AGILE: GIScience Series, 7, 25, 2026. https://doi.org/10.5194/agile-giss-7-25-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

A Fractional Raster–Vector Clipping Operator for Boundary-Consistent Multi-Scale Spatial Analysis

Younes Hamdani [1,2,3], Urs Albert Treier [1,2,3], and Signe Normand [1,2,3]

[1]Section for Ecoinformatics & Biodiversity, Department of Biology, Aarhus University, Aarhus, Denmark
[2]Center for Landscape Research in Sustainable Agricultural Futures (Land-CRAFT), Aarhus University, Aarhus, Denmark
[3]Center for Sustainable Landscapes Under Global Change (SustainScapes), Department of Biology, Aarhus University, Aarhus, Denmark

Correspondence: Younes Hamdani (younes.hamdani@bio.au.dk)

Abstract.

Raster–vector clipping is a fundamental operation in geospatial analysis. In practice, however, it is typically implemented as a binary rule that assigns pixels as either fully inside or outside a target geometry. Although widely used, clipping is rarely defined as a generic spatial operator with explicit inclusion semantics. This paper introduces a Fractional Raster–Vector Clipping Operator that models pixel inclusion through proportional areal overlap and formalizes clipping as a mathematically specified operator. The formulation is independent of any specific software environment and is accompanied by a generic algorithm. The algorithm avoids costly universal containment tests by restricting explicit geometric evaluation to a boundary region derived from pixel dimensions. The operator is implemented natively in PostgreSQL/PostGIS and evaluated using Denmark's national land-cover dataset (Basemap04, 2021) aggregated across multiple spatial resolutions. The results show that conventional binary clipping produces systematic and resolution-dependent underestimation relative to the fractional reference. Explicit boundary semantics therefore establish a consistent and resolution-aware basis for raster–vector aggregation in multi-scale spatial analysis.

Submission Type. Spatial Databases and Data Management; Data fusion and uncertainty reduction.

BoK Concepts. [AM4] Basic analytical operations; [AM2] Query operations and query languages; [AM2-3] Spatial queries.

Keywords. Raster–Vector Integration; Fractional Clipping; Boundary Effects; Multi-Resolution Analysis; Areal Weighting; Spatial databases

1 Introduction

Raster–vector integration is a core operation in GIScience and underpins a wide range of environmental and ecological analyses (Goodchild et al., 2007; Cova, 2016; Hamdani et al., 2021). Estimating land-cover composition within administrative units, quantifying habitat extent inside protected areas, computing vegetation indicators per watershed, or aggregating environmental variables for municipal reporting all require overlaying raster fields with vector geometries (Galton, 2001; Cova and Goodchild, 2002; Cova, 2016; Hamdani et al., 2023). In practice, this integration is most commonly performed through binary clipping rules implemented in GIS software and spatial databases, where a raster cell is either fully included or fully excluded based on a geometric predicate (e.g., center-in-polygon). While computationally efficient, this binary decision model implicitly assumes that boundary effects are negligible, an assumption that becomes increasingly problematic at coarser resolutions or along complex geometries.

For ecologists and environmental scientists, the implications of this simplification are non-trivial. Many indicators such as total habitat area, proportional land-cover composition, or categorical transitions along edges are sensitive to boundary configurations (Goodchild, 2011; Olofsson et al., 2014). When a raster cell partially overlaps a polygon, treating it as entirely inside or outside introduces systematic over- or under-estimation. The magnitude and direction of this distortion depend on spatial resolution, polygon morphology (e.g., fragmentation and perimeter-to-area ratio), and the spatial distribution of thematic classes near boundaries. Changes in spatial resolution and aggregation have been shown to alter spatial statistics and structural metrics in spatial data (Nalej et al., 2026), while mixed focal–zonal approaches emphasize the challenge of capturing heterogeneity in aggregated raster statistics (Ren et al., 2025). As raster resolution decreases, each pixel represents a larger spatial support, and misclassification at edges propagates into aggregated statistics, with documented scale dependence of error across multiple components (Rigge et al., 2025). Consequently, boundary bias is not merely a technical artifact but a scale-dependent source of uncertainty in spatial ecological inference.

Despite the centrality of raster–vector overlays, the inclusion semantics of partially intersecting pixels remain weakly formalized in database-centric GIS workflows. Most spatial database systems expose binary clipping operators that evaluate pixel inclusion through universal containment predicates, yet do not quantify the degree of pixel–polygon overlap. In remote sensing and areal interpolation research, sub-pixel and fractional approaches are well established; however, these methods are often implemented as specialized preprocessing routines rather than as formally defined, reusable operators integrated into general-purpose geospatial systems. This separation limits transparency, reproducibility, and conceptual clarity in large-scale analytical pipelines.

In this paper, we treat raster–vector clipping as a generic geospatial operator with explicit mathematical semantics. We introduce a *Fractional Raster–Vector Clipping Operator* that defines pixel inclusion as a continuous function of areal overlap between a raster cell and a target geometry. We implement the operator natively in PostgreSQL/PostGIS and evaluate its analytical behavior using national Danish land-cover data (2021) (Levin, 2022) aggregated over municipal boundaries. Two experiments are conducted. First, we quantify total area bias across multiple spatial resolutions by comparing the standard center-based clipping rule against a fully fractional reference model. Second, we assess class-specific boundary bias by examining how categorical land-cover areas vary across thematic classes and resolutions. These experiments allow us to disentangle scale effects from thematic heterogeneity and to demonstrate that boundary bias is both resolution-dependent and class-sensitive.

