Technical Note

Transformation of a Classified Image from Pixel Clutter to Land Cover Map Using Geometric Generalization and Thematic Self-Enrichment

Geir-Harald Strand [1,*], Eva Solbjørg Flo Heggem [1], Linda Aune-Lundberg [1], Agata Hościło [2] and Adam Waśniewski [3]

1 Division of Survey and Statistics, Norwegian Institute of Bioeconomy Research,
P.O. Box 115, 1431 Ås, Norway; eva.heggem@nibio.no (E.S.F.H.); linda.aune-lundberg@nibio.no (L.A.-L.)
2 National Centre for Emissions Management, Institute of Environmental Protection—National Research Institute, 02-170 Warsaw, Poland; agata.hoscilo@ios.edu.pl
3 Centre of Applied Geomatics, Institute of Geodesy and Cartography, 02-679 Warszawa, Poland; adam.wasniewski@igik.edu.pl
* Correspondence: ghs@nibio.no

Abstract: Land cover maps are frequently produced via the classification of satellite imagery. There is a need for a practicable and automated approach for the generalization of these land cover classification results into scalable, digital maps while minimizing information loss. We demonstrate a method where a land cover raster map produced using the classification of Sentinel 2 imagery was generalized to obtain a simpler, more readable land cover map. A replicable procedure following a formal generalization framework was applied. The result of the initial land cover classification was separated into binary layers representing each land cover class. Each binary layer was simplified via structural generalization. The resulting images were merged to create a new, simplified land cover map. This map was enriched by adding statistical information from the original land cover classification result, describing the internal land cover distribution inside each polygon. This enrichment preserved the original statistical information from the classified image and provided an environment for more complex cartography and analysis. The overall accuracy of the generalized map was compared to the accuracy of the original, classified land cover. The accuracy of the land cover classification in the two products was not significantly different, showing that the accuracy did not deteriorate because of the generalization.

Keywords: land cover; pixel clutter; cartography; generalization; enrichment

1. Introduction

Land cover data are frequently produced using the classification of satellite imagery. The primary results are datasets where the entities are raster pixels assigned to land cover classes. The classification may or may not be correct and verification is required before a dataset can be used for any particular purpose. The raster image format is another confounding element. Raster structures rarely exist on the ground and the raster often appears cluttered because the spatial resolution is high and the amount of information is excessive [1].

The spatial distribution of land cover classes is generally perceived by users as a partition of the earth surface represented as non-overlapping polygons. Raster images can be converted to polygon (vector) data using two processes, often combined: segmentation and generalization. Both can be strengthened through data enrichment [2].

The generalization of spatial data is an activity that originated in cartography. The topic was intensively researched when geographic information systems were in their infancy [3,4]. The term cartographic generalization describes the changes made when a large-scale map (with many details) is transformed into a smaller-scale map (with fewer details).

Originally a design technique carried out manually by a cartographer, computer cartography has changed generalization into a set of algorithms allowing spatial data with a high level of detail to be rendered as a map with a lower level of detail. With spatial data stored and managed as digital datasets, elementary generalization has become a process where data are adjusted to a suitable visual representation by leaving out some of the details in the original dataset. Thus, even naïve generalization can be defined as a special variant of spatial modeling [5]. More recent research has widened the topic to provide for dynamic web mapping and support for modeling and analysis [6].

According to Bertin [7], there are two different types of cartographic generalization: structural generalization and conceptual generalization. Structural generalization is concerned with the geometry and the choice of cartographic symbols, while conceptual generalization is concerned with ontology. Structural generalization changes how a feature is represented in the map while conceptual generalization involves changing the identity and substance of the features. This is described by Wolf [8] as feature modification based on geometry vs. feature modification based on semantics.

Steiniger and Weibel [9] demonstrate how generalization procedures can be designed using a typology of the relations between map objects. According to these authors, structural generalization handles horizontal relations, including geometric and topological properties in the dataset. Structural generalization is thus carried out on single features individually and must be embedded in procedures addressing "update relations". These "update relations" handle the relationship between features when more than a single feature is involved.

Land cover maps with several land cover classes can be compiled using remote sensing data classification but will always present a simplified reflection of the physical world [10]. These maps should be customized to meet the user requirements. Raw land cover classification results are usually cluttered. Map generalization is needed to emphasize the relevant land cover information while omitting less important features, with respect to map scale and purpose [11]. Generalization allows the removal of classification errors or misclassification, offsetting the "salt and pepper' effect, which is common in pixel-based classifications of fine spatial resolution imagery [12]. Tailored sequences of cartographic operations must be applied to handle geometry, topology and attribute data. The procedures can involve basic operations like reclassification, aggregation, amalgamation, displacement, elimination, enlargement, exaggeration, symbolization, simplification and smoothing [13].

To operationalize generalization as a process, a land cover map can be separated into a stack of binary images, one for each land cover class (creating vertical relations between the classes). Structural generalization is subsequently carried out separately on each binary image. The results can later be combined using the vertical relationships between the layers to produce a new, generalized land cover map. Several structural generalization processes addressing horizontal relations are thus embedded in a contextual process using vertical relations by first splitting the original map into binary layers and later merging the generalized layers into a single, generalized land cover map.

The structural generalization process is faced with at least two closely related challenges; data loss due to simplification and inaccuracies due to a mismatch between class categories and reality. The simplified geometry is believed to be less accurate than the original with respect to information concerning where a category is present. The delineation of each class is also assumed to be less precise than in the original data. Furthermore, the class names inherited from the original maps may describe the actual content of the resulting polygons less accurately.

A possible solution is to take advantage of the vertical relations between the original data and the generalized dataset. This can be implemented through data enrichment [14–16] and conceptual generalization [17]. Data enrichment consists of creating a land cover profile of each generalized feature (polygon) based on the original raster data. Conceptual generalization involves the creation of new classes and reassigning the features to these classes based on the auxiliary information obtained via enrichment.

Enrichment also creates opportunities for the new and alternative use of the generalized dataset. An enriched dataset is stocked with supplementary information and a land cover class can be separated into new classes based on the profiles of the features. Generalization followed by self-enrichment is not only an act of simplification, but also a framework for the conceptual diversification of the dataset.

