# Lecture 9: Sampling, sweeps, & uncertainty

Complex Systems 530

# Logistical things

- Lab 2 degree distribution comment
- Networks lab check in
- Final project details
- Plan for the next few weeks

# Final Project: Presentation

- **Still sorting out plan!**
- **Dates**
- **Length**: still sorting out due to larger class size, but historically has been 10-15 min presentations
- Everyone in the group must present
- Ask at least one question of another group (or more!)

# Final Project: Paper

- **Due date: Final exam day**, via Canvas (grades are due 72 h later, so this is a hard deadline!)

- **Write-up:** roughly 8-12 pages, but shorter or longer is fine so long as you fully cover the motivation, problem, methods/model, results, etc.

- **Model code:** I should be able to both review the code and run it myself if in NetLogo or Python.
  - Your code must be **documented, clear, and readable**.
  - Be sure to also document the version of python and any packages you used.

# Paper components

- **Introduction.** problem overview and literature review—what gap or question are you addressing? What has been done before?

- **Methods: model description.** Describe how your model works in terms of its: agents, interactions, environment, model schedule/timing

  - You can use the PARTE and/or ODD frameworks as a guide

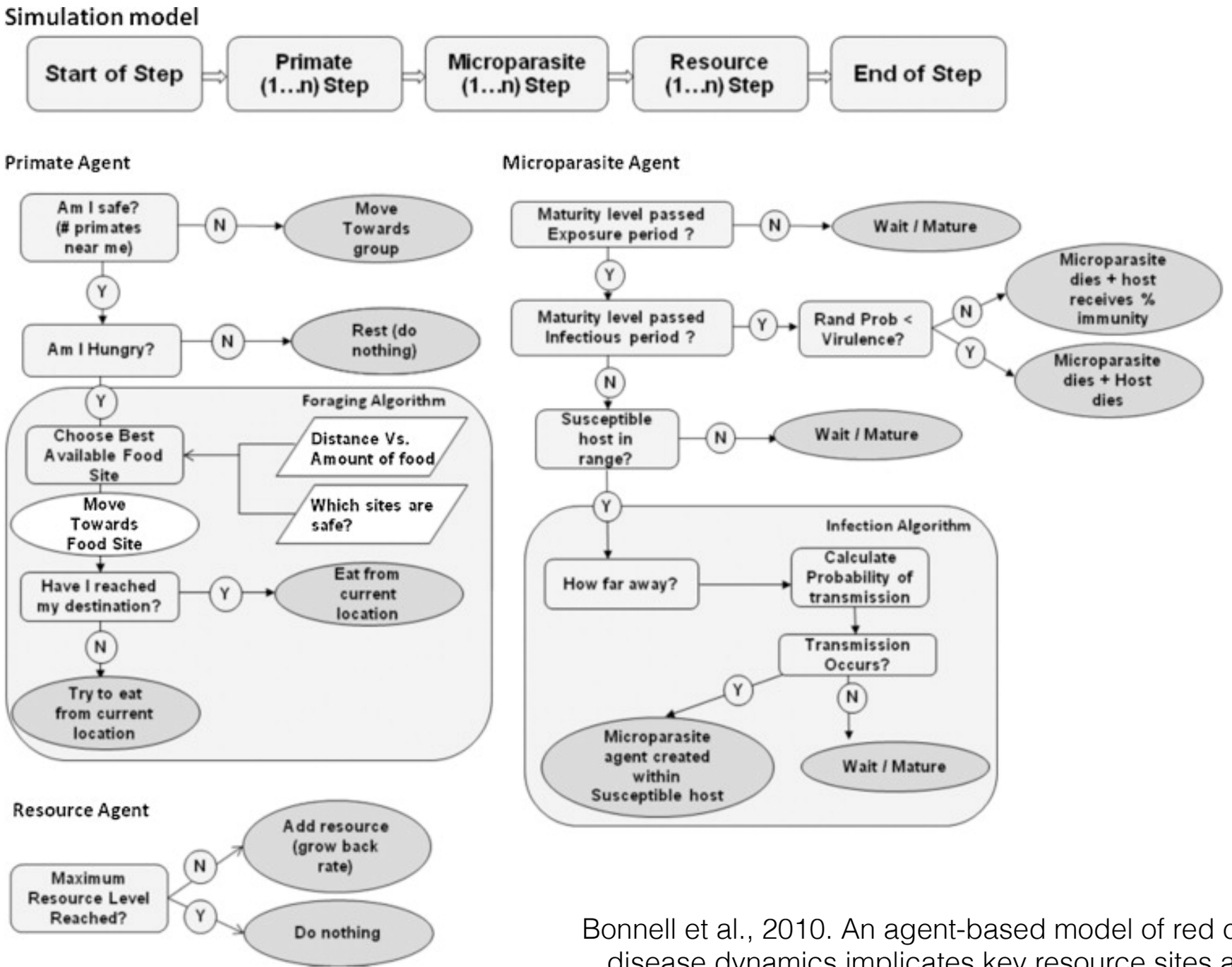
  - Flow charts & visuals are good!

# How to write up an ABM

- PARTE - Properties, Actions, Rules, Time, Environment
    - Agents are defined by their properties, actions, and rules
- ODD Protocol - Overview, Design concepts, and Details
    - Describes the model purpose, structure, and design concepts around which the model was built (e.g. emergence, adaptation, etc.)
- You will get more practice on the next lab!

PARTE: https://www.ncbi.nlm.nih.gov/books/NBK305917/ ODD: https://www.sciencedirect.com/science/article/pii/S030438001000414X

# Flow Chart Example




Bonnell et al., 2010. An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission. https://www.vetmed.wisc.edu/goldberglab/pdf/P071.pdf

# Paper components

- **Methods: model analysis.** Describe the parameter settings you swept through, the analyses you ran. If your model is stochastic, you may need to run multiple trials at the same parameter settings.

- **Results.** Provide qualitative and quantitative summaries of how your model behaves. Provide graphs and plots of model outcomes at different settings as needed.

# Paper components

- **Discussion.** Return to the question or problem in your intro—what do your results say about this problem? Put this in a broader context. Describe the strengths, limitations, and potential future directions of your work.

  - If you think there are still some bugs driving your model's behavior, this is the place to discuss this

  - Also a good place to talk about how you might verify, validate, or extend the model in the future

# A note about verification as you build your model

- Your model will have bugs, be wrong, do things you didn't intend/think through, etc.

- Unit testing and thinking and building with small examples and modules is key

- Figure out small test cases where almost the entire model is either absent or set to a constant and you know what the one piece you're working with should do. Then move on to the next piece/ submodel, etc.

# Plan for the next few weeks

- Sensitivity analysis and exploring the parameter space
- Visualizing and interpreting results
- Different levels of agent cognition - game theory & decision theory
- Model calibration & estimating parameters from data
- Model comparison & inference robustness
- More complex environments (e.g. GIS)

# ABM complexity & dimensionality

# ABM complexity & issues of dimensionality

- So far, our ABMs have obeyed fairly simple rules—reasonably easy to describe with a bulleted list (even if this can leave a lot of ambiguity still)

- This already leads to a high dimensional input/parameter space

- But many ABMs used in practice include more detailed mechanisms, more complicated processes to capture realistic details or more intricate questions

# Parameter/input space

- Voting model
- Grid size (width, height)
- Initial fraction/probability of yes/no votes
- Neighborhood size

# SIR model on a network

- Network generation parameters (e.g. total number of nodes and probability of an edge for an Erdös-Renyi graph)
- Probability of transmission
- Probability of recovery
- Initial conditions (numbers of S, I, and R individuals)

# Power grid network model

- Network parameters: number of nodes, number of connections for each new node (scale free graph)

- Node capacity distribution parameters (in this case the mean and SD for a normal distribution)

- Initial electrical load parameters (in this case the bounds for a uniform distribution)

- Initial node states (e.g. fraction of failed nodes at start time)

# Ants Model

- Food sources: number of sources, sizes, locations
- Nest location, nest scent gradient parameters
- Population size
- Chemical (pheromone) parameters: how much, diffusion rate, evaporation rate
- Ant movement parameters (angles to wiggle), chemical detection upper & lower bounds, chemical and nest scent angle bounds, how much chemical to drop, etc.

