Romania SEM

Marissa Dyck

2024-08-08

Contents

Before you begin

Notes

A few notes about this script.

If you are want to run through the full analysis with the published data make sure you download the whole (Romania_SEM repository)[https://github.com/marissadyck/Romania_SEM] from the author's (Marissa A. Dyck) GitHub. This will ensure you have all the files, data, and proper folder structure you will need to run this code and associated analyses.

Also make sure you open RStudio through the R project (Romania_SEM.Rproj) this will automatically set your working directory to the correct place (wherever you saved the repository) and ensure you don't have to change the file paths for the data.

If you have question please email the author,

Marissa A. Dyck
Postdoctoral research fellow
University of Victoria
School of Environmental Studies
Email: marissadyck17@gmail.com

R and RStudio

Before starting you should ensure you have the latest version of R and RStudio downloaded. This code was generated under R version 4.2.3 and with RStudio version 2024.04.2+764.

You can download R and RStudio HERE

R markdown

This script is written in R markdown and thus uses a mix of coding markup languages and R. If you are planning to run this script with new data or make any modifications you will want to be familiar with some basics of R markdown.

Below is an R markdown cheatsheet to help you get started,
R markdown cheatsheet

Install packages

If you don't already have the following packages installed, use the code below to install them. *Note this code chunk will NOT automatically run when the file is knit since eval=FALSE is set, because I already have the packages and don't want to reinsall them every time. You will need to run this chunk separately if you need to install these packages.* You only need to do this once.

```r
install.packages('tidyverse')
install.packages('piecewiseSEM')
install.packages("PerformanceAnalytics")
install.packages("purrrlyr")
```

Load libraries

Then load the packages to your library for this current session.

```r
library(piecewiseSEM) # used to run piecewise structural equation models
library(tidyverse)  # data tidying, visualization, and much more; this will load all tidyverse packages
library(PerformanceAnalytics) # generates correlation chart
library(purrrlyr) # data manipulation
```

Data

Import

Let's read in the data that we cleaned and formatted with the earlier code '1_RO_SEM_data_formatting'

In this code chunk I will also add columns with the scaled the predictor variables fo analysis

```r
ro_sem_dat <- read_csv('data/processed/ro_sem_data_effort_2018-2019.csv') %>%

  # scale predictor variables
  mutate(z.s = scale(z),
         denslocalr.s = scale(denslocalr),
         tri5.s = scale(tri5),
         clc_forest.s = scale(clc_forest))
```

```
## Rows: 138 Columns: 30
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr  (1): trapcode
```

```
## dbl (29): session, trap_effort, hare, fox, wolf, badger, lynx, human, red_de...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Summary statistics

We will want to report some summary statistics for our data, the code chunks below will extract/calculate those details

Total occurrences

Let's calculate the total occurrences for 8 focal species by session/season

```r
ro_sem_dat %>%

  # group by session
  group_by(session) %>%

  # calculate occurrences for species of interest (8)
  summarise(total_8 = sum(hare,
                          fox,
                          lynx,
                          wolf,
                          red_deer,
                          roe_deer,
                          wild_boar,
                          wild_cat))
```

```
## # A tibble: 2 x 2
##   session total_8
##     <dbl>   <dbl>
## 1       2    1626
## 2       3    1973
```

Species occurrences

We will also want some summary information for each species

```r
# total species occurrence summaries

ro_sem_dat %>%

  # select species of interest
  select(badger:otter) %>%

  # use purrr map to calculate summary stats for all columns listed
  dmap(~c(sum(.x),
          min(.x),
          max(.x),
          (sum(.x)/3599*100))) %>%

  # create columns for new data to report
  mutate(stat = c('sum',
                  'min',
                  'max',
```

```
                    'percent')) %>%
  relocate(stat)
```

```
## # A tibble: 4 x 17
##   stat    badger   lynx  human red_deer  bear roe_deer wild_boar mustelid chamois
##   <chr>    <dbl>  <dbl>  <dbl>    <dbl> <dbl>    <dbl>     <dbl>    <dbl>   <dbl>
## 1 sum       208    332  18         772   604      386       694       69      11
## 2 min         0      0   0           0     0        0         0        0       0
## 3 max        83     20   4          81    49       24        79        9       6
## 4 perce~   5.78   9.22  0.500      21.5  16.8     10.7      19.3     1.92   0.306
## # i 7 more variables: bird <dbl>, wild_cat <dbl>, dog <dbl>, squirrel <dbl>,
## #   livestock <dbl>, hedgehog <dbl>, otter <dbl>
```

Environmental variables

And we will want to report some summary info on the environmental variables we are using in the analysis

