figure 2 shows a synthetic view
of the 100 instances of size 300×300 of the "A" library and Figure 3 shows an individual view of 4 instances among the 100
30 by means of duly annotated satellite imagery.

To provide a brief historical context, this catalogue has its origins in work carried out on the optimal configuration of active
Multistatic Sonar Networks (MSNs) in order to maximise the insonification (coverage) of a geographical Area of Interest (AoI).
An active MSN is made up of a set of sonar systems in a monostatic and/or bistatic configuration, where each sonar system is a
35 pair consisting of a transmitter (source) and a receiver. Monostatism refers to the case where the two sensors are colocated and
bistatism to the case where they are delocated (Cox (1989); Urick (1983)) (potentially operating several kilometres apart and on
two disjoint immersions planes). These sonars can be of several different types: sonobuoys parachuted from an airborne carrier
(e.g. Maritime Patrol Aircraft (MPA), helicopter or drone), dipping sonar from a helicopter, towed array or hull mounted sonar
on a frigate, etc. (Avcioglu et al. (2022)). For further information on the subject of optimal configuration of MSN networks,
40 readers are invited to consult Craparo and Karatas (2018); Craparo et al. (2019); Craparo and Karatas (2020); Fügenschuh
et al. (2020); Thuillier et al. (2024a). In this context, it was therefore usefull to have a variety of AoIs in order to carry out
numerical experiments on the optimisation methods developed to configure such networks. As a result, we decided to generate
a large number of instances distributed around the world over coastal areas and accompanied by elevation data (bathymetric
and topographic), as this is an important aspect of sonar performance. Indeed, until now there has been no benchmark available
45 for this and we think it might be beneficial to have a reference with which to compare in the literature. Furthermore, we believe
that these instances could find a use outside the framework in which they were devised, i.e. with applications not necessarily
related to the configuration of sonar networks.

This article is structured as follows. First, in Section 2, we provide essential background informations about GEBCO through
50 original visualisations specially derived for the occasion, as it constitutes the centrepiece of this catalogue. Then, in Section 3,
we formally define the instances, including their intrinsic features, the file format and the naming convention adopted. Subsequently, in Section 4, we provide in-depth coverage of the generation procedure designed to populate the present catalogue so
as to ensure reproductibility. Finally, we conclude and pinpoint a number of identified prospects for further work in Section 5.

---

[2]Bathymetry refers to the study of underwater relief such as oceans (seabed topography), lakes and rivers, while topography refers to the study of land surface relief.

[3]The data used for the figures in this article will, however, be based on the up-to-date 2023 grid.

Library A

![Library A world map showing distribution of instances]

Library B

![Library B world map showing distribution of instances]

Library C

![Library C world map showing distribution of instances]

Figure 1. Synthetic overview showing the distribution of the various instances across the three catalogue libraries using a plate carrée projection (special case of an equirectangular projection centred on the equator) based on contour lines derived from Natural Earth data (public-domain).
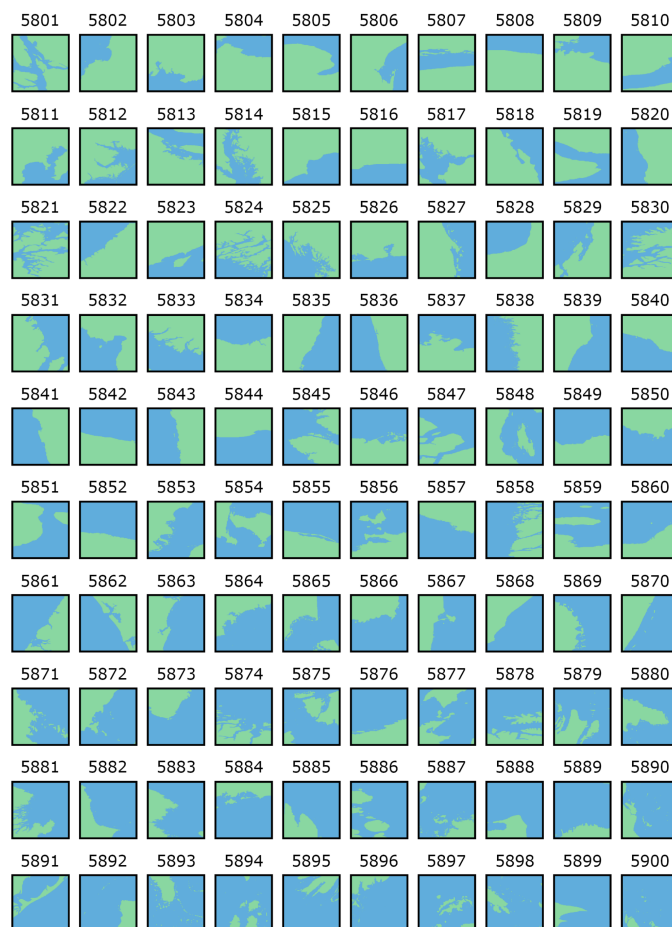
Figure 2. Synthetic thumbnail mosaic displaying all 300×300 instances (DEMs) of library 'A' in a bi-colour format.

2 Background information

55 GEBCO is a project that was instigated at the beginning of the 20th century by Prince Albert 1 of Monaco during a commission meeting when the need for a standardised international nomenclature and terminology for bathymetric contour charts was expressed. As well as working on nomenclature, this same commission, appointed at the 7th International Geographic Congress in Berlin in 1899, was responsible for publishing a general bathymetric chart in paper format, the first edition of which was printed in Paris in 1905 and called the "carte générale bathymétrique des océans" which literally translates to GEneral Bathy-

60 metric Chart of the Oceans, or GEBCO. The aim was straightforward: to provide state-of-the-art information on the shape and depth of the world's seabed. As a result, between 1904 and 1982, there were a total of 5 paper editions:

- 1st edition in 1903,

- 2nd edition from 1910 to 1930,
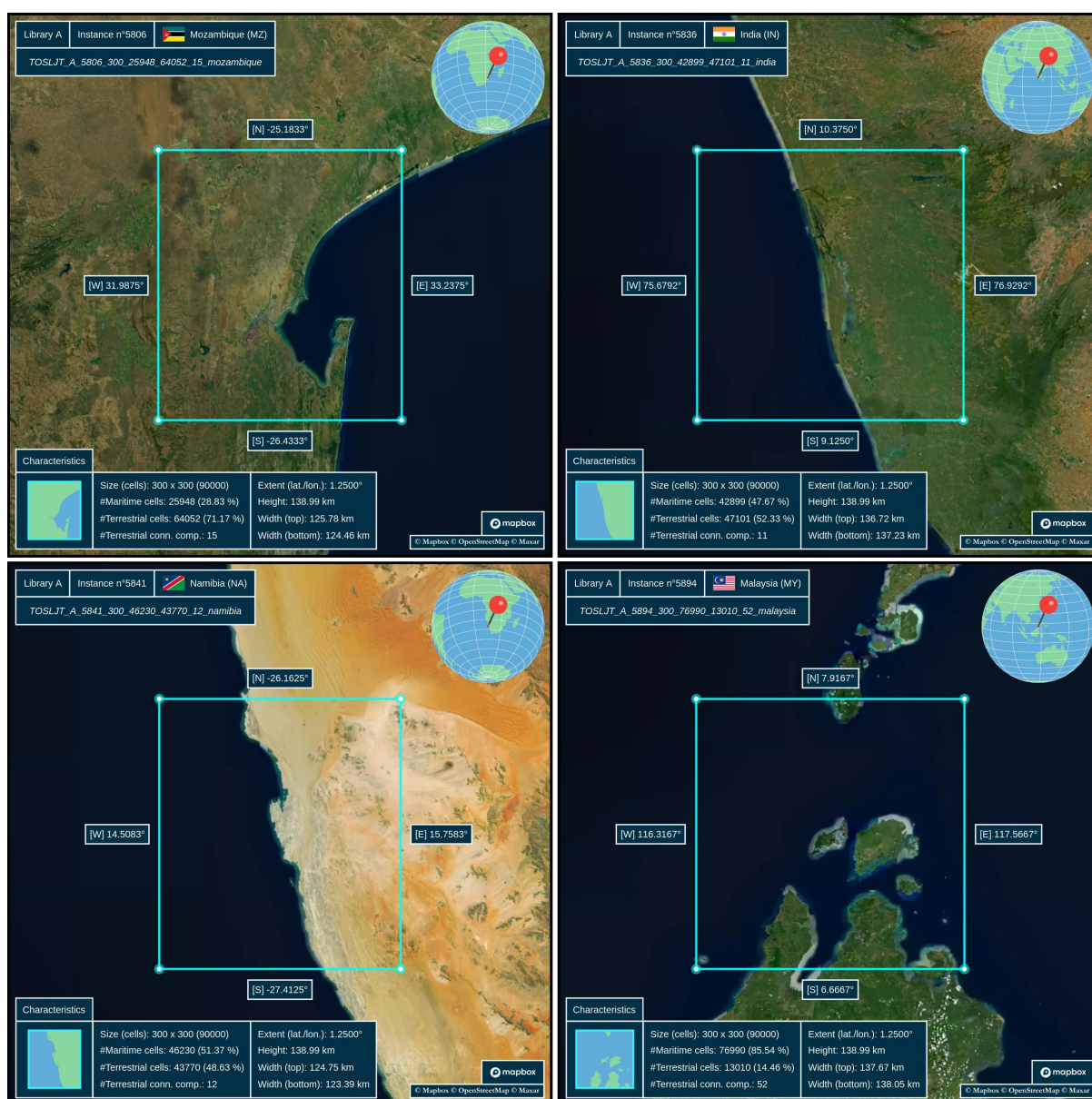
Figure 3. Individual view through annotated satellite imagery of the following 300 × 300 instances of the "A" library: n°5806, 5836, 5841 and 5894.

    – 3rd edition from 1932 to 1966,

65   – 4th edition from 1958 to 1973,

    – 5th and final edition from 1973 to 1982.

Nonetheless, for the third and fourth editions, there was a major reorganisation. Indeed, following the death of Prince Albert 1 and the dismantling of his scientific team, the project was transferred to the International Hydrographic Bureau (IHB), henceforth known as the International Hydrographic Organisation (IHO). Nowadays, GEBCO operates under the joint auspices of
70 the IHO and the Intergovernmental Oceanographic Commission (IOC) of UNESCO, the latter having been invited to take part in the 5th edition.

The first digital version of GEBCO, named GEBCO Digital Atlas (GDA) on CD-ROM, was published in 1994 on the grounds of the 5th edition with a second updated version in 1997[4]. Building on the success of the GDA, a version called GEBCO One
75 Minute Grid was published in 2003, largely based on the data in the GDA and thus marking the real beginning of the online version of GEBCO. In 2009, an updated version (GEBCO_08 Grid) was put online with a resolution of 30 arc-seconds. The datapoints in the latter grid were calculated by combining depths obtained by seabed echography and interpolation between measured points. This interpolation between mesured points was guided by satellite derived gravity data. Updated versions of this grid were put online in 2010 and 2014 respectively. Then, a milestone was reached in 2016 when Mr. Sasakawa, Chairman
80 of The Nippon Foundation[5], decided to partner with GEBCO to work cooperatively to map 100 % of the world's seafloor topography by 2030. The project is called Seabed 2030 and its stated ambition is to produce the definitive, most authoritative and publicly acessible high-resolution bathymetric map of the entire world ocean. Indeed, as of 2023, it is estimated that only 20 % of the oceans have been accurately mapped, the rest being based on interpolation of varying degrees of coarseness. More specifically, the latest GEBCO products are produced by the five centres that make up Seabed 2030, including one global centre
85 and 4 regional centres. The distribution of these centres is detailed hereafter.

    – Southern Ocean, Alfred Wegener Institute (AWI), Germany.

    – South and West Pacific Ocean, National Institute of Water and Atmospheric Research (NIWA), New Zealand.

    – Atlantic and Indian Oceans, Lamont Doherty Earth Observatory (LDEO), Columbia University, USA.

    – Arctic and North Pacific Oceans, Stockholm University (SU), Sweden and the Center for Coastal and Ocean Mapping at
90   the University of New Hampshire (UNH), USA.

    – Global Data Center, British Oceanographic Data Centre (BODC), National Oceanography Centre (NOC), UK.

---

[4]The last version of the GDA in CD-ROM format dates back to 2015, this medium having been abandoned in favour of exclusive online hosting.

[5]Japan's Nippon Foundation is a non-profit philanthropic organisation active throughout the world.

The overall organisation is as follows. The four regional centres compile the various bathymetric data (largely based on multibeam echosounders data) for their respective areas of interest and supply them in grid form to the Global centre, which is responsible for delivering the global grid by splicing the pieces together. Thus, since the Seabed 2030 initiative, there have been the 2019, 2020, 2021, 2022 and 2023 versions with a resolution of 15 arc-seconds for all these grids. From 2019 onwards, the grids have:

- a horizontal extent from 179° 59' 52.5" W (-179.9979167°) to 179° 59' 52.5" E (179.9979167°),

- a vertical extent from 89° 59' 52.5" S (-89.9979167°) to 89° 59' 52.5" N (89.9979167°),

- 43 200 (rows) × 86 400 (columns) = 3 732 480 000 unique pixels (i.e. cells).

All the pixels in the GEBCO grid are referenced according to the horizontal CRS WGS84 (EPSG:4326) and to the vertical CRS MSL height (EPSG:5714). For more information on CRS, please refer to Section A. There is also a reference for the combination of these two CRS (compound CRS), which can be found under the code ESPG:9705 (WGS 84 + MSL height) as discussed in Section A. Besides, as a reminder, the elevation data are "pixel-centre" registered in the GEBCO grid, meaning that it is the elevation at the centre of a cell with a resolution of 15 arc-seconds (approximately 460 metres by 460 metres at equatorial level). This resolution of 15 arc-seconds makes the plate carrée projection the most intuitive for visualisation (see Section A for more details on projections) and which will therefore be used throughout this article. Indeed, the meridians are projected onto regularly-spaced vertical lines and the parallels onto regularly-spaced horizontal lines, giving us a subdivision of space into cells (pixels) of equal size, i.e. a grid. This projection thus facilitates the conversion between geographical coordinates and pixels, which makes it a standard for a large number of global raster datasets including GEBCO. The PCS associated with this projection has the code ESPG:54001 and depends on the GCS WGS84.

Finally, note that the latest grids are based on version 2.5.5 of the SRTM15+ (Shuttle Radar Topography Mission with 15 arc-resolution) dataset[6] Becker et al. (2009); Olson et al. (2014); Tozer et al. (2019) augmented with bathymetric data from the Seabed 2030 regional centres. The SRTM dataset is a fusion of terrestrial topography and an estimate of seabed topography using altimetry (depth predicted from gravity). Version 2.5.5 is similar to version 2.1 (Tozer et al. (2019)) but includes predicted depths based on the V32 (Sandwell et al. (2021)) gravity model.

Figure 4 shows the difference between the areas that have been mapped accurately and the global GEBCO grid as of 2023 with inteprolation on the unmapped areas (around 80 %). Figure 5 focuses on two given zones in order to show the heterogeneity of the input data from which the global GEBCO grid is produced. Finally, Figure 6 shows a 3-dimensional view of the GEBCO 2023 grid in various representations including a detailed close-up on two areas of interest.