# Example: Walking & Health

## A Spatial Agent-Based Model for the Simulation of Adults' Daily Walking Within a City

Yong Yang, PhD, Ana V. Diez Roux, PhD, MD, Amy H. Auchincloss, PhD,
Daniel A. Rodriguez, PhD, Daniel G. Brown, PhD

- Models walking behaviors in an Ann Arbor-sized city as a function of health, SES, neighborhood safety, commuting patterns, and other features

- Time is discrete steps, with 1 step = 1 day

*Yang et al / Am J Prev Med 2011;40(3):353–361*

# Environment

- Environment: 800 x 800 grid
  - Each cell = 10 m x 10 m (64 km city ~ Ann Arbor)
- Neighborhoods - 40 x 40 cells
  - Neighborhood properties: safety, asthetics, other environmental elements

# Environment properties

- Elements in the environment (based on Ann Arbor census)
    - 200 groceries
    - 800 non-food shops
    - 1500 social places
    - 12000 workplaces
    - 60000 households

# Agents

- 108,000 individuals
  - 48,000 couples, and 12,000 single people
    - Range of properties related to walking, SES, health, etc.
    - E.g. walking ability is determined by:

$$A_b = U^4(0,1) \times \left( \frac{\mathbf{Min}(|137 - \mathbf{Age}|, 100)}{100} \right)$$

**Table 1.** Properties of individuals and selected model parameters

| Individual-level properties | Values and meaning |
|---|---|
| **Gender** | Assigned as male or female with equal probability |
| **Age** | Random integer from the uniform distribution ranging between 18 and 87 years; the difference between a couple is no more than 3 years |
| **SES** | Integer value ranging between 1 and 5, with higher values indicating higher SES. Members of a family are assumed to have the same SES. In select scenarios, SES is assumed to be positively correlated with increasing distance from the city's center |
| **Family size** | Family size can be 1 (with probability of 20%) or 2 (with probability of 80%) |
| **Friends** | Each individual has 3 to 5 friends who can influence her/his walking attitude, randomly selected from the people with the same or similar SES value (difference no more than 2) |
| **Dog ownership** | Each individual has a 20% probability of having a dog (dog owners have a higher probably of walking within the neighborhood) |
| **Household** | Each person is randomly assigned to a household (except in scenarios involving residential segregation by SES) |
| **Work and workplace** | If aged ≤69 years, the probability of working is 95%; people aged >69 years are assumed to not work. Working people are randomly assigned a workplace in the city. This can be a grocery store, a non-food shop, a social place, or another workplace in the city |
| **Walking ability ($A_b$)** | Value ranges from 0 to 1, the higher the value, the longer distance an individual can walk |
| **Attitude toward walking ($A_t$)** | Value ranges from 0 to 1, the higher the value, the higher probability an individual will walk |

**CALIBRATED MODEL PARAMETERS**

| Activity | Daily probability of performing the activity | Maximum walking distance for the activity (miles) |
|---|---|---|
| Work[a] | 1 for individuals with a job; 0 for others | 1.125 |
| Food shopping | 0.4 | 2 |
| Other shopping | 0.25 | 1.5 |
| Visiting a social place | 0.2 | 2.5 |
| Leisure within neighborhood | 0.33 | 5.5 |

[a]Only among employed

# Actions & Rules

- Individuals can walk for three purposes: (1) to work; (2) for basic needs (such as food shopping, other shopping, and visits to social places); and (3) for leisure.

- Walk with a certain probability each day (e.g., non-food shopping takes place every 4 days on average, resulting in a daily probability of 0.25).

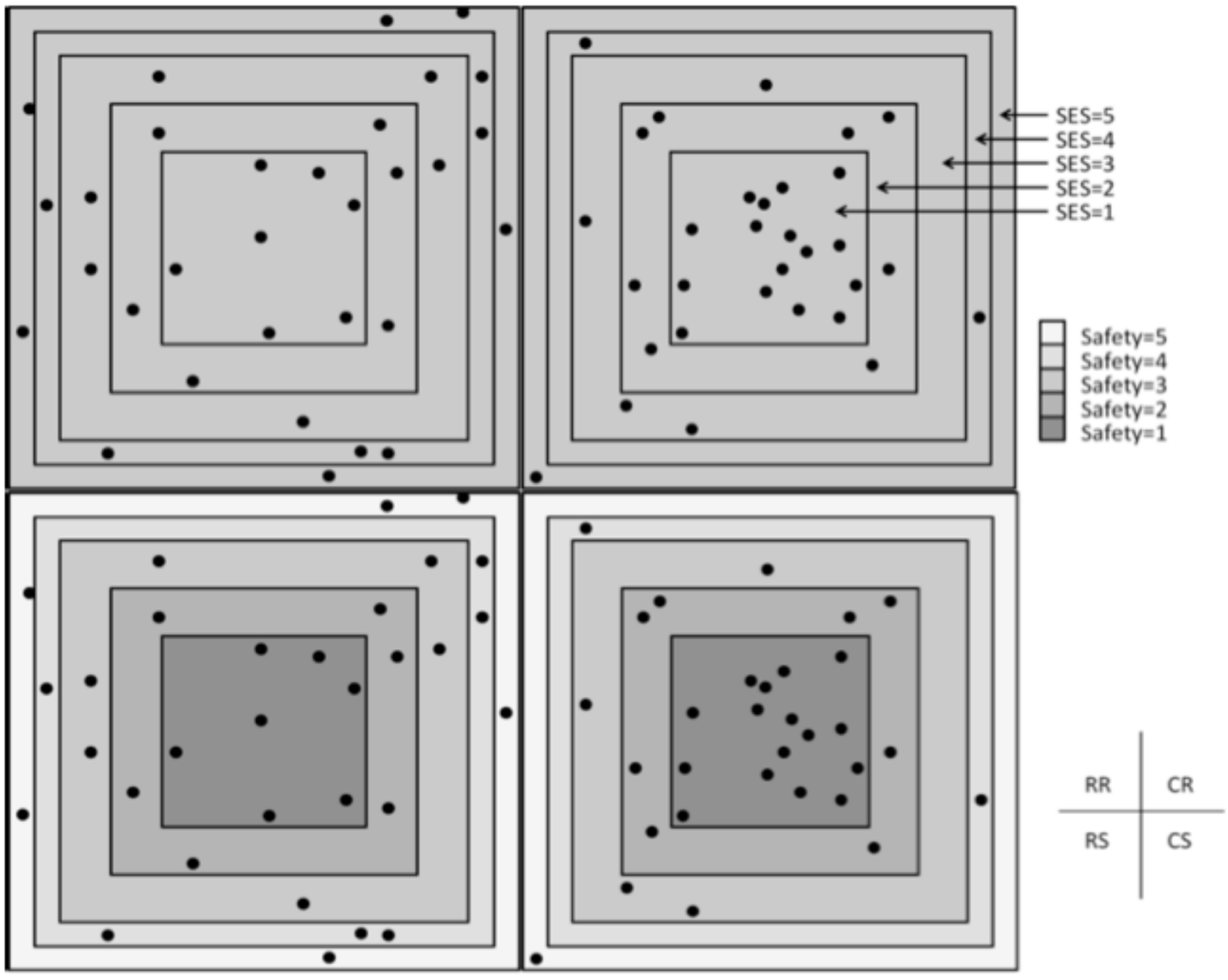
- Maximum walking distance for each type of activity. Person-specifıc maximum distances are calculated as the product of maximum distances and the person-specifıc ability (Ab).

# Actions & Rules

- Example: Work. If the distance between the person's household and workplace is less than the person-specific maximum walking distance for work, then the decision to walk is a random draw with probability equal to attitude At.

- Attitudes update each time step with feedback from household members, safety, aesthetics of the route, etc.

# Example: Walking & Health



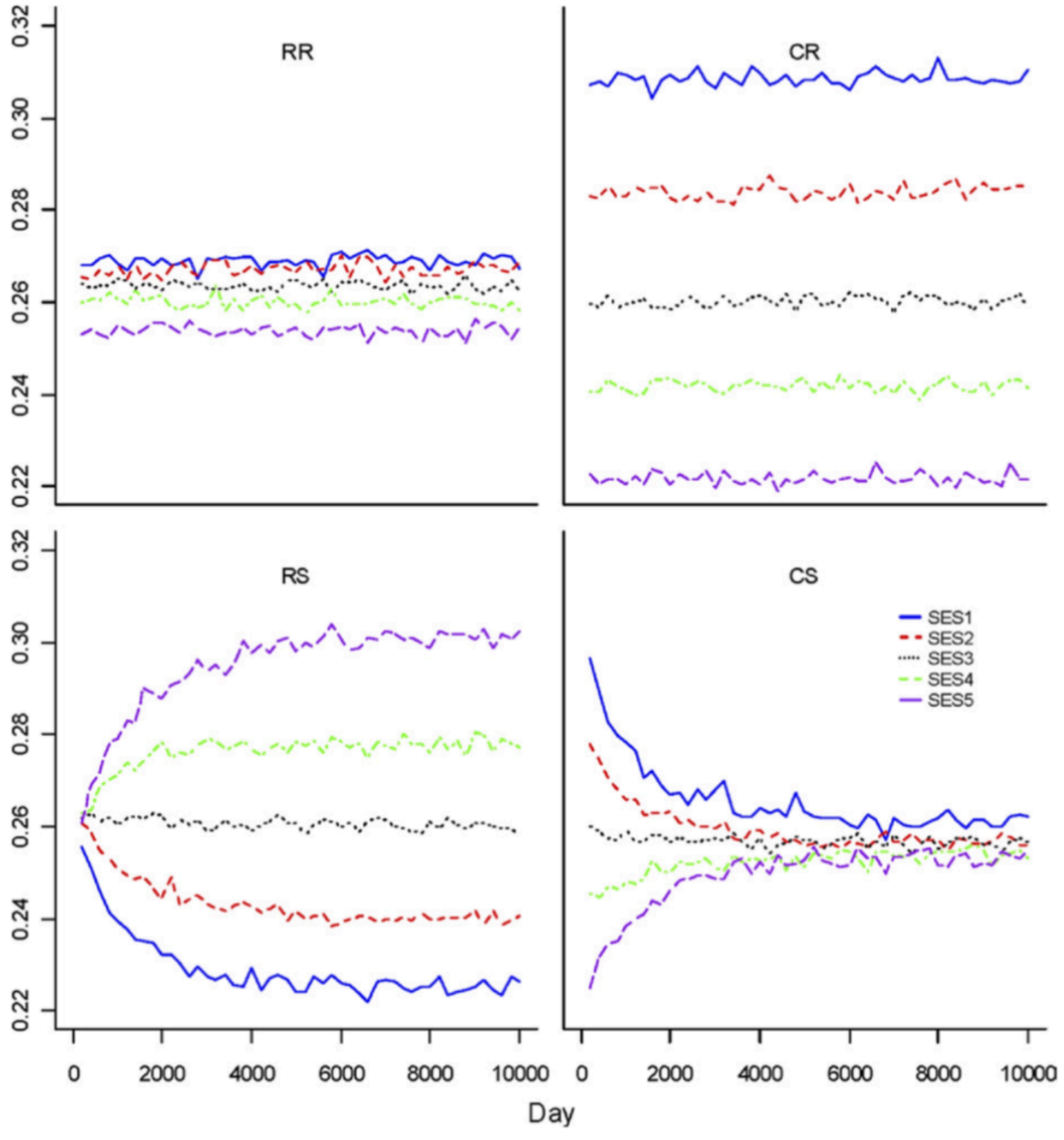



**Figure 1.** Average walking trips per day per person for different SES groups over time for four scenarios

CR, more non-household locations in the core and random safety level; CS, more non-household locations and lower safety level in the core; RR, randomly distributed non-household locations and random safety level; RS, randomly distributed non-household location and lower safety level in the core