```
ro_sem_dat %>%

  # select columns of raw data (not scaled)
  select(z, denslocalr, tri5, clc_forest) %>%

  # print summary info for each
  summary()
```

```
##        z           denslocalr          tri5          clc_forest
##  Min.   : 663   Min.   :0.2127   Min.   : 66.38   Min.   :0.08505
##  1st Qu.:1025   1st Qu.:0.2413   1st Qu.:178.43   1st Qu.:0.64883
##  Median :1168   Median :0.2705   Median :221.61   Median :0.83402
##  Mean   :1166   Mean   :0.2702   Mean   :220.27   Mean   :0.76541
##  3rd Qu.:1293   3rd Qu.:0.2939   3rd Qu.:260.65   3rd Qu.:0.91221
##  Max.   :1617   Max.   :0.3434   Max.   :494.01   Max.   :1.00000
```

Correlations

Before we run any models we need to test an assumption of Independence and see if any of our potential variables are correlated.
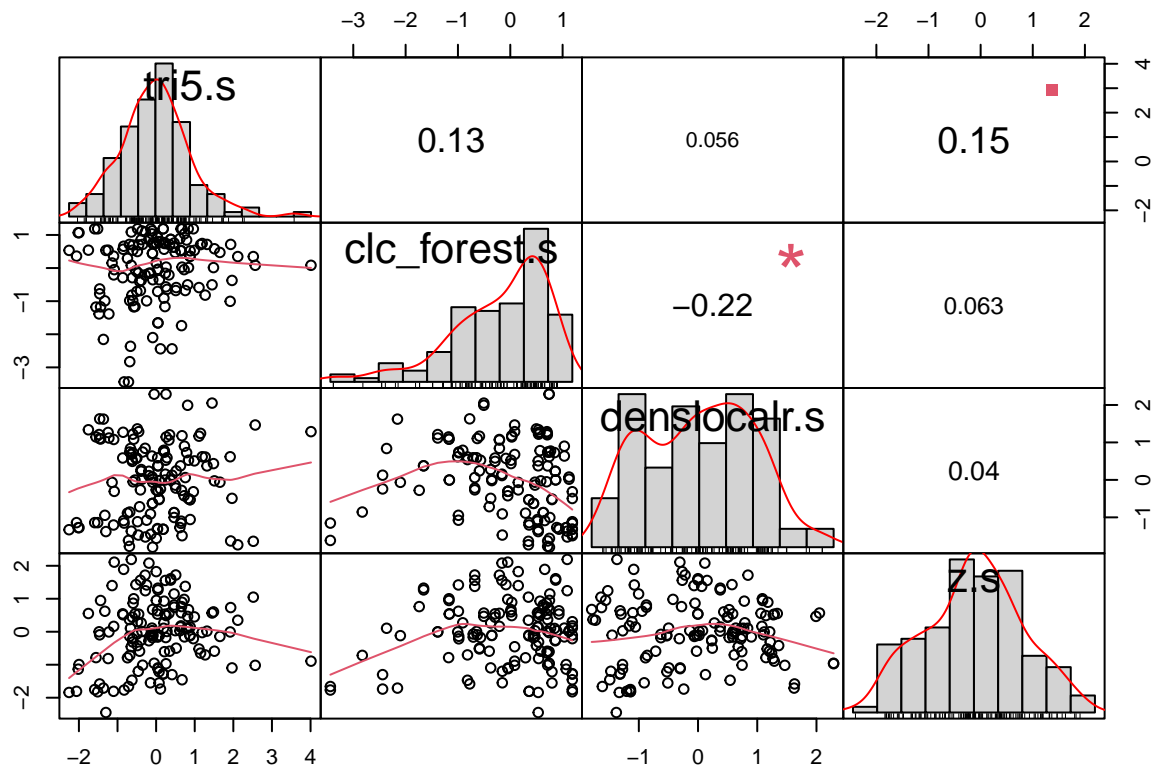
```
# subset data to just environmental variables for correlation tests

ro_sem_env <- ro_sem_dat %>%
  select(tri5.s,
         clc_forest.s,
         denslocalr.s,
         z.s)

