EPA Optimization Project Plan

Kalyanmoy Deb, Pouyan Nejadhashemi, Gregorio Toscano, and Sebastian Hernandez-Suarez

6 OCTOBER 2020

MICHIGAN STATE UNIVERSITY

Agenda

- Introduction to Michigan State University (MSU) Team
- Principles of Optimization Methods
- CAST System and Optimization (current status)
  - Interior Point Optimizer (Ipopt)
  - Base Genetic Algorithm (BGA)
  - Hybridizing BGA with Ipopt
  - CAST System Integration Plan
- Overall Project Plan

Introduction to MSU Team

- Kalyanmoy Deb, Professor
- Pouyan Nejadhashemi, Professor
- Gregorio Toscano, Post-doc Researcher
- J. Sebastian Hernandez-Suarez, Doctoral Student

Kalyanmoy Deb

- **Title:**
  - Koenig Endowed Chair Professor
    - Dept of Electrical and Computer Engineering
    - Dept of Computer Science and Engineering
    - Dept of Mechanical Engineering

- **Expertise and Achievements:**
  - Optimization, Multi-objective optimization, Machine Learning, Modeling

  - 33 years of experience in optimization and its applications
  - Author of popular evolutionary optimization methods: NSGA-II, NSGA-III
  - Author of two text-books on optimization, 548 research papers
  - 148,000 Google Scholar citations, h-index: 122
  - Director, Computational Opt. and Innovation (COIN) Lab at MSU

Pouyan Nejadhashemi

- **Title:**
  - University Foundation Professor
    - Department of Biosystems and agricultural Engineering
    - Department of Plant, Soil and Microbial Sciences
  - Elected board member
    - International Environmental Modelling & Software Society
- **Expertise and Achievements:**
  - Soft computing applications in water resources management
  - Computational Ecohydrology
  - Evaluation and development of watershed and water quality models
  - Received over $37M in grant funding
  - Over 100 peer-reviewed publications
  - Over 150 scientific presentations

Gregorio Toscano

- **Title:**
  - CBPO CAST optimization researcher
  - Associate Professor - Center for Research and Advanced Studies, Mexico
  - PhD in Evol. Multi-Criterion Optimization, 2005
- **Expertise and Achievements:**
  - Multi-objective optimization, Computational Intelligence, and Machine Learning
  - Multi-objective Micro-GA, Multi-objective PSO
  - Full Stack
  - Programming Languages

  - 5,907 Google Scholar citations

J. Sebastian Hernandez-Suarez

- **Title:**
  - PhD Candidate, Biosystems Engineering, MSU
  - MS Water Resources Engineering
  - BS Civil Engineering

- **Expertise:**
  - Watershed/water quality modeling
  - Environmental flow
  - Multi-objective optimization
  - Uncertainty quantification

A Brief Introduction to Optimization

- Optimization methods: Point versus Population approaches
- Hybrid and Customized optimization principle
  - Population-based methods and structured point-based methods
  - A case study on a billion-variable optimization
- Multi-objective optimization
  - Evolutionary multi-objective optimization
  - A large-scale land use management problem

Optimization Methods: Point and Population Based

Point-Based

Best for Convex and simplistic problems

Best for local search

Nonlinear Programming

$$\nabla f(x) - \sum_{j=1}^{J} u_j \nabla g_j(x) - \sum_{k=1}^{K} v_k \nabla h_k(x) = 0$$

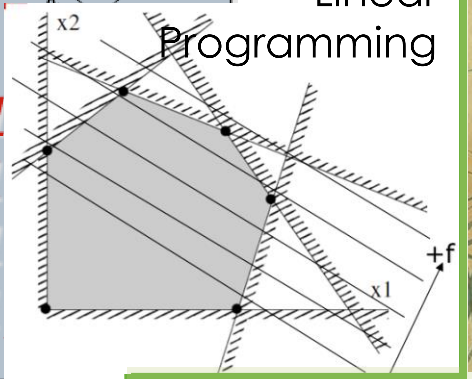
$g_j(\mathbf{x}) \geq 0 \quad j = 1$

$h_k(\mathbf{x}) = 0 \quad k = 1$

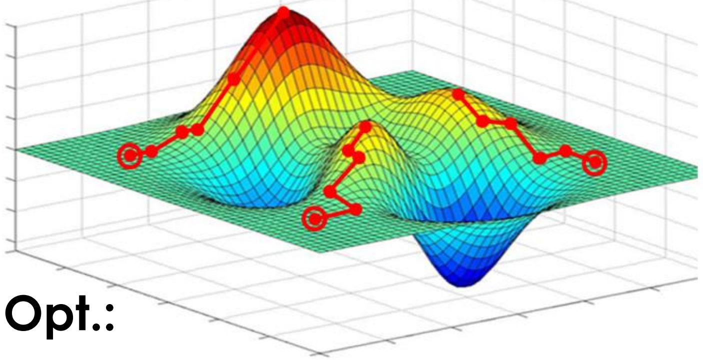
$u_j g_j(\mathbf{x}) = 0 \quad j = 1$

$u_j \geq 0 \quad j = 1$

Linear Programming



Math. optimization



Single-obj Opt.:

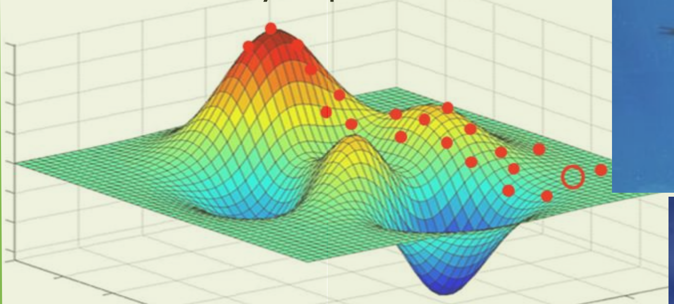
Search for a point which minimizes an objective function satisfying constraints

Population-Based

Best for non-convex and complex problems

Best for global search

Evolutionary optimization

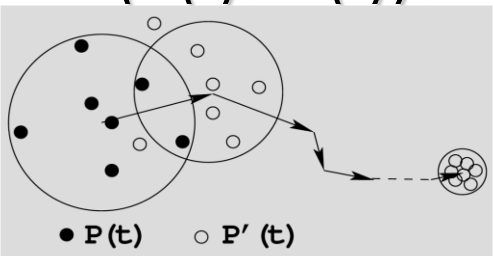


Other Meta-heuristics based methods

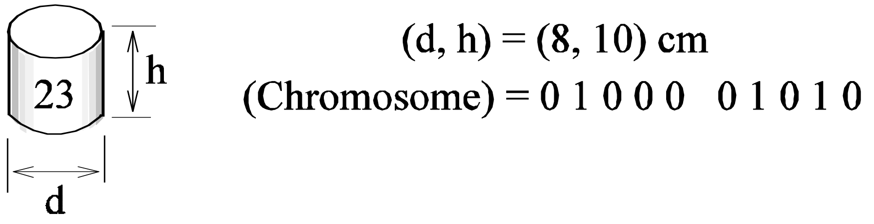
Goldberg, D. E. (1989). *Genetic algorithms for search, optimization, and machine Learning*, Reading, MA: Addison-Wesley.

Evolutionary Optimization: A Population-based Approach – Binary-coded Genetic Algorithm (BGA)

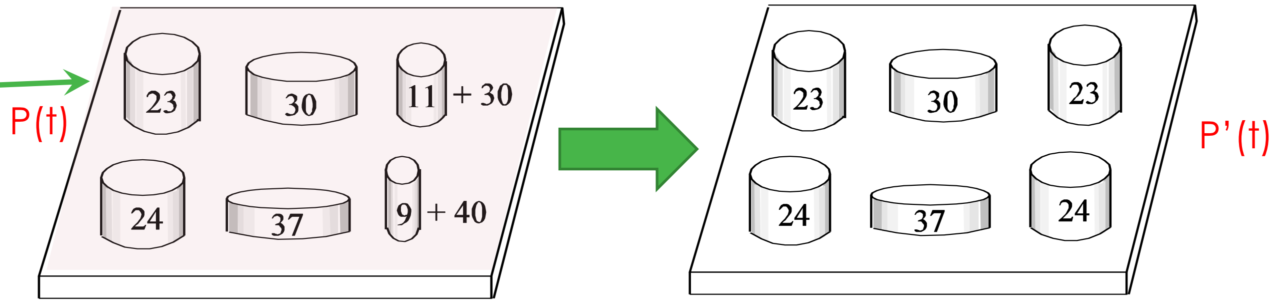
begin
Solution Representation
t := 0;
Initialization P(t);
Evaluation P(t);
while not Termination
do
  P'(t) := Selection (P(t));
  P''(t) := Variation (P'(t));
  Evaluation P''(t);
  P(t+1):= Survivor (P(t),P''(t));
  t := t+1;
od
end



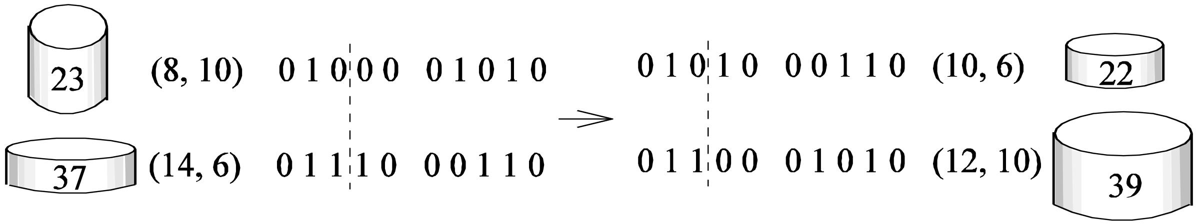
Can Design:



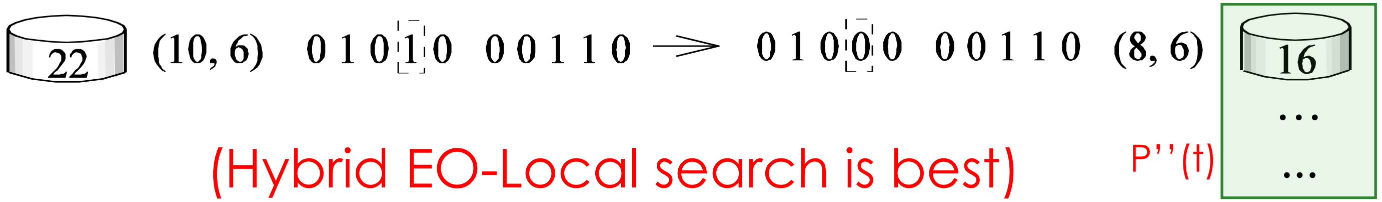
Selection:



Recombination:



Mutation:



(Hybrid EO-Local search is best)

A Fact of Optimization Methods

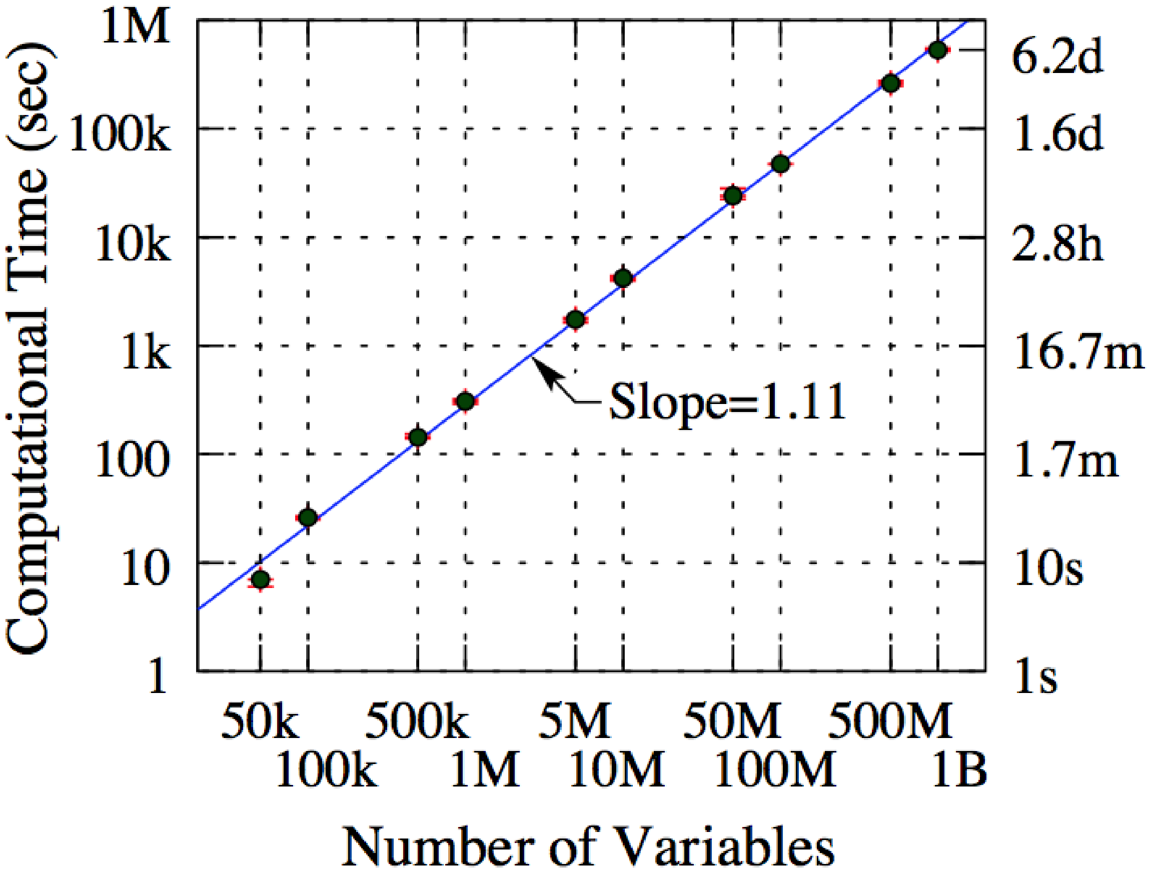
No-Free Lunch (NFL) theorem:

- Wolpert and Macready (1997)
- Algorithms A1 and A2
- All possible problems **F**
- Performances P1 and P2 using A1 and A2 for a fixed number of evaluations
- P1 = P2
- NFL breaks down for a narrow class of problems or algorithms
- Suggests to develop and use a **Customized Optimization** method, rather than a generic method

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1):67–82, 1997.

