



HAL
open science

Catalogue of Coastal-Based Instances

Owein Thuillier, Nicolas Le Josse, Alexandru-Liviu Olteanu, Marc Sevaux,
Hervé Tanguy

► **To cite this version:**

Owein Thuillier, Nicolas Le Josse, Alexandru-Liviu Olteanu, Marc Sevaux, Hervé Tanguy. Catalogue of Coastal-Based Instances. 2024, 10.5281/zenodo.10530246 . hal-04427796

HAL Id: hal-04427796

<https://hal.science/hal-04427796v1>

Submitted on 31 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Catalogue of Coastal-Based Instances: User Guide

Owein THUILLIER^{a,b,*}, Nicolas LE JOSSE^a, Alexandru-Liviu OLTEANU^b,
Marc SEVAUX^b, Hervé TANGUY^a

^aThales, Defence and Mission Systems, Brest, France

{[owein.thuillier](mailto:owein.thuillier@fr.thalesgroup.com), [nicolas.lejosse](mailto:nicolas.lejosse@fr.thalesgroup.com), [herve.tanguy](mailto:herve.tanguy@fr.thalesgroup.com)}@fr.thalesgroup.com

^bUniversité Bretagne-Sud, Lab-STICC, UMR CNRS 6285, Lorient, France

{[owein.thuillier](mailto:owein.thuillier@univ-ubs.fr), [alexandru.olteanu](mailto:alexandru.olteanu@univ-ubs.fr), [marc.sevaux](mailto:marc.sevaux@univ-ubs.fr)}@univ-ubs.fr

Abstract

This document provides a comprehensive user's guide to the use of this catalogue, which consists of 17 700 coastal-based instances distributed around the world and divided equally between three libraries: "A", "B" and "C". These instances, or Digital Elevation Models (DEMs), given in the form of regular rectangular grids (rasters), contain topographic and bathymetric data retrieved from the 2022 global grid made publicly available by the General Bathymetric Chart of the Oceans (GEBCO).

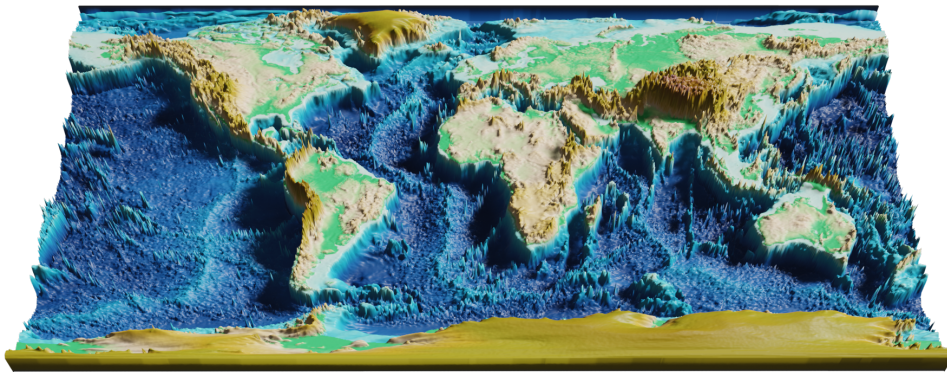


Figure 1: GEBCO 2022 global grid with z-axis exaggeration on an equirectangular projection.

Keywords: catalogue; geographical areas; coastlines; elevation data; raster grids; topography; bathymetry; digital elevation models

*Corresponding author.

Email address: owein.thuillier@fr.thalesgroup.com (Owein THUILLIER)

Table of contents	Page
Disclaimer	3
1 Introduction	4
2 Catalogue Organisation	4
2.1 Overall Organisation	4
2.2 Data	5
2.3 Documentation	5
2.4 Images	6
2.5 License	8
3 Projections	8
3.1 General Remarks	9
3.2 At Local Scales	9
4 Instances	11
4.1 Intrinsic Characteristics	11
4.2 Naming Convention	11
4.3 File Format	13
5 Instance Selection	16
5.1 Library Overviews	17
5.2 Individual Views	17
6 Visualisation Guidelines	21
6.1 Two-Dimensionnal	21
6.2 Three-Dimensionnal	27
7 Re-Sampling	30
7.1 Up-Sampling Procedure	30
7.2 Down-sampling procedure	35
8 Conclusion	38

Disclaimer

The instances provided in this catalogue are based on bathymetric and topographic data made available in the public domain by the General Bathymetric Chart of the Oceans (GEBCO) as of 2022. This organisation operates under the joint auspices of the International Hydrographic Organisation (IHO) and the Intergovernmental Oceanographic Commission (IOC) of UNESCO with the objective of providing the most reliable and publicly accessible bathymetry of the world’s oceans.

Nevertheless, to date, these data originates from a compilation of diverse sources with varying degrees of quality and their limited accuracy (15 arc-seconds) means that they are not suitable for hiking, sailing or any other activity that might jeopardise the personal safety of people, whether at sea or on land.

Furthermore, the GEBCO grid is considered to be an information product due to the use of algorithms to process measured data in order to derive useful values, albeit with uneven levels of precision. In fact, as of 2023, the consensus is that less than 20 % of the oceans has been accurately mapped through direct measurements such as singlebeam and multibeam echosounders technologies while the rest is based on interpolation.

Therefore, in the event of use or misuse of these instances, no responsibility will be endorsed by the authors or, by extension, by GEBCO, IHO, IOC and all those involved in any way in the creation and distribution of these data. To ensure the highest safety standards, people are advised to contact their local authorities for up-to-date charts, advices or any other relevant information.



Figure 2: Planet Earth, made using textures from NASA Visible Earth.

1. Introduction

This catalogue contains 17 700 instances distributed on the world’s shorelines and divided equally between three libraries “A”, “B” and “C”, i.e. 5 900 instances per library. More precisely, an instance correspond to a Digital Elevation Models (DEM) supplied in raster form (a regular grid), i.e. a discretisation of a given Area of Interest (AoI) using cells of regular angular extent¹ and accompanied by bathymetric (depth) and topographic (height) data. These elevation data originate from the 2022 global grid made publicly available by the GEneral Bathymetric Chart of the Oceans (GEBCO)², with a resolution of 15 arc-seconds, i.e. approximately 460 m by 460 m at the equator³. Besides, the GEBCO grid being pixel-centre registered, it means that the elevations, obtained either by interpolation or by direct measurement, correspond to the point at the centre of each cell. In addition, the 5 900 instances of a given library are subdivided into 59 groups of 100 instances, each group corresponding to a single dimension ranging from 10 × 10 cells to 300 × 300 cells, with an increment of 5.

The aim of this document is to provide a guide on how to use this catalogue. To this end, in the first section, we will begin by describing how the catalogue is organised to make it easier to find your way around. Then, in a second section, we will briefly discuss the two types of projection encountered in this catalogue through the documentation and images provided. In the third section, we will formally present the instances, including their intrinsic characteristics, the naming convention adopted and the file format. In the fourth section, we discuss how it is possible to select instances of interest using the available documentation. The fifth section will be devoted to visualisation guidelines for the various instances, and, the sixth section will explain how the data can be re-sampled *ex post facto*. Finally, we give a conclusion in the 7th and last section.

2. Catalogue Organisation

This section aims to detail the hierarchical structure of the catalogue, focusing on the organisation and classification of folders and files, given their significant quantity.

2.1. Overall Organisation

First of all, at the root level, there are four main folders: “data”, “documentation”, “images”, and “license” (refer to Figure 3).

¹In other words, the grid is made of cells arranged at equal intervals of longitude and latitude.

²<https://www.gebco.net/>

³The width narrows as you head further towards the poles.

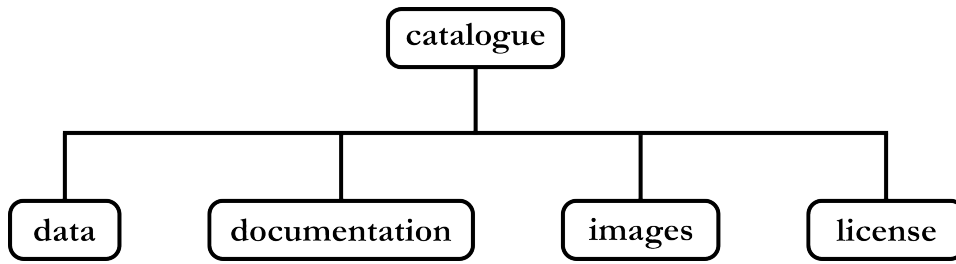


Figure 3: Overall folder organisation.

2.2. Data

Within the “data” folder, the complete set of instances is spread across three sub-folders corresponding to the three libraries “A”, “B”, and “C”. Each of these library folders encompasses 59 sub-folders, one per instance dimension, each containing 100 instances. These instances are delivered in the text-based ESRI ASCII grid format whose extension is “.asc” (see Section 4.3 for more details). This organisational structure is depicted in Figure 4.

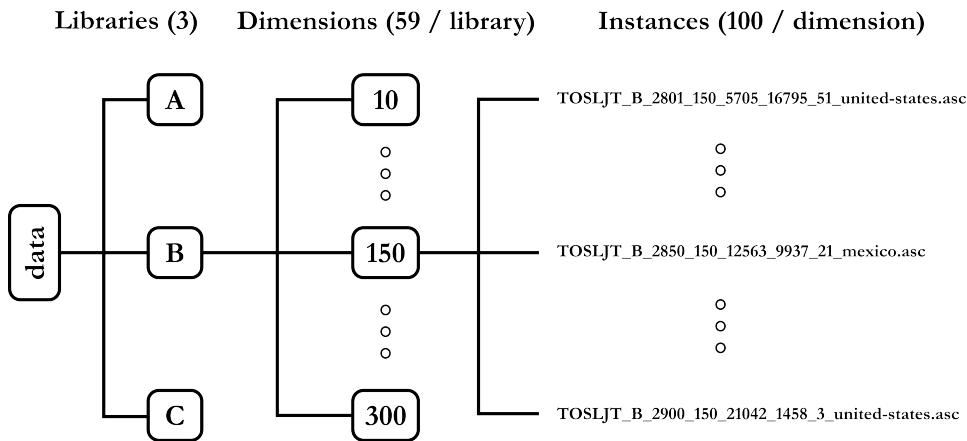


Figure 4: Organisation of the “data” folder.

2.3. Documentation

The “documentation” folder contains the current PDF file “user_guide.pdf” as well as the “library_overviews” sub-folder. The “library_overviews” folder is divided into two sub-folders for the following two projections: “Mercator” and “Equirectangular”. For both of these particular projections and for each library, there is a PDF file presenting a summary of the different instances in a mosaic layout. There is one mosaic for every dimension, each comprising 100 instances (see Section 5.1 for more details about these PDF files). Figure 5 below summarises the organisation of this folder.