---

[6]The "plus" (+) indicates the addition of ocean bathymetry from shipboard soundings and satellite-derived predicted depths.
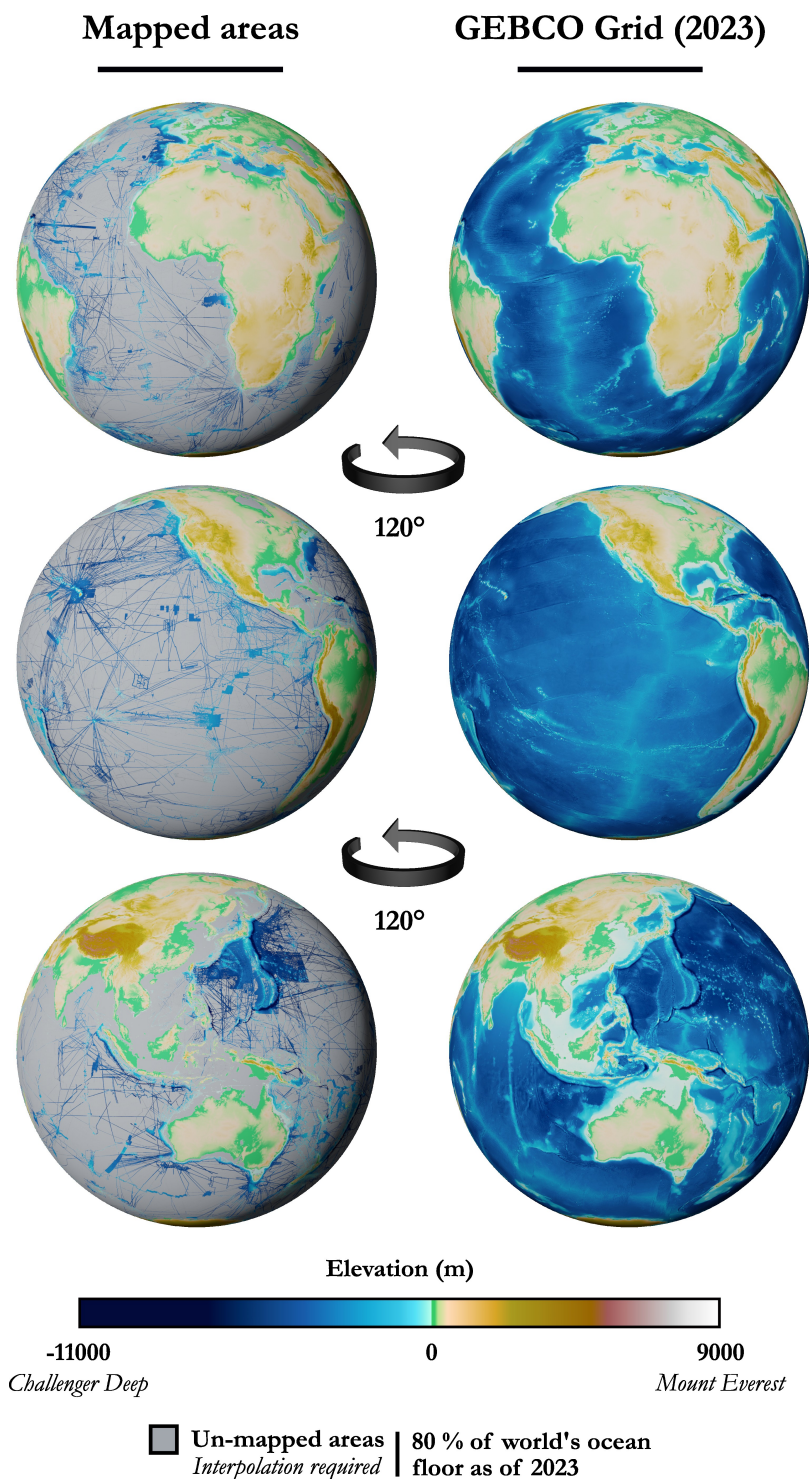
Figure 4. Areas mapped accurately through measured depth values on the left and complete GEBCO 2023 grid on the right for comparative purposes (flat geometry, textures colour-coded for elevation, no shaded relief).
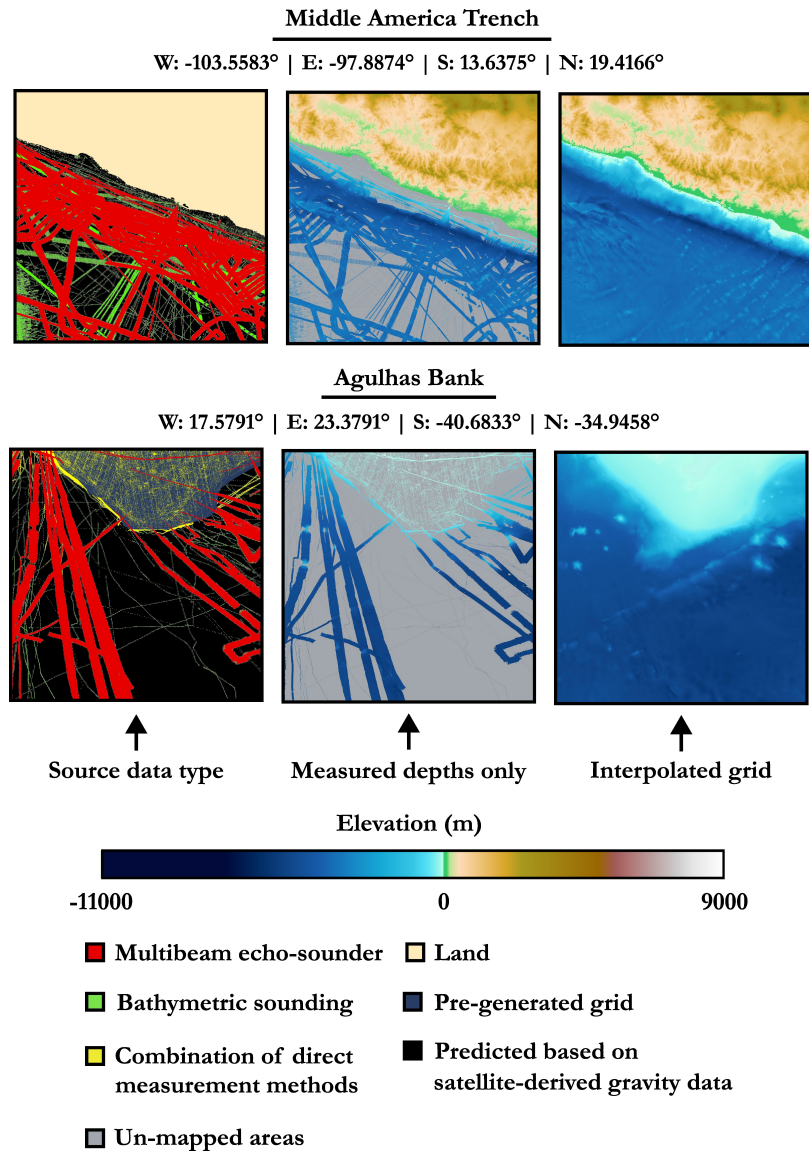
Figure 5. Focus on two Areas of Interest (AoIs) with details of the source data types (not exhaustive list) in order to demonstrate the inherent heterogeneity of the GEBCO global elevation model.
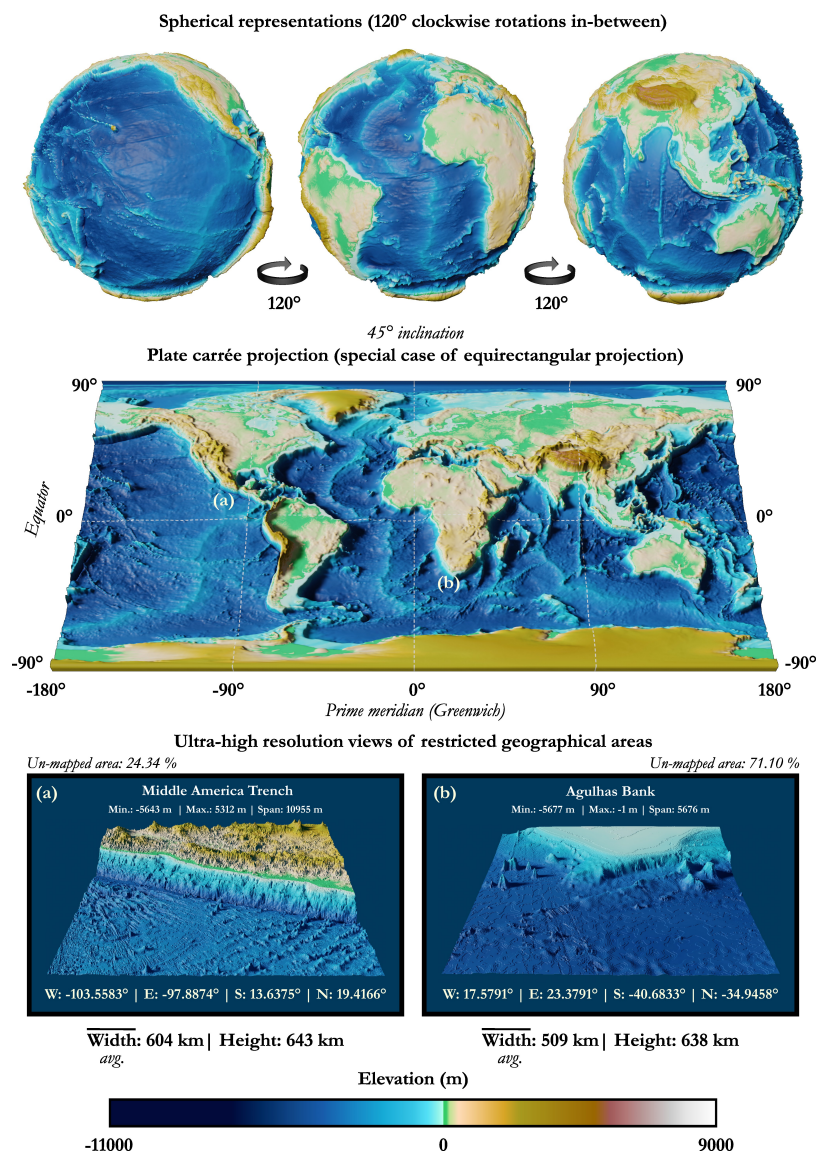
Figure 6. Visualisation of the global Digital Elevation Model (DEM) made available by GEBCO as of 2023 using spherical representations, a planar projection (90° graticule) and ultra-high resolution views of restricted geographical areas (textures colour-coded for elevation and relief exaggeration on the vertical axis for illustrative purposes).

3 Instances

In this third section, we are going to take a closer look at the formal description of the instances, including their intrinsic features, file format, naming convention, overall catalogue organisation and some visualisation guidelines.

125 ### 3.1 Intrinsic features

The instances in the catalogue have a number of common intrinsic features, which we will highlight in the following. First of all, the different instances have a unique maritime connected component and have a ratio of maritime cells ranging from 25 % to 95 %, so as to cover a whole spectrum of different coastline geometric configurations. It should be noted that the decision to have only a single maritime connected component stems from the application for which these instances were originally

130 intended: the configuration of MSNs. Furthermore, there are terrestrial cells that are actually inland waters which have been artificially filled in, because for the intended application there was no reason to consider them. Nevertheless, we were keen not to lose the information corresponding to inland waters that had been artificially filled in, and so we opted for a compromise that would allow us to retain the elevation data associated with them[7]. For example, if an isolated maritime cell had an elevation of -13 metres, then we would replace this elevation with 999913, which would then be decoded as a character string, with

135 the number 9999 encoding that this is inland water (this works because there are no summits above this altitude on Earth). In particular, it is useful to have access to the elevation data of isolated maritime cells if one wishes to carry out re-sampling (up or down).

3.2 File format

The various instances (DEMs) are delivered in the text-based ESRI ASCII grid format (".asc"), which is proving to be a

140 suitable medium for distributing such DEMs because:

- it is human-readable,

- it is not dependent on any particular hardware (plain-text), which makes it easier to transfer across platforms (highly portable),

- it can be used (export/import) in most GIS softwares[8] such as QGIS, gvSIG or SAGA GIS, to name just a few of the
145 open alternatives,

- it is compatible with any horizontal CRS, GCS or PCS, and vertical CRS as lon as this is specified elsewhere and non-ambiguous,

- it is easy to parse through a script if required,

---

[7]Indeed, it may be useful to have access to this data for other applications, and it can also be used for graphical display.

[8]If one wishes to import any of these instances into a GIS software, do not forget to reset the isolated maritime cells if there are any (e.g. "999916" ⟶ "-16").

  – it is reasonably compact when files are compressed.

150 An alternative to this type of text-based format is to use binary formats such as GeoTIFF[9] (Tag Image File Format: ".tiff") or NetCDF[10] (Network Common Data Form: ".nc") as proposed natively by GEBCO. The main advantages of using such formats are an increased speed of read/write (I/O) operations, as there is no need for two-way conversion[11], and a reduced storage footprint[12]. Indeed, for the global coverage provided by GEBCO in 2022, it takes around 8 GB for the binary format against 20 GB for the text format without compression (however the gap narrows to 4 GB and 5 GB respectively after compression).

155 Another advantage of NetCDF and GeoTIFF formats is the possibility of having multiple bands, or strips (i.e. dimensions), whereas this is impossible with the ESRI ASCII format: one file per band is required. However, the added efficiency of binary files comes with the major drawback that it is not directly readable and therefore not very user-friendly when it comes to obtaining an overview of the data, plus it can be difficult to fix corrupted files. In the end, since performance is not paramount, storage a lesser problem and that we only need a single band for the instances herein under consideration (i.e. elevation), we

160 have therefore opted for this text-based format for the above-mentioned advantages.

A file in the ESRI ASCII grid format consists of two consecutive parts:

  – A **header** containing a certain amount of metadata about the grid, such as the geographical extent[13] and the grid resolution (i.e. the cell size).

165 – The **data part**, where an integer or floating-point numerical value is associated to each cell of regular angular extent (cf. "cellsize" field), which here corresponds to the elevation taken at the centre-point of the latter (see section "Introduction" for further details).

More formally, this takes the following form:

---

[9]GeoTIFF is used to associate geo-referencing data (e.g. projection and geographic extent) with TIFF images.

[10]The NetCDF files supplied by GEBCO are in version 4 and follow the Climate and Forecast (CF) Metadata convention in version 1.6 (http://cfconventions.org/).

[11]For example, converting a 32-bit integer into a string of characters during the write phase and then do the reverse conversion during the read phase is a time-consuming process.
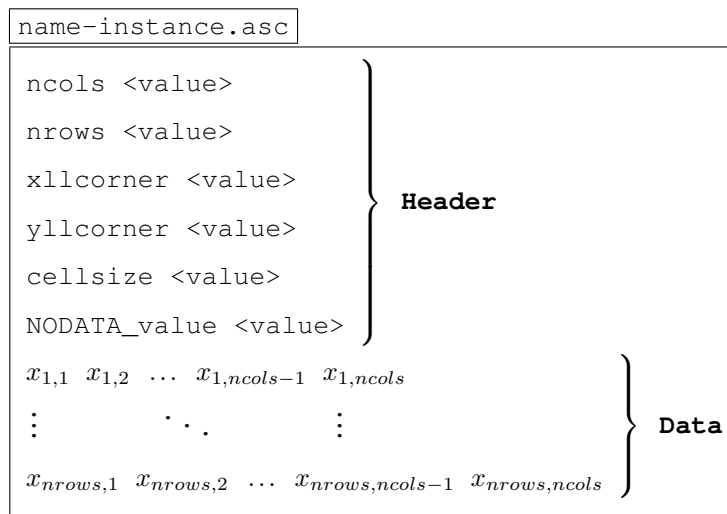
[12]For example, for an integer encoded in 32-bits (4 bytes) such as 1234567890, this would require exactly 4 bytes to be written to a binary file, compared with 10 bytes for a text file (1 byte per character). The cost is even higher when you take into account any spaces and delimiters used to separate data within the file.

[13]There are two main ways of giving a geographical extent: using a Bounding BOX (BBOX) defining West, East, South and North, or using an origin coupled with a horizontal and vertical extents. The latter is the case for files in ESRI ASCII format.

```
┌─────────────────────────┐
│ name-instance.asc       │
└─────────────────────────┘
```

ncols <value>

nrows <value>

xllcorner <value>

yllcorner <value>   **Header**

cellsize <value>

NODATA_value <value>

$x_{1,1}$ $x_{1,2}$ $\cdots$ $x_{1,ncols-1}$ $x_{1,ncols}$

$\vdots$ $\ddots$ $\vdots$ **Data**

$x_{nrows,1}$ $x_{nrows,2}$ $\cdots$ $x_{nrows,ncols-1}$ $x_{nrows,ncols}$

where:

- ncols (*integer*) and nrows (*integer*) correspond respectively to the **number** of **columns** and **rows** in the grid.

- xllcorner (*float*) and yllcorner (*float*) are respectively the **x-coordinate** (i.e. longitude) and the **y-coordinate** (i.e. latitude) of the **grid origin located** at the **lower left corner** of the lower **left cell**, expressed in the horizontal CRS WGS84 (EPSG:4326[14]) with the Decimal Degrees (DD) notation.

- cellsize (*float*) is the **length** of **one side** of a **square cell** (i.e. both **height** and **width**) in the same reference unit as the origin (DD) and which corresponds to approximately 0.00417° in the case of the GEBCO grid (i.e. 15 arc-seconds).