The objective of this study was to develop and implement a practicable, mechanistic approach to the generalization of a land cover classification result and to support the later analytical use of the land cover map via the self-enrichment of the resulting dataset. We expected that the new map would provide for more diverse applications but also assumed that the generalization would cause a distortion of the information and introduce statistical bias. The latter hypothesis was examined by comparing the content and accuracy of the original classified land cover raster to the generalized map product.

2. Materials and Methods

The study used part of an existing land cover map of the former Viken and Oslo counties in south-eastern Norway derived from the classification of Sentinel 2 imagery. The former Viken county is located in the south-eastern part of Norway, surrounding Oslo, the capital of Norway. The county extended from the Swedish border and the Oslo Fiord with a flat coastal landscape up to the mountainous areas of Hardangervidda (appr. 1900 m above sea level) in the north-west. The southern region is characterized by farmland and forest, changing into a mountain and valley landscape towards the north and north-west.

The land cover classification was carried out using a Random Forest (RF) algorithm. The detailed description of the land cover classification process is found in [18]. A subset (12,056 km^2) with nine land cover classes (Table 1) was used in the study of the generalization methodology. We are not concerned with the details of the classification method here since our project only used the classification results as a testbed.

Table 1. The nine landcover classes found in the land cover map of the study area.

Code	Name	Definition
1	Sealed surfaces	Land covered by buildings, roads and other human-made structures such as railroads. Buildings include both residential and industrial built-up areas
2	Woodland broadleaved	Land cover dominated by trees with cover of 10% or more; more than 50% covered by broadleaved trees
3	Woodland coniferous	Land cover dominated by trees with cover of 10% or more; more than 50% covered by coniferous trees
4	Permanent herbaceous	Land covered with pastures
5	Periodically herbaceous	Land covered with annual cropland sowed/planted and harvested at least once within the 12 months after the sowing/planting date.
6	Mosses	Peat bogs covered by mosses
7	Water	Area covered by water, including the following formations: lakes, ponds, rivers and ocean
8	Low vegetation	Land covered with intermediate and vigorous low mountain vegetation
9	Sparsely vegetated	Land covered with more than 70% bare rock and with less than 30% vegetation (including lichens)

Field observations from the Norwegian area frame survey of land cover and outfield land resources survey [19,20] were used in the accuracy assessment. This is a field survey with sample points scattered throughout Norway. In total, 362 sample points fell within the study area.

The study consisted of five components: (1) development of a generalization methodology (establishing a practicable, mechanistic approach to generalization of a land cover), (2) enrichment of the generalized map (to preserve the information in the original map), (3) examination of the content of the generalized land cover map, (4) conceptual improvement of the final product and (5) comparing statistics and accuracy.

(1)   Generalization methodology

The development of a practicable, mechanistic approach to the generalization of a land cover classification was carried out by dividing the process into tasks composed of sequences of smaller steps. The initial land cover map was transposed into a set of binary maps representing the individual land cover classes. Each binary map was filtered and cleaned and the final, generalized land cover map was created by merging the refined binary layers. The methodology was implemented using the Open Source GDAL library for raster and vector geospatial data. The script was written as a standard Shell script running on a Linux computer.

A prominent feature in the generalization methodology was the use of morphological filters [21,22]. The main morphological filters used in the study are illustrated using a simplified example in Figure 1. Figure 1a is a binary, cluttered pixel map. A growth filter is applied, where each pixel is extended into all eight pixels in its neighborhood. The new pixels included in the map are shown with a violet color in Figure 1b. A contraction filter is then applied, removing every pixel with a non-committed neighbor in any of the four cardinal directions (north, east, south, west). The removed pixels are shown with a beige color in Figure 1c; the red and violet pixels are retained. Finally, small areas (single pixels in the example) were removed, resulting in the generalized binary map in Figure 1d.
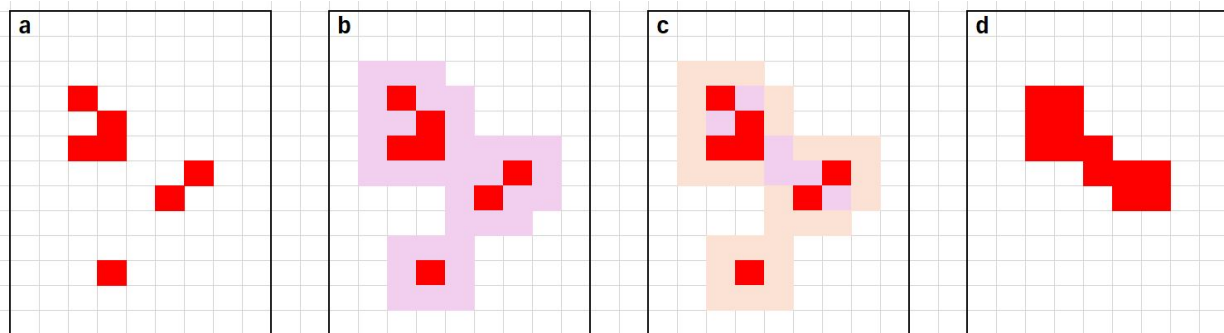

Figure 1. Morphological filters (simplified example). (**a**) Binary pixel map; (**b**) growth; (**c**) contraction; (**d**) removal and result. Detailed explanation (including the use of colors) is found in the text.

(2)   Enrichment

The enrichment of the generalized polygons was conducted by linking a vectorized version of the original land cover classification to the polygons of the generalized map to create a statistical profile of each polygon. The statistical profile is simply a list of the relative size of each land cover class inside the polygons. The enriched product has at least two applications. It can be used to create overall descriptive statistics for the classes by summarizing the profiles for the entire map. It can also be used to create new thematic map products.

(3)   Examination

The results of the generalization were studied using descriptive statistics. The statistical software IBM® SPSS® Version 27 was used for this purpose. Area statistics for the land

cover classes based on polygon area and enrichment were tabulated and compared using the Aggregate and Summarize functions. We also created profiles and described the composition of the land cover types present in each polygon and each class after generalization. Examples are found in the Results section below.

(4) Conceptual improvement

The distribution of land cover types in each polygon was examined. Polygons where at least 80% of the surface corresponded with the assigned land cover class were characterized as pure. Polygons where 50–80% of the surface corresponded with the land cover class were characterized as "heterogeneous". Polygons where no single class covered at least 50% of the polygon were characterized as "complex". This characterization was stored as a separate attribute.

(5) Comparing statistics and accuracy

We used a random sample of 362 field observations from the Norwegian area frame survey of land cover and outfield land resources survey [19] to quantify the impact of the generalization on the land cover statistics. The land cover at these sample plots was verified by field observations and cross-checked on recent orthophoto (2020 +/− two years). The land cover types reported in the remote sensing product before and after generalization were compared to the ground truth and the overall accuracy was calculated as the correctly classified proportion. The accuracy before and after generalization was compared using a paired sample *t*-test.

3. Results

The workflow in the generalization methodology developed in the project is illustrated in Figure 2. The initial step in the geometric generalization was to split the raw land cover classification map into binary layers, each representing a single land cover class. The nine land cover classes were thus separated into nine binary land cover layers, each representing a single land cover class.

This step was followed by structural generalization applied individually to each binary layer. The structural generalization is represented as a single element (described as "Expansion, Contraction and Removal" in Figure 2), but consisted of several sub-steps, each using horizontal relations in the dataset. These details are illustrated in Figure 3. The procedure is based on a method developed for vector maps [23] but is also applicable for raster datasets by using morphological filters. The main sub-steps were as follows:

Sub-step a: Expansion (growth). A filter operation focused on those pixels where the class was present in the binary layer. All eight neighbors of these pixels were assigned to the class. See Figure 1b. The sub-step can be repeated several times to strengthen the presence of a class.

Sub-step b: Contraction. A filter operation focused on each pixel where the said class was present after sub-step a. The pixels were removed from the class if the class was absent in any neighbor in a cardinal direction from the pixel. See Figure 1c. This sub-step can be repeated several times to debilitate the presence of a class.

Sub-step c: Removal. Any pixel belonging to a continuous group of pixels smaller than a predefined threshold value was removed from the class. See Figure 1d. The threshold is adjustable. The threshold in this exercise was set to 150 pixels (1.5 hectare).

The third main step was to merge the generalized binary layers. The binary layers were loaded according to a predefined priority sequence (see Table 2) and the layer representing the least important class was used as the starting layer. The remaining layers were incorporated into this layer according to priority, with the most important layer entered at the end. Water bodies had the highest priority and were added last to preserve streams and the outline of lakes. Finally, any pixel belonging to a group of identical pixels but smaller than a predefined threshold value was removed.
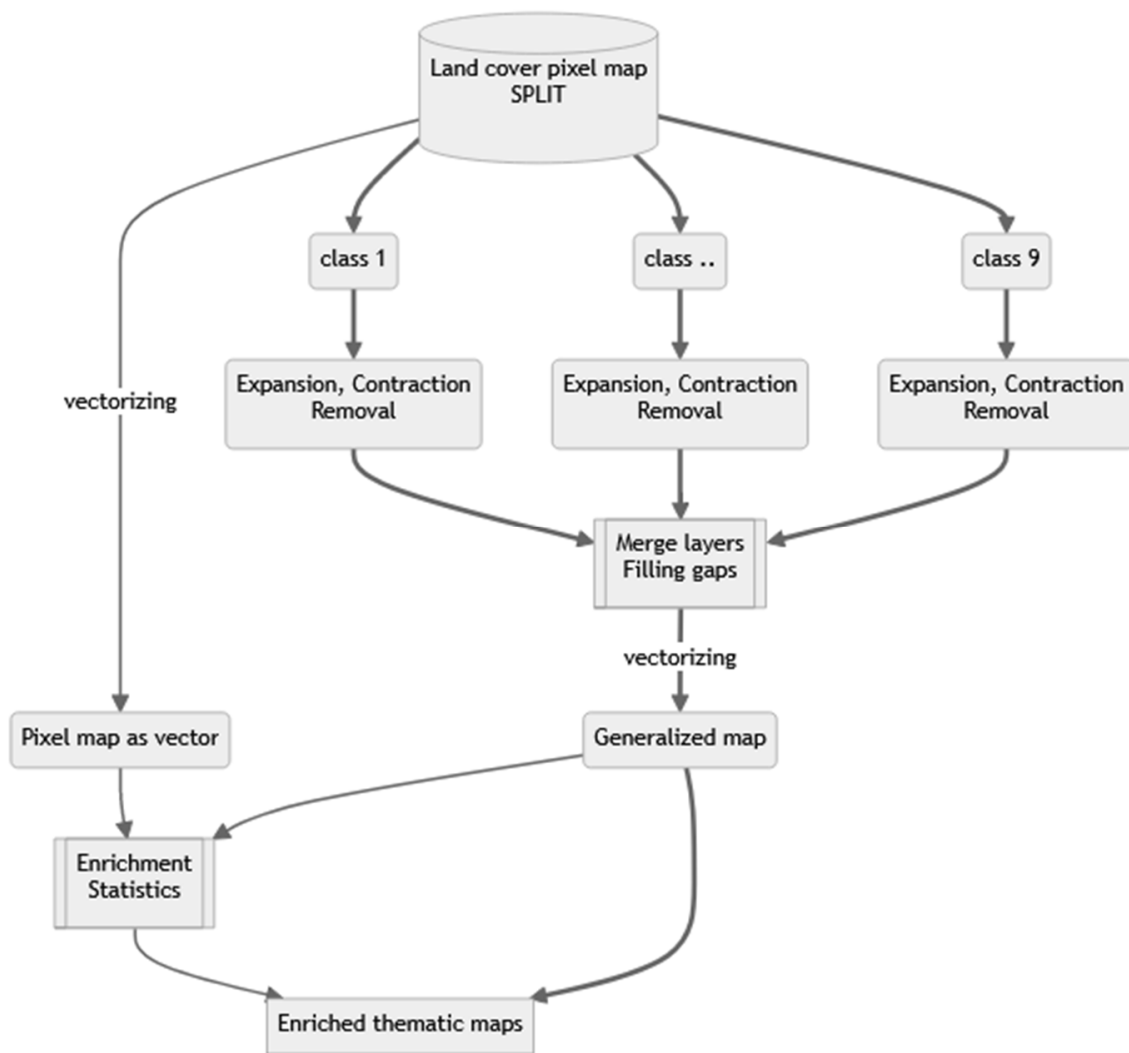
Figure 2. The generalization process. The content of the "Expansion, Contraction, Removal" step is described in more detail in Figure 3.

The resulting layer contained gaps, partly due to the merging process and partly because of the removal of small clusters of pixels. These gaps were filled by expanding classes from the edges using the value of the largest juxtaposed neighbor cluster. The gap filling was iterated until no gap remained.

