

## Metadata

**Note:** All of the files in Data S1 must be within your working directory for the analysis to work. Further, within your working directory there should be a number of subfolders. These include:

- plots
- model\_output
- jags\_models
- data

**Data S1 consists of 4 scripts used for analysis, model summary, and plotting. They are located within the working directory. These scripts include:**

**fit\_softmax\_model.R:** This script 1) reads in the raw detection non-detection data for coyote, opossum, and raccoon and converts it to community states at each site that can be supplied to the dynamic co-occurrence model, 2) reads in the covariate data and generates the urbanization covariate URB, 3) fits the 6 different models outlined in the manuscript, and 4) compares the fit of each of these models.

**fit\_models\_utility\_functions.R:** This script contains a number of useful functions that are sourced for fit\_softmax\_model.R. The functions are commented out within this script and describe what they do. It include three functions: two to generate initial values for JAGS models (those with and without species interactions) and one to generate the vectors to organize the species interactions within a matrix in JAGS.

**calculate\_cpo.R:** This function calculates the conditional predictive ordinate for each data point within a model when given the model output as a matrix and the data list supplied to JAGS. It returns the summary statistic:  $-\sum_{k,t} \log(CPO_{kt})$  for site  $k$  and time  $t$ , where lower values indicate better model fit. It is sourced within fit\_softmax\_model.R

**plotting\_script.R:** This is the code used to generate the figures in the manuscript. Some additional calls to imageMagick were done in the console to generate the final figures.

**plotting\_script\_utility\_functions.R:** This script included functions to help generate the figures in plotting\_script.R. It includes 4 functions: one to generate a species colonization and extinction rate given the presence or absence of another species, two to label the x and y axes for the figures, and one to generate each sub-figure for figure 3 (the occupancy plot).

**calculate\_steady\_state.R:** This function calculates the expected occupancy rate along an environmental gradient for the urbanization covariate with the best fit model and the data list used to fit this model. It is sourced and called within plotting\_script.R to generate figure 3.

**Data S1 also has 2 JAGS dynamic co-occurrence occupancy models that are used in this analysis. They should be placed within the jags\_models sub-folder of the working directory. These include:**

**dcom\_inxs.R:** This is the dynamic co-occurrence occupancy model that is used for models 1 through 4 and includes interactions between species. It is general enough such that any number of covariates could be fit to each process (e.g., include 2 covariates on all species interactions for colonization). The parameters within each linear predictor are named as they are within the manuscript.

**dcom\_no\_inxs.R:** This is the dynamic co-occurrence occupancy model that is used for models 5 and 6 and does not include interactions between species. It is general enough such that any number of covariates could be fit to each process. The parameters within each linear predictor are named as they are within the manuscript.

**There are two data files within the data sub-folder which are used in this analysis. They include:**

**fidino\_sp\_data.csv:** This is the detection / non-detection data for coyote, opossum, and raccoon in long format. It is 44772 records and 9 columns. The ordering of this data frame is an important part of how the data gets sorted and summarized within **fit\_softmax\_model.R**. In R, these data should be read such that character strings are not automatically turned into factors:

```
data <- read.csv("fidino_sp_data.csv", header = TRUE, stringsAsFactors = FALSE)
```

The structure of fidino_sp_data.csv		
Column header	Data type	Description
Season	Categorical	This contains information on which season and year the detection / non-detection data is associated it. Each element has four characters. The first two characters describe which season (SP = spring, SU = summer, FA = fall, WI = winter) while the final two characters describe the year (10 = 2010, 11 = 2011, etc.). The data is currently sorted so that the season and year continue temporally (i.e., SP10 data, SU10 data, FA10 data, WI11 data, etc.)
Week	Categorical	Which week of sampling do the 7 detection / non-detection days fall within per season. Represented as 'week 1', 'week 2', 'week 3' and 'week 4'.
Date	Date	The day associated to the detection / non-detection data. Format is '%Y-%M-%D'
SeasonWeek	Categorical	The data from the Season column and Week column have been concatenated and separated by a space. For example, the first record is 'SP10 week 1'.
StationID	Categorical	Which site the detection / non-detection data is associated to. Within each season the StationID is sorted alphabetically (and therefore in the same order across seasons).
SurveyID	Categorical	StationID concatenated Season and separated by a '-'. For example, the first record is 'D02-BMT0-SP10'. Because Season = SP10 and SurveyID = D02-BMT0.

Coyote	Binary	If a coyote was detected at a site on a given day the associated element in this column is a 1. If coyote were not detected it is a 0. If sampling did not occur it is NA.
Opossum	Binary	If an opossum was detected at a site on a given day the associated element in this column is a 1. If opossum were not detected it is a 0. If sampling did not occur it is NA.
Raccoon	Binary	If a raccoon was detected at a site on a given day the associated element in this column is a 1. If raccoon were not detected it is a 0. If sampling did not occur it is NA.

The ordering of the rows within **fidino\_sp\_data.csv** is sorted hierarchically:

1. Season (progresses temporally)
2. StationID (alphabetical within each season).

**fidino\_covariate\_data.csv:** This is the three covariates used to create the urbanization PCA for each site. These covariates include: housing density, proportion tree cover, and proportion impervious cover all within 1000 meters of a site. It is alphabetically ordered by StationID and therefore has the same order for each season within **fidino\_sp\_data.csv** (after the sites that have fewer than 2 seasons of data are removed). This dataset has 103 rows, one for each site that was included in the analysis.

The structure of fidino_covariate_data.csv		
Column header	Date type	Description
StationID	Categorical	Which site the collected covariate data is associated to. The whole data frame is sorted alphabetically based off of this column.
House	Numeric	Housing density (number of units within a 1000 meter buffer of a site)
Tree	Numeric	Proportion of tree cover (the percent of tree cover within a 1000 m buffer of a site). Can only range from 0 to 1.
Imp	Numeric	Proportion impervious cover (the percent of impervious cover within a 1000 m buffer of a site). Can only range from 0 to 1.

Housing density was collected from the Silvis lab website and is from the 2010 census.

Website: [http://silvis.forest.wisc.edu/data/housing/block\\_change](http://silvis.forest.wisc.edu/data/housing/block_change)

Tree cover and Impervious cover were calculated from the 2011 NLCD data.

Website: <https://www.mrlc.gov/nlcd2011.php>