Predicting Drive Time to Care Sites using Great Circle Distance

Christopher L Richmond, Chase W Gatlin,
and Mark F Guagliardo, PhD

Planning Systems Support Group
Office of the Assistant Deputy Undersecretary
for Health for Policy and Planning
Veterans Health Administration

ESRI Health GIS Conference
October 18-19, 2010, Denver, CO

Overview

- Background
  - Travel time vs. straight line distance
  - VHA special case
- Goal
- Methods
- Results
  - Best national regression model
  - Maps of regression residuals

Background

- Network-based travel costs are best
- Not always available or attainable
- Various non-network estimates used
  - Cartesian distances
    - Euclidean distance (straight line)
    - Manhattan distance
  - Great circle distances
    - Haversine
    - Vincenty

Background (cont.)

- Time – direct distance correlation gets mixed reviews in literature

Goal

Develop the best possible regression model to predict drive time to VHA primary care stations from any residential point in the U.S. using x,y coordinates and easily obtainable residential area statistics.

Goal (cont.)

Rationale:

- Need quicker turnaround for travel time analysis in the VHA
- Smaller ad hoc analyses by other VA offices

Methods

Data

- Origination points
  - 500,000 geocoded VHA enrollee addresses (x,y)
- Destination points
  - 932 VHA primary care sites (x,y)
- County and census tract attributes bundled by ArcGIS 9.3

Methods

Travel time estimation

- Software Used: ESRI ArcMap Version 9.3
- Street Reference Data: StreetMap Premium North America Tele Atlas
- Centrus Group 1 Geocoder
- Drive time and drive distance calculated to the nearest VA Health Care Facility and included in each enrollee record

Methods

Great circle distance calculation

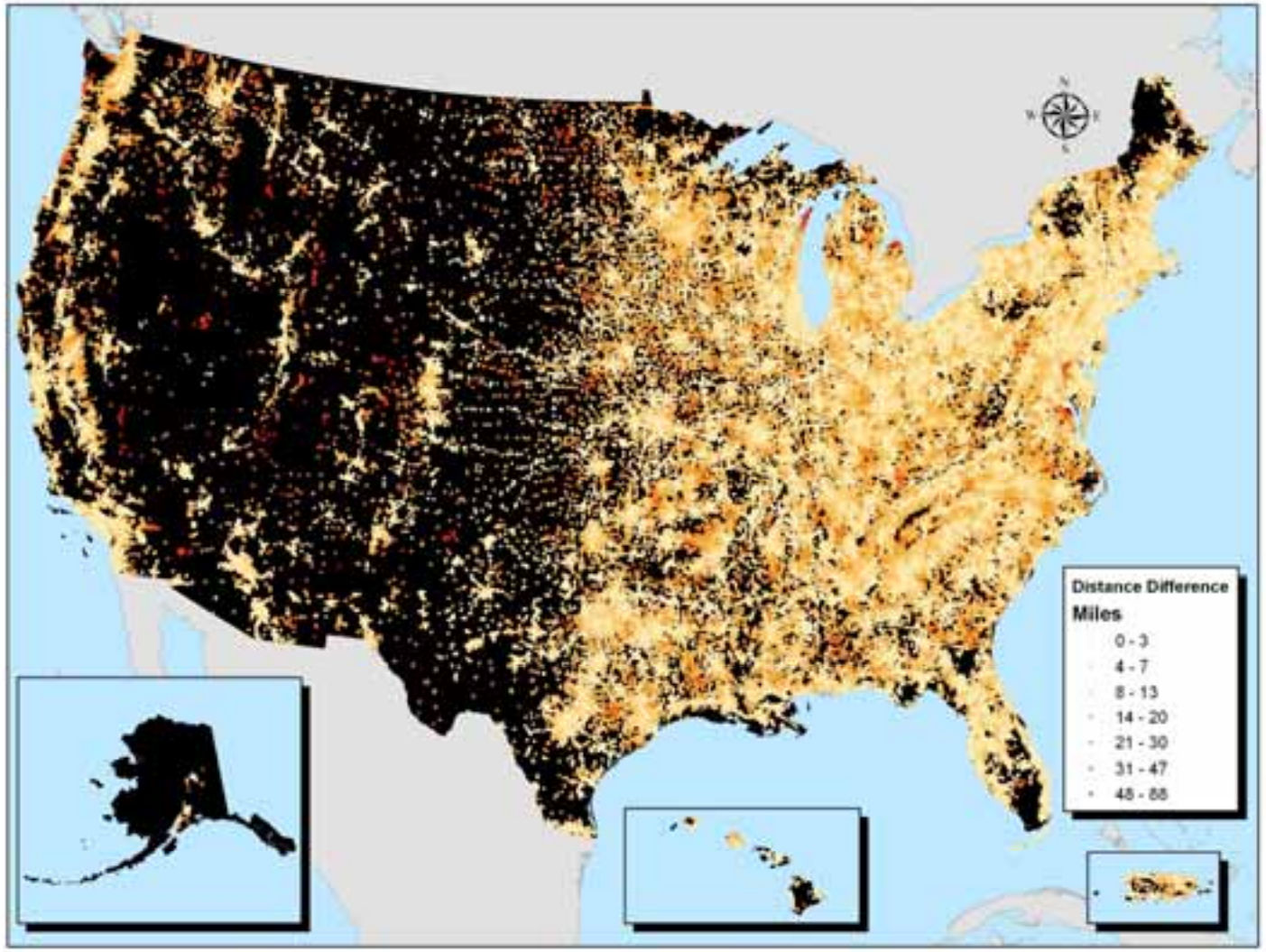
- Used SAS variable array processing to discover nearest of 932 sites
- distance = geodist(person_lat, person_lon, site_latit(i), site_lon(i), 'DM')

Methods

Great circle distance calculation

- 9.3% of the 500,000 enrollees had a different “nearest site” by vincenty distance versus drive time estimation
- Drive *distance* - vincenty distance
  - mean = 4.0 miles
  - s.d. = 4.8 miles
  - min = 0.0 miles
  - max = 88 miles

Distance Type Differences Mapped



Methods

Predictor variables

- URH - a VHA indicator of rurality
- Census tract population
- Census tract population density
- Census tract percent minority
- Census tract percent owning home
- County crop acreage

Methods

Building the regression model

- SAS PROC GLMSELECT
  - Allows stepwise selection of predictor variables
  - Makes use of categorical predictors easy
  - Allows split-sample development of models

Results

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	198082620	33013770	831211	<.0001
Error	498512	19799731	39.71766		
Corrected Tot	498518	217882351			

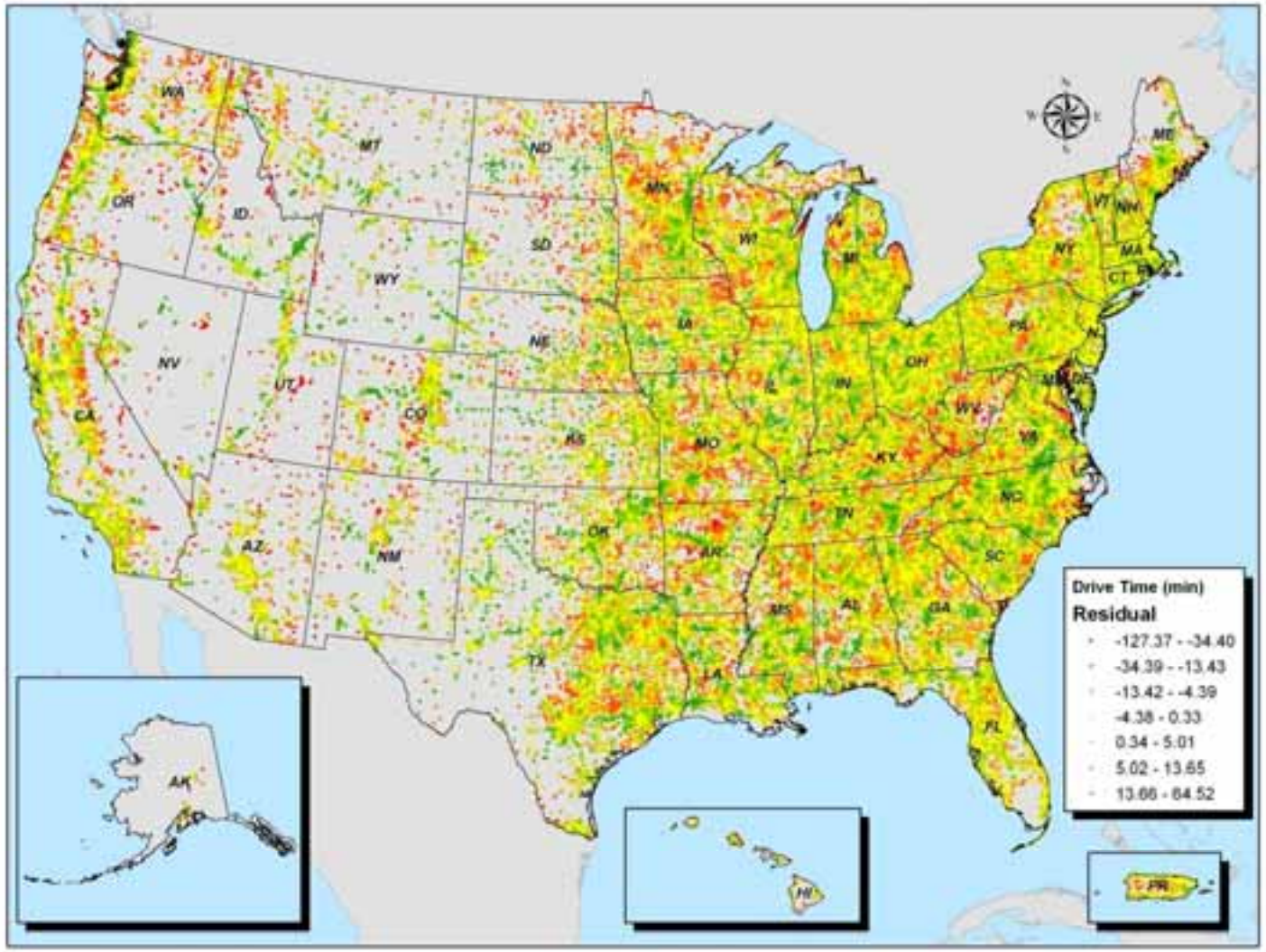
Results

Root MSE	6.30220	R-Square	0.9091
Dependent Mean	22.55120	Adj R-Sq	0.9091
Coeff Var	27.94616		

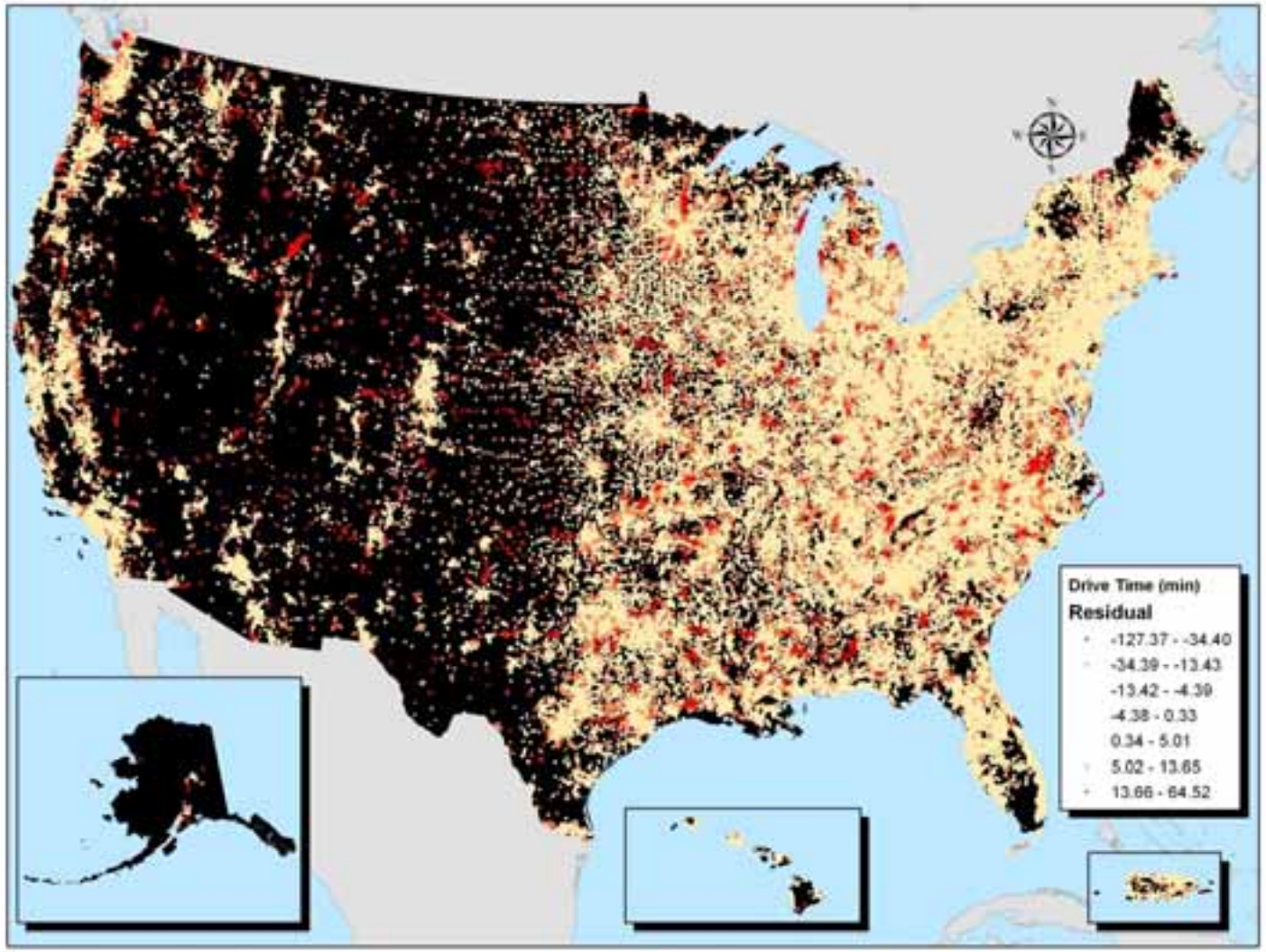
Results

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.64645	0.03787	17.07	<.0001
vincentydist	1	1.55105	0.00090043	1722.56	<.0001
URH_R	1	2.54448	0.02242	113.51	<.0001
URH_H	1	10.42304	0.08092	128.80	<.0001
PercentMin	1	-0.00385	0.00039966	-9.64	<.0001
percentowned	1	0.02777	0.00044578	62.30	<.0001
CROP_ACR97	1	-0.00000355	6.488861E-8	-54.74	<.0001

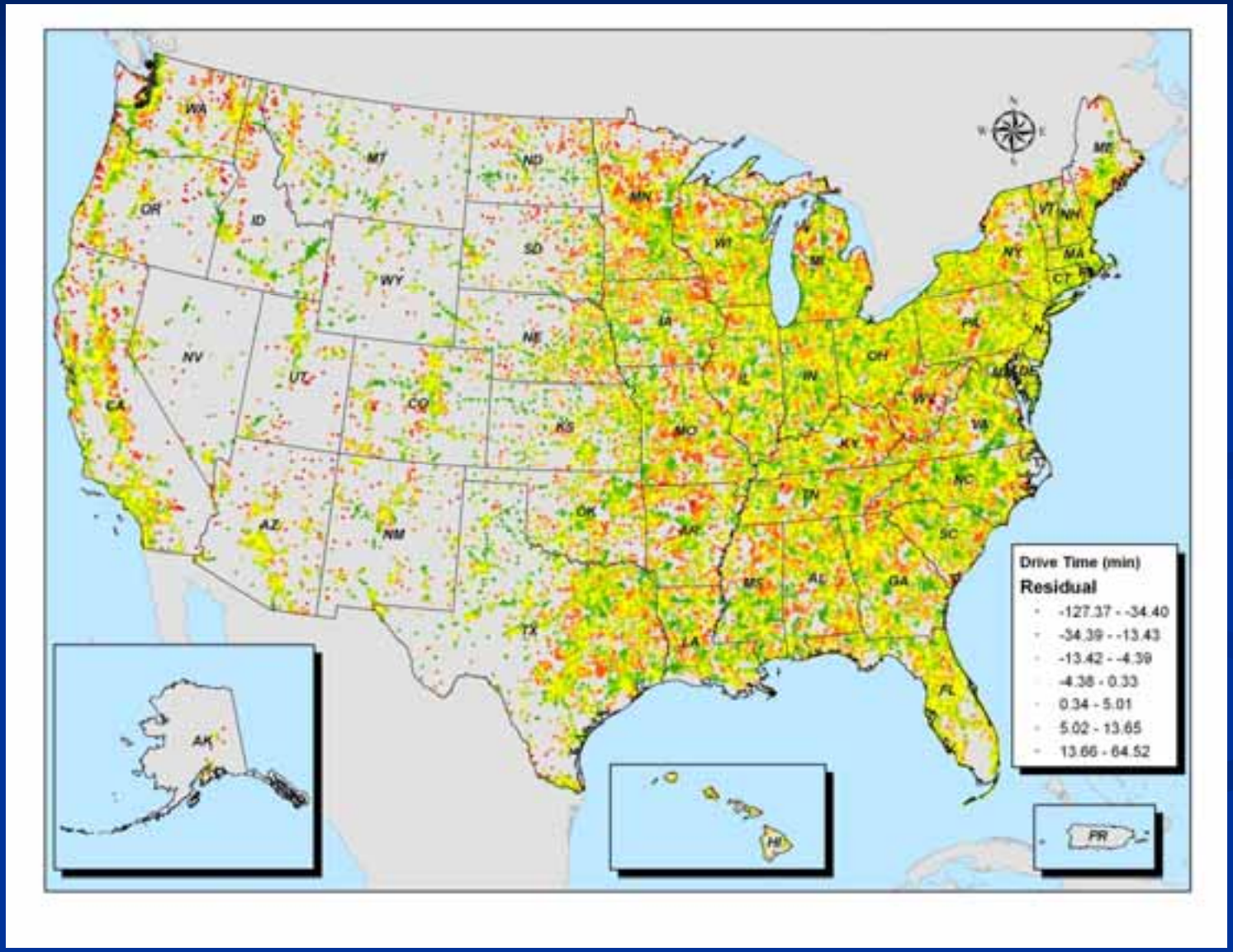
Residuals Mapped



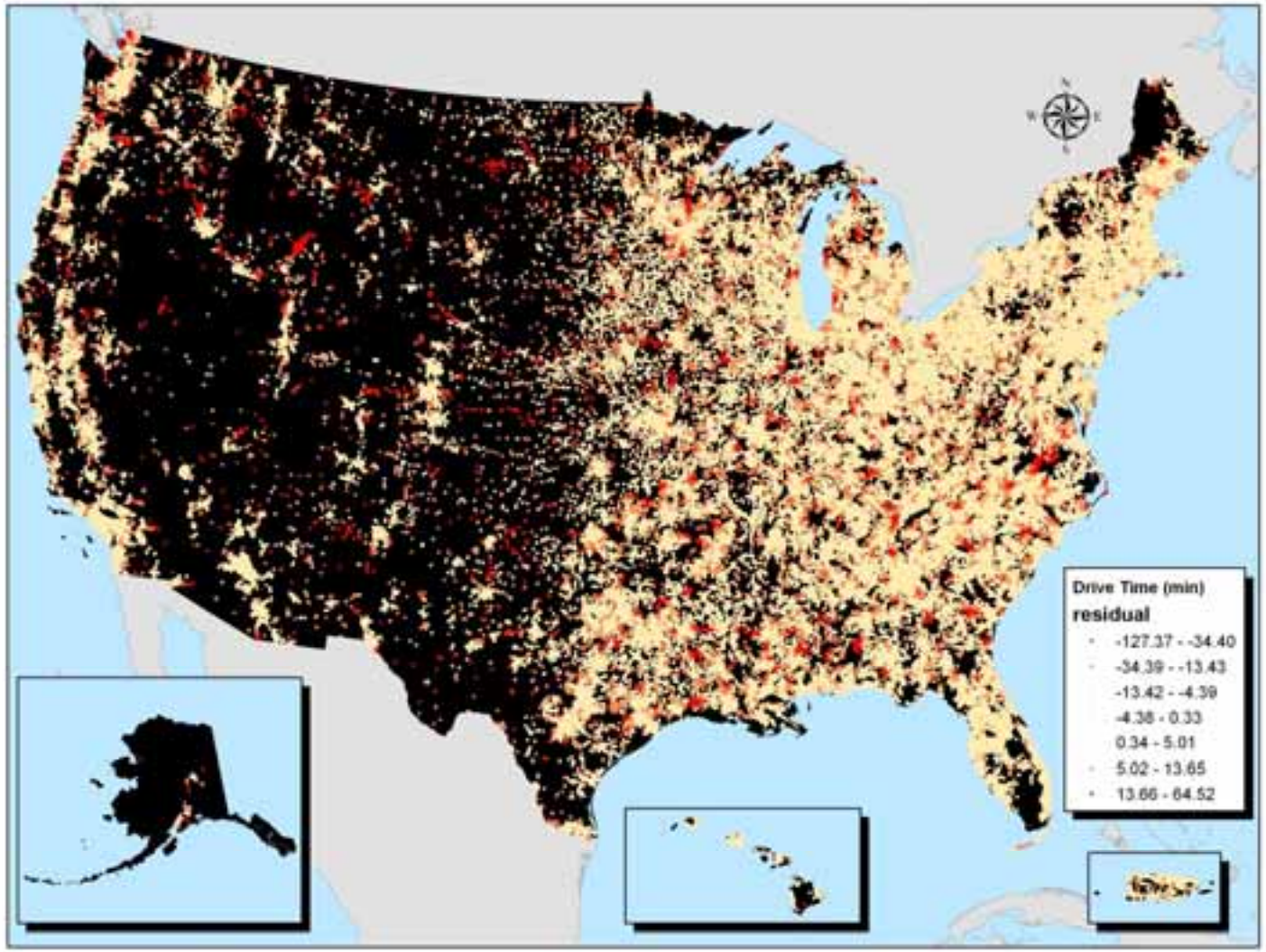
Residuals Mapped



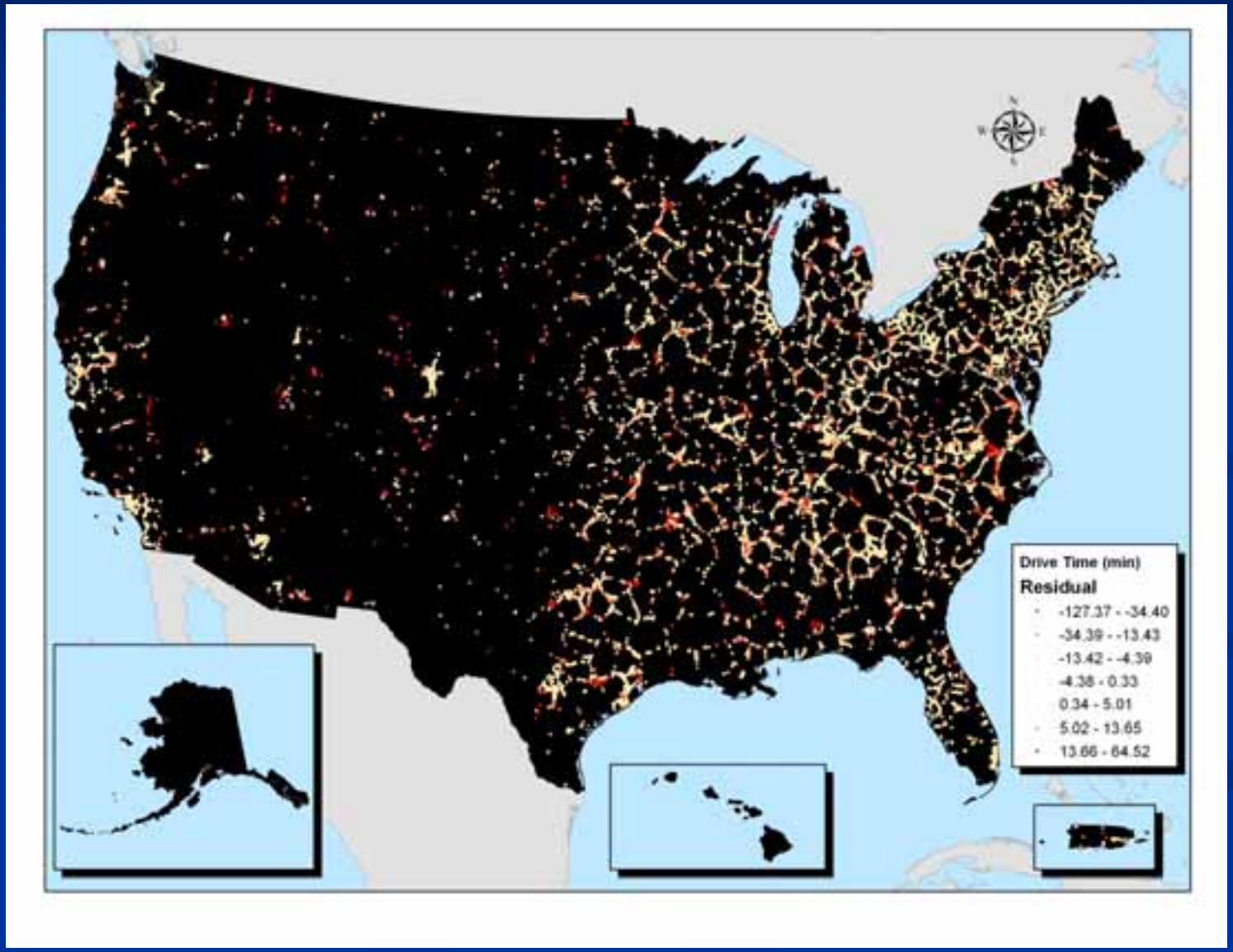
Residuals Mapped of Same Site Matches



Residuals Mapped of Same Site Matches



Residuals Mapped of Unmatched Site Assignments



Geographically Weighted Regression

- Using tool in ArcGIS 9.3
- Same independent variables and predictors
- Fixed kernel search distance of 100 miles
- Default bandwidth method (Akaike information criterion)

Geographically Weighted Regression

- R-square = 0.9377
- Regression parameters vary spatially, so not practical for our purposes

Next Steps

- Try some interaction terms
- Seek additional predictor variables
- Experiment with other GWR options

Limitations

- VHA regression formula may not be appropriate for other applications

Thank you.