# get pairwise correlation matrix

chart.Correlation(ro_sem_env,
                  histogram = TRUE,
                  pch = 19)
```

all correlations for chosen environmental variables >0.25

Camera activity check

Since we aren't accounting for the camera activity/effort because we have a poisson response variable that we want to keep as poisson; I want to check that we aren't introducing bias by using the raw occurrence data. I'm going to check this by seeing if there are correlations in camera activity and number of detections per species first, and if there are see if the cameras that have higher detections are biased in any way (correlated) with the environmental variables in our model

```
#
ro_sem_dat %>%

  select(trap_effort,
         wolf,
         lynx,
         wild_cat,
         fox,
         red_deer,
         roe_deer,
         wild_boar,
         hare) %>%

  chart.Correlation(.,
                    histogram = TRUE,
                    pch = 19)
```
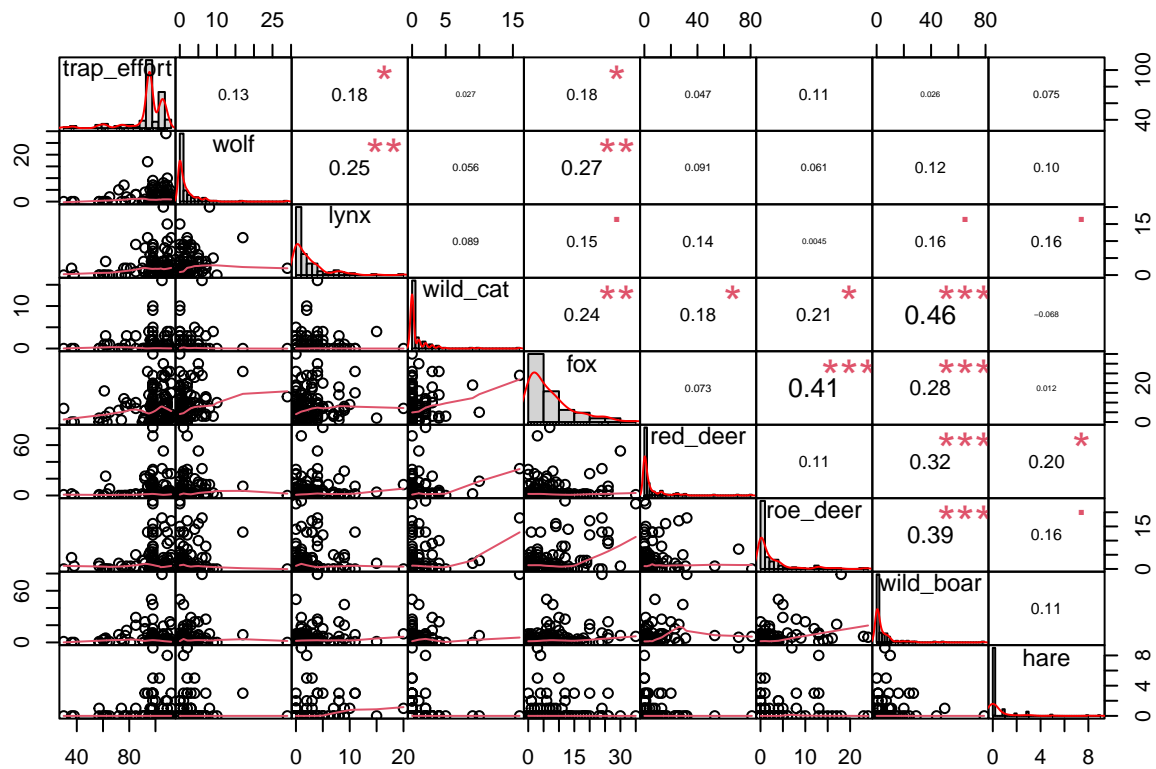
Nothing is over 0.5 so doesn't seem to be a huge concern that cameras that were active longer have more detections although some positive relationships as you'd expect

Now we will check if cameras that were active longer are biased towards any environmental variables

```
ro_sem_dat %>%

  select(trap_effort,
         z,
         distlocalr,
         tri5,
         clc_forest) %>%

  chart.Correlation(.,
                    histogram = TRUE,
                    pch = 19)
```
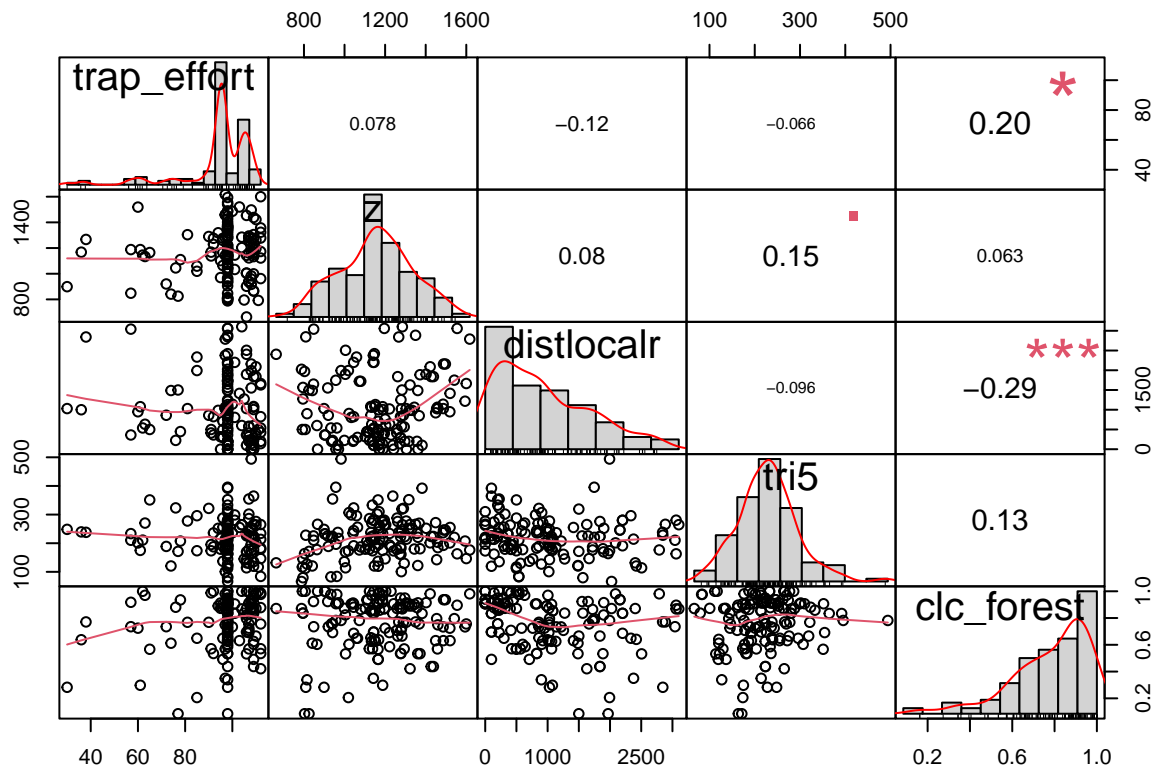
Also no biases here for camera activity, which is good.

Analysis

The code chunks below define the models for our three hypotheses including the hypothesized model and final optimized models

- top-down

- bottom-up

- combined

  The formal analysis was trasnferred to a basic r script as a bug with the latest version of *piceweiseSEM* package and the summary doesn't print tests of directed separation. This may have been corrected after publication of this analysis

Top-down

Hypothesis

This first chunk is our initial model as we defined it a priori based on previous research and the literature review

Optimized   Now, based on the tests of directed separation, we will add significant pathways to model that make biological sense and fit hypotheses (top-down) and remove non-significant pathways and pathways with coefficient estimates that do not fit the hypothesis (e.g. positive estimate for a hypothesized negative causal pathway)

Bottom-up

This section will define our bottom-up model

Hypothesis   Again, this first chunk is our initial model as we defined it a priori based on previous research and the literature review

Optimized   And again, based on the tests of directed separation, we will add significant pathways to model that make biological sense and fit hypotheses (top-down) and remove non-significant pathways and pathways with coefficient estimates that do not fit the hypothesis (e.g. positive estimate for a hypothesized negative causal pathway)

Combined

Lastly we will define a combined model that includes both bottom-up and top-down processes/interactions. This model will be based on the optimized models for both top-down and bottom-up

Based on the Chi-squared and fisher's C the combined model fits the data best