

A GIS-Based Comparison of Vector and Raster Data for Landscape Environmental Assessment

Rizky Ahmad Laitupa ^{1,*}, Halim ², Ahmad Rifai ³, Heinrich Rakuasa ^{4,**}

¹ Departement of Regional and Rural Development Planning, IPB University, Indonesia

² Departement of Aquatic Resources Management, IPB University, Indonesia

³ Departement of Environmental Science, Faculty of Graduate School of Sustainable Development, Universitas Indonesia Jakarta, Indonesia

⁴ Departement of Geoecology, National Research Tomsk State University, Russian Federation

* Correspondence: rizkylaitupa10@gmail.com, heinrich.rakuasa@yandex.ru

Received : 30 June 2026

Accepted: 29 July 2026

Published: 31 July 2026

Abstract: Landscape environmental assessment using Geographic Information Systems (GIS) requires spatial data that accurately represent land-cover characteristics. However, differences in spatial resolution and vector-to-raster conversion methods can introduce quantitative and geometric distortions. This study compares vector and raster data representations for landscape environmental assessment in Tomsk City, Russia, focusing on the effects of conversion methods and spatial resolutions on land-cover area estimates. PlanetScope imagery with a native spatial resolution of 3 m, acquired on August 14, 2026, was classified using the Maximum Likelihood Classification (MLC) method in ArcGIS Pro into six land-cover classes: Built Area, Crops, Flooded Vegetation, Rangeland, Trees, and Water. The resulting vector data were converted into raster format using three cell-assignment methods: Cell Center, Maximum Area, and Maximum Combined Area, at spatial resolutions of 3 m, 30 m, and 300 m. The results show that the 30 m Cell Center method produced minimal area deviations, generally below 0.1% compared with the 3 m reference. In contrast, increasing the spatial resolution to 300 m generated substantial distortions, particularly for small and fragmented land-cover classes. Under the Maximum Area and Maximum Combined Area methods at 300 m, Flooded Vegetation decreased by 87.80%, Rangeland by 77.23%, Crops by 71.58%, and Built Area by 64.10%, whereas Water increased by 158.53%. These findings demonstrate that spatial resolution and cell-assignment methods strongly influence land-cover representation through class aggregation and the loss of small spatial features. The findings support SDG 11 and SDG 15 by informing spatially accurate urban planning and terrestrial land-cover monitoring, without directly measuring SDG performance.

Keywords: GIS; Land-Cover; Landscape; PlanetScope

1. Introduction

In landscape environmental assessment applications, such as calculating land-cover proportions, analyzing road-network density, and modeling riparian vegetation corridors, the combination of vector and raster data is often unavoidable. Rasterization redistributes polygon information into discrete cells and can alter area, shape, and topological properties, <https://jurnal.istekaisyiyah.id/index.php/ijsth>

particularly when cell size becomes large relative to the mapped features (Huo et al., 2022). In this study, land-cover information was extracted from high-resolution PlanetScope satellite imagery using the Maximum Likelihood Classification (MLC) algorithm implemented in ArcGIS Pro. MLC is a well-established parametric pixel-based classification method that assigns pixels to classes according to their spectral probability distributions (Dash et al., 2023; Rakuasa & Khromykh, 2025). Because heterogeneous urban landscapes commonly contain small features and mixed boundaries, the choice of spatial resolution can strongly affect their representation (Ganesh et al., 2023). Raster-to-vector and vector-to-raster conversion studies further show that conversion can alter quantitative LULC estimates (Auradkar et al., 2021), while vector-raster integration is relevant to environmental feature extraction (Chen & Qian, 2022).

To address conversion-induced distortion, previous research has compared Cell Center, Maximum Area, and Maximum Combined Area approaches across multiple cell sizes. Giofandi et al. (2023), for example, evaluated polygon-to-raster conversion at several resolutions and found that the Cell Center approach at a fine grid could closely reproduce vector-based area estimates. More recent rasterization research likewise shows that increasing cell size can produce measurable losses in area, shape, and topology, and that no single allocation rule is universally optimal (Huo et al. 2022; Han et al. 2026). The present study therefore uses a contrasting multi-scale scenario consisting of 3 m, 30 m, and 300 m cell sizes to examine how strongly land-cover area estimates change across micro-, meso-, and macro-scale representations. Related landscape-assessment studies demonstrate the importance of maintaining consistent spatial representation when LULC data are used for ecological-risk and urban-environment analyses (Zhao et al., 2020; Yankovich et al., 2023)

The availability of very-high-resolution land-cover data at approximately 3 m provides substantial spatial detail, but the present study does not directly quantify computational performance (Rakuasa et al., 2025). Although finer grids generally increase the number of cells and potential storage and processing requirements, empirical statements about CPU time, peak memory, or file size require direct measurements from the processing environment. Accordingly, computational efficiency is treated here as a methodological consideration rather than a measured outcome (Haag et al., 2020). The central analytical focus is the reproducible comparison of land-cover area estimates across conversion methods and cell sizes. (Oliveira et al., 2020)

This study aims to compare vector and raster representations of land cover for landscape environmental assessment in the Tomsk City study area, Russia. Specifically, it evaluates area deviations among three polygon-to-raster Cell Assignment Type methods at 3 m, 30 m, and 300 m cell sizes. The study's contribution is a multi-scale empirical comparison that distinguishes directly measured area distortion from geometric and computational indicators that require additional spatial or processing records. The findings are intended to provide a methodological reference for selecting rasterization configurations that preserve land-cover area information while making the limitations of coarse-scale representation explicit.

Relevance to Sustainable Development Goals. The study contributes methodologically to SDG 11 (Sustainable Cities and Communities) by improving the consistency of spatial information used in urban and peri-urban environmental assessment. It also supports SDG 15 (Life on Land) by highlighting how coarse rasterization can obscure small terrestrial and

riparian land-cover features. These are methodological contributions and do not constitute direct measurements of SDG target performance.

2. Methods

2.1. Research Location, Data, and Software

The research was conducted within the study boundary identified as Tomsk City, Russia (Figure 1). The mapped area covers a heterogeneous urban-taiga landscape containing built-up areas, the Tom River corridor, agricultural/open land, and forest vegetation. The boundary shown in Figure 1, rather than the administrative name alone, defines the analytical extent. Because the area values in the original table are substantially larger than the expected footprint of the mapped urban study boundary, the revised manuscript treats the numerical area values as dataset outputs that require final verification against the polygon boundary, projected coordinate reference system, and area-calculation units before publication. No unverified conversion to a different area unit is introduced here.

