Improving Genetic Algorithm Calibration of Probabilistic Cellular Automata for Modeling Mining Permit Activity

Sam Talaie

This project is partially supported by the USGS and builds on previous work done by
Sushil J. Louis, Gary L Raines, Ryan E. Leigh

Genetic Algorithms

- Developed by John Holland
  - Natural selection – survival of the fittest
  - Natural genetics
- Used when problems are poorly defined, hard.
  - Multimodal
  - Discontinuous
  - Non-linear

Robotics	**How do we catch a ball, navigate, play basketball**
User Interfaces	**Predict next command, adapt to individual user**
Medicine	**Protein structure prediction, Is this tumor benign, design drugs**
Design	**Design bridge, jet engines, Circuits, wings**
Control	**Nonlinear controllers**

Genetic Algorithms

- Solutions are encoded as binary chromosomes
- A set of operators acts on a population of chromosomes to evolve better solutions
  - Selection
  - Crossover
  - Mutation
- **Quickly produces good (usable) solutions**
- **Not guaranteed to find optimum**

Searching for Optima

- Searching for optima
  - Traditional Methods
    - Calculus
      - Depend on existence of derivatives
      - Most real-world functions are not unconstrained, smooth, calculus friendly functions.
    - Hill Climbing
      - Fails when reaches local optima

Search as a solution to hard problems

- Strategy: generate a potential solution and see if it solves the problem
- Make use of information available to guide the generation of potential solutions
- How much information is available?
  - Very little: We know the solution when we find it
  - Lots: linear, continuous, …
  - Modicum: Compare two solutions and tell which is "better"

Algorithm

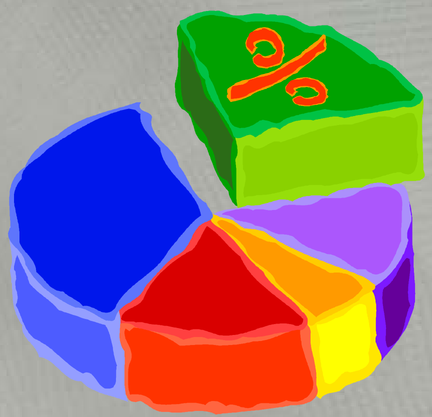
- Generate pop(0)
- Evaluate pop(0)
- T=0
- While (not converged) do
  - Select pop(T+1) from pop(T)
  - Recombine pop(T+1)
  - Evaluate pop(T+1)
  - T = T + 1
- Done

GA – Evaluation

Application dependent fitness function

GA - Selection

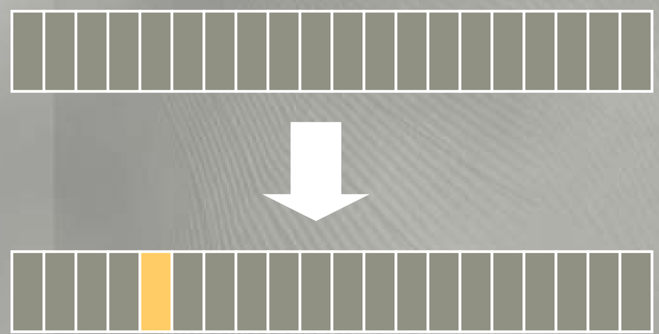
- Each member of the population gets a share of the pie proportional to fitness relative to other members of the population



- Spin the roulette wheel pie and pick the individual that the ball lands on

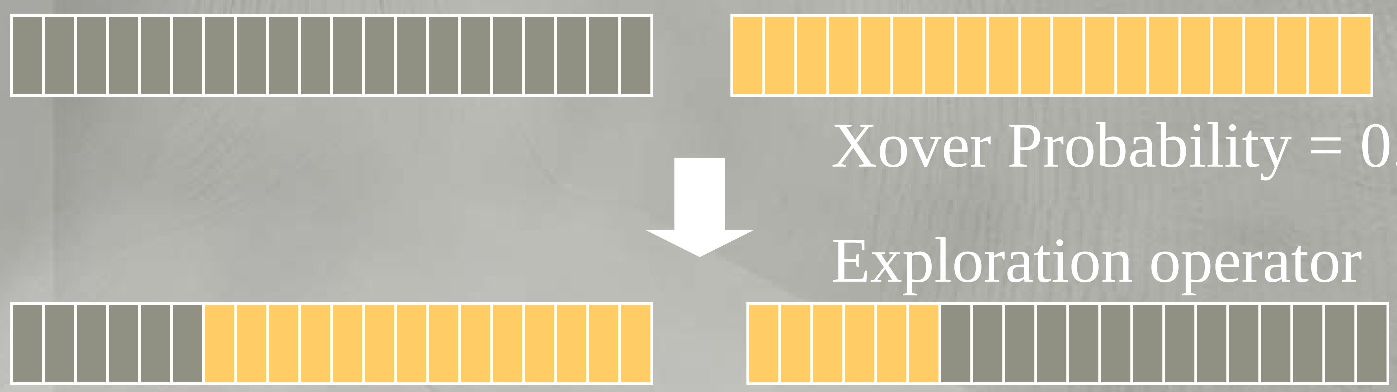
- Focuses search in promising areas

Crossover and Mutation



Mutation Probability = 0.001

Insurance



Xover Probability = 0.7

Exploration operator

GA – Exploration vs. Exploitation

- More exploration means
  - Better chance of finding solution (more robust)
  - Takes longer

- More exploitation means
  - Less chance of finding solution, better chance of getting stuck in a local optimum
  - Takes less time

GA - Example

String	decoded	f(x^2)	fi/Sum(fi)	Expected	Actual	
01101	13	169	0.14	0.58	1	
11000	24	576	0.49	1.97	2	
01000	8	64	0.06	0.22	0	
10011	19	361	0.31	1.23	1	
Sum		1170	1.0	4.00	4.00	
Avg		293	.25	1.00	1.00	
Max		576	.49	1.97	2.00	

GA - Example

String	mate	offspring	decoded	f(x^2)		
0110|1	2	01100	12	144		
1100|0	1	11001	25	625		
11|000	4	11011	27	729		
10|011	3	10000	16	256		
Sum				1754		
Avg				439		
Max				729		

Research

A harmonious marriage between

Cellular Automata

&

Genetic Algorithms

What is the project about?

- **What is the problem?**
  - Calibrating a CA
- **What is the technique?**
  - Genetic Algorithm
- **What are the issues?**
  - Encoding
  - Evaluation
- **What are our results ?**

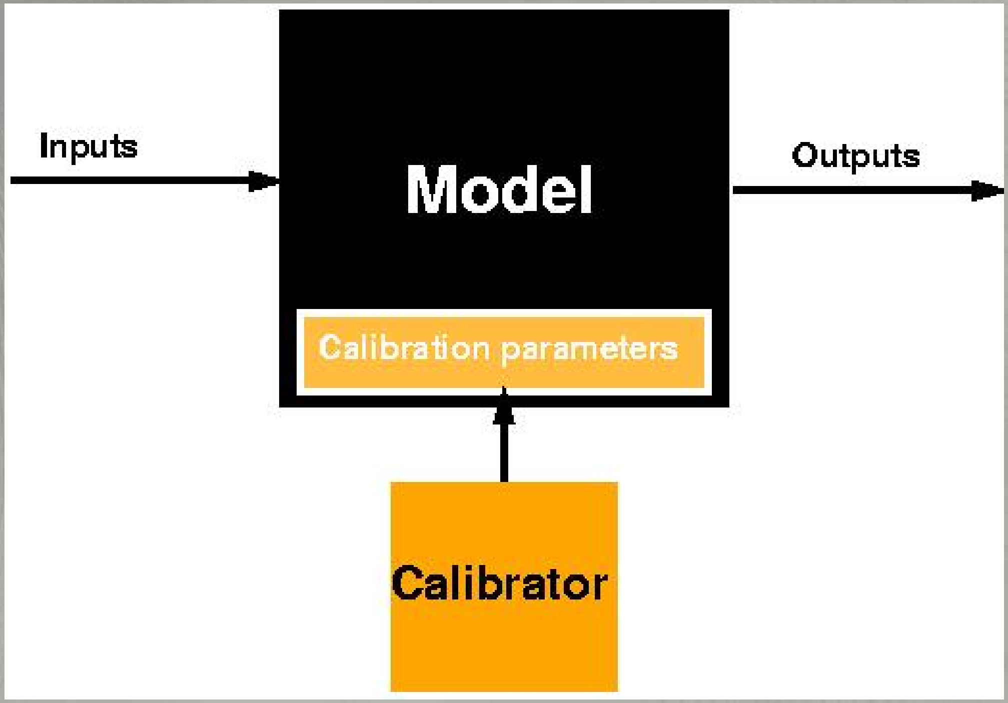
Problem

- Project mineral-related activity on public land to 2010
  - Predicting permit activity in an area
    - Spatially explicit
    - USGS
      - permit activity from 1989 – 1998
      - natural resources
    - Use cellular automata to model (predict) mining activity over next ten years
  - Problem: Takes weeks to tune CA rules to match available data

Problem

- **Can we automate calibrating a cellular automaton**
    - As good as CA calibrated by human
    - In the same or less time

Problem



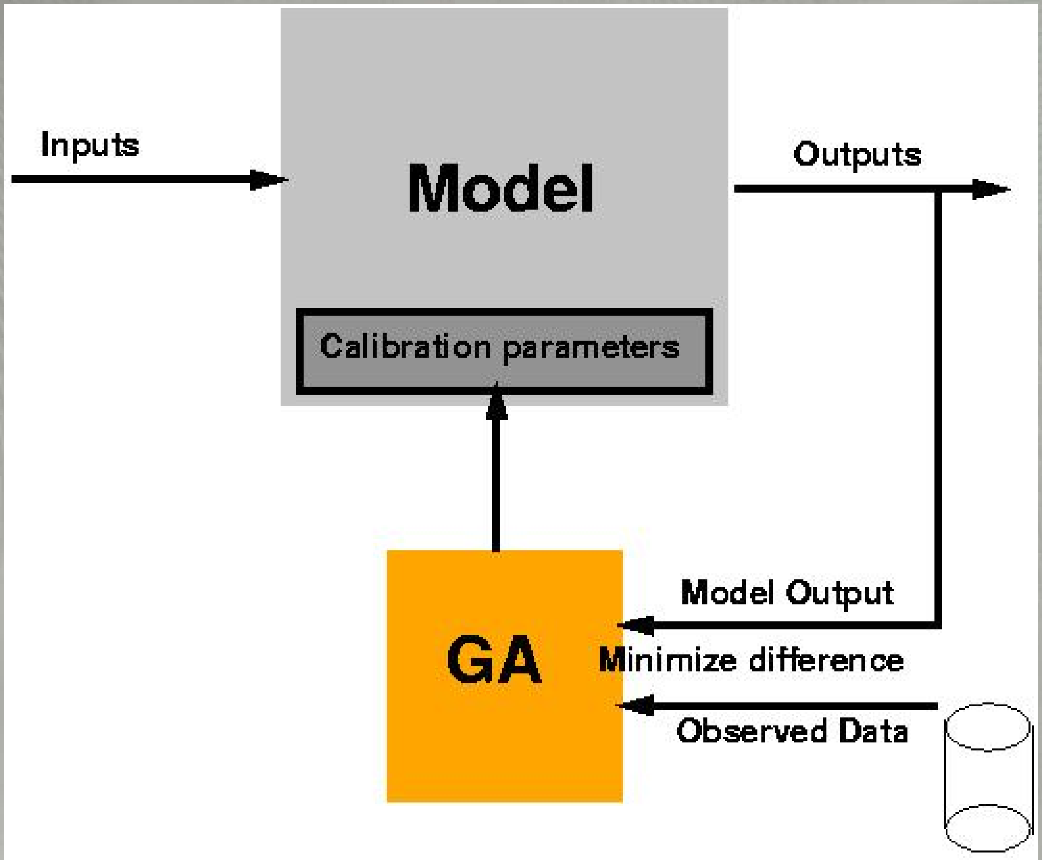
Model Parameters

- 496 X 503 = 249,488 cell CA
- 5 years (iterations)
- Average over 3 runs
- Roughly 4 Million computations.

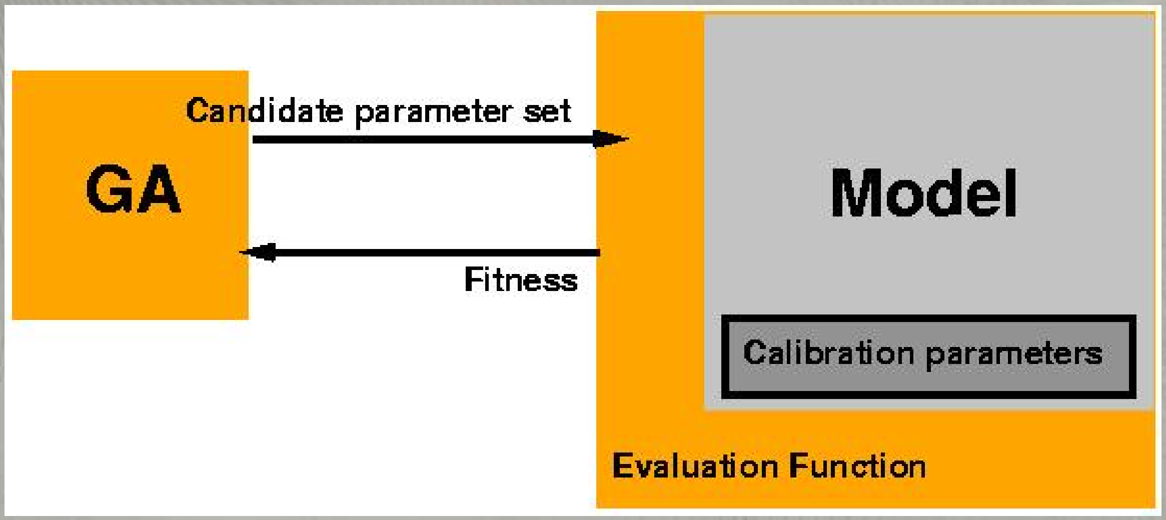
GA Calibration

- Empirical evidence to support their use in this kind of problem
  - **Physics models**
    - *Physical Review Letters, Volume 88, Issue 4*
    - *Journal of Quantitative Spectroscopy and Radiative Transfer. Volume 75, 2002, Pgs. 625 - 636*
  - **Seismic models**
    - *Congress on Evolutionary Computing 1999, pages 855 - 861*
  - **Hydrology models**
    - *In progress*
  - Proceedings of GECCO, CEC, …

GA Calibration



GA Evaluation



Modified Annealed Voting Rule

Probability of Life in Next Generation

Number of Live Neighbors	Status of Center Cell	
	Alive	Dead
> Annealing Window	**Very Likely**	**Likely**
Annealing Window	**Likely**	**Somewhat Likely**
< Annealing Window	**Very Somewhat Likely**	**Unlikely**

CA Parameters

Parameters	Definition
Very Likely	Square root of Likely (Larger)
Likely	A high probability of life.
Somewhat Likely	An intermediate probability of life
Very Somewhat Likely	Square root of Somewhat Likely (Larger)
Unlikely	A low probability of life
Resource Threshold	Minimum fuzzy membership defining where a reasonable explorationist would explore
Anneal Window	Position and width control response of CA

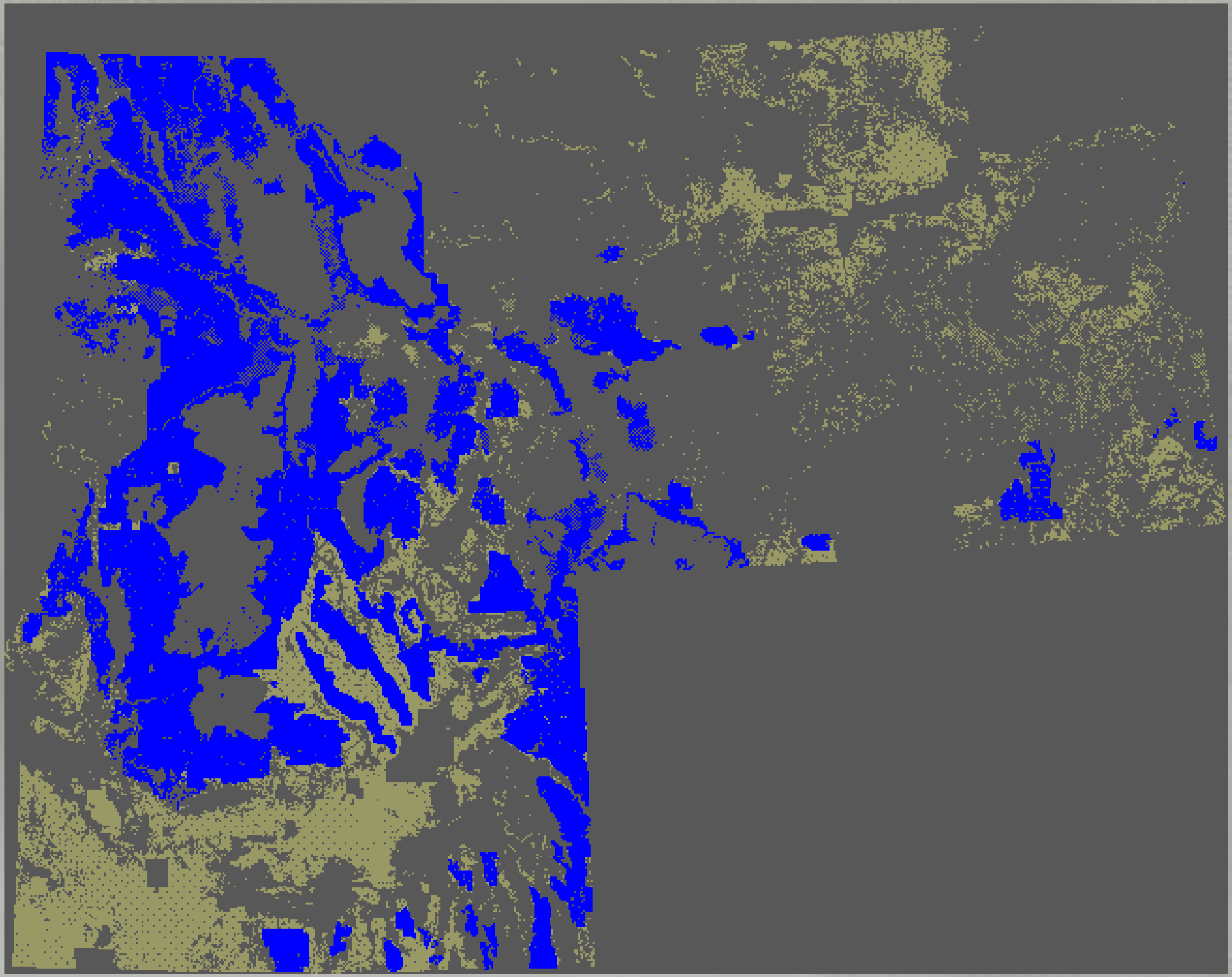
GA Encoding

top	Bottom	likelyInactive	likelyActive	veryLikely	somewhatLikely	verySomewhatLikely	unlikelyProb	ResourceThreashold
4	4	7	7	7	7	7	7	7

CHC Benefits

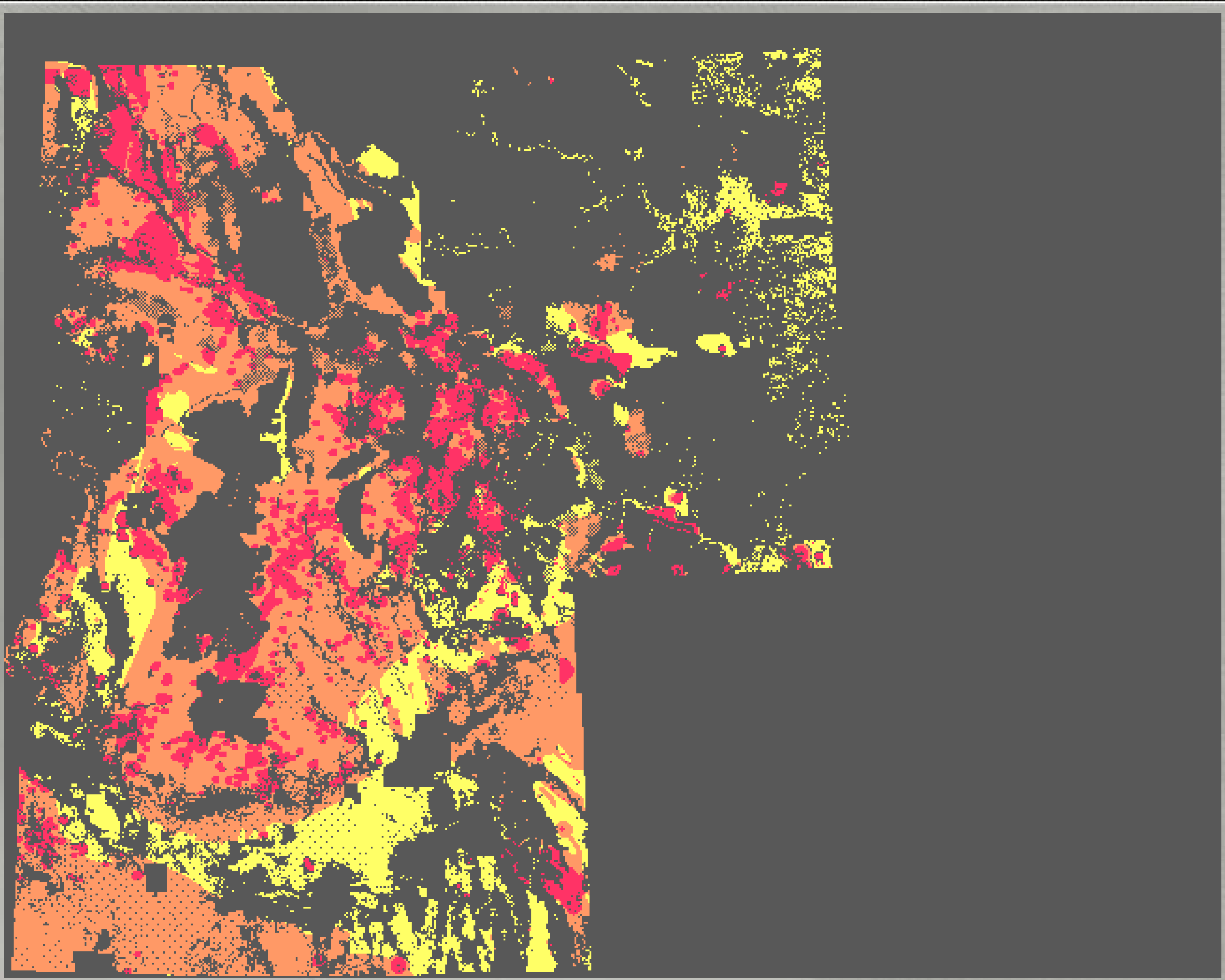
- Outperforms traditional GA as function optimizer
- Smaller population size needed to maintain same diversity as traditional GA
- Very effective for parameter optimization (Darrel Whitley)

Visualization of Data



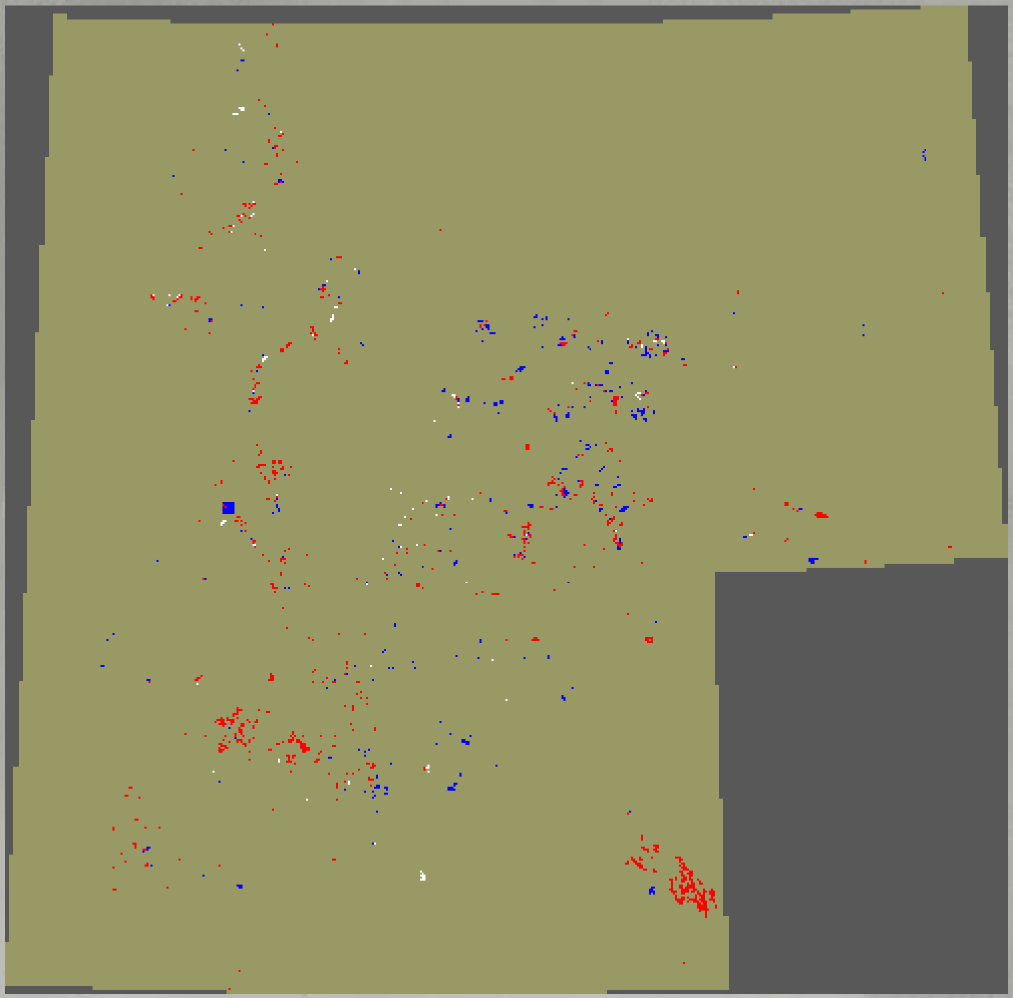
* Public Land

Visualization of Data



* Resources

Visualization of Data



*** CA Activity Model**

Evolution of the project

- **TCSC: Total Cell State Count**

$$\text{Minimize } \; g = \sum_{j=0}^{\text{nyears}} \sum_{i=0}^{\text{nstates}} \left(100 \times \frac{\mid M_{ij} - O_{ij} \mid}{M_{ij} + O_{ij}} \right)$$

  - Mij : predicted number of cells in state i in year j
  - Oij : actual number of cells in state i in year j
  - 4 Types of Cells:
    - Alive
    - Dead
    - Just Born
    - Just Died

Evolution of the project

- **Kappa statistic**
  - *Kappa is a measure of agreement normalized for chance agreement*

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

  - ***Where P(A) is the percentage agreement (e.g., between your classifier and ground truth) and P(E) is the chance agreement. K=1 indicates perfect agreement, K=0 indicates chance agreement.***

Evolution of the project

- **NSCP: Number of Spatially Correct Predictions**

$$\sum_{j=0}^{nyears} \sum_{i=0}^{nstates} w_i M_{ij}$$

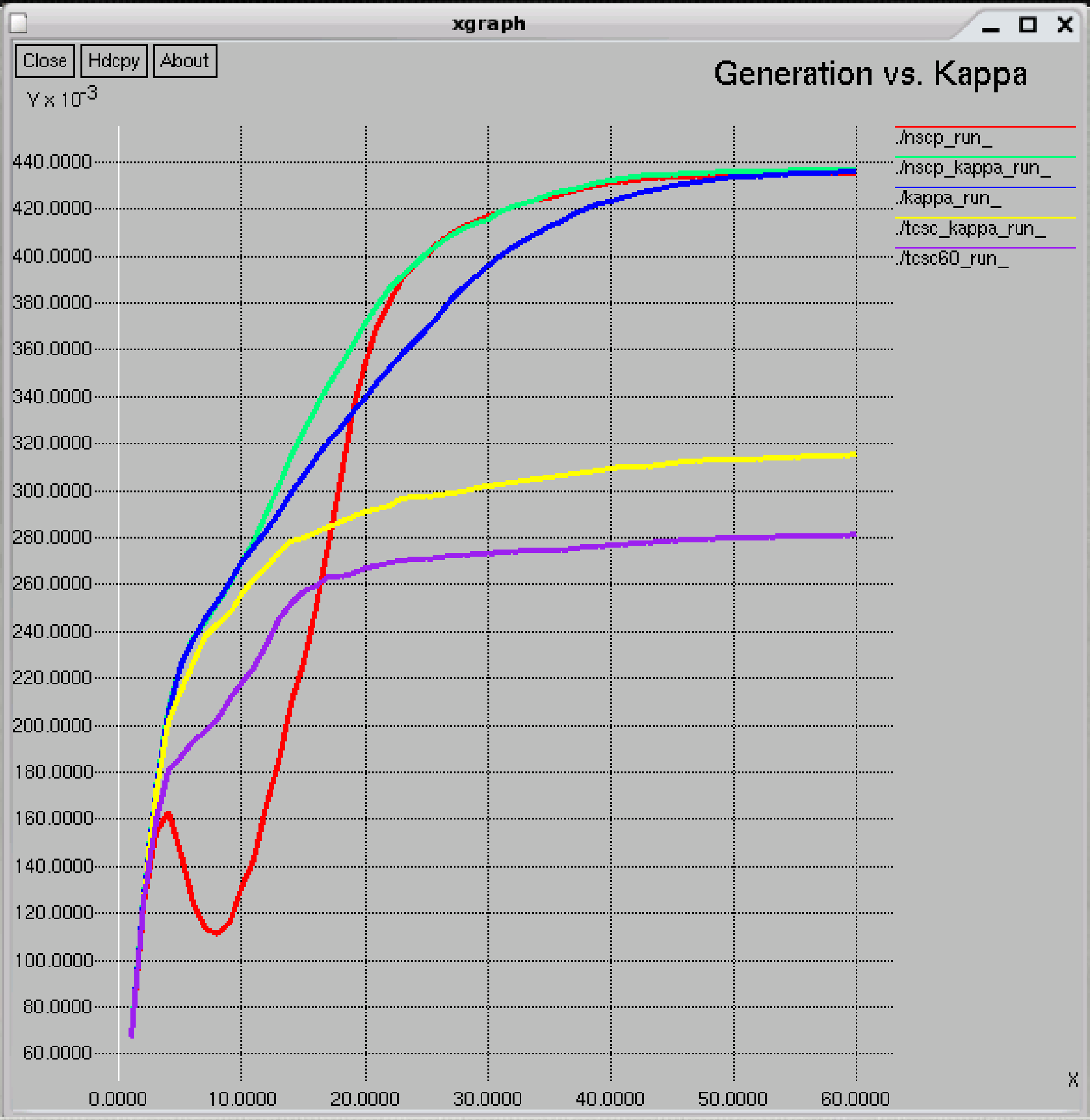
- M_{ij} : NSCP in state i in year j
- w_i : weight of state i

Results

- **Different Evaluation methods tested**
  - Population : 60
  - Generations : 60
  - Crossover Rate : 0.99
  - Mutation : 0.05
  - Runs : 10 with different seeds

  - 4 Million Computations * 60 * 60 =
  - 14.4 Billion Computations
  - On average, 0.3 seconds / evaluation

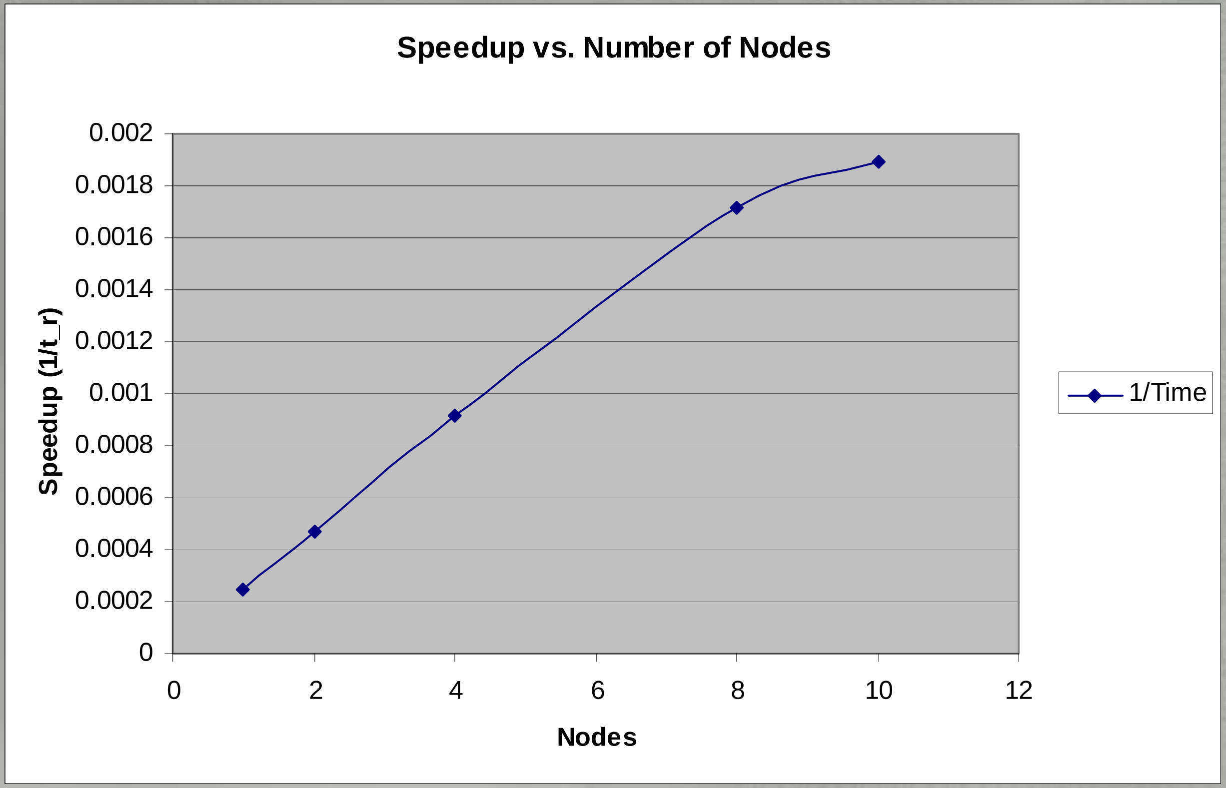
Results


xgraph
Close
Hdcpy
About
Generation vs. Kappa
Y x 10-3
./nscp_run_
./nscp_kappa_run_
./kappa_run_
./tcsc_kappa_run_
./tcsc60_run_
440.0000
420.0000
400.0000
380.0000
360.0000
340.0000
320.0000
300.0000
280.0000
260.0000
240.0000
220.0000
200.0000
180.0000
160.0000
140.0000
120.0000
100.0000
80.0000
60.0000
0.0000
10.0000
20.0000
30.0000
40.0000
50.0000
60.0000
X

Results

- Kappa results for fitness defined by
  - **TCSC**
    - Avg : 0.2814
  - **Kappa**
    - Avg : 0.4362
  - **TCSC** and **Kappa**
    - Avg : 0.3154
  - **NSCP**
    - Avg : 0.4356
  - **NSCP** and **Kappa**
    - Avg : 0.4366

Parallel GA



Speedup vs. Number of Nodes
0.002
0.0018
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
Speedup (1/t_r)
0
2
4
6
8
10
12
Nodes
1/Time

Conclusion

- 0.437 = Absolute Barrier
- Using Kappa Statistic in evaluation improves performance in both NSCP and TCSC
- Using NSCP results in reaching higher Kappa values more quickly
- Unfortunately NSCP was not able to break the 0.437 barrier

Future Work

- Evolve different rules for different sub-regions of the grid
- Encode and evolve rules instead of just rule parameters
- Explore different measurements of “success”
- Visualize Results