Complex Systems 530: Computer Modeling of Complex Systems

Lecture 4: Case Study - the Artificial Anasazi Project
1/21/20

Today

- Case study: the Artificial Anasazi Project
- Continue Lab 1
  - Ants model in PyCX

Artificial Anasazi Project

- Model of the **Ancestral Puebloan** people
- Developed in 2002 (original version 1998)
- Example of a more complex, empirically-based model
- Illustrates a few key parts of ABM analysis
  - Parameter space exploration, sensitivity analysis, parameter sweeps, calibration
  - Replication, verification & validation

Axtell, Robert L., et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." *PNAS* 99.suppl 3 (2002): 7275-7279.

Parameter space exploration

- Models (ABMs and otherwise) can have a wide range of behaviors depending on the initial values, parameter values, etc.

- Importance of exploring parameter space - random sampling, parameter sweeps

- Parameter sensitivity analysis

- Model calibration - tuning, parameter estimation, optimization

Verification & Validation

- **Verification**: confirming the mapping between your conceptual model and your computational model
  - Relates to replication, debugging, documentation
- **Validation**: evaluating the match/connection between your model and reality

Replication

- Replication studies are more important (and difficult) than you might expect—particularly for more complex models such as ABMs
- Many examples of difficulties in recreating modeling results
- Code availability
- Critical to have clear & complete documentation - of both code and the model as a whole

Models & Archeology

- Archeology has advantages in studying long-term patterns of human behavior due to having data over very long periods of time
- However, difficult to conduct reproducible experiments and test potential explanations
    - Also true for many other fields - astronomy, geophysics, public health (e.g. disease epidemics), paleontology, etc.

Modeling counterfactuals

- Modeling provides a way for researchers to “rewind and rerun the tape of history”
- Creates an *in silico* “laboratory” with which to test hypotheses, explore counterfactuals, and adjudicate between competing explanations

ABMs & Archeology

- ABM in particular also allows the for the development of “generative” explanations
  - “If you didn’t grow it, you didn’t explain it” (Epstein 2006)
- It also allows for explorations of complex interactions between physical and social landscapes, as well as an ability to explore the role of heterogeneity

The Challenge of Validation

- A challenge of this use of ABMs involves validation & connecting with data
    - Potential links at both the micro and macro levels - initial conditions, parameters, model behaviors
- Often involves a substantial increase in model complexity and calibration of large parameter spaces
- In best case, big payoff though: parsimony of explanation and adjudication between hypotheses

Artificial Anasazi Project

- Model of the Ancestral Puebloan people
- The AAP grew out of Sugarscape (Epstein & Axtell)
- Gumerman, Dean et al. had extensive archeological database for the region

Long House Valley

- 96 km^2 area region of northeastern Arizona
- Inhabited by the Ancestral Puebloan people starting 1800 BCE
- Depended on maize cultivation (starting around 200 CE)
- ~1300 CE major population collapse
- Rich data on paleoenvironment and archeological record of settlement patterns and insights into social life

Fig. 1. Long House Valley, looking to the South.

Axtell, Robert L., et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." *PNAS* 99.suppl 3 (2002): 7275-7279.

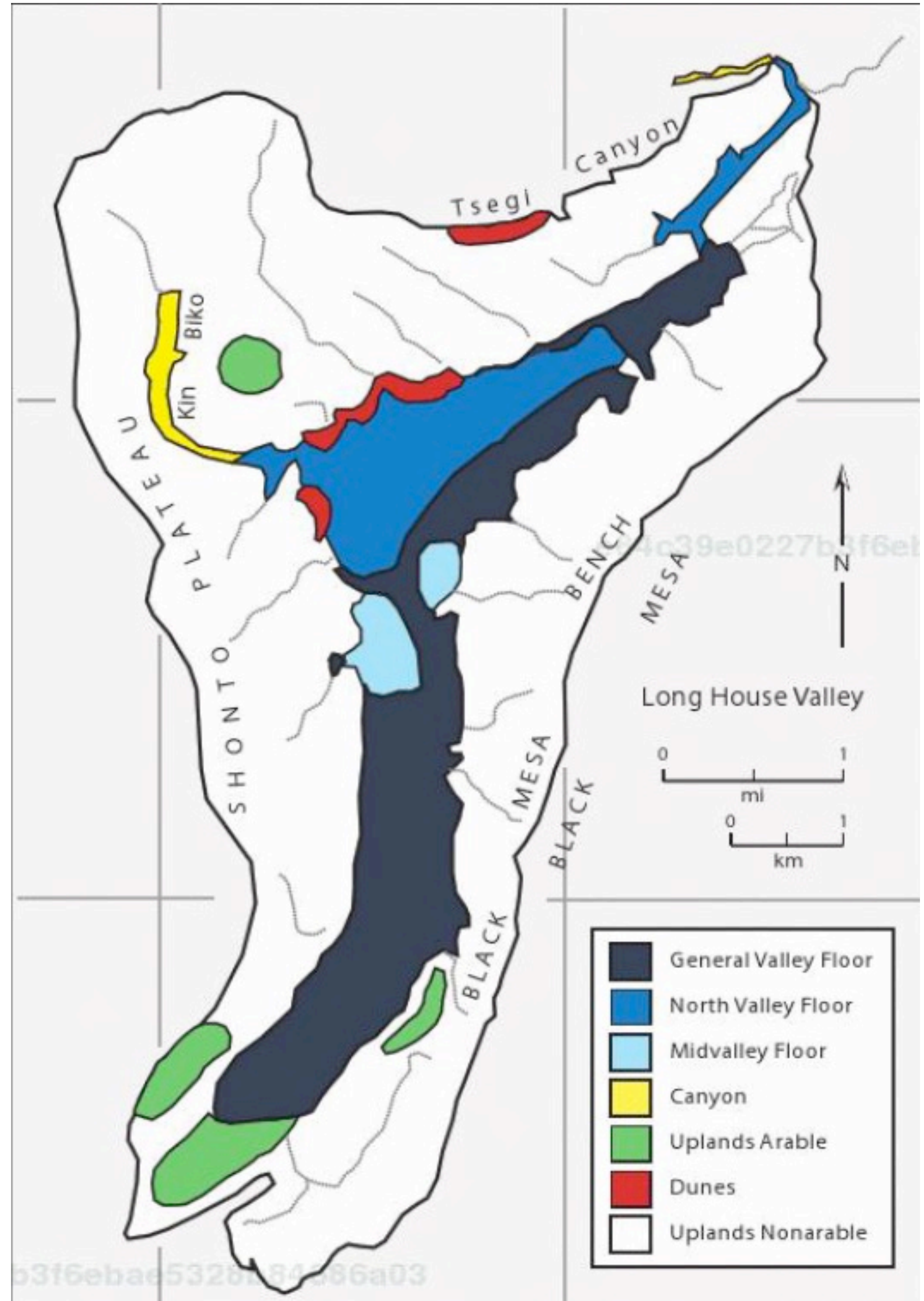
Artificial Anasazi Project

- Goal: Build an empirically grounded ABM using environmental record and “anthropologically feasible” social rules to replicate the spatiotemporal settlement history of Long House Valley between 200 – 1300 C.E.

- Generate insight into the cause of the population collapse?

AAP: the Environment

- 80 × 120 grid cells each representing 100 m x 100 m
- 7 zones of land, with different levels of agricultural productivity
- Focus on annual maize yields by region
- Input data on paleoenvironmental variability



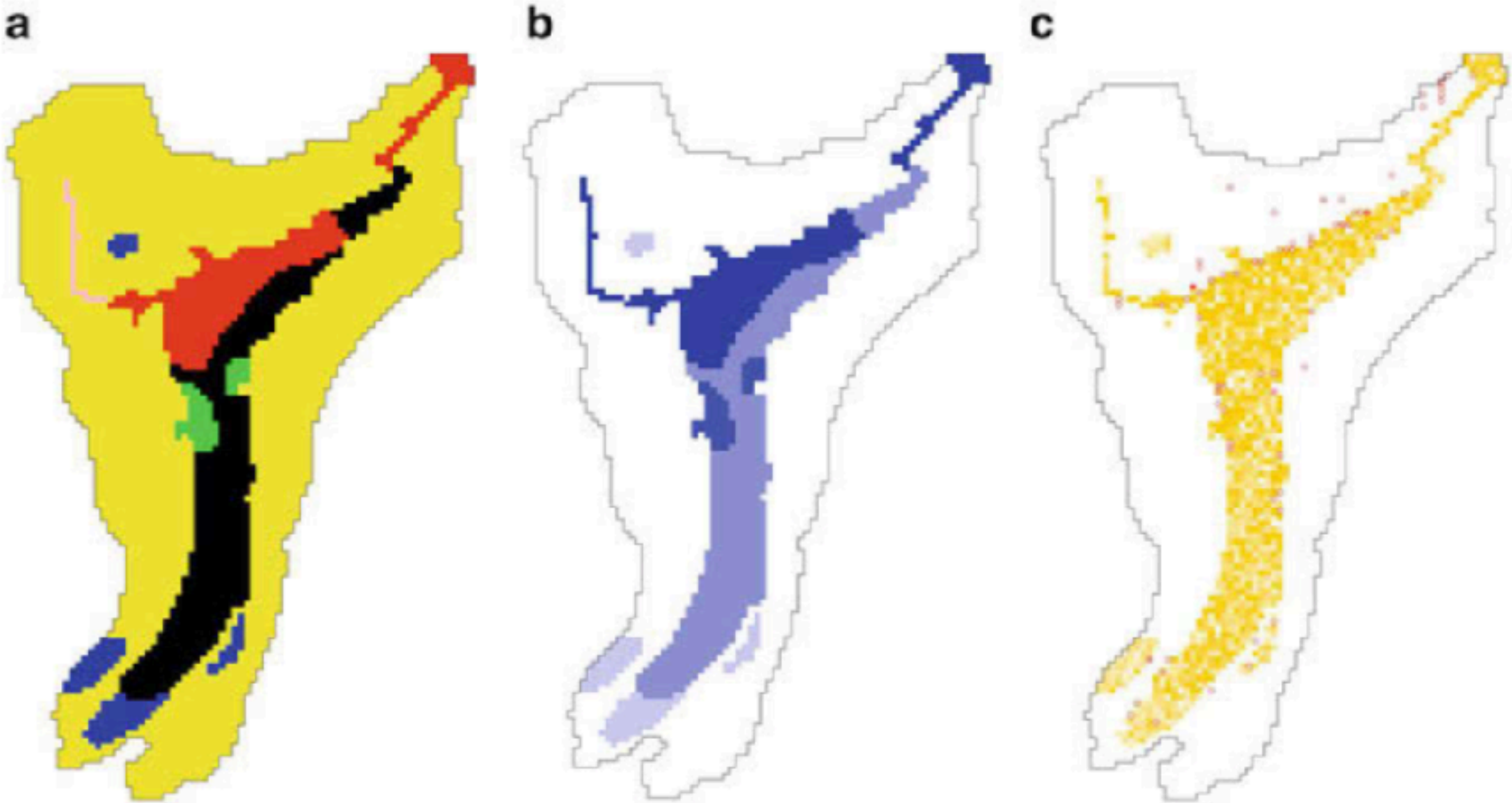


Fig. 2.2 (**a**) Environmental zones. (**b**) Hydrology. (**c**) Plot yields and historic settlements

Swedlund, Alan C., et al. "Modeling archaeology: origins of the artificial Anasazi project and beyond." *Agent-based modeling and simulation in archaeology*. Springer, Cham, 2015. 37-50.

AAP: the Environment

- Estimates of maize yield in regions based on range of factors including:
    - Adjusted Palmer Drought Severity Index (PDSI)
    - Rise and fall of groundwater
- Overall: $BY = y \times q \times H_a$
- where y(t) = yield based on data, q = soil quality, Ha = harvest adjustment (scaling parameter)

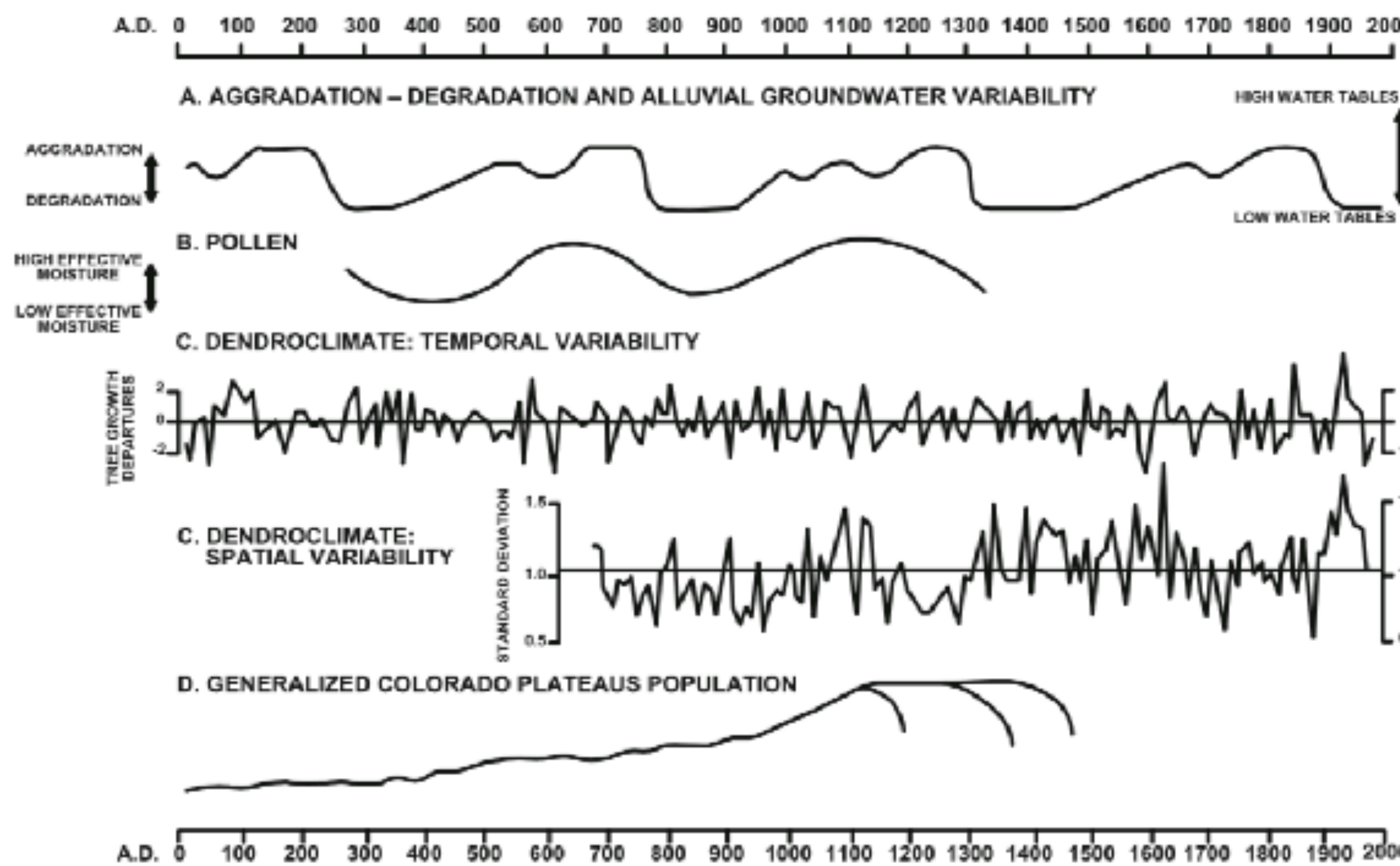


Fig. 2.1 Data used in the Artificial Anasazi model

Swedlund, Alan C., et al. "Modeling archaeology: origins of the artificial Anasazi project and beyond." *Agent-based modeling and simulation in archaeology*. Springer, Cham, 2015. 37-50.

AAP: Time

- Annual time steps from 800 to 1350 CE
- In this period the valley population increased to about 250 households and then declined
- No evidence of new constructions of houses in the valley or continued households after ~1300

AAP: Agents

- Agents in the model represent households of 5 individuals
- Archeological record used to establish settlement patterns (important for initialization & validation)
- Regional ethnographies used for generating “anthropologically feasible” behavioral rules

AAP: Agents

- Each household requires a certain amount of food, for which they farm
- Each agent farms a cell (distinct from its house)
- Each household makes annual decisions on where to farm and where to settle
- Properties: each household has an age, food stock level, farming location, etc.

AAP: Agents

Table 1. Household (agent) attributes

1. Five surface rooms or one pithouse is considered to represent a single household.
2. Each household that is both matrilineal and matrilocal consists of 5 individuals. Only female marriage and residence location are tracked, although males are included in maize-consumption calculations.
3. Each household consumes 160 kg of maize per year per individual.
4. Each household can store a maximum of 2 years' total corn consumption (1,600 kg), i.e., if at harvest 800 kg of corn remains in storage and additional 800 kg can be added to that from the current crop.
5. Households use only 64% of the total potential maize yield. (The unutilized production is attributed to fallow, loss to rodents, insects, and mildew, and seed for the next planting.)

Axtell, Robert L., et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." *PNAS* 99.suppl 3 (2002): 7275-7279.

AAP: Rules & Interactions

- Complicated set of agent rules and attributes (but still vastly simpler than they could be…)

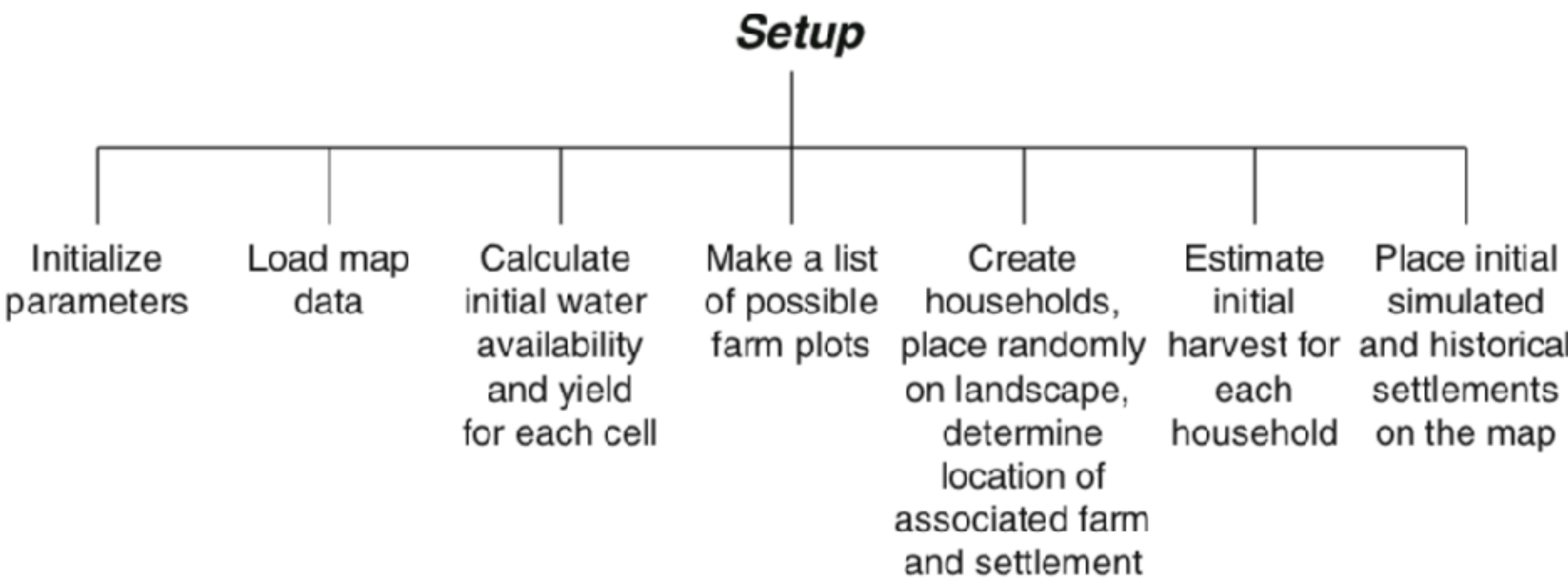
AAP: Rules & Interactions

Table 2. Household (agent) rules

1. A household fissions when a daughter reaches the age of 15.
2. A household moves when the amount of grain in storage in April plus the current year's expected yield (based on last year's harvest total) falls below the amount necessary to sustain the household through the coming year.
   A. Identification of agricultural location:
      The location must be currently unfarmed and uninhabited.
      The location must have potential maize production sufficient for a minimum harvest of 160 kg per person per year (22). Future maize production is estimated from that of neighboring sites.
      If multiple sites satisfy these criteria the location closest to the current residence is selected.
      If no site meets the criteria the household leaves the valley.
   B. Identification of a residential location:
      i. The residence must be within 1 km of the agricultural plot.
      ii. The residential location must be unfarmed (although it may be inhabited, i.e., multihousehold sites permitted).
      iii. The residence must be in a less productive zone than the agricultural land identified in A.
      If multiple sites satisfy these above criteria the location closest to the water resources is selected.
      If no site meets these criteria they are relaxed in order of iii then i.

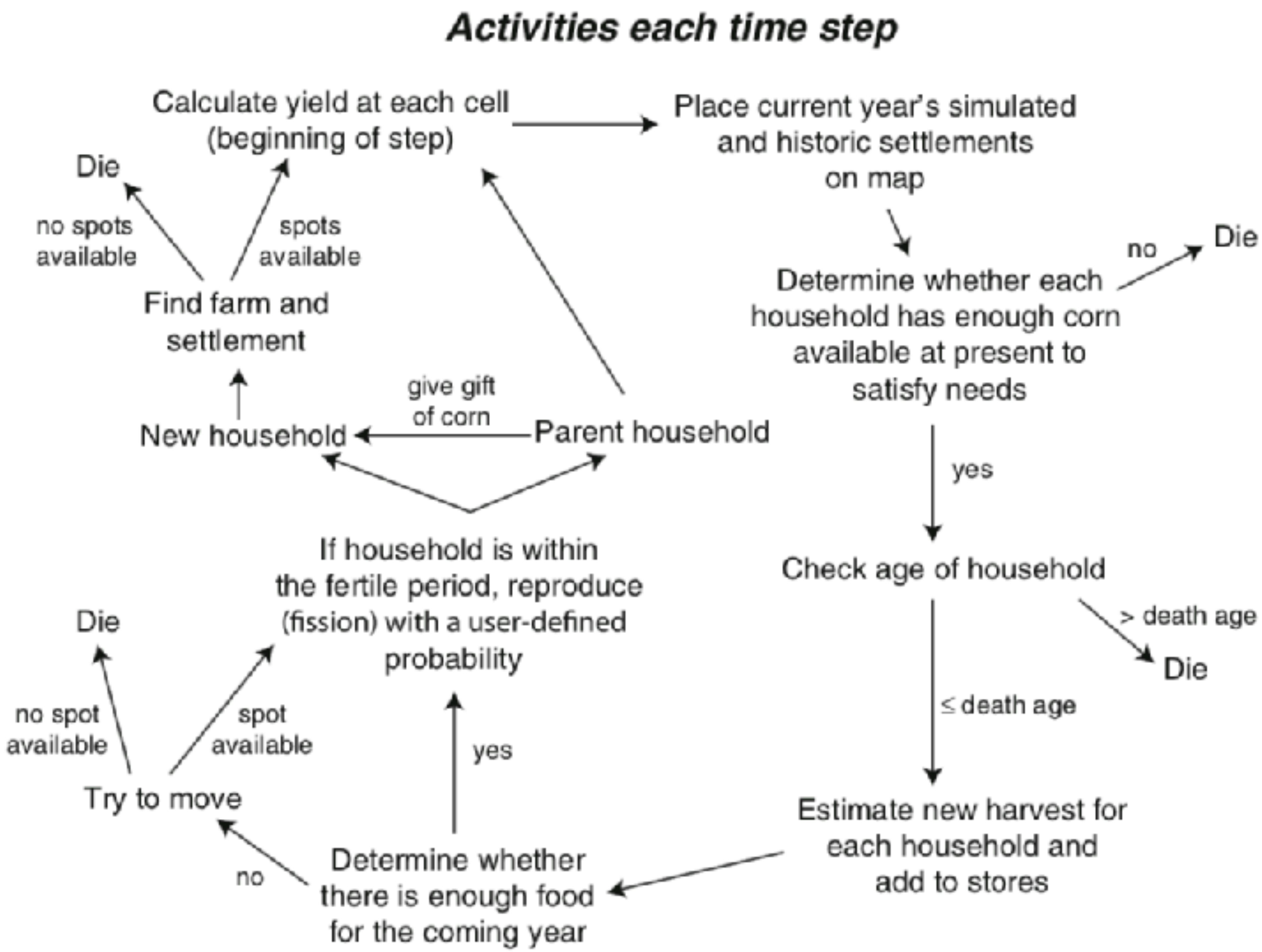
Axtell, Robert L., et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." *PNAS* 99.suppl 3 (2002): 7275-7279.

AAP: Overall Model



Swedlund, Alan C., et al. "Modeling archaeology: origins of the artificial Anasazi project and beyond." *Agent-based modeling and simulation in archaeology*. Springer, Cham, 2015. 37-50.

AAP: Overall Model



Swedlund, Alan C., et al. "Modeling archaeology: origins of the artificial Anasazi project and beyond." *Agent-based modeling and simulation in archaeology*. Springer, Cham, 2015. 37-50.

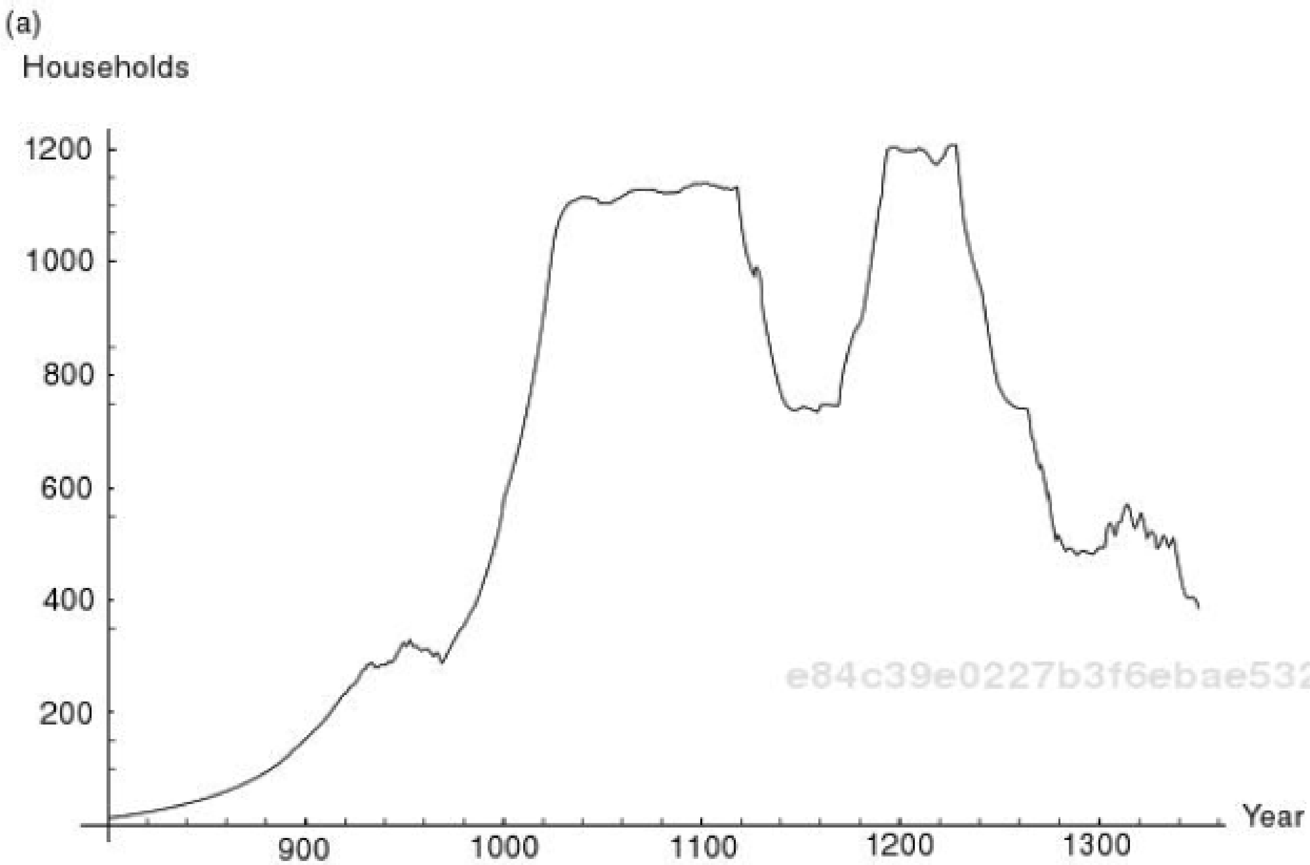
AAP: Base case parameterization

Table 4. Base case parameterization of the model

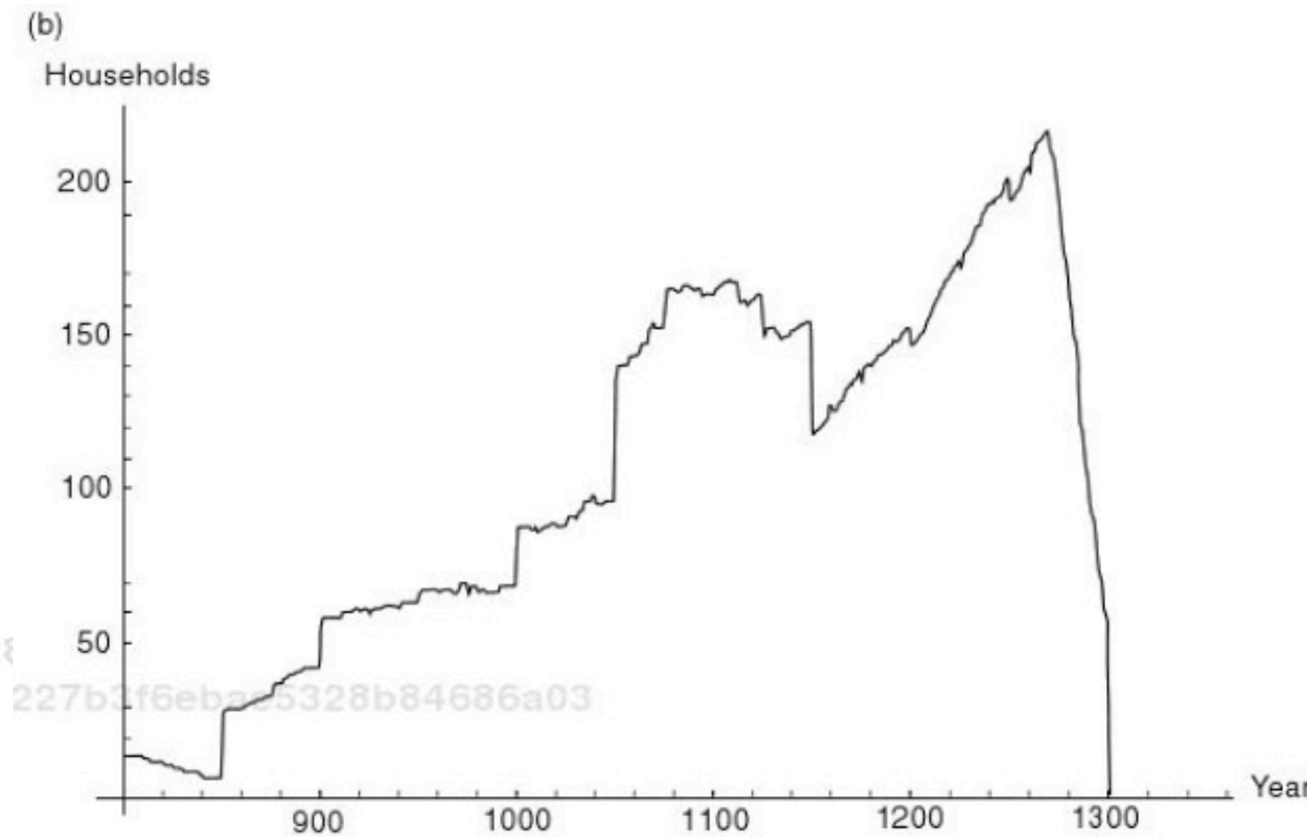
Parameter	Value
Random seed	Varies
Year at model start	A.D. 800
Year at model termination	A.D. 1350
Nutritional need per individual	800 kg
Maximum length of grain storage	2 years
Harvest adjustment	1.00
Annual variance in harvest	0.10
Spatial variance in harvest	0.10
Household fission age	16 years
Household death age	30 years
Fertility (annual probability of fission)	0.125
Grain store given to new household	0.33
Maximum farm to residence distance	1,600 m
Initial corn stocks, minimum	2,000 kg
Initial corn stocks, maximum	2,400 kg
Initial household age, minimum	0 years
Initial household age, maximum	29 years

Axtell, Robert L., et al. "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley." *PNAS* 99.suppl 3 (2002): 7275-7279.

AAP: Base case results

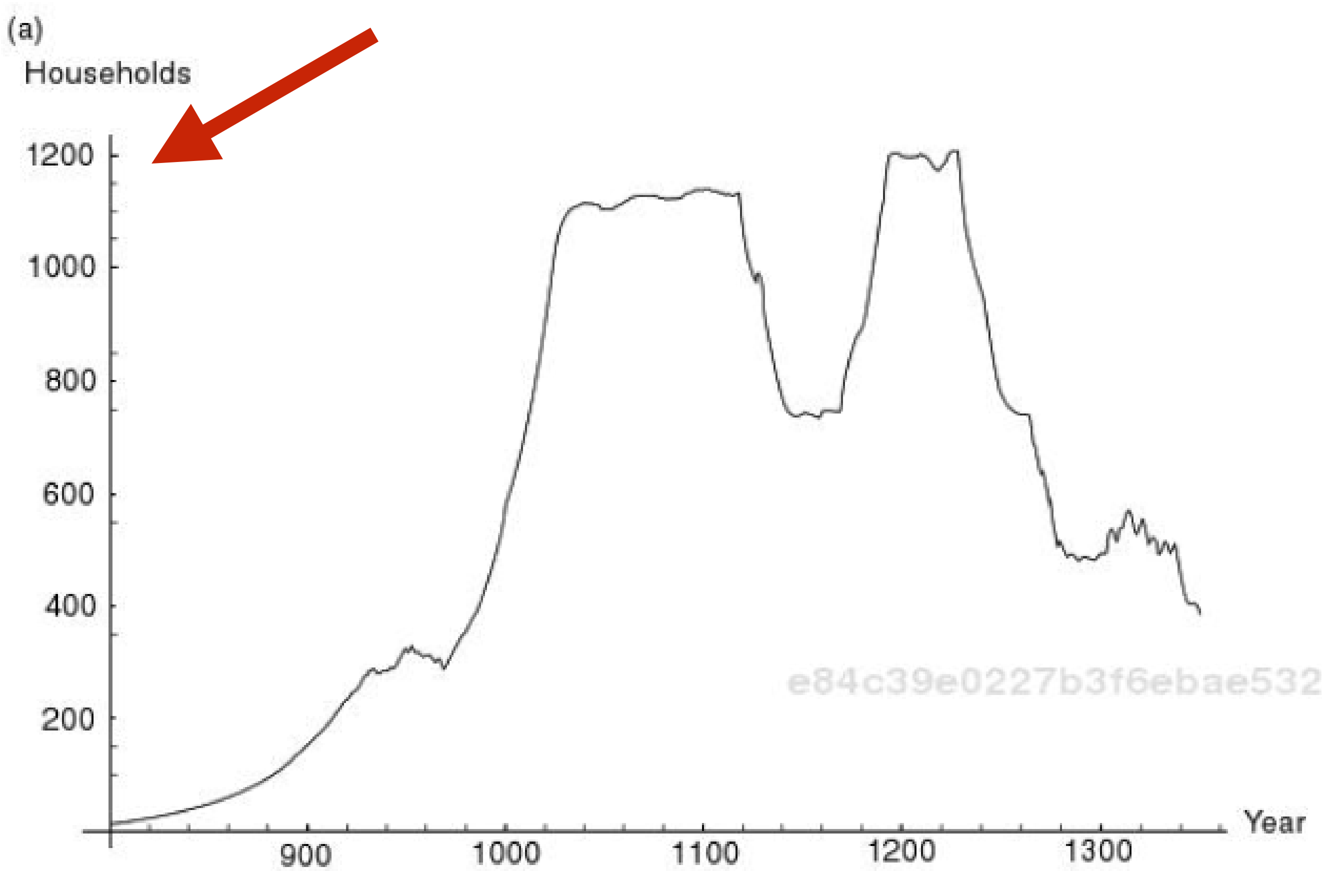


Model

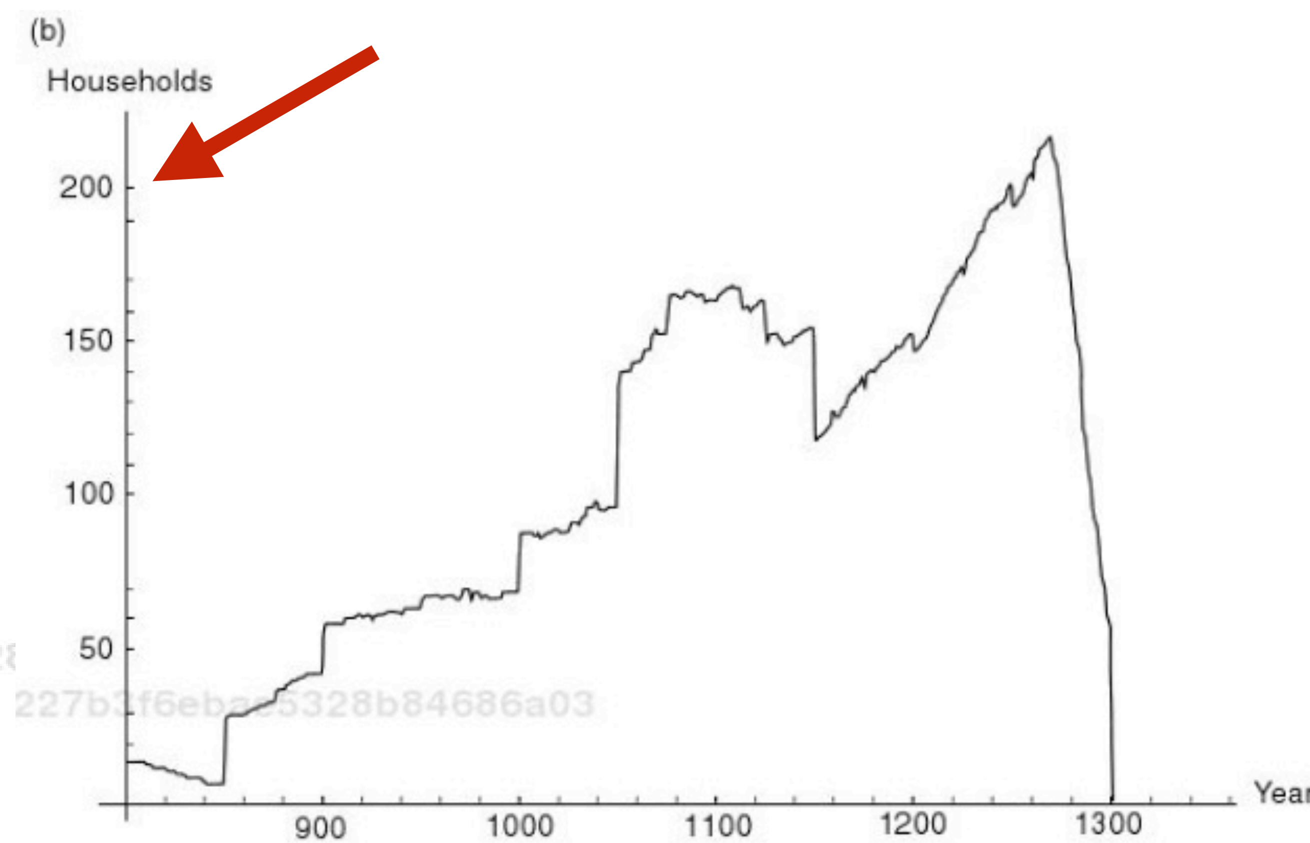


Data

AAP: Base case results



Model



Data

AAP: Base case results

- Qualitative match of some features
  - Captured spatial settlement patterns of households
  - General patterns of aggregation, dispersal, population dynamics
- Quantitative match? Not so much
  - Attempts to improve fit by hand-adjusting agent parameters resulted in premature population collapse

Model calibration & adding heterogeneity

- Introduced free parameters: 6 for agents, 2 for environment

- Allows for increased heterogeneity of agents and landscape by introducing draws from uniform distributions

- ‘Systematic search’ of parameter space to optimize fit to data based on L^1, L^2, or L^∞ norms

  - Fit based on either 1 simulation or average of 15

Model calibration

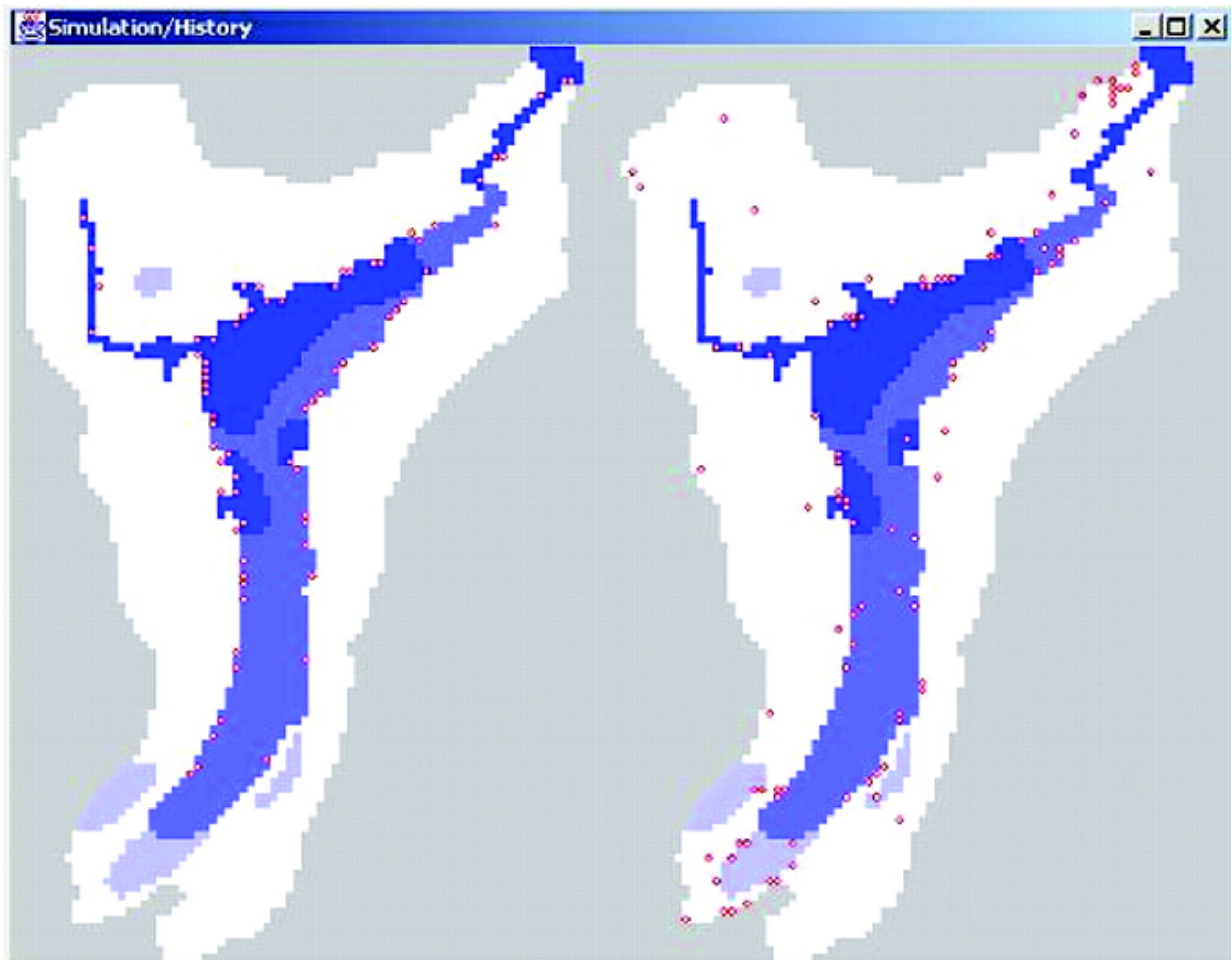
Table 5. Optimized parameter settings based on single "runs" of the model

Parameter/norm	L^1	L^2	L^∞
Minimum death age	26	30	25
Maximum death age	32	39	34
Minimum age, end of fertility	30	28	30
Maximum age, end of fertility	32	30	30
Minimum fission probability	0.125	0.120	0.125
Maximum fission probability	0.129	0.125	0.125
Average harvest	0.60	0.62	0.60
Harvest variance	0.41	0.40	0.40

Table 6. Optimized parameter settings based on the average over 15 runs of the model

Parameter/norm	L^1, L	L^2
Minimum death age	30	25
Maximum death age	36	38
Minimum age, end of fertility	30	30
Maximum age, end of fertility	32	38
Minimum fission probability	0.125	0.125
Maximum fission probability	0.125	0.125
Average harvest	0.6	0.6
Harvest variance	0.4	0.4

AAP: Qualitative match of settlement patterns



AAP: Improved quantitative match with time-series data

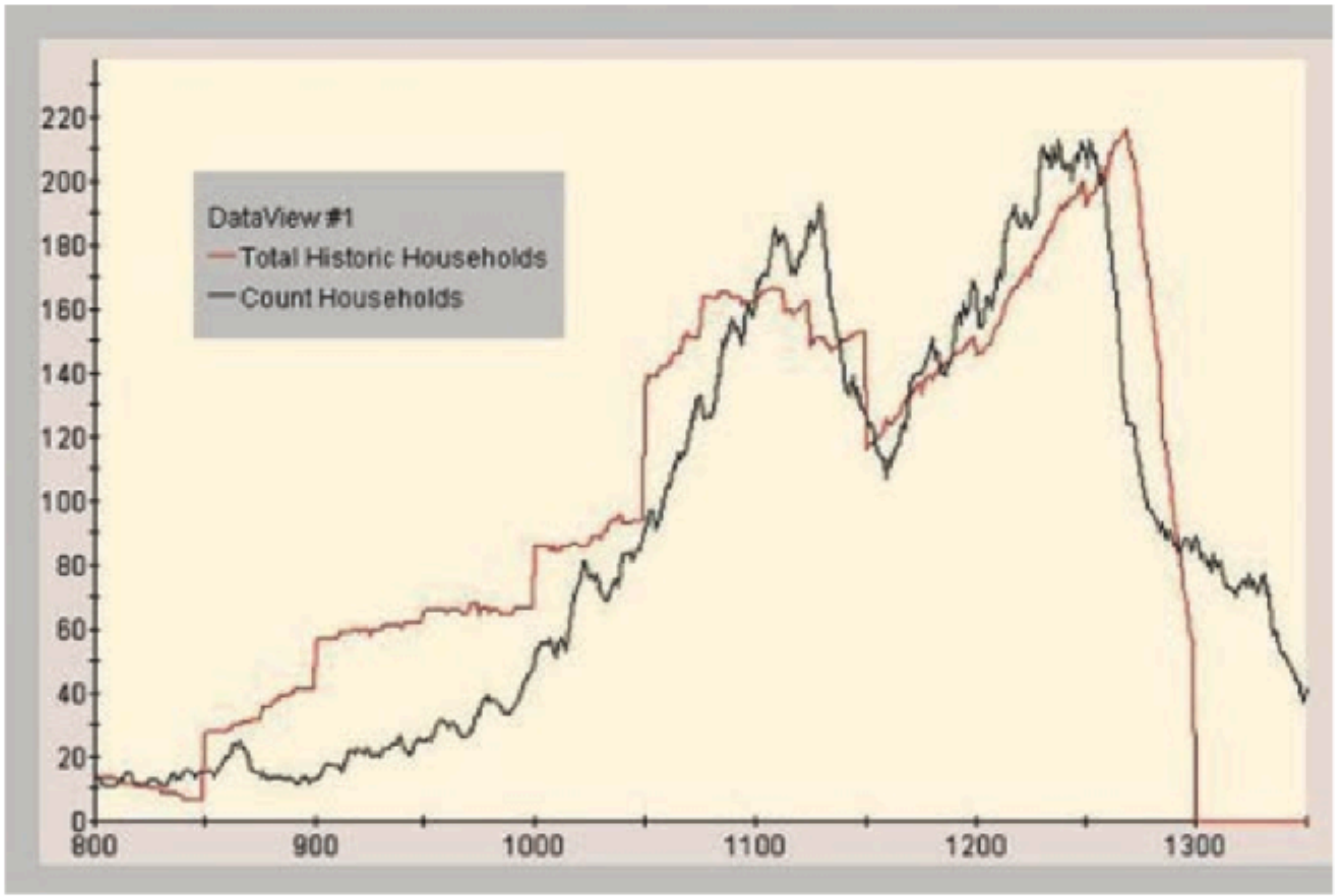
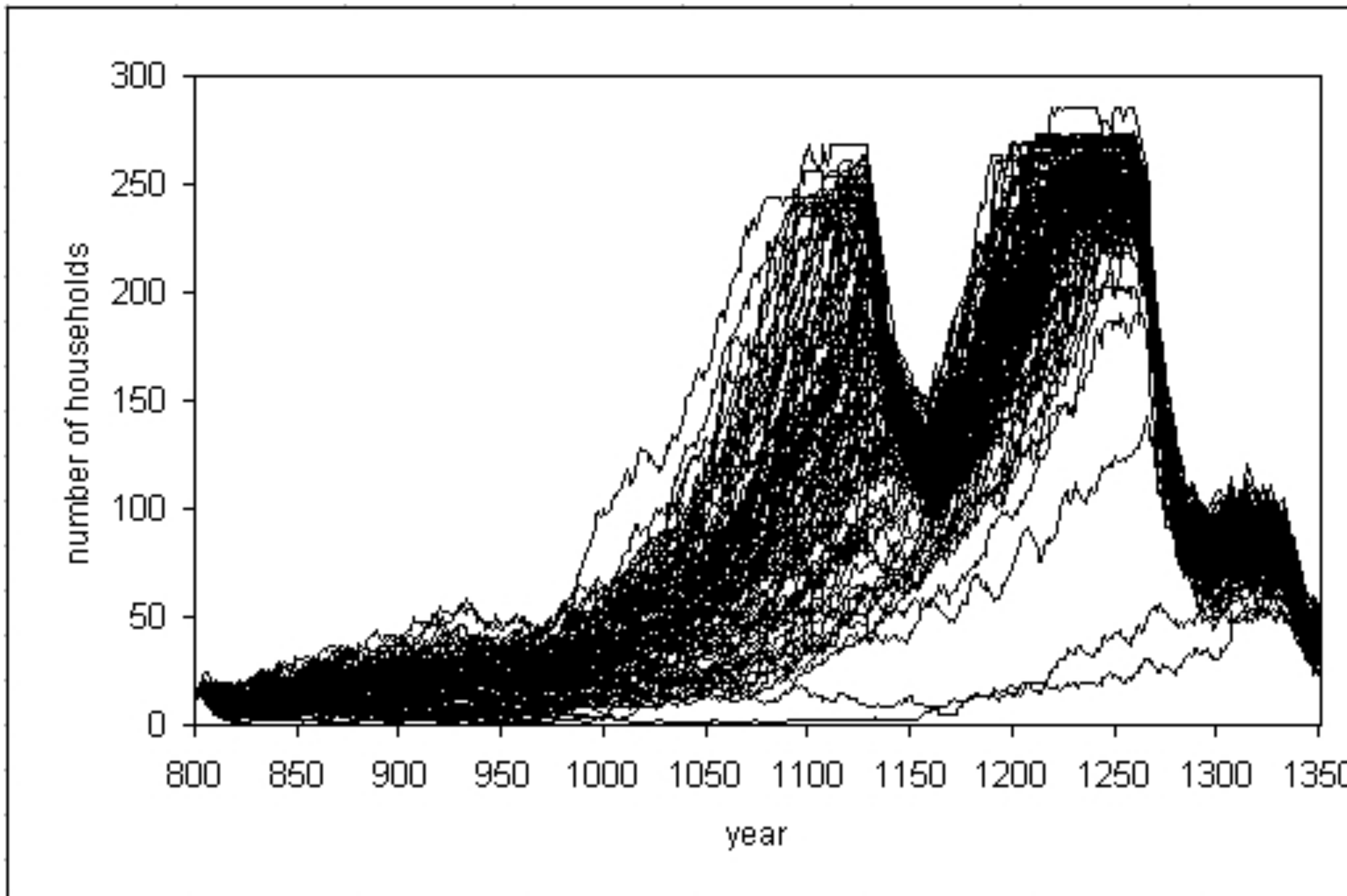


Fig. 2. Best single run of the model according to the L^1 norm. Other best runs based on other norms yield very similar trajectories. The average run, produced by averaging over 15 distinct runs, looks quite similar to this one as well.

Variation across 100 runs



What happened to the population ~1300 CE?

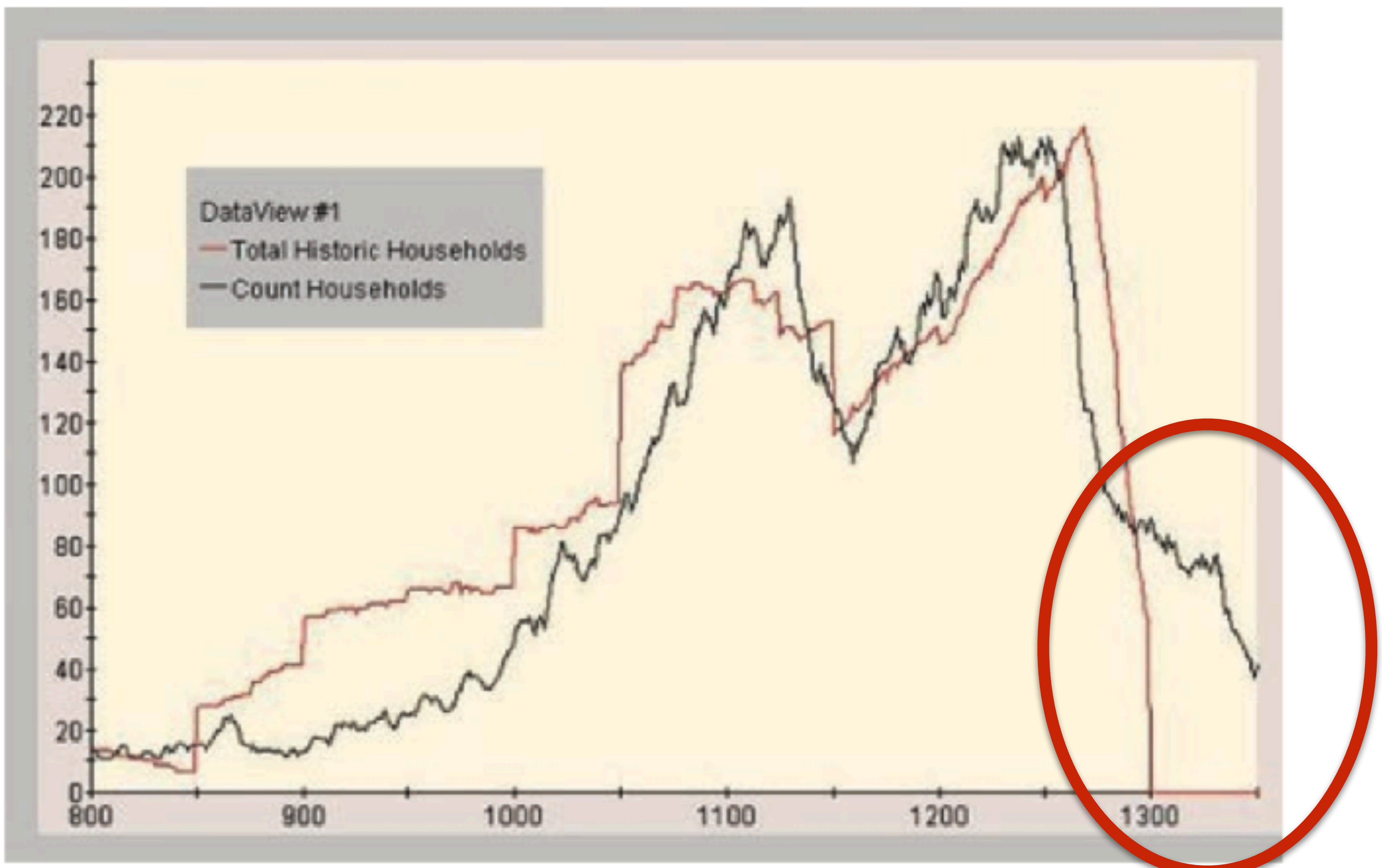


Fig. 2. Best single run of the model according to the L^1 norm. Other best runs based on other norms yield very similar trajectories. The average run, produced by averaging over 15 distinct runs, looks quite similar to this one as well.

