



University of Zagreb

University of Zagreb

Faculty of Geodesy

Frane Gilić

**DEVELOPMENT OF A GEOSPATIAL  
DATA MODEL FOR GLOBAL LAND  
COVER DATA STORAGE  
AND HANDLING**

DOCTORAL THESIS

Zagreb, 2025



University of Zagreb

University of Zagreb

Faculty of Geodesy

Frane Gilić

**DEVELOPMENT OF A GEOSPATIAL  
DATA MODEL FOR GLOBAL LAND  
COVER DATA STORAGE  
AND HANDLING**

DOCTORAL THESIS

Mentor:

Assoc. Prof. Mateo Gašparović, PhD

Zagreb, 2025



Sveučilište u Zagrebu

Sveučilište u Zagrebu

Geodetski fakultet

Frane Gilić

**RAZVOJ GEOPROSTORNOG MODELA  
PODATAKA ZA POHRANU I  
RUKOVANJE GLOBALNIM  
PODATCIMA O POKROVU ZEMLJIŠTA**

DOKTORSKI RAD

Mentor:

izv. prof. dr. sc. Mateo Gašparović

Zagreb, 2025.

## **General information about the doctoral candidate**

Full name: **Frane Gilić**

Date and place of birth: **September 22, 1994, Split, Croatia**

Current employment: **Research and Teaching Assistant at the Faculty of Civil Engineering, Architecture and Geodesy, University of Split, Split, Croatia**

## **General information about the doctoral thesis**

Title: **Development of a Geospatial Data Model for Global Land Cover Data Storage and Handling**

Number of pages: **144**

Number of figures: **51**

Number of tables: **14**

Number of references: **159**

Scientific area: **Technical Sciences**

Scientific field: **Geodesy**

Scientific branch: **Photogrammetry and Remote Sensing**

Mentor: **Assoc. Prof. Mateo Gašparović, PhD**



## **Doctoral thesis evaluation and defense**

Date of the public defence of the doctoral thesis proposal:  
**July 19, 2024**

Date of the Decision on the approval of the doctoral thesis proposal  
by the Faculty Council of the University of Zagreb Faculty of  
Geodesy: **September 12, 2024**

Date of the Decision on the approval of the doctoral thesis proposal  
by the Senate of the University of Zagreb: **November 20, 2024**

Appointed Thesis Evaluation Committee:

**Prof. Robert Župan, PhD**

**Assoc. Prof. Mario Miler, PhD**

**Assist. Prof. Ivan Medved, PhD**

Date of the Decision on the acceptance of the positive report of the  
Thesis Evaluation Committee on the doctoral thesis by the Faculty  
Council of the University of Zagreb Faculty of Geodesy:

Appointed Thesis Defence Committee:

Date of the public defence of the doctoral thesis:

**UNIVERSITY OF ZAGREB**

**FACULTY OF GEODESY**



Pursuant to the Article 19 of the Code of Ethics of the University of Zagreb and the Decision No. 1\_349\_11 from the Faculty Council of the University of Zagreb Faculty of Geodesy from 26 October 2017 (CLASS: 643-03/16-07/03), it is required to give the “Statement of Originality” of the or doctoral thesis that is being evaluated on the Postgraduate Doctoral Study of Geodesy and Geoinformatics, which serves as confirmation that the thesis is a result of the student’s original work and contains no other sources besides those that are referenced within it.

### **STATEMENT OF ORIGINALITY**

I, Frane Gilić (OIB: 88642078124), hereby declare that this doctoral thesis represents my original work and that I have used no other sources beside those that are referenced within it.

Zagreb, \_\_\_\_\_  
(date)

Signature

\_\_\_\_\_

# Table of Contents

Acknowledgements .....	iii
Information about the mentor .....	iv
Abstract .....	vi
Sažetak.....	vii
List of figures and tables.....	viii
Acronyms.....	xiv
1 Introduction .....	1
1.1 Examples of global land cover products .....	2
1.1.1 ESA WorldCover .....	2
1.1.2 Dynamic World .....	4
1.1.3 Esri Land Cover .....	6
1.1.4 GLC_FCS30D land cover data.....	7
1.1.5 GLAD GLCLUC land cover data.....	9
1.2 Problem statement and research hypotheses .....	10
1.3 Structure of the doctoral thesis .....	14
2 Overview of global grids.....	16
2.1 Direct surface tessellation .....	17
2.2 Map-projection-based tessellation.....	20
2.2.1 Single projection plane .....	20
2.2.2 Multiple projection planes .....	21
2.3 Discrete global grid system (DGGS) .....	24
3 Development of an enhanced DGGS for land cover mapping .....	29

3.1 Deficiencies of current approaches in land cover data storage and handling .....	29
3.2 Choosing a DGGS for global land cover data.....	34
3.3 Modifying the rHEALPix DGGS.....	36
3.3.1 Enhancing authalic latitude calculation .....	43
3.3.2 Modifying shapes of the rHEALPix DGGS cells.....	48
3.4 Evaluating geometric properties of rHEALPix and QPix DGGSs.....	58
3.4.1 Calculating area of DGGS cells.....	58
3.4.2 Calculating perimeter of DGGS cells .....	62
3.4.3 Calculating compactness of DGGS cells .....	63
3.5 Results .....	64
3.5.1 Authalic latitude calculation for the rHEALPix DGGS .....	64
3.5.2 Compactness of the rHEALPix and QPix DGGS cells.....	70
4 Discussion.....	76
4.1 Authalic latitude calculation .....	76
4.2 Comparison of rHEALPix and QPix DGGS geometrical properties .....	79
4.2.1 Compactness of the rHEALPix and QPix DGGS cells.....	85
4.3 Comparison of QPix DGGS and “traditional” global gridding approaches.....	88
5 Application of DGGS on land cover data .....	94
5.1 Methodology .....	95
5.2 Results .....	100
5.3 Discussion and conclusions.....	102
6 Conclusion and scientific contribution .....	106
References.....	109
CV of the doctoral candidate.....	143

## Acknowledgements

## Information about the mentor

Mateo Gašparović, Ph.D. was born in 1986 in Rijeka, Republic of Croatia. He is Associate Professor at the Chair of Photogrammetry and Remote Sensing at the Faculty of Geodesy, University of Zagreb where he is currently appointed as the Vice Dean for Education and Students. He graduated from the Faculty of Geodesy, University of Zagreb in 2009. Assoc. prof. Gašparović obtained the academic degree of Doctor of Technical Sciences in 2015 at the Faculty of Geodesy, University of Zagreb with the doctoral thesis entitled “Radiometric Equalization of Textures on the Photorealistic 3D Models”. His field of expertise is remote sensing algorithm development, remote sensing application in forestry, GIS, spatial data analysis, etc. He has participated in more than 10 scientific projects of the Faculty of Geodesy and is currently a leader of the scientific project "Assessment of the Long-term Climatic and Anthropogenic Effects on the Spatio-temporal Vegetated Land Surface Dynamics in Croatia using Earth Observation Data" (ALCAR) funded by the Croatian Science Foundation (HRZZ), as well as various projects funded by the European Space Agency (ESA). As an author and co-author, he has published more than 30 scientific papers in high-quality scientific journals indexed in the Web of Science Core Collection (WoSCC), 15 of which in Q1 journals according to the Journal Citations Reports (JCR). He was a reviewer of over 150 papers in 25 scientific journals, over 100 of which in Q1 journals. Assoc. prof. Gašparović is a member of the European Geosciences Union (EGU) and the International Society for Photogrammetry and Remote Sensing (ISPRS).

ORCID iD: [0000-0003-2345-7882](https://orcid.org/0000-0003-2345-7882)

CroRIS profile: <https://www.croris.hr/osobe/profil/29967>

## List of selected scientific papers

1. Gašparović, M., Dobrinić, D., & Pilaš, I. (2023). Mapping of allergenic tree species in highly urbanized area using PlanetScope imagery—A case study of Zagreb, Croatia. *Forests*, 14(6), 1193. <https://doi.org/10.3390/f14061193>
2. Gašparović, M., & Singh, S. K. (2022). Urban surface water bodies mapping using the automatic k-means based approach and sentinel-2 imagery. *Geocarto International*, Advanced online publication. <https://doi.org/10.1080/10106049.2022.2148757>

3. Gašparović, M., & Dobrinić, D. (2020). Comparative assessment of machine learning methods for urban vegetation mapping using multitemporal Sentinel-1 imagery. *Remote Sensing*, 12(12), 1952. <https://doi.org/10.3390/rs12121952>
4. Gašparović, M., Zrinjski, M., Barković, Đ., & Radočaj, D. (2020). An automatic method for weed mapping in oat fields based on UAV imagery. *Computers and Electronics in Agriculture*, 173, 105385. <https://doi.org/10.1016/j.compag.2020.105385>
5. Gašparović, M., Zrinjski, M., & Gudelj, M. (2019). Automatic cost-effective method for land cover classification (ALCC). *Computers, Environment and Urban Systems*, 76, 1-10. <https://doi.org/10.1016/j.compenvurbsys.2019.03.001>

## Abstract

Land cover data are regularly delivered in a planar raster geospatial data model. The raster data model was developed at a time when geospatial data handling on the surface of the ellipsoid was demanding and, therefore, all data were projected to the plane. This approach in the modern computer environment is not necessary. Discrete global grid systems (DGGSs) have emerged as a new geospatial data model and georeferencing framework over the past couple of decades. DGGSs represent hierarchically organized grids that enable progressively finer tessellation of the sphere or ellipsoid. The advantage of this data model over the planar raster one is, first and foremost, its better adaptation to the actual shape of the earth's surface. The goal of this thesis is to develop DGGS that is suitable for storing and handling global land cover data and that has quadrangular cells of equal area. Additionally, this DGGS has to exhibit high tessellation geometrical uniformity, i.e., high stability of cells' shapes. We started from already available DGGS implementation, specifically, rHEALPix DGGS, on which we applied two modifications. The first was related to increasing computational accuracy and speed for converting geodetic to authalic latitude and vice versa—a crucial calculation component that enables defining rHEALPix DGGS not only on spheres, but also on ellipsoids of revolution. The second modification was related to applying different approach in defining DGGS grids on ellipsoids and spheres; the approach that we expected would increase stability of cells' shapes. Results confirmed that this expectation was fulfilled. We also quantitatively compared this modified version of the rHEALPix DGGS that we refer to as QPix DGGS, with the traditional approaches of tessellating the earth's surface by raster model georeferenced in projected or geographic 2D coordinate reference system. Results showed that from the three analyzed tessellation approaches, QPix DGGS provided tessellation that has the highest uniformity. Finally, in the last part of the thesis, we successfully applied rHEALPix and QPix DGGSs on land cover data. Therefore, we have overall demonstrated the applicability of DGGSs for storing and handling global land cover data, while eliminating or at least reducing some of the challenges of raster models for such purposes.

**Key words:** authalic latitude, compactness, DGGS, discrete global grid systems, ellipsoid, GIS, global data, land cover, map projections, rHEALPix, sphere



## Sažetak

Podatci o pokrovu zemljišta najčešće se isporučuju u obliku ravninskog rasterskog geoprostornog modela podataka. Rasterski model podataka je razvijen u vrijeme kada je rukovanje geoprostornim podacima na površini elipsoida bilo zahtjevno pa su se svi podatci projicirali u ravninu. Takav pristup u modernom računalnom okruženju nije nužan. Već nekoliko desetljeća, teoretski i praktično, razvijaju se sustavi diskretnih globalnih mreža (DGGS) koji se mogu promatrati kao geoprostorni model podataka, ali kao i model za georeferenciranje. DGGS predstavlja hijerarhijski organizirane mreže koje omogućuju progresivno finiju teselaciju sfere ili elipsoida. Prednost ovog modela u odnosu na ravninski rasterski model je prije svega to što je bolje prilagođen stvarnom obliku Zemljine površine. Cilj ovog rada je razvoj DGGS-a koji je prikladan za pohranu i rukovanje globalnim podacima o pokrovu zemljišta i koji ima četverokutne ćelije jednakih površina. Osim toga, dodatan zahtjev je i da se DGGS treba odlikovati visokom geometrijskom pravilnošću teselacije, odnosno visokom stabilnošću oblika ćelija. Polazeći od rHEALPix DGGS kao već dostupne implementacije DGGS-a, uveli smo dvije izmjene. Prva se odnosila na povećanje točnosti i brzine preračunavanja geodetske u ekvivalentnu širinu i obrnuto, što je ključne komponente koja omogućuje definiranje rHEALPix DGGS-a ne samo na sferama, već i na rotacijskim elipsoidima. Druga izmjena se odnosila na primjenu drugačijeg pristupa u definiranju DGGS mreža na elipsoidima i sferama. Očekivali smo da će ovaj pristup povećati stabilnost oblika ćelija, što su rezultati i potvrdili. Također smo kvantitativno usporedili ovu izmijenjenu verziju rHEALPix DGGS-a, koju smo nazvali QPix DGGS, s tradicionalnim pristupima teselacije Zemljine površine pomoću rasterskog modela koji je georeferenciran u projekcijskom ili geografskom 2D koordinatnom referentnom sustavu. Rezultati su pokazali da od tri analizirana pristupa, QPix DGGS definira teselaciju s najvećom geometrijskom pravilnošću. Konačno, u završnom dijelu rada, uspješno smo primijenili smo rHEALPix i QPix DGGS-ove na podatke o pokrovu zemljišta. Tako smo sveukupno pokazali primjenjivost DGGS-a za pohranjivanu i rukovanje globalnim podacima o pokrovu zemljišta, uz istovremeno eliminiraje ili barem reduciranje nekih od izazova rasterskog modela za takve svrhe.

**Ključne riječi:** DGGS, ekvivalentna širina, elipsoid, GIS, globalni podaci, kartografske projekcije, kompaktnost, pokrov zemljišta, rHEALPix, sfera, sustavi diskretnih globalnih mreža

## List of figures and tables

Figure 1.1 Visualization of the ESA WorldCover land cover map for 2020 (Zanaga et al., 2021) over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 and Sentinel-1 images. Sentinel-2 (red – green – blue composite: band 4 – band 3 – band 2) and Sentinel-1 image (red – green – blue composite: VV – VH – VH/VV; Kumar, 2021) on this figure were both acquired on July 28, 2020. ....	3
Figure 1.2 Visualization of the Dynamic World land cover map (Brown et al., 2022) over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 images. The Sentinel-2 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on July 28, 2020, from which Dynamic World land cover map on this figure was derived. ....	5
Figure 1.3 Visualization of the Esri Land Cover data (Karra et al., 2021) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 images. The Sentinel-2 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on July 28, 2020.....	7
Figure 1.4 Visualization of the GLC_FCS30D land cover data (X. Zhang et al., 2024) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Landsat mission images (Landsat 8 image in this case). The Landsat 8 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on August 26, 2020.....	8
Figure 1.5 Visualization of the GLAD GLCLUC land cover data (Potapov et al., 2022) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Landsat mission images (Landsat 8 image in this case). The Landsat 8 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on August 26, 2020.....	10
Figure 1.6 DGGS constructed by projecting grid from the faces of the cube to the sphere/ellipsoid. ....	14
Figure 1.7 Global grids with (left to right): triangular, quadrangular, and hexagonal/pentagonal cells. ....	14
Figure 2.1 Classification of global grids (adapted from (R. Gibb et al., 2021; J. A. Kimerling et al., 1999)).....	16

Figure 2.2 Platonic polyhedrons. ....	19
Figure 2.3 Different refinements of a parent cell (tick boundary) with children cells (pink boundary): (a) congruent, center- and vertex-aligned 1-to-4 refinement of a triangular cell, (b) not- congruent, center-aligned aperture 3 refinement of a hexagonal cell, (c) congruent, vertex- aligned 1-to-4 refinement of square cell, and (d) congruent, vertex-aligned 1-to-9 refinement of a square cell (adapted from Alderson et al. (2020)). ....	26
Figure 2.4 The first three steps of Morton (upper row) and Hilber (lower row) space-filling curves (adapted from Amiri et al. (2015) and Moon et al. (2001)). ....	28
Figure 3.1 Reprojecting raster. When axes-aligned pixel edges are transformed from one projected CRS into another projected CRS, they are, in general, some arbitrary curves. However, pixel edges of a raster in another CRS are again axes-aligned and thus resampling is required. ...	31
Figure 3.2 Relationship between $30 \times 30$ m quadrangle (e.g., GIFOV of satellite sensor detector) on the WGS 84 ellipsoid and $1'' \times 1''$ geographical grid.....	33
Figure 3.3 Projection of unit sphere and the resolution 0 rHEALPix DGGS cells in the plane by the (0, 0)-rHEALPix map projection. ....	37
Figure 3.4 (0, 0)-rHEALPix map projection of the sphere/ellipsoid and resolution 1, $N_{\text{side}} = 2$ rHEALPix DGGS cells. ....	38
Figure 3.5 (3, 2)-rHEALPix map projection of the sphere/ellipsoid and resolution 1, $N_{\text{side}} = 3$ rHEALPix DGGS cells. The same cells visualized in Robinson projection are available in Appendix A.5. ....	39
Figure 3.6 Resolution-0 (left) and resolution 1 (right) rHEALPix DGGS cells on the ellipsoid for the $N_{\text{side}} = 3$ .....	40
Figure 3.7 rHEALPix DGGS grids and cell identifiers: Japan (left), Split, Croatia (right).....	40
Figure 3.8 Intersecting unit sphere with diagonal planes of the axis-aligned cube to obtain spherical quadrangles.....	49
Figure 3.9 LAEA projection of spherical quadrangle (curved square) and rectangular square of the same area. ....	49
Figure 3.10 Rotating the cube by $45^\circ$ around z-axis. ....	51
Figure 3.11 Mapping curved square to square (Rořca-Plonka projection). ....	52

Figure 3.12 Projection of the unit sphere and the resolution 0 QPix DGGS cells in the plane by the (0, 0)-QPix map projection. ....	53
Figure 3.13 (0, 0)-QPix map projection of the sphere/ellipsoid and resolution 1, $N_{\text{side}} = 2$ QPix DGGS cells. ....	54
Figure 3.14 (0, 3)-QPix map projection of the sphere/ellipsoid and resolution 1, $N_{\text{side}} = 2$ QPix DGGS cells. The same cells visualized in Robinson projection are available in Appendix A.5. ....	55
Figure 3.15 Resolution-0 (left) and resolution-1 (right) QPix DGGS cells on the ellipsoid for the $N_{\text{side}} = 3$ . ....	57
Figure 3.16 QPix DGGS grids and cell identifiers: Japan (left), Split, Croatia (right). ....	57
Figure 3.17 Box plots of the observed calculation errors related to calculating authalic latitude from the geodetic. ....	65
Figure 3.18 Box plots of the observed calculation errors related to calculating geodetic latitude from the authalic. ....	65
Figure 3.19. Spatial distribution of the observed calculation errors related to calculating authalic latitude from the geodetic. ....	66
Figure 3.20. Distribution of observed calculation errors across latitudes when calculating authalic latitude from the geodetic. ....	66
Figure 3.21. Spatial distribution of the observed calculation errors related to calculating geodetic latitude from the authalic. ....	67
Figure 3.22. Distribution of observed calculation errors across latitudes when calculating geodetic latitude from the authalic. ....	67
Figure 3.23. Processing times across latitudes. ....	69
Figure 3.24. Total processing times for different number of executions (i.e., number of input latitudes). ....	70
Figure 3.25 Compactness of rHEALPix and QPix DGGS cells for resolutions from 0 to 5 for a 1-to-4 refinement ( $N_{\text{side}} = 2$ ). ....	71
Figure 3.26 Compactness of rHEALPix and QPix DGGS cells for resolutions from 0 to 5 for a 1-to-9 refinement ( $N_{\text{side}} = 3$ ). ....	72
Figure 3.27 rHEALPix and QPix DGGS compactness ranges for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ). ....	73

Figure 3.28 rHEALPix and QPix DGGS compactness standard deviations for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).....	73
Figure 3.29 rHEALPix and QPix DGGS compactness means for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).....	74
Figure 3.30 Compactness of the rHEALPix DGGS cells.....	74
Figure 3.31 Compactness of the QPix DGGS cells. ....	75
Figure 3.32 Cell-wise absolute differences between cell compactness and mean compactness for the rHEALPix and QPix DGGSs. Mean compactness as well as differences are calculated separately for the rHEALPix and QPix DGGSs.....	75
Figure 4.1 Graph of the <i>arcsin</i> function and its derivative. ....	77
Figure 4.2 Absolute value function of the function given by (5).....	78
Figure 4.3 Distribution of the standardized cell area of the EGG grid. ....	90
Figure 4.4 Box plots of cell compactness for three analyzed global gridding approaches: gridding based on interrupted Goode Homolosine projection, QPix DGGS grid, and equiangular geographic grid. ....	91
Figure 4.5 Cell-wise absolute differences between cell compactness and mean compactness for the EGG (equiangular geographic grid) and QPix grid. Mean compactness as well as differences are calculated separately for the EGG and for QPix grids.....	91
Figure 4.6 Cell-wise absolute differences between cell compactness and mean compactness for the IGH-based (IGH–interrupted Goode Homolosine) and QPix grid. Mean compactness as well as differences are calculated separately for the IGH-based and QPix grids.....	92
Figure 5.1 Pixel in a raster that is georeferenced in the equiangular geographical grid is bound by two meridians with longitudes $\lambda_1$ and $\lambda_2$ and two parallels with latitudes $\varphi_1$ and $\varphi_2$ . ....	95
Figure 5.2 Initial GLC_FCS30D land cover data in WGS 84 CRS (on top) and the same data after reprojecting and resampling in the LAEA projection ( $22 \text{ m} \times 30 \text{ m}$ pixels), and rHEALPix and QPix $N_{\text{side}} = 13$ , resolution 5 grids. Figure shows Sveti Ivan island near Rovinj, Croatia in the transverse Mercator projection with central meridian $13^\circ 37' \text{ E}$ .....	97

Table 1.1 CRSs used for georeferencing analyzed land cover products in Section 1.1 as well as raster pixel sizes and data formats used for data storage.....	12
Table 2.1 Latitudinal zones and corresponding factors for determining angular spacing across parallel (INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1, 2023).....	18
Table 3.1 Number of cells, theoretical areas, and square roots of theoretical areas across rHEALPix DGGS resolutions in the 1-to-4 refinement ( $N_{\text{side}} = 2$ ) for the WGS 84 ellipsoid.....	41
Table 3.2 Number of cells, theoretical areas, and square roots of theoretical areas across rHEALPix DGGS resolutions in the 1-to-9 ( $N_{\text{side}} = 3$ ) refinement for the WGS 84 ellipsoid.....	42
Table 3.3 Unique identifiers of the highest resolution cells used for assessing variability in cell areas because of the calculation errors in conversions between authalic and geodetic latitudes. For each following lower resolution, parents of these cells were used for area calculations. ....	47
Table 3.4 Absolute differences between theoretical and calculated areas for different number of densification points along each edge of the rHEALPix DGGS resolution-1 cells in 1-to-4 refinement (WGS 84 ellipsoid). ....	60
Table 3.5 Absolute differences between theoretical and calculated areas for different number of densification points along each edge of the rHEALPix DGGS resolution-1 cells in 1-to-9 refinement (WGS 84 ellipsoid). ....	60
Table 3.6 Maximum recorded differences between theoretical and calculated area of the ellipsoidal rHEALPix DGGS cells across resolutions in the 1-to-4 refinement. ....	68
Table 3.7 Maximum recorded differences between theoretical and calculated area of the ellipsoidal rHEALPix DGGS cells across resolutions in the 1-to-9 refinement. ....	69
Table 3.8. Total processing times for old statements (both, <i>NumPy</i> - and <i>math</i> -based) and new statements for different number of executions. The number of executions is, in fact, the number of input latitudes on which statements were executed. These latitudes include the whole range from $-90^\circ$ to $90^\circ$ with varying difference between consecutive latitudes. ....	70
Table 4.1 Comparison of rHEALPix and QPix DGGS based on Goodchild criteria (as recorded by Kimerling et al. (1999)). Green tick means that criterion is fully met, red cross that it was not met, either fully or not at all, and for some criteria, number of stars correspond to the level of criterion fulfillment (five stars are maximum).....	84

Table 5.1 Results of built-up area calculations obtained by WGS84-, LAEA-, rHEALPix-, and QPix-based methods. All areas are given in square kilometers. ....	101
Table 5.2 Friedman test statistics and corresponding p-values for three datasets (built-up for 2010, built-up for 2020, and built-up change), each containing six built-up area calculations (one for each country) obtained by four different methods. ....	102
Table 5.3 Wilcoxon test statistics and corresponding p-values for pairwise comparison between WGS84-based method and other three applied methods. Pairwise comparisons were performed separately within each dataset.....	102

## Acronyms

<b>AEAC</b>	Albers equal-area conic
<b>CGLS-LC100</b>	Copernicus Global Land Service Land Cover at 100 m
<b>CNN</b>	convolution neural network
<b>COG</b>	Cloud Optimized GeoTIFF
<b>CRS</b>	coordinate reference system
<b>DEM</b>	digital elevation model
<b>DGG</b>	discrete global grid
<b>DGGRID</b>	Discrete Global Grid
<b>DGGS</b>	discrete global grid system
<b>DL</b>	deep learning
<b>DTED</b>	Digital Terrain Elevation Data
<b>EASE-Grid</b>	Equal-Area Scalable Earth Grid
<b>EGG</b>	equiangular geographic grid
<b>EO</b>	earth observation
<b>EPSG</b>	European Petroleum Survey Group
<b>ESA</b>	European Space Agency
<b>EU</b>	European Union
<b>FAO</b>	Food and Agriculture Organization of the United Nations
<b>FCNN</b>	fully convolutional neural network
<b>GEDI</b>	Global Ecosystem Dynamics Investigation Lidar
<b>GEE</b>	Google Earth Engine
<b>GeoTIFF</b>	Geographic Tagged Image File Format
<b>GIFOV</b>	ground projected instantaneous field of view
<b>GIS</b>	geographic information system
<b>GLAD GLCLUC</b>	Global Land Analysis and Discovery group Global Land Cover and Land Use Change
<b>GLC_FCS30D</b>	Global 30-m dynamic land cover with fine classification system
<b>GOF</b>	grid oversampling factor



<b>GRD</b>	ground range detected
<b>GZD</b>	grid zone designation
<b>HEALPix</b>	Hierarchical Equal Area Isolatitude Pixelation
<b>HTM</b>	Hierarchical Triangular Mesh
<b>IEEE</b>	Institute of Electrical and Electronics Engineers
<b>IGFOV</b>	instantaneous geometric field of view
<b>IGH</b>	interrupted Goode Homolosine
<b>INSPIRE</b>	Infrastructure for Spatial Information in the European Community
<b>IPQ</b>	Iso-Perimetric Quotient
<b>IQR</b>	interquartile range
<b>ISEA</b>	icosahedral Snyder equal-area
<b>L1C</b>	Level-1C
<b>L2A</b>	Level-2A
<b>LAEA</b>	Lambert Azimuthal Equal Area
<b>LC</b>	land cover
<b>LCCS</b>	Land Cover Classification System
<b>LU</b>	land use
<b>LUCAS</b>	Land Use/Cover Area Frame Survey
<b>MGRS</b>	Military Grid Reference System
<b>MMU</b>	minimum mapping unit
<b>NRT</b>	near real-time
<b>OSM</b>	OpenStreetMap
<b>OGC</b>	Open Geospatial Consortium
<b>OpenEAGGR</b>	Open Equal Area Global Grid
<b>O-QTM</b>	Octahedral Quaternary Triangular Mesh
<b>ppt</b>	parts-per-trillion
<b>RF</b>	random forest
<b>rHEALPix</b>	Rearranged Hierarchical Equal Area Isolatitude Pixelation
<b>QSC</b>	Quadrilateralized Spherical Cube
<b>QTM</b>	quaternary triangular mesh

<b>RGB</b>	red, green, blue
<b>ulp</b>	unit in the last place
<b>UN</b>	United Nations
<b>UN-GGIM</b>	United Nations Global Geodetic Centre of Excellence
<b>UTM</b>	Universal Transverse Mercator
<b>WGS 84</b>	World Geodetic System 1984

# 1 Introduction

Before the development of artificial Earth satellites and remote sensing technologies, acquiring data over large areas of the surface of the Earth was a tedious manual process. Data acquired by sensors onboard Earth observation (EO) satellites and the development of methods for processing those data have marked the beginning of an era of the (semi)automatic production of accurate datasets related to the Earth's surface. When discussing data related to the Earth's surface, two terms that are usually first introduced are: land cover and land use. Di Gregorio and Jansen (1998) define land cover (LC) as “observed (bio)physical cover on the Earth's surface” and land use (LU) as “activities and inputs people undertake in a certain land cover type to produce, change or maintain it”. Therefore, LC can be “observed” from the satellite images, either by human visual interpretation or by means of some automatic computerized method, while LU, in general, cannot. However, these terms are often used interchangeably, especially because some land use information can be extracted from satellite images (e.g., vegetated areas used for agriculture). LC and LU data are usually classified in multiple classes, and it is thus common for a single LC/LU dataset to include classes that are related to both land cover and land use. In this case, generally only a fraction of all classes is related to the land use and most of them are associated with land cover. Because of this and for the sake of simplicity, LC/LU datasets are most often referred to as land cover data in this thesis, although it is possible that they also include land use classes.

United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) included land cover and land use as one of the 14 global fundamental geospatial data themes (UN-GGIM, 2019). These themes serve as a foundation that is supporting the United Nations Integrated Geospatial Information Framework (UN-IGIF) and other global geospatial information management initiatives (UN-GGIM, 2019). UN-GGIM considers land cover and land use data as needed to support meeting 11 out of 17 Sustainable Development Goals (UN-GGIM, 2019) which emphasizes its importance and vast demand for such data.

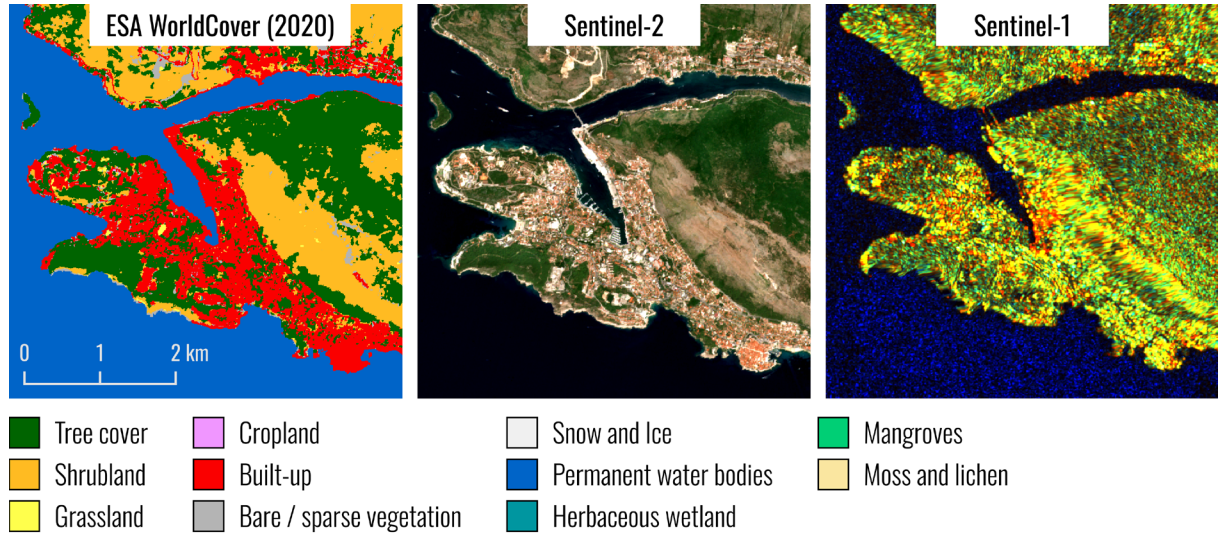
## 1.1 Examples of global land cover products

LC data can be general or thematically focused. General LC data classify (bio)physical cover of the earth's surface in multiple classes, without focusing on particular type of land cover, while thematic LC data are aimed at specific LC, for example, built-up land (García-Álvarez and Nanu, 2022). Various general and thematic LC datasets that have global coverage are becoming available in recent years. Some of the most prominent general LC datasets with global coverage and pixel sizes of less than 30 m are: ESA (European Space Agency) WorldCover (Zanaga et al., 2021, 2022), Dynamic World (Brown et al., 2022), Esri Land Cover (Karra et al., 2021), GLC\_FCS30D (GLC stands for global land cover, FCS for fine classification system, 30 for 30 m pixels size, and D for dynamic; X. Zhang et al., 2024), and GLAD GLCLUC (Global Land Cover and Land Use Change dataset from the Global Land and Discovery group; Potapov et al., 2022). In the following sections, each of these datasets is briefly described.

### 1.1.1 ESA WorldCover

ESA WorldCover is a global land cover map (Figure 1.1) that is available for years 2020 (Zanaga et al., 2021) and 2021 (Zanaga et al., 2022) and is mainly produced from the optical multispectral Sentinel-2 Level-2A (L2A) images and radar Sentinel-1 ground range detected (GRD) products (Van De Kerchove et al., 2020). Additionally, a range of auxiliary datasets were used either during classification or for post-classification application of expert rules, including Copernicus 30-m digital elevation model (DEM; European Space Agency and Airbus, 2022), Ecoregions2017 (Dinerstein et al., 2017), TerraClimate data (Abatzoglou et al., 2018), OpenStreetMap data (<https://www.openstreetmap.org/>), Global Surface Water Explorer data (Pekel et al., 2016), Global Mangrove Watch data (Bunting et al., 2022) and some others were also used (Van De Kerchove et al., 2020, 2022). Zanaga et al. (2020, 2022) state that land cover was classified by CatBoost algorithm into 11 land cover classes (listed on Figure 1.1) that were defined by the UN Food and Agriculture Organization (FAO) Land Cover Classification System (LCCS; Di Gregorio and Leonardi, 2016). ESA WorldCover maps for years 2020 and 2021 do not use completely the same input data and classification procedures, which means that they are not suitable for analyzing changes in land cover since some changes might be a result of these differences and not of a real change in land cover. ESA

WorldCover land cover map with discrete classes is accompanied with the quality raster that indicates per-pixel quality of the input Sentinel-1 and Sentinel-2 observation (Van De Kerchove et al., 2022).



**Figure 1.1** Visualization of the ESA WorldCover land cover map for 2020 (Zanaga et al., 2021) over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 and Sentinel-1 images. Sentinel-2 (red – green – blue composite: band 4 – band 3 – band 2) and Sentinel-1 image (red – green – blue composite: VV – VH – VH/VV; Kumar, 2021) on this figure were both acquired on July 28, 2020.

Regarding the classification quality, Tsendbazar et al. (2021, 2022) report that overall classification accuracy on a continental level ranges from 68% for Australia and Oceania to 81% for Asia and globally is around 74% for 2020 and for 2021 it ranges from 73% for Australia and Oceania and for Eurasia to 82% for Asia on a continental level and is around 77% on a global level. For assessing classification quality they used an adapted validation dataset that was developed for validating Copernicus Global Land Service Land Cover at 100 m resolution (CGLS-LC100; Buchhorn et al., 2020) that has a global coverage and is available as an annual land cover dataset from 2015 to 2019. Venter et al. (2022) also performed validation of the ESA WorldCover land cover data (only for 2020), but with different validation datasets. The first validation dataset they used is the dataset provided by Brown et al. (2021) that was created for validating the Dynamic World land cover (Brown et al., 2022) and the second is LUCAS (Land Use/Cover Area Frame Survey) data (d’Andrimont et al., 2024) for 2018 that are limited to European Union (EU) countries. Based on the first validation dataset, they calculated an overall accuracy of 65%, and based on the second dataset, overall accuracy for the EU countries is 71%. It should be noted however that one of the aims of Venter et al. (2022) was to

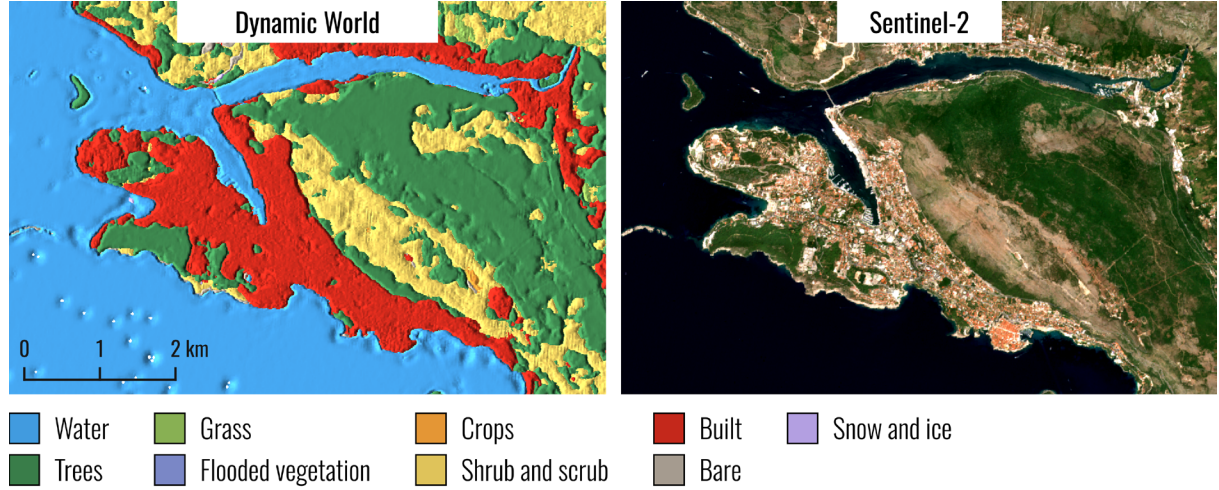
compare classification accuracies of ESA World Cover, Dynamic World and Esri Land Cover datasets so before calculating classification accuracy they have aggregated some of the ESA WorldCover classes so that they correspond to the Dynamic World and Esri Land Cover classes. Zhao et al. (2023) developed a stratified random sampling dataset for global land cover validation and found that overall classification accuracy for ESA WorldCover for 2020 is around 71%.

ESA WorldCover dataset is delivered as a COG (Cloud Optimized GeoTIFF (Geographic Tagged Image File Format)) raster tiles covering  $3^\circ \times 3^\circ$  georeferenced in the WGS 84 (World Geodetic System 1984) geographic coordinate reference system (CRS), referenced under code 4326 in the EPSG (European Petroleum Survey Group) Geodetic Parameter Dataset (EPSG:4326; EPSG, n.d.-c). Rasters tiles with land cover data are compressed with lossless DEFLATE compression (Deutsch, 1996) and pixel values are stored as unsigned 8-bit integers. Each pixel has a dimension of  $0.3'' \times 0.3''$  which at the equator for the WGS 84 ellipsoid (EPSG:7030; EPSG, n.d.-a) corresponds to approximately 9.28 m in east-west direction and 9.21 m in north-south direction.

### 1.1.2 Dynamic World

In contrast to all other LC datasets that are being analyzed, Dynamic World (Figure 1.2) is the only one that is not an aggregated annual product but is produced from every Sentinel-2 image as it is being acquired, provided that it has less than 35% of cloud coverage (Brown et al., 2022). This makes the Dynamic World a near real-time (NRT) land cover product. Classification is performed by a deep learning (DL) model, namely, fully convolutional neural network (FCNN) model that was trained on approximately 24 000 tiles of  $510 \times 510$  pixels (Tait et al., 2021) extracted from the Sentinel-2 images and which were at least 70% labeled by drawing vector polygons that refer to one of the nine classes (listed on Figure 1.2), with a minimum mapping unit (MMU) of  $5 \times 5$  pixels (Brown et al., 2022). A single model obtained in the training phase that does not only account for spectral, but also for spatial characteristics of each class, is used to classify LC on existing and newly acquired Sentinel-2 images. Brown et al. (2022) further explain that the result of the application of the trained DL model on Sentinel-2 images are land cover prediction for each pixel and for each class. Final discrete LC map is produced by taking the class with the highest probability as a final class for each pixel. As can be seen on Figure 1.2, Brown et al. (2022) suggested an innovative approach in visualizing discrete LC map by overlaying a hillshade produced from the class probability values of the corresponding class

with the highest probability. This makes visible features within a specific LC class patch that otherwise would not be visible. Production of the Dynamic World LC data is performed on the Google Earth Engine (GEE) platform (<https://earthengine.google.com/>) were also all products (LC map with discrete classes and per-class predictions) are available.



**Figure 1.2** Visualization of the Dynamic World land cover map (Brown et al., 2022) over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 images. The Sentinel-2 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on July 28, 2020, from which Dynamic World land cover map on this figure was derived.

Brown et al. (2022) assessed the quality of land cover class predictions by the deep learning model and thus the capability of model generalization to new, never-before-seen data, by comparing model predictions with hand-labeled 409 tiles of  $510 \times 510$  pixels (Brown et al., 2021) that were extracted from the training dataset and were not used in the model training phase. Each tile was labeled by three experts and overall accuracy for labels that at least two experts assigned in the same land cover class is found to be 73.80%. Venter et al. (2022) also performed classification accuracy assessment for the Dynamic World data. They first made an annual Dynamic World composite for 2020 by taking the land cover class with the highest annual frequency as land cover class for specific pixel. By using the first validation dataset (the same validation dataset used by Brown et al. (2022)), they calculated an overall accuracy of 72%, and by using the second one (LUCAS dataset) they found overall accuracy to be 66%.

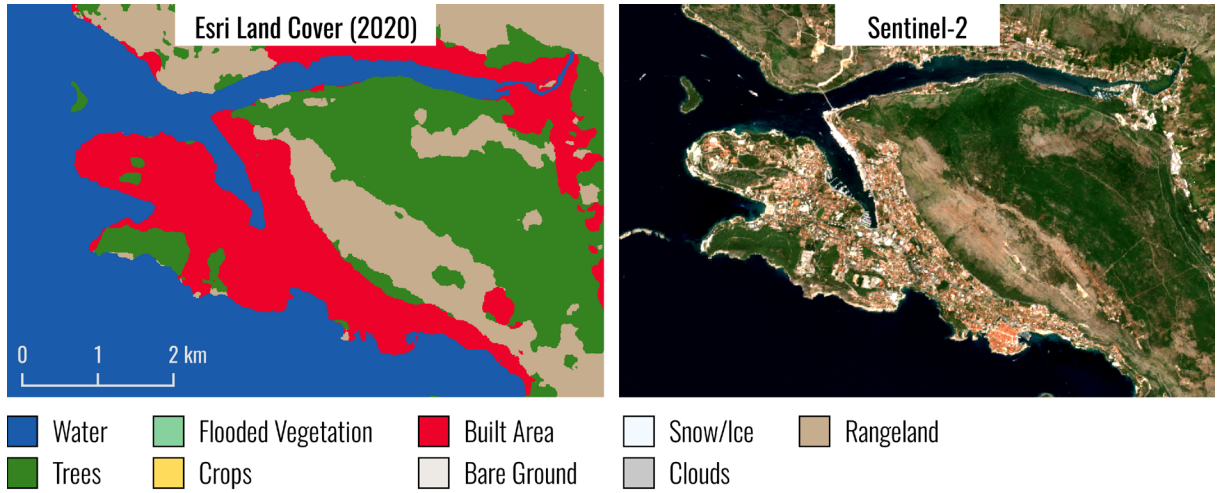
As already mentioned, Dynamic World data are available for download through GEE. Since Dynamic World data are produced from the Sentinel-2 images, the default download option provides Dynamic World tiles in the same CRS that is used to georeference Sentinel-2 images, i.e., one of the 60 projected CRSs based on WGS 84 ellipsoid (EPSG, n.d.-a) and Universal Transverse Mercator (UTM) projection (Gascon et al., 2014). Again, by default settings, raster data are provided as an uncompressed GeoTIFF file with pixel values stored as single-precision (32 bits) floating-point numbers. Pixel size is the same as the pixel size for the highest resolution Sentinel-2 bands, which is  $10 \times 10$  m in a projected UTM-based CRS.

### 1.1.3 Esri Land Cover

Esri Land Cover dataset is produced from the Sentinel-2 images as an annual product from 2017 to, currently, 2024. Land cover is classified in nine discrete land cover classes (Figure 1.3) by applying convolutional neural network (CNN) deep learning model that was trained using the same training dataset used for training the classification model of the Dynamic World dataset (Karra et al., 2021). As Karra et al. (2021) explain, classification model that was applied includes segmentation, which reduces the salt and paper effect that is often present in land cover maps. For some applications, this characteristic of Esri Land Cover data might be undesirable as it might be perceived as too generalized, especially where higher frequency of mixed pixels is present. On the other hand, this dataset is especially interesting for land cover change analysis because the same model is used for classifying land cover for different years and thus interannual classification stability is expected to be high.

Kara et al. (2021) performed validation for their first land cover product for 2020 by using the same 409 labeled tiles of  $510 \times 510$  pixels that were used for validating Dynamic World data and report overall accuracy of 85%. Venter et al. (2022) also calculated accuracy of Esri Land Cover for 2020 using the same dataset, but they obtained accuracy of 75%. This difference of 10% can be partly attributed to the fact that Kara et al. (2021) used only those labels from the validation dataset for which all three experts selected the same land cover class (approximately 27 million pixels according to Brown et al. (2022)), while Venter et al. (2022) used a dataset that consists of 72 million pixels, obtained from validation labels by allowing some level of disagreement between experts' annotations. Based on LUCAS data, Venter et al. (2022) calculated overall accuracy of 63% for land cover data covering EU countries.





**Figure 1.3** Visualization of the Esri Land Cover data (Karra et al., 2021) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Sentinel-2 images. The Sentinel-2 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on July 28, 2020.

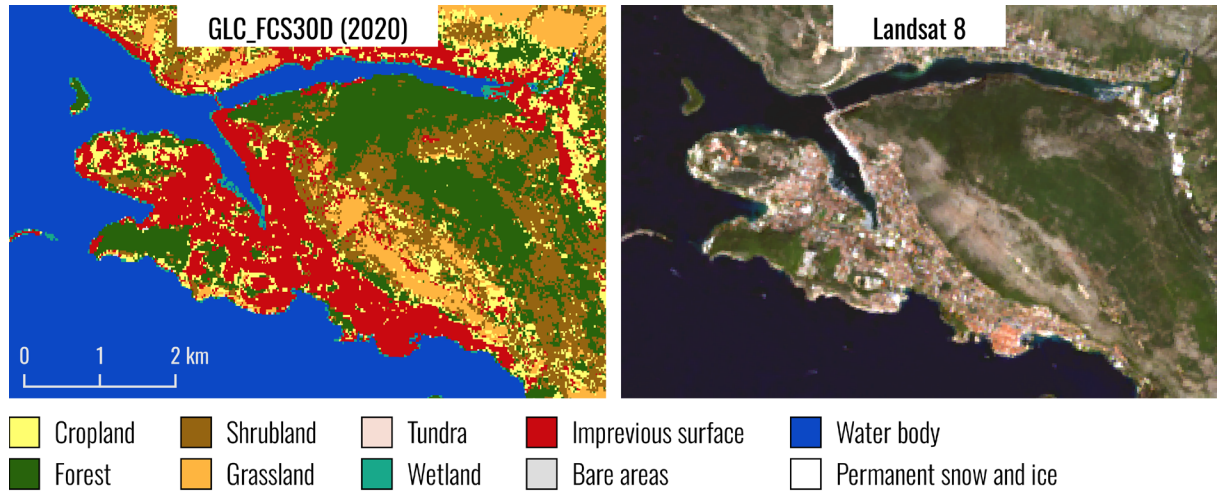
Esri Land Cover data are available for download through Esri’s Living Atlas Land Cover Explorer (<https://livingatlas.arcgis.com/landcoverexplorer>). Data are provided as GeoTIFF tiles with extent corresponding to the grid zone designation (GZD) areas of the Military Grid Reference System (MGRS), covering, in general, 6° across parallels of latitude and 8° across meridians of longitude (NGA Office of Geomatics, 2014). Tiles are georeferenced in the same CRS as Sentinel-2 images – projected CRS based on WGS 84 ellipsoid (EPSG, n.d.-a) and UTM projection, with pixel dimensions of 10 × 10 m. Pixel values are stored as 8-bit unsigned integers and GeoTIFF rasters are compressed with the lossless LZW (Lempel–Ziv–Welch) compression (Dheemanth, 2014).

#### 1.1.4 GLC\_FCS30D land cover data

GLC\_FCS30D is a global land cover dataset that is available for every five years from 1985 to 2000 and annually from 2000 to 2022. Land cover data were generated on the GEE, mainly from the Landsat mission satellite images and additionally from DEM data and an auxiliary datasets for impervious surface and wetlands land cover class that were used in the post-classification phase to increase final classification accuracy for these two land cover classes (X. Zhang et al., 2024). As Zhang et al. (2024) explain, land cover classification algorithm is quite advanced; it includes mechanisms for detecting changes in pixel values over the years and ensuring temporal classification stability (i.e., avoiding false changes), and it also uses locally-adapted (per each 5° × 5° tiles) random forest (RF)

classification models. Land cover is classified by employing hierarchical classification system, with 35 classes on the third hierarchical level, 17 on the second, and 10 (shown on Figure 1.4) on the first one (X. Zhang et al., 2024).

Regarding the classification accuracy, Zhang et al., (2024), based on a dataset they specifically created for independent validation purposes, calculated an overall accuracy of 80.9% for 10 first-level classes and 73.0% for 17 second-level classes, both for 2020. They also used LUCAS data from 2006, 2009, 2012, 2015, and 2018 to assess classification accuracy for EU countries and found that for all years overall accuracy is around 82% for 10 first-level classes. To the best of our knowledge, GLC\_FCS30D can thus be considered as the highest quality long-term global land cover product that is available from 1985 to 2022 in a 30-m resolution.



**Figure 1.4** Visualization of the GLC\_FCS30D land cover data (X. Zhang et al., 2024) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Landsat mission images (Landsat 8 image in this case). The Landsat 8 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on August 26, 2020.

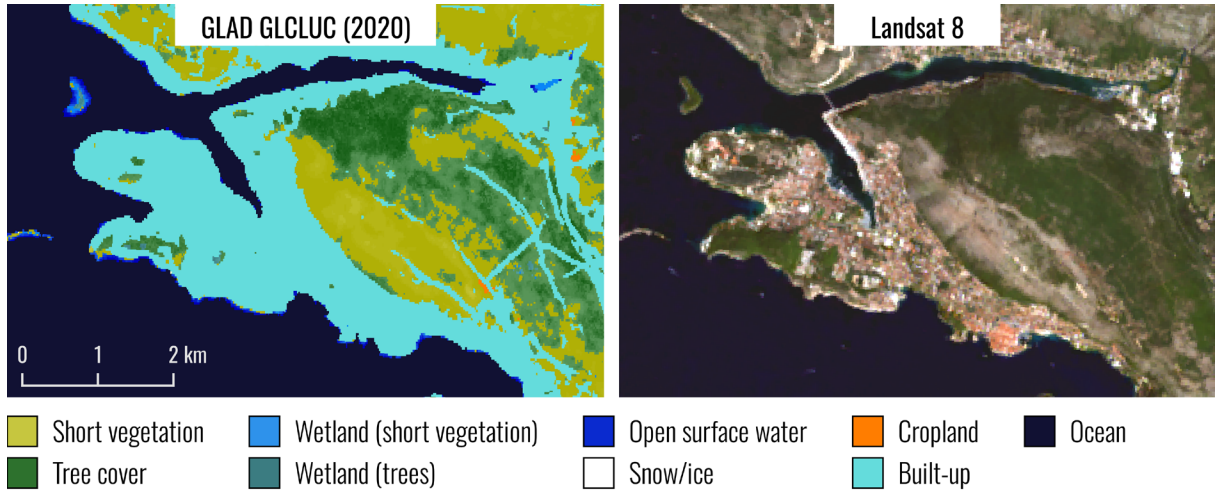
GLC\_FCS30D data are provided by Liu et al. (2025) as  $5^\circ \times 5^\circ$  LZW-compressed GeoTIFF tiles with 8-bit unsigned integer pixel values. Each GeoTIFF raster consists of multiple bands, one for each year for which data are available. Data are georeferenced in 2D geographic WGS 84 CRS (EPSG:4326; EPSG, n.d.-c), with pixel size of approximately  $0.97'' \times 0.97''$ , which at the equator for the WGS 84 ellipsoid (EPSG:7030; EPSG, n.d.-a) corresponds to approximately 30.00 m in east-west direction and 29.80 m in north-south direction.

