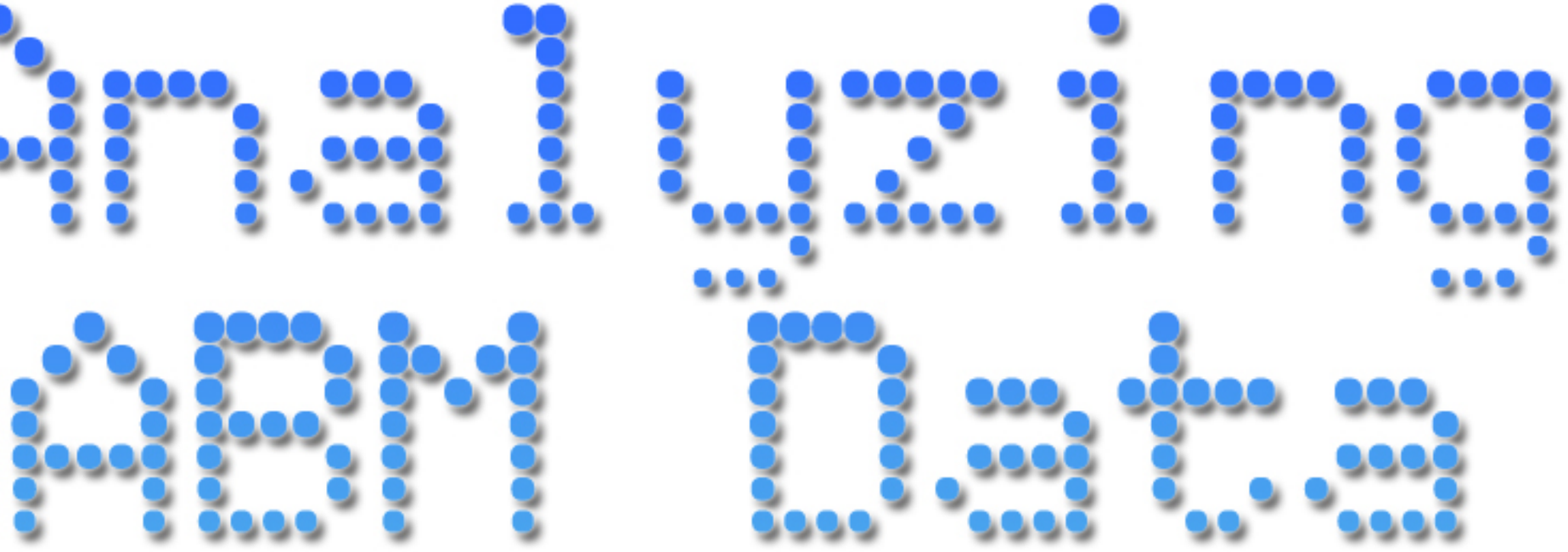
Introduction to Agent-Based Modeling in Netlogo

Aaron L Bramson



Setting and Getting a Random Seed

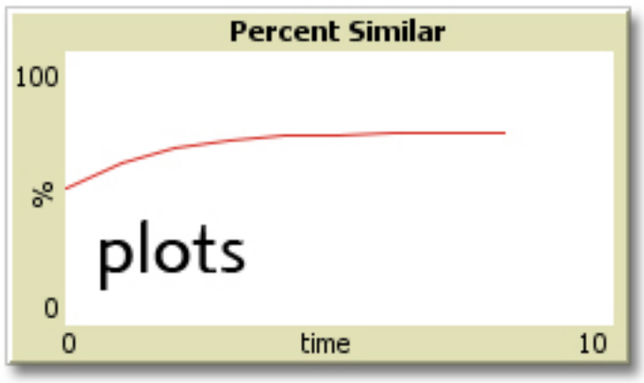
- If you want to reproduce a run, you need to know the random seed.
- You can set the random seed yourself: `random-seed #`
- You can get the next random seed Netlogo picks: `new-seed`
- Code in monitors and sliders do not affect the random numbers.
- More info in the Programming Guide and Code Example.

