STAT 510: Spatiotemporal Stats

Getting Started and Introduction

Prof. Taylor-Rodriguez

Welcome: what's for today?

1. Brief introductions
2. Course Overview
3. Vote on drop-in hours
4. Getting Started
5. D2L 2 minute tour
6. Course motivation and intro

Brief Introductions

- What is your name? your program? any coding experience?
- Your motivation for taking this course?
- Something you like about remote learning
- Something you really dislike about remote learning
- Anything else you'd like to share???

Course Overview

Description

This course will introduce you to basic concepts and approaches for the analysis of dependent data in space, time or both, and will provide you with the computational tools needed to do so.

VERY BIG area of research, so we'll stick to a breadth of well-seasoned approaches, emphasizing the intuition behind the methods worked out through applications in R.

Prerequisites

STAT 562 and STAT 565, some matrix algebra is useful but not necessary

- *if you feel like you could use a refresher (or a crash course) on the bare bone basics of matrix algebra needed for the course, check out Appendix A in the textbook*.
- R experience not required but would help A LOT!!! I will provide some resources for you to catch up with the basics, and we may use some drop-in hour time to help you out

Textbook

(Mandatory) Spatio-Temporal Statistics with R, by Chris Wikle, Andrew Zammit-Mangion and Noel Cressie. This book is freely available online at https://spacetimewithr.org.

(Optional) Statistics for Spatio-Temporal Data. Noel Cressie and Chris Wikle.

Objectives

1. Introduce descriptive statistics and visualization techniques to discover spatial and temporal patterns.
2. Learn basic modeling approaches as tools for exploratory analysis to aid in identifying patterns and validate modeling assumptions.
3. Build suitable inferential models based on a rigorous preliminary exploration of the data and its characteristics.

Student Learning Outcomes

By the end of the course, you should be able to:

1. Select appropriate methods for common spatial, temporal or spatiotemporal problems.
2. Have polished your computational skills in R.
3. Critically evaluate the resulting statistical models.
4. Communicate to lay audience steps and results of a spatio-temporal statistical analysis.

Course Format

- Mix of brief lectures (20-30 mins) with short practice problems (a.k.a. in-class problems, are to be submitted once weekly)
- Once we have made enough progress on each chapter we'll tackle an R lab
- You will work on a project throughout the whole term (more on this later)

Assessments

Assessment	Weight
Class participation problems	20%
Homework and lab assignments	25%
Project proposal	5%
Midterm project update	10%
Project report	20%
Final presentation	20%

Assessments

Class participation problems (20%)

- Each lecture I will assign short exercises and provide time in-class for you to work on them as groups and ask questions.
- These exercises will be due that same week before Sunday at noon in D2L (upload to Activities⇒Assignments⇒In-class Exercises Week n).
- Grade based on submission.

Assessments

Homework and lab assignments (25%)

- Labs associated to the topics we cover in class during class time, and it is your responsibility to work on these and submit your solution to D2L.
- Whenever I consider it essential, in addition to the labs I will ask you to solve a few more theoretical exercises along with your lab solution.

Assessments

Final Project Competition

- By end of week 2 (individually or in groups of 2), choose a spatio-temporal data set (could be one of the datasets I posted). I will meet with each group to discuss your topic.
- You will work with these data throughout the term, attempting to answer a question or set of questions you formulate in your proposal.
- You may alternatively use your data to delve into a topic that we do not cover in class and introduce it to the group.

Assessments

Final Project Competition

Project proposal (5%): one or two page description of the problem, the data and the questions/hypotheses

Midterm project update (10%): short presentation introducing the problem, a brief overview of the data, and describing methodology

Project report (20%) write-up ($\leq$ 10 pages) describing the problem, the data, the methodology and the findings of your analysis.

