Lavaan Option for Adjusting for Spatial Autocorrelation

Jim Grace

U.S. Department of the Interior
U.S. Geological Survey

No explicit citation for the methodological details presented here is currently available. References that illustrate the procedures are available and include:

Harrison, S. and Grace, JB. 2007. Biogeographic affinity contributes to our understanding of productivity-richness relationships at regional and local scales. American Naturalist. 170:S5-S15.

Matteson, K.C., Grace, J.B., and Minor, E.S. 2012. Direct and indirect effects of land use on floral resources and flower-visiting insects across an urban landscape. *Oikos* 122:682-694.

Notes: IP-056512; Support provided by USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge the major contribution by Jarrett Byrnes, Univ. Mass. – Boston for the `lavSpatialCorrect` function used in this module. Appreciation also to Darren Johnson for technical advice. Formal review of the material from which this tutorial was derived was provided by Jesse Miller and Phil Hahn, Univ. Wisconsin. Any use of trade, form, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Questions about this material can be sent to sem@usgs.gov. Last revised 15.06.17.

What do we mean by the phrase "Spatial Autocorrelation"?

- When spatial effects are not fully incorporated in models, we may have spatially autocorrelated residuals.

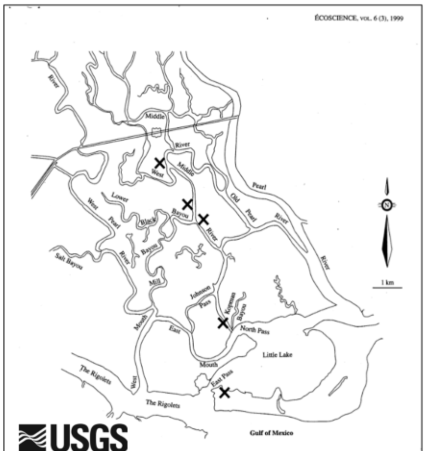


FIGURE 1. The data presented in this study were from the lower Pearl River area. The X symbol marks each of the five study areas. Here the zero point for distance from the sea is at the junction of the Rigolets waterway with the Gulf of Mexico.

Issues:

- Sample sites close to each other in space may be correlated because of spatial processes not included in our models.
- Standard errors, and associated probabilities, depend on an assumption of independence among residuals.
- This module illustrates how to test and correct for spatial autocorrelation when using `lavaan`.

2

There are a number of good references on the topic of spatial autocorrelation, e.g.,

Dale, MRT and Fortin, MJ (2014) Spatial Analysis: A Guide for Ecologists. Cambridge University Press

Bivand, RS, Pebesma, E, and Gomez-Rubio, V (2013) Applied Spatial Data Analyses with R. Springer Verlag

Bordard, D, Gillet, F, Legendre, P (2011) Numerical Ecology with R. Springer Verlag

What is the consequence of spatially-correlated residuals?

- According to Naroll (1961), Sir Francis Galton, one of the earliest pioneers in statistics, raised the question of spatial autocorrelation in 1889. Paraphrasing,

  "Positive spatial dependence can reduce the amount of information in the observations because proximate observations can partly predict each other."

- The primary consequence of such reduced information can be expressed as a reduced "effective" sample size.

- One solution is to estimate the magnitude of residual autocorrelation, compute the effective sample size, then replace the value of n (sample size) used for calculating standard errors and p-values with the "effective n".

USGS 3

A historic reference on this topic is.

Raoul Naroll (1961). "Two solutions to Galton's Problem". Philosophy of Science 28: 15–29. doi:10.1086/287778.

I. The Basic Procedures:

1. Obtain model residuals.
2. Test for autocorrelation using global Moran's I test.
3. Determine the effective sample size using Moran's I.
4. Adjust the standard errors for model parameters affected by the autocorrelation.

USGS 4

Moran's I index quantifies the degree of spatially-structured correlation in a dataset.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17-23.

Moran's *I* is defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

where N is the number of spatial units indexed by i and j; X is the variable of interest; $\bar{X}$ is the mean of X; and w_{ij} is an element of a matrix of spatial weights.

The expected value of Moran's *I* under the null hypothesis of no spatial autocorrelation is

$$E(I) = \frac{-1}{N-1}$$

http://en.wikipedia.org/wiki/Moran's_I.