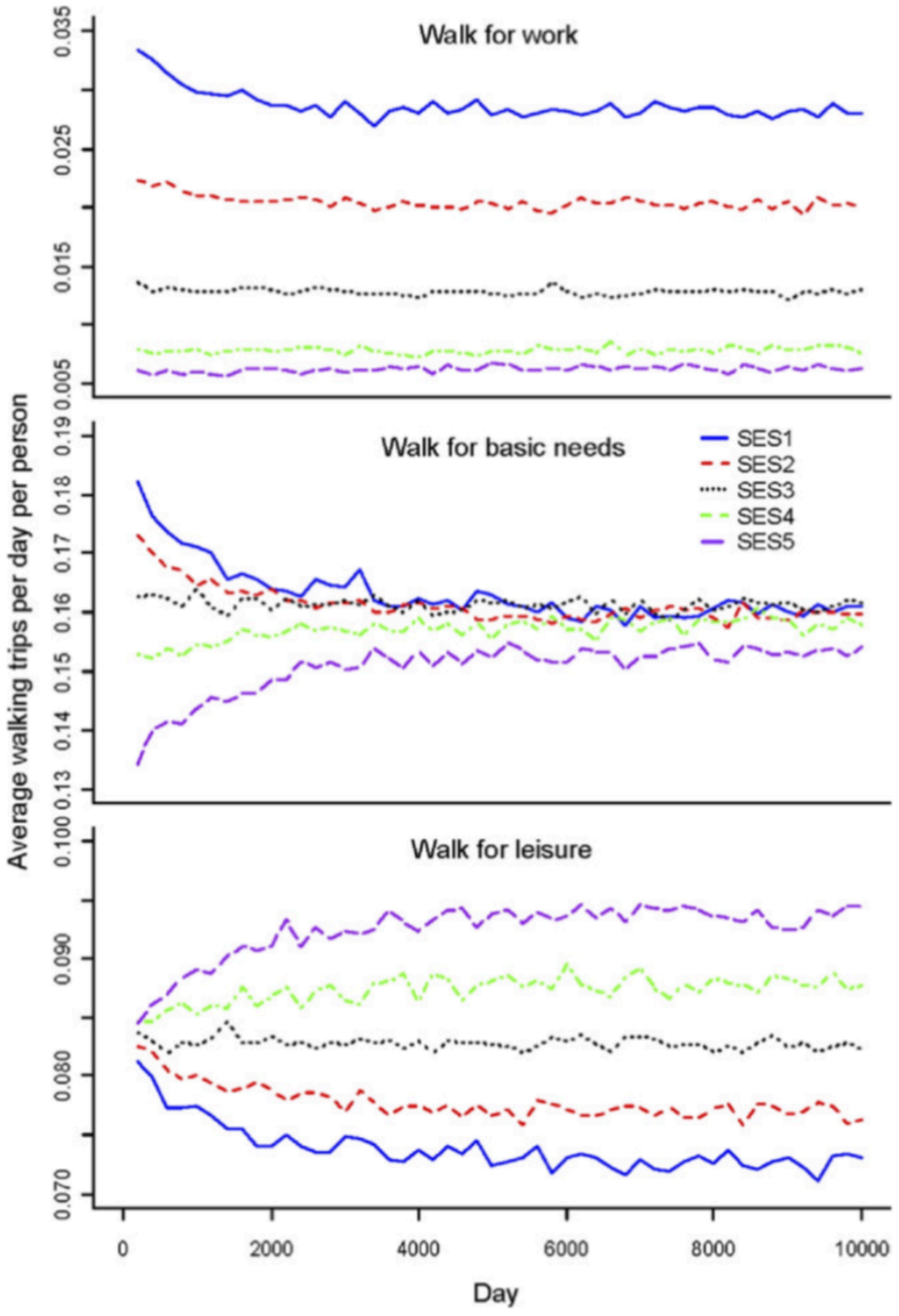



**Figure 2.** Average walking trips per day per person for different purposes and for different SES groups over time in the CS scenario
CS, more non-household locations and lower safety level in the core

# Complexity & high-dimensionality with ABMs

- Many different inputs and possible outputs
- Too complicated to explore by hand
- Need ways to:
  - Visualize and interpret the variation in behaviors
  - Explore which parameters/inputs have the most effect on model outputs of interest

# Complexity & high-dimensionality with ABMs

- Also underscores the importance of being targeted and thoughtful in your question!

- Freedom of design means you can easily end up with way too many variables to analyze effectively

- Decide what scenarios to run (e.g. four city setups), what questions to answer, and then keep it as simple as possible

# High dimensionality of ABM

- Benefit: ability to explore relationships between large spaces of system inputs & system outputs

- Also allows for exploration of stochastic processes and evolution of systems across time and space

- However, very high dimensionality of data to deal with in analysis

# High dimensionality of ABM

| Level | Inputs/Parameters | Outputs |
|---|---|---|
| **Agent** | Initial distributions of agent types, attributes, preference structure, direction, number, density | Final distribution of resources, density, proportion in a given state, position |
| **Interaction** | Type of strategy, connection topology, adaptation | Strategy prevalence, network structure |
| **Environment** | Resources, type, layout | Resources, percent of environment in a given state |

# Sensitivity, uncertainty, & sampling

# Uncertainty quantification & sensitivity analysis

- **Uncertainty quantification (UQ)**: understand the variation in model outputs & behaviors

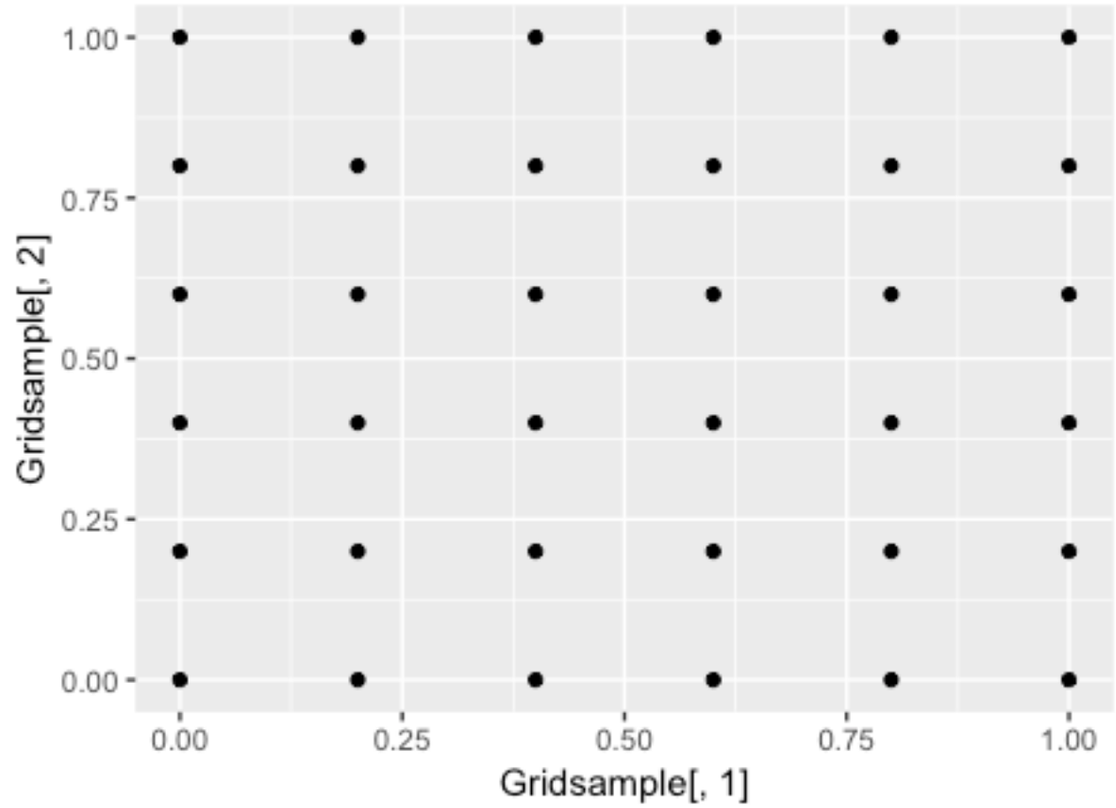
- **Sensitivity analysis**: understand how the inputs/ parameters affect the variation in output/behavior

