Proposed sequence for CSULA STEM Workshop, March 2, 2019

Application example: Wildfires

Covered topics: dealing with messy and incomplete data, analyzing statistical patterns in data, developing a simulation model, analyzing model results

Approximate schedule, 9:30 AM – 2:30 PM:

1. [9:30] Pandas DataFrames, Redux, Redux (10 min.)
2. [9:40] Historical wildfire data from Cal Fire (15 min.)
   1. Overview of datasets
   2. Aside on xlsx vs csv – data structure vs. data format
   3. Reading data from Excel spreadsheet (xlsx) into DataFrame
   4. Initial examination of data
3. [9:55] Dealing with messy data (15 min.)
   1. Types of messiness in data
   2. Missing data (incl. specifying what values constitute “missing”)
   3. Converting data types of interest
4. [10:10] Adding and inspecting derived data for quantities of interest (15 min.)
   1. Computing wildfire durations from alarm and containment dates
   2. Identifying and dropping erroneous data
5. [10:25] Statistics of wildfire sizes, durations, and jurisdictions (20 min.)
   1. Plotting relationship between size and duration, including jurisdictions
   2. Plotting histograms of sizes and durations
   3. Power laws, heavy-tailed distributions, and “crackling noise” in nature

10:45

AM break, 15 min.

1. [11:00] On modeling (10 min.)
   1. Review of wildfire data
   2. Can we build a simulation to model these fires?
   3. Mechanistic models: state variables and rules for their transformation
      1. Some examples
   4. What would you need to describe to model wildfires?
      1. Brainstorming
2. [11:10] Simple “forest fire” model (10 min.)
   1. Reference to papers
   2. Assumptions of the model
   3. Schematic of model ingredients and rules
3. [11:20] Beginning program development for simulation model (30 min.)
   1. Reviewing Spyder IDE
   2. Function to initialize Forest array
   3. Function to plot configuration
   4. Function using NumPy operations to query neighboring trees
   5. Function to update empty forest sites
   6. Function to update forest sites with trees
   7. Function to update forest sites with burning trees

11:50

Lunch break, 1 hr.

1. [12:50] Continuing development of simulation model (30 min.)
   1. Function to update entire forest by one time step
   2. Function to generate a random lightning strike
   3. Function to initiate a fire and let it burn to completion
   4. Function to burn many fires and generate statistics
   5. Function to plot a sequence of fires
2. [1:20] Run simulation (15 min.)
   1. Assemble distributions of fire sizes and durations
   2. Plot simulation data, compare to real data
3. [1:35] Revisiting the model (10 min.)
   1. Extensions: Wind? Drought? Suppression strategies? Climate change?
   2. Connections to infectious disease models
   3. The prevalence of lattices in simulations (incl. discretization of continuous systems)

1:45

PM break, 10 min.

1. [1:55] Running simulation on Stampede2 (35 min.)
   1. How might we use large-scale computer such as Stampede2?
      1. Increased memory for large lattices
         1. Memory estimates? Distributed memory computing…
      2. Multiple independent runs to get better statistics
         1. Like the pi calculation, although runs not strictly independent
   2. Revisting ssh and file transfer
   3. Running multiple jobs to get better statistics