Formula for corrected sample size (Neff) is

Neff = N*((1-I)/(1+I))

Procedures for Addressing Spatial Autocorrelation Using "`lavSpatialCorrect`" Function (for `lavaan` objects).

- On January 14, 2015, Jarrett Byrnes described on his blog a new R function he developed to correct for spatial autocorrelation in lavaan objects*.
- To access Jarrett's function and associated materials go to "https://github.com/jebyrnes/spatial_correction_lavaan". There you can find both the function itself "`lavSpatialCorrect.R`" and additional information related to his example.
- In this module, I first give a generic presentation of the method. Then I provide an example from one of our published studies.

* (see presentation at http://www.imachordata.com/space-and-sems-2/

USGS

5

First, tip of the hat to Jarrett Byrnes for developing another very helpful R function for use in SEM. You can find more of his materials at

http://jarrettbyrnes.info/sem.shtml

and

http://byrneslab.net/ (look especially at his tab "Teaching")

II. Generic Illustration:

```
### Generic Code for Implementing lavSpatialCorrect
# Create lavaan model (see Lavaan Basics Tutorial).
model <- 'y2 ~ x1 + y1
          y1 ~ x1 + x2'

# Load lavaan library.
library(lavaan)
library(ape)

# Fit the model using sem function.
model.fit <- sem(model, data, meanstructure=T)

# Adjust for correlations in the residuals.
lavSpatialCorrect(model.fit, xcoordinate, ycoordinate)
```

≋USGS

6

Note that this illustration assumes basic knowledge about working with R. For example, you will want to set your working directory, input your data, etc. Basic knowledge of modeling with lavaan is also assumed. A module "Introduction to lavaan" can be found at

http://www.nwrc.usgs.gov/test/SEM.html.

Citation for lavaan package:

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Citation for ape package:

Paradis E., Claude J. & Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

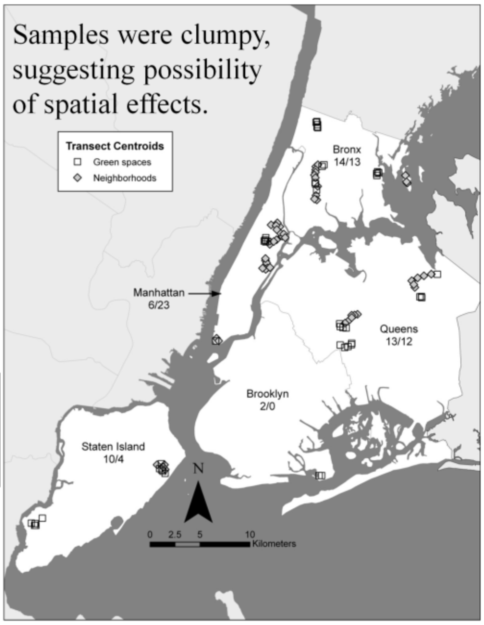
III. Example Application: Insect samples in a spatial landscape.

Matteson, Grace, and Minor (2012) Direct and indirect effects of land use on floral resources and flower-visiting insects across an urban landscape. *Oikos* 122:682-694.

http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0706.2012.20229.x/abstract

Note that "Green spaces" refers to city parks while "Neighborhoods" refers to yards etc.

Samples were clumpy, suggesting possibility of spatial effects.



USGS

Here is a published study in which procedures were used to estimate and correct for impacts of spatial autocorrelation. In this tutorial, I revisit those data using the `lavaan` package and the `lavSpatialCorrect` function.

Matteson et al. – Initial structural equation meta-model.