# Exploring the space of model behaviors

- How to explore model behaviors?
- So far, we have focused on adjusting specific variables and parameters, and seeing how the model behavior changed by running the model a few times and visually assessing
- We need a more comprehensive, systematic approach

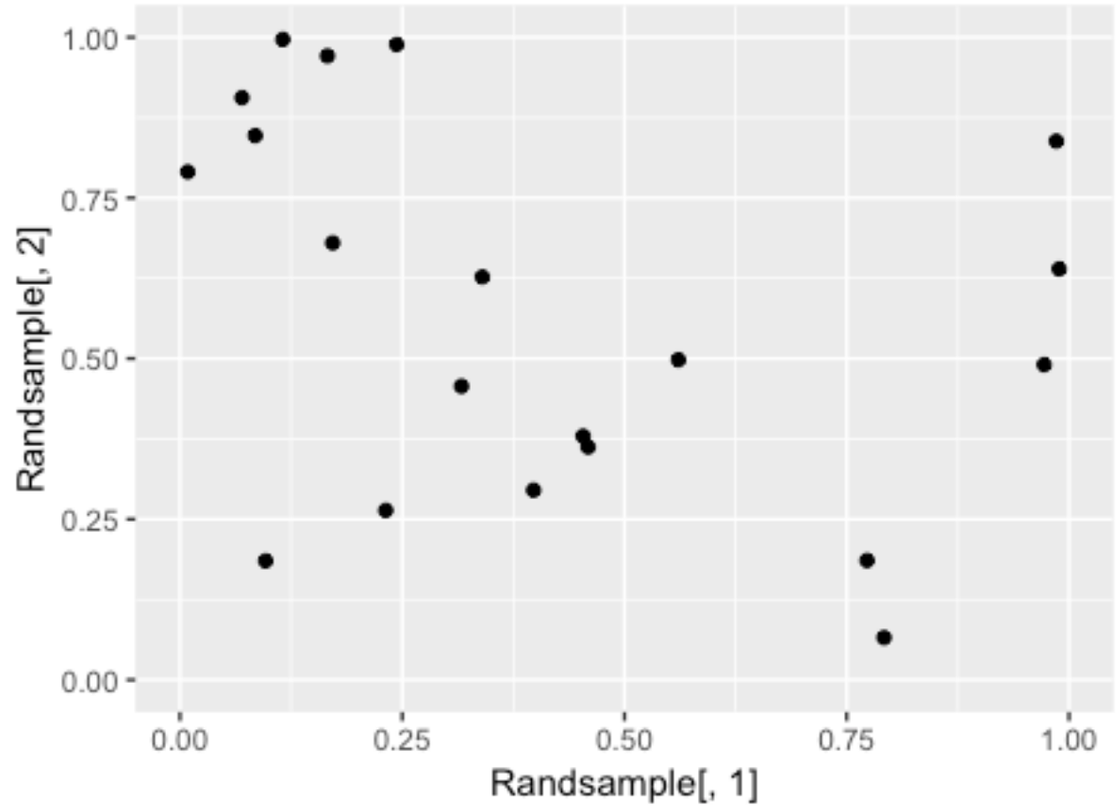
# Sampling parameter space

- **Grid sampling**
- Typically choose uniform distribution of points
- Good coverage of space
- Computationally expensive! Becomes infeasible as dimension increases



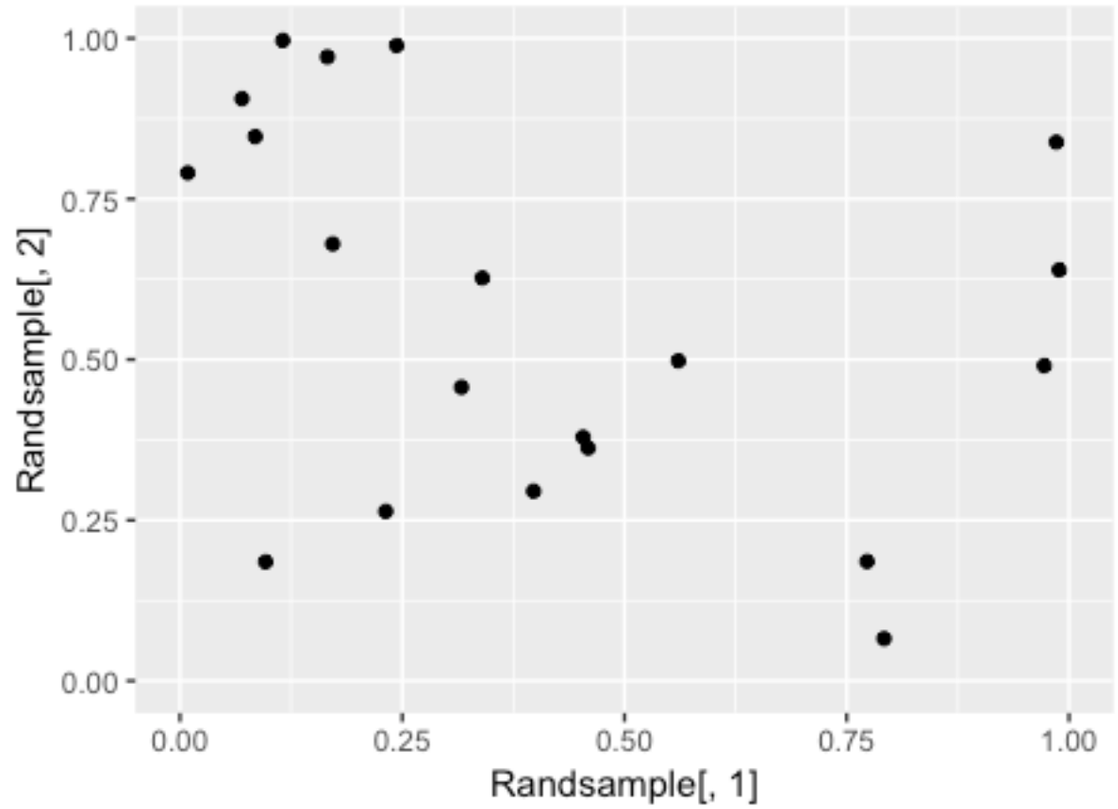
# Sampling parameter space

- **Random sampling** (Monte Carlo)
- Often done with uniform distribution, but can choose any distribution
- However, may leave big blank spots, require many samples to fully explore the space



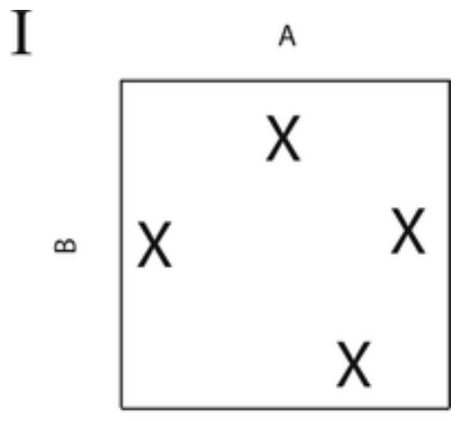
# Sampling parameter space

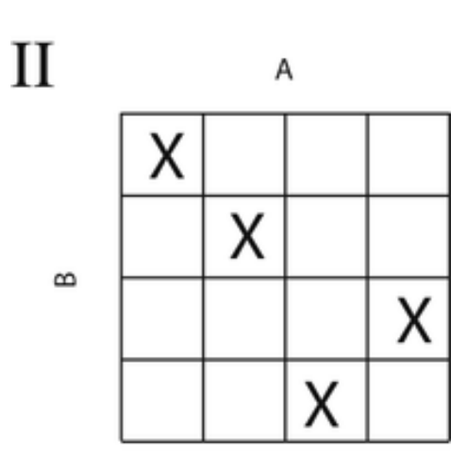
- More efficient ways to explore the space?
  - Latin hypercube sampling (& variants, orthogonal, etc.)
  - Sobol sampling (& other low-discrepancy sequences)

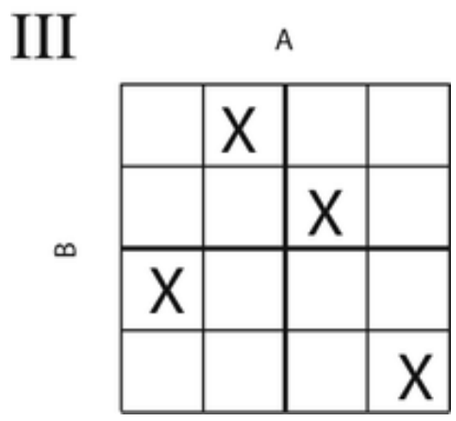


# Latin hypercube sampling

- Kind of like sudoku
- Divide space into a grid of rows & columns
- Choose one square in each row and each column
- Choose a random point within that square