- NODATA_value (*float* or *integer*) is the **default value** assigned when an **input** is **missing** or **unknown**.

- $x_{i,j}$ (*float* or *integer*) is the **elevation value** assigned to the **cell** (i,j) at the intersection of row $i \in [\![1, nrows]\!]$ and column $j \in [\![1, ncols]\!]$, given in the vertical CRS MSL height (EPSG:5714[15]), i.e. in metres referenced to mean sea level (positive upwards).

Note that in our case, we will only have integer values for elevations ($x_{i,j}$) and the NODATA_value, as it is the case for the data retrieved from the GEBCO global grid. A visual representation of the different elements defining a grid in ESRI ASCII format is available in Figure 7. In addition, a number of general remarks about this particular format are listed hereafter if anyone would like to modify or propose new instances:

- The values assigned to the various cells are given in row-major order (i.e. left-to-right and top-to-bottom) and must be separated by a single space character. Moreover, in theory, no carriage return is required at the end of each row of the grid, because the number of columns is known, but in practice this makes it more legible. In either case, the number of

---

[14]https://epsg.io/4326
[15]https://epsg.io/5714
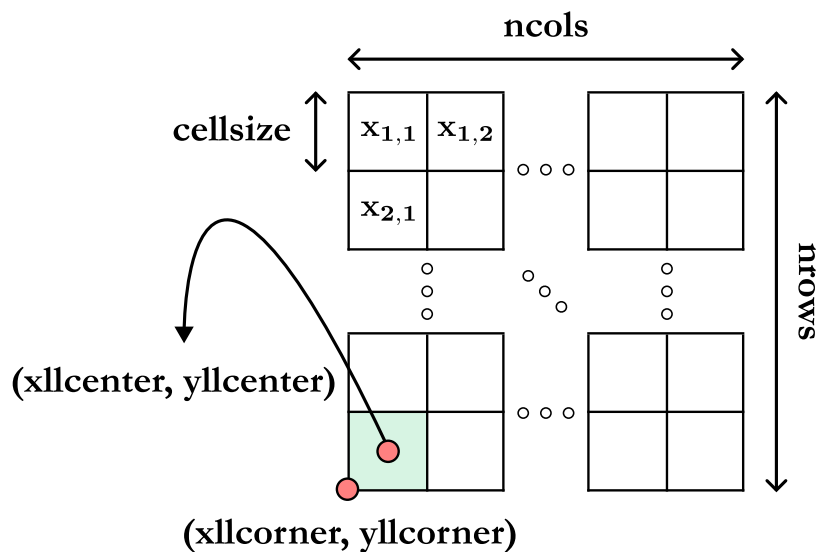
Figure 7. Schematic view of the structure of an ESRI ASCII grid.

values must be equal to the number of rows (`nrows`) times the number of columns (`ncols`), i.e. one single value per
190   cell.

– It is possible to define the origin of the grid alternatively by using the keywords `xllcenter` (*float*) and `yllcenter`
(*float*), which then locates it in the centre of the lower left cell.

– The `NODATA_value` keyword is optional but strongly recommended.

– The various keywords in the header are not case-sensitive, although it is recommended that they be consistent for ease
195   of use.

For illustrative purposes, a didactic example of a 20×20 grid showing the structure of all the catalogue instances in this
particular format is shown hereafter:

```
didactic-instance.asc

ncols 15
nrows 15
xllcorner 14.320833333333
yllcorner 38.512500000000
cellsize 0.004166666667
NODATA_value -32767
-959 -878 -811 -753 -719 -684 -666 -665 -680 -712 -760 -806 -848 -897 -952
-877 -776 -684 -591 -537 -510 -488 -481 -499 -539 -599 -659 -716 -782 -854
-783 -674 -561 -441 -332 -285 -273 -265 -284 -348 -439 -522 -593 -668 -752
-656 -536 -413 -276 -167 -104 -87 -79 -92 -155 -253 -362 -455 -553 -660
-529 -379 -236 -110 -53 -41 -36 -31 -31 -46 -96 -178 -317 -457 -593
-418 -244 -100 -41 -20 -11 89 176 216 210 150 -82 -233 -401 -547
-364 -183 -63 -24 -8 110 349 444 411 258 110 -95 -211 -378 -527
-362 -175 -54 -21 -5 185 445 512 389 201 30 -114 -252 -403 -538
-399 -204 -70 -24 -6 151 337 390 284 85 -70 -139 -307 -447 -576
-461 -278 -128 -41 -17 113 130 136 64 -53 -92 -191 -350 -495 -632
-520 -388 -250 -114 -62 -74 -62 -57 -64 -86 -164 -298 -436 -566 -683
-621 -508 -393 -273 -187 -169 -159 -159 -165 -197 -295 -414 -532 -644 -748
-733 -633 -542 -466 -416 -404 -392 -394 -393 -397 -456 -542 -635 -738 -833
-826 -750 -694 -637 -593 -574 -564 -567 -568 -578 -616 -674 -745 -837 -923
-912 -866 -830 -783 -737 -710 -707 -717 -728 -743 -778 -820 -871 -938 -1010
```

More specifically, this is the volcanic island of Alicudi, part of the Aeolian archipelago in the Tyrrhenian Sea and lying to the north of Sicily, in Italy. Although a bi-colour thumbnail gives a rough idea of the geographical configuration of the instance herein under consideration, Figure 8 presents a much more detailed visualisation through the compilation of numerous data from various sources. This figure is structured as follows.

Firstly, the texture in the upper plane corresponds to a satellite view taken from a compilation of data from the Sentinel-2 mission of the Copernicus program operated by the ESA[16] in the three bands of the visible spectrum (B2, B3 and B4 respectively) over a sliding year in order to obtain an image with as few clouds as possible. This texture was then post-processed using artificial intelligence in order to remove certain artefacts such as boat tracks or dwellings, the idea being to obtain the purest texture possible, representative of the island's topography and free from anthropogenic activity. A DEM is then added to this base texture, using topographic data from NASA's SRTMGL1[17] version 3 dataset, with a resolution of around 30 m × 30 m. Note that the horizontal CRS is WGS84 (EPSG:4326), like the GEBCO grid, but that the vertical CRS here is EGM96[18] height (EPSG:5773) and so there may be a difference of several metres with the GEBCO elevations, which are referenced to

---

[16]European Space Agency
[17]Shuttle Radar Topography Mission (SRTM) GLobal 1 arc-second (GL1).
[18]Earth Gravitational Model from 1996.

the vertical CRS MSL height (EPSG:5714). At equivalent resolution, however, this error would be negligible, as these are two gravity-related models linked to the geoid, but here the GEBCO data (15 arc-seconds) has a much lower resolution than the

215 SRTMGL1 data (1 arc-second) and this can lead to large differences (negligible, nevertheless, for our use). In addition, these elevation data have been interpolated afterwards to obtain a rendering as close as possible to reality, avoiding the asperities associated with the data resolution. It should also be noted that the proportions have not been preserved and that the z axis (vertical) has been distorted in order to achieve a better rendering (the same applies to the bottommost plane). Also, for the three parts of the figure, these representations are based on an equirectangular (or equidistant cylindrical) projection of the globe, i.e. forming

220 a Cartesian grid made up of identical squares (pixels), a projection which is used for global raster datasets such as GEBCO or SRTMGL1 because it facilitates conversions between geographical coordinates and pixels (as discussed earlier in Section 2).

In the middle section, we have the grid with the bathymetric and topographic data supplied by GEBCO. As a reminder, the GEBCO grid is pixel-centre registered, so the elevations (in metres) correspond to the points at the centre of each of the cells

225 (pixels) with a resolution of 15 arc-seconds, i.e. approximately 460 m × 460 m at equatorial level. A vertical column showing the elevations in plain-text at GEBCO resolution is also provided to show the transition between the top and middle planes for an area of 3 × 3 pixels.

On the lower plane, we have this same grid, but this time with a 3-dimensional representation, with the elevation represented

230 on the z-axis (proportions not respected, same distortion as for the upper plane). As discussed above, the resolution is lower than with the data from SRTMGL1 and this can now be seen (interpolation aside), but this time we have the bathymetric data as well, which is the key feature of the GEBCO grid.

In addition, a picture of the island with an East-West approach (i.e. a heading of 270°) is also available at the bottom left of

235 the figure. Finally, a map on the bottom right-hand side gives a rough location for the island of Alicudi, whose GPS coordinates in DMS format are also shown on the right-hand side of the base in the horizontal CRS WGS84 (EPSG:4326). This map was produced using Blue Marble Next-Gen images, in this case the December texture with topography and bathymetry, made available by NASA Visible Earth.

3.3 Naming convention

240 This Section describes the naming convention used to designate all the different instances of the catalogue spread across the three libraries "A", "B" and "C". First of all, an instance is formally referred to under the following full format:

$$\texttt{"TOSLJT\_library\_id\_n\_m\_t\_nb-tcc\_country"},$$

where, respectively, we have:

– `TOSLJT` which refers to the **initials** of the **authors** (Thuillier, Olteanu, Sevaux, Le Josse and Tanguy).
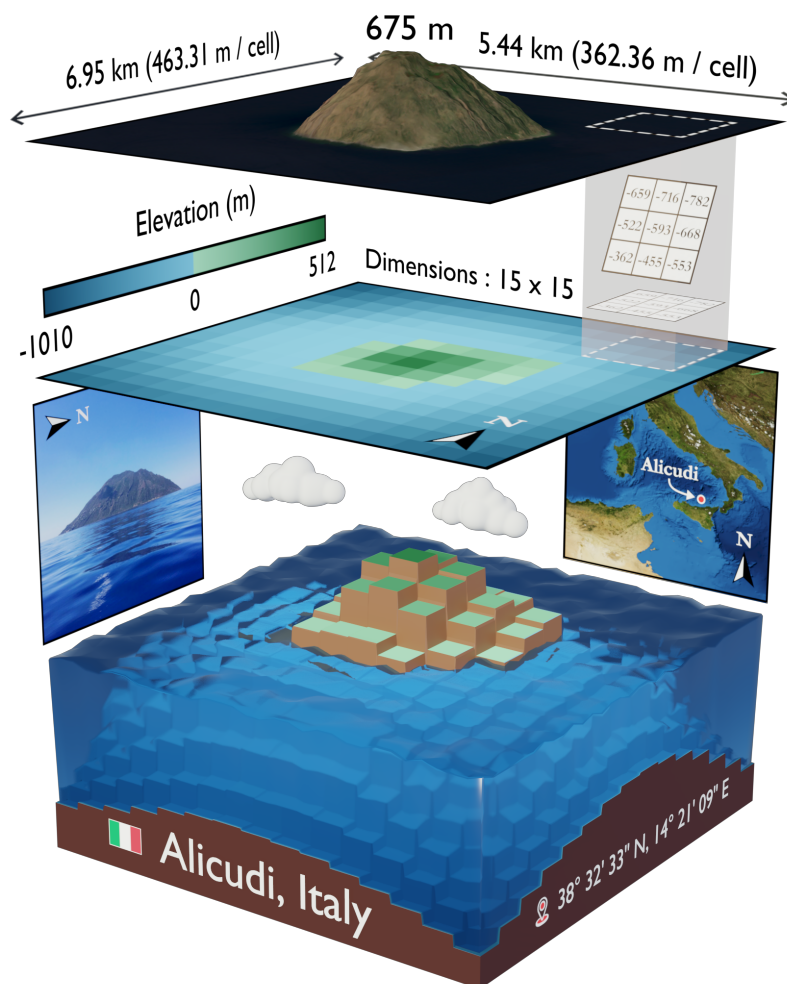
Figure 8. Detailed visualisation of the didactic instance.

245   – library (*string*) corresponding to the **parent library**.

     – id (*string*) which is the **unique identifier** of an instance within the parent library. This unique identifier is coded on a string of 4 characters (examples: 0001, 0015, 0108, 1204 for instances no. 1, no. 15, no. 108 and no. 1204 respectively). Furthermore, for a given dimension, the 100 instances are ranked in ascending order of the number of maritime cells, so the instance with the smallest identifier corresponds to the instance with the fewest maritime cells. That said, it is

250   possible for a lower-dimensional instance to have more maritime cells.

     – n (*integer*) corresponding to the **dimensions** of the instance (example: $n = 10$ means that we are dealing with an instance of dimensions 10×10). Note that we only need one integer, as we are dealing exclusively with square grids.

- `m` (*integer*) and `t` (*integer*) corresponding respectively to the **number** of **maritime cells** and **terrestrial cells**.

- `nb-tcc` (*integer*) the **number** of **terrestrial connected components**.

255   – `country` (*string*) the **country** of **affiliation**.

For example, the instance with the name

$$\texttt{"TOSLJT\_A\_0108\_15\_69\_156\_1\_australia"}$$

reveals that it is instance no. 108 from the A library, which is a 15×15 grid with 69 maritime cells, 156 terrestrial cells, 1 terrestrial connected component and located on a coastline of Australia.

260

Finally, a particular instance of the catalogue can be uniquely and unambiguously identified simply by the following short format:

$$\texttt{"TOSLJT\_library\_id"},$$

which can be useful for referring to a set of instances during numerical experiments. For the above example, this would 265 therefore be `"TOSLJT_A_0108"`.

4   Generation procedure

In this fourth section, we will take an in-depth look at the procedure for generating the various instances, which will be subdivided into key sub-steps.

4.1   Seed selection

270 The first step consists of randomly drawing a geographical point (a seed) from one of the world's coastlines, i.e. on a natural physical boundary on the edge of an ocean or sea, either on a continental landmass or on an ocean island. Indeed, we filter out lakes, rivers, islands on lakes and ponds on islands within lakes, because these cases actually exist and were artefacts: there is no (or partial) bathymetric data available for these places[19] in the global grid provided by GEBCO. To achieve this, we have retrieved a dataset listing all the coastlines on the globe in the form of polylines made up of a set of points with geographical 275 coordinates in the horizontal CRS WGS84 (EPSG:4326), and by arranging them in such a way as to retain only the coastlines of interest to us (see Figures 9 and 10 for an illustration of the different levels of coastline mentioned, case n°1 is the one adopted here). This dataset is the "Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG)" made available by Wessel and Smith (1996) and which is based on three public-domain datasets: World Vector Shorelines (WVS)[20] (Soluri and Woodson (1990)), CIA World Data Bank II (WDBII)[21] (Gorny (1977)) and Atlas of the Cryosphere

---

[19] There are exceptions, however, such as the Caspian Sea, which is actually a lake and has therefore been excluded here (arbitrary choice).

[20] Basis for land masses and some ocean islands.

[21] Basis for lakes, islands on lakes, ponds on islands within lakes, rivers, political borders (not shown here) and some ocean islands.

280   (AC)[22] (Bohlander and Scambos (2007)). It should also be noted that we have excluded the Antarctic from this catalogue, as
it is treated differently in the GEBCO dataset: there is both a part corresponding to the elevations above and below ice. The
difference between grounding line and ice front boundary for the coastlines shown in Figure 10 is illustrated in Figure 11.

We take this opportunity to mention a pitfall we encountered when producing this catalogue: the spatial distribution of in-
285   stances was originally extremely unbalanced. The reason is illustrated in Figure 12 and is as follows. The resolution of the
coastlines was far too excessive and there are places in the world made up of tens of thousands of small islands (e.g. the
Archipelago Sea[23] with an estimated 50 000 islets and the Canadian Arctic Archipelago with nearly 40 000 islets), which led
to a genuine imbalance in the spatial distribution of instances. Indeed, there was a greater probability of drawing a seed from
one of these regions. One of the solutions envisaged to overcome this problem was to (*step 1*) filter out islands below a certain
290   size and (*step 2*) weight the probability of drawing a seed from a landmass or an island according to its size and therefore de-
pending on the number of segments (by extension, points) used for discretisation. This solution has been preferred, providing
a fairer spatial distribution of instances throughout the globe. Another solution to this issue lays in the use of a geometric data
compression algorithm such as Douglas-Peucker (Douglas and Peucker (1973)), sometimes called Ramer-Douglas-Peucker,
which simplifies a polygon or broken line (polyline) by removing some of the points. This would have had the effect of elim-
295   inating the smallest islands and thus reducing the number of points from which to choose a seed (in particular by eliminating
geographically neighbouring points). However, such an algorithm tends to deteriorate the quality of the coastline contouring,
as illustrated in Figure 13, and this solution was not chosen for this reason.