Data Analysis for Agent-Based Models

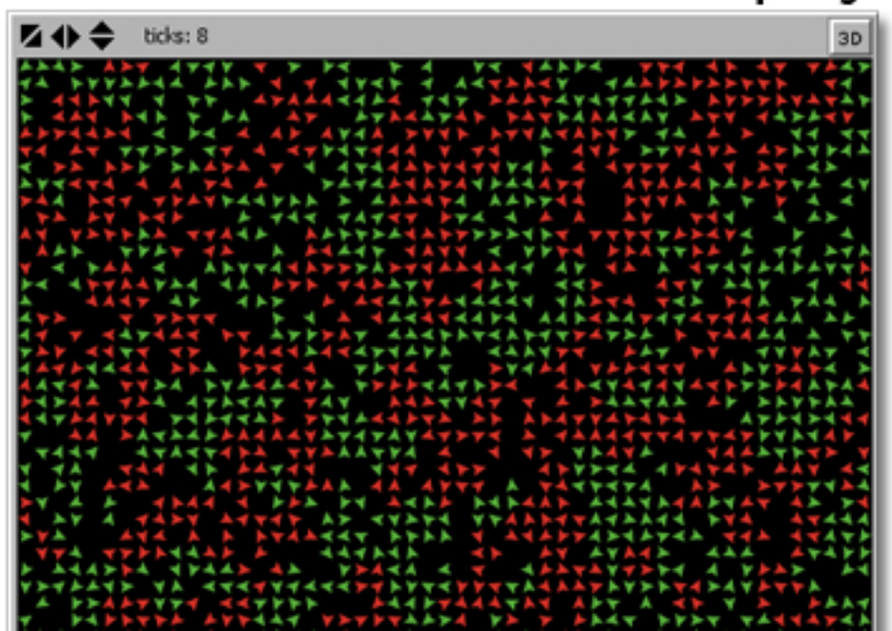
- First stage is immediate model feedback (display, plots, monitors).
- Data collection using Behavior Space.
- Data plotting and analysis in other software (e.g., Excel, Stata).
- Statistical considerations for ABM data.
- Informative comparisons and not just summaries.

Getting Data from Models

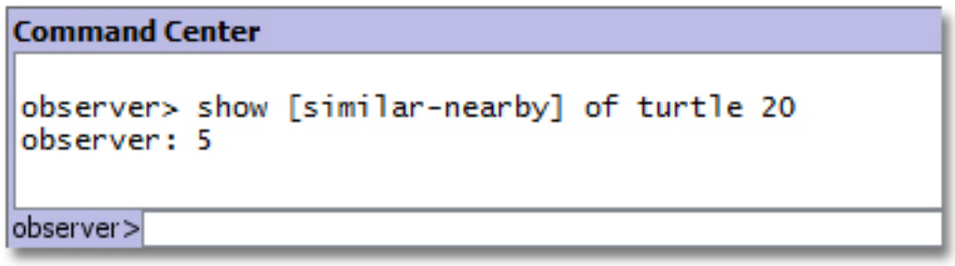
- Runtime Feedback:



display



command center



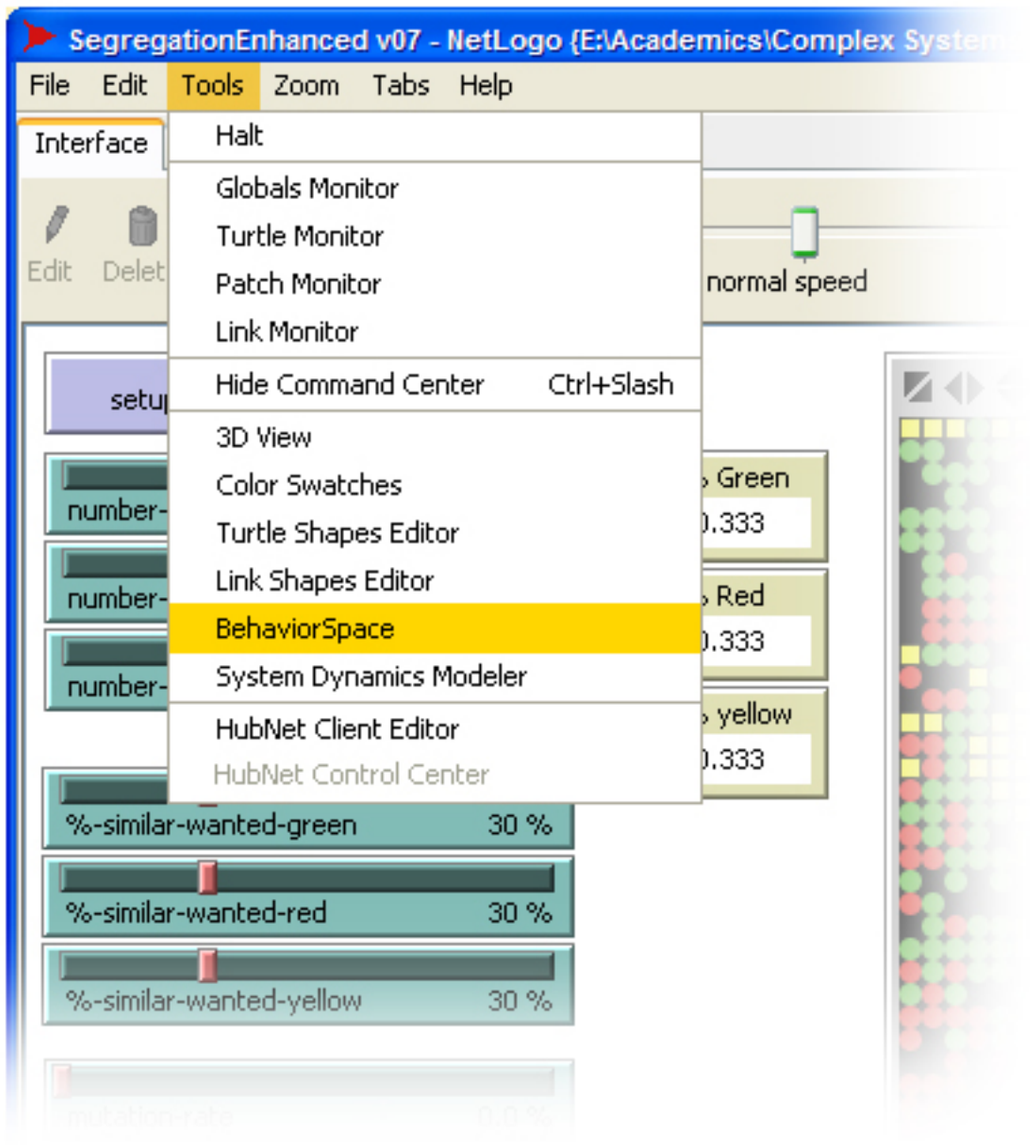
Getting Data from Models

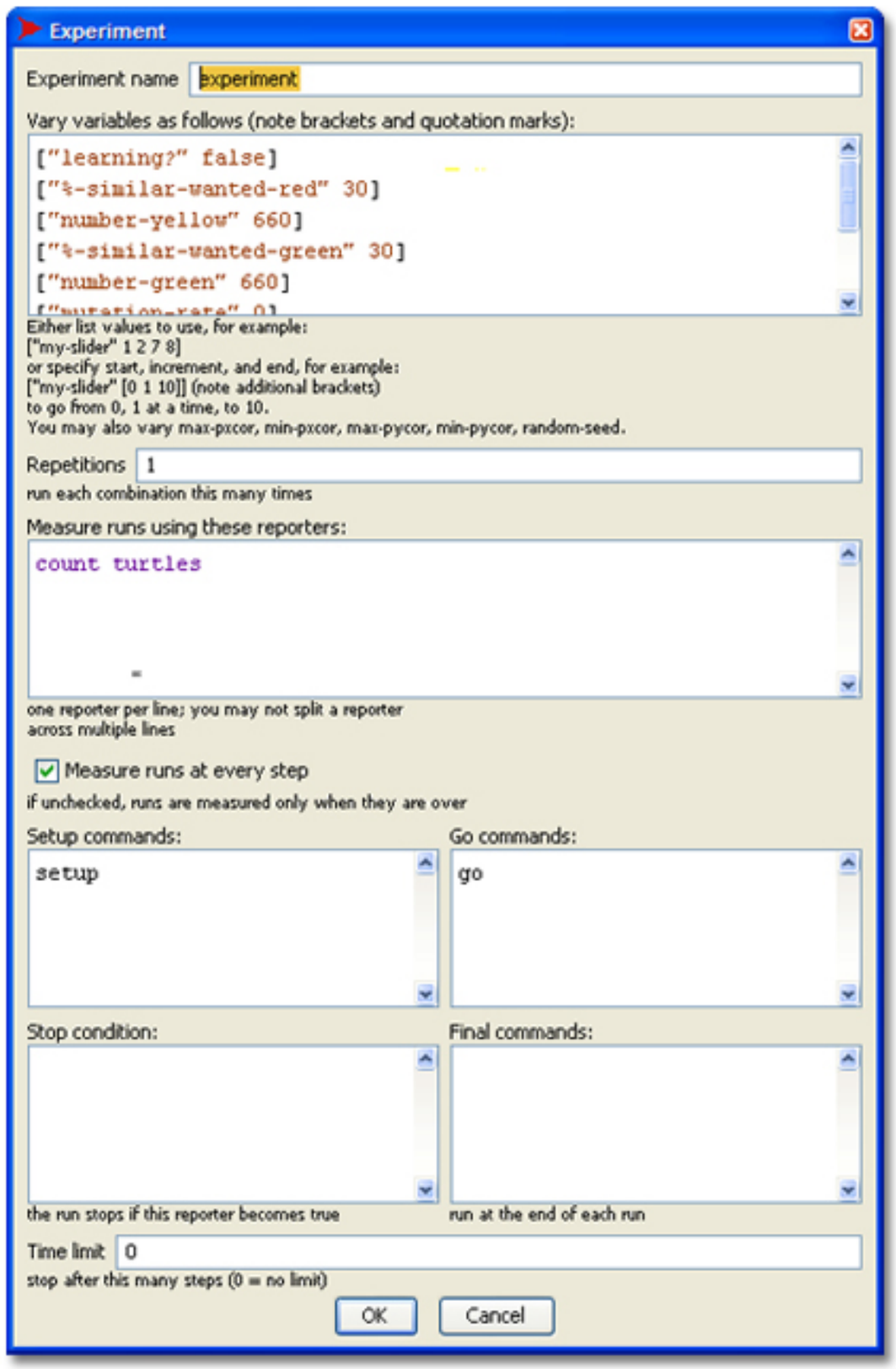
- Using the Behavior Space Tool for "drone" runs:

- Sweeps Variables Setable from Interface
- Can Run Different Experiments on the Same Model
- Specify How Many Repetitions for each Variable Combo
- Specify Output Variables
- Specify Setup, Go, Stop and Post-Run Commands
- A Time Limit Can Be Specified Separately

- NOTE: for Excel specify "Table" format for output

Getting Data from Models





Getting Data from Models

- How much data should you collect? As much as you can!

- The more runs you do for each parameter combination the the more characteristic your sample is and the more confident your analysis will be on measuring the features.

- The more parameters combinations you run the more robust your analysis and the more likely you'll find interesting features.