All models are wrong, but this can be useful in many ways

- “All models are wrong, but some are useful” - George E. P. Box

- May capture a simplistic key underlying premise/ mechanism that explains what you see, even though the model is wrong (e.g. Lotka-Volterra, SIR, Hooke's law)

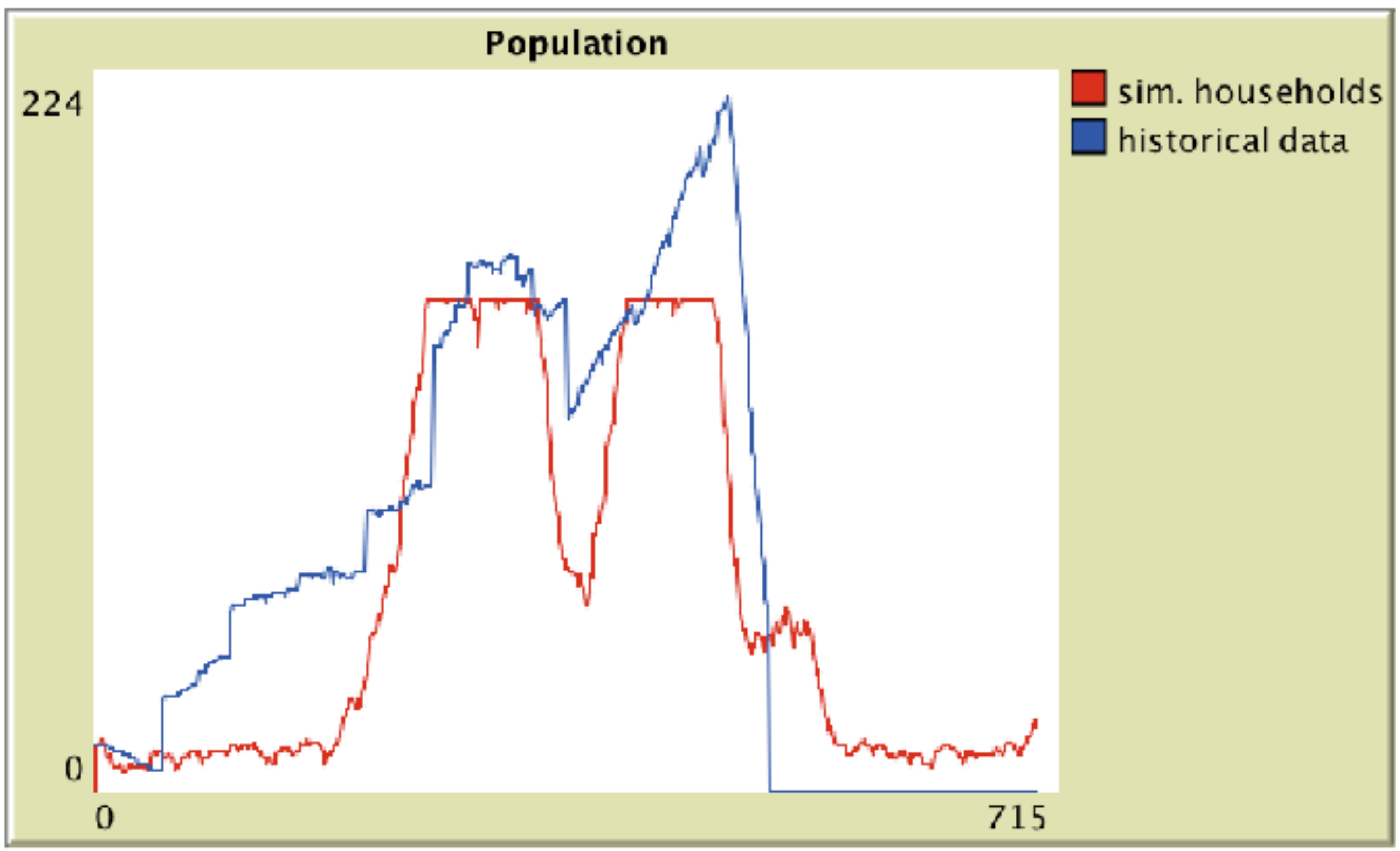
- May also disprove a hypothesis and prompt you to find new ways of thinking (William Harvey)

What happened to the population ~1300 CE?

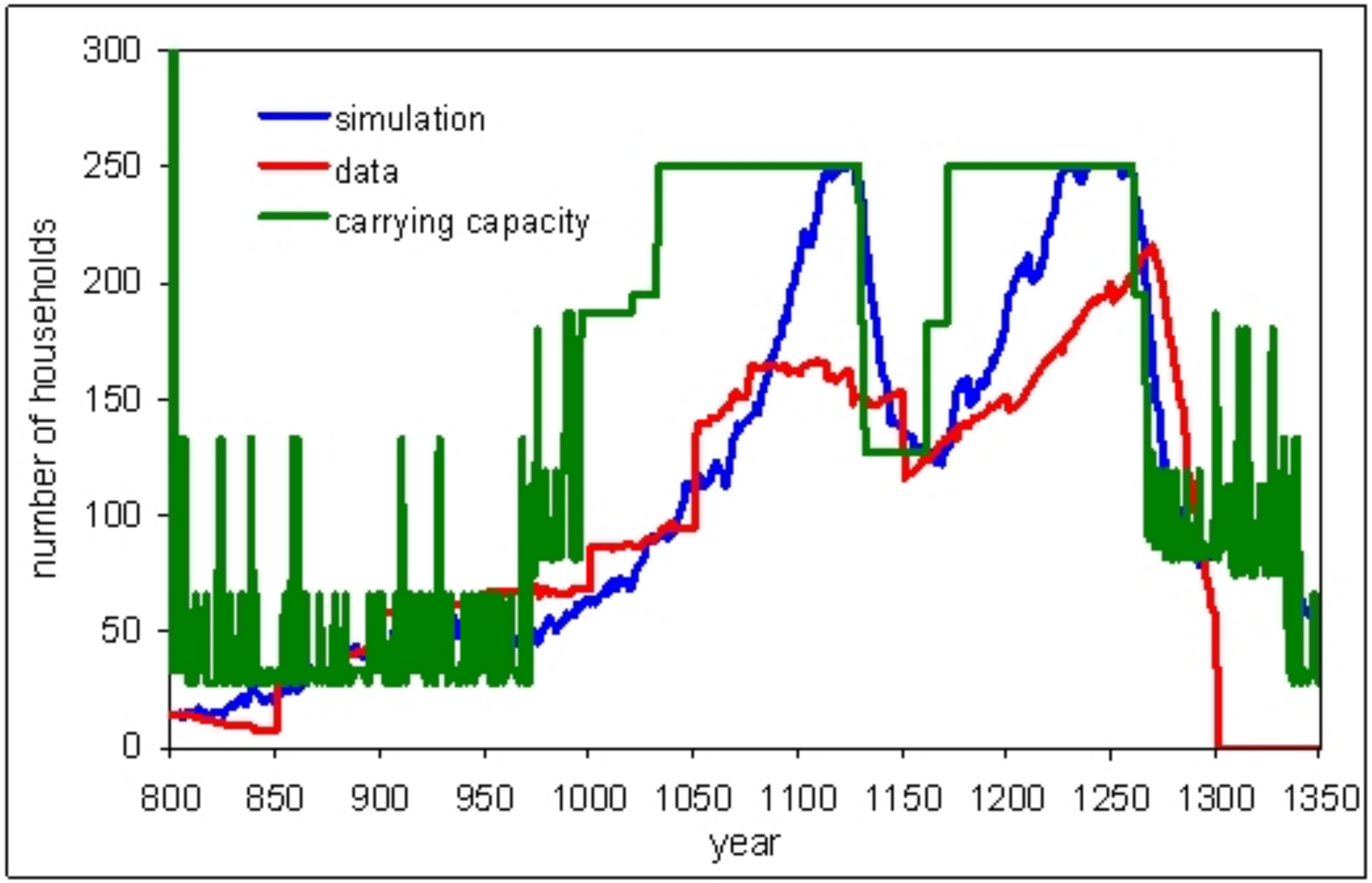
- The divergence after 1250 of the model and record constitutes an “informative failure”
- Model suggests that the area could potentially have continued to support a smaller population
- History shows the valley being abandoned
- Model suggests the inclusion of additional non-environmental, potentially sociocultural forces—hypothesis generation

What would happen if we ran the model longer?

Try it! NetLogo implementation of AAP in the models library



Some criticisms of the AAP



Carrying capacity = total cells with base yield ≥ nutritional needs
Recall base yield depends on input data & scaling factor

Janssen, Marco A. "Understanding artificial anasazi." Journal of Artificial Societies and Social Simulation 12.4 (2009): 13.

Artificial Anasazi Project

- Demonstrates the potentially powerful use of modeling to “rewind the tape of history” and provide an experimental laboratory for the past
- Shows the challenges and advantages of computational models with strong empirical grounding
- Example of pursuing “generative sufficiency” to understand behavior of a system
- Has also sparked many more advanced & interesting models!