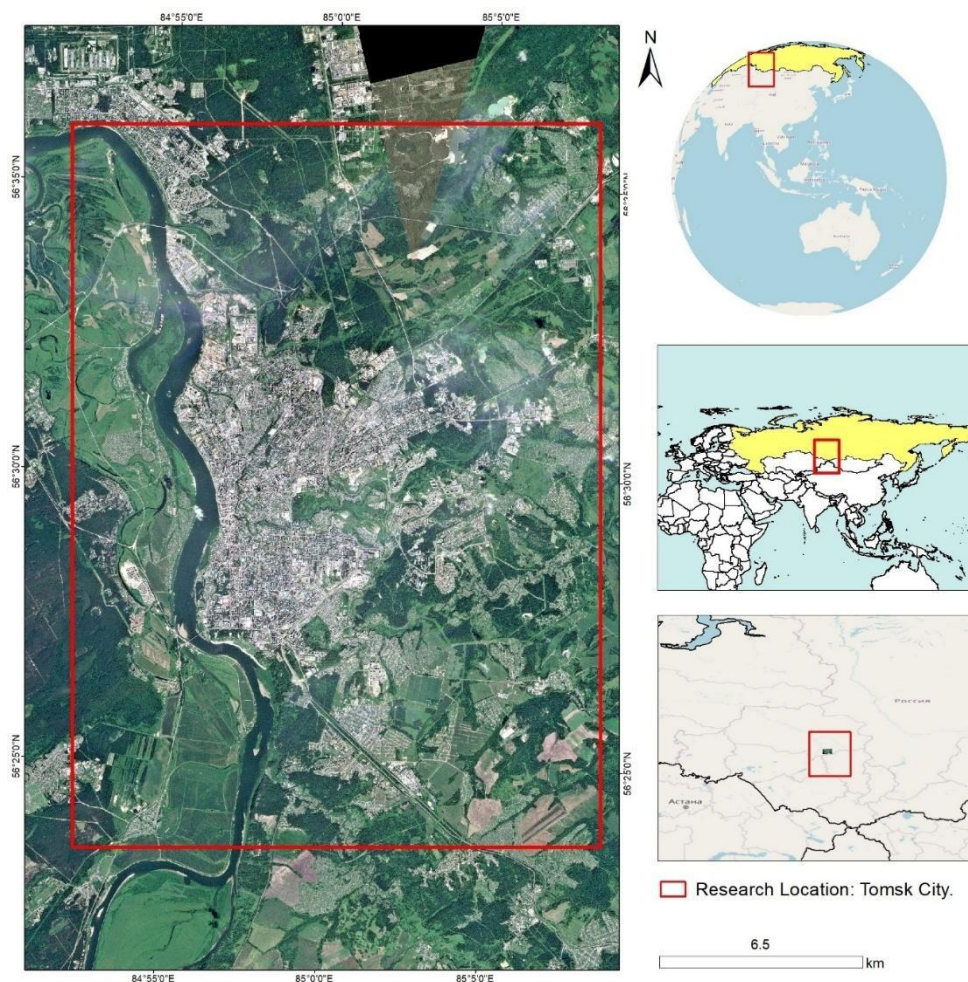


Figure 1. Study area: Tomsk City, Russia.

The primary spatial data were PlanetScope satellite imagery with a native spatial resolution of approximately 3 m, acquired on August 14, 2026. The imagery was obtained through the Planet Explorer/Data API provided by Planet Labs PBC. The analytical boundary was the study-area polygon shown in Figure 1. All area calculations should be performed in an appropriate projected coordinate reference system with linear units in metres, and the same study-area boundary and processing extent should be retained across all rasterization

scenarios. PlanetScope has also been applied for detailed land-cover transformation analysis in heterogeneous landscapes (Quevedo et al., 2023).

Spatial data processing was conducted using ArcGIS Pro for image preprocessing, MLC classification, polygon-to-raster conversion, raster-to-polygon comparison, and area extraction. The rasterization scenarios were designed so that the input polygon layer, class/value field, study-area boundary, processing extent, and raster alignment remained constant, while Cell Assignment Type and cell size were varied. For reproducibility, the ArcGIS Pro environment settings should use a common processing extent and a common snap raster/alignment reference across scenarios (Singla et al., 2021). The manuscript does not claim direct computational-performance measurements because CPU time, peak memory, and file-size records were not retained in the analytical results.

$$g_i = -12 \ln |i| - 12 (x - \mu_i)^T i^{-1} (x - \mu_i) + \ln P(i) \quad (1)$$

where x is the spectral value vector of a pixel, μ_i is the spectral mean of class i , i is the covariance matrix of class i , and P_i is the prior probability of class i . Each pixel is assigned to the class with the highest discriminant-function value:

$$\omega(x) = \arg \max_i g_i(x) \quad (2)$$

Thus, each pixel is classified into the land-cover class with the highest membership probability based on its spectral characteristics.

2.2. Land Cover Classification and Research Workflow

Land-cover extraction from the 3-m PlanetScope imagery was performed using the Maximum Likelihood Classification (MLC) method in ArcGIS Pro. The MLC algorithm is a parametric, pixel-based classification method that assigns each pixel to a class based on the estimated probability of membership from training samples. Six land-cover classes were used: Rangeland, Built Area, Trees, Crops, Flooded Vegetation, and Water. The resulting polygon layer was treated as the vector baseline for subsequent area comparisons. Because the validity of all rasterization comparisons depends on the baseline classification, the manuscript should report the classification-accuracy assessment associated with the MLC training and validation data if such information is available.

The research workflow began with acquisition and preprocessing of 3-m PlanetScope imagery captured on August 14, 2026. Pixel-based classification was then performed using MLC in ArcGIS Pro to produce the six land-cover classes. The classified polygons were converted to raster using Cell Center, Maximum Area, and Maximum Combined Area at cell sizes of 3 m, 30 m, and 300 m. The 3-m Cell Center output was used as the operational raster reference for the reported percentage comparisons. This designation is analytical and should not be interpreted as proof that the 3-m raster is geometrically identical to the original vector layer.