The contribution of this work is threefold. (i) We formalize fractional raster–vector clipping as a generic geospatial operator with explicit mathematical semantics. (ii) We provide an implementation-agnostic algorithm that transforms universal inclusion tests into boundary-localized intersection evaluation and demonstrate its realization as a database-native operator. (iii) We empirically quantify how boundary semantics influence land-cover aggregation across spatial resolutions and thematic classes. The aim of such work is to advance boundary-consistent raster–vector integration and to strengthen the theoretical and practical foundations of scale-aware spatial analysis in GIScience.

The remainder of this paper is structured as follows. Section 2 reviews related work on raster–vector integration, scale effects, and overlap-aware methods. Section 3 introduces the formal model of the fractional clipping operator, including its mathematical definition and inclusion semantics. Section 4 presents the implementation of the operator in a database environment, outlining how the method is applied in practice and discussing its computational behavior. Section 5 describes the experimental design and evaluates the impact of boundary semantics across spatial resolutions and thematic classes. Finally, Section 6 concludes the paper and outlines directions for future research.

2 Related Work

Raster–vector overlay and zonal statistics are fundamental operations in GIS and spatial analysis. They are implemented in desktop software, spatial databases, geospatial libraries, and distributed processing frameworks for large-scale spatial data (Yu et al., 2015; Singla and Eldawy, 2018). In most environments, raster clipping is based on binary per-pixel inclusion rules, most commonly center-in-polygon or all-touched criteria, whereby each pixel is treated as either entirely inside or entirely outside a target geometry. Although zonal tools frequently provide a “majority” option, this refers to selecting the most frequent raster value among the cells assigned to a zone and therefore concerns thematic aggregation rather than geometric inclusion. Majority statistics operate after pixels have been included, whereas clipping rules determine whether and how pixels contribute in the first place. Binary inclusion strategies provide computational efficiency and scalability, but they discretize spatial support and disregard partial pixel overlap. Consequently, boundary effects arise, yet these effects are rarely formalized at the level of the clipping operator itself.

The broader GIScience literature has extensively examined scale effects and aggregation sensitivity, most prominently through the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). Numerous studies demonstrate that spatial resolution and zoning influence aggregated statistics and thematic proportions (Chen et al., 2022). However, MAUP research primarily addresses variations in aggregation units and zoning configurations rather than the geometric interaction between raster cells and vector boundaries. As a result, the contribution of partial cell overlap to area bias is typically embedded within general discussions of scale dependence rather than treated as a distinct computational mechanism.

In remote sensing, sub-pixel and soft-classification approaches explicitly represent fractional class membership within pixels. Spectral unmixing and continuous field mapping methods estimate proportional land-cover fractions to account for intra-pixel heterogeneity (Shi and Wang, 2014). While these approaches enhance thematic representation at the classification stage, they are generally not integrated into generic raster–vector overlay operators used for zonal aggregation across arbitrary geometries. Fractional representation at the classification stage therefore does not automatically ensure boundary-consistent aggregation during subsequent spatial analysis.

Per-pixel fractional and area-weighted extraction methods that compute exact pixel–polygon overlap have been developed in specialist libraries and within the areal interpolation literature. For example, the `exactextractr` package in R provides precise overlap-based weighted summaries for raster data (Baston et al., 2021). Similarly, efficient algorithms have been proposed to compute exact pixel–polygon coverage while restricting geometric intersection tests to boundary cells (Xie et al., 2017). These contributions provide computationally rigorous and practically useful solutions for overlap-aware extraction and rasterization. However, they are generally implemented as procedural extraction or transformation workflows rather than as formally specified clipping operators with explicit and controllable inclusion semantics. In particular, they do not define clipping itself as a parameterized spatial operator that makes the relationship between geometric overlap and categorical inclusion analytically explicit.

Across database systems, geospatial libraries, and distributed processing environments, raster–vector clipping therefore remains predominantly implemented through implicit binary inclusion rules. While weighted aggregation can be achieved through custom intersection procedures, clipping itself is seldom defined as a mathematically specified spatial operator with explicit boundary interpretation. In contrast, we formalize clipping as a parameterized operator in which pixel inclusion is governed directly by proportional geometric overlap. The proposed fractional raster–vector clipping operator models pixel–polygon interaction explicitly and supports threshold-based inclusion, independent of any specific software environment.

3 Formal Model

This section presents a mathematical formulation of raster–vector clipping as an operator with explicit and controllable inclusion semantics.

3.1 Spatial Setting

Let $G \subset \mathbb{R}^2$ denote a planar polygon representing a target spatial unit. Let the raster domain $R \subset \mathbb{R}^2$ be partitioned into a finite collection of non-overlapping rectangular pixels $\mathcal{R} = \{P_i\}_{i \in I}$ such that

$$R = \bigcup_{i \in I} P_i, \qquad P_i \cap P_j = \varnothing \ \text{ for } i \neq j. \tag{1}$$

Each pixel P_i has constant width w, height h, and area

$$\mathrm{Area}(P_i) = w \cdot h. \tag{2}$$

Raster–vector clipping seeks to determine how the spatial support of each pixel P_i should be interpreted relative to the target polygon G when computing aggregated quantities. The essential challenge lies in specifying inclusion semantics for pixels that may only partially intersect G, thereby determining how geometric overlap is translated into area estimates and categorical aggregation in a resolution-consistent manner.