The generalized map was vectorized. The result was a vector land cover dataset with the same classes as the original raster dataset, but with a simplified geometrical structure and larger, homogeneous areas.

An extract of the map of the study area before and after generalization is shown in Figures 4 and 5. These figures compare the cluttered pixel map, created via the classification of Sentinel imagery, to the generalized vector map. Figure 4 shows a larger section covering an area with two valleys, a forest and two mountains. Figure 5 shows the enlargement of a settlement and a surrounding agricultural area. Both figures show the cluttered pixel map at the bottom (Figures 4a and 5a) and the generalized land cover map of the same area at the top (Figures 4b and 5b). The effect of the generalization is not only visual, as seen in the maps, but it is also statistical. Table 2 is a summary of the area statistics generated from the two maps.
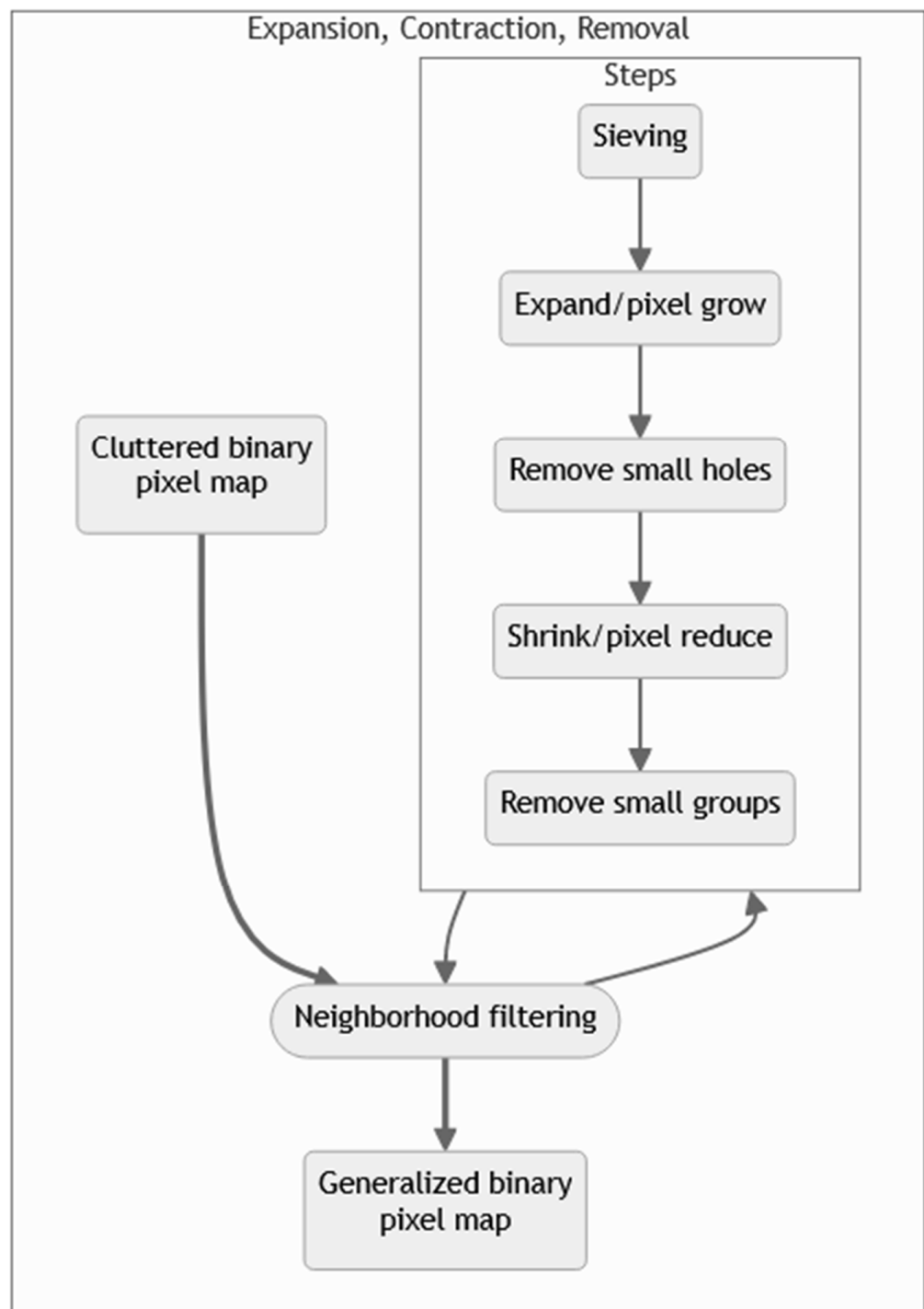
Figure 3. The detailed operations and workflow embedded in the "Expansion, Contraction and Removal" step of the generalization process that was described in Figure 2.

The generalization caused considerable change in the gross area of several classes (Table 2). The area covered by small classes like permanent herbaceous and sealed surfaces was reduced, despite the relatively high priority assigned to these classes. The priorities modified but did not change this general trend. There was a sizable (almost 40,000 hectare) growth in the area classified as woodland coniferous, irrespective of the low priority (priority 7) assigned to this class. Woodland broadleaved, permanent herbaceous, non- and sparse vegetation and mosses were reduced the most, probably because they do not constitute large, continuous areas.

Table 2. Area statistics generated from the land cover classification—raster map and the generalized vector map of Viken and Oslo. The column "Priority" shows the priority of each theme in the merging process.

| Code | Name | Priority | Hectare | | | % |
			Raster Map	Vector Map	Diff.	Diff.
1	Sealed surfaces	2	61,248	60,264	−984	−1.6
2	Woodland broadleaved	6	35,795	23,464	−12,331	−34.5
3	Woodland coniferous	7	740,892	779,906	39,015	5.3
4	Permanent herbaceous	4	17,313	6210	−11,103	−64.1
5	Periodically herbaceous	3	107,910	112,471	4561	4.2
6	Mosses	5	73,315	60,206	−13,109	−17.9
7	Water	1	113,242	111,602	−1641	−1.5
8	Low vegetation	8	47,100	47,189	89	0.2
9	Sparsely vegetated	9	9619	5122	−4497	−46.8
	Total		1,206,434	1,206,434	0.0	0.0

The distribution of pixels from the original land cover raster map found inside each class in the generalized vector map, as obtained via self-enrichment, is found in Table 3. Each row in the table summarizes to 100%, except for round-off errors. All classes in the generalized map are dominated by pixels from the same class, but some classes appear to be more purified than others. Water (class 7) is the purest class, probably because it has the highest priority in the merging process and has a high degree of accuracy in the pixel map classification. Woodland coniferous (class 3) is also a class containing limited amounts of noise. This is probably because the acreage of the class is large and the noise is small compared to the overall extent.

The proportion (in percent) of the polygons in each land cover class according to the amount of the polygon surface covered with pixels assigned to the same class is listed in Table 4. For example, 70% of the polygons in class 1, sealed surfaces, contain more than 80% pixels classified as sealed surface. Another 29% of the polygons in this land cover class have 50–80% coverage classified as sealed surface. Clearly, these polygons represent less homogeneously sealed areas. The final 1% of the polygons in this class contain less than 50% pixels classified as sealed surface.

Table 3. Content of the generalized land cover classes, shown as percentage of the area assigned to the class. See Table 2 for class names.

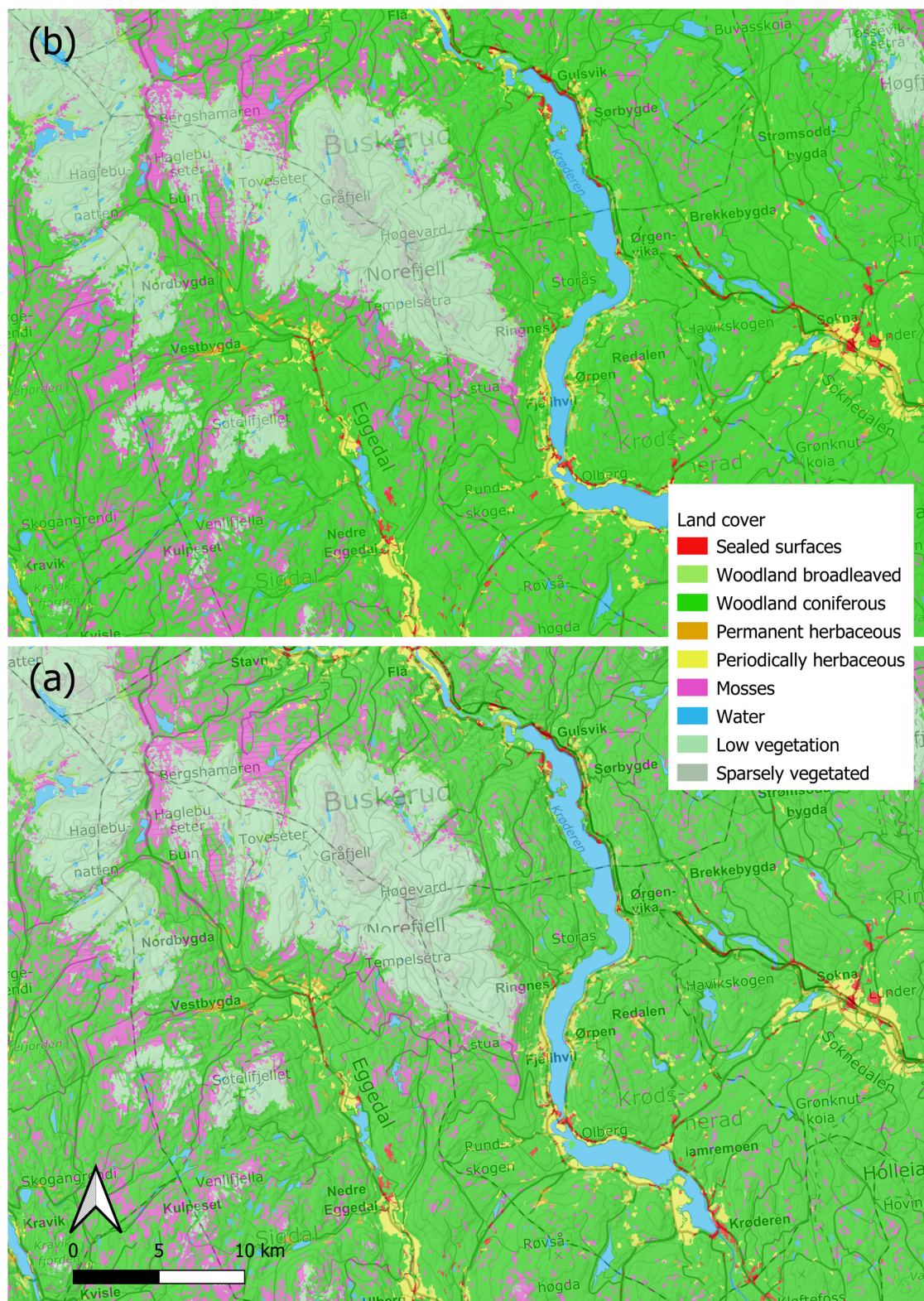| | | Percent Land Cover Content Based on the Original Cluttered Pixel Map | | | | | | | | |
		1	2	3	4	5	6	7	8	9
	1	80	3	7	1	7	1	<1	<1	<1
	2	1	78	11	4	4	1	<1	1	<1
	3	1	1	93	1	1	3	<1	<1	<1
Land cover class in the generalized map	4	1	6	3	77	10	2	<1	<1	<1
	5	4	3	2	5	85	1	<1	<1	<1
	6	1	1	8	1	2	78	<1	7	3
	7	<1	<1	<1	<1	<1	<1	100	<1	<1
	8	<1	2	2	<1	<1	5	<1	87	4
	9	1	1	1	<1	2	3	<1	4	90

Figure 4. Land cover map of an area around Norefjell mountain and lake Krøderen in former Viken county, Norway. The pixel clutter of the classified image ((**a**), bottom) and the generalized land cover map ((**b**), top). Location: 60.30 north, 9.46 east. Background: N50/Norge digital.
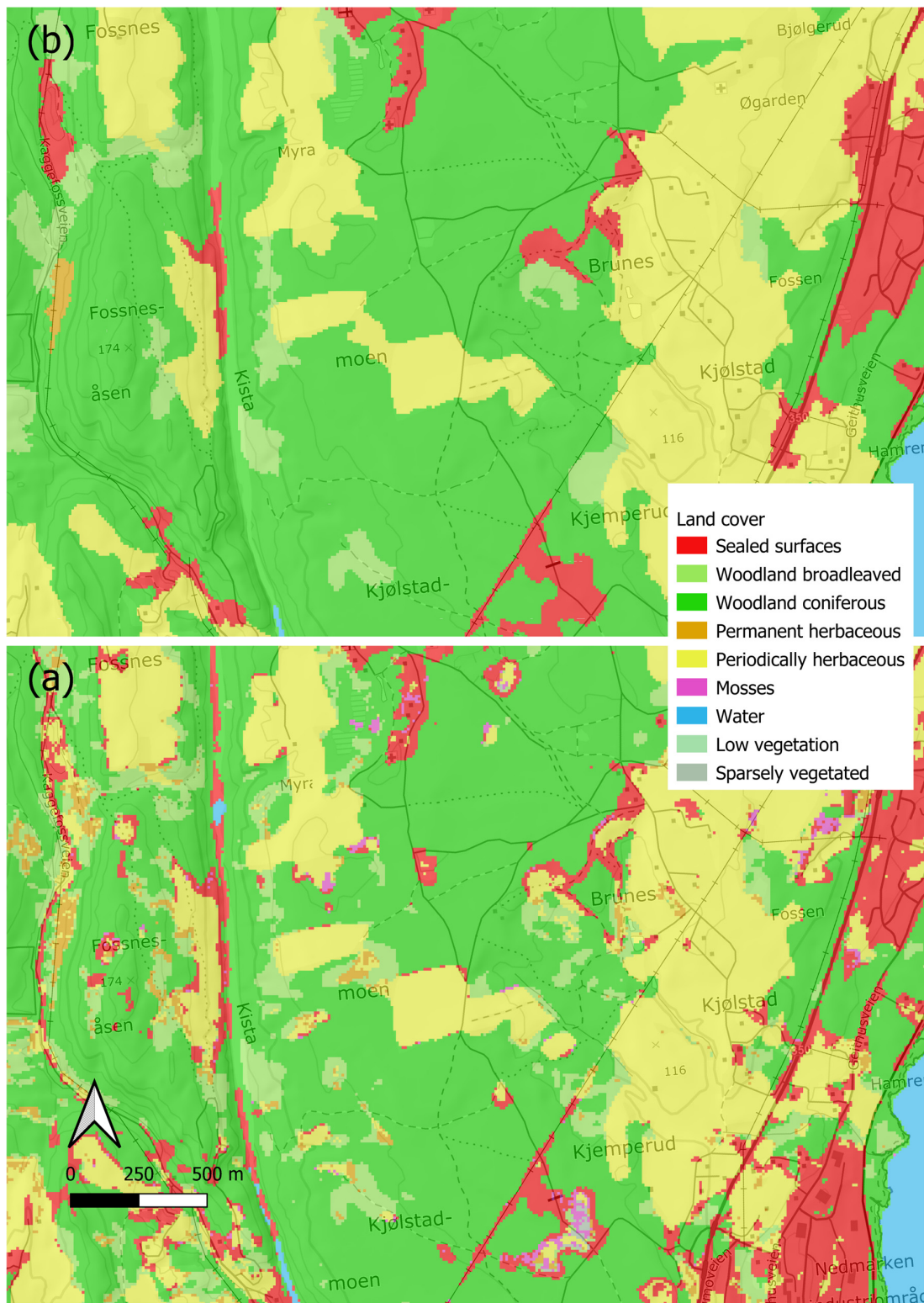
Figure 5. Land cover map of an area south of Vikersund in former Viken county, Norway. The pixel clutter of the classified image ((**a**), bottom) and the generalized land cover map ((**b**), top). Location: 59.96 north, 9.95 east. Notice that this region is outside (south of) the region shown in Figure 4. Background: N50/Norge digital.

Table 4. Proportion (%) of the polygons in each land cover class according to the amount of the polygon surface covered with pixels assigned to the same class.

Code	Name	Proportion of Polygons with Coverage		
		Over 80%	**50–80%**	**<50%**
1	Sealed surfaces	70	29	1
2	Woodland broadleaved	41	59	1
3	Woodland coniferous	100	0	0
4	Permanent herbaceous	42	56	2
5	Periodically herbaceous	84	16	0
6	Mosses	41	58	1
7	Water	100	0	0
8	Low vegetation	95	5	0
9	Sparsely vegetated	86	12	2

The land cover polygons were separated into 19 new classes. The homogenous polygons come from each of the nine original classes, while the heterogeneous polygons come from each of the eight classes containing heterogeneous polygons. Finally, a class with complex polygons emerges from all nine classes. The result was a richer, more diversified land cover map.

Other reclassifications are also possible. Thematic maps of individual land cover classes can be drawn with a symbology showing the amount of the surface belonging to that particular class. An example could be a thematic map of mosses, where each polygon is colored according to the proportion of the pixels inside the polygon that is classified as mosses.

The effect of the generalization on the overall accuracy of the map product was examined using 362 field observations. The overall accuracy of the land cover classification pixel map (before generalization) was 85.9% correctly classified. After generalization, the accuracy increased to 88.1%. The generalized map thus appeared to be more correct than the classified satellite pixel map. The difference was examined using a paired sample *t*-test. The increased accuracy (2.2%) was not significant (at the 95% level).

4. Discussion

The objective of this study was to develop and implement a practicable, mechanistic approach to the generalization of a land cover classification. The model-based split, expand and contract algorithm using morphological filters allowed a reasonable control of the generalization procedure, avoiding direct operator intervention. The parameters for expansion and contraction as well as the priority of each class could be determined anteriorly and the process was able to be implemented as a batch job. The entire procedure is reproducible. A visual comparison of the original pixel map and the generalized vector map (as in Figures 4 and 5) also showed (by subjective judgment) that the overall spatial structure of the map was preserved.

The generalization emphasized the broad spatial structures; the details were removed and the spatial statistics changed accordingly. Substantial areas of woodland broadleaved, permanent herbaceous vegetation, mosses and sparse vegetation were absorbed, mainly into the coniferous forest dominating in the area but also, to some extent, into patches of periodically herbaceous vegetation. This was probably due to the effect of the specific composition of land cover classes in this region. The relationship between the classes is most likely not the same in other regions where the spatial composition and structure of the classes is different. The extent of classes with a small or scattered presence was susceptible to decrease in the process, being encroached on by more dominant categories in their neighborhoods. Classes given low priority in the reassembling of the map are

most prone to such changes. The exchange of pixels between classes did not have a serious impact on the overall accuracy of the map but could affect the representation of specific features.

Figure 5b demonstrates a reduction in sealed surfaces (construction land) when compared to the original classified image (Figure 5a). This is evidenced by the disappearance of several road sections, e.g., in the center of the image. The observation raises questions about the limitations of these methods in addressing linear features. Water and sealed surfaces were both given high priority in the merging of the binary layers, but this is not an expedient measure when the goal is to preserve linear features.

Cartographic generalization can be graphical or semantic. The graphical approach is the simplification of geometry, filtering objects and curtailing symbols. The semantic approach is to aggregate units and create new classes. Our approach was essentially graphical.

Graphical generalization has statistical consequences (changing the number of points, the length of lines and the area of polygons). The total area of each land cover class changed, in some cases substantially. This statistical corollary can, as in this project, be managed through self-enrichment. A second objective of this study was therefore to preserve much of the original information and support the later analytical use of the land cover map via the self-enrichment of the resulting dataset. Self-enrichment was carried out by attaching statistical information from the original, cluttered map as an attribute vector.

Figures 4b and 5b also show that the generalization removes small features when compared to the cluttered originals (Figures 4a and 5a) and should not be applied when it is important to preserve small or narrow features in the map. The presence and extent of small features can also be documented through self-enrichment, where the information is represented as an attribute vector linked to each polygon in the generalized map, but the exact location of the features is lost. The decision to apply or not to apply generalization therefore depends on the information required in the final map product.

The self-enrichment of the generalized map allowed us to preserve the original statistical information and differentiate polygons according to composition. This is closely linked to the third objective of the study. We expected that the new, self-enriched map would provide for more diverse applications. This was demonstrated by the description of a new thematic map with more detailed land cover classes according to the homogeneity of the land cover in the polygons.

The example is an arbitrary result. It demonstrates that the expectation was correct, but does not explore the possibilities and limitations regarding new and more creative thematic maps that exploit the information in the self-enriched map. Further research is needed to explore this issue.

The final objective of the study was to examine the statistical distortion and possible bias introduced by the generalization. The classification of mosses and permanent herbaceous classes was found to be less accurate in the initial land cover classification [18]. The permanent herbaceous areas were often misclassified as periodically herbaceous. The self-enrichment also showed that just 42% of the generalized polygons in this class can be counted as pure, 56% as heterogeneous and 2% as complex. A similar pattern was observed for mosses, as by their nature they are quite heterogenous and difficult to classify using remote sensing, as they can be covered by forest or mixed with heathland [24]. The area of mosses decreased by nearly 14% during generalization. More attention needs to be paid to the effect of generalization on the heterogenous and complex land cover classes.

The comparison of the accuracy before and after generalization showed that the percentage that was correctly classified increased by 2.2% percent (from 85.9 to 88.1%) when the map was simplified through generalization. Obviously, the occasionally large changes in area statistics for some classes corrected some errors while creating others. An overall accuracy around 85 to 90% is, from our experience, common in land cover classification products derived from remote sensing. Misclassified pixels can be interpreted as commission errors (from the point of view of the assigned class) or omission errors (from the point of view of the class they should have been assigned to). Omission errors appearing

as noise in a classified image can to some extent be eliminated during generalization. The statistical effect of generalization is therefore not necessarily a corruption of the data since the original classification of the pixels is also prone to classification errors.

The dominating class in our study was woodland coniferous, a class that increased substantially during generalization. Omission errors were probably scattered misclassified pixels, small forest gaps and recent clear-cuts. These were absorbed by the surrounding forest class during the generalization process. Conifer forest, periodically herbaceous land and land with low vegetation were all improved by the generalization. Sealed surfaces and woodland broadleaved were slightly impaired. These were small, dispersed classes. The accuracy of the remaining classes remained unchanged. The important result is, in our opinion, that the generalization had only a minor impact on the overall accuracy of the map.

5. Conclusions

Generalization via filtering information and simplifying graphical representation to facilitate communication is a constituent part of cartography. It is also requisite with respect to visualization and user interaction in geographic information science [25]. Databases can hold large amounts of detailed, high resolution spatial data, but too much information can effectively impede decision making and users may be better off with a digested and simplified representation of the world. Generalization, supporting the use of the information at multiple scales, is needed to provide data for multiple purposes and tasks [26].

Generalization is a transformation undertaken to facilitate the visualization and interpretation of complex data through simplification. The generalization of the original classified pixel image in our study created a smoother, simplified map where details were removed. The result accentuated the broader spatial structures of the land cover in the region. Generalization, in this respect, takes a broader view, examining the data from a larger distance and with an eye for the whole rather than the parts.

This appraisal of the result is subjective and not substantiated by a formal investigation. The smoother, simplified map is regarded as an improved cartographic product by the project team. Other users, with different preferences or objectives, may assess the generalized map otherwise.

The study leaves several open questions to be addressed by systematic, structured and well-designed future studies. There is a need for a broad survey of how users representing different user communities assess the resulting map. The study should address the simplified map compared to the original cluttered map, as well as the possible benefit from the self-enrichment of the polygons. Another research issue is the need to compare generalization methodologies. There are several alternative approaches to the structural generalization of pixel maps, including filters and segmentation [27–29]. Our split-generalize and merge (SGM) approach based on a modification of the method from [23] was fast and reliable; however, further analysis is needed to compare the result to other structural techniques, with respect to the ease of implementation, information loss and user acceptance.

The generalization did not reduce the accuracy of the map. The measured accuracy increased slightly, but the improvement was not statistically significant. The example shows that the corruption of the map via generalization does not need to be a major concern. The study did, however, demonstrate a loss of linear features, in particular roads or transport lines among the sealed surfaces. Tools could be developed and added to preserve these features in situations where narrow linear features are an important component of the map.

The self-enrichment achieved by populating the polygons in the generalized map with statistics using the original pixel map maintained the statistical information from the original map and allowed for a more flexible use of the data by combining the land cover classification with the information about the assumed distribution of land cover classes inside the polygons. Further studies are needed to explore the new analytical and cartographic possibilities created by the self-enrichment of the generalized land cover map, irrespective of the generalization methodology used.

References

1. Touya, G.; Hoarau, C.; Christophe, S. Clutter and map legibility in automated cartography: A research agenda. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **2016**, *51*, 198–207. [CrossRef]
2. van der Knaap, W.G.M. The vector to raster conversion: (mis)use in geographical information systems. *Int. J. Geogr. Inf. Syst.* **1992**, *6*, 159–170. [CrossRef]
3. Spiess, E. The need for generalization in a GIS environment. In *GIS and Generalisation: Methodology and Practice*; Lagrange, J.-P., Weibel, R., Müller, J.-C., Eds.; Taylor and Francis: Oxfordshire, UK, 1995; pp. 31–46. ISBN 978-0748403196. [CrossRef]
4. Joao, E. *Causes and Consequences of Map Generalization*; Taylor and Francis: Oxfordshire, UK, 1998; ISBN 978-0748407767. [CrossRef]
5. Brassel, K.E.; Weibel, R. A review and conceptual framework of automated map generalization. *Int. J. Geogr. Inf. Syst.* **1998**, *2*, 229–244. [CrossRef]
6. Kronenfeld, B.J.; Buttenfield, B.P.; Stanislawski, L.V. Map generalization for the future: Editorial comments on the special issue. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 468. [CrossRef]
7. Bertin, J. *Semiology of Graphics*; The University of Wisconsin Press: Madison, WI, USA, 1983; ISBN 978-1589482616.
8. Wolf, E.B. Ontology-driven generalization of cartographic representations by aggregation and dimensional collapse. In *The Semantic Web—ISWC 2009*; Bernstein, A., Karger, D.R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., Thirunarayan, K., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5823, pp. 990–997. [CrossRef]
9. Steiniger, S.; Weibel, R. Relations among map objects in cartographic generalization. *Cartogr. Geogr. Inf. Sci.* **2007**, *34*, 175–197. [CrossRef]
10. Wästfelt, A. Ambiguous Use of Geographical Information Systems for the Rectification of Large-Scale Geometric Maps. *Cartogr. J.* **2020**, *57*, 209–220. [CrossRef]
11. Galanda, M. Modelling constraints for polygon generalization. In Proceedings of the 5th ICA Workshop on Progress in Automated Map Generalization, Paris, France, 28–30 April 2003; pp. 28–30.
12. Kelly, M.; Blanchard, S.D.; Kersten, E.; Koy, K. Terrestrial Remotely Sensed Imagery in Support of Public Health: New Avenues of Research Using Object-Based Image Analysis. *Remote Sens.* **2011**, *3*, 2321–2345. [CrossRef]
13. Shea, K.S.; McMaster, R.B. Cartographic generalization in a digital environment: When and How to generalize. In Proceedings of the Auto Carto 9th International Symposium on Computer-Assisted Cartography, Baltimore, Maryland, 2–7 April 1898; pp. 56–67.
14. Ruas, A.; Plazanet, C. Strategies for automated generalization. In Proceedings of the 7th International Symposium on Spatial Data Handling (Advances in GIS Research II), Delft, The Netherlands, 12–16 August 1996; Taylor & Francis: London, UK, 1996; p. 6.
15. Neun, M.; Weibel, R.; Burghardt, D. Data enrichment for adaptive generalisation. In Proceedings of the ICA Workshop on Generalisation and Multiple Representation, Leicester, UK, 20–21 August 2004; pp. 20–21.
16. Neun, M.; Steiniger, S. Modelling cartographic relations for categorical maps. In Proceedings of the 22nd International Cartographic Conference, A Coruña, Spain, 9–16 July 2005; Volume 11.
17. Steward, H.J. Cartographic Generalisation Some Concepts and Explanation. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **1974**, *11*, 1–50. [CrossRef]
18. Waśniewski, A.; Hościło, A.; Aune-Lundberg, L. The impact of selection of reference samples and DEM on the accuracy of land cover classification based on Sentinel-2 data. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101035. [CrossRef]
19. Strand, G.H. The Norwegian area frame survey of land cover and outfield land resources. *Nor. J. Geogr.* **2013**, *67*, 24–35. [CrossRef]
20. Bryn, A.; Strand, G.H.; Angeloff, M.; Rekdal, Y. Land cover in Norway based on an area frame survey of vegetation types. *Nor. J. Geogr.* **2018**, *72*, 131–145. [CrossRef]
21. Serra, J.; Vincent, L. An overview of morphological filtering. *Circuits Syst. Signal Process.* **1992**, *11*, 47–108. [CrossRef]
22. Dimov, D.; Palomar, J.M.; Ruiz, L.A. Automated generalization of land-use data with GIS-based programming. *Gis. Sci. Z. Geoinformatik* **2014**, *3*, 109–120.

23. Strand, G.H.; Moum, S.O. Compilation and evaluation of a small-scale land resource map. *Nor. J. Geogr.* **2000**, *54*, 148–156. [CrossRef]

24. Aune-Lundberg, L.; Strand, G.H. The content and accuracy of the CORINE Land Cover dataset for Norway. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *96*, 102266. [CrossRef]

25. Cartwright, W.; Crampton, J.; Gartner, G.; Miller, S.; Mitchell, K.; Siekierska, E.; Wood, J. Geospatial information visualization user interface issues. *Cartogr. Geogr. Inf. Sci.* **2001**, *28*, 45–60. [CrossRef]

26. Müller, J.C.; Weibel, R.; Lagrange, J.P.; Salgé, F. Generalization-state of the art and issues. In *GIS and Generalisation: Methodology and Practice*; Lagrange, J.P., Weibel, R., Müller, J.C., Eds.; Taylor & Francis: Oxfordshire, UK, 1995; pp. 3–17. ISBN 978-0748403196. [CrossRef]

27. Martinez-Casasnovas, J.A. A cartographic and database approach for land cover/use mapping and generalization from remotely sensed data. *Int. J. Remote Sens.* **2000**, *21*, 1825–1842. [CrossRef]

28. Smith, G.; Beare, M.; Boyd, M.; Downs, T.; Gregory, M.; Morton, D.; Brown, N.; Thomson, A. UK land cover map production through the generalisation of OS MasterMap®. *Cartogr. J.* **2007**, *44*, 276–283. [CrossRef]

29. Costa, H.; Carrão, H.; Bação, F.; Caetano, M. Combining per-pixel and object-based classifications for mapping land cover over large areas. *Int. J. Remote Sens.* **2014**, *35*, 738–753. [CrossRef]