### 1.1.5 GLAD GLCLUC land cover data

GLAD GLCLUC is global land cover product produced by Potapov et al. (2022), mainly from the Landsat mission ARD (analysis ready data) satellite images, also produced by GLAD (Potapov et al., 2020), and is available for every five years from 2000 to 2020. They used a different classification models for five main land cover classes: tree cover (including wetlands and height of vegetation), cropland, built-up area, water, and permanent snow and ice. Built-up areas, croplands and snow/ice are classified as binary yes/no classes, while vegetation and wetland vegetation are classified into multiple classes; short vegetation into different classes based on coverage percentage, and higher vegetation (i.e., trees) into different classes based on their height, from 3 to 25 m with 1 m resolution. Water areas are classified in eight classes, based on the percentage of year they really are covered with water. This approach to classification gives users flexibility in generating derived land cover maps with discrete classes based on class definitions that are suitable for specific applications. Figure 1.5 shows GLAD GLCLUC land cover data for 2020, however, classes in the legend are aggregated for the vegetation and wetland vegetation.

In addition to Landsat ARD data, for determining tree cover height, Potapov et al. (2022) used Global Ecosystem Dynamics Investigation LiDAR (GEDI) measurements to build a regression tree model. For classifying cropland, they applied ensemble of decision trees, for built-up areas, CNN model trained on OpenStreetMap (OSM) data (<https://www.openstreetmap.org/>), and for water and snow and ice classes, again ensemble of regression trees.

Potapov et al. (2022) performed classification accuracy validation, separately for each of the five main classes and they report overall accuracies that are higher than 97% for the extent of the cropland, permanent snow and ice, and vegetation classes. Permanent water is also classified with high accuracy; 99% user's and 86% producers accuracy, while dynamic water achieved relatively low user's accuracy of 72%, but high producer's accuracy of 95%. Finally, for the extent of built-up areas, achieved accuracy is the lowest. User's accuracy is approximately 64%, and producer's is around 39%. Potapov et al. (2022) attribute this low accuracy to the fact that built-up areas are heterogeneous within a  $30 \times 30$  m Landsat pixel.



**Figure 1.5** Visualization of the GLAD GLCLUC land cover data (Potapov et al., 2022) for 2020 over city of Dubrovnik, Croatia, along with the primary input data used for its creation: Landsat mission images (Landsat 8 image in this case). The Landsat 8 image on this figure (red – green – blue composite: band 4 – band 3 – band 2) was acquired on August 26, 2020.

GLAD GLCLUC land cover data are available for download through GLAD website (<https://glad.umd.edu/dataset/GLCLUC2020>) as  $10^\circ \times 10^\circ$  LZW-compressed GeoTIFF tiles with 8-bit unsigned integer pixel values. Different GeoTIFF files are provided for different years. Data are georeferenced in 2D geographic WGS 84 CRS (EPSG:4326; EPSG, n.d.-b), with pixel size that correspond to the GLAD’s Landsat ARD data (Potapov et al., 2020), which is  $0.00025^\circ \times 0.00025^\circ$ . At the equator for the WGS 84 ellipsoid (EPSG:7030; EPSG, n.d.-a), this pixel size correspond to approximately 27.64 m in east-west direction and 27.83 m in north-south direction.

## 1.2 Problem statement and research hypotheses

From the previous sections, it is clear that the domain of developing global land cover data is under significant advancements. The foundation for producing this data, especially when generating maps for vast regions or the entire planet, rests upon three key elements: freely accessible imagery from remote sensing satellites, advanced classification algorithms, and powerful cloud computing infrastructures (Cuyper et al., 2023; Venter et al., 2022). Each of these components is currently in a state of rapid evolution. Among the most remarkable of these advancements are improvements in the spatial, spectral, and temporal resolution of satellite images (C. Zhang and Li, 2022), as well as the progress of deep learning methodologies. As a direct consequence, the overall quality and reliability of land cover datasets are consistently rising.

An underlying consideration for land cover data is the specific format in which they are organized and spatially referenced. As previously noted, satellite images are in general a primary source for generating land cover data. Since they are typically provided in a raster data model, land cover data are also predominantly distributed in raster form. Emerging formats like Cloud-optimized GeoTIFF (COG) or GeoZarr are being developed to better align with today's large data volumes and cloud computing environments (Abernathy et al., 2021), however they are still built upon the flat raster data model. Despite the substantial technological progress made since the initial development of GIS, many of the foundational design choices that were directly shaped by the technological limitations of that era, such as the significant challenges in handling 3D data, continue to be embedded within the modern GIS (Goodchild, 2018). For example, Tobler and Chen (1986) recognized that traditional analog maps had a dual function; they were a medium for visualizing spatial data and a medium for spatial data storage. In a digital, computerized environment, although these two functions can be separated, they are often treated jointly. They also emphasize that for spatial data storage, the concept of map projections that is essential when dealing with analog maps, is not necessary in GIS. Nevertheless, projecting spatial data from the Earth's surface (i.e., an ellipsoid) onto a plane is still a common practice applied when handling spatial data. This approach tolerates deformations that map projections introduce for the benefit of simpler calculations in the plane (Goodchild, 2019).

Table 1.1 summarizes data formats and georeferencing approach (including pixel sizes) applied in the global land cover product that were analyzed in Section 1.1. It can be concluded that global raster land cover data are georeferenced in geographical 2D CRS, which uses ellipsoidal coordinates (ellipsoidal latitude and longitude) or a projected 2D CRS, which transforms ellipsoidal coordinates into a flat, two-dimensional grid of perpendicular affine coordinates. Despite their different approaches, both coordinate systems ultimately serve the same purpose: to represent the spherical/ellipsoidal surface of the earth in a planar, 2D form.

When working with projected CRSs, a general challenge is that all map projections introduce distortions of area and/or shape as they flatten the Earth's curved surface. Nevertheless, GIS analysts prefer projected CRSs, as computations within a flat plane are typically simpler to execute than those on an ellipsoid. This benefit is especially evident in the case of land cover analysis when the data is

referenced in a projected CRS specifically designed to be area-preserving. Using such a system makes the calculation of areal statistics, which is often the main goal of land cover analysis, exceptionally simple (Thompson et al., 2022). The critical caveat, however, is that if the data is not initially given in an area-preserving CRS, it must be reprojected. This conversion requires a regridding of the raster, a step that inevitably compromises the original data quality and is almost guaranteed to result in either the loss or the replication of initial data (A. J. Kimerling, 2002; Lu et al., 2018).

**Table 1.1** CRSs used for georeferencing analyzed land cover products in Section 1.1 as well as raster pixel sizes and data formats used for data storage.

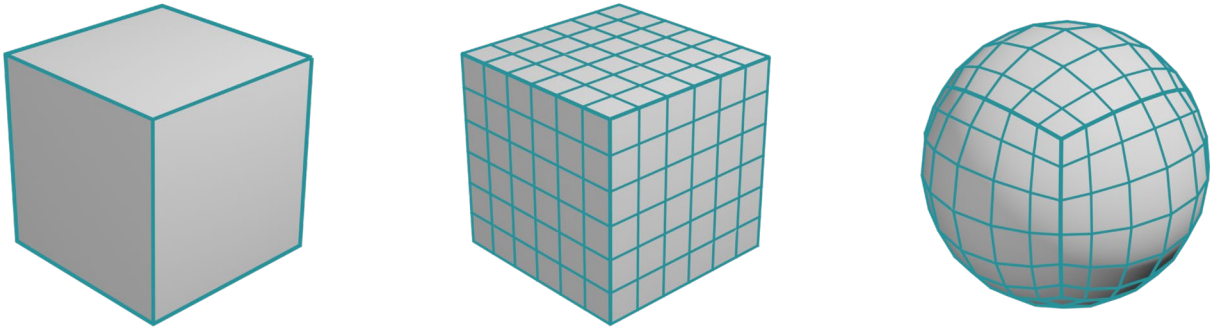
<b>Land cover product</b>	<b>CRS for georeferencing</b>	<b>Pixel size</b>	<b>Data format</b>
<b>ESA WorldCover</b>	Geographic 2D, WGS 84 (EPSG:4326)	0.3" × 0.3" (at the equator approx. 9.28 × 9.21 m)	DEFLATE-compressed GeoTIFF (unsigned 8-bit integers)
<b>Dynamic World</b>	Projected CRS based on WGS 84 ellipsoid and UTM projection	10 × 10 m in projection plane	LZW-compressed GeoTIFF (32-bits floating-point)
<b>Esri Land Cover</b>	Projected CRS based on WGS 84 ellipsoid and UTM projection	10 × 10 m in projection plane	LZW-compressed GeoTIFF (unsigned 8-bit integer)
<b>GLC_FCS30D</b>	Geographic 2D, WGS 84 (EPSG:4326)	0.97" × 0.97" (at the equator approx. 30.00 × 29.80 m)	LZW-compressed GeoTIFF (unsigned 8-bit integer)
<b>GLAD GLCLUC</b>	Geographic 2D, WGS 84 (EPSG:4326)	0.90" × 0.90" (at the equator approx. 27.64 × 27.83 m)	LZW-compressed GeoTIFF (unsigned 8-bit integer)

While raster data in a geographic CRS does link data points to the ellipsoid, the underlying data model is still fundamentally a two-dimensional flat grid (Goodchild, 2019; M. Li and Stefanakis, 2020). This grid is formed by dividing the Earth's surface (modeled as an ellipsoid) with meridians

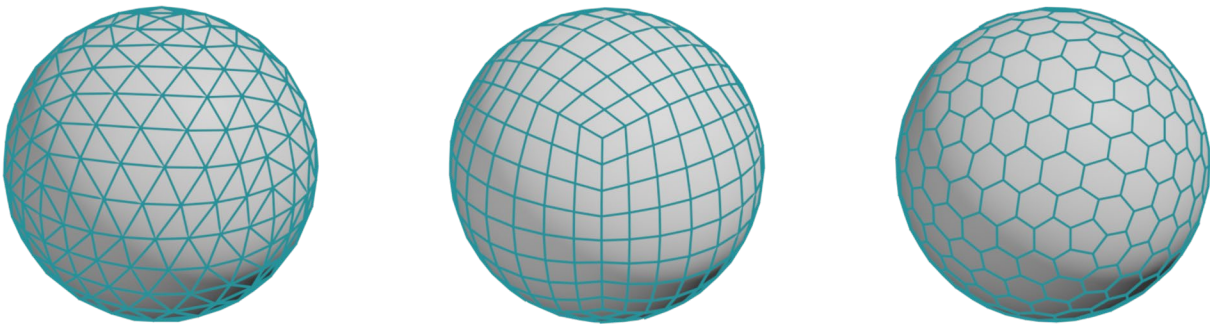
and parallels at fixed angular intervals. This structure is resulting in pixels on the surface of ellipsoid that are not uniform; they vary in both area and shape, largely because the lines of longitude converging towards poles. This poses a problem when integrating data from satellite sensors, which are designed with sensor elements that capture a more-or-less constant square area on the ground. To fit these uniform measurements into the non-uniform cells of the raster, some level of data replication must occur. An established approach to address this issue is to create latitudinal zones that use different east-west pixel sizes (INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1, 2023). The greatest flaw of this approach is data handling issues at the zone boundaries and the need to use multiple raster files for representing area that crosses a zonal boundary. This undermines modern cloud-optimized concepts where all data are usually stored in one file.

Another georeferencing approach that can be applied to land cover data and to geospatial data in general, is based on discrete global grid systems (DGGSs). Although it is not conceptually new, it remains largely unadopted by the GIS community. DGGS is a system of hierarchically organized discrete global grids (DGGs) that cover the entire sphere or ellipsoid that have grid elements (i.e. cells) that are highly regular. Cells are usually triangles, quadrangles or hexagons that are tessellating the earth's surface. Each cell has a unique identifier, meaning that position on the earth's surface can be assigned not in a traditional form of coordinates, but in a form of cell identifier. In theory, dimensions of cell can be adapted to a particular application needs and can be defined so that it, for example, matches the accuracy of a positioning system. More on DGGSs is provided in Chapter 2, thus here we only provide the most common method of constructing a DGGS. Construction starts from sphere- or ellipsoid-inscribed polyhedron, usually a regular one. Projecting edges of the polyhedron, for example, radially, to the sphere/ellipsoid generates initial spherical/ellipsoidal tessellation. Next, on each face of the polyhedron a hierarchical grid is defined and then projected to the sphere/ellipsoid (Alderson et al., 2020). This approach is for the cube (i.e., regular hexahedron) as a base polyhedron shown on Figure 1.6. Figure 1.7 shows a global grid with triangular, quadrangular, and hexagonal cells.





**Figure 1.6** DGGS constructed by projecting grid from the faces of the cube to the sphere/ellipsoid.



**Figure 1.7** Global grids with (left to right): triangular, quadrangular, and hexagonal/pentagonal cells.

The objective of this doctoral thesis is the development of DGGS as a geospatial data model for storage and handling of global land cover data. In terms of objective, research hypothesis are as follows:

1. The discrete global grid system can be used to store and handle global land cover data while eliminating the shortcomings of the traditionally used two-dimensional raster geospatial data model.
2. It is possible to develop a DGGS in which cells are quadrangles of constant area and whose shapes are more stable than that in the currently developed systems.

### 1.3 Structure of the doctoral thesis

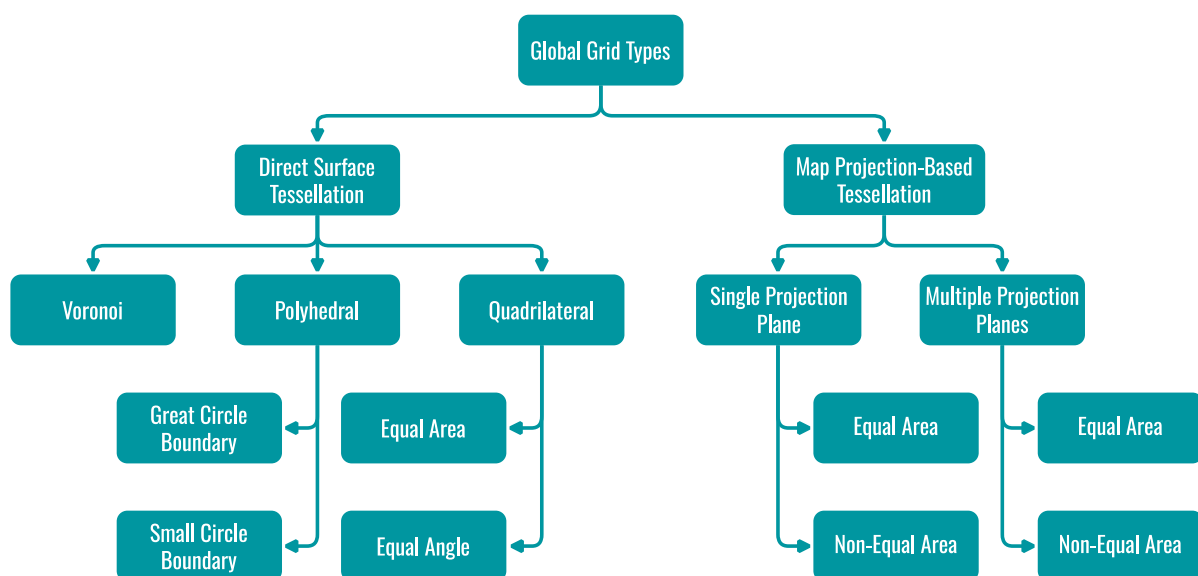
In addition to this introductory chapter, this doctoral thesis is further divided into five main chapters. In Chapter 2 we give an overview of various methods that are applied for tessellating sphere/ellipsoid that are used as models for earth's surface and define main concepts that are related to DGGSs. In Chapter 3, we explain the main deficiencies related to the current land cover data georeferencing approaches and propose enhancements in the currently available DGGS implementations that



should make them more suitable for land cover data. Namely, these enhancements are applied to the rHEALPix DGGS and they include increasing the accuracy of area-preserving mapping from sphere to ellipsoid and vice versa and increasing stability of shapes of rHEALPix DGGS cells. In this chapter, we explain methodology for such enhancements and provide results that enable assessment and comparison between current and enhanced DGGS versions. In Chapter 4, we discuss these results and confirm whether the suggested modifications did in fact make rHEALPix DGGS more suitable for land cover data storage and handling. At the end of this chapter we also perform comparison between modified rHEALPix DGGS version and traditional global gridding approaches. In Chapter 5, we apply rHEALPix DGGS and its modified version for performing land cover change calculations and make comparison with results obtained from the same land cover data but that are georeferenced in geographic and projected CRSs. Finally, in Chapter 6 we draw the main conclusions.

## 2 Overview of global grids

Global land cover data are almost exclusively provided in a gridded/raster form. This is not a surprise since satellite images that are commonly used as the primary input for generating global land cover data are also provided in a raster form. Each raster element, i.e., pixel or cell, corresponds to the specific part of the earth surface to which land cover class is assigned. This grid, or raster, essentially tessellates the earth's surface in order to discretize it into finite number of areal elements. Tessellation approaches can be different and adapted so that they have characteristics that are suitable for specific application. Kimerling et al. (1999) provide classification of various tessellation approaches and they divide them into two categories: those that tessellate sphere or ellipsoid directly and those that use map projections. Slightly modified classification by Kimerling et al. (1999) is provided by Gibb et al. (2021) and shown on Figure 2.1. These global grids are referred to as discrete global grids (DGG), and if a particular tessellation approach establishes a recursive partitioning resulting in a hierarchically organized grids that with finer tessellation in each subsequent hierarchical level, then it establishes a discrete global grid system (DGGS). It should be emphasized here that although DGGS is in broad sense any discrete tessellation with established grid hierarchy, the term DGGS is usually reserved only for tessellations that use sphere- or ellipsoid-inscribed polyhedron for its definition (Section 2.2.1).



**Figure 2.1** Classification of global grids (adapted from (R. Gibb et al., 2021; J. A. Kimerling et al., 1999))

## 2.1 Direct surface tessellation

The most prominent tessellation approach is the latitude-longitude grid (Kelly and Šavrič, 2021) which is formed by a grid of meridians and parallels on the sphere or ellipsoid. This tessellation, according to Figure 2.1, falls under quadrilateral equal angle direct surface tessellation group and its projection to the plane with the equidistant cylindrical projection having an equator as a standard parallel (sometimes referred to as lat/long projection or Plate Carrée projection; Kerkovits, 2023, p. 108) forms a rectangular grid in which cells are squares if angular spacing between meridians and parallels is the same (i.e., equiangular grid). Since raster data model is based on a rectangular grid and Plate Carrée projection of latitude-longitude grid is also rectangular, this is the main reason for popularity of this tessellation. The main disadvantage of equiangular latitude-longitude grid is the fact that cells do not have the same area (Mahdavi-Amiri et al., 2015b). The area is the largest for the cells near the equator and is the smallest for cells near the poles. To account for this issue to a certain degree, it is possible to define latitudinal zones with different angular spacing between meridians. This approach is adapted, for example, by the INSPIRE (Infrastructure for Spatial Information in Europe) Technical Guidelines on Geographic Grid Systems (INSPIRE Temporary MIWP 2021-2024 subgroup 2.3.1, 2023) which refers to this grid as a zoned geographical grid. INSPIRE guidelines define five latitudinal zones in each hemisphere, as shown in Table 2.1. Cell (i.e., pixel) angular size across meridian is constant and angular size across parallel is for each zone obtained by multiplying spacing across meridian by factor in Table 2.1. This way, cell size shrinking in east-west direction by going polewards from the equator because of meridian convergence is mitigated by widening cell sizes in this direction. This approach is taken from global digital elevation model (DEM) datasets, more specifically from the Digital Terrain Elevation Data (DTED; National Imagery and Mapping Agency, 2000). Copernicus DEM (European Space Agency and Airbus, 2022), for example, is delivered as DTED and also as DGED (Defence Gridded Elevation Data; Defence Geospatial Information Working Group, 2020) format, which defines six latitudinal zones with different factors.

**Table 2.1** Latitudinal zones and corresponding factors for determining angular spacing across parallel (INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1, 2023).

Zone	Latitude range	Factor for spacing across parallel
1	0°–50°	1
2	50°–70°	2
3	70°–75°	3
4	75°–80°	4
5	80°–90°	6

A quadrilateral direct surface tessellation that is equal area (i.e., constant cell area) is proposed by Tobler and Chen (1986). This tessellation is also based on a grid formed by meridians and parallels. Angular cell size along parallels is constant, and along meridians it varies to account for meridian convergence. This means that angular sizes of cells in the north-south direction are being enlarged by going polewards from the equator in order to ensure that all cells have equal area. As Seong (2005) noticed, this approach produces cells that are narrow in east-west direction and long in north-south direction. Therefore, as he wanted to produce square-like tessellation that is also based on meridians and parallels, he proposed that metric distance along meridians is kept constant and distance along parallels changes so that all cells have the same area. He first provides a formula for calculating the latitudes of meridians and then a formula for calculating longitude difference within each zone between two consecutive parallels. One disadvantage of this approach is the fact that cell boundaries in north-south direction are not aligned, which for some applications might not be acceptable.

Polyhedral direct surface tessellation begins with projecting edges of the polyhedron that is inscribed in the sphere to the sphere. Most often, one of the five Platonic polyhedrons (Figure 2.2) are used: tetrahedron (four triangular faces), cube/hexahedron (six square faces), octahedron (eight triangular faces), dodecahedron (twelve pentagonal faces), and icosahedron (twenty triangular faces). These are the only five regular polyhedra, meaning that their faces are all regular and mutually congruent polygons (Kerkovits, 2023, p. 253). This initial tessellation is then further refined by arcs of great or small circles of the sphere.



**Figure 2.2** Platonic polyhedrons.

One example that uses arcs of great circles is proposed by Szalay et al. (2007), which they refer to as Hierarchical Triangular Mesh (HTM). This surface discretization starts by connecting vertices of the octahedron that is inscribed into sphere with arcs of great circles. Each of the eight initial spherical triangles are recursively decomposed into four triangles by connecting midpoints of spherical triangle edges, again with the great circles arcs (Szalay et al., 2007). As Szalay et al. (2007) further explain, this tessellation does not produce cells of equal area; cells closer to the vertices of the octahedron are the smallest, while the largest are the ones near the center point of the initial spherical triangles. Tessellation by Dutton (2016) has a similar approach. He also started from the octahedron and eight initial spherical triangles that are in each hierarchical level refined to four triangles by connecting midpoints of edges of predecing triangles, but with a difference that cell edges obtained by connecting midpoints are not arcs of great, but of small circles (some cell edges are arcs of parallels of latitude). Dutton (1989) refers to this tessellation as Octahedral Quaternary Triangular Mesh (O-QTM, or often just QTM), and it again does not produce a tessellation with cells of constant area. Tessellation that produces cells of equal area that have small circles as their boundaries and that have only slightly shape variations is developed by Song et al. (2002). This approach starts from an icosahedron to define initial tessellation of sphere into twenty equal area spherical triangles. Each initial triangle is then divided into four or nine triangles by dividing each triangle edge into two or three parts, respectively, and then connecting them by arcs of small circles. As Song et al. (2002) state, there are infinite number of small circle arcs that can be used to connect two points on the sphere and thus it is possible to define small circle in such a way that ensures that all subdivision triangles have the same area. Since this small circle subdivision method is based on spherical trigonometry, it is computationally complex, however, it is possible to optimize it by adapting it to vector algebra and thus make it more suitable for wider adaptation (Song et al., 2002).

Direct surface tessellation approach that is based on spheroidal Voronoi polygons, which is a component of a Hipparchus Geopositioning Model, is developed by Lukatela (1987). As Lukatela (1987) explains, this spheroidal surface partitioning is based on a set of points on the surface of the spheroid that become center points of the Voronoi polygons defined by all points that are closer to the specific center point than to any other point from the initial set of points. The distribution of the center points can be based on various criteria, such as density of spatial data, i.e., where data are more dense, center points can also be distributed more densely (Lukatela, 1987, 2002). Clearly, this tessellation is not intended to produce cells of equal area or of regular shapes and is also not aimed at defining hierarchical tessellation. However, it has many benefits for geopositioning in computerized environment over the commonly used ellipsoidal coordinates.

## **2.2 Map-projection-based tessellation**

The map-projection-based tessellation can be divided into two categories: those that define tessellation in map projection plane and then map the tessellation back to the sphere or ellipsoid (i.e., single projection plane approach), and those that define tessellation on the faces of the polyhedron centered in the center of the sphere ellipsoid and then map that tessellation to the sphere or ellipsoid (i.e., multiple projection planes approach). Each of these approaches can be further divided into those that ensure that all tessellation cells have the same area and those that do not (J. A. Kimerling et al., 1999).

### **2.2.1 Single projection plane**

Single projection plane tessellation approach is arguably the most common and traditionally the most accepted one. First, the whole surface of the sphere or ellipsoid is mapped in plane, then tessellation is established in plane and then mapped back to the ellipsoid. The final tessellation properties would, of course, be directly linked to the properties of a used map projection.

Plate Carrée projection is often used for this purpose because of already discussed reasons; equiangular grid of meridians and parallels is mapped to grid of squares, which is especially suitable from the perspective of data storage in a form of a raster geospatial data model. This surface partitioning is an example of a non-equal-area one. Another non-equal-area tessellation that is

especially popular by web-map providers is based on the so-called Web Mercator projection (EPSG:3857). This tessellation maps the whole globe except polar areas beyond latitudes of approximately  $\pm 85^\circ$  to square in plane that is then recursively decomposed into, in general, four smaller equal-sized squares (Stefanakis, 2017). As Stefanakis (2017) explain, web-map providers generate (i.e., render) server-side vector and raster tiles based on this tessellation and then deliver them upon web-client requests. Each hierarchical level in this tessellation corresponds to a specific zoom level, meaning that in each subsequent level data are less generalized.

From the equal area projections there are various options to choose from. For example, Seong et al. (2002) compared Mollweide, Hammer-Aitoff, Eckert IV and sinusoidal equivalent projection. The rectangular grid in any of these projection planes, when mapped back to the sphere or ellipsoid, defines a global tessellation. If cells in rectangular grid have the same area, then tessellation of sphere or ellipsoid would also be equal area. Although the cell area is constant, their shape on sphere or ellipsoid gets distorted. Seong et al. (2002) were actually interested in determining which of these four projections is the most suitable for archiving global image data and have identified sinusoidal (sometimes referred to as Sanson-Flamsteed or Mercator-Sanson projection; Kerkovits, 2023, p. 177) as such. Sinusoidal projection demonstrated the lowest level of data replication and omission when reprojecting a tessellation that is based on a locally-accurate projection (in this instance, the Universal Transverse Mercator, UTM). INSPIRE Data Specification of Geographical Grids (INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1, 2023) defines a tessellation that is based on Lambert Azimuthal Equal Area (LAEA) map projection. This tessellation, however, addresses only Europe region and is not suitable for global applications. EASE-Grid (Equal-Area Scalable Earth Grid) on the other hand also uses LAEA, but it separately projects the north and separately the south hemisphere (for polar areas) and additionally uses equal-area cylindrical projection (for areas closer to the equator) and thus addresses the whole globe (Brodzik et al., 2012).

### 2.2.2 Multiple projection planes

Planes in multiple projection planes approach are generally faces of one of the five Platonic polyhedra (J. A. Kimerling et al., 1999) shown on Figure 2.2, but there are also examples of using other polyhedra, such as truncated icosahedron (Tong et al., 2005), disdyakis triacontahedron (Catalan solid; Hall et al., 2020), rhombic triacontahedron (also Catalan solid; Huang et al., 2024) or

cuboctahedron (Archimedean solid; Gray, 1994). To map sphere to the faces of the inscribed polyhedron, the first choice might be using the gnomonic projection (Snyder, 1992), i.e., central azimuthal perspective projection, however, there are many other suggested approaches. Projections on the faces of the polyhedron can further be unfolded into plane with unavoidable interruptions. There is also an example of the so-called polyhedric projection in which ellipsoidal quadrangles bounded by two meridians and two parallels of maximum dimensions of  $1^\circ \times 1^\circ$  are separately mapped to the plane (Frančula, 2000, p. 135). These  $1^\circ \times 1^\circ$  planar trapezium can be folded into polyhedron that Kerkovits (2023, p. 185) describes as ‘disco ball’. Somewhat similar approach but that does not depend on latitude-longitude grid is developed by Wecker et al. (2024). They propose a high-resolution polyhedron (i.e., polyhedron with many faces) instead of usually employed Platonic or Archimedean solids. Since high-resolution polyhedron better resembles sphere, distortions when projecting from polyhedron faces to the corresponding spherical polygons and vice versa are significantly reduced in comparison to, for example, Platonic solids (Wecker et al., 2024).

As already mentioned, the gnomonic projection is a straightforward solution for mapping a sphere to the faces of the inscribed polyhedron or vice versa. While it is computationally very efficient and it maps polyhedron edges to geodesics on sphere (straight line in projection plane, i.e., faces of polyhedron), it is not an equal area projection (Wecker et al., 2024), which is problematic for some applications. Gnomonic projection is, for example, used in hexagonal-cell, icosahedron-based H3 DGGs by Uber (Uber, n.d.) and as a part of a double mapping (i.e., composition of gnomonic projection and transformation aimed at making cells more uniform) in quadrangle-cell, cube-based S2Geometry DGGs by Google (Google, n.d.). Arguably the most widely recognized projection of a sphere that is based on a polyhedron is the Fuller’s Dymaxion Map by Buckminster Fuller. It first used cuboctahedron (polyhedron with eight triangular and six square faces) and later regular icosahedron, actually their spherical counterparts (spherical polyhedra), to unfold the globe into the plane (Gray, 1994). As Gray (1994) state, this projection does not preserve areas, but when areal distortions are compared to other map projections that are also suitable for global visualizations, namely Robinson and van der Grinten projections, this distortions are much less observable (i.e., areal distortion variability is lower).



Regarding the equal area projections from sphere to the faces of the inscribed polyhedron and vice versa, one of the most prominent examples is the work of Snyder (1992) that can be applied to any Platonic polyhedron and truncated icosahedron. Snyder uses LAEA projection with projection center corresponding to the latitude and longitude of the center of the spherical polygon and introduces additional transformations that ensure there are no overlaps or gaps between spherical polygons when planar polyhedron faces are mapped to sphere. Continuing on the work of Snyder (1992), van Leeuwen and Strebe (2006) developed a methodology for defining various equal area projections from faces of the polyhedron with regular polygonal faces to the sphere and vice versa. They also provide implementation for two such projections; one that does not introduce cusps when projecting meridian, parallels and other lines (which Snyder's projection does) and another that preserves lengths of polygon edges. As Harrison et al. (2011) state, inverse of the Snyder's projection (from faces of the polyhedron to the sphere) includes iterative calculations since it is not possible to derive formulas in the closed form. Therefore, they optimized inverse projection to make it more computationally efficient by employing different techniques, such as polynomial approximations to reduce the number of iterations or to abandon iteration altogether. Based on the Snyder's projection, Carr et al. (1997) developed the ISEA (icosahedral Snyder equal area) discrete global grid by projecting sphere to the faces of the icosahedron (20 triangles), then constructing regular hexagon within each triangle followed by projecting hexagons back to sphere. This procedure creates 20 spherical hexagons and 12 spherical pentagons (centered at the vertices of icosahedron) on the sphere, which establish an initial tessellation of the sphere (Carr et al., 1997). As Carr et al. (1997) further explain, in each hierarchical level of recursive decomposition, at each hexagon's center and each of its vertices, one hexagon with three times smaller area is centered to define planar tessellation in subsequent level. This planar tessellation can again be mapped to the sphere by the ISEA projection. Carr et al. (1997) also developed a way of assigning unique identifier to each spherical hexagon and 12 pentagons in each hierarchical level.

Roşca and Plonka (2011) applied a similar approach for the cube. They also use LAEA projection for mapping faces of the cube to the sphere, followed by the additional area preserving transformation that tackles overlaps which occur when each cube face is mapped to the sphere with the LAEA projection (more in Section 3.3.2). Chan and O'Neill (1975) also developed an equal area

mapping from the faces of the cube to the sphere, which they refer to as Quadrilateralized Spherical Cube (QSC). Both of these approaches define quadrangular partitioning of the sphere.

UTM projection should also be mentioned here. It is essentially a series of transverse Mercator projections that are used to map 6° degrees wide longitudinal zones into the plane, consecutively from the antemeridian (i.e., meridian with longitude of -180°). Strictly speaking, ellipsoids or sphere is this way mapped into 60 different planes. However, there is no geometrical relationship between these planes and sphere or ellipsoid, such as in case when planes are faces of the inscribed polyhedron, and these 60 planes can be in fact treated as a single plane. In that sense, UTM-based tessellation should be treated as single-projection-plane-base.

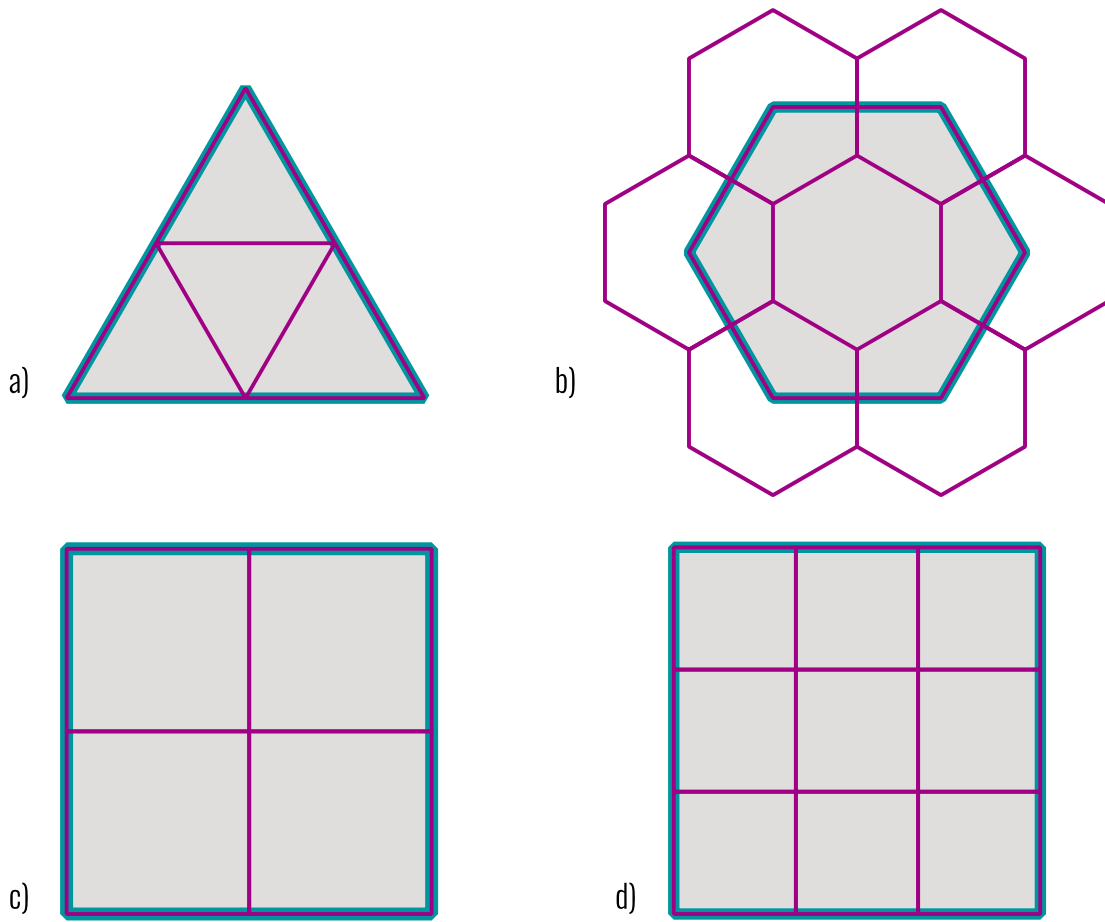
### **2.3 Discrete global grid system (DGGS)**

As already stated, any grid that partitions earth's surface (i.e., sphere or ellipsoid of revolution) into finite number of areal elements can be considered as discrete global grid (DGG), and a system of (hierarchically organized) DGGs defines a discrete global grid system (DGGS). In previous sections, an overview of various approaches that can be applied in generating DGGs and subsequently DGGS was reviewed. Since we are examining DGGS from the land cover data perspective, DGGS that has cells of constant area is a basic requirement since it makes areal statistics calculations simple. From all tessellation approaches listed on the Figure 2.1 and examined in previous sections, only those that are based on small circle, polyhedral, direct surface tessellation and those that are based on equal area projections on a single or multiple projection planes, enable generating DGGS with cells having equal areas in a single hierarchical level (the so-called equal area DGGSs). From these, the most promising and most often applied approach is the one that is based on the area map projection from the faces of the polyhedron to the sphere or ellipsoid (Sahr et al., 2003; Wecker et al., 2024). Mahdavi-Amiri et al. (2015b) defined five parameters that characterize DGGS that are based on polyhedron: the choice of polyhedron, subdivision or refinement method applied to the polyhedron faces, shapes of cells, mapping approach for transferring cells from faces of the polyhedron to the sphere or ellipsoid, and indexing method for assigning unique identifiers to each of these cells. In the following paragraphs we give a brief overview of each of these design choices.

As already discussed, the most often choice for the regular polyhedron that defines the initial tessellation is one of the five Platonic polyhedrons (Figure 2.2), although other polyhedrons can also be used. The reason behind choosing Platonic polyhedrons is the fact that there is no other way to construct tessellation of sphere into regular, congruent (i.e., equal in shape and size) polygons beside radially projecting edges of the sphere-inscribed Platonic polyhedrons to the sphere (Lukatela, 2002). Any further refinements produce inevitable distortions in shape and/or area of tessellation cells (J. A. Kimerling et al., 1999). If number of polyhedral faces is larger, the polyhedron better represents the sphere and lower level of distortions are introduced (Wecker et al., 2024). This is the reason why icosahedron is the most often used from the Platonic polyhedrons (Sahr et al., 2003) and other polyhedrons that have a larger number of faces. Wecker et al. (2024) go the farthest in this regard by developing an approach for constructing high resolution polyhedrons, theoretically with unconstrained number of faces. After polyhedron has been selected, additional step also includes orienting it in relation to the equator and prime meridian (Sahr et al., 2003). There are various ways of orienting polyhedron, which is usually application-related. If shape and/or cell area distortions are largest for cells near the vertices of Polyhedron, and if DGGS is intended for applications on land, then it might be beneficial of placing polyhedron vertices so that they are placed into the sea. In some cases, it might be suitable placing vertices so that they coincide with poles of the geographic grid which might simplify establishing a link between geographical coordinates and DGGS cells.

Refinement method defines how initial tessellation is further subdivided into finer cells. Refinement is defined in the plane (on each face of the polyhedron or in its unfolded form) and then mapped back to the sphere or ellipsoid. As Mahdavi-Amiri et al. (2015b) explain, refinement approach can be described by aperture, congruency, and alignment. Aperture, or sometimes refinement ratio (R. Gibb et al., 2021), defines number of cells (i.e., children cells) that each cell in one hierarchical level (i.e., parent cell) is decomposed into in the subsequent hierarchical level. Aperture is in some instances (for example, by Sahr et al. (2003)) defined as ratio of area between parent and children cells. Hierarchical levels are also sometimes referred to as refinement levels, recursion levels or resolutions. Apertures of, for example, 4 or 9 are commonly denoted as 4-fold or 9-fold, or similarly as 1-to-4 or 1-to-9, respectively. Lower aperture values are usually desired since they provide greater flexibility in selecting the most appropriate cell size for specific application and

also provide spatially finer transition between hierarchical levels (Amiri et al., 2015). If refinement is congruent, then spatial union of children cells is fully encompassed by their parent cell (Mahdavi-Amiri et al., 2015a) and thus hierarchical relationships between children and parent cell can be easily established. Otherwise, hierarchical relationship might, for example, be established based on rules that are defined during the DGGs design phase. Alignment can be vertex- and center-aligned, based on the fact whether children cells inherit vertices (vertex-aligned) and/or center points (center-aligned) from their parents (Mahdavi-Amiri et al., 2015a). Refinements of various apertures, congruency, and alignment properties are shown on Figure 2.3.



**Figure 2.3** Different refinements of a parent cell (tick boundary) with children cells (pink boundary): (a) congruent, center- and vertex-aligned 1-to-4 refinement of a triangular cell, (b) not-congruent, center-aligned aperture 3 refinement of a hexagonal cell, (c) congruent, vertex-aligned 1-to-4 refinement of square cell, and (d) congruent, vertex-aligned 1-to-9 refinement of a square cell (adapted from Alderson et al. (2020)).

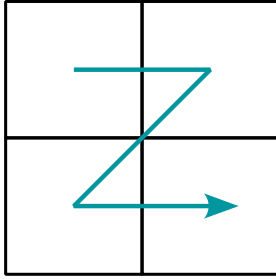
Regarding the cell shapes, the most common ones are triangles, quadrangles, and hexagons. In hexagonal tessellation, often exactly 12 pentagons are introduced because it is not possible to

tessellate sphere or ellipsoid with hexagonal cells only (Saff and Kuijlaars, 1997). Characteristics of refinement (congruency and alignment) are directly related to the cell shape. For example, refinement in a DGGS with hexagonal cells can never be congruent (Alderson et al., 2020), while for triangular and quadrangular cells congruent refinement can be easily established for some apertures. On the other hand, hexagonal cells have the largest compactness and have uniform adjacency; each hexagon has six neighboring hexagons with which it shares one edge, while triangles and quadrangles have some neighbors with which they share one edge and some with which they share one vertex (Sahr et al., 2003). Because of these reasons, as Sahr et al. (2003) point out, single resolution hexagonal DGG is usually preferred over triangular and quadrangular, but if clear hierarchical relationship is required between grids of different refinement, then triangular and quadrangular ones are more suitable.

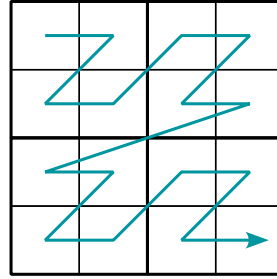
Projections used to transfer hierarchical grid from the faces of the polyhedron to the sphere or ellipsoid have already been discussed in Section 2.2.2. The last characteristic that enables distinction between different DGGSs according to Mahdavi-Amiri et al. (2015a) is the indexing method, i.e., assigning a unique identifier/index/address to each cell. These unique identifiers can be viewed as a substitution for traditionally used coordinates, i.e., DGGS can in this sense be viewed as a distinct georeferencing model. In general, one of the four indexing methods are used: hierarchical indexing, space filling curve indexing, axes-based indexing, or encoded addresses schemas (Amiri et al., 2015; R. Gibb et al., 2021). In hierarchical indexing, as its name suggests, cell indices are directly related to their “position” in the established hierarchical structure. As Amiri et al. (2015) explain, identifier of a specific cell is formed by adding an integer to the identifier of its parent cell. Space-filling curves are often used for digitally storing multidimensional data in a 1D structure. In this case, different space-filling curves (Hilbert and Morton space-filling curves are shown on Figure 2.4) can be used to go (i.e., traverse) through all cells at all hierarchical levels and assign unique identifier to each cell. These identifiers are generally integers that increase by one for each cell in a traversal order and are given in a numeral system with base that corresponds to the DGGS aperture or its square root (Amiri et al., 2015). In axis-based indexing, identifier of a cell is determined based on a coordinate system whose axes are aligned in a way that enables addressing each cell with an vector of integers, each representing equal steps needed along each axis (Alderson et al., 2020). This vector can be 2D for grid that is made of quadrangular cells, or 3D for grids made of hexagonal cells (L. Zhao et al.,

2022). If indexing is not based on any of these approaches, then it can be classified as an indexing that is based on some schema for encoding addresses. An example of this type of indexing is used in the What3Words where each cell of about  $3 \times 3$  m is uniquely addressed by three words separated by period (Arthur, 2023).

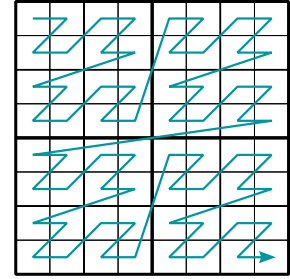
### Morton space-filling curve



First step

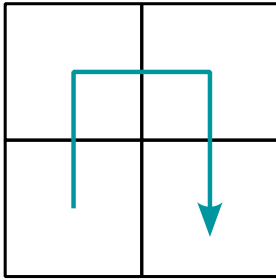


Second step

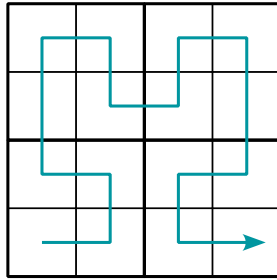


Third step

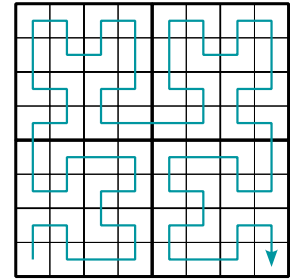
### Hilbert space-filling curve



First step



Second step



Third step

**Figure 2.4** The first three steps of Morton (upper row) and Hilber (lower row) space-filling curves (adapted from Amiri et al. (2015) and Moon et al. (2001)).

### **3 Development of an enhanced DGGS for land cover mapping**

From the analysis of several prominent global land cover datasets in Section 1.1, it is clear that images acquired by remote sensing optical multispectral and radar sensors onboard earth observation satellites are the primary data source for creating land cover maps. Once the data has been collected, one of the pre-processing steps before delivering it to users typically involves projecting the data into a plane using a map projection and storing it in a raster data format. Such raster files then become part of geographic information system (GIS), where they are processed, analyzed, and visualized. Although GIS development began in the 1960s and was in many ways conditioned with the computer capabilities of that time, it seems that some of the fundamental concepts are still omnipresent in modern-day GIS (Goodchild, 2019). As recognized by Goodchild (2019), basic concept of early GIS as a collection of flat two-dimensional raster and vector layers has basically remained unchanged. This concept has various drawbacks that are not related to land cover data only, but to georeferencing spatial data in general. Some of these drawbacks are examined in the following section.

#### **3.1 Deficiencies of current approaches in land cover data storage and handling**

We firstly start from the approach of acquiring data by sensors on earth observation satellites. Each sensor detector receives electromagnetic energy that is emitted or reflected from the particular part of the earth's surface and then stores a value that is proportional to the received energy (Cracknell, 1998). This area of the Earth's surface is known as the ground projected instantaneous field of view (GIFOV; Markham et al., 2020) or the instantaneous geometric field of view (IGFOV; Joseph, 2020). These raw observations are stored in the form of rectangular matrix in what is often called, sensor geometry, as opposed to cartographic geometry after performing orthorectification (Gascon et al., 2014). Kimerling et al. (1999) state that one of the fundamental issues related to global raw observations data storage is the issue of their conversion to the global regular grid. Almost without exception, satellite observations and therefore land cover products as well are typically delivered in the form of a flat raster geospatial data model. This is not because this model is the most suitable for

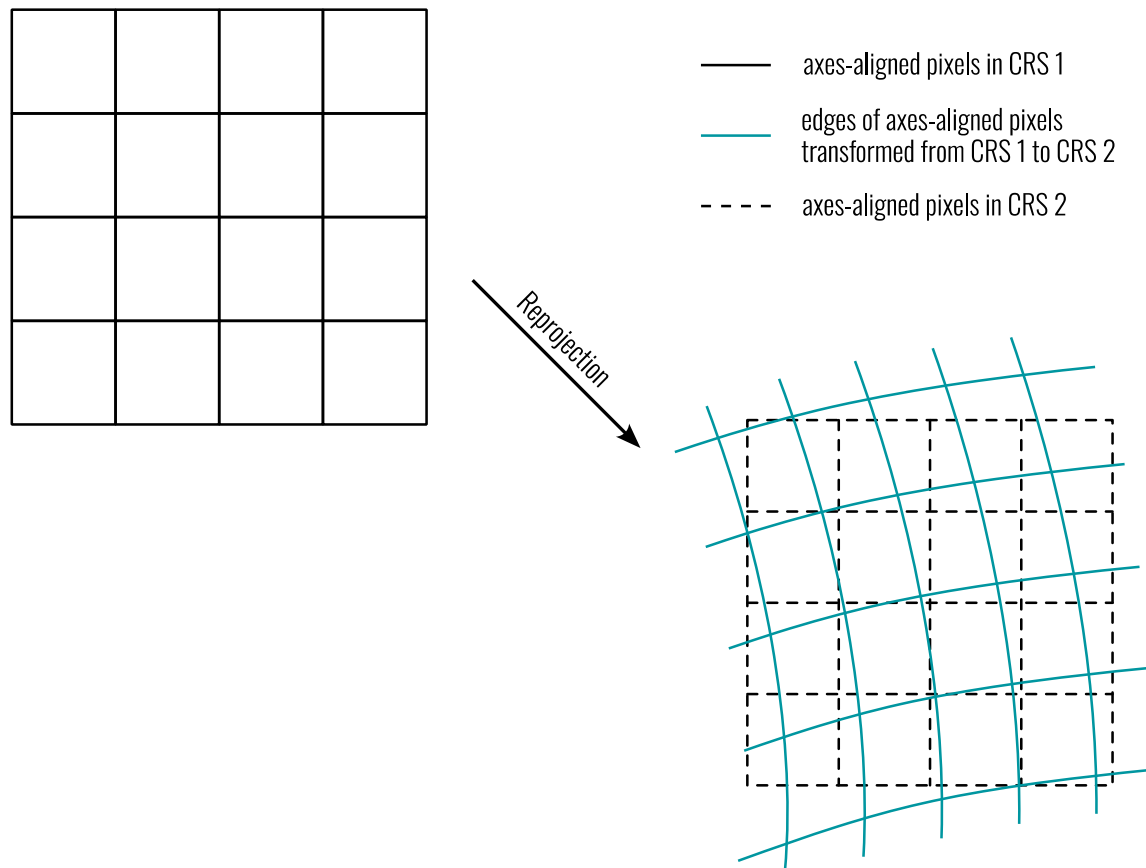
such purposes, but rather because it is a widely accepted model that was used in GIS from its beginnings.

As already stated, global raster (land cover) data are usually stored in a geographic 2D or projected CRS. Raster data in geographic 2D CRS essentially tessellates sphere/ellipsoid in a way that, according to Chapter 2 and Figure 2.1 is classified as a quadrilateral equal angle direct surface tessellation. In projected CRS on the other hand, raster data define tessellation that can be classified as map-projection-based. Map-projection-based approaches that are currently widely used are based on a single map projection plane (unless UTM-based tessellation is considered as a multiple-projection-planes-based). Some of the issues related to these tessellation approaches have already been briefly established, however, here they will be given into more detail.

When using projected CRS, raw observations are transferred from the curved surface of the earth to the plane by using a particular map projection. In this process, along with a well-known problem of deformations that map projections introduce, an additional issue is the reprojection—changing the CRS in which raster data are georeferenced (Figure 3.1; reprojection also occurs when transferring remotely sensed data from sensor to cartographic geometry). A change in the CRS requires the resampling of raster data, as the pixel grid must be realigned to the new coordinate axes. Although various reprojection and resampling methods are available, all introduce some form of data alteration (Lu et al., 2018). For instance, nearest neighbor resampling preserves the original pixel values but reprojection itself inevitably introduces a spatial shifts. On the other hand, resampling methods such as bilinear and cubic convolution introduce data smoothing, which may be advantageous for certain applications, but they modify the initial pixel values (Dwyer et al., 2018). According to Steinwand et al. (1995), one of the main problems related to reprojection is determining the target pixel size. Another challenging aspect of reprojection is the potential pixel loss or replication (Moreira de Sousa et al., 2019; White, 2006). Bauer-Marschallinger et al. (2014) even developed an indicator—grid oversampling factor (GOF) as a measure of pixel duplication that occurs when satellite images are projected onto a regular grid. The fact that issues related to the raster data reprojection are not negligible, can be distinguished from the research by Dwyer et al. (2018). Mainly to minimize the negative effects of raster data reprojection when producing Landsat ARD images, they started from the raw Landsat data, and not from the generally available Landsat data



georeferenced in the UTM-based CRS. Their aim was to generate images in the Albers equal area conic projection and by using raw Landsat data instead of data in UTM, they have avoided double reprojection.