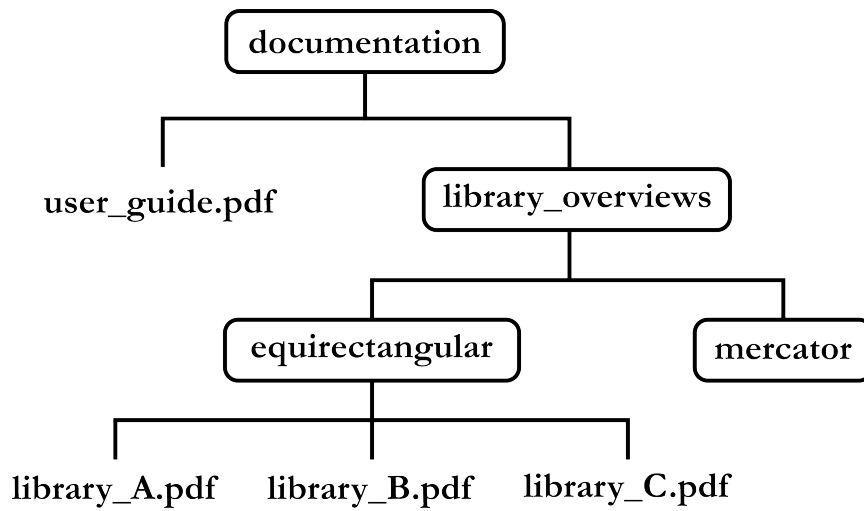


Figure 5: Organisation of the “documentation” folder.

2.4. Images

The “images” folder contains the following three sub-folders: “maps”, “individual_views” and “mosaics” (see Figure 6).

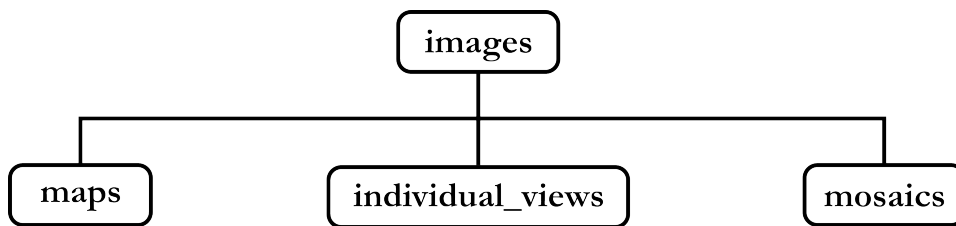


Figure 6: Organisation of the “images” folder.

The first sub-folder, “maps”, includes for both projections and all of the three libraries, a representation of the geographical distribution of 5 900 instances from a given library. The organisation of this subfolder is depicted in Figure 7.

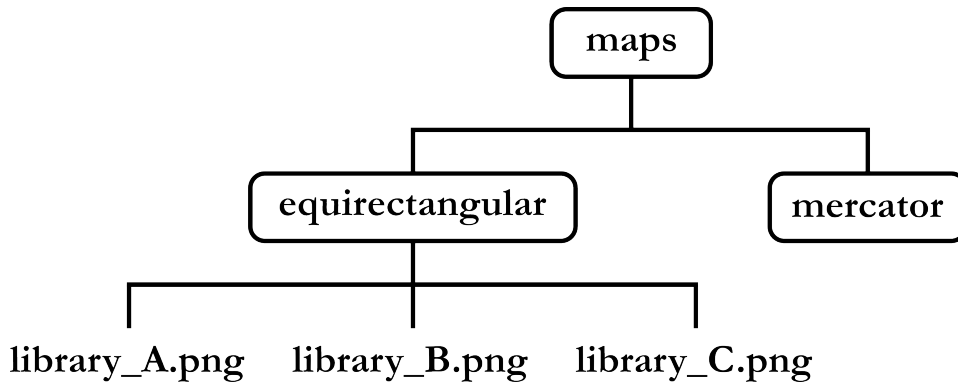


Figure 7: Organisation of the “maps” subfolder.

Then, in the “individual_views” sub-folder, we find for each instance a detailed, individual view, based on satellite imagery and accompanied by additional information such as the geographical extent and a thumbnail of the DEM (Mercator projection). Readers are referred to Section 5.2 for more details on the usefulness of these images. The structure of this sub-folder is shown in Figure 8 below.

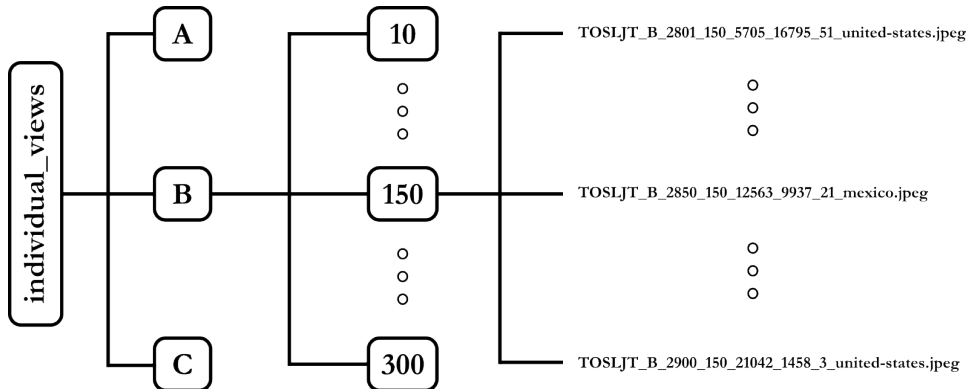


Figure 8: Organisation of the “individual_views” subfolder.

Finally, in the “mosaics” sub-folder, we find for both of the two projections and for each of the three libraries, a synthetic image representing the 100 instances of a given dimension in the form of a mosaic⁴. The arborescence is displayed in Figure 9 hereunder.

⁴These mosaics are used in the PDF files that are part of the included documentation (refer to the documents in “documentation/library_overviews/”).

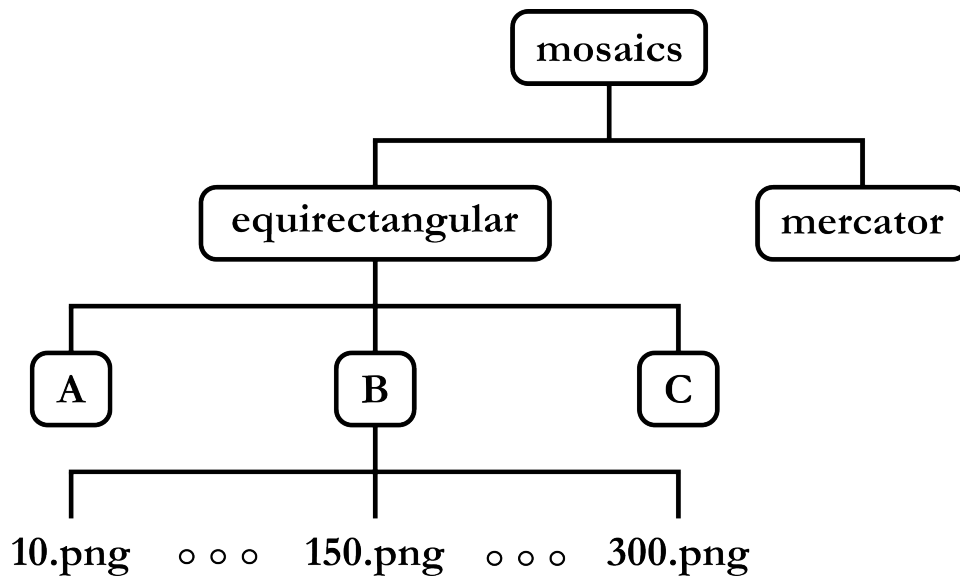


Figure 9: Organisation of the “mosaics” subfolder.

2.5. License

The last folder, named “license”, contains details about the catalogue’s licensing type. Specifically, it includes two files: “license.txt”, which outlines the chosen license, and “CC.by_4-0.txt”, providing the official text of this license. Additionally, the folder contains the logo of the chosen license: “CC_BY.png”. Figure 10 hereafter illustrates this structure.

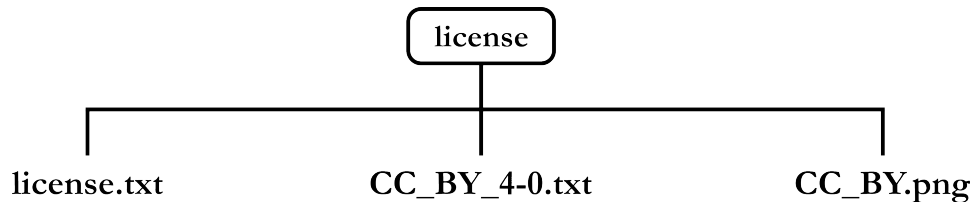


Figure 10: Organisation of the “license” folder.

3. Projections

In this catalogue, we have two co-existing cylindrical projections: Mercator⁵ and equirectangular⁶. The latter is also known as “plate carrée” when

⁵This projection was invented by Gerardus Mercator, a Flemish geographer and cartographer. It was a major innovation for maritime navigation, as it greatly simplified the delineation of maritime routes. Indeed, straight lines on the Mercator projection correspond to trajectories of constant bearing on the globe (i.e. rhumb lines or loxodromes).

⁶This projection is often attributed to Marinus of Tyre, a Greek geographer of the 1st century A.D. Marinus of Tyre is also known for his work which influenced Claudius Ptolemy,

it is centred on the equator. They both have their advantages and disadvantages, and the aim of this section is to outline them in order to guide the reader in choosing one or the other.

3.1. General Remarks

The Mercator projection, which is more familiar and popular than the equirectangular projection due to its predominance in world maps, presents the meridians and parallels as perpendicular straight lines with regular spacing for the first and irregular spacing for the latter. Indeed, this projection stretches the distances towards the poles in order to maintain aspect ratios and thus preserve local angles: the projection is said to be conformal. This means that the shapes are preserved and the visualisation tends to be faithful to reality in this sense. However, the regions close to the poles appear to be much larger than they really are: there is a surface distortion. For example, Greenland appears to be almost the same size as Africa, when in fact it is around 14 times smaller. Furthermore, the Mercator projection cannot represent extreme latitudes (the distortion tends towards infinity since the aspect ratio must be preserved), which makes it unsuitable for a complete representation of the world. Figure 11 below shows a world map in Mercator projection.

The equirectangular projection, which is not as widely used, is one of the simplest to calculate and consists of projecting the meridians and parallels in the form of regularly-spaced perpendicular straight lines. This makes it useful for applications requiring quick and easy transformation of coordinates. It also has the advantage of being able to represent the entire globe, including the poles. However, this projection has two major drawbacks. The first is that it causes significant distortion of surfaces, especially at extreme latitudes. The second is that, unlike conformal projections (such as the Mercator projection), the equirectangular projection does not preserve local angles. This leads to an alteration of shapes and thus additional distortion, especially away from the equator. Figure 12 below shows a world map with a “plate carrée” projection.

3.2. At Local Scales

Now that we have made a few general remarks about the two types of projection, we need to temper our observations in the light of the geographical extent of the instances present in the catalogue. In fact, we are working here on restricted areas, up to a maximum extent of $1.25^{\circ 7}$ in latitude and

whose writings had a considerable influence on cartography for many centuries.

⁷The largest instances are 300×300 cells with a resolution of 15 arcseconds per cell, i.e. approximately 0.00417° , giving us a geographical extent of 1.25° (i.e. $300 \cdot 0.00417$).