From now on, we have a seed with coordinates $s = (lat_s, long_s) \in [-90, 90] \times [-180, 180]$ representing the centre of the
300   current grid of dimension $n \in \mathbb{N}$ (square here) and whose bounding box is defined by the bounds:

   – $north = lat_s + (\frac{n}{2} \cdot cellsize)$,

   – $south = lat_s - (\frac{n}{2} \cdot cellsize)$,

   – $west = long_s - (\frac{n}{2} \cdot cellsize)$,

   – $east = long_s + (\frac{n}{2} \cdot cellsize)$.

305   As a reminder, cellsize corresponds to the size of a cell, i.e. 15 arc-seconds for the GEBCO global grid, or approximately
$0.00417°$. Figure 14 shows a current grid determined by the drawing of a seed, with the GEBCO grid graticule in the back-
ground (in grey). Note that at this stage, the two do not necessarily coincide, and the probability of this happening is actually
very low.

---

[22]Basis for Antarctica (both grounding line and ice front boundaries).
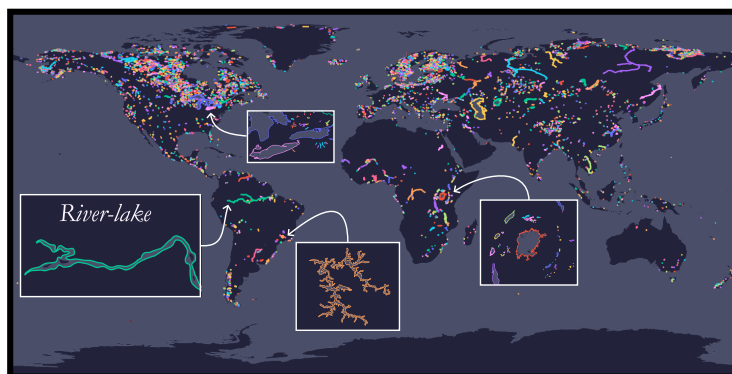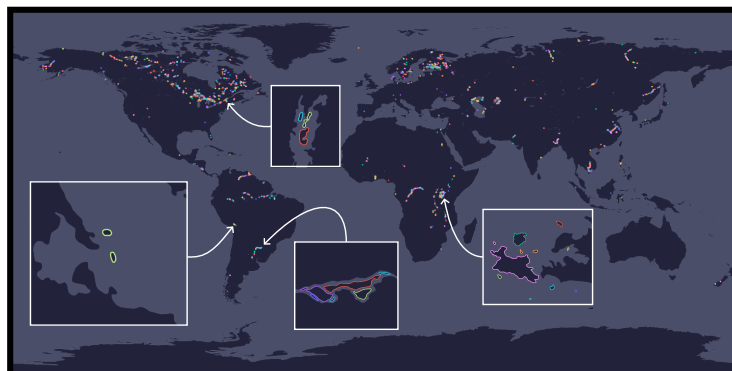
[23]In Finnish Saaristomeri and in Swedish Skärgårdshavet.

Continental land masses and ocean islands (no Antarctica)
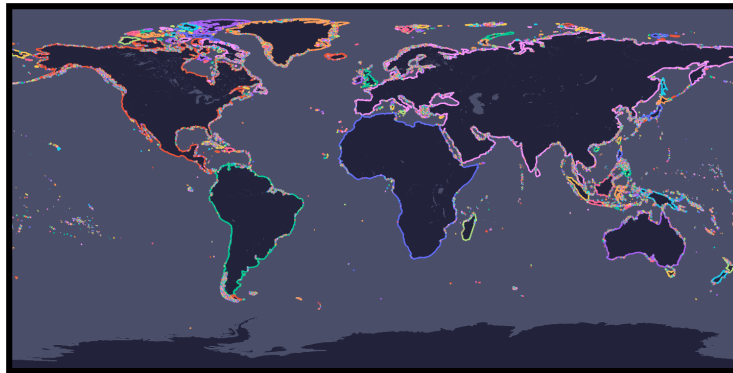
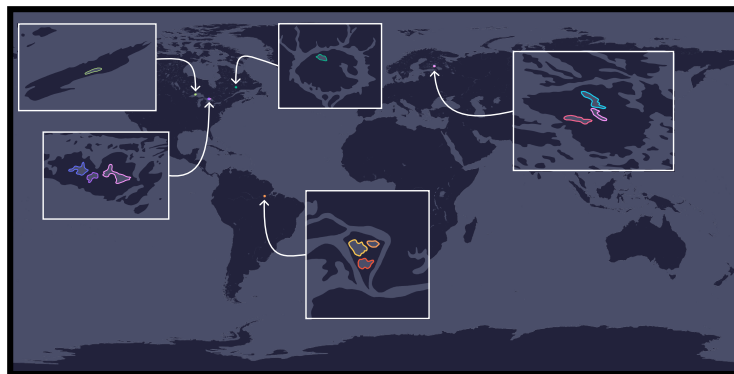**Lakes and double-lined rivers (river-lakes)**

Islands in lakes

Figure 9. Illustration of the different levels of coastlines using a plate carrée projection (special case of an equirectangular projection centred on the equator), part 1/2.
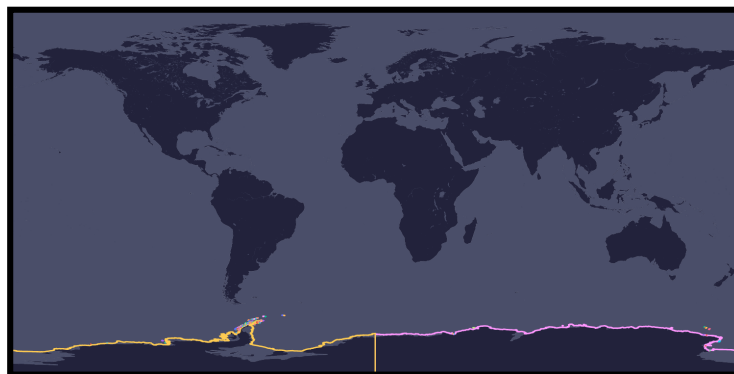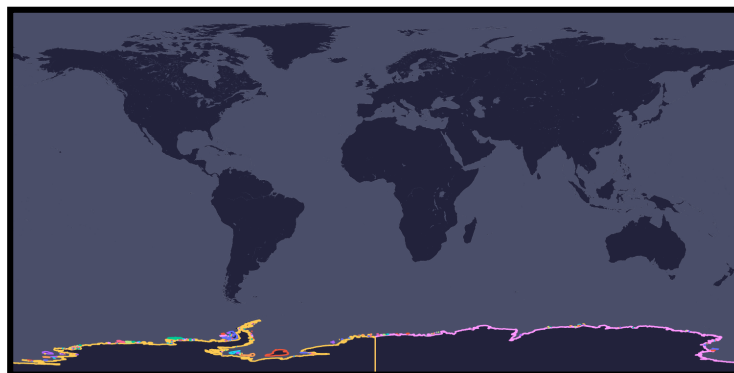
Ponds in islands within lakes
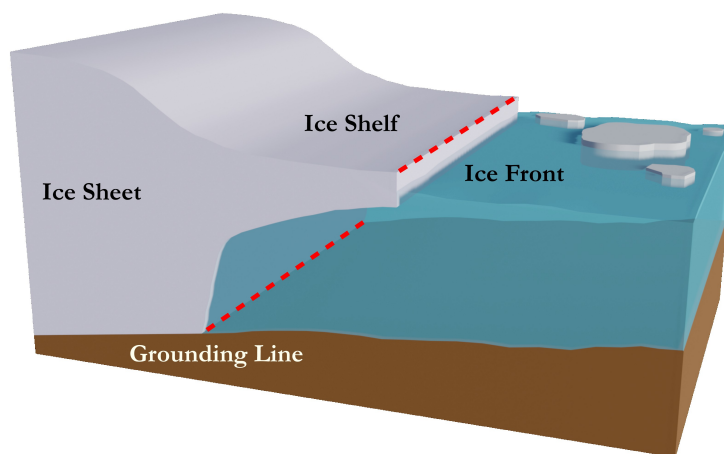

Antarctica based on ice front boundary

Antarctica based on grounding line boundary

Figure 10. Illustration of the different levels of coastlines using a plate carrée projection (special case of an equirectangular projection centred on the equator), part 2/2.

Figure 11. Illustration of the difference between ice front boundary and grounding line boundary for the Antarctic.

4.2 Offset

310 Once we have a seed located on a coastline (continental landmass or ocean island), we are going to offset it in such a way as to introduce more or less maritime cells and a little more variability[24]. To do this, we are going to offset the seed by randomly drawing:

- a vertical displacement $\lambda^{\updownarrow} \in [-\frac{n}{\delta} \cdot cellsize, \frac{n}{\delta} \cdot cellsize]$,

- a horizontal displacement $\lambda^{\leftrightarrow} \in [-\frac{n}{\delta} \cdot cellsize, \frac{n}{\delta} \cdot cellsize]$,

315 where $\delta \in \mathbb{R}$ is a parameter. For this catalogue, we choose $\delta = 3$ so that there is a minimum overlap between the old and new zones. The seed is therefore moved to the point:

$$s' = \left(lat_s + \lambda^{\updownarrow}, lon_s + \lambda^{\leftrightarrow}\right). \tag{1}$$

The principle of offset is illustrated in Figure 15. In addition, it was also possible to envisage a displacement in a circular region around the seed and not in a square region as has been done here. The difference between these two possibilities is 320 illustrated in Figure 16.

4.3 Recalibration on the GEBCO grid

Now that we have offset the seed, all that remains is to align the current grid with the GEBCO grid graticule (if this is not the case, otherwise we move on to the next step). To do this, we need to distinguish between two cases depending on whether n is even or odd.

---

[24]This reinforces the choice made in the previous section to retain a high resolution for the coastlines, as it makes the offset relevant
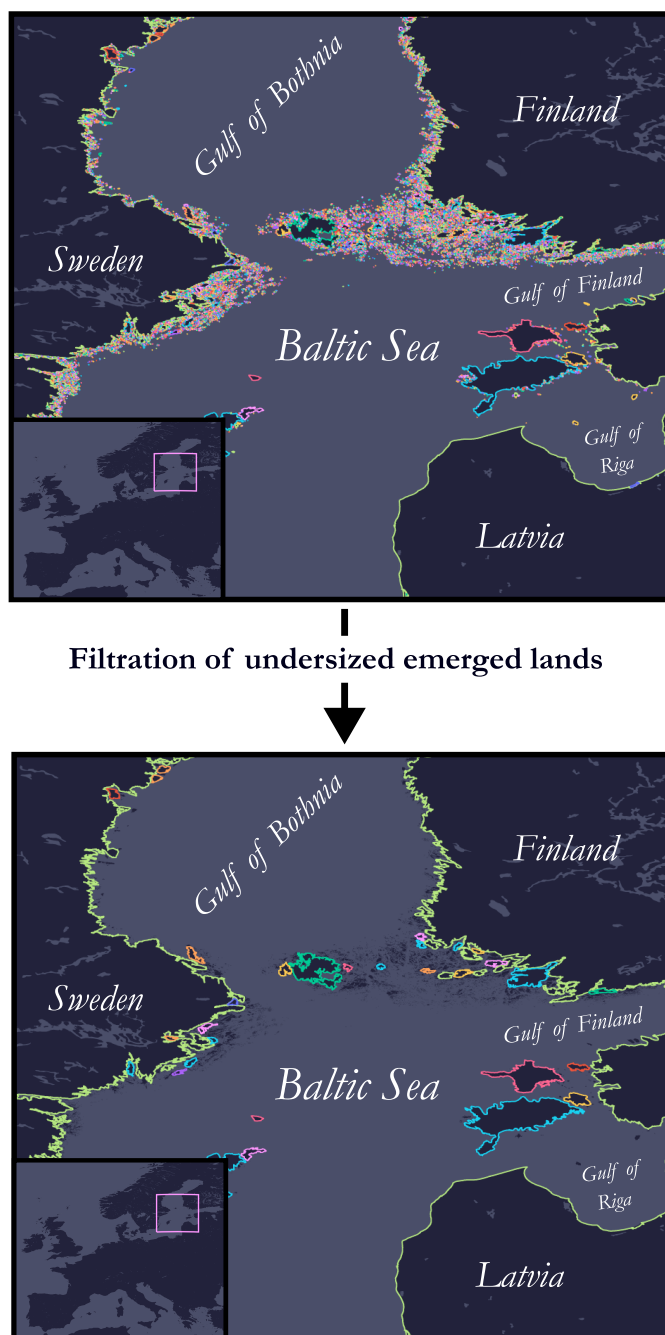
Figure 12. Illustration of the pitfall encountered and the filtration procedure put in place for undersized emerged lands (focus on part of the Baltic Sea subject to numerous islets).

Figure 13. Illustration of the Douglas-Peucker geometric data compression algorithm for coastline simplification at three levels of resolution.
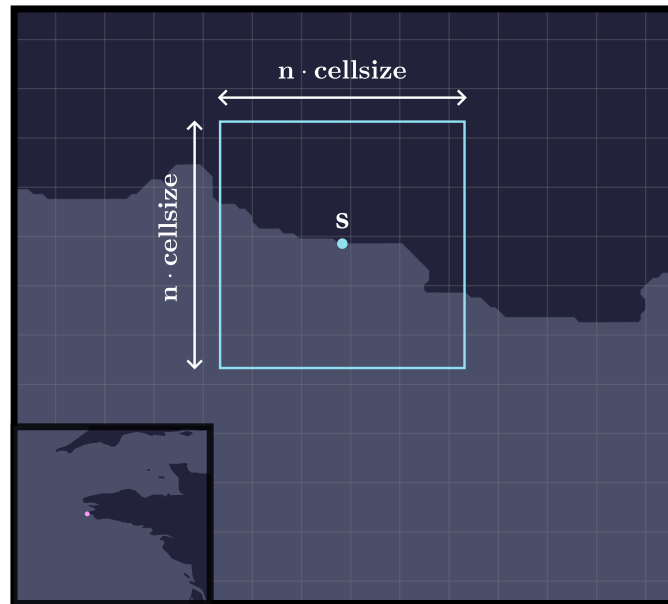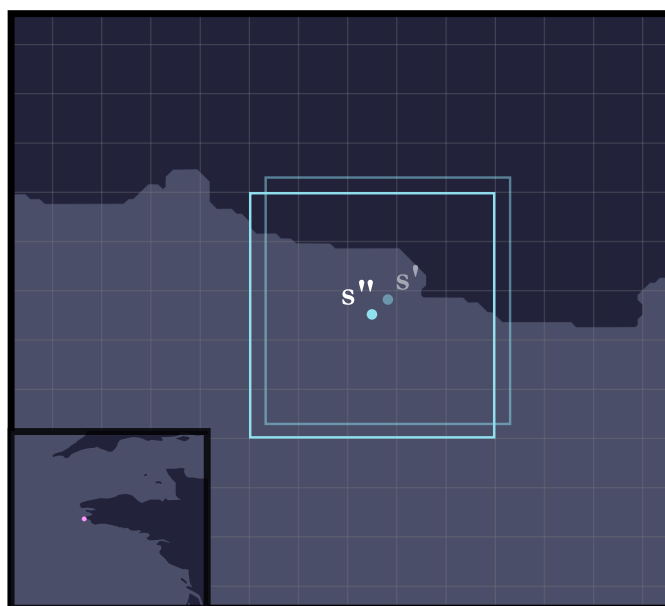
Figure 14. Illustration of a current grid of dimension n determined by the drawing of a seed on a coastline (GEBCO graticule in grey, i.e. 15 arc-seconds).
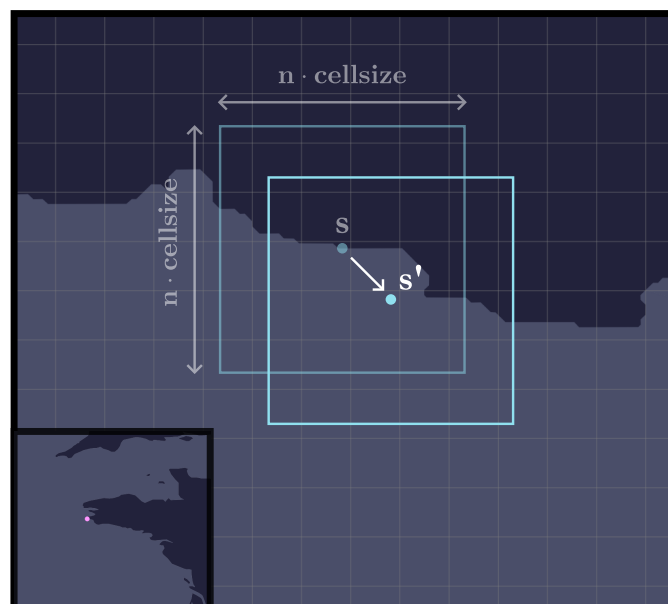
Figure 15. Illustration of the seed offset mechanism (GEBCO graticule in grey, i.e. 15 arc-seconds).
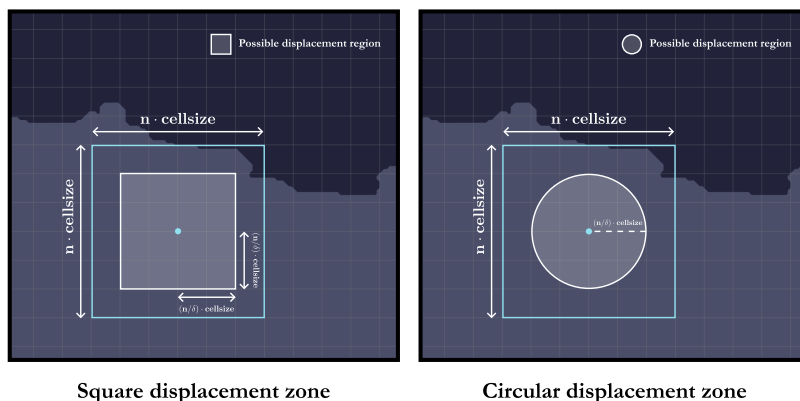
Figure 16. Two possibilities for the offset: circular or square displacement (GEBCO graticule in grey, i.e. 15 arc-seconds).

