PIECEWISE REGRESSION ANALYSIS ESTABLISHES A BREAKPOINT FOR THE BOTANIST EFFECT

BOTANIST EFFECT

Justin K. Williams

Department of Biological Sciences, Sam Houston State University, Huntsville Texas 77341-2116, U.S.A.

What is the Botanist Effect?

What is the Botanist Effect?

Journal of Biogeography (*J. Biogeogr.*) (2006) **33**, 1969–1974

ORIGINAL ARTICLE

The botanist effect: counties with maximal species richness tend to be home to universities and botanists

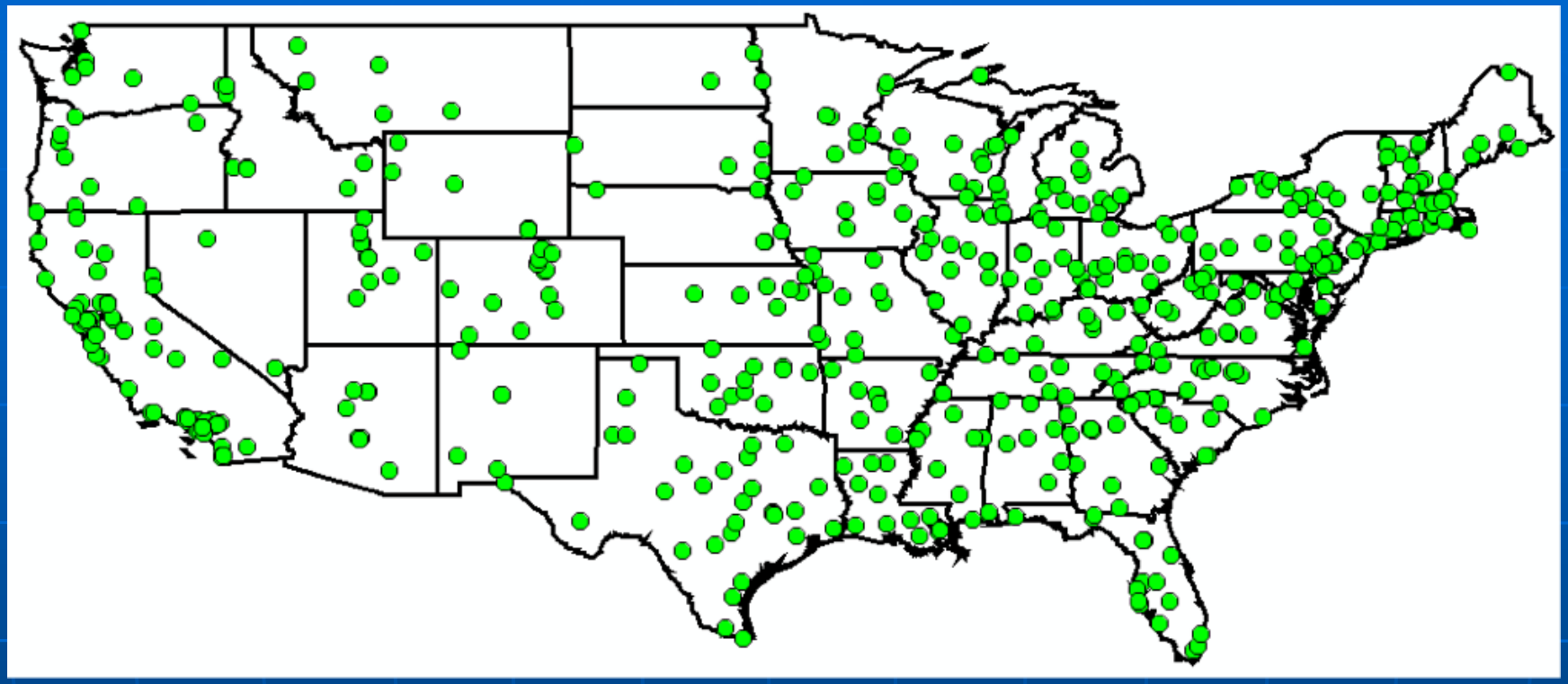
Daniel E. Moerman[1]* and George F. Estabrook[2]

[1]*Behavioral Sciences, University of Michigan-Dearborn, Dearborn, MI,* [2]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA*

ABSTRACT

Aim To investigate the distribution of local flowering plant species richness in areas surrounding American universities.

Methods Species richness in university counties was compared with neighbouring counties. Data were derived from *Synthesis of the North American Flora* (http://www.phylosystems.com/prepublication). Probabilities of the resultant distribution were calculated, and the results were also simulated.



Moerman and Estabrook (2006) demonstrated for 80 herbaria distributed throughout the US that plant species richness was highest in counties with herbaria and significantly lower in neighboring counties without herbaria. They coined this phenomenon "The Botanist Effect".

Moerman, D. E., and G. F. Estabrook. 2006. The botanist effect: counties with maximal species richness tend to be home to universities and botanists. *Journal of Biogeography* 33: 1969-74.

What is the Botanist Effect?

The Botanist Effect Revisited: Plant Species Richness, County Area, and Human Population Size in the United States

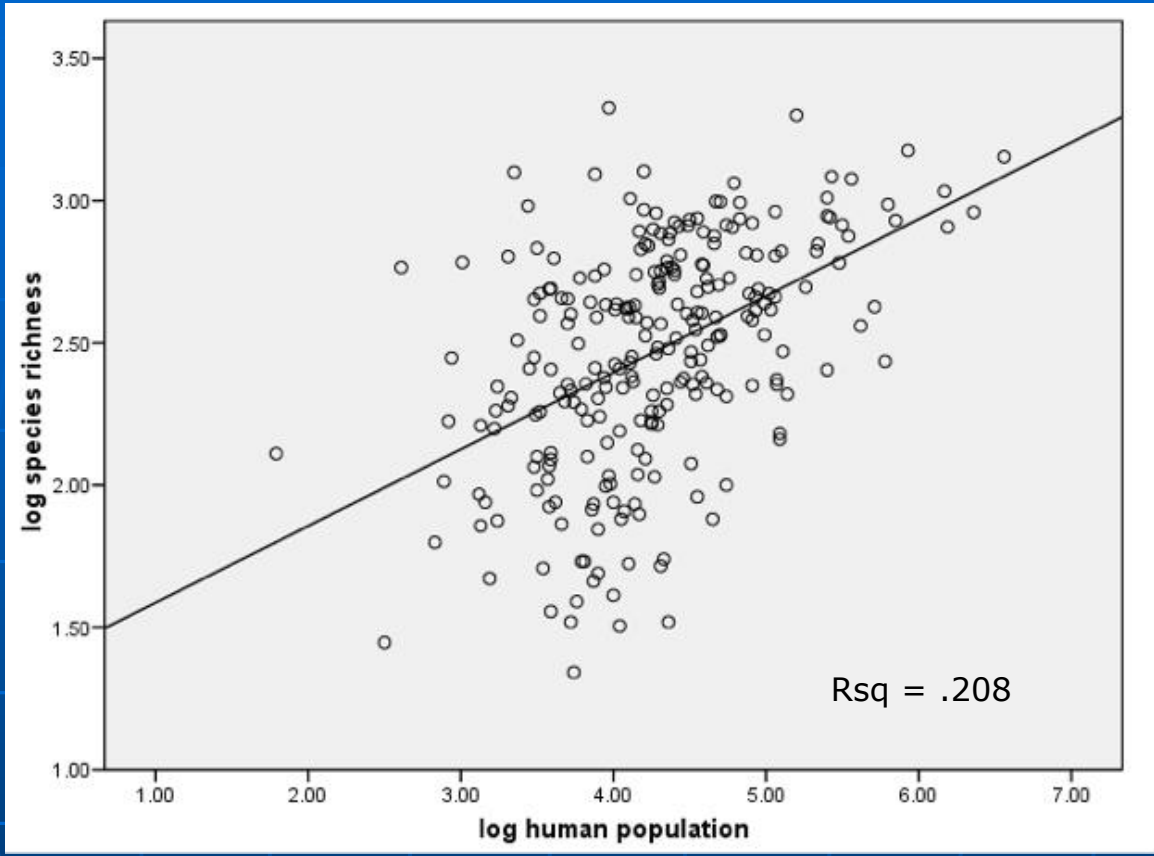
MARCO PAUTASSO* AND MICHAEL L. McKINNEY†

*Division of Biology, Imperial College London, Wye Campus, High Street, Wye, Kent, TN25 5AH, United Kingdom, email m.pautasso@ic.ac.uk

†Department of Earth and Planetary Sciences, 1412 Circle Drive, 306 Earth and Planetary Sciences Building, The University of Tennessee, Knoxville, TN 37996-1410, U.S.A.

Abstract: *The "botanist effect" is thought to be the reason for higher plant species richness in areas where botanists are disproportionately present as an artefactual consequence of a more thorough sampling. We examined whether this was the case for U.S. counties. We collated the number of species of vascular plants, human population size, and the area of U.S. counties. Controlling for spatial autocorrelation and county area, plant species richness increased with human population size and density in counties with and without universities and/or botanical gardens, with no significant differences in the relation between the two subsets. This is consistent with previous findings and further evidence of a broad-scale positive correlation between species richness and human population presence, which has important consequences for the experience of nature by inhabitants of densely populated regions. Combined with the many reports of a negative correlation between the two variables at a local scale, the positive relation between plant species richness in U.S. counties and human population presence stresses the need for the conservation of seminatural areas in urbanized ecosystems and for the containment of urban and suburban sprawl.*

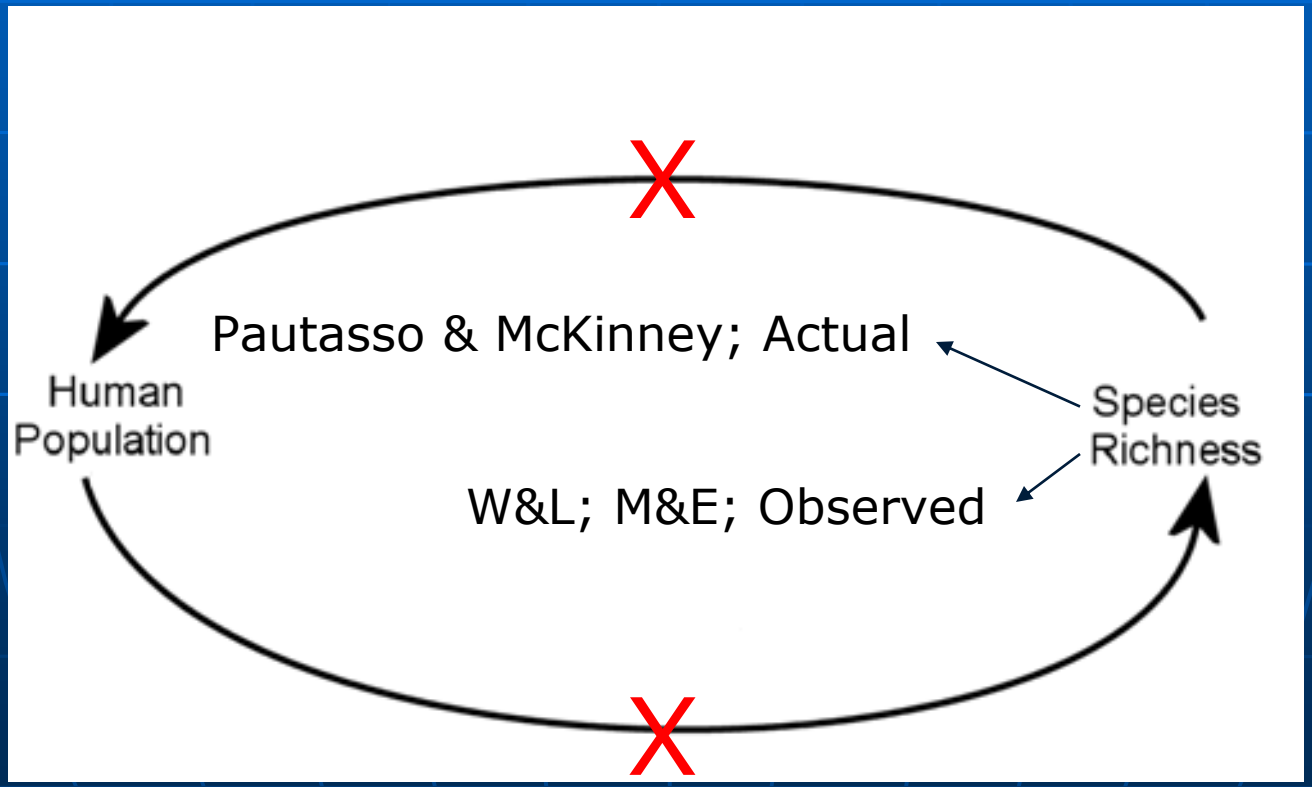
Conserv Biol. 2007 Oct;21(5):1333-40.