3.2 Fractional Pixel Overlap

To capture partial pixel support, we define the fractional overlap between a pixel P_i and the geometry G as

$$f(P_i, G) = \frac{\mathrm{Area}(P_i \cap G)}{\mathrm{Area}(P_i)}. \tag{3}$$

The function $f(P_i, G)$ defined in Equation 3 quantifies the proportion of the pixel's spatial support that lies inside G and satisfies

$$0 \leq f(P_i, G) \leq 1. \tag{4}$$

The three geometrically distinct cases are:

- $f(P_i, G) = 0$ if $P_i \cap G = \varnothing$,
- $f(P_i, G) = 1$ if $P_i \subseteq G$,
- $0 < f(P_i, G) < 1$ if P_i partially intersects ∂G.

This definition separates geometric support from thematic content: $f(P_i, G)$ depends only on spatial intersection and provides the structural basis for boundary-consistent aggregation. Figure 1 illustrates the conceptual workflow of the fractional overlap construction.

3.3 Thresholded Inclusion Operator

While fractional overlap provides a continuous measure of pixel support, many analytical workflows require categorical inclusion. We therefore define a parameterized inclusion operator

$$\mathcal{T}_\tau(P_i; G) = \begin{cases} 1 & \text{if } f(P_i, G) \geq \tau, \\ 0 & \text{otherwise,} \end{cases} \qquad \tau \in [0, 1]. \tag{5}$$

The parameter τ controls the minimum proportion of overlap required for inclusion. Special cases clarify its semantics:

- $\tau = 0$ includes all pixels intersecting G,
- $\tau = 1$ includes only pixels fully contained in G,

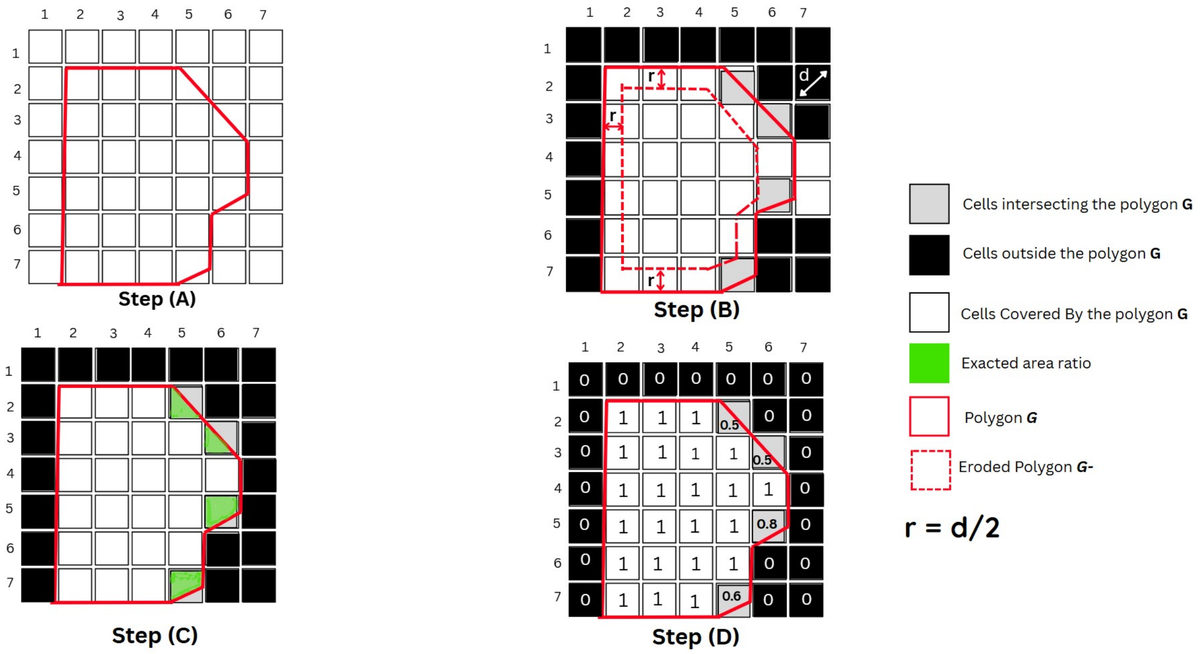

Figure 1. Conceptual workflow of the fractional raster–vector clipping operator: (Step A) input polygon G and raster grid $\mathcal{R}$; (Step B) interior–boundary classification using erosion radius r and eroded geometry G^-; (Step C) computation of areal overlap fractions $f(P_i, G)$ for boundary pixels; (Step D) generation of the fractional raster clipped to G according to a user-defined threshold τ. For example, if $\tau = 0.6$, only pixels satisfying $f(P_i, G) \geq 0.6$ are retained, together with fully contained pixels for which $f = 1$. In the illustrated case, pixels with indices $(5,7)$ and $(6,5)$ are included in Step D, as their respective overlap fractions are 0.6 and 0.8, in addition to all fully contained pixels with $f = 1$.