Figure 17. Illustration of the procedure for recalibrating the current grid on the GEBCO graticule: odd case (GEBCO graticule in grey, i.e. 15 arc-seconds).

4.3.1 Odd case

In the odd case illustrated in Figure 17, the seed is moved to the centre of the cell in which it is located, i.e. at the coordinates:

$$s" = \left(\left\lfloor \frac{lat_{s'}}{cellsize} \right\rfloor cellsize + \frac{cellsize}{2}, \left\lfloor \frac{lon_{s'}}{cellsize} \right\rfloor cellsize + \frac{cellsize}{2} \right). \tag{2}$$

4.3.2 Even case

In the even case illustrated in Figure 18, there is a choice to be made: the seed (i.e. the centre of the current grid) can be moved
330 to one of the 4 corners of the cell in which it is located: south-west, north-west, north-east or south-east. Here, we have chosen
to shift the seed to the corner that minimises the displacement distance (i.e. to the corner closest to the seed after offset). The
seed is therefore moved to the point

$$s" = arg\,min_{x \in X}(d(s', x)), \tag{3}$$

where $X = \{south\_west, north\_west, north\_east, south\_east\}$, d is the great-circle (orthodromic) distance between two
335 geographical points and:

- $south\_west = \left(\left\lfloor \frac{lat_{s'}}{cellsize} \right\rfloor cellsize, \left\lfloor \frac{lon_{s'}}{cellsize} \right\rfloor cellsize \right)$,

- $north\_west = \left(\left\lceil \frac{lat_{s'}}{cellsize} \right\rceil cellsize, \left\lfloor \frac{lon_{s'}}{cellsize} \right\rfloor cellsize \right)$,

- $north\_east = \left(\left\lceil \frac{lat_{s'}}{cellsize} \right\rceil cellsize, \left\lceil \frac{lon_{s'}}{cellsize} \right\rceil cellsize \right)$,

- $south\_east = \left(\left\lfloor \frac{lat_{s'}}{cellsize} \right\rfloor cellsize, \left\lceil \frac{lon_{s'}}{cellsize} \right\rceil cellsize \right)$.

340 4.4 Retrieving elevations

At this stage, all we need to do is go through all the cells (pixels) in the current grid and retrieve their elevations using the global
GEBCO grid. This involves mapping a pixel (i, j) in the current grid to a pixel (i', j') in the GEBCO global grid reference data
file in ESRI ASCII format. The main difficulty arises from the fact that the global GEBCO grid is subdivided into several files,
or tiles. Hence, for each pixel (i, j) in the current grid, we need to know in which tile (sub-file) it is located and to which pixel
345 (i', j') it corresponds. Indeed, the same current grid can straddle several tiles at the same time.

Using grid notation for tiles, i.e. with (x, y) the tile located on row x and column y, a pixel (i, j) is then located on:

- tile $(x, y) = \left(\left\lfloor \frac{90 - lat_{(i,j)}}{h} \right\rfloor + 1, \left\lfloor \frac{lon_{(i,j)} + 180}{w} \right\rfloor + 1 \right)$,

- pixel $(i', j') = \left(\left\lfloor \frac{((90 - (x-1)h)) - lat_{(i,j)}}{cellsize} \right\rfloor, \left\lfloor \frac{(lon_{(i,j)} - (-180 + (y-1)w))}{cellsize} \right\rfloor \right)$,

350 where h and w correspond respectively to the width and height (in decimal degrees) of each of the tiles (regular sub-division)
and $lat_{(i,j)}$ and $lon_{(i,j)}$ correspond respectively to the latitude and longitude of the point at the centre of cell (i, j). The principle
is summarised in Figure 19.
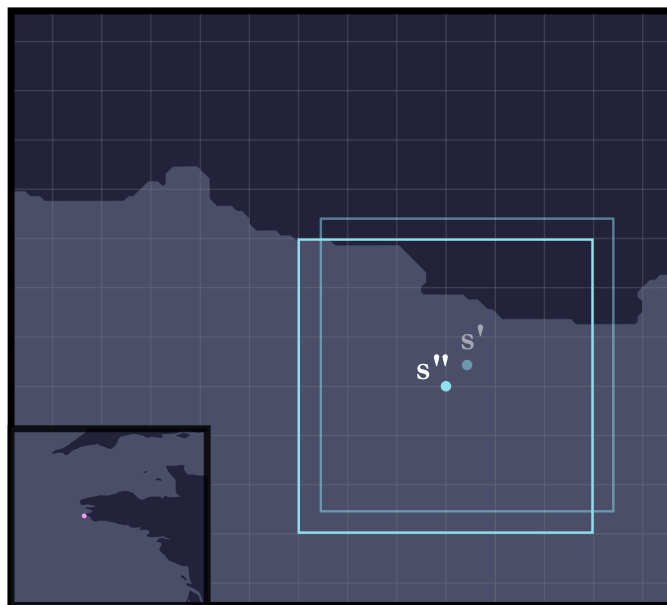
Figure 18. Illustration of the procedure for recalibrating the current grid on the GEBCO graticule: even case (GEBCO graticule in grey, i.e. 15 arc-seconds).

4.5 Selection criteria

Now that we have retrieved elevations for the entire grid, we need to ensure that the instance meets certain criteria in order to decide whether or not to accept it. In Section 3.1, we stated that an instance must have a unique maritime connected component and a ratio of maritime cells lying between 25 % and 95 %. However, it is essential to exercise increased accuracy in this regard, due to the presence of a subtlety. Indeed, in the case of instances generated from real geographical locations, it is highly unlikely to have a unique maritime connected component, so we had to resort to a few tricks to overcome this technical challenge. This is particularly the case where there are inland waters such as ponds, lakes or rivers. Therefore, in practice, the true criteria for selecting an instance are that:

- the weight of the main maritime connected component must be of at least 90 % of the total maritime cells (inland waters included),

- there must be a ratio of maritime cells of between 25 % and 95 % after inland waters, if any, have been artificially filled in (no more than 10 %).

The reason for choosing a minimum of 90 % for the weight of the main connected component is that we did not want to deteriorate the instances excessively by artificially filling in the inland waters, it was therefore necessary to impose an upper bound on their presence (i.e. 10 % in this catalogue).
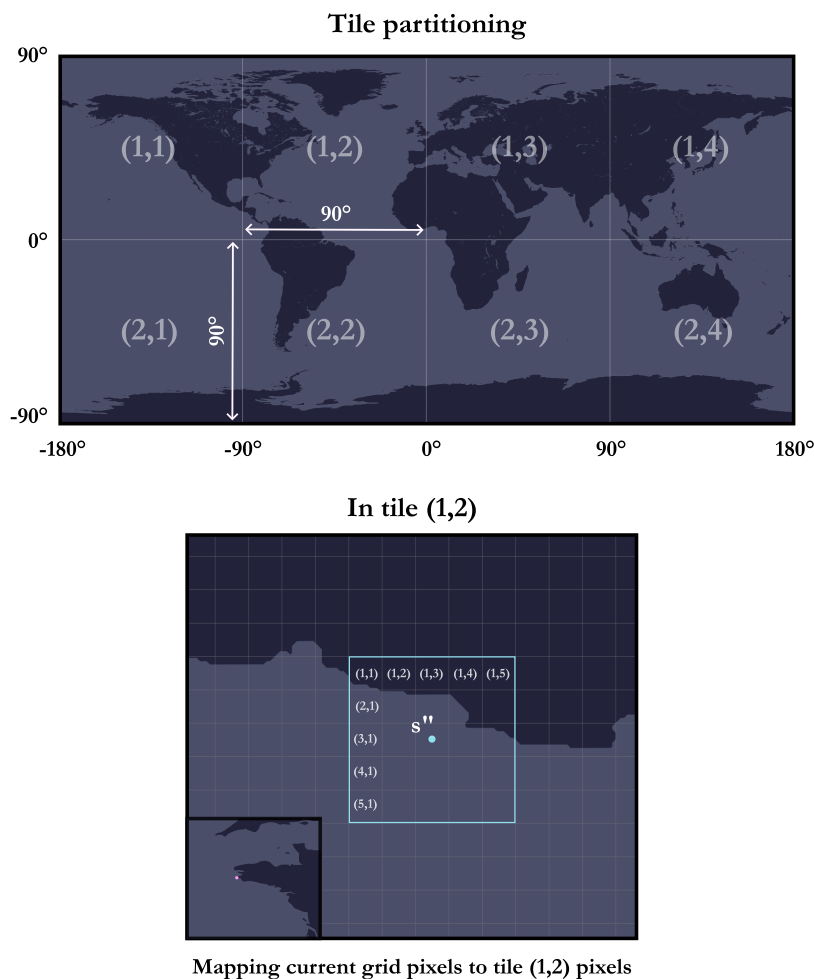
Figure 19. Illustration of the procedure for retrieving elevations of the different cells (pixels) in a current grid using a global grid sub-divided into tiles.

To achieve this, a given instance (DEM) has been transformed into a graph where each maritime cell represents a vertex
370 and the edges embody the connections between adjacent cells. We have opted for the case where a given maritime cell has 4
direct neighbours: west, east, south and north. This is an arbitrary implementation choice, and we could also have considered
the alternative of taking into account the diagonal cells (see Figure 20 for an overview of the differences).

With such a graph, it is now possible to determine the maritime connected components, see He et al. (2017) for more details.
375 If a given instance is accepted, the set of isolated maritime cells (inland waters) that are not in the main connected component
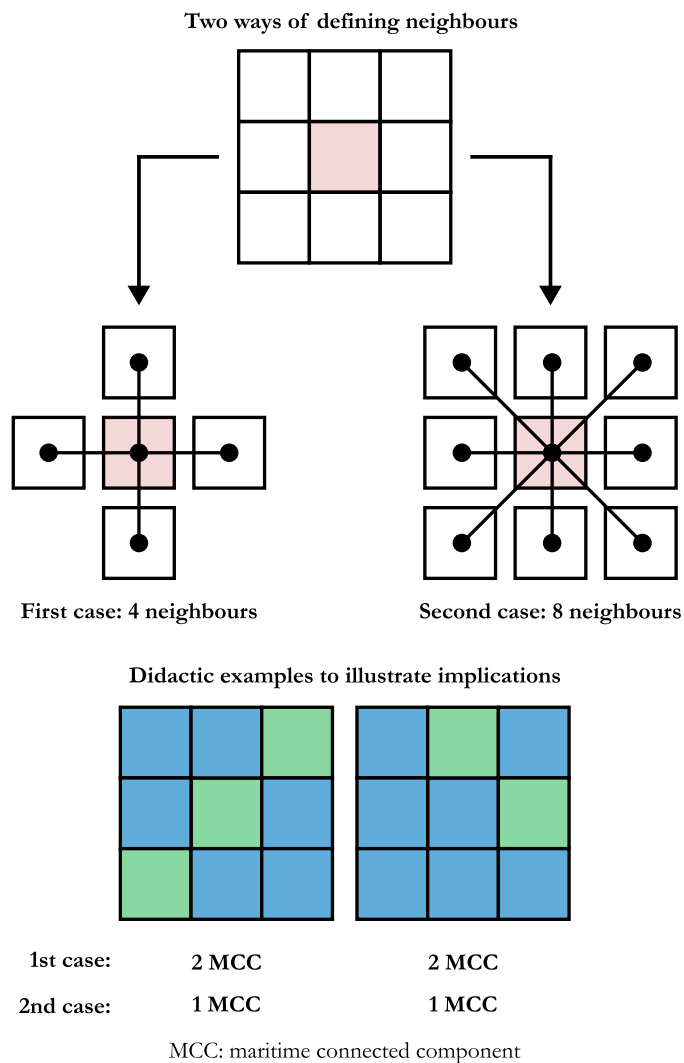
Figure 20. Illustration of the two alternative ways of defining the neighbours of a given cell and the possible implications when this is implemented on a particular instance.
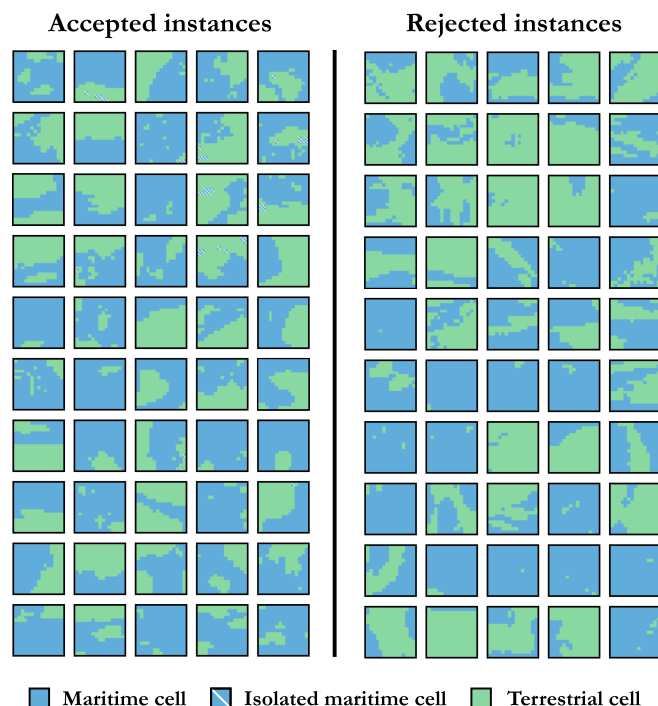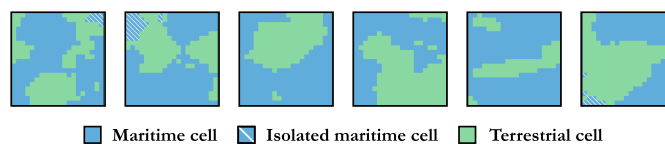
Figure 21. Examples of instances (n=15) that were accepted as meeting the selection criteria and instances that were rejected.

are then artificially filled in as discussed in Section 3.1. Finally, Figure 21 illustrates this with two columns: one for accepted instances (meeting the criteria) and one for rejected instances[25].

4.6 Reverse geocoding

Once the instance has been approved, we will try to determine the country to which it belongs so that we can proceed with
380 the naming process, in accordance with the convention discussed in more detail in Section 3.3. For this, the approach is to use a reverse geocoding Application Programming Interface (API) to query an appropriate database using the seed, i.e. the centre of the grid. It may happen that the query returns no result for the seed, in which case it is generally sufficient to poll all the terrestrial cells until there is a match (this can happen for very large instances). If no country of affiliation is detected by reverse geocoding (this has happened 3 times out of all the instances generated), the instance is simply rejected automatically. Note
385 that in this catalogue, we have chosen to use only the country, but it is entirely possible to use the smallest subdivision for the geographical location in question. This has not been done, because the subdivisions are unequal depending on the location and this can lead to instance names that are excessively verbose.

