Income Indicators based on Electricity Consumption:

A Geostatistical Approach

Eduardo Francisco eduardo.francisco@aes.com

Francisco Aranha francisco.aranha@fgv.br

Felipe Zambaldi zambaldi@yahoo.com

Rafael Goldszmidt rafael.goldszmidt@fgv.br

Peter Whigham PWhigham@infoscience.otago.ac.nz

Antoni Moore AMoore@infoscience.otago.ac.nz

2008 ESRI International User Conference

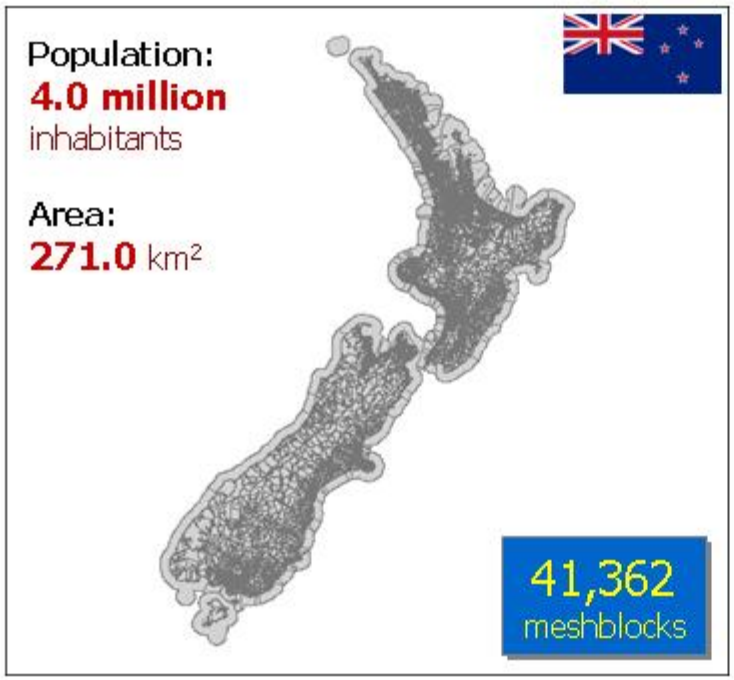
San Diego, CA, USA - August 5th 2008

Income and Electricity Consumption

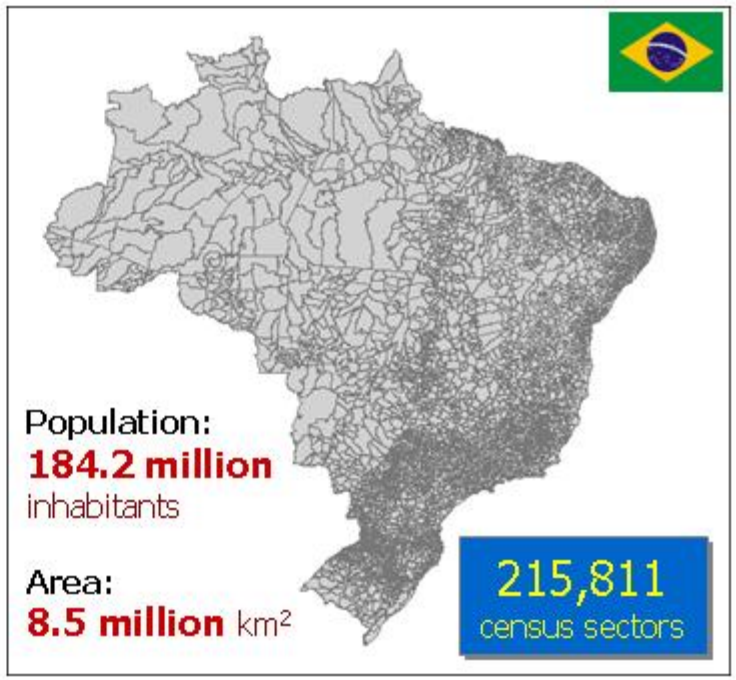
- **Income**
  - **Indicator usually adopted in studies of Poverty, Living Conditions, Finance and Marketing**
    - Estimates purchasing power of urban people and families
- **Difficulty in the collection of this information:**

1 Wide-ranging and high coverage (very expensive)

- Depends on Demographic Census or large surveys to be representative of census tracts, census sectors or meshblocks (areas with 100 to 400 households)



Census cycle: **Quinquennial** / Last: **2006**



Census cycle: **Decennial** / Last: **2000**

Income and Electricity Consumption

- **Income**
  - **Indicator usually adopted in studies of Poverty, Living Conditions, Finance and Marketing**
    - Estimates purchasing power of urban people and families
- **Difficulty in the collection of this information:**

1 Wide-ranging and high coverage (very expensive)

- Depends on Demographic Census or large surveys to be representative of census tracts, census sectors or meshblocks (areas with 100 to 400 households)

2 Accurate data on Income is difficult for some low income (and low educational level) population

- Altered declaration, seasonal changes, refusal etc.
- Social and Economic Classification or Purchasing Power based on indicators – e.g., ownership of goods – **Needs constant update**

Goods	Number of goods				
	0	1	2	3	4 or more
Television	0	2	3	4	5
Radio	0	1	2	3	4
Bathroom	0	2	3	4	4
Automobile	0	2	4	5	5
Domestic Employee	0	2	4	4	4
Vacuum Cleaner	0	1	1	1	1
Washing Machine	0	1	1	1	1
Videocassette and/or DVD	0	2	2	2	2
Refrigerator or Freezer	0	2	2	2	2
Freezer (independent machine)	0	1	1	1	1

+

Head of family's educational level	
Unfinished Basic Education	0
Basic Education	1
Unfinished High School	2
Finished High School	3
Undergraduate	5

=

Brazilian Criterion (1996)

Economic Class	Points
A1	30-34
A2	25-29
B1	21-24
B2	17-20
C	11-16
D	6-10
E	0-5

INTRO
METHODS
RESULTS
CONCLUSION

Income and Electricity Consumption

- **Income**
  - **Indicator usually adopted in studies of Poverty, Living Conditions, Finance and Marketing**
    - Estimates purchasing power of urban people and families
- **Difficulty in the collection of this information:**

1. **Wide-ranging and high coverage (very expensive)**
   - Depends on Demographic Census or large surveys to be representative of census tracts, census sectors or meshblocks (areas with 100 to 400 households)
2. **Accurate data on Income is difficult for some low income (and low educational level) population**
   - Altered declaration, seasonal changes, refusal etc.
   - Social and Economic Classification or Purchasing Power based on indicators – e.g., ownership of goods – **Needs constant update**

Consumption of Electric Energy can be a good indicator to better assist process of characterize customers

- **Easy to get - Monthly Collected**
- **Essential Utility, Wide-ranging and Coverage**
- **Could be published in aggregate areas** (census sectors, districts, municipalities)

Income and Electricity Consumption

OBJ: Analyze the relationship between Residential Electricity Consumption and Household Income

Create an Income indicator based on Electricity consumption

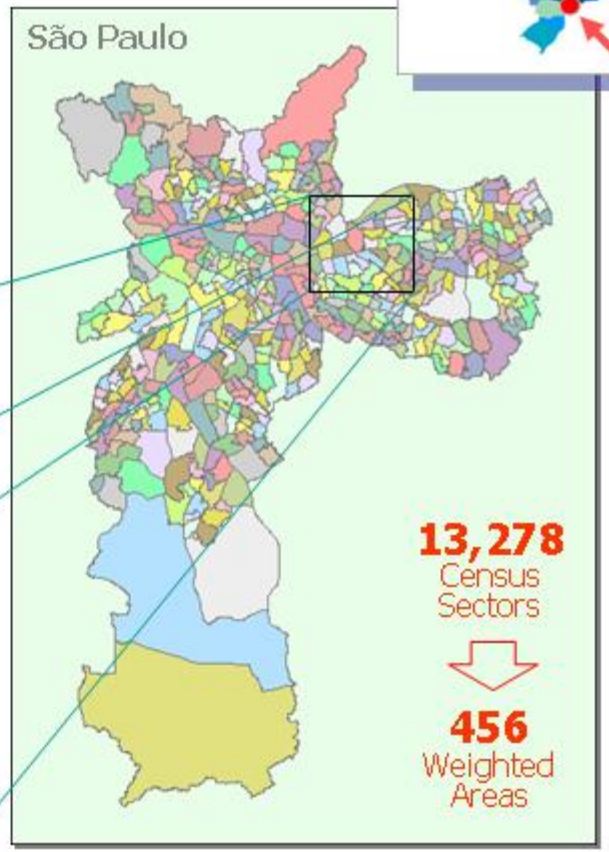
Income-predicting (regression) models: Electric Energy → + → HH Income

Object: City of São Paulo, Brazil

(more than 10 million inhabitants)

Data:

Brazilian Demographic Census 2000 +
Customers Database of AES Eletropaulo
(São Paulo's Power Distribution company) (full access)
Aggregate in weighted census areas (polygons)



Income and Electricity Consumption

OBJ: Analyze the relationship between Residential Electricity Consumption and Household Income

Create an Income indicator based on Electricity consumption

Income-predicting (regression) models: Electric Energy → + → HH Income

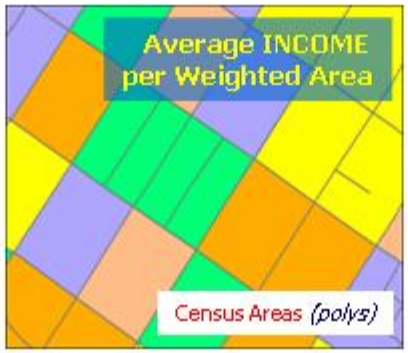
Object: City of São Paulo, Brazil

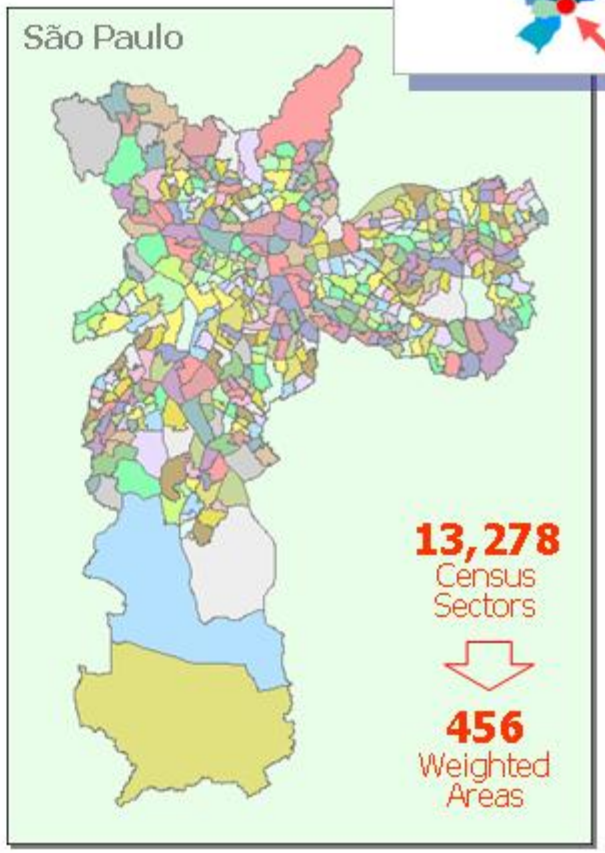
(more than 10 million inhabitants)

Data:

Brazilian Demographic Census 2000 +
Customers Database of AES Eletropaulo
(São Paulo's Power Distribution company) (full access)
Aggregate in weighted census areas (polygons)

+





Income and Electricity Consumption

OBJ: Analyze the relationship between Residential Electricity Consumption and Household Income

Create an Income indicator based on Electricity consumption

Income-predicting (regression) models: Electric Energy → (+) HH Income

Traditional Linear Regression:

$$\hat{y} = \beta_0 + \beta_1 x + \varepsilon$$

GLOBAL MODEL

SAR (*Spatial Auto-Regression*):

$$\hat{y} = \beta_0 + \beta_1 x + \rho W y + \varepsilon$$

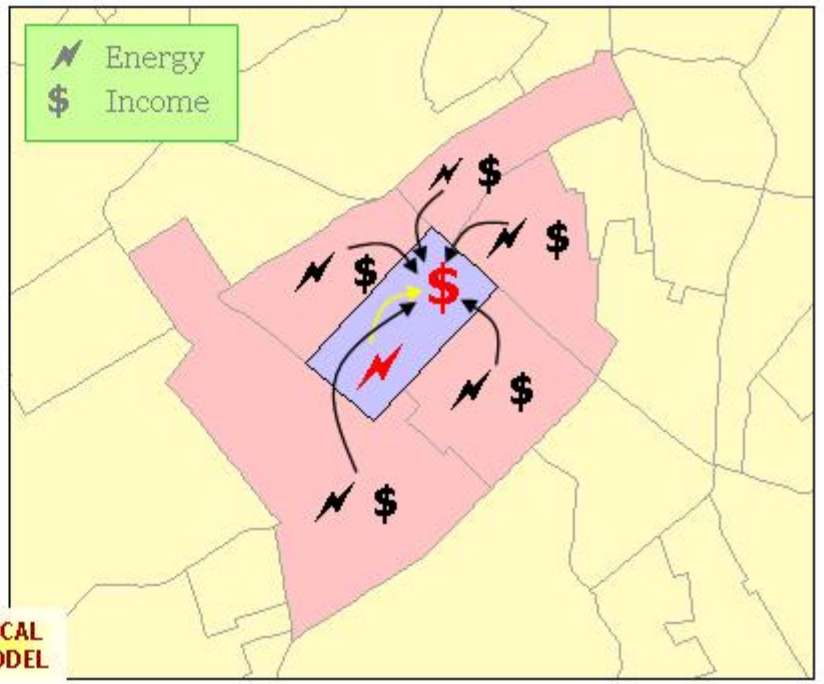
GLOBAL MODEL

Neighbourhood Matrix
(based on k nearest neighbours)

GWR (*Geographically Weighted Regression*):

$$\hat{y}_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\, x_i$$

LOCAL MODEL



Different regressions for each weighted census areas (i) considering a local sample based on k nearest neighbours

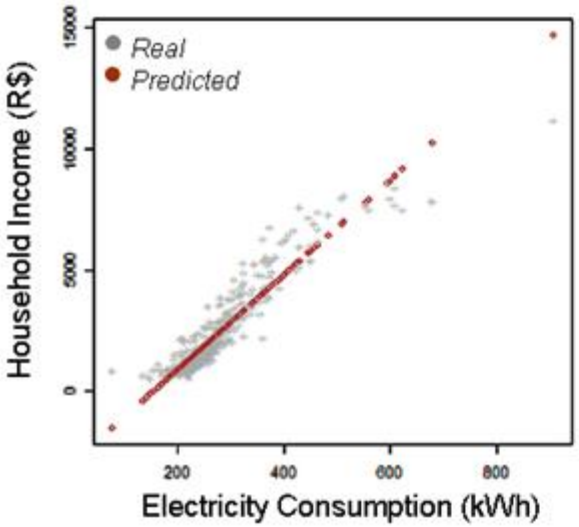
GWR: Different β parameters and different local R^2 for each i (and for each local sample size k)

Results of Predictive Models

Traditional Linear Regression

$R^2 = 86.80\%$

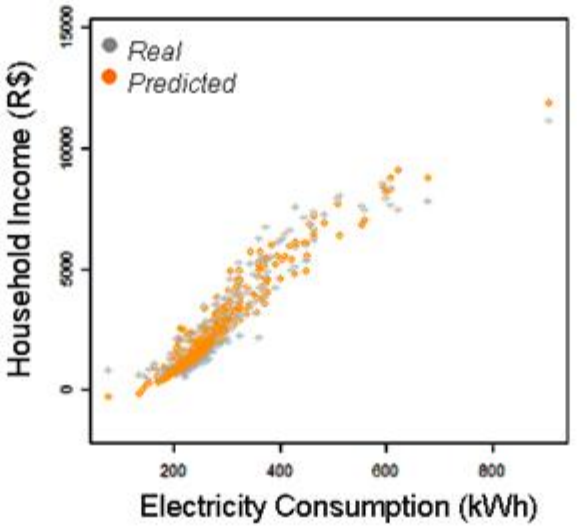
$\hat{y} = -3{,}034.71 + 19.55 \cdot x$



SAR Model (*Spatial Auto-Regression*)

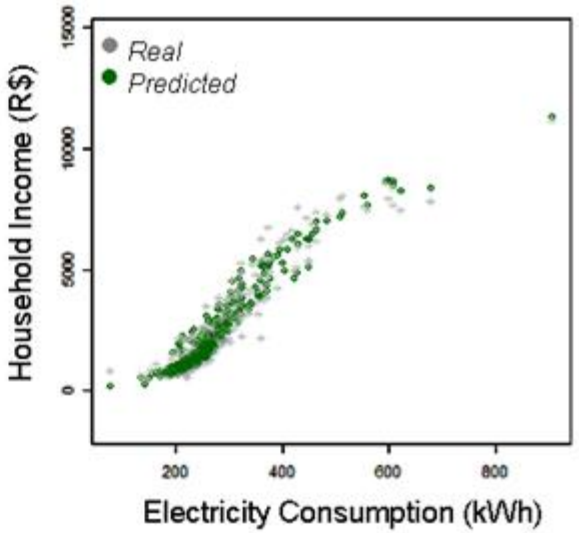
$R^2 = 94.48\%$

$\hat{y} = -2{,}303.64 + 12.73 \cdot x + 0.499\, W\, y$



GWR Model (*Geographic Weighted Regression*)

$R^2 = 96{,}80\%$



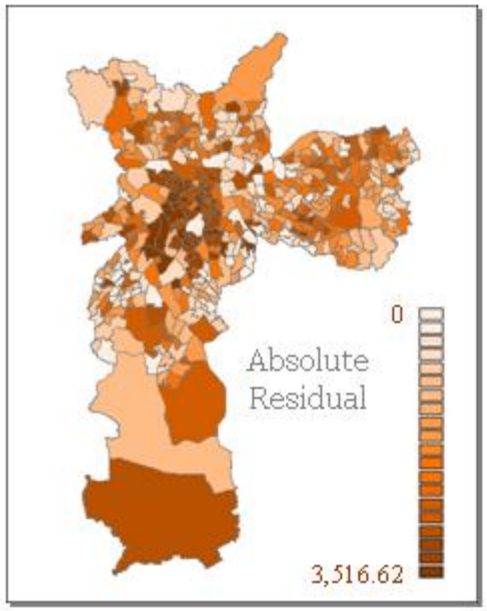
Obs: $k = 9$ in Spatial Statistics models

Results of Predictive Models

Traditional Linear Regression

$R^2 = 86.80\%$

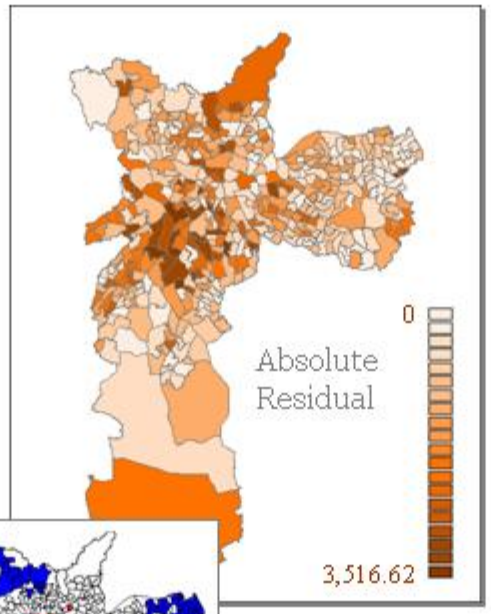
$\hat{y} = -3{,}034.71 + 19.55 \cdot x$

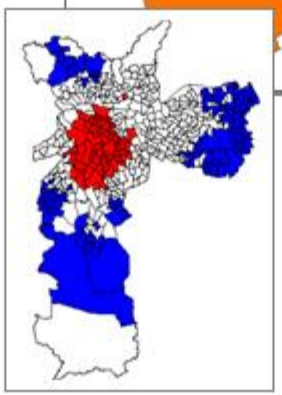


SAR Model (*Spatial Auto-Regression*)

$R^2 = 94.48\%$

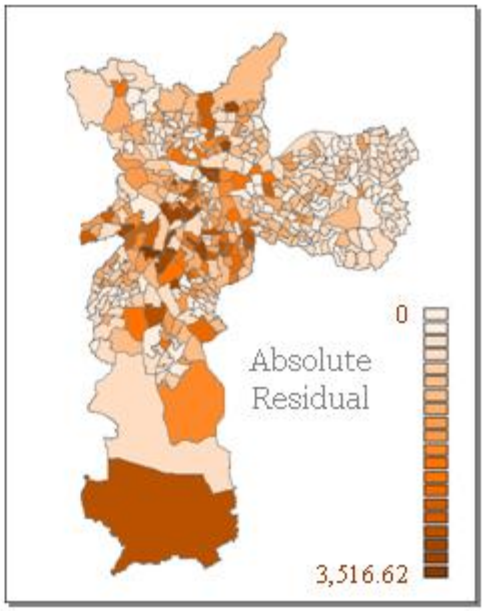
$\hat{y} = -2{,}303.64 + 12.73 \cdot x + 0.499\, W\, y$

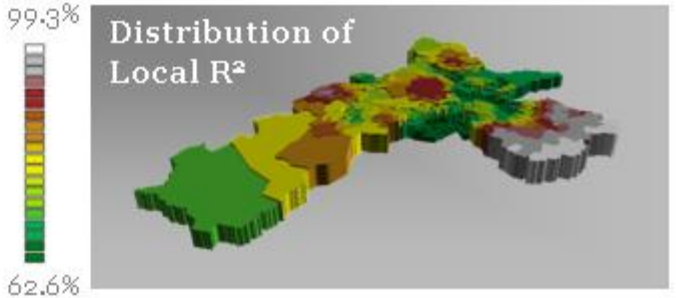




GWR Model (*Geographic Weighted Regression*)

$R^2 = 96{,}80\%$



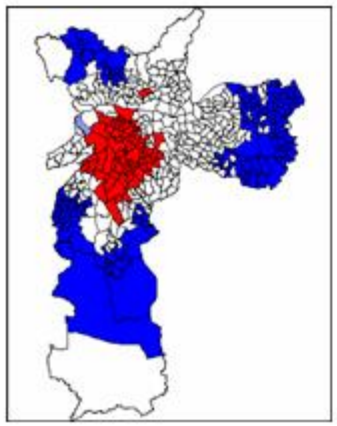


Obs: $k = 9$ **in Spatial Statistics models**

SAR: Neighbourhood Graphs

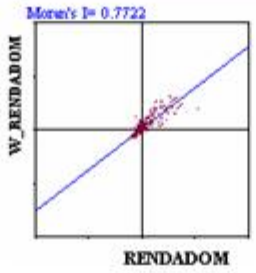
- For different neighborhood matrix, Energy Moran's I showed high values (0.77+)
- It suggests high influence of neighborhood in Household Income behavior
- **LISA maps:** Increase of income concentration in direction Suburbs-Centre. The same for Electricity consumption

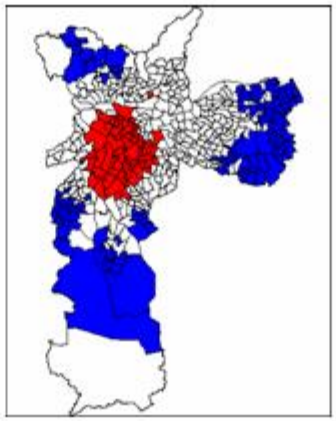
Not Significant
High-High
Low-Low
Low-High
High-Low



Neighborhood Graph:
Queen Graph

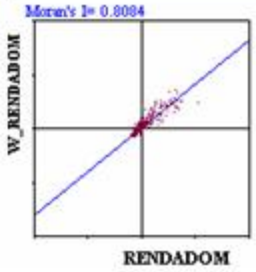
Moran's I = **0.7722**

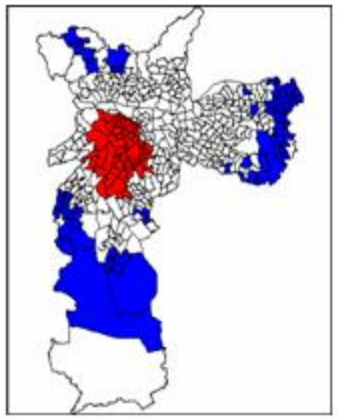




Neighborhood Graph:
Gabriel's Graph

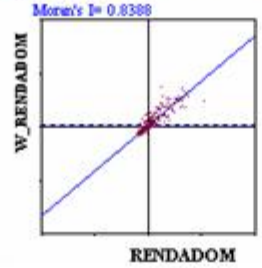
Moran's I = **0.7722**

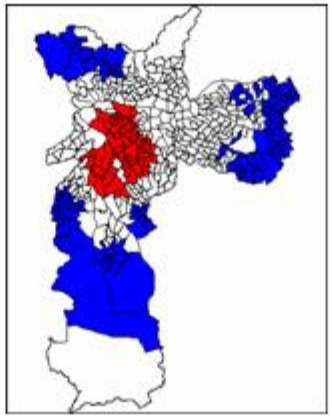


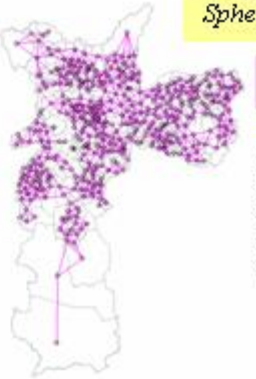


Neighborhood Graph:
Relative Neighbor Graph

Moran's I = **0.8388**

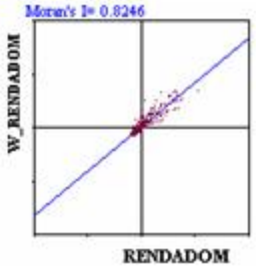






Neighborhood Graph:
Sphere of Influence Graph

Moran's I = **0.8246**

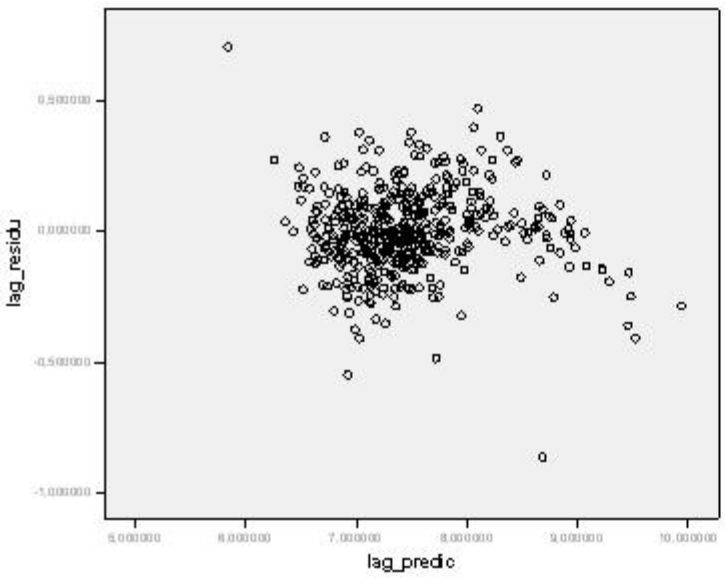


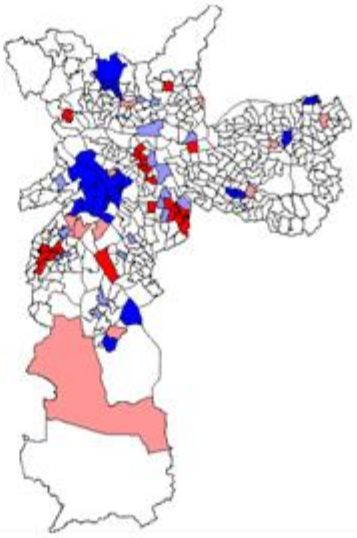
SAR: Absence of Spatial Dependence in Residuals

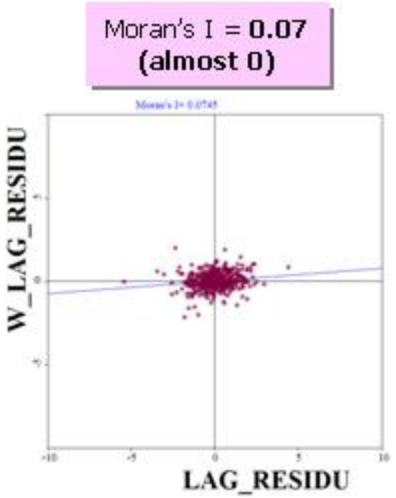
Spatial Auto-regressive Model

```
Data set            : electric energy
Spatial Weight      : areaqueen1.GAL (Queen Graph)
Dependent Variable  :     LNINCOME  Number of Observations:  456
Mean dependent var  :      7.46738  Number of Variables   :    3
S.D. dependent var  :     0.633242  Degrees of Freedom    :  453
Lag coeff.   (Rho)  :     0.607507

R-squared           :     0.94484  Log likelihood        :  171.909
Sq. Correlation     : -            Akaike info criterion :  -337.818
Sigma-square        :    0.0253932 Schwarz criterion     :  -325.451
S.E of regression   :     0.159352
```

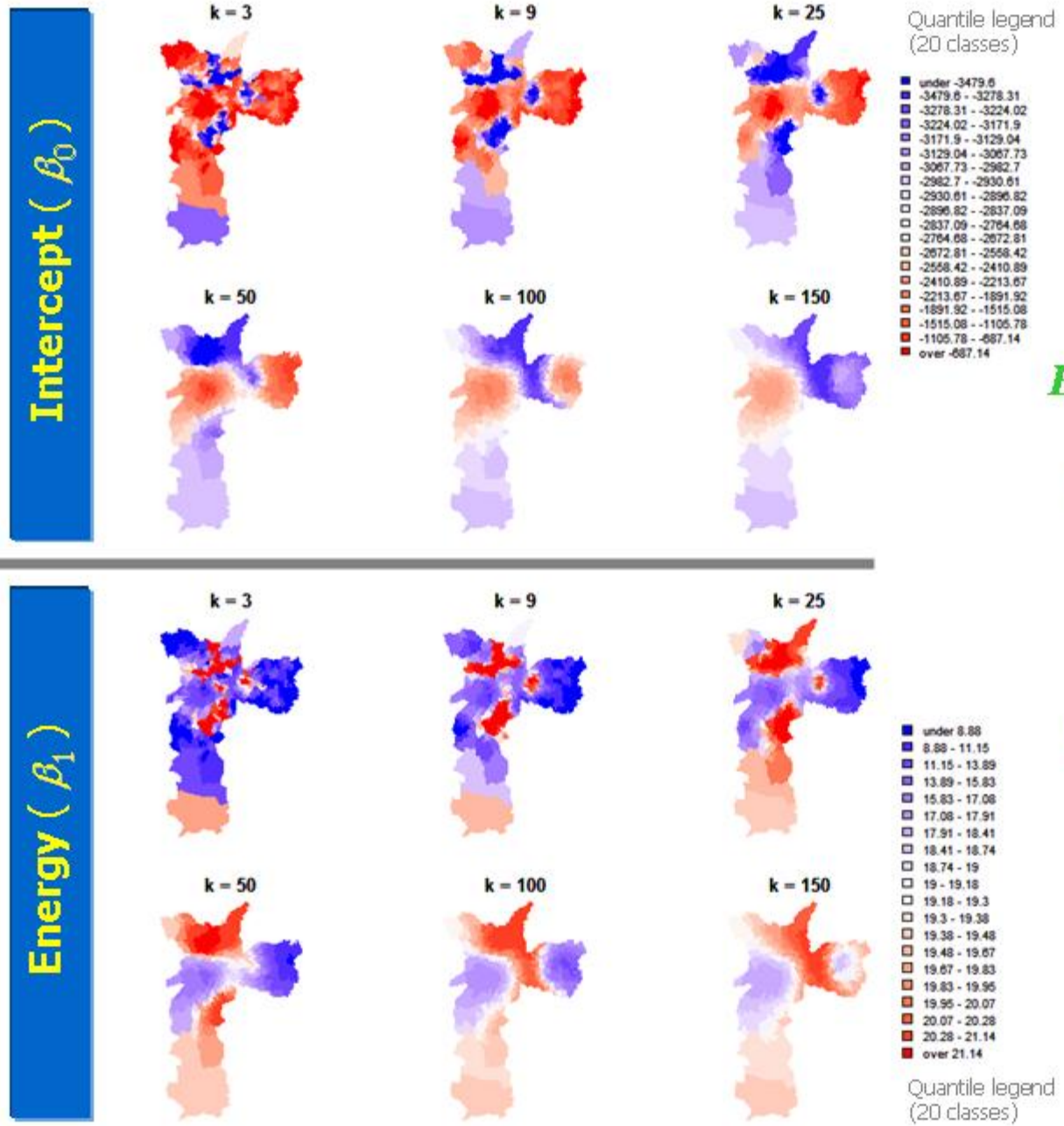






- Use of Neperian Logarithms of dependent and independent variables
- Residual error of this model assumed normal distribution pattern and homoskedasticity - **Absence of spatial dependence in residuals**

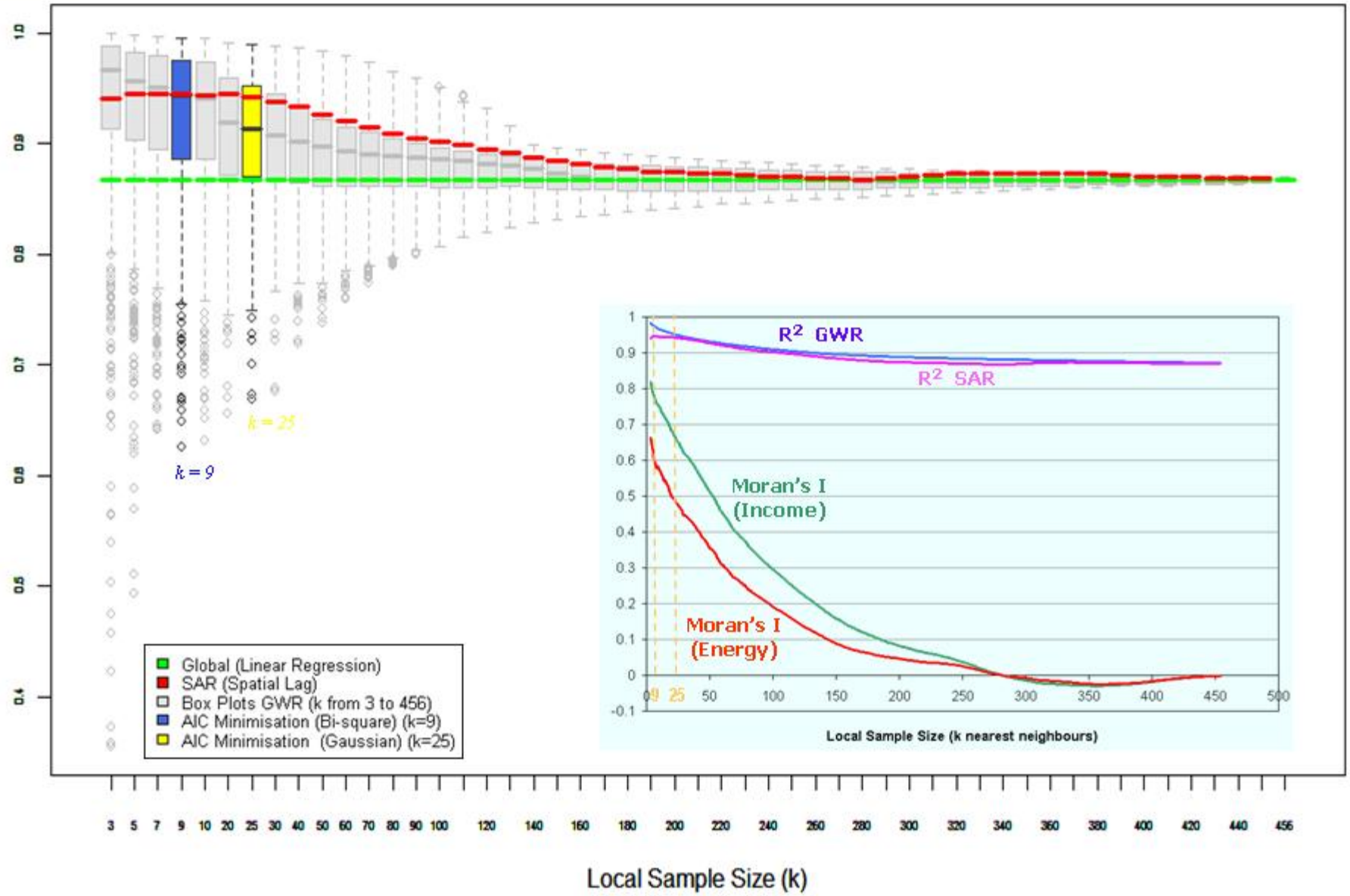
GWR: Distribution of β_0 and β_1 parameters (varying k)



In central areas, ***Energy parameter*** *has a lower contribution than in suburbs.*

Vice-versa for Intercept (initial ***Income*** *considered in prediction).*

Distribution of global R^2 (GWR and SAR models)

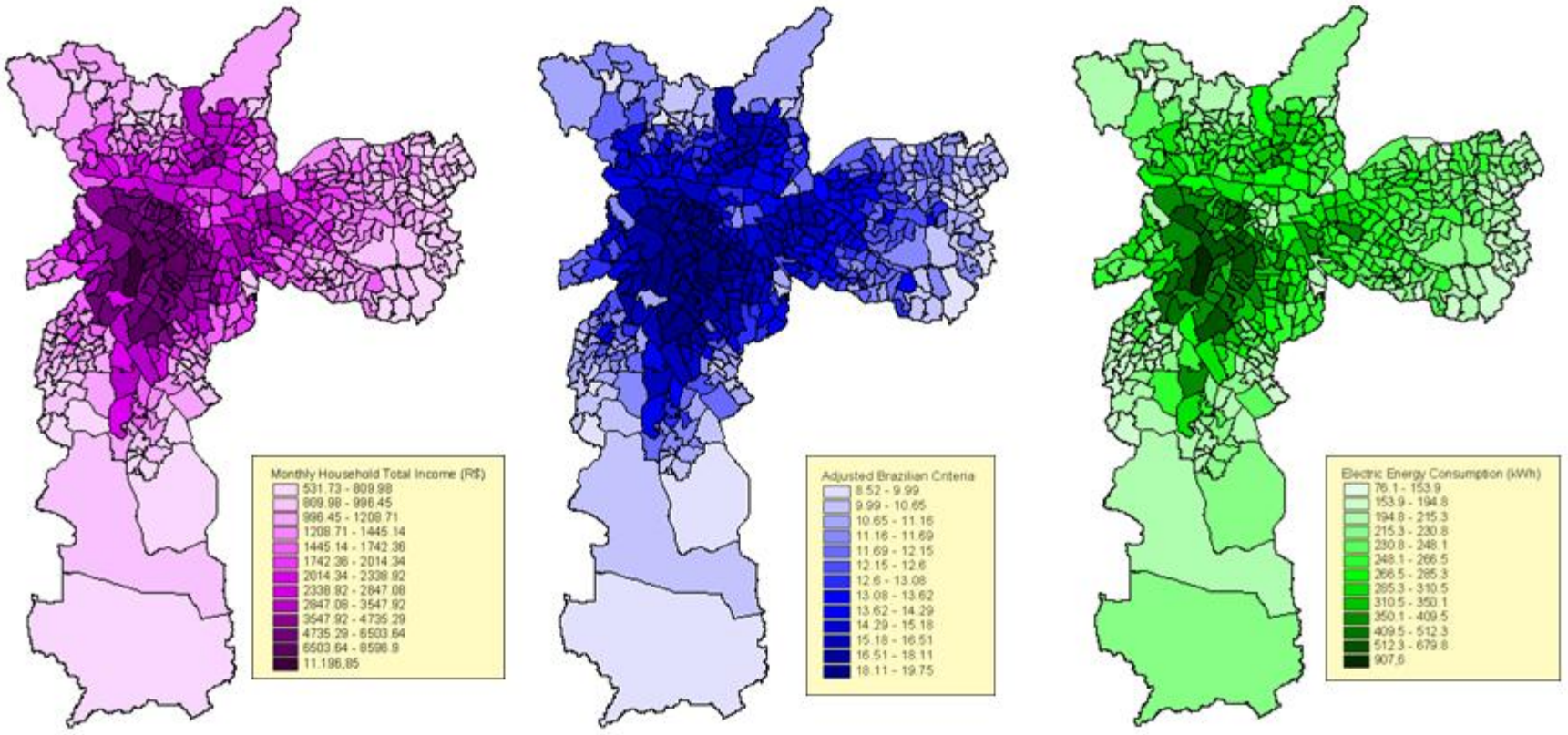


Tools

- **ArcGIS 9.2 and ArcView GIS 3.1**
  - Exploratory analysis, Spatial Join and Summaries
- **ArcGIS Spatial Analyst & Geostatistical Analyst**
  - Exploratory analysis
- **R 2.6.1 Statistical Tool**
  - SPDEP package: Moran's I and SAR models
  - SPGWR package: GWR models (including AIC optimization)
- **GeoDA 0.95i**
  - LISA Maps, SAR models
- **GWR3X**
  - GWR models
- **Next Steps:**
  - **Promote comparative studies**
    (Latin America, USA/North America, Europe, Asia, Australia/NZ)
  - **Evaluate temporal perspective and Household Level**
  - **Explore GWR and Moran's I in ArcGIS 9.3 !!!**

Conclusions

- Use of the mean **household electricity consumption**, at a territorial aggregated level, **is an excellent regional indicator of income concentration in the city of São Paulo**



Household Income

Brazilian Economic Status

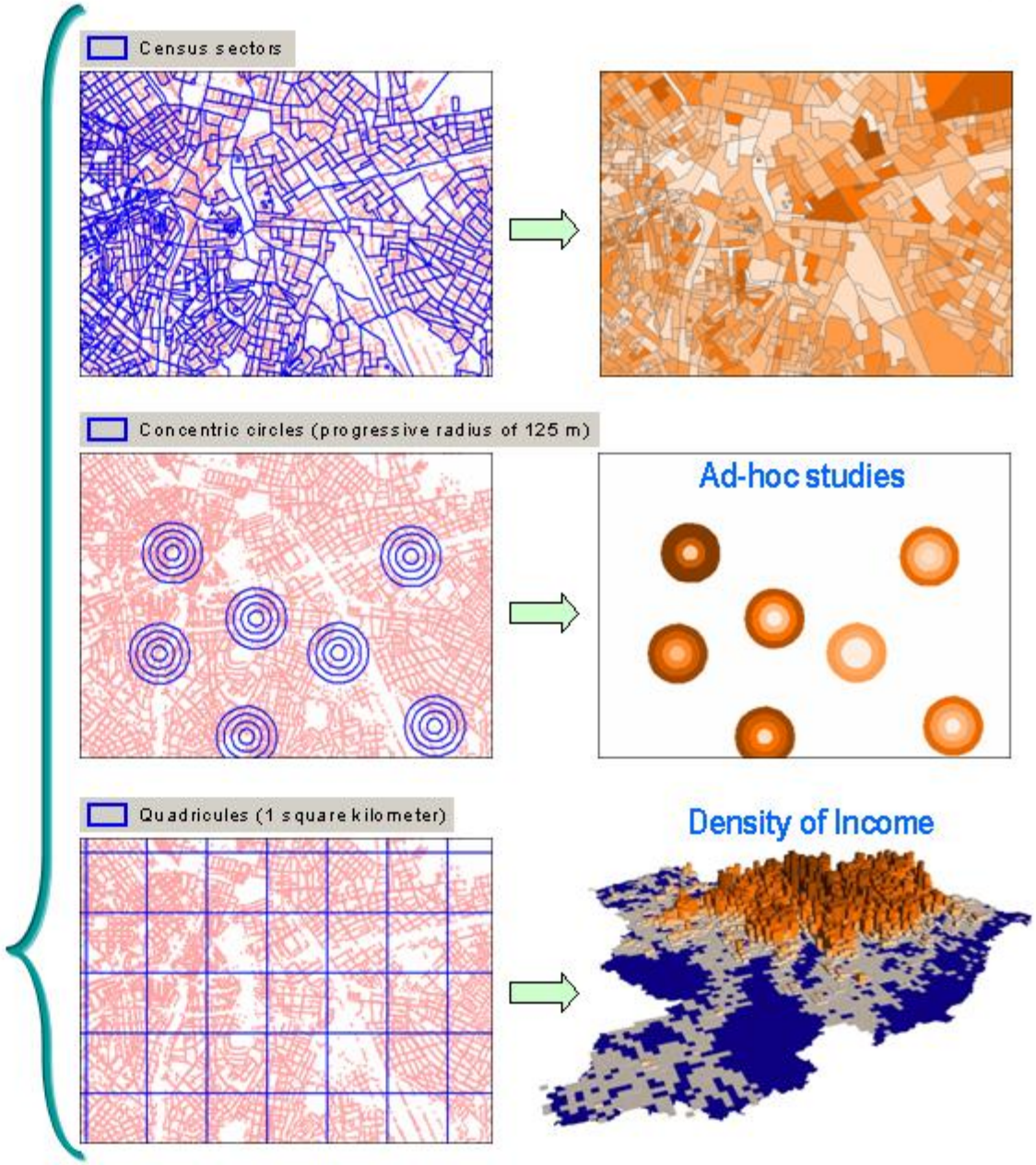
Electric Energy Consumption

Managerial Implications – A Potential New Business

Income indicators based on Electric Energy consumption

- Should be published widely by power distribution companies, Energy commissions or agencies
- Useful for strategy formulation and decision making
  - **Household classification, Concentration analysis, Prediction, clustering**

Households



Eletropaulo

Social-Economical Variables & Energy Losses (Fraud)

Losses per Station (%)

+

Income Density
(Electric Consumption)

+

Percentage of
Disconnected Customers

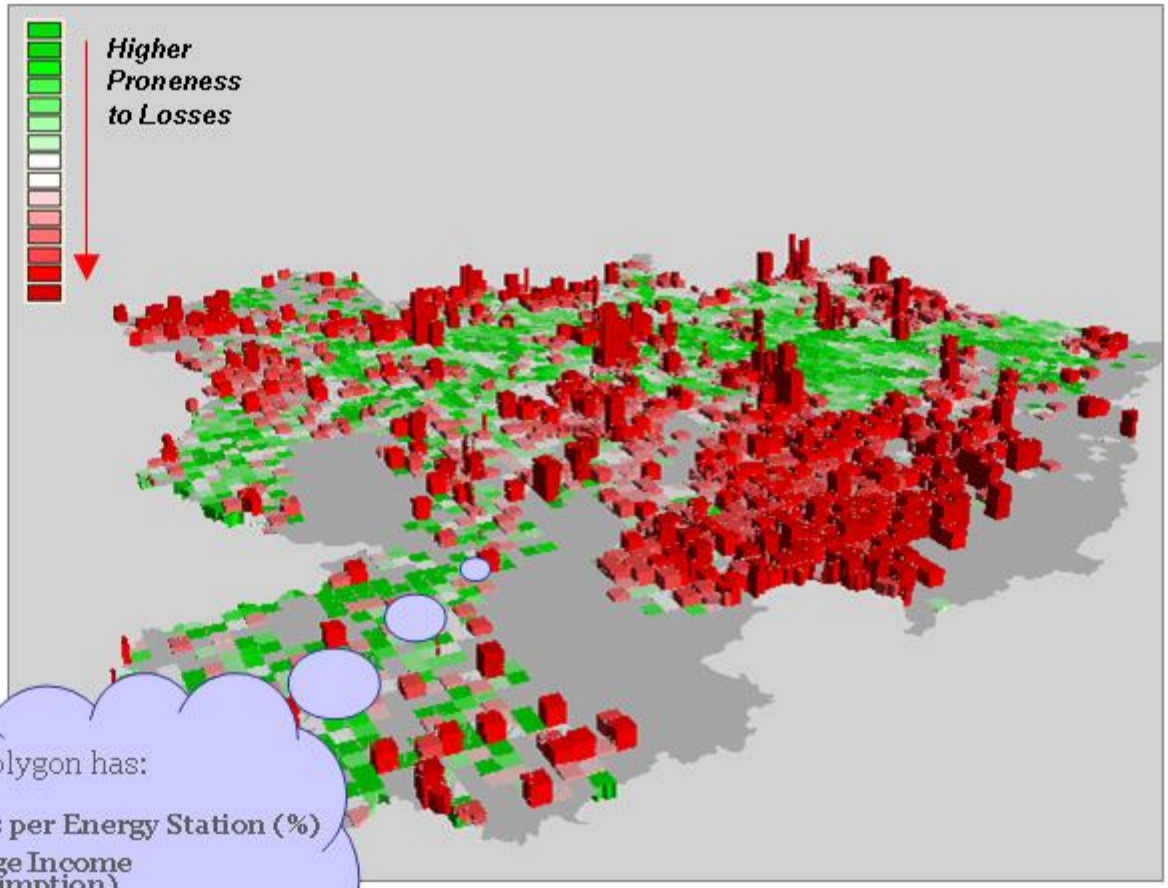
+

Slum Coverage

+

Geographical Superposition
Spatial Overlay

Application of Energy Consumption as Income indicator



Each Polygon has:

- Losses per Energy Station (%)
- Average Income (Consumption) per residence
- Slum Coverage Index (%)
- Cut-off Clients (%)

Map of Areas with Higher Proneness to Energy Losses (FRAUD)

Thank You !!!

Income Indicators based on Electricity Consumption:
A Geostatistical Approach

Eduardo de Rezende Francisco
eduardo.francisco@aes.com

2008 ESRI International
User Conference

San Diego, CA, USA - August 5th 2008