- $0 < \tau < 1$ defines intermediate inclusion rules based on minimum areal support.

Conventional binary clipping methods correspond to implicit threshold choices that do not explicitly compute $f(P_i, G)$. In contrast, the operator $\mathcal{T}_\tau$ defined in Equation 5 makes the inclusion rule explicit and controllable, allowing categorical aggregation to be derived consistently from fractional geometric support.

3.4 Geometric Decomposition into Interior and Boundary Regions

Distinguishing interior pixels from boundary pixels requires evaluating the full-containment topological predicate between each raster pixel P_i and the target polygon G, that is, verifying the condition $P_i \subseteq G$. This is a universal geometric predicate: every point of P_i must lie inside G. Direct evaluation of this condition for all pixels is computationally expensive and unnecessary if containment can be guaranteed through a sufficient geometric criterion. Let w and h denote the pixel width and height, and define the pixel diagonal as $d = \sqrt{w^2 + h^2}$. Any point within a pixel lies at a distance of at most $r = \frac{d}{2}$ from some point inside that pixel. The quantity r therefore provides an upper bound on the spatial extent of a pixel. This bound can be used to derive a sufficient condition for containment that avoids evaluating the universal predicate $P_i \subseteq G$ directly.

We therefore define the *eroded geometry*:

$$G^- = \{x \in G \mid \operatorname{dist}(x, \partial G) \geq r\}, \tag{6}$$

which corresponds to the Minkowski erosion of G by radius r defined above. The set G^- in Equation 6 contains all points of G that are at least distance r away from its boundary, i.e., points that lie sufficiently deep inside the geometry so that a disk of radius r centered at any such point remains entirely within G. This construction yields the following sufficient condition for containment:

$$P_i \cap G^- \neq \varnothing \implies P_i \subseteq G. \tag{7}$$

Indeed, if a pixel intersects G^-, then at least one point of the pixel lies at distance greater than or equal to r from ∂G (i.e., boundary of G). Because every point of the pixel is within distance r of that point, no part of the pixel can reach the boundary. The entire pixel must therefore lie inside G, and consequently $f(P_i, G) = 1$. The geometry G can thus be decomposed as:

$$G = G^- \cup \left(G \setminus G^-\right), \tag{8}$$

where G^- defines an interior certainty region and $G \setminus G^-$ defines a boundary band of thickness r. Within the clipped raster domain, pixels intersecting G^- are guaranteed to be fully contained and require no further geometric evaluation. Pixels intersecting $G \setminus G^-$ may partially overlap the boundary and therefore require

Algorithm 1 Fractional Raster–Vector Clipping Operator

Input: Raster $\mathcal{R}$, polygon G, threshold $\tau \in [0,1]$, mode m
Output: Two-band raster $\mathcal{R}^*$

Align G to raster reference system
Restrict $\mathcal{R}$ to the spatial support of G
Compute pixel dimensions (w,h) and radius $r = \frac{1}{2}\sqrt{w^2+h^2}$
Initialize weight band to 1 for all retained pixels
Construct interior region $G^- = \text{erosion}(G,r)$
Define boundary region $G_b = G \setminus G^-$
If G_b is non-empty then:
  For each pixel P_i intersecting G_b:
    Compute overlap fraction $f = \dfrac{\text{Area}(P_i \cap G)}{\text{Area}(P_i)}$
    If inclusion condition (based on τ and m) is not satisfied:
      Set value and weight of P_i to NULL
    Else:
      Set weight of P_i to f
Enforce consistency: pixels with NULL value receive NULL weight
Return $\mathcal{R}^*$

explicit computation of fractional overlap $f(P_i,G)$. This decomposition transforms universal containment testing into an existential intersection test localized to a geometrically defined boundary region. As a result, containment is guaranteed for interior pixels, while fractional computation is restricted to pixels that can partially overlap the boundary.

4 Operator Implementation and Computational Performance

The Fractional Raster–Vector Clipping Operator is implemented as a deterministic spatial procedure that faithfully translates the formal model into computation while maintaining efficiency (Algorithm 1). The implementation relies on the geometric interior–boundary decomposition: pixels guaranteed to be fully contained are included without explicit testing, and only pixels that may intersect the boundary are evaluated fractionally.

After restricting computation to the raster region overlapping the target geometry, a resolution-dependent erosion distance is derived from pixel dimensions. This distance defines an interior certainty region in which containment is guaranteed. Computation is then localized to the complementary boundary band, where partial overlap is possible. Interior pixels are implicitly included through initialization of a weight band, whereas boundary pixels undergo explicit overlap computation, with inclusion controlled by a user-defined threshold.

The operator maintains two raster bands: the original thematic values and a weight band encoding geometric support. The weight band is initialized to full inclusion and updated only for boundary pixels. Because the erosion distance is derived from pixel geometry, the procedure adapts naturally to spatial resolution. The resulting implementation is resolution-aware, mathematically consistent, and suitable for integration in database-centric and general geospatial processing environments.

We evaluated computational performance against the default PostGIS `ST_Clip` operation using spatial SQL queries on increasing raster sizes. Runtime (seconds) was measured as a function of total pixel count, ranging from approximately 1.8×10^4 to 1.5×10^8 pixels (Figure 2).