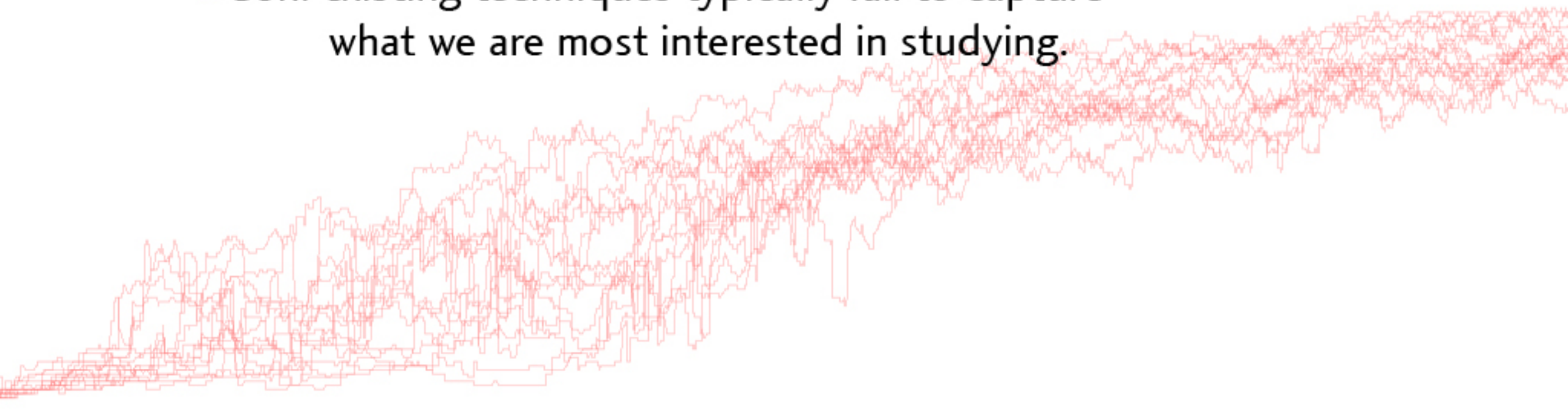
- Recording at every time allows you to measures dynamics rather then just outcomes.

Getting Data from Models

- Constrained by the runtime of your model.
- Constrained by the time to analyze the data.
- Constrained by you ability to analyze data.
- Informed by your purpose for collecting the data.
- Informed by how you are going to use and present the data.

Data Analysis for ABMs

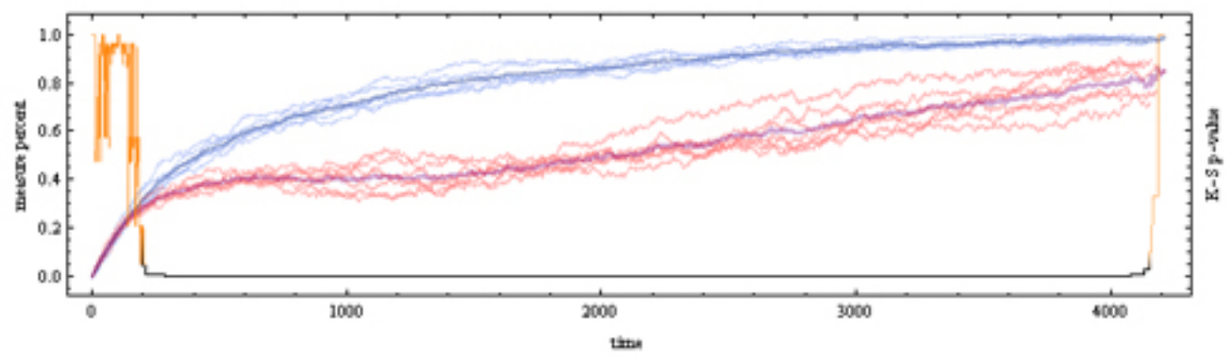
- Agent-based models are like empirical experiments.
- The output is time-series observational data.
- Pro: there already exists a plethora of techniques to analyze this sort of data.
- Con: existing techniques typically fail to capture what we are most interested in studying.

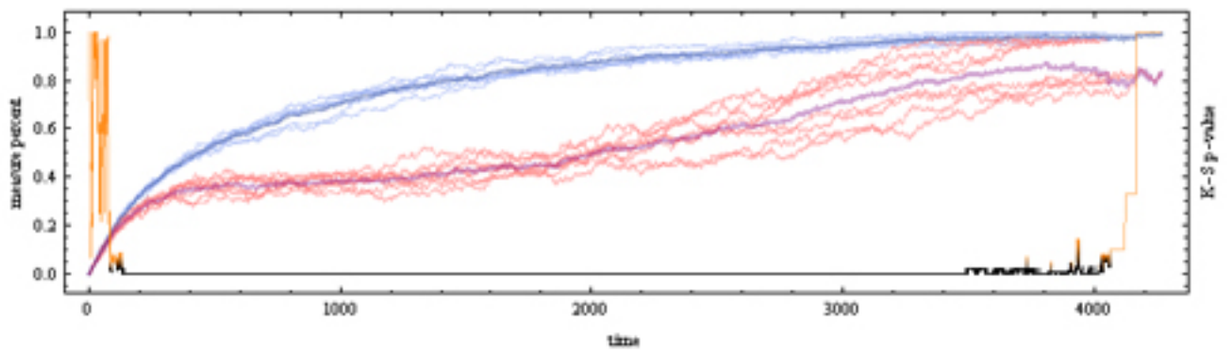


Data Analysis for ABMs

- Before you know what analysis is appropriate for your data you may need to do some exploratory analysis.