In the Cell Center method, the cell value is determined by the polygon containing the cell's center point:

$$R(C) = L(P_i), \text{ if Center}(C) \in P_i \quad (3)$$

In the Maximum Area method, the cell value is determined by the polygon with the largest intersecting area within the cell:

$$R(C) = L(P_k), k \arg \max_i A(C \cap P_i) \quad (4)$$

In the Maximum Combined Area method, the intersecting areas of polygons belonging to the same class are first combined:

$$A_j(C) = \sum_{i: L(P_i)=j} A(C \cap P_i) \quad (5)$$

The class with the largest combined area is then assigned as the cell value:

$$R(C) = j^*, j^* = \arg \max_j A_j(C) \quad (6)$$

Each conversion method was tested at three cell sizes: 3 m, 30 m, and 300 m. The primary evaluation indicator was quantitative area distortion, calculated as the percentage difference between each raster-derived class area and the 3-m Cell Center reference. Geometric effects such as boundary stair-stepping and feature loss were interpreted qualitatively from the mapped outputs. Centroid displacement, boundary-shape indices, and other geometric distances were not calculated because the required centroid and geometry comparison dataset was not retained in the reported analytical table. Computational efficiency was likewise not treated as a measured indicator because processing time, peak memory, and file-size measurements were not available. These limitations are explicitly reflected in the revised objectives, Results, Discussion, and Conclusions. The use of regular grids for polygonal statistics also emphasizes the need for consistent cell geometry and processing extent when comparing rasterized areas (Haag et al., 2020).

2.3. Polygon-to-Raster Conversion Method Testing and Spatial Resolution Scenarios

The testing of Cell Assignment Type during Polygon to Raster conversion compared three allocation approaches. Cell Center assigns the class of the polygon containing the centre of each raster cell. Maximum Area assigns the class associated with the largest intersecting polygon area within the cell (Congalton, 1997). Maximum Combined Area first aggregates intersecting areas belonging to the same class and then assigns the class with the largest combined area. The three methods can produce different results in cells containing several polygons of different classes (Chen & Qian, 2022). However, if the polygon with the largest individual intersection is also from the class with the largest combined intersection, the two area-based methods can produce the same class assignment. Therefore, identical outputs are possible for a particular dataset without implying that the two algorithms are mathematically identical (Huang et al., 2024).

Each conversion method was tested at three cell-size levels: 3 m, 30 m, and 300 m. The 3-m cell size represents the native spatial detail of the PlanetScope input; 30 m represents a meso-scale generalization; and 300 m represents a deliberately coarse macro-scale scenario. Throughout the experiment, the study-area boundary, class/value field, processing extent, and raster alignment should remain constant so that differences can be attributed to Cell Assignment Type and cell size. The term “cell size” is used consistently in the remainder of the manuscript to avoid ambiguity between spatial resolution and pixel dimension.

A comparative evaluation of the nine scenarios (3 Cell Assignment Types $\times$ 3 cell sizes) was conducted using quantitative area deviation as the primary metric. Visual inspection of the maps was used to describe boundary stair-stepping, loss of small patches, and class aggregation, but these observations were not converted into numerical geometric metrics. The equality of Maximum Area and Maximum Combined Area results was retained as an observed dataset-specific result. Because the source table alone cannot verify the full ArcGIS environment configuration, the final reproducibility check should confirm the input feature layer, value field, priority field, snap raster, processing extent, cell size, and coordinate reference system in the ArcGIS Pro project.

3. Results and Discussion

3.1. Results

3.1.1. Baseline Vector Land-Cover Classification Results (3 m PlanetScope MLC)

The Maximum Likelihood Classification (MLC) based on 3 m-resolution PlanetScope imagery produced six main land-cover classes: Built Area, Crops, Flooded Vegetation, Rangeland, Trees, and Water. Based on quantitative area statistics, the 3 m Cell Center scenario produced areas of 34,555.60 ha for Built Area, 39,156.81 ha for Crops, 1,821.81 ha for Flooded Vegetation, 70,029.29 ha for Rangeland, 766,316.84 ha for Trees, and 16,799.76 ha for Water. With a total area of 928,680.10 ha, Trees was the dominant class, whereas Flooded Vegetation had the smallest area. The dominance of Trees indicates that the land-cover structure in the study area is strongly influenced by woody vegetation or forested areas. Land-cover areas are presented in Table 1.

Table 1. Land-Cover Areas in 2026

LULC Class	Area (ha)	Proportion (%)
Built Area	34,555.60	3.72
Crops	39,156.81	4.22
Flooded Vegetation	1,821.81	0.20
Rangeland	70,029.29	7.54
Trees	766,316.84	82.52
Water	16,799.76	1.81

Spatially, the distribution of the six classes represents a heterogeneous landscape consisting of built-up areas, vegetation, open land or rangeland, agricultural areas, water bodies, and vegetation associated with aquatic environments (Duan et al., 2026). In the context of the urban landscape of Tomsk, the Built Area class represents built surfaces, while Trees and Rangeland constitute the vegetation components, which are quantitatively much larger than the other classes. Flooded Vegetation and Water represent landscape components closely associated with river corridors and wetland areas. However, detailed spatial patterns, such as the Tom River corridor, patch fragmentation, and the distribution of class boundaries, need to be presented using the classification map because the spatial attribute table primarily provides quantitative information on area.

The classification results subsequently served as the baseline for evaluating changes in area after the vector polygon data were converted into raster format at 30 m and 300 m resolutions. In this study, the 3 m Cell Center scenario was used as the operational reference because it represents the highest-resolution raster in the analytical dataset. The comparison

shows that changes in area at 30 m resolution were relatively small when using the Cell Center method, whereas the changes became substantially greater at 300 m resolution. Therefore, the land-cover statistics indicate that increasing pixel size, or degrading spatial resolution, causes the representation of land-cover areas to become increasingly distorted from the reference condition at 3 m resolution, particularly for classes with smaller areas or narrow spatial dimensions (Auradkar et al., 2021).