Both methods exhibit increasing runtime with input size. The default PostGIS clipping shows near-linear scaling, while the fractional operator exhibits a steeper growth trend in the log–log representation (Figure 2), reflecting the additional boundary-level processing required for fractional evaluation.

Despite this additional computation, absolute runtimes remain within operationally practical limits across all tested scales. The observed overhead is therefore modest relative to the methodological benefit of reducing boundary-induced aggregation bias.

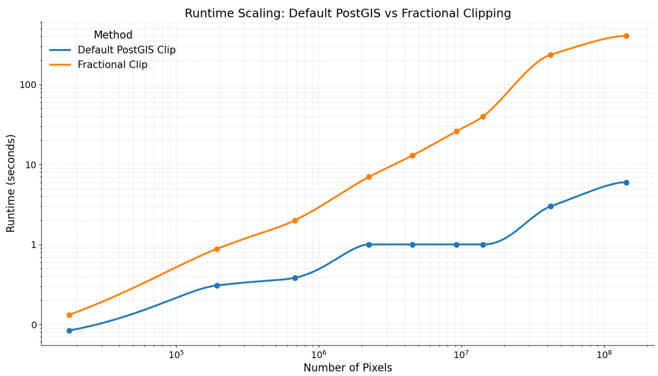

Figure 2. Log–log runtime comparison between Default PostGIS Clip and Fractional Clip across raster sizes ranging from 10^4 to 10^8 pixels. Markers indicate measured runtimes.

5 Experimental Design and Results

The evaluation uses *Basemap04* (2021) (Levin, 2022), Denmark's most recent national land-use and land-cover raster. Basemap is a harmonized product derived from publicly available spatial datasets and is used by Statistics Denmark for official land-use statistics and green national accounts. The dataset has a native spatial resolution of 10 m × 10 m and preserves original land-use classes, enabling flexible thematic aggregation.

Municipal boundaries serve as vector aggregation units, representing realistic reporting geometries with varying perimeter complexity. The experiments focus on Aarhus municipality as an illustrative case study (Figure 3).

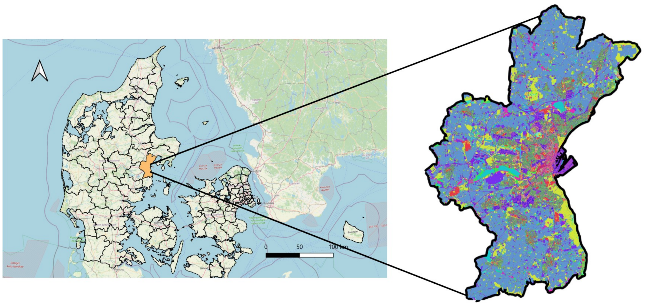

Figure 3. Aarhus municipal polygon within Denmark (left, orange) and the corresponding Basemap04 land-cover raster clipped to the polygon boundary (right).

5.1 Experiment 1: Total Area Bias Across Resolutions

The first experiment evaluates how raster resolution influences total aggregated area under different clipping semantics. The native 10 m raster was resampled to coarser resolutions (20, 50, 100, 200, 500, and 1000 m) using nearest-neighbor resampling to preserve categorical integrity. For each municipality and each resolution, total land-cover area was computed using (1) conventional binary clipping (PostGIS `ST_Clip`), (2) fractional clipping with threshold parameter τ (with $\tau = 0.7$), and (3) fully fractional clipping ($\tau = 0$), which serves as the geometric reference.

The reference area is defined as the weighted sum of pixel areas using fractional overlap:

$$A_{\text{ref}}(G) = \sum_i \text{Area}(P_i)\, f(P_i, G), \tag{9}$$

For any evaluated method M, the estimated area $A_M(G)$ is obtained by summing the areas of included pixels under that method. Area bias is quantified as:

$$\text{Bias}_M(\%) = 100 \cdot \frac{A_M(G) - A_{\text{ref}}(G)}{A_{\text{ref}}(G)}. \tag{10}$$

This experiment isolates the effect of spatial resolution on total area estimation when different inclusion rules are applied.

5.2 Experiment 2: Class-Level Area Bias

The second experiment follows the same design but evaluates area estimates separately for selected land-cover classes. For each class and each resolution, class-specific area was computed within the municipality using the same three clipping strategies. The reference area for each class is calculated using fractional weighting restricted to pixels belonging to that class. Bias is computed using the same formula as above in equation 10, replacing total area with class-specific area.

This experiment assesses how boundary effects propagate differently across land-cover types when aggregation is performed at varying spatial resolutions.

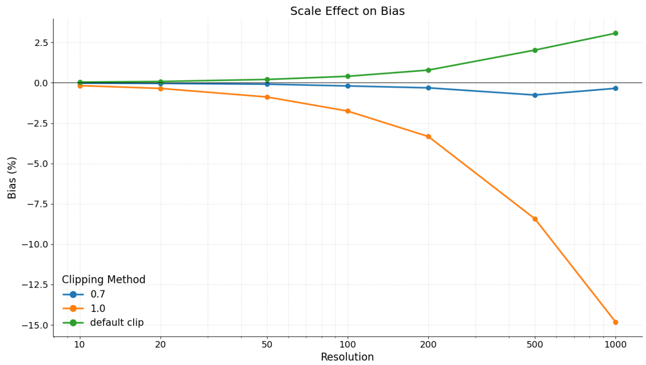

Figure 4. Total area bias (%) relative to the fully fractional reference across spatial resolutions (10–1000 m) for default clipping and fractional thresholds ($\tau = 0.7$, $\tau = 1.0$).

