A scalable multi-resolution framework for connectivity-based biodiversity indicators

Roozbeh Valavi[1,*], Karel Mokany[2], Chris Ware[3], Mathew Vickers[2], Katherine M. Giljohann[1], and Simon Ferrier[2]

[1]CSIRO, Clayton, VIC 3168, Australia
[2]CSIRO, PO Box 1700, Canberra, ACT, 2601, Australia
[3]CSIRO, Sandy Bay, TAS 7005, Australia
[*]Corresponding author: roozbeh.valavi@csiro.au

August 2026

Abstract

1. Landscape connectivity links local habitat condition to broader ecological processes including dispersal, recolonisation, range shifts, and species persistence, that underpin biodiversity responses to land-use and climate change. Several established biodiversity indicators — habitat connectedness (including as an input to the Biodiversity Habitat Index, BHI), Protected Area Connectedness Index (commonly referred to as PARC-connectedness), and the Bioclimatic Ecosystem Resilience Index (BERI) — formalise these processes as least-cost, condition-weighted representations of connected habitat. Applying these landscape connectivity formulations consistently across large spatial extents, multiple dispersal scales, protected-area configurations, and climate futures remains constrained by the computational cost of fine-resolution connectivity over broad landscapes.

2. We present a multi-resolution analytical framework that addresses this constraint by extending established cost-benefit connectivity logic into a scalable, indicator-oriented architecture. Building on the long-recognised principle that spatial precision matters most for areas close to each focal cell, formalised in earlier implementations through distance-dependent aggregation, we reformulate the calculation around a hierarchical graph derived from globally anchored, pre-computed raster overviews. This design preserves fine-grain detail where it most influences connectivity, represents more distant landscape context at progressively coarser resolution, and produces reusable path-level outputs that are computed once and applied across dispersal scales and climate scenarios without repeating graph traversal.

3. From these shared path distances, habitat connectedness, PARC-connectedness, and BERI are each derived as condition-weighted, dispersal-decayed integrals over accessible habitat, differing only in how destination nodes are weighted and what ecological quantity is being assessed. The unified derivation ensures consistency across indicators while eliminating redundant computation.

In an example application for Tasmania, Australia (approximately 68,400 km^2; more than 300,000 cells at about 900 m resolution), all three indicators were computed across three dispersal scales and six climate scenarios in under a minute on a standard laptop.

4. The framework makes no claim to a new definition of connectivity. Its contribution is analytical and architectural, enabling established and well-validated connectivity concepts to be applied reproducibly and efficiently at the scales demanded by contemporary biodiversity monitoring and conservation planning. It is distributed as a Python package with a high-performance Rust engine, designed for integration into open geospatial workflows and large-scale indicator applications.

KEYWORDS: *BERI, landscape connectivity; least-cost path; multi-resolution analysis; habitat connectedness; climate-change resilience; conservation planning*

1 Introduction

Landscape connectivity is a foundational concept in ecology and conservation, describing the extent to which landscapes facilitate or impede movement among resource patches (Taylor et al., 1993). Movement underpins a wide range of ecological processes, including dispersal, recolonisation following disturbance, gene flow, and the ability of populations to persist in fragmented environments (Calabrese and Fagan, 2004; Tischendorf and Fahrig, 2000). As such, connectivity links local habitat conditions to broader population and community dynamics, and is central to understanding how biodiversity responds to land-use change and environmental gradients.

Despite its importance, connectivity is not a single, universal property of landscapes. It can refer to the physical arrangement of habitat, the modelled permeability of a landscape, or the realised movement of organisms. These distinctions are often discussed in terms of structural connectivity and functional connectivity (Tischendorf and Fahrig, 2000), but the boundary between them is not always clear when connectivity is inferred from landscape data alone. In this paper, we use connectivity to mean modelled potential connectedness: a landscape-derived estimate of how habitat condition and spatial configuration may facilitate or impede movement under generalised dispersal assumptions.

Much of the connectivity literature has focused on species-specific modelling, where resistance surfaces and dispersal parameters are calibrated for particular taxa (Adriaensen et al., 2003; Zeller et al., 2012). While this approach is appropriate for targeted conservation problems, it is often impractical for large-scale biodiversity assessments. Regional to global analyses require methods that can be applied consistently across diverse ecosystems and taxa, often in the absence of species-specific movement data (Ware et al., 2026). In these contexts, connectivity is better treated as a landscape-scale property derived from habitat condition, spatial configuration, and generalised movement assumptions, rather than as a species-specific prediction. This shift aligns with the increasing use of *coarse-filter* approaches in conservation (Noss, 1987), where indicators are designed to capture broad ecological processes relevant to many species simultaneously. Complementary indicator families based on graph-theoretic patch-level metrics, such as the integral index of connectivity and the probability of connectivity, have also been widely applied at landscape and regional scales (Saura and Pascual-Hortal, 2007; Saura and Torné, 2009).

A further development is the growing use of connectivity as a component of biodiversity indicators and monitoring frameworks. Conservation policy increasingly emphasises the need for well-connected systems of protected areas and the maintenance of ecological networks beyond reserves (IPBES, 2019; CBD, UN, 2022), most notably through Target 3 of the Kunming-Montreal Global Biodiversity Framework, which calls for a well-connected system of protected and conserved areas covering 30 % of terrestrial, inland water, and coastal and marine areas by 2030 (CBD, UN, 2022). This has motivated the development of a family of indicators including habitat connectedness (used, among other applications, as an input to the Biodiversity Habitat Index, BHI), Protected Area Connectedness Index (commonly referred to as PARC-connectedness), and the Bioclimatic Ecosystem Resilience Index (BERI), which quantify connectivity in a form that can be aggregated, compared across regions, and tracked through time (Drielsma et al., 2007a; Ferrier et al., 2020; Harwood et al., 2022). These indicators share a common structure: they combine condition-weighted representations of habitat with modelled potential connectedness, typically based on least-cost paths and dispersal kernels, to estimate the effective availability of habitat within a landscape. BERI extends this structure by incorporating compositional similarity under current and future climate scenarios, derived from generalised dissimilarity modelling (Ferrier et al., 2007; Mokany et al., 2022), providing an estimate of the capacity of landscapes to retain biodiversity under climate change (Ferrier et al., 2020).

Across these applications, a consistent challenge is the tension between spatial resolution and computational tractability. Connectivity is inherently multi-scale: landscapes that appear connected at coarse resolution may be fragmented at finer resolution, and different dispersal processes operate over different spatial extents (Calabrese and Fagan, 2004; Urban and Keitt, 2001). Most implementations rely on single-resolution raster representations, which can become prohibitively expensive when applied at high spatial resolution over large extents, or when calculations must be repeated across scenarios, time steps, and dispersal parameters (Tang and Dou, 2023). Alternative methods such as circuit theory or resistant-kernel approaches address some conceptual limitations of least-cost paths but often introduce additional computational complexity or require specialised implementations (McRae et al., 2008; Kervellec et al., 2024; Compton et al., 2007; Ocampo-Peñuela et al., 2023). For indicator-based workflows, where connectivity must be computed repeatedly and consistently across large datasets, this gap between well-established connectivity formulations and their practical application at scale remains a significant barrier (Drielsma et al., 2022).

The analytical foundations for the framework we describe here are well established. The framework extends the raster-based cost-benefit approach to assessing habitat configuration (Drielsma et al., 2007a), which integrates the benefits of access to habitat with the costs of movement through heterogeneous landscapes using continuously varying habitat and permeability surfaces. A key efficiency principle of that earlier implementation was distance-dependent aggregation, in which neighbourhood cells were grouped into progressively coarser units at greater distances from the focal cell, reducing computation where spatial precision has least bearing on the outcome (Drielsma et al., 2007a). Subsequent developments include the Spatial Links Tool and the General Landscape Connectivity Model, which extended these ideas for mapping habitat linkages, cross-scale neighbourhood habitat, and operational landscape connectivity across large regions (Drielsma et al., 2007b, 2022), and provided the conceptual and mathematical basis for the connectivity calculations used in BERI and PARC-connectedness (Ferrier et al., 2020; Harwood et al., 2022; Ware et al., 2026). A corresponding software implementation exists in the NSW Biodiversity Tools package (`https://github.com/EM-MAF-NSW-DCCEEW/BiodiversityTools`), but is built as a C++/CUDA console application dependent on NVIDIA hardware and ESRI floating-point raster formats; requirements that limit portability, accessibility, and integration with open-source geospatial workflows. The present implementation therefore contributes not by replacing the underlying ecological framework, but by operationalising it in a cross-platform Python package with a high-performance Rust engine, internally generated multi-resolution arrays, polygon-tile processing, and reusable path-level outputs for multiple indicators.

Here, we present a multi-resolution, raster-based framework for landscape connectivity that operationalises established cost-benefit connectivity concepts within a scalable, indicator-oriented architecture. Building on the distance-dependent aggregation principle of earlier implementations, the approach reformulates the calculation around a hierarchical graph derived from globally anchored, pre-computed raster overviews that remain consistent across independently processed spatial tiles. The framework produces standardised path-level outputs that can be directly and efficiently transformed into habitat connectedness, PARC-connectedness, and BERI across multiple dispersal scales and climate scenarios without repeating graph traversal. The contribution is analytical and architectural: enabling existing, well-validated connectivity concepts to be applied more efficiently, transparently, and reproducibly at the scales demanded by contemporary biodiversity monitoring and conservation planning.

2 Conceptual framework

The framework is a least-cost formulation operating on a multi-resolution graph, building on the raster-based cost-benefit approach of Drielsma et al. (2007a) and subsequent developments by Drielsma et al. (2022). In that approach, habitat is represented as a continuous condition or permeability surface, rather than a binary patch-matrix system, and connectivity is estimated by combining the benefits of accessible habitat with the costs of movement through heterogeneous landscapes. Our new framework retains this logic but reformulates its distance-dependent aggregation principle within a scalable graph architecture: globally anchored raster overviews represent progressively coarser landscape context with increasing distance from the focal cell, enabling consistent tile-based processing and reuse of path distances across indicators and dispersal scales. Formal definitions of the multi-resolution representation, graph structure and edge costs are provided in Section 3, and indicator derivations in Section 4.

Within this framework, potential connectivity is represented using effective distance through heterogeneous landscapes. We adopt a least–cost formulation, in which movement between locations is expressed as the accumulated cost of traversing intervening cells, with costs determined by habitat condition (Harwood et al., 2022). This approach provides an interpretable and computationally tractable approximation of landscape permeability, and forms the basis of several widely used biodiversity indicators accounting for connectivity (Ware et al., 2026). Least-cost distances are computed using shortest-path algorithms on raster-derived graphs, allowing condition-weighted connectivity to be quantified consistently across large spatial domains.

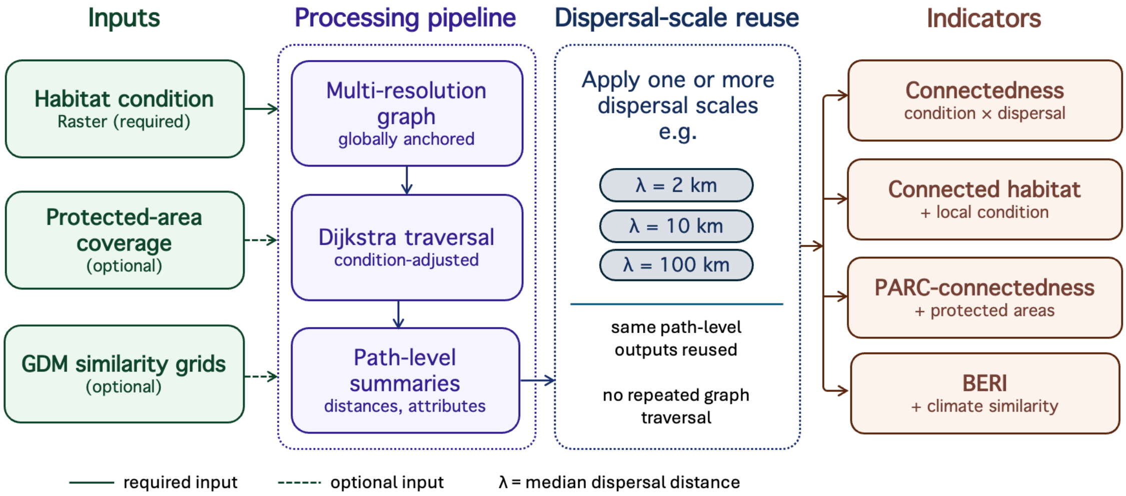

Figure 1: Schematic overview of the analytical framework. A habitat condition raster is required for all indicators; protected-area coverage and generalised dissimilarity modelling (GDM)–derived compositional similarity grids are additional inputs for PARC-connectedness and BERI, respectively. The processing pipeline constructs a multi-resolution graph from globally anchored raster overviews, computes condition-adjusted and intact-reference path distances by Dijkstra traversal, and stores path-level summaries for each focal cell. Each indicator applies a different ecological weighting to these shared summaries. Path distances are computed once per focal cell and reused across dispersal scales and climate scenarios without repeating graph traversal.