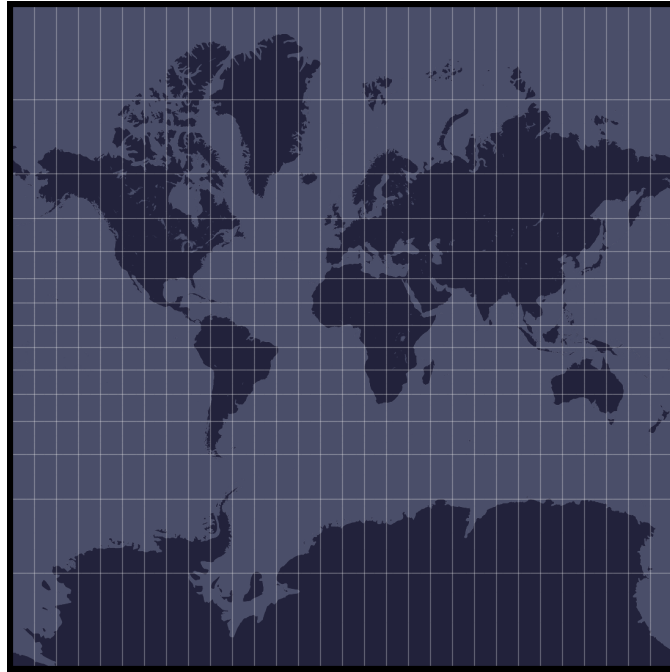


Figure 11: Map at global scale with the Mercator projection (graticule with 12° spacing).

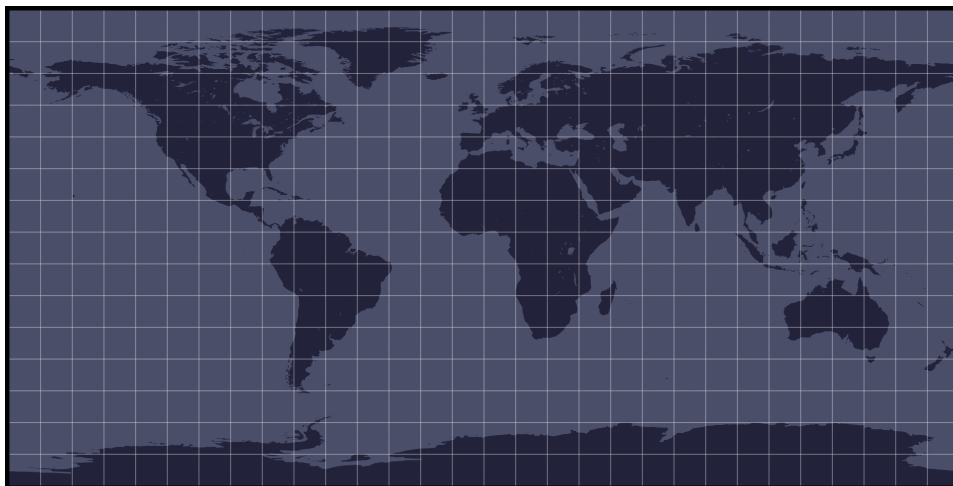


Figure 12: Map at global scale with the “plate carrée” projection, a special case of the equirectangular projection centred on the equator (graticule with 12° spacing).

longitude. As a result, there is very little distortion in the Mercator projection, even for the largest instances, which makes it very faithful to reality, unlike the equirectangular projection, which induces major distortions as you get closer to the poles. That said, the disadvantage of Mercator projection compared with equirectangular projection is that it is more difficult to process digitally, particularly when you want to plot over it. In conclusion, the two projections have their own interests and their own place when it comes to visualising the data provided.

4. Instances

In this section, we are going to take a closer look at the formal description of the instances, including their intrinsic characteristics, naming convention and file format.

4.1. Intrinsic Characteristics

The instances in the catalogue have a number of common intrinsic features. First of all, the different instances have a unique maritime connected component⁸, which was obtained by artificially filling in inland waters such as lakes, ponds and rivers whenever they occurred. Then, all instances have a ratio of maritime cells ranging from 25 % to 95 %, so as to cover a whole spectrum of different coastline geometric configurations. In a specific library and for a particular dimension, the 100 instances are arranged from the one with the smallest maritime area to the one with the largest, i.e. they are ordered based on the increasing number of maritime cells (see Section “Naming Convention” for more information on how they are classified). 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 intended: the configuration of sonar networks. For this application, there was no need to consider inland waters, however, we decided not to lose the information about the cells that had been filled in, so we used a special encoding to preserve them (see Section 4.3 for more details).

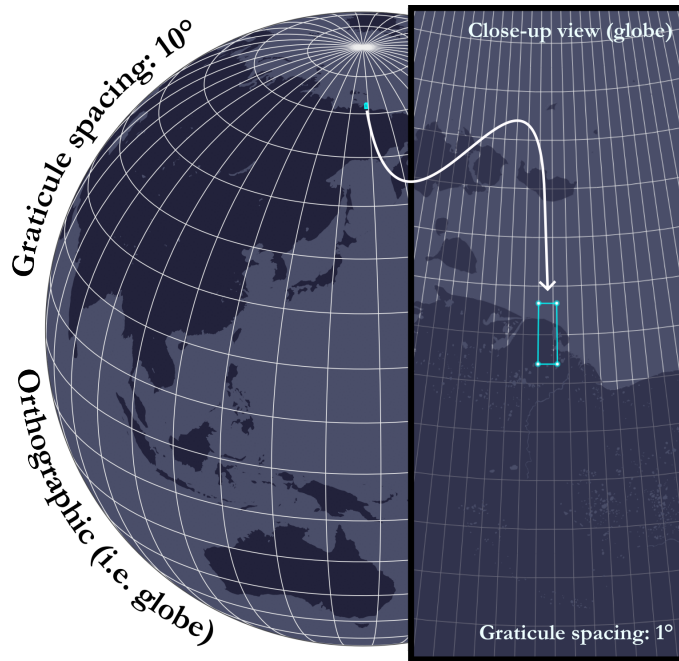
4.2. Naming Convention

An instance is formally referred to under the following full format:

"TOSLJT_library_id.n.m.t.nb-tcc.country",

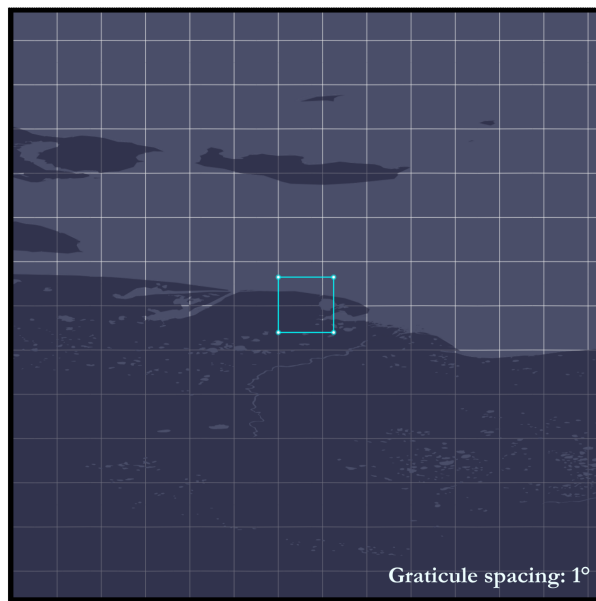
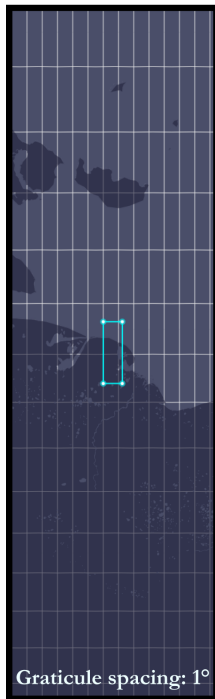
where, respectively, we have:

⁸In graph theory, a connected component is a maximal set of vertices such that each pair of vertices within the set is linked by a path in the graph, and there are no paths connecting vertices in the set to vertices outside of it.



Mercator

Equirectangular



Instance: TOSLJT_A_5808_300_26309_63691_1_russia

Extent (latitude/longitude): 1.25°

Figure 13: Instance “TOSLJT_A_5808_300.26309_63691_1_russia” represented with orthographic (i.e. globe), Mercator and equirectangular (“plate carrée”) projections.

- TOSLJT which refers to the **initials** of the **authors** (Thuillier, Olteanu, Sevaux, Le Josse and Tanguy).
- 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 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**.
- country (*string*) the **country of affiliation**.

For example, the instance with the name

"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.

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

"TOSLJT_library_id",

which can be useful for referring to a set of instances during numerical experiments. For the example above, this would therefore be "TOSLJT_A.0108".

4.3. File Format

Having discussed the file naming convention, we will now describe the format of instances. The latter are delivered in the text-based ESRI ASCII grid format which has the ".asc" extension.

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 extent⁹ and the grid resolution (i.e. the cell size).
- 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:

```

name-instance.asc
ncols <value>
nrows <value>
xllcorner <value>
yllcorner <value>
cellsize <value>
NODATA_value <value>
x1,1 x1,2 ... x1,ncols-1 x1,ncols
⋮           ⋮           ⋮
xnrows,1 xnrows,2 ... xnrows,ncols-1 xnrows,ncols
    
```

} Header

} Data

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:4326¹⁰) 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).

⁹There 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.

¹⁰<https://epsg.io/4326>

- `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 $i \in \llbracket 1, n_{rows} \rrbracket$ and column $j \in \llbracket 1, n_{cols} \rrbracket$, given in the vertical CRS MSL height (EPSG:5714¹¹), 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 14. 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 values must be equal to the number of rows (`nrows`) times the number of columns (`ncols`), i.e. one single value per 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 of use.

Finally, in order to illustrate this in more concrete terms, we show below the contents of the file associated with the instance "TOSLJT_A_0108_15_69_156_1_australia" along with a thumbnail added in the top right-hand corner to make it easier to grasp the instance morphology on an equirectangular projection (square cells of the same size throughout the grid, regardless of the true size at the given latitude).