Pautasso and McKinney (2007) argued against "The Botanist Effect" demonstrating instead that higher SR is positively correlated with human population size. Their suggestion, humans are driven to settle in areas (counties) with higher species richness.

Pautasso, M., and McKinney M.L. 2007. The botanist effect revisited: Plant species richness, county area, and human population size in the United States. *Conservation Biology* 21, no. 5: 1333-40.

Chicken and Egg. Which came first SR values or Human Population

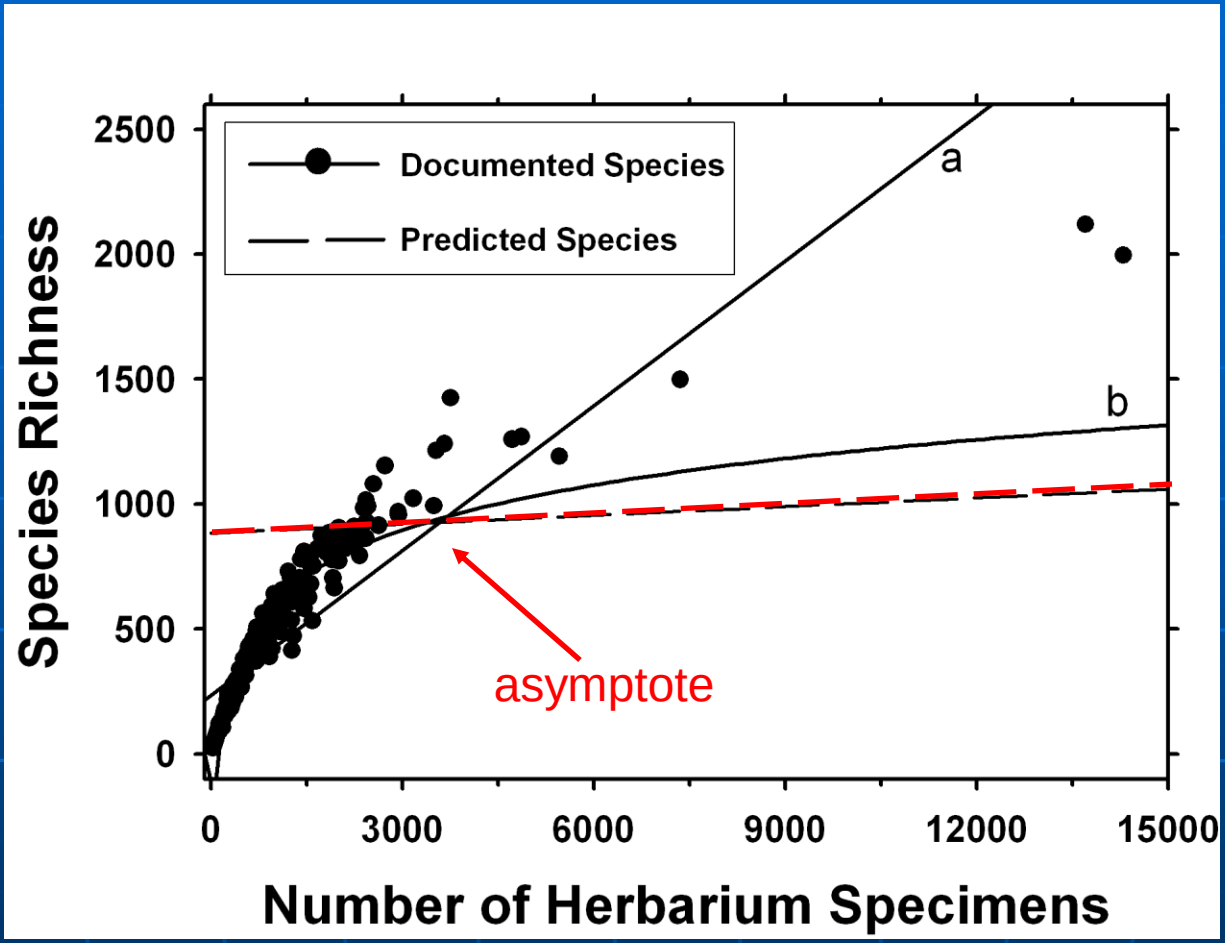


How best to address the Chicken and Egg argument?

- The phenomenon in question is coined the “Botanist Effect” and the data used to support this effect is species richness.

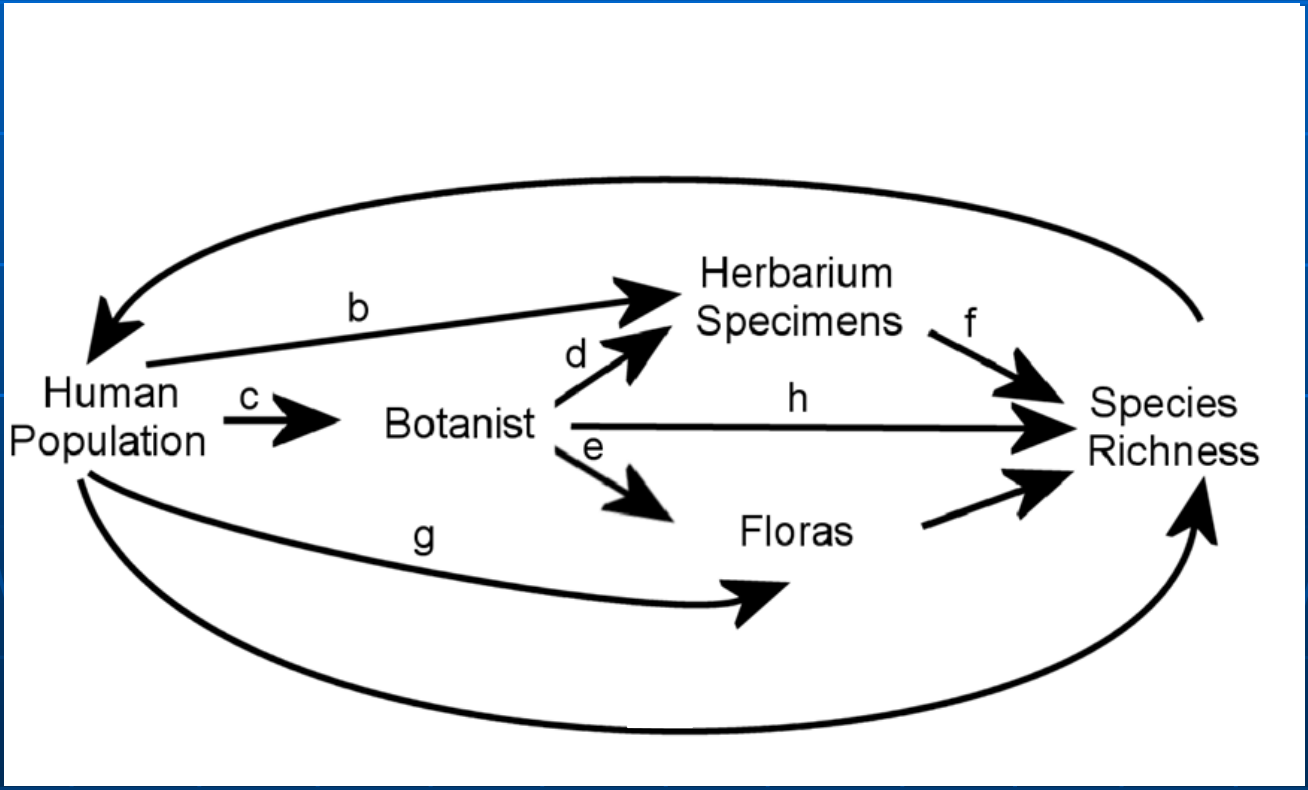
- However, botanist do not collect species richness, instead they collect herbarium specimens, that are in turn used to calculate species richness.

Preston's or collectors curve.



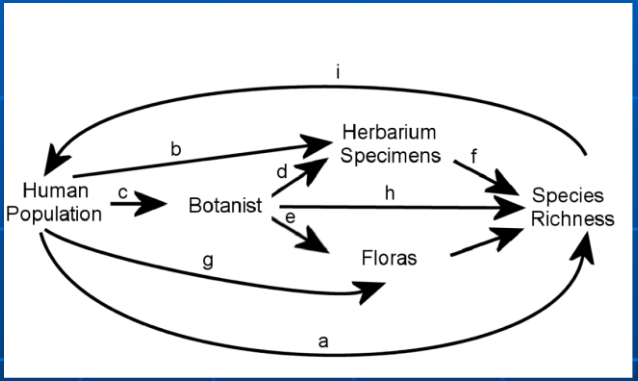
The collector's curve indicates that SR values increase per county with an increase in herbarium specimens or sampling effort. However, once the asymptote is reached the exponential increase in SR wanes with additional sampling effort.

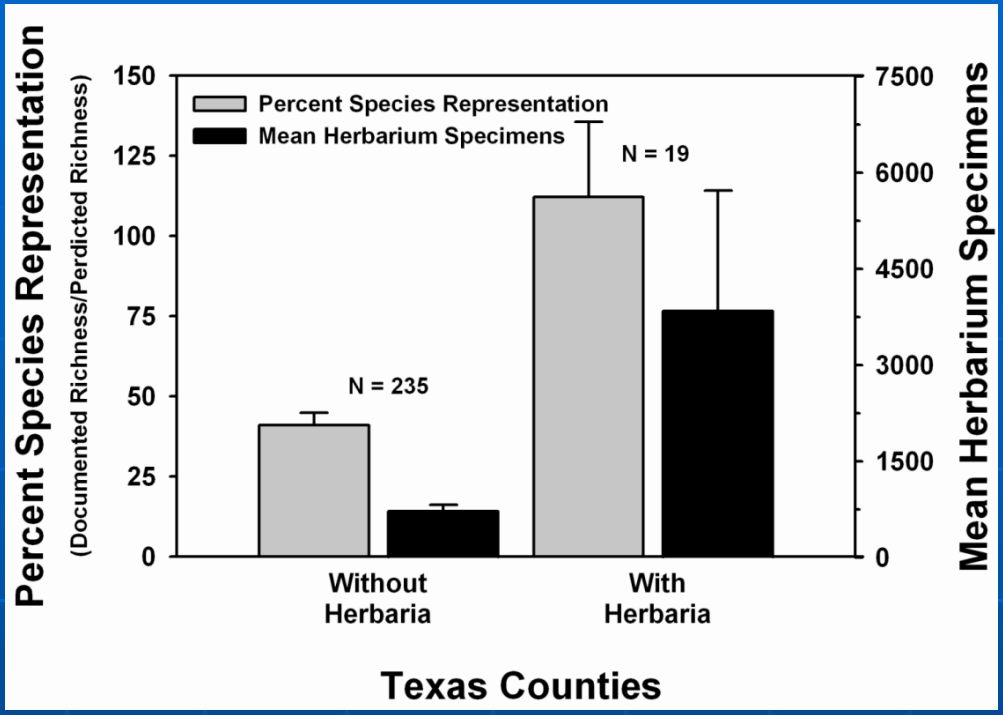
Chicken and Egg. Which came first SR values or Human Population?

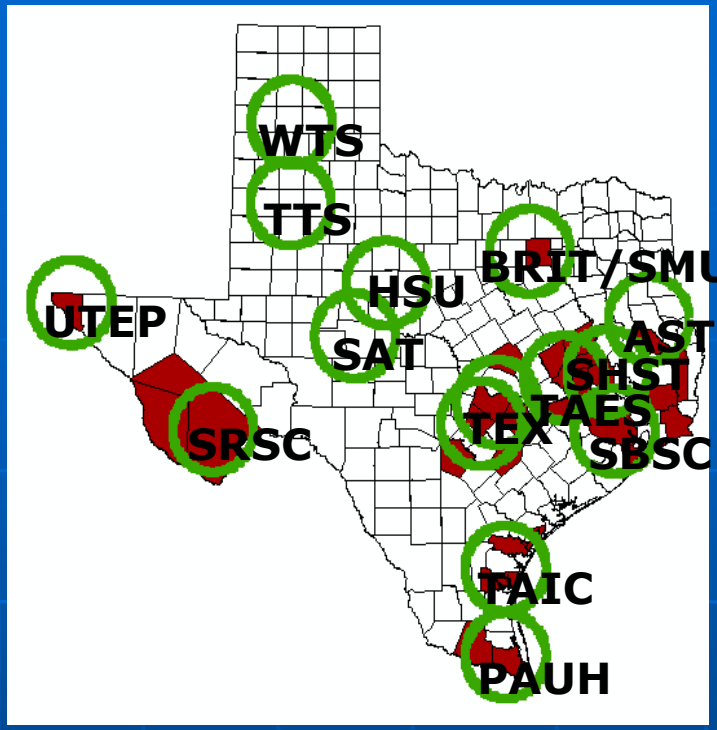


	Log-log	What dose this imply?
Herbarium to Specimens	.627	The more herbaria in a state the more herbarium specimens
SR to specimens	.616	Species Richness values increase with the accumulation of herbarium specimens.
All herbaria spec. to pop	.397	The larger the human population, the more herbarium specimens that are collected.
SR to Biomes	.372	The more Biomes within an area the higher the Species Richness.
GIS specimens to pop	.304	The larger the human population, the more herbarium specimens that are georeferenced.
Log Population to Herbaria	.287	As population increases the number of herbaria increases
SR to population	.243	SR increase as population increase or Human population increase with an increase in Species Richness
SR to area	.215	Species Richness values increase as area increases.
Biomes to population	.115	Human population increases with an increase in Biomes.
All specimens to Biomes	.0000	There is no correlation between specimens collected and the number of Biomes.
Specimens to area	-.025	The number of herbarium specimens collected within an area tends to slightly decrease as area increases.
Biomes to herbaria	.000	There is no correlation between the number of herbaria in a state and the number of its Biomes.
Herbaria to area	-.051	As area of a State increases the number of herbaria decreases.