The objective is not to represent realised movement of specific taxa, nor to claim functional connectivity in the empirical sense; rather, the method estimates potential connectedness under generalised assumptions about dispersal limitation and landscape resistance, providing a consistent and repeatable basis for biodiversity indicators across ecosystems and taxa.

Alternative frameworks including circuit theory, resistant kernels, and graph-theoretic approaches each address limitations of single least-cost paths, particularly regarding pathway redundancy and diffuse movement (McRae et al., 2008; Urban and Keitt, 2001; Compton et al., 2007; Zeller et al., 2012), and are more appropriate where those properties are the primary focus. The least-cost formulation adopted here is most appropriate where the aim is consistent, repeatable estimation of potential connected habitat availability across large extents and multiple scenarios, rather than prediction of realised movement by particular species. In this context, effective distance provides a practical and interpretable basis for deriving habitat connectedness, PARC-connectedness, and BERI, while enabling integration with the habitat condition surfaces and compositional similarity models that those indicators require (Ferrier et al., 2020; Harwood et al., 2022).

3 Multi-resolution connectivity algorithm

This section describes the multi-resolution, raster-based algorithm for estimating landscape connectivity from continuous habitat-condition surfaces (with binary surfaces as a special case). The method constructs a hierarchical representation of the landscape from base-resolution raster data, defines a multi-scale graph that captures both local and long-range movement opportunities, and computes least-cost connectedness using Dijkstra's algorithm. Shortest-path distances and node attributes are stored as reusable path-level outputs, from which habitat connectedness, PARC-connectedness, and the Bioclimatic Ecosystem Resilience Index (BERI) are derived (Section 4). Computational aspects of the implementation, including tile-based processing and boundary handling, are described in Section 5.

3.1 Multi-resolution raster representation

The method begins with a base-resolution raster $R^{(0)}$, typically representing habitat condition as a continuous variable $h_i \in [0, 1]$, or a binary habitat surface. Coarser resolution representations are generated internally by aggregating $R^{(0)}$, rather than relying on precomputed overview layers. This approach provides flexibility in the construction of multi-resolution neighbourhoods, supports consistent processing across independent spatial tiles, and avoids dependence on external overview formats.

Each coarser level $R^{(l)}$ is defined using an aggregation factor f, where f is a power of two (e.g. $f \in \{2, 4, 8, 16, 32\}$). At each level, a coarse cell corresponds to a block of $f \times f$ base-resolution cells. Aggregation blocks are defined using global pixel indices:

$$\mathrm{row}^{(f)} = \left\lfloor \frac{r}{f} \right\rfloor, \qquad \mathrm{col}^{(f)} = \left\lfloor \frac{c}{f} \right\rfloor. \tag{1}$$

Where r and c are row and column of the global base-resolution grid, respectively. Global anchoring ensures that aggregation boundaries are consistent across independently processed tiles, preventing spatial discontinuities when results are combined.

For habitat condition data, aggregation is performed using arithmetic averaging over all valid cells within each block:

$$h_{i,j}^{(f)} = \frac{1}{|B_{i,j}|} \sum_{(r,c)\in B_{i,j}} h_{r,c}^{(0)}, \tag{2}$$

where $B_{i,j}$ is the set of valid base-resolution cells contributing to coarse cell $(i,\, j)$ (missing values are excluded), and $|B_{i,j}|$ denotes the cardinality (count of valid cells) of that set. Compositional similarity surfaces used in BERI calculations (Section 4.3) are represented as aligned multi-band arrays and aggregated using the same scheme to maintain consistency across resolution levels.

Alongside habitat condition, each level $R^{(f)}$ carries a cell-area weight $A_{i,j}^{(f)}$ representing the effective base-resolution area contained within each coarse cell. For projected coordinate systems with uniform pixel area, $A_{i,j}^{(f)}$ equals the number of valid base-resolution cells aggregated into (i, j). For geographic coordinate systems, $A_{i,j}^{(f)}$ is latitude-weighted using $\cos\phi$ to account for variation in physical cell area along latitude. The cell-area weight serves as the foundation for the effective area assigned to nodes in the multi-resolution graph (Section 3.2), and is later combined with a window-overlap fraction to give each node its final area contribution.

This hierarchical representation provides a scale-consistent abstraction of the landscape, in which fine-scale heterogeneity is preserved locally while broader spatial structure is represented through aggregated cells (Figure 2).

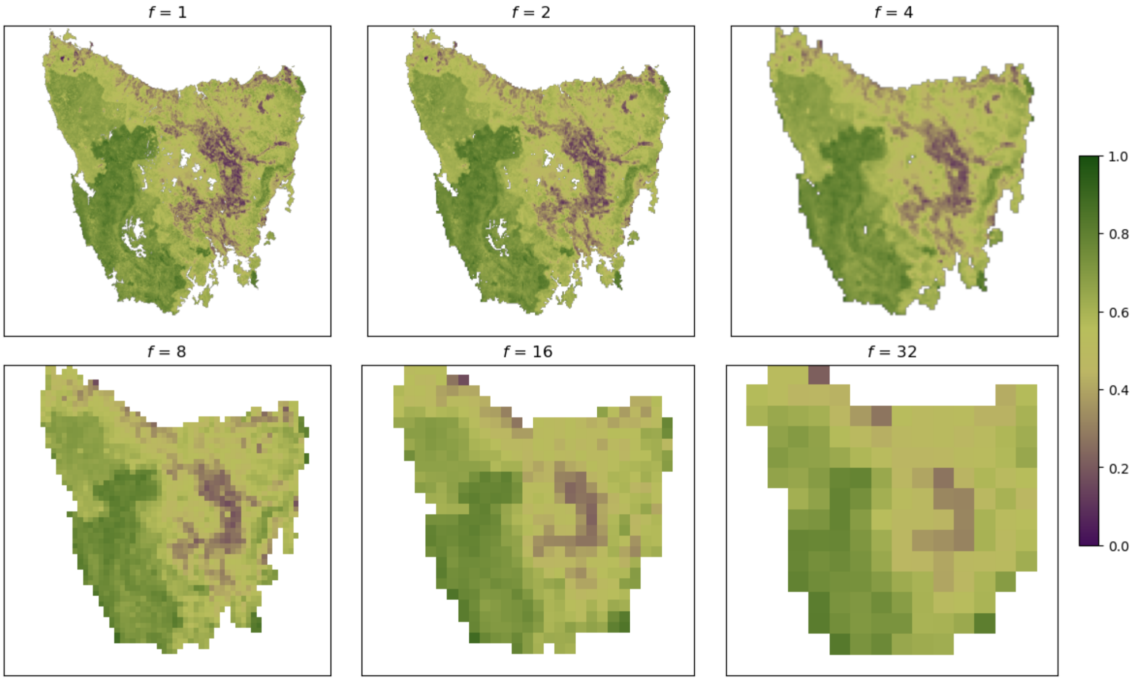

Figure 2: Example of multi-resolution representation of habitat condition used in the connectivity analysis, for Tasmania, Australia. Panels show the base-resolution ($f = 1$) and progressively coarser aggregation levels (aggregation factors $f = 2, 4, 8, 16, 32$), generated internally from the base-resolution raster. Each coarse cell represents the mean condition of the corresponding block of base-resolution cells.

3.2 Neighbourhood construction across spatial scales

For each focal cell at the base spatial resolution, the algorithm constructs a multi-resolution neighbourhood that defines the candidate cells used in the connectivity graph. The neighbourhood retains fine spatial detail near the focal cell while representing more distant parts of the landscape using progressively coarser overview cells. The aim is to preserve local heterogeneity while avoiding redundant representation of the same area at multiple resolutions.

The default neighbourhood construction is a source-centred circular annulus scheme. At each resolution level, indexed by aggregation factor f, neighbours are selected relative to the focal base-resolution cell. Here, $f = 1$ denotes the base raster, larger powers of two denote coarser overview levels, and $f_{\max}$ denotes the largest aggregation factor used in the analysis. Let $s = (r_s + 0.5,\, c_s + 0.5)$ denote the centre of the focal base-resolution cell in global base-pixel coordinates.

At each level f, the neighbourhood is defined as a circular annulus $\mathcal{A}^{(f)}$ centred on s. The inner radius is

$$\rho_{\text{in}}^{(f)} = \begin{cases} 0, & f = 1 \\ \dfrac{fw}{2}, & f > 1 \end{cases} \tag{3}$$

and the outer radius is

$$\rho_{\text{out}}^{(f)} = \begin{cases} fw, & f < f_{\max} \\ fw_{\text{outer}}, & f = f_{\max} \end{cases} \tag{4}$$

where w is window-size and w_{outer} is outer-window parameters set by the user, after enforcing $w_{\text{outer}} = \max\{w_{\text{outer}}, w\}$. The annulus at level f is

$$\mathcal{A}^{(f)} = \{\mathbf{x} : \rho_{\text{in}}^{(f)} \leq \|\mathbf{x} - \mathbf{s}\|_2 \leq \rho_{\text{out}}^{(f)}\} \tag{5}$$

with distances measured in base-resolution pixel units. The radius scales with f because w is expressed in level-f cells, and each level-f cell spans f base-resolution pixels. At the finest level $\rho_{\text{in}}^{(1)} = 0$, so $\mathcal{A}^{(1)}$ is a full disc rather than a ring. At the coarsest level, $w_{\text{outer}} \geq w$ allows the search radius to extend beyond the standard local window, decoupling long-range reach from the local resolution. This enables coarse-level context to be captured efficiently, without forcing a uniformly larger window at finer levels or requiring a higher maximum aggregation factor. Because aggregation factors follow a powers-of-two sequence, adjacent levels meet exactly at their annulus boundaries, and apart from zero-area shared boundaries, each point inside the search disc belongs to exactly one resolution level. Figure 3 illustrates this source-centred circular multi-resolution neighbourhood. The focal cell is surrounded by nested annuli corresponding to successive aggregation levels, so fine-resolution cells are retained close to the focal location and progressively coarser cells represent more distant landscape context. Cells intersecting annulus boundaries contribute according to their geometric overlap.

Each overview cell f has a rectangular footprint $F_{ij}^{(f)}$ in base-pixel coordinates. The algorithm computes the geometric overlap fraction:

$$\alpha_{i,j}^{(f)} = \frac{\left|F_{i,j}^{(f)} \cap \mathcal{A}^{(f)}\right|}{\left|F_{i,j}^{(f)}\right|} \in [0,\, 1] \tag{6}$$

which captures the proportion of each cell's footprint that lies within the annulus. For circular mode, the rectangle-disc intersection is evaluated analytically using circular-segment geometry and quadrant-based inclusion-exclusion; no Monte Carlo sampling or raster resampling is used. A cell is retained when $\alpha_{i,j}^{(f)} > 0$ and the aggregated habitat condition value is finite. Cells whose footprints lie entirely within $\mathcal{A}^{(f)}$ receive full weight ($\alpha = 1$), whereas cells crossing an inner or outer annulus boundary receive a fractional weight $\alpha \in (0,\, 1)$.

The effective area assigned to a retained node combines the cell-area weight $A_{i,j}^{(f)}$ defined in Section 3.1 with this overlap fraction:

$$\widetilde{A}_{i,j}^{(f)} = \alpha_{i,j}^{(f)} \cdot A_{i,j}^{(f)} \tag{7}$$

Cells fully inside the annulus contribute their full cell-area weight, while cells crossing a boundary contribute proportionally. This combined weighting ensures that the neighbourhood's spatial extent is faithfully represented at every level, even where annulus boundaries cut across coarse-cell footprints. Habitat condition $h_{i,j}^{(f)}$ remains the finite-cell mean of the overview cell. When compositional similarity layers are used, these are also represented by overview-level means.

The annuli are circular in base-pixel coordinates. For projected rasters with square pixels, this corresponds to circular sampling in physical space. For geographic rasters or rasters with non-square cells, the sampled region is circular in raster-index space rather than physical space. Physical distances used in graph edge weights are computed separately from the raster transforms, using Euclidean distances for projected rasters and great-circle distances for geographic rasters.