¹¹<https://epsg.io/5714>

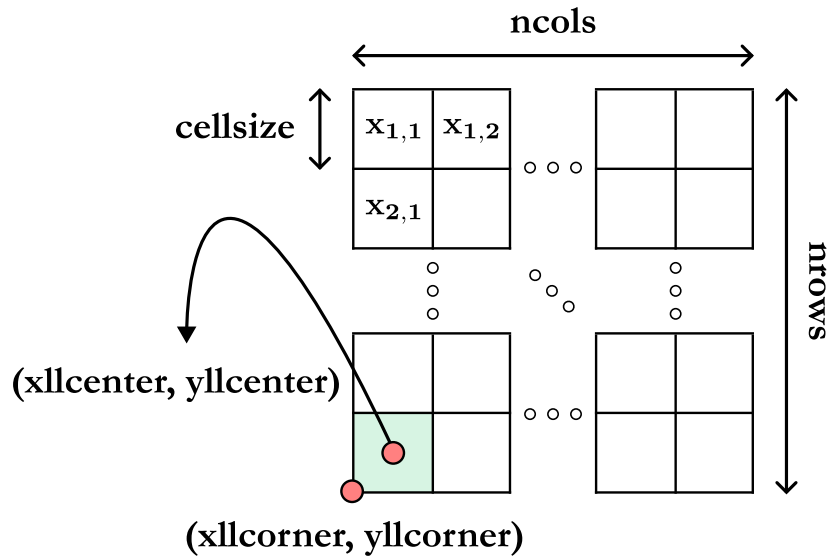


Figure 14: Schematic view of the structure of an ESRI ASCII grid.

```
TOSLJT_A_0108_15_69_156_1_australia.asc
ncols 15
nrows 15
xllcorner 134.10833333333082
yllcorner -32.87500000000025
cellsize 0.004166666667
NODATA_value -32767
-32 -27 -23 -17 -12 5 18 26 36 42 46 57 61 62 66
-34 -31 -26 -20 -14 1 20 28 30 34 43 55 60 56 55
-35 -32 -29 -24 -17 -8 26 41 41 40 43 55 50 42 38
-34 -32 -29 -25 -20 -11 31 45 47 43 36 40 37 31 33
-34 -32 -29 -25 -21 -13 10 43 46 38 26 23 22 26 30
-33 -32 -29 -25 -21 -15 8 43 47 39 27 19 18 22 29
-32 -30 -28 -24 -21 -16 6 38 38 32 23 18 22 24 24
-31 -29 -26 -24 -21 -14 5 30 24 24 18 15 16 16 13
-28 -26 -25 -23 -18 -11 23 20 19 23 22 18 12 9 7
-24 -23 -22 -20 -15 2 17 23 29 29 29 28 20 10 5
-20 -20 -19 -17 2 7 15 21 27 30 30 25 18 9 5
-16 -15 -14 2 8 14 11 11 19 25 29 26 19 9 10
-11 -10 -8 16 15 14 8 6 12 19 25 24 14 11 15
-5 -5 10 16 18 11 5 4 7 15 21 18 13 8 8
3 6 10 14 17 11 6 5 4 10 12 11 8 5 3
```

5. Instance Selection

In this section, we will look at the way in which it is possible to select instances of interest. Indeed, because of the large number of instances, it

might be difficult to identify an instance with a particular topography on the basis of the raw data alone, unless one wishes to retrieve a batch of instances without concern for individual detail. To this extent, we have therefore derived additional documentation and dedicated images to make it easier to navigate through all these instances.

5.1. Library Overviews

First of all, we have a documentation including overviews of the different libraries in the form of synthetic mosaics, on the basis of one mosaic per instance size. Each of these mosaics consists of a compilation of 100 thumbnails, one per instance, for all of the 59 unique sizes available, i.e. from 10×10 cells to 300×300 cells with a step of 5. More specifically, these thumbnails correspond to a bi-colour visualisation of the DEM, with blue for the maritime cells and green for the terrestrial squares. The mosaic is displayed from left to right and top to bottom (i.e. row-major order), with the instance with the smallest unique identifier at the top left corner. As a reminder, within an instance size, instances are ranked in ascending order of the number of maritime cells. Furthermore, we provide this documentation for both Mercator and equirectangular projections (see Section 3 to find out the reasons for this dual existence). Thus, it is now simple to navigate through the entire catalogue. Note also that all the mosaics are available in the "images/mosaics" folder in the form of raw images.

For example, in Figure 15 below, suppose we are in the documentation associated with library A in the equirectangular projection and we are interested in the instance which has the unique identifier 0108¹². Next, go to the "data" folder, then to the "A" folder corresponding to the "A" library, then to the "15" folder corresponding to the 15×15 instance size, and finally retrieve the instance starting with "TOSLJT_A_15_108" (see Section 2 for details of how the folders are organised if necessary). As indicated in Section 4.2, the previous short format allows a given instance to be uniquely and unambiguously identified. In fact, simply with the previous information, we know that it is instance n°108 of library A and of size 15×15 , whose full name is "TOSLJT_A_0108_15_69_156_1_australia".

5.2. Individual Views

After seeing how it is possible to search efficiently for instances from a large pool, we will now see that it is possible to be more refined, in particular with the help of satellite imagery compiled especially for the occasion. Indeed, for each of the 17 700 instances in this catalogue, we have retrieved the best satellite imagery of the areas in question, which we have annotated

¹²Instance used in sections 4.2 and 4.3.

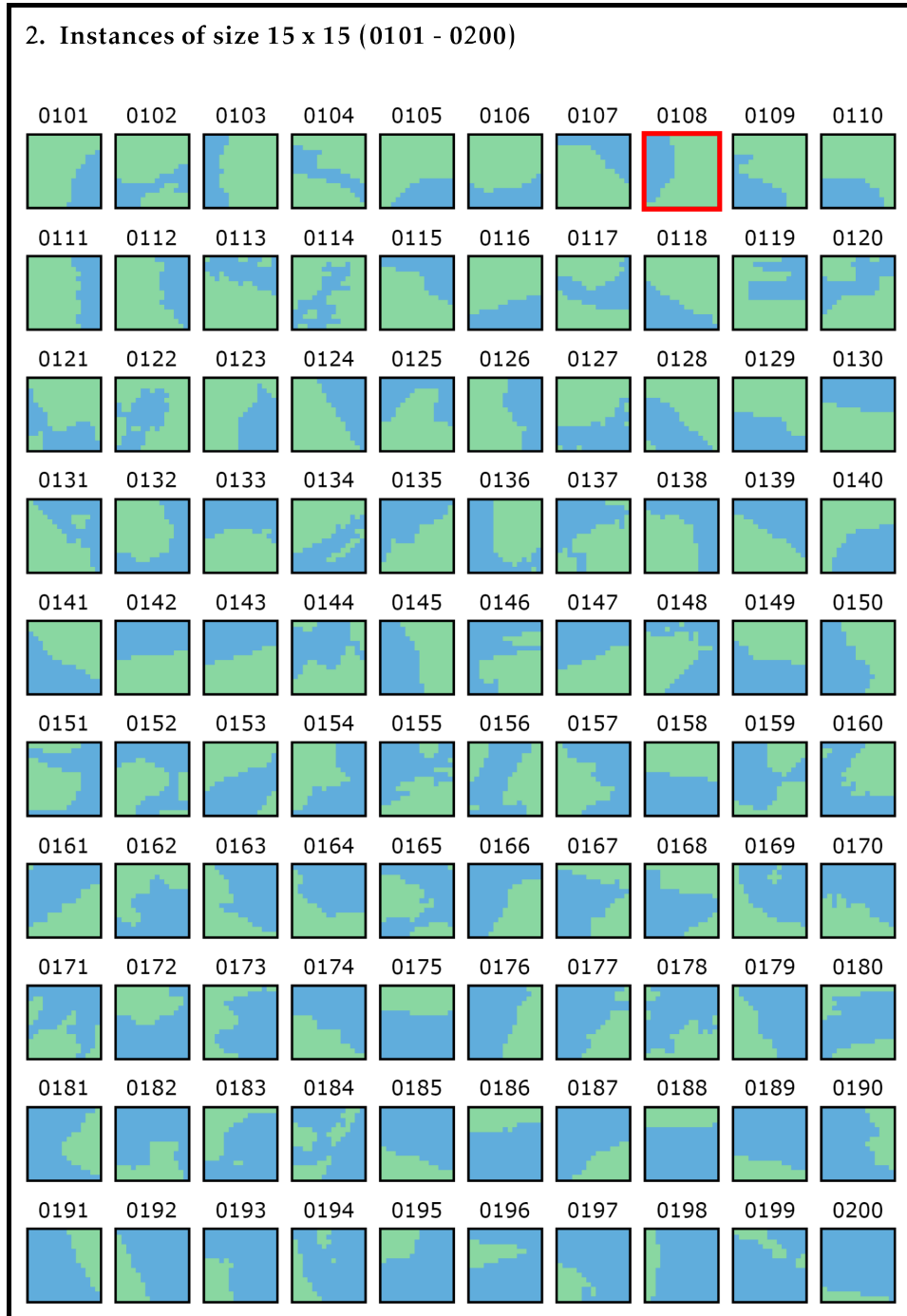


Figure 15: Example of instance selection using the documentation provided, here for library A with the Mercator projection.

and supplemented with a certain amount of information about the instances (DEMs). It should be noted that here we are only providing these images in Mercator projection, and this for two reasons. The first is that, as mentioned in Section 3, for the areas of interest considered which are limited in size, this representation is fairly faithful to reality. The second, these images represent a certain amount of storage space despite all the efforts made to reduce their size as much as possible, so it made little sense to duplicate them with the equirectangular projection.

Hence, for each of these images, the top left-hand corner first shows:

- the originating library,
- the unique identifier of the instance within that library,
- the country in which it is located along with the two-letter country code (ISO 3166-1 alpha-2) as well as the country flag,
- the full name of the instance.

Then, in the top right-hand corner, there is a globe (orthographic projection) showing the rough location of the coastline where the instance is located. The centre of the figure is occupied by a bounding box highlighting the zone covered by the instance, with the west, east, south and north boundaries annotated on all four sides. In the bottom left-hand corner, we have a first group containing the DEM bi-colour thumbnail. Then, in a second group, we have respectively:

- the size of the instance as well as the total number of cells,
- the number of maritime cells with corresponding ratio,
- the number of land cells with corresponding ratio,
- the number of connected terrestrial components.

Next, in a third and last group, we have:

- the extent of the geographical area covered by the instance in decimal degrees,
- the height (km),
- the width (km), at the top,
- the width (km), at the bottom.

Finally, in the bottom right-hand corner are the various attributions for satellite imagery. Figure 16 shows an example of instance "TOSLJT_A_0108_15_69_156_1_australia".

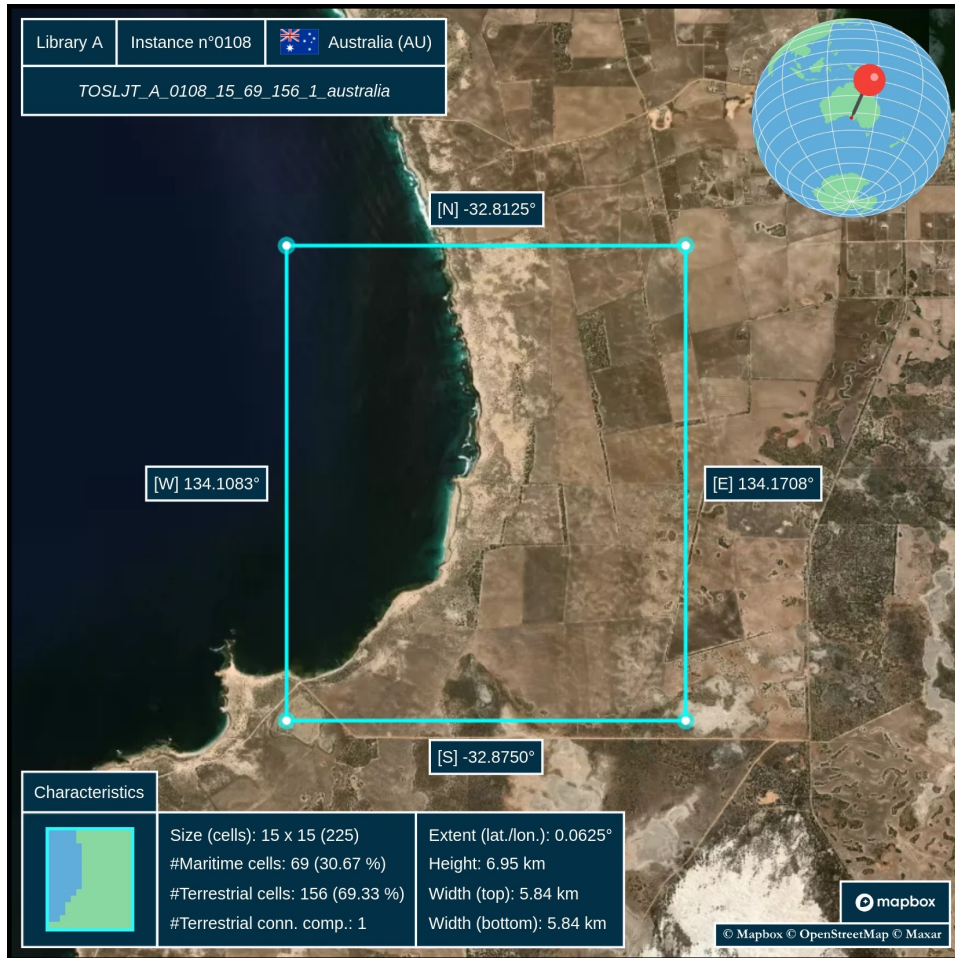


Figure 16: Detailed view of instance "TOSLJT_A_0108_15_69_156_1_australia" with satellite imagery and various annotations (Mercator projection).

6. Visualisation Guidelines

In this section, we will give some suggestions for displaying the various instances provided in this catalogue in the form of 2D heatmaps as well as 3D representations with a set of optional additional layers.

6.1. Two-Dimensionnal

The first thing to do is to have a base layer, i.e. a 2D heatmap with a suitable colour palette. For this purpose, we provide in the repository located at <https://doi.org/10.5281/zenodo.10530296>¹³ a collection of unique palettes specially derived for DEMs and which we invite the reader to consult if necessary. Figure 17 shows some examples with various colour palettes taken from this repository.

Then, once we have a base layer, i.e. a 2D heatmap with a given colour palette, we can add additional layers to improve the visualisation of a given instance. First, it is possible to display in a special way isolated maritime cells that have been artificially filled in to make them more prominent. Indeed, we have kept the actual elevations of the isolated maritime cells using a specific encoding. This means that they can be displayed as true maritime cells, but with a slanting line, for example, indicating that they still count towards the total number of terrestrial cells. Figure 18 shows examples of isolated maritime cells displayed in this way.

It is also possible, and even recommended in some applications, to display the graticule (i.e. the regular grid). This makes easier to assess the coordinates and it is then immediately clear which cell a point belongs to, and therefore by extension the associated elevation. Figure 19 shows examples of graticules.

Then, it may be useful to display level lines such as isobaths (isodepth) and isohypses (isoheight), i.e. contour lines at a constant elevation level. The number of contours to be displayed is then a parameter to be chosen wisely so as not to overload the instance being viewed more than necessary. More precisely, these contour lines reveal the rate of change in elevation in a given area: the closer the lines are to each other, the faster the change in elevation. Figure 20 shows examples of level lines.

An alternative to level lines is hill shading, used as a way of simulating the interaction between a terrain and a given light source. This process

¹³Thuillier, O., Le Josse, N., Olteanu, A.-L., Sevaux, M., & Tanguy, H. (2024). Colour Palettes for Digital Elevation Models [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.10530296>

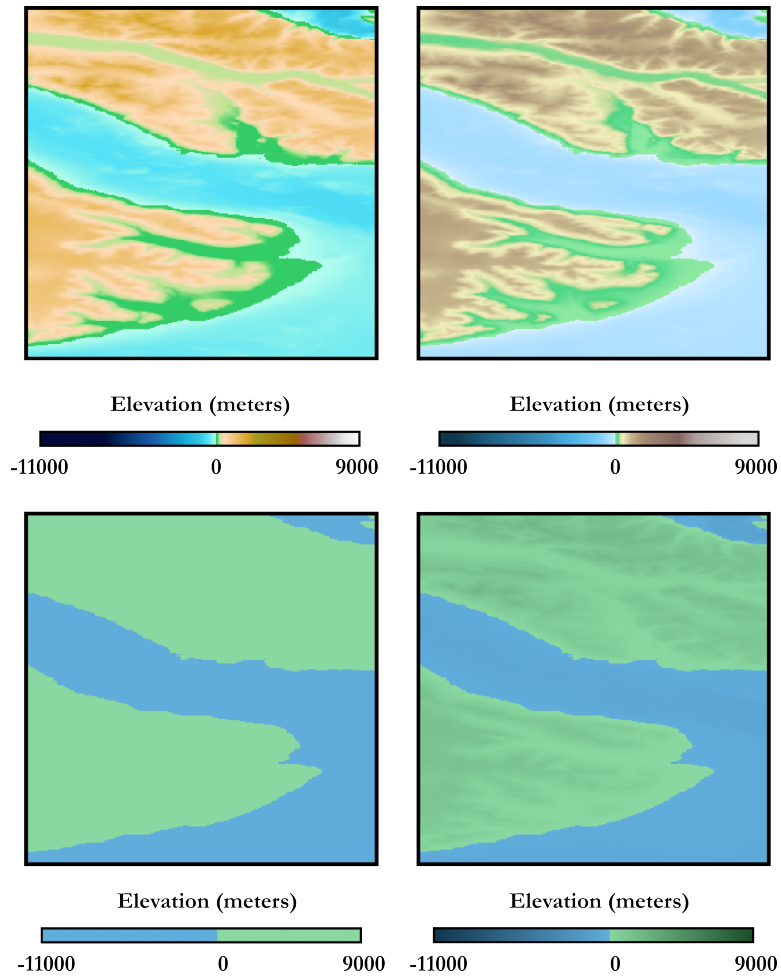


Figure 17: Base layer for visualizing a given instance on a 2D heatmap with a selection of colour palettes.

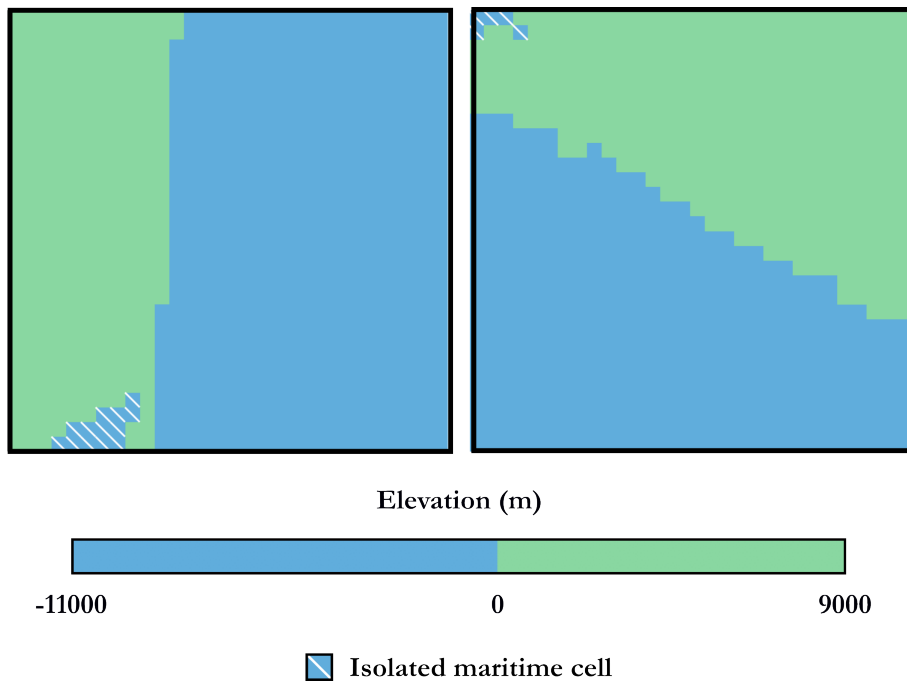


Figure 18: Illustration of the treatment of isolated maritime cells.

can make it easier to recognise the shapes and intrinsic characteristics of a terrain, especially as it is possible to vary the position of the source (more specifically the orientation). Hill shading involves rendering the different cells (pixels) of the DEM based on calculations of the slope and aspect (i.e. orientation) of each of them. An example is shown in Figure 21¹⁴ with different orientations for the light source.

Also, in some colour palettes, it can be difficult to distinguish between shallow maritime cells and low-lying terrestrial cells, as the colours used may be bright. One way of improving this is to add a layer reinforcing this demarcation by adding extra segments of sufficient thickness to all the coastlines. Figure 22 shows examples of accentuated coastline demarcations.

Moreover, when viewing instances over local (restricted) areas, it is possible, and even recommended, to readjust the bounds of a given colour palette in order to provide more contrast, i.e. to make it easier to appreciate local elevation variations. Figure 23 shows concrete examples where palettes A and B have been readjusted to match the bounds of the instance under con-

¹⁴In this example, a better rendering of the shading was achieved through an up-sampling procedure, which is presented in Section 7.

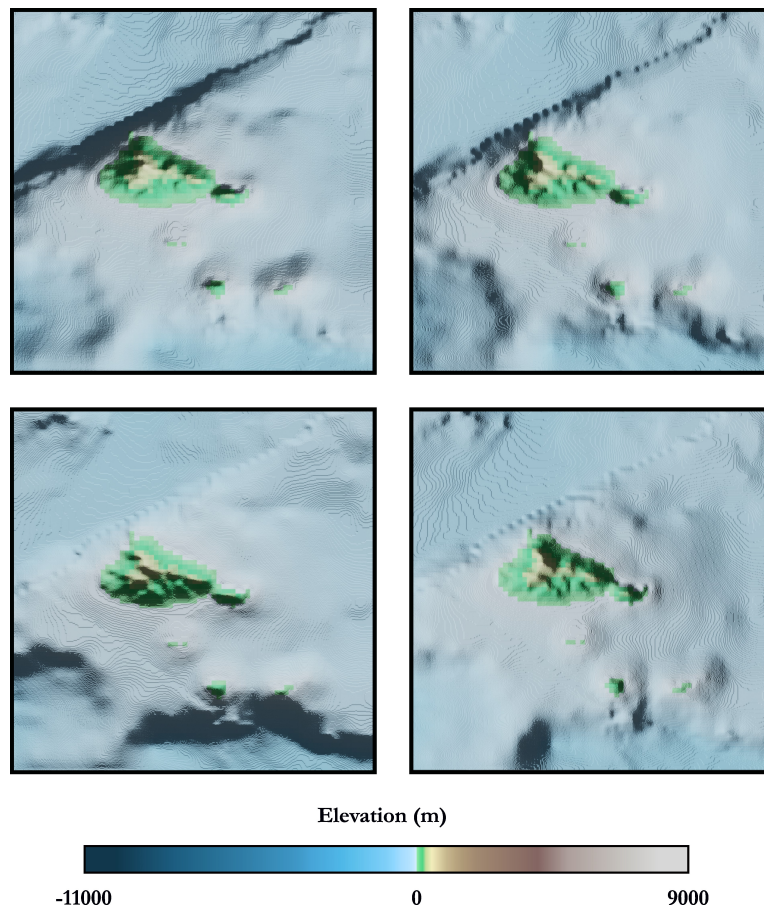


Figure 21: Illustration of the hill shading process with different positions for the light source.

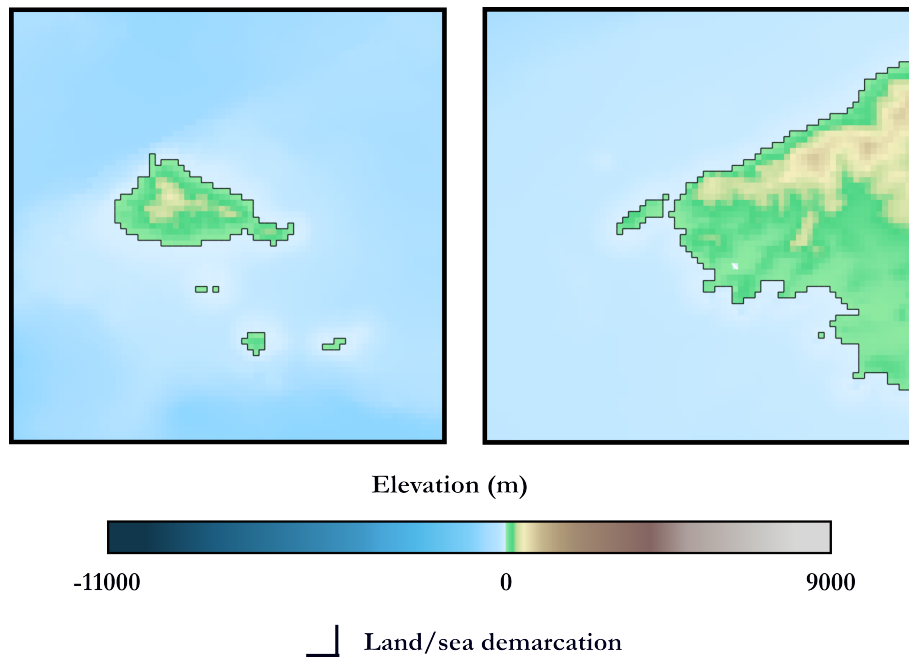


Figure 22: Illustration of the demarcation between terrestrial and maritime cells.

sideration. It is clear that this makes it easier to appreciate local elevation variations, without requiring the use of contour lines. For the illustrations in this article, we systematically use the colour palette “TOSLJT_E_continuous” with this approach.

Finally, Figure 24 shows examples combining several of the layers mentioned above. Note that it is not necessarily a judicious idea to combine all the layers, as this could have the opposite effect and lead to poor visualisation. We therefore need to proceed on a case-by-case basis depending on the area in question, there are no real rules of thumb other than common sense. An extension of this section to visualise instances in three dimensions is later discussed in Section 6.2.

6.2. Three-Dimensionnal

We extend the two-dimensional visualisation presented earlier in Section 6.1. The idea is based on a combination of two elements: a base texture and a texture representing the DEM in grayscale. The base texture can be:

- the base layer only (i.e. heatmap with one of the colour palettes),
- the base layer and a number of additional layers (e.g. demarcations, level lines),

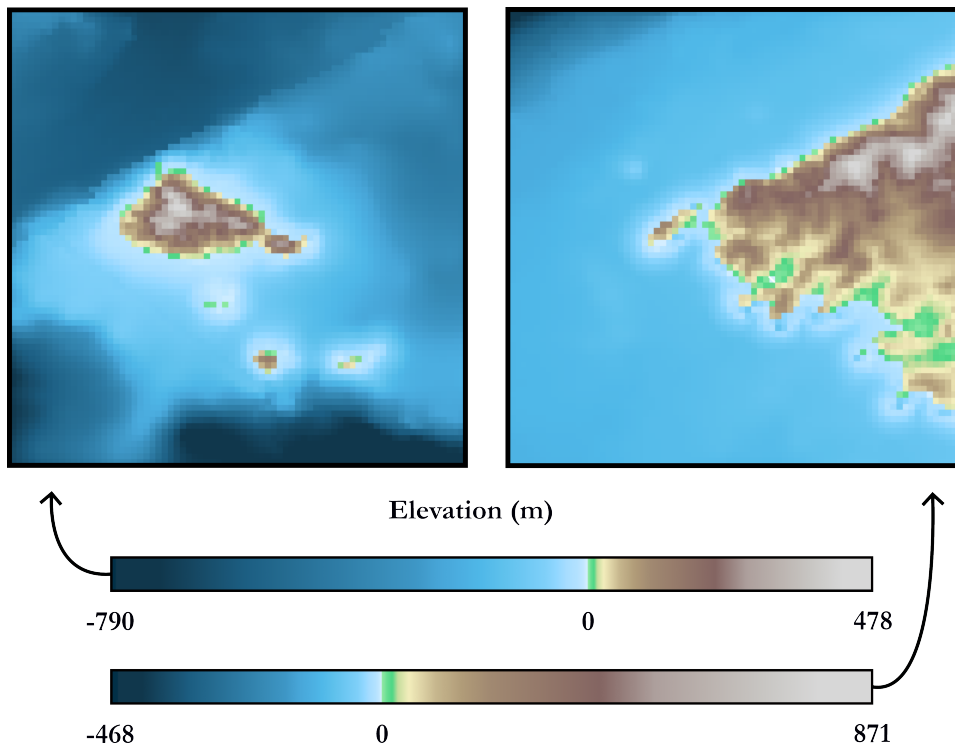


Figure 23: Illustration of bounds readjustment for a given colour palette.

- a satellite view of the area under consideration.

In addition, it is also possible to up-sample the grid of the considered instance if a better resolution is required for the three-dimensional modelling (refer to Section 7 for more details). This gives us a raw three-dimensional model, i.e. just a DEM applied to a base texture. However, it is possible to go further by enhancing this modelling, for example by adding a bedrock and a sea surface. Figures 25 and 26 show the process on a concrete example, with raw modelling in the former and enhancements in the latter.

7. Re-Sampling

In this section, we will look at methods for re-sampling instances (DEMs), which can be useful for specific applications. It can be done either through up-sampling (data creation) or down-sampling (data aggregation).

7.1. Up-Sampling Procedure

A first approach consists of increasing the resolution of an instance by artificially increasing the number of cells (pixels) it contains. The basic scheme is to subdivide each pixel in the parent grid into 4 equal-sized pixels, and then there are two main methods for assigning values (i.e. elevations) to the child pixels: by interpolation or by duplication. The duplication method is the simplest and most naïve way of proceeding: the child pixels are assigned the same elevation as the parent pixel (also known as nearest neighbor interpolation). The interpolation method, on the other hand, is more sophisticated and is the one we will introduce here, since it delivers more convincing results. In particular, we will dwell on bilinear interpolation (4-pixel neighbourhood), although cubic interpolation (16-pixel neighbourhood) is also a possibility. In a nutshell, once all the cells in the grid have been subdivided into 4 child cells, all we need to do is go through all the child cells and perform a bilinear interpolation by considering the 4 parent cells around this child cell. However, there is a special case for child cells at the edge of the grid and therefore having only 1 or 2 parent cells. For these cells, we have chosen the option of adding fictitious margins on each side of the grid. In other words, we have an extra row at the top and bottom and an extra column on the left and right but the latter will only be used for interpolation and subsequently removed. Also, an important point to have in mind when performing interpolations is that the elevations of the cells correspond to the central point (GEBCO grid convention). The procedure is illustrated generically in Figure 27 and applied to a didactic example in Figure 28.

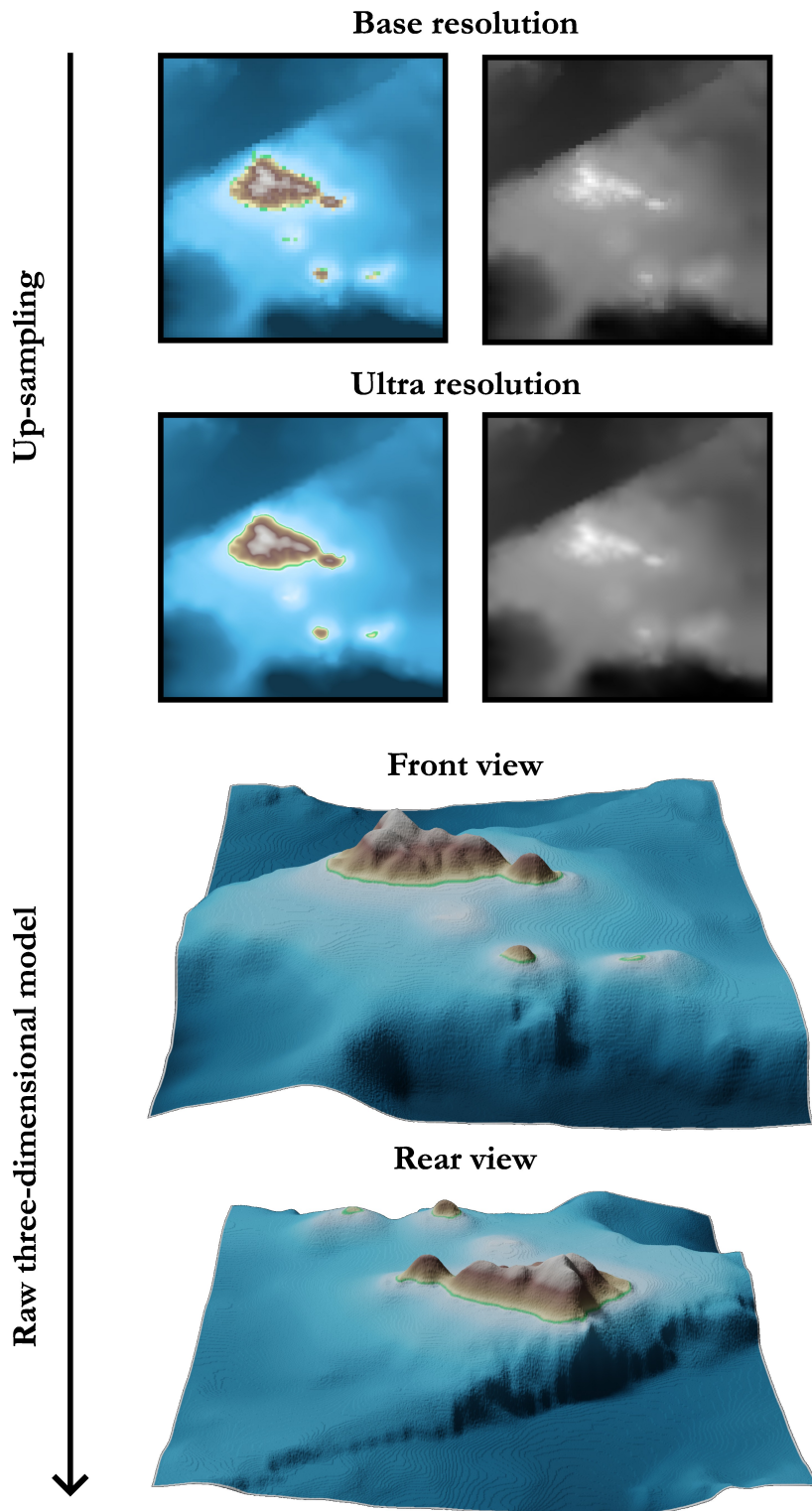


Figure 25: Three-dimensional visualisation process for a given instance: raw modelling.

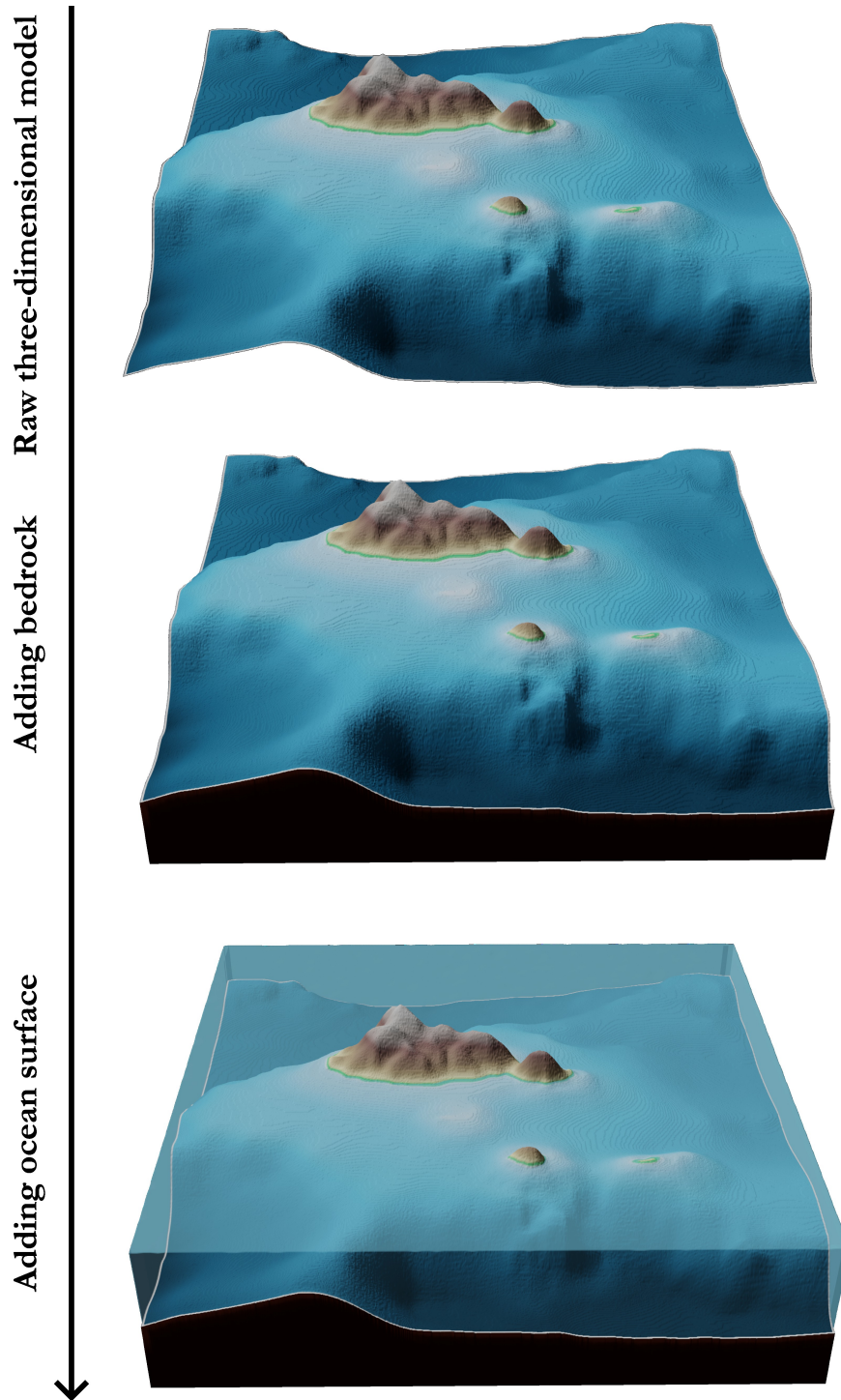
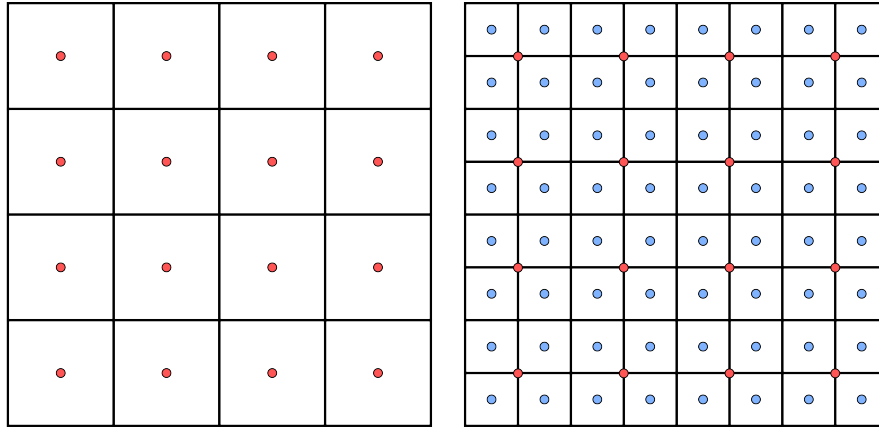
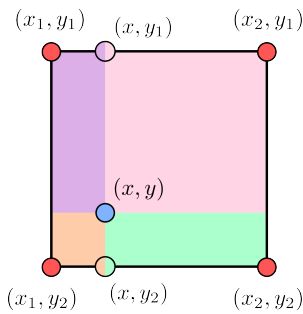


Figure 26: Three-dimensional visualisation process for a given instance: enhancements.



● Elevation of parent grid cells ● Elevation of child grid cells (interpolation required)

Bilinear interpolation



x-direction

$$f(\text{pink dot}) = f(x, y_1) = f(x_1, y_1) + \frac{x-x_1}{x_2-x_1}(f(x_2, y_1) - f(x_1, y_1))$$

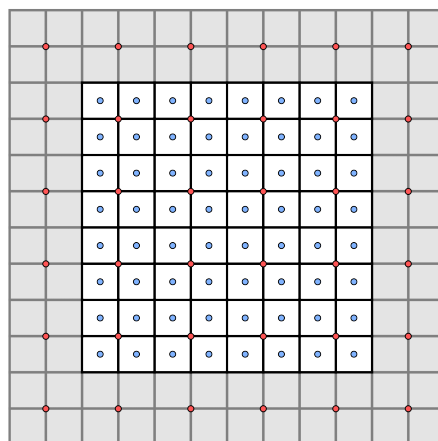
$$f(\text{orange dot}) = f(x, y_2) = f(x_1, y_2) + \frac{x-x_1}{x_2-x_1}(f(x_2, y_2) - f(x_1, y_2))$$

y-direction

$$f(\text{blue dot}) = f(x, y) = f(x, y_1) + \frac{y-y_1}{y_2-y_1}(f(x, y_2) - f(x, y_1))$$

$$f(\text{purple dot}) \quad f(\text{green dot}) \quad f(\text{pink dot})$$

Edge pixels management



Adding extra columns and rows (margins)

Added parent points are assigned the elevation of the nearest existing parent point

Figure 27: Illustration of the up-sampling procedure.

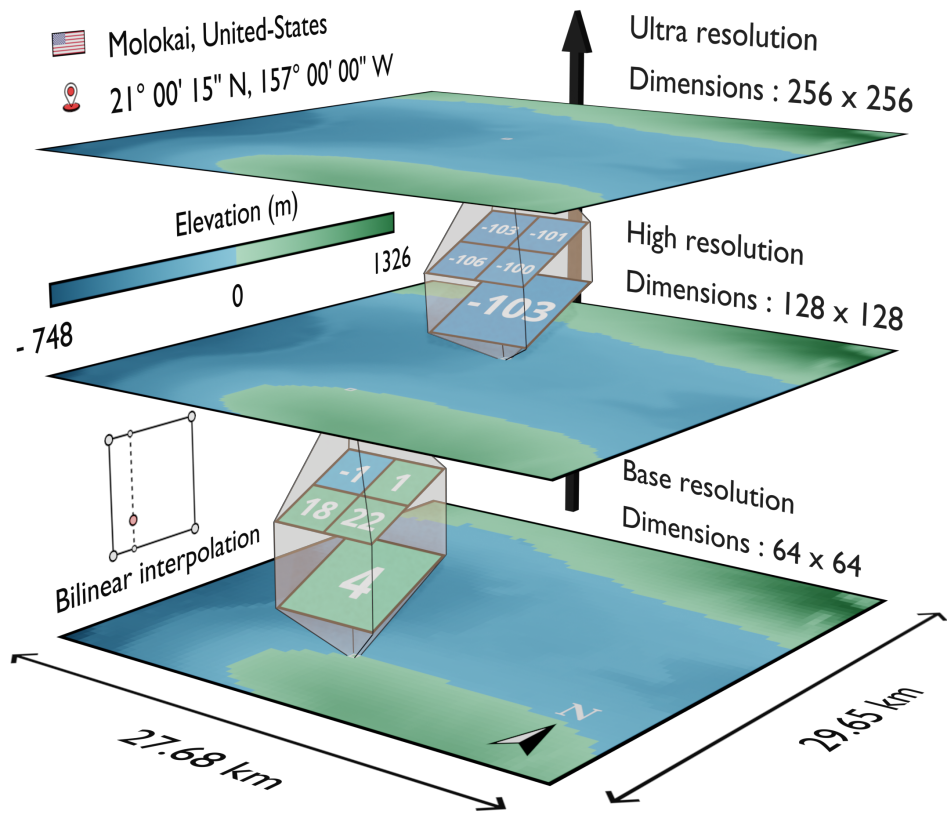


Figure 28: Didactic example of up-sampling.

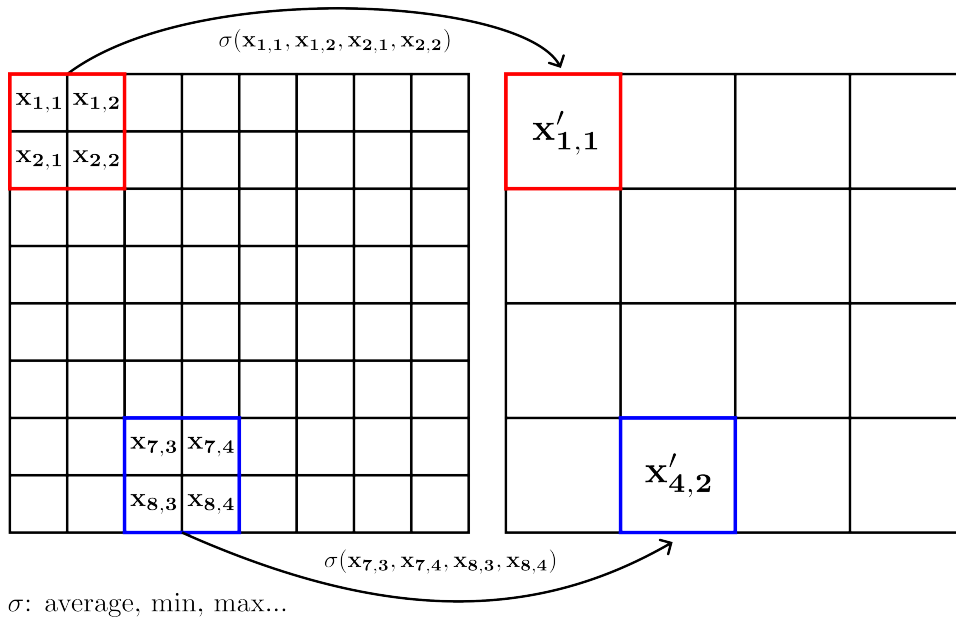


Figure 29: Illustration of the down-sampling procedure.

7.2. Down-sampling procedure

The other alternative is to decrease the resolution of an instance (DEM) by reducing the number of cells (pixels) it contains: this is known as down-sampling. To achieve this, the idea is to aggregate a certain number of cells by dragging a filter of size $n' \times n'$ ($n' < n$) over the entire grid and aggregating the elevations using a given operator (e.g. average, max or min). Note that if the filter extends beyond the grid, it is sufficient to consider only those cells actually covered by the filter (or to copy the last row/column a certain number of times). The procedure is illustrated generically in Figure 29 and applied to a didactic example in Figure 30. An alternative way of down-sampling is to perform bilinear interpolation in the same way as illustrated above in Section 7.1.

Finally, Figure 31 shows a mosaic where 3 reference instances have been iteratively up-sampled and down-sampled by a factor of 2 base resolution on the centre row). In this figure, we have chosen to have one colour palette per instance¹⁵ (i.e. column) in order to get a better appreciation of the re-sampling (depth variations stand out more clearly in this manner).

¹⁵More precisely, it is the same colour palette, but with different end points (i.e. maximum depth and maximum altitude).

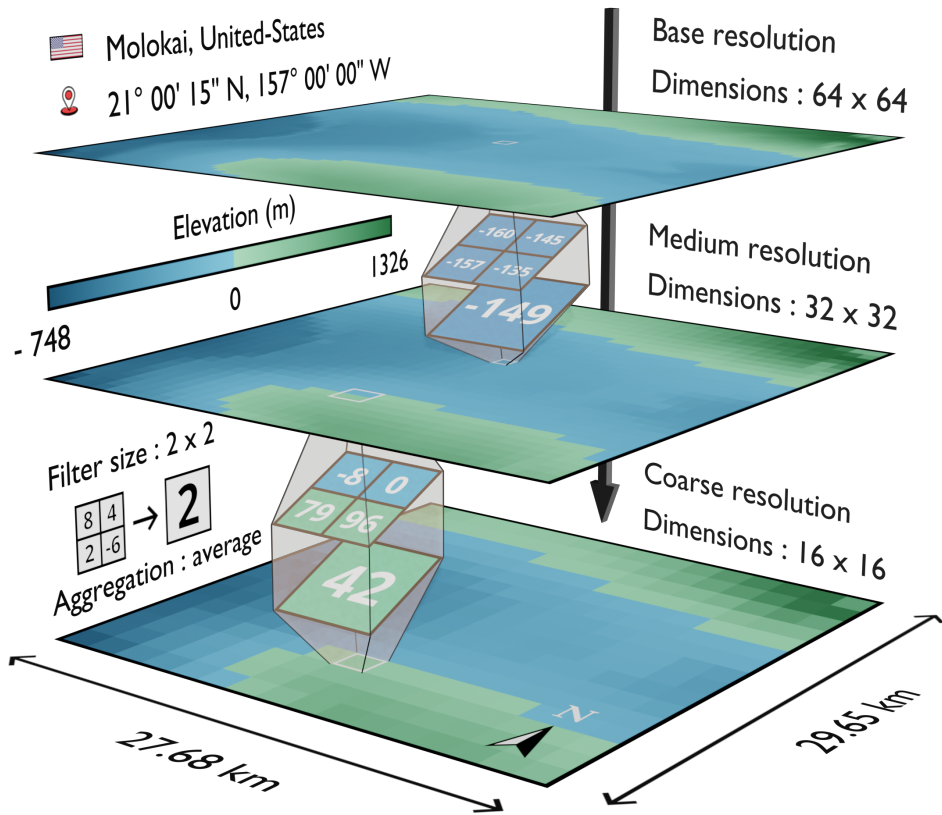


Figure 30: Didactic example of down-sampling.

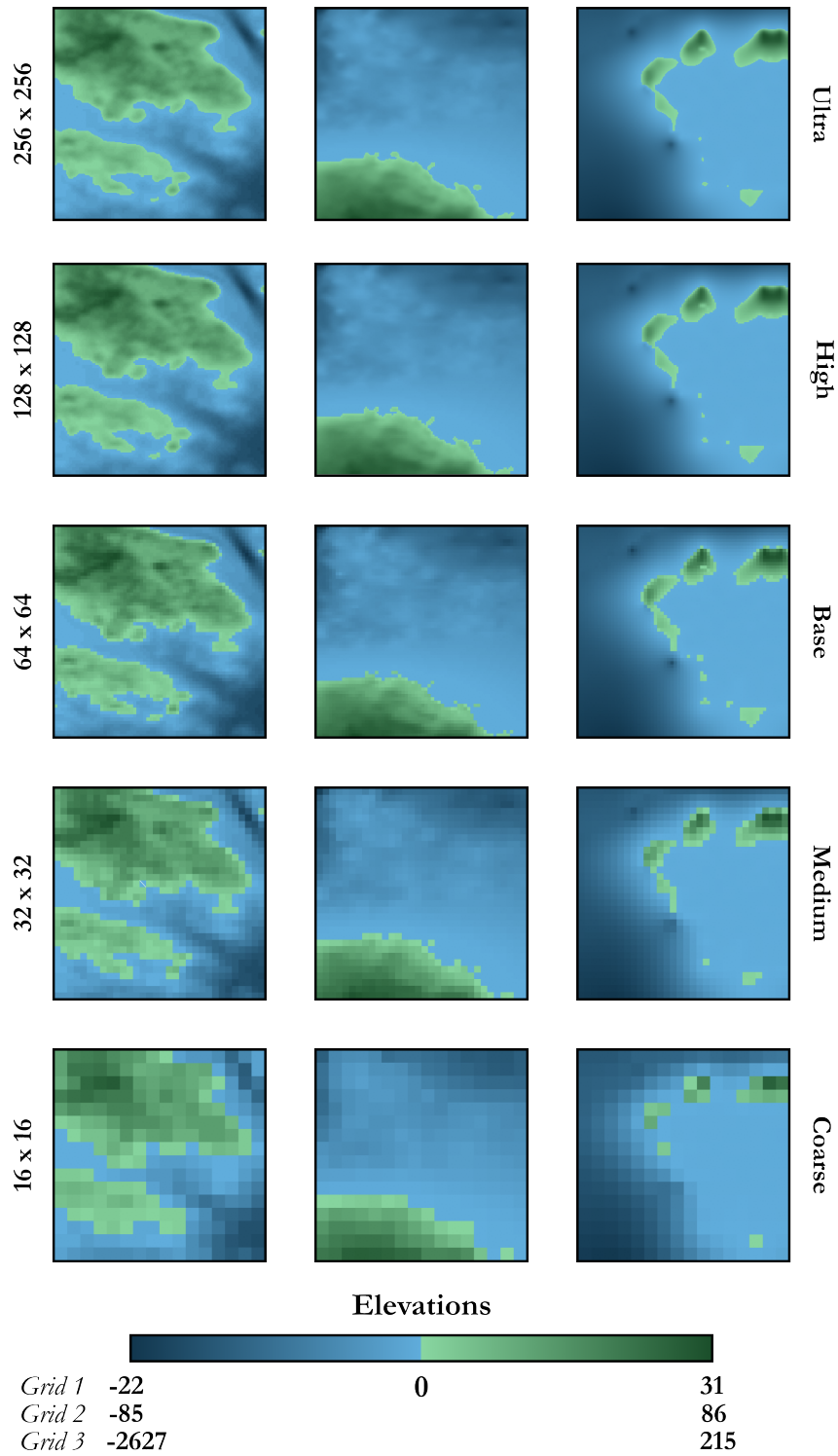


Figure 31: Mosaic of 3 reference instances that have been iteratively up-sampled and down-sampled by a factor of 2 (base resolution on the centre row).

8. Conclusion

To conclude, we provide open-access to a unique catalogue of 17 700 coastal instances, or Digital Elevation Models (DEMs), based on elevation data collected from the GEneral Bathymetric Chart of the Oceans (GEBCO) 2022 global grid and which are distributed uniformly over the Earth's shorelines. These instances, supplied in the form of raster grids, are equally divided between 3 libraries "A", "B" and "C", each containing 5 900 instances. Furthermore, for each library, these 5 900 instances are divided into 59 groups of different sizes according to the number of cells present in the DEM, ranging from 10×10 cells to 300×300 cells in 5-step increments. The resolution of these DEMs is 15 arc seconds, which roughly corresponds to $460 \text{ m} \times 460 \text{ m}$ at equatorial level, with the width gradually decreasing the further away one gets towards the poles. Concerning the intrinsic characteristics of these instances, they all have a single maritime connected component and a ratio of maritime cells lying between 25 % and 95 %, so as to cover a wide range of coastlines configurations. Within a given size, the 100 instances are then ranked in ascending order of number of maritime cells.

Furthermore, we also provide a certain amount of additional documentation to ensure efficient browsing of the entire catalog. The objective is to be able to select instances of interest on the basis of mosaics of two-colour thumbnails representing DEMs (one mosaic per instance size). Additionally, we systematically offer the alternative of visualisation via Mercator or equirectangular projection, depending on the user's requirements. On top of this, we also provide a set of 17 700 images, 1 per instance, composed of satellite imagery duly annotated to visualise the area of the instance under consideration as well as its immediate vicinity. It should be noted that all the productions produced here can be freely reused, provided that the chosen license is respected.

To complement the above points, we also propose methods for resampling DEMs. For example, a given DEM (grid) can be oversampled if greater precision is required, or undersampled if less precision is needed (e.g. for memory reasons). Finally, we give some advice on how to visualise these instances, both in 2 and 3 dimensions.