3.1.2. Evaluation of Quantitative Land-Cover Area Distortion

Comparison of the nine scenarios revealed clear differences in area estimates across cell sizes and Cell Assignment Types. Using Cell Center, the change from 3 m to 30 m was very small: Built Area decreased by 2.62 ha (0.0076%), Crops decreased by 4.56 ha (0.0117%), Flooded Vegetation increased by 1.68 ha (0.0925%), Rangeland increased by 4.39 ha (0.0063%), Trees increased by 92.29 ha (0.0120%), and Water decreased by 0.54 ha (0.0032%). These results show that the 30-m Cell Center configuration closely reproduced the 3-m reference in terms of class area, although this does not establish equivalent geometric boundaries.

In contrast, the change from 3 m to 300 m under Cell Center produced greater area deviations. Built Area decreased by 1,615.60 ha (4.68%), Crops by 1,977.81 ha (5.05%), Flooded Vegetation by 345.81 ha (18.98%), and Rangeland by 2,682.29 ha (3.83%). Trees increased by 6,297.16 ha (0.82%), while Water increased by 1,011.24 ha (6.02%). Flooded Vegetation therefore showed the largest relative sensitivity among the Cell Center classes, consistent with the greater vulnerability of small and narrow features to coarse cell sizes.

Table 2. Percentage change in land-cover area from 3 m to 30 m cell size, relative to the 3-m Cell Center reference

LULC Class	Cell Center (%)	Maximum Area (%)	Maximum Combined Area (%)
Built Area	-0.0076	-6.6281	-6.6281
Crops	-0.0117	-7.9961	-7.9961
Flooded Vegetation	0.0925	-21.0946	-21.0946
Rangeland	0.0063	-9.3586	-9.3586
Trees	0.0120	1.3955	1.3955
Water	-0.0032	9.7418	9.7418

The Maximum Area and Maximum Combined Area methods produced identical tabulated values at every tested cell size in this dataset. This equality is interpreted as a dataset-dependent outcome. In cells containing multiple polygons, the class occupying the largest individual polygon intersection may also be the class with the largest combined intersection of all polygons of that class. Under that condition, both allocation rules assign the same class. The equality should therefore not be interpreted as evidence that the algorithms are intrinsically equivalent. For reproducibility, the ArcGIS Pro project should be checked to confirm that the same input polygon layer, class/value field, processing extent, cell alignment or snap raster, and cell-size settings were used, with only the Cell Assignment Type changed between the two scenarios.

At 300 m, the changes became more extreme: Built Area decreased by 64.10%, Crops by 71.58%, Flooded Vegetation by 87.80%, and Rangeland by 77.23%, while Trees increased by 10.25% and Water increased by 158.53%. These results indicate a tendency toward class

aggregation as pixel size increases, particularly the expansion of Trees and Water at the expense of other classes. Small or narrow land-cover patches are more likely to be excluded or absorbed during rasterization when their areas are insufficient to dominate a larger cell. ArcGIS documentation notes that, under the Maximum Area method, a polygon generally must occupy the largest portion of a cell and in some conditions at least 50% of the cell to determine the cell value; otherwise, the cell may be assigned to another class or NoData.

Table 3. Percentage change in land-cover area from 3 m to 300 m cell size, relative to the 3-m Cell Center reference.

LULC Class	Cell Center (%)	Maximum Area (%)	Maximum Combined Area (%)
Built Area	-4.68	-64.10	-64.10
Crops	-5.05	-71.58	-71.58
Flooded Vegetation	-18.98	-87.80	-87.80
Rangeland	-3.83	-77.23	-77.23
Trees	0.82	10.25	10.25
Water	6.02	158.53	158.53

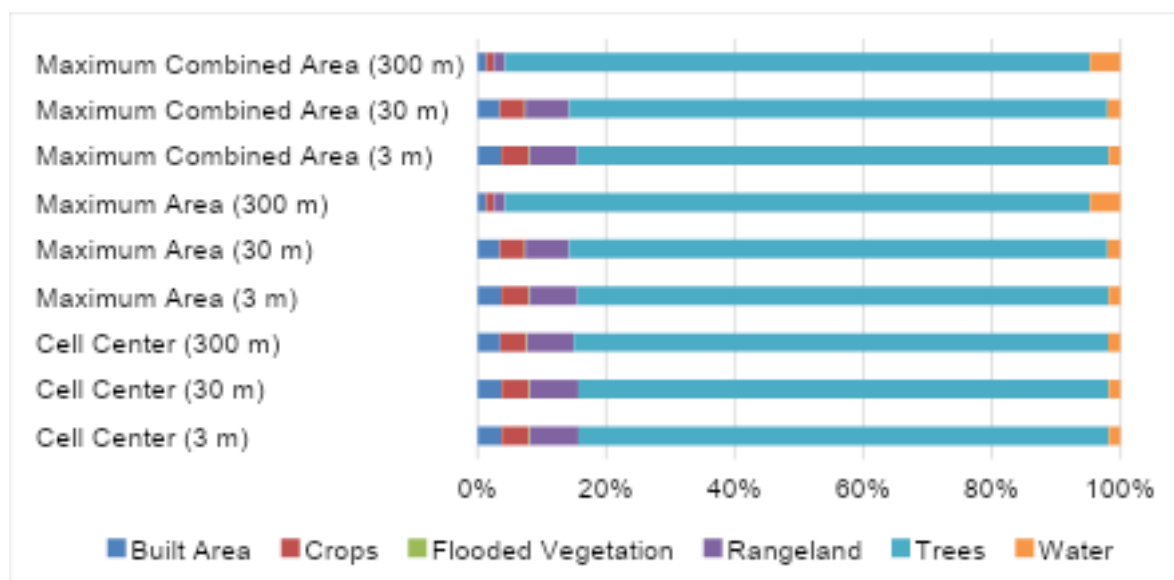


Figure 2. Comparison of land-cover class areas across Cell Assignment Types and cell sizes.

Overall, the results demonstrate that quantitative distortion is influenced by both spatial resolution and the pixel-allocation rule. The Cell Center method preserved the baseline area most effectively at 30 m, whereas Maximum Area and Maximum Combined Area generated greater changes, particularly at 300 m (Figure 2). The results also show that classes with small, fragmented, or narrow spatial distributions especially Flooded Vegetation are more vulnerable to area loss during raster conversion, while larger or spatially dominant classes tend to absorb neighboring areas (Dimitriadou & Nikolakopoulos, 2022).