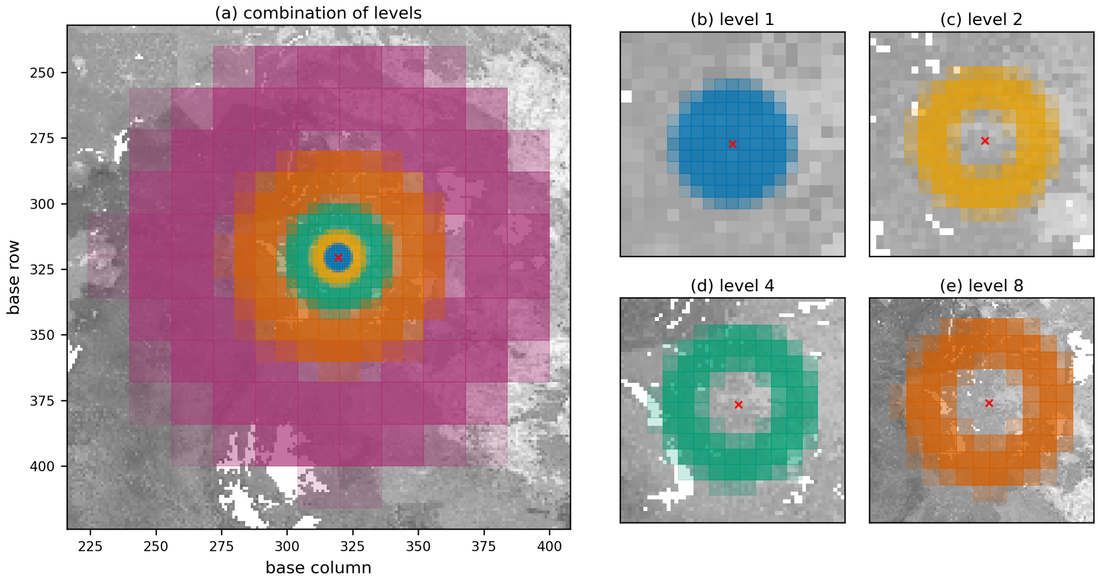

Figure 3: Source-centred circular multi-resolution neighbourhood on a raster of habitat condition (low = white, high = black). Panel (a) shows the assembled neighbourhood around a focal cell up to level 16, with annuli at successive levels shown in distinct colours. Panels (b)–(e) show selected level-specific contributions (note: level 16 is not shown). The blocky appearance arises from the rectangular footprints of aggregated raster cells, whereas neighbourhood membership is defined by the underlying circular annuli in base-pixel coordinates. Cells crossing annulus boundaries contribute fractionally according to their geometric overlap.

The retained cells form the nodes of the multi-resolution graph, each carrying its overview habitat condition $h_{i,j}^{(f)}$ and effective area $\widetilde{A}_{i,j}^{(f)}$. Within each resolution level, neighbouring retained cells are connected using queen adjacency. Cross-scale edges between adjacent resolution levels are added through boundary cells, identified geometrically as cells whose footprints intersect the outer circle of their annulus, so cross-scale links follow the curved perimeter of the circular neighbourhood.

The framework also supports a square neighbourhood mode, which uses the same source-centred fractional weighting approach but replaces circular annuli with axis-aligned square annuli based on the Chebyshev norm (see Supporting Information). The circular construction is the default because its geometry is independent of the focal cell's position within the aggregation grid and avoids the corner emphasis of square annuli, while preserving the non-redundant multi-resolution partition used to build the connectivity graph.

3.3 Graph formulation and edge costs

The multi-resolution neighbourhood is represented as a directed, weighted graph $G = (V, E)$, where each node $v \in V$ corresponds to a spatial unit (a base or aggregated cell retained in the neighbourhood) and edges $e \in E$ represent potential movement pathways. This formulation provides a discrete approximation of movement through a continuous landscape, in which spatial heterogeneity is represented through node attributes and edge weights.

Each node v is associated with three attributes carried from the neighbourhood construction (Section 3.2):

- habitat condition $h_v \in [0, 1]$;
- effective area $\widetilde{A}_v$, combining the cell-area weight $A_{i,j}^{(f)}$ with the window-overlap fraction $\alpha_{i,j}^{(f)}$;
- and, for BERI applications, compositional similarity values $s_v^{(k)}$ for each climate scenario k.

Nodes at coarser resolutions represent larger spatial extents, reflected directly in $\widetilde{A}_v$ and used to weight contributions in the indicator calculations (Section 4).

Edges capture both local and cross-scale movement. Within each resolution level, nodes are connected using 8-neighbour (queen) adjacency, representing local movement through the landscape. To enable efficient representation of longer-distance connectivity, additional cross-scale edges are defined between nodes at adjacent resolution boundaries. These cross-scale connections are restricted to boundary cells and link each node to a small set of neighbouring cells in the next coarser level, providing a sparse approximation of movement across aggregated spatial units. This design reduces graph density while preserving connectivity pathways across scales.

Geographic distance d_{uv} between two connected nodes u and v is computed from their centroid coordinates. For projected coordinate systems with planar geometry, this is the Euclidean distance $d_{uv} = \|\mathbf{x}_u - \mathbf{x}_v\|_2$. For geographic coordinates, d_{uv} is computed as the great-circle (Haversine) distance over the spherical Earth approximation.

To represent the reduced permeability of degraded habitat, each edge carries a condition-adjusted cost obtained by scaling the geographic distance according to the condition of the destination node v:

$$w_{uv}^{\mathrm{adj}} = d_{uv} \cdot [(1 - \mu)\, h_v + \mu] \tag{8}$$

where $\mu \geq 1$ is the maximum-cost parameter that sets the cost ratio between fully degraded and fully intact habitat. When $h_v = 1$, the bracketed term equals 1 and traversal cost reduces to the geographic distance d_{uv}. When $h_v = 0$, the bracketed term equals μ and traversal cost equals $\mu \cdot d_{uv}$. The implementation uses a default of $\mu = 2$, so movement through a fully degraded cell incurs twice the cost of equivalent traversal through intact habitat. This linear form provides a numerically stable, parameter-light approximation of habitat resistance suitable for large-scale analyses, avoiding the extreme cost values that arise from non-linear or threshold-based resistance functions.

An intact reference cost is defined for use in the connectedness denominators (Section 4):

$$w_{uv}^{\text{int}} = d_{uv} \tag{9}$$

which corresponds to traversal in a hypothetical intact landscape where $h_v = 1$ everywhere, reducing w_{uv}^{int} to the geographic distance.

Together, these definitions yield a multi-resolution graph in which movement is constrained by both spatial distance and habitat condition, and long-range connectivity is captured through a combination of fine- and coarse-resolution pathways.

3.4 Least-cost traversal and path-level outputs

For each focal cell, the algorithm computes shortest-path distances to all reachable nodes in the multi-resolution graph using Dijkstra's algorithm. Two traversals are performed over the same graph structure, one using the condition-adjusted edge weights w_{uv}^{adj} and one using the intact reference weights $w_{uv}^{\text{int}} = d_{uv}$ (Section 3.3). For a focal node i, these traversals yield the minimum cumulative path distance to each reachable node j:

$$D_{ij}^{r} = \min_{p \in P_{ij}} \sum_{(u,v) \in p} w_{uv}^{r}, \qquad r \in \{\text{adj, int}\} \tag{10}$$

where P_{ij} is the set of all paths from i to j. The adjusted traversal represents movement through the observed landscape, while the intact traversal represents movement in a hypothetical fully intact landscape and provides the reference quantity used in the denominators of the connectivity indicators (Section 4).

For each reachable node j, the algorithm records the pair of path distances D_{ij}^{adj} and D_{ij}^{int}, together with the attributes of the destination node carried from the neighbourhood construction (Section 3.2): habitat condition h_j, effective area $\widetilde{A}_j$, and, where required, scenario-specific GDM predicted values $s_j^{(k)}$ enabling prediction of pairwise compositional similarity. These quantities form the path-level outputs from which all connectivity indicators are derived in Section 4.

Two design properties of this stage are important. Connectivity is evaluated in terms of accessibility to destination nodes, not as cumulative properties along paths; nodes at coarser resolutions represent larger spatial extents and contribute proportionally through $\widetilde{A}_j$. And shortest-path computation is performed once per focal cell, so the resulting distances can be reused across multiple dispersal-scale parameters within each indicator run, avoiding repeated graph traversal. The separation between path computation and metric aggregation is a core efficiency feature of the framework and is exploited directly in the indicator derivations below.

3.5 From path distances to connectivity metrics

The outputs of the multi-resolution connectivity algorithm consist of path-level distances between each focal cell and all reachable destination nodes – the condition-adjusted distance D_{ij}^{adj} and intact reference distance D_{ij}^{int} – together with the destination-node attributes recorded in Section 3.4. These outputs provide a general representation of landscape accessibility, independent of any specific ecological indicator.

Connectivity indicators are derived by applying dispersal-decay functions and ecological weighting schemes to this shared set of path-level outputs. The algorithm itself does not impose a specific decay function or indicator formulation; it produces a single representation of accessibility that can be transformed into different indicators depending on the application. The specific derivations of habitat connectedness, PARC-connectedness, and BERI are presented in Section 4.

4 Derivation of connectivity indicators

The multi-resolution connectivity algorithm described in Section 3 produces a shared set of path-level outputs for each focal cell i: condition-adjusted distances, intact reference distances, and destination-node attributes such as habitat condition, effective area and, where required, scenario-specific compositional similarity. These quantities describe the accessibility of surrounding habitat from each focal location and provide a common computational basis for deriving multiple connectivity indicators.

In this section, we formalise how these path-level quantities are transformed into three biodiversity indicators: habitat connectedness (Drielsma et al., 2007a), protected-area connectedness (PARC-connectedness), and the Bioclimatic Ecosystem Resilience Index (BERI; Ferrier et al., 2020; Harwood et al., 2022). Each indicator expresses connectivity as the effective availability of habitat, weighted by dispersal limitation and, where relevant, habitat condition, protected-area coverage or compositional similarity under current or future environmental conditions.

This unified formulation allows the same computed paths to be reused across indicators and dispersal-scale parameters, reducing redundant computation while ensuring that all indicators are derived from an internally consistent representation of landscape accessibility.

4.1 Habitat connectedness

Habitat connectedness C_i quantifies the proportion of potential connected habitat retained around a focal cell i, relative to a hypothetical intact landscape (Drielsma et al., 2007a; Ferrier et al., 2020). For each reachable destination node j, habitat contribution is discounted by condition-adjusted path distance and weighted by the condition and effective area of the destination node:

$$S_i^{\mathrm{adj}} = \sum_j \exp\left(-\frac{D_{ij}^{\mathrm{adj}}}{\lambda}\right) h_j\, \widetilde{A}_j, \tag{11}$$

where D_{ij}^{adj} is the condition-adjusted least-cost distance from i to j, h_j is the habitat condition of destination node j, $\widetilde{A}_j$ is its effective area, and λ controls the rate of distance decay. The sum is taken over all reachable destination nodes.

The intact reference is obtained by replacing condition-adjusted distances with intact reference distances and implicitly setting destination condition to $h_j = 1$ (intact condition everywhere):

$$S_i^{\mathrm{int}} = \sum_j \exp\left(-\frac{D_{ij}^{\mathrm{int}}}{\lambda}\right) \widetilde{A}_j, \tag{12}$$

Habitat connectedness is then the ratio

$$C_i = \frac{S_i^{\mathrm{adj}}}{S_i^{\mathrm{int}}}. \tag{13}$$

with $C_i \in [0, 1]$ when habitat condition values are bounded between 0 and 1 and condition-adjusted distances are greater than or equal to their intact-reference counterparts. Values near 1 indicate that current habitat condition and spatial configuration impose little constraint on connectivity relative to the intact reference, whereas lower values indicate increasing degradation, isolation or fragmentation.

Because the kernel evaluates accessibility to destination nodes rather than cumulative resistance along paths, habitat condition enters through both the adjusted effective distance D_{ij}^{adj} and the destination weight h_j, jointly representing landscape permeability and the availability of suitable habitat. The negative exponential form $(\exp(-D/\lambda))$ follows widely used parameterisations in dispersal modelling (Nathan et al., 2012), and λ retains its interpretation as the median single-generation dispersal distance (Figure 4).