- Visualizations are often helpful guides to where to look closer.





Data Analysis for ABMs

- Don't Aggregate Your Data.

- The mean is useful to smooth out random perturbations in your data, but variation in ABM output isn't just random.

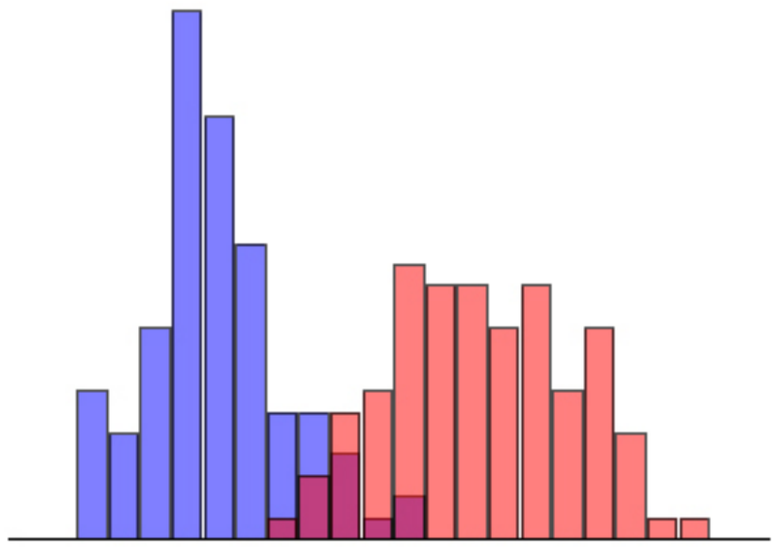
- If there were good statistical models for complex systems, then you wouldn't have needed to build an ABM.

- Simplifying to "representative" cases or averages throws away data critical to understanding model complexity.