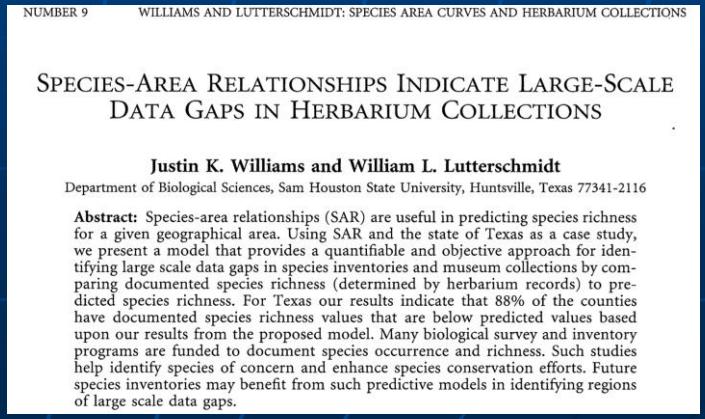
Mexico







Williams and Lutterschmidt (2006) demonstrated, for the state of Texas, that counties with herbaria had higher documented species richness and herbarium specimens than counties without herbaria.



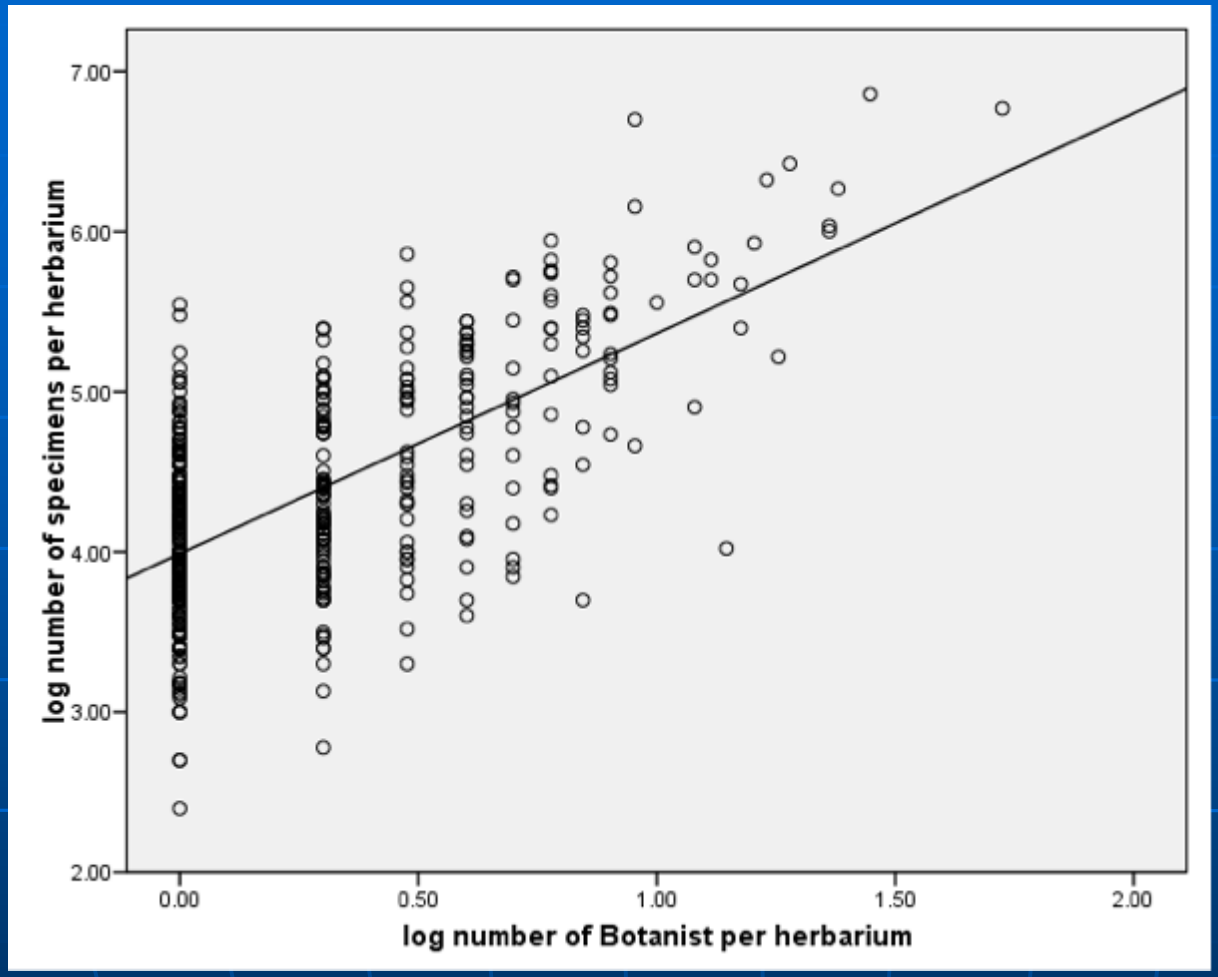
NUMBER 9 WILLIAMS AND LUTTERSCHMIDT: SPECIES AREA CURVES AND HERBARIUM COLLECTIONS

SPECIES-AREA RELATIONSHIPS INDICATE LARGE-SCALE DATA GAPS IN HERBARIUM COLLECTIONS

Justin K. Williams and William L. Lutterschmidt
Department of Biological Sciences, Sam Houston State University, Huntsville, Texas 77341-2116

Abstract: Species-area relationships (SAR) are useful in predicting species richness for a given geographical area. Using SAR and the state of Texas as a case study, we present a model that provides a quantifiable and objective approach for identifying large scale data gaps in species inventories and museum collections by comparing documented species richness (determined by herbarium records) to predicted species richness. For Texas our results indicate that 88% of the counties have documented species richness values that are below predicted values based upon our results from the proposed model. Many biological survey and inventory programs are funded to document species occurrence and richness. Such studies help identify species of concern and enhance species conservation efforts. Future species inventories may benefit from such predictive models in identifying regions of large scale data gaps.

Williams, J. K., and W. I. Lutterschmidt. 2006. Species area relationships indicate large-scale data gaps in herbarium collections.



Rsq = .433

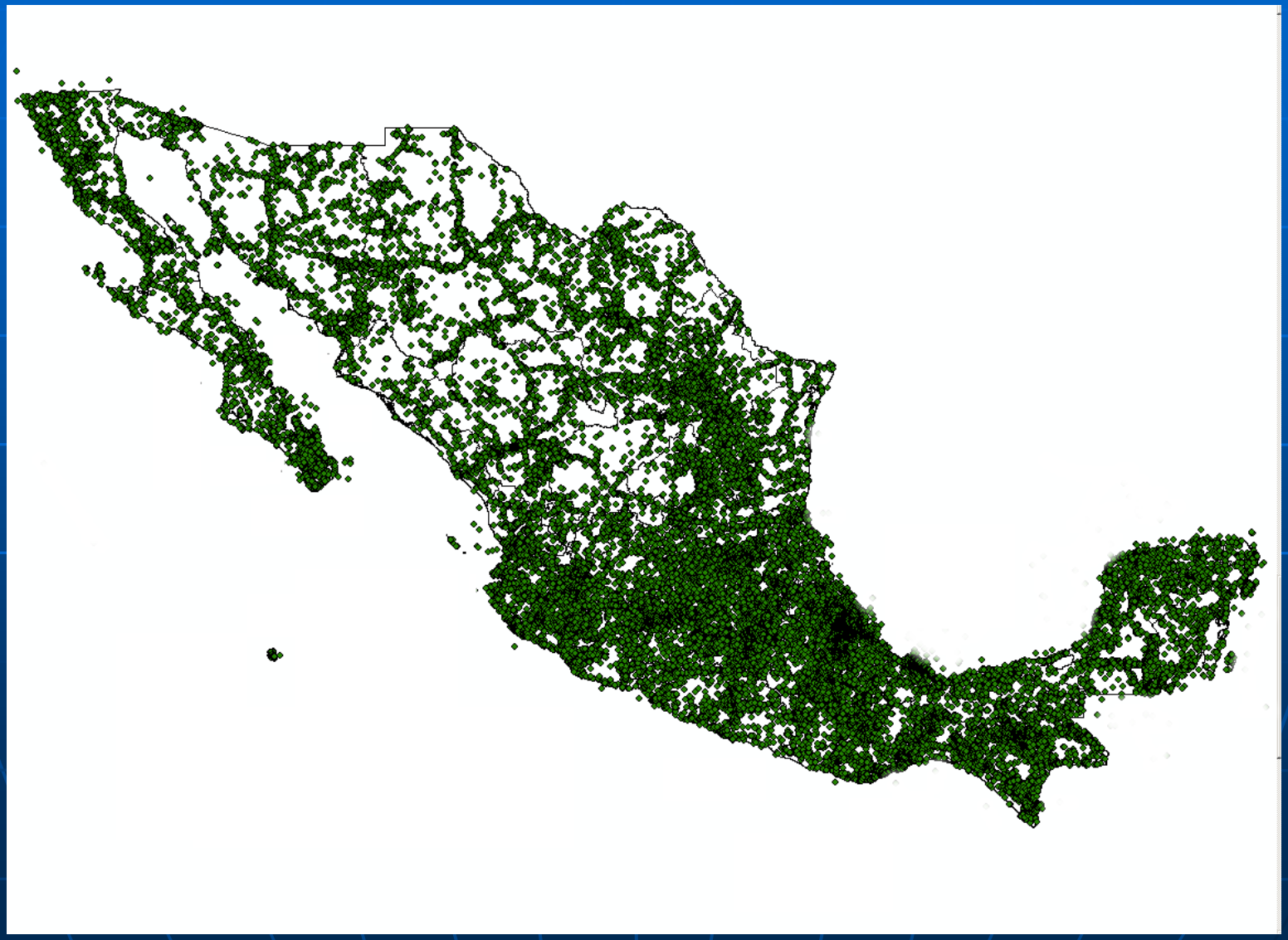
- Regression curve of the number of herbarium specimens per herbarium to the number of Botanist per herbarium. Curve shows a positive correlation, i.e. with an increase in Botanist there is an increase in the number of Herbarium specimens. Distribution of 510 herbaria from the Continental United States.

When addressing the Botanist Effect it is perhaps best to evaluate Herbarium Specimens and not Species Richness.

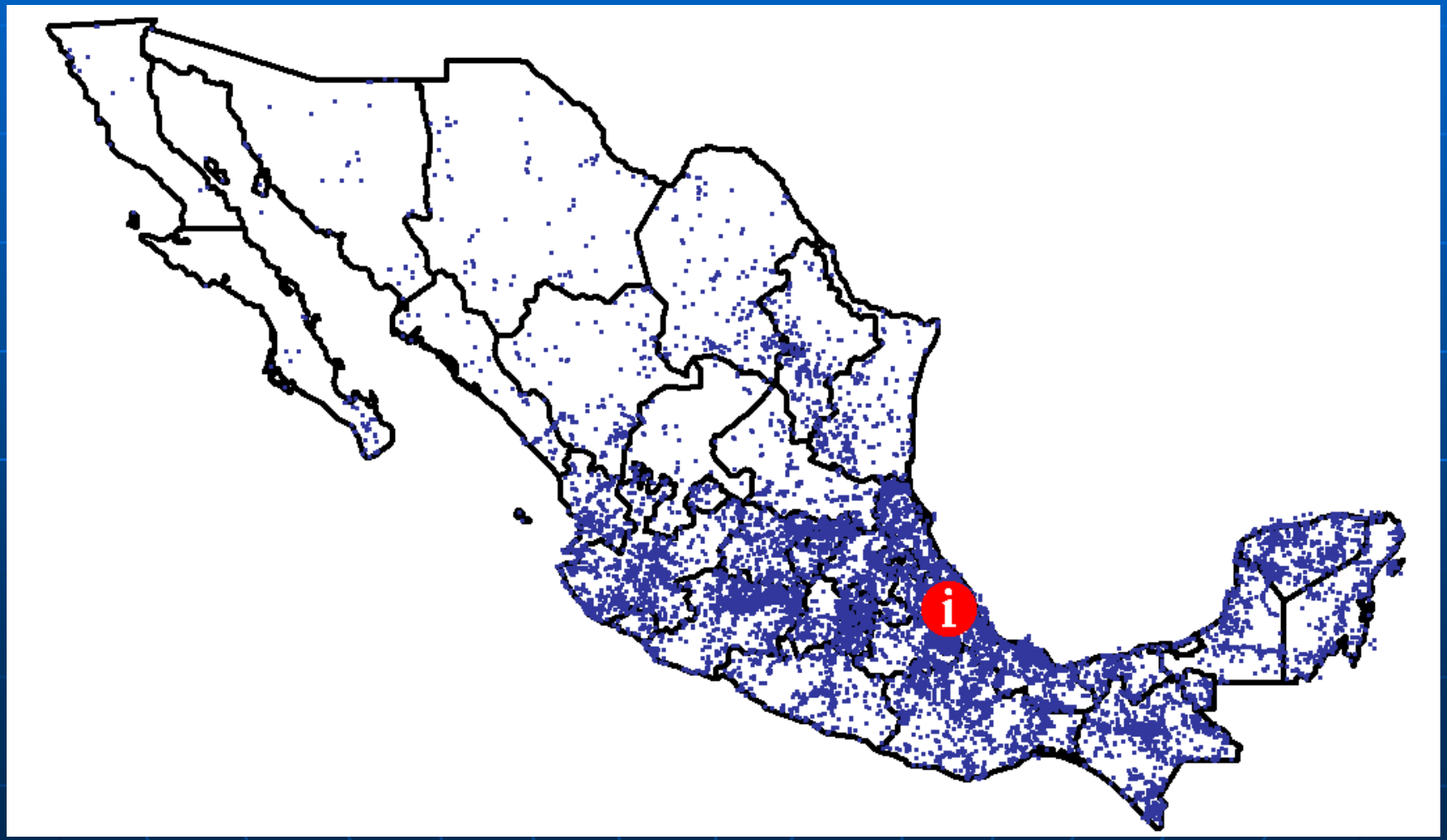
Will test the Botanist Effect by measuring the proximal effect of herbarium distance on herbarium collecting.

- Will need to locate herbaria that have geo-referenced their herbarium specimens.
- Test the hypothesis:
  - If the “botanist effect” is an actual phenomenon then the number of herbarium specimens will decrease as distance from the herbarium increases.”

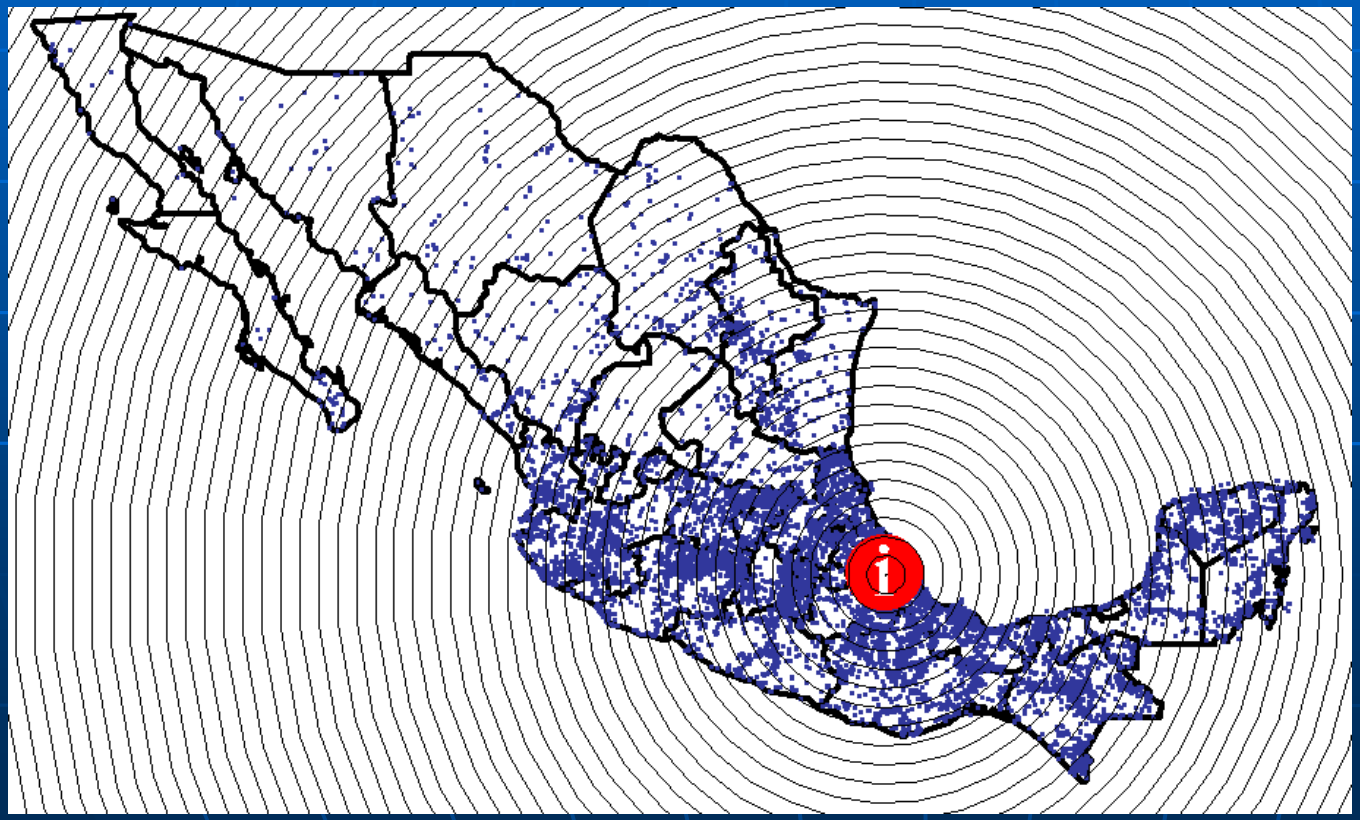
Mexico has geo-referenced ca 700,000 herbarium specimens from eight different herbaria.



Isolated only the specimens from XAL; herbarium in Xalapa.
Then measure the number of specimens at 10 km increments.

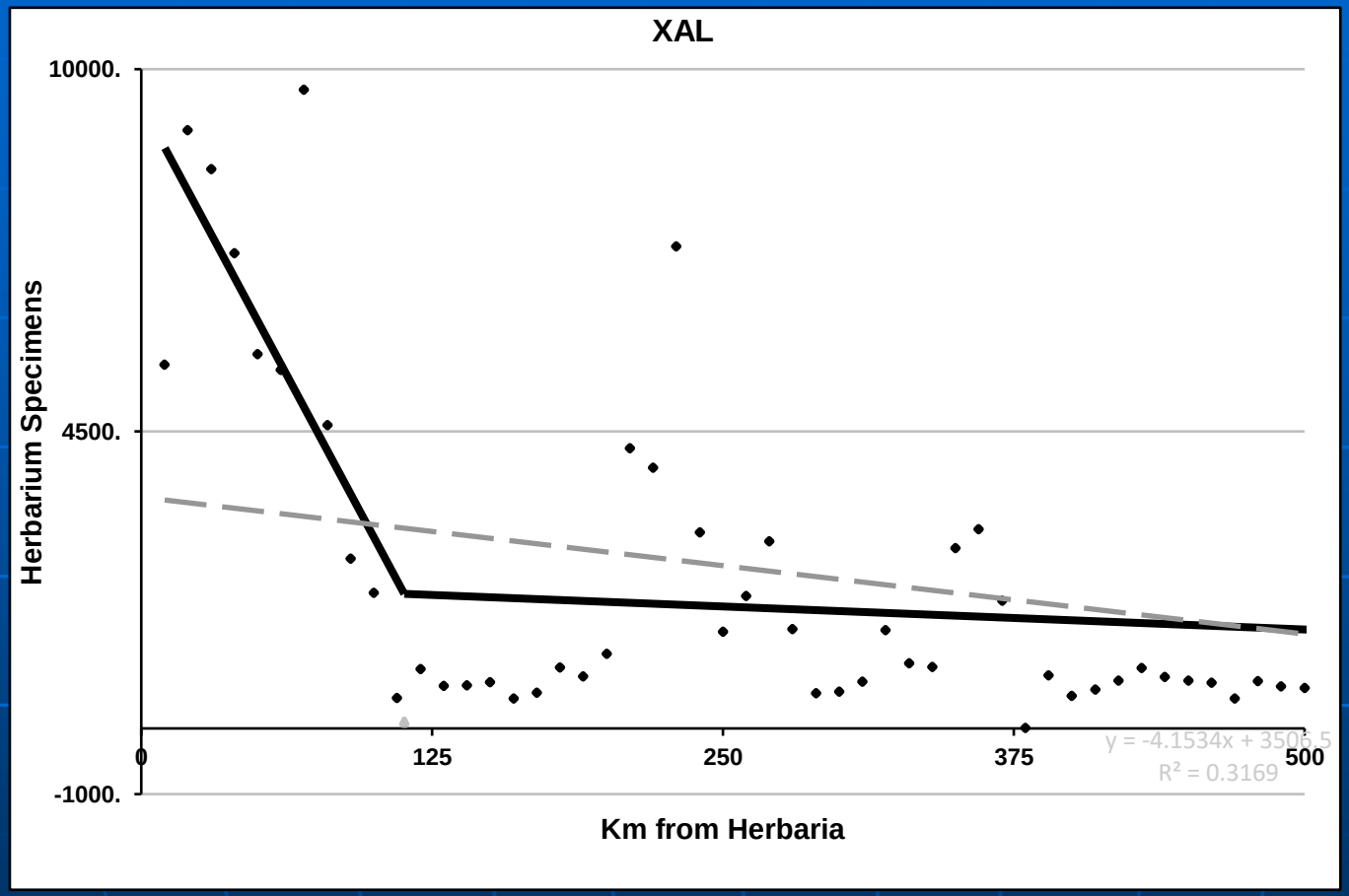


A total of 150,629 specimens, approximately 21% of all data-based Mexican specimens. The furthest is 1300 km from XAL.



Distance KM	# of specimens
10	5509
20	9067
30	8482
40	7201
50	5673
60	5430
70	9679
80	4592
90	2574
100	2050
110	461
120	897
130	639
140	653
150	701
160	449
170	538
180	917
190	782
200	1127
210	4248
220	3948

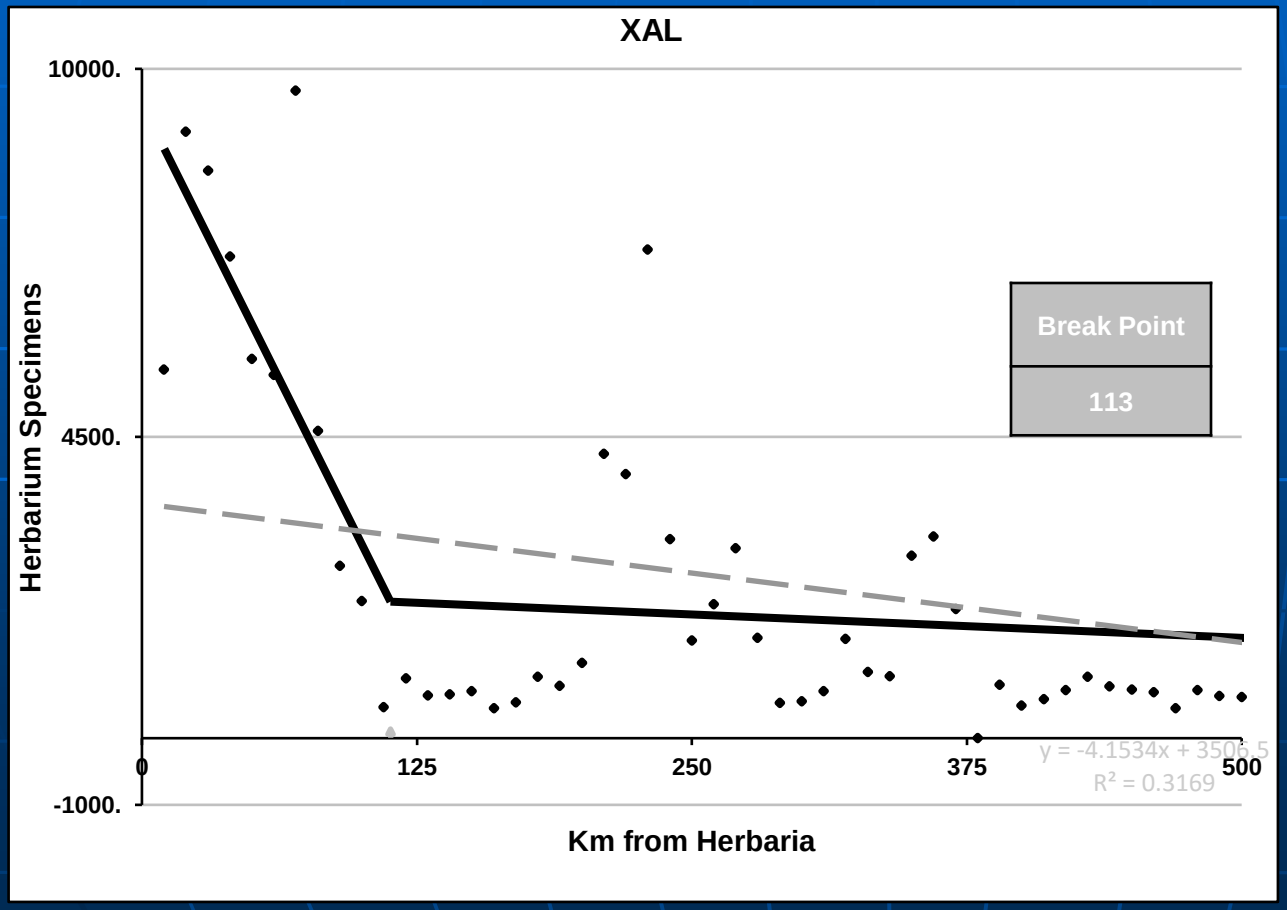
Segmented - Piecewise Regression



$$Y = a_1 + b_1 x$$
$$Y = a_2 + b_2 x$$
$$a_1 + b_1 c = a_2 + b_2 c$$
$$a_2 = a_1 + c(b_1 - b_2)$$

Piecewise regression is useful when studying independent variables that partition into clusters or separate line segments. When the partitioned data is analyzed the results of the piecewise analysis can identify the edge between the different segments. This edge is often referred to as the *threshold* or *breakpoint*. In the analysis presented here the breakpoint that is identified is the distance where a significant decrease in herbarium collections occurs.

Both Linear and Piecewise regression indicate that indeed the number of specimens collected decreases as distance from the herbarium increases.



Distance KM	# of specimens
10	5509
20	9067
30	8482
40	7201
50	5673
60	5430
70	9679
80	4592
90	2574
100	2050
110	461
120	897
130	639
140	653
150	701
160	449
170	538
180	917
190	782
200	1127
210	4248
220	3948

Interactive Piecewise Regression

Adjust scroll bar, watch change in r2 to find Break Point Year that maximizes overall r^2

	Break Point	Row
8	**113**	11

Y Axis Title	**Day of Year**

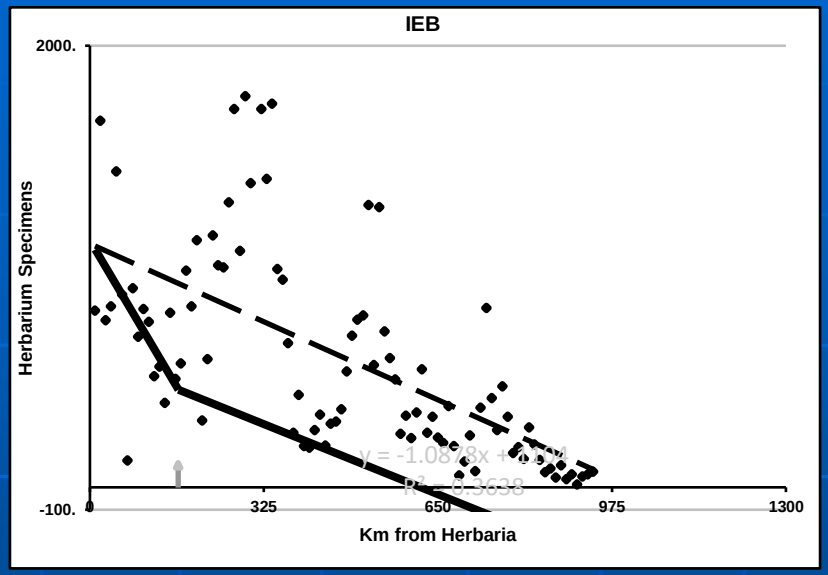
113;r2 = 0.72; Sig-0.00000% **5.0000%**

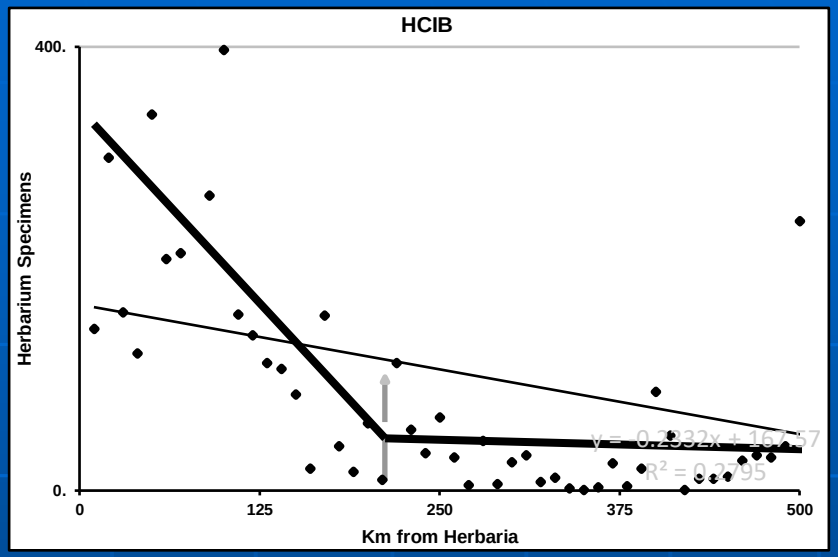
Piecewise Regression Results

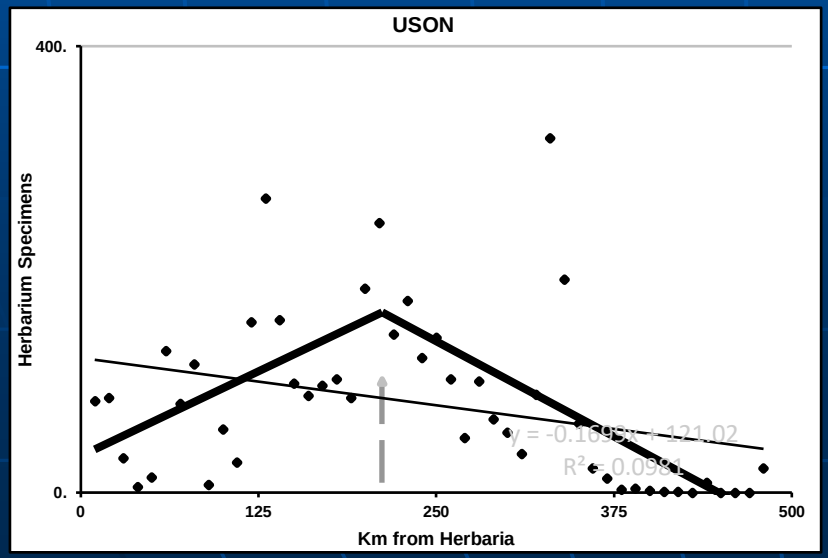
Par	All Data	Line 1	Line 2	PW Regr
Intercept - a	3026.19	9460.40	2196.85	
Slope - b	-2.851	-65.676	-1.397	
No. Obs	130	11	120	130
SS Reg	148,821,802	47,447,232	28,105,988	304,856,691
SS Res	277,291,394	41,236,573	80,855,955	121,256,506
SS - Total	426,113,196	88,683,806	108,961,943	426,113,196
r^2	0.3493	0.5350	0.2579	0.7154
df	128	9	118	128
F	68.70	10.36	41.02	321.81
Sig of F	0.000000%	1.1%	0.000000%	0.000000000%
Fcrit	3.9	5.1	3.9	3.9

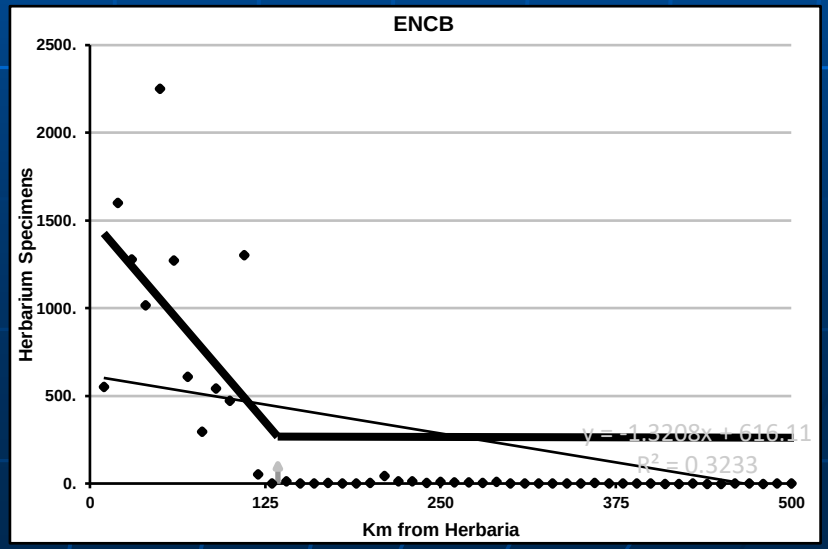
Est - Line 1	
10	8803.6
113	2039.0
Est - Line 2	
113	2039.0
1300	380.6

This phenomenon is observed in 4 other herbaria in Mexico.

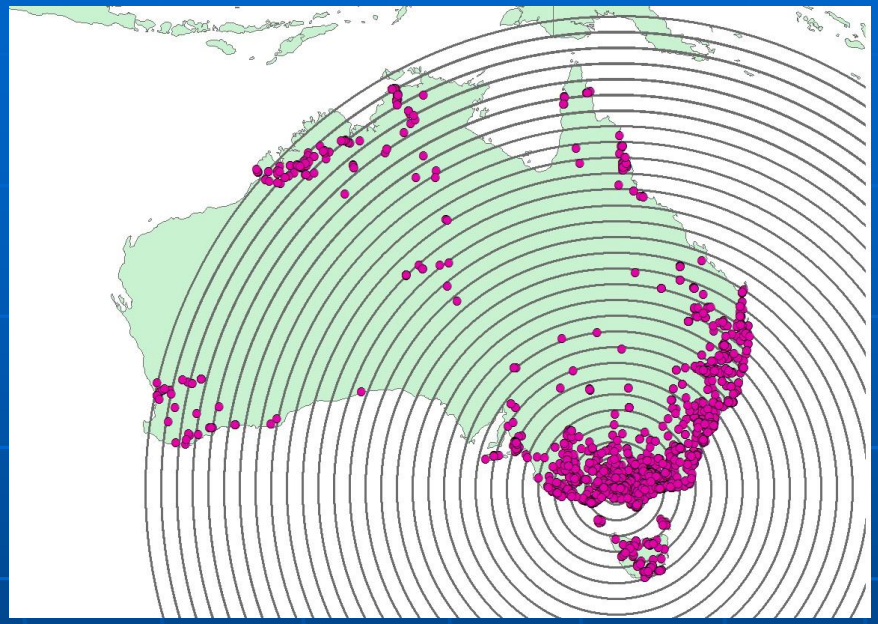


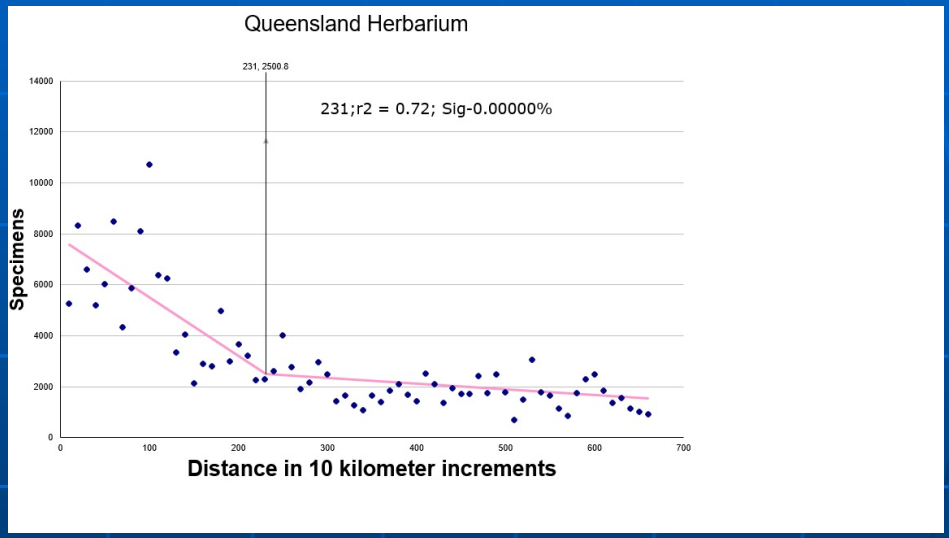


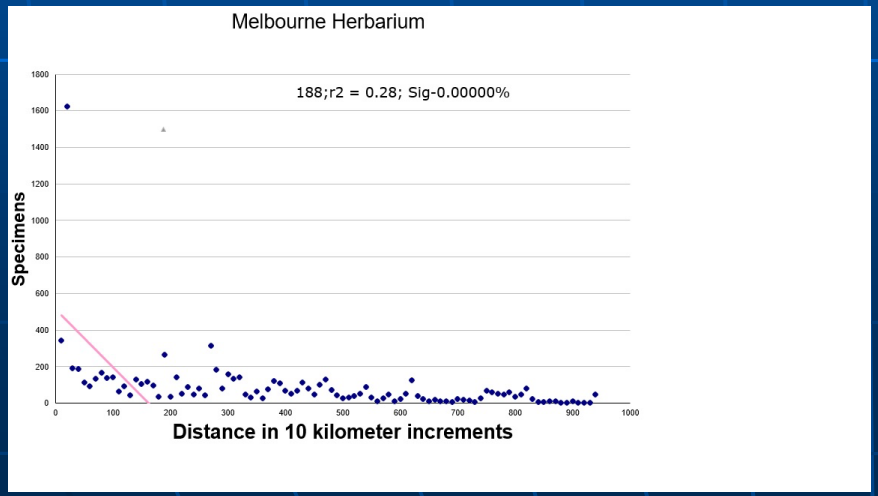




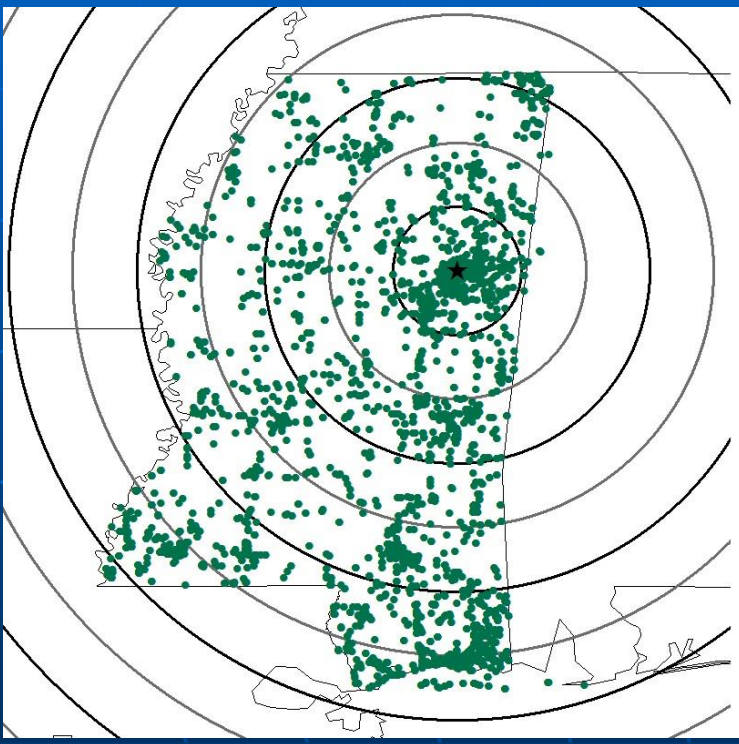
AUSTRALIA

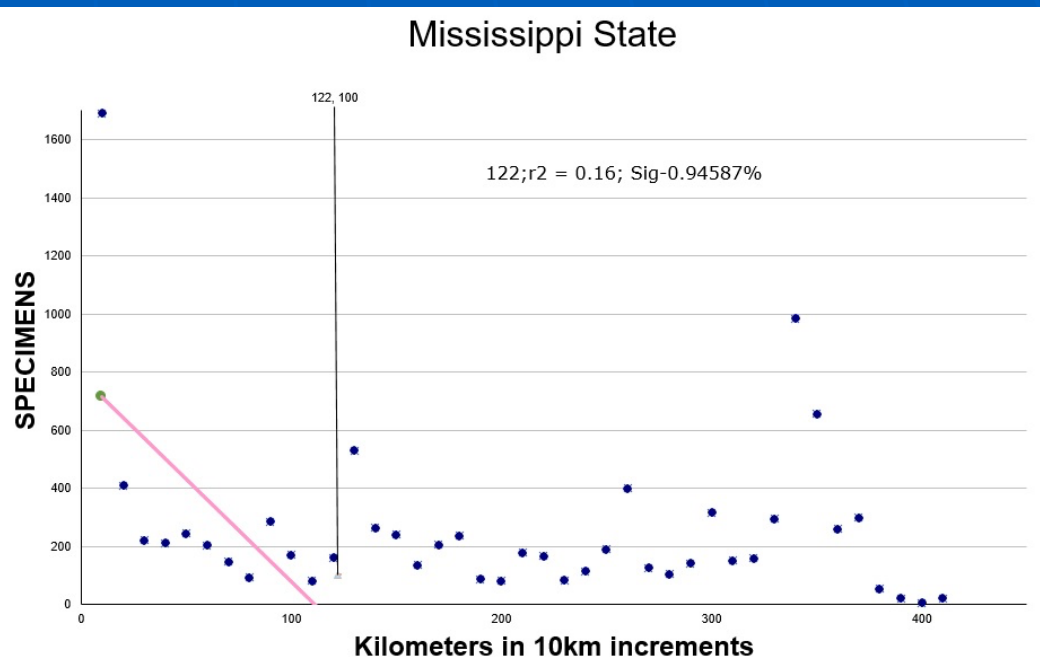



Queensland Herbarium
231, 2500.8
231;r2 = 0.72; Sig-0.00000%
Specimens
Distance in 10 kilometer increments

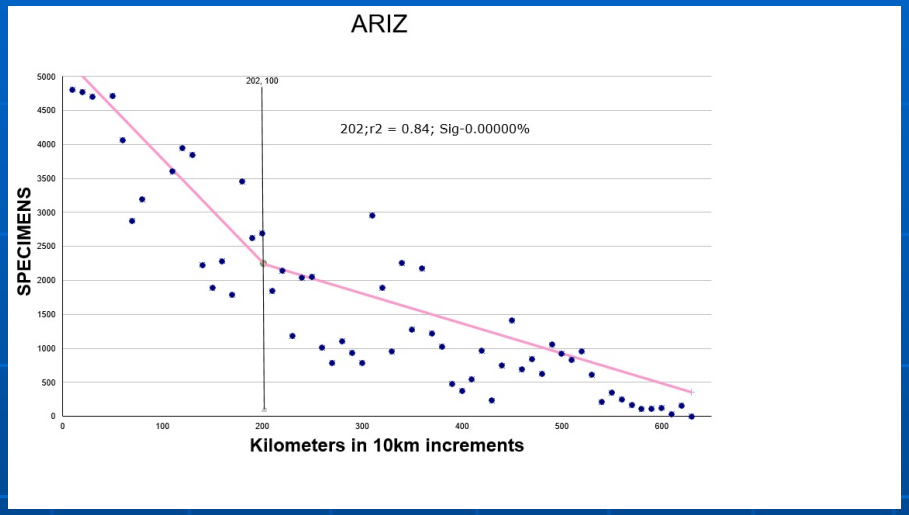

Melbourne Herbarium
188;r2 = 0.28; Sig-0.00000%
Specimens
Distance in 10 kilometer increments

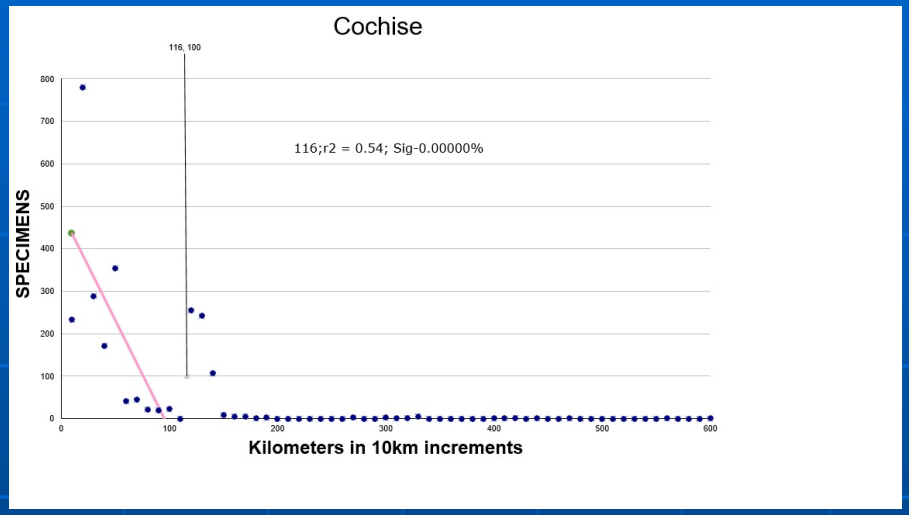
MISSISSIPPI

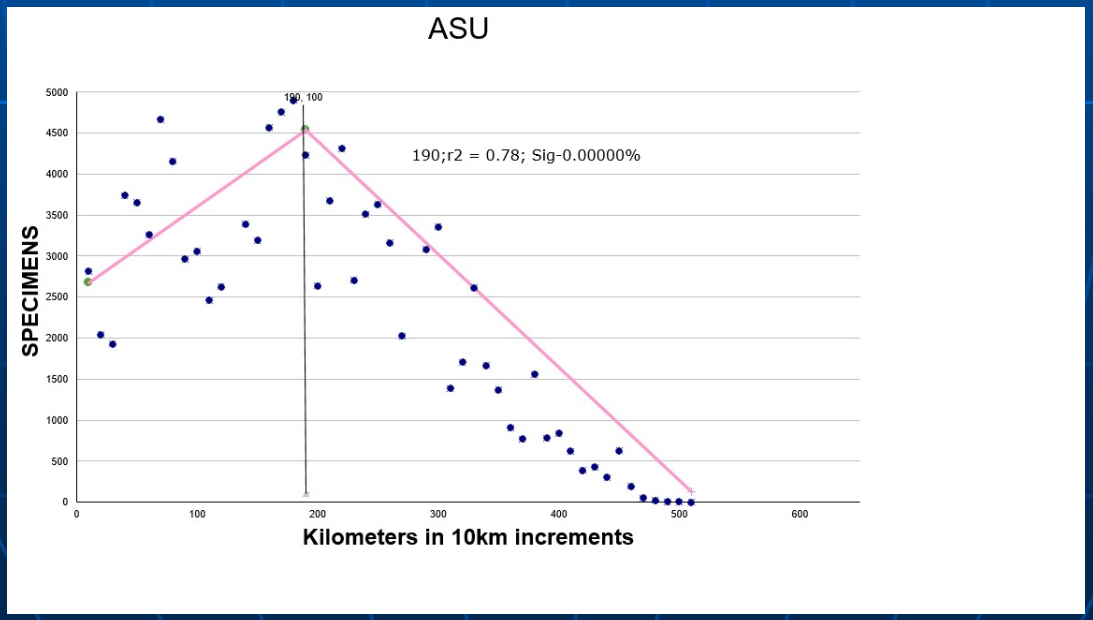




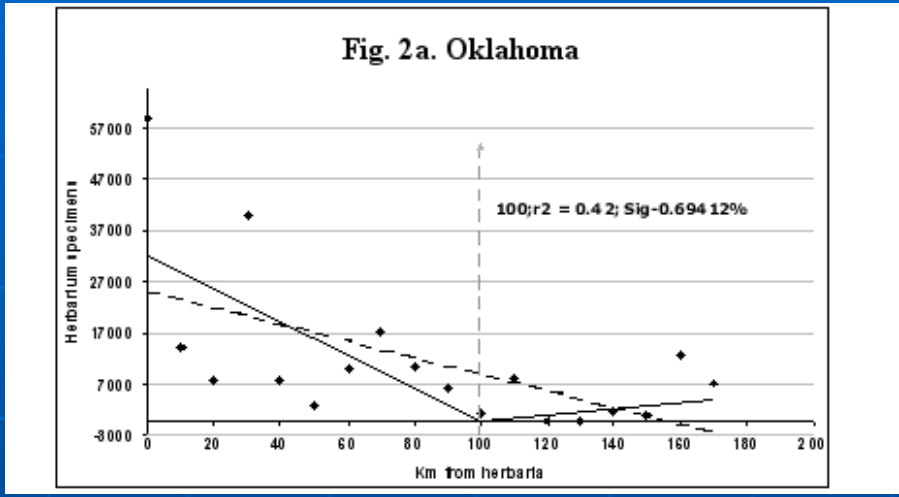
ARIZONA

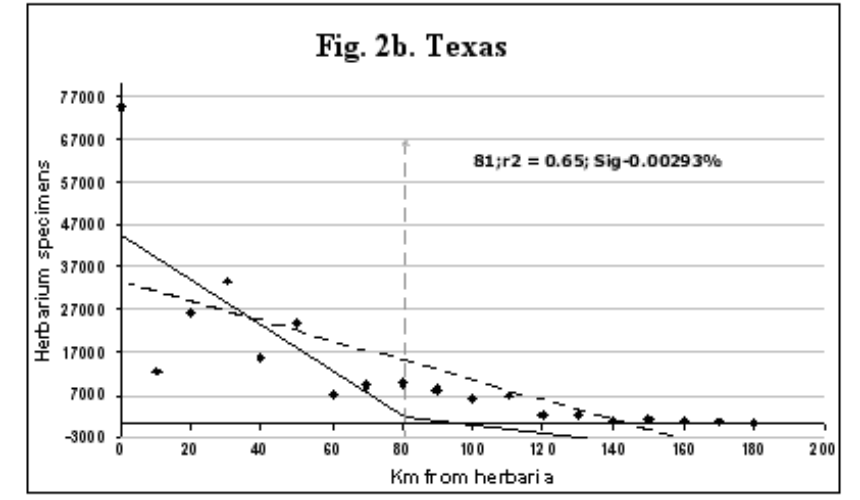

ARIZ
202, 100
202;r2 = 0.84; Sig-0.00000%
SPECIMENS
Kilometers in 10km increments

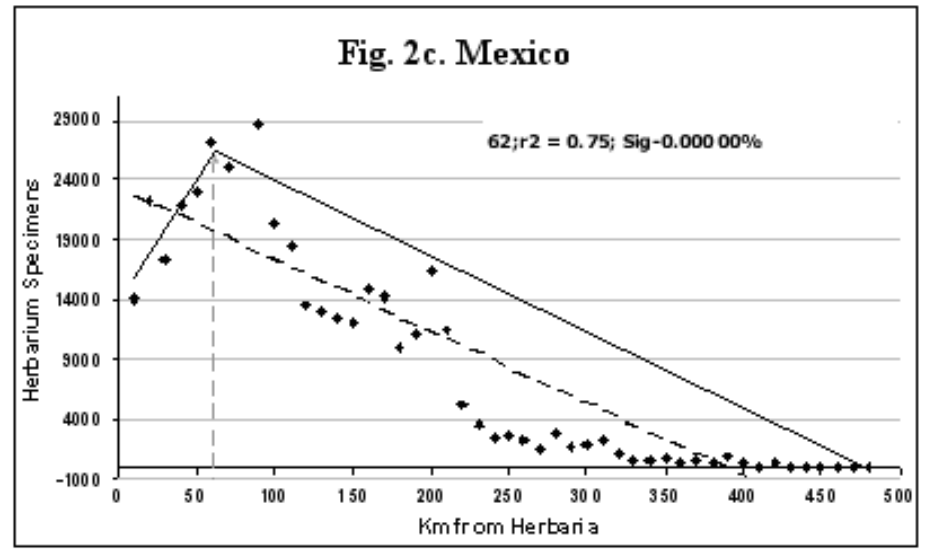

Cochise
116, 100
116;r2 = 0.54; Sig-0.00000%
SPECIMENS
Kilometers in 10km increments


ASU
190, 100
190;r2 = 0.78; Sig-0.00000%
SPECIMENS
Kilometers in 10km increments

ENTIRE COLLECTIONS







RESULTS

Location	Break Point km
ASU	190
ARIZ	202
COCHISE	116
HCIB	212
IEB	165
Melbourne	188
MISS	122
Queensland	231
USON	212
XAL	113
Oklahoma	106
Texas	81
Mexico	62

Range

Entire region: 62-106 Km
Individual Herbarium: 113-231 Km

Average

Entire region: 154 Km
Individual Herbarium: 175 km

Conclusions

- Data from numerous geographic areas and sources confirms that indeed a spatial bias in herbarium collecting exists.

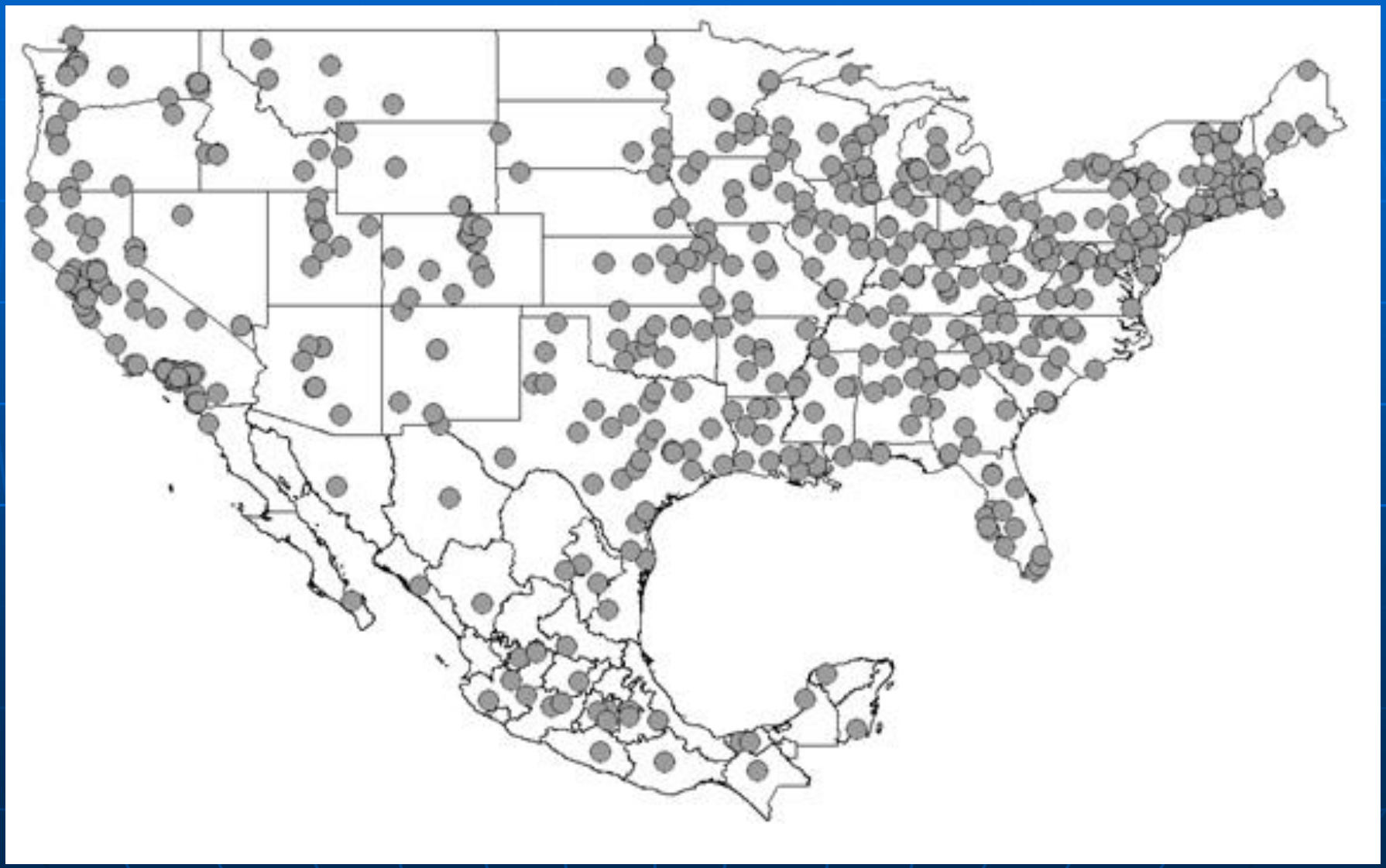
Conclusions

- Data from numerous geographic areas and sources confirms that indeed a spatial bias in herbarium collecting exists.
- Botanist clearly, show a bias towards collecting closer to herbaria.

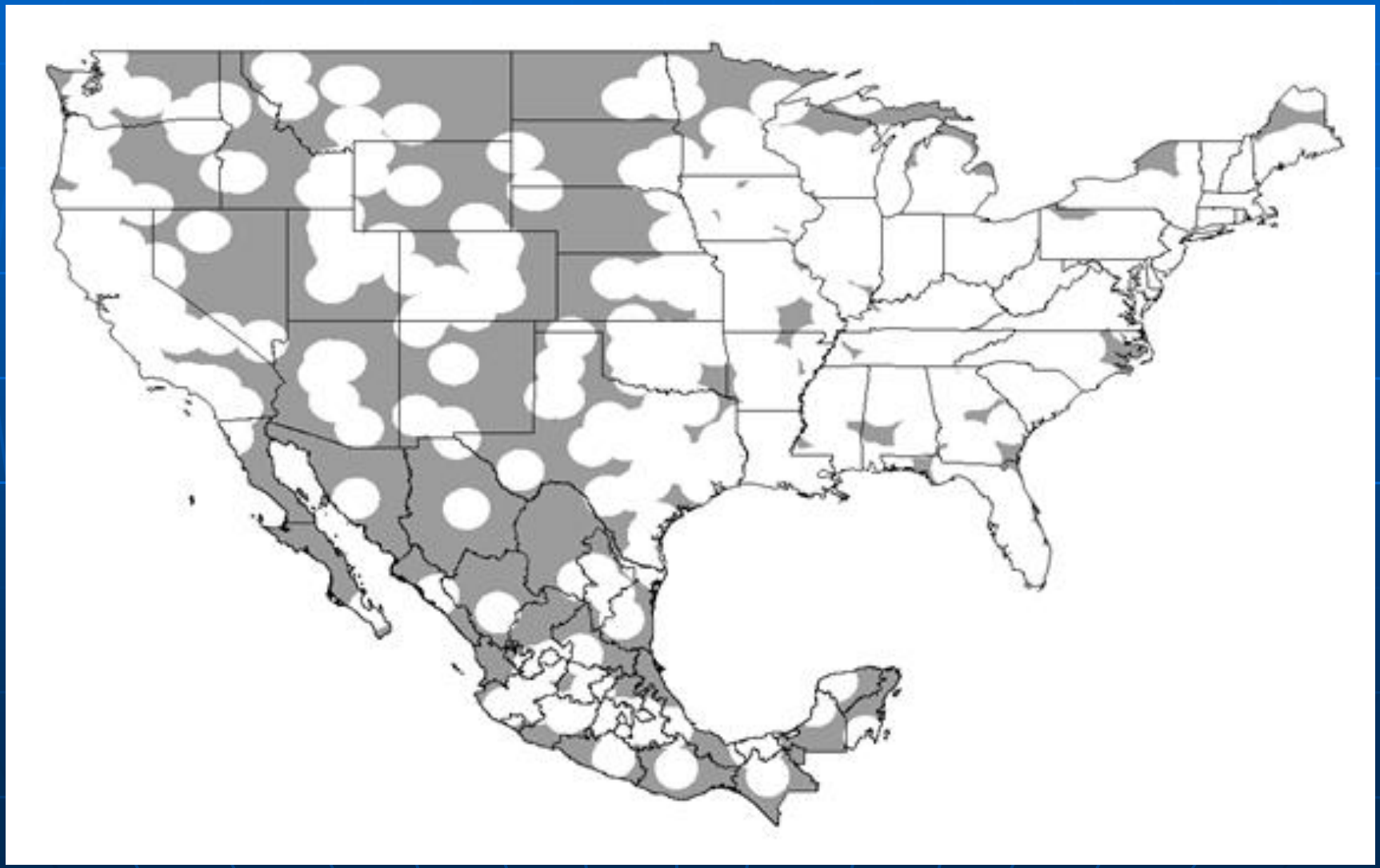
Conclusions

- Data from numerous geographic areas and sources confirms that indeed a spatial bias in herbarium collecting exists.
- Botanist clearly, show a bias towards collecting closer to herbaria.
- The threshold distance in collecting from herbaria averages to about ca. 180 km.

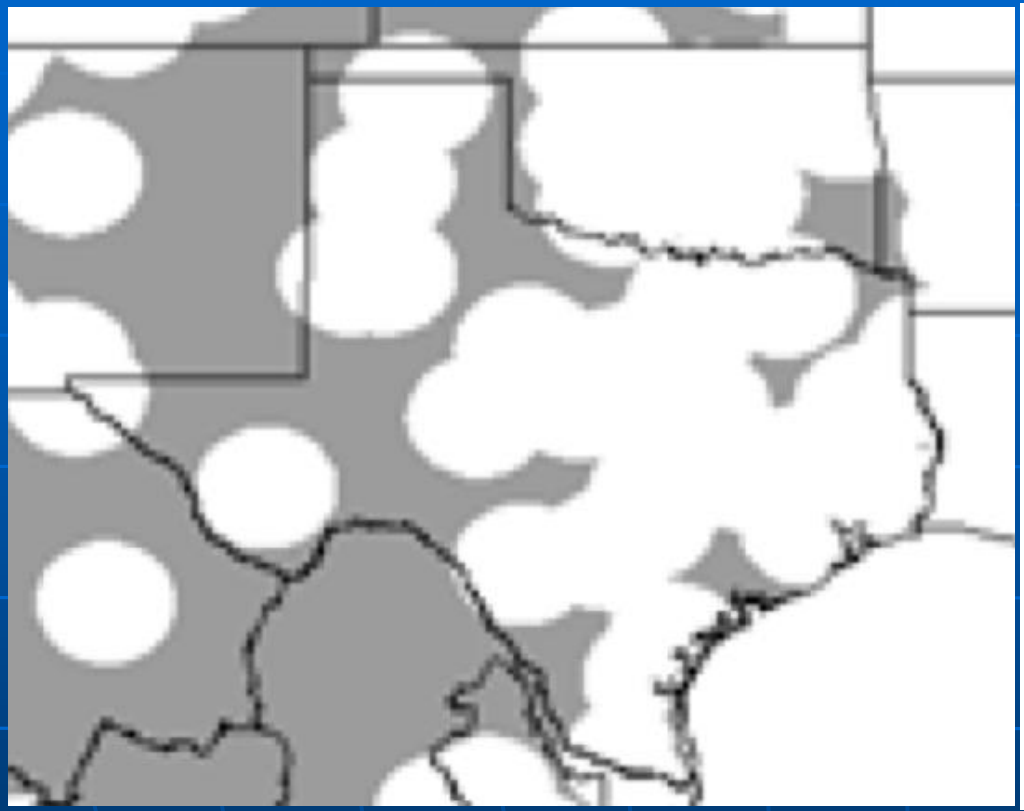
Herbaria in the Continental US and Mexico.

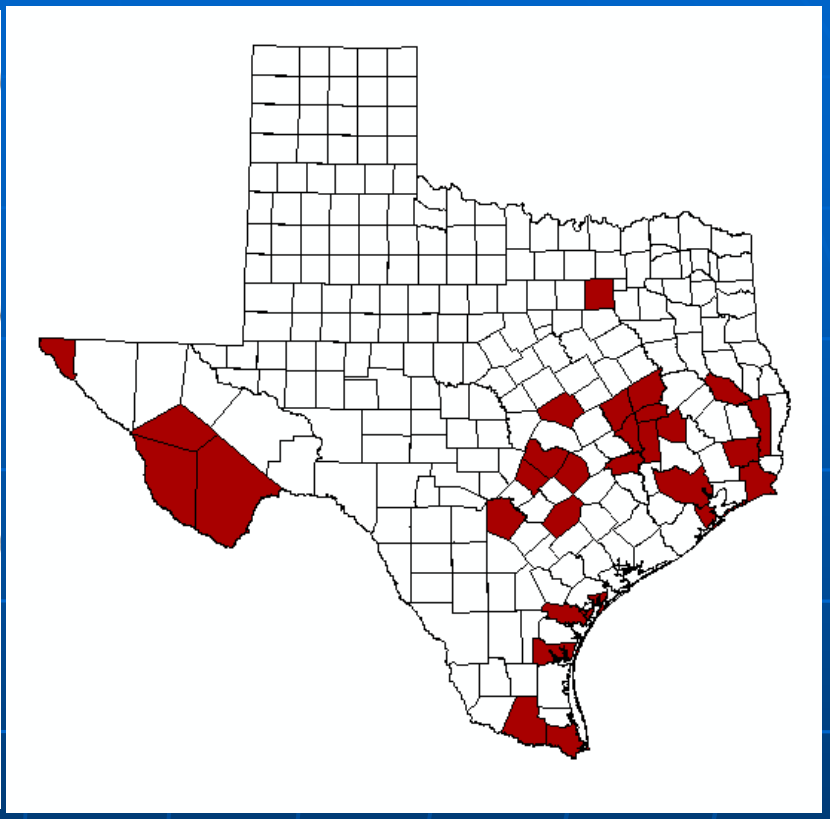


Areas 180 km from herbaria in the Continental US and Mexico.



DATA GAP COMPARISON





SPECIES-AREA RELATIONSHIPS INDICATE LARGE-SCALE DATA GAPS IN HERBARIUM COLLECTIONS

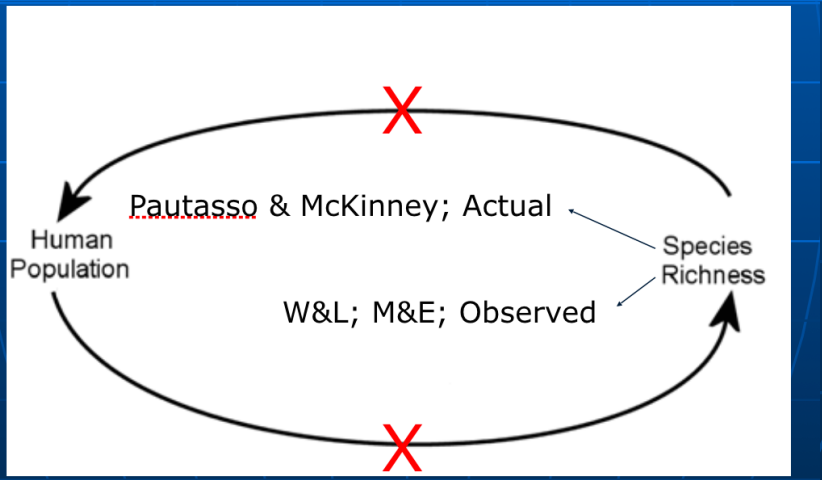
Justin K. Williams and William L. Lutterschmidt
Department of Biological Sciences, Sam Houston State University, Huntsville, Texas 77341-2116

Abstract: Species-area relationships (SAR) are useful in predicting species richness for a given geographical area. Using SAR and the state of Texas as a case study, we present a model that provides a quantifiable and objective approach for identifying large scale data gaps in species inventories and museum collections by comparing documented species richness (determined by herbarium records) to predicted species richness. For Texas our results indicate that 88% of the counties have documented species richness values that are below predicted values based upon our results from the proposed model. Many biological survey and inventory programs are funded to document species occurrence and richness. Such studies help identify species of concern and enhance species conservation efforts. Future species inventories may benefit from such predictive models in identifying regions of large scale data gaps.

Counties where reported species richness matches, exceeds or approximates predicted species richness (red). Note that for the majority of counties reported species richness does not match predicted richness.

Solutions?

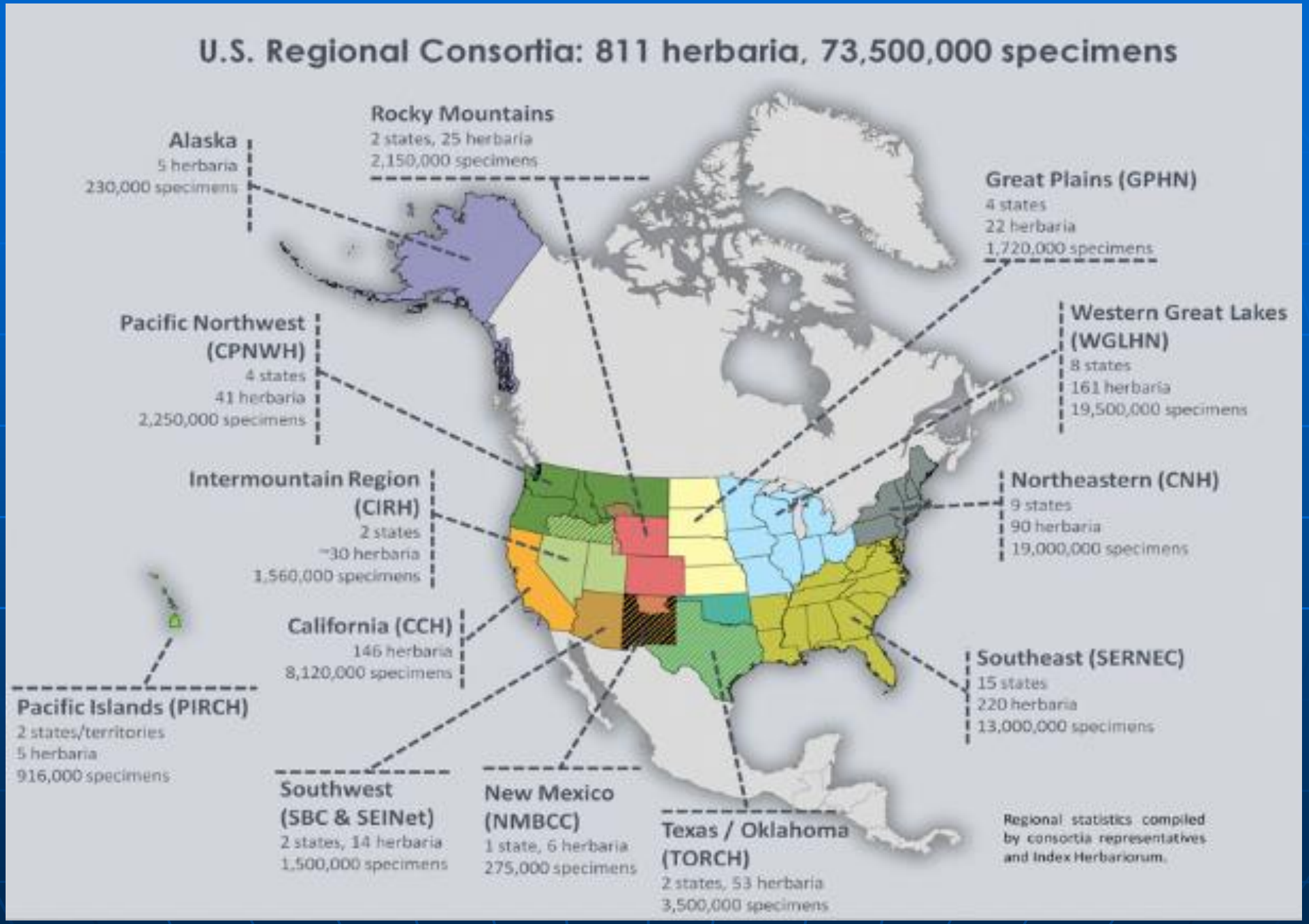
- Encourage students to collect. Perhaps as part of an undergraduate research project.
- Promote collecting in areas outside the 180 km threshold.
- Encourage state wide herbaria to data-base collections.
- Run models in other areas to identify spatial gaps.
- Redefine the definition of the Botanist Effect.



Thank You

The earliest suggestion of spatial bias was made in 1768 by the famed naturalist Gilbert White when he wrote:

"It is, I find, in zoology as it is in botany: all nature is so full, that that district produces the greatest variety which is the most examined."


U.S. Regional Consortia: 811 herbaria, 73,500,000 specimens
Alaska
5 herbaria
230,000 specimens
Rocky Mountains
2 states, 25 herbaria
2,150,000 specimens
Great Plains (GPHN)
4 states
22 herbaria
1,720,000 specimens
Pacific Northwest (CPNWH)
4 states
41 herbaria
2,250,000 specimens
Western Great Lakes (WGLHN)
8 states
161 herbaria
19,500,000 specimens
Intermountain Region (CIRH)
2 states
~30 herbaria
1,560,000 specimens
Northeastern (CNH)
9 states
90 herbaria
19,000,000 specimens
California (CCH)
146 herbaria
8,120,000 specimens
Pacific Islands (PIRCH)
2 states/territories
5 herbaria
916,000 specimens
Southeast (SERNEC)
15 states
220 herbaria
13,000,000 specimens
Southwest (SBC & SEINet)
2 states, 14 herbaria
1,500,000 specimens
New Mexico (NMBCC)
1 state, 6 herbaria
275,000 specimens
Texas / Oklahoma (TORCH)
2 states, 53 herbaria
3,500,000 specimens
Regional statistics compiled by consortia representatives and Index Herbariorum.