---

[25]These instances are used for didactic purposes and have not been included in the catalogue.

Figure 22. Examples of instances ($n = 15$) withdrawn from the catalogue due to special bottleneck situations.

Specifically, those country names were obtained using Nominatim (from the Latin, meaning "by name"), which is a search
390   engine for OpenStreetMap (OSM) data. It is important to note that OSM recognizes as countries only those political entities
listed in the ISO-3166-1 standard with the attribute "Independent=Yes". Moreover, the borders depicted in OSM are those that
are broadly recognized at the international level and most accurately represent the on-the-ground realities, often implying phys-
ical control. In regions where the borders lack precise definition, their representation is only an approximation. The objective
of the OSM community is to create a map that reflects the current state of the world, rather than an idealized version of it.

395   **4.7   Manual post-processing**

Finally, a last step involves manually reviewing the instances, because occasionally there are a few undesirables despite the
criteria put in place (refer to Section 4.5). This applies, for example, to instances where bottlenecks occur, e.g. with a narrow
passage bridging two maritime areas. This is a subjective criterion, i.e. left to personal discretion, but these instances were
considered less relevant than the others to the field of application initially targeted, and were filtered out for this reason. Figure
400   22 shows some instances of this type.

5   Conclusions and perspectives

In this paper, we have provided free access to the first catalogue, to the best of our knowledge, consisting of 17 700 coastal
instances spread throughout the globe and distributed equally between three libraries "A", "B" and "C". These instances, or
Digital Elevation Models (DEMs), are delivered in the form of raster grids with bathymetric and topographic data originating
405   from the GEneral Bathymetric Chart of the Oceans (GEBCO) as of 2022 with a resolution of 15 arc-seconds, i.e. approximately
460 m $\times$ 460 m at equatorial level. These instances cover a wide range of dimensions, from 10×10 to 300×300 in steps
of 5 (100 instances per dimension for a given library), and a broad spectrum of different coastlines geometries. Furthermore,
the common feature between these different instances is the presence a unique maritime connected component with a ratio
of maritime cells lying between 25 % and 95 %. These instances are sorted in ascending order of maritime cells within each
410   dimension of a given library. In a nutshell, this catalogue was created with the intention of constituting a reference benchmark
for eventual numerical experiments or applications that require working on such Areas of Interest (AoIs). Moreover, we have
provided in-depth details of the automated generation procedure for these instances, and we have proposed a number of visu-
alisations in 2 and 3 dimensions as well as adapted colour palettes compiled specially for the occasion.

415 As far as the future is concerned, we can already say that it may be necessary to update this catalogue in a few years' time, once the ocean floor has been mapped more accurately. Indeed, in the period 2022-2023, the scientific consensus is that only 20 % of the oceans have been accurately mapped by direct measurement, the rest having been interpolated. It would also be interesting to propose an analogous catalogue made up solely of purely maritime instances (i.e. without coastlines), as this could be of interest in many application domains. It would therefore be wise to define criteria to classify these instances and

420 differentiate them from one another (type of seabed, underwater topography, etc.). Another approach would be to supplement these instances with the GPS coordinates of the various coastlines that we know with greater precision. This could provide additional information compared with the elevation data we currently have at our disposal, the resolution of which can be a hindrance depending on the application being addressed. Finally, it would also be interesting to provide a new library comprised solely of Antarctic instances, which has so far been completely excluded from the catalogue.

425 ## 6 Data availability

The catalogue produced as part of this research is accessible for download on Zenodo, a general-purpose repository operated by CERN (European Organisation for Nuclear Research) and developed within the European OpenAIRE initiative, at the following persistent link: https://doi.org/10.5281/zenodo.10530247 (last access: February 15th, 2024) (Thuillier et al. (2024b)). In particular, the repository contains the 17 700 instances supplied in ESRI ASCII format, as well as a comprehensive docu-

430 mentation, including a user guide detailing the entire dataset and its organisation. In addition to this user guide, we also provide visualisations of the 17 700 instances in the form of duly annotated satellite images, and in the form of mosaics made up of two-colour thumbnails, with both Mercator and equirectangular projection. In addition to this catalogue, a set of 18 colour palettes dedicated to the visualisation of DEMs has also been derived for the occasion and is available at the following persistent address: https://doi.org/10.5281/zenodo.10530296 (last accessed: February 15th, 2024) (Thuillier et al. (2024c)). This

435 collection of palettes also come with an extensive documentation. The organisation of these two repositories is illustrated in Figures 23 and 24 respectively.

Appendix A: Coordinate Reference System (CRS)

First of all, a CRS, sometimes called a Spatial Reference System (SRS), enables a point to be uniquely located in a two- or three-dimensional space. There are two main types of CRS: horizontal CRS (2D) used to locate a point on the surface of the

440 globe (i.e. the horizontal component) and vertical CRS (1D) used to give the elevation of a point (i.e. the vertical component). Although a horizontal CRS is essential for defining a point, a vertical CRS is not mandatory. It should be noted that we also refer to a Compound Coordinate Reference System (CCRS) when we combine a horizontal CRS and a vertical CRS (3D).
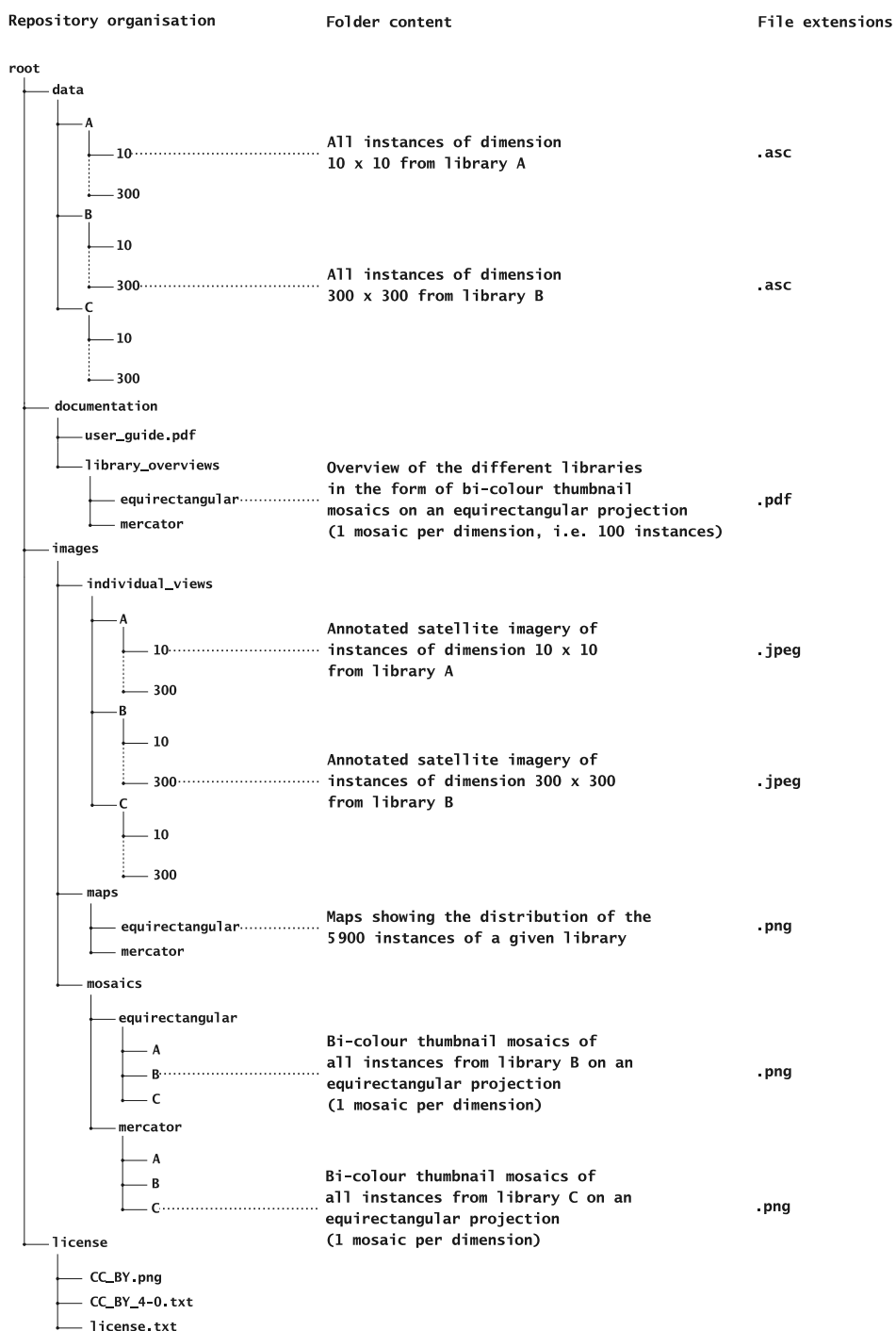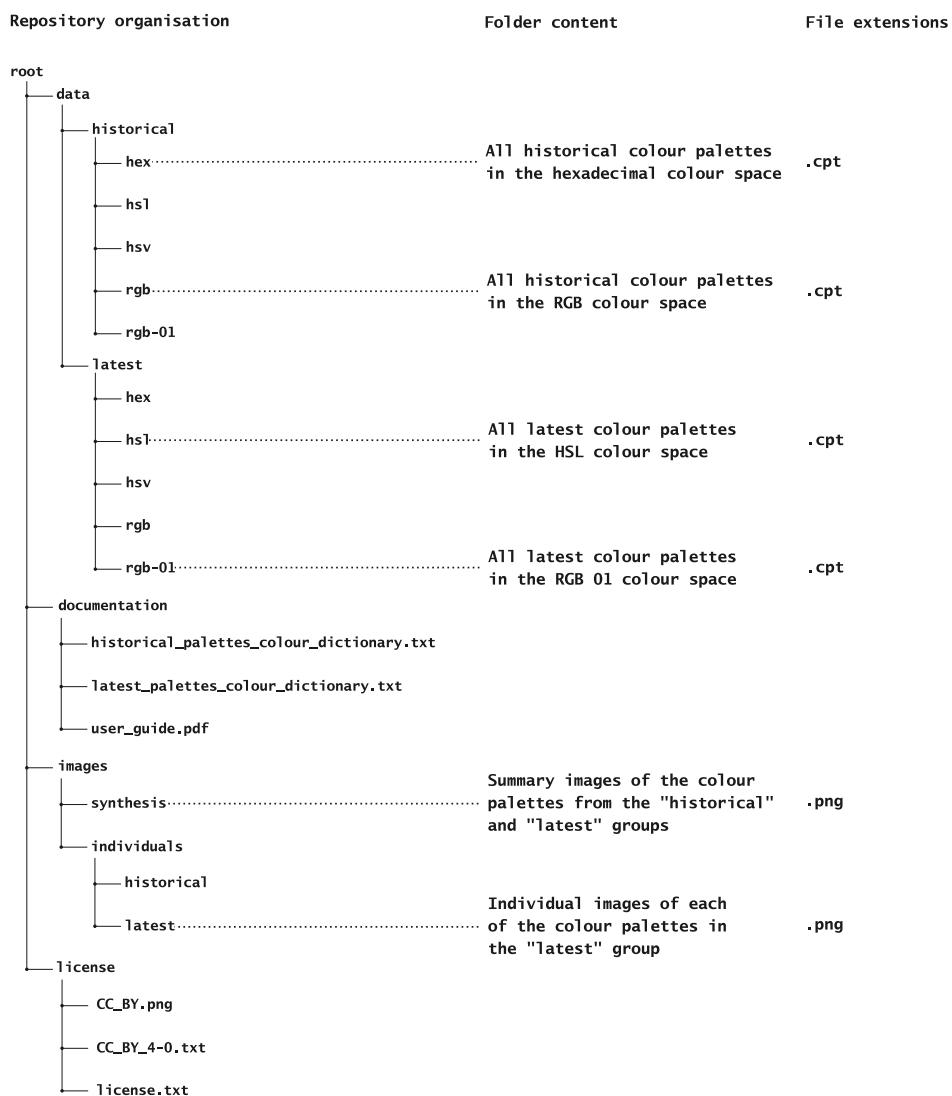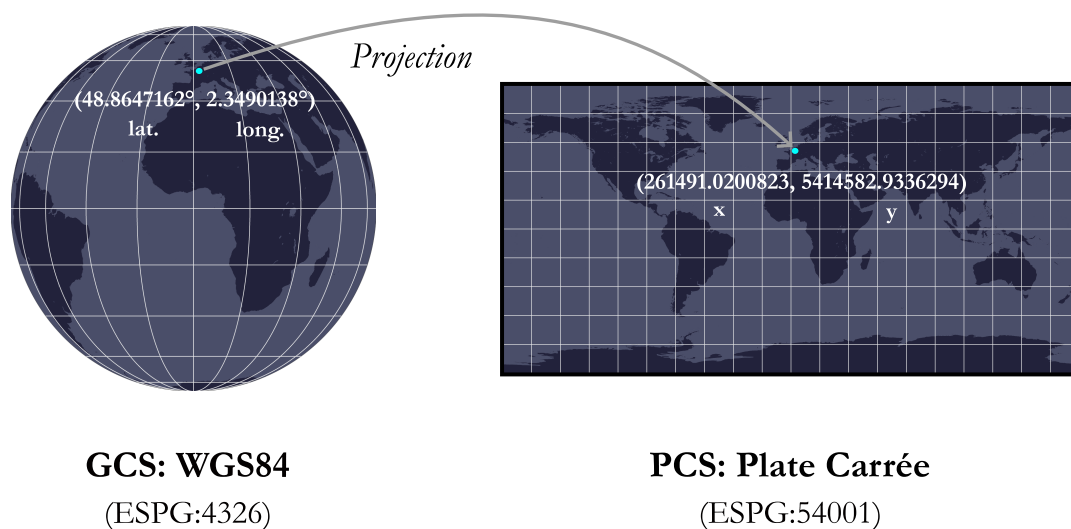
Figure 23. Organisation of the repository for the catalogue of instances.

Figure 24. Organisation of the repository for the colour palettes.

GCS: WGS84

(ESPG:4326)

PCS: Plate Carrée

(ESPG:54001)

Figure A1. Synthetic difference between a GCS and a PCS.

A1 Horizontal CRS

There are two main types of horizontal CRS: Geographic (geodetic) Coordinate Systems (GCS) and Projected Coordinate
Cystems (PCS). The former is used to define the location of points on a model (i.e. an approximation) of the Earth's surface,
while the latter is planar and necessarily contains a GCS from which the projection is made. In short, the GCS describes where
points are on the Earth's surface and the PCS describes how to represent them on a flat surface (see Figure A1).

More specifically, a GCS is made up of:

– a reference spheroid, or ellipsoid[26]), used to approximate the shape of the Earth,

– a (horizontal) datum used to position the reference spheroid at a certain point relative to the Earth on a so-called anchor
point (horizontal reference point)[27],

– a reference meridian to locate the 0° of longitude,

– an angular unit, often degrees (in Decimal Degrees (DD) or Degrees Minutes Seconds (DMS) notation), or even radians
in some cases.

As a reminder, the latitude of a point is the angle formed by the normal to the plane tangent to this point with the equatorial
plane. It is an angular value expressing the north or south position of a given point (relative to equator). The longitude of a
point corresponds to the angle at the centre that the plane passing through this point and through the earth's axis of rotation
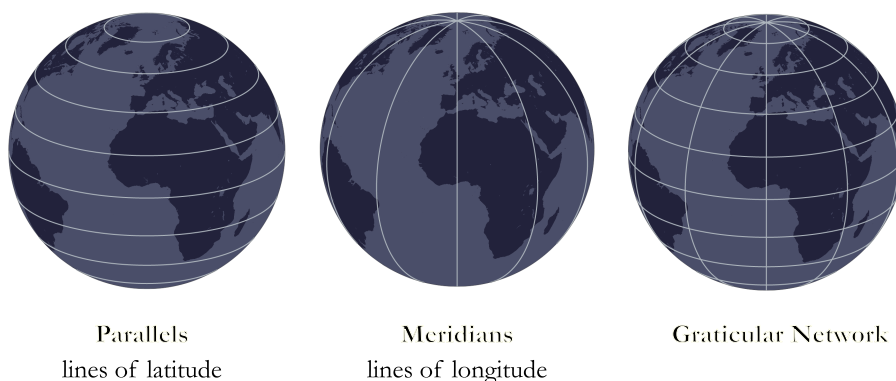
---

[26]The terms spheroid and ellipsoid are regularly used interchangeably in the GIS community, so we will use spheroid here for the sake of consistency.