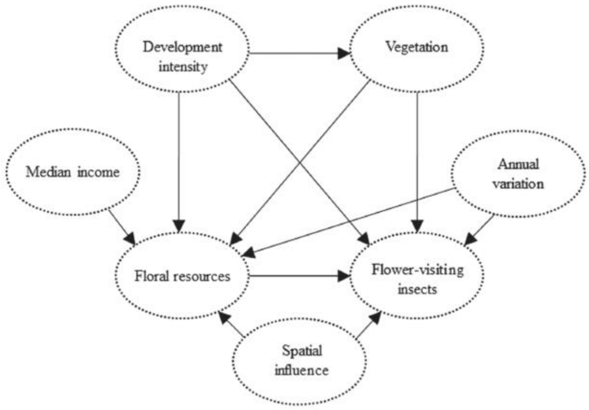
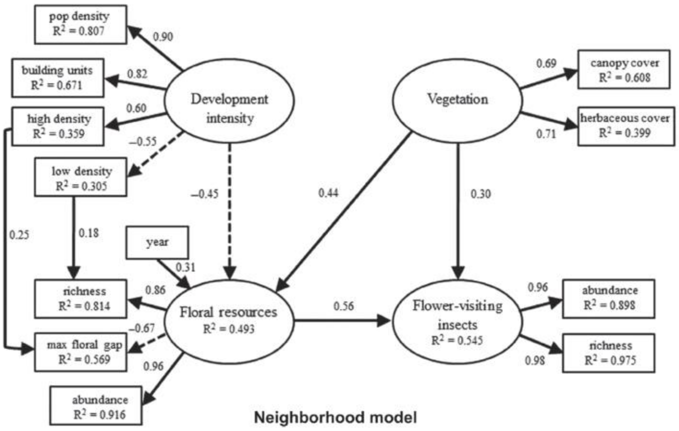


Figure 1. Original conceptual model for the relationship between floral resources and flower-visiting insects in urbanized landscapes. The link between floral resources and flower-visiting insects is well established. It is less certain how the other variables interact and influence flower-visiting insects directly or indirectly in urbanized landscapes.

8

Matteson et al. (2012) adopted an initial structural equation meta-model that defined their general network hypothesis. The focus in this study was to understand the urban conditions influencing pollinators (flower-visiting insects). The final models for the green space and neighborhood samples were subsets of this broader set of initial possibilities.

Published Structural Equation Model for Neighborhoods



Reviewers asked, “Are the results influenced by spatial autocorrelation?”

USGS

Here I show just the final Neighborhood model. In the review process, the authors were asked about the possible role of spatial autocorrelation. Since the model includes latent variables, we first estimated latent variable scores (using procedures described in the original publication).

Procedural steps for our example using `lavSpatialCorrect`:

1. Develop, fit, and test model.

2. Obtain latent variable scores and their residuals for analysis.

3. Adjust the standard errors for model parameters affected by the autocorrelation.

USGS

10

For our current illustration, we use `lavaan` and the latent variable scores to create a model so we can illustrate the "`lavSpatialCorrect`" function.

Additional details related to steps 1 and 2 can be found in Matteson et al. 2012.

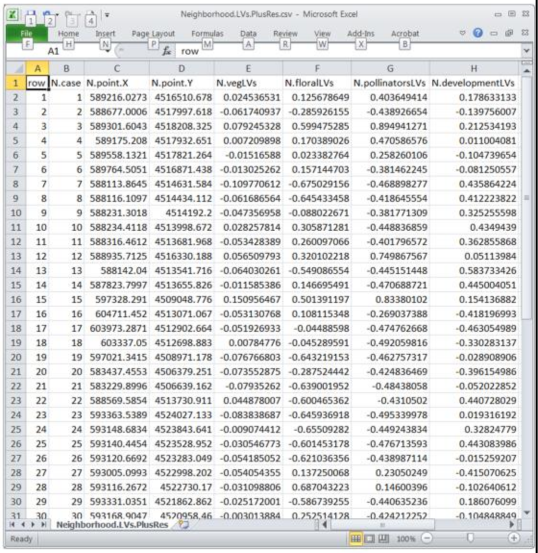
View of the data.

In this data (csv) file we see:

<u>Columns</u>
A & B – sample number (52 samples),

C & D – x, y coordinates (latitude and longitude),

E through H – Derived scores for the four latent variables in the model.



	A	B	C	D	E	F	G	H
1	row	N.case	N.point.X	N.point.Y	N.vegLVs	N.floralLVs	N.pollinatorsLVs	N.developmentLVs
2	1	1	589216.0273	4516510.678	0.024536531	0.125678649	0.403649414	0.178633133
3	2	2	588677.0006	4517997.618	-0.061740937	-0.285926155	-0.438926654	-0.139756007
4	3	3	589301.6043	4518208.325	0.079245328	0.599475285	0.894941271	0.212534193
5	4	4	589175.208	4517932.651	0.007209898	0.170389026	0.470586576	0.011004081
6	5	5	589558.1321	4517821.264	-0.01516588	0.023382764	0.258260106	-0.104739654
7	6	6	589764.5051	4516871.438	-0.013025262	0.157144703	-0.381462245	-0.081250557
8	7	7	588113.8645	4514631.584	-0.109770612	-0.675029156	-0.468898277	0.435864224
9	8	8	588116.1097	4514434.112	-0.061686564	-0.645433458	-0.418645554	0.412223822
10	9	9	588231.3018	4514192.2	-0.047356958	-0.088022671	-0.381771309	0.325255598
11	10	10	588234.4118	4513998.672	0.028257814	0.305871281	-0.448836859	0.4349439
12	11	11	588316.4612	4513681.968	-0.053428389	0.260097066	-0.401796572	0.362855868
13	12	12	588935.7125	4516330.188	0.056509793	0.320102218	0.749867567	0.05113984
14	13	13	588142.04	4513541.716	-0.064030261	-0.549086554	-0.445151448	0.583733426
15	14	14	587823.7997	4513655.826	-0.011585386	0.146695491	-0.470688721	0.445004051
16	15	15	597328.291	4509048.776	0.150956467	0.501391197	0.83380102	0.154136882
17	16	16	604711.452	4513071.067	-0.053130768	0.108115348	-0.269037388	-0.418196993
18	17	17	603973.2871	4512902.664	-0.051926933	-0.04488598	-0.474762668	-0.463054989
19	18	18	603337.05	4512698.883	0.00784776	-0.045289591	-0.492059816	-0.330283137
20	19	19	597021.3415	4508971.178	-0.076766803	-0.643219153	-0.462757317	-0.028908906
21	20	20	583437.4553	4506379.251	-0.073552875	-0.287524442	-0.424836469	-0.396154986
22	21	21	583229.8996	4506639.162	-0.07935262	-0.639001952	-0.48438058	-0.052022852
23	22	22	588569.5854	4513730.911	0.044878007	-0.600465362	-0.4310502	0.440728029
24	23	23	593363.5389	4524027.133	-0.083838687	-0.645936918	-0.495339978	0.019316192
25	24	24	593148.6834	4523843.641	-0.009074412	-0.65509282	-0.449243834	0.32824779
26	25	25	593140.4454	4523528.952	-0.030546773	-0.601453178	-0.476713593	0.443083986
27	26	26	593120.6692	4523283.049	-0.054185052	-0.621036356	-0.438987114	-0.015259207
28	27	27	593005.0993	4522998.202	-0.054054355	0.137250068	0.23050249	-0.415070625
29	28	28	593116.2672	4522730.17	-0.031098806	0.687043223	0.14600396	-0.102640612
30	29	29	593331.0351	4521862.862	-0.025172001	-0.586739255	-0.440635236	0.186076099
31	30	30	593168.9047	4520958.46	-0.003013884	0.252514128	-0.424212252	-0.104848849

≋USGS

11

Using the `lavSpatialCorrect` function requires that we have the data for building a lavaan model, along with the x-y spatial coordinates (e.g., latitudes and longitudes).

Variable code names include "N." because this is the neighborhood sample.

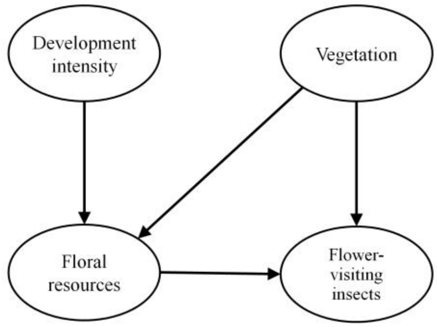
Development Intensity = N.developmentLVs

Vegetation = N.vegLVs

Floral resources = N.floralLVs

Flower-visiting insects = N.pollinatorsLVs

Simplified model we are using in place of the published model.



We use LV scores as the observed values in this simplification.

USGS

12

Note that we are actually using latent variable scores for the named variables, which is why ovals are shown instead of rectangles.

Lavaan code – part 1: Read data and create data objects.

```
### Read data
dat <- read.csv("Spatial Autocorrelation Illustration
Data.csv")
names(dat)
summary(dat)
dim(dat)

### Rename some variables
x       <- dat$N.point.X
y       <- dat$N.point.Y
dev     <- dat$N.developmentLVs
veg     <- dat$N.vegLVs
floral  <- dat$N.floralLVs
insects <- dat$N.pollinatorsLVs

sem.dat <- data.frame(x, y, dev, veg, floral, insects)
```

USGS

13

Here I present the code found in the file

“Spatial Autocorrelation Matteson lavaan Illustration.R”. Here I only show the preliminaries, The next slide shows the meat of the code.

Lavaan code – part 2: Specify lavaan model, fit, and correct.

```
### Load needed library
library(lavaan)

### lavaan modeling
# Specify model
neigh.mod <- 'floral ~ dev + veg
              insects ~ veg + floral'

# Fit model
neigh.mod.fit <- sem(neigh.mod, sem.dat, meanstructure=T)

# Examine model with uncorrected parameters
summary(neigh.mod.fit, rsq=T)

### Correct for spatial autocorrelation
source(“./lavSpatialCorrect.R") # access the function
library(ape) # load supporting library

# Execute correction function
lavSpatialCorrect(neigh.mod.fit, x, y)
```

14

USGS

Here is the code for specifying and fitting a model, then correcting for spatial autocorrelation in the residuals.

Corrected Output

No spatial autocorrelation for floral, but some for insects.

Sample size not changed for floral, but reduced for insects

```
$Morans_I$floral
    observed    expected         sd  p.value n.eff
1 0.05273638 -0.01960784 0.06067637 0.233145    52

$Morans_I$insects
   observed    expected          sd     p.value    n.eff
1 0.1645226 -0.01960784 0.06098543 0.002533982 37.30698
```

Key parameters (in red boxes) still clearly different from zero.

```
$parameters
$parameters$floral
                   Parameter     Estimate n.eff    Std.err   Z-value      P(>|z|)
floral~dev        floral~dev -0.450226622    52 0.12856366 -3.501974 4.618241e-04
floral~veg        floral~veg  3.959315392    52 0.72681502  5.447487 5.108644e-08
floral~~floral floral~~floral 0.099187174    52 0.01945221  5.099020 3.414174e-07
floral~1            floral~1  0.004880643    52 0.04402965  0.110849 9.117361e-01

$parameters$insects
                       Parameter    Estimate    n.eff    Std.err    Z-value      P(>|z|)
insects~veg          insects~veg  2.76068127 37.30698 1.13815889  2.4255675 0.0152844711
insects~floral    insects~floral  0.57748186 37.30698 0.15946032  3.6214768 0.0002929261
insects~~insects insects~~insects 0.11628244 37.30698 0.02692366  4.3189687 0.0000156760
insects~1              insects~1 -0.02236798 37.30698 0.05608586 -0.3988168 0.6900281735
```

USGS

15

The output generated by the `lavSpatialCorrect` function gives revised estimates for

n.eff – the effective sample size,

Std.err – the adjusted standard errors,

Z-value – the adjusted Z-values (maximum likelihood t-values),

and

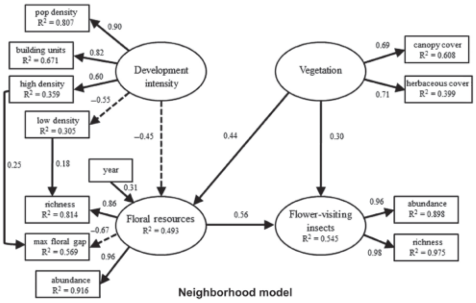
P-values – based on the new standard errors and n.eff.

Note that the other output associated with the model can be obtained using the

`summary(neigh.model.fit)`

command.

What can we conclude about Matteson et al.'s model?



Regarding methods - Modest spatial correlation amongst residuals was detected. Adjustments of effective sample sizes led to no changes to the model or conclusions.

Regarding results - In both neighborhoods and in green spaces, vegetation had a positive effect on flower-visiting insects (1) through increased floral resources and (2) independent of floral resources.

USGS

16

So, in this case residual spatial autocorrelation does not influence our major conclusions.

More information can be found at
http://www.nwrc.usgs.gov/SEM

USGS

I hope this overview has been useful. For more information, go to our webpage or search for examples involving your subject of interest. Questions and comments can be sent to sem@usgs.gov. Please note I cannot guarantee responses to individual inquiries, but will definitely incorporate suggestions in future tutorials. – Thanks!