**Figure 3.1** Reprojecting raster. When axes-aligned pixel edges are transformed from one projected CRS into another projected CRS, they are, in general, some arbitrary curves. However, pixel edges of a raster in another CRS are again axes-aligned and thus resampling is required.

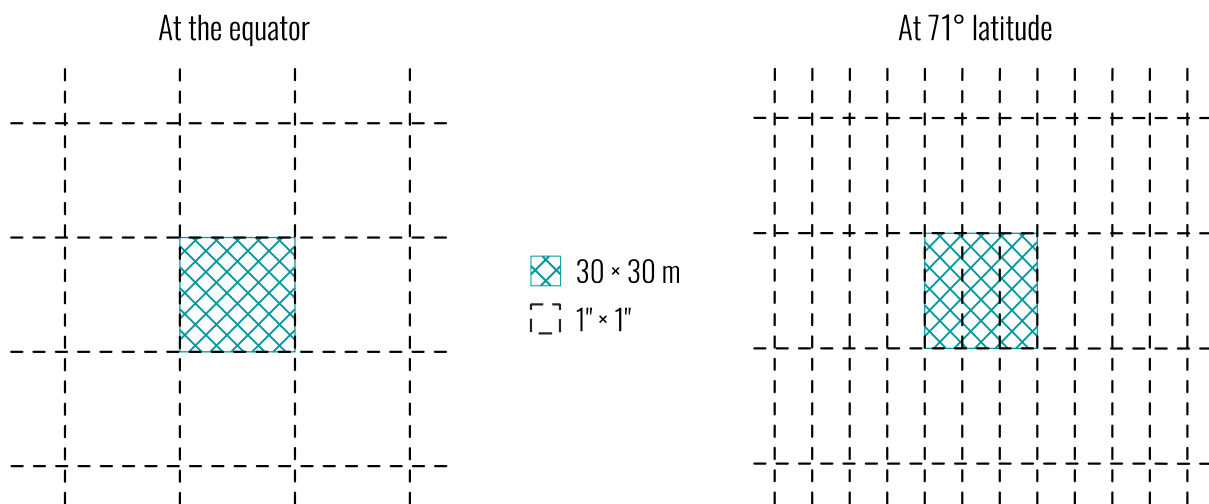
Publicly available Sentinel-2 images are also delivered in the UTM-based CRSs. Although linear projection deformations in UTM-based CRSs are not significant for most applications and can thus be ignored, they virtually introduce 60 different CRSs, one for each 6° wide longitudinal zone. As a consequence, using Sentinel-2 and Landsat images for global-scale analysis might be challenging since not all data are provided in the same CRS. Data can be either analyzed on a zone-by-zone basis, or can be reprojected to some common CRS, in which case reprojection-related issues will again be introduced. Additionally, Sentinel-2 images that are acquired within the same datatake, but were georeferenced in the different UTM zones, will partly overlap. It might therefore be challenging to determine which pixel value in the overlapping region to choose during the reprojection. Roy et al.

(2016) developed a resampling approach for such overlapping parts of Sentinel-2 images and concluded that if all available data are used from the overlapping parts of Sentinel-2 images, reprojection results will be of higher quality. On the other hand, Bauer-Marschallinger and Falkner (2023) found out that because of the overlap between Sentinel-2 images, not only between images in adjacent UTM zones, but also between images within the same UTM zone, a total of about one petabyte of data is redundantly stored for L1C and L2A images in a single Sentinel-2 data storage instance every year.

From the global land cover data perspective, equal area projections are usually the only preferred ones since they enable performing simple areal statistics calculations (Steinwand et al., 1995; Thompson et al., 2022). Steinwand et al. (1995) have identified four equal area projections that are suitable for data with global coverage: interrupted Goode Homolosine, interrupted Mollweide, Wagner IV and Wagner IV. Moreira de Sousa et al. (2019) also concentrated on equal area projections that are suitable for global analyses and are implemented in widely used PROJ (PROJ contributors, 2025) and other free and open-source libraries. They have compared sinusoidal, Mollweide, Hammer, Eckert IV, and interrupted Goode Homolosine (IGH) projections and have identified the last one as the one that, on average, introduces the lowest angular and distance deformations. Although interrupted Goode Homolosine projection introduces the lowest deformations, it still introduces them, mainly because of the attempt to deal with global data in a flat surface, rather than on the sphere or ellipsoid directly. Additionally, IGH projection also introduces the problem of data replication (observed by Yang et al. (1996), as cited by Usery et al. (2002)).

Even though raster data in geographic 2D CRS directly tessellate sphere or ellipsoid, they also introduce data replication, firstly in data acquired by remote sensing satellites, and then subsequently in land cover data. This is visible from example on Figure 3.2. Because of the convergence of meridians, in the equal angle quadrilateral tessellation, cell/pixel size in east-west direction decreases by going from the equator towards north or south pole. In the north-south direction, in case of sphere, cell size is constant, and in the case of ellipsoids that are used as a model for earth's surface, cell sizes are slightly varying, but with a much lower magnitude than in east-west direction. An example on Figure 3.2 shows  $30 \times 30$  m quadrangle that can be perceived as a GIFOV of a satellite sensor detector. For simplicity, it can be taken that  $30 \times 30$  GIFOV has approximately constant size

and shape regardless of geographical location. If equiangular geographic grid (EGG) is defined so that cells have approximately the same size as GIFOV at the equator, by going polewards, one GIFOV can, depending on the latitude, be mapped to more than one cell. For the WGS 84 ellipsoid, as visible on Figure 3.2, at approximately  $71^\circ$  north or south latitude,  $30 \times 30$  GIFOV corresponds to three equiangular quadrilateral cells. In this case, redundancy, or multiplied storage of original pixels, occurs. Because of the recent trend of delivering land cover data in geographical 2D CRS (Table 1.1) it is very likely that all input data used for land cover classification are integrated in the common equiangular quadrilateral grid prior to classification. This again means that, for example, Landsat or Sentinel-2 data will be partly replicated when reprojecting from UTM-based CRS to grid in geographic 2D CRS. Possible solution was already mentioned—defining latitudinal zones with varying cell sizes in east-west direction (Table 2.1). However, this approach in essence produces the same issue as using UTM-based CRSs does. Regular raster data formats, such as GeoTIFF, cannot have different grid resolutions (i.e., pixel sizes) within the same file, which hinders data analysis, especially in era of cloud-optimized data formats that enable storing and managed all data from a single file, even on a global level.



**Figure 3.2** Relationship between  $30 \times 30$  m quadrangle (e.g., GIFOV of satellite sensor detector) on the WGS 84 ellipsoid and  $1'' \times 1''$  geographical grid.

As a disadvantage of georeferencing spatial data in geographic 2D CRS instead of in projected CRS, often a more complex data analysis is mentioned, especially calculating areas and distances in geographic CRS. Regardless of this and data replication issue, we still more support geographic CRS when choosing between geographic and projected CRS for global land cover data georeferencing.

Especially because modern GIS applications are starting to natively support working with data in geographic CRSs (Frančula et al., 2021; more on this in Chapter 5). Chrisman (2017) advocates for abandoning map projections for most (but not all) calculations on the ellipsoid and using them for visualization purposes, while also tuning them data representation needs rather than for performing calculations (e.g., using equal area projection for calculating areas).

All the previously mentioned challenges stem from the effort to define tessellation in a flat surface (i.e., plane) and then transfer it to curved earth's surface (or more precisely, the ellipsoid or sphere that approximates it). The most obvious problem here is that flat surface is not homeomorphic (i.e., topologically equivalent) to the sphere/ellipsoid (Chrisman, 2017). Even the equiangular tessellation by meridians and parallels, that is classified as direct surface tessellation approach, is defined by mapping a 2D domain ( $[-180^\circ, 180^\circ] \times [-90^\circ, 90^\circ]$ ) to the sphere or ellipsoid (Guo et al., 2020). Therefore, tessellation approaches that are based on polyhedra (geodesic DGGS) that are homeomorphic to the sphere are a more efficient way of discretizing sphere or ellipsoid into finite number of highly regular cells.

### 3.2 Choosing a DGGS for global land cover data

From the global land cover data perspective, only equal area DGGSs come into account. Equal area DGGSs enable trivial calculations of land cover areal statistics—it is required to count number of DGGS cells with assigned specific land cover class and then multiply it with constant area of cells. Additionally, equal area DGGSs ensure that all cells have a uniform probability of contributing to analysis (White et al., 1992).

Next requirement would be that DGGS defines hierarchical tessellation of ellipsoids of revolution and not only spheres. Although for some applications using sphere as a proxy for the surface of the earth is justified, in order to achieve a more accurate analysis based on gridded data, ellipsoid must be used (Kelly and Šavrič, 2021). Process of land cover data classification already includes various sources of uncertainty and thus introducing any additional element of variability, such as using sphere instead of ellipsoid, should be avoided. If DGGS is an equal area one and is defined for spheres only, it can be easily extended to ellipsoid of revolution by using the so called authalic latitude (more on this in Section 3.3.1). This way, equal area property of the DGGS will be

retained on ellipsoid as well. Therefore, for equal area DGGSs, this requirement is preferred, but not mandatory.

An additional requirement on the choice of DGGS that we consider important from the land cover data perspective is that it has a quadrangular cell. As noted earlier, DGGS cells are most often triangles, quadrangles and hexagons. Hexagonal grids are often considered as superior to quadrangular and triangular for achieving a highly regular discretization of sphere or ellipsoid. Hexagons have the greatest compactness (Section 3.4.3), and hexagonal grids have uniform adjacency, clear nearest neighbors (in hexagonal grid each hexagon shares a joint edge with their neighbors, while in quadrangular and triangular grids, some neighboring cells share a joint edge while others a joint vertex) and more-less constant distances between centers of neighboring cells (Birch et al., 2007; Sahr et al., 2015). However, the greatest drawback of hexagonal (or triangular) cells as well as triangular ones, as recognized by Sahr et al. (2011), is that sensor detectors onboard remote sensing satellites, that are the most valuable data source for producing land cover data, do not have hexagonal (or triangular), but rather quadrangular shapes. Therefore, choosing DGGS with quadrangular cells would be the most sensible from this perspective. DGGS with quadrangular cells are most often based on a cube as a base polyhedron, however, there are also examples that are based on octahedron (Luo et al., 2023). Often, the necessity for development of software that will enable advanced handling of data referenced with the DGGS is underscored as one of the greatest downsides of DGGSs from the applicative standpoint. However, Béjar et al. (2023) showed that existing software for managing raster data can be directly applied to data in DGGS that has quadrangular cells. This is another reason why we decided to go with a DGGS with quadrangular cells for managing land cover data.

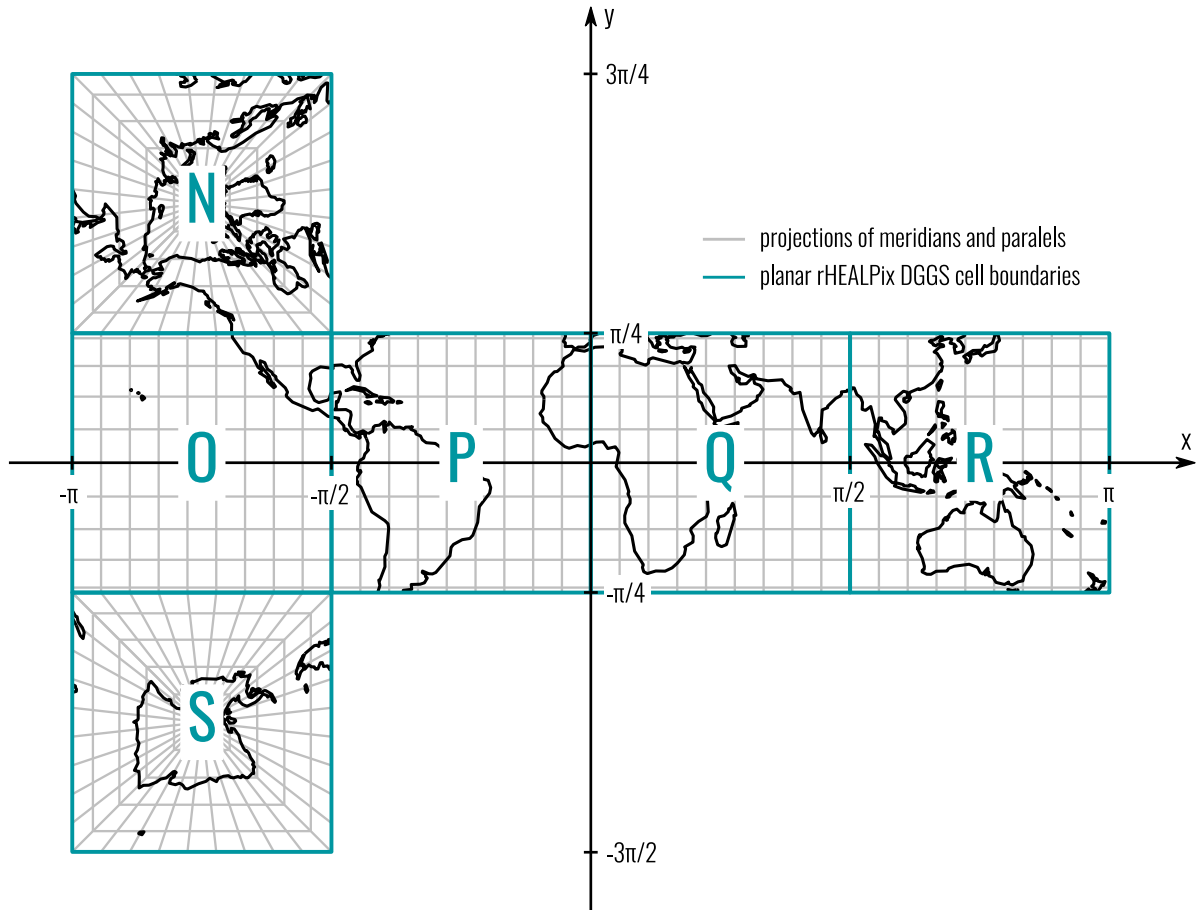
To summarize, a DGGS that is equal area, supports tessellation of both spheres and ellipsoids and has quadrangular cells is, based on the above discussion, an optimal choice for land cover data storage and handling. From the currently available DGGS implementations that were examined by Kmoch et al. (2022) and that fulfill established requirements, rHEALPix (Rearranged Hierarchical Equal Area Isolatitude Pixelation) DGGS was identified as the one that has the most stable shapes of cells. It should be noted however, that all cells of the rHEALPix DGGS are not ellipsoidal quadrangles—some cells are triangles. Nevertheless, the number of triangular cells is relatively small and one of the enhancements that we examine is particularly dealing with this issue. In the following

sections an overview of the rHEALPix DGGS is given and then two enhancements are introduced. The first one deals with numerical stability of calculating authalic latitude, a crucial aspect of the rHEALPix DGGS that enables equal-area tessellation of ellipsoid (Section 3.3.1), and the second one is focused on increasing stability of shapes of DGGS cells by applying different approach for transferring grid from the faces of the cube to the sphere/ellipsoid (Section 3.3.2). This second modification results in a tessellation of ellipsoid that is composed of cells that all have quadrangular shape. Results of the quantitative evaluation of rHEALPix DGGS before and after these modifications are provided in Section 3.5, while their interpretation and implications are discussed in Chapter 4.

### 3.3 Modifying the rHEALPix DGGS

The rHEALPix DGGS is a DGGS that is based on the HEALPix (Hierarchical Equal Area Isolatitude Pixelation) DGGS introduced by Górski et al. (1998; 2005). HEALPix DGGS provides an equal-area discretization of sphere and was initially intended for analysis of astronomical data (Górski et al., 2005). An interesting characteristic of HEALPix DGGS is that the center points of adjacent cells (or pixels, according to the HEALPix terminology) in east-west direction lie on the same parallel of latitude and are equally spaced on each parallel, which reduces complexity of spherical harmonics calculations (Gorski et al., 1998). Based on work of Roukema and Lew (2004), Calabretta and Roukema (2007) further derive an HEALPix spherical map projection that is suitable for projecting a sphere to the plane in a way that will map all 12 resolution-0 HEALPix DGGS cells to squares. As explained by Calabretta and Roukema (2007), this projection is a combination of Lambert's cylindrical equal-area map projection for spherical zone between latitudes  $\phi = \pm \arcsin(2/3) \approx \pm 41.8103^\circ$  (the so-called equatorial region) and interrupted equal-area Collignon projection for areas outside of this spherical zone (the so-called north and south polar regions). Since HEALPix DGGS was originally developed for spheres only, Gibb et al. (2013) modified it so that it can be applied to ellipsoids of revolution as well (further explained in Section 3.3.1). They also defined the rearranged HEALPix (i.e., rHEALPix) map projection and developed an rHEALPix DGGS that is based on it.

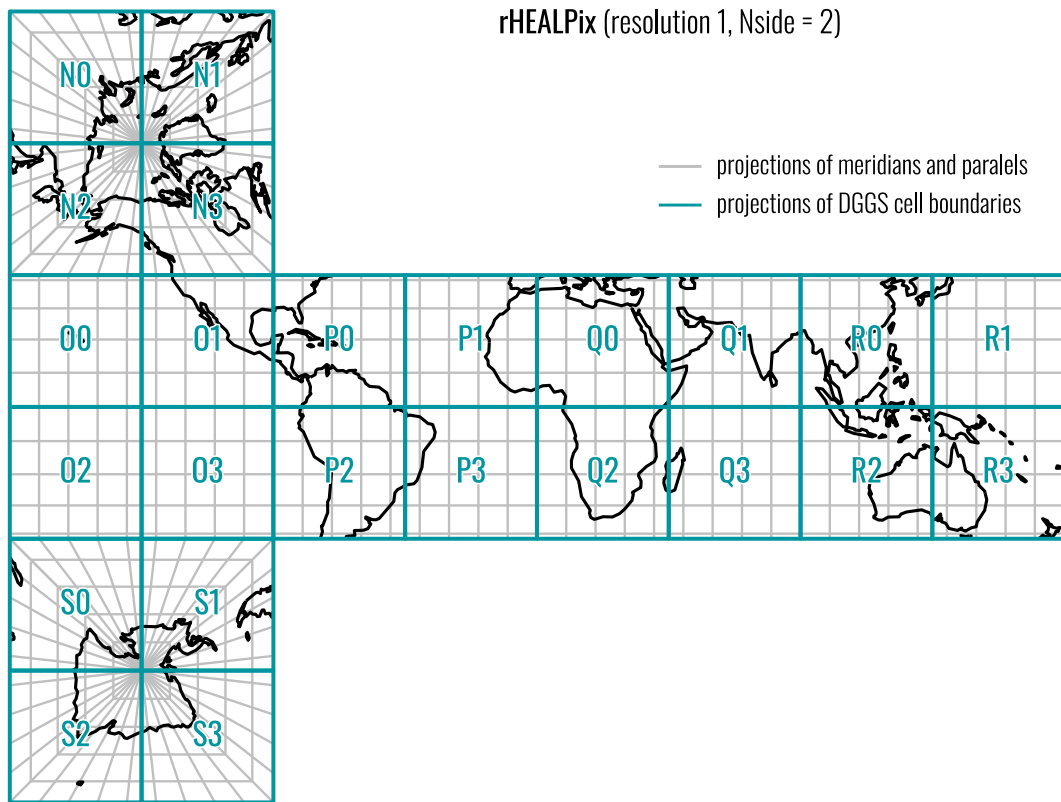
The rHEALPix map projection projects sphere or ellipsoid of revolution into six squares of equal size. Each polar region is projected into one square, and equatorial region is projected into four squares, as shown on Figure 3.3. Squares that are projection of the north and south polar region are denoted as N and S, respectively, while squares that are obtained by slicing projection of the equatorial region by projection of meridians that have longitudes  $-90^\circ$ ,  $0^\circ$ , and  $90^\circ$  are denoted as O, P, Q, and R, from left to right (Figure 3.3). These squares, or more accurately, their corresponding cells on the sphere or ellipsoid (spherical or ellipsoidal areas) define the first hierarchical level (i.e., resolution-0) of the rHEALPix DGGS, while assigned letters N, O, P, Q, R, and S are their identifiers. The rHEALPix map projection has two required parameters;  $n$  and  $s$  that are integers between 0 and 3 and that define position of the N and S squares (projection of the N and S cells). For the N square, if it is placed above O square then  $n = 0$ , if it is above P then  $n = 1$ , if it is above Q then  $n = 2$ , and if it is above R then  $n = 3$ . Value of the parameter  $s$  analogously defines the position of square S in



**Figure 3.3** Projection of unit sphere and the resolution 0 rHEALPix DGGS cells in the plane by the (0, 0)-rHEALPix map projection.

relation to the squares O, P, Q, and R. Therefore, rHEALPix map projection is usually denoted as  $(n, s)$ -rHEALPix, which means that Figure 3.3 shows the  $(0, 0)$ -rHEALPix projection of the unit sphere. From this definition of the initial tessellation of the sphere or ellipsoid it can be concluded that the rHEALPix DGGS is a cubic geodesic DGGS (Gibb, 2016) that can be applied to spheres and ellipsoids of revolution.

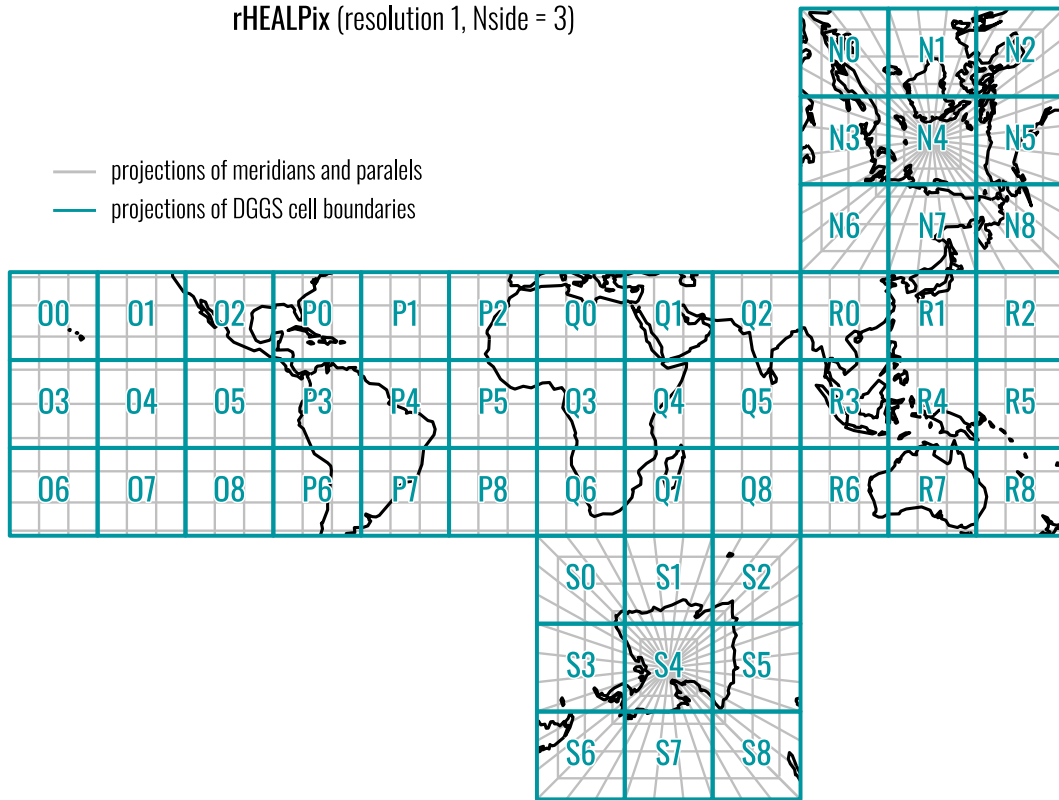
Aperture or refinement of the rHEALPix DGGS is defined by the integer parameter  $N_{side}$  which determines the number of segments of the same length in which each planar cell edge is divided in order to construct the cells on a subsequent hierarchical level. This means that aperture is  $N_{side}^2$  for parameter  $N_{side}$ , or refinement ratio can be expressed as 1-to- $N_{side}^2$ . Initial planar squares at the first hierarchical level (i.e. resolution-0), that correspond to the initial tessellation of the sphere or ellipsoid into six cells of equal area, are thus divided into  $N_{side}^2$  squares each in the following level. Figure 3.4 shows example for resolution 1 for  $N_{side} = 3$  and Figure 3.5 for  $N_{side} = 3$  (Appendix A.4 additionally shows resolution 2). All these squares have the same area and are in subsequent hierarchical levels again recursively decomposed into  $N_{side}^2$  squares. These planar square grids are mapped back to the sphere or ellipsoid with the inverse  $(n, s)$ -rHEALPix map projection which



**Figure 3.4**  $(0, 0)$ -rHEALPix map projection of the sphere/ellipsoid and resolution 1,  $N_{side} = 2$  rHEALPix DGGS cells.

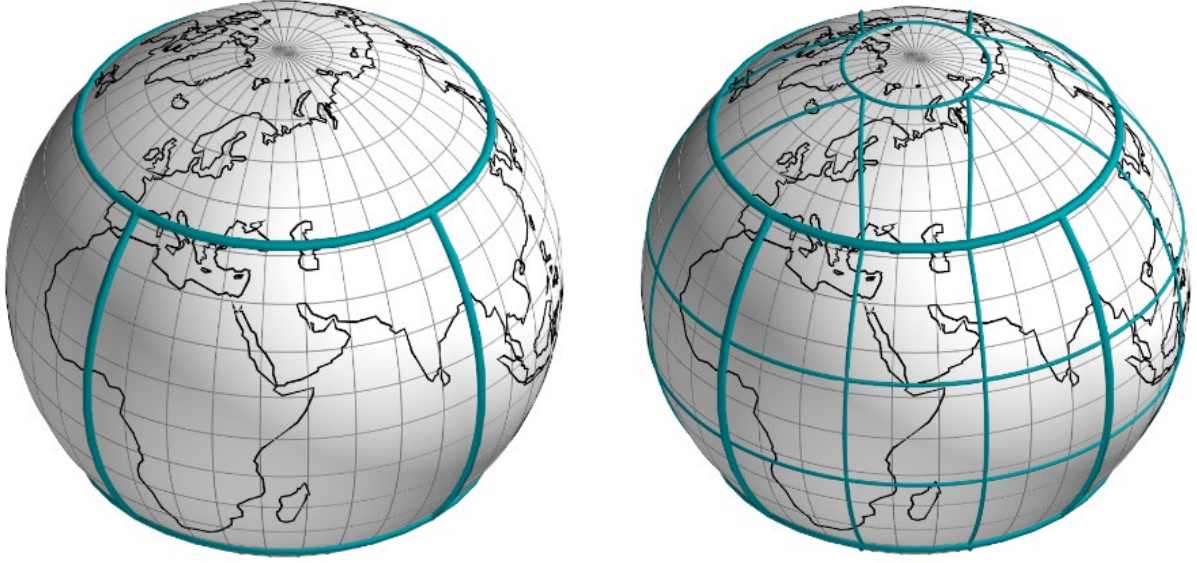


produces the rHEALPix DGGs on the sphere or ellipsoid (Figure 3.6; more visualizations are available in Appendices A.1 and A.2). Since  $(n, s)$ -rHEALPix map projection is an equal-area projection and since all squares have the same area in each hierarchical level, rHEALPix DGGs is an equal-area DGGs. Also, from the approach on how cells from one hierarchical level are decomposed into finer cells on the following level, it is clear that rHEALPix DGGs is congruent and for odd values of  $N_{side}$  it is center-aligned while for even values of  $N_{side}$  it is vertex-aligned.



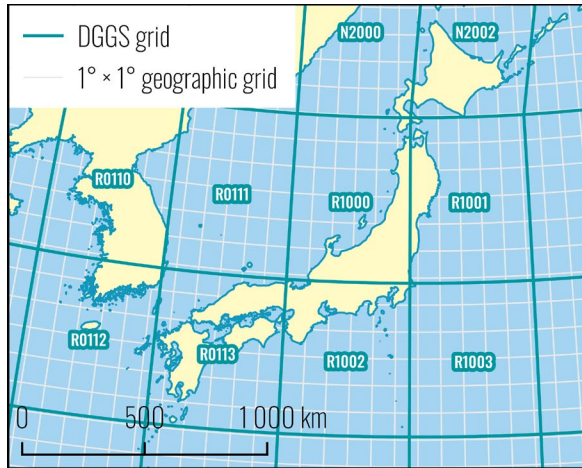
**Figure 3.5** (3, 2)-rHEALPix map projection of the sphere/ellipsoid and resolution 1,  $N_{side} = 3$  rHEALPix DGGs cells. The same cells visualized in Robinson projection are available in Appendix A.5.

Each cell has a unique identifier that is assigned to it. As already stated, resolution-0 cells have identifiers N, O, P, Q, R, and S. When each resolution-0 cell is decomposed into  $N_{side}^2$  cells, their identifiers are determined by adding an integer from 0 to  $N_{side}^2 - 1$  to the identifier of the corresponding resolution-0 cell from left to right and top to bottom in a z-order filling curve manner. Example of assigning identifiers to resolution-1 cells for  $N_{side} = 3$  is visible on Figure 3.5. This same procedure is followed on each hierarchical level, which means that each cell identifier is determined



**Figure 3.6** Resolution-0 (left) and resolution 1 (right) rHEALPix DGGs cells on the ellipsoid for the  $N_{side} = 3$ .

by adding one of the integers from 0 to  $N_{side}^2 - 1$  to the identifier of their parent cell. Gibb (2016) developed a mechanism of determining topological relationships between cells using only their identifiers. This means that it is not required to know cell coordinates on sphere or ellipsoid, but only their identifiers to be able to find out which cells are neighbors or does one cell contain another. Figure 3.7 shows rHEALPix DGGs grids for resolution 4 and 11 in 1-to-4 and 1-to-9 refinements respectively, along with cell identifiers (more figures are available in Appendix A.3).



rHEALPix (1-to-4, resolution 4)



rHEALPix (1-to-9, resolution 11)

**Figure 3.7** rHEALPix DGGs grids and cell identifiers: Japan (left), Split, Croatia (right).

Number of cells at each hierarchical level (i.e., resolution) is calculated as  $6N_{side}^{2res}$ , where  $res$  is the resolution/hierarchical level that starts from 0 for the initial tessellation of a sphere or ellipsoid.

Tables 3.1 and 3.2 contain the number of cells across different resolutions of the rHEALPix DGGS for the 1-to-4 ( $N_{side} = 2$ ) and 1-to-9 ( $N_{side} = 3$ ) refinement, respectively. Tables also contain theoretical areas of cells as well as their square roots for the WGS 84 ellipsoid (EPSG, n.d.-a). Square

**Table 3.1** Number of cells, theoretical areas, and square roots of theoretical areas across rHEALPix DGGS resolutions in the 1-to-4 refinement ( $N_{side} = 2$ ) for the WGS 84 ellipsoid.

Resolution	Theoretical area (m <sup>2</sup> )	Square root of theoretical area (m)	Number of cells
0	85 010 936 954 014.70	9 220 137.58	6
1	21 252 734 238 503.60	4 610 068.79	24
2	5 313 183 559 625.92	2 305 034.39	96
3	1 328 295 889 906.48	1 152 517.20	384
4	332 073 972 476.62	576 258.60	1 536
5	83 018 493 119.16	288 129.30	6 144
6	20 754 623 279.79	144 064.65	24 576
7	5 188 655 819.95	72 032.32	98 304
8	1 297 163 954.99	36 016.16	393 216
9	324 290 988.75	18 008.08	1 572 864
10	81 072 747.19	9 004.04	6 291 456
11	20 268 186.80	4 502.02	$2.52 \times 10^7$
12	5 067 046.70	2 251.01	$1.01 \times 10^8$
13	1 266 761.67	1 125.51	$4.03 \times 10^7$
14	316 690.42	562.75	$1.61 \times 10^9$
15	79 172.60	281.38	$6.44 \times 10^9$
16	19 793.15	140.69	$2.58 \times 10^{10}$
17	4 948.29	70.34	$1.03 \times 10^{11}$
18	1 237.07	35.17	$4.12 \times 10^{11}$
19	309.27	17.59	$1.65 \times 10^{12}$
20	77.32	8.79	$6.60 \times 10^{12}$
21	19.33	4.40	$2.64 \times 10^{13}$
22	4.83	2.20	$1.06 \times 10^{14}$
23	1.21	1.10	$4.22 \times 10^{14}$
24	0.30	0.55	$1.69 \times 10^{15}$
25	0.08	0.27	$6.76 \times 10^{15}$
26	0.02	0.14	$2.70 \times 10^{16}$
27	$4.7 \times 10^{-3}$	0.07	$1.08 \times 10^{17}$
28	$1.2 \times 10^{-3}$	0.03	$4.32 \times 10^{17}$
29	$2.9 \times 10^{-4}$	0.02	$1.73 \times 10^{18}$
30	$7.4 \times 10^{-5}$	0.01	$6.92 \times 10^{18}$

**Table 3.2** Number of cells, theoretical areas, and square roots of theoretical areas across rHEALPix DGGS resolutions in the 1-to-9 ( $N_{side} = 3$ ) refinement for the WGS 84 ellipsoid.

Resolution	Theoretical area (m <sup>2</sup> )	Square root of theoretical area (m)	Number of cells
0	85 010 936 954 014.70	9 220 137.58	6
1	9 445 659 661 557.20	3 073 379.19	54
2	1 049 517 740 173.02	1 024 459.73	486
3	116 613 082 241.45	341 486.58	4 374
4	12 957 009 137.94	113 828.86	39 366
5	1 439 667 681.99	37 942.95	354 294
6	159 963 075.78	12 647.65	3 188 646
7	17 773 675.09	4 215.88	$2.87 \times 10^7$
8	1 974 852.79	1 405.29	$2.58 \times 10^8$
9	219 428.09	468.43	$2.32 \times 10^9$
10	24 380.90	156.14	$2.09 \times 10^{10}$
11	2 708.99	52.05	$1.88 \times 10^{11}$
12	301.00	17.35	$1.69 \times 10^{12}$
13	33.44	5.78	$1.53 \times 10^{13}$
14	3.72	1.93	$1.37 \times 10^{14}$
15	0.41	0.64	$1.24 \times 10^{15}$
16	0.05	0.21	$1.11 \times 10^{16}$
17	$5.1 \times 10^{-3}$	0.07	$1.00 \times 10^{17}$
18	$5.7 \times 10^{-4}$	0.02	$9.01 \times 10^{17}$
19	$6.3 \times 10^{-5}$	0.01	$8.11 \times 10^{18}$

roots of areas serve as an approximate measure of sizes of cells, provided they are considered to be squares. In both tables, values are given up to a resolution for which cell size can be approximated as  $1 \times 1$  cm. The same values, except number of cells, are for more parameter  $N_{side}$  values provided in Appendix B.

Although in plane all rHEALPix DGGS cells are squares, Gibb et al. (2013) classify their equivalents on the sphere or ellipsoid in four different categories, based on their shapes: quad cells in equatorial region, and cap, dart, and skew quad cells in the north and south polar regions (visible on Figure 3.6). Quad cells have four edges; two are arcs of parallels, and two are arcs of meridians. Cap cells are bounded by one parallel, and they are centered at the north or south pole. If  $N_{side}$  is even, then there are no cap cells besides N and S resolution-0 cells. Dart cells have three edges, one of them

is arc of a parallel and the other two are directed towards the corresponding pole, while all other cells in the polar region are skew quad cells whose southern and northern edges are arcs of a parallel.

### 3.3.1 Enhancing authalic latitude calculation

Although spherical calculations are less complex than those on an ellipsoid of revolution (Dimitrijević et al., 2023), when high accuracy is required, substituting a sphere for an ellipsoid is usually not acceptable. Nevertheless, certain problems involving ellipsoids of revolution can be addressed by first transforming the ellipsoid to a sphere, followed by calculations on the spherical model. This transformation typically involves retaining the longitude while computing an auxiliary latitude on the sphere that is derived from the geodetic latitude on the ellipsoid. Various forms of auxiliary latitudes exist, each with distinct properties that are suitable for specific application (Karney, 2024). Among them, the authalic latitude is particularly interesting for equal area DGGSs, as it enables mapping from an ellipsoid to a sphere while preserving areas. Of course, when performing area-preserving mapping from sphere to ellipsoid, they both must have the same area. Sphere that has the same area as ellipsoid is referred to as its authalic sphere, and radius of such sphere as authalic radius.

From the available DGGS implementations, rHEALPix DGGS is one of the few that, in addition to spheres, has a built-in support for ellipsoids of revolution. DGGS implementations such as Uber's hexagonal H3 (Uber, n.d.) or Google's quadrangular S2 Geometry (Google, n.d.) DGGSs are currently available for spherical surfaces only. As already stated, rHEALPix extends the HEALPix DGGS so that it can define DGGS on ellipsoids of revolution. If DGGS has to be defined on an ellipsoid of revolution, rHEALPix DGGSs is first defined on its authalic sphere and then grids are mapped to the ellipsoid by calculating geodetic latitudes from the corresponding authalic latitudes. Since all cells on a particular refinement level of the rHEALPix DGGS have the same area on the sphere and since mapping from sphere to ellipsoid is area-preserving, equal-area property of rHEALPix DGGS is also retained on ellipsoid. This means that the  $(n, s)$ -rHEALPix map projection that enables constructing rHEALPix DGGS on ellipsoid, defines a double mapping, first from plane (i.e., faces of the cube) to sphere and then from sphere to ellipsoid.

Currently, rHEALPix DGGS Python package (Raichev et al., 2023) uses equations that are provided by Snyder (1987). Equation for obtaining authalic latitude  $\phi_A$  from the geodetic latitude  $\phi$  is:

$$\phi_A = \arcsin \frac{(1 - e^2) \left[ \frac{\sin \phi}{1 - e^2 \sin^2 \phi} - \frac{1}{2e} \ln \frac{1 - e \sin \phi}{1 + e \sin \phi} \right]}{1 - \frac{1 - e^2}{2e} \ln \frac{1 - e}{1 + e}}, \quad (1)$$

where  $e$  is eccentricity calculated as  $e = \sqrt{(a^2 - b^2)/a^2}$  from semi-major axis  $a$  and semi-minor axis  $b$  of the ellipsoid of revolution. For calculating geodetic latitude  $\phi$  from authalic latitude  $\phi_A$ , Snyder (1987) defines power series expansion in terms of eccentricity  $e$ , up to  $e^6$ :

$$\begin{aligned} \phi \approx \phi_A &+ \left( \frac{1}{3} e^2 + \frac{31}{180} e^4 + \frac{517}{5040} e^6 \right) \sin 2\phi_A \\ &+ \left( \frac{23}{360} e^4 + \frac{251}{3780} e^6 \right) \sin 4\phi_A \\ &+ \left( \frac{761}{45360} e^6 \right) \sin 6\phi_A. \end{aligned} \quad (2)$$

Arguably the most widely used number representation in modern computers, IEEE (Institute of Electrical and Electronics Engineers) double-precision (64-bit) floating-point, enables storing numbers with approximately 16 significant decimal digits (Overton, 2001, p. 24). This in general means that all calculations on double-precision floating-point numbers will generate a hardware-related round off-error (Press et al., 1992, p. 30). This round-off error is then being accumulated with each additional operation and if not taken care of, can, in some circumstances, drastically degrade calculation accuracy. Accuracy of results obtained by Equations 1 and 2 are therefore both affected by round-off errors. While Equation 1 is provided in closed-form, Equation 2 gives an approximate solution that is based on truncated power series expansion. This means that results obtained by Equation 2, in addition to round-off errors, which is hardware-related, also suffer from the truncation error, which can be treated as implementation-related. Although rHEALPix DGGS on ellipsoid is analytically equal area DGGS, because of round-off and truncation errors, it will never be numerically equal-area. However, it is possible to reduce these errors in calculating authalic latitude from geodetic and vice versa. Making these calculations more accurate will bring

$(n, s)$ -rHEALPix map projection closer to being numerically area-preserving and then subsequently rHEALPix DGGS closer to being numerically equal area DGGS.

Karney (2024) provides equations for calculating authalic latitude from geodetic and vice versa in a form of power-series expansion in terms of third flattening  $n$ . Third flattening is calculated as  $n = (a - b)/(a + b)$ . He provides power-series expansions up to  $n^6$  and confirms that these equations enable fully accurate double-precision calculations for flattenings  $|f| \leq 1/150$  (flattening of the WGS 84 ellipsoid (EPSG, n.d.-a) is  $1/298.257223563$ ). Karney (2024) additionally suggests optimizing calculations of power-series expansions by applying Horner method and Clenshaw summation (Brenner et al., 1955). These optimizations are aimed at further reducing round-off errors and reducing the execution time. After applying Horner method, equation for calculating authalic latitude  $\phi_A$  from geodetic latitude  $\phi$  is given by:

$$\begin{aligned} \phi_A \approx & \phi + n \left( -\frac{4}{3} + n \left( -\frac{4}{45} + n \left( \frac{88}{315} + n \left( \frac{538}{4725} + n \left( \frac{20824}{467775} - n \frac{44732}{2837835} \right) \right) \right) \right) \right) \sin 2\phi \\ & + n \left( n \left( \frac{34}{45} + n \left( \frac{8}{105} + n \left( -\frac{2482}{14175} + n \left( -\frac{37192}{467775} - n \frac{12467764}{212837625} \right) \right) \right) \right) \right) \sin 4\phi \\ & + n \left( n \left( n \left( -\frac{1532}{2835} + n \left( -\frac{898}{14175} + n \left( \frac{54968}{467775} + n \frac{100320856}{1915538625} \right) \right) \right) \right) \right) \sin 6\phi \\ & + n \left( n \left( n \left( n \left( \frac{6007}{14175} + n \left( \frac{24496}{467775} - n \frac{5884124}{70945875} \right) \right) \right) \right) \right) \sin 8\phi \\ & + n \left( n \left( n \left( n \left( n \left( -\frac{23356}{66825} - n \frac{839792}{19348875} \right) \right) \right) \right) \right) \sin 10\phi \\ & + n \left( n \left( n \left( n \left( n \left( n \left( \frac{570284222}{1915538625} \right) \right) \right) \right) \right) \right) \sin 12\phi \end{aligned} \quad (3)$$

and for the inverse direction, calculating geodetic latitude  $\phi$  from authalic latitude  $\phi_A$ :

$$\begin{aligned} \phi \approx & \phi_A + n \left( \frac{4}{3} + n \left( \frac{4}{45} + n \left( -\frac{16}{35} + n \left( -\frac{2582}{14175} + n \left( \frac{60136}{467775} + n \frac{28112932}{212837625} \right) \right) \right) \right) \right) \sin 2\phi_A \\ & + n \left( n \left( \frac{46}{45} + n \left( \frac{152}{945} + n \left( -\frac{11966}{14175} + n \left( -\frac{21016}{51975} + n \frac{251310128}{638512875} \right) \right) \right) \right) \right) \sin 4\phi_A \\ & + n \left( n \left( n \left( \frac{3044}{2835} + n \left( \frac{3802}{14175} + n \left( -\frac{94388}{66825} - n \frac{8797648}{10945935} \right) \right) \right) \right) \right) \sin 6\phi_A \\ & + n \left( n \left( n \left( n \left( \frac{6059}{4725} + n \left( \frac{41072}{93555} - n \frac{1472637812}{638512875} \right) \right) \right) \right) \right) \sin 8\phi_A \\ & + n \left( n \left( n \left( n \left( n \left( \frac{768272}{467775} + n \frac{455935736}{638512875} \right) \right) \right) \right) \right) \sin 10\phi_A \\ & + n \left( n \left( n \left( n \left( n \left( n \left( \frac{4210684958}{1915538625} \right) \right) \right) \right) \right) \right) \sin 12\phi_A. \end{aligned} \quad (4)$$

We also tried applying Clenshaw summation, but experimental assessment of our implementation showed that execution time was slightly prolonged with almost no significant

increase in calculation accuracy. Therefore, in the final implementation of Equations 3 and 4 we did not use Clenshaw summation.

In the current rHEALPix DGGS Python package, we replaced statements that were based on Equations 1 and 2 (hereafter, old statements) with statements based on Equations 3 and 4 (hereafter, new statements). Afterward, we assessed old and new statements from the perspective of calculation accuracy and processing time. To assess calculation accuracy, we employed the *mpmath* Python package (The mpmath development team, 2023) that enables floating-point calculations with arbitrary precision. This arbitrary precision was set to be much higher than the 64-bit double-precision so that results obtained by *mpmath* package can be used as a reliable baseline for comparisons. Regarding accuracy of conversion from geodetic to authalic latitude, we first performed regular double-precision floating-point calculations based on Equations 1 and 3. Then we again performed calculations based on Equation 1 but with *mpmath* package. Differences between authalic latitude obtained by the *mpmath* package and each of the two double-precision floating-point-based calculated authalic latitudes is considered as a calculation error. We carried out these calculations across all latitudes from  $-90^\circ$  to  $90^\circ$  with a resolution of  $0.1^\circ$ .

For the inverse direction, we applied a similar approach. For the input authalic latitude we used all authalic latitudes obtained by the *mpmath* in the previous step. This authalic latitude is at a double-precision level without calculation errors and for each of these latitudes we know what the expected error-free resulting geodetic latitudes are—those that were used as inputs for assessment in the opposite conversion direction (values from  $-90^\circ$  to  $90^\circ$  with resolution of  $0.1^\circ$ ). We had to apply this approach since there is no available closed-form equation for converting authalic to geodetic latitude. Therefore, for assessing calculation accuracy of Python statements that are based on Equations 2 and 4 we did not need to perform any calculations with *mpmath* package, but only regular double-precision floating-point calculations. Calculation errors were in this case calculated by subtracting obtained geodetic latitudes from geodetic latitudes that were used as an input for assessing calculation accuracy in geodetic to authalic conversion direction.

To examine the impact of calculation errors in converting geodetic to authalic latitude and vice versa on the rHEALPix DGGS cell areas, we calculated areas of cells on the WGS 84 ellipsoid by following procedure from Section 3.4.1. We did not perform comprehensive analysis for all cells on



[illegible]

47

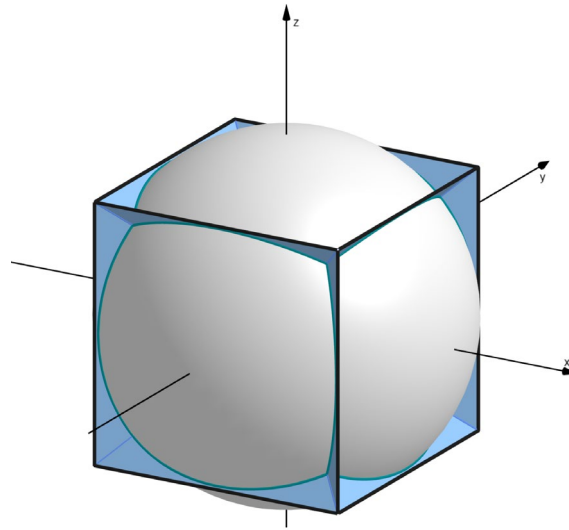
also two different versions of old statements. First was the unaltered version that uses functions from the *NumPy* Python package (Harris et al., 2020). In the second version, all *NumPy* functions were replaced by functions from the *math* module without any other modifications. *NumPy* functions are optimized for working on arrays while statements for converting authalic to geodetic latitude take a single numerical value as input. Therefore, it is expected that *math* functions will be more time efficient since they natively perform calculations on single values (Sundnes, 2020, p. 85). This means that processing time assessment was conducted for three groups of statements—unaltered old statements, old statements in which *NumPy* functions were replaced with corresponding *math* functions, and new statements. For each group of statements, processing times were measured in both conversion directions. In order to examine whether input latitude has any influence on the processing time, we measured all these processing times across all altitudes with resolution of 1°. Additionally, we measured a total processing time that is required for converting all latitudes from -90° to 90°, with varying steps between them: 1°, 10', 1'', 10'', 1'', 0.1'', and 0.01''. These different steps mean that statements were executed different number of times, from 181 for step of 1° to 64 800 001 for step of 0.01''.

Here we also provide the main characteristics of the workstation that was used for all these calculations as it might be particularly relevant from the processing time perspective—Windows workstation with an Intel® Xeon® Silver 4216 CPU and 64 GB of DDR4 RAM. Also, as a proxy for the surface of the Earth, WGS 84 ellipsoid (EPSG, n.d.-a) was used in all calculations. Results obtained by all these assessments are provided in Section 3.5.1 and corresponding discussion in Section 4.1.

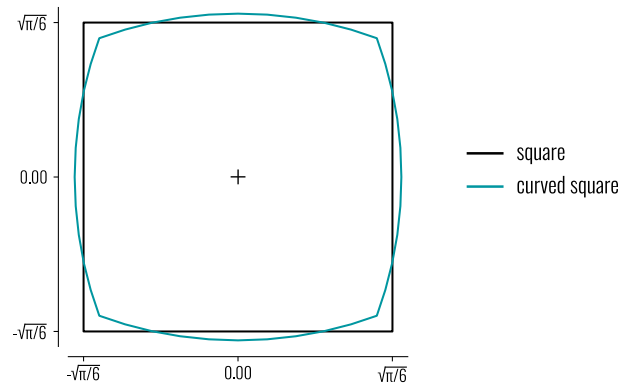
### 3.3.2 Modifying shapes of the rHEALPix DGGS cells

One of the disadvantages of the rHEALPix DGGS is the fact that different projections are used for equatorial and for polar regions. For equatorial region, Lamber cylindrical equal area map projection was used, and for polar regions Collignon projection (R. Gibb et al., 2013). This means that even on the first hierarchical level shapes of the cells on sphere/centroid are not the same. Roşca and Plonka (2011) developed a different approach. Unit sphere centered in the origin of the 3D Cartesian coordinate system is first divided into six equal parts (i.e., spherical quadrangles of equal area) by diagonal planes  $y = \pm x$ ,  $z = \pm y$ , and  $z = \pm x$  (Figure 3.8). Each of these quadrangles is

separately projected in plane with Lambert azimuthal equal-area (LAEA) map projection where intersection between a particular coordinate axis and quadrangle being projected is taken as a projection center (Figure 3.9). The projection of each of these spherical quadrangles are, as Roşca and Plonka (2011) put it, ‘curved squares’ that have the same area as spherical quadrangles. These curved squares are then mapped to a rectangular square also by applying area-preserving mapping (we refer to this mapping as Roşca-Plonka projection). This double mapping (from spherical quadrangles to curved square in plane and finally to rectangular square in plane) can be geometrically interpreted as mapping spherical quadrangles to the corresponding faces of the cube that is centered in the origin of the 3D Cartesian coordinate system (Figure 3.8). The area of each face of the cube is the same as the area of spherical quadrangles, which means that the length of the edge  $a$  of the cube is  $\sqrt{2\pi/3}$ . This cube is oriented in a way that its diagonal planes correspond to the diagonal planes used to obtain spherical quadrangles (Figure 3.8).



**Figure 3.8** Intersecting unit sphere with diagonal planes of the axis-aligned cube to obtain spherical quadrangles.

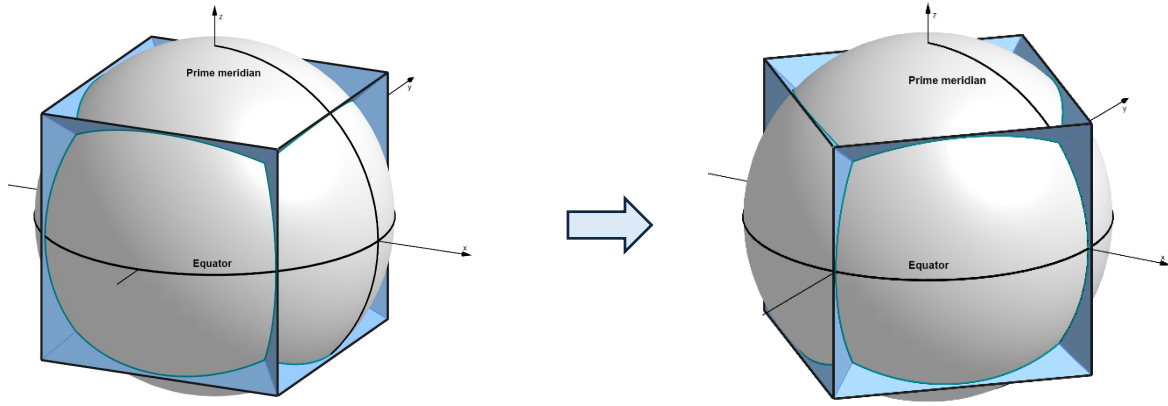


**Figure 3.9** LAEA projection of spherical quadrangle (curved square) and rectangular square of the same area.

The inverse of this projection can geometrically be interpreted as area-preserving mapping from faces of the cube to the sphere. Each face of the cube is mapped to the corresponding spherical quadrangle. This means that if we decompose each face of the cube with a rectangular grid in which grid elements are square cells of equal area, we can map it back to the sphere in an area-preserving manner, while ensuring that the whole sphere is covered with this grid. This procedure essentially produces DGG on the sphere. If we further define a system of hierarchically organized grids that are recursively partitioning (J. A. Kimerling et al., 1999) each face of the cube, then we get geometrical component of the DGGS on the sphere. To consider these hierarchically organized grids as a DGGS, additionally at least a mechanism for assigning unique identifiers to each grid element (i.e., cell) is needed. Since rHEALPix DGGS cells in the plane are also squares and since it already has a mechanism for assigning unique identifiers to each cell and other various functions for generating DGGS and performing some spatial analysis with cells and grids, we decided to modify rHEALPix DGGS by modifying a way of how square grid is mapped from the plane to the sphere. The main aim of this modification is to make shapes of the cells on sphere or ellipsoid more stable. We refer to this modified DGGS as QPix DGGS and below we explain steps that were taken to implement it.

We first consider a unit sphere which can later easily be generalized to any sphere of arbitrary radius, or any spheroid (i.e., ellipsoid of revolution, oblate or prolate). For the rHEALPix DGGS, hierarchical grids, unique identifier assigning mechanism, and many other functions are defined in plane, and then, if needed, transferred to the sphere or ellipsoid. Therefore we also start from plane in order to make use of as much as possible of the current rHEALPix DGGS implementation written in Python. Figure 3.3 shows projection of the spheroid to the plane using (0, 0)-rHEALPix projection. Projection of the north polar region is considered as resolution-0 cell N, while projection of the south polar region is considered as resolution-0 cell S. Projection of the equatorial region is divided in four resolution-0 cells: O, P, Q, and R. All resolution-0 cells are squares of equal size and in each subsequent resolution (i.e., hierarchical level of the DGGS) they are being uniformly partitioned in square grid and mapped back to the sphere. Using analogous notation, we define  $(n, s)$ -QPix projection by arranging Roşca-Plonka projections of each previously defined spherical quadrangle so that they are organized in the same way as  $(n, s)$ -rHEALPix projection of resolution-0 cells of the rHEALPix DGGS.

In more practical terms, the procedure is the following. Cube is rotated by  $45^\circ$  around z-axis in negative direction (Figure 3.10) for reasons discussed later. Each spherical quadrangle constructed by intersecting diagonals of the cube with sphere is separately projected in plane by LAEA map projection where corresponding quadrangle centroid is taken as projection center. Spherical quadrangle projection is curved square (Figure 3.9).

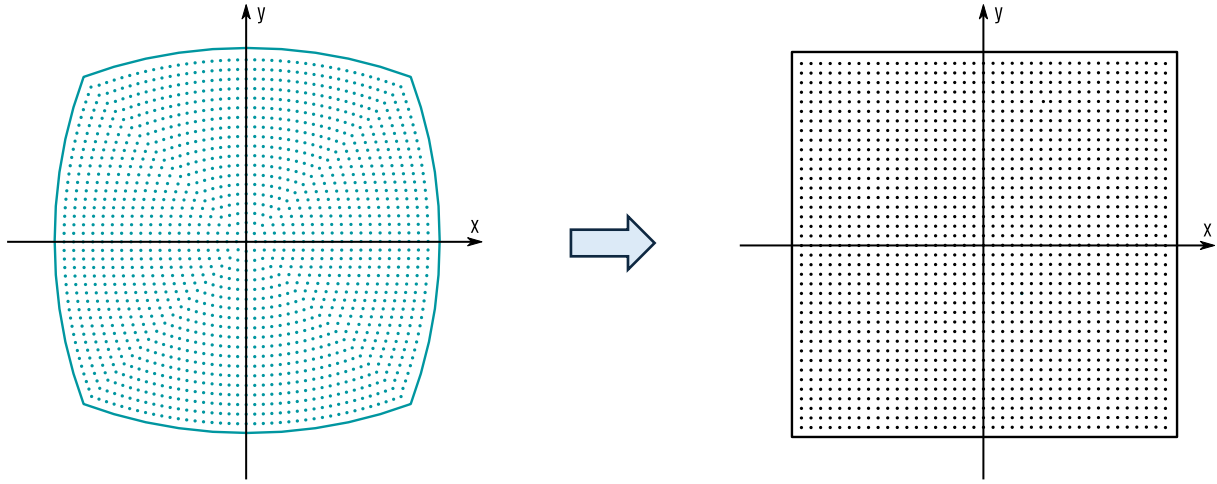


**Figure 3.10** Rotating the cube by  $45^\circ$  around z-axis.

Point  $(x_c, y_c)$  related to the curved square is mapped to a point  $(x_s, y_s)$  related to the square (Figure 3.11) using equations derived by Roşca and Plonka (2011):

$$\left\{ \begin{array}{l} x_s = \frac{\beta}{\sqrt{2}} \operatorname{sgn}(x_c) \sqrt[4]{2x_c^2 + y_c^2} \sqrt{|x_c| + \sqrt{2x_c^2 + y_c^2}} \\ y_s = \frac{\sqrt{2}}{\beta} \sqrt[4]{2x_c^2 + y_c^2} \sqrt{|x_c| + \sqrt{2x_c^2 + y_c^2}} \left( \operatorname{sgn}(x_c) \arctan \frac{y_c}{x_c} - \arctan \frac{y_c}{\sqrt{2x_c^2 + y_c^2}} \right), \text{ if } |y_c| \leq |x_c| \\ x_s = \frac{\sqrt{2}}{\beta} \sqrt[4]{x_c^2 + 2y_c^2} \sqrt{|y_c| + \sqrt{x_c^2 + 2y_c^2}} \left( \operatorname{sgn}(x_c) \arctan \frac{x_c}{y_c} - \arctan \frac{x_c}{\sqrt{x_c^2 + 2y_c^2}} \right), \text{ if } |x_c| \leq |y_c| \\ y_s = \frac{\beta}{\sqrt{2}} \operatorname{sgn}(y_c) \sqrt[4]{x_c^2 + 2y_c^2} \sqrt{|y_c| + \sqrt{x_c^2 + 2y_c^2}} \end{array} \right. \quad (5)$$

where  $\beta = \sqrt{\pi/6}$ . We refer to this projection as Roşca-Plonka projection.



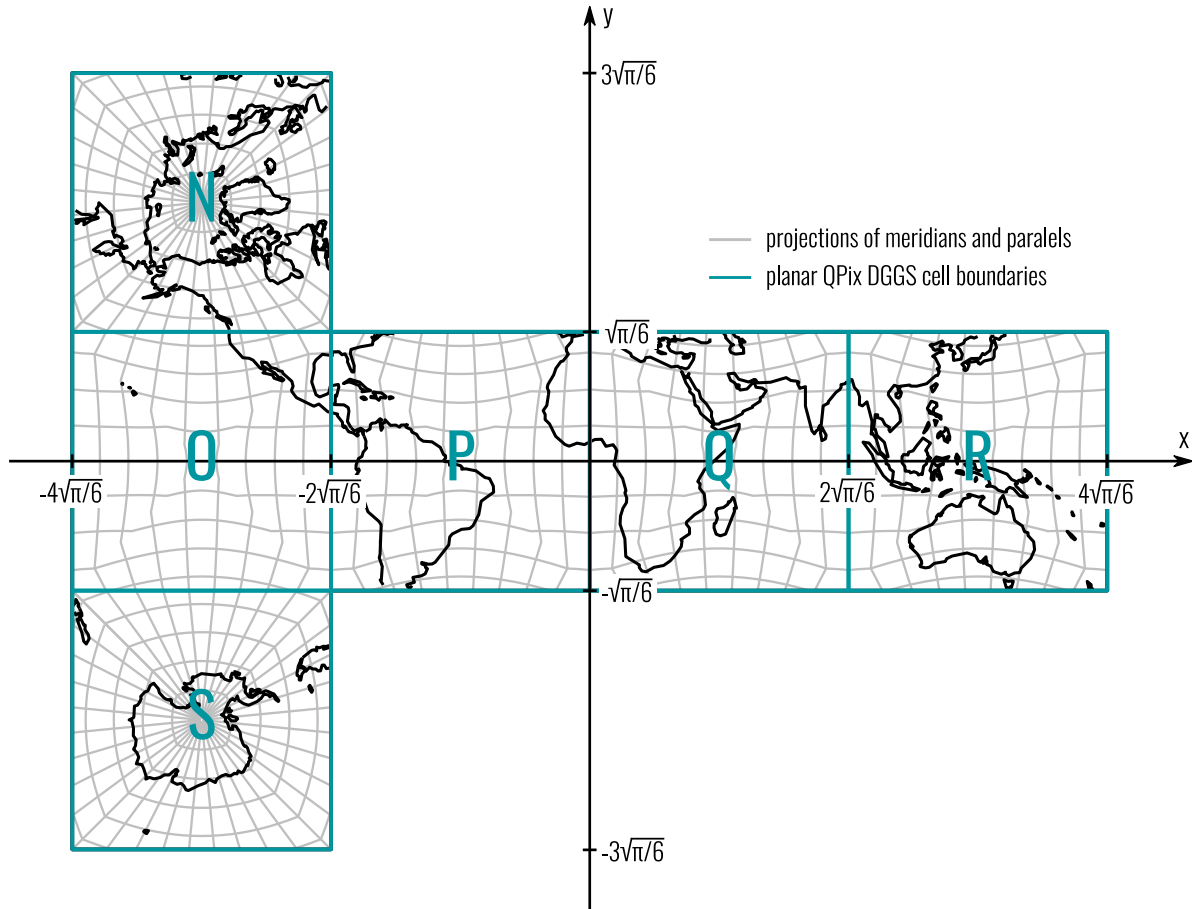
**Figure 3.11** Mapping curved square to square (Roşca-Plonka projection).

Each square is then translated within the  $xy$  coordinate system in order to ensure that the whole sphere is projected within the single coordinate system:

- square that is a projection of spherical quadrangle that has a centroid on the equator with latitude of  $-135^\circ$  is translated by  $-3\sqrt{\pi/6}$  along x-axis and is denoted as square  $O$ ,
- square that is a projection of spherical quadrangle that has a centroid on the equator with latitude of  $-45^\circ$  is translated by  $-\sqrt{\pi/6}$  along x-axis and is denoted as square  $P$ ,
- square that is a projection of spherical quadrangle that has a centroid on the equator with latitude of  $45^\circ$  is translated by  $\sqrt{\pi/6}$  along x-axis and is denoted as square  $Q$ ,
- square that is a projection of spherical quadrangle that has a centroid on the equator with latitude of  $135^\circ$  is translated by  $3\sqrt{\pi/6}$  along x-axis and is denoted as square  $R$ ,
- square that is a projection of spherical quadrangle that has a centroid that matches the north pole is translated by  $2\sqrt{\pi/6}$  along y-axis and by  $(2n - 3)\sqrt{\pi/6}$  ( $n \in \{1, 2, 3, 4\}$ ) along x-axis is denoted as square  $N$ , and
- square that is a projection of spherical quadrangle that has a centroid that matches the south pole is translated by  $-2\sqrt{\pi/6}$  along y-axis and by  $(2s - 3)\sqrt{\pi/6}$  ( $s \in \{1, 2, 3, 4\}$ ) along x-axis is denoted as square  $S$ .

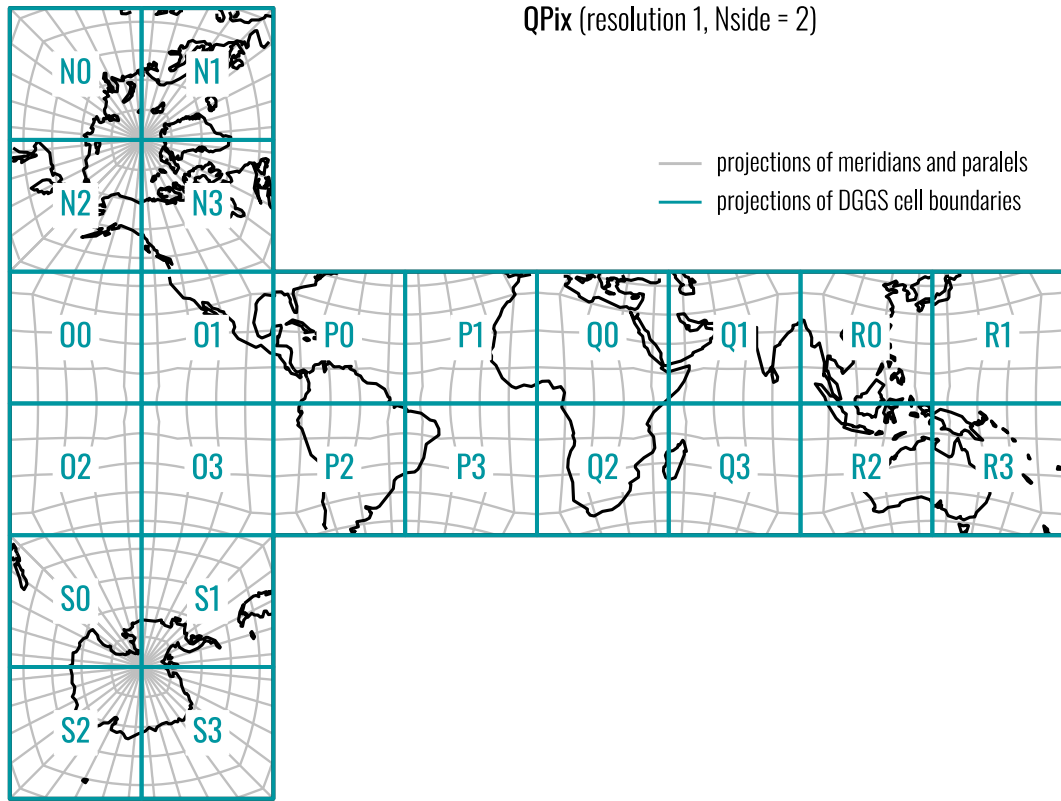
The above procedure defines the  $(n, s)$ -QPix map projection of a unit sphere to the plane (Figure 3.12). Value  $n$  defines above which of the squares that contain projection of equator ( $O$ ,  $P$ ,  $Q$ , and  $R$ ) the  $N$  square will be placed. If  $n = 0$ ,  $N$  is placed above square  $O$ , if  $n = 1$ ,  $N$  is placed above square  $P$ , if  $n = 2$ ,  $N$  is placed above square  $Q$ , and if  $n = 3$ ,  $N$  is placed above square  $R$ . Value  $s$  analogously defines below which of these squares the  $S$  square will be placed. These squares denoted as  $N$ ,  $O$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$  are considered as planar projections of resolution-0 QPix DGGs cells. Within each of these squares, a hierarchical square grid can be defined and mapped back to the sphere.

In order to make the most of the rHEALPix DGGs implementation in Python, we had to replace the hard-coded extents of the  $(n, s)$ -rHEALPix projection of the sphere with the extents of the  $(n, s)$ -QPix (extents of the  $(0, 0)$ -rHEALPix map projection of the unit sphere are marked on Figure 3.3, and extents of the  $(0, 0)$ -QPix projection on Figure 3.12). After this step, the same Python functions that are used to generate planar version of the rHEALPix DGGs can be used to generate



**Figure 3.12** Projection of the unit sphere and the resolution 0 QPix DGGs cells in the plane by the  $(0, 0)$ -QPix map projection.

the planar version of the QPix DGGs (resolution 1 cells for  $N_{side} = 2$  shown on Figure 3.13 and for  $N_{side} = 3$  on Figure 3.14; resolution 2 cells are shown in Appendix A.4). Regarding the initial rotation of the cube around z-axis by  $-45^\circ$ , this step was performed so that the projection of prime meridian by the  $(n, s)$ -QPix projection coincides with the y-axis in the planar coordinate system, which is also the case with the  $(n, s)$ -rHEALPix projection. This step is not necessary, but it was performed so that arrangement of the resolution-0 cells in relation to the spherical/geodetic coordinates is for the QPix DGGs kept in resemblance with the rHEALPix DGGs.



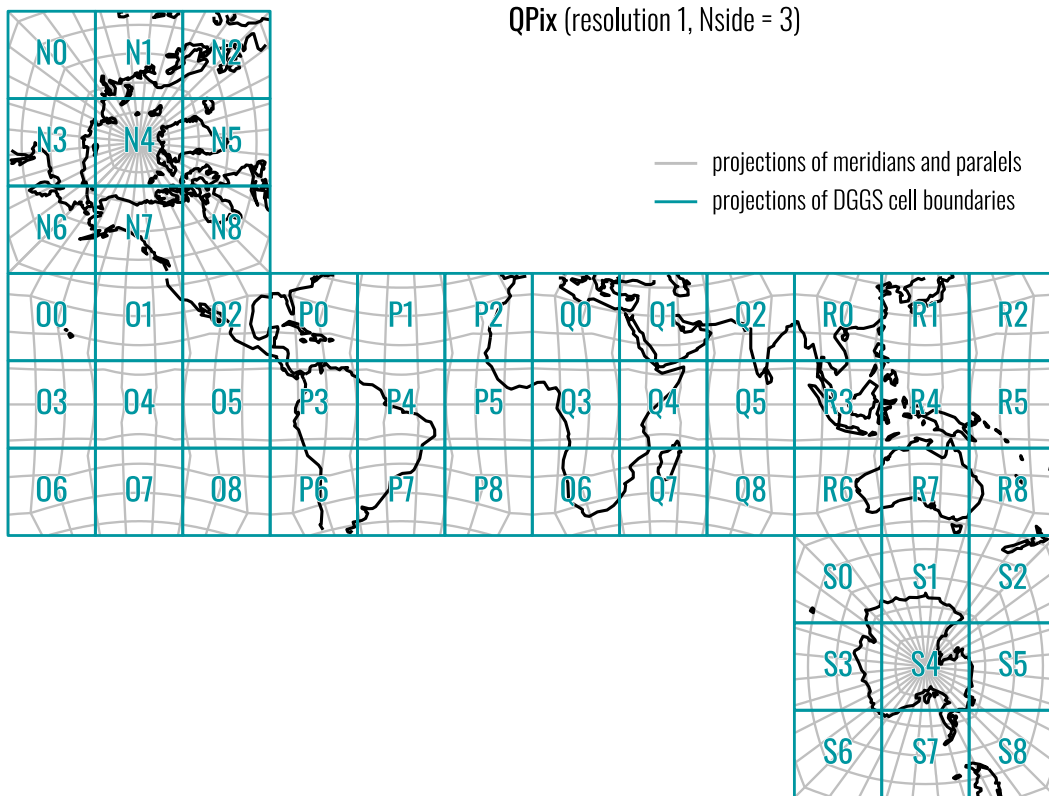
**Figure 3.13** (0, 0)-QPix map projection of the sphere/ellipsoid and resolution 1,  $N_{side} = 2$  QPix DGGs cells.

To map a square grid from each of the six squares back to the sphere, the inverse of the mapping from the curved square to square has to be applied for each square and then curved squares have to be mapped to the sphere by the inverse LAEA projection. Rosca and Plonka (2011) give equations for mapping a point  $(x_s, y_s)$  that is related to the square to a point  $(x_c, y_c)$  that is related to the curved square (Figure 3.11, but in opposite direction):



$$\left\{ \begin{array}{l} x_c = \frac{2^{1/4} x_s}{\beta} \frac{\sqrt{2} \cos\left(\frac{y_s \pi}{12 x_s}\right) - 1}{\sqrt{\sqrt{2} - \cos\left(\frac{y_s \pi}{12 x_s}\right)}}, \text{ if } |y_s| \leq |x_s| \\ y_c = \frac{2^{1/4} x_s}{\beta} \frac{\sqrt{2} \sin\left(\frac{y_s \pi}{12 x_s}\right)}{\sqrt{\sqrt{2} - \cos\left(\frac{y_s \pi}{12 x_s}\right)}} \\ \\ x_c = \frac{2^{1/4} y_s}{\beta} \frac{\sqrt{2} \sin\left(\frac{x_s \pi}{12 y_s}\right)}{\sqrt{\sqrt{2} - \cos\left(\frac{x_s \pi}{12 y_s}\right)}}, \text{ if } |x_s| \leq |y_s| \\ y_c = \frac{2^{1/4} y_s}{\beta} \frac{\sqrt{2} \cos\left(\frac{x_s \pi}{12 y_s}\right) - 1}{\sqrt{\sqrt{2} - \cos\left(\frac{x_s \pi}{12 y_s}\right)}} \end{array} \right. \quad (6)$$

where again  $\beta = \sqrt{\pi/6}$ . We refer to this projection as inverse Rořca-Plonka projection.



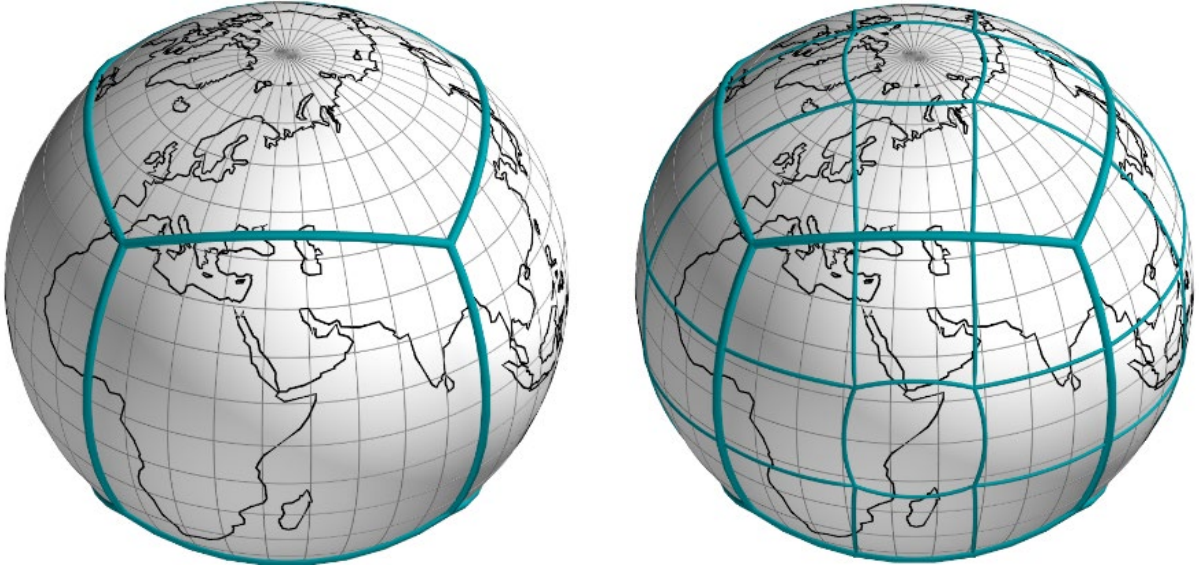
**Figure 3.14** (0, 3)-QPix map projection of the sphere/ellipsoid and resolution 1,  $N_{\text{side}} = 2$  QPix DGGs cells. The same cells visualized in Robinson projection are available in Appendix A.5.

The mapping described above is related to the unit sphere, however, it can be easily extended to spheres of arbitrary radius or to arbitrary spheroids. Extension to the spheres is straightforward. If

$R_A$  is the radius of the sphere and if the point on this sphere has a spherical coordinates  $(\lambda, \phi)$ , then its  $(n, s)$ -QPix projection is point with coordinates  $(R_A x_s, R_A y_s)$ . Here  $x_s$  and  $y_s$  are coordinates obtained by mapping a point  $(\lambda, \phi)$  to the plane by the  $(n, s)$ -QPix projection defined for the unit sphere. In essence, this means that coordinates  $(x_s, y_s)$  obtained by  $(n, s)$ -QPix projection defined on a unit sphere need to be scaled by  $R_A$ . For the inverse direction, point coordinates  $(R_A x_s, R_A y_s)$  in a map projection plane first need to be ‘scaled down’ by  $R_A$  to  $(x_s, y_s)$  and then by applying inverse of the  $(n, s)$ -QPix map projection for unit sphere mapped to the spherical coordinates  $(\lambda, \phi)$ . These spherical coordinates are the same for the unit sphere as they are for sphere with arbitrary radius  $R_A$ .

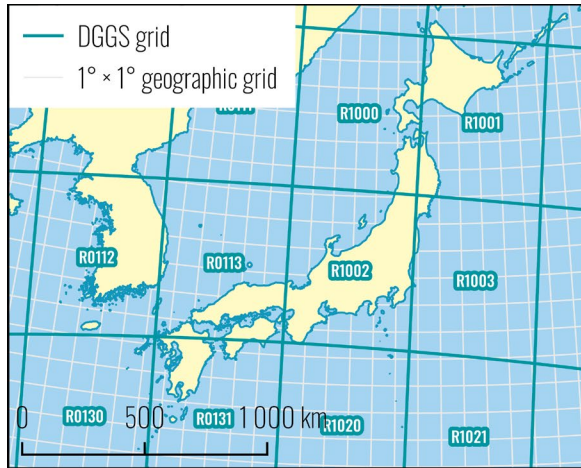
Extension to ellipsoids is also simple and is the same for QPix DGGs as it is for rHEALPix DGGs—spheroid can be mapped to a sphere by retaining longitude and calculating an auxiliary latitude on sphere (Karney, 2024). Since area preservation is crucial, as an auxiliary latitude we chose an authalic latitude. Point on an ellipsoid is first mapped to the sphere with radius  $R_A$  by calculating authalic latitude from its geodetic latitude and then it is mapped from sphere to the plane as described before. Here  $R_A$  is the radius of the authalic sphere of the ellipsoid, i.e., radius of the sphere that has the same area as ellipsoid. If point is being mapped by the  $(n, s)$ -QPix projection from the map projection plane to the ellipsoid, then after it is being mapped on the sphere, it is mapped to the ellipsoid by calculating geodetic from authalic latitude. The  $(n, s)$ -QPix map projection of ellipsoid to the plane can thus be considered as triple mapping: (1) mapping ellipsoid to sphere, (2) mapping each of the six spherical quadrangles to the plane by LAEA projection, and (3) mapping each curved square obtained by LAEA projection to square by the Rořca-Plonka projection followed by translating all squares so that they do not overlap within the same coordinate system. QPix DGGs cells that are mapped from the plane to the spheroid by the  $(n, s)$ -QPix projection are visualized on Figure 3.15 (more visualizations are available in Appendices A.1 and A.2). Figure 3.16 shows QPix DGGs grids for resolution 4 and 11 in 1-to-4 and 1-to-9 refinements respectively, along with cell identifiers (more figures are available in Appendix A.3).

Regarding the area of QPix DGGs cells on the sphere or ellipsoid, they are the same as for the rHEALPix DGGs, which means that values in Tables 3.1 and 3.2 are applicable to the QPix DGGs.



**Figure 3.15** Resolution-0 (left) and resolution-1 (right) QPix DGGs cells on the ellipsoid for the  $N_{\text{side}} = 3$ .

We compared QPix DGGs and rHEALPix DGGs from the perspective of stability of cell shapes. In the following section, Section 3.4, we first explain how areas and perimeters of cells were calculated, which are needed for calculating compactnesses of cells. Results for rHEALPix and QPix DGGs are provided in Section 3.5.2, and related discussion in Section 4.2. Discussion in Section 4.2 is not limited to compactness assessment only, but also includes an overall comparison between grids in rHEALPix and QPix DGGs.



QPix (1-to-4, resolution 4)



QPix (1-to-9, resolution 11)

**Figure 3.16** QPix DGGs grids and cell identifiers: Japan (left), Split, Croatia (right).

### 3.4 Evaluating geometric properties of rHEALPix and QPix DGGSs

DGGS should hierarchically tessellate the earth's surface (i.e., sphere or ellipsoid) in a highly regular way (Alderson et al., 2020, p. 26). Full regularity, i.e., tessellation with polygons that are congruent and regular, is not achievable on a curved surface (R. Gibb et al., 2021), however, some DGGSs exhibit higher, and others lower degree of regularity. Regularity in this sense can be observed as a level of distortion of areas and shapes of a DGGS cells (J. A. Kimerling et al., 1999). Kimerling et al. (1999) suggest that geometric structure of a DGGSs should be compared by calculating surface area and compactness of DGGS cell, as well as other metrics, across multiple refinement levels (i.e. resolutions). Compactness (defined in Section 3.4.3) is used as a measure of DGGS cell shape distortion. The higher the stability of area and compactness across cells on a specific refinement level, the higher the tessellation regularity. In order to compare compactness stability across different DGGS refinement levels (i.e., resolutions), White et al. (1998) calculated ranges (i.e., difference between maximum and minimum value), standard deviations, and means of compactness values of all cells on a single level. We also followed their approach and have calculated these values for up to resolution 8 in 1-to-4 refinement and resolution 5 in 1-to-9 refinement for rHEALPix and QPix DGGSs. Results of these calculations are available in Section 3.5.2 and associated discussion in Section 4.2.

Since rHEALPix and QPix DGGS are analytically equal-area DGGSs, calculating cell areas might be interesting only to determine numerical stability of calculations involved in generating DGGS cells on a sphere or ellipsoid (Section 3.3.1) and much less as a means of determining geometrical properties of a DGGS. Nevertheless, since we wanted to keep our calculations free from theoretical assumptions as much as possible, we still performed cell area calculations from the generated cell geometries on the spheroid. For calculating compactness of a cell, along with area, its perimeter is also required. Therefore, in the following sections, we describe a methodology applied for calculating area, perimeter, and compactness of cells across refinement levels for rHEALPix and QPix DGGS cells.

#### 3.4.1 Calculating area of DGGS cells

DGGS cells are polygons on a sphere or spheroid. To determine their area, different approaches can be applied. Since analogous calculations on spheres are generally simpler than on spheroids, one

approach would be to first map spheroid to the sphere by using authalic latitude and then perform areal calculations on sphere. We dismiss this approach since, among other things, our aim is to determine how increase of accuracy of calculating authalic latitude is affecting cell areas. Therefore, we need to rely on a method to determine cell areas that does not depend on authalic latitude.

If DGGS cells are quadrangles bounded by two meridians and two parallels, calculating their area directly on the surface of the spheroid is straightforward (Kelly and Šavrič, 2021; Lapaine and Lapaine, 1991). If cells are bounded by geodesic lines, then equations for calculating their area directly from geodetic coordinates of their vertices are also provided by various authors (Karney, 2023; Lukatela, 2000; Nowak and Nowak Da Costa, 2022). Although for some DGGSs, at least some of the cells are bounded by meridians and parallels or geodesic lines, this is not the case for the QPix DGGS. In these situations, Chrisman (2017) suggests projecting polygon (i.e., DGGS cell) from spheroid to a plane by an equal-area projection and then calculating area from a planar coordinates. Based on a definition of a DGGS cell boundary and based on a choice of a projection, projections of a DGGS cell boundaries can be straight lines. In general case, however, these projections will be curves map projection plane that diverge from a straight line. Therefore, besides vertices of a DGGS cells (three for triangles, four for quadrangles, etc.), Chrisman (2017) emphasizes the need to densify edges of a DGGS cell with sufficient number of intermediate points before projecting them to the plane. Of course, densification reduces calculation error, but certain level of error stemming from approximation of a curved line with straight segments remains. Gillissen (1993) suggests a procedure for eliminating this approximation error.

Since QPix DGGS cell edges are not geodesic lines or meridians or parallels, we decided to calculate cell areas by first projecting cells to a plane by equal-area map projection and then performing calculations in plane. Cells in the rHEALPix DGGS and in QPix DGGS are highly regular and their extent is approximately the same in all directions from the corresponding center point, therefore we decided to use LAEA projection, as suggested by Kelly and Šavrič (2021). Each cell was separately projected with the LAEA projection with cell centroid as projection center. The LAEA was also used by Kmoch et al. (2022) when calculating areas of DGGS cells as a part of determining area and shape distortions for various available DGGS implementations. Results obtained by Berk and Ferlan (2018) when comparing effect of using different projections on cadastral

parcel areas give preference to Albers equal-area conic (AEAC) projection over LAEA. We performed preliminary testing with LAEA and AEAC and we obtained better results with the LAEA projection, presumably because Berk and Ferlan (2018) analysis was limited to Slovenia, while our analysis is global.

The number of densification points along edges of DGGS cells was determined empirically. Kmoch et al. (2022) did not explicitly state how they performed densification, but from the supplementary material they have published (Kmoch and Uuemaa, 2024), it seems each edge was densified with 10 points. White et al. (1998) found that for their analysis,  $2^{16}$ -resolution points along the edge was sufficient to reach needed accuracy (resolution is hierarchical level). We performed analysis for the resolution-1 rHEALPix DGGS cells on WGS 84 ellipsoid for a 1-to-4 and 1-to-9 refinements. We calculated differences between theoretical area and calculated areas for each cell for different number of densification points along each edge, starting from  $2^{16}$  and doubling it until  $2^{20}$  for 1-to-4

**Table 3.4** Absolute differences between theoretical and calculated areas for different number of densification points along each edge of the rHEALPix DGGS resolution-1 cells in 1-to-4 refinement (WGS 84 ellipsoid).

Number of densification points	Min. difference (m <sup>2</sup> )	Max. difference (m <sup>2</sup> )	Max. difference / theoretical area* (ppt)
$2^{16}$	183.90	616.00	28.98
$2^{17}$	44.90	156.50	7.36
$2^{18}$	9.90	41.50	1.95
$2^{19}$	1.30	11.00	0.52
$2^{20}$	0.00	4.90	0.23

\* Theoretical area = 21 252 734 238 503.60 m<sup>2</sup>

**Table 3.5** Absolute differences between theoretical and calculated areas for different number of densification points along each edge of the rHEALPix DGGS resolution-1 cells in 1-to-9 refinement (WGS 84 ellipsoid).

Number of densification points	Min. difference (m <sup>2</sup> )	Max. difference (m <sup>2</sup> )	Max. difference / theoretical area* (ppt)
$3^9$	175.68	10 027.13	1 061.56
$3^{10}$	19.37	1 114.18	117.96
$3^{11}$	1.65	123.76	13.10
$3^{12}$	0.68	13.49	1.43
$3^{13}$	0.09	2.44	0.26

\* Theoretical area = 9 445 659 661 557.20 m<sup>2</sup>

refinement and from  $3^9$  and tripling it until  $3^{13}$  for 1-to-9 refinement. Maximum and minimum observed differences between theoretical and calculated areas are shown in Table 3.5 and Table 3.8.

In tables above, maximum differences are observed for cap cells in case of 1-to-9 refinement and for dart cells in case of 1-to-4 refinement. To a certain level, this is an expected finding since approximation of a curved cell edges with straight segments would yield higher areal error for cells with higher edge curvature. Northmost/southmost dart cells and cap cells have the highest edge curvature on the spheroid and this characteristic is obviously being kept after projection to the plane with LAEA. Since differences between calculated and theoretical area for cap and northmost/southmost dart cells are outliers and since there are the least cap and dart cells from four possible cell shapes, they cannot be considered as a good representation for all cells. Therefore, we decided to concentrate on quad cells in the equatorial region because there are twice as many cells with quad shape than all other cells' shapes combined.

Based on a logic that in order to achieve a comparable cell area calculation accuracy across DGGS resolutions, distance between densification points should be stable across resolutions, and on results from Table 3.5 and Table 3.8, and also on work by White et al. (1998), we propose an equation for calculating number of densification points for 1-to-4 refinement:

$$\max(2, 2^{17-\text{resolution}}); \quad (7)$$

and for 1-to-9 refinement:

$$\max(3, 3^{11-\text{resolution}}). \quad (8)$$

These equations make sense only if curvature and length of cell edges in LAEA projection are similar. Since cell edges of cells in equatorial region are meridians and parallels, near the equator they have similar curvature and length. By going polewards, curvature of north and south cell edges is becoming larger, and their length is getting smaller. According to Equations 7 and 8, the number of densification points is kept the same for a particular resolution. This means that in this case increased curvature will, to a certain level, be compensated with smaller distances between densification points. Exponents in Equations 7 and 8 for resolution 0 were chosen based on results from Table 3.5 and Table 3.8 so that cell areas are calculated with at least 10 significant figures. We confirmed empirically that for quad cells near the equator and near the border between equatorial and polar regions, if

densification points are determined from Equations 7 and 8, cell areas will be calculated with at least 10 significant figures. The same empirical analysis was performed for the QPix DGGS, and it was concluded that number of densification points along cell edges determined by Equations 7 and 8 would for most cells achieve comparable cell area calculation accuracy than is achieved by the rHEALPix DGGS.

### 3.4.2 Calculating perimeter of DGGS cells

For calculating a perimeter of a DGGS cell on a spheroid, based on review by Chrisman (2017), two approaches are feasible. The first approach again includes projecting the cell edges to the plane. Edges again need to be densified before projecting with a sufficient number of intermediate points. Projections of cell edges, which are in general curves, will thus be approximated with straight segments. Length of a cell edge is then calculated by adjusting length of each segment by a local map projection scale factor and then summing all adjusted lengths. To make calculations of scale factor as simple as possible, Snyder (1987, p. 21) states that a choice should be made between cylindrical, conic, or azimuthal projections with normal aspect (as defined by Lapaine and Frančula (2019)).

The second approach deals with the problem of perimeter calculation directly on the surface of a spheroid. If edges of DGGS cells are meridians, parallels, or geodesics, then equations for calculating their length are well known. Otherwise, each cell edge should be densified with intermediate points and then the length of cell edge can be approximated with a sum of lengths between each neighboring densification points. Since geodesic line is the line with the shortest length between two points on a spheroid, it makes the most sense to use it as an approximation of a cell edge between two densification points. We decided to go with this approach. Based on discussion and findings in Section 3.4, we performed a similar analysis and empirically came to the equations for required number of densification points along each edge of the rHEALPix DGGS quad cells that will ensure that perimeter is calculated with at least 10 significant digits across all resolutions. For 1-to-4 refinement, the number of densification points should be:

$$\max(2, 2^{16-\text{resolution}}); \quad (9)$$

and for 1-to-9 refinement:

$$\max(3, 3^{10-\text{resolution}}). \quad (10)$$



For calculating length of a geodesic, we used the Python implementation of a C++ library *GeographicLib* (Karney et al., 2022) which implements findings from Karney (2013). Again, the same empirical analysis was performed for the QPix DGGs, and it was concluded that number of densification points along cell edges determined by Equations 9 and 10 would for most cells achieve comparable cell perimeter calculation accuracy than is achieved for the rHEALPix DGGs.

### 3.4.3 Calculating compactness of DGGs cells

The rHEALPix and QPix DGGs cells are squares in the  $(n, s)$ -rHEALPix or  $(n, s)$ -QPix map projection plane and when they are projected to the spheroid by inverse of these projections, their area is, at least analytically, preserved, but their shape is being distorted. As a measure for shape distortion, often and angular distortion is used, however, for determining DGGs cell shape distortion, compactness is more common (J. A. Kimerling et al., 1999; Kmoch, Vasilyev, et al., 2022; White et al., 1998). Compact shapes are considered as those that have maximum accessibility to all parts of the shape and thus circle is considered as the most compact planar shape (W. Li et al., 2013). The Iso-Perimetric Quotient (IPQ) is the most often used measurement for compactness of a planar shapes and is calculated as (defined by Osserman (1978) as cited by Li et al. (2013)):

$$C_{IPQ} = \frac{4\pi A}{P^2}, \quad (11)$$

where  $A$  is the area of shape, and  $P$  is its perimeter. For circle,  $C_{IPQ}$  is equal to 1, and as shape is more diverging for a circular shape,  $C_{IPQ}$  is decreasing towards 0.  $C_{IPQ}$  was used by Kmoch et al. (2022) to compare various DGGs implementations from the cell shape stability point of view. Since  $C_{IPQ}$  is used for planar shapes, they applied Equation 11 on a projection of each DGGs cell to the plane by the LAEA map projection. We decided not to follow this direction since LAEA deforms shapes as well as length of the cell perimeter which can thus have a negative impact on compactness calculation accuracy. Instead, we decided to calculate compactness that is more suitable for shapes on the spheroid. Basaraner and Cetinkaya (2019) define spherical compactness as:

$$C_{sp} = \left(4\pi - \frac{A}{r^2}\right) \frac{A}{P^2}, \quad (12)$$

where  $A$  is the area of shape,  $P$  is its perimeter and  $r$  is the radius of the sphere. By using Equation 12, a spherical cap will have compactness equal to 1, meaning that it is considered as the most compact

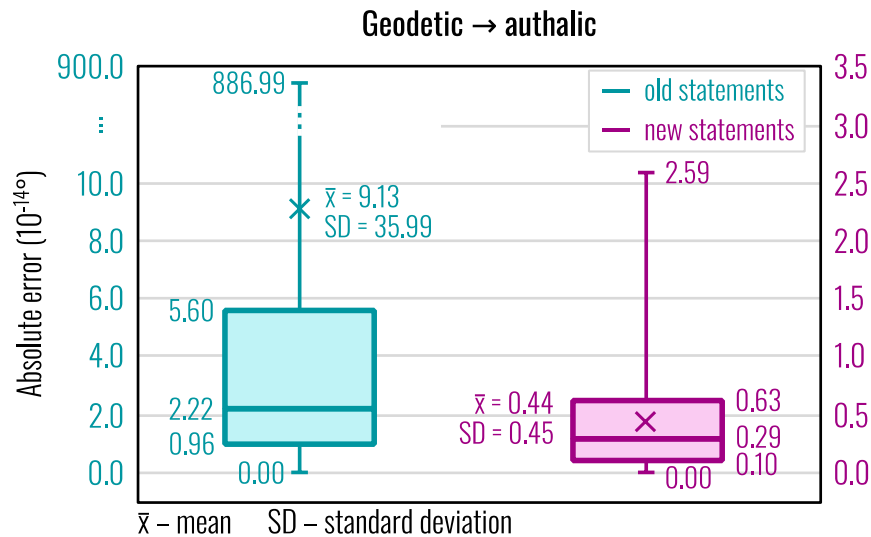
shape on the sphere. Adapting this equation to the arbitrary spheroid would be complex. To a certain degree, the simplest solution would be to define a polar cap as the most compact shape. However, if the most compact shape is the one that has a property of maximum accessibility, because curvature of a spheroid is not constant, the most compact shape varies with the location on a spheroid. One way to approach this problem is to determine the centroid of the shape for which compactness is being calculated and then define a buffer zone around that centroid by geodesic lines of constant length as the most compact shape for that location. Since we are performing calculations on a WGS 84 ellipsoid of revolution, to avoid this complexity we decided to use area and perimeter of a DGGS cell that were calculated on the spheroid and radius of the authalic sphere of the ellipsoid to calculate compactness using Equation 12. We decided to use the radius of the authalic sphere because coefficient before  $A/P^2$  in Equation 12 is determined in a way to ensure that the area of a shape being considered on a sphere is the same as the area of the spherical cap (i.e., the most compact shape). Although this is not a perfect solution, it is still better than using Equation 11 for calculating compactness of a shape on a spheroid.

### 3.5 Results

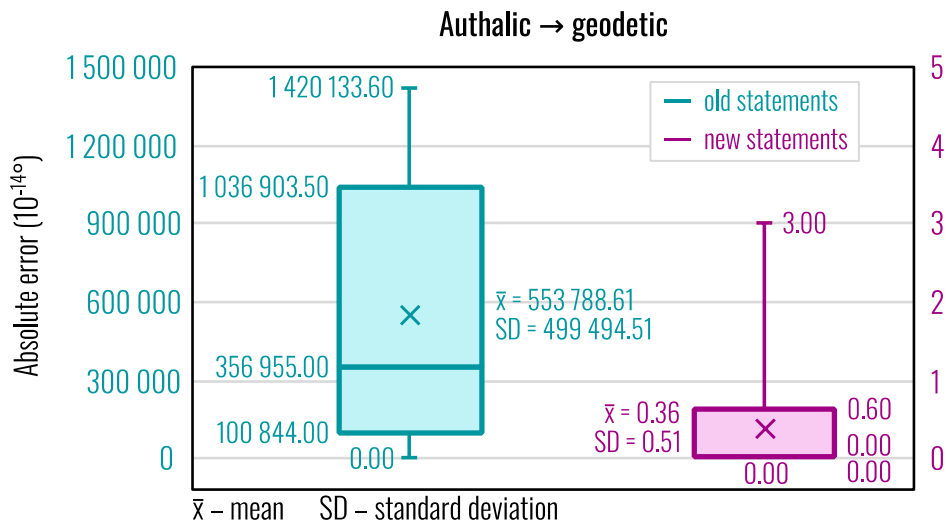
In this section, we first provide results of the accuracy and processing time assessment that are related to modifying Python statements used for converting geodetic to authalic latitude and vice versa (explained in Section 3.3.1). Then, we present results of compactness calculations for the rHEALPix and QPix DGGSs (explained in Section 3.4).

#### 3.5.1 Authalic latitude calculation for the rHEALPix DGGS

All calculation errors, i.e., differences between baseline latitudes and latitudes that were calculated using old and new code, are in this section reported as their absolute values. Box plots on Figures 3.17 and 3.18 show distribution of these calculation errors using old and new Python statements for converting geodetic to authalic and authalic to geodetic latitude, respectively. We have to note that on these box plots we did not show outliers but have extended whiskers to minimum and maximum observed errors. Also, because we wanted to clearly show distribution of errors produced by the old and new statements and because magnitudes of these errors significantly differ, we used two axes on each graph and colored them the same way as the corresponding box plots.

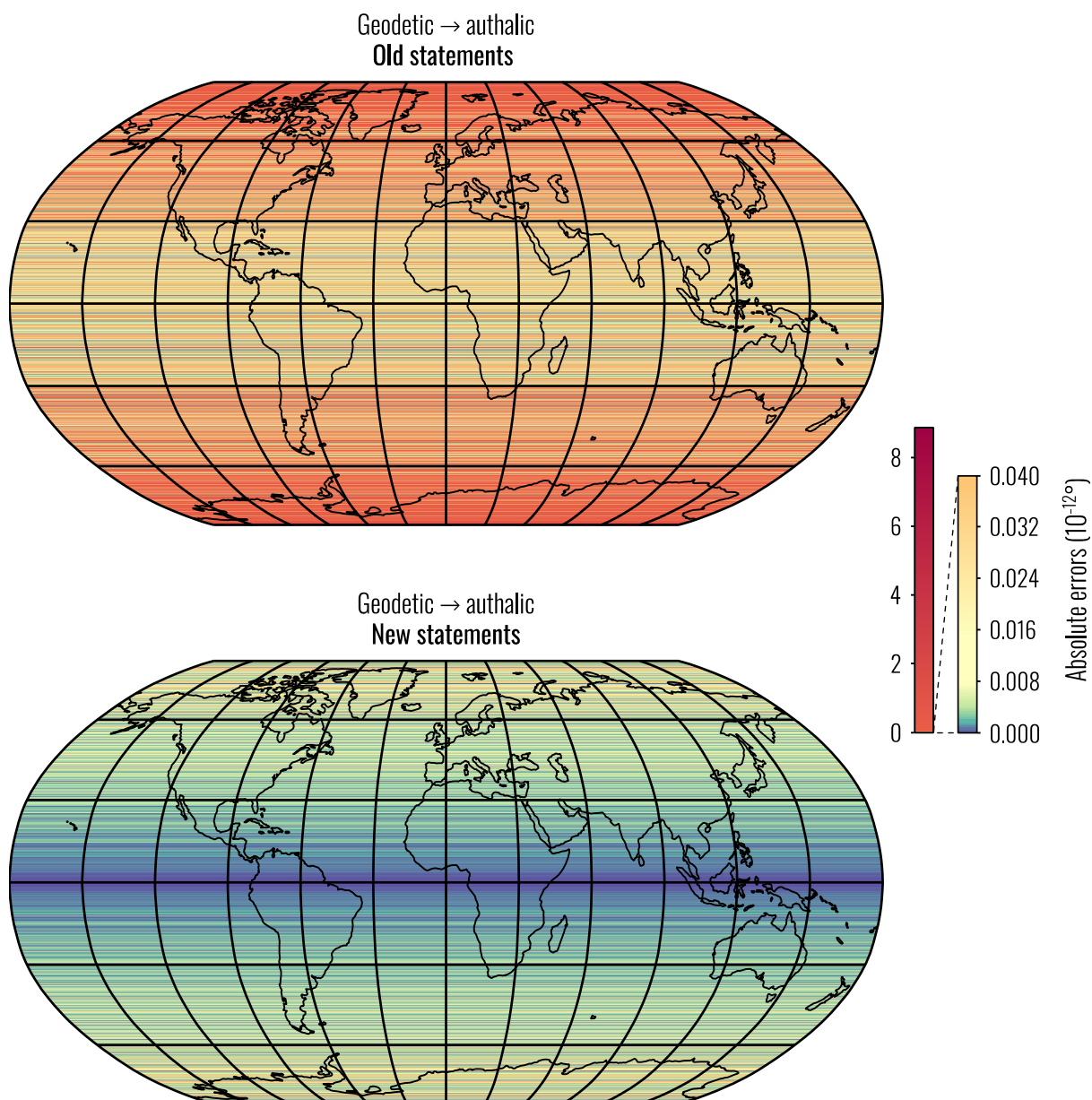


**Figure 3.17** Box plots of the observed calculation errors related to calculating authalic latitude from the geodetic.

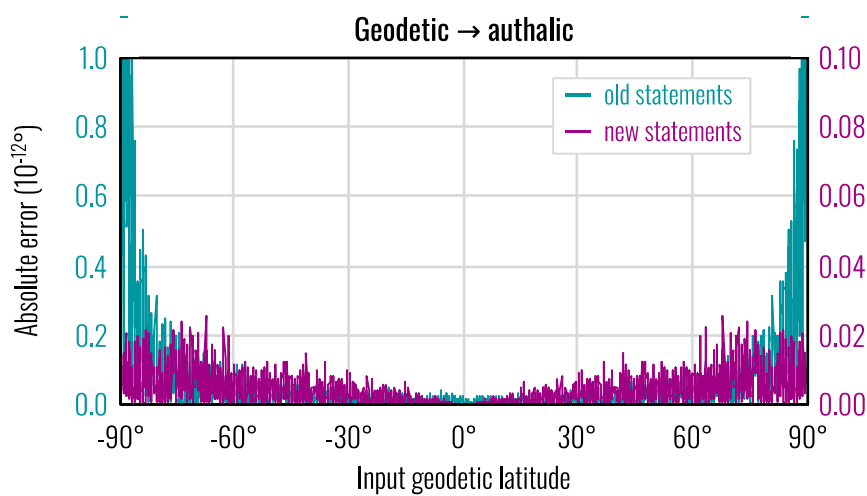


**Figure 3.18** Box plots of the observed calculation errors related to calculating geodetic latitude from the authalic.

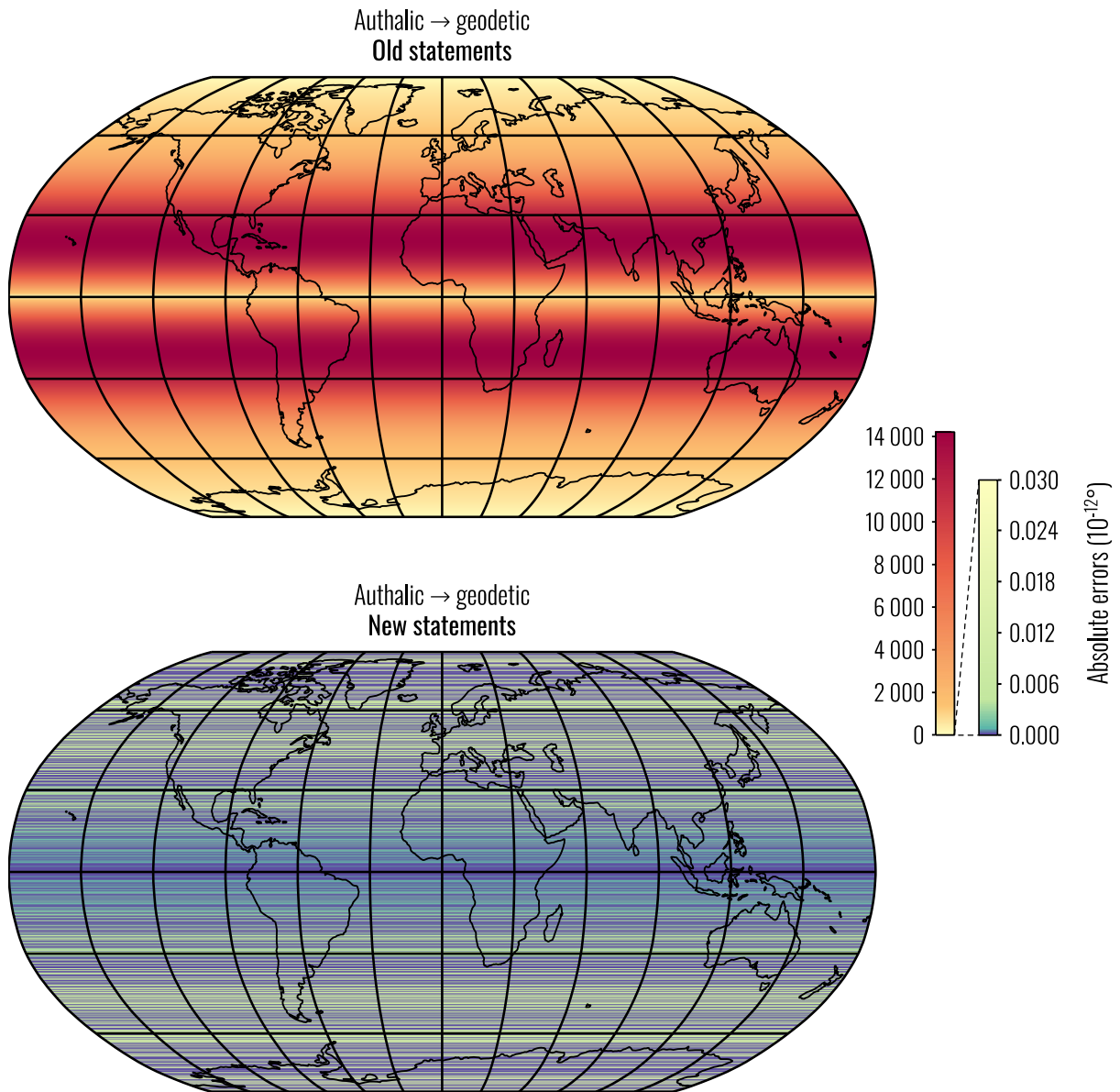
Figure 3.19 shows maps with spatial distribution of calculation errors generated by old and new statements in geodetic to authalic latitude calculation direction. Figure 3.20 show identical distribution across latitudes in the form of a graph. For the authalic to geodetic latitude conversion, these distributions are shown on Figures 3.21 and 3.22. On graphs on Figures 3.20 and 3.22, again two axes with different scales were used. To clearly show variability in calculation errors, colors in the color bar for maps on Figures 3.19 and 3.21 were arranged so that approximately the same number of parallels is shown with the same or similar color. These are the parallels for which calculation errors were determined (all parallels with latitudes from  $-90^\circ$  to  $90^\circ$  and with a difference of  $0.1^\circ$  between them).