3.1.3. Qualitative Assessment of Spatial Representation and Boundary Effects

Changes in spatial resolution indicate an increased potential for geometric distortion as pixel size increases from 3 m to 30 m and 300 m. According to the table, using the Cell Center method, the Built Area remains relatively stable, from 34,555 ha at 3 m to 34,553 ha at 30 m, but decreases to 32,940 ha at 300 m. A clearer pattern is observed for Flooded Vegetation,

which changes from 1,822 ha at 3 m to 1,823 ha at 30 m, then drops to 1,476 ha at 300 m. These changes indicate that the 300 m resolution has a lower ability to preserve the spatial detail of relatively small objects. In contrast, the Trees class maintains a very large area: 766,317 ha, 766,409 ha, and 772,614 ha at 3 m, 30 m, and 300 m resolutions, respectively. This differential response among classes indicates that the effect of increasing pixel size is not uniform and depends strongly on the size, shape, and fragmentation level of surface features (Figure 3,4,5).

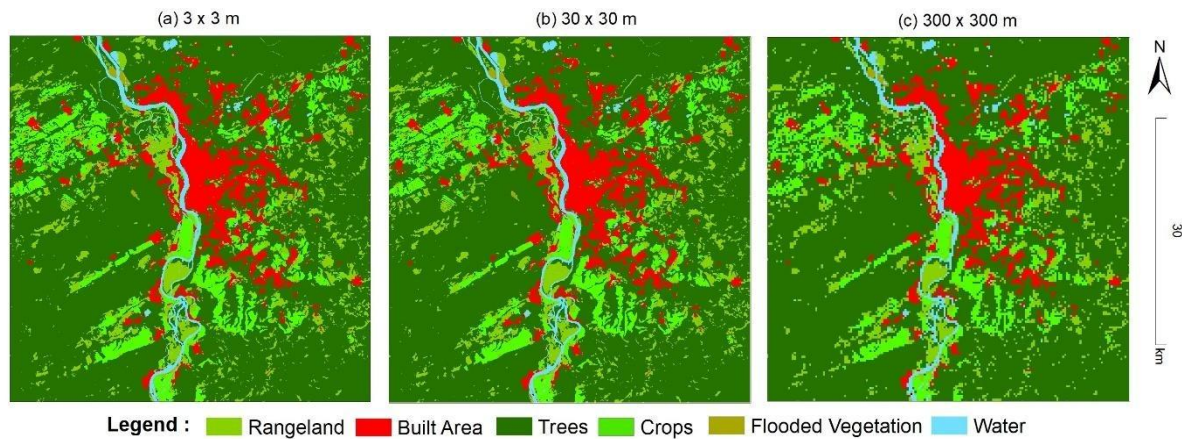


Figure 3. Land-cover representation at 3 m, 30 m, and 300 m cell sizes using Cell Center

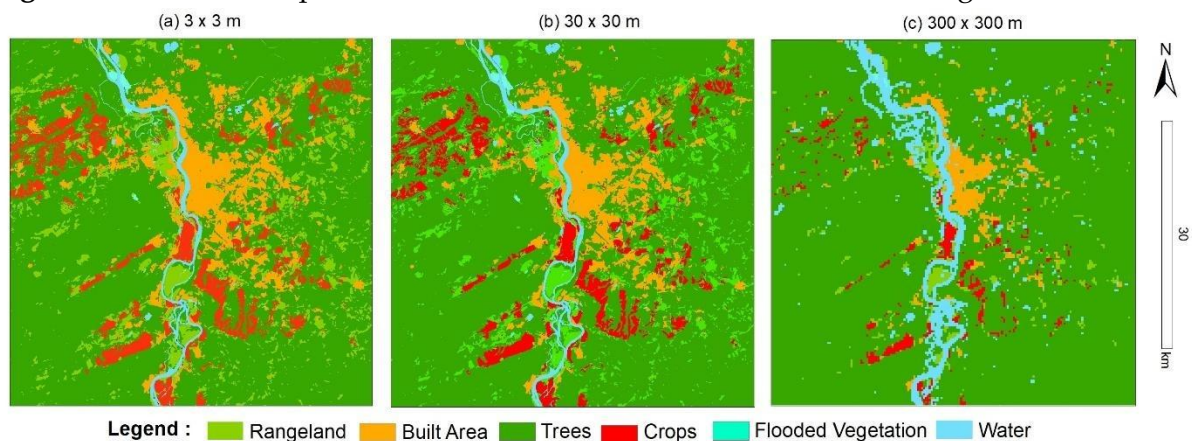


Figure 4. Land-cover representation at 3 m, 30 m, and 300 m cell sizes using Maximum Area.

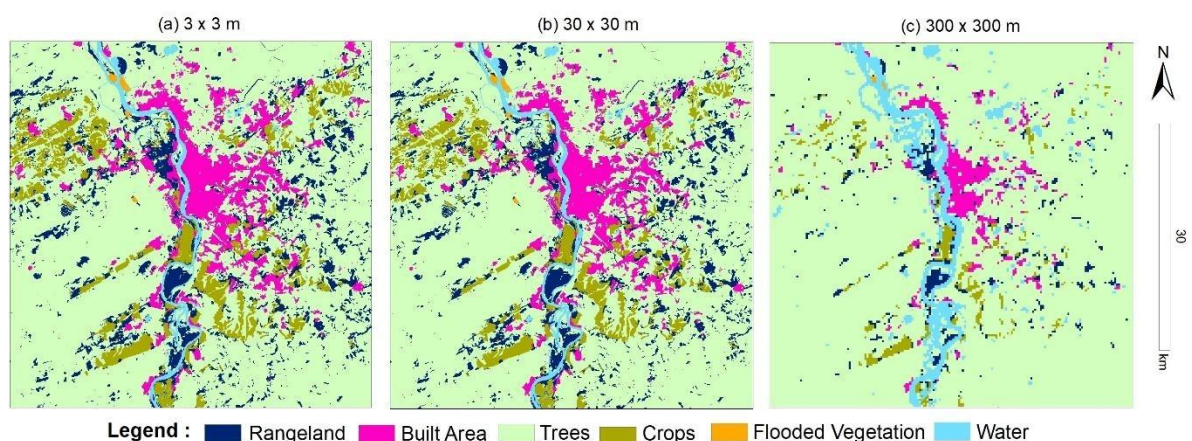


Figure 5. Land-cover representation at 3 m, 30 m, and 300 m cell sizes using Maximum Combined Area.

The cut-cell phenomenon and loss of small polygons become more evident in the Maximum Area and Maximum Combined Area methods, especially at 300 m resolution. For example, in Maximum Area, Built Area decreases from 34,296 ha at 3 m to 32,023 ha at 30 m, <https://jurnal.istekaisyiyah.ac.id/>

and further to only 12,312 ha at 300 m. Crops also show a strong decline from 38,787 ha to 35,685 ha, and then to 11,025 ha, while Flooded Vegetation changes from 1,771 ha to 1,397 ha and drops to only 216 ha at 300 m resolution. At the same time, Trees increase from 767,550 ha at 3 m to 778,262 ha at 30 m and 846,198 ha at 300 m. This pattern indicates a shift in cell allocation at class boundaries as pixel size increases, so that portions of smaller or more fragmented classes may be reassigned to the class dominating the cell (Figure 3,4,5). Interestingly, the Maximum Combined Area values are identical to the Maximum Area values across all classes and resolutions in this dataset, meaning both methods yield the same spatial response based on the available conversion results (Haag et al., 2020).

Boundary shift (positional shifting) can be inferred from changes in area across resolutions, but centroid shift values cannot be directly calculated from this table because it provides only pixel counts and areas, without centroid coordinates or the geometry of the converted polygons. Nevertheless, the area changes provide strong evidence of shifts in position and boundary shape at coarser resolutions. The most extreme changes occur in the Maximum Area/Maximum Combined Area methods at 300 m resolution, particularly for Built Area, which decreases by approximately 64.1% compared to 3 m resolution; Crops by about 71.6%; and Flooded Vegetation by about 87.8%. In contrast, Trees increase by approximately 10.2%, indicating that larger cells tend to allocate more area to the dominant class. Thus, based on the available table, increasing pixel size from 3 m to 300 m results in progressively greater changes in spatial representation, especially for smaller and more fragmented classes. To obtain quantitative centroid shift values (in meters), further analysis would need to use the original polygon layer and the raster-to-polygon converted polygons to compute centroid-to-centroid distances for each Cell Assignment Type (Teng et al., 2021).

3.2. Discussion

Evaluation of the Cell Assignment Types indicates that Cell Center produced the most stable class-area estimates when the cell size increased from 3 m to 30 m. Deviations for all classes remained below 0.1% relative to the 3-m reference. This stability indicates that, for this dataset, sampling the class at the cell centre preserved aggregated class areas better than the two area-dominance approaches at 30 m. It should not, however, be interpreted as proof that Cell Center always preserves geometric boundaries more accurately. This interpretation is also consistent with prior GIS conversion research emphasizing the sensitivity of landscape measures to rasterization rules and scale (Auradkar et al., 2021). The stability of Cell Center at 30 m is consistent with the idea that a fine-to-intermediate grid can preserve aggregate area when class boundaries are sampled at cell centres. Published work shows that rasterization error depends on cell size and polygon characteristics, and that area, shape, and topology can respond differently to the same conversion process (Huo et al., 2022). In the present dataset, the empirical evidence is limited to area deviation and therefore supports a conclusion about area preservation rather than a universal ranking of geometric accuracy. Related work on raster and vector processing likewise emphasizes that computational and spatial consequences should be evaluated together when conversion is used in environmental workflows (Silva-Coira et al., 2020).

Maximum Area and Maximum Combined Area exhibited larger quantitative deviations as cell size increased. At 30 m, the Maximum Area result showed a 9.36% reduction in Rangeland and a 1.40% increase in Trees, while at 300 m the corresponding changes reached -77.23% and +10.25%. The identical outputs of Maximum Area and Maximum Combined

Area in this dataset are best understood as the consequence of coincident class rankings within the raster cells. This result highlights the importance of documenting ArcGIS Pro environment settings and inspecting cell-level class intersections when two allocation methods unexpectedly produce identical outputs. The same issue is relevant to topology-sensitive generalization, where changes in cell size can affect spatial structure beyond simple area totals (Duan et al., 2026). At 300 m, Cell Center also produced substantial area deviations, including an 18.98% reduction in Flooded Vegetation and a 6.02% increase in Water. These results indicate that the advantage of Cell Center at 30 m does not extend to a deliberately coarse 300-m grid. The dominant control is therefore not the allocation rule alone, but the interaction between cell size and the spatial structure of the land-cover polygons. Based on area deviation, the 30-m Cell Center configuration most closely approximated the 3-m reference in this dataset.

The 3-m cell size provides the finest spatial detail in the analytical dataset and retains more small landscape features than the coarser scenarios. At 30 m, the Cell Center method maintained very small class-area deviations, whereas the Maximum Area and Maximum Combined Area methods produced larger changes. At 300 m, all methods showed substantial generalization, particularly for small or elongated classes. These results are consistent with studies showing that rasterization errors increase with cell size and are strongly affected by polygon characteristics (Duan et al., 2026; Huo et al., 2022). The broader literature on rasterization also demonstrates that cell size can control the preservation of spatial information in categorical land-cover maps (Díaz-Varela et al., 2010; Huo et al., 2022). The 30-m scenario represents a meso-scale generalization rather than a direct substitute for the original 3-m classification. For broad land-cover proportions in this dataset, Cell Center at 30 m produced deviations below 0.1%. However, this numerical stability should not be interpreted as preservation of every boundary or small feature, because no quantitative boundary or centroid metric was calculated. The distinction between area accuracy and geometric accuracy is important when selecting a rasterization scale. At 300 m, spatial generalization becomes much stronger. Maximum Area and Maximum Combined Area produced area losses of up to 87.80% for Flooded Vegetation, 77.23% for Rangeland, 71.58% for Crops, and 64.10% for Built Area, while Water increased by 158.53%. These changes are consistent with coarse-grid aggregation and the loss of minority classes. The 300-m outputs should therefore not be used for analyses that depend on accurate representation of small or fragmented land-cover features in the study area.