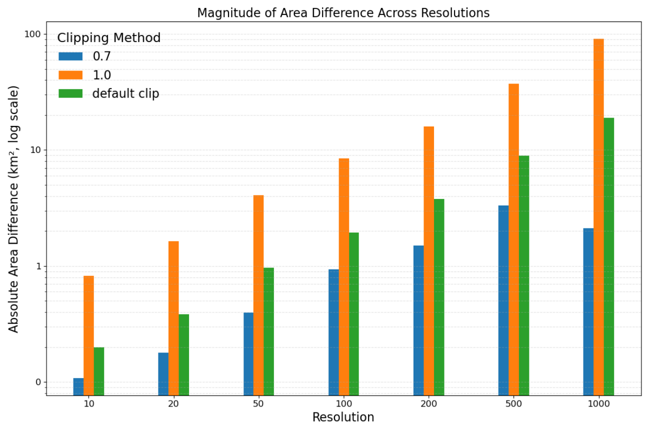

Figure 5. Absolute area difference (km^2) relative to the fully fractional reference across spatial resolutions (10–1000 m) for default clipping and fractional thresholds ($\tau = 0.7$, $\tau = 1.0$).

5.3 Results

5.3.1 Total Area Bias Across Resolutions.

Figure 4 and 5 quantify total aggregated area bias relative to the fully fractional reference model across spatial resolutions from 10 m to 1000 m. Clear resolution-dependent behavior is observed for all inclusion strategies.

The conventional binary clipping (default clip) exhibits small positive deviations at fine resolutions (0.04% at 10 m; 0.21% at 50 m), indicating slight overestimation relative to the fractional reference . However, bias increases systematically as spatial resolution coarsens, reaching 0.79% at 200 m, 2.03% at 500 m, and 3.07% at 1000 m (see Figure 4). In absolute terms (see Figure 5), this corresponds to nearly 19 km^2 overestimation at 1000 m resolution. These results demonstrate that binary clipping introduces scale-dependent aggregation effects

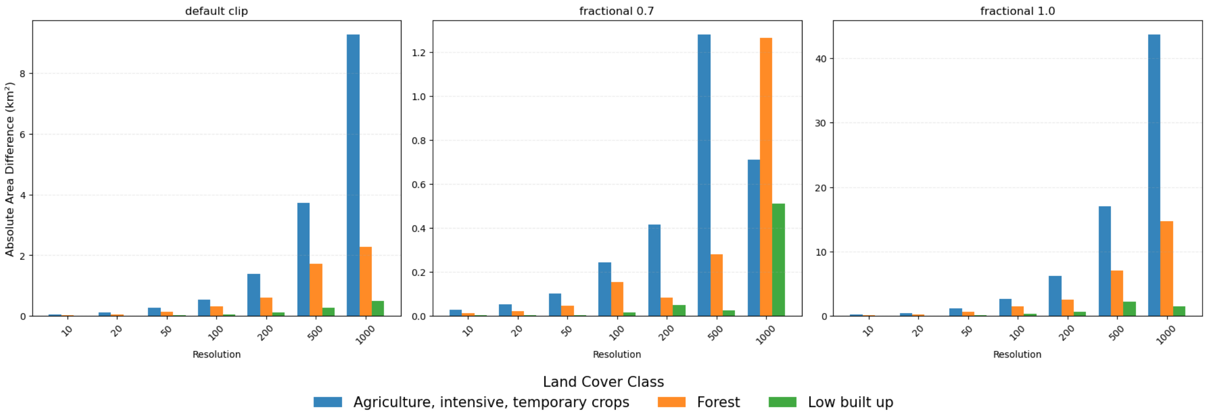

Figure 6. Absolute area difference (km^2) relative to the fully fractional reference across spatial resolutions (10–1000 m) for selected land-cover classes (Agriculture, intensive temporary crops; Low built up; Forest) under default clipping and fractional thresholds ($\tau = 0.7$, $\tau = 1.0$).

that become increasingly pronounced as pixel support grows.

The strict inclusion strategy ($\tau = 1.0$), which admits only fully contained pixels, produces the opposite pattern. Bias is already negative at fine resolutions (-0.17% at 10 m) and increases in magnitude nonlinearly with spatial resolution, reaching -8.43% at 500 m and -14.81% at 1000 m. At the coarsest resolution, this corresponds to an underestimation exceeding 90 km^2. This behavior confirms that strict containment does not converge toward the fractional reference at coarse scales; instead, boundary exclusion effects amplify rapidly with pixel size.

The intermediate threshold ($\tau = 0.7$) exhibits comparatively stable behavior across resolutions. Deviations remain below 1% for all tested scales and are substantially smaller in magnitude than both the default and strict strategies at coarse resolutions (-0.34% at 1000 m). Notably, at 500 m resolution, the absolute area difference is 3.33 km^2 under $\tau = 0.7$, compared to 8.92 km^2 for the default rule and 37.08 km^2 for $\tau = 1.0$.