To combine local condition with surrounding connectedness, connected habitat is defined as geometric mean:

$$H_i^{\mathrm{conn}} = \sqrt{h_i\, C_i}, \tag{14}$$

where h_i is the habitat condition at the focal cell and C_i is the connectedness defined above. This formulation ensures that both components contribute symmetrically, such that low values in either habitat condition or connectivity reduce the overall connected habitat score. This combined measure of local habitat condition and surrounding connectedness is used as the connectivity-adjusted ecosystem condition input to the BHI (Ware et al., 2026).

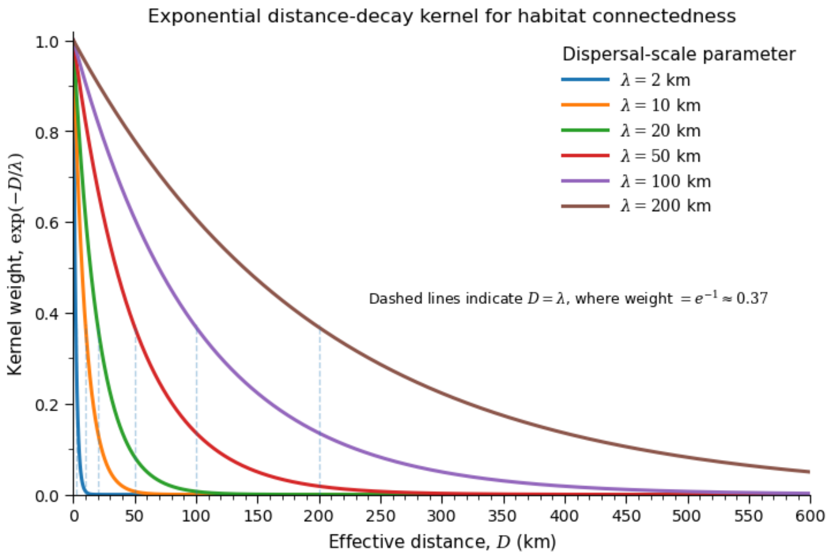

Figure 4: Negative exponential distance-decay kernel used for habitat connectedness and PARC-connectedness. Curves show the kernel weight $\exp(-D/\lambda)$ as a function of effective distance D for different values of the scale parameter λ. Dashed vertical lines indicate $D = \lambda$, where the kernel weight equals $e^{-1} \approx 0.37$. Larger values of λ give slower decay and increase the influence of distant habitat on connectedness estimates.

4.2 Protected Area Connectedness Index

Protected Area Connectedness Index (PARC-connectedness) quantifies the degree to which protected areas and other effective area-based conservation measures (OECMs) are connected to one another and to surrounding intact habitat (UNEP-WCMC, 2023; Ware et al., 2026). It uses the same kernel and ratio structure as habitat connectedness, but applies them to a modified habitat surface that incorporates protected-area coverage.

Let p_i denote the proportion of cell i covered by protected areas or OECMs, and let h_i denote its habitat condition. The effective habitat quantity used in the PARC calculation is

$$\varphi_i = \max(h_i, p_i). \tag{15}$$

This definition ensures that cells with protected-area coverage contribute at least that amount of effective habitat, while unprotected cells in high condition can still contribute as intact habitat in the surrounding landscape. The modified surface φ replaces h in both the edge-cost calculation and the destination-weighting term.

PARC-connectedness is reported only for focal cells with $p_i > 0$, but least-cost paths are computed through the full landscape. Unprotected cells are therefore not excluded: they may act as traversable matrix and, where condition is high, as contributing habitat.

For a protected focal cell i, the condition-adjusted and intact reference sums are

$$S_i^{\mathrm{adj},\mathbf{PA}} = \sum_j \exp\left(-\frac{D_{ij}^{\mathrm{adj}}}{\lambda}\right) \varphi_j \, \widetilde{A}_j, \tag{16}$$

and

$$S_i^{\mathrm{int},\mathbf{PA}} = \sum_j \exp\left(-\frac{D_{ij}^{\mathrm{int}}}{\lambda}\right)\widetilde{A}_j. \tag{17}$$

where $D_{ij}^{\mathrm{adj},\mathbf{PA}}$ is the least-cost distance calculated using the modified surface φ. PARC-connectedness is then

$$\mathrm{PARC}_i^{\mathrm{con}} = \frac{S_i^{\mathrm{adj},\mathbf{PA}}}{S_i^{\mathrm{int},\mathbf{PA}}}. \tag{18}$$

Values approach 1 where protected cells are embedded in well-connected networks of protected or intact habitat, and decline where they are isolated by fragmentation or degraded matrix. For regional reporting, cell-level values can be aggregated by weighting each protected cell by its protected-area coverage and expressing the result relative to all cells in the reporting unit, yielding a connectivity-adjusted estimate of protected area extent (Ware et al., 2026).

4.3 Biodiversity Ecosystem Resilience Index (BERI)

The Bioclimatic Ecosystem Resilience Index (BERI) quantifies the capacity of each focal location to retain biological diversity under climate change by weighting accessible habitat according to its projected compositional similarity to the focal cell under alternative climate futures (Ferrier et al., 2020; Harwood et al., 2022). BERI extends the habitat connectivity formulation in two ways: it introduces a destination weight based on projected ecological similarity, and it replaces the single-generation kernel of Sections 4.1–4.2 with a multi-year Gaussian kernel that reflects integrated dispersal over the climate-scenario horizon.

Climate scenarios used here span several decades, over which members of a species may undergo many successive dispersal events. As Harwood et al. (2022) show analytically, iterating a two-dimensional negative exponential dispersal kernel over g generations produces, by the Central Limit Theorem, a kernel that converges to a Gaussian whose variance grows linearly with g. Adopting this multi-generation form for BERI, the kernel applied at effective distance D is

$$K_g(D) = \exp\left[-\frac{(D/\lambda)^2}{\delta}\right], \tag{19}$$

with dimensionless scaling constant $\delta = 5.785\,(g-1)$. Here $g = 50$ generations is chosen to match a 50-year horizon of the climate scenarios, giving $\delta \approx 283.5$ (Ferrier et al., 2020; Harwood et al., 2022). The parameter λ retains its meaning as the median single-generation dispersal distance, while the Gaussian form captures the integrated effect of repeated dispersal over the scenario time horizon. This change from the single-generation exponential of Sections 4.1–4.2 is a refinement appropriate to the multi-decadal scenario horizon rather than the defining distinction between the indicators; the central methodological difference is BERI's weighting of accessible habitat by its projected compositional similarity to the focal cell under current and future climate scenarios.

Let s_{ijk} denote the predicted compositional similarity between the present-day assemblage of focal cell i and the assemblage projected for destination cell j under climate scenario k, with $k = 0$ denoting the current-climate case and $k = 1, \ldots, m$ the future scenarios. For each scenario, the amount of connected, compositionally similar habitat accessible from i is:

$$B_{ik} = \sum_j K_g\left(D_{ij}^{\text{adj}}\right) h_j \, \widetilde{A}_j \, s_{ijk}. \tag{20}$$

Destinations contribute more strongly when they are more accessible, in better habitat condition, larger in effective area, and more compositionally similar to the focal cell under scenario k.

The reference quantity represents the maximum connected habitat the focal cell could access in the absence of degradation or climatic change, i.e. an intact landscape under current climate:

$$B_i^{\text{int}} = \sum_j K_g\left(D_{ij}^{\text{int}}\right) \widetilde{A}_j \, s_{ij0}. \tag{21}$$

Here, D_{ij}^{int} is the intact-reference distance, habitat condition is implicitly set to $h_j = 1$, and compositional similarity is evaluated under the current climate.

To summarise performance across the $m+1$ scenarios while penalising poor worst-case outcomes, scenario-specific values are combined using the limited-degree-of-confidence (LDC) statistic (Bryan et al., 2014; Ferrier et al., 2020), defined as the unweighted mean of the across-scenario mean and minimum:

$$\text{LDC}(C_{i0}, \ldots, C_{im}) = \frac{1}{2}\left[\frac{1}{m+1}\sum_{k=0}^{m} C_{ik} + \min_k C_{ik}\right]. \tag{22}$$

BERI is then the ratio of this aggregate to the intact reference:

$$\text{BERI}_i = \frac{\text{LDC}(C_{i0}, \ldots, C_{im})}{C_i^{\text{int}}}. \tag{23}$$

Higher values indicate that the focal cell has access to substantial habitat that is both well connected under the multi-year dispersal kernel and projected to support compositionally similar assemblages across plausible climate futures. Lower values indicate that retention of present-day biological composition at i depends either on fragile, scenario-specific connectivity or on habitat that remains accessible only under the current climate (Ferrier et al., 2020; Harwood et al., 2022).

5 Computational implementation

5.1 Multi-resolution efficiency gains

The present implementation reformulates the distance-dependent aggregation principle (Section 3) around globally anchored raster overviews, generated internally before focal-cell processing begins. Habitat condition, cell-area weights, and, where required, scenario-specific compositional similarity layers are aggregated once into aligned multi-resolution arrays and reused across all focal cells. This avoids repeated on-the-fly aggregation and ensures that focal-cell neighbourhoods draw from a consistent set of resolution levels regardless of which spatial tile is being processed.

Efficiency gains follow directly from this design. Shortest-path computations scale with the size of the multi-resolution neighbourhood rather than with the geographic extent of the search radius. Because coarser cells represent larger spatial extents, long-distance landscape context is captured using relatively few graph nodes, keeping computational cost and memory bounded even when the analysed neighbourhood extends across tens or hundreds of kilometres. An extended outer-window parameter at the coarsest level (Section 3.2) decouples search radius from local resolution, so distant context can be captured cheaply where it has least influence on connectivity. Path distances are computed once per focal cell and reused across multiple dispersal-scale parameters within each indicator run, further reducing redundancy without compromising consistency. In practice, the window size and maximum level should be chosen so that the neighbourhood covers the effective support of the largest λ used in the analysis; values of λ much smaller than the reachable distance add cost without changing results, while values approaching the reachable distance risk truncation.

5.2 Implementation and software architecture

The core algorithm is implemented in Rust (Klabnik and Nichols, 2023), with multi-resolution neighbourhood construction, graph generation, and Dijkstra traversal written in a compiled, strongly typed environment that supports efficient processing of large raster datasets. Parallelisation is applied at the focal-cell level: because connectivity calculations for individual focal cells are independent, focal cells are distributed across threads, with graph construction and traversal performed locally within each thread to avoid contention and minimise synchronisation overhead. Edge weights are represented internally as scaled integer types, reducing arithmetic overhead during traversal while preserving sufficient precision.

A Python interface exposes the Rust engine to standard geospatial analysis environments. Users supply raster inputs, set model parameters (dispersal scale λ, aggregation levels, window sizes, boundary mode, etc.), and retrieve indicator outputs through Python, while the Rust backend handles all computationally intensive operations. This separation allows the framework to integrate directly with established geospatial libraries and analysis pipelines without compromising on the performance of the underlying computation.

5.3 Memory management, tiling, and scalability

To support application to large spatial domains, the implementation can operate on spatial polygon tiles provided by the user, allowing large rasters to be processed in separate spatial units rather than requiring the full dataset to be loaded into memory. Each polygon tile defines the target region for which output values are produced. For each tile, the implementation reads an expanded area around the target region before running the connectivity calculation; these internally managed padded cells provide additional landscape context for multi-resolution neighbourhood construction and least-cost traversal, ensuring that paths near polygon boundaries are not artificially truncated. After computation, outputs are cropped back to the target polygon extent.

To ensure consistency among independently processed tiles, coarser-resolution levels are generated using globally anchored aggregation windows. Aggregation is aligned to the original raster grid rather than to the local extent of each tile, so the same base-resolution cell contributes to the same coarser-resolution block regardless of which tile is being processed. This prevents discontinuities that would otherwise arise if adjacent tiles generated their multi-resolution representations using tile-local origins.

Two boundary modes are supported. Under a **closed-boundary**, cells outside the polygon tile are excluded from graph construction and least-cost traversal; this is appropriate when the polygon represents a true analytical or ecological boundary (e.g. a national park or administrative unit). Under an **open-boundary**, padded cells outside the polygon are available for neighbourhood construction and path traversal; this mode is recommended when processing large continuous rasters as separate polygon tiles, because it reduces edge effects and preserves continuity across tile boundaries.

By combining internally managed padding, tiled processing, and the multi-resolution representation, the framework enables efficient processing of very large rasters while maintaining consistency with full-domain connectivity calculations. Memory usage scales with tile size and neighbourhood extent rather than with the total spatial extent of the dataset, allowing application to continental or global analyses (e.g. Ware et al., 2026), provided tile padding is sufficiently large relative to the neighbourhood and dispersal parameters used.