At the macro-scale of 300 m, spatial information degradation becomes extremely pronounced, and the resulting area distortion is no longer tolerable for precision analysis. The use of Maximum Area at 300 m resulted in an area loss of up to 87.80% for Flooded Vegetation, a 77.23% reduction for Rangeland, a 71.58% decrease for Crops, and a 64.10% reduction for Built Area, while the Water class exhibited an artificial expansion of 158.53%. These massive distortions occur because 300 m pixels absorb small patches and narrow linear corridors, transforming the heterogeneous land-cover pattern of Tomsk City into coarsely generalized homogeneous pixel blocks. Spatial accuracy and computational efficiency involve a general trade-off, but computational efficiency was not directly measured in this study. The 3-m scenario contains substantially more cells than the 30-m and 300-m scenarios for the same extent, so a reduction in cell size will normally increase data volume and processing requirements. However, because the current dataset does not contain empirical CPU time, peak RAM, or raster file-size measurements, no numerical claim about computational savings

is made (Rakuasa & Budnikov, 2024). Future work should record these parameters directly if computational efficiency is to be evaluated as a formal performance indicator.

Based on the measured area deviations alone, the Cell Center scenario at 30 m represents the closest approximation to the 3-m reference in this dataset, with all class deviations below 0.1%. It may therefore be considered an area-preserving compromise for analyses where 30-m cells are acceptable. The term “sweet spot” is avoided because computational cost and geometric fidelity were not measured directly. A complete optimization among accuracy, geometry, and computational cost would require additional empirical metrics. In contrast, simplifying the data to a 300 m resolution provides very high computational efficiency through a drastic reduction in the pixel matrix and file size. However, this efficiency comes at the cost of losing critical spatial information, degrading landscape boundaries, and reducing the representation of minority patches through the cut-cell phenomenon (Duan et al. 2026). Although the current quantitative dataset does not record empirical computational-performance parameters such as CPU run time, peak RAM usage, and raster file size, the quantitative changes in area demonstrate that decisions regarding spatial-resolution degradation must carefully consider the risk of spatial-information distortion (Chen & Qian, 2022).

Based on the quantitative evaluation of area deviations, the Cell Center scenario at 30 m resolution emerges as the reference sweet spot that most closely approximates the 3 m baseline. This scenario provides an optimal compromise by minimizing area-estimation errors to less than 0.1% while substantially reducing the computational burden compared with the 3 m raster. Nevertheless, determining an overall operational sweet spot for landscape analysis in urban Tomsk still requires direct verification using spatial-geometric metrics, such as centroid displacement and aliasing, as well as empirical measurements of computational costs (Duan et al. 2026). The findings have methodological implications for environmental assessment within the mapped Tomsk City study area. At 300 m, Maximum Area and Maximum Combined Area produced large reductions in Built Area, Crops, Flooded Vegetation, and Rangeland and increases in Trees and Water. Such changes can alter the apparent composition of an urban-taiga landscape and may bias subsequent analyses that depend on class proportions. The magnitude of this effect is especially important for small or elongated features associated with river corridors and fragmented vegetation.

Flooded Vegetation showed the highest relative sensitivity, with an area reduction of 87.80% under the 300-m Maximum Area and Maximum Combined Area scenarios. This result indicates that coarse rasterization can remove or absorb small wetland and riparian features. Therefore, coarse-resolution outputs should be used cautiously for local ecological assessment, corridor mapping, or analyses where the spatial persistence of small classes is important. These implications are based on observed area changes and visual map interpretation, not on independently measured ecological impacts. For analyses requiring consistency with the 3-m reference, the quantitative results support the use of Cell Center at 30 m in this dataset because all reported class deviations were below 0.1%. The 300-m outputs should be reserved for applications where strong spatial generalization is acceptable. Selection of a final operational resolution should also consider the intended analytical scale and should be verified against the study-area boundary and projected CRS before the reported areas are interpreted as administrative statistics. The study therefore distinguishes three levels of evidence: quantitative area deviation is directly measured; geometric effects are qualitatively interpreted from mapped outputs; and computational efficiency is discussed

conceptually rather than empirically measured. This distinction reduces the risk of over-interpreting the available results and provides a clearer basis for reproducibility in future vector-to-raster assessments.

Conclusions

This study demonstrates that polygon-to-raster Cell Assignment Type and cell size affect land-cover area representation in the Tomsk City study area. The 3-m Cell Center raster served as the operational reference, while increasing cell size produced progressively greater area deviations, particularly for small, narrow, and fragmented classes. Cell Center at 30 m produced deviations below 0.1% for all six classes relative to the 3-m reference. At 300 m, Maximum Area and Maximum Combined Area produced the largest changes, including an 87.80% decrease in Flooded Vegetation, a 77.23% decrease in Rangeland, a 71.58% decrease in Crops, and a 64.10% decrease in Built Area, while Water increased by 158.53%. The identical Maximum Area and Maximum Combined Area outputs are interpreted as a dataset-dependent result and should be reproduced only after confirming identical input, value-field, extent, snap-raster/alignment, and cell-size settings in ArcGIS Pro. The study does not claim quantitative centroid displacement, boundary-distance metrics, or computational-performance measurements because those data were not available. Instead, geometric effects are reported qualitatively and computational efficiency is identified as a future measurement need. Finally, the reported total area in Table 1 must be reconciled with the mapped study boundary and projected CRS before the values are used as definitive statistics for Tomsk City. Within these limitations, the results provide a reproducible basis for selecting rasterization settings according to the required analytical scale and tolerance for area distortion.

