Parallel PySAL

Autoregression and Complex System Framework Integration

Jason Laura, Robert Pahle, Sergio Rey, Luc Anselin

GeoDa Center for Geospatial Analysis and Computation
Arizona State University

August 13, 2014

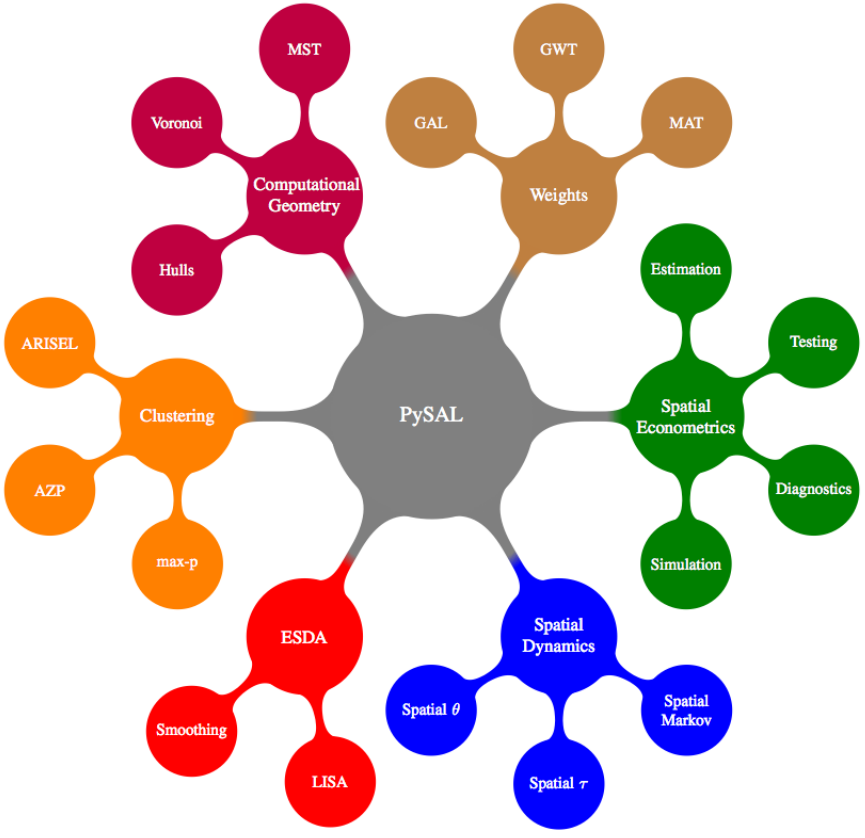
PySAL

Substantive Application: Spatial Econometrics

Implementation

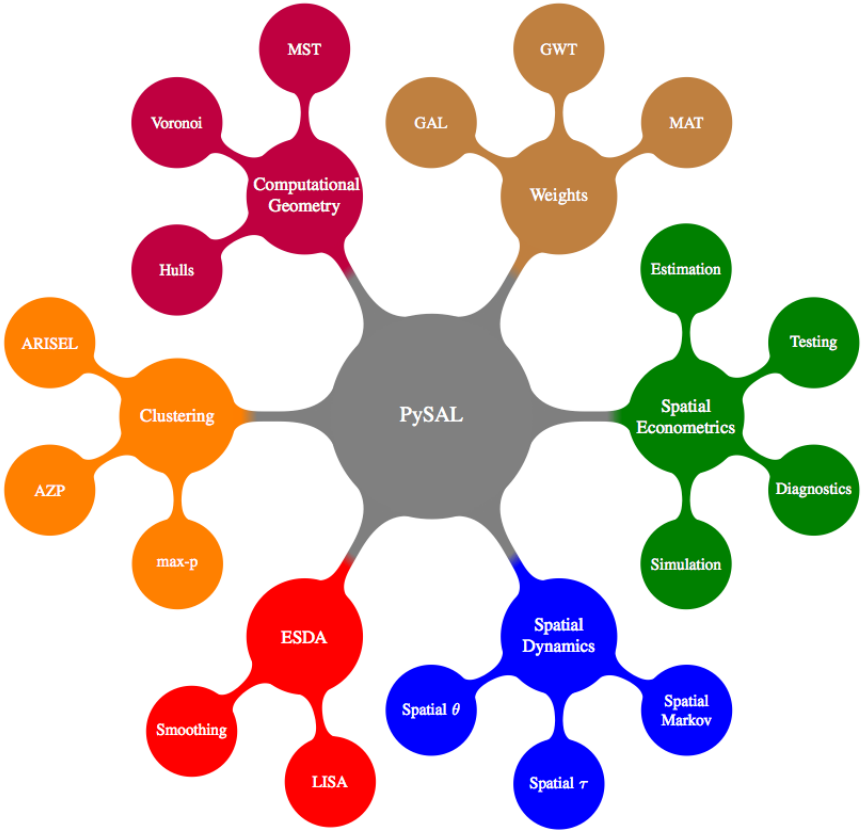
PySAL

- Spatial analysis library
- Big data world
- v 1.8 July 2014



pPySAL

- contiguity builder
- max-p region
- p-lisa
- fisher jenks
- spatial regimes



Lessons Learned

- Hardware dependence
- No holy grail of automatic parallelization
- Need a roadmap = Taxonomy
  - Guidance on "best practice"
  - Identify dead ends

Spatial Econometrics

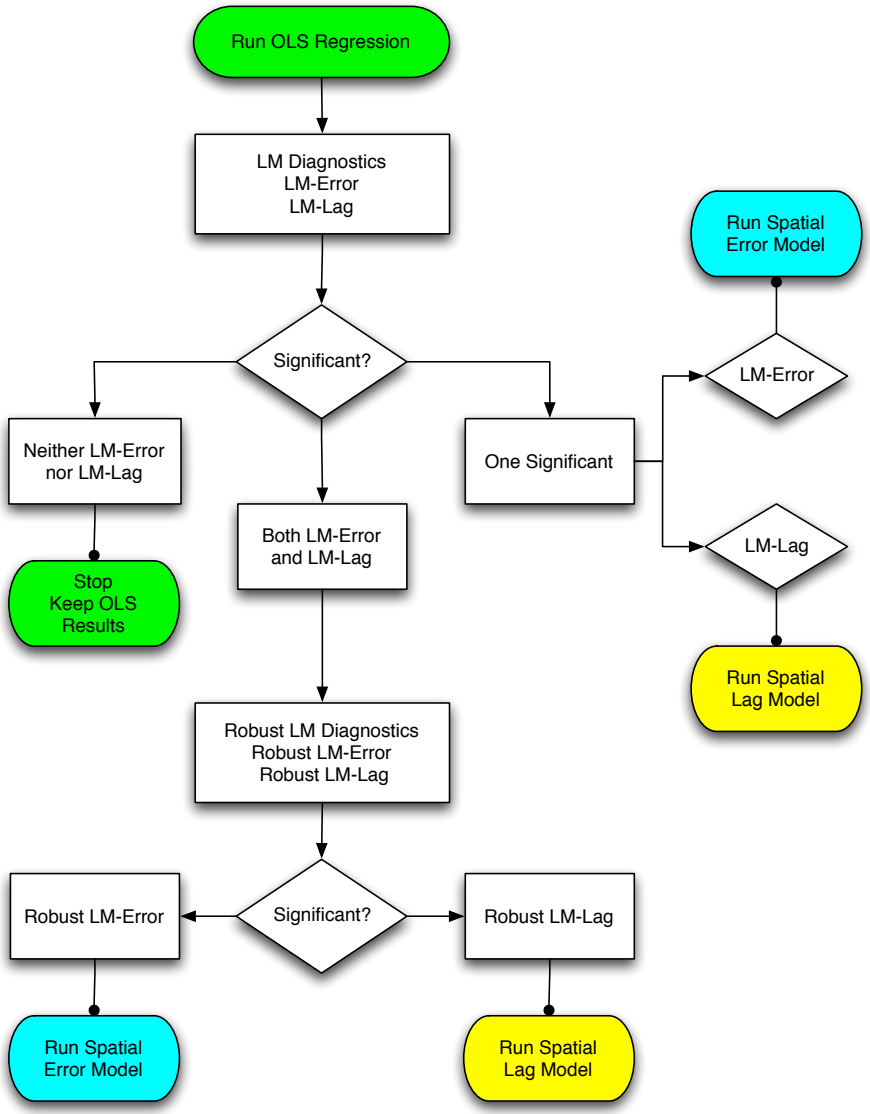
GeoDaSpace

- GUI ontop of spreg
- Subset of spreg functionality
- Cross-platform

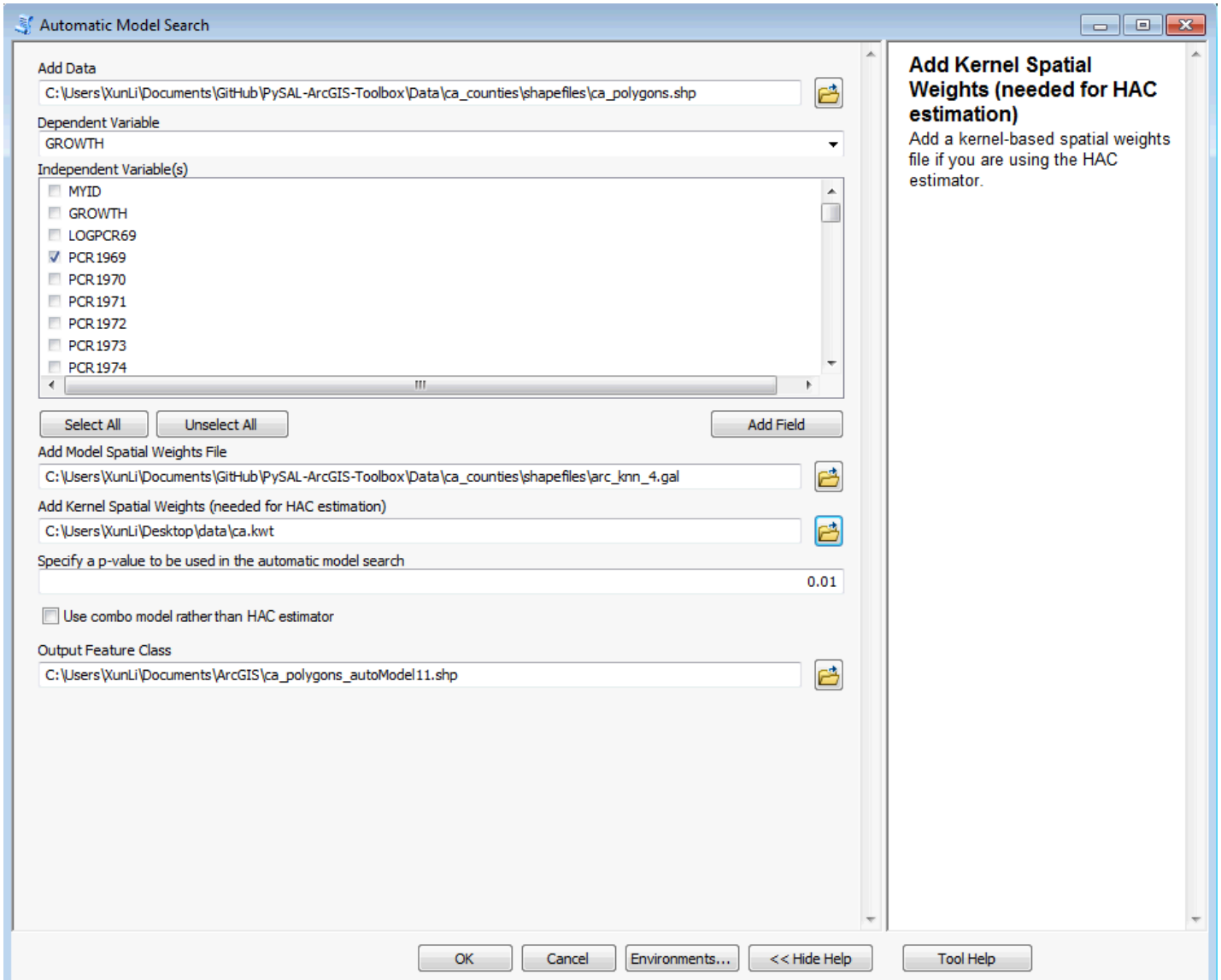
Specification Searches

- Specific to General
  - $y = X\beta + \epsilon$
  - OLS + Lagrange Multiplier Tests
- General to Specific
  - $y = \rho Wy + X\beta + (I - \lambda W)^{-1}\nu$
  - ML + Restrictions

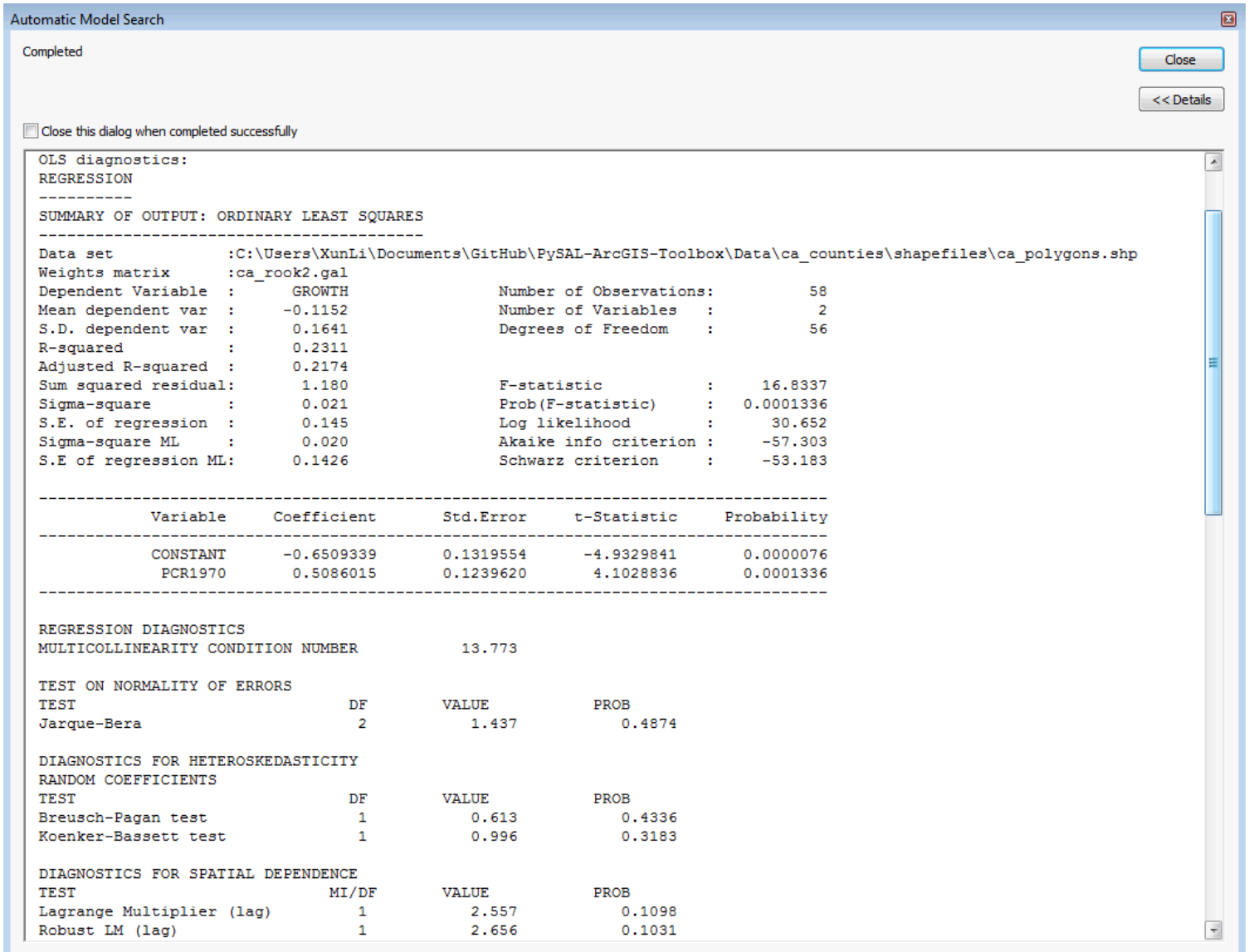
LM Based Specification



ArcGIS Toolbox



ArcGIS Toolbox




```
OLS diagnostics:
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :C:\Users\XunLi\Documents\GitHub\PySAL-ArcGIS-Toolbox\Data\ca_counties\shapefiles\ca_polygons.shp
Weights matrix      :ca_rook2.gal
Dependent Variable  :      GROWTH                Number of Observations:          58
Mean dependent var  :     -0.1152                Number of Variables   :           2
S.D. dependent var  :      0.1641                Degrees of Freedom    :          56
R-squared           :      0.2311
Adjusted R-squared  :      0.2174
Sum squared residual:       1.180                F-statistic           :     16.8337
Sigma-square        :       0.021                Prob(F-statistic)     :   0.0001336
S.E. of regression  :       0.145                Log likelihood        :      30.652
Sigma-square ML     :       0.020                Akaike info criterion :     -57.303
S.E of regression ML:      0.1426                Schwarz criterion     :     -53.183

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      -0.6509339       0.1319554      -4.9329841       0.0000076
             PCR1970       0.5086015       0.1239620       4.1028836       0.0001336
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           13.773

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2           1.437          0.4874

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                1           0.613          0.4336
Koenker-Bassett test              1           0.996          0.3183

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Lagrange Multiplier (lag)         1           2.557          0.1098
Robust LM (lag)                   1           2.656          0.1031
```

Hardware

Parallel Strategy

Complex Systems Framework

Components for Autoreg