

Spatial Data Analysis and Econometrics with PySAL

Pedro Amaral, Luc Anselin, Elijah Knaap, and Sergio J. Rey

NARSC 2024

Description

A unique feature of this tutorial is the use of Python based software tools for spatial data analysis and spatial econometrics. Python is an object oriented scripting language that is gaining rapid adoption in the computational sciences. Since its initial release in July 2010, PySAL has been downloaded over 1.7 million times. This two-part tutorial will first provide participants with an introduction to Python and related tools for exploratory spatial data analysis. In the second part of the tutorial the focus is on spatial econometric methods in PySAL.

Instructors

Pedro Amaral is Professor at the Department of Economics at the Federal University of Minas Gerais, Brazil (Cedeplar/Brazil), Visiting Professor at the University of Chicago, and Fellow of the Center for Spatial Data Science at the University of Chicago. He holds a PhD from the University of Cambridge and serves as co-editor of the journals Spatial Economic Analysis, Journal of Spatial Econometrics, and Regional Studies, Regional Science. He also serves on the scientific council of the Brazilian Regional Science Association, on the Honors Committee of the Regional Science Association International, and as ambassador for the Regional Studies Association in Brazil.

Amaral is a core developer of the spatial econometrics package spreg in PySAL and his research interest is in applied spatial economics and spatial econometrics, working mainly with health equity and regional disparities.

Luc Anselin is Stein-Freiler Distinguished Service Professor and Founding Director of the Center for Spatial Data Science at the

University of Chicago. He previously held appointments at Arizona State University, the University of Illinois, Urbana-Champaign, the University of California, Santa Barbara, and The Ohio State University, among others. He holds a PhD in Regional Science from Cornell University.

Anselin developed the original SpaceStat software for spatial data analysis in the early 1990s, and leads the the GeoDa software project for spatial data science. He also was a co-founder of the PySAL project. His 1988 Spatial Econometrics text has been cited more than 16,000 times.

His research has been recognized by a number of awards, including the Walter Isard and Alonso prizes of NARSC, and the Tobler award of the Austrian Academy of Sciences. He is a Fellow of RSAI, member of the National Academy of Sciences and of the American Academy of Arts and Sciences.

Elijah Knaap is the Associate Director of the Center for Open Geographical Science at San Diego State University. He holds a PhD in Urban Studies and Planning from the University of Maryland and is a core developer for the PySAL ecosystem of spatial analysis software. His work examines question of urban and regional inequality, and he leads development of geosnap, a Python package for applied neighborhood analyses.

Sergio J. Rey is Professor of Geography and Founding director of the Center for Open Geographical Science at San Diego State University. Prior to joining SDSU, he held professor and director positions at the University of California Riverside and Arizona State University. He is a Fellow of the Spatial Econometrics Association, the Regional Science Association International, and the recipient of the 2016 University Consortium of Geographical Information Science Research Award.

Rey is project director for PySAL and the creator of the open source package STARS: Space-Time Analysis of Regional Systems. His research interests are in the area of geocomputation, exploratory space-time data analysis, spatial econometrics and urban/regional modeling with substantive applications to problems in regional economics, criminology, epidemiology, spatial demography and urban dynamics, among others. Rey has served as editor of the *International Regional Science Review* (1999-2015), editor of *Geographical Analysis* (2014-17) and as an editorial board member of *Computers, Environment and Urban systems*, *Geographical Analysis*, *Papers in Regional Science*, *Professional Geographer*, *Region et Developpement*, and *Spatial Demography*.

Format

The tutorial is planned for a full day, broken into two half-day sessions. Each session is organized into two 80-minute components separated by a 20 minute break.

Objectives

This tutorial will offer participants the following:

- Introduction to Python for spatial data and regional analysis
- Introduction to PySAL for exploratory spatial data analysis
- Introduction the PySAL for spatial econometric modeling

Outline

AM Session (9-12)

1. Software and Tools Installation (30 min)
 - a. Anaconda Python Distribution
 - b. IPython/Jupyter Notebooks
2. PySAL for ESDA (80 min)
 - a. PySAL Overview and Setup (20 min)
 - b. Spatial data processing with PySAL (30 min)
 - i. Processing spatial data with PySAL
 - ii. Spatial weights in PySAL
 - c. ESDA with PySAL (30 min)
 - i. Global spatial autocorrelation analysis
 - ii. Local spatial autocorrelation analysis
3. PySAL for Regional Analysis (60 min)
 - a. Regionalization (20 minutes)
 - i. clustering
 - ii. spatially constrained clustering
 - b. Segregation (20 min)
 - i. segregation measures
 - ii. comparative segregation analysis
 - c. GeoSNAP (20 min)
 - i. neighborhood definition
 - ii. neighborhood dynamics

PM Session (1:30-4:30)

1. Introduction to spatial econometrics in PySAL (20 min)
 - a. Overview of functionality in spreg
 - i. cross-sectional models
 - ii. spatial regimes
 - iii. spatial panel data
 - b. Logic of spatial econometric analysis in spreg
 - i. spatial weights
 - ii. model specification
 - iii. estimation
 - iv. diagnostics

2. Classic cross-sectional spatial regression - 1 (60 min)
 - a. OLS with spatial diagnostics
 - b. Spatial lag and spatial error model
 - i. ML estimation
 - ii. IV/GMM estimation

Break (20 min)

3. Classic cross-sectional spatial regression - 2 (30 min)
 - a. Spatial Durbin and SLX model
 - b. Direct and indirect effects
 - c. Spatial probit

4. Spatial regimes (30 min)
 - a. Exogenous regimes
 - b. Endogenous regimes

5. Spatial panels using SUR (20 min)
 - a. Classic SUR models
 - b. Spatial lag and spatial error SUR models

Audience

Regional scientists, GIScientists, researchers and students interested in learning Python and using PySAL for computational scripting in spatial analysis and spatial econometrics.

Prerequisites

This tutorial is geared towards individuals who have a basic

understanding of exploratory spatial data analysis and econometrics. Background materials will be sent to participants prior to the workshop.

Required Materials

- The tutorial will consist of computationally-based instruction. This will include basic scientific scripting with Python and an introduction to PySAL.
 - Participants are encouraged to bring their own laptops. We will cover installation of all required software in the morning session.
 - Software Requirements
 - Software will be made available on site.
 - Sample data sets will be provided with the packages. ●
- Workshop repository:
<https://github.com/sjsrey/workshop-pysal-narsc24>

Expected Number of Participants

We have taught similar workshops to audiences of up to 75.