A Case Study: Customized Optimizatio

- Casting scheduling with large dimensions
  - 50k variables to start
- CPLEX cannot solve ≥2,000 variables
- Custom recombination and mutation operators
- Custom GA solves one **Billion-variable** problem in polynomial time

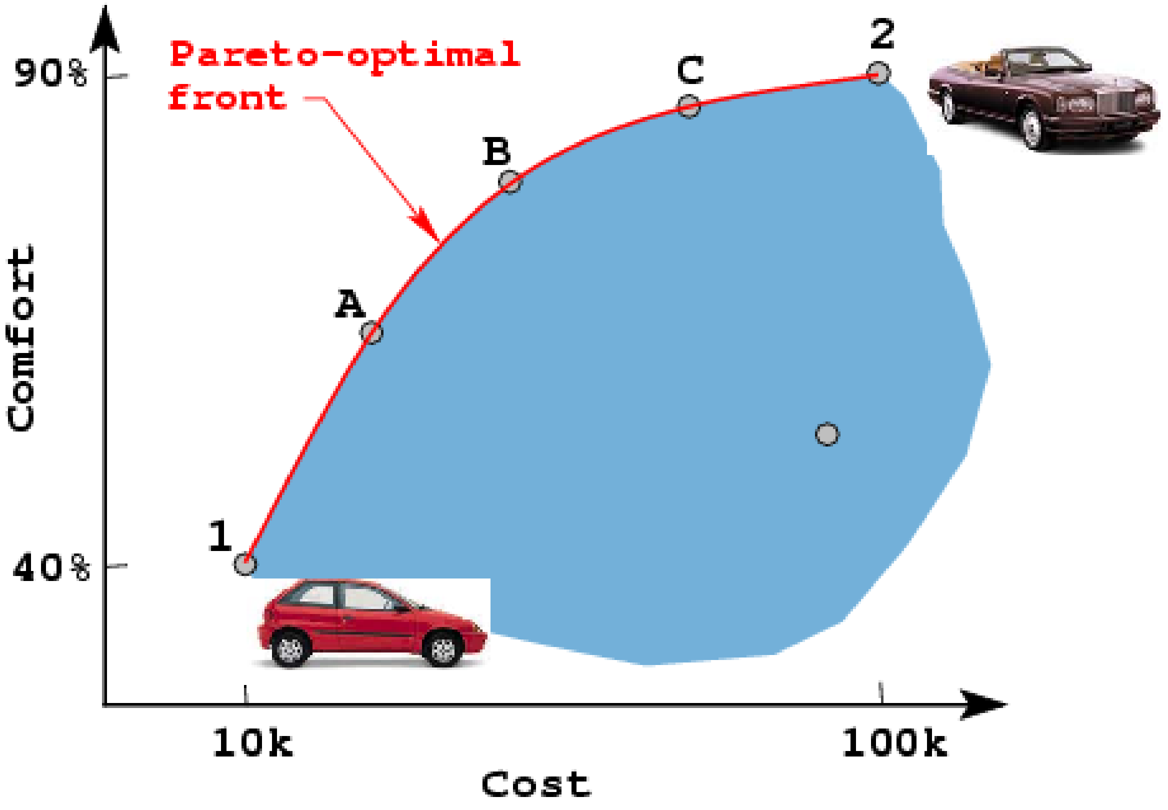


Deb, K. and Myburgh, C. (2017). A Population-Based Fast Algorithm for a Billion-Dimensional Resource Allocation Problem with Integer Variables. *European Journal of Operational Research, 261*(2), 460–474.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, London: Wiley,

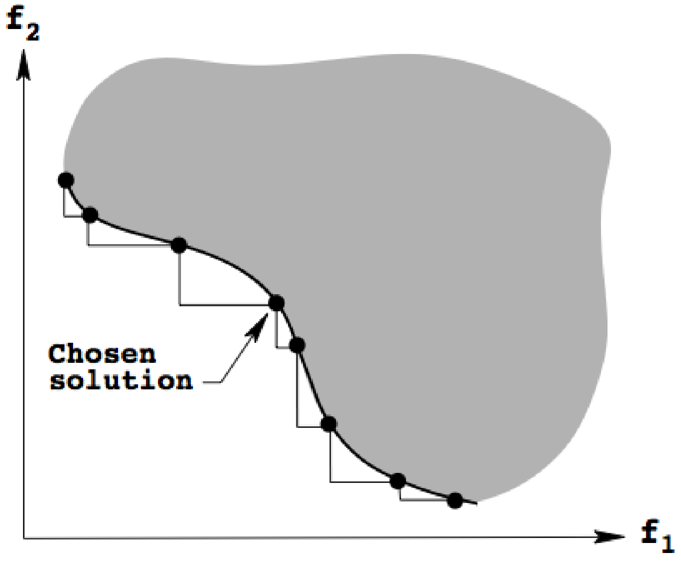
Multi-Objective Optimization

- Results in a set of Pareto-optimal solutions
- EO's population can store multiple solutions
  - Implicit parallelism helps
- NSGA-II, NSGA-III (commercialized) solve 2-15 objectives with constraints
- **Custom Evol. MO** (EMO) for applied problems



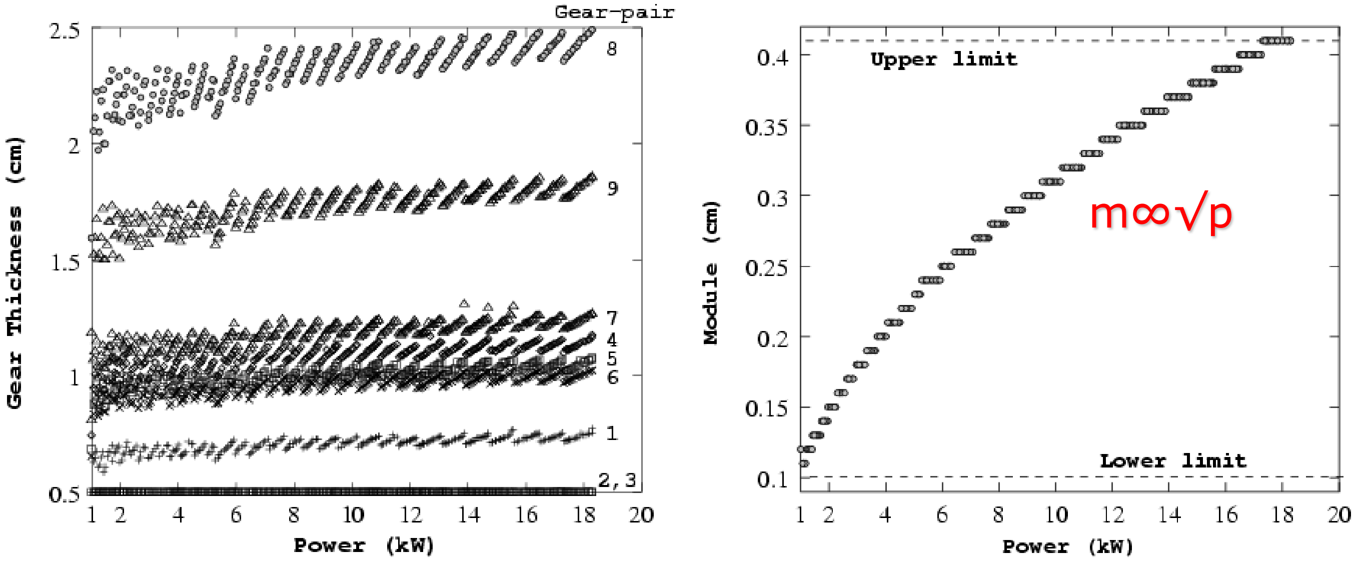
EMO and Decision-Making

- **Decision-making**: A systematic algorithmic approach
  - Trade-off analysis to choose a preferred solution
  - Interactive EMO-MCDM methods exist
- **"Innovization"**
  - Finding patterns in Pareto-optimal solution for knowledge discovery using AI/ML methods



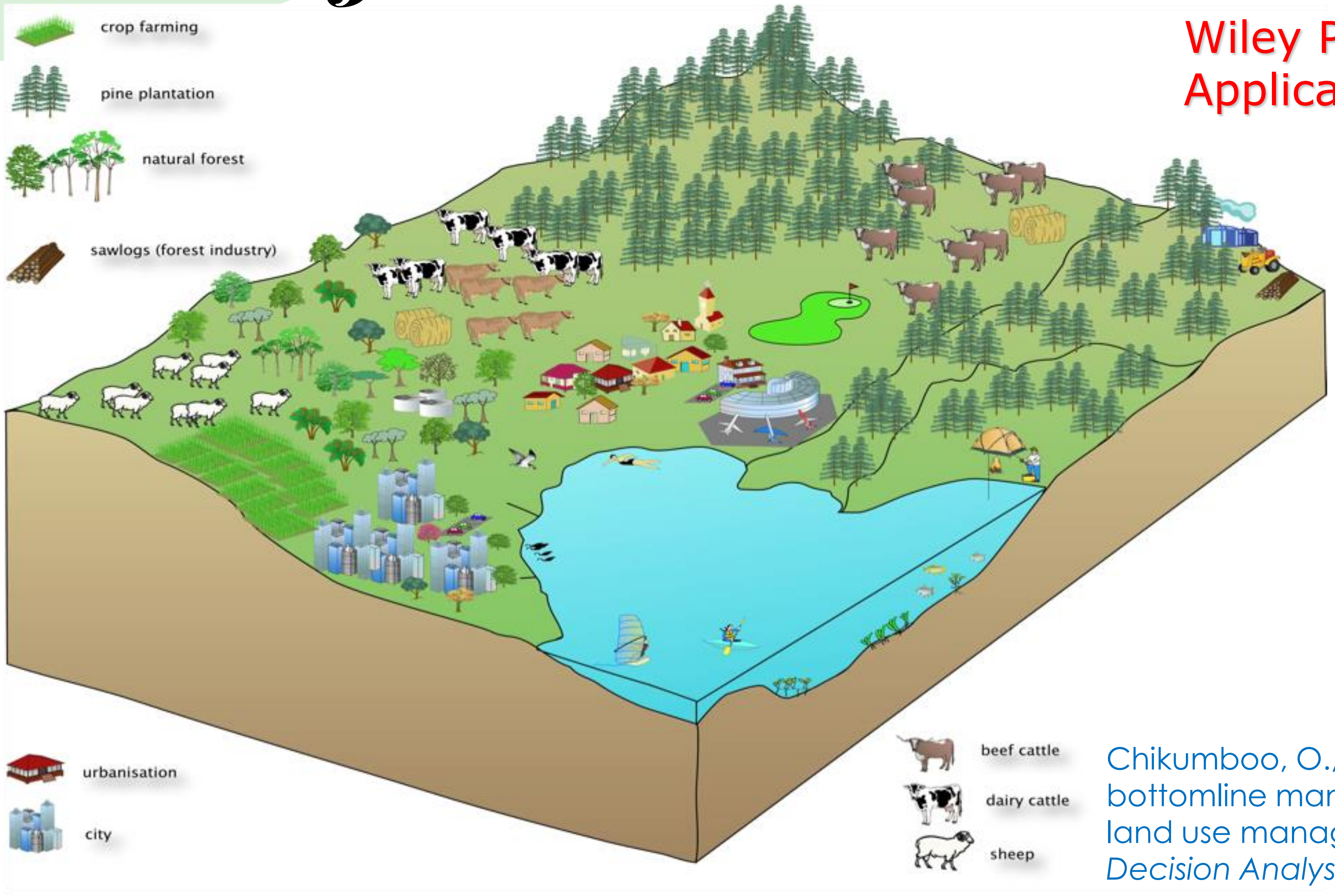
Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, London: Wiley,

A Gear-box design:



Deb, K. and Srinivasan, A. (2006). *Innovization*: Innovating design principles through optimization. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2006)*, New York: The Association of Computing Machinery (ACM), (pp. 1629–1636)

Case Study: New Zealand Land Use Management

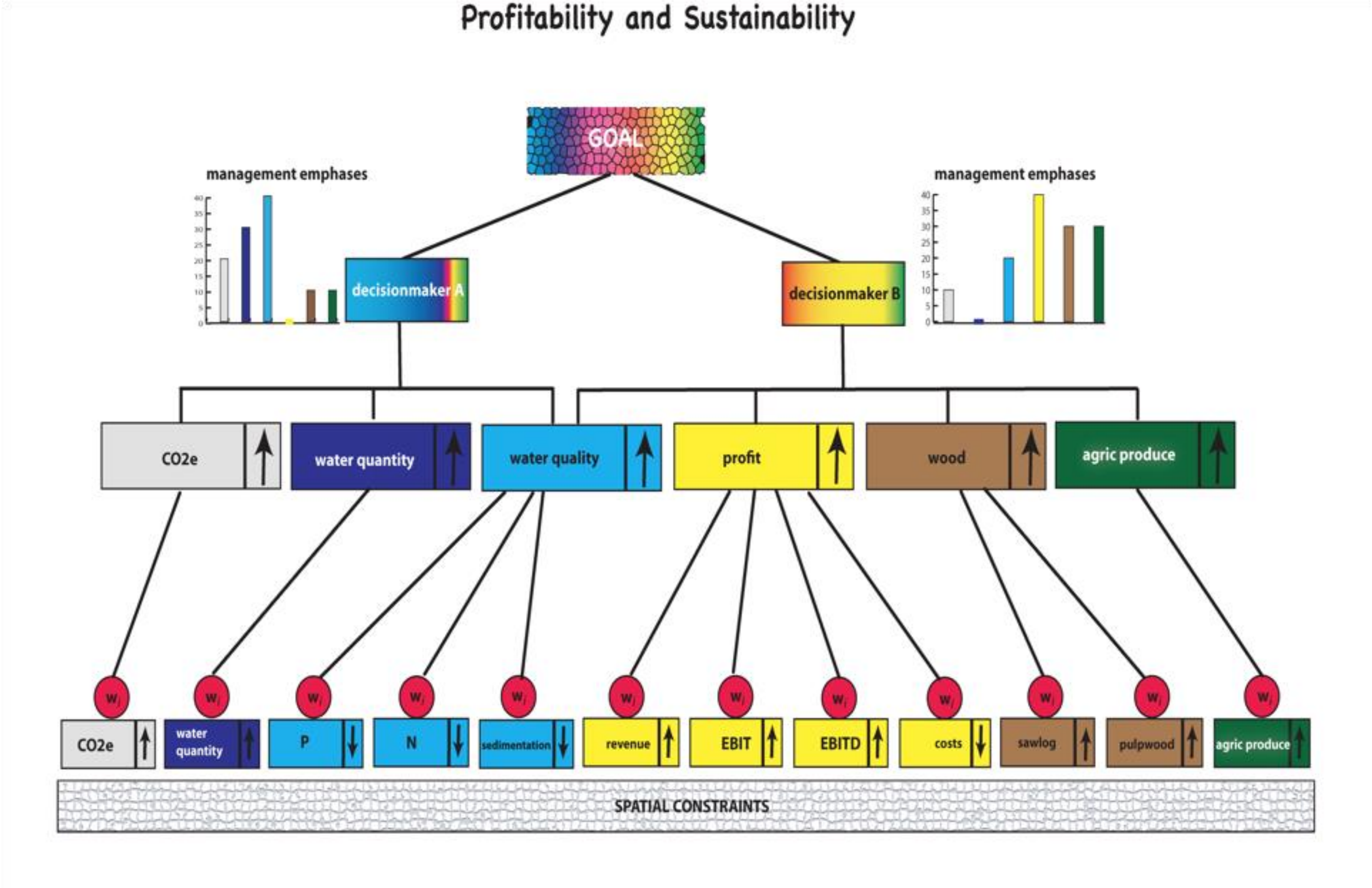


Wiley Practice Prize Winning Application (June 2013)

Principle applicable to other "Wicked" societal/industrial problems

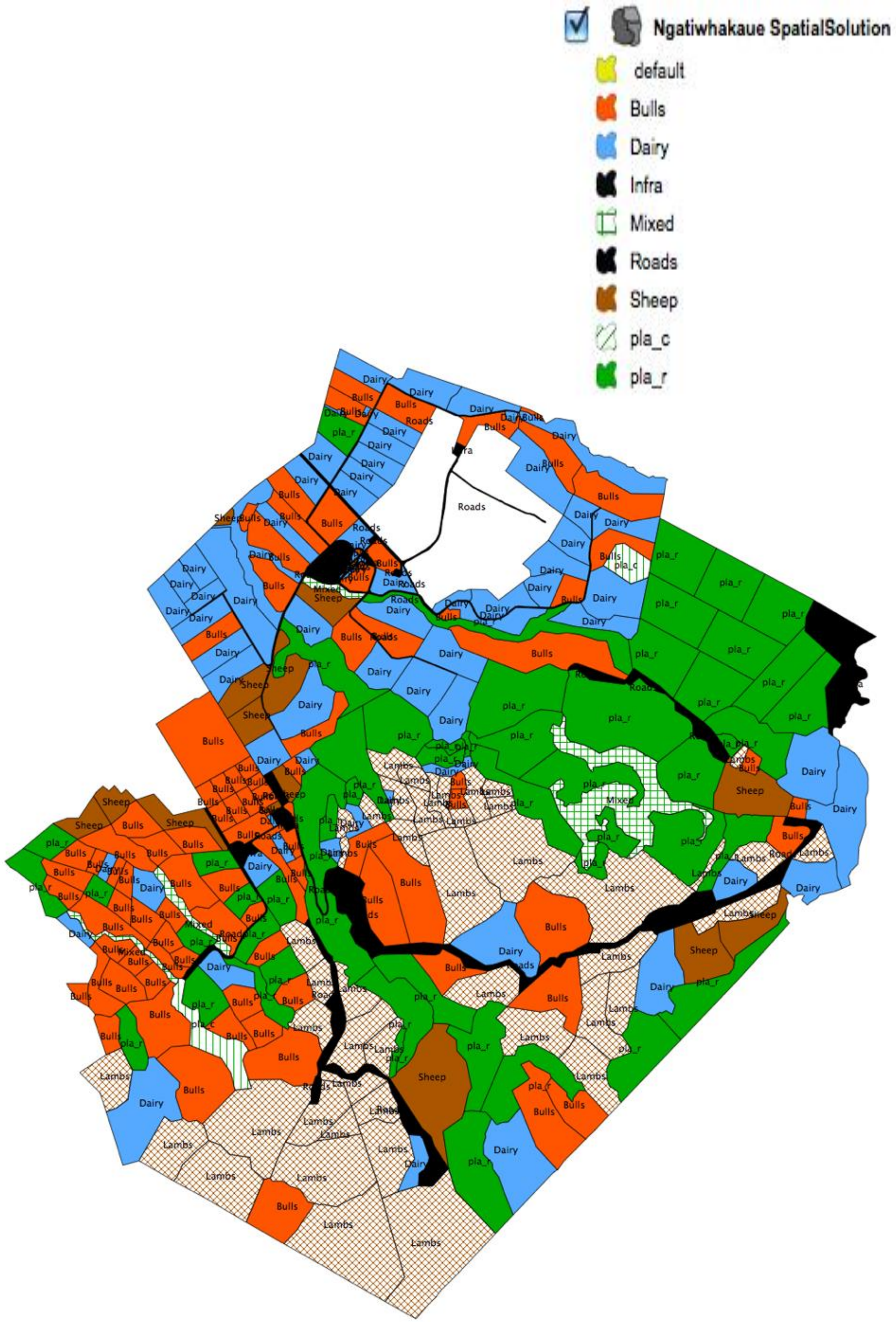
Chikumboo, O., Goodman, E., Deb, K. (2015). Triple bottomline many-objective-based decision making for a land use management problem. *Journal of Multi-Criterion Decision Analysis, 22*(3-4), 133–159.

14 Objectives – about profitability, production, and environment



Astronomical Search Space

- The space of land-use plans:
  - 315 independently manageable paddocks
  - Each having around 100 choices
  - Planning for 10 years
- Huge search space
  - $(100^{315})^{10}$ solutions
  - $\sim 10^{82}$ atoms in the universe
- Large number of objectives



Financials

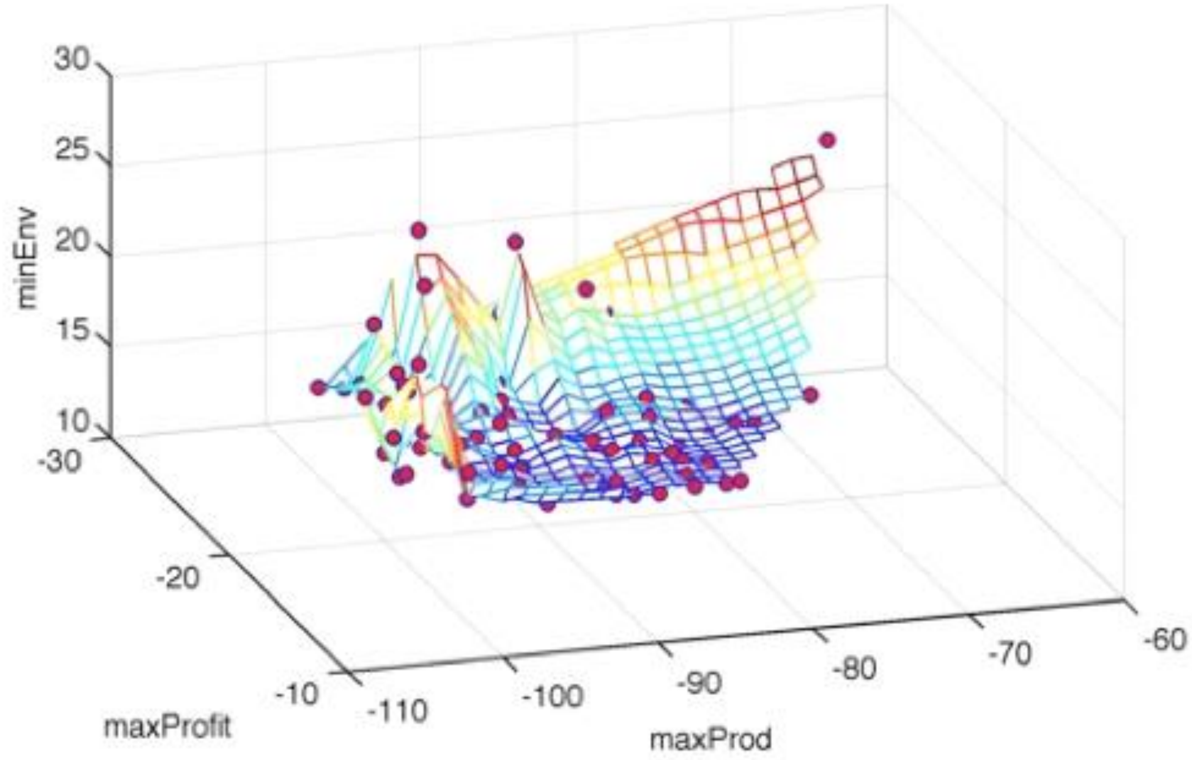
Productivity

Environment

Pareto front =

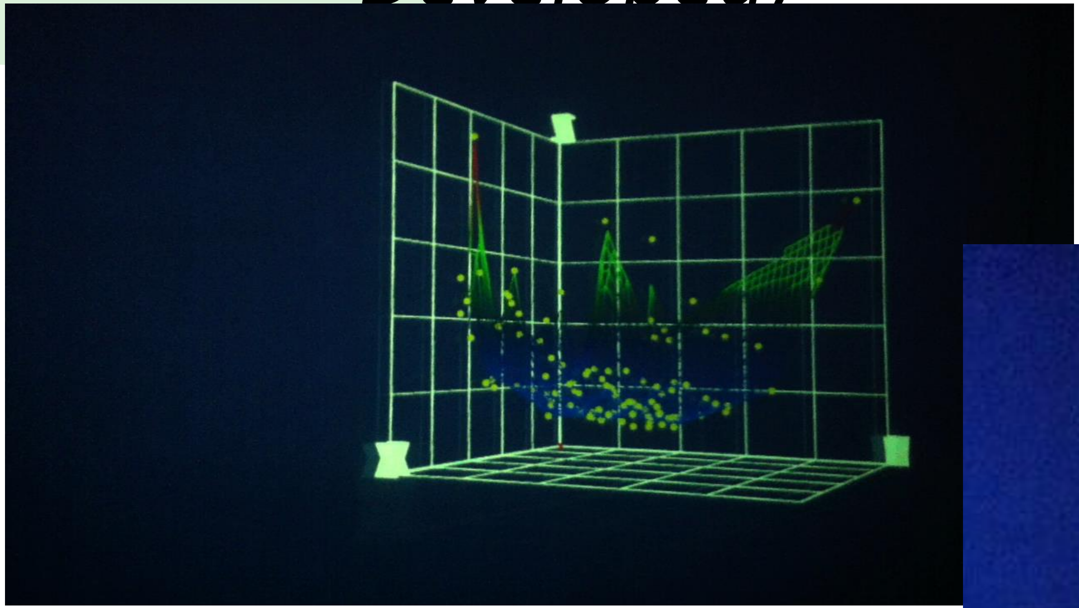
f(MaxProfit, MaxProd, MinEnvImpact)

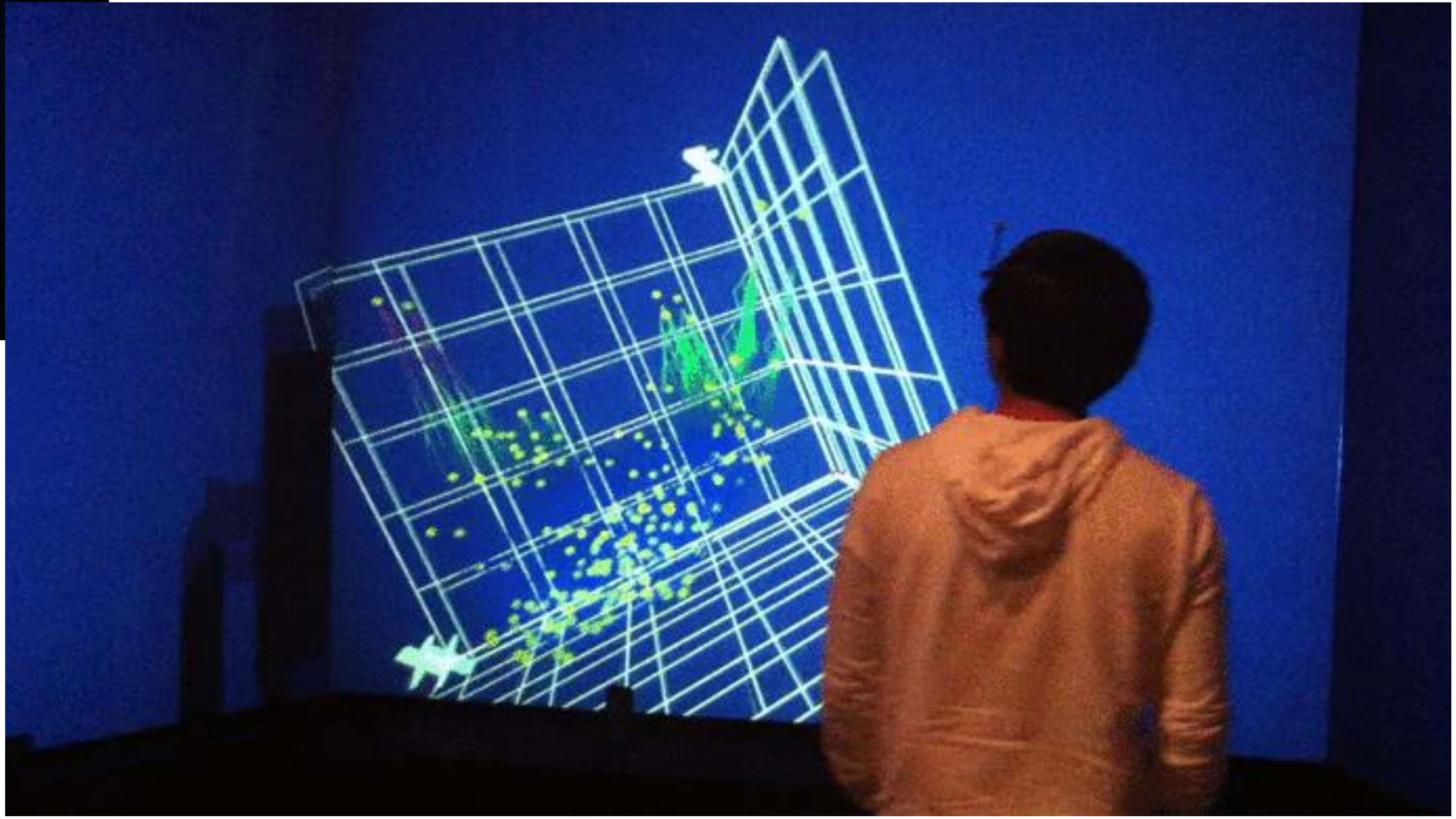
Developed WISDOM Approach using a customized EMO
A Typical Result:

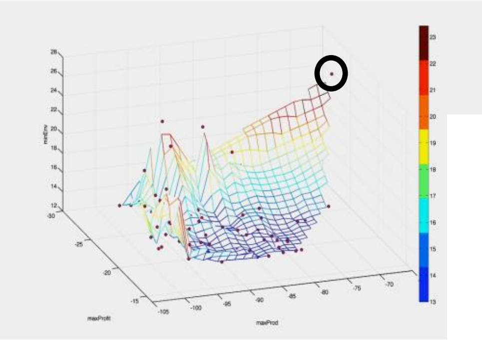


Decision-making later

Visualization Tools to Help Understand Trade-off (A VR-based System Developed)

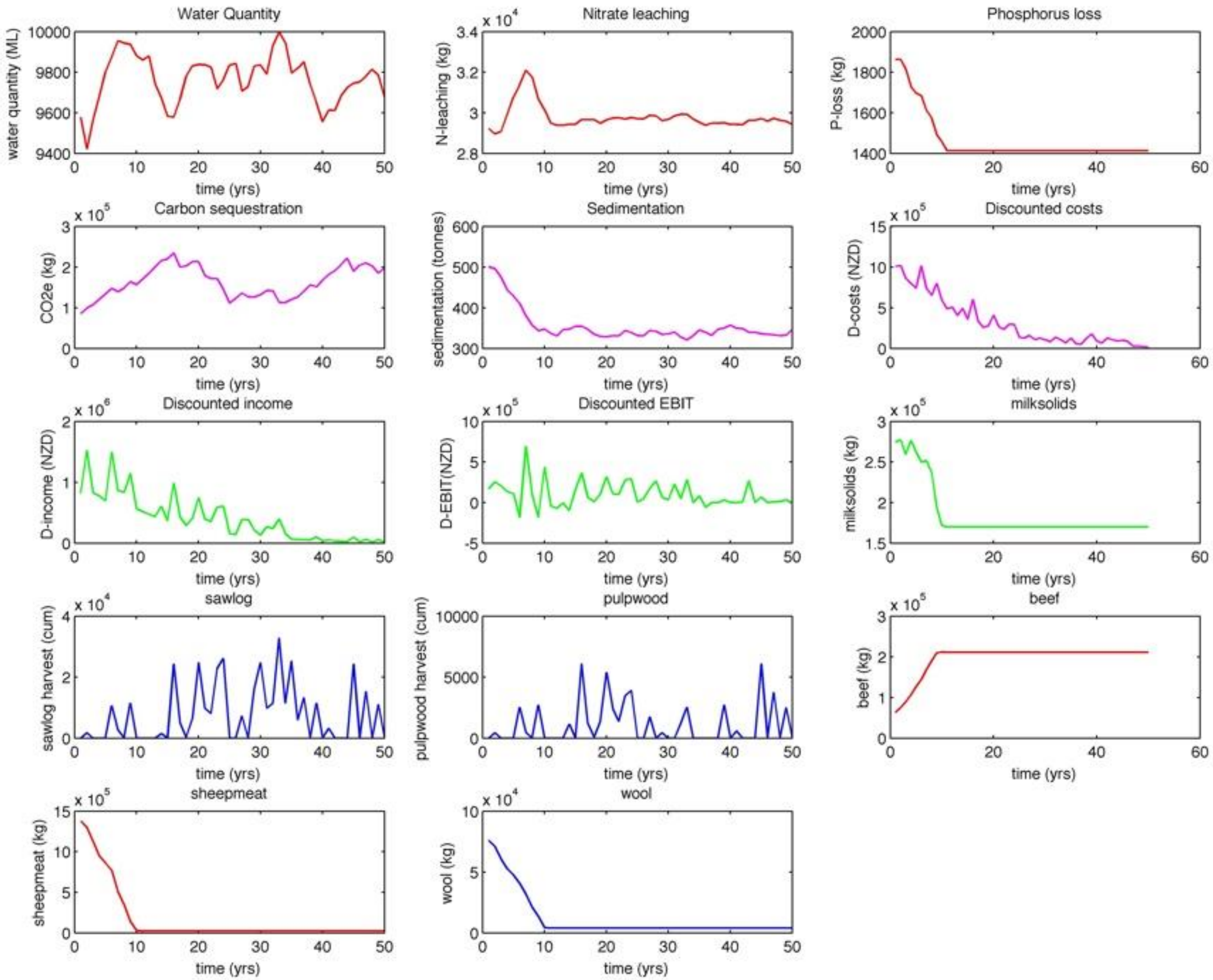






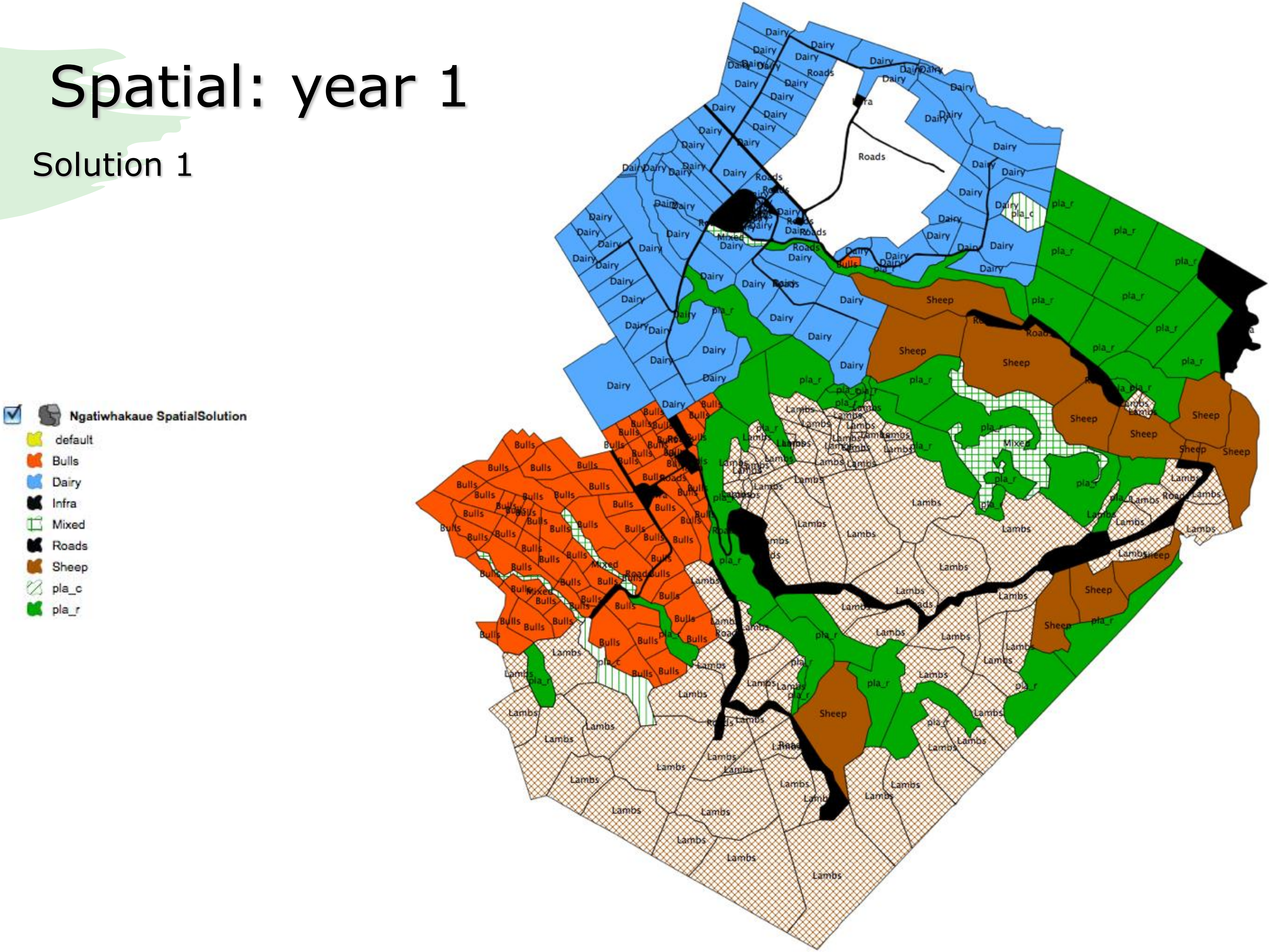
Environment-friendly Solution

- Sheep substantively lower;
- Dairy drastically reduced;
- More and less frequent forest harvesting;
- Substantial increase in beef cattle.

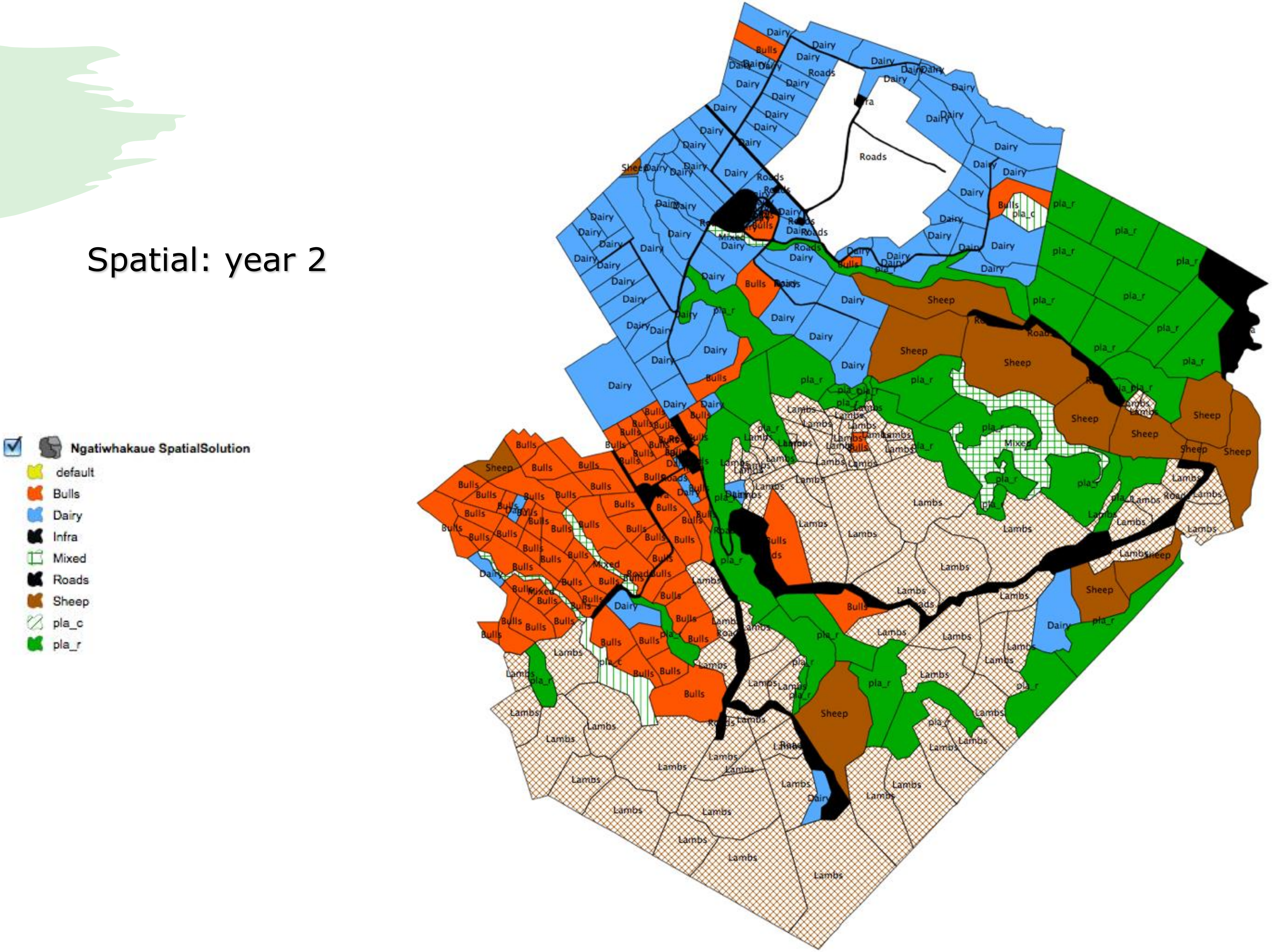


Spatial: year 1

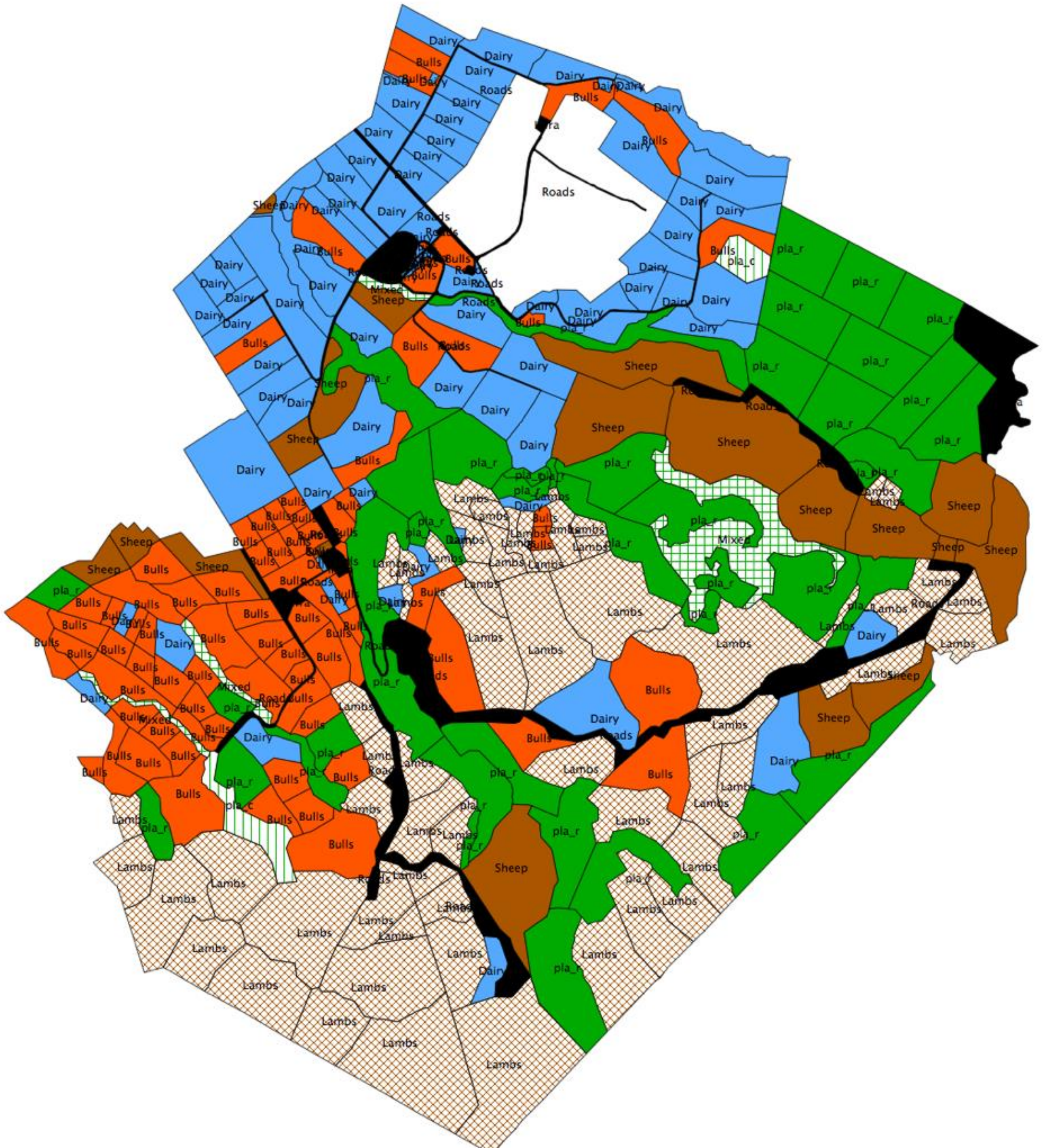
Solution 1



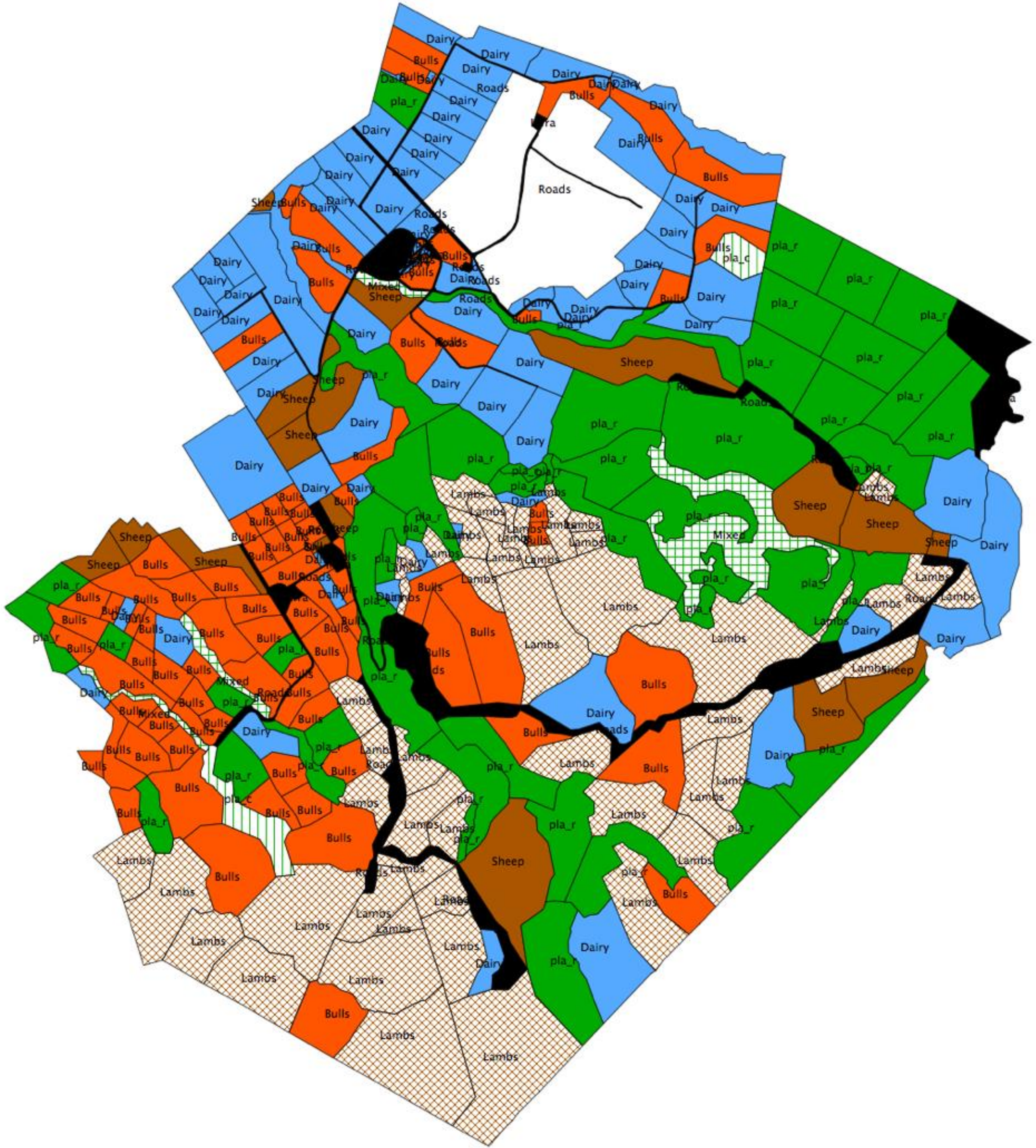
Spatial: year 2



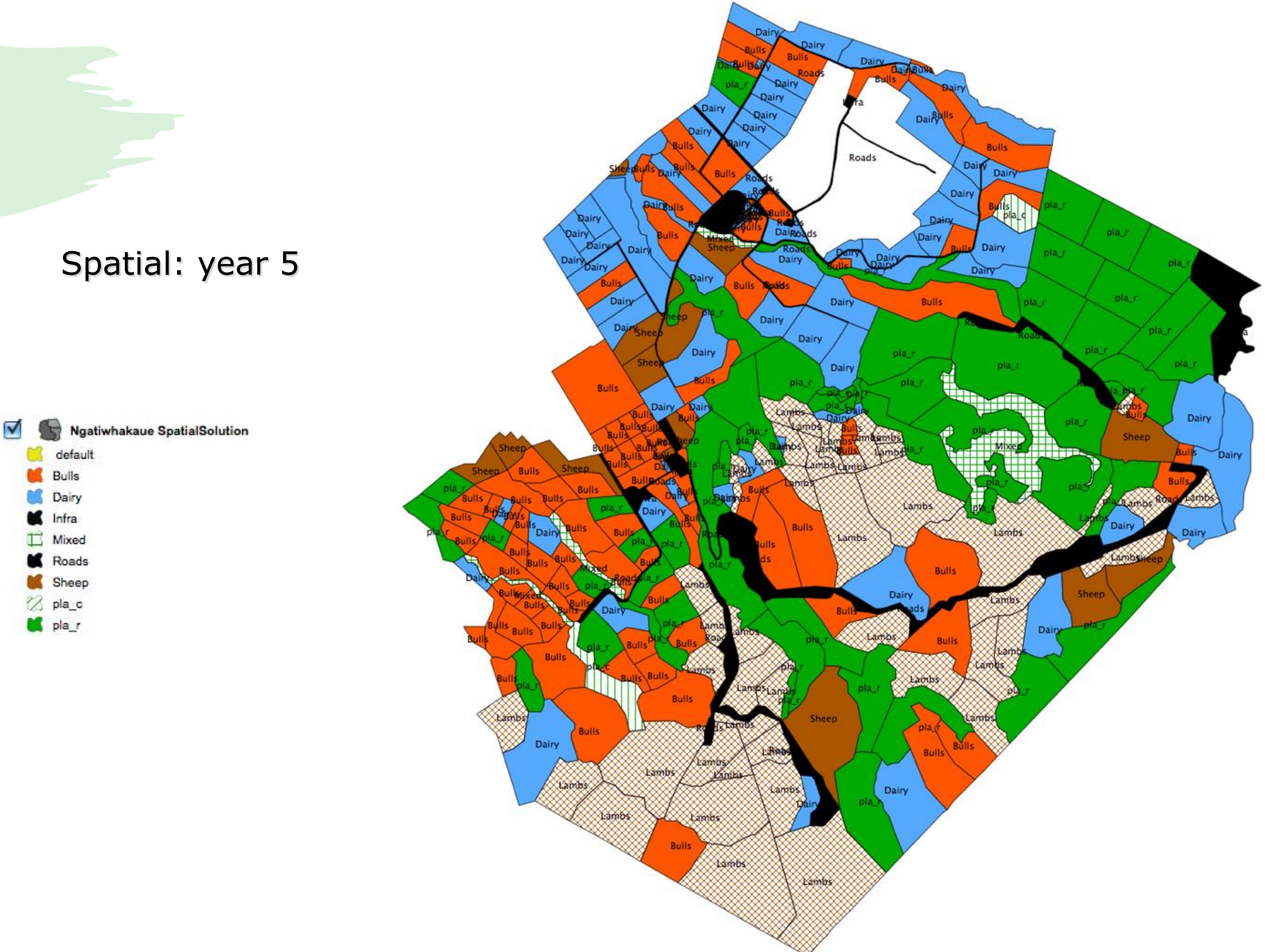
Spatial: year 3



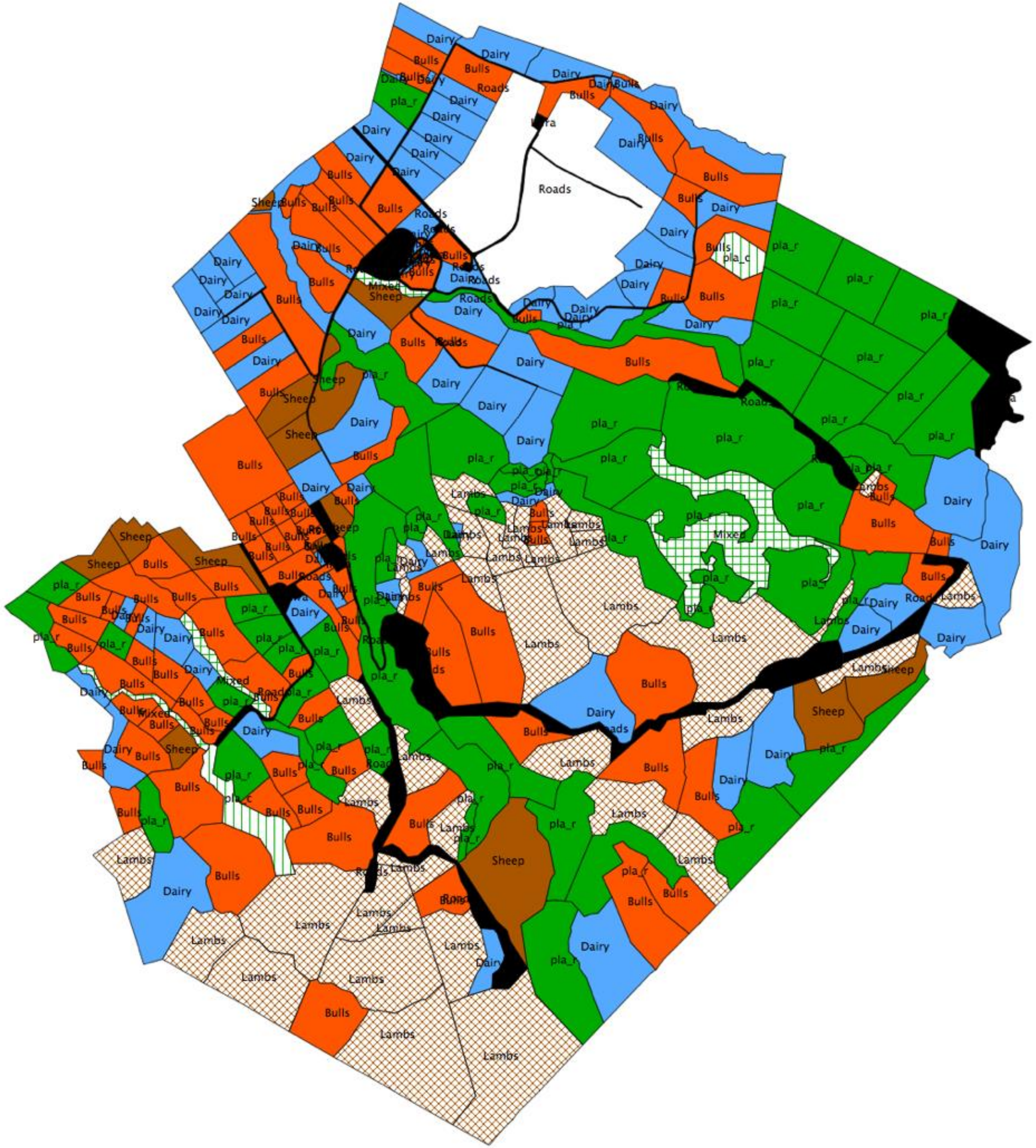
Spatial: year 4



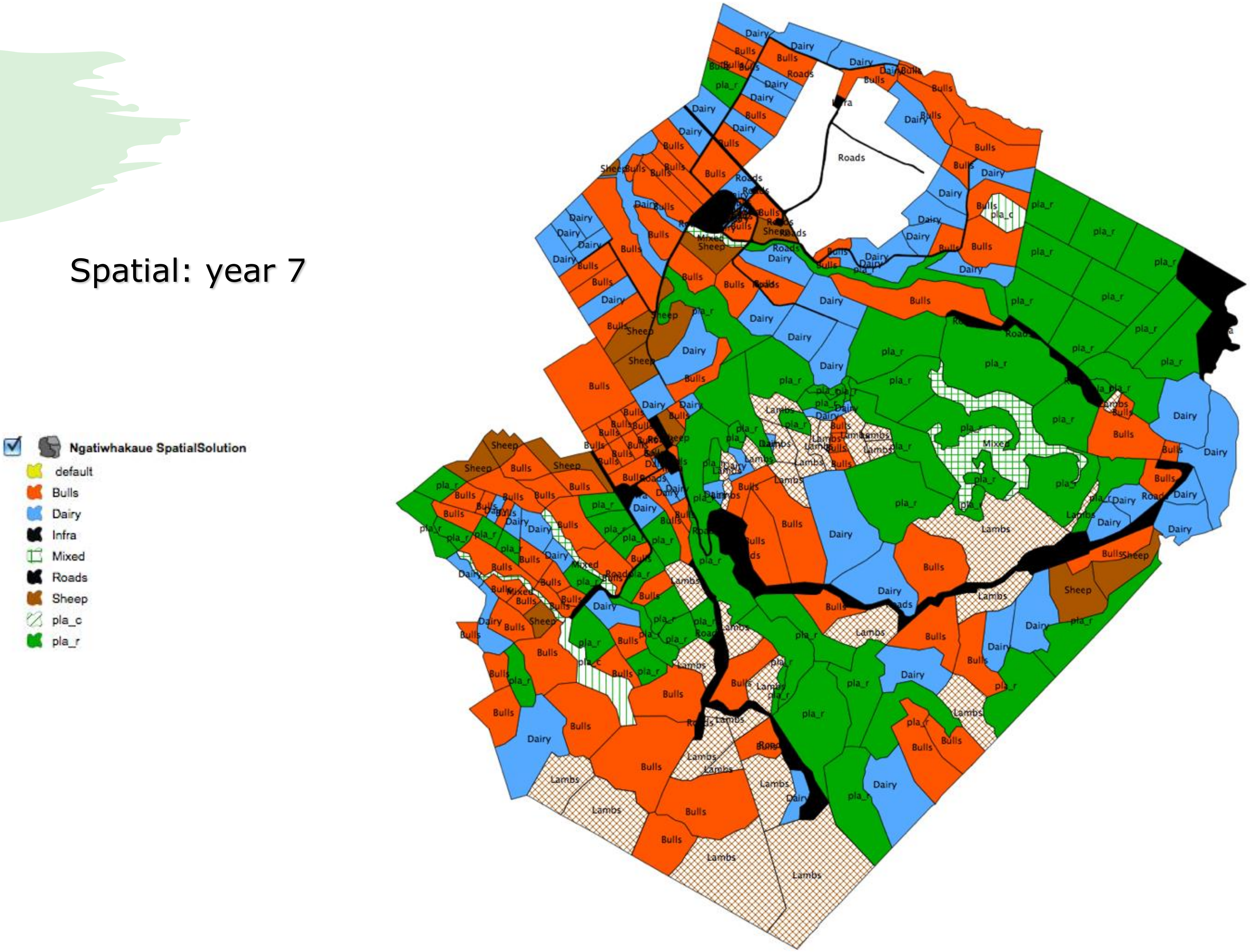
Spatial: year 5



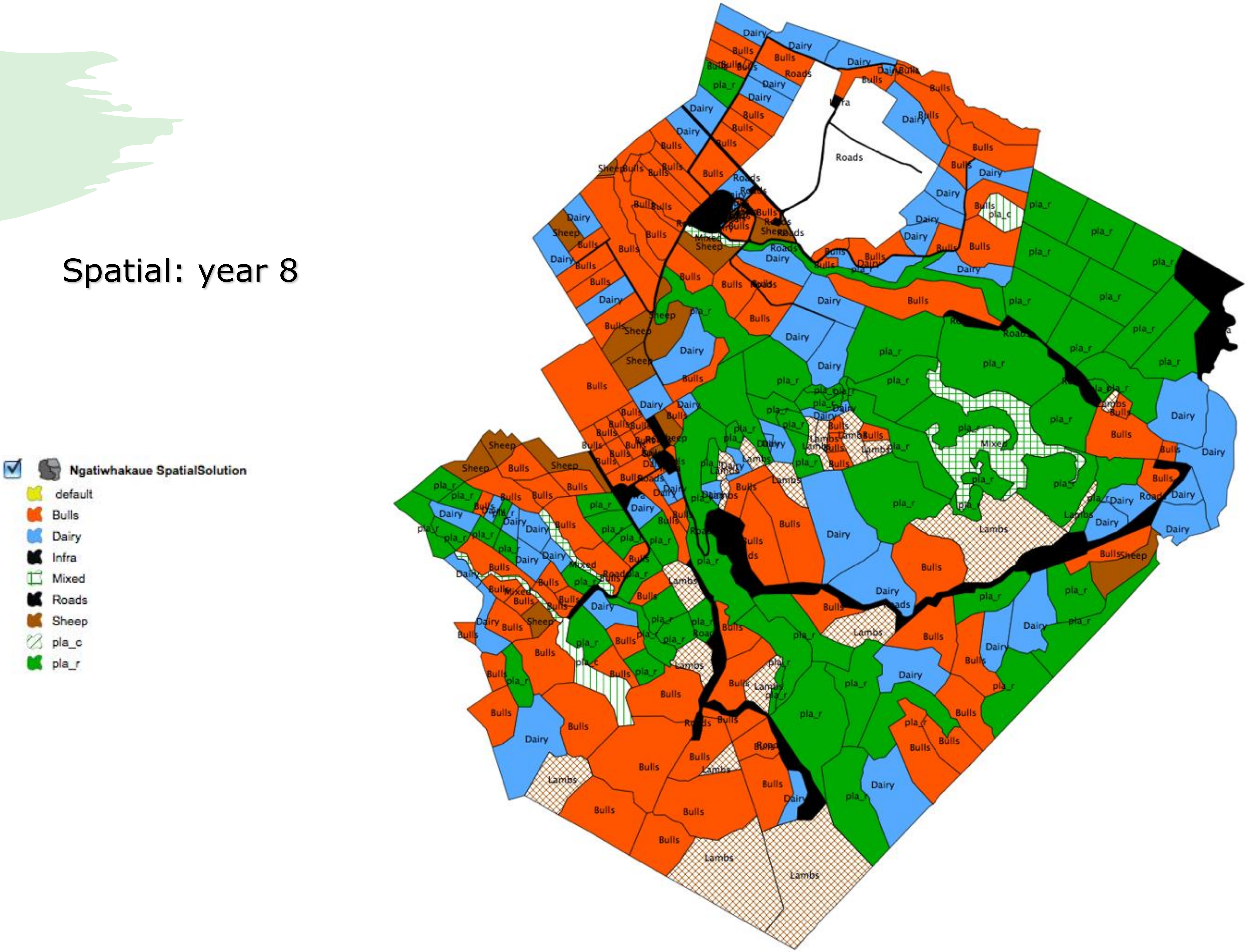
Spatial: year 6



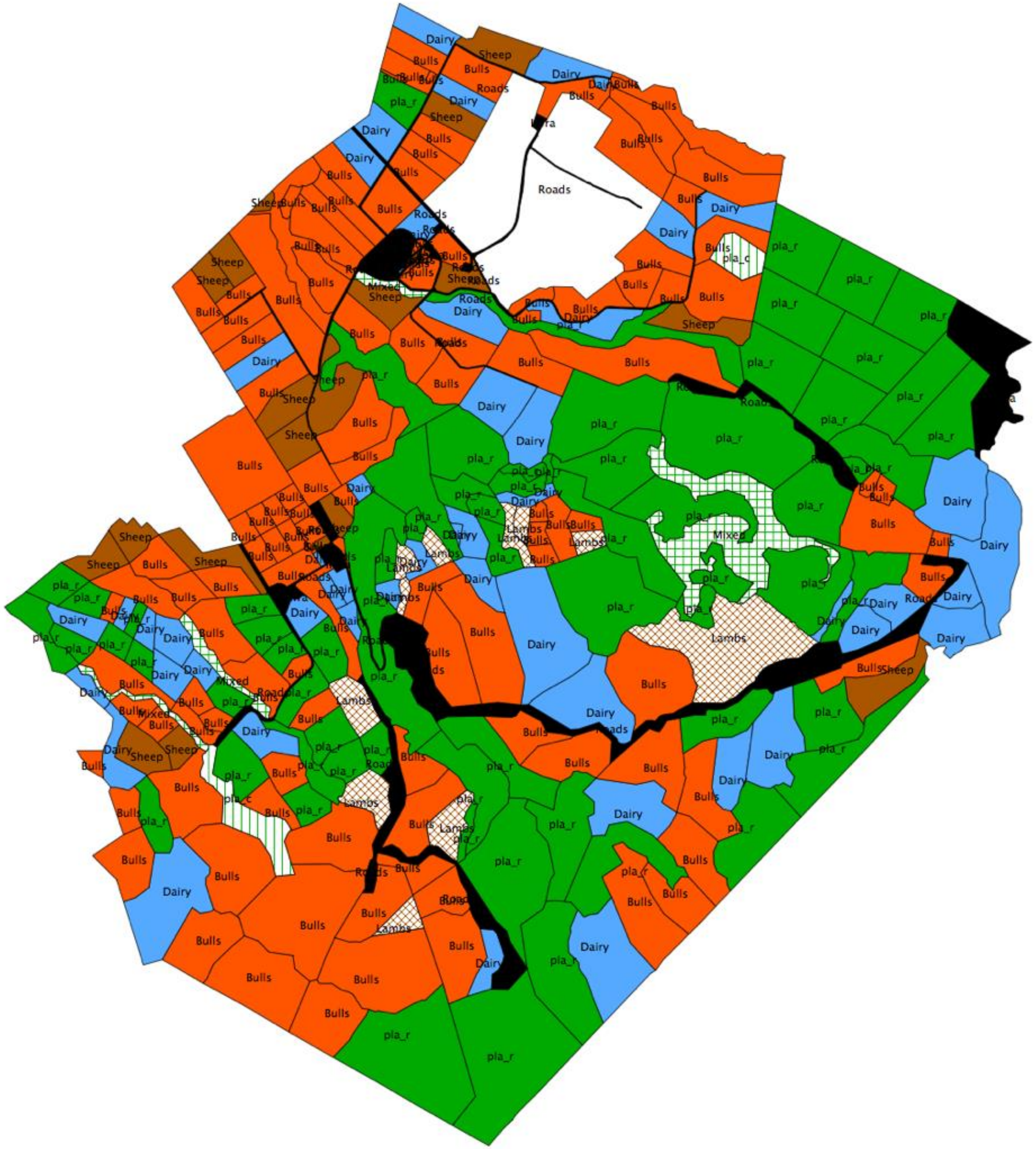
Spatial: year 7



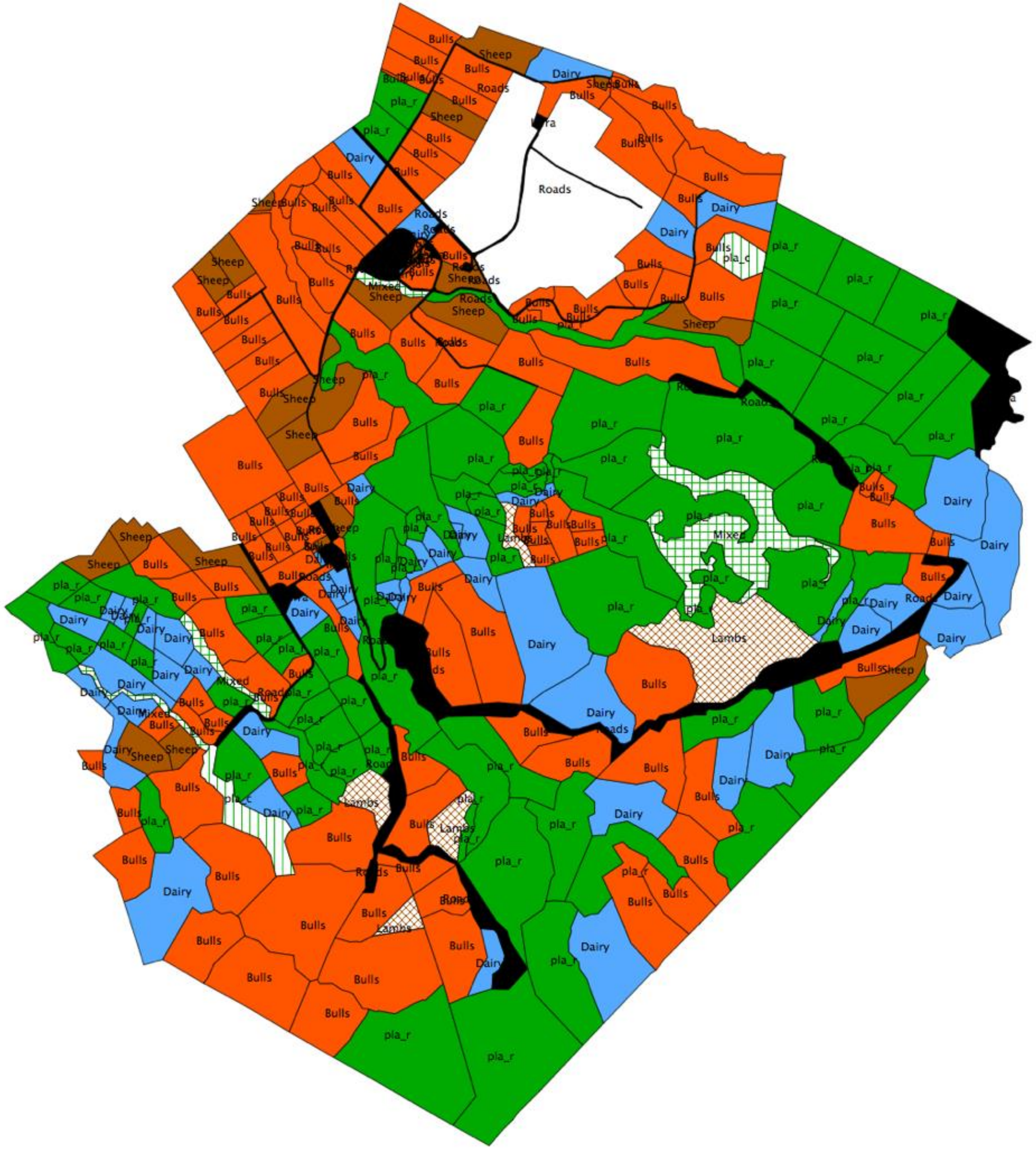
Spatial: year 8



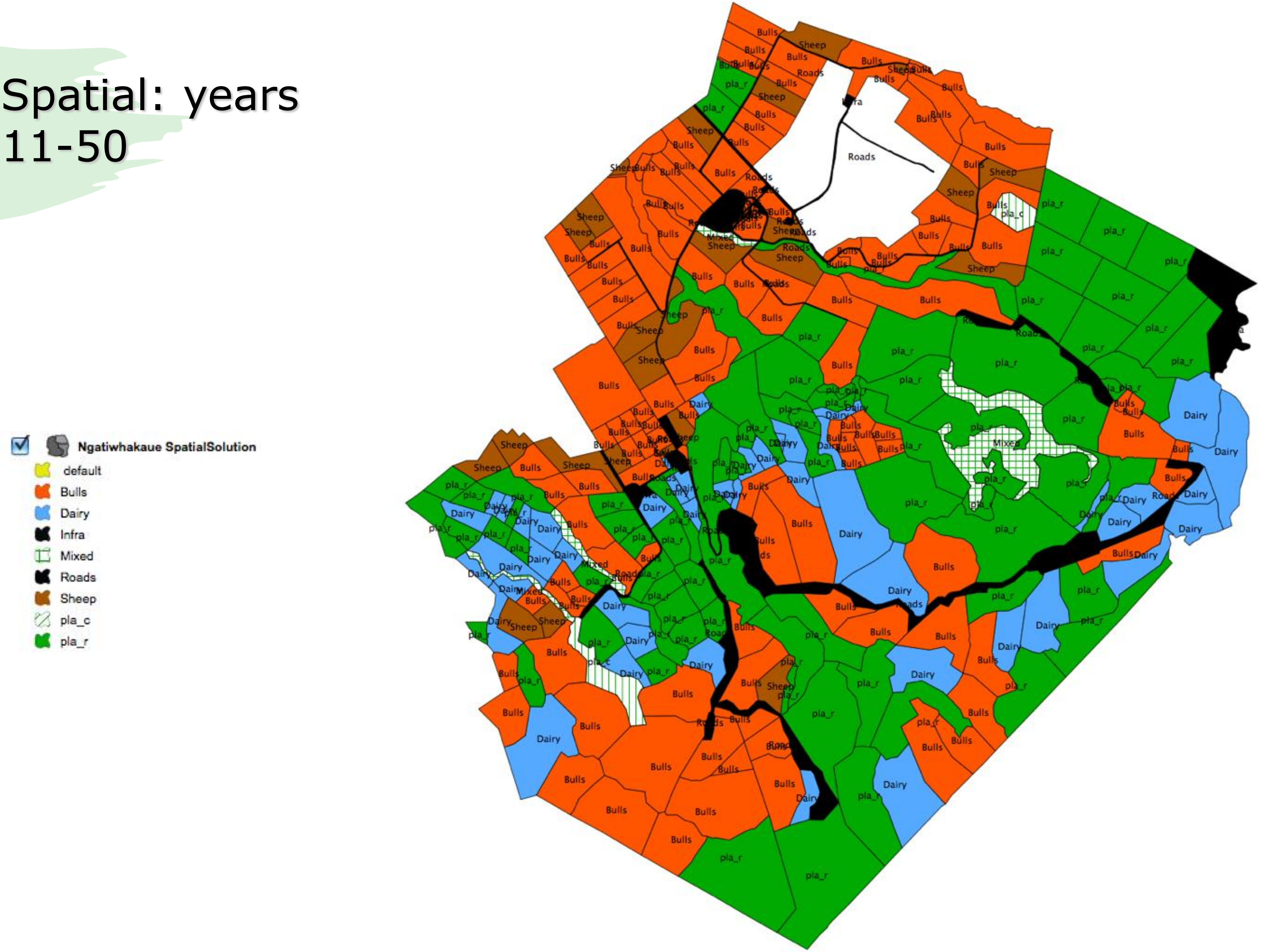
Spatial: year 9



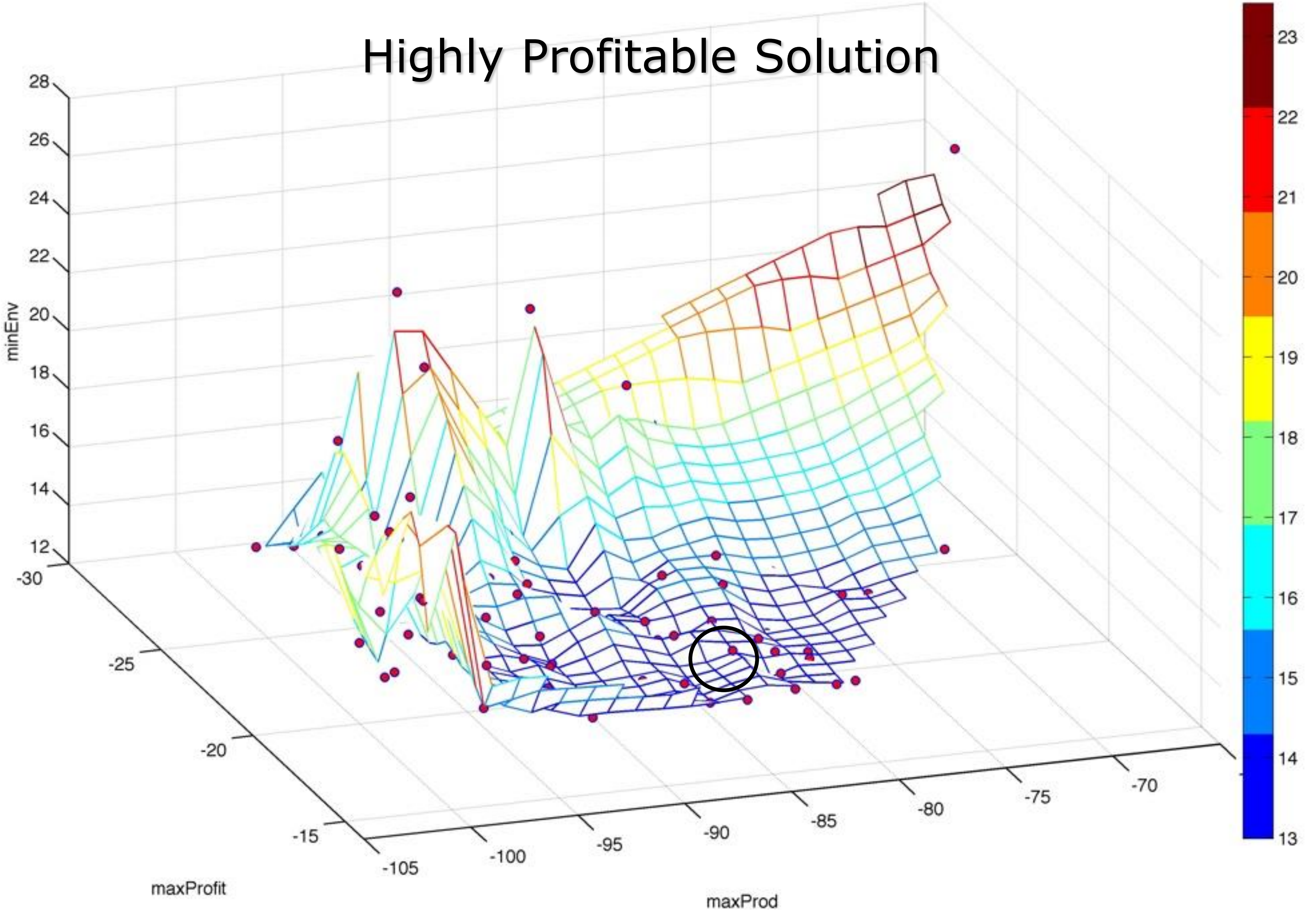
Spatial: year 1(



Spatial: years 11-50

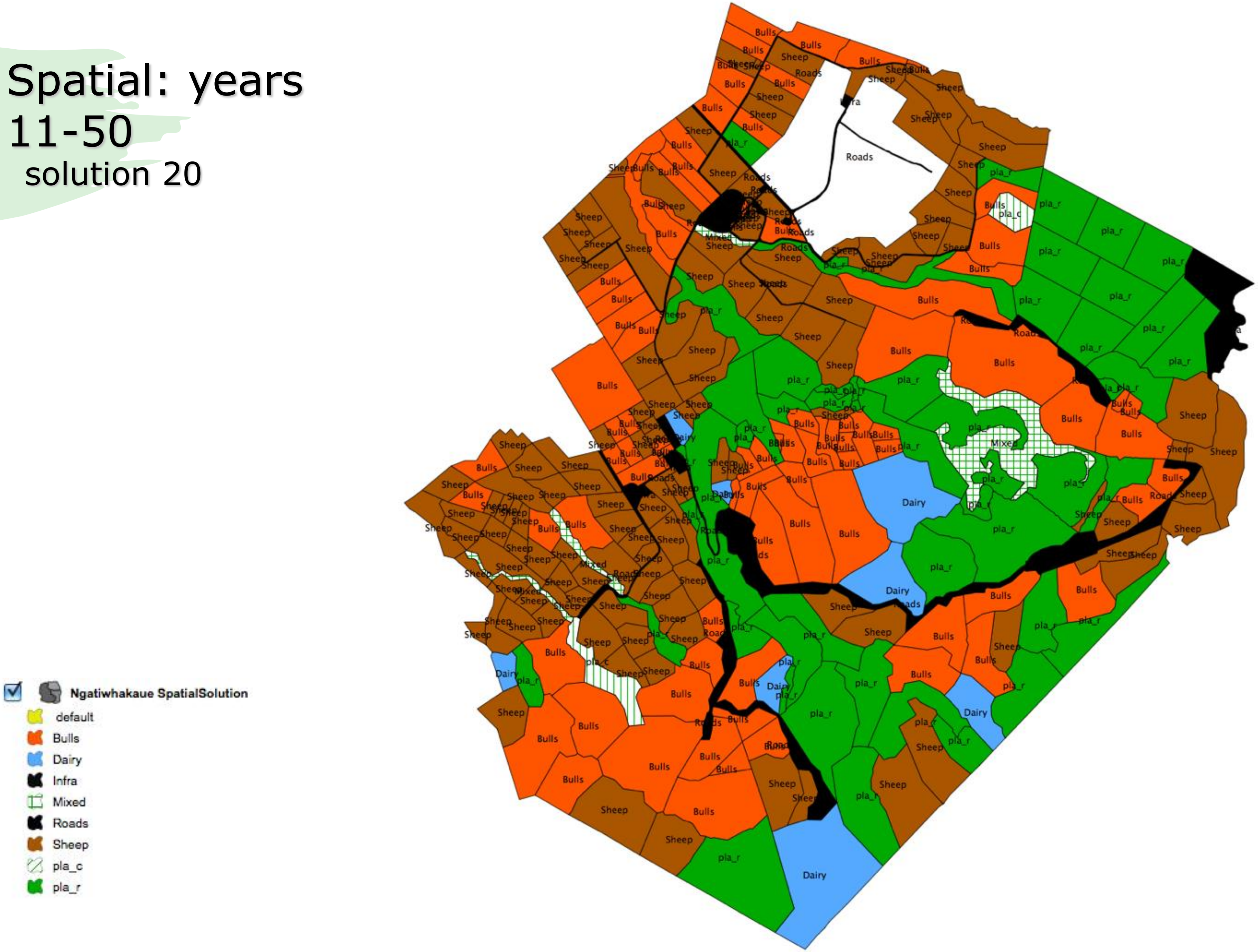


Highly Profitable Solution

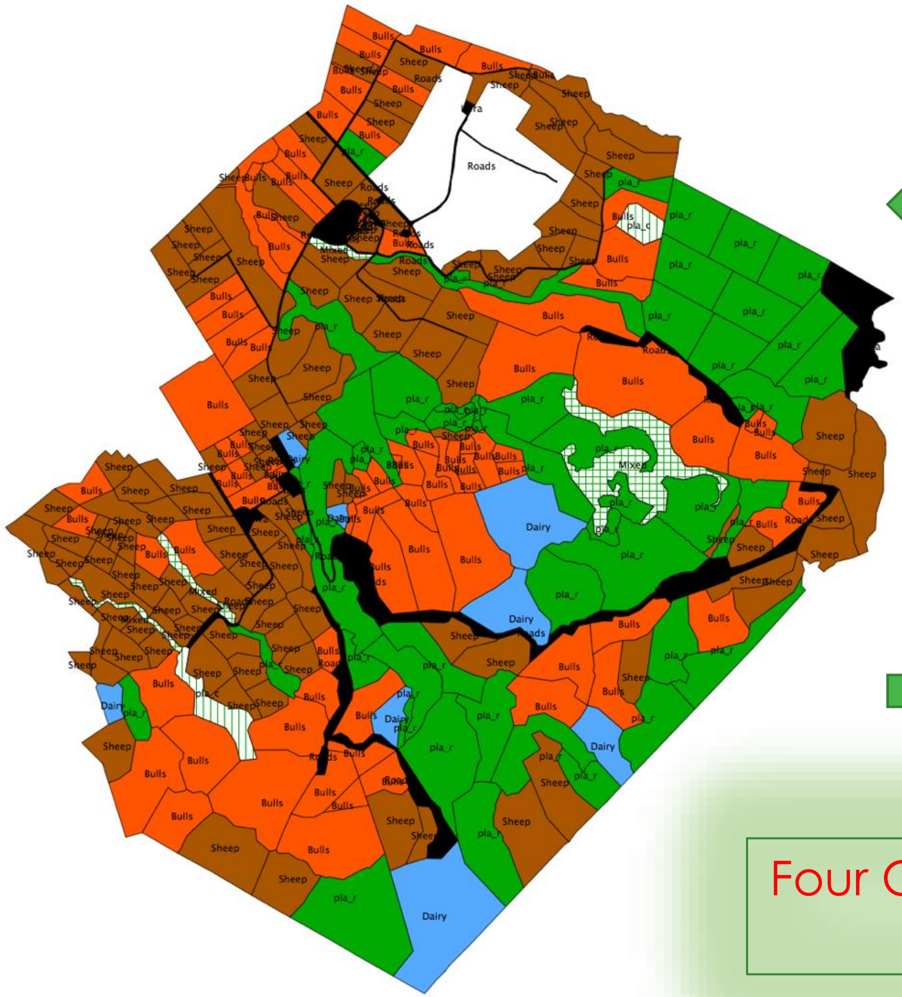


Spatial: years 11-50