[27]This also includes the orientation of the spheroid.
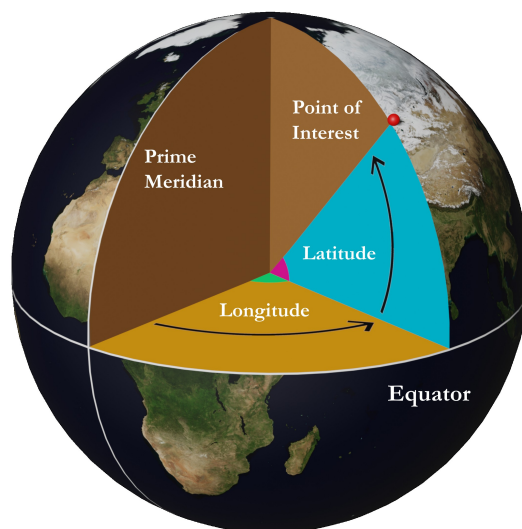
Figure A2. Latitude and longitude: a visual interpretation.

Figure A3. Illustration of the graticular network made up of imaginary lines known as parallels and meridians.

forms with the plane of the reference meridian (i.e. prime meridian). It is an angular value expressing the east or west position
460 (relative to the prime meridian). See Figure A2 for a visual explanation. A GCS is therefore based on these imaginary lines of
latitude (parallels) and longitude (meridians), which are structured into a so-called graticule to enable all the points on Earth to
be identified (with respect to a given spheroid and a datum) as illustrated in Figure A3.

Given that the earth is not perfectly round, there is a plethora of datums (spheroid + anchor point) that are more or less
465 adapted to certain parts of the world depending on the reference spheroid on which they are based. Nevertheless, a more ac-
curate representation of the Earth's surface than spheroidal models is the geoid, which corresponds to an equipotential in the
Earth's gravity field. More precisely, the geoid is, at all points, perpendicular to the direction of the gravity vector (i.e. the
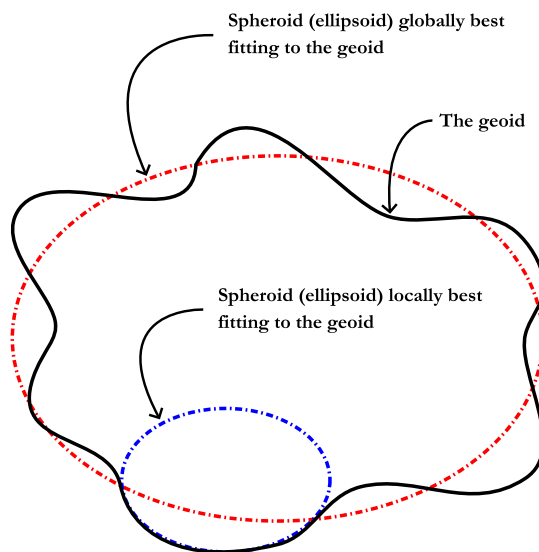
Figure A4. The geoid, the ellipsoid that best fits it at global level and the ellipsoid that best fits it at regional level, for a resricted region.

direction of the plumb line). And, as the mass of the Earth is not uniform at all points and the direction of gravity changes, the shape of the geoid is thus irregular (bumpy), which is why it is easier to manipulate spheroids that simply approximate it as closely as possible, either locally or globally (see Figure A4 for an illustration). In simpler terms, the geoid can be described as the surface that coincides with the undisturbed mean level of the oceans (e.g. without storms and tides) and with its imaginary extension through the continental masses. For a continental mass, the geoid could be physically described at any point by digging a narrow channel to connect it with an ocean, thus causing the water in the channel to settle at the level of the geoid (Lambert (1926)). In particular, there are several geoid models, such as the Earth Gravitational Models (EGM) of 1984 (EGM84), 1996 (EGM96) and 2008 (EGM2008). Figure A5 shows a view of the EGM2008 geoid. Back to the spheroids, it is possible to distinguish one from another by the lengths of the semimajor and semiminor axes. For example, consider the following table with four particular spheroids.

Spheroid	Semimajor axis (m)	Semiminor axis (m)
Clarke 1866	6378206.4	6356583.80000
International 1924 (Hayford)	6378388.0	6356911.94612
GRS80	6378137.0	6356752.31414
WGS84	6378137.0	6356752.31424

Table A1. A few examples of spheroids.
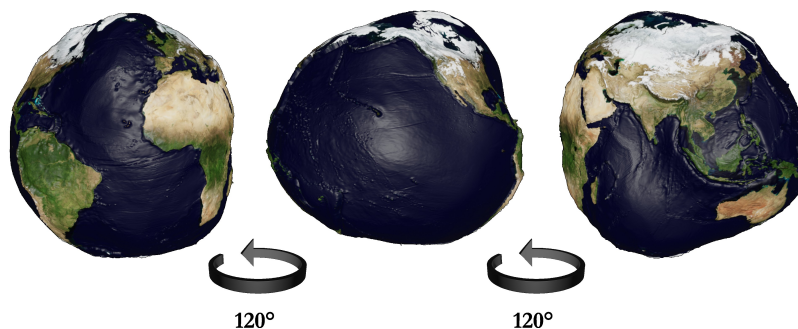
Figure A5. Visualisation of the 2008 Earth Gravitational Model (EGM2008) geoid with vertical exaggeration and background texture from NASA Visible Earth (Blue Marble Next-Gen, December version).

480　We note that depending on whether we use the Clarke 1866 spheroid, International 1924 (Hayford), GRS80 or WGS84 to define a datum and therefore, by extension, a GCS, we will not obtain the same locations for a single point in space. One spheroid may be more suitable than another for a given region, because it approximates the geoid for that region as closely as possible (but it may be a poor approximation for another part). The best known is the WGS84 spheroid, which is a reliable global approximation and is therefore used as the standard for the Global Positioning System (GPS), among other things. It should also be noted that there are some possible confusions, as there is the WGS84 GCS based on the WGS84 datum, itself
485　based on the WGS84 spheroid. The WGS84 GCS uses the Greenwich meridian as its reference meridian and the angular unit is the degree.

Thus, the choice of GCS and therefore, by extension, of datum and reference spheroid, is crucial, as the coordinates will not be the same depending on the approximation used. An example is illustrated below with the details of the city of Paris in
490　various GCSs.

GCS	Latitude (°)	Longitude (°)
OSGB36	48.8647162	2.3490138
ED50	48.8656272	2.3502878
Kalianpur 1975	48.8637586	2.3391549
WGS84	48.8647162	2.3490138

Table A2. Differences in coordinates for the city of Paris depending on the GCS used.

Now that we know where our points are, we can decide to supplement our GCS with a PCS in order to project our points onto a flat surface (i.e. in 2 dimensions). A PCS is made up as follows:

– a GCS defining the Earth model on which the PCS is based (input details),

    – a type of projection, strictly speaking,

495   – a linear unit such as a metre or a kilometre, for example.

The projection corresponds to the algorithm used to transform the Earth according to the model defined in the GCS into a flat surface (do not confuse projection and PCS, the former is included in the latter). Note that depending on the projection used, there may be a certain number of additional parameters (e.g. false easting, central meridian, standard parallel, etc.) enabling the PCS to be centred on a certain part of the world. Nevertheless, as there is no way to transpose a curved surface (spheroid) into a flat surface without inducing any distortion, there are a multitude of projections with different properties (more than 100 projections). Some projections preserve local angles, others the surface, while some preserve specific distances or directions. In general, the choice of projection is guided by the extent, location and property that needs to be preserved. Figure A6 shows some reference projections. For the GEBCO grid, and as we will see later in Section 2, the preferred projection is the flat square projection, which is a special case of an equirectangular projection centred on the equator.
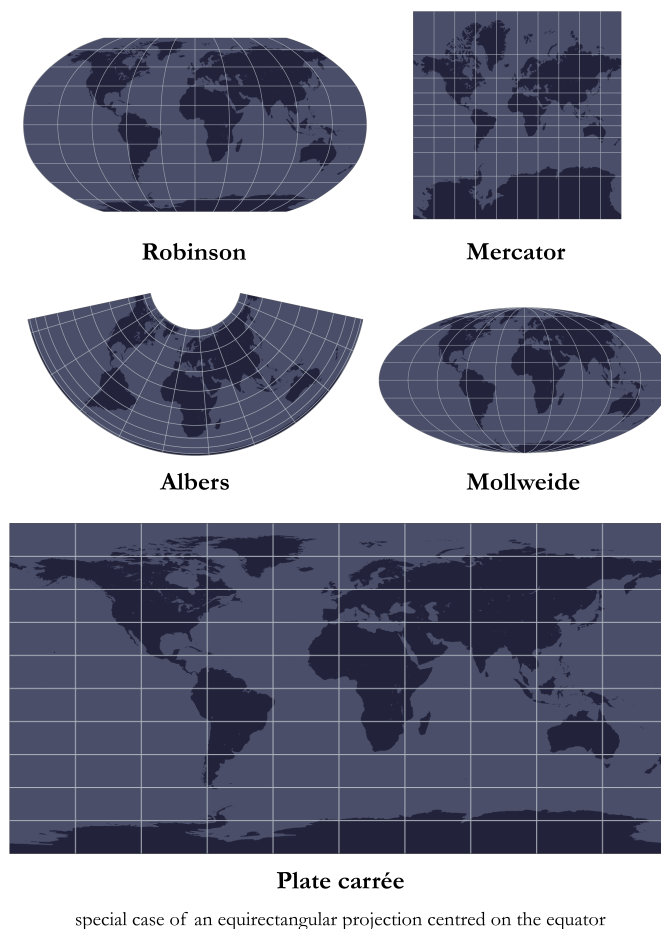
A2   Vertical CRS

There are the vertical CRSs, which can be coupled with various horizontal CRSs to create a complete 3D system (i.e. elevation included). The Geodetic Glossary (2009) defines elevation[28] as "the distance, measured along a perpendicular, between a point and a reference surface". Although this definition is precise and concise, it leaves some ambiguity as to the reference surface used, which is an essential element of a vertical CRS. A vertical CRS, sometimes called a VCS (Vertical Coordinate System), is formally made up of:

    – A datum or vertical datum representing the reference point (surface) from which to measure an elevation (depth or altitude): this is the zero-point of elevation. There are two main types of VCS: those based on geometric models (ellipsoidal, spheroidal), which are referred to as datum (similar to those in the horizontal CRSs), and those based on gravity-related models (geoidal), which are referred to as vertical datum.

515   – A direction for the main axis, which can be directed upwards or downwards depending on the quantity being measured (height or depth).

   – A unit of measurement, necessarily linear, such as a metre or a foot.

In the case of a VCS based on a geometric model, the elevations are therefore measured directly from a reference spheroid (e.g. the WGS84 spheroid), whereas in the case of a gravity-related model, the elevations are measured from a geoid (e.g. Earth Gravitational Model (EGM) 2008) or an aproximation thereof. It should be noted that Mean Sea Level (MSL) has long been considered a satisfactory approximation to the geoid and can therefore be used as a reference surface for determining elevation. Nowadays, we know that the mean sea level can deviate from the geoid by one metre, but the exact difference is proving

---

[28]This actually corresponds to the strict definition of "height" in the glossary, but elevation and height are regularly used as synonyms. To be precise, we should be talking about elevation in the case of a geoid as a reference surface (or geodetic height) and height if we are using an ellipsoid.

Robinson   **Mercator**

Albers   **Mollweide**

Plate carrée

special case of an equirectangular projection centred on the equator

Figure A6. Some examples of common projections.

difficult to determine. Despite this, much of the data is still referenced to MSL. At its simplest, the MSL corresponds to the average position of the ocean surface measured over time. The aim is to minimise as far as possible the random and periodic

525   variations caused, for example, by tides or storm surges. The time window for measuring these variations in the ocean surface has been set at 19 years by the U.S. National Ocean Surface. These measurements can then be combined to form various tidal datums such as (in addition to MSL) the Mean Low Water (MLW), the Mean Tide Level (MTL) or even the Mean Lower Low Water (MLLW), among many others. Nevertheless, in this form, the MSL is not necessarily the most suitable global vertical reference, as this computation only exists at a given measuring station and in its immediate vicinity (e.g. the Marseille tide

530   gauge to measure elevation in France). There is also a point to be borne in mind when coupling a VCS with a horizontal CRS if one wishes to use a geometric model for the elevation: care must be taken to ensure that the reference spheroid is identical for the VCS and the GCS (or the PCS, if applicable). These different types of information are summarised in Figure A7.
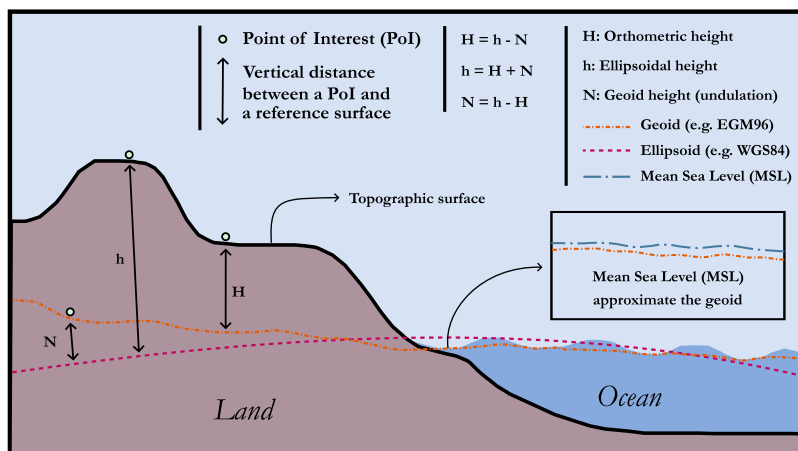
Figure A7. Overview of the difference between the two types of vertical datum: geometric model (spheroid or ellipsoid) and gravity-related model (geoid).

A3 Unique reference code: European Petroleum Survey Group (EPSG)

The EPSG Geodetic Parameter Dataset is a register of spheroids, datums, horizontal and vertical CRSs, units, etc. This register was initiated by the European Petroleum Survey Group (EPSG) in 1985 and assigns to each feature an EPSG code between 1024 and 32767 as well as a textual (human-readable) description in the "Well-Known Text representation of Coordinate Reference Systems (WKT or WKT-CRS)" format. The register is currently maintained by the International Association of Oil & Gas Producers (IOGP) Committee. These ESPG codes are used extensively in GIS software. For example, the WGS84 horizontal CRS is identified by EPSG:4326, while the WGS84 spheroid is identified by EPSG:7030 and the WGS84 datum by EPSG:6326. For vertical CRS based on MSL, we find, for example, MSL Height with ESPG:5714 (positive upwards) or MSL Depth with ESPG:5715 (positive downwards). The content of a file in the WKT format for the WGS 84 CRS is shown below:

```
world_geodetic_system_1984.wkt
GEOGCS["WGS 84",
  DATUM["WGS_1984",
    SPHEROID["WGS 84",6378137,298.257223563,
      AUTHORITY["EPSG","7030"]],
    AUTHORITY["EPSG","6326"]],
  PRIMEM["Greenwich",0,
    AUTHORITY["EPSG","8901"]],
  UNIT["degree",0.0174532925199433,
    AUTHORITY["EPSG","9122"]],
  AUTHORITY["EPSG","4326"]]
```

Appendix B: Digital Elevation Model (DEM)

545 A digital elevation model (DEM) is a three-dimensional approximation of a terrain surface by means of a discrete set of points (3D), or point cloud, expressed in a horizontal CRS (2D) and to which is associated an elevation in a vertical CRS (1D) (see Section A for the definition of these CRSs) (Hirt (2014)). DEMs are digital in the sense that it is produced, distributed and used in digital format (Croneborg et al. (2020)) and are a model because the elevations must be available at all points in the AoI (Pike et al. (2009); Hengl and Evans (2009); Szypuła (2017)). Indeed, if the elevation is not available at every point,

550 then these are simply samples of heights at discrete locations (points or lines such as contour lines), and not models of a land surface (Hengl and Evans (2009)). It should also be noted that a DEM can represent the dry land surface (topography) and/or submerged surfaces (bathymetry) of the surface of a terrestrial area (Earth-based) as well as that of solid celestial bodies such as asteroids, sattelites or planets (telluric or gaseous) (Guth et al. (2021)). In addition, when a DEM covers an entire area, such as the earth's surface, it is also referred to as a Discrete Global Grid (DGG), as is the case with the GEBCO global grid

555 (see Section 2). More specifically, DGGs are a partition of space, i.e. a division into non-overlapping regions (e.g. pixels): a mosaic. Sometimes, DEMs are also referred to as '2.5D' rather than true 3D models, as some terrain features (e.g. caves and overhanging cliffs) cannot be properly represented (Hirt (2014)).