6 Case study: connectivity-based indicators for Tasmania

To demonstrate the application of the framework, we calculated habitat connectedness, PARC-connectedness, and BERI for Tasmania, Australia. Tasmania provides a useful test case because it contains strong spatial gradients in habitat condition, protected-area coverage, topographic complexity, and projected climate-related compositional change, while remaining compact enough for clear visualisation. The global implementation of these indicators is presented in Ware et al. (2026).

All indicators were derived from the same habitat-condition raster, which was used as the base condition surface for connectivity calculations. PARC-connectedness additionally used a raster representing the proportion of each cell covered by protected areas (UNEP-WCMC and IUCN, 2026), with the effective condition surface adjusted as described in Section 4.2. BERI additionally used transformed environmental grids derived from a generalised dissimilarity model (GDM) fitted to baseline climate averages centred on 1990 and projected to multiple future climate scenarios (Ware et al., 2026). The future scenario layers used here were four representative concentration pathways (RCPs) under the IPSL-CM5A-LR general circulation model (GCM) (2.6, 4.5, 6.0, 8.5), and an additional two GCMs under RCP 8.5: ACCESS1-0 and GFDL-CM3 (as per Ferrier et al., 2020).

The example was run using a local window size of 5 cells, an outer window size of 5 cells, and multi-resolution levels $f = 2, 4, 8, 16$, and 32. Given a base raster resolution of approximately 0.9 km, this corresponds to a maximum neighbourhood radius of approximately

$$0.9 \times 32 \times 5 = 144 \text{ km}. \tag{24}$$

This radius defines the broadest landscape context considered by the multi-resolution neighbourhood. Connectivity metrics were calculated using three dispersal-scale parameters, $\lambda = 2$, 20 and 50 km. The model evaluates all three indicators within the same run and returns a single output raster representing the average connectedness across these dispersal scales. This provides a simple multi-scale summary of connectivity, rather than a result tied to a single assumed dispersal distance. Figure 5 shows the habitat-condition input surface and four derived connectivity outputs: habitat connectedness, connected habitat, PARC-connectedness, and BERI.

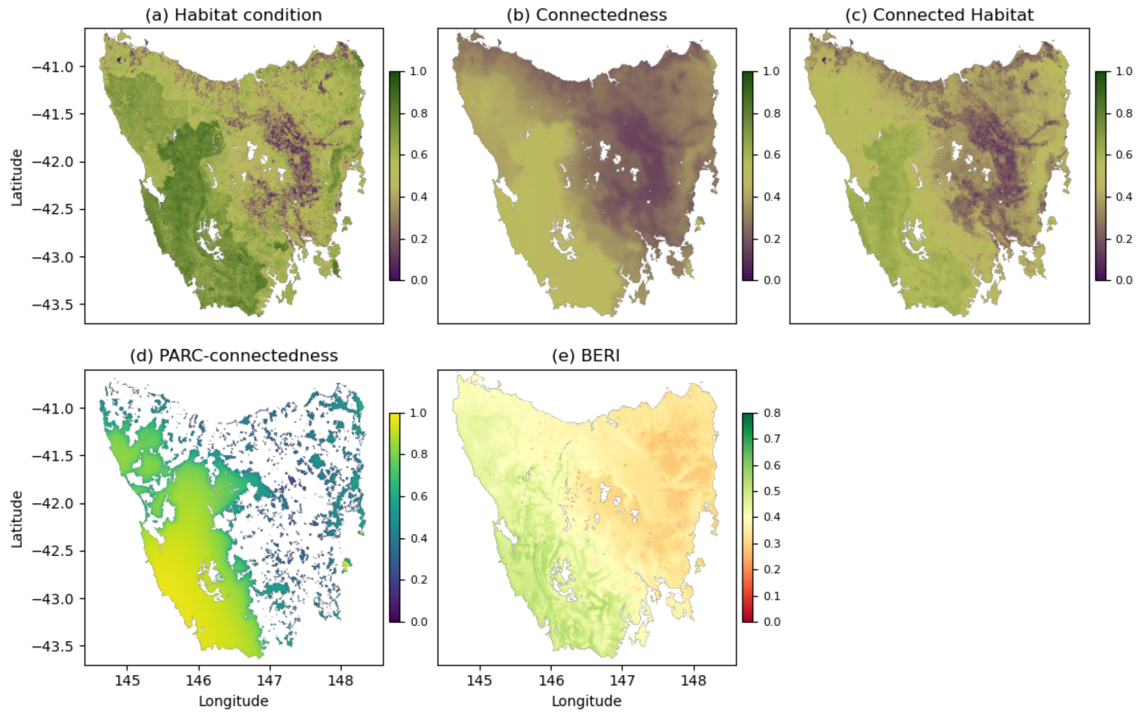

Figure 5: Example input and indicator outputs for Tasmania. The habitat condition raster (panel a) is the primary input surface used in all connectivity calculations. Derived outputs are shown for connectedness (b), connected habitat (c), PARC-connectedness (d), and BERI (e). Connected habitat combines local habitat condition with connectedness using the geometric mean defined in Section 4.1. PARC-connectedness additionally uses protected-area proportion, and BERI uses GDM-derived compositional similarity grids across future climate scenarios. Outputs were calculated using window size = 5, outer window = 5, levels $f = 2, 4, 8, 16, 32$, and $\lambda = 2, 10, 50$ km. Each derived connectivity output represents the average across the three λ values.

The habitat condition input surface shows strong spatial structure, with higher condition across much of western and south-western Tasmania and lower values across more modified eastern and central landscapes. Connectedness produces a smoother spatial pattern than raw condition because each focal cell is evaluated in relation to accessible surrounding habitat, rather than only its local condition. Connected habitat combines these two components, retaining fine-scale variation from the condition surface while reducing scores where surrounding connectedness is low.

PARC-connectedness highlights a different aspect of the landscape. Because the metric is evaluated only for cells with non-zero protected-area coverage and uses protected-area proportion to modify the habitat surface, high values are concentrated in and around the protected-area network. However, least-cost paths are still computed through the full landscape, so unprotected areas can influence protected-cell connectedness by acting as permeable or degraded matrix. The resulting pattern therefore reflects both protected-area configuration and the condition of the surrounding landscape.

BERI adds a climate-resilience interpretation. Values are highest where habitat condition, connectedness and projected compositional similarity jointly support the capacity of ecosystems to retain biodiversity under the evaluated climate scenarios. Lower values indicate areas where this capacity is limited by lower condition, weaker connectedness, lower projected similarity, or stronger dependence on favourable individual scenarios. The BERI surface should therefore not be interpreted as habitat condition or connectivity alone, but as an integrated indicator of condition-weighted accessibility to climatically and compositionally relevant habitat.

For the Tasmania case study, habitat connectedness and BERI were computed at approximately 900 m resolution across more than 300,000 raster cells, requiring approximately 6 s and 10 s processing time, respectively, on a personal laptop. These timings are intended only as indicative case-study runtimes; more systematic performance benchmarks, including a continental implementation, are provided in the Supporting Information.

7 Discussion

The framework presented here advances indicator-based connectivity assessment by making established cost-benefit connectivity formulations practical for repeated large-scale application. Its main advances are architectural: a globally anchored multi-resolution graph that preserves fine local detail while representing distant landscape context efficiently, reusable path-level outputs that avoid repeated shortest-path computation across dispersal scales and scenarios, and an open Python–Rust implementation that can be integrated into reproducible geospatial workflows. These advances expand the spatial extent, scenario breadth, and update frequency over which habitat connectedness, PARC-connectedness, and BERI can be applied consistently.

7.1 Interpretation and scope

The framework produces estimates of potential landscape connectivity, not predictions of realised movement. Outputs should be interpreted as repeatable, landscape-derived indicators of how habitat condition and spatial configuration may facilitate or constrain broad ecological processes, rather than as empirical measures of functional connectivity for specific taxa (Tischendorf and Fahrig, 2000; Zeller et al., 2012). This framing is particularly important for biodiversity indicator applications, where the objective is consistent, comparable assessment across ecosystems and taxa rather than species-specific movement prediction (Drielsma et al., 2022; Ware et al., 2026).

The Tasmania case study illustrates how the same processing pipeline produces related but ecologically distinct indicator surfaces. Habitat connectedness adjusts the raw condition surface by integrating accessible surrounding habitat. PARC-connectedness emphasises the protected-area network while still reflecting matrix condition, since traversal is calculated through the full landscape. BERI provides a climate-resilience interpretation by weighting accessible habitat according to projected compositional similarity from the present to alternative future scenarios. These differences confirm the value of separating path computation from indicator aggregation: a single representation of landscape accessibility supports multiple ecological interpretations without redundant computation.

7.2 Assumptions and limitations

The use of least-cost paths provides deterministic and interpretable effective distances but does not capture pathway redundancy, diffuse movement, or simultaneous use of multiple routes (McRae et al., 2008). Applications concerned with these properties may be better served by circuit-theory approaches, which represent movement as a diffusion process across all pathways (McRae et al., 2008), or resistant-kernel methods, which estimate diffuse movement potential without restricting connectivity to optimal paths (Compton et al., 2007; Zeller et al., 2012). The present framework is most appropriate where repeatable, large-scale assessment of potential connectivity is the goal rather than detailed modelling of movement behaviour.

A related limitation is sensitivity to the condition-to-resistance transformation. Habitat condition influences both effective distance and destination weighting, providing internal consistency but also a modelling assumption. Ecological responses to degradation and land-cover composition may be nonlinear, threshold-dependent, or strongly taxon-specific (Adriaensen et al., 2003; Zeller et al., 2012). Future applications should evaluate alternative transformations, including steeper nonlinear functions or empirically calibrated resistance surfaces where movement data are available.

For BERI specifically, compositional similarity under future climate scenarios is incorporated, but habitat condition and resistance surfaces are treated as fixed under current conditions (Ferrier et al., 2020). The dispersal kernel, the median dispersal distance λ, and the multi-generation parameter g are user-specified rather than fixed by the method, so alternative kernels or dispersal assumptions can be applied to suit the application and time horizon. Outputs should therefore be interpreted as the capacity of the current landscape configuration to support compositional redistribution under climate change, rather than as a forecast of future realised connectivity. Where future condition surfaces are available, the framework could be extended to evaluate scenario-specific resistance and path distances.

Finally, parameterisation requires care. The maximum reachable distance is determined jointly by base-resolution, aggregation levels, and neighbourhood size. Sensitivity analysis across these parameters and dispersal-scale values is important, particularly when indicators are used for regional comparison or policy reporting (Drielsma et al., 2022).

7.3 Future directions

Three areas warrant further development. First, systematic benchmarking against single-resolution implementations is needed to quantify the trade-off between computational efficiency and approximation error across different landscape types, dispersal scales, and neighbourhood configurations. Second, additional dispersal kernels and resistance transformations should be incorporated and evaluated; because path computation is separated from metric aggregation, these can be tested without recomputing graph traversal, lowering the cost of comparing alternative ecological assumptions. Third, while the framework has already been applied at global scale in operational indicator workflows (Ware et al., 2026), further development of time-series and scenario-based applications would strengthen its role in tracking connectivity change through time, supporting adaptive management and periodic biodiversity reporting under land-use and climate change.

8 Conclusion

We present a multi-resolution analytical framework that reformulates established cost-benefit connectivity logic around globally anchored, pre-computed raster overviews, and separates path computation from indicator aggregation. This design enables habitat connectedness, PARC-connectedness, and BERI to be derived consistently from a common processing pipeline, with path distances reused across dispersal scales and climate scenarios within each indicator run. The framework makes well-validated connectivity concepts applicable at the spatial extents and indicator-update frequencies demanded by contemporary biodiversity monitoring, and is distributed as an open Python package with a high-performance Rust engine to support reproducible use, independent scrutiny, and extension. Its application within the global indicator suite of Ware et al. (2026) demonstrates operational viability at planetary scale.

Code Availability

The connectivity library is openly available on GitHub at `https://github.com/csiro/connectivity`. The data used in the examples are available in the library. The code and data for producing the figures in this study are provided in the Supporting Information.

Acknowledgements

This research is supported by the Science and Industry Endowment Fund.

References

Frank Adriaensen, J. Paul Chardon, Geert De Blust, Erik Swinnen, Sabine Villalba, Hubert Gulinck, and Erik Matthysen. The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, 64(4):233–247, 2003.

B. A. Bryan, M. Nolan, T. D. Harwood, J. D. Connor, J. Navarro-Garcia, D. King, D. M. Summers, D. Newth, Y. Cai, N. Grigg, I. Harman, N. D. Crossman, M. J. Grundy, J. J. Finnigan, S. Ferrier, K. J. Williams, K. A. Wilson, E. A. Law, and S. Hatfield-Dodds. Supply of carbon sequestration and biodiversity services from australia's agricultural land under global change. *Global Environmental Change*, 28:166–181, September 2014. doi: 10.1016/j.gloenvcha.2014.06.013.