5.3.2 Class-Level Boundary Bias Across Resolutions.

Figure 6 presents the absolute area differences for the three dominant land-cover classes in Aarhus municipality, namely *Agriculture areas*, *Low built up*, and *Forest*, across spatial resolutions ranging from 10 m to 1000 m.

At fine resolutions (10–50 m), deviations from the fractional reference remain very small for all classes and methods. For example, at 10 m resolution, the default clipping overestimates agricultural area by only 0.054 km^2 (0.027%), while the strict threshold ($\tau = 1.0$) underestimates it by 0.222 km^2 (-0.11%). Similar magnitudes are observed for *Low built up* and *Forest*, indicating minimal boundary impact at high spatial resolution.

As spatial resolution becomes coarser, class-level deviations increase in magnitude and exhibit distinct behavior across inclusion rules. At 500 m resolution, the default rule overestimates agricultural area by 3.72 km^2 (2.03%) and forest area by 1.72 km^2 (3.93%). In contrast, the strict threshold ($\tau = 1.0$) produces substantial underestimation: -17.03 km^2 (-9.28%) for agriculture and -7.03 km^2 (-16.06%) for forest. The *Low built up* class exhibits smaller absolute deviations, but the same directional pattern is observed.

At the coarsest resolution (1000 m), boundary effects become pronounced. The default rule overestimates agricultural area by 9.29 km^2 (3.59%) and forest area by 2.27 km^2 (3.35%). The strict threshold produces strong underestimation, reaching -43.71 km^2 (-16.90%) for agriculture and -14.73 km^2 (-21.75%) for forest. For *Low built up*, deviations remain smaller in magnitude but still increase with resolution (0.49 km^2 overestimation under the default rule; -1.51 km^2 under $\tau = 1.0$).

The intermediate threshold ($\tau = 0.7$) consistently yields smaller absolute deviations than both the default and strict strategies at coarse resolutions. For instance, at 1000 m, agricultural deviation under $\tau = 0.7$ is -0.71 km^2 (-0.28%), compared to +9.29 km^2 under the default rule and -43.71 km^2 under $\tau = 1.0$. Similar stabilization is observed for forest and low built-up classes.

Overall, the experiments show that boundary-induced deviations increase with spatial resolution and are strongly influenced by the chosen inclusion rule. Strict containment amplifies underestimation at coarse scales, while conventional binary clipping produces systematic overestimation. Intermediate overlap thresholds consistently reduce deviation magnitude but do not remove scale dependence. The relationship between inclusion strictness and deviation is therefore not monotonic. These patterns are observed both for total area and across individual land-cover classes, with larger effects in spatially extensive classes such as agriculture and forest. Together, the results demonstrate that aggregation outcomes are sensitive to boundary semantics and that threshold selection provides a practical means to evaluate the magnitude of boundary effects in multi-resolution analysis.

5.4 Data and Software Availability

The PostgreSQL/PostGIS implementation of the Fractional Raster–Vector Clipping Operator, together with validation test cases and the scripts used to reproduce the experimental results, is publicly available at: https://github.com/Sustainscapes/Fractional-Raster-Vector-Clipping-Operator-for-Multiscale-Spatial-Analysis (accessed 20 February 2026).

The land-cover dataset used in this study is Basemap04 (2021), Denmark's national land use and land cover map (Levin, 2022), available from Aarhus University: https://envs.au.dk/fileadmin/envs/Hjemmeside_2018/Zip_filer/Basemap04_public_geotiff.zip (accessed on 10 February 2026).

Danish municipal boundary geometries were obtained from the public repository:

https://github.com/magnuslarsen/geoJSON-Danish-municipalities

(accessed on 10 February 2026).

All experiments were conducted using PostgreSQL with the PostGIS spatial extension. The repository above contains the complete functions and processing workflows required to reproduce the results.

6 Conclusion

This paper formalizes a Fractional Raster–Vector Clipping Operator as a generic geospatial operator with explicit inclusion semantics. By defining pixel contribution through proportional areal overlap, clipping is elevated from an implicit binary routine to a mathematically specified component of spatial analysis. The proposed algorithm replaces universal containment testing with boundary-localized intersection evaluation, ensuring computational efficiency while preserving consistent geometric interpretation.

The database implementation demonstrates that the operator can be integrated within existing geospatial infrastructures without sacrificing scalability. Empirical results using Danish national land-cover data show that conventional binary clipping introduces systematic, resolution-dependent deviations relative to the fractional reference. These findings highlight that boundary semantics are not merely implementation details but analytical assumptions that directly influence aggregated spatial statistics. The present study focuses on categorical land-cover aggregation within administrative units. Sensitivity may differ for other thematic domains, complex geometries, or continuous variables. The fractional model also assumes proportional areal contribution, which does not account for sub-pixel thematic heterogeneity. Although boundary-localized evaluation improves efficiency, further optimization may be required for large-scale distributed environments.

Future work may extend the operator to probabilistic rasters, multi-temporal datasets, as well as investigate analytical bounds on aggregation error under varying geometric configurations. Formalizing clipping as an explicit operator opens a path toward more transparent, reproducible, and scale-aware raster–vector integration in GIScience.

Acknowledgements

This work was funded by Land-CRAFT—Pioneer Center for Landscape Research in Sustainable Agricultural Futures (DNRF grant number P2) and SustainScapes—Center for Sustainable Landscapes under Global Change (grant NNF20OC0059595 to SN).

Author Contributions (CRediT taxonomy)

YH: Conceptualization; Methodology; Formal analysis; Data curation; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. **UAT:** Conceptualization; Investigation; Writing – review & editing. **SN:** Supervision; Project administration; Funding acquisition; Writing – review & editing.

Declaration of Generative AI in writing

The authors declare that Generative AI tools were used solely to improve the readability, grammar, and language of this manuscript. The AI-assisted technologies were applied with human oversight and control, and the authors carefully reviewed and edited all outputs. The tools were not used to generate scientific content, research data, analysis, interpretation, or substantive conclusions.

References

Baston, D., ISciences, L., and Baston, M. D.: Package 'exactextractr', R Foundation for Statistical Computing. URL: https://cran. r-project. org, https://doi.org/10.32614/CRAN.package.exactextractr, 2021.

Chen, X., Ye, X., Widener, M. J., Delmelle, E., Kwan, M.-P., Shannon, J., Racine, E. F., Adams, A., Liang, L., and Jia, P.: A systematic review of the modifiable areal unit problem (MAUP) in community food environmental research, Urban Informatics, 1, 22, https://doi.org/10.1007/s44212-022-00021-1, 2022.

Cova, T. J.: Data Model, F-Objects and O-Fields, International Encyclopedia of Geography: People, the Earth, Environment and Technology, pp. 1–5, https://doi.org/10.1002/9781118786352.wbieg0481, 2016.

Cova, T. J. and Goodchild, M. F.: Extending geographical representation to include fields of spatial objects, International Journal of geographical information science, 16, 509–532, https://doi.org/10.1080/13658810210137040, 2002.

Galton, A.: A formal theory of objects and fields, in: International conference on spatial information theory, pp. 458–473, Springer, https://doi.org/10.1007/3-540-45424-1_31, 2001.

Goodchild, M. F.: Scale in GIS: An overview, Geomorphology, 130, 5–9, https://doi.org/10.1016/j.geomorph.2010.10.004, 2011.

Goodchild, M. F., Yuan, M., and Cova, T. J.: Towards a general theory of geographic representation in GIS, International journal of geographical information science, 21, 239–260, https://doi.org/10.1080/13658810600965271, 2007.

Hamdani, Y., Thibaud, R., and Claramunt, C.: A hybrid data model for dynamic GIS: application to marine geomorphological dynamics, International Journal of Geographical Information Science, 35, 1475–1499, https://doi.org/10.1080/13658816.2020.1829628, 2021.

Hamdani, Y., Xiao, G., Ding, L., and Calvanese, D.: An ontology-based framework for geospatial integration and querying of raster data cube using virtual knowledge graphs, ISPRS International Journal of Geo-Information, 12, 375, https://doi.org/10.3390/ijgi12090375, 2023.

Levin, G.: Basemap04: Documentation of the Data and Method for the Elaboration of a Land Use and Land Cover Map for Denmark, Tech. Rep. Technical Report No. 252, Aarhus University, DCE – Danish Centre for Environment and Energy, Aarhus, Denmark, https://dce2.au.dk/pub/TR252.pdf, 2022.

Nalej, M. et al.: Scale and Aggregation Effects of the Modifiable Areal Unit Problem on Built-Up Area Metrics, ISPRS International Journal of Geo-Information, 15, 72, https://doi.org/10.3390/ijgi15020072, 2026.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment, 148, 42–57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

Openshaw, S.: The Modifiable Areal Unit Problem, Concepts and Techniques in Modern Geography, Geo Books, Norwich, UK, 1984.

Ren, N., Zhang, D., and Cheng, Q.: FZStats v1.0: a raster statistics toolbox for simultaneous management of spatial stratified heterogeneity and positional dependence, Geoscientific Model Development, 18, 7165–7187, https://doi.org/10.5194/gmd-18-7165-2025, 2025.

Rigge, M., Bunde, B., McCord, S. E., Harrison, G., Assal, T. J., and Smith, J. L.: Spatial scale dependence of error in fractional component cover maps, Rangeland Ecology & Management, 99, 77–87, https://doi.org/10.1016/j.rama.2025.01.004, 2025.

Shi, C. and Wang, L.: Incorporating spatial information in spectral unmixing: A review, Remote Sensing of Environment, 149, 70–87, https://doi.org/10.1016/j.rse.2014.03.034, 2014.

Singla, S. and Eldawy, A.: Distributed zonal statistics of big raster and vector data, in: Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 536–539, https://doi.org/10.1145/3274895.3274985, 2018.

Xie, P., Liu, Y., He, Q., Zhao, X., and Yang, J.: An efficient vector-raster overlay algorithm for high-accuracy and high-efficiency surface area calculations of irregularly shaped land use patches, ISPRS International Journal of Geo-Information, 6, 156, https://doi.org/10.3390/ijgi6060156, 2017.

Yu, J., Wu, Z., and Sarwat, M.: GeoSpark: A cluster computing framework for processing large-scale spatial data, in: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 1–4, https://doi.org/10.1145/2820783.2820860, 2015.