



Sử dụng GIS để nghiên cứu tác động môi trường

Lê Việt Phú

Chương trình Giảng dạy Kinh tế Fulbright

5-2015

Phần 2 – Mô hình không gian

- Spatial interpolation
- Kiểm định tự tương quan trong không gian - Spatial autocorelation tests
- Mô hình kinh tế lượng không gian - Spatial econometric modeling

Spatial modeling

- Tạo thư mục Spatial Modeling ở Desktop, copy các files sau đây:
 - counties.sdc (gồm có 4 files có cùng một tên nhưng khác đuôi file) có các thông tin về nhân khẩu học.
 - PRISM_ppt, PRISM_tmean, PRISM_dem là các file ảnh (raster) độ phân giải cao (4km cho mỗi pixel) về lượng mưa, nhiệt độ trung bình hàng năm, và địa hình của 48 bang nước Mỹ. Đơn vị là độ C, mm, và m. PRISM được xem là dữ liệu khí tượng tốt nhất ở Mỹ, do trường Oregon cung cấp.
 - farmValue.xls là file chứa thông tin giá đất nông nghiệp (đơn vị: \$/acre) tại mỗi county (hạt).
 - station_data.xls là file chứa thông tin về các quan sát khí tượng (nhiệt độ trung bình, lượng mưa trung bình) của các trạm khí tượng khắp nước Mỹ từ cuối thế kỷ 18 đến 2014. Mỗi trạm có thông tin về tọa độ và độ cao (longitude/latitude/altitude) và năm quan sát.

Bước 1 – khám phá dữ liệu

- Đọc dữ liệu vào ArcGIS
- Properties -> Open Attribute Table
- Lọc phần khu vực quan tâm (48 bang liền kề – contiguous states), lưu thành file riêng.
- Đọc dữ liệu PRISM, lựa chọn màu hiển thị phù hợp (nhiệt độ cao -> red, thấp -> blue; lượng mưa cao -> blue, thấp -> red; độ cao thấp -> trắng, cao -> đen ...)
- Có nhận xét gì về các bản đồ đã vẽ? Có hợp lý không?

Bước 2 – spatial interpolation

- Dùng dữ liệu thu nhập được từ các trạm khí tượng để vẽ bản đồ khí tượng cho cả nước
- Đọc dữ liệu station_data.xls: dùng ArcCatalog để import dữ liệu bằng tọa độ (X,Y,Z) và hệ tọa độ NAD 1983, lưu thành file mới có tên station_data.shp cùng một thư mục. Add station_data vào ArcMap
- Nhận xét? Dữ liệu có đầy đủ không?

Bước 2 – spatial interpolation (cont)

- Vào ArcToolbox/Spatial Analyst Tools/Interpolation/IDW
- Chọn input feature là station_data, Z value là tm hay prec (giá trị muốn tính), output raster là US_tm, rồi OK
- File vừa tạo ra là temperature surface – nhiệt độ trung bình bằng cách lấy bình quân gia quyền, với quyền số là nghịch đảo khoảng cách từ mỗi trạm khí tượng đến địa điểm cần tính (tại sao?)

$$\bar{T} = \frac{\sum_i \frac{1}{d_i} T_j}{\sum_i \frac{1}{d_i}}$$

- Có nhiều cách tích trọng số khác nhau.

So sánh giữa bản đồ tự tạo ra với dữ liệu của PRISM

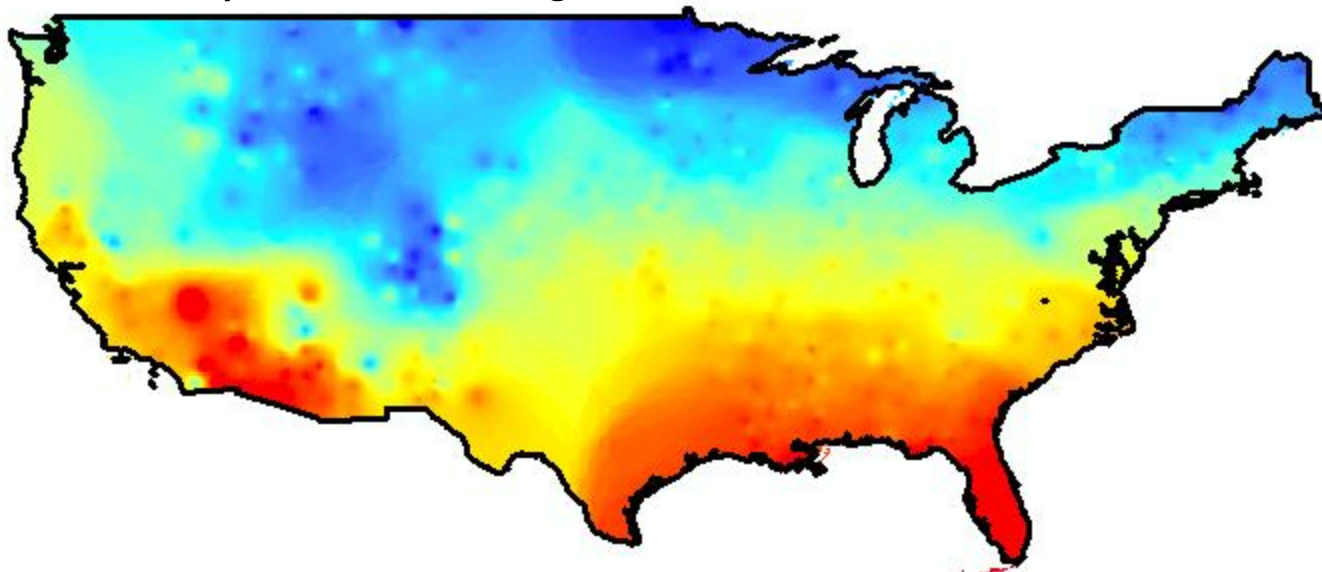
- Cắt lọc phần dữ liệu nằm trong biên giới nước Mỹ:
 - Dùng công cụ Data Management/Generalization/Dissolve để tạo biên giới nước Mỹ từ file counties, save vào file US_border.shp
 - Dùng file US_border để cắt lọc dữ liệu cần quan tâm từ file US_tm bằng công cụ Spatial Analyst/Extraction/Extract by mask. Input raster là US_tm, mask là US_border, output raster là US_tm_selfie.
 - Kiểm tra đơn vị: bản đồ US_tm_selfie là 1/10 Fahrenheit (F), cần phải chuyển thành độ C theo công thức