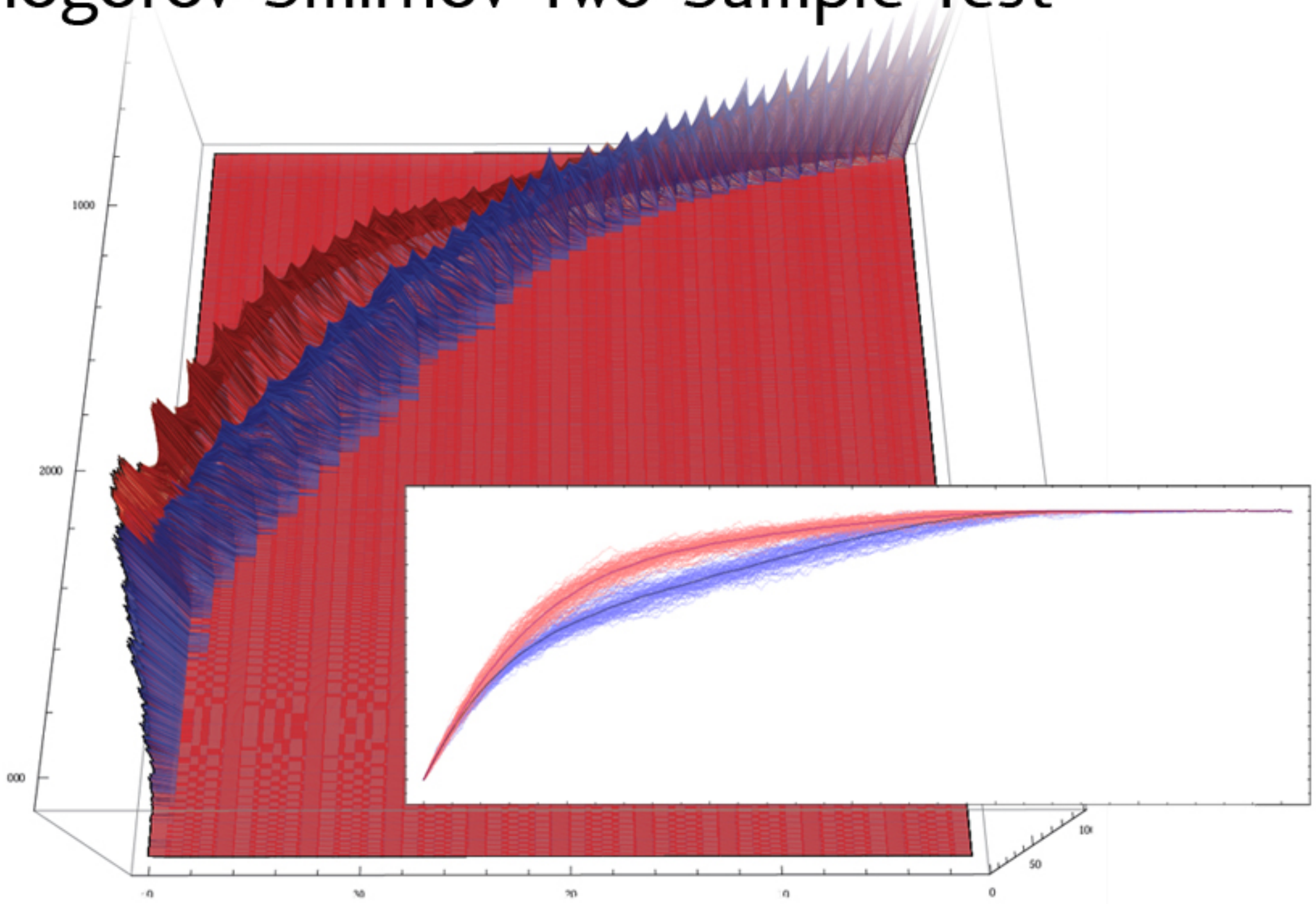
- Instead, use visualization and descriptive techniques that utilize all your precious output data.

Kolmogorov-Smirnov Two-Sample Test

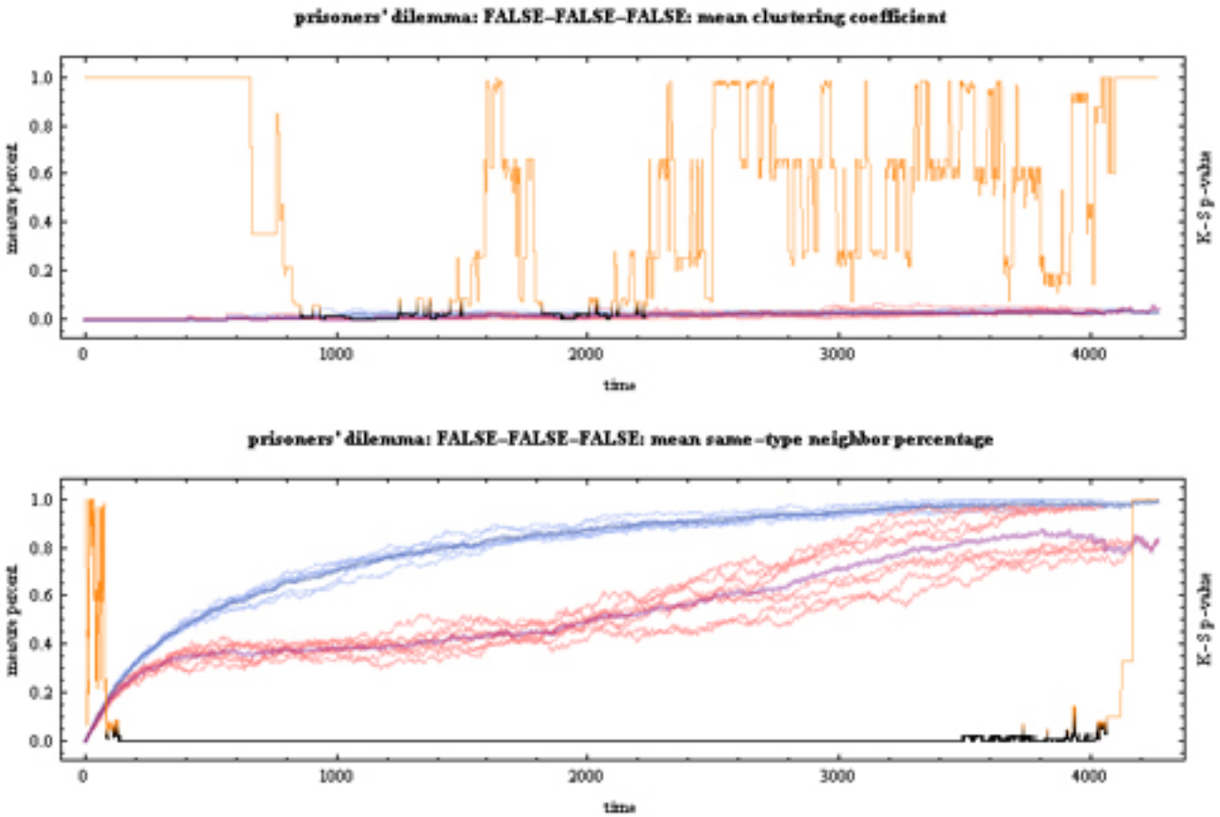
- Nonparametric statistical technique.
- Generates the empirical cumulative distribution functions from each data sample.
- Compares each point from each sample to the other sample's distribution to determine membership probability.
- Yields the probability that the two samples could have been generated from the same continuous distribution.