In the literature, a DEM is often used as an umbrella term for either a Digital Surface Model (DSM) or a Digital Terrain

560 Model (DTM) without further information about the surface (Masini et al. (2011a); Hirt (2014); Guth et al. (2021)). A DSM takes into account all the objects and structures present on the ground, whether they belong to the biosphere (vegetation) or the antroposphere, i.e. all man-made structures (e.g. power lines, buildings, bridges) (Hirt (2014); Guth et al. (2021)). A DTM filters out the biosphere and anthroposphere, leaving only a representation of the bare ground (bare earth surface) (Hirt (2014); Guth et al. (2021)). In other words, a DSM corresponds to the highest surface (radar reflective), while the DTM is a DSM

565 stripped of all surface objects, either natural or man-made. The difference between these terms is illustrated schematically in Figure B1. It should be noted, however, that there is no consensus on the definition of the terms DEM, DTM and DSM in the scientific literature (Hirt (2014)). Thus, sometimes DEM and DTM are used as synonyms (Arundel et al. (2015); Croneborg et al. (2020)), DEM and DSM as synonyms (Graham et al. (2007)) or DEM is defined as a subset[29] of DTM (Zhou (2017)). As in Section 2, although we call the GEBCO global grid a DEM, it would be entirely possible and appropriate to call it a DTM,

570 as it is more precise.

Going further, according to Hengl and Evans (2009), all DEMs can be classified into the following two groups (see Figure B2 for a schematic view of the differences):
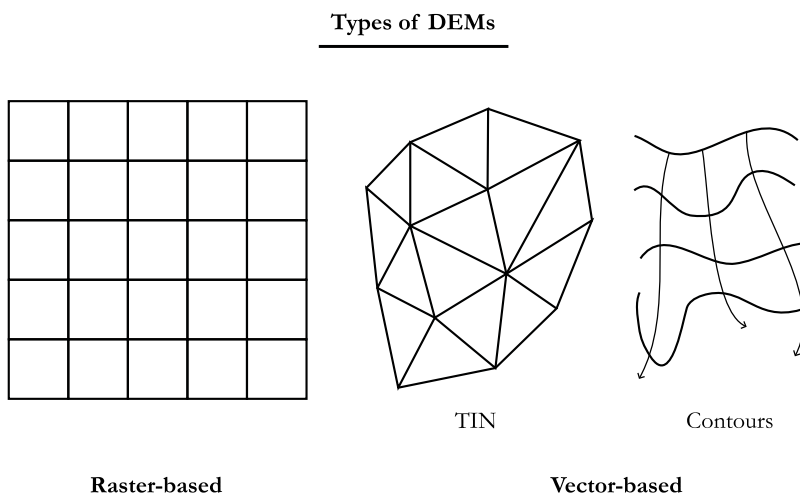
– vector-based, or irregular,

---

[29]In this case, a DTM is an enhanced-DEM with one or more types of information about the terrain, such as morphological features, drainage patterns or soil properties.
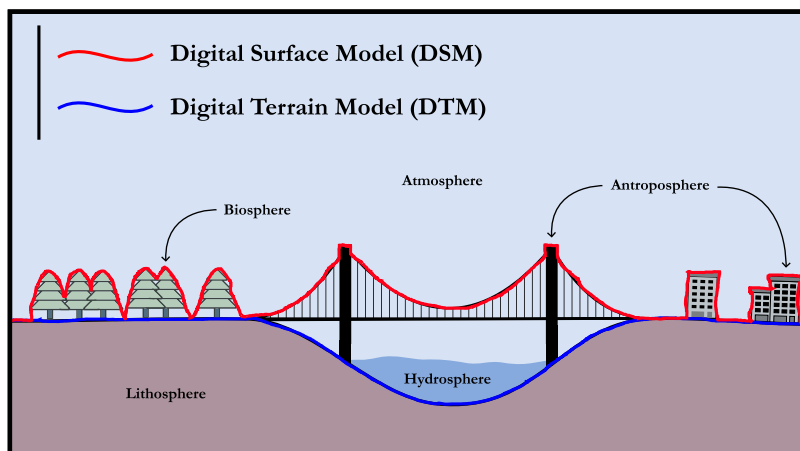
Figure B1. Differences between a DSM and a DTM.

Types of DEMs

TIN   Contours

Raster-based   **Vector-based**

Figure B2. The two different types of DEMS: raster-based and vector-based.

575   – raster-based (grid-based[30]), or regular.

In the vector-based or irregular group, there are two main sub-categories (Masini et al. (2011b)): Triangular Irregular Networks (TINs) and contours (level lines). The former involves deriving a DEM using a triangulation algorithm to pave a surface with a set of non-overlapping contiguous triangles (irregulars). This can be done using Delaunay triangulation, where the vertices of each triangle correspond to points in the initial sample and where we try to maximise the minimum of all the angles

580   of the triangles in the triangulation (so as to avoid sliver triangles). In this way, in each triangle, the surface (facet) is represented by a plane between the three vertices and the values of the points inside can be calculated through interpolation. As

---

[30]As pointed out in Hengl and Evans (2009), the term grid is probably more appropriate than raster, the former being a mathematical concept and the latter more related to a technology, although favoured by the GIS community.

for contour-based irregular DEMs, they are derived using lines connecting points of same elevation, i.e. isohypses (isoheights) and isobaths (isodepths), as well as their orthogonals in order to produce surfaces (Hengl and Evans (2009); Masini et al. (2011a)). The advantage of these irregular DEMs is that it is possible to accurately capture sudden changes in elevation in the area concerned with a number of points equivalent to a DEM in raster form (de Sousa et al. (2006)). To be more precise, in Mark (1975), the author was one of the first to compare TINs with DEMs in raster form and to conclude that it took twice as much memory to obtain as good an estimate as with TIns using DEMs. The major drawback is that it is difficult to manipulate DEMs in these vector forms compared with the raster DEMs presented thereafter.

The raster-based or regular group includes DEMs based on regular tiling (i.e. a grid) of space using polygons such as triangles, squares or hexagons[31], the only regular shapes that can be used to tile (tesselate) a given space (i.e. no overlaps or gaps). In this group, the DEMs are then stored in the form of a matrix whose values correspond to the elevations. The advantages of this regular tiling of space are that the conversion between the coordinates of a geographic point in a given CRS and the coordinates of the corresponding pixel in the matrix is trivial and the structure is simple enough to permit visualisation.

Overall, there is only one important characteristic for this type of DEM: the size of a cell (pixel), which defines the so-called resolution of the DEM. In addition, because of their ease of use, models based on a regular tiling of space using squares remain a standard for DEMs (see in particular Section 2). However, the major disadvantage of raster DEMs is that they tend to under-sample areas with complex topography and over-sample areas with less complex (i.e. smooth) topography, due to the regular spacing of the polygons. Unlike vector models such as TINs or contour-based DEMs, abrupt changes in elevation are not particularly well described using raster DEMs.

It should also be noted that the grid of a DEM does not necessarily have to be regular and stored in the form of a two-dimensional array (matrix); it can be irregular and stored using quadtrees, for example. Similarly, vector-based DEMs can be regular, for example and as stated in Hengl and Evans (2009), a regular lattice of points stored as a vector layer is regular. The regular/irregular distinction for DEMs can therefore be a source of confusion, and it is better to retain the notion of raster or vector for DEMs (Hengl and Evans (2009)).

---

[31]See for example the work of de Sousa et al. (2006) for a comparison between square-based and hexagon-based rasters.

References

615 Arundel, S. T., Archuleta, C.-A. M., Phillips, L. A., Roche, B. L., and Constance, E. W.: 1-Meter Digital Elevation Model specification, Techniques and Methods, 2015.

Avcioglu, A., Bereketli, A., and Bay, O. F.: Three Dimensional Volume Coverage in Multistatic Sonar Sensor Networks, IEEE Access, 10, 123 560–123 578, https://doi.org/10.1109/ACCESS.2022.3223714, 2022.

620 Becker, J., Sandwell, D., Smith, W., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S. H., Ladner, R., and Marks, K.: Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS, Marine Geodesy - MAR GEODESY, 32, 355–371, https://doi.org/10.1080/01490410903297766, 2009.

Bohlander, J. and Scambos, T.: Antarctic coastlines and grounding line derived from MODIS Mosaic of Antarctica (MOA), National Snow and Ice Data Center, Boulder, CO, USA, 2007.

625 Cox, H.: Fundamentals of Bistatic Active sonar, Springer, Dordrecht, ISBN 978-94-009-2289-1, https://doi.org/10.1007/978-94-009-2289-1_1, 1989.

Craparo, E. M. and Karatas, M.: A Method for Placing Sources in Multistatic Sonar Networks, Tech. Rep. NPS-OR-18-001, Naval Postgraduate School, Monterey, California, 2018.

Craparo, E. M. and Karatas, M.: Optimal source placement for point coverage in active multistatic sonar networks, Naval Research Logistics, 630 67, 63–74, https://doi.org/10.1002/nav.21877, 2020.

Craparo, E. M., Fügenschuh, A., Hof, C., and Karatas, M.: Optimizing source and receiver placement in multistatic sonar networks to monitor fixed targets, European Journal of Operational Research, 272, 816–831, https://doi.org/10.1016/j.ejor.2018.02.006, 2019.

Croneborg, L., Saito, K., Matera, M., McKeown, D., and van Aardt, J.: Digital Elevation Models, https://doi.org/10.1596/34445, 2020.

de Sousa, L., Nery, F., Sousa, R., and Matos, J.: Assessing the accuracy of hexagonal versus square tilled grids in preserving DEM surface flow 635 directions, in: Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2006), pp. 191–200, Citeseer, 2006.

Douglas, D. H. and Peucker, T. K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartographica: The International Journal for Geographic Information and Geovisualization, 10, 112–122, https://doi.org/10.3138/FM57-6770-U75U-7727, 1973.

640 Fügenschuh, A. R., Craparo, E. M., Karatas, M., and Buttrey, S. E.: Solving multistatic sonar location problems with mixed-integer programming, Optimization and Engineering, 21, 273–303, https://doi.org/10.1007/s11081-019-09445-2, 2020.

GEBCO Bathymetric Compilation Group: GEBCO 2023 Grid - a continuous terrain model of the global oceans and land., https://doi.org/10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b, April 18, 2023.

GEBCO Bathymetric Compilation Group: GEBCO 2022 Grid - a continuous terrain model of the global oceans and land., 645 https://doi.org/10.5285/e0f0bb80-ab44-2739-e053-6c86abc0289c, June 22, 2022.

Gorny, A.: World Data Bank II General User GuideRep, Central Intelligence Agency: Washington, DC, USA, 1977.

Graham, A., Kirkman, N. C., and Paul, P. M.: Mobile radio network design in the VHF and UHF bands: a practical approach, John Wiley & Sons, 2007.

Guth, P. L., Van Niekerk, A., Grohmann, C. H., Muller, J.-P., Hawker, L., Florinsky, I. V., Gesch, D., Reuter, H. I., Herrera-Cruz, V., 650 Riazanoff, S., López-Vázquez, C., Carabajal, C. C., Albinet, C., and Strobl, P.: Digital Elevation Models: Terminology and Definitions, Remote Sensing, 13, https://doi.org/10.3390/rs13183581, 2021.

He, L., Ren, X., Gao, Q., Zhao, X., Yao, B., and Chao, Y.: The Connected-Component Labeling Problem: A Review of State-of-the-Art Algorithms, Pattern Recognition, 70, https://doi.org/10.1016/j.patcog.2017.04.018, 2017.

Hengl, T. and Evans, I.: Chapter 2 Mathematical and Digital Models of the Land Surface, Developments in Soil Science, 33, 31–63, https://doi.org/10.1016/S0166-2481(08)00002-0, 2009.

Hirt, C.: Digital Terrain Models, pp. 1–6, Springer International Publishing, Cham, ISBN 978-3-319-02370-0, https://doi.org/10.1007/978-3-319-02370-0_31-1, 2014.

Kelso, N. V. and Patterson, T.: Introducing natural earth data - naturalearthdata.com, Geographia Technica, 5, 25, 2010.

Lambert, W. D.: The Figure of the Earth and the New International Ellipsoid of Reference, Science, 63, 242–248, http://www.jstor.org/stable/1650705, 1926.

Mark, D. M.: Computer Analysis of Topography: A Comparison of Terrain Storage Methods, Geografiska Annaler Series A-physical Geography, 57, 179–188, 1975.

Masini, N., Coluzzi, R., and Lasaponara, R.: On the Airborne Lidar Contribution in Archaeology: from Site Identification to Landscape Investigation, ISBN 978-953-307-205-0, https://doi.org/10.5772/14655, 2011a.

Masini, N., Coluzzi, R., and Lasaponara, R.: On the Airborne Lidar Contribution in Archaeology: from Site Identification to Landscape Investigation, ISBN 978-953-307-205-0, https://doi.org/10.5772/14655, 2011b.

Olson, C. J., Becker, J. J., and Sandwell, D. T.: A new global bathymetry map at 15 arcsecond resolution for resolving seafloor fabric: SRTM15_PLUS, in: AGU Fall Meeting Abstracts, vol. 2014, pp. OS34A–03, 2014.

Pike, R. J., Evans, I. S., and Hengl, T.: Chapter 1 Geomorphometry: A Brief Guide, Developments in soil science, 33, 3–30, 2009.

Sandwell, D. T., Harper, H., Tozer, B., and Smith, W. H.: Gravity field recovery from geodetic altimeter missions, Advances in Space Research, 68, 1059–1072, https://doi.org/doi.org/10.1016/j.asr.2019.09.011, 25 Years of Progress in Radar Altimetry, 2021.

Soluri, E. and Woodson, V.: World vector shoreline, The International Hydrographic Review, LXVII(1), 27–35, 1990.

Szypuła, B.: Digital Elevation Models in Geomorphology, ISBN 978-953-51-3573-9, https://doi.org/10.5772/intechopen.68447, 2017.

Thuillier, O., Le Josse, N., Olteanu, A.-L., Sevaux, M., and Tanguy, H.: An improved two-phase heuristic for active multistatic sonar network configuration, Expert Systems with Applications, 238, 121 985, https://doi.org/10.1016/j.eswa.2023.121985, 2024a.

Thuillier, O., Le Josse, N., Olteanu, A.-L., Sevaux, M., and Tanguy, H.: Catalogue of Coastal-Based Instances [data set], https://doi.org/10.5281/zenodo.10530247, 2024b.

Thuillier, O., Le Josse, N., Olteanu, A.-L., Sevaux, M., and Tanguy, H.: Colour Palettes for Digital Elevation Models [data set], https://doi.org/10.5281/zenodo.10530296, 2024c.

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., and Wessel, P.: Global Bathymetry and Topography at 15 Arc Sec: SRTM15+, Earth and Space Science, 6, 1847–1864, https://doi.org/doi.org/10.1029/2019EA000658, 2019.

Urick, R. J.: Principles of underwater sound, McGraw-Hill, New York, 3rd ed edn., ISBN 9780070660878, 1983.

Wessel, P. and Smith, W. H. F.: A global, self-consistent, hierarchical, high-resolution shoreline database, Journal of Geophysical Research: Solid Earth, 101, 8741–8743, https://doi.org/doi.org/10.1029/96JB00104, 1996.

Zhou, Q.: Digital Elevation Model and Digital Surface Model, ISBN 9780470659632, https://doi.org/10.1002/9781118786352.wbieg0768, 2017.