Justin M. Calabrese and William F. Fagan. A comparison-shopper's guide to connectivity metrics. *Frontiers in Ecology and the Environment*, 2(10):529–536, 2004. doi: 10.1890/1540-9295(2004)002[0529:ACGTCM]2.0.CO;2.

CBD, UN. Kunming-montreal global biodiversity framework. Fifteenth Meeting of the Conference of the Parties to the Convention on Biological Diversity (Part Two) Decision 15(4), 2022.

Bradley W. Compton, Kevin McGarigal, Samuel A. Cushman, and Lloyd R. Gamble. A resistant-kernel model of connectivity for amphibians that breed in vernal pools. *Conservation Biology*, 21(3):788–799, 2007. doi: 10.1111/j.1523-1739.2007.00674.x.

Michael Drielsma, Simon Ferrier, and Glenn Manion. A raster-based technique for analysing habitat configuration: The cost–benefit approach. *Ecological Modelling*, 202(3–4):324–332, 2007a. doi: 10.1016/j.ecolmodel.2006.10.016.

Michael Drielsma, Glenn Manion, and Simon Ferrier. The spatial links tool: Automated mapping of habitat linkages in variegated landscapes. *Ecological Modelling*, 200(3–4):403–411, 2007b. doi: 10.1016/j.ecolmodel.2006.08.017.

Michael J. Drielsma, Jamie Love, Subhashni Taylor, Rajesh Thapa, and Kristen J. Williams. General landscape connectivity model (glcm): A new way to map whole of landscape biodiversity functional connectivity for operational planning and reporting. *Ecological Modelling*, 465:109858, March 2022. doi: 10.1016/j.ecolmodel.2021.109858.

Simon Ferrier, Glenn Manion, Jane Elith, and Karen Richardson. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13(3):252–264, 2007. doi: 10.1111/j.1472-4642.2007.00341.x.

Simon Ferrier, Thomas D. Harwood, Chris Ware, and Andrew J. Hoskins. A globally applicable indicator of the capacity of terrestrial ecosystems to retain biological diversity under climate change: The bioclimatic ecosystem resilience index. *Ecological Indicators*, 117:106554, October 2020. doi: 10.1016/j.ecolind.2020.106554.

Tom Harwood, Jamie Love, Michael Drielsma, Clare Brandon, and Simon Ferrier. Staying connected: Assessing the capacity of landscapes to retain biodiversity in a changing climate. *Landscape Ecology*, 37(12):3123–3139, 2022. doi: 10.1007/s10980-022-01534-5.

IPBES. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. IPBES Secretariat, Bonn, Germany, 2019.

Maëlis Kervellec, Thibaut Couturier, Sarah Bauduin, Delphine Chenesseau, Pierre Defos du Rau, Nolwenn Drouet-Hoguet, Christophe Duchamp, Julien Steinmetz, Jean-Michel Vandel, and

Olivier Gimenez. Bringing circuit theory into spatial occupancy models to assess landscape connectivity. *Methods in Ecology and Evolution*, 15(11):2141–2152, 2024. doi: https://doi.org/10.1111/2041-210X.14418.

Steve Klabnik and Carol Nichols. *The Rust Programming Language*. No Starch Press, 2023.

Brad H. McRae, Brett G. Dickson, Timothy H. Keitt, and Viral B. Shah. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, 89(10):2712–2724, 2008. doi: 10.1890/07-1861.1.

Karel Mokany, Chris Ware, Skipton N. C. Woolley, Simon Ferrier, and Matthew C. Fitzpatrick. A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, 31(4):802–821, 2022. doi: 10.1111/geb.13459.

Ran Nathan, Etienne Klein, J. J. Robledo-Arnuncio, and Eloy Revilla. Dispersal kernels: Review. In *Dispersal Ecology and Evolution*. January 2012.

Reed F. Noss. From plant communities to landscapes in conservation inventories: A look at the nature conservancy (usa). *Biological Conservation*, 41(1):11–37, 1987. doi: 10.1016/0006-3207(87)90045-0.

Natalia Ocampo-Peñuela, Luca de Alfaro, Jasmine Tai, Natalie Valett, Coen Adler, and Artie Nazarov. Ecoscape: A model for species-specific functional habitat connectivity. *Research Square (Preprint)*, November 2023. doi: 10.21203/rs.3.rs-3650354/v1.

Santiago Saura and Lucía Pascual-Hortal. A new habitat availability index to integrate connectivity in landscape conservation planning: Comparison with existing indices and application to a case study. *Landscape and Urban Planning*, 83(2–3):91–103, 2007. doi: 10.1016/j.landurbplan.2007.03.005.

Santiago Saura and Josep Torné. Conefor sensinode 2.2: A software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Modelling & Software*, 24(1):135–139, 2009. doi: 10.1016/j.envsoft.2008.05.005.

Qiuling Tang and Wanfeng Dou. An effective method for computing the least-cost path using a multi-resolution raster cost surface model. *ISPRS International Journal of Geo-Information*, 12(7):287, 2023. doi: 10.3390/ijgi12070287.

Philip D. Taylor, Lenore Fahrig, Kringen Henein, and Gray Merriam. Connectivity is a vital element of landscape structure. *Oikos*, 68(3):571–573, 1993. doi: 10.2307/3544927.

Lutz Tischendorf and Lenore Fahrig. On the usage and measurement of landscape connectivity. *Oikos*, 90(1):7–19, 2000. doi: 10.1034/j.1600-0706.2000.900102.x.

UNEP-WCMC. Indicators for the Kunming-Montreal Global Biodiversity Framework: Metadata factsheet for the Protected Area Representativeness and Connectedness (PARC) Indices, 2023. URL https://gbf-indicators.org/metadata/other/A-5-C.

UNEP-WCMC and IUCN. Protected planet: The world database on protected areas (wdpa). UNEP-WCMC and IUCN, Cambridge, UK, 2026.

Dean Urban and Timothy Keitt. Landscape connectivity: A graph-theoretic perspective. *Ecology*, 82(5):1205–1218, 2001. doi: 10.1890/0012-9658(2001)082[1205:LCAGTP]2.0.CO;2.

Chris Ware, Roozbeh Valavi, Mathew Vickers, et al. Global indicator data tracking changes in species-level biodiversity persistence, ecosystem resilience under climate change, and protected-area representativeness and connectedness. *EcoEvoRxiv (Preprint)*, 2026. doi: 10.32942/X26Q2G.

Katherine A. Zeller, Kevin McGarigal, and Andrew R. Whiteley. Estimating landscape resistance to movement: A review. *Landscape Ecology*, 27(6):777–797, 2012. doi: 10.1007/s10980-012-9737-0.

Supporting Information for
A scalable multi-resolution framework for connectivity-based biodiversity indicators

Roozbeh Valavi Karel Mokany Chris Ware Mathew Vickers
Katherine M. Giljohann Simon Ferrier

S1 Performance benchmarking

The indicative runtimes reported for the Tasmania case study in the main text (habitat connectedness and BERI computed over roughly $3{\times}10^5$ cells at about 900 m resolution in a few seconds each) demonstrate feasibility on a single landscape, but do not show how cost grows with the parameters that control per-focal-cell work, or with the size of the analysed area. This section describes a small, reproducible benchmarking protocol that characterises that scaling for both pure connectedness and the Bioclimatic Ecosystem Resilience Index (BERI). The benchmarks are deliberately simple and interpretable: each suite varies one parameter family against an otherwise fixed baseline, so that the measured change in runtime can be attributed to that parameter alone. Full continental and global indicator results are reported by Ware et al. (2026); the purpose here is to document performance behaviour, not to reproduce those results.

S1.1 Indicators, study area, and input data

Two entry points are benchmarked: pure connectedness (the `connectedness()` function with `option=1`, returning connectedness rather than connected habitat) and BERI (the `beri()` function). Connectedness requires only a habitat-condition raster; BERI additionally requires a current-climate transform grid and one or more future-scenario transform grids derived from generalised dissimilarity modelling (GDM), as described in the main text.

All benchmarks use a continental habitat-condition raster for Australia, masked to the Australian land area, as the base condition surface. For BERI, the current climate is represented by a baseline transform grid centred on 1990, and the future climate by up to six scenario transform grids: four representative concentration pathways under the IPSL-CM5A-LR general circulation model (RCP 2.6, 4.5, 6.0, 8.5) and two further models under RCP 8.5 (ACCESS1-0 and GFDL-CM3), matching the scenario set used in the main text. Using the full continental extent ensures that the benchmark reflects realistic input sizes rather than a small illustrative tile.

S1.2 Benchmark design

All runs share a single baseline configuration, against which each suite varies one parameter family (Table 1). The baseline uses internally generated overview levels $f \in \{2, 4, 8, 16, 32\}$, three dispersal-scale parameters $\lambda = \{2,\ 20,\ 200\}$ km, a local window of 5 cells with an outer window of 7 cells, a maximum-cost parameter $\mu = 2$, and the circular window mode. Path distances are computed once per focal cell and reused across the three λ values within each run, consistent with the algorithm described in the main text.

Suite	Parameter varied	Values
Baseline	— (fixed reference)	levels $\{2, 4, 8, 16, 32\}$; $\lambda = \{2, 20, 200\}$ km; window 5, outer window 7; $\mu = 2$; circular mode
Overview levels	maximum overview level	8, 16, 32, 64, 128 (cumulative from $f = 2$)
Dispersal kernels	number of λ values	1, 3, 5, 10, 20 (log-spaced over 1–200 km)
Window size	(window, outer window)	3×3, 5×5, 7×7, 9×9, 3×9, 5×11, 7×15, 9×21
Analysed area	square side length	100, 250, 500, 1000, 2000 km, and full Australia
Climate scenarios	number of future scenarios	1, 2, 4, 6 (BERI only)

Table 1: Benchmark suites. Each suite varies a single parameter family while holding all other parameters at the baseline. The *overview levels* suite uses cumulative level sets (e.g. a maximum level of 64 corresponds to $f \in \{2, 4, 8, 16, 32, 64\}$); the *dispersal kernels* suite keeps the three-value baseline when the count is three and otherwise uses log-spaced values spanning 1–200 km; and the *climate scenarios* suite applies only to BERI, since connectedness does not use scenario transform grids.

For the area-scaling suite, the analysed region is a square centred on the continental centroid, constructed in an equal-area coordinate reference system (EPSG:6933) and transformed back to the raster's coordinate system, with side lengths of 100 to 2000 km, plus the full Australian mask (Figure S1). This isolates the effect of analysed extent while keeping the landscape context comparable across sizes.

Fixed-area protocol. Every benchmark case analyses only the raster area inside the polygon mask, using `closed_border=True` and `margin_px=0`. This is essential for clean attribution: under the default open-boundary behaviour, the read window is padded by an amount that grows with the outer window and the maximum overview level, so increasing `levels` or `outer_window` would silently enlarge the area actually processed and confound the windows, levels, and area suites. Closing the boundary and removing the margin fixes the assessed area to the masked polygon, so that a swept parameter changes only the per-cell work and not the size of the problem. Note that this fixed-area setting is a benchmarking control; for production tiled analyses the open-boundary mode remains recommended (Section S1 notwithstanding, see the tiled-analysis guidance below Section S4.4).

Recording problem size. Because bounding-box side length and nominal area can be misleading measures of computational work for irregular masks, study-area size is reported using the number of valid output cells actually analysed inside the polygon (`valid_cells`). This cell count provides a direct measure of the raster workload and avoids conflating runtime with empty or masked regions outside the analysed area. The nominal study-area width and area descriptors are retained as metadata, but the study-area scaling results are shown against analysed-cell count.

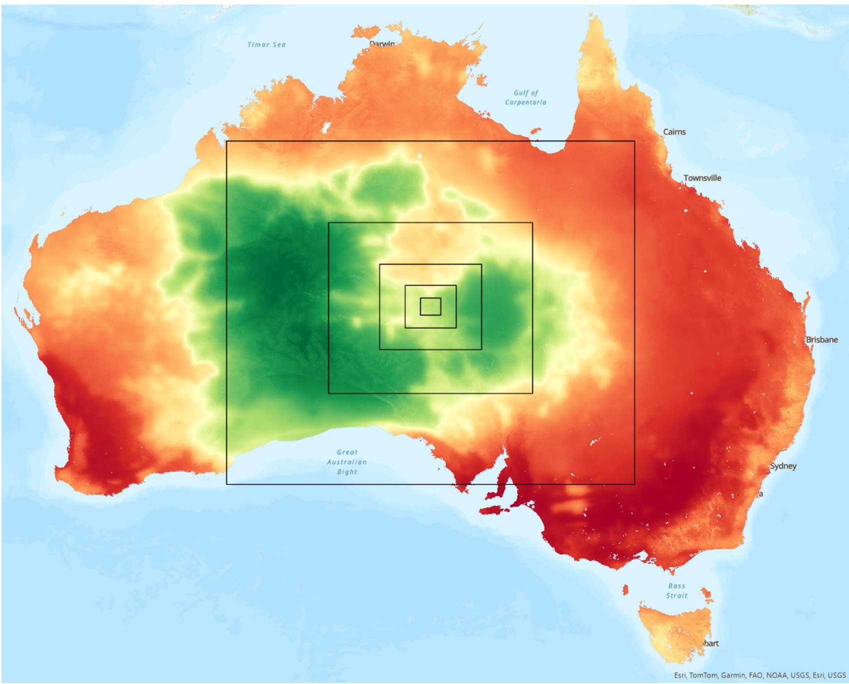

Figure S1: Continental Australian study area used for the performance benchmarks, shown as an example indicator surface. The benchmark suites analyse the full Australian land mask; the area-scaling suite additionally uses centred square study areas with side lengths of 100 to 2000 km. The background map shows BERI values across Australia, with darker green indicating higher values and darker red indicating lower values.

S1.3 Measurement protocol

For each case the runner records wall-clock runtime (elapsed seconds and minutes), peak resident set size, the thread count used , and the output raster shape, together with the parameter settings and the area descriptors above. Each case is repeated 3 times, so points show the mean and error bars show one standard deviation across repeats.

Parallelism is applied at the focal-cell level (main text), and the thread count is held fixed within each sweep and recorded for every run (`n_threads`); the benchmarks reported here do not include a separate thread-scaling experiment. Peak resident memory is reported below (Expected scaling and interpretation).

Peak resident memory was governed by the analysed extent and the number of input grids rather than by the per-cell parameters. For pure connectedness it stayed at or below approximately 1 GB across every configuration, rising from about 0.2 GB for the smallest study areas to about 0.9 GB for the full continental extent (∼10.2 million analysed cells). BERI uses substantially more memory because it additionally loads the current- and future-climate transform grids: per-case memory scaled from about 0.4 GB for the smallest study area to about 39 GB for the full continental extent with six future scenarios, with an overall peak of 42 GB across all benchmarked BERI configurations. All runs therefore remained within the 64 GB allocated per job.

Study area	Analysed cells	Connectedness (GB)	BERI (GB)
100 km square	13,230	0.17	0.44
250 km square	81,432	0.20	0.52
500 km square	324,583	0.21	1.08
1000 km square	1,296,234	0.23	2.76
2000 km square	4,798,530	0.33	8.81
Full Australia	10,162,074	0.91	39.2

Table 2: Peak resident memory as a function of analysed extent, for pure connectedness and BERI, on the 64 core / 64 GB benchmark node. Analysed cells are the finite output cells inside each mask (`valid_cells`); area cases are squares centred on the continental centroid. Memory scales with analysed extent and input size rather than with the per-cell parameters, and BERI is far heavier than connectedness because it loads the climate transform grids. Because each benchmark process executed several cases and `ru_maxrss` reports a process-lifetime maximum, per-case memory is taken as the minimum value recorded for each case across the three repeats, which removes the contribution of any heavier case run earlier in the same process. All runs remained within the 64 GB allocation.

All benchmarks were run on a single node of a high-performance computing (HPC) cluster, with each job allocated 64 CPU cores and 64 GB of RAM; the connectivity engine was configured to use all 64 allocated cores (`n_threads` = 64). Exact node hardware may vary between scheduler allocations.

S1.4 Expected scaling and interpretation

Three aspects of the framework's design predict the observed scaling. First, per-focal-cell cost is governed by the number of nodes in the multi-resolution neighbourhood, which grows with the window and outer-window sizes and with the number of overview levels; because coarser levels represent larger areas with few nodes, extending reach through additional levels or a larger outer window adds comparatively little work. Second, because shortest-path distances are computed once per focal cell and reused, increasing the number of dispersal kernels λ adds only kernel-aggregation work rather than repeated graph traversal, so runtime is expected to grow weakly with the number of λ values; for BERI, each additional climate scenario adds per-destination transform-grid work and is expected to contribute a larger marginal cost. Third, because focal cells are independent, total runtime is expected to grow approximately in proportion to the number of analysed cells (and hence analysed area) at fixed per-cell parameters, while memory scales with neighbourhood and tile size rather than total extent.

Figure S1 shows the continental study area and the polygon bounding box of study ares used for the benchmarks. Figure S2 summarises runtime scaling for connectedness and BERI across the benchmarked parameters; comprehensive continental and global outputs are reported by Ware et al. (2026).

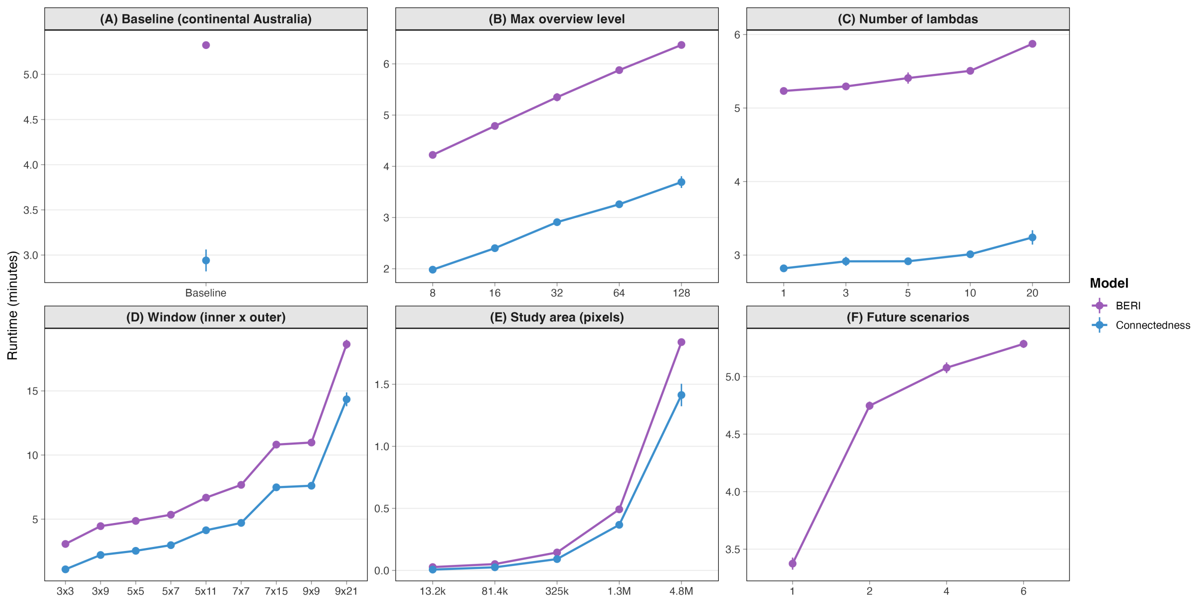

Figure S2: Runtime scaling of connectedness and BERI across the benchmarked parameters. Each panel shows mean wall-clock runtime (points; error bars show ± 1 standard deviation across repeats) as a function of (A) the baseline configuration, (B) the maximum overview level, (C) the number of dispersal kernels λ, (D) the window/outer-window size, (E) the analysed area (equal-area km^2), and (F) the number of future climate scenarios (BERI only). All other parameters are held at the baseline, and every case analyses a fixed area (`closed_border=True`, `margin_px=0`).

S2 Least-cost paths within the multi-resolution graph

For each focal cell, the algorithm evaluates all reachable cells or nodes within the specified multi-resolution search window. The neighbourhood is represented as a graph, with fine-resolution nodes close to the focal cell and progressively coarser nodes at greater distances. This structure allows broad spatial context to be included while avoiding the computational cost of evaluating all cells at the base raster resolution.

Least-cost paths from the focal cell to all reachable nodes in this graph are calculated using Dijkstra's algorithm. Edge costs are derived from the habitat condition surface, so movement through low-condition cells has a higher cost than movement through high-condition cells. The shortest path to each reachable node is then used in the connectedness, PARC-connectedness, or BERI metric calculation. Thus, the final value for a focal cell is not based on a single path, but on the contribution of all reachable cells in the multi-resolution window, weighted by their least-cost distance and the relevant metric parameters.

The figure illustrates one example least-cost path from a focal cell to an outer node in the search window. The displayed path is only one destination-specific path; in the actual calculation, equivalent shortest paths are calculated from the focal cell to all reachable nodes within the window, and all of these reachable nodes contribute to the final metric value.

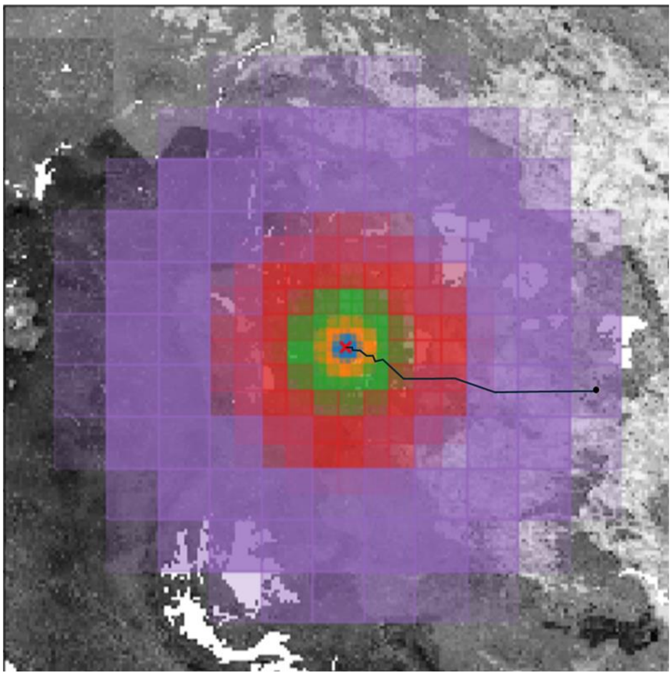

Figure S3: Example of least-cost routing within the multi-resolution search window. The focal cell is shown near the centre, with surrounding graph nodes represented at increasingly coarser resolution away from the focal cell. The black line shows one shortest path from the focal cell to an outer node, calculated through the multi-resolution graph using Dijkstra's algorithm. Although only one path is shown for clarity, the metric calculation uses the shortest paths from the focal cell to all reachable nodes in the search window.

S3 Behaviour at coastlines and domain boundaries

A common concern with neighbourhood-based connectivity metrics is the appearance of edge artefacts where the analysis window is truncated by a coastline, a nodata region, or the boundary of the analysis domain. In this framework such truncations do not bias the indicators, because connectedness, PARC-connectedness, and BERI are each expressed as a ratio of a condition-adjusted sum to an intact-reference sum evaluated over the *same* set of reachable nodes.

For a focal cell i, connectedness (main text, "Habitat connectedness") is

$$C_i = \frac{S_i^{\mathrm{adj}}}{S_i^{\mathrm{int}}}, \qquad S_i^{\mathrm{adj}} = \sum_j e^{-D_{ij}^{\mathrm{adj}}/\lambda}\, h_j\, \widetilde{A}_j, \qquad S_i^{\mathrm{int}} = \sum_j e^{-D_{ij}^{\mathrm{int}}/\lambda}\, \widetilde{A}_j, \tag{1}$$

where both sums run over the same reachable destination nodes j in the multi-resolution neighbourhood. Cells that fall on ocean or nodata are simply not part of the graph, so they are absent from both the numerator and the denominator. The intact reference S_i^{int} therefore represents the maximum connectedness attainable at that location given the same configuration of valid cells – that is, the same coastline or boundary – with habitat condition set to one everywhere.

Because the missing area is excluded from both terms, it cancels in the ratio: a focal cell on the coast is evaluated relative to its own achievable maximum, not relative to a hypothetical complete neighbourhood of land. Connectedness near a coastline thus reflects the condition and configuration of the habitat that is actually present, and does not decline simply because part of the neighbourhood is ocean (Figure S4). This distinguishes the relative formulation used here from absolute measures of reachable habitat area, which necessarily decline towards edges purely because less land is within reach; no coastline buffering or edge correction is required.

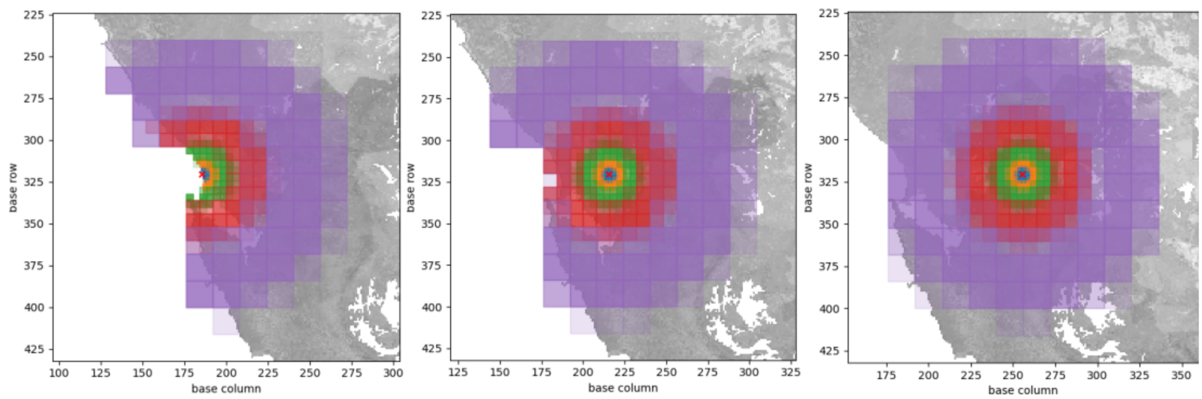

Figure S4: Multi-resolution neighbourhood at a coastline, shown for a focal cell (red marker) at increasing distance from the coast over the habitat basemap (grey; ocean and nodata in white). In the left panel the focal cell lies on the coast and the neighbourhood is clipped by the ocean; in the middle and right panels it is progressively complete. Because connectedness is a ratio evaluated over the same valid nodes in numerator and denominator, the clipped neighbourhood in the left panel still yields a correctly scaled value.

This property applies to genuine domain boundaries, where the absent cells represent a real absence of habitat (ocean, nodata, or a true ecological boundary). It is distinct from the truncation of paths at the edge of a processing tile that cuts through continuous land, which is handled separately by reading a padded margin around each tile under the open-boundary mode (Section S4.4), so that paths crossing a tile boundary are not artificially severed.

S4 Practical guidance for running analyses

S4.1 How `window_size`, `outer_window`, and `levels` combine

The multi-resolution neighbourhood (main text, "Neighbourhood construction across spatial scales") is controlled by three settings that together fix how far the analysis reaches and at what resolution each part of that reach is represented.

- `levels` – the aggregation factors f (powers of two) used for the coarser bands. The base level $f = 1$ is always added automatically, so `levels=[2,4,8,16,32]` builds six nested bands $f \in \{1, 2, 4, 8, 16, 32\}$. The number of entries sets the number of bands, and the largest entry, $f_{\max}$, sets the coarsest resolution used.
- `window_size` (w) – sets the radial thickness of every band. At level f, the band occupies base-pixel radii $[\,fw/2,\ fw\,]$ from the focal cell, except the base level ($f = 1$), which is a filled disc of radius w. Equivalently, each band is the ring between radius $w/2$ and w measured in that band's *own* cells; because both the radius and the cell size scale with f, every band contains a comparable, bounded number of cells regardless of resolution.
- `outer_window` ($w_{\text{outer}} \geq w$) – applies only to the coarsest band. It replaces w with w_{outer} at $f_{\max}$, extending the outer radius of the final ring from $f_{\max}w$ to $f_{\max}w_{\text{outer}}$. This lets the neighbourhood reach further using only coarse-resolution cells, without enlarging every band.

The per-band radii are therefore

$$\rho_{\text{in}}^{(f)} = \begin{cases} 0, & f = 1 \\ fw/2, & f > 1 \end{cases} \qquad \rho_{\text{out}}^{(f)} = \begin{cases} fw, & f < f_{\max} \\ fw_{\text{outer}}, & f = f_{\max}, \end{cases} \tag{2}$$

in base-pixel units, and consecutive bands meet exactly because $\rho_{\text{out}}^{(f)} = fw = \rho_{\text{in}}^{(2f)}$. The maximum one-sided search reach is set jointly by the coarsest level and the outer window,

$$\text{maximum reach} = f_{\max} \times w_{\text{outer}} \times \text{base raster resolution}, \tag{3}$$

so it is *not* `outer_window` alone: `outer_window` controls how many coarsest-level cells the final ring spans, while `levels` controls how large those cells are. Both `window_size` and `outer_window` are odd cell counts.

For example, with a 1 km raster, `levels=[2,4,8,16,32]`, `window_size=5`, and `outer_window=11`, the bands are (Table 3) and the maximum reach is $32 \times 11 \times 1$ km $= 352$ km.

Level f	Cell size (km)	Inner radius ρ_{in} (km)	Outer radius ρ_{out} (km)
1	1	0	5
2	2	5	10
4	4	10	20
8	8	20	40
16	16	40	80
32	32	80	352

Table 3: Worked example of how `window_size`, `outer_window`, and `levels` combine, for a 1 km raster with `levels=[2,4,8,16,32]`, `window_size=5`, and `outer_window=11`. Each band is a nested ring around the focal cell; the base level ($f = 1$) is a filled disc. The coarsest band ($f = 32$) is widened by `outer_window`, extending the maximum one-sided reach to 352 km. Radii equal kilometres here because the base resolution is 1 km.

In practice: increase `outer_window` to extend reach cheaply by adding distant, coarse context; increase `window_size` for finer-grained context within every band, at higher cost; and add or remove entries in `levels` to change how many resolution steps span that reach. The transition distance from one resolution to the next always occurs at fw base pixels.

S4.2 Selecting resolution levels and maximum search reach

Before running a full analysis, we recommend using `resolution_info()` to inspect the raster dimensions, the candidate internally generated levels, and the resulting maximum search reach. Supplying the same `outer_window` planned for the analysis returns the reach that the chosen `levels` and `outer_window` will actually produce (Section S4.1), so the configuration can be checked before committing to a long run.

```python
from connectivity import resolution_info

condition_file = "./Tas_site_condition.tif"
resolution_info(condition_file,  outer_window=7)
```

✓ 0.0s Python

```
File: ./Tas_site_condition.tif
Base raster: 588 x 516 cells (303.41K total)
Base resolution: 0.008333 degrees
CRS: EPSG:4326
Bands: 1
Data type(s): float32
Nodata: -9999.0
Outer window: 7
Small-dimension warning threshold: < 16 cells
Distance note: approximate km values use centre latitude -41.550

Candidate internally generated levels:
  Level    Width   Height      Cells   % base          Resolution          Reach     Reach km  Note
      1      588      516    303.41K 100.000%    0.008333 degrees  0.05833 degrees     6.494 km
      2      294      258     75.85K  25.000%     0.01667 degrees   0.1167 degrees     12.99 km
      4      147      129     18.96K   6.250%     0.03333 degrees   0.2333 degrees     25.97 km
      8       74       65      4.81K   1.585%     0.06667 degrees   0.4667 degrees     51.95 km
     16       37       33      1.22K   0.402%      0.1333 degrees   0.9333 degrees     103.9 km
     32       19       17        323   0.106%      0.2667 degrees    1.867 degrees     207.8 km
     64       10        9         90   0.030%      0.5333 degrees    3.733 degrees     415.6 km  small grid; review
    128        5        5         25   0.008%       1.067 degrees    7.467 degrees     831.2 km  small grid; review
    256        3        3          9   0.003%       2.133 degrees    14.93 degrees     1.66K km  small grid; review
    512        2        2          4   0.001%       4.267 degrees    29.87 degrees     3.32K km  small grid; review
```

Figure S5: Example output from `resolution_info()` showing raster properties, candidate internally generated levels, and approximate search reach.

The main thing to watch for is that very coarse levels can leave only a small number of cells across the study area, which makes the broad-scale neighbourhood less spatially informative. Levels flagged as `small grid; review` should therefore be used cautiously. In practice, we choose the finest set of levels that still gives adequate ecological reach while avoiding very small coarse grids.

S4.3 Preparing protected-area proportions for `pa_file`

PARC-connectedness needs a protected-area surface in which each cell holds the *proportion* of that cell covered by protected areas, rather than a simple inside/outside mask. The package provides a helper, `pixel_coverage()`, that takes protected-area polygons together with a reference raster and returns, for every pixel, the fraction (between 0 and 1) of that pixel's area covered by the polygons. When the reference raster is the condition raster, the result is aligned to the same grid, extent, transform, and coordinate system, so it can be written out and used directly as the protected-area input. A minimal workflow is:

```
from connectivity import connectedness, pixel_coverage

# 1. Fraction (0-1) of each condition-raster pixel covered by
#    protected-area polygons, aligned to the condition grid.
pa_fraction = pixel_coverage("protected_areas.gpkg", "condition.tif")

# 2. Save pa_fraction as a GeoTIFF on the condition raster's grid
#    (same shape, transform, and CRS), e.g. "pa_fraction.tif".

# 3. Run PARC-connectedness with that raster as the protected-area input.
result = connectedness("condition.tif", pa_file="pa_fraction.tif")
```

Using fractional coverage rather than a presence/absence mask matters most when raster cells are large relative to protected-area boundaries, where a binary classification would over- or under-count partially protected cells. The essential requirement is simply that the protected-area input is (or be scaled to) a 0–1 coverage surface on the same grid as the condition raster; the exact helper name and call signature are secondary, and any equivalent rasterisation of protected-area coverage onto the condition grid can be used.

S4.4 Tiled analysis considerations

For large rasters, analyses can be run by passing a tile polygon to `polygon_mask`. We recommend using `closed_border=False`, which allows the function to read neighbourhood context beyond the tile boundary before cropping the result back to the tile. This eliminates edge effects at tile joins.

For tiled workflows, all tiles should use the same `levels`, `outer_window`, `window_mode`, `lambdas`, condition raster, and protected-area raster. A small overlap between neighbouring tile polygons is useful when mosaicking outputs, and `margin_px` can be increased, for example to 32, if visible seams occur at tile boundaries. Tile joins should be checked by inspecting narrow strips around boundaries after mosaicking.

Tile polygons can be generated with `make_tiles()` (or `make_tile()` for a single tile), which divides the raster into a non-overlapping `nrows`×`ncols` grid of core tiles with boundaries snapped to a multiple of `align_to` pixels (setting `align_to` to the largest aggregation factor keeps tile edges on the globally anchored overview grid). By default the tiles are equal in size; setting `balanced=True` instead places the tile breaks using a deterministic raster-mask cost estimate so that each tile carries a similar computational load. This is useful for large masks where valid cells are unevenly distributed – for example where some tiles are mostly ocean or nodata while others are dense land – and equal-size tiles would otherwise give very uneven per-tile runtimes (an `io_weight` term ensures mostly empty regions still carry some cost). The default (`balanced=False`) is sufficient when equal aligned tiles already give comparable loads.

S5 Square neighbourhood mode

The framework supports an optional square neighbourhood mode as an alternative to the default circular construction described in the main text (Section "Neighbourhood construction across spatial scales"). The two modes are identical in every respect – the same source-centred nested annuli, the same inner and outer radii $\rho_{\text{in}}^{(f)}$ and $\rho_{\text{out}}^{(f)}$ at each level f, the same fractional area weighting, and the same graph construction (queen adjacency within levels and cross-scale edges at annulus boundaries) – and differ only in how distance from the focal cell is measured.

The circular mode defines annulus membership using the Euclidean (L_2) norm, so the rings are circular. The square mode replaces this with the Chebyshev (L_∞) norm,

$$\|\mathbf{x} - \mathbf{s}\|_\infty = \max\Big(|\Delta\text{row}|,\ |\Delta\text{col}|\Big), \tag{4}$$

so the annuli become nested, axis-aligned square rings centred on the focal cell. As a result, each cell's overlap fraction $\alpha_{i,j}^{(f)}$ is computed as the intersection of its rectangular footprint with two concentric squares (an axis-aligned box intersection), rather than the rectangle–disc (circular-segment) calculation used in circular mode.

The circular mode is the default because, as noted in the main text, its geometry is independent of the focal cell's position within the aggregation grid and avoids the corner emphasis of square annuli, while the square mode is retained mainly for comparison and for compatibility with grid-aligned conventions. Both modes preserve the non-redundant multi-resolution partition of the neighbourhood.