Kolmogorov-Smirnov Two-Sample Test

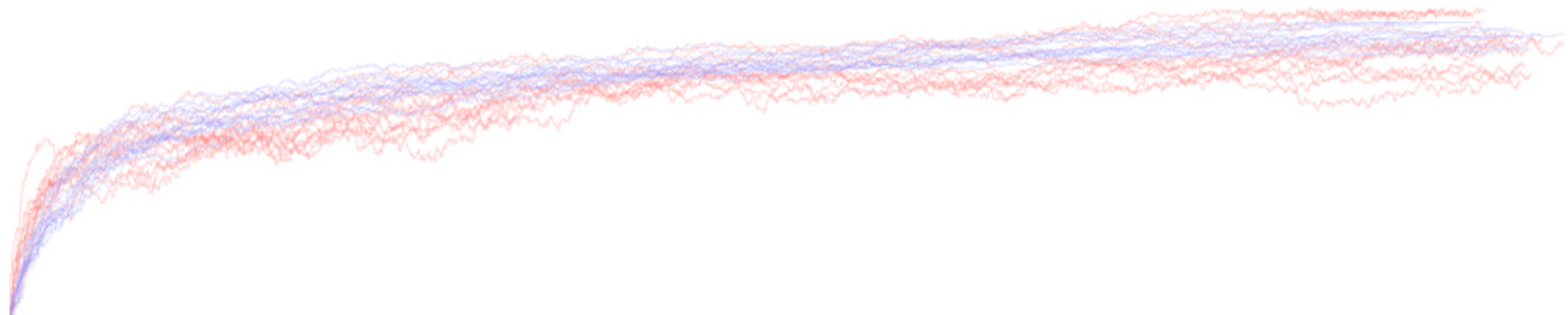


Kolmogorov-Smirnov Two-Sample Test



Siegel-Tukey Test of Differences of Variances

- Ranks values from two samples by how extreme they are.
- Sum the ranks within each sample.
- Lower score is more extreme (highly variable).
- Higher score means values clumped near collective median.
- Nearly same score (accounting for size) means no difference.



Compare Variables over Time

- For validation, verification, and calibration compare the same feature in ported/docked models or empirical data.
- Same characteristic of two or more populations in a model.
- Effect of a parameter change on a model variable, e.g. sensitivity through parameter sweeps.
- And of course there are more tests, measures, and applications already available and under development.

Data Visualization

- Even given the same data, the presentation of the data is important.
- Excel line plot with default options: