Habitat node clustering and link identification methods write up:

To process and visualize LANDIS-II outputs we used both ArcGIS 10.6 and the raster library in the statistical software package R (v3.4.3) (ESRI 2018, Hijmans and van Etten 2012, R Development Core Team 2017). To analyze connectivity, we used the gDistance, spatstat, igraph, and Conefor command line (v2.6) R packages (Van Etten 2014, Baddeley et al., Csardi and Nepusz 2006, Saura and Torne 2012).

We used structural and functional graph theoretic metrics to quantify potential connectivity across the landscape over time. Graph theory network analysis efficiently assesses landscape connectivity across large study areas, incorporating both landscape structure and organism movement information to provide indices of connectivity well-suited for assessing landscape-level trends over time (Galpern et al. 2011, Tulbure et al. 2014, Saura et al. 2011, Minor and Urban 2008). In this approach, the landscape is represented as a set of discrete habitat patches, or nodes, connected by links that represent the ability of an organism to disperse between nodes (Urban and Keitt, 2001; Calabrese and Fagan 2004). Multiple links coalesce to form dispersal paths (McIntyre et al. 2018).

While links can be defined by the Euclidean distance between habitat nodes, where nodes are considered connected if they fall within an organism’s plausible dispersal distance, we defined links using least-cost paths as a measure of effective distance between nodes (Bishop-Taylor et al. 2015). Studies of potential and functional connectivity have commonly used least-cost paths to represent the influence of a heterogeneous landscape matrix on species dispersal ability and to more realistically reflect landscape permeability than Euclidean distance measures (Dilts et al. 2016). Least-cost paths use cost-distance surfaces to account for landscape resistance to movement (Bunn et al. 2000). These cost-distance surfaces reflect the difficulty for an organism to move through each cell on the landscape and the mortality risk associated with land cover type in relative terms (Adriaensen et al. 2003, Theobald et al. 2006, Zeller et al. 2012). We defined resistance values based on the assumption that forest species, especially specialist species, will face greater dispersal difficulty as they move through land cover types with characteristics increasingly disparate from those of the forested areas where they reside, similar to Saura et al. (2011). Our definition accounted for potential barriers to dispersal including major roads, urban infrastructure, and water bodies. A rating scale of 1-100 was used for resistance values, consistent with the magnitude of possible resistances found in previous studies (e.g. Stevenson-Holt et al. 2014, Shirk et al. 2015, Blazquez-Cabrera et al. 2016, Greenwald et al. 2009). As gDistance calculates least-cost paths using conductance instead of resistance surfaces, every cell on the landscape was assigned a conductance value (1/resistance) based on whether the species was a habitat specialist or generalist. Least-cost distances were calculated between the centroids of habitat patches similar to Theobald et al. (2012) and Dickson et al. (2017) and reflect the minimum cost accumulated along the shortest path between two habitat nodes (van Etten 2017, Bishop-Taylor et al. 2015). Using habitat patch centroids reduced computation time and allowed us to simulate species movement between and within habitat patches as a continuous process rather than assuming within-patch homogenization or an abrupt end to species movement at a patch edge (Dickson et al. 2017). For habitat specialists, a cell on the landscape was considered habitat if at least 25% of total biomass was longleaf pine and the median stand age was at least 21. Habitat cells for generalist species were defined as any cell with at least 25% of biomass from longleaf pine OR at least 65% of biomass from any mix of pine species AND with the same median stand age requirements. Cells with 90% or more of their total biomass from loblolly pine were considered to be loblolly plantations and not considered habitat.

In our study area, especially for habitat specialist species, habitat is currently clumped in the southeastern portion of the landscape, on and around Ft. Bragg. Such clumping can result in the ‘mega-patch’ problem as documented in Cavanaugh et al. (2014). Mega-patches mask the relevant ecological processes and connectivity dynamics taking place on the landscape. For example, a simple contiguity approach to group habitat pixels into habitat patches at every time step creates several mega-patches of tens of thousands of hectares and inflates the connectivity of the starting landscape. Because habitat for specialist species was originally confined to a small subset of the landscape, grouping contiguous habitat cells into patches creates a network graph in the earliest time step(s) that appears highly connected. With no habitat currently existing outside of this small area, effective distances between habitat patches are very low and average patch size is fairly large, primarily due to habitat on Ft. Bragg. This approach results in an apparent decline in connectivity over time even after land is restored and habitat is added, because the effective distance between habitat patches increases as the habitat network expands geographically across the landscape and the number of habitat patches remains fairly static as individual patches began to coalesce as habitat is added. To resolve mega-patches, we created a comprehensive raster stack of habitat pixels for each conservation strategy, from every time step and every model replicate. To increase graph modularity based on habitat asynchrony, we then grouped habitat pixels using a quintiles of occurrence approach. For each individual time step and replicate, we grouped contiguous habitat pixels that co-occurred in at least 80% of rasters in the comprehensive raster stack, in 60-79% of rasters, between 40-59%, between 20-39%, and in at least one raster using an eight neighbor approach for each species type to accommodate changes to connectivity across time and space. Doing so allowed us to track habitat network changes throughout model runs and to treat habitat added over time as discrete habitat patches, which more realistically reflects how an animal would perceive spatiotemporal habitat fluctuations (Wimberly 2006). Species colonization of new, or newly suitable, habitat patches does not happen instantaneously and modularizing our habitat network enabled a better representation of intra- and inter-patch movement dynamics (Hanski 1999). Habitat nodes were weighted by area. Links were restricted to pairs of nodes with a pair-wise Euclidean distance that fell below each species group’s maximum dispersal distance. The probability of direct movement between patches was obtained using a negative exponential function of inter-patch least cost path value.

**Contiguity habitat clustering approach**

-Combine LANDIS-II biomass and median stand age output maps with current land use map to create one map of landscape composition for each time step/replicate of each conservation strategy.

-Identify habitat cells at each time step that are longleaf pine community type and at least 21 years old median stand age. Cluster any contiguous habitat cells for that time step using an 8-neighbor approach. These become habitat nodes.

-Nodes fluctuate over time as new habitat is added through land acquisition and restoration and habitat patches coalesce.

-Nodes vary in location, size, and total number for each time step.

-Nodes are weighted by habitat area.

-Problems that arise: patches coalesce as new habitat is added, meaning the number of habitat patches does not change significantly over time. In early time steps, landscape connectivity *decreases* as patches are added to other parts of the landscape while habitat patches near Ft. Bragg being to coalesce 🡪 no marked increase in the number of patches while increasing average least cost path distance as landscape graph “expands”

*Habitat at time t, 5 nodes*



*Habitat at time t +1, 5 nodes*

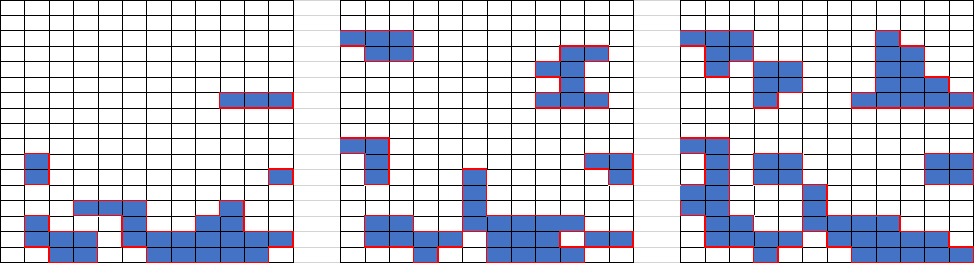


*Habitat at time t + 2, 4 nodes*



Overall view of habitat nodes over time:

Time t, 5 nodes Time t +1, 5 nodes Time t + 2, 4 nodes

****

**Modularity habitat clustering approach**

-Combine LANDIS-II biomass and median stand age output maps with current land use map to create a map of landscape composition for each time step/replicate of each conservation strategy.

-Identify habitat cells at each time step that are longleaf pine community type and at least 21 years old median stand age.

-Create a raster stack of all habitat created across time steps and reps for a single conservation strategy.

-For each individual time step in every replicate, cluster habitat cells using occurrence quintiles. Habitat cells that occur in at least 80% of time steps and reps are clustered together if contiguous (8-neighbor). Habitat cells that occur in 60-79% of time steps/reps are clustered, as are those that occur in 40-59%, 20-39%, and those in at least one habitat raster if contiguous.

-Nodes fluctuate and generally increase in number over time as new habitat is added through land acquisition and restoration and habitat patches coalesce.

-Nodes vary in location, size, and total number for each time step.

-Nodes are weighted by habitat area.

-Potential problem: is this a statistically valid way of clustering?

Habitat cells at time t Habitat cells at time t +1 Habitat cells at time t + 2



Raster stack of habitat cells from time steps t, t + 1, and t + 2, are summed to create one master habitat raster (color equates to occurrence: darkest blue = occurs in all three time step habitat raster, medium blue = occurs in two habitat rasters, lightest blue = occurs in only one)



This is used to cluster habitat cells in the individual habitat rasters.

*Habitat at time t, 13 nodes*



*Habitat at time t + 1, 18 nodes*



*Habitat at time t + 2, 21 nodes*



Overall view of habitat nodes over time:

Time t Time t + 1 Time t + 2



**Other possible solutions**:

1. At time t, cluster all contiguous habitat into nodes. At time t +1, cluster all habitat cells identical to time t together, then cluster all habitat cells unique to time t+1 together. For time t+2, cluster all habitat cells from time t, then cluster all cells different from time t but in time t+1, then cluster all remaining cells in time t+2, and so on.
2. Bishop-Taylor et al. 2017 approach (not sure how to practically do this in R, specifically using pearson correlation coefficient to come up with correlation between individual cells over time for binary habitat/nonhabitat):
   1. To resolve this “mega-patch problem,” we used a simplified version of a method based on habitat asynchrony and graph theory community detection previously used to identify temporally consistent subregions in remotely sensed giant kelp forest time series (Cavanaugh et al. 2014). We initially generated an m × n matrix of n observations (surface water or non-water for each of the 99 seasons) for each raster pixel m in the study area (Fig. 2A). We converted this matrix into a network graph by connecting surface-water pixels (nodes) to their immediate eight spatial neighbors, and weighted the resulting graph edges by the pairwise Pearson correlation in temporal dynamics between each pixel (i.e., high correlation if the two pixels consistently occurred together as either surface water or non-water throughout the time series; Fig. 2B).
   2. We used the Igraph Python package (Csardi and Nepusz 2006) “community\_multilevel” implementation of the Louvain community detection algorithm to divide “mega-patches” into discrete asynchronous potential habitats (Cavanaugh et al. 2014). The Louvain algorithm efficiently partitions large connected graphs into unique “communities” by maximizing the graph’s modularity, a measure of the density of within-community links compared to between-community links (Newman 2006, Blondel et al. 2008). As the Louvain algorithm does not accept negative weights, we truncated graph edge weights (correlations ranging from −1 to 1) to between 0 and 1 prior to analysis. The community membership identifiers produced by the algorithm were finally converted back into a raster containing 277874 temporally consistent unique IDs informed by both the spatial structure and temporal synchrony of habitats throughout the entire 1987–2011 time series (Fig. 2C). These unique IDs served as individual graph nodes in subsequent network analyses.