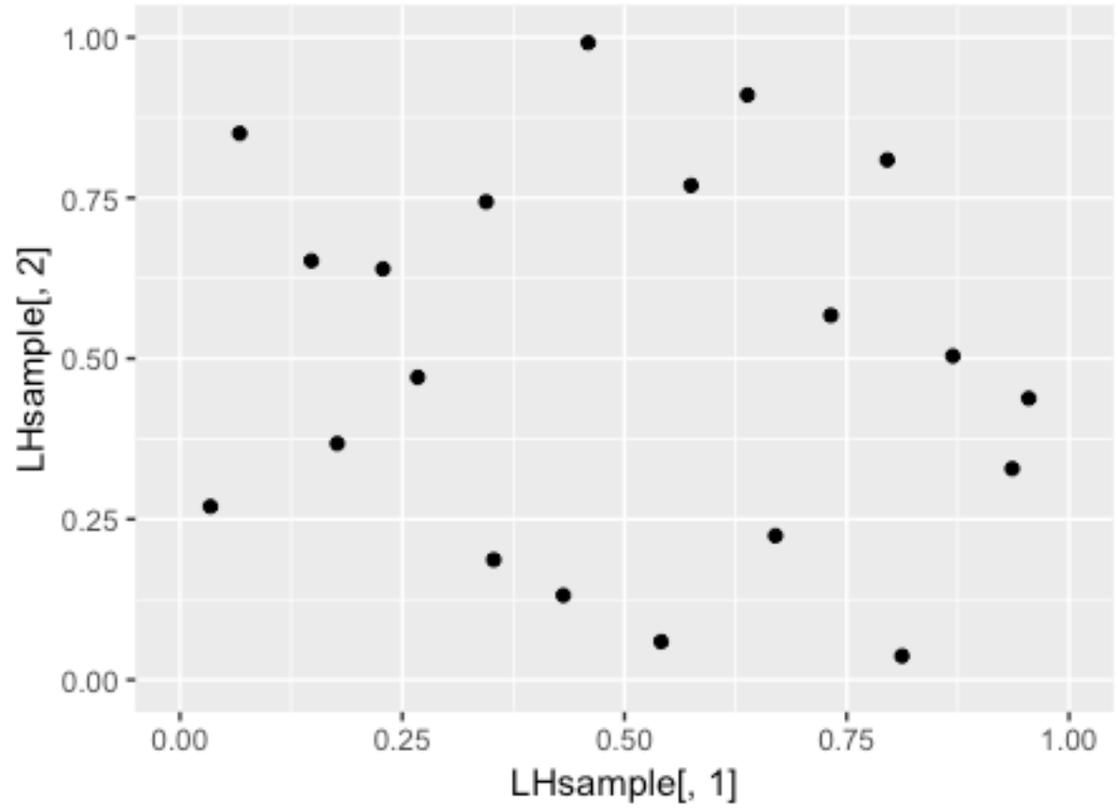


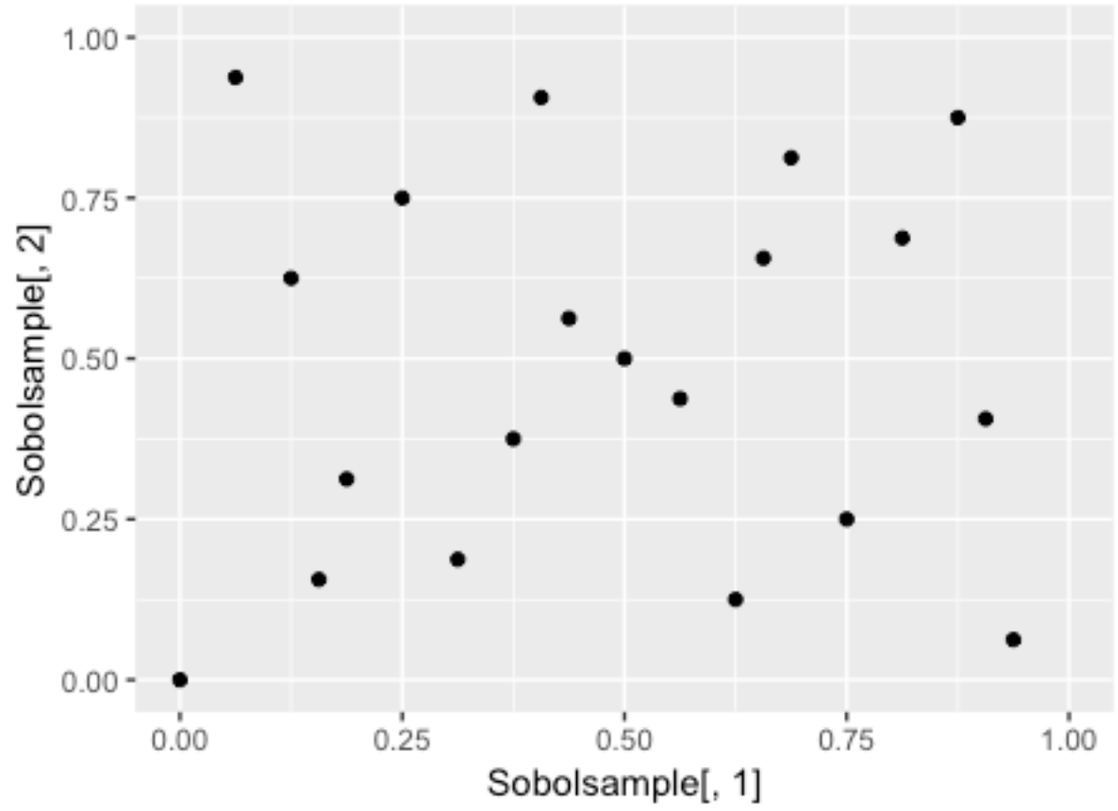
# Latin hypercube sampling

- Still an element of randomness
- Ensures better coverage of the space/faster convergence to the sampled distribution



# Sobol sampling

- Low-discrepancy sequence (see also Halton, Faure)
- Generates a sequence that samples the space evenly but requires few points
- Convergence can be better than LHS




Kucherenko et al., Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. https://arxiv.org/abs/1505.02350

# How many samples to take?

- Tough to say! Balance computation time with good coverage (often >>100 samples, e.g. in the 1K to 10K range depending on number of parameters)

- May need to run more than one sample for a given point due to stochasticity (since different runs may give different behaviors)

- For LHS, $N_S > (4/3) \times N_P$ has been proposed, but you will often want much more than this bound

- General guide: you want enough samples that further sampling won't change the outcomes of your analysis

Blower and Dowlatabadi (1994), McKay et al. (1979)

# Choosing outputs

- Given that we'll be running large numbers of samples, evaluating the model behavior visually will be difficult!

- Choose your model outputs based on your question of interest—be thoughtful about this!

- Can often be hard to pick metrics that accurately capture the behaviors you're interested in examining

# Visualizing & interpreting results

- Histograms and scatterplots—simple but useful!
- Boxplots, violin plots, ridgeline plots, heatmaps
- Basic descriptive stats (means, medians, SD, quantiles)
- We will explore other/fancier options as we go

# Uncertainty Quantification

# Epistemic vs. aleatory uncertainty

- **Aleatory uncertainty** - uncertainty due to randomness/ stochasticity/error, etc. Can be process-related noise, e.g. stochastic mixing of indivdiuals in a population, measurement error, e.g. variation in a given assay, etc. (comes from the Latin *alea*, meaning dice)/

- **Epistemic uncertainty** - systematic or model uncertainty, often characterized as things we could in principle know but do not—e.g. we might not know the true underlying process for a system, resulting in multiple alternate models (or potentially a need to measure further)

  - How do parameter values fit in here?

# Example: Uncertainty Quantification

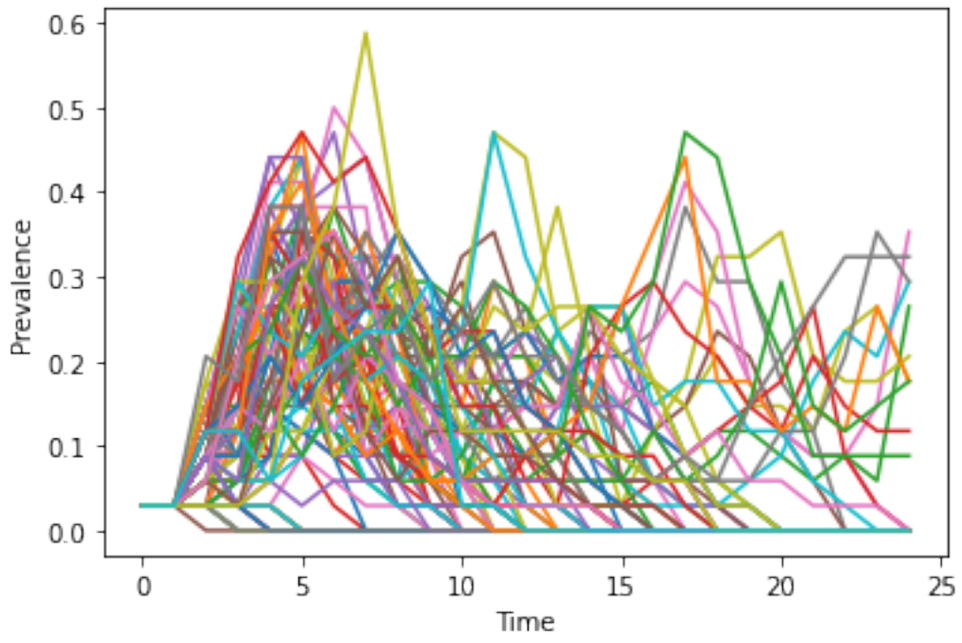
- **Model**: SIS on the karate club network, with edge breaking and regrowth