$$C = (F - 32) * 5 / 9$$

- Sử dụng công cụ Spatial Analyst/Map algebra/raster calculator, chuyển đổi từ F sang C, lưu dưới file US_tm_self_f

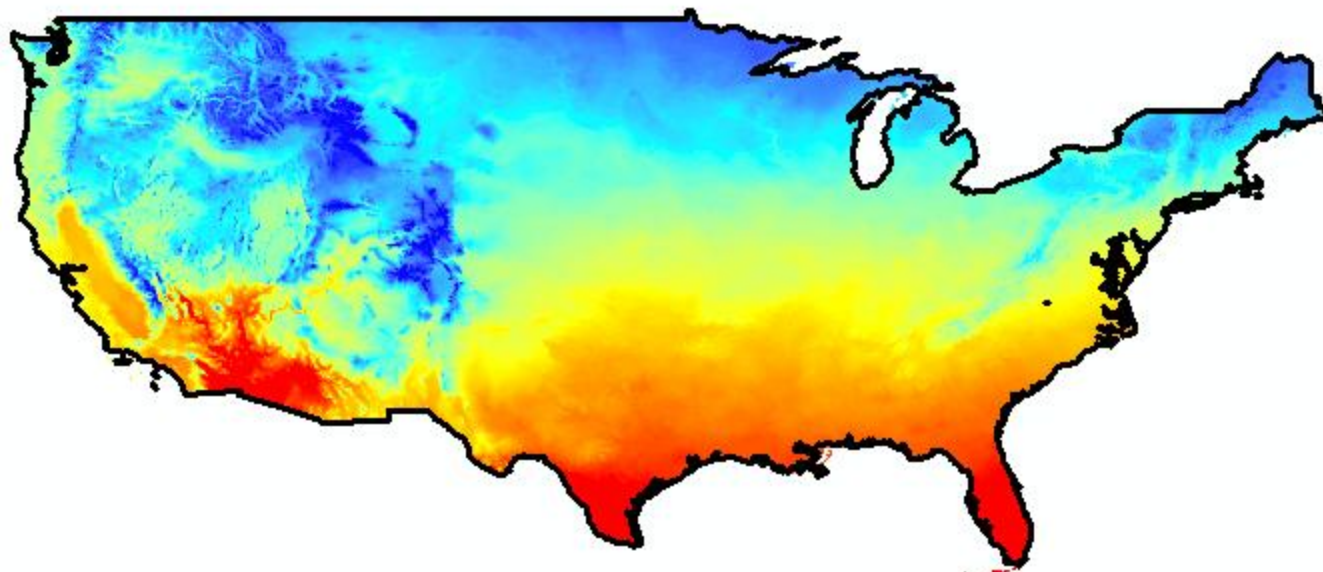
Đây là bản đồ chúng ta vừa tạo ra, \$0.

☒ us_tm_self_F
Value
High : 24.7442
Low : 0.523762



Đây là bản đồ của PRISM, \$m's.

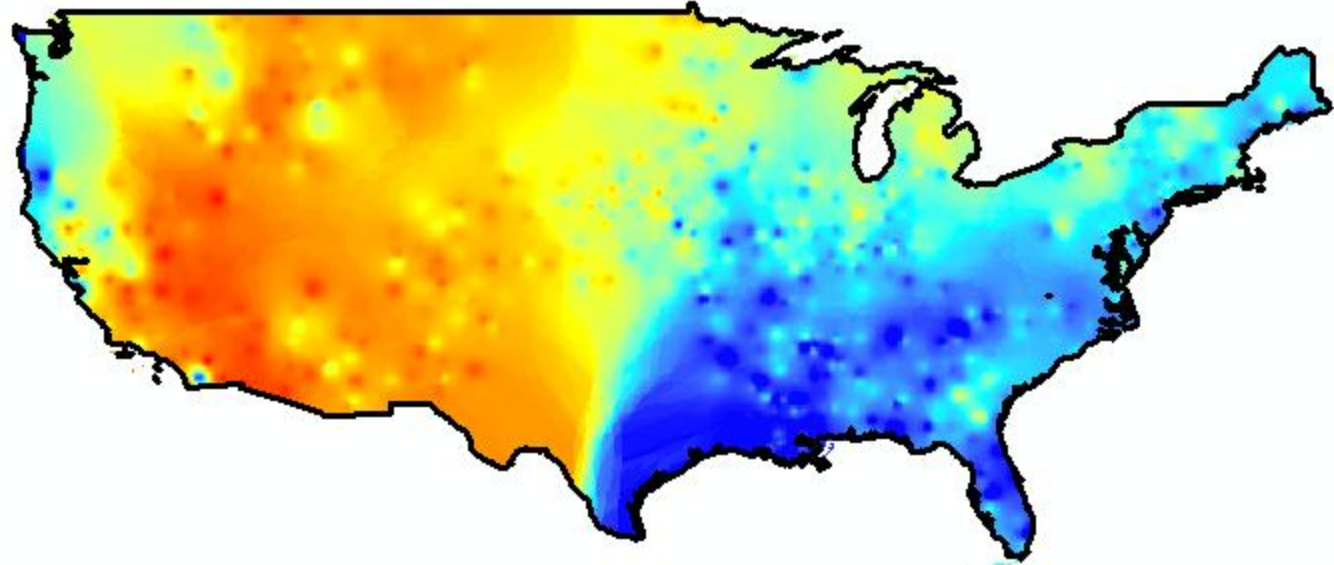
☒ PRISM_tmean_30yr_normal_
Value
High : 25.5633
Low : -5.47292



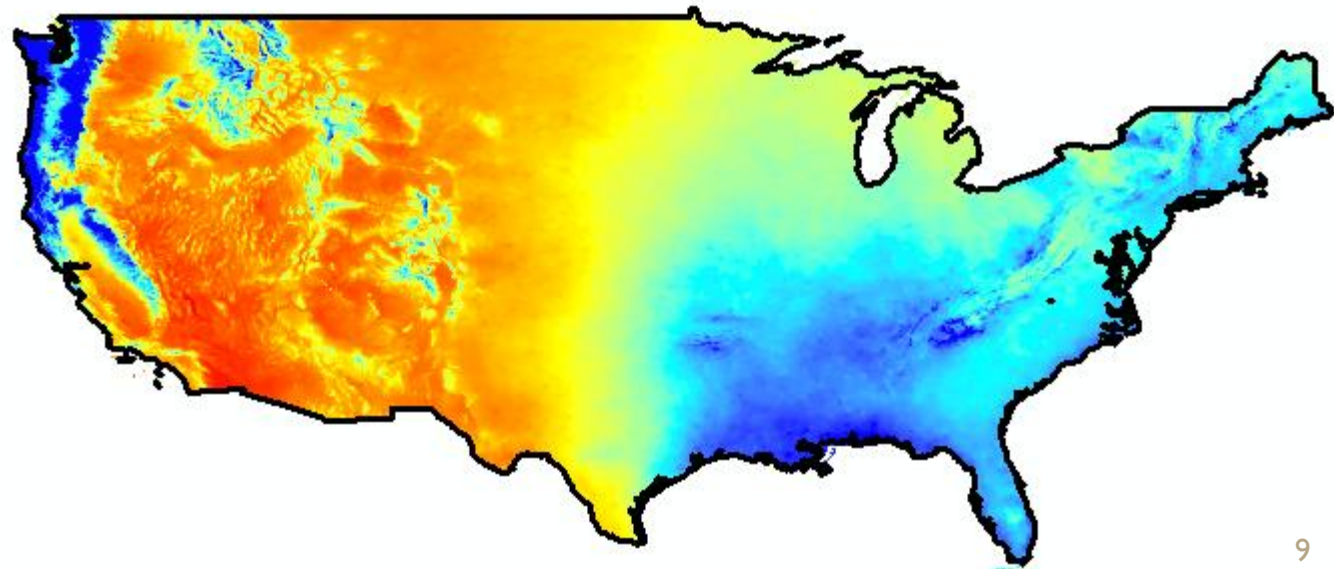
Thực hiện các bước tương tự với lượng mưa

- Tạo file US_prep_self
- Chuyển từ 1/100 của 1 inch (station data) sang 1/10 mm (PRISM)

☒ us_prep_s_mmm
Value
High : 3310.24
Low : 39.3976



☒ PRISM_ppt_30yr_normal_4kml
Value
High : 5983.99
Low : 46.75



Not bad!

Spatial econometrics – kinh tế lượng không gian

- Ước lượng giá trị đất nông nghiệp bằng điều kiện tự nhiên và nhân khẩu học.
 - Điều kiện khí hậu tốt (nhiệt độ, lượng mưa) -> năng suất cây trồng cao -> sản lượng cao -> lợi nhuận cao -> giá đất cao. Lưu ý tính phi tuyến của khí hậu.
 - Khu vực đông dân hay có thu nhập trung bình cao -> giá đất cao
 - Các đặc tính khác của đất như độ phì nhiêu, độ ẩm, khả năng chống sủi mòn...

$$\text{Log(Value)} = F(T, P, D) = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 P + \beta_4 P^2 + \gamma_1 D_1 + \dots + \gamma_k D_k + \varepsilon$$

- Giải thích ý nghĩa của các hệ số beta, gamma?

Spatial econometrics

- Vấn đề: các quan sát gần nhau ảnh hưởng lẫn nhau (spatial spill-over) -> vi phạm giả định của CLRM -> OLS estimator?
- Những khu đất có giá trị cao thường nằm gần nhau, nếu không xử lý vấn đề tương quan không gian sẽ dẫn đến ước lượng sai.
- Moran's I test: kiểm tra liệu các quan sát có tự tương quan theo chiều không gian hay không.
$$\begin{cases} H_0 : \text{no spatial autocorrelation} \\ H_1 : \text{reject } H_0 \end{cases}$$
- Kiểm định Moran's I có phân phối chuẩn

Bước 1 – tạo file dữ liệu từ các nguồn khác nhau

- Merge file farmValue (giá đất nông nghiệp trung bình tại các county) vào file counties_lower bằng mã fips
 - Add field “cntyfips” trong file counties_lower, dùng field calculator, chọn Long integer, rồi chọn FIPS. Lý do là FIPS là string (biến chuỗi), phải tạo biến cntyfips là biến số để merge với file farmValue
 - Add farmValue vào ArcGIS, kích phải chuột vào counties_lower, chọn Join and Relates/Join/chọn cntyfips ô thứ nhất, dữ liệu giá đất ô thứ 2, và mã county (fips) ô thứ 3, OK

Bước 1 – tạo file dữ liệu từ các nguồn khác nhau (cont)

- Kích phải chuột vào
counties_lower/Data/export data/ save
dữ liệu dưới tên county_value.shp. Kiểm
tra xem file này đã có dữ liệu giá đất
chưa?

Bước 2 – Nhập dữ liệu khí tượng (nhiệt độ) từ các file mới tạo (US_tm_self_f và US_prep_mmm)

- 2 hệ thống định dạng file khác nhau: giá đất là polygon (file có định hình), còn khí tượng là raster (ô điểm).
- Tạo điểm trọng tâm (centroid) đại diện cho mỗi county, sau đó gán giá trị khí tượng từ các file raster khác vào mỗi centroid
- Data management/features/feature to point. Input feature là county_value, output feature là county_value_centroid, OK

Bước 2 – nhập dữ liệu (cont)

- Kiểm tra xem có phải mỗi county đã được đại diện bằng một điểm chưa?
- Spatial Analyst/Extraction/Extract multi values to point. Input feature là county_value_centroid, input raster là US_prep_s_mmm, OK
- Spatial Analyst/Extraction/Extract multi values to point. Input feature là county_value_centroid, input raster là US_tm_self_f, OK

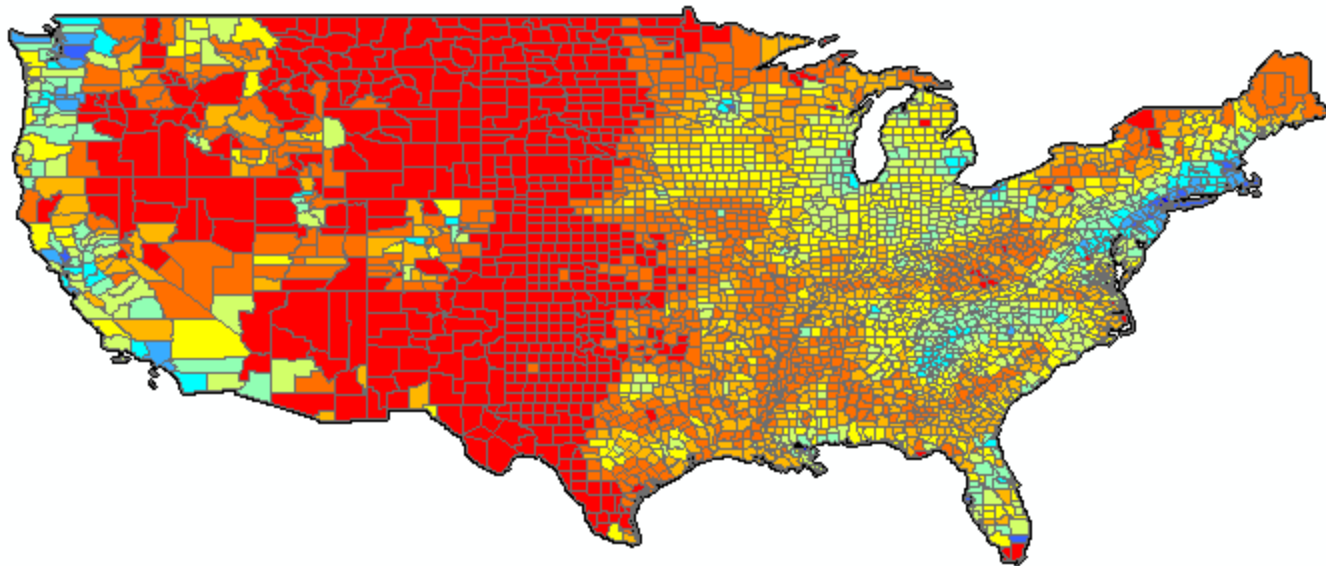
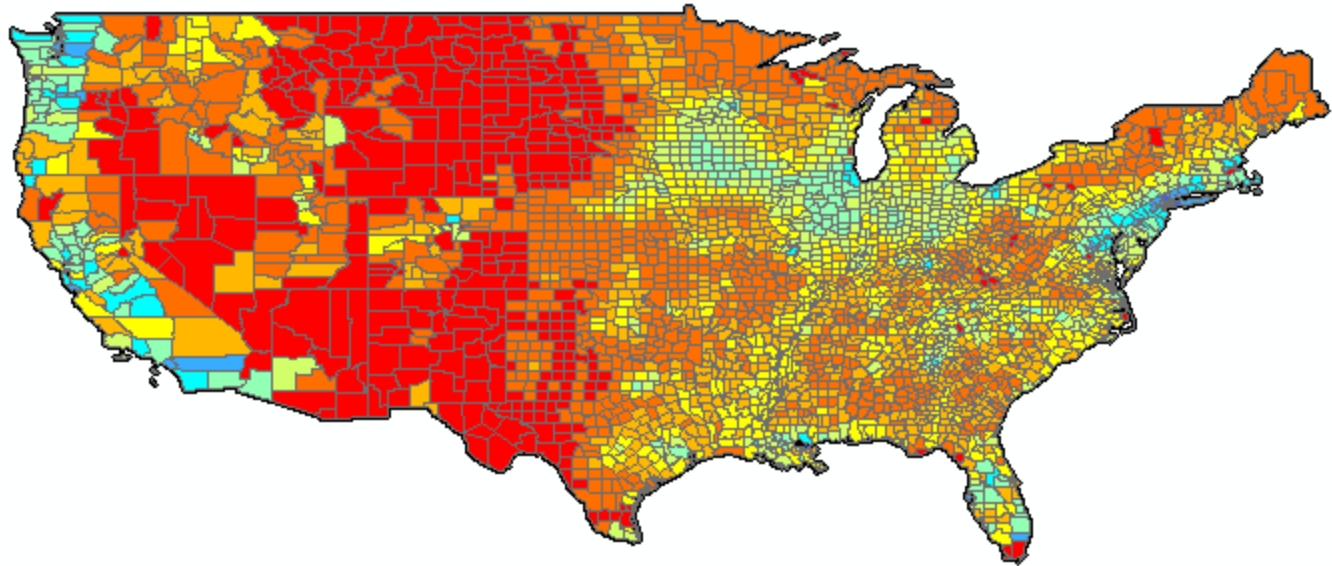
Bước 2 – nhập dữ liệu (cont)

- Kiểm tra xem file county_value_centroid đã có đủ dữ liệu chưa?
 - Lưu ý tên biến có thể bị thay đổi tự động khi merge nếu bạn đặt tên dài quá

Spatial autocorrelation tests

- Kiểm tra xem giá đất nông nghiệp có tương quan không gian hay không
- Spatial statistics/analyzing patterns/spatial autocorrelation. Input feature là county_value_centroid, input field là giá đất (farm | 982...2002). Còn lại là mặc định. OK.
- For farm | 982: Z-score=44.5, p-value=0.0000
-> reject H_0 -> có hiện tượng spatial autocorrelation. Làm tương tự cho năm 92, 97, 2002.

Bản đồ giá đất năm 1987 (trên) và 2002 (dưới)



Spatial econometrics

$$\text{Log(Value)} = F(T, P, D) = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 P + \beta_4 P^2 + \gamma_1 D_1 + \dots + \gamma_k D_k + \varepsilon$$

- Tạo biến $\log(\text{value})$, T^2 , P^2
 - Add field/lvalue82, chọn float, 6, 12, OK.
 - Lưu ý lọc những quan sát có $\text{farmvalue} > 0$ bằng cách Select by Attribute/chọn $\text{farml982} > 0$

$T2 = [\text{us_tm_self}] * [\text{us_tm_self}]$

$P2 = [\text{us_prep_s_}] * [\text{us_prep_s_}]$

Spatial econometrics (cont)

- Spatial Statistics/modeling spatial relationship/Ordinary least square:
 - Input: county_value_centroid
 - Unique ID: cntyfips
 - Output feature: farm82_ols
 - Dependent variable: lvalue82
 - Explanatory: us_tm_self, us_prep_s_, T2, P2, POP00_SQMI (mật độ dân số), vv, OK

Giải thích ý nghĩa?



Summary of OLS Results



Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF [1]
----------	-------------	----------	-------------	-------------	-----------	----------	-----------	---------



Intercept	6.229272	0.043403	143.522551	0.000000*	0.047892	130.070206	0.000000*	-----
-----------	----------	----------	------------	-----------	----------	------------	-----------	-------



POP00_SQMI	0.000641	0.000028	22.603587	0.000000*	0.000125	5.130302	0.000001*	1.017476
------------	----------	----------	-----------	-----------	----------	----------	-----------	----------



US_PREP_S_	0.000602	0.000023	25.757874	0.000000*	0.000033	18.417034	0.000000*	1.662718
------------	----------	----------	-----------	-----------	----------	-----------	-----------	----------



US_TM_SELF	-0.024470	0.003180	-7.694253	0.000000*	0.004140	-5.910177	0.000000*	1.809047
------------	-----------	----------	-----------	-----------	----------	-----------	-----------	----------



T2	0.020922	0.004429	4.724283	0.000004*	0.004669	4.481427	0.000010*	1.099210
----	----------	----------	----------	-----------	----------	----------	-----------	----------



P2	-0.055677	0.005468	-10.182146	0.000000*	0.006573	-8.470243	0.000000*	1.049072
----	-----------	----------	------------	-----------	----------	-----------	-----------	----------



OLS Diagnostics



Number of Observations:	3018	Number of Variables:	6
-------------------------	------	----------------------	---



Degrees of Freedom:	3012	Akaike's Information Criterion (AIC) [2]:	5032.485932
---------------------	------	---	-------------



Multiple R-Squared [2]:	0.339735	Adjusted R-Squared [2]:	0.338639
-------------------------	----------	-------------------------	----------



Joint F-Statistic [3]:	309.961001	Prob(> F), (5,3012) degrees of freedom:	0.000000*
------------------------	------------	---	-----------



Joint Wald Statistic [4]:	667.715559	Prob(> chi-squared), (5) degrees of freedom:	0.000000*
---------------------------	------------	--	-----------



Koenker (BP) Statistic [5]:	234.489655	Prob(> chi-squared), (5) degrees of freedom:	0.000000*
-----------------------------	------------	--	-----------

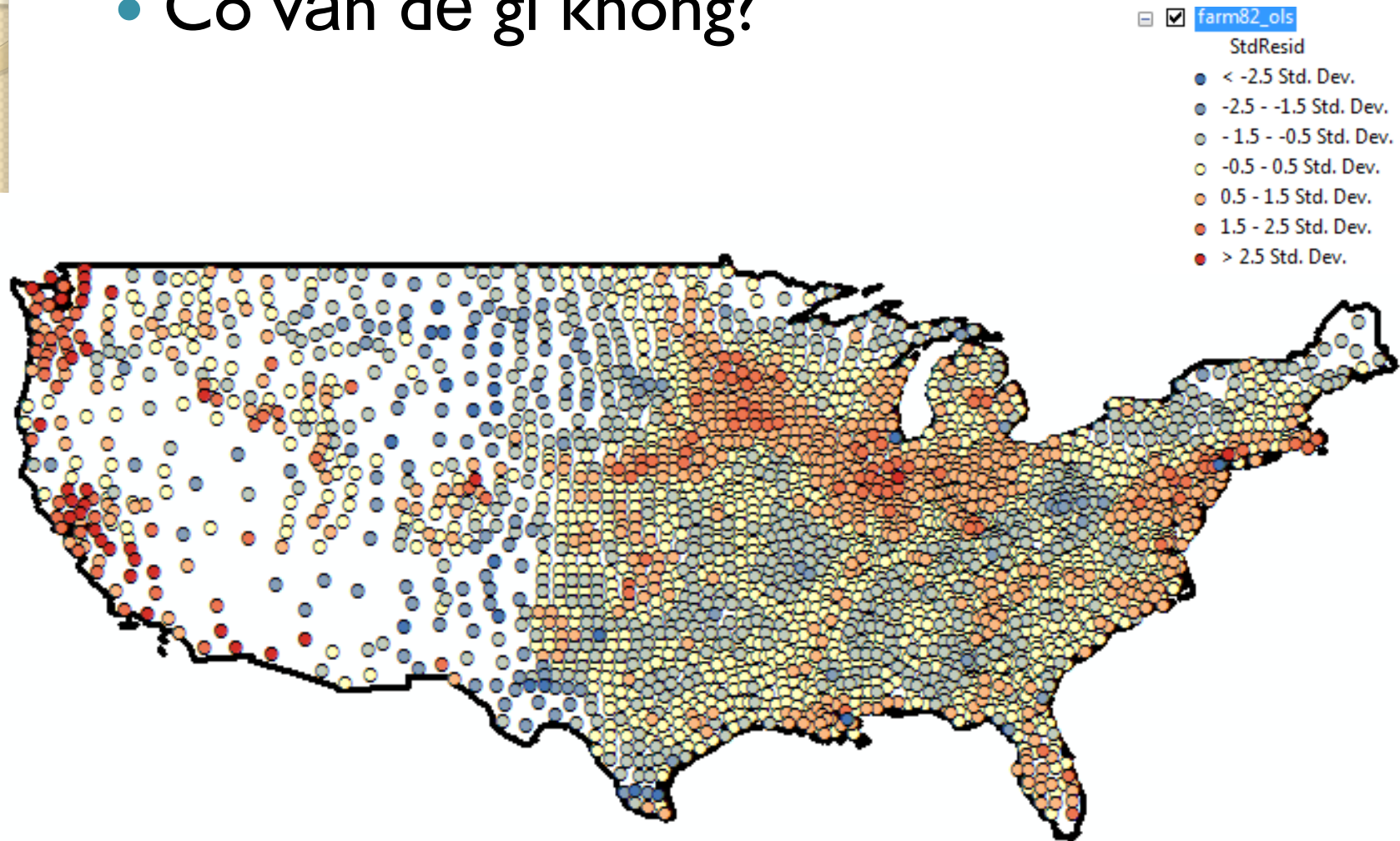


Jarque-Bera Statistic [6]:	122839.853005	Prob(> chi-squared), (2) degrees of freedom:	0.000000*
----------------------------	---------------	--	-----------



Phân phối không gian của sai số

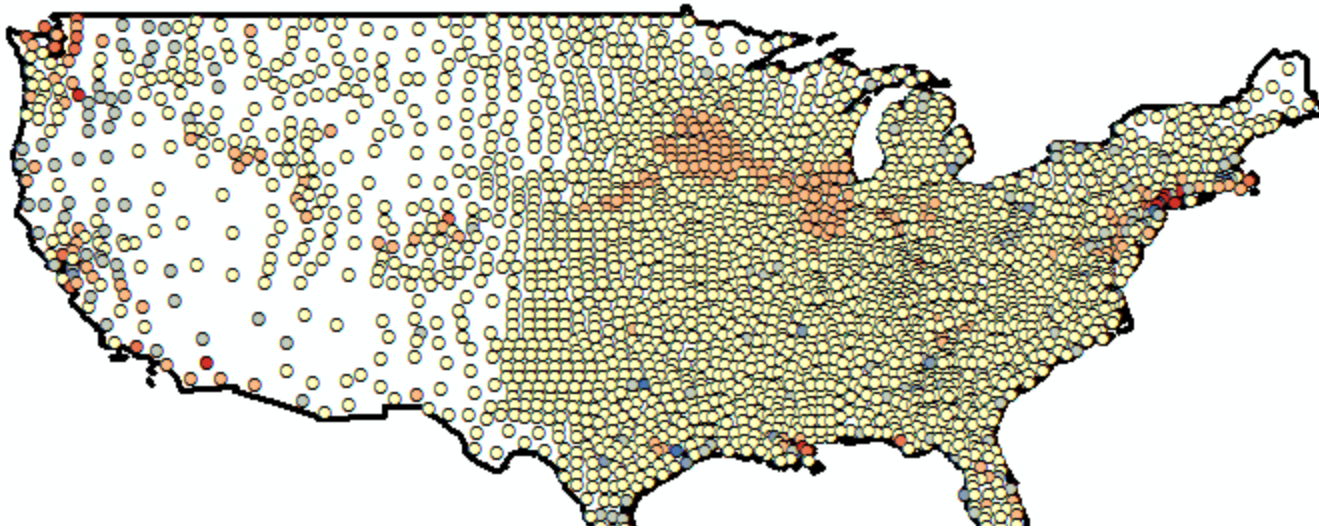
- Có vấn đề gì không?



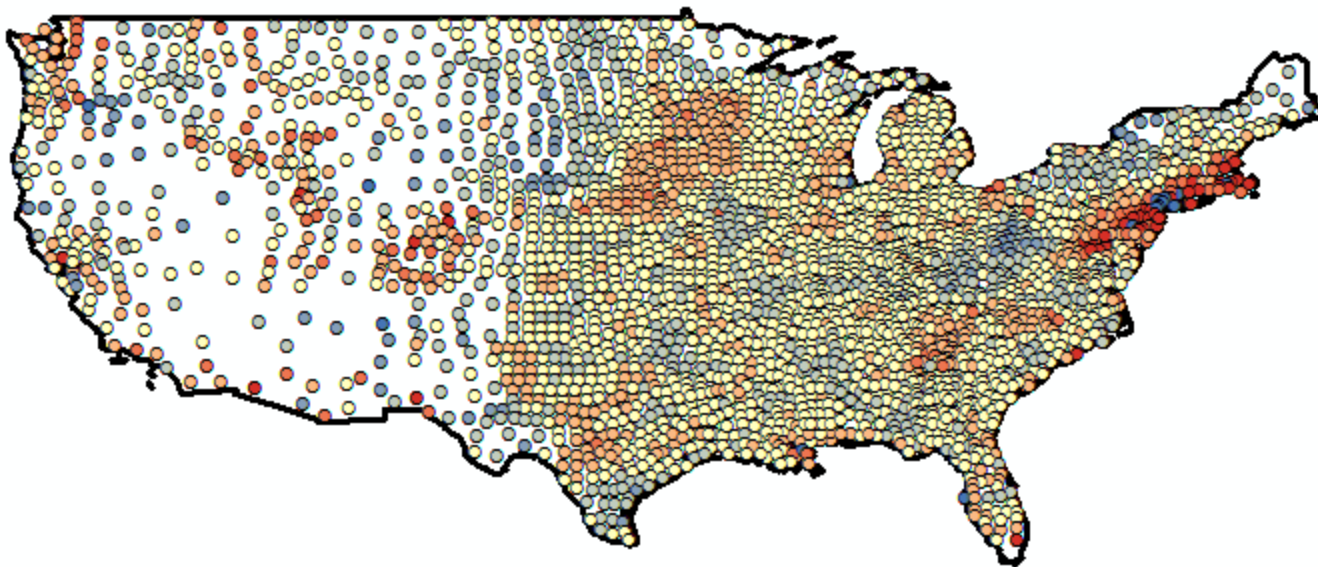
Geographically weighted regression (GWR)

- Spatial statistics/modeling spatial relation/GWR/..., save dưới file farm82_gwr
- Bandwidth parameter/distance, 5, OK
- Thay đổi cấu trúc hàm, tùy biến các lựa chọn để kiểm nghiệm GWR

Phân phối không gian của sai số sau khi đã điều chỉnh tương quan không gian



1982



2002

Tác động biên của thay đổi khí hậu đến giá đất

$$\text{Log(Value)} = F(T, P, D) = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 P + \beta_4 P^2 + \gamma_1 D_1 + \dots + \gamma_k D_k + \varepsilon$$

$$\frac{d\text{Log}(\text{value})}{dT} = \beta_1 + 2\beta_2 T$$

- Tác động biên tùy thuộc vào nhiệt độ hay lượng mưa hiện nay (baseline), có thể (-) hoặc (+)
- Lấy nhiệt độ và lượng mưa trung bình -> tác động biên trung bình

Dự báo tác động của BĐKH đến giá đất nông nghiệp

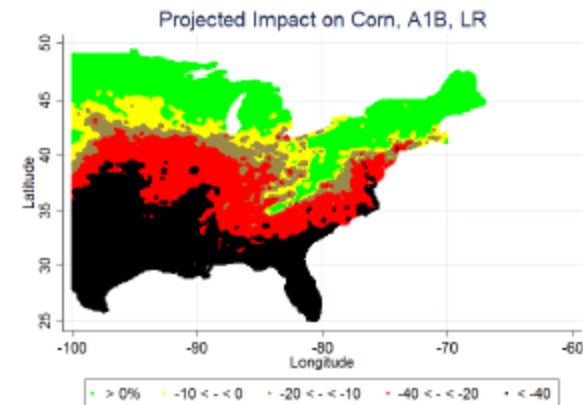
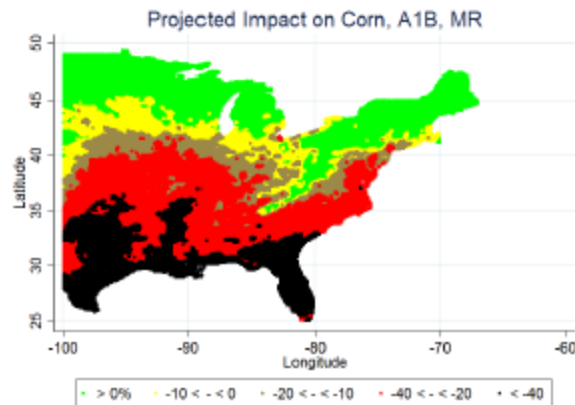
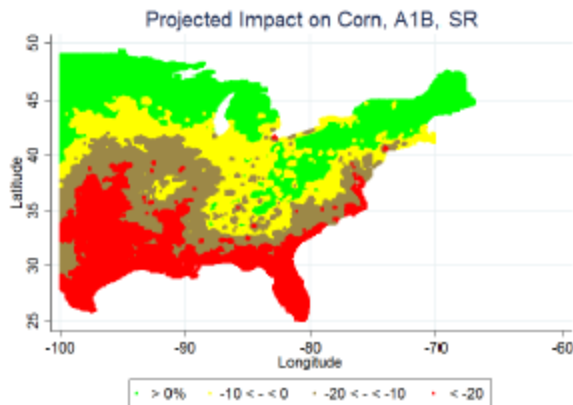
- Giả định 3 kịch bản nhiệt độ tăng đều 1, 2, 3 độ C vào cuối thế kỷ 21 ở tất cả mọi nơi
- Giả định lượng mưa tăng 10, 20, 30% so với giá trị trung bình hiện tại ở tất cả mọi nơi
- Tác động đến giá đất nông nghiệp như thế nào? Thể hiện như thế nào trên bản đồ?
 - Back-of-the-envelope method: Lấy tác động biên nhân với mức độ thay đổi

$$\frac{d\text{Log}(\text{value})}{dT} * \Delta T$$

- Cộng gộp tác động của mưa và nhiệt độ

Tác động của BĐKH lên năng suất cây nông nghiệp tại Hoa Kỳ

Projected Impacts - A1B. Corn



Wheat