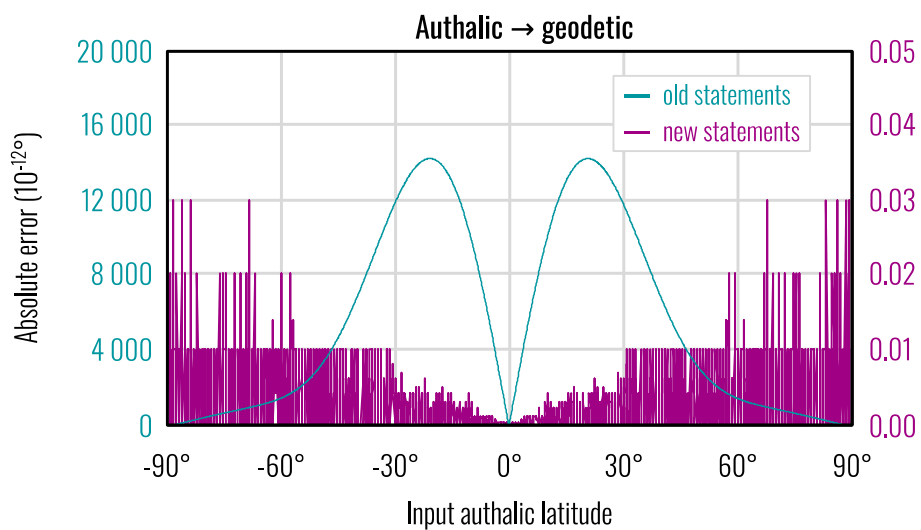
**Figure 3.19.** Spatial distribution of the observed calculation errors related to calculating authalic latitude from the geodetic.



**Figure 3.20.** Distribution of observed calculation errors across latitudes when calculating authalic latitude from the geodetic.



**Figure 3.21.** Spatial distribution of the observed calculation errors related to calculating geodetic latitude from the authalic.



**Figure 3.22.** Distribution of observed calculation errors across latitudes when calculating geodetic latitude from the authalic.

Regarding the influence of the calculation errors when performing conversions between authalic and geodetic latitudes in both ways, Tables 3.6 and 3.7 list the maximum recorded differences between theoretical and calculated rHEALPix DGGS cell areas on the WGS 84 ellipsoid, across the resolutions in the 1-to-4 and 1-to-9 refinements, respectively. Maximum differences for each resolution were determined from all differences that were calculated on cells starting from those listed in Table 3.3 and then continuing on their parents on each subsequent lower resolution.

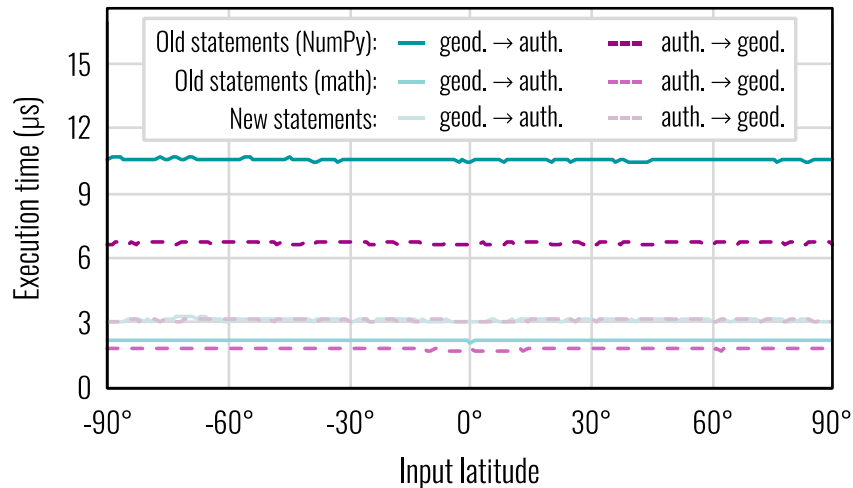
**Table 3.6** Maximum recorded differences between theoretical and calculated area of the ellipsoidal rHEALPix DGGS cells across resolutions in the 1-to-4 refinement.

Resolution (i.e., refinement level)	Theoretical area (m <sup>2</sup> )	Old Python statements		New Python statements	
		Max. cell area difference from theoretical area (m <sup>2</sup> )	Share of area difference in total area (ppb)	Max. cell area difference from theoretical area (m <sup>2</sup> )	Share of area difference in total area (ppb)
0	85 010 936 954 014.70	9 871.80	0.12	17.94	
1	21 252 734 238 503.60	2 465.17	0.12	4.59	
2	5 313 183 559 625.92	3 698.64	0.70	3.36	
3	1 328 295 889 906.48	1 325.45	1.00	1.20	
4	332 073 972 476.62	360.99	1.09	0.24	
5	83 018 493 119.16	92.17	1.11	0.05	
6	20 754 623 279.79	23.17	1.12	0.01	< 0.01
7	5 188 655 819.95	5.80	1.12		
8	1 297 163 954.99	1.45	1.12		
9	324 290 988.75	0.36	1.12		
10	81 072 747.19	0.09	1.12		
11	20 268 186.80	0.02	1.12		
12	5 067 046.70	0.01	1.12		
13	1 266 761.67		1.12		0.01
14	316 690.42		1.12		0.01
15	79 172.60		1.12		0.00
16	19 793.15		1.09		0.05
17	4 948.29		1.18		0.05
18	1 237.07		1.09		0.07
19	309.27		1.03		0.48
20	77.32		1.05		0.26
21	19.33		1.49		0.13
22	4.83	< 0.01	2.85		0.89
23	1.21		5.93		3.89
24	0.30		22.40		16.03
25	0.08		28.05		10.13
26	0.02		83.40		12.33
27	0.0047		76.24		33.19
28	0.0012		153.11		88.01
29	0.0003		295.17		94.10
30	0.0001		597.80		208.88

**Table 3.7** Maximum recorded differences between theoretical and calculated area of the ellipsoidal rHEALPix DGGS cells across resolutions in the 1-to-9 refinement.

Resolution (i.e., refinement level)	Theoretical area (m <sup>2</sup> )	Old Python statements		New Python statements	
		Max. cell area difference from theoretical area (m <sup>2</sup> )	Share of area difference in total area (ppb)	Max. cell area difference from theoretical area (m <sup>2</sup> )	Share of area difference in total area (ppb)
0	85 010 936 954 014.70	9 871.80	0.12	17.94	
1	9 445 659 661 557.20	8 609.40	0.91	5.09	
2	1 049 517 740 173.02	1 147.71	1.09	1.32	
3	116 613 082 241.45	130.07	1.12	0.12	
4	12 957 009 137.94	14.48	1.12	0.01	< 0.01
5	1 439 667 681.99	1.61	1.12		
6	159 963 075.78	0.18	1.12		
7	17 773 675.09	0.02	1.12		
8	1 974 852.79		1.12		0.02
9	219 428.09		1.12		0.01
10	24 380.90		1.12		0.03
11	2 708.99		1.13		0.36
12	301.00		1.10	< 0.01	0.29
13	33.44	< 0.01	2.74		2.45
14	3.72		9.29		0.19
15	0.41		71.42		21.31
16	0.05		52.74		52.78
17	0.01		588.80		527.72
18	0.0006		1 195.57		186.06
19	0.0001		168.94		1 041.95

Finally, we provide results that are related to processing time assessment. Processing time across latitudes, for both conversion directions and for old (based on *NumPy* and based on *math*) and new Python statements are shown on Figure 3.23. As already explained, processing time was measured for each latitude from -90° to 90° with steps of 1°, in order to determine whether input latitude value has any influence on processing time.

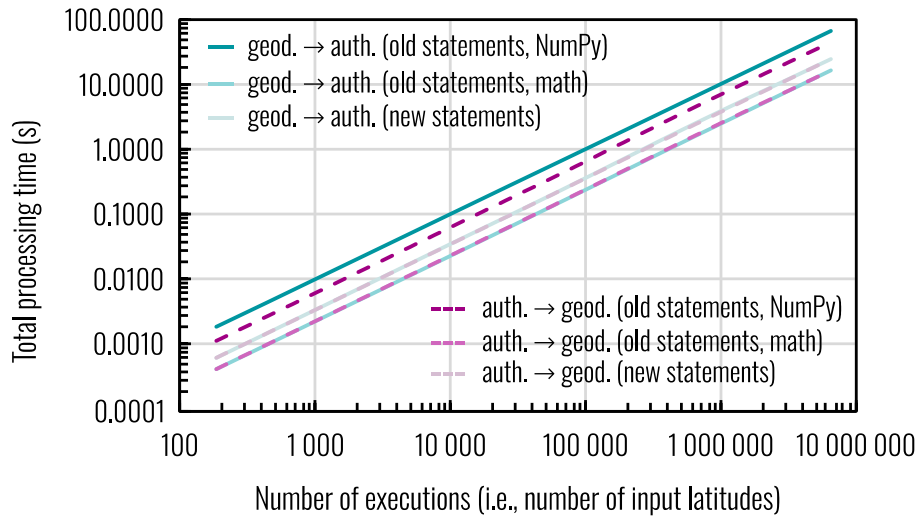


**Figure 3.23.** Processing times across latitudes.

Total processing times for different number of executions (i.e., number of input latitudes) of all statement variants are listed in Table 3.8. From this data, we generated graph on Figure 3.24 which shows the relationship between different statement variants in terms of processing time and thus enables their direct comparison.

**Table 3.8.** Total processing times for old statements (both, *NumPy*- and *math*-based) and new statements for different number of executions. The number of executions is, in fact, the number of input latitudes on which statements were executed. These latitudes include the whole range from  $-90^\circ$  to  $90^\circ$  with varying difference between consecutive latitudes.

Number of input latitudes	Difference between consecutive latitudes	Cumulative execution time (s)					
		Geodetic $\rightarrow$ authalic latitude			Authalic $\rightarrow$ geodetic latitude		
		Old statements		New statements	Old statements		New statements
		NumPy	math		NumPy	math	
181	1°	$2 \times 10^{-3}$	$4 \times 10^{-4}$	$6 \times 10^{-4}$	$1 \times 10^{-3}$	$4 \times 10^{-4}$	$6 \times 10^{-4}$
1 081	10'	0.01	$3 \times 10^{-3}$	$4 \times 10^{-3}$	0.01	$3 \times 10^{-3}$	4
10 801	1'	0.11	0.03	0.04	0.07	0.03	0.04
64 801	10"	0.66	0.16	0.24	0.41	0.15	0.24
648 001	1"	6.90	1.61	2.43	4.19	1.64	2.47
6 480 001	0.1"	67.99	15.92	24.09	41.92	15.77	24.33
64 800 001	0.01"	685.90	163.15	245.48	429.99	159.05	246.38



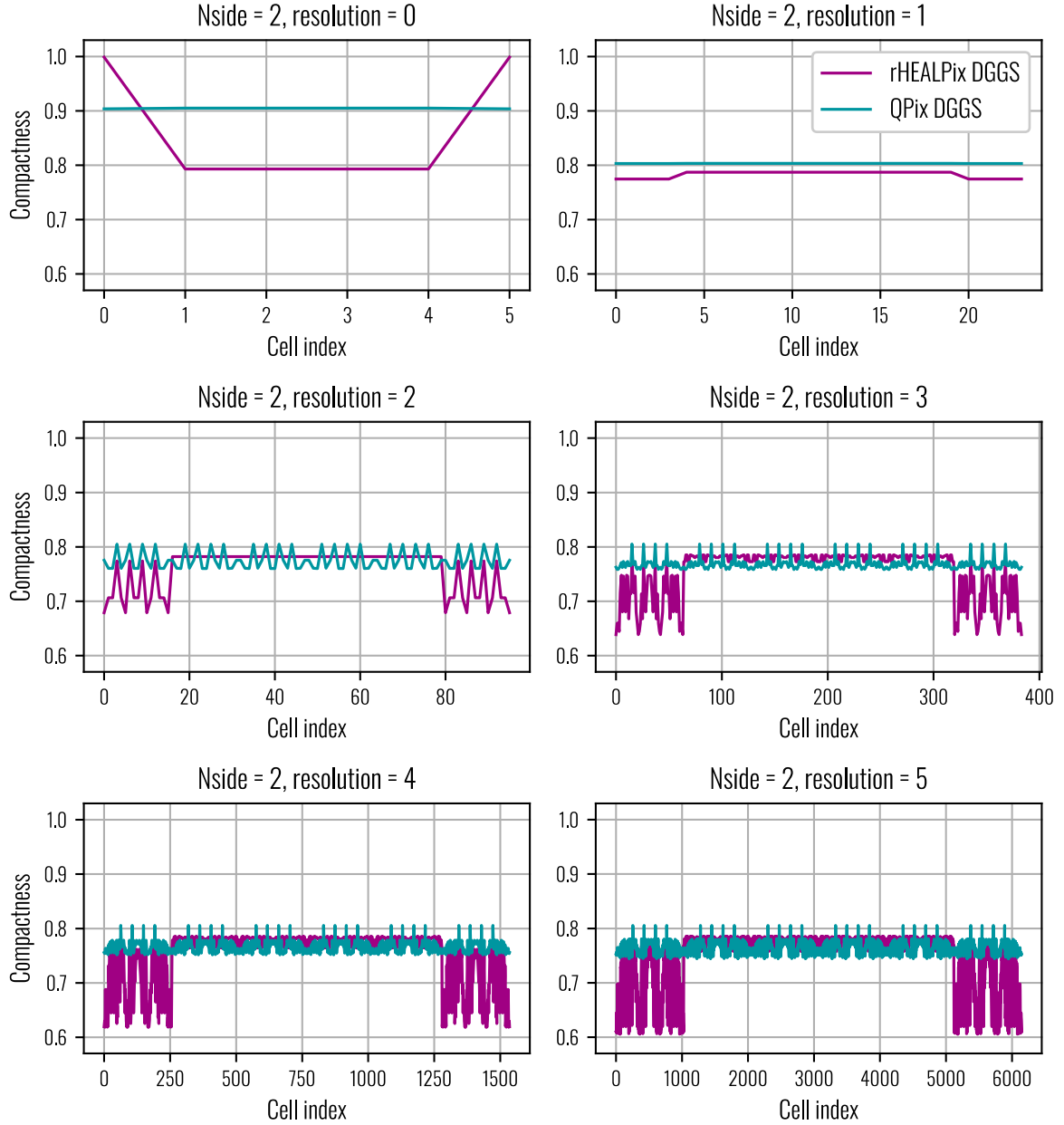
**Figure 3.24.** Total processing times for different number of executions (i.e., number of input latitudes).

### 3.5.2 Compactness of the rHEALPix and QPix DGGs cells

This section presents results of calculating compactness of the rHEALPix and QPix DGGs cells. Compactness was calculated for all cells in the rHEALPix and QPix DGGs for resolution up to eight (first nine refinement levels) in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and up to 5 (first six refinement levels)

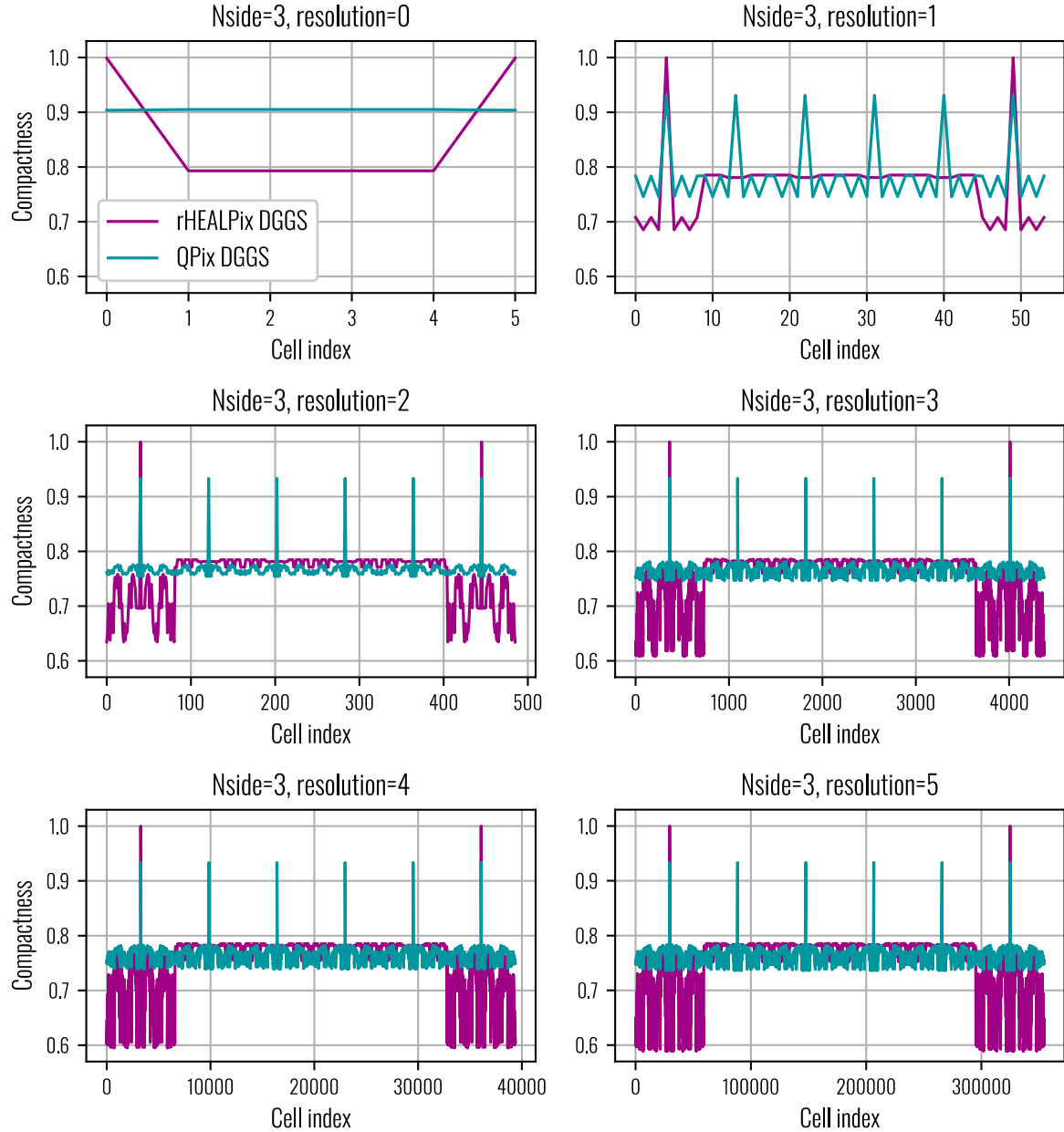


in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ). Figure 3.25 shows calculated compactness values for all cells in the rHEALPix and QPix DGGs in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) up to resolution five, while Figure 3.26 also shows compactness, but for 1-to-9 refinement ( $N_{\text{side}} = 3$ ).



**Figure 3.25** Compactness of rHEALPix and QPix DGGs cells for resolutions from 0 to 5 for a 1-to-4 refinement ( $N_{\text{side}} = 2$ ).

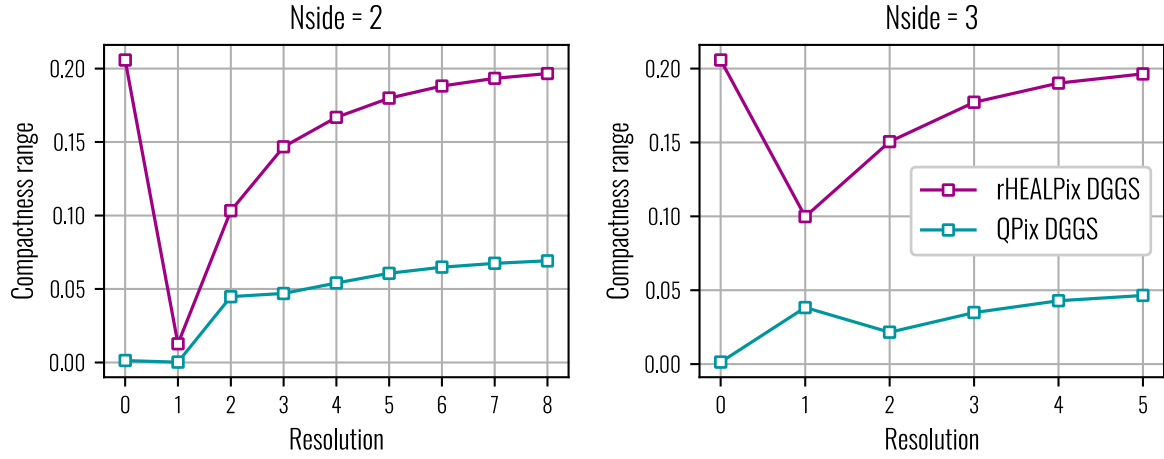
On Figure 3.25 and Figure 3.26, values on the abscissa are cell indices. Cell indices are determined by starting with index 0 from the utmost top left cell in the (0, 0)-rHEALPix or (0, 0)-QPix projection plane (Figure 3.5 and Figure 3.14) and then increasing by 1 for each by following Morton space-filling curve (left-to-right and top-to-bottom, Figure 2.4).



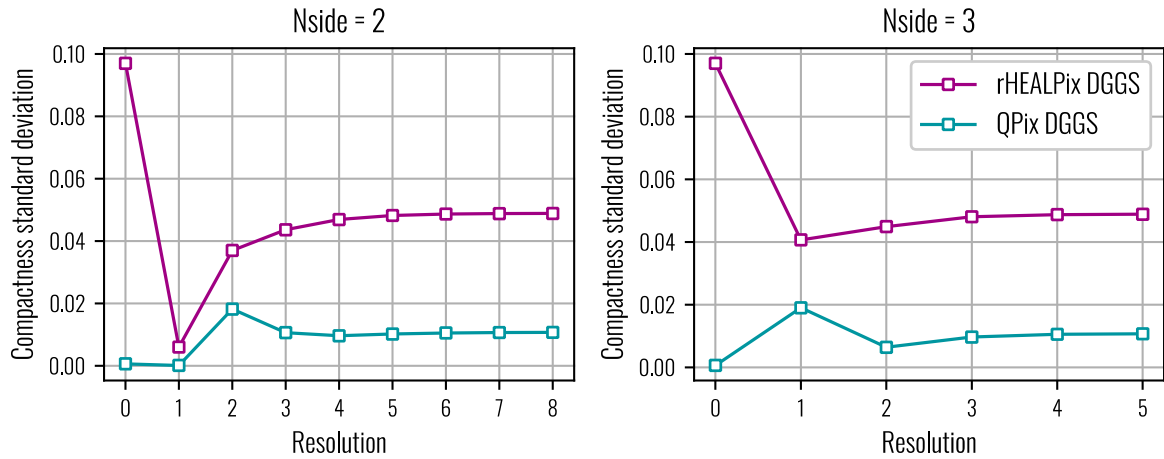
**Figure 3.26** Compactness of rHEALPix and QPix DGGs cells for resolutions from 0 to 5 for a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).

Figures 3.27, 3.28, and 3.29 show compactness ranges, standard deviations, and means, respectively, for the rHEALPix and QPix DGGs cells up to a resolution eight for 1-to-4 refinement and up to resolution five for 1-to-9 refinement. For the 1-to-9 refinement ( $N_{\text{side}} = 3$ ), as can be observed on Figure 3.26, for all resolutions except the first one (resolution 0), there are six compactness values that are clearly an outliers for the QPix DGGs, and two for the rHEALPix DGGs. We decided to remove these outliers prior to calculating means, ranges, and standard deviations for the 1-to-9 refinement in order to make them better represent compactness stability for

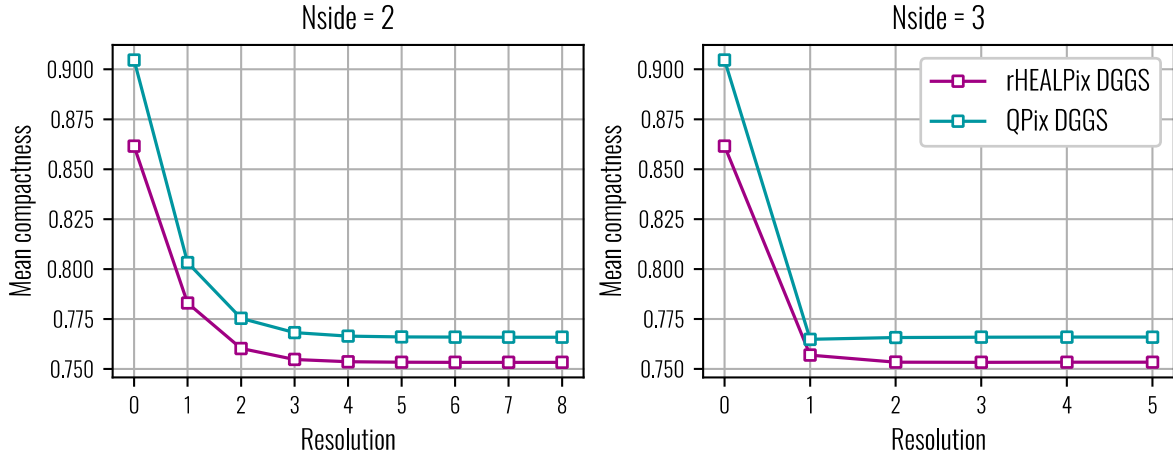
the whole population of cells on a specific refinement level. Since range and standard deviations are measures of variability in a dataset, they can, in this case, be considered as measures of cell shape stability. The lower the compactness standard deviation and range, the higher the cell shape stability and DGGS regularity.



**Figure 3.27** rHEALPix and QPix DGGS compactness ranges for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).

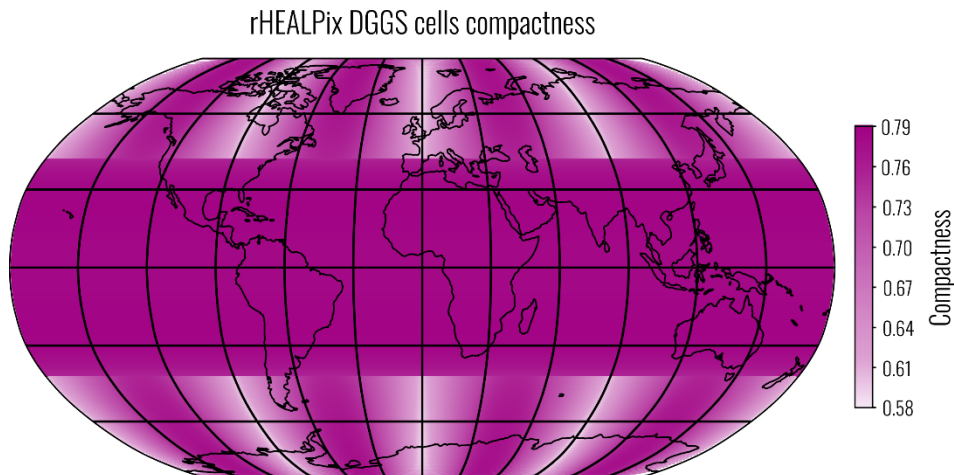


**Figure 3.28** rHEALPix and QPix DGGS compactness standard deviations for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).

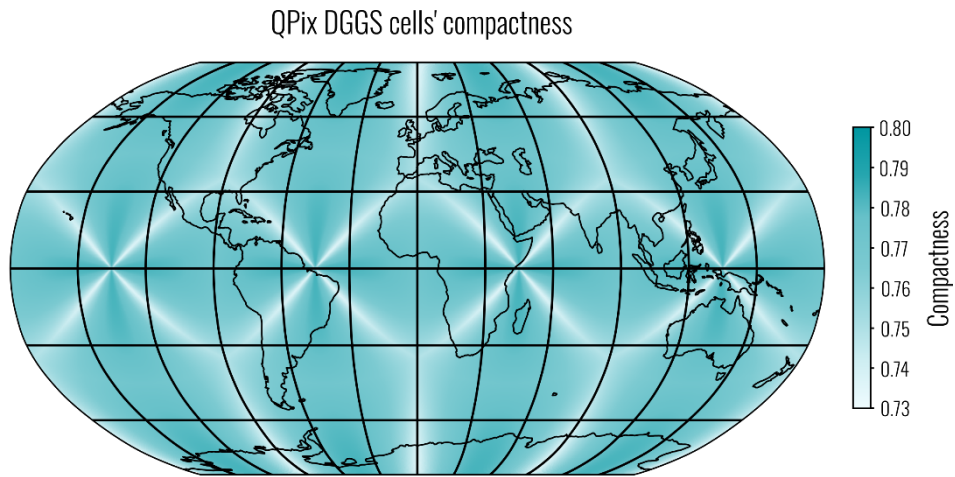


**Figure 3.29** rHEALPix and QPix DGGs compactness means for resolutions 0–8 in a 1-to-4 refinement ( $N_{\text{side}} = 2$ ) and resolution 0–5 in a 1-to-9 refinement ( $N_{\text{side}} = 3$ ).

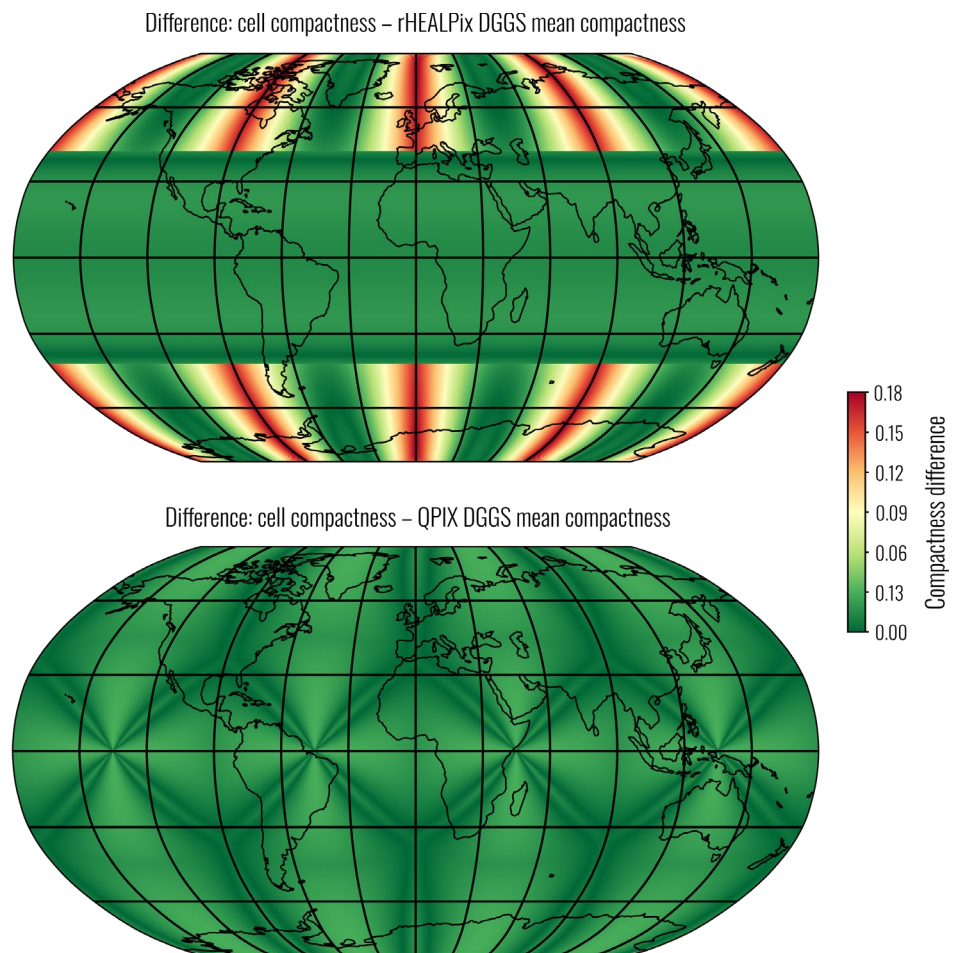
Figures 3.30 and 3.31 show spatial distribution of cell compactness on a global level for the rHEALPix and QPix DGGs respectively. To get a better insight into variations in distribution and magnitude of compactness for the rHEALPix and QPix DGGs, we calculated mean compactness and cell-wise absolute differences between cell compactness and corresponding mean compactness. Mean compactness values as well as cell-wise differences were determined separately for the rHEALPix and for QPix DGGs. These differences are shown on Figure 3.32 with a joint color bar. Color bar was defined so that approximately the same number of cells will be shown using the same color. Lower values are, of course, preferred on this figure from the high compactness stability standpoint.



**Figure 3.30** Compactness of the rHEALPix DGGs cells.



**Figure 3.31** Compactness of the QPix DGGs cells.



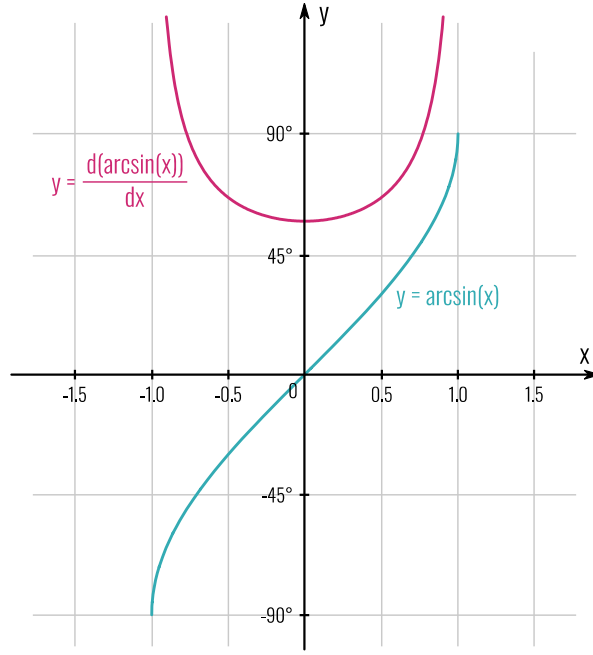
**Figure 3.32** Cell-wise absolute differences between cell compactness and mean compactness for the rHEALPix and QPix DGGs. Mean compactness as well as differences are calculated separately for the rHEALPix and QPix DGGs.

## 4 Discussion

### 4.1 Authalic latitude calculation

When observing Figures 3.17 and 3.18 with box plots of calculation errors produced when converting geodetic to authalic latitude and vice versa, first thing that can be noticed is that overall magnitude of errors are for both conversion directions significantly lower for new statements that are based on Equations 3 and 4, than for old statements based on Equations 1 and 2. These box plots also show that calculation errors are not normally distributed but are in all cases right skewed with outliers that are the most extreme for old statements when converting geodetic to authalic latitude. Outliers are not present only for old statements when converting authalic to geodetic latitude, however, in this case also, distribution of calculation errors is right skewed. This means that mean values of calculation errors and standard deviations, that are sensitive to outliers, are not a good representation of all calculation errors. The interquartile range (IQR) is therefore a better representation. When converting geodetic to authalic latitude, new statements reduce IQR by approximately 8.8 times, while for the authalic to geodetic conversion direction, this reduction is at level of six orders of magnitude for new statements.

If we concentrate on distribution of calculation errors across latitudes for converting geodetic to authalic latitude based on old statements, we can observe that errors increase as input geodetic latitude increases, with extreme values observed near the poles. Since Equation 1, that is used in old statements, is provided in closed form, computation errors consist only of round-off errors. Round-of error is initially introduced when converting input geodetic latitude from decimal to binary representation. It is then further transferred with every following calculation and finally accumulated in the argument of the  $\arcsin$  function. By observing the graph of the  $\arcsin$  function and of its derivative on Figure 4.1, it can be concluded that small changes in argument value near  $\pm 1$  (function value near  $\pm 90^\circ$ ) results in excessive change in function value. Even more, as argument approaches  $\pm 1$ , derivative of the  $\arcsin$  tends to  $+\infty$ . This means that  $\arcsin$  function is extremely numerically unstable for arguments near  $\pm 1$ , which correspond to function values near  $\pm 90^\circ$ , or latitudes near the poles. Therefore, it is not a surprise that observed calculation errors when converting geodetic to authalic latitude by using old statements are the largest near the poles.

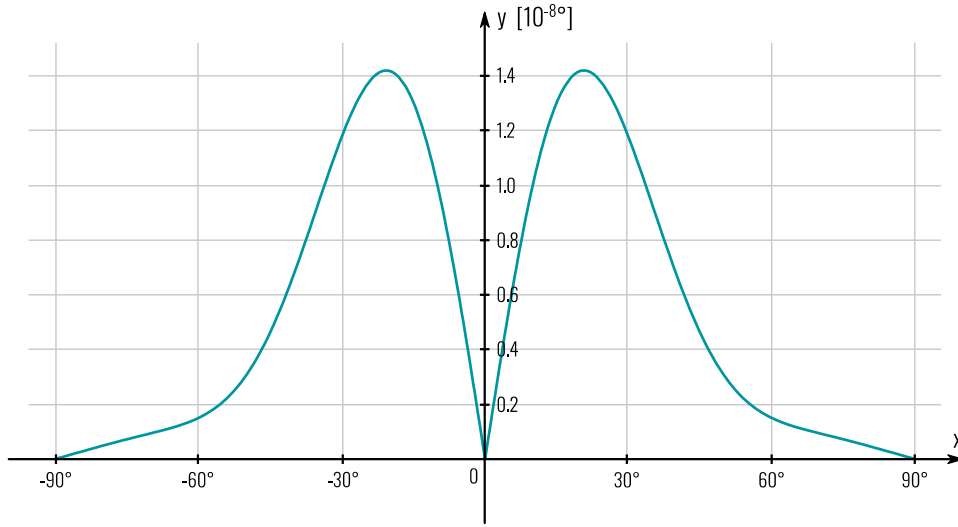


**Figure 4.1** Graph of the *arcsin* function and its derivative.

For the opposite conversion direction (geodetic to authalic), old statements are based on Equation 2. This equation is provided as a power-series expansion in terms of eccentricity  $e$  of ellipsoid of revolution, extending up to  $e^6$ . As a result, in addition to round-off errors, it can be expected that a cumulative calculation error is also affected by truncation errors. As already stated, Karney (2013) provided equation for performing the same calculation but that is based on power-series expansion in terms of third flattening  $n$ , up to  $n^6$ . He also confirmed that this power-series expansion results in truncation errors that are at the ulp (unit in the last place) level. Li et al. (2022) compared expansions based on  $e$  and  $n$  and concluded that expansion in  $e$  has to be extended up to  $e^{10}$  in order to match the accuracy of expansion in  $n$  up to  $n^5$ . This suggests calculation errors in this case can be mostly attributed to truncation errors. To confirm this assumption, we subtracted power-series expansion in Equation 2 from power-series expansion in terms of  $e$  that was provided by Bian et al. (2012, Equation 48) up to  $e^{10}$  and got truncation error expressed by Equation 13.

The graph of the absolute value function of this function (Equation 13) with argument  $\phi_A$  is shown on Figure 4.2. It is visible that this graph is in close resemblance to the absolute calculation errors that were observed for converting authalic to geodetic latitude using the old statements that are

$$\begin{aligned}
\text{truncation error} = & \left( \frac{120\,389}{181\,400} e^8 + \frac{1\,362\,253}{29\,937\,600} e^{10} \right) \sin 2\phi_A \\
& + \left( \frac{102\,287}{1\,814\,400} e^8 + \frac{450\,739}{997\,920} e^{10} \right) \sin 4\phi_A \\
& + \left( \frac{47\,561}{1\,814\,400} e^8 + \frac{434\,501}{14\,968\,800} e^{10} \right) \sin 6\phi_A \\
& + \left( \frac{6\,059}{1\,209\,600} e^8 + \frac{625\,511}{59\,875\,200} e^{10} \right) \sin 8\phi_A \\
& + \left( \frac{48\,017}{29\,937\,600} e^{10} \right) \sin 10\phi_A
\end{aligned} \tag{13}$$



**Figure 4.2** Absolute value function of the function given by (5).

based on Equation 2 and are visible on Figure 3.22. This proves that in this case the overall calculation error is mostly related to the truncation error, while round-off errors get obscured.

Regarding the statements that are based on Equations 3 and 4, they can be considered as truncation-error-free at the double-precision floating point level (as shown by Karney (2024)). The only remaining source of calculation errors are the round-off errors. To minimize their effect, Karney (2024) suggests applying Horner method and Clenshaw summation for evaluating power-series expansions. As already stated, we applied Horner method, but did not use Clenshaw summation because it, or at least our implementation approach, slightly prolonged execution time.

Concerning the impact of calculation errors of conversions between authalic and geodetic latitude on the area of rHEALPix DGGS cells, data in Tables 3.6 and 3.7 suggest that this impact is not significant from applicative point of view. When maximum observed errors in cell areas are put



into relation with the corresponding theoretical cell area, it turns out that error is on a ppb (parts per billion) level of the cell area. Although these errors can be considered as insignificant for most (if not all) applications, from data in Tables 3.6 and 3.7 it is visible that new statements that are based on Equations 3 and 4 have a result in a lower level of errors in cell areas. This means that new statements brought rHEALPix DGGs closer to being numerically equal area DGGs. This also means that  $(n, s)$ -rHEALPix map projection is also brought closer to being numerically area-preserving, which increases its reliability.

By observing processing times on Figures 3.23 and 3.24, a couple of conclusions can be emphasized. Regardless of the value of the input latitude, processing time is stable for all six statement variants. As expected, when in old statements *NumPy* functions are replaced with *math* functions, significant drop in processing time is observed. In fact, these statements exhibit the shortest processing time. The longest processing time is observed for calculating authalic latitude from geodetic by using old statements. This suggests that statements that are based on Equation 1 are computationally the most demanding. The processing time of new statements is between *NumPy*- and *math*-based old statements. Although *NumPy*-based old statements exhibit the shortest processing time, since we deem calculation accuracy as more important, the final implementation of statements for bidirectional conversions between authalic and geodetic latitude are chosen those that we have been referring to as “new statements”.

## 4.2 Comparison of rHEALPix and QPix DGGs geometrical properties

Different DGGs can be compared against each other with a focus on a particular characteristic that is important for a specific DGGs use-case. One of these characteristics is geometrical and topological properties of grids and their cells. As recorded by Kimerling et al. (1999), prof. Michale Goodchild formulated initial version of criteria for evaluating geometrical and topological characteristics of a global grid (i.e., ‘Goodchild criteria’; also listed in Table 4.1): (1) cells cover the entire earth model (sphere or ellipsoid) without overlapping, (2) all cells on a specific DGGs resolution (i.e., hierarchical level) have the same area, (3) all cells’ boundaries have the same number of vertices and edges (i.e., they are topologically equivalent), (4) all cells have the same shape, (5) cells are compact, (6) in a map projection plane, cell edges are straight lines, (7) when the center points of two neighboring cells are

connected with arc, its midpoint matches the midpoint of their joint edge, (8) grids of varying resolutions are organized hierarchically, (9) each cell has only one reference point, (10) reference points are centered as much as possible within the cells, (11) distances between reference points of adjacent cells are equal, (12) grid of reference points and cells is regular, which enables their simple addressing, (13) grid can be easily put into relationship with grid formed by meridians and parallels, and (14) spatial resolution of grids within the grid system can be defined arbitrarily.

Kimerling et al. (1999) also state that it is not possible to fulfill all of these criteria at the same time and that some criteria might be important, while some completely irrelevant for a specific application. Also, some of these criteria can be either fulfilled or not (i.e., qualitative criteria, according to Wang et al. (2021)), while for some it is possible to determine a level of fulfillment (i.e., quantitative criteria, according to Wang et al. (2021)). Wang et al. (2021) categorize criteria 1, 3, 8, 9, 10, 13, and 14 as qualitative and the rest as quantitative. Some authors (Luo et al., 2023; Wang et al., 2021) argue that some of the Goodchild criteria are interdependent and redundant and can lead to unreliable DGGs comparison results. Wang et al. (2021) performed analysis of suitability of Goodchild criteria for evaluating geometrical and topological aspects of the QTM DGGs (Section 2.1) and concluded that fourteen Goodchild criteria can be reduced to only six, four for qualitative evaluation and two for quantitative. Luo et al. (2023) concentrated on the seven quantitative Goodchild criteria for evaluating five DGGs with quadrangular (i.e., diamond) cells, some of which are equal-area and some are not. For the analyzed equal-area DGGs, they concluded that seven criteria can be reduced to only two, and for non-equal-area DGGs, to three. Although it is clear that some of the Goodchild criteria are redundant, since these reductions in number of criteria cannot be directly applied to all discrete global grids (as suggested by Luo et al. (2023)) and since we only want to make comparison between rHEALPix and QPix DGGs, we decided to stick with the initial Goodchild criteria.

From the quantitative Goodchild criteria, rHEALPix DGGs, as noted by Gibb et al. (2013), fulfills criteria 1, 8, 9, and 14. This means that rHEALPix DGGs addresses the whole globe with the hierarchically organized cells, without overlaps between cells or constraints related to the minimum size of the cell. Each cell also contains a unique reference point which Gibb et al. (2013) refer to as nucleus point to emphasize that it is not a cell's centroid. In the  $(n, s)$ -rHEALPix projection plane,

cells' nuclei points and centroids coincide, but when they are projected to the ellipsoid by the inverse of the  $(n, s)$ -rHEALPix projection, they are not centroids of the ellipsoidal cells and are thus called nuclei. As a direct consequence, criterion 10 is not fulfilled. Gibb et al. (2013) provide mathematical formulations for calculating ellipsoidal cells' centroids coordinates, however, they are not defined as reference points and require numerical integration. Along with criterion 10, criteria 3 and 13 are also not fulfilled. Both criteria are fulfilled in the equatorial region of the rHEALPix DGGS, but they fail at the polar regions. In the polar regions, skew quad cells have four edges, dart cells three, and polar cells only one. This means that cells do not have the same topology and are thus not fulfilling criterion 3. Criterion 13, which requires simple link between DGGS grid and geographical grid of meridians and parallels, is again not achieved in the polar region (Figure 3.6).

From the seven quantitative criteria, rHEALPix DGGS fulfills criteria 2, 6, and 12, as reported by Gibb et al. (2013). Wang et al. (2021) categorized as quantitative those criteria that can be evaluated numerically and they also proposed various indicators for such calculations. These calculated values should then serve as an indication on how good specific DGGS conform to the criterion requirements. However, criterion 2, which requires that all cells have the same area, does not require calculations in the case of equal-area DGGS. Evaluating cell area without considering any theoretical assumptions makes sense only for determining whether a specific equal-area DGGS implementation is not deviating significantly from being truly equal-area (as we have done at the end of Section 4.1). Therefore, for the rHEALPix DGGS, criterion 2 can be considered as qualitative since it is analytically fulfilled. Situation is similar with criterion 6, which requires straight cell edges in the projection. It is also fulfilled for the rHEALPix DGGS since cells in the plane are constructed as squares (that have straight edges) and then projected to the ellipsoid/sphere. Calculations for determining the level of conformance to this criterion requirement are thus again not required in this case. Criterion 12 can be a little problematic, depending on the perspective from which it is approached. Kimerling et al. (1999) put an emphasis in this criterion on grid cell and reference point regularity from the indexing perspective, and point out that this regularity should ensure that some basic operations on grids, such as determining cell neighbor or traversing through resolutions, are simple. From this perspective, since cell indexing for the rHEALPix DGGS is performed in a plane where grid is rectangular and formed of isomorphic squares, it can be concluded that grid is highly

regular and enables conducting mentioned basic operations efficiently. Luo et al. (2023) and Wang et al. (2023) observe this criterion more from the perspective of the cell geometry on the sphere/ellipsoid and they propose calculating landscape shape index and fractal dimension in order to quantify it. Although for some other DGGs this might be an interesting indicator for the highly regular rHEALPix DGGs and from the indexing perspective we do not find it important.