Final presentation(20%) A 25-30 minute presentation of your project. Your peers, a small group of Faculty, and obviously myself will attend your final presentations and vote for our favorite project

Tentative Schedule

Week	Tuesday		Thursday	
	Topic	Readings	Topic	Readings
1	Setup and Intro	Sec.s 1, 2.1-2.2	Exploratory analysis	Sec.s 2.3-2.4
2	Lab 1	Ch. 2 Lab	Spatiotemporal modeling (part 1)	Sec.s 3.1-3.2
3	Spatiotemporal modeling (part 2)	Sec.s 3.2-3.5	Lab 2	Ch. 3 Lab
4	Descriptive Models (part 1)	Sec.s 4.1-4.2	Descriptive Models (part 2)	Sec.s 4.2-4.3
5	Lab 3	Ch. 4 Labs	Project Progress Update	
6	Descriptive Models (part 3)	Sec.s 4.4-4.5	Lab 4	Ch. 4 Labs
7	Dynamic Models (part 1)	Sec.s 5.1-5.2	Dynamic Models (part 2)	Sec.s 5.3-5.4
8	Lab 5	Ch. 5 Labs	Model Evaluation (part 1)	Sec.s 6.1-6.2
9	Model Evaluation (part 2)	Sec.s 6.2-6.3	Model Evaluation (part 3)	Sec.s 6.3-6.4
10	Lab 6	Ch. 6 Labs	Final Presentations (part 1)	
11	Final Presentations (part 2)			

Grading Policy

Late assignments will be accepted for:

- no penalty if a valid excuse is communicated to the instructor before deadline
- 10% deduction during the 48 hours following deadline
- 20% deduction if submitted between 48 and 96 hours after deadline.
- After that, no grade.

Grade	Range (%)
A	(95-100]
A-	(89-95]

Letter	Range (%)
B+	(86-89]
B	(82-86]
B-	(79-82]

Letter	Range (%)
C+	(76-79]
C	(72-76]
C-	(69-72]

Letter	Range (%)
D+	(66-69]
D	(62-66]
D-	(59-62]
F	(0-59]

Vote on drop-in hours

Getting Started (and other organizational stuff)

R and RStudio

- Download both R and RStudio (instructions in D2L)
- You'll also need SEVERAL packages, they are all listed in page 351 of the book.
- I included a folder with many useful R links.
- If there is enough interest, I am happy to devote one (or part of it) of my two drop-in hours teaching R basics.

Virtual Meeting Space (just for you – mostly)

I thought you might want to try something fun, which will perhaps help you connect outside of class, and I'll most likely make use of it for a few activities (e.g., final presentations).

- Follow **this link to gather town** to register
- You can join at agreed times or just to check if anyone is in to discuss lab, homework, projects, or just to interact with classmates, vent, talk about life, etc.

I am really curious about your feedback on this!!!

D2L two minute tour

Intro

Why study spatio-temporal processes

- Spatio-temporal processes are common in the real world. These can result from interactions across many processes and scales
- Snapshots of spatial events at a fixed time are informative, but provide an incomplete picture about the process
- Looking at a time series at a single location might miss out on the influence nearby locations exert

there is no history without geography and vice-versa

Why should it be statistical?

Processes mostly follow deterministic (even if sometimes chaotic) rules, but...

There is often incomplete data and knowledge about mechanisms driving phenomena

$\Rightarrow$ uncertainty in data, model and parameter values

Why should these models be statistical?

Statistical spatio-temporal models allow:

- capturing the notion of uncertainty without obscuring important trends
- building-in system components that appear random – even if they are not, models are useful if predictions are accurate
- parameter estimation and process prediction (conditional on observed data)
- can be based on our physical understanding of a process (i.e., on a mechanistic model)

Why spatio-temporal modeling?

To characterize processes with uncertain and (too often) incomplete data and system knowledge to:

- make predictions in space and time (smoothing and filtering)
- make predictions in time (forecasting)
- data assimilation with mechanistic models
- conduct parameter inference
- other (computer-model emulation, monitoring network design)

The two approaches

Descriptive (or maginal) approach

Characterize the first and second-moment (covariance) behavior of the process

- Many different processes can generate same marginal form
- More useful if knowledge is limited about driving mechanisms behind process

The two approaches

Dynamical (or conditional) approach

Process values at a location evolve from past values at many locations

- Relate more closely to causal (mechanistic) explanations of the process
- Most useful when good prior knowledge about process is available

Both approaches can be connected through their covariance functions

Hierarchical Statistical Models

- The likelihood of marginal probability models for processes with complex dependencies are hard (sometimes impossible) to compute
- These also struggle with the fact that data are noisy imperfect measurements of what we are usually interested in
- Alternatively, complexity can be built-in gradually through conditioning