Acknowledgments

We extend our gratitude to all parties who have assisted in this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Auradkar, P. K., Raykar, A., Agarwal, I., Sitaram, D., & R., M. (2021). Accuracy assessment and performance analysis of raster to vector conversions on LULC data – India. *Journal of Engineering, Design and Technology*, 20(6), 1787–1809. <https://doi.org/https://doi.org/10.1108/JEDT-04-2021-0224>
- Chen, G., & Qian, H. (2022). Extracting Skeleton Lines from Building Footprints by Integration of Vector and Raster Data. *ISPRS International Journal of Geo-Information*, 11(9). <https://doi.org/10.3390/ijgi11090480>
- Congalton, R. G. (1997). Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion. *Photogrammetric Engineering and Remote Sensing*, 63(4), 425–434.
- Dash, P., Sanders, S. L., Parajuli, P., & Ouyang, Y. (2023). Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data in an Agricultural Watershed. *Remote Sensing*, 15(16). <https://doi.org/10.3390/rs15164020>
- Dimitriadou, S., & Nikolakopoulos, K. G. (2022). Development of the Statistical Errors Raster Toolbox with Six Automated Models for Raster Analysis in GIS Environments. *Remote Sensing*, 14(21). <https://doi.org/10.3390/rs14215446>
- <https://jurnal.istekaisyiyah.ac.id/>

-
- Duan, Y., Zhao, R., Xu, X., & Zhang, J. (2026). Topology-Aware Field Parcel Delineation: Bridging Deep Semantic Features and Geometric Constraints. *Remote Sensing*, 18(11). <https://doi.org/10.3390/rs18111783>
- Ganesh, B., Vincent, S., Pathan, S., & Benitez, S. R. G. (2023). Utilizing LANDSAT data and the Maximum Likelihood Classifier for Analysing Land Use Patterns in Shimoga, Karnataka. *Journal of Physics: Conference Series*, 2571(1), 12001. <https://doi.org/10.1088/1742-6596/2571/1/012001>
- Giofandi, E., Munibah, K., Kraugusteeliana, K., Novalinda, A., & Sekarrini, C. (2023). The Comparison of Vector and Raster Data for The Calculation of Landscape Environment Using a Geographic Information System Approach. *IT Journal Research and Development*, 7(2), 209–219. <https://doi.org/10.25299/itjrd.2023.10878>
- Haag, S., Tarboton, D., Smith, M., & Shokoufandeh, A. (2020). Fast summarizing algorithm for polygonal statistics over a regular grid. *Computers & Geosciences*, 142, 104524. <https://doi.org/https://doi.org/10.1016/j.cageo.2020.104524>
- Han, B., Qi, X., Qi, X., & Wang, Y. (2026). Refining Public DEMs for Urban Waterlogging Simulation via Vector–Raster Integration. *Remote Sensing*, 18(7). <https://doi.org/10.3390/rs18071080>
- Huang, Y., Zheng, M., Li, T., Xiao, F., & Zheng, X. (2024). An Integrated Framework for Landscape Indices' Calculation with Raster–Vector Integration and Its Application Based on QGIS. *ISPRS International Journal of Geo-Information*, 13(7). <https://doi.org/10.3390/ijgi13070242>
- Huo, X., Zhou, C., Xu, Y., & Li, M. (2022). A methodology for balancing the preservation of area, shape, and topological properties in polygon-to-raster conversion. *Cartography and Geographic Information Science*, 49(2), 115–133. <https://doi.org/10.1080/15230406.2021.1991478>
- Oliveira, A., Lopes, A., & Niza, S. (2020). Local climate zones classification method from Copernicus land monitoring service datasets: An ArcGIS-based toolbox. *MethodsX*, 7, 101150. <https://doi.org/https://doi.org/10.1016/j.mex.2020.101150>
- Pacheco Quevedo, R., Velastegui-Montoya, A., Montalván-Burbano, N., Morante-Carballo, F., Korup, O., & Daleles Rennó, C. (2023). Land use and land cover as a conditioning factor in landslide susceptibility: a literature review. *Landslides*, 20(5), 967–982. <https://doi.org/10.1007/s10346-022-02020-4>
- Rakuasa, H., & Budnikov, V. V. (2024). Assessing Settlement Suitability Using Road Network Analysis for Sustainable Urban Planning in Ambon City, Indonesia. *Applied Engineering, Innovation, and Technology*, 1(2), 68–74. <https://doi.org/10.62777/aeit.v1i2.39>
- Rakuasa, H., & Khromykh, V. V. (2025). Simulation of Urban Growth in Ternate Island using Cellular Automata Markov Chains Models. *Asian Journal of Environmental Research*, 2(1), 101–114. <https://doi.org/10.69930/ajer.v2i1.310>
- Rakuasa, H., Ahmat Rifai, A. R., Latulanit, A. K., Amahoru, A. H., Laitupa, R. A., Diaz Astiza, D. A., & Rajani, L. M. (2025). Detection of Land Cover Change in Nickel Mining Areas, in Weda Tengah Sub-district, Halmahera Island, Indonesia using Planet-Scope Satellite Images. *Frontiers in Sustainable Science and Technology*, 2(1), 41–50. <https://doi.org/10.69930/fsst.v2i1.379>

-
- Silva-Coira, F., Paramá, J. R., Ladra, S., López, J. R., & Gutiérrez, G. (2020). Efficient processing of raster and vector data. *PLoS ONE*, 15(1), 1–35. <https://doi.org/https://doi.org/10.1371/journal.pone.0226943>
- Singla, S., Eldawy, A., Diao, T., Mukhopadhyay, A., & Scudiero, E. (2021). Experimental Study of Big Raster and Vector Database Systems. *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2243–2248. <https://doi.org/10.1109/ICDE51399.2021.00231>
- Teng, D., Baig, F., Sun, Q., Kong, J., & Wang, F. (2021). IDEAL: a Vector-Raster Hybrid Model for Efficient Spatial Queries over Complex Polygons. *2021 22nd IEEE International Conference on Mobile Data Management (MDM)*, 99–108. <https://doi.org/10.1109/MDM52706.2021.00024>
- Yankovich, E. P., Yankovich, K. S., & Baranovskiy, N. V. (2023). Dynamics of Forest Vegetation in an Urban Agglomeration Based on Landsat Remote Sensing Data for the Period 1990–2022: A Case Study. *Remote Sensing*, 15(7). <https://doi.org/10.3390/rs15071935>
- Zhao, D.-M., Jiao, Y.-M., Wang, J.-L., Ding, Y.-P., Liu, Z.-L., Liu, C.-J., Qiu, Y.-M., Zhang, J., Xu, Q.-E., & Wu, C.-R. (2020). Comparative performance assessment of landslide susceptibility models with presence-only, presence-absence, and pseudo-absence data. *Journal of Mountain Science*, 17(12), 2961–2981. <https://doi.org/10.1007/s11629-020-6277-y>
-

CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International).

This license allows users to share and adapt an article, even commercially, as long as appropriate credit is given and the distribution of derivative works is under the same license as the original. That is, this license lets others copy, distribute, modify and reproduce the Article, provided the original source and Authors are credited under the same license as the original.