The four remaining quantitative criteria that Gibb et al. (2013) did not find to be fulfilled are 4, 5, 7, and 11. Criteria 7 and 11, that are related to cell centroids and reference points, are not met by the rHEALPix DGGs. This is clear from observing cells in the polar rHEALPix DGGs region on Figure 3.6 and in Appendices A.1 and A.2. We did not examine these criteria further since we do not find them relevant from the land cover data analysis perspective. Criterion 4 is certainly a quantitative one since it is not possible to tessellate sphere, and especially not ellipsoid, into a set of spherical polygons that have completely the same shapes, i.e., that are completely regular and congruent, except from the spherical versions of the five Platonic polyhedra (Lukatela, 2002). This means that besides Platonic spherical polyhedra, this criterion cannot be fully met and is therefore meaningful calculating quantitative indicators that should enable determining a level of conformity. Under the same cell topology (i.e., same number of edges), cell compactness (Section 3.4.3), or more specifically, its stability can be used as a good measure of stability of cells' shapes. Stable compactness should correspond to stable cell shapes. Although topology of all rHEALPix DGGs cells is not the same, since most of the cells are quadrangles (quad and skew quad cells), at most two are circular (cap cells), and only a fraction of all cells are triangles (dart cells; according to Gibb et al. (2013), only  $4 \lfloor N_{side}^i / 2 \rfloor$  of them in resolution  $i$ ). This means that for the rHEALPix DGGs, we can merge criterion 4 with criterion 5, which is specifically related to compactness. This is also true for the QPix DGGs that has all cells as that are quadrangles. Detailed analysis of compactness and comparison between rHEALPix and QPix DGGs are given in Section 4.2.1.

Since QPix DGGs is a modification of the rHEALPix DGGs, it inherited some of its properties, while some are not directly inherited but are also shared with QPix DGGs. From the qualitative Goodchild criteria, fulfillment of criteria 8, 9, and 14, and from the quantitative criteria, criterion 12 (that should, as already discussed, be considered as qualitative in this case) are directly inherited. This is due to planar grid versions being geometrically equivalent (apart from scaling) for

rHEALPix and QPix DGGs and the same recursive subdivision method that is also the same. Criterion 1 (grid addresses global domain without overlapping) is also fulfilled by the QPix DGGs, this time it is not inherited from the rHEALPix DGGs but is a result of QPix DGGs design approach. Criteria 2 and 6 are also fulfilled; grid of congruent squares organized in the same way as for the rHEALPix DGGs is being projected from the plane by the inverse  $(n, s)$ -QPix map projection. This means that QPix DGGs is an equal-area DGGs (criterion 2) and that planar cells have straight lines for their edges (criterion 6).

From the rest of criteria, since all ellipsoidal cells for the QPix DGGs are quadrangles and thus have the same topology, criterion 3, which was not met by the rHEALPix DGGs because of cap and dart cells, is for the QPix DGGs fully met. Criterion 10 is not satisfied by the QPix DGGs, and it was not intended to be satisfied by design. Although Wang et al. (2021) classify this criterion as qualitative, we believe that for some applications it can make sense to calculate distance between cell's reference point and its centroid in order to compare various DGGs. We decided not to perform this analysis since we do not see its direct value from the perspective of land cover data. The situation is the same for criterion 7 and 11, they are not met for the QPix DGGs but evaluating a level of their conformance we do not find significantly important for the land cover data. Although criterion 13 is not fully met for polar regions of the rHEALPix DGGs, it is still met for majority of cells (all cells in the equatorial region). In the polar regions, this criterion is only partly met. For skew quad cells, north and south cell edges follow parallels of latitude, south edge of dart cells in the north polar region and north edge in the south polar region also follow parallels, but the rest of the edges do not follow neither parallels nor meridians (figures in Appendices A.1, A.2, and A.5). For the QPix DGGs, this criterion is in general not met. Ellipsoidal cell edges for the QPix DGGs are not geodesic lines or lines from some other known class of lines on the ellipsoid but are rather some arbitrary lines (figures in Appendices A.1, A.2, and A.5). Only edges that are in plane laying on the axis-aligned straight lines drawn from the centers of each resolution 0 cells will be mapped to the meridians or parallels. This is because of the characteristics of the Rořca-Plonka projection, which, as explained by Rořca and Plonka (2011), is invariant in these directions. Additionally, edges can have this property of lying on axis-aligned lines only for even values of  $N_{side}$  parameter (Figure 3.13, figures in Appendix A.1).

The above findings across Goodchild criteria are summarized in Table 4.1. As already discussed, for some of these criteria it can be hard to determine whether they can be considered as qualitative (i.e., those that should use ticks and crosses in Table 4.1) or quantitative (i.e., those that should be computed and then use star symbols in Table 4.1). Therefore, we did not stick to the strict classification by Wang et al. (2021), but rather intended Table 4.1 to allow comparison between rHEALPix and QPix DGGs from the aspects that we consider important and that we find relevant for the land cover data. In this table, criteria 4 and 5 are also listed, however, their more thorough analysis is provided in Section 4.2.1.

**Table 4.1** Comparison of rHEALPix and QPix DGGs based on Goodchild criteria (as recorded by Kimerling et al. (1999)). Green tick means that criterion is fully met, red cross that it was not met, either fully or not at all, and for some criteria, number of stars correspond to the level of criterion fulfillment (five stars are maximum).

Criterion	Description	rHEALPix	QPix
1	Addresses the global domain without overlapping.	✓	✓
2	Cells have the same area.	✓	✓
3	Cells are topologically equivalent (i.e., have the same number of edges and vertices).	✗ (★★★)	✓ (★★★★★)
4 and 5	Cells have the same shape and are compact.	★★★★	★★★★★★
6	In map projection plane, cell edges are straight lines.	✓	✓
7	When the center points of two neighboring cells are connected with arc, its midpoint matches the midpoint of their joint edge.	✗	✗
8	Hierarchy is established between grids of various resolutions.	✓	✓
9	Each cell has only one reference point.	✓	✓
10	Reference points are centered as much as possible within the cell.	✗	✗
11	Distance between reference points of adjacent cells is equal.	✗	✗
12	Grid is regular and can thus be efficiently indexed.	✓	✓
13	Grid can be easily related to the grid of meridians and parallels.	✗ (★★★★)	✗ (★)
14	Spatial resolution of grids within the grid system can be defined arbitrarily.	✓	✓

Before moving to the compactness comparison between rHEALPix and QPix DGGs, we would like to mention that the isolatitude property of the rHEALPix DGG cell nuclei is not retained in the QPix DGG. This property is lost because of the properties of projections used for mapping grids from plane to the sphere/ellipsoid. rHEALPix DGG uses Lambert cylindrical equal area projection for equatorial region and pseudocylindrical Collignon projection for polar regions, both of which map parallels (and meridians) to straight lines in normal aspect (Kerkovits, 2023). In plane, cell nuclei (equivalent to cell centroids) that lie on the same straight line that is parallel to abscissa (i.e., projection of parallel) will thus all lie on the same parallel when mapped to the sphere/ellipsoid. Simple proof that confirms this property was lost for the QPix DGG can be drawn from Figures 3.8 and 3.9. Composition of Rořca-Plonka and LAEA projection that is used in the QPix DGG maps intersecting line between sphere and diagonal planes of the sphere-inscribed cube to straight lines. This interesting line is by definition a great circle and thus not a parallel, which means that not a significant number of cell nuclei on a specific DGG refinement level lie on the same parallel. Isolatitude property, as already stated, enables efficient calculations of spherical harmonics. From the perspective of land cover data analysis, which most often includes determining areal statistics, we do not believe this fact can be perceived as a significant drawback of the QPix DGG.

#### 4.2.1 Compactness of the rHEALPix and QPix DGG cells

As already discussed in Section 3.5.2, stability of compactness is a good measure of cell shape stability. The higher the compactness stability, the higher the shape stability. As explained by Wecker et al. (2024, Figure 2), equal area geodesic DGG (both, rHEALPix and QPix DGG are geodesic DGGs that use cube as base polyhedron) ensure constant cell area, but at the cost of distortions in cell shapes. However, not all DGGs introduce the same levels of shape distortions.

In the rHEALPix and QPix DGG, cell areas are constant, and will thus high compactness stability mean that cell shapes are highly regular. Although high compactness values are desirable from the land cover data perspective (in relation to the Tobler’s first law of geography, as recognized by Kimerling et al. (1999)), because of the reasons stated in the beginning of the Chapter 3, having quadrangular cells is more important. Quadrangular cells in general clearly deviate from the spherical cap that is on the sphere considered as the most compact shape, so we are not aiming at cell shapes

being as compact as possible (compactness values being as close to 1 as possible), but rather to low dispersion of compactness values.

Regarding the initial tessellation (i.e., first hierarchical level, resolution 0), QPix DGGS effectively uses gnomonic projection to map edges of the sphere-inscribed cube to the sphere and is therefore creating one of the five possible tessellations of the sphere that are composed of regular and mutually congruent spherical polygons. Because of this, compactness of resolution 0 cells of the QPix DGGS should be constant, as can be observed on Figures 3.25 and 3.26 (top left graphs). On these two figures compactness is not fully constant (i.e., line is slightly bended) because these figures are related to the DGGS on the WGS 84 ellipsoid and not sphere. When cells are mapped from the sphere to the ellipsoid by using authalic latitude, their shape gets slightly deformed. Since WGS 84 is an oblate ellipsoid of revolution and has the highest curvature at the equator, four cells that intersect the equator come closer to the circular shape and become slightly more compact. On the second level (resolution 1) and for  $N_{\text{side}} = 2$ , cells of the QPix DGGS have lower, but again almost constant compactness since they all have almost the same (but not regular) shape (this tessellation would be the same as with gnomonic projection; Appendix A.1). For all other refinements, compactness gets dispersed. For the rHEALPix DGGS, initial tessellation is not regular, and it was not designed to be. The north and south polar regions of the rHEALPix DGGS are cells themselves in the first level and are spherical/ellipsoidal caps. Therefore, as can be observed on Figures 3.25 and 3.26, they have the highest possible compactness value of 1. When  $N_{\text{side}}$  parameter is odd number, on each refinement level rHEALPix DGGS has two cap cells that have compactness value equal two 1 and this compactness values can be taken as outliers since they are significantly different than the rest of the values (Figure 3.26). Similar is the case for the odd values of  $N_{\text{side}}$  parameter for QPix DGGS for cells that are centered within each of the six initial cells. As visible on figures in Appendix A.2, these cells are highly circular and thus have high compactness values. The same, but with a lower magnitude, is also true for the four QPix DGGS cells that have one vertex corresponding to the centroid of the one of the six initial cells in a refinement with even  $N_{\text{side}}$  values. Additionally, from the Figures 3.25 and 3.26, it is visible that compactness values for the rHEALPix DGGS significantly deviate from the compactness values of cells in the equatorial region. This is the main reason why compactness ranges



(Figure 3.27) and standard deviations (Figure 3.28) are consistently larger and means (Figure 3.29) are consistently lower for the rHEALPix DGGS than they are for the QPix DGGS.

Compactness ranges on Figure 3.27 and standard deviations on Figure 3.28 again show that cell shapes are the most stable (compactness ranges and standard deviations are near zero) at resolution 0 for the QPix DGGS, while for the rHEALPix DGGS at the same resolution they are found to be the largest overall. The reason behind these large values has already been discussed; it is related to the fact that for the initial rHEALPix DGGS tessellation two cells are polar spherical/ellipsoidal caps that have maximum compactness value of 1, while other cells are spherical/ellipsoidal quadrangles that are bound by two meridians and two parallels and thus have significantly lower compactness. For the rHEALPix DGGS resolution 1, range and standard deviation experience a steep fall in comparison to the resolution 0. In case of 1-to-4 refinement, this is due to absence of highly compact cap cells in resolution 1, and in case of 1-to-9 refinement it is related to the fact that compactness values for cap cells were filtered out as they are identified as outliers. As the resolution increases, for both analyzed refinement ratios and for both DGGSs, compactness ranges and standard deviations are stabilizing. There is no reason to expect that some irregularities might occur on the higher resolutions that were not analyzed. Therefore, from Figures 3.27 and 3.28, it can be undoubtedly concluded that cell compactness is consistently more stable across resolutions for the QPix DGGS than it is for the rHEALPix DGGS.

In addition, Figure 3.29 shows that mean compactness values are also consistently larger for the QPix DGGS. This means that QPix DGGS cells are in general more compact. Also, if consider that compactness of square is approximately 0.8, this means that from data on Figure 3.29, QPix DGGS cells are on average closer to being square-shaped than rHEALPix DGGS cells are. This can also be perceived as some kind of advantage of the QPix DGGS if the aim is ensuring that cell shapes are in resemblance with earth observation satellite sensor detector footprints, which are, in general, square-shaped.

Regarding the geographical distribution of cell compactness values, the following can be observed from the Figures 3.30, 3.31, and 3.32. On Figure 3.30, it is visible that compactness of the rHEALPix DGGS is the most stable in the equatorial region and becomes unstable in the polar regions. In polar regions, deviations are becoming the largest when approaching dart cells. This is

expected since the shape of dart cells is triangular and thus notably different from quadrangular shape of other cells. For the QPix DGGS (Figure 3.31), deviations are the largest for cells that in the plane lie on the diagonals of the initial resolution 0 squares. From figures in the Appendices A.1 and A.2 it is visible that these cells have the most distorted shapes. However, by observing values in the color bars of these two figures, it is visible that magnitude of compactness deviations is lower for the QPix DGGS. This is also confirmed on Figure 3.32, where it is, once again, clearly visible that QPix DGGS has cells with more stable shapes.

### 4.3 Comparison of QPix DGGS and “traditional” global gridding approaches

In Section 3.1, the main drawback of currently applied approaches for global gridding were examined. To reiterate, global raster land cover data are currently being delivered in projected or geographic CRSs (Table 1.1). Global grids that are defined by rasters in projected CRSs are burdened with deformations of shapes and/or areas, they introduce the need for resampling of initial raw observations resulting in degradation of data quality and possibly loss or replication of data. Even the grid defined in the projected UTM-based CRSs, which exhibits deformations that are negligible for most applications, brings challenges for global data handling since this grid is effectively defined in 60 different CRSs. In geographic CRS, the greatest problem from the land cover data is that equiangular geographic grid defines tessellation that is not equal area. It further introduces replication of data because of meridian convergence. Geodesic DGGSs, on the other hand, since they treat earth’s surface as being homeomorphic to a sphere/ellipsoid (Tobler, 1993), solve, or at least reduce some of these issues.

In addition to these qualitative challenges of dealing with grids defined in projected and geographic CRSs, in this section our aim is to briefly examine quantitative indications of performance in achieving uniform global tessellation between DGGS, namely QPix DGGS, and these “traditionally” applied approaches. As a uniform global tessellation, we consider the one that exhibits high stability of cell areas and shapes. We examined three global grids: QPix DGGS, equiangular geographic grid (EGG), and grid that is based on interrupted Goode Homolosine (IGH) map projection. We chose IGH map projection because it is often recommended choice from the equal

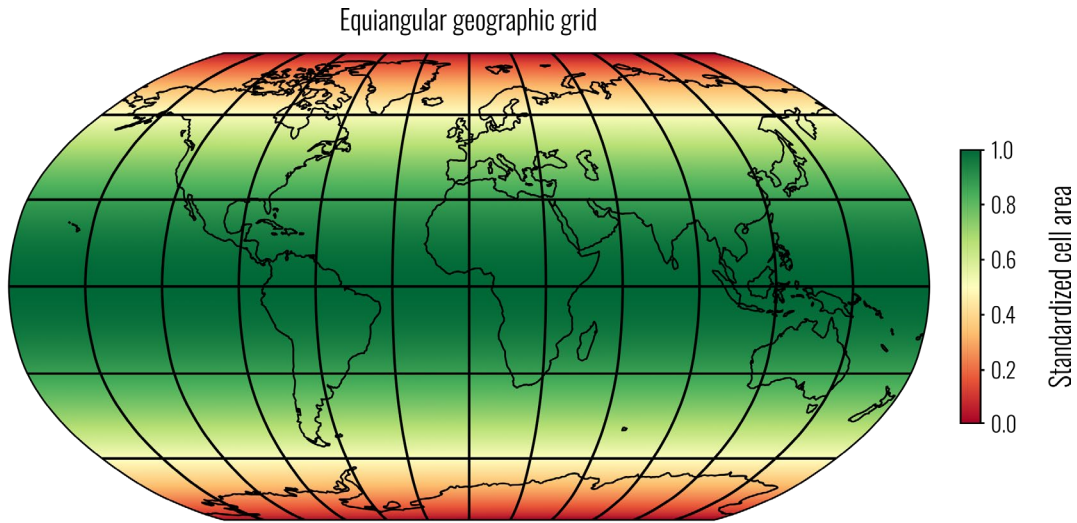
area projections that are suitable for performing global analyses (Moreira de Sousa et al., 2019; Simoes et al., 2024). All these three grids were again constructed on the WGS 84 ellipsoid (EPSG, n.d.-a).

The QPix DGGS grid was fixed to resolution 5 with  $N_{\text{side}}$  parameter set to 3 (1-to-9 refinement; hereafter referred to as QPix grid). According to Table 3.2, on WGS 84 ellipsoid, this QPix grid has cells that have an area of  $1\,439\,667\,681.99\text{ m}^2$ . The square root of this value is  $37\,942.95\text{ m}$ , and it can be taken as a rough estimate for length of an edge of quadrangular QPix cells. EGG and IGH-based grids were formed with these values in mind in order to ensure that all three grids are in resemblance regarding the cell sizes. Therefore, EGG grid was defined so that cell sizes are  $0.3155'' \times 0.3155''$ , which at the equator corresponds to cells of approximately  $34\,886.20\text{ m}$  in east-west direction and  $35\,121.30\text{ m}$  in north-south direction. Constructing an IGH-based grid included more steps. IGH projection is a spherical projection, so we needed to establish a mapping from sphere to ellipsoid and vice versa. We again used authalic latitude for this purpose since we want to retain equal area property of IGH projection. The first step was thus projecting the authalic sphere of the WGS 84 ellipsoid (authalic sphere radius is approximately  $6\,371\,007.1809\text{ m}$ ) to the plane by the IGH projection. This way, we determined the co-domain of the projection, i.e., the extent for defining tessellation in the plane. We defined cells as squares having sizes of  $38\text{ km} \times 38\text{ km}$  in plane, meaning that their area on authalic sphere and on WGS 84 ellipsoid will be  $1\,444\text{ km}^2$ , again relatively close to the area of QPix grid cells. In order to avoid issues in subsequent steps, only cells that were fully within the co-domain of the IGH projection were selected, and all other cells were discarded. These cells were projected to the authalic sphere by the inverse IGH projection and then by applying authalic latitude to the WGS 84 ellipsoid.

For all these three grids, we calculated areas and perimeters of all cells on the WGS 84 ellipsoid. We applied the same procedures as explained in Sections 3.4.1 and 3.4.2 for these calculations, with a difference that we densified each cell edge with only 20 vertices, mainly to speed up processing, while also not expecting any significant impact on comparison results between these grids. From cell perimeter and area, we also calculated spherical compactness of each cell, in line with Equation 12 in Section 3.4.3.

Again, we consider uniformity of global grid, which essentially defines global sampling, from the perspective of stability of cell areas and shapes (higher stability–higher grid uniformity). As

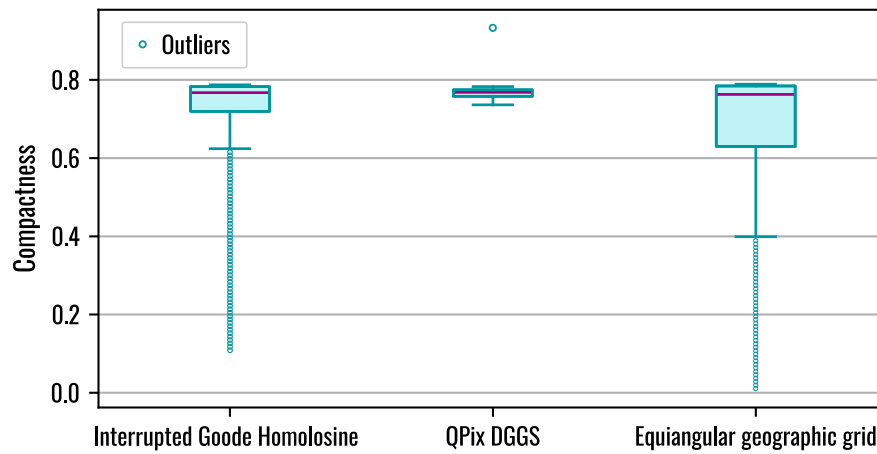
already discussed, stability of cell shapes is examined through stability of cell compactnesses. Regarding the cell area, QPix grid and IGH-based grid define equal area tessellation, i.e., cells have constant area. In other words, and in an analytical sense, these two grids achieve full stability of area. Variations in area can only occur because of the numerical instability related to computations in the grid construction process or because of the flaws in the area calculation approach (e.g., insufficient number of densification points). The EGG grid on the other hand does not. Only cells that are intersected with the same parallel and its antiparallel (i.e., parallel symmetrical with respect to the equator) have the same area. This is, of course, a well-known fact and is visible on Figure 4.3. This figure shows standardized cell area which was calculated in line with the suggestion by White et al. (1998)—as a cell-wise ratio between cell area and maximum observed cell area. Clearly, from the perspective of stability of cell area, QPix grid and IGH-based grids are a better choice than the EGG grid.



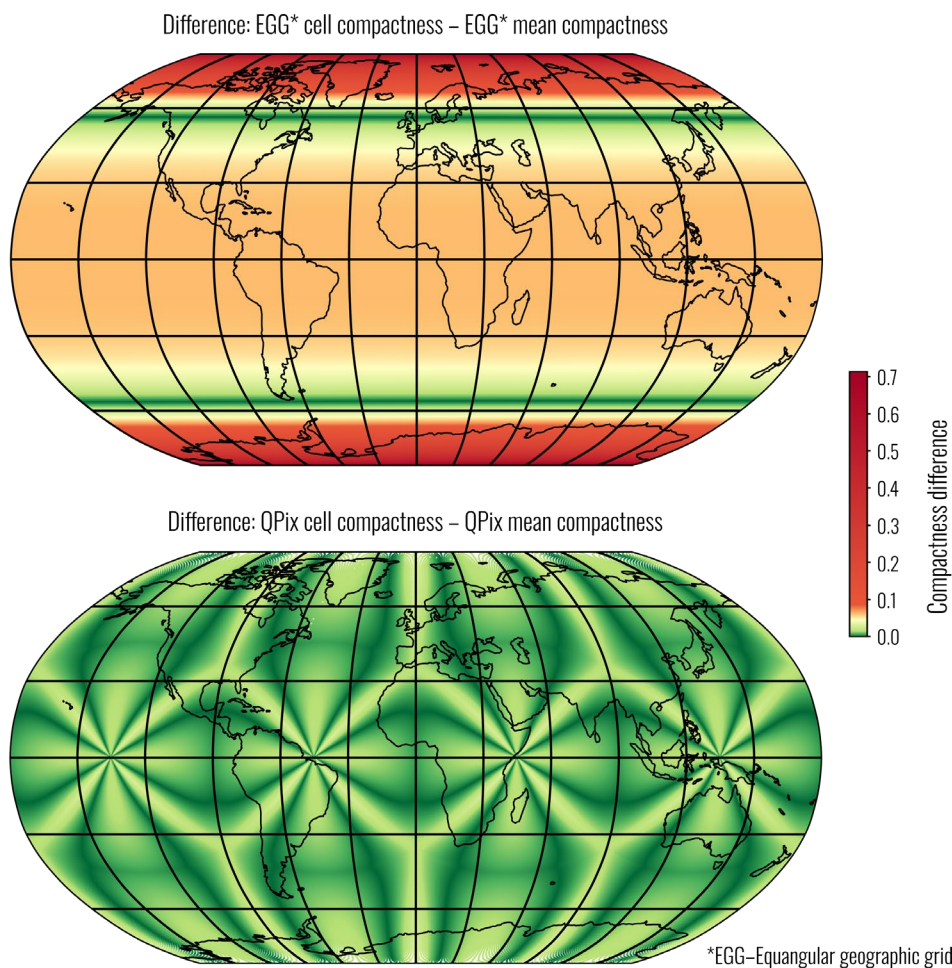
**Figure 4.3** Distribution of the standardized cell area of the EGG grid.

Regarding the stability of cell compactness, Figure 4.4 shows box plots of calculated compactness values. QPix grid has the lowest range and interquartile range, IGH-based grid has larger, and EGG has the largest. This means that cell shapes are the least dispersed, or the most stable for the QPix grid.

Figure 4.5 shows spatial distribution of cell-wise absolute differences between mean compactness value and each cell compactness for the EGG grid and QPix grid. Outlier-adjusted mean compactness values and then subsequent absolute differences were calculated separately for the EGG



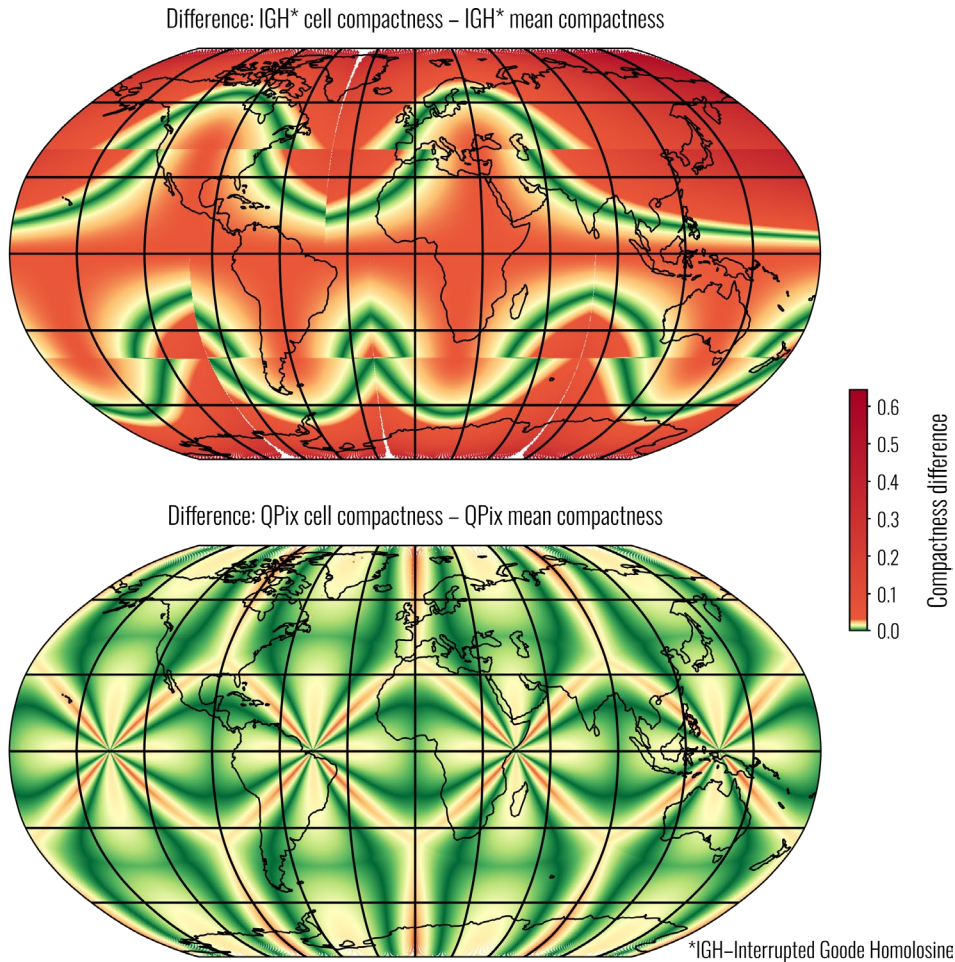
**Figure 4.4** Box plots of cell compactness for three analyzed global gridding approaches: gridding based on interrupted Goode Homolosine projection, QPix DGGS grid, and equiangular geographic grid.



**Figure 4.5** Cell-wise absolute differences between cell compactness and mean compactness for the EGG (equiangular geographic grid) and QPix grid. Mean compactness as well as differences are calculated separately for the EGG and for QPix grids.

grid and for the QPix grid, however, they were colored using the same color bar to make them directly comparable. Color bar was defined so that approximately the same number of cells will be shown using the same color. The same was done for the comparison between compactness differences of IGH-based and QPix grids and results are shown on Figure 4.6.

On Figures 4.5 and 4.6, lower values are preferred if the aim is to ensure high stability of cell shapes. Distribution of compactness differences is for QPix grid, as expected, the same as on Figure 3.32. For EGG grid on Figure 4.5, cells that are near  $\pm 65^\circ$  latitude have compactness that is near the mean of all cells compactness values. This clearly indicates that for EGG grid, compactness is not linearly distributed along meridians and gets highly dispersed by approaching poles (also visible from distribution of standardized cell areas on Figure 4.3). Distribution of compactness differences for



**Figure 4.6** Cell-wise absolute differences between cell compactness and mean compactness for the IGH-based (IGH–interrupted Goode Homolosine) and QPix grid. Mean compactness as well as differences are calculated separately for the IGH-based and QPix grids.

IGH-based grid on Figure 4.6 shows some clear lines of discontinuity that are related to the IGH map projection definition. This projection is a combination of interrupted Sanson's sinusoidal projection for latitudes up to  $\pm 40^\circ 44' 11.8''$ , and interrupted Mollweide's projection for latitudes beyond (Goode, 1925). Therefore, a line of discontinuity in compactness difference distribution on Figure 4.6 is following parallels where switch between two projections is made. Additional lines of discontinuity follow meridians along which projections are interrupted. According to Goode (1925) and Steinwand (1994), for the northern hemisphere, meridian that has longitude of  $-40^\circ$ , and for the southern, meridians with longitudes of  $-100^\circ$ ,  $-20^\circ$ , and  $80^\circ$  are meridians along which interruptions are introduced. In addition to discontinuities, along these meridians on Figure 4.6 can also be observed gaps that are a result of excluding cells from the IGH-based grid that were not fully within the extent of the co-domain of the IGH projection of the authalic sphere.

From these two figures above, and from box plots on Figure 4.4, it is clear that cells in the QPix grid have a higher shape stability than cells in the IGH-based and EGG grid. Therefore, an overall conclusion is that (QPix) DGGS solves or at least reduces issues that are related to handling global raster/gridded data in geographic and projected CRS, while also, specifically in the case of the QPix DGGS, provides a more uniform global tessellation/sampling. Lastly, we would like to acknowledge a couple of limitations related to this conclusion. Since we concentrated only on global tessellation approaches that are currently employed in relation to land cover data, this conclusion should also be viewed from that perspective. For some other application areas, aspects that we see as disadvantageous might not be disadvantageous at all. Then, we disregarded UTM-based tessellation, mainly because it uses 60 different CRSs to georeference global data. We see this as a deficiency since it, in general, does not allow global data to be stored within the same data file, which is something that is highly desirable in respect to cloud-optimized data formats. However, in some processing pipelines, this might not pose an obstacle. And finally, it is certainly possible that there is some global tessellation that exhibits higher uniformity than that of QPix DGGS, however, our goal was not to achieve the most uniform tessellation, but rather to show that it is possible to overcome some of the issues that are related to current approaches related to the land cover data storage and handling.



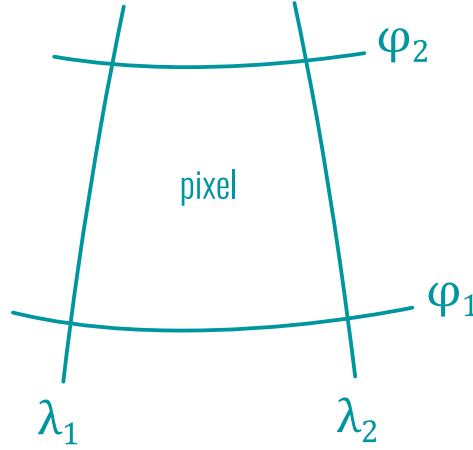
## 5 Application of DGGS on land cover data

In this chapter, we used the GLC\_FCS30D global land cover dataset (X. Zhang et al., 2024) for calculating land cover change. As already stated in Section 1.1.4, GLC\_FCS30D land cover product includes 35 land cover classes and is suitable for land cover change analysis because it applies a mechanism that ensures temporal stability of land cover classification. In these land cover change calculations, we applied two constraints in order to reduce computational overhead but that will not reduce relevance of the results for the aim of this chapter—demonstrating that it is possible to use DGGS as a georeferencing system for land cover data. The first constraint was to focus only on built-up land cover class, or more specifically, land cover class that is referred to as “impervious surface” in GLC\_FCS30D. As X. Zhang et al. (2024) state, they have developed a separate land cover product that focuses specifically on this land cover class and have subsequently superimposed it with GLC\_FCS30D data to ensure high classification accuracy of impervious surfaces. Additionally, this land cover class is interesting because observing its changing over time is often perceived as an indicator of impact of human activities on the environment and vice versa. The second constraint was to limit analysis only on the coastal zone of the Adriatic Sea. This coastal zone is defined as a land that goes 10 km from the coastline (hereafter referred to as 10-km coastal zone). This zone was chosen because it is often perceived as an area with a high level of urbanization, which means that, again, results might be interesting for examining human impact on environment. However, we acknowledge that the choice of the geographical extent for the analysis has no relevance in demonstrating application of DGGS for land cover data. Finally, we decided to calculate built-up land cover change in this 10-km coastal zone for the 10-year period, between 2010 and 2020.

GLC\_FCS30D data are delivered as GeoTIFF files georeferenced in WGS 84 geographic 2D CRS (EPSG, n.d.-c). In all GeoTIFF files, pixels (or grid cells) have size of  $0.97'' \times 0.97''$  (Table 1.1). This means that land cover data in GLC\_FCS30D are provided in an equiangular geographic grid and that pixel edges correspond to meridians of longitude and parallels of latitude. Since the 10-km coastal zone approximately spans between  $39^\circ 38' \text{ N}$  and  $45^\circ 53' \text{ N}$ , the cell sizes vary from 20.92 to 23.13 m along parallel (i.e., east–west) and from 29.92 to 29.95 m along meridian (i.e., north–south). These distances can be easily calculated by using equations that are available in (Lapaine, 1994) and



in (Rapp, 1991, Equation 3.115 and 3.122). Regarding an area of the pixels, since they are bound by two parallels and two meridians, a closed-form equation is available (Kelly and Šavrič, 2021; Lapaine and Lapaine, 1991). By using these equations, it can be calculated that pixel areas within the geographical extent of land cover change calculation vary from 626.40 m<sup>2</sup> to 692.00 m<sup>2</sup>.



**Figure 5.1** Pixel in a raster that is georeferenced in the equiangular geographical grid is bound by two meridians with longitudes  $\lambda_1$  and  $\lambda_2$  and two parallels with latitudes  $\varphi_1$  and  $\varphi_2$ .

## 5.1 Methodology

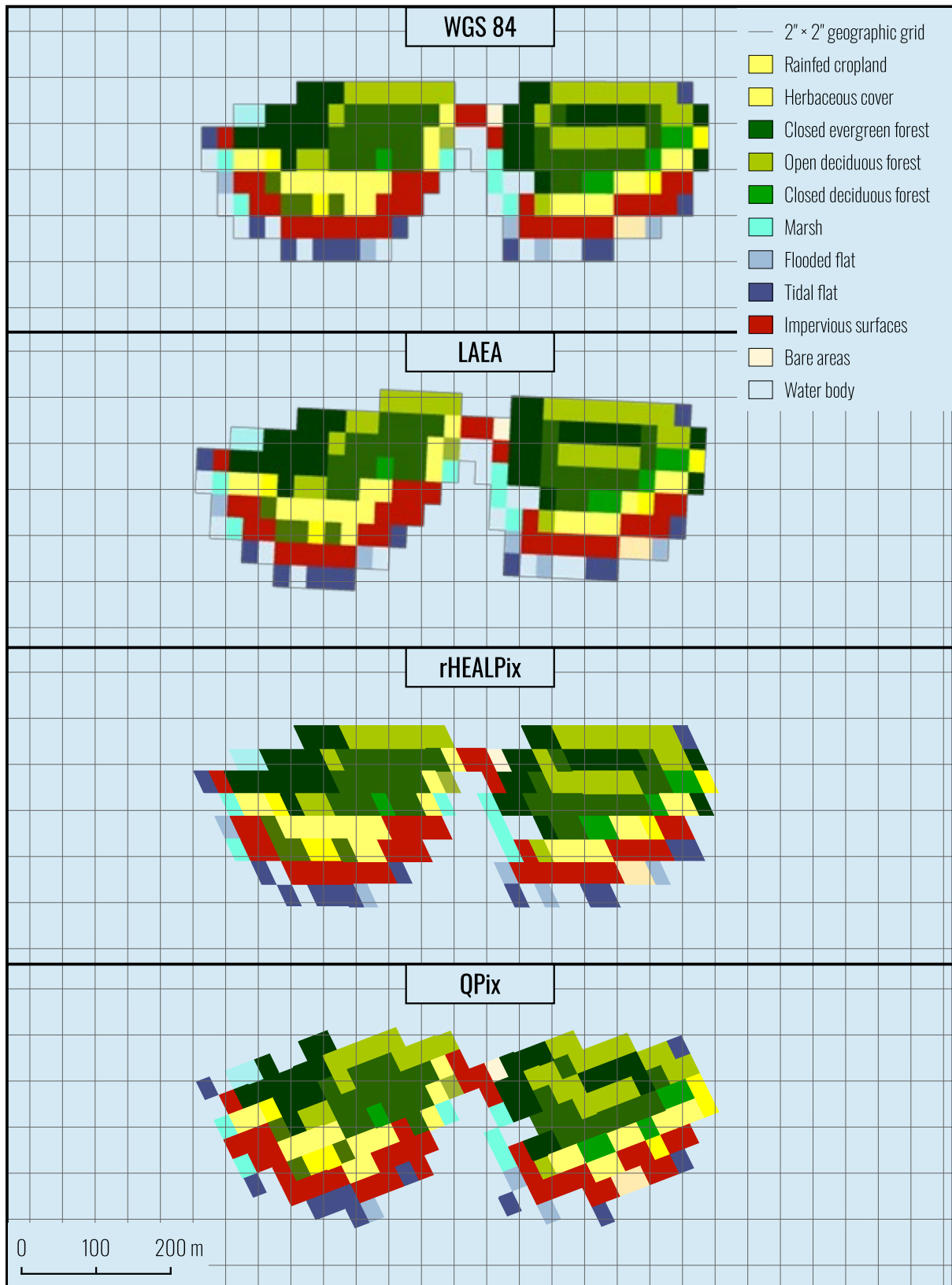
The first step of calculating land cover change was generating a vector layer with polygon corresponding to the 10-km coastal zone. Coastline was extracted from the OpenStreetMap data (<https://www.openstreetmap.org/>), transformed to the WGS 84 geographic 2D CRS and then a 10-km geodesic buffer zone was generated on the ellipsoid on the inland side of the coastline. Country administrative borders were also extracted from the OpenStreetMap data and after transformation to the WGS 84 CRS, intersected with the 10-km coastal zone. This produced six vector polygons, one for each country (Albania, Bosnia and Herzegovina, Croatia, Italy, Montenegro, and Slovenia; San Marino was also partly within the 10-km coastal zone but was excluded because of relatively small area in comparison to other countries), that were used for aggregating built-up land cover data.

The area of built-up land cover class for 2010 and 2020, as well as their change was calculated by using four methods. The main difference between methods was related to their adaptation to the way land cover data were georeferenced. In the first method, areas of built-up land cover class were calculated directly from the original, unaltered GLC\_FCS30D data, meaning that data were

georeferenced in the WGS 84 geographic CRS (hereafter, WGS84-based method). In the second method, original GLC\_FCS30D data were reprojected to the “ETRS89-extended / LAEA Europe” projected CRS (EPSG, n.d.-b; hereafter LAEA-based method). We chose this CRS because it is suggested by the INSPIRE Data Specification of Geographical Grids (INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1, 2023) for spatial statistical reporting within continental Europe. The third and fourth method were based on the rHEALPix (hereafter, rHEALPix-based method) and QPix (hereafter, QPix-based method) DGGSs, respectively. Figure 5.2 shows initial GLC\_FCS30D data georeferenced in the WGS 84 geographic CRS, as well as the same data after reprojecting and resampling (i.e., regridding) into LAEA-based projected CRS and rHEALPix and QPix grids. The rHEALPix and QPix grids are shown on this figure for visualization purposes only, since only pixels that were classified as built-up were regridded from the initial GLC\_FCS30D data to the rHEALPix and QPix grids, as explained later on.

For the second, LAEA-based method, in order to maintain similar sizes between initial pixels and pixels in reprojected raster, we chose pixel sizes in LAEA-based CRS to be 22 m along abscissa and 30 m along ordinate. As a result, all pixels in the LAEA-based CRS represent an area of 660 m<sup>2</sup> and correspond to cells on ellipsoid that have that same area. Since initial GLC\_FCS30D pixels have an area between 626.40 m<sup>2</sup> and 692.00 m<sup>2</sup> within the geographical scope of analysis, we believe that pixels with an area of 660 m<sup>2</sup> are their good approximation. As a resampling method, the nearest neighbor method was applied, which is the only sensible resampling method for categorical land cover data with discrete classes.

For the rHEALPix-based and QPix-based methods, the first step was determining resolution (i.e., refinement level) and value of the  $N_{\text{side}}$  rHEALPix and QPix DGGS parameter (i.e., refinement ratio) that will yield cells of approximately the same area as initial GLC\_FCS30D cells. Based on areas of rHEALPix and QPix DGGS cells on WGS 84 ellipsoid for various resolutions and values of  $N_{\text{side}}$  parameter that are provided in Annex B, resolution 5 grid for the  $N_{\text{side}}$  parameter value of 13 yields cells whose area most closely match those of original GLC\_FCS30D pixels. For both the rHEALPix and QPix DGGSs at this resolution, each cell has an area of 616.65 m<sup>2</sup>. Although this area is lower than the minimum area of the GLC\_FCS30D pixels within the extent of analysis,



**Figure 5.2** Initial GLC\_FCS30D land cover data in WGS 84 CRS (on top) and the same data after reprojecting and resampling in the LAEA projection (22 m × 30 m pixels), and rHEALPix and QPix N<sub>side</sub> = 13, resolution 5 grids. Figure shows Sveti Ivan island near Rovinj, Croatia in the transverse Mercator projection with central meridian 13° 37' E.

we do not find it as problematic as in opposite situation—area of DGGS cells being larger than the maximum GLC\_FCS30D area. If DGGS cell area is larger than initial raster pixel area, then, as was also recognized by Kmoch, Matsibora et al. (2022), certain level of data loss during resampling might occur. On the other hand, if DGGS cell area is smaller, then data replication, which results in additional processing demands, is expected. Since DGGS cells do not have area that is drastically lower than the GLC\_FCS30D pixel area, we do not expect this issue will have impact on final results. To map GLC\_FCS30D data to these rHEALPix and QPix grids, we again applied the nearest neighbor resampling method. For each rHEALPix and QPix cell we determined its reference point (i.e., nucleus according to the rHEALPix terminology) and then extracted the value of the GLC\_FCS30D pixel (i.e., land cover class) that contains that reference point. If pixel was classified as built-up land cover class, we stored the corresponding cell's reference point in GeoParquet (<https://geoparquet.org/>) file, along with information whether pixel is classified as built-up for 2010 and/or 2020. This way we generated two GeoParquet files, one for the rHEALPix and another one for the QPix grid.

After this processing, calculating land cover change for the LAEA-, rHEALPix-, and QPix-based methods is straightforward. For the LAEA-based method, for each country we counted how many pixels that are classified as built-up fall within the 10-km coastal zone, separately for 2010 and 2020. These numbers of pixels are then multiplied by the area of each pixel, which is 660 m<sup>2</sup>, and the result is the built-up area for corresponding country and year. For the rHEALPix- and QPix-based methods, procedure is in essence the same, with the difference that instead of pixels, points are being counted. Of course, the number of points is in this instance multiplied by 616.65 m<sup>2</sup> as it is the area of rHEALPix and QPix cells.

For the WGS84-based method, procedure requires a couple of additional steps. As already mentioned, pixels in initial GLC\_FCS30D raster data are bound by two meridians and two parallels and their area can be easily calculated (Kelly and Šavrič, 2021; Lapaine and Lapaine, 1991). Therefore, the first step was to generate an additional raster that is aligned to GLC\_FCS30D raster, has the same pixel size as the pixels in the GLC\_FCS30D raster, and in which each pixel value corresponds to that pixel's area (hereafter, area raster). Concretely, for this purpose we used a *r.mapcalc.simple* command that is available in the GRASS GIS application (GRASS Development Team, 2024). In the

subsequent step, two new rasters are generated—one for 2010 and one for 2020—each retaining only the values (i.e., pixel areas) from the area raster that overlap with pixels classified as built-up land cover in the GLC\_FCS30D dataset for the respective year, while all other pixels are set to 0 or a *nodata* value. Finally, to calculate the area of built-up land cover class for each year, we just needed to sum-up all pixel values in these two rasters that are within each country's 10-km coastal zone.

To reiterate, we have calculated area of built-up land cover class within the 10-km coastal zone for six countries by applying four different calculation methods. We performed these calculations twice, for years 2010 and 2020, and then we additionally calculated changes in built-up areas between these two years. Our goal is to determine whether results obtained by four different methods differ significantly or not. If we concentrate on one year only, results obtained by four different methods can be treated as four dependent groups of measurements, each having six observations, one for each country. Since there are less than 10 observations in each group of measurements, performing a non-parametric statistical test, in this case namely Friedman test, is advised (Corder, 2014; Warner, 2013). We thus performed Friedman test for comparing results of all four methods, separately for 2010, 2020, and for calculated changes in built-up area between 2010 and 2020. We performed these calculations in Python, using SciPy package (Virtanen et al., 2020), in line with procedure explained by Corder (2014). This procedure also includes performing post hoc Wilcoxon test in cases when Friedman test suggests rejecting the null hypothesis of no significant differences between methods. Therefore, we performed the Wilcoxon test pairwise between WGS84-based method and three other methods. This means that WGS84-based method was used as a baseline method, mainly because it uses unaltered GLC\_FCS30D data, while all other methods introduce modifications of initial GLC\_FCS30D data. The Wilcoxon test will thus show whether there are significant differences between WGS84-method and any of the other methods. Corder (2014) also suggest performing Bonferroni procedure to reduce the rate of faulty rejections of the stated null hypothesis by reducing the level of significance. We chose level of significance  $\alpha$  to be 0.05 for Friedman tests, which, after applying Bonferroni procedure reduced to  $\alpha_B = 0.017$  for Wilcoxon tests (again in line with Corder (2014)).

For the Wilcoxon test, we used the SciPy function *wilcoxon*, that returns Wilcoxon test statistic and p-value. This function, by default, ensures that p-value is calculated using the method

that will, based on the characteristics of the two input groups of measurements (e.g., number of observations in each group), yield the most reliable result. Therefore, we used p-value to determine whether there are significant pairwise differences between methods or not. For the Friedman test, we used the SciPy function *friedmanchisquare*, which again returns Friedman test statistic and p-value. The p-values are based on the assumption of chi-squared distribution, which means that p-values are not reliable in cases when there are less than 10 observations in each group of measurements (The SciPy community, 2025). Therefore, for the Friedman test, we primarily used the Friedman test statistic, and compared it against critical values from the lookup table (Corder, 2014, Table B.5) in order to determine whether differences between methods are significant.

## 5.2 Results

All results of calculating area of built-up land cover class for 2010 and 2020, as well as their changes, are available in Table 5.1. Changes in built-up area in this table are calculated by subtracting built-up area in 2010 from built-up area in 2020. This means that positive values correspond to the increase of built-up area, and negative to its decrease. Changes are also expressed as percentages, calculated as the ratio of the change in built-up area to the built-up area in 2010. Table 5.1 additionally provides absolute differences between WGS84-based built-up areas and their changes and other three methods. Differences are also expressed as percentage of the corresponding WGS84-based area.

Table 5.2 lists test statistics and corresponding p-values obtained by the Friedman test, separately applied on each of the three datasets (i.e., on calculated built-up areas for 2010, 2020, and on the changes in built-up areas between 2010 and 2020). Test statistics and corresponding p-values obtained by the Wilcoxon tests for pairwise comparison between WGS84-based method and other three methods within each dataset, are available in Table 5.3. From the data presented in Table 5.2, it is evident that for all three datasets, the Friedman test statistic values fall below the critical value of 7.600, and all p-values, although potentially unreliable, exceed the level of significance  $\alpha = 0.05$ . These results indicate no statistically significant differences among the four applied methods. Wilcoxon tests support this conclusion—all calculated p-values are above the Bonferroni-reduced level of significance  $\alpha_B = 0.017$ . Therefore, based on both tests, it can be concluded that there are no significant differences between results obtained by the four analyzed methods.

**Table 5.1** Results of built-up area calculations obtained by WGS84-, LAEA-, rHEALPix-, and QPix-based methods. All areas are given in square kilometers.

Country (10-km coastal zone)	WGS84-based		LAEA-based		rHEALPix-based		QPix-based	
	Area		Area	Difference	Area	Difference	Area	Difference
<b>Albania</b>	2010	182.819	182.740	0.079 (0.04%)	182.922	0.103 (0.06%)	182.729	0.090 (0.05%)
	2020	192.229	192.142	0.087 (0.05%)	192.264	0.035 (0.02%)	192.152	0.077 (0.04%)
	Change	9.410	9.402	0.009 (0.09%)	9.342	0.069 (0.73%)	9.424	0.013 (0.14%)
	Change (%)	5.15%	5.14%	0.00%	5.11%	0.04%	5.16%	0.01%
<b>Bosnia and Herzegovina</b>	2010	2.433	2.432	0.001 (0.03%)	2.431	0.002 (0.08%)	2.430	0.003 (0.11%)
	2020	2.625	2.625	0.001 (0.03%)	2.624	0.000 (0.01%)	2.627	0.002 (0.08%)
	Change	0.192	0.193	0.001 (0.77%)	0.194	0.002 (0.90%)	0.197	0.005 (2.50%)
	Change (%)	7.89%	7.95%	0.06%	7.97%	0.08%	8.09%	0.21%
<b>Croatia</b>	2010	545.190	544.404	0.787 (0.14%)	544.130	1.060 (0.19%)	544.467	0.724 (0.13%)
	2020	569.574	568.787	0.786 (0.14%)	568.453	1.121 (0.20%)	568.753	0.821 (0.14%)
	Change	24.383	24.384	0.000 (0.00%)	24.323	0.061 (0.25%)	24.286	0.097 (0.40%)
	Change (%)	4.47%	4.48%	0.01%	4.47%	0.00%	4.46%	-0.01%
<b>Italy</b>	2010	1 326.481	1 325.528	0.953 (0.07%)	1 326.470	0.012 (0.00%)	1 326.270	0.211 (0.02%)
	2020	1 337.604	1 336.560	1.044 (0.08%)	1 337.453	0.151 (0.01%)	1 337.265	0.339 (0.03%)
	Change	11.123	11.032	0.091 (0.82%)	10.983	0.140 (1.26%)	10.996	0.128 (1.15%)
	Change (%)	0.84%	0.83%	0.01%	0.83%	0.01%	0.83%	-0.01%
<b>Montenegro</b>	2010	45.352	45.279	0.072 (0.16%)	45.147	0.205 (0.45%)	45.329	0.023 (0.05%)
	2020	53.741	53.658	0.083 (0.15%)	53.579	0.161 (0.30%)	53.645	0.095 (0.18%)
	Change	8.389	8.379	0.010 (0.12%)	8.432	0.043 (0.51%)	8.316	0.073 (0.87%)
	Change (%)	18.50%	18.50%	0.01%	18.68%	0.18%	18.35%	-0.15%
<b>Slovenia</b>	2010	33.522	33.500	0.022 (0.06%)	33.508	0.014 (0.04%)	33.532	0.010 (0.03%)
	2020	33.361	33.343	0.018 (0.05%)	33.395	0.034 (0.10%)	33.381	0.021 (0.06%)
	Change	-0.161	-0.158	0.004 (2.21%)	-0.113	0.048 (29.65%)	-0.150	0.011 (6.72%)
	Change (%)	-0.48%	-0.47%	0.01%	-0.34%	0.14%	-0.45%	0.03%

**Table 5.2** Friedman test statistics and corresponding p-values for three datasets (built-up for 2010, built-up for 2020, and built-up change), each containing six built-up area calculations (one for each country) obtained by four different methods.

	Dataset		
	Built-up for 2010	Built-up for 2020	Built-up change
<b>Test statistic</b>	7.000	2.600	0.000
<b>p-value</b>	0.072	0.457	1.000

Critical test statistic value = 7.600 (for 4 groups of measurements with 6 observations in each group, under  $\alpha = 0.05$ )

**Table 5.3** Wilcoxon test statistics and corresponding p-values for pairwise comparison between WGS84-based method and other three applied methods. Pairwise comparisons were performed separately within each dataset.

		Dataset		
		Built-up 2010	Built-up 2020	Built-up change
<b>WGS84 vs. LAEA</b>	<b>Test statistic</b>	0.000	1.000	6.000
	<b>p-value</b>	0.031	0.063	0.438
<b>WGS84 vs. rHEALPix</b>	<b>Test statistic</b>	4.000	5.000	6.000
	<b>p-value</b>	0.219	0.313	0.438
<b>WGS84 vs. QPix</b>	<b>Test statistic</b>	2.000	3.000	6.000
	<b>p-value</b>	0.094	0.156	0.438

$\alpha = 0.05$ , after Bonferroni reduction  $\alpha_B = 0.017$

### 5.3 Discussion and conclusions

The aim of this chapter was to demonstrate application of DGGS on a task that is common when working with land cover data—calculating area of a specific land cover class for two years and then determining magnitude of change of that specific land cover class. These calculations were performed directly from the original data (WGS84-based method) which might, in general, be considered as traditionally not preferred one. On the other hand, the second method that included reprojecting raster data from the geographic to the LAEA-based projected CRS, would probably be the first choice by most of the GIS professionals, without considering possible effects of such approach on obtained results. The last two methods (rHEALPix- and QPix-based methods) are to a certain degree similar to the LAEA-based method. The main difference is that instead of the planar raster model, the last two methods use grid from a specific DGGS refinement level.



Regardless of the calculation method used, results show there are no statistically significant differences between obtained results. To a certain degree, this was expected. Although regridding results in various interrelated issues, such as, data loss or replication, spatial shifts, and change in pixel/cell sizes, if regridding is performed carefully, then on a large dataset these issues tend to have minimal effect on results, i.e., they tend to get obscured by the volume of data. Under carefully performed regridding, here we consider choosing a suitable pixel/cell size in a target grid and ensuring that target grid will not get too distorted in comparison to the source grid within the geographical scope of analysis. The fact that no statistically significant differences between methods were observed in this case confirms that regridding parameters were chosen appropriately. When performing small-scale analysis, there are many projected CRS-based options to choose from, but when analysis becomes large- or global-scale, then various benefits of DGGS comes into play (Sections 3.1 and 4.3).

Here we are not interested in the interpretation and implications of the obtained built-up areas and their changes from the environmental perspective, thus we omit them. We also do not find meaningful extensive examination of the data in Table 5.1 since the four analyzed methods clearly do not produce significantly different results. However, we would like to point out two interesting details. First is related to the high percentage in change difference for Slovenia based on rHEALPix-based method. This percentage of 25% is higher than in any other instance mainly because WGS84-based change of  $-0.16 \text{ km}^2$  that was used as a baseline for calculating percentage of 25% is very small. By “very small” we mean that area of  $0.16 \text{ km}^2$  is not sufficiently larger than the average cell/pixel area and thus regridding-related issues get exaggerated. This should be kept in mind when analyzing results. The second one is related to the differences between built-up area changes between methods. They are almost without an exception lower than differences in built-up areas for each year. This means that specific method almost always either overestimates or underestimates areas for each year and when change between years is calculated, these biases tend to cancel each other out.

It is interesting to notice that after regridding original GLC\_FCS30D data to the rHEALPix and QPix DGGSs, we decided to store land cover data as points in GeoParquet files. Of course, instead of points, we could use polygons that represent DGGS cell edges. At first sight, it might seem that we stored DGGS-based data by using a vector data model. However, this is only partly the case. We had to store DGGS cells either as polygons or as their representative points because we had to

perform calculations that included aggregation on the 10-km coastal zone that is provided as a polygonal vector. We decided to use points mainly because of the more efficient calculation workflow; counting points in polygons is computationally efficient than, for example, intersecting many polygonal cells with 10-km polygonal coastal zone. Geometry, whether in form of points, lines, or polygons, is essential from the perspective of a vector data model. Geometry is required because vector data model is based on coordinates for storing location of, firstly points, and then lines and polygons. From the DGGS perspective, reference to cell geometry in that sense is unnecessary. Location in DGGS is not defined by coordinates but by unique cell indices, without any mention to traditional geometric primitives. For example, the `raster2dgs` Python package (Ardo and Law, 2024), that can be used to convert data that are georeferenced by using raster data model to DGGS data model, outputs results in a Parquet files (not GeoParquet) without a reference to cell geometry. One of the columns in this Parquet file is reserved for cell indices. If required, and if some specific DGGS implementation supports it, cell geometry, either polygonal or as a cell's reference points, can be generated from cell index at any time. This means that if 10-km coastal zone was georeferenced by using DGGS data model (based on `rHEALPix` or `QPix`), analysis could have been performed without reference to points or polygons. Some of the approaches of transferring points, lines, and polygons to DGGS are suggested by, for example, Hojati et al. (2022) and Kmoch, Matsibora, et al. (2022). In our case, the same `rHEALPix` or `QPix` grid to which `GLC_FCS30D` data were regridded, could be used for georeferencing the 10-km coastal zone for each country. If, for example, a reference point of a particular cell is within the polygon that represents 10-km coastal zone, for that cell in the Parquet file a categorical attribute with country name can be saved. Now in the Parquet file we would have one column with cell indices, one column for 2010 and one for 2020 with binary attribute representing whether that cell corresponds to the pixel that was in `GLC_FCS30D` classified as built-up land cover class for 2010 or for 2020, respectively, and one column that contains country name if that cell's reference point is within that country's 10-km coastal zone. From data in this table, and without a direct reference to cell geometry, it is now possible to efficiently aggregate data on a country level and perform calculations of built-up area change. Robertson et al. (2020) suggested a hybrid relational/key-value database model for storing data that are georeferenced in DGGS. They have also shown that common GIS operations such as constructing buffer zones, or interesting vector polygons, are more efficient when performed on their model than when performed according to

standard procedures on a vector data model. Martin-Segura et al. (2024) proposed a method for storing rHEALPix DGGS data in a GeoTIFF raster data format. This GeoTIFF raster is georeferenced in a  $(n, s)$ -rHEALPix map projection plane, which means that if it is aligned with the rHEALPix grid, each square pixel corresponds to one DGGS square cell. The benefit of this approach is that output GeoTIFF file can be directly used with advanced algorithms for raster data processing, but with a drawback that GeoTIFF format does not allow storing some more complex attribute data types that, for example, Parquet format support.

From this discussion, it is evident that DGGS is a distinct geospatial data model; it is different from vector and raster models, but it also has some characteristics of both. It is similar to a raster model because they both represent spatial data using a grid structure and it is similar to the vector model because each cell can be viewed as a vector polygon with assigned various attributes. It is also different from the vector and raster models as it can additionally be viewed as a georeferencing framework in which cell identifiers are used for positioning instead of coordinates. This means that each cell can be treated as a placeholder for storing spatial data. Various categories of spatial data (e.g., land cover data, administrative boundaries, land surface temperature data, point-object (for example, communication antennas) locations, etc.) can all be stored within a single file (or table in a database) without the need to separate them into different layers. Therefore, in a sense, we can transition from layer-oriented way of thinking to a cell-oriented. With DGGS, we can also transition from viewing GIS as a collection of flat digital maps to viewing it as a digital globe (Goodchild, 2018), where each cell contains various thematic data.

## 6 Conclusion and scientific contribution

In this thesis, we first gave an overview of various approaches to generating global tessellations or grids, with emphasis on approaches that are currently widely used. We then identified the main challenges when it comes to handling geospatial, namely land cover data, by using these conventional approaches. We further introduced a discrete global grid system (DGGS) as a georeferencing framework and geospatial data model that has a potential of eliminating or at least reducing some of these challenges. Subsequently, we narrowed our focus on one particular DGGS implementation, rHEALPix DGGS, which we found suitable for global land cover data storage and handling.

We introduced two modifications in the rHEALPix DGGS that we considered important. The first modification is aimed at increasing calculation accuracy and reducing processing time when converting geodetic to authalic latitude and vice versa. This conversion is crucial for constructing the rHEALPix grid on the ellipsoids of revolution. The second modification was related to the enhancement of the rHEALPix DGGS cell shape stability. This modification results in grid structures on the sphere or ellipsoid that are different from the rHEALPix DGGS grids and thus we refer to it as QPix DGGS.

The results indicate that both modifications enhanced the rHEALPix DGGS, at least from our perspective. Furthermore, we also quantitatively compared QPix DGGS and currently widely used, traditional global gridding approaches and found that QPix DGGS enables global gridding that exhibits higher uniformity than traditional approaches. Finally, we demonstrated the application of the rHEALPix and QPix DGGSs on calculating land cover change between two years.

Regarding the research hypothesis that were stated in Section 1.2, we state the following. The first hypothesis was: “The discrete global grid system can be used to store and handle global land cover data while eliminating the shortcomings of the traditionally used two-dimensional raster geospatial data model.” This hypothesis is confirmed. The fact that DGGS can be used to store and handle global land cover data is shown in Chapter 5. Although in this chapter analysis was not performed on a full global scale, calculations that were performed on a global level in some other parts of the thesis, for example calculations of cell compactness values (Sections 3.4 and 3.5.2), indicates that

extending analysis to the global level will not bring any differences, except for, of course, a drastically higher demand for computational resources. Regarding the second part of this hypothesis—addressing the limitations of the traditionally used two-dimensional raster geospatial data model, it was qualitatively discussed in Section 3.1 and to a certain level in Sections 4.3 and 5.3. Section 4.3 additionally provides a quantitative comparison between DGGS-based (specifically, QPix DGGS) and two traditional global gridding approaches that are based on flat raster data models. One is based on a raster model that is georeferenced in a geographic 2D CRS and thus defines an equiangular quadrilateral global tessellation, and another that is based on a raster model in projected CRS that is based on the interrupted Goode Homolosine map projection. The results showed that DGGS-based global gridding approach provides a more uniform global tessellation. Therefore, DGGS not only solves some of the problems that are related to handling geospatial (land cover) data in the form of a 2D raster data model, but it also provides a framework for more uniform global gridding, i.e., more uniform global sampling. This is mainly because DGGS georeferences data directly on the sphere or ellipsoid and thus enables storing and handling geospatial data in the form of digital globes, rather than in the form of digital maps.

The second hypothesis stated: “It is possible to develop a DGGS in which cells are quadrangles of constant area and whose shapes are more stable than that in the currently developed systems.” This hypothesis is also confirmed. We imposed a requirement of quadrangular, constant area cells because we found global grid with such characteristics suitable for storage and handling land cover data. Briefly, constant, or equal area cells, simplify areal statistics calculations and ensure that all cells have the same probability of contributing the analysis. DGGS grids that have quadrangular cells are interesting for two main reasons—first, remote sensing satellite sensor detectors in general have quadrangular shape and second, already developed advanced algorithms for processing raster data can be directly or with slight modifications applied on these grids. Starting from the research by Knoch, Vasilyev et al. (2022), we identified rHEALPix DGGS as the only DGGS implementation that is, according to our requirements, suitable for global land cover data storage and handling (i.e., almost all of its cells are quadrangles of equal area). DGGS that we refer to as QPix DGGS is in fact a modified version of the rHEALPix DGGS. In QPix DGGS not just most, but all cells are quadrangles on sphere or ellipsoid and they all have, at least analytically, the same area (Section 3.3.2). Based on results

in Section 3.5.2 and discussion in Section 4.2.1, it was confirmed that QPix DGGS exhibits a more stable cell shapes than rHEALPix DGGS. This means that QPix DGGS meets the criteria of the second hypothesis.

The main scientific contribution of this doctoral thesis is the development of the DGGS as a geospatial data model and georeferencing framework that is suitable for global land cover data storage and handling. This scientific contribution can be further divided into three distinct components:

1. Modification of the mathematical basis of the existing DGGS implementation (specifically, rHEALPix DGGS) and its improvement by increasing the accuracy and speed of the area-preserving mapping of an ellipsoid to a sphere and vice versa. (Sections 3.3.1, 3.5.1, and 4.1)
2. Development of a DGGS that has cells that are quadrangles of equal area and whose cell-shape stability is overall higher than in currently available DGGS implementations. (Sections 3.3.23.5.2, and 4.2)
3. Improvement of a currently common approach in global land cover data storage and handling by transition from the two-dimensional raster data model to the DGGS. (Sections 3.1 and 4.3 and Chapter 5, especially Section 5.3)

## References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(1), 170191. <https://doi.org/10.1038/sdata.2017.191>
- Abernathey, R. P., Augspurger, T., Banihirwe, A., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., Robinson, N. H., and Signell, R. P. (2021). Cloud-Native Repositories for Big Scientific Data. *Computing in Science & Engineering*, 23(2), 26–35. *Computing in Science & Engineering*. <https://doi.org/10.1109/MCSE.2021.3059437>
- Alderson, T., Purss, M., Du, X., Mahdavi-Amiri, A., and Samavati, F. (2020). Digital Earth Platforms. In H. Guo, M. F. Goodchild, and A. Annoni (Eds.), *Manual of Digital Earth* (pp. 25–54). Springer. [https://doi.org/10.1007/978-981-32-9915-3\\_2](https://doi.org/10.1007/978-981-32-9915-3_2)
- Amiri, A. M., Samavati, F., and Peterson, P. (2015). Categorization and Conversions for Indexing Methods of Discrete Global Grid Systems. *ISPRS International Journal of Geo-Information*, 4(1), Article 1. <https://doi.org/10.3390/ijgi4010320>
- Ardo, J., and Law, R. (2024). *raster2dggs* (Version 0.3.0) [Computer software]. Retrieved June 11, 2025, from <https://github.com/manaakiwhenua/raster2dggs>
- Arthur, R. (2023). A critical analysis of the What3Words geocoding algorithm. *PLOS ONE*, 18(10), e0292491. <https://doi.org/10.1371/journal.pone.0292491>
- Basaraner, M., and Cetinkaya, S. (2019). New measures for analysis and comparison of shape distortion in world map projections. *Cartography and Geographic Information Science*, 46(6), 518–531. <https://doi.org/10.1080/15230406.2019.1567394>
- Bauer-Marschallinger, B., and Falkner, K. (2023). Wasting petabytes: A survey of the Sentinel-2 UTM tiling grid and its spatial overhead. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202, 682–690. <https://doi.org/10.1016/j.isprsjprs.2023.07.015>

- Bauer-Marschallinger, B., Sabel, D., and Wagner, W. (2014). Optimisation of Global Grids for High-Resolution Remote Sensing Data. *Computers & Geosciences*, 72. <https://doi.org/10.1016/j.cageo.2014.07.005>
- Béjar, R., Lacasta, J., Lopez-Pellicer, F. J., and Nogueras-Iso, J. (2023). Discrete Global Grid Systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes. *Environmental Modelling & Software*, 162, 105656. <https://doi.org/10.1016/j.envsoft.2023.105656>
- Berk, S., and Ferlan, M. (2018). Accurate area determination in the cadaster: case study of Slovenia. *Cartography and Geographic Information Science*, 45(1), 1–17. <https://doi.org/10.1080/15230406.2016.1217789>
- Bian, S.-F., Li, H.-P., Bian, S.-F., and Li, H.-P. (2012). Mathematical Analysis in Cartography by Means of Computer Algebra System. In *Cartography - A Tool for Spatial Analysis*. IntechOpen. <https://doi.org/10.5772/50159>
- Birch, C. P. D., Oom, S. P., and Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206(3), 347–359. <https://doi.org/10.1016/j.ecolmodel.2007.03.041>
- Brenner, J. L., Reitwiesner, G. W., Clenshaw, C. W., and Good, I. J. (1955). Technical Notes and Short Papers. *Mathematical Tables and Other Aids to Computation*, 9(51), 117–121. <https://doi.org/10.2307/2002068>
- Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. *ISPRS International Journal of Geo-Information*, 1(1), Article 1. <https://doi.org/10.3390/ijgi1010032>
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., and Tait, A. M. (2021). *Dynamic World Test Tiles* (Version v2) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.4766508>



- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., and Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1), Article 1. <https://doi.org/10.1038/s41597-022-01307-4>
- Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S. (2020). *Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe* (Version V3.0.1) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.3939050>
- Bunting, P., Rosenqvist, A., Hilarides, L., Lucas, R. M., Thomas, N., Tadono, T., Worthington, T. A., Spalding, M., Murray, N. J., and Rebelo, L.-M. (2022). Global Mangrove Extent Change 1996–2020: Global Mangrove Watch Version 3.0. *Remote Sensing*, 14(15), Article 15. <https://doi.org/10.3390/rs14153657>
- Calabretta, M. R., and Roukema, B. F. (2007). Mapping on the HEALPix grid. *Monthly Notices of the Royal Astronomical Society*, 381(2), 865–872. <https://doi.org/10.1111/j.1365-2966.2007.12297.x>
- Carr, D., Kahn, R., Sahr, K., and Olsen, T. (1997). ISEA discrete global grids. *Statistical Computing and Statistical Graphics Newsletter*, 8(2/3), 31–39.
- Chan, F. K., and O'Neill, E. M. (1975). *Feasibility Study of a Quadrilateralized Spherical Cube Earth Data Base*. (No. ADA010232). Computer Sciences Corp., Silver Spring, Md. System Sciences Div; Environmental Prediction Research Facility (Navy), Monterey, Calif. Retrieved April 6, 2024, from <https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/ADA010232.xhtml>
- Chrisman, N. R. (2017). Calculating on a round planet. *International Journal of Geographical Information Science*, 31(4), 637–657. <https://doi.org/10.1080/13658816.2016.1215466>
- Corder, G. W. (2014). *Nonparametric Statistics: a Step-by-Step Approach*. Wiley.
- Cracknell, A. P. (1998). Review article Synergy in remote sensing-what's in a pixel? *International Journal of Remote Sensing*, 19(11), 2025–2047. <https://doi.org/10.1080/014311698214848>

- Cuypers, S., Nascetti, A., and Vergauwen, M. (2023). Land Use and Land Cover Mapping with VHR and Multi-Temporal Sentinel-2 Imagery. *Remote Sensing*, 15(10), Article 10. <https://doi.org/10.3390/rs15102501>
- d'Andrimont, R., Yordanov, M., Sedano, F., Verhegghen, A., Strobl, P., Zachariadis, S., Camilleri, F., Palmieri, A., Eiselt, B., Rubio Iglesias, J. M., and van der Velde, M. (2024). Advances in LUCAS Copernicus 2022: enhancing Earth observations with comprehensive in situ data on EU land cover and use. *Earth System Science Data*, 16(12), 5723–5735. <https://doi.org/10.5194/essd-16-5723-2024>
- Defence Geospatial Information Working Group. (2020). *Defence Gridded Elevation Data Product Implementation Profile* (No. DGIWG 250; Version 1.2.1). Retrieved May 19, 2025, from <https://portal.dgiwg.org/files/71215>
- Deutsch, L. P. (1996). *DEFLATE Compressed Data Format Specification version 1.3* (Request for Comments No. RFC 1951). Internet Engineering Task Force. <https://doi.org/10.17487/RFC1951>
- Dheemanth, H. N. (2014). LZW Data Compression. *American Journal of Engineering Research*, 3(2), 22–26.
- Di Gregorio, A., and Jansen, L. (1998). *Land Cover Classification System (LCCS): Classification Concepts and User Manual*. FAO. Retrieved December 19, 2022, from [https://www.researchgate.net/publication/229839605\\_Land\\_Cover\\_Classification\\_System\\_LCCS\\_Classification\\_Concepts\\_and\\_User\\_Manual](https://www.researchgate.net/publication/229839605_Land_Cover_Classification_System_LCCS_Classification_Concepts_and_User_Manual)
- Di Gregorio, A., and Leonardi, U. (2016). *Land Cover Classification System: User Manual, software version 3*. Food and Agriculture Organisation of the United Nations. Retrieved February 27, 2025, from <https://openknowledge.fao.org/server/api/core/bitstreams/1a16e147-892a-47da-89f3-2dddf6c3c703/content>
- Dimitrijević, A., Milosavljević, A., and Rančić, D. (2023). Efficient Distortion Mitigation and Partition Reduction in Mapping Global Geodata: Dual Orthogonal Equidistant Cylindrical Projection Approach. *ISPRS International Journal of Geo-Information*, 12(7). <https://doi.org/10.3390/ijgi12070289>

- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., ... Saleem, M. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), 534–545. <https://doi.org/10.1093/biosci/bix014>
- Dutton, G. (1989). *Planetary Modelling via Hierarchical Tessellation*. 426–471. Retrieved May 21, 2025, from <https://cartogis.org/docs/proceedings/archive/auto-carto-9/pdf/planetary-modeling-via-hierarchical-tessellation.pdf>
- Dutton, G. (2016). The making of a Global Grid - Remembering my escape from flatland. *IOP Conference Series: Earth and Environmental Science*, 34, 012007. <https://doi.org/10.1088/1755-1315/34/1/012007>
- Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., and Lymburner, L. (2018). Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sensing*, 10(9), Article 9. <https://doi.org/10.3390/rs10091363>
- EPSG. (n.d.-a). *EPSG Geodetic Parameter Dataset record for ellipsoid WGS 84*. Retrieved January 30, 2025, from [https://epsg.org/ellipsoid\\_7030/WGS-84.html](https://epsg.org/ellipsoid_7030/WGS-84.html)
- EPSG. (n.d.-b). *EPSG Geodetic Parameter Dataset record for the ETRS89-extended / LAEA Europe projected CRS*. Retrieved June 17, 2025, from [https://epsg.org/crs\\_3035/ETRS89-extended-LAEA-Europe.html](https://epsg.org/crs_3035/ETRS89-extended-LAEA-Europe.html)
- EPSG. (n.d.-c). *EPSG Geodetic Parameter Dataset record for WGS 84 geographic 2D CRS*. Retrieved June 7, 2025, from [https://epsg.org/crs\\_4326/WGS-84.html](https://epsg.org/crs_4326/WGS-84.html)
- European Space Agency and Airbus. (2022). *Copernicus DEM* [Dataset]. <https://doi.org/10.5270/ESA-c5d3d65>
- Frančula, N. (2000). *Kartografske projekcije*. Sveučilište u Zagrebu Geodetski Fakultet. Retrieved May 27, 2025, from [https://www.researchgate.net/publication/279885540\\_Kartografske\\_projekcije](https://www.researchgate.net/publication/279885540_Kartografske_projekcije)

- Frančula, N., Lapaine, M., Župan, R., Kljajić, I., Poslončec-Petrić, V., Vinković, A., and Cibilić, I. (2021). Determining Areas from Maps. *Geodetski List*, 75(4), 365–379. Scopus. Retrieved July 22, 2022, from <https://hrcak.srce.hr/clanak/394586>
- García-Álvarez, D., and Nanu, S. F. (2022). Land Use Cover Datasets: A Review. In D. García-Álvarez, M. T. Camacho Olmedo, M. Paegelow, and J. F. Mas (Eds.), *Land Use Cover Datasets and Validation Tools: Validation Practices with QGIS* (pp. 47–66). Springer International Publishing. [https://doi.org/10.1007/978-3-030-90998-7\\_4](https://doi.org/10.1007/978-3-030-90998-7_4)
- Gascon, F., Cadau, E., Colin, O., Hoersch, B., Isola, C., Fernández, B., and Martimort, P. (2014). Copernicus Sentinel-2 Mission: Products, Algorithms and Cal/Val. In *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 9218). <https://doi.org/10.1117/12.2062260>
- Gibb, R., Bell, J., Cheng, C., Peterson, P., Purss, M., Ren, M., Samavati, F., and Strobl, P. (2021). *Topic 21 - Discrete Global Grid Systems - Part 1 Core Reference system and Operations and Equal Area Earth Reference System*. Retrieved March 21, 2023, from <http://www.opengis.net/doc/AS/dggs/2.0>
- Gibb, R. G. (2016). The rHEALPix Discrete Global Grid System. *IOP Conference Series: Earth and Environmental Science*, 34(1), 012012. <https://doi.org/10.1088/1755-1315/34/1/012012>
- Gibb, R., Raichev, A., and Speth, M. (2013). *The rHEALPix Discrete Global Grid System*. <https://doi.org/10.7931/J2D21VHM>
- Gilić, F., Baučić, M., and Samah, T. (2024). Global Land Cover Mapping – Need for Discrete Global Grid System. *International Conference of Environmental Remote Sensing and GIS, 2024*, 167–170. <https://doi.org/10.5281/zenodo.11621507>
- Gilić, F., and Gašparović, M. (in press). *Calculating Land Cover Change: DGGS Perspective*. Mediterranean Geosciences Union 4th Annual Meeting MedGU-24, Barcelona, Spain.
- Gillissen, I. (1993). AREA COMPUTATION OF A POLYGON ON AN ELLIPSOID. *Survey Review*. <https://doi.org/10.1179/sre.1993.32.248.92>

- Goodchild, M. F. (2018). Reimagining the history of GIS. *Annals of GIS*, 24(1), 1–8. <https://doi.org/10.1080/19475683.2018.1424737>
- Goodchild, M. F. (2019). Preface. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 54(1), 1–3. <https://doi.org/10.3138/cart.54.1.preface>
- Goode, J. P. (1925). The Homolosine Projection: A New Device for Portraying the Earth's Surface Entire. *Annals of the Association of American Geographers*, 15(3), 119–125. <https://doi.org/10.2307/2560812>
- Google. (n.d.). *S2 Geometry*. S2Geometry. Retrieved October 7, 2024, from <http://s2geometry.io/>
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622(2), 759. <https://doi.org/10.1086/427976>
- Gorski, K. M., Hivon, E., and Wandelt, B. D. (1998). *Analysis Issues for Large CMB Data Sets* (No. arXiv:astro-ph/9812350). arXiv. <https://doi.org/10.48550/arXiv.astro-ph/9812350>
- GRASS Development Team. (2024). *Geographic Resources Analysis Support System (GRASS) Software* (Version 8.2) [Computer software]. Open Source Geospatial Foundation. <https://doi.org/10.5281/zenodo.5176030>
- Gray, R. W. (1994). Fuller's Dymaxion™ Map. *Cartography and Geographic Information Systems*, 21(4), 243–246. <https://doi.org/10.1559/152304094782540628>
- Guo, H., Goodchild, M. F., and Annoni, A. (Eds.). (2020). *Manual of Digital Earth*. Springer. <https://doi.org/10.1007/978-981-32-9915-3>
- Hall, J., Wecker, L., Ulmer, B., and Samavati, F. (2020). Disdyakis Triacontahedron DGGS. *ISPRS International Journal of Geo-Information*, 9(5), Article 5. <https://doi.org/10.3390/ijgi9050315>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array

- programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Harrison, E., Mahdavi-Amiri, A., and Samavati, F. (2011). Optimization of Inverse Snyder Polyhedral Projection. *2011 International Conference on Cyberworlds*, 136–143. <https://doi.org/10.1109/CW.2011.36>
- Hojati, M., Robertson, C., Roberts, S., and Chaudhuri, C. (2022). GIScience research challenges for realizing discrete global grid systems as a Digital Earth. *Big Earth Data*, 6, 358–379. <https://doi.org/10.1080/20964471.2021.2012912>
- Huang, X., Dai, J., Ben, J., Zhou, J., and Ding, J. (2024). Bidirectional mapping between rhombic triacontahedron and icosahedral hexagonal discrete global grid systems. *International Journal of Digital Earth*, 17. <https://doi.org/10.1080/17538947.2024.2324952>
- INSPIRE Temporary MIWP 2021-2024 sub-group 2.3.1. (2023). *D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines*. INSPIRE Maintenance and Implementation Group. Retrieved May 19, 2025, from <https://github.com/INSPIRE-MIF/technical-guidelines/tree/main/data/gg>
- Joseph, G. (2020). How to Specify an Electro-optical Earth Observation Camera? A Review of the Terminologies Used and its Interpretation. *Journal of the Indian Society of Remote Sensing*, 48(2), 171–180. <https://doi.org/10.1007/s12524-020-01105-8>
- Karney, C. F. F. (2013). Algorithms for geodesics. *Journal of Geodesy*, 87(1), 43–55. <https://doi.org/10.1007/s00190-012-0578-z>
- Karney, C. F. F. (2023). Geodesics on an arbitrary ellipsoid of revolution. *Journal of Geodesy*, 98(1), 4. <https://doi.org/10.1007/s00190-023-01813-2>
- Karney, C. F. F. (2024). On auxiliary latitudes. *Survey Review*, 56(395), 165–180. <https://doi.org/10.1080/00396265.2023.2217604>
- Karney, C. F. F., pwm1234-sri, and Ferrara, J. (2022). *GeographicLib (Python implementation)* (Version 2.0) [Computer software]. Retrieved April 24, 2025, from <https://github.com/geographiclib/geographiclib-python>

- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P. (2021). Global land use / land cover with Sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4704–4707. <https://doi.org/10.1109/IGARSS47720.2021.9553499>
- Kelly, K., and Šavrič, B. (2021). Area and volume computation of longitude–latitude grids and three-dimensional meshes. *Transactions in GIS*, 25(1), 6–24. <https://doi.org/10.1111/tgis.12636>
- Kerkovits, K. A. (2023). *Map projections: Lecture notes for students in Cartography and Geoinformatics MSc programme*. Eötvös Loránd University, Faculty of Informatics.
- Kimerling, A. J. (2002). Predicting Data Loss and Duplication when Resampling from Equal-Angle Grids. *Cartography and Geographic Information Science*, 29(2), 111–126. <https://doi.org/10.1559/152304002782053297>
- Kimerling, J. A., Sahr, K., White, D., and Song, L. (1999). Comparing Geometrical Properties of Global Grids. *Cartography and Geographic Information Science*, 26(4), 271–288. <https://doi.org/10.1559/152304099782294186>
- Kmoch, A., Matsibora, O., Vasilyev, I., and Uemaa, E. (2022). Applied open-source Discrete Global Grid Systems. *AGILE: GIScience Series*, 3, 1–6. <https://doi.org/10.5194/agile-giss-3-41-2022>
- Kmoch, A., and Uemaa, E. (2024). *Landscape Geoinformatics DGGS comparison materials (v4)* (Version v4.0.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.11125815>
- Kmoch, A., Vasilyev, I., Virro, H., and Uemaa, E. (2022). Area and shape distortions in open-source discrete global grid systems. *Big Earth Data*, 6(3), 256–275. <https://doi.org/10.1080/20964471.2022.2094926>
- Kumar, D. (2021). Urban objects detection from C-band synthetic aperture radar (SAR) satellite images through simulating filter properties. *Scientific Reports*, 11(1), 6241. <https://doi.org/10.1038/s41598-021-85121-9>
- Lapaine, M. (1994). Osnovni geodetski zadaci uzduž meridijana i uzduž paralele na rotacijskom elipsoidu. *Geodetski list*, 48 (71)(1), 43–52. Retrieved June 17, 2025, from <https://hrcak.srce.hr/292334>

- Lapaine, M., and Frančula, N. (2019). Polar and Equatorial Aspects of Map Projections? *Proceedings of the ICA*, 2, 1–6. <https://doi.org/10.5194/ica-proc-2-71-2019>
- Lapaine, M., and Lapaine, M. (1991). Površina elipsoidnog trapeza. *Geodetski list*, 45 (68)(10–12), 367–373. <https://www.bib.irb.hr/1125596>
- Li, M., and Stefanakis, E. (2020). Geospatial Operations of Discrete Global Grid Systems—a Comparison with Traditional GIS. *Journal of Geovisualization and Spatial Analysis*, 4(2), 26. <https://doi.org/10.1007/s41651-020-00066-3>
- Li, W., Goodchild, M. F., and Church, R. (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, 27(6), 1227–1250. <https://doi.org/10.1080/13658816.2012.752093>
- Li, X., Li, H., Liu, G., Bian, S., and Jiao, C. (2022). Simplified Expansions of Common Latitudes with Geodetic Latitude and Geocentric Latitude as Variables. *Applied Sciences*, 12(15). <https://doi.org/10.3390/app12157818>
- Liu, L., Zhang, X., and Li, Z. (2025). *GLC\_FCS30D: the first global 30-m land-cover dynamic monitoring product with fine classification system from 1985 to 2022* [Dataset]. Zenodo. Retrieved June 8, 2025, from <https://zenodo.org/records/15063683>
- Lu, M., Appel, M., and Pebesma, E. (2018). Multidimensional arrays for analysing geoscientific data. *ISPRS International Journal of Geo-Information*, 7(8). Scopus. <https://doi.org/10.3390/ijgi7080313>
- Lukatela, H. (1987). *Hipparchus geopositioning model: an overview*. 87–96. Retrieved May 22, 2025, from <https://cartogis.org/docs/proceedings/archive/auto-carto-8/pdf/hipparchus-geopositioning-model-an-overview.pdf>
- Lukatela, H. (2000, March). *Ellipsoidal Area Computations of Large Terrestrial Objects*. International Conference on Discrete Global Grids, Santa Barbara, USA. Retrieved April 18, 2025, from <https://geodyssey.github.io/papers/ggelare.html>



- Lukatela, H. (2002). A Seamless Global Terrain Model in the Hipparchus System. In M. F. Goodchild and A. J. Kimerling (Eds.), *Discrete Global Grids: A Web Book* (pp. 2–44). UC Santa Barbara. Retrieved May 21, 2025, from <https://escholarship.org/uc/item/9492q6sm>
- Luo, F., Shupeng, G., Xinpeng, W., Aimei, C., Zheng, W., and Li, Y. (2023). Construction of quality evaluation indicator system for diamond discrete global grid systems. *International Journal of Digital Earth*, 16(1), 3637–3660. <https://doi.org/10.1080/17538947.2023.2218115>
- Mahdavi-Amiri, A., Alderson, T., and Samavati, F. (2015a). A Survey of Digital Earth. *Computers & Graphics*, 53. <https://doi.org/10.1016/j.cag.2015.08.005>
- Mahdavi-Amiri, A., Alderson, T., and Samavati, F. (2015b). *A Survey of Digital Earth Representation and Visualization*. <https://doi.org/10.13140/RG.2.1.3877.9049>
- Markham, B. L., Jenstrom, D., Sauer, B., Pszcolka, S., Dulski, V., Hair, J., McCorkel, J., Kvaran, G., Thome, K., Montanaro, M., Pedelty, J., Anderson, C., Choate, M., Barsi, J., Kaita, E., and Miller, J. (2020). Landsat 9 mission update and status. *Earth Observing Systems XXV*, 11501, 96–102. <https://doi.org/10.1117/12.2569748>
- Martin-Segura, S., Béjar, R., and Zarazaga-Soria, F. J. (2024). DGGSTools: An Open Source Python Package for the Manipulation of Vector and Raster Datasets in the rHEALPix Discrete Global Grid System. *AGILE: GIScience Series*, 5, 1–6. <https://doi.org/10.5194/agile-giss-5-40-2024>
- Moon, B., Jagadish, H. V., Faloutsos, C., and Saltz, J. H. (2001). Analysis of the clustering properties of the Hilbert space-filling curve. *IEEE Transactions on Knowledge and Data Engineering*, 13(1), 124–141. <https://doi.org/10.1109/69.908985>
- Moreira de Sousa, L., Poggio, L., and Kempen, B. (2019). Comparison of FOSS4G Supported Equal-Area Projections Using Discrete Distortion Indicatrices. *ISPRS International Journal of Geo-Information*, 8(8), Article 8. <https://doi.org/10.3390/ijgi8080351>
- National Imagery and Mapping Agency. (2000). *Performance Specification: Digital Terrain Elevation Data* (No. MIL-PRF-89020B). Retrieved October 2, 2024, from [http://everyspec.com/MIL-PRF/MIL-PRF-080000-99999/MIL-PRF-89020B\\_25316/](http://everyspec.com/MIL-PRF/MIL-PRF-080000-99999/MIL-PRF-89020B_25316/)

- NGA Office of Geomatics. (2014). *Universal Grids and Grid Reference Systems* (Standardization Document No. NGA.STND.0037\_2.0.0\_GRIDS; Version 2.0.0).
- Nowak, E., and Nowak Da Costa, J. (2022). Theory, strict formula derivation and algorithm development for the computation of a geodesic polygon area. *Journal of Geodesy*, 96(4), 20. <https://doi.org/10.1007/s00190-022-01606-z>
- Osserman, R. (1978). The isoperimetric inequality. *Bulletin of the American Mathematical Society*, 84(6), 1182–1238. Retrieved April 24, 2025, from <https://projecteuclid.org/journals/bulletin-of-the-american-mathematical-society/volume-84/issue-6/The-isoperimetric-inequality/bams/1183541466.full>
- Overton, M. L. (2001). *Numerical Computing with IEEE Floating Point Arithmetic*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718072>
- Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. <https://doi.org/10.1038/nature20584>
- Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A., and Ying, Q. (2020). Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sensing*, 12(3), 426. <https://doi.org/10.3390/rs12030426>
- Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, X.-P., Baggett, A., Kommareddy, I., and Kommareddy, A. (2022). The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. *Frontiers in Remote Sensing*, 3, 856903. <https://doi.org/10.3389/frsen.2022.856903>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing* (Second Edition). Cambridge university press. <https://www.grad.hr/nastava/gs/prg/NumericalRecipesinC.pdf>
- PROJ contributors. (2025). *PROJ coordinate transformation software library* [Computer software]. Open Source Geospatial Foundation. <https://doi.org/10.5281/zenodo.5884394>

- Raichev, A., Gibb, R., Car, N., De Maio, N., Gilić, F., and Law, R. (2023, December). *rHEALPixDGGS*. Retrieved July 10, 2024, from <https://github.com/manaakiwhenua/rhealpixdggs-py>
- Rapp, R. H. (1991). *Geometric Geodesy Part I*. Ohio State University Department of Geodetic Science and Surveying. Retrieved March 26, 2025, from <https://kb.osu.edu/items/5cfd1585-a92a-5933-9469-366a5d6b6894>
- Robertson, C., Chaudhuri, C., Hojati, M., and Roberts, S. A. (2020). An integrated environmental analytics system (IDEAS) based on a DGGS. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 214–228. <https://doi.org/10.1016/j.isprsjprs.2020.02.009>
- Roşca, D., and Plonka, G. (2011). Uniform spherical grids via equal area projection from the cube to the sphere. *Journal of Computational and Applied Mathematics*, 236(6), 1033–1041. <https://doi.org/10.1016/j.cam.2011.07.009>
- Roukema, B. F., and Lew, B. (2004). *A Solution to the Isolatitute, Equi-area, Hierarchical Pixel-Coordinate System* (No. arXiv:astro-ph/0409533). arXiv. <https://doi.org/10.48550/arXiv.astro-ph/0409533>
- Roy, D. P., Li, J., Zhang, H. K., and Yan, L. (2016). Best practices for the reprojection and resampling of Sentinel-2 Multi Spectral Instrument Level 1C data. *Remote Sensing Letters*, 7(11), 1023–1032. <https://doi.org/10/gnmk87>
- Saff, E. B., and Kuijlaars, A. B. J. (1997). Distributing many points on a sphere. *The Mathematical Intelligencer*, 19, 5–11. <https://doi.org/10.1007/BF03024331>
- Sahr, K. (2011). Hexagonal Discrete Global Grid Systems for Geospatial Computing. *Archives of Photogrammetry, Cartography and Remote Sensing*, 22, 363–376. Retrieved January 15, 2025, from [https://www.researchgate.net/publication/266415994\\_Hexagonal\\_discrete\\_global\\_grid\\_systems\\_for\\_geospatial\\_computing](https://www.researchgate.net/publication/266415994_Hexagonal_discrete_global_grid_systems_for_geospatial_computing)
- Sahr, K., Dumas, M., and Choudhur, N. (2015). *The PlanetRisk Discrete Global Grid System*. Retrieved June 15, 2025, from <https://www.discreteglobalgrids.org/wp-content/uploads/2016/10/PlanetRiskDGGS.pdf>

- Sahr, K., White, D., and Kimerling, A. J. (2003). Geodesic Discrete Global Grid Systems. *Cartography and Geographic Information Science*, 30(2), 121–134.  
<https://doi.org/10.1559/152304003100011090>
- Seong, J. C. (2005). Implementation of an Equal-area Gridding Method for Global-scale Image Archiving. *Photogrammetric Engineering & Remote Sensing*, 71(5), 623–627.  
<https://doi.org/10.14358/PERS.71.5.623>
- Seong, J. C., Mulcahy, K., and Usery, E. L. (2002). The Sinusoidal Projection: A New Importance in Relation to Global Image Data. *The Professional Geographer*, 54(2), 218–225.  
<https://doi.org/10.1111/0033-0124.00327>
- Simoës, R., Hengl, T., and Crăciunescu, V. (2024). *OpenLandMap technical documentation and user manual*. Retrieved June 17, 2025, from <https://docs.openlandmap.org/>
- Snyder, J. P. (1987). Map projections: A working manual. In *Map projections: A working manual* (USGS Numbered Series No. 1395; Professional Paper, Vol. 1395). U.S. Government Printing Office. <https://doi.org/10.3133/pp1395>
- Snyder, J. P. (1992). An Equal-Area Map Projection For Polyhedral Globes. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 29(1), 10–21.  
<https://doi.org/10.3138/27H7-8K88-4882-1752>
- Song, L., Kimerling, A. J., and Sahr, K. (2002). Developing an Equal Area Global Grid by Small Circle Subdivision. In M. F. Goodchild and A. J. Kimerling (Eds.), *Discrete Global Grids: A Web Book* (pp. 2–44). UC Santa Barbara. Retrieved May 21, 2025, from <https://escholarship.org/uc/item/9492q6sm>
- Stefanakis, E. (2017). Web Mercator and Raster Tile Maps: Two Cornerstones of Online Map Service Providers. *Geomatica*, 71(2), 100–109. <https://doi.org/10.5623/cig2017-203>
- Steinwand, D. R. (1994). Mapping raster imagery to the Interrupted Goode Homolosine projection. *International Journal of Remote Sensing*, 15(17), 3463–3471.  
<https://doi.org/10.1080/01431169408954340>

- Steinwand, D. R., Hutchinson, J. A., and Snyder, J. P. (1995). Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. *Photogrammetric Engineering and Remote Sensing*, 61(12), 1487–1497. USGS Publications Warehouse. Retrieved May 10, 2024, from <http://pubs.er.usgs.gov/publication/70187049>
- Sundnes, J. (2020). *Introduction to Scientific Programming with Python* (Vol. 6). Springer International Publishing. <https://doi.org/10.1007/978-3-030-50356-7>
- Szalay, A. S., Gray, J., Fekete, G., Kunszt, P. Z., Kukol, P., and Thakar, A. (2007). *Indexing the Sphere with the Hierarchical Triangular Mesh* (No. arXiv:cs/0701164). arXiv. <https://doi.org/10.48550/arXiv.cs/0701164>
- Tait, A. M., Brumby, S. P., Hyde, S. B., Mazzariello, J., and Corcoran, M. (2021). *Dynamic World training dataset for global land use and land cover categorization of satellite imagery* [Dataset]. PANGAEA. <https://doi.org/10.1594/PANGAEA.933475>
- The mpmath development team. (2023). *mpmath: a Python library for arbitrary-precision floating-point arithmetic* (Version 1.3.0) [Computer software]. <http://mpmath.org/>
- The SciPy community. (2025). *SciPy v1.15.3 Manual – friedmanchisquare*. Retrieved June 19, 2025, from <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.friedmanchisquare.html#scipy.stats.friedmanchisquare>
- Thompson, J. A., Brodzik, M. J., Silverstein, K. A. T., Hurley, M. A., and Carlson, N. L. (2022). EASE-DGGS: a hybrid discrete global grid system for Earth sciences. *Big Earth Data*, 6(3), 340–357. <https://doi.org/10.1080/20964471.2021.2017539>
- Tobler, W. (1993). Global Spatial Analysis. In *Three Presentations On Geographical Analysis And Modeling*. Retrieved June 17, 2025, from <https://escholarship.org/uc/item/05r820mz>
- Tobler, W., and Chen, Z. (1986). A Quadtree for Global Information Storage. *Geographical Analysis*, 18(4), 360–371. <https://doi.org/10.1111/j.1538-4632.1986.tb00108.x>
- Tong, X., Ben, J., and Zhang, Y. (2005). Expression of spherical entities and generation of Voronoi diagram based on truncated icosahedron DGG. *ISPRS - International Archives of the*

- Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-2/W25. Retrieved May 27, 2025, from [https://www.isprs.org/proceedings/xxxvi/2-W25/source/THE\\_EXPRESSION\\_OF\\_SPHERICAL\\_ENTITIES\\_AND\\_GENERATING\\_OF\\_VORONOI\\_DIAGRAM\\_BASED\\_ON\\_TRUNCATED\\_ICOSAHE.pdf](https://www.isprs.org/proceedings/xxxvi/2-W25/source/THE_EXPRESSION_OF_SPHERICAL_ENTITIES_AND_GENERATING_OF_VORONOI_DIAGRAM_BASED_ON_TRUNCATED_ICOSAHE.pdf)
- Tsendbazar, N., Li, L., Koopman, M., Carter, S., Herold, M., Georgieva, I., and Lesiv, M. (2021). *ESA WorldCover Product Validation Report V1.1*. Retrieved November 22, 2023, from [https://esa-worldcover.s3.eu-central-1.amazonaws.com/v100/2020/docs/WorldCover\\_PVR\\_V1.1.pdf](https://esa-worldcover.s3.eu-central-1.amazonaws.com/v100/2020/docs/WorldCover_PVR_V1.1.pdf)
- Tsendbazar, N., Xu, P., Herold, M., Lesiv, M., and Duerauer, M. (2022). *ESA WorldCover Product Validation Report V2.0*. Retrieved November 22, 2023, from [https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover\\_PVR\\_V2.0.pdf](https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover_PVR_V2.0.pdf)
- Uber. (n.d.). *H3*. Retrieved October 7, 2024, from <https://h3geo.org/>
- UN-GGIM. (2019). *The Global Fundamental Geospatial Data Themes*. United Nations. Retrieved February 25, 2025, from [https://ggim.un.org/documents/Fundamental\\_Data\\_Publication\\_08July2024.pdf](https://ggim.un.org/documents/Fundamental_Data_Publication_08July2024.pdf)
- Usery, E. L., Seong, J. C., and Steinwand, D. (2002). Methods to achieve accurate projection of regional and global raster databases. In *Methods to achieve accurate projection of regional and global raster databases* (USGS Numbered Series Nos. 2001–181; Open-File Report, Vols. 2001–181). U.S. Geological Survey. <https://doi.org/10.3133/ofr01181>
- Van De Kerchove, R., Zanaga, D., De Keersmaecker, W., Linlin, L., Tsendbazar, N., and Lesiv, M. (2020). *ESA WorldCover Product User Manual V1.0*. Retrieved November 22, 2023, from [https://esa-worldcover.s3.eu-central-1.amazonaws.com/v100/2020/docs/WorldCover\\_PUM\\_V1.0.pdf](https://esa-worldcover.s3.eu-central-1.amazonaws.com/v100/2020/docs/WorldCover_PUM_V1.0.pdf)
- Van De Kerchove, R., Zanaga, D., Panpan, X., Tsendbazar, N., and Lesiv, M. (2022). *ESA WorldCover Product User Manual V2.0*. Retrieved November 22, 2023, from [https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover\\_PUM\\_V2.0.pdf](https://esa-worldcover.s3.eu-central-1.amazonaws.com/v200/2021/docs/WorldCover_PUM_V2.0.pdf)

- van Leeuwen, D., and Strebe, D. (2006). A “Slice-and-Dice” Approach to Area Equivalence in Polyhedral Map Projections. *Cartography and Geographic Information Science*, 33(4), 269–286. <https://doi.org/10.1559/152304006779500687>
- Venter, Z. S., Barton, D. N., Chakraborty, T., Simensen, T., and Singh, G. (2022). Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sensing*, 14(16), Article 16. <https://doi.org/10.3390/rs14164101>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wang, Z., Zhao, X., Sun, W., Luo, F., Li, Y., and Duan, Y. (2021). Correlation Analysis and Reconstruction of the Geometric Evaluation Indicator System of the Discrete Global Grid. *ISPRS International Journal of Geo-Information*, 10(3), Article 3. <https://doi.org/10.3390/ijgi10030115>
- Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques*, 2nd ed (pp. xxxiii, 1172). Sage Publications, Inc.
- Wecker, L., Hall, J., and Samavati, F. F. (2024). Constructing Efficient Mesh-Based Global Grid Systems with Reduced Distortions. *ISPRS International Journal of Geo-Information*, 13(11), Article 11. <https://doi.org/10.3390/ijgi13110373>
- White, D. (2006). Display of Pixel Loss and Replication in Reprojecting Raster Data from The Sinusoidal Projection. *Geocarto International*, 21(2), 19–22. <https://doi.org/10.1080/10106040608542379>
- White, D., Kimerling, A., Sahr, K., and Song, L. (1998). Comparing Area and Shape Distortion on Polyhedral-Based Recursive Partitions of the Sphere. *International Journal of Geographical Information Science*, 12, 805–827. <https://doi.org/10.1080/136588198241518>

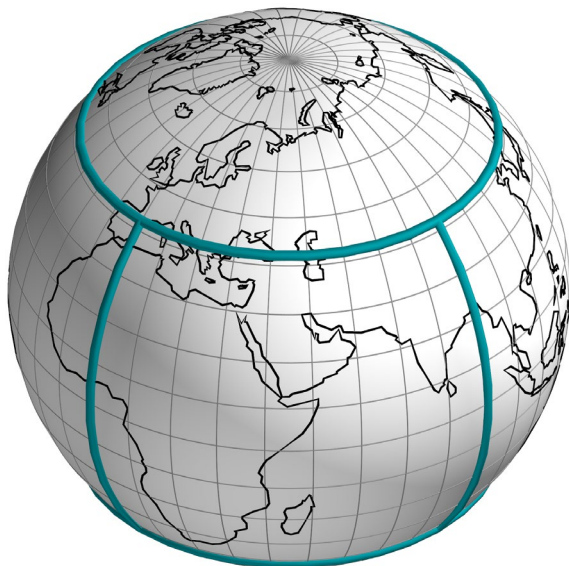
- White, D., Kimerling, J. A., and Overton, S. W. (1992). Cartographic and Geometric Components of a Global Sampling Design for Environmental Monitoring. *Cartography and Geographic Information Systems*, 19(1), 5–22. <https://doi.org/10.1559/152304092783786636>
- Yang, L., Zhu, Z., Izaurrealde, J., and Merchant, J. (1996). *Evaluation of North and South America AVHRR 1-Km data for global environmental modelling*. Third International Symposium on Integrating GIS and Environmental Modeling Proceedings, Santa Fe, NM, USA.
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., and Arino, O. (2022). *ESA WorldCover 10 m 2021 v200* (Version v200) [Dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.7254221>
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., ... Arino, O. (2021). *ESA WorldCover 10 m 2020 v100* (Version v100) [Dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.5571936>
- Zhang, C., and Li, X. (2022). Land Use and Land Cover Mapping in the Era of Big Data. *Land*, 11(10), Article 10. <https://doi.org/10.3390/land11101692>
- Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., and Liu, L. (2024). GLC\_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Earth System Science Data*, 16(3), 1353–1381. <https://doi.org/10.5194/essd-16-1353-2024>
- Zhao, L., Li, G., Yao, X., Ma, Y., and Cao, Q. (2022). An optimized hexagonal quadtree encoding and operation scheme for icosahedral hexagonal discrete global grid systems. *International Journal of Digital Earth*, 15(1), 975–1000. <https://doi.org/10.1080/17538947.2022.2088871>
- Zhao, T., Zhang, X., Gao, Y., Mi, J., Liu, W., Wang, J., Jiang, M., and Liu, L. (2023). Assessing the Accuracy and Consistency of Six Fine-Resolution Global Land Cover Products Using a



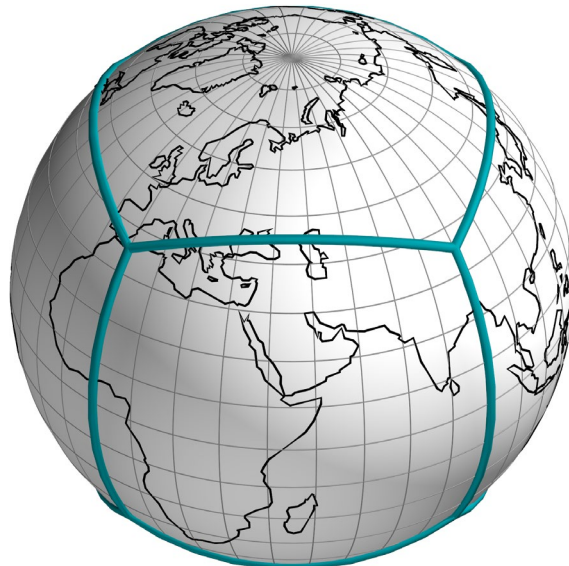
Novel Stratified Random Sampling Validation Dataset. *Remote Sensing*, 15(9). Scopus.  
<https://doi.org/10.3390/rs15092285>

## Appendix A: Visualizations

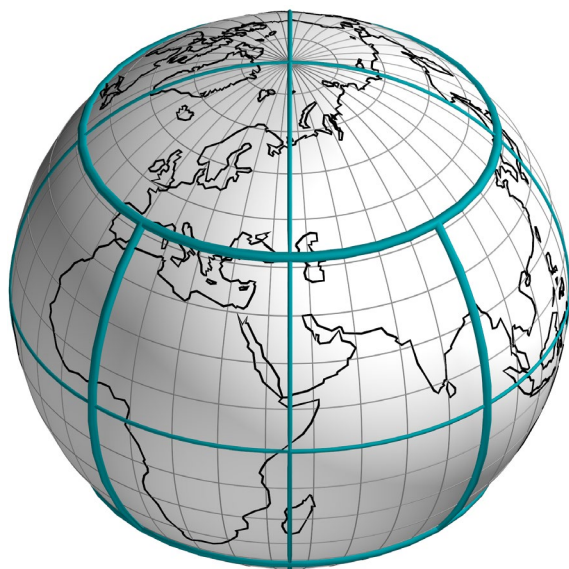
### A.1 Orthographic visualizations of 1-to-4 refinement of the rHEALPix and QPix DGGs



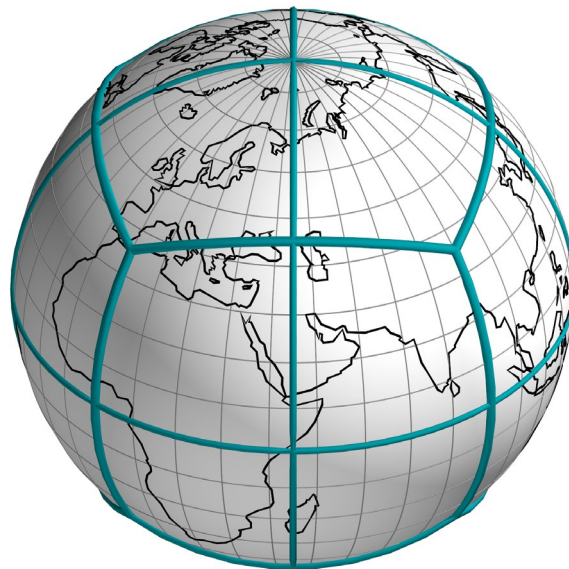
rHEALPix (resolution 0)



QPix (resolution 0)



rHEALPix (resolution 1)



QPix (resolution 1)



rHEALPix (resolution 2)



QPix (resolution 2)



rHEALPix (resolution 3)

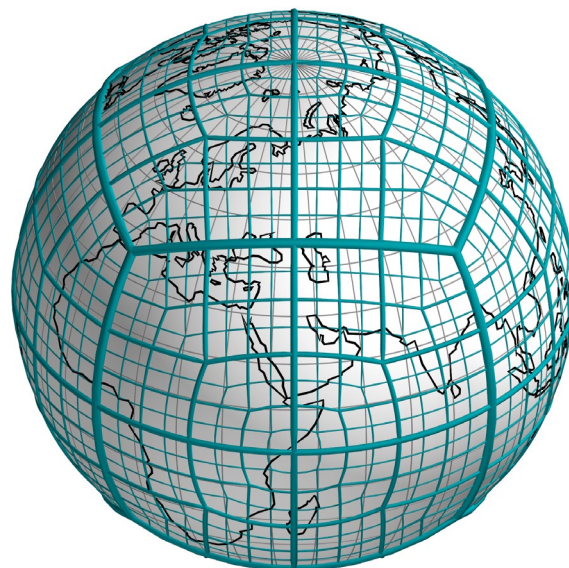


QPix (resolution 3)



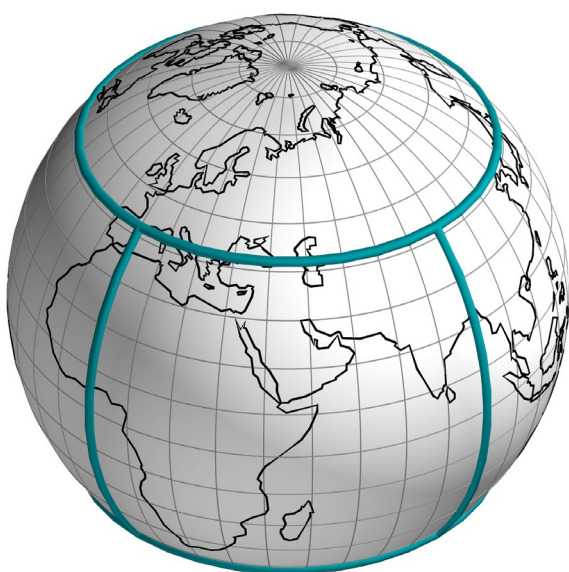


rHEALPix (resolution 4)



QPix (resolution 4)

## A.2 Orthographic visualizations of 1-to-9 refinement of the rHEALPix and QPix DGGs



rHEALPix (resolution 0)



QPix (resolution 0)



rHEALPix (resolution 1)



QPix (resolution 1)



rHEALPix (resolution 2)



QPix (resolution 2)



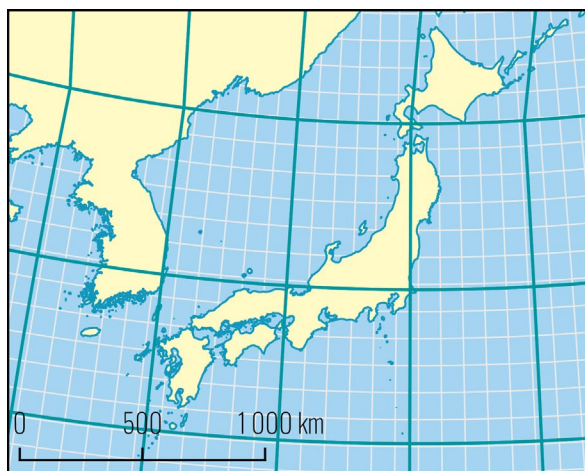


rHEALPix (resolution 3)

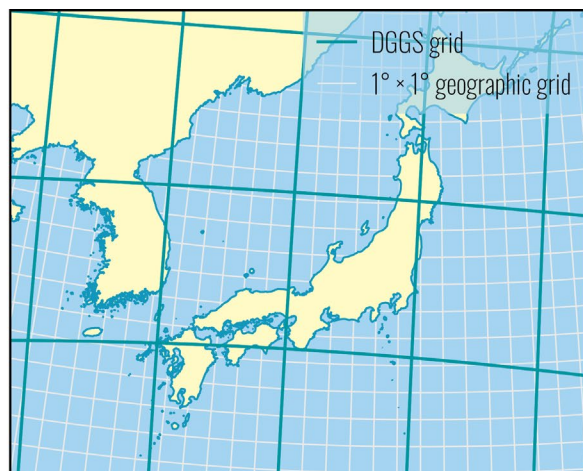


QPix (resolution 3)

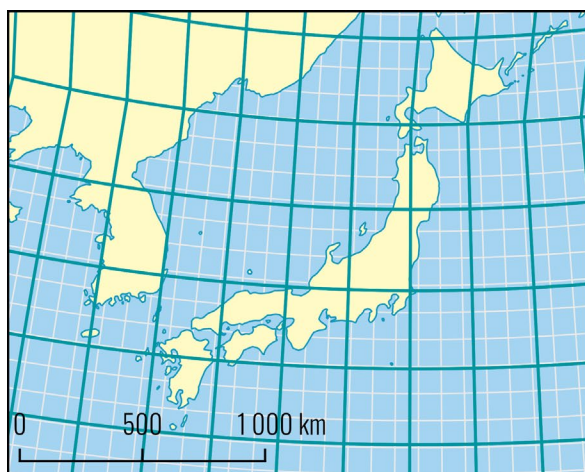
### A.3 Small-scale visualizations of the rHEALPix and QPix DGGS grids



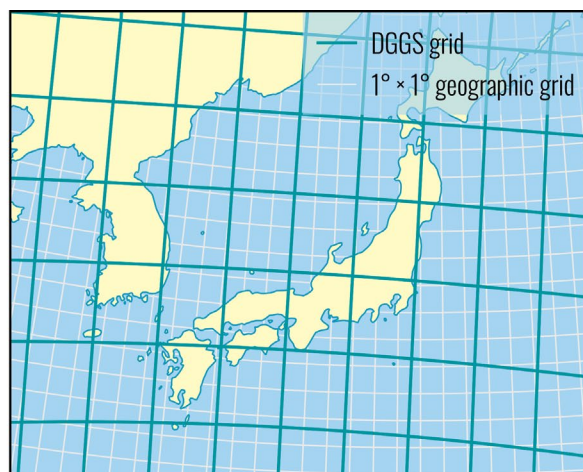
rHEALPix (1-to-4, resolution 4)



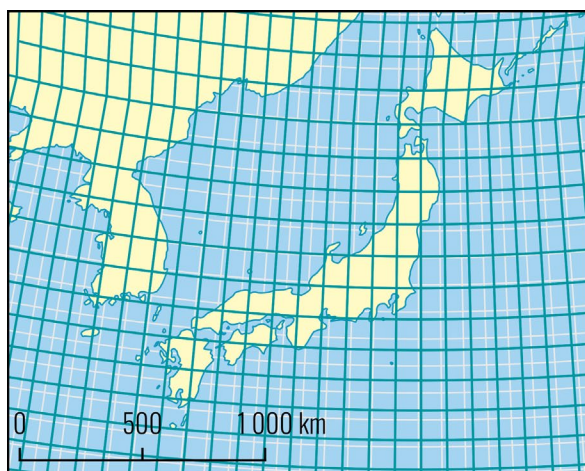
QPix (1-to-4, resolution 4)



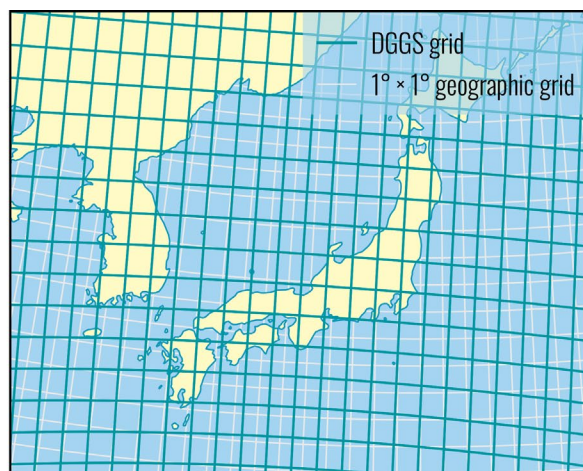
rHEALPix (1-to-4, resolution 5)



QPix (1-to-4, resolution 5)

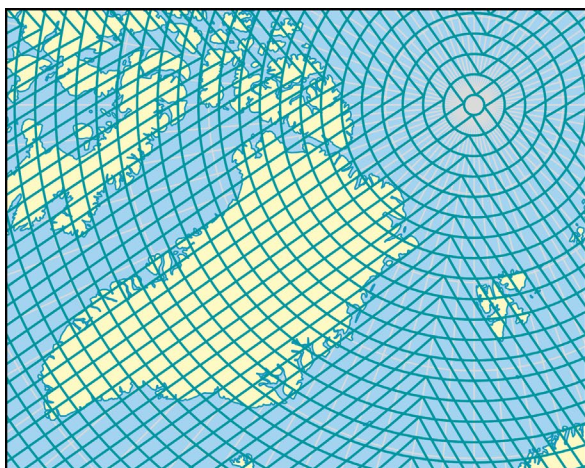


rHEALPix (1-to-9, resolution 4)

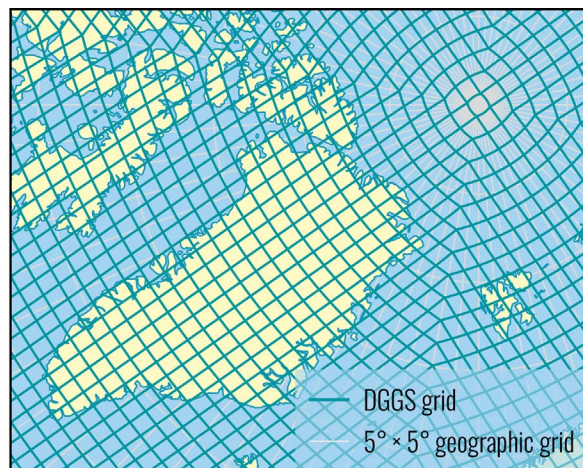


QPix (1-to-9, resolution 4)

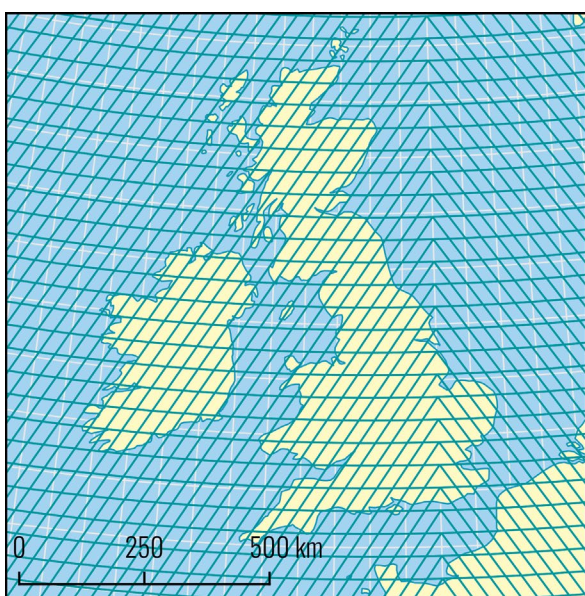




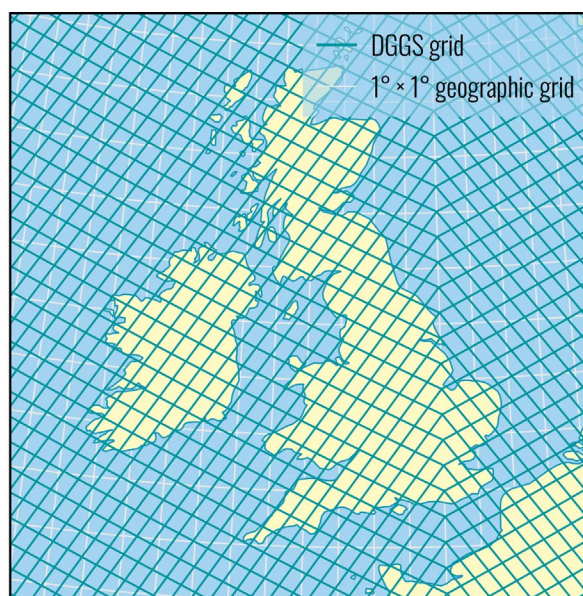
rHEALPix (1-to-9, resolution 4)



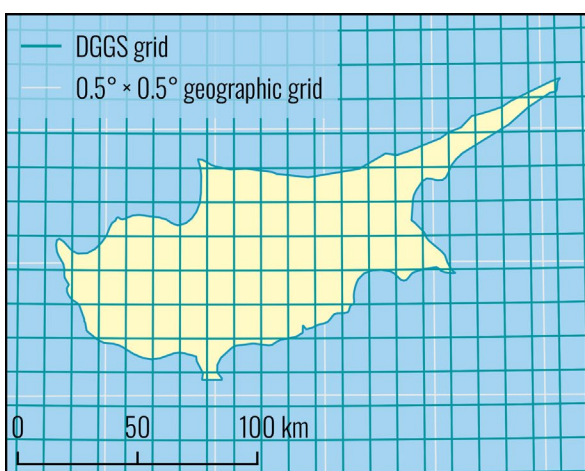
QPix (1-to-9, resolution 4)



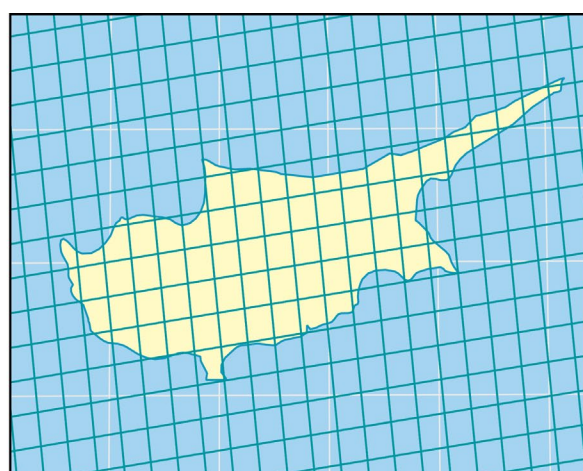
rHEALPix (1-to-9, resolution 5)



QPix (1-to-9, resolution 5)

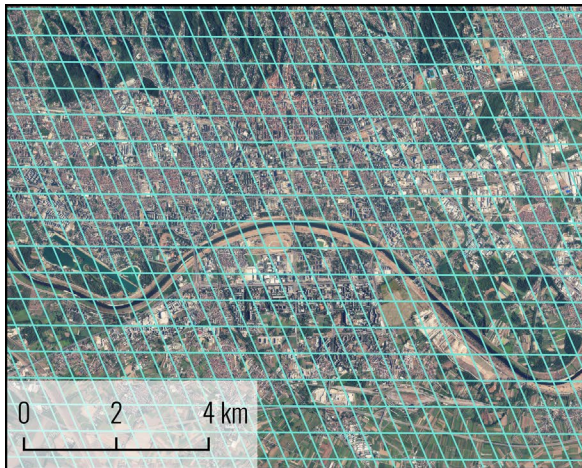


rHEALPix (1-to-9, resolution 5)



QPix (1-to-9, resolution 5)

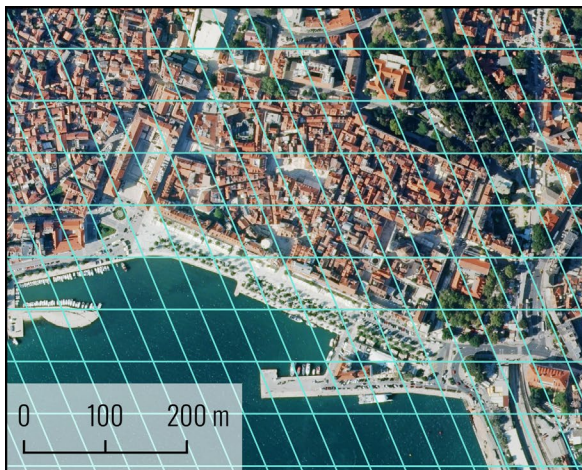




rHEALPix (1-to-9, resolution 9)



QPix (1-to-9, resolution 9)



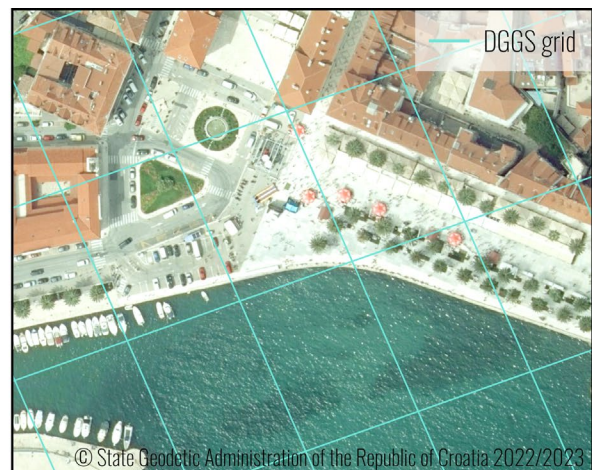
rHEALPix (1-to-9, resolution 11)



QPix (1-to-9, resolution 11)



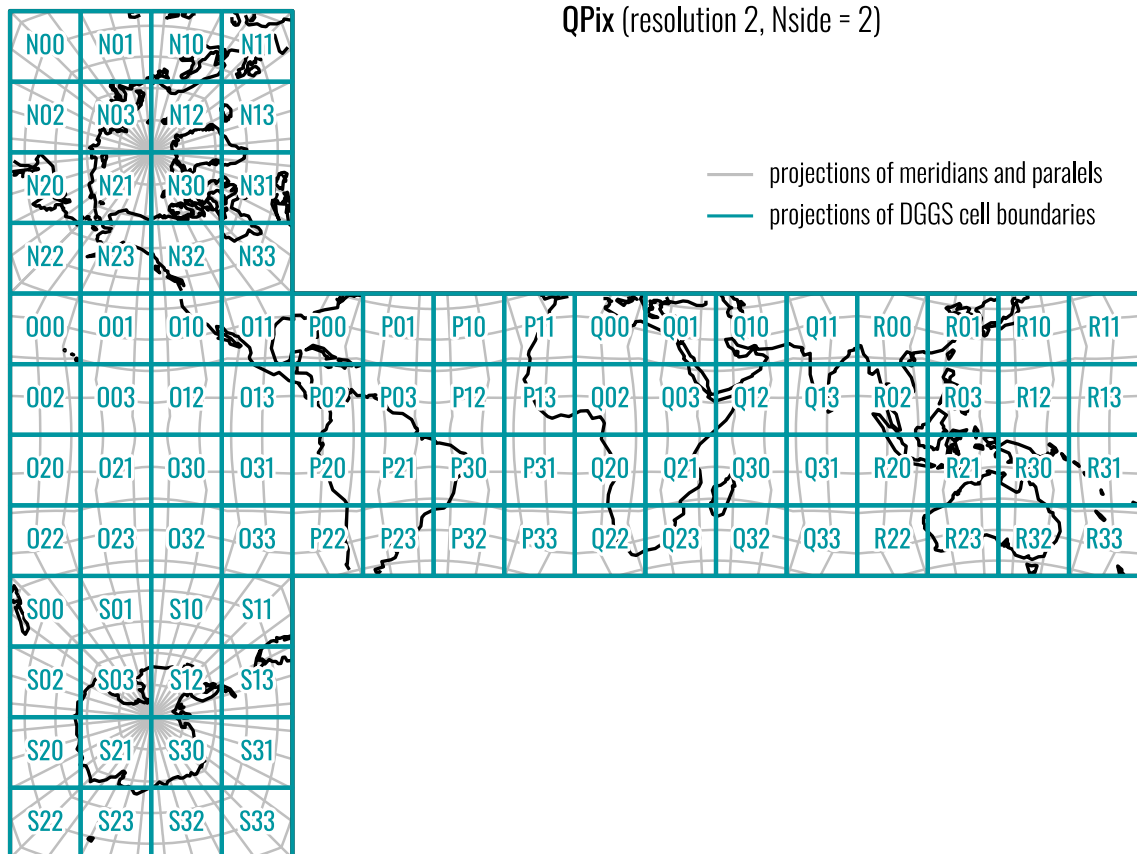
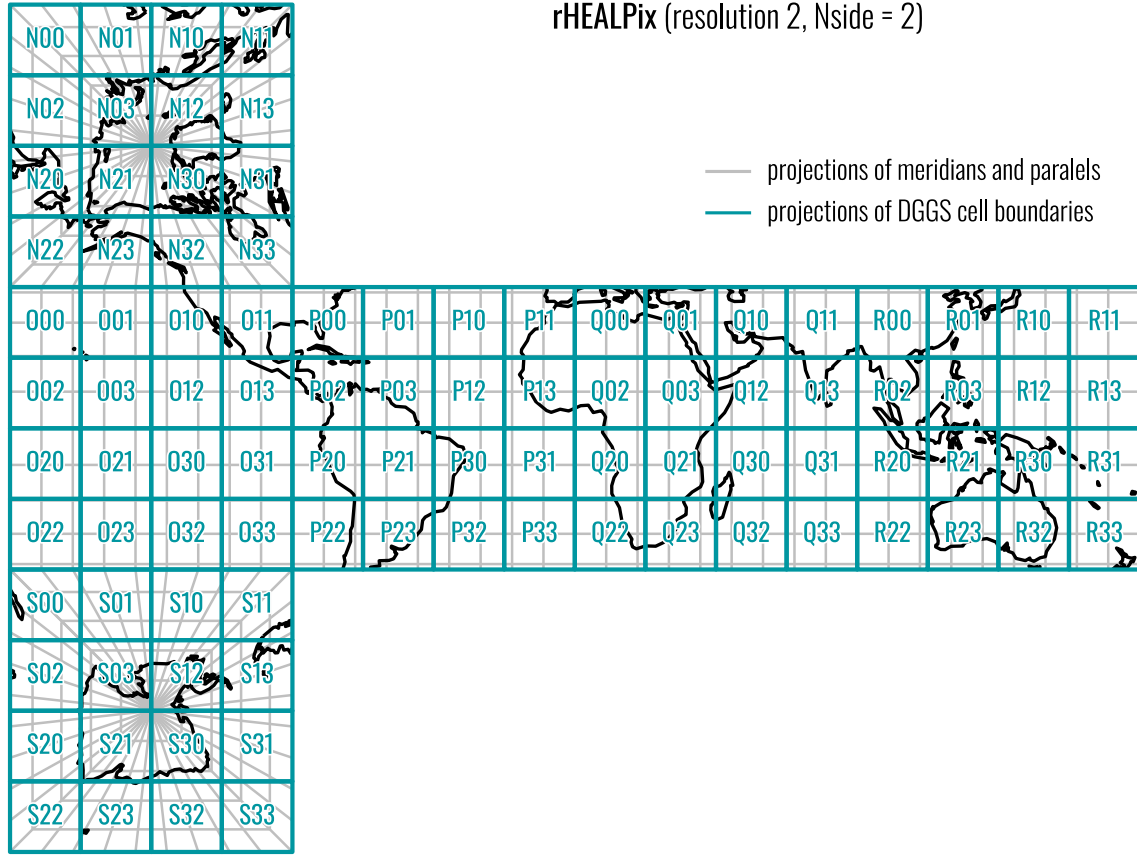
rHEALPix (1-to-9, resolution 11)

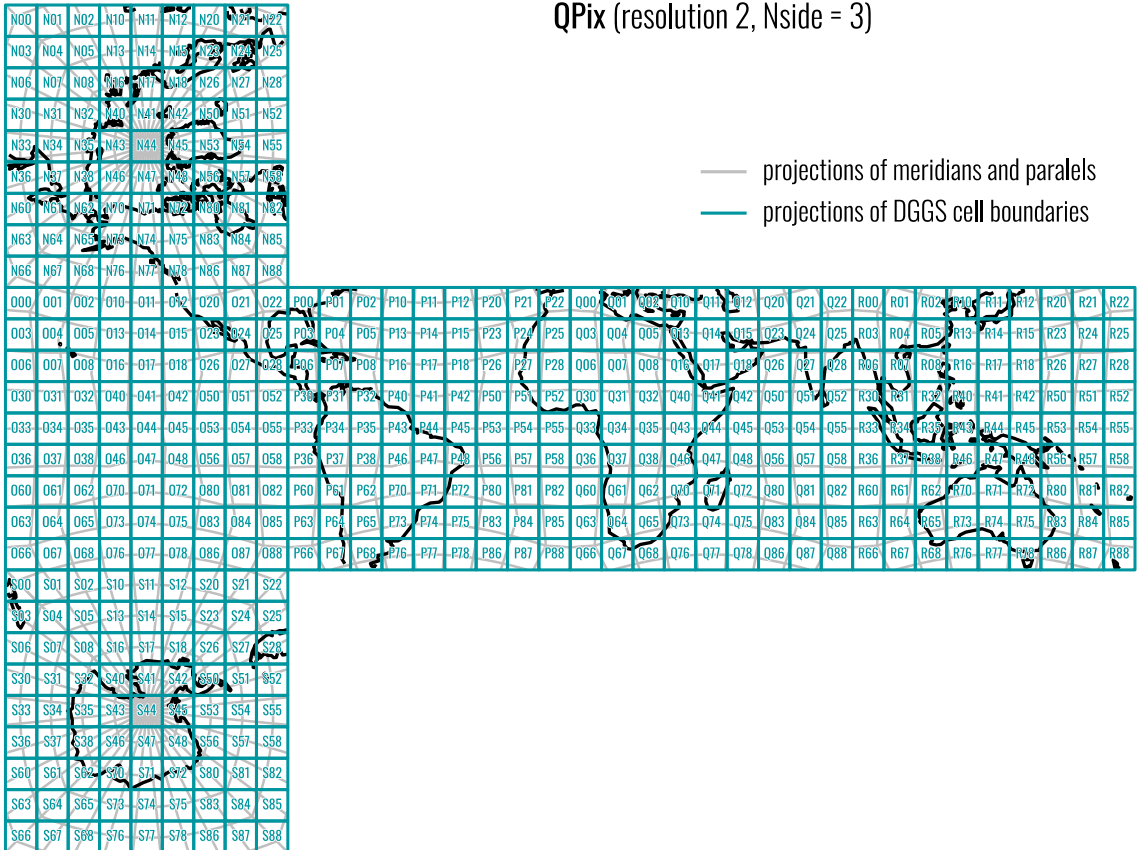
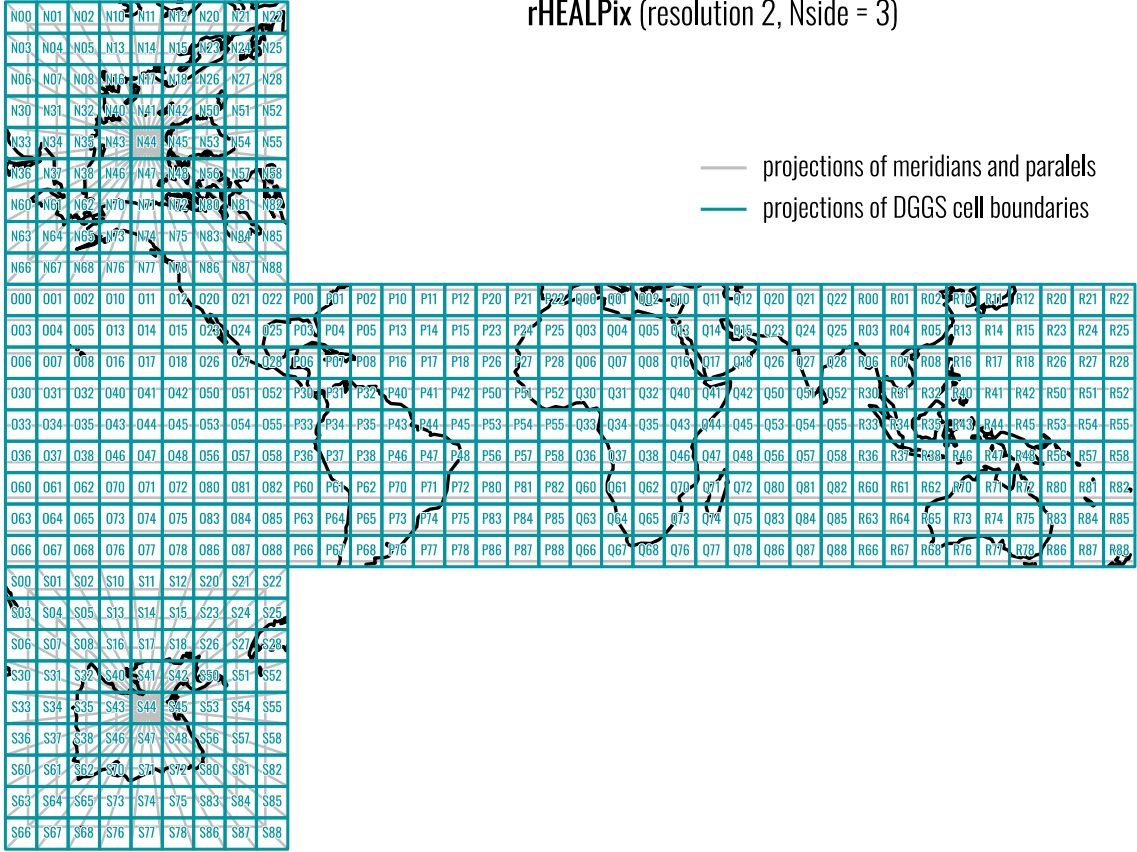


QPix (1-to-9, resolution 11)

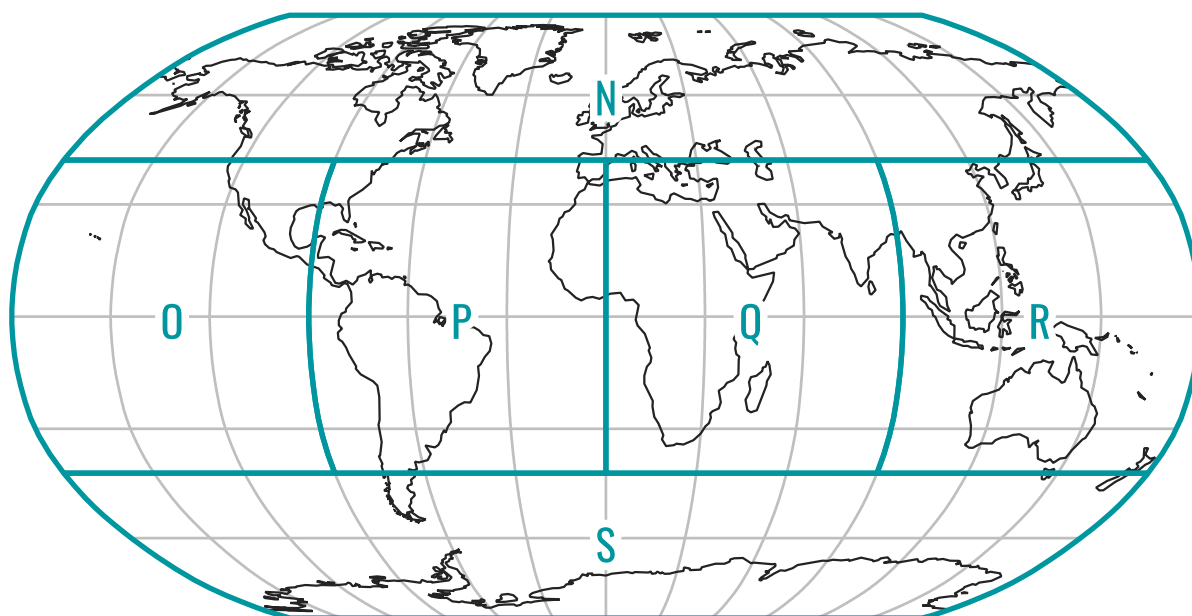


#### A.4 rHEALPix and QPix DGGS cells in a corresponding map projection plane

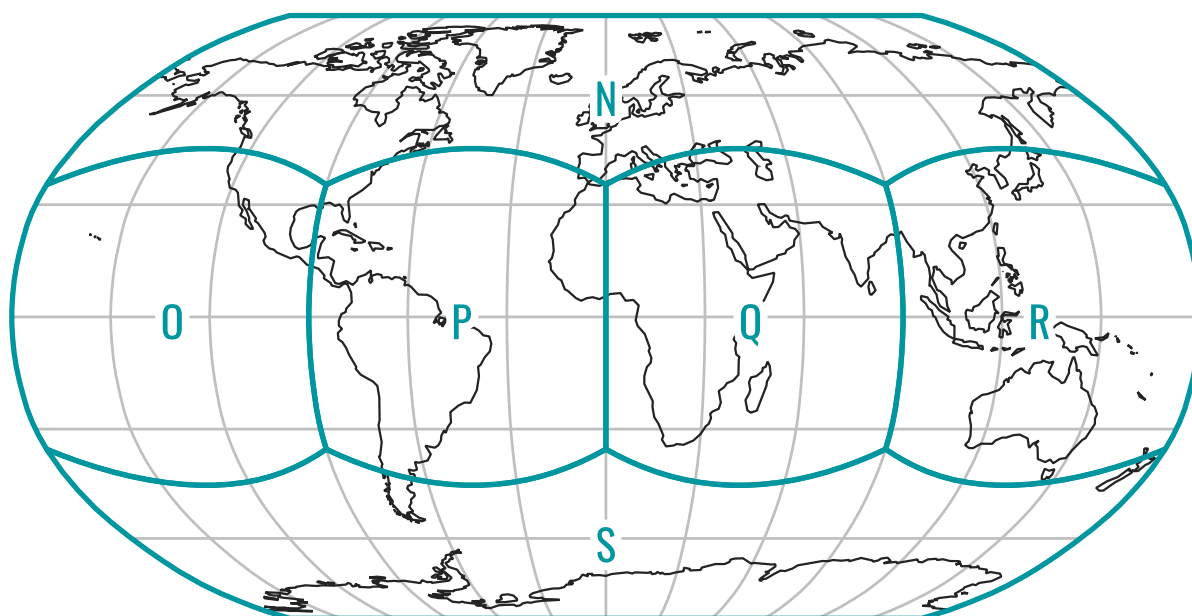




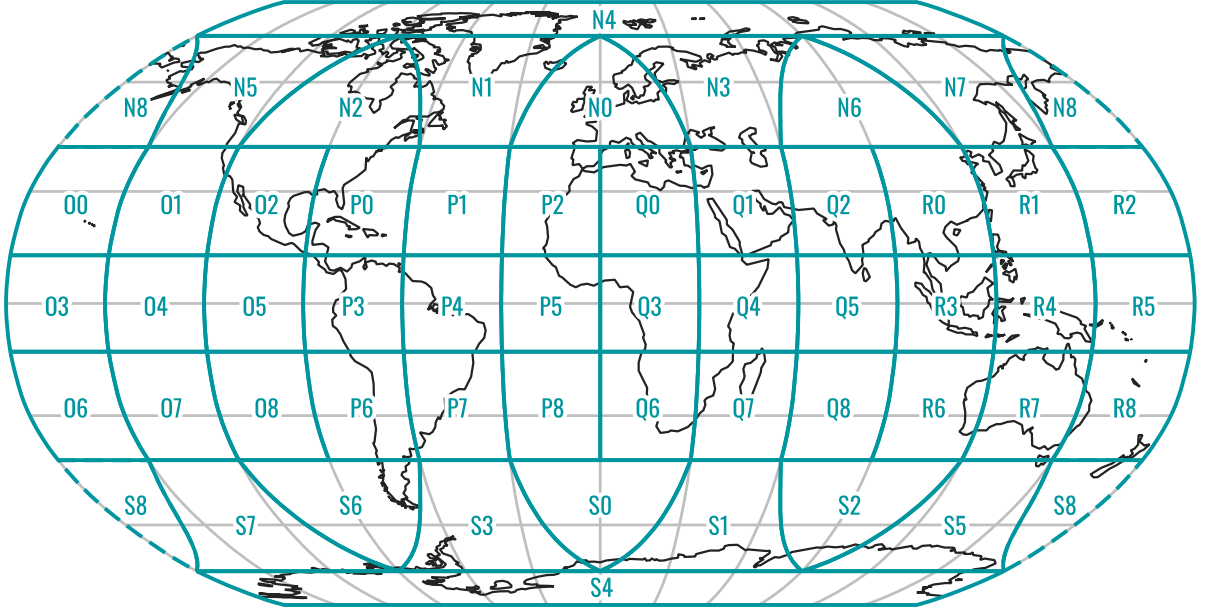
## A.5 rHEALPix and QPix DGGS cells in Robinson map projection



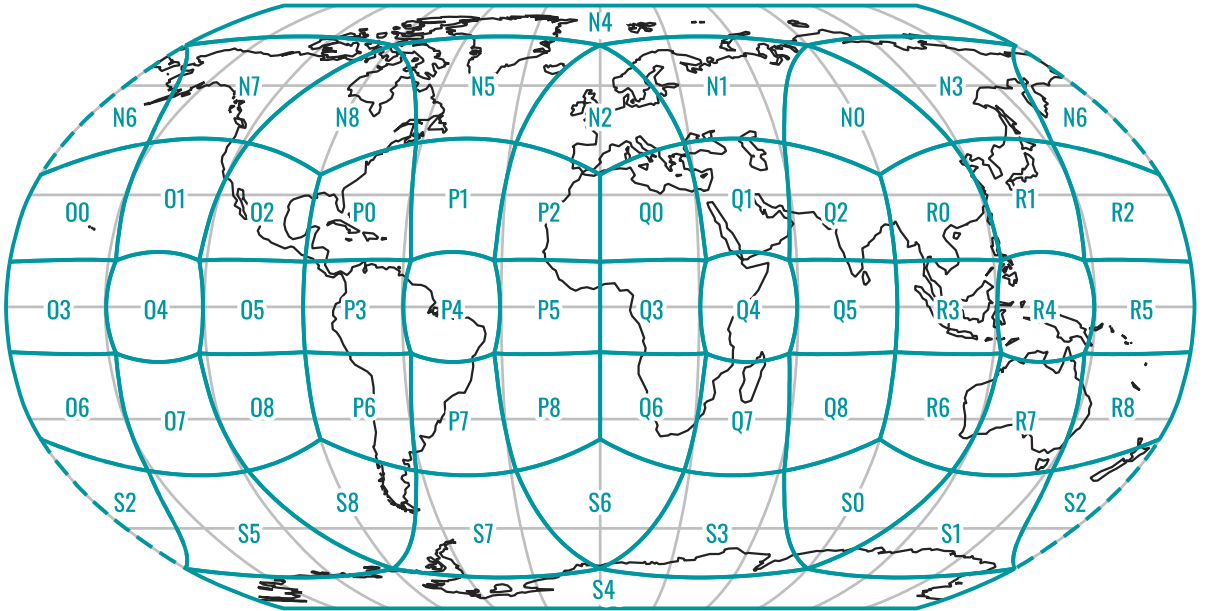
rHEALPix (resolution 0)



QPix (resolution 0)



rHEALPix (resolution 1, Nside = 3)  
(cell identifiers according to (3, 2)-rHEALPix map projection)



QPix (resolution 1, Nside = 3)  
(cell identifiers according to (0, 3)-QPix map projection)

## B DGGs cell areas for various refinement ratios

The table below shows areas of the rHEALPix and QPix DGGs cells for various refinement ratios (parameter  $N_{\text{side}}$  from 2 to 20) across resolutions. Along with the area, square root of area is also provided, which can serve as a means for approximately determining cell sizes. Data are provided only for resolutions that include cells of sizes between 10 km and 1 cm. We did not include larger cells because we do not find them relevant from the land cover data perspective. The 1-cm cells are in general also not relevant from the land cover data perspective, especially not global one, however, we still included those cells since it might be interesting to check what would be the combination of resolution and parameter  $N_{\text{side}}$  that will yield cells that are suitable for some more precise application. All data in table below are based on the WGS 84 ellipsoid (EPSG, n.d.-a).

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
<b><math>N_{\text{side}} = 2</math></b>		
9	324 290 988.75	18 008.08
10	81 072 747.19	9 004.04
11	20 268 186.80	4 502.02
12	5 067 046.70	2 251.01
13	1 266 761.67	1 125.51
14	316 690.42	562.75
15	79 172.60	281.38
16	19 793.15	140.69
17	4 948.29	70.34
18	1 237.07	35.17
19	309.27	17.59
20	77.32	8.79
21	19.33	4.40
22	4.83	2.20
23	1.21	1.10
24	0.30	0.55
25	0.08	0.27
26	0.02	0.14
27	0.0047	0.07
28	0.0012	0.03
29	0.0003	0.02
30	0.0001	0.01

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
<b><math>N_{\text{side}} = 3</math></b>		
6	159 963 075.78	12 647.65
7	17 773 675.09	4 215.88
8	1 974 852.79	1 405.29
9	219 428.09	468.43
10	24 380.90	156.14
11	2 708.99	52.05
12	301.00	17.35
13	33.44	5.78
14	3.72	1.93
15	0.41	0.64
16	0.05	0.21
17	0.01	0.07
18	0.0006	0.02
19	0.0001	0.01
<b><math>N_{\text{side}} = 4</math></b>		
4	1 297 163 954.99	36 016.16
5	81 072 747.19	9 004.04
6	5 067 046.70	2 251.01
7	316 690.42	562.75
8	19 793.15	140.69
9	1 237.07	35.17
10	77.32	8.79
11	4.83	2.20

# DGGS CELL AREAS FOR VARIOUS REFINEMENT RATIOS

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
12	0.30	0.55
13	0.02	0.14
14	0.0012	0.03
15	0.0001	0.01
<b>N<sub>side</sub> = 5</b>		
4	217 627 998.60	14 752.22
5	8 705 119.94	2 950.44
6	348 204.80	590.09
7	13 928.19	118.02
8	557.13	23.60
9	22.29	4.72
10	0.89	0.94
11	0.04	0.19
12	0.0014	0.04
13	0.0001	0.01
<b>N<sub>side</sub> = 6</b>		
3	1 822 079 410.02	42 685.82
4	50 613 316.95	7 114.30
5	1 405 925.47	1 185.72
6	39 053.49	197.62
7	1 084.82	32.94
8	30.13	5.49
9	0.84	0.91
10	0.02	0.15
11	0.00065	0.03
12	0.00002	0.004
<b>N<sub>side</sub> = 7</b>		
3	722 581 041.52	26 880.87
4	14 746 551.87	3 840.12
5	300 950.04	548.59
6	6 141.84	78.37
7	125.34	11.20
8	2.56	1.60
9	0.05	0.23
10	0.00107	0.03
11	0.00002	0.005
<b>N<sub>side</sub> = 8</b>		
3	324 290 988.75	18 008.08
4	5 067 046.70	2 251.01

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
5	79 172.60	281.38
6	1 237.07	35.17
7	19.33	4.40
8	0.30	0.55
9	0.0047	0.07
10	0.0001	0.01
<b>N<sub>side</sub> = 9</b>		
3	159 963 075.78	12 647.65
4	1 974 852.79	1 405.29
5	24 380.90	156.14
6	301.00	17.35
7	3.72	1.93
8	0.05	0.21
9	0.00057	0.02
10	0.00001	0.003
<b>N<sub>side</sub> = 10</b>		
2	8 501 093 695.40	92 201.38
3	85 010 936.95	9 220.14
4	850 109.37	922.01
5	8 501.09	92.20
6	85.01	9.22
7	0.85	0.92
8	0.01	0.09
9	0.0001	0.01
<b>N<sub>side</sub> = 11</b>		
2	5 806 361 379.28	76 199.48
3	47 986 457.68	6 927.23
4	396 582.29	629.75
5	3 277.54	57.25
6	27.09	5.20
7	0.22	0.47
8	0.00185	0.04
9	0.00002	0.004
<b>N<sub>side</sub> = 12</b>		
2	4 099 678 672.55	64 028.73
3	28 469 990.78	5 335.73
4	197 708.27	444.64
5	1 372.97	37.05
6	9.53	3.09

DGGs CELL AREAS FOR VARIOUS REFINEMENT RATIOS

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
7	0.07	0.26
8	0.000460	0.02
9	0.000003	0.002
<b>N<sub>side</sub> = 13</b>		
2	2 976 469 204.65	54 557.03
3	17 612 243.81	4 196.69
4	104 214.46	322.82
5	616.65	24.83
6	3.65	1.91
7	0.02	0.15
8	0.0001	0.01
<b>N<sub>side</sub> = 14</b>		
2	2 212 904 439.66	47 041.52
3	11 290 328.77	3 360.11
4	57 603.72	240.01
5	293.90	17.14
6	1.50	1.22
7	0.01	0.09
8	0.00004	0.01
<b>N<sub>side</sub> = 15</b>		
2	1 679 228 384.28	40 978.39
3	7 463 237.26	2 731.89
4	33 169.94	182.13
5	147.42	12.14
6	0.66	0.81
7	0.00291	0.05
8	0.00001	0.004
<b>N<sub>side</sub> = 16</b>		
2	1 297 163 954.99	36 016.16
3	5 067 046.70	2 251.01
4	19 793.15	140.69
5	77.32	8.79
6	0.30	0.55
7	0.001180	0.03
8	0.000005	0.002

Res.	Area (m <sup>2</sup> )	Sqrt of area (m)
<b>N<sub>side</sub> = 17</b>		
2	1 017 839 069.86	31 903.59
3	3 521 934.50	1 876.68
4	12 186.62	110.39
5	42.17	6.49
6	0.15	0.38
7	0.000505	0.02
8	0.000002	0.001
<b>N<sub>side</sub> = 18</b>		
2	809 813 071.12	28 457.21
3	2 499 423.06	1 580.96
4	7 714.27	87.83
5	23.81	4.88
6	0.07	0.27
7	0.000227	0.02
8	0.000001	0.001
<b>N<sub>side</sub> = 19</b>		
2	652 319 556.74	25 540.55
3	1 806 979.38	1 344.24
4	5 005.48	70.75
5	13.87	3.72
6	0.04	0.20
7	0.0001	0.01
<b>N<sub>side</sub> = 20</b>		
2	531 318 355.96	23 050.34
3	1 328 295.89	1 152.52
4	3 320.74	57.63
5	8.30	2.88
6	0.02	0.14
7	0.0001	0.01



## CV of the doctoral candidate

Frane Gilić was born in 1994 in Split, Croatia where he finished primary school and gained a qualification of Land Surveying Technician from the Technical School of Construction and Geodesy Split in 2013. In 2016, he obtained an academic title of the Bachelor of Engineering in the field of geodesy and geoinformatics from the Faculty of Civil Engineering, Architecture and Geodesy, University of Split and in 2019 he received an academic title of the Master of Engineering in the field of geodesy and geoinformatics from the Faculty of Geodesy, University of Zagreb. The same year he enrolled in the doctoral study in the field of geodesy and geoinformatics at the Faculty of Geodesy, University of Zagreb and got employed at the Faculty of Civil Engineering, Architecture and Geodesy, University of Split as a Teaching and Research Assistant, where he is also currently employed. At his workplace, he participates in conducting exercises in several courses in the field of geodesy and geoinformatics at undergraduate university studies.

He has participated as a presenter at several international scientific conferences, he is a co-author of one university educational manual, and an author and co-author of three scientific papers published in international scientific journals and four papers published in the scientific conference proceedings. He participated in multiple international projects and is currently a team member on one Horizon Europe scientific project.

## List of papers published in scientific journals

- Duplančić Leder, T., Baučić, M., Leder, N., and Gilić, F. (2023). Optical Satellite-Derived Bathymetry: An Overview and WoS and Scopus Bibliometric Analysis. *Remote Sensing*, 15(5), 1294. <https://doi.org/10.3390/rs15051294>
- Gilić, F., and Gašparović, M. (2025). Enhancing Authalic Latitude Calculation for the rHEALPix DGGS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 12675–12683. <https://doi.org/10.1109/JSTARS.2025.3567839>

Gilić, F., Gašparović, M., and Baučić, M. (2023). Current State and Challenges in Producing Large-Scale Land Cover Maps: Review Based on Recent Land Cover Products. *Geocarto International*, 38(1), 2242693. <https://doi.org/10.1080/10106049.2023.2242693>

### **List of papers published in scientific conference proceedings**

Baučić, M., Gilić, F., Bačić, S., and Duplančić-Leder, T. (2020). Open Geospatial Data for Urban Green Areas. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, VI-4/W2-2020, 17–24. <https://doi.org/10.5194/isprs-annals-vi-4-w2-2020-17-2020>

Baučić, M., Gilić, F., Morić-Španić, A., and Povh Škugor, D. (2024). Global Open Maps for the Integrated Monitoring and Assessment Programme of the Mediterranean Sea and Coast. *International Conference of Environmental Remote Sensing and GIS, 2024*, 101–104. <https://doi.org/10.5281/zenodo.11621508>

Gilić, F., Baučić, M., and Termos, S. (2024). Global Land Cover Mapping – Need for Discrete Global Grid System. *International Conference of Environmental Remote Sensing and GIS, 2024*, 167–170. <https://doi.org/10.5281/zenodo.11621508>

Gilić, F., and Gašparović, M. (in press). Calculating Land Cover Change: DGGS Perspective. Mediterranean Geosciences Union 4<sup>th</sup> Annual Meeting (MedGU-24), Barcelona, Spain.