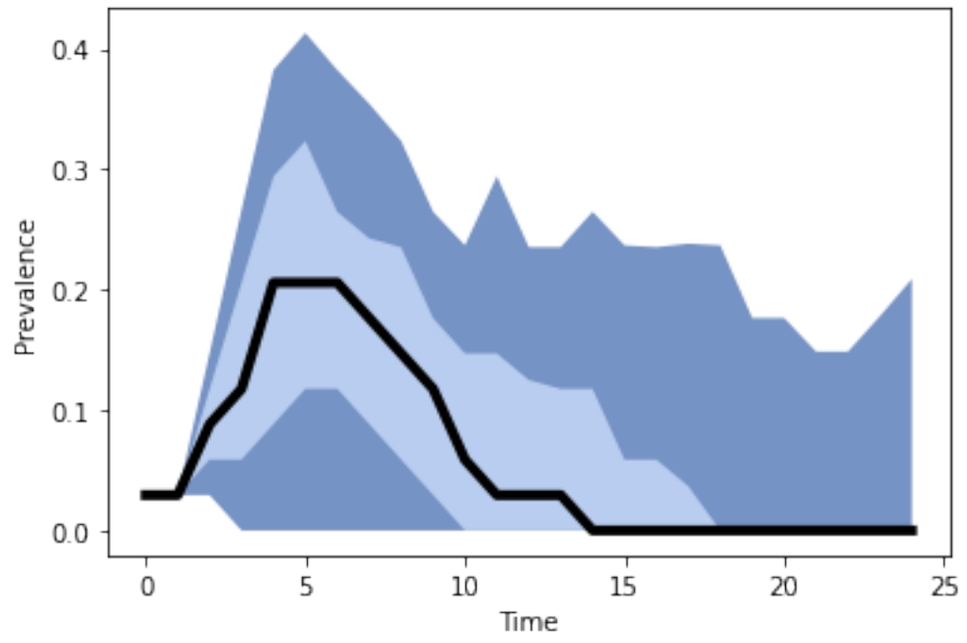
- **Parameters to sample**: infection probability, recovery probability, tie-breaking probability, tie-regrowth probability

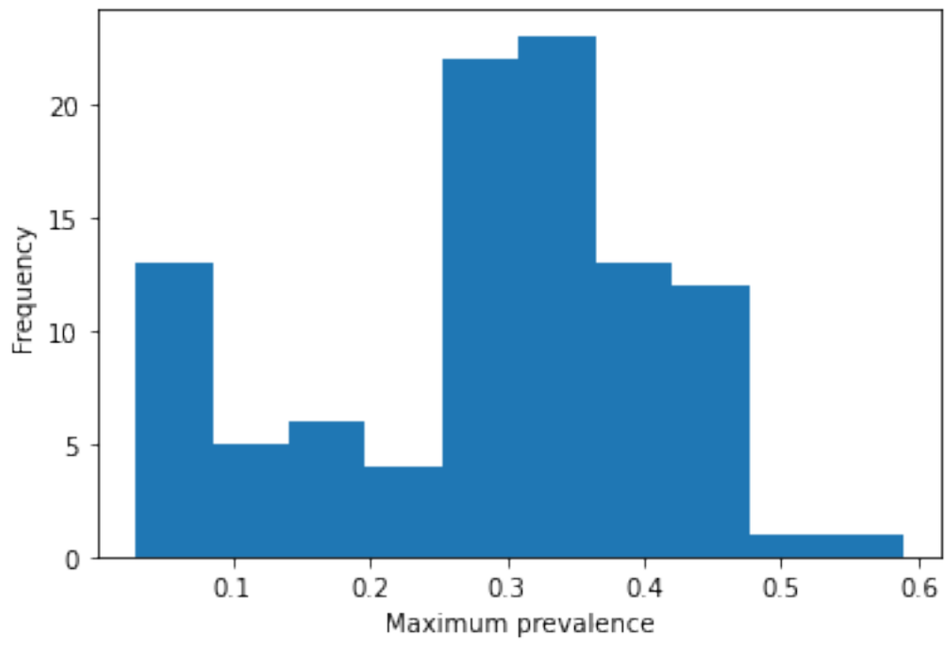
- **Outputs**: maximum prevalence (fraction infected), duration of the epidemic (within the 25 timesteps simulated)

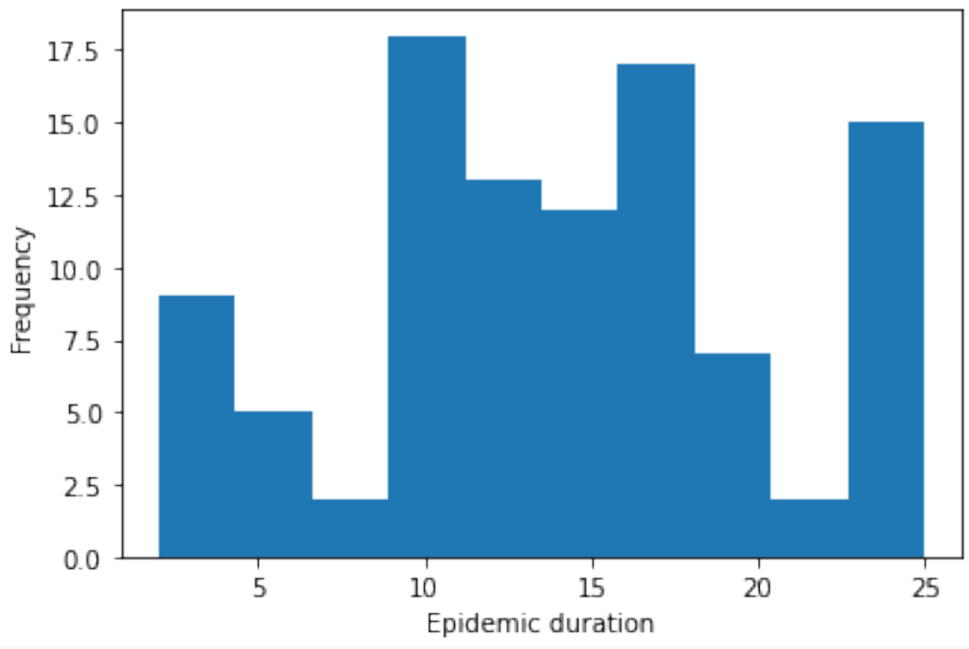
- Pause to work with code

# Stochastic variation only



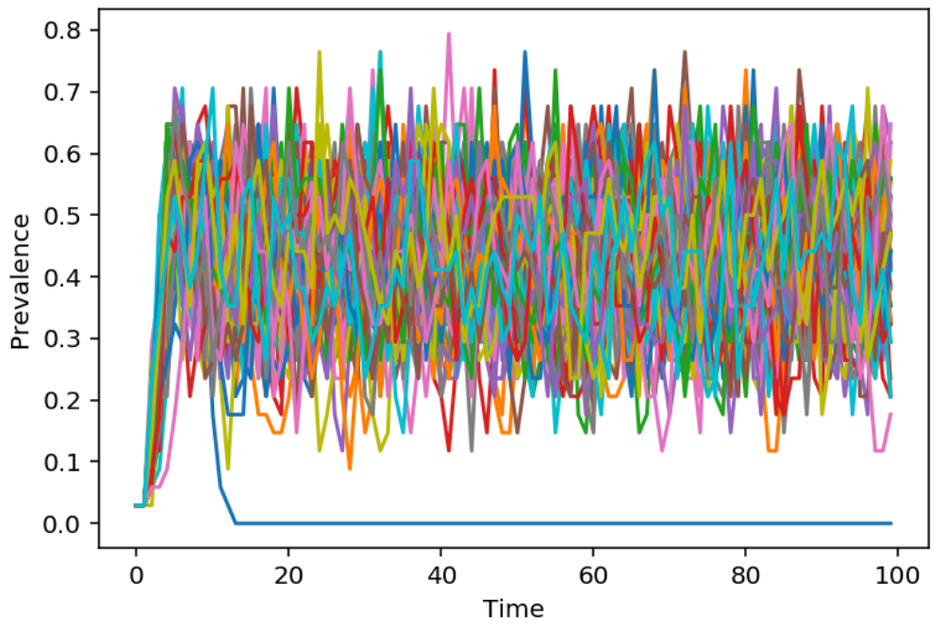



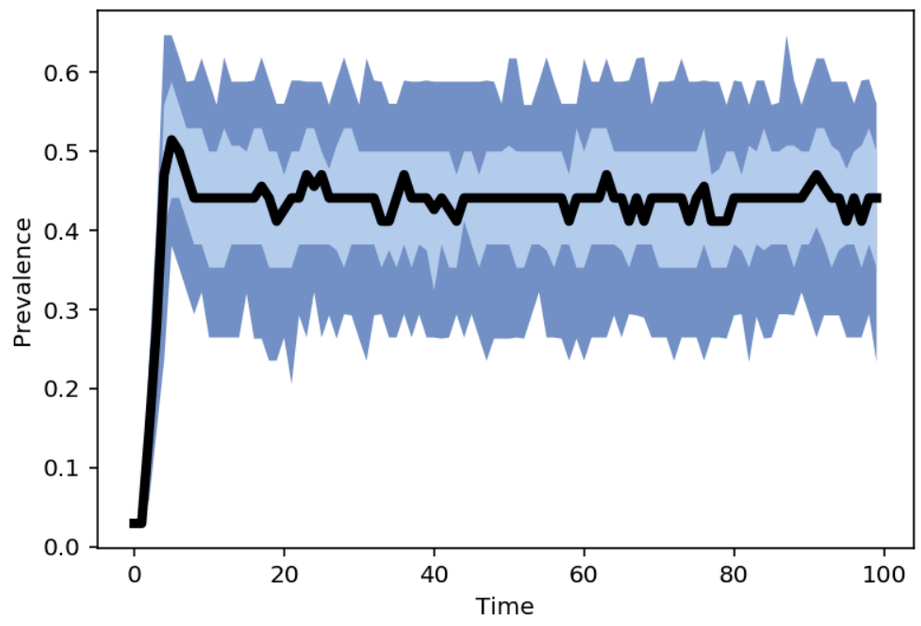



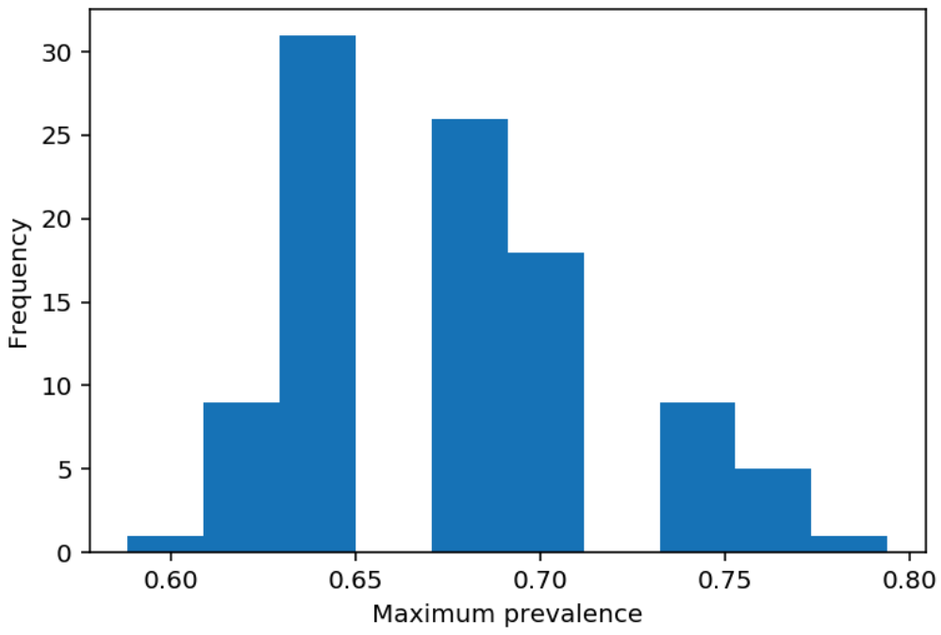



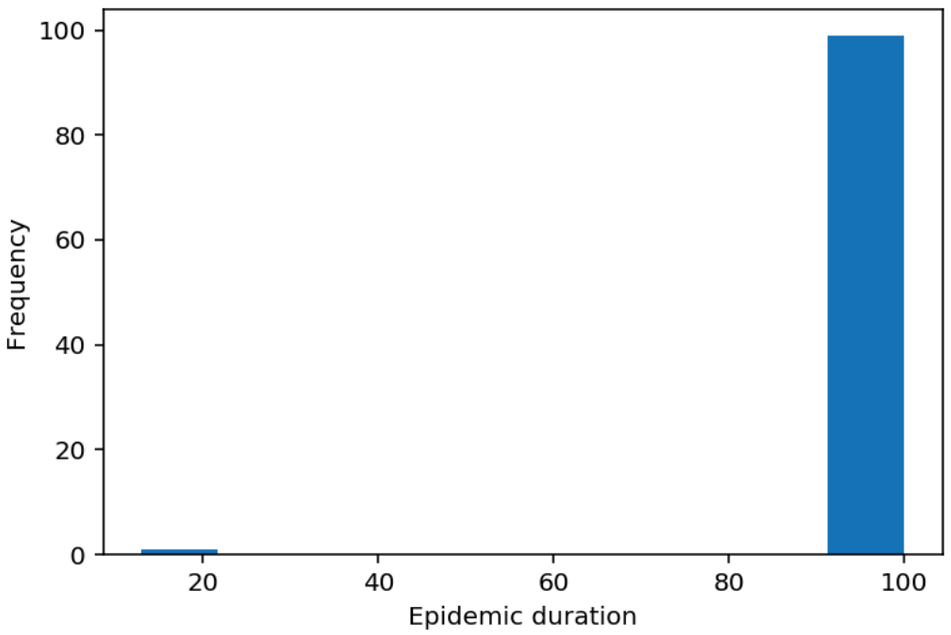

```
p_i = 0.3 # infection probability per contact
p_r = 0.5 # recovery probability
p_0 = 0.75 # probability of breaking tie if infected neighbor
p_1 = 0.25 # probability of regrowing tie if non-infected
```

# Stochastic variation may be different in different parts of parameter space



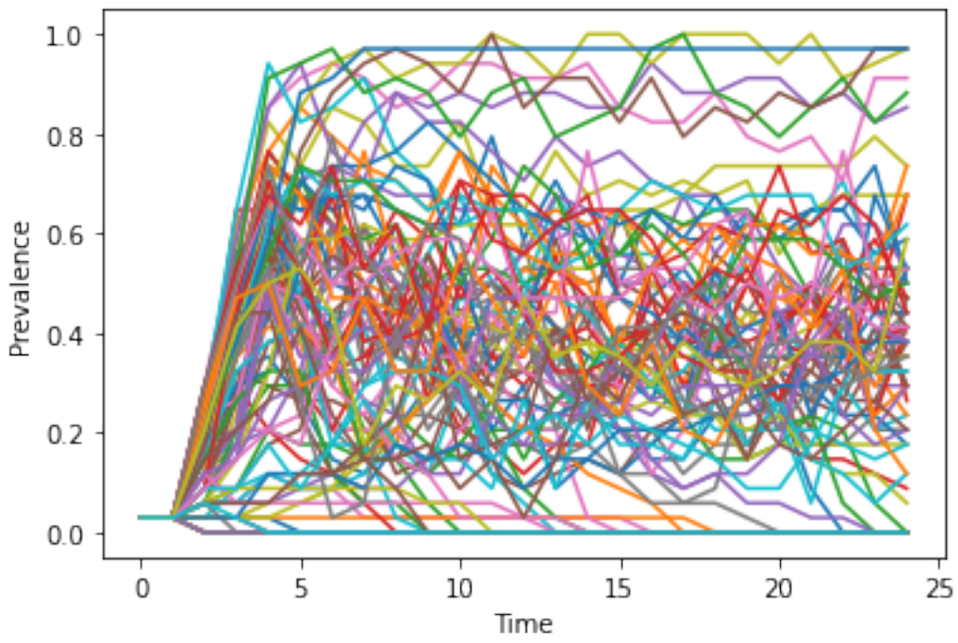



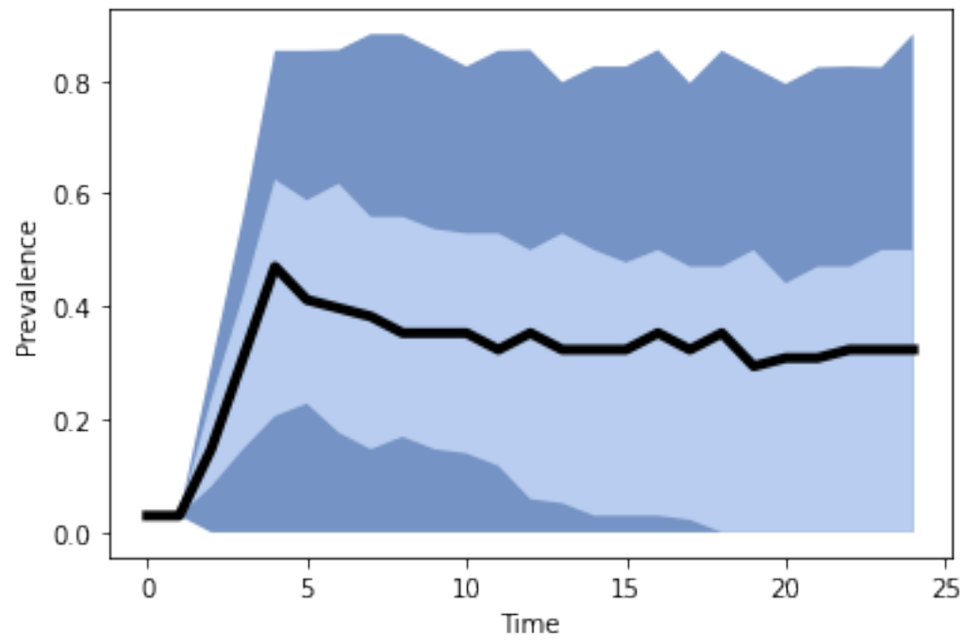



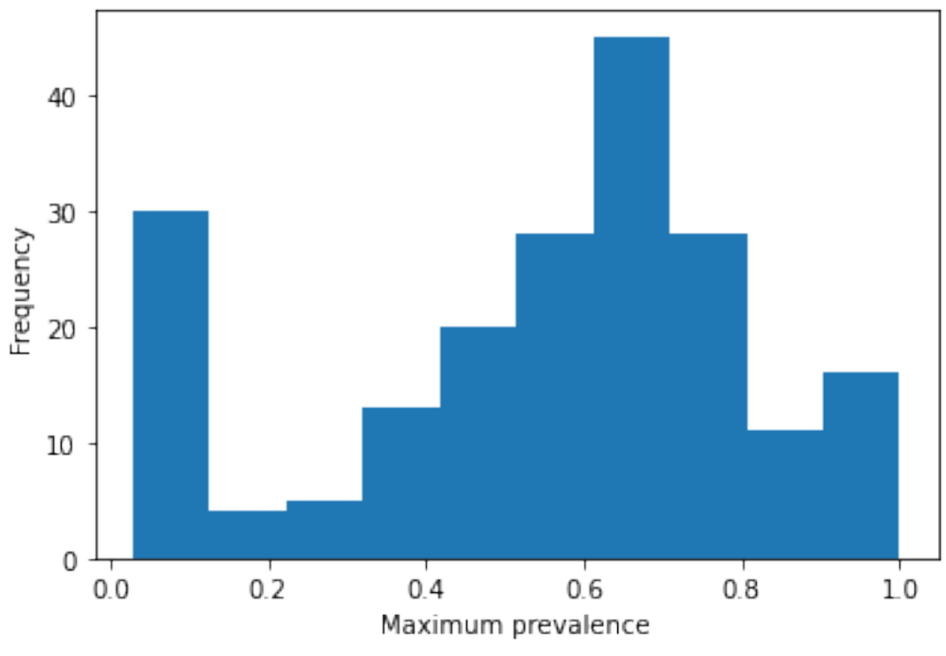



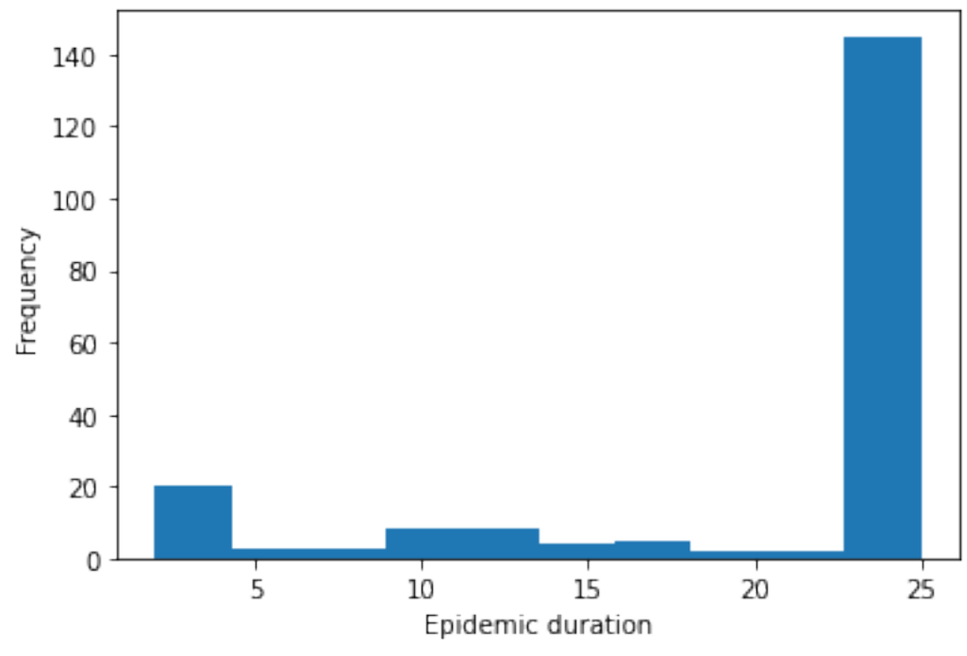

```
p_i = 0.5 # infection probability per contact
p_r = 0.5 # recovery probability
p_0 = 0.5 # probability of breaking tie if infected neighbor
p_1 = 0.5 # probability of regrowing tie if non-infected
'''
```

# Sampling parameter space: parameter and stochastic variation

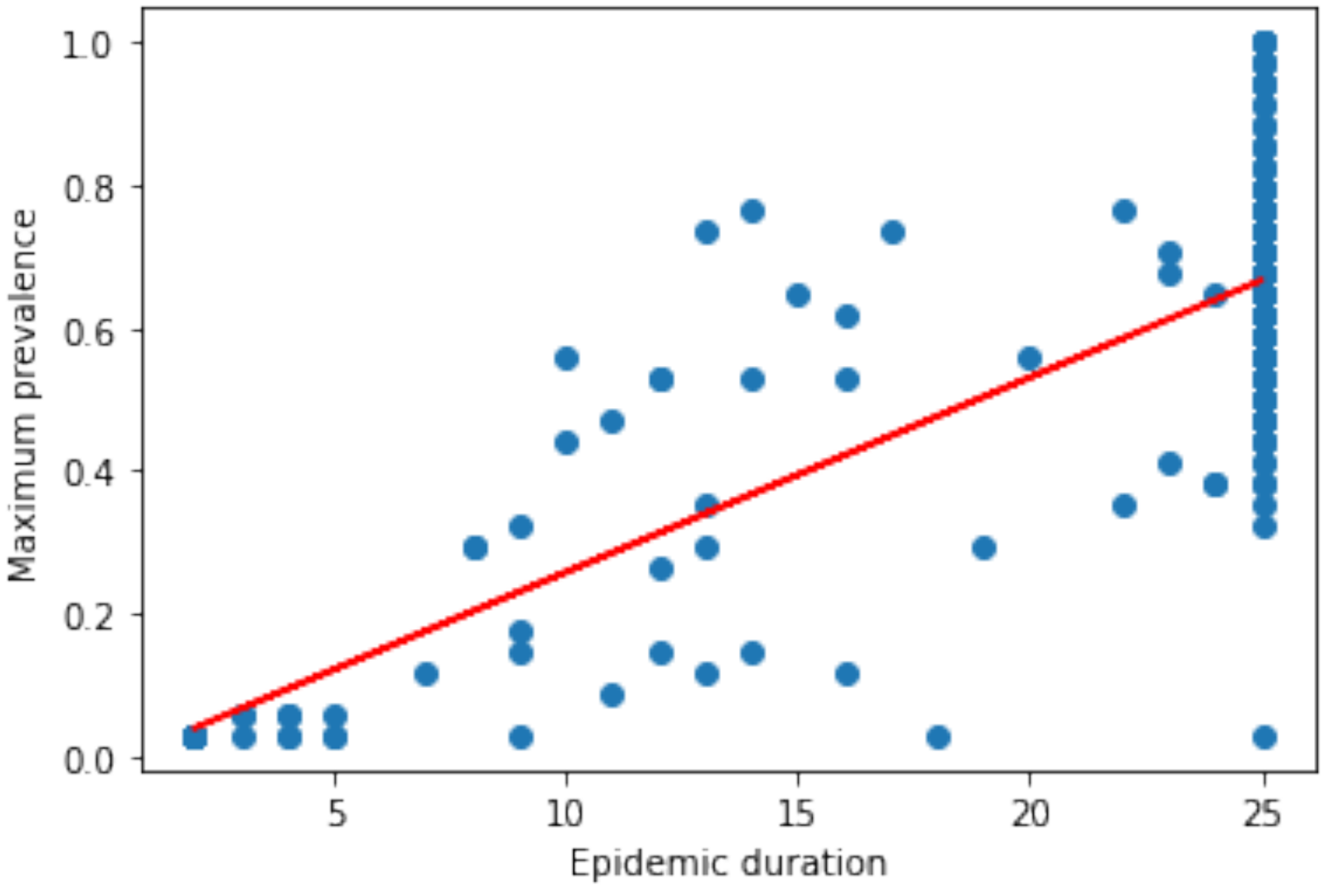


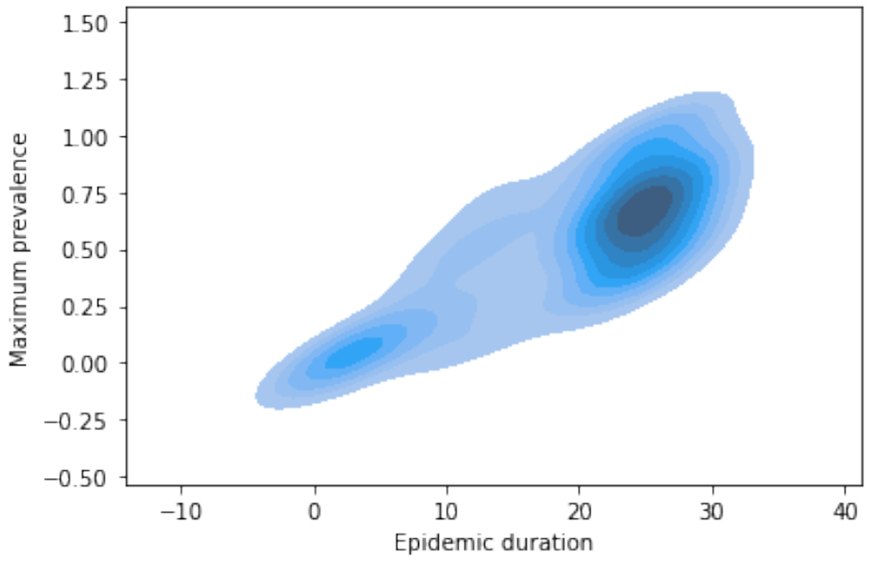



# Sampling parameter space



# For next time…

- Readings
- PARTE framework
- ODD protocol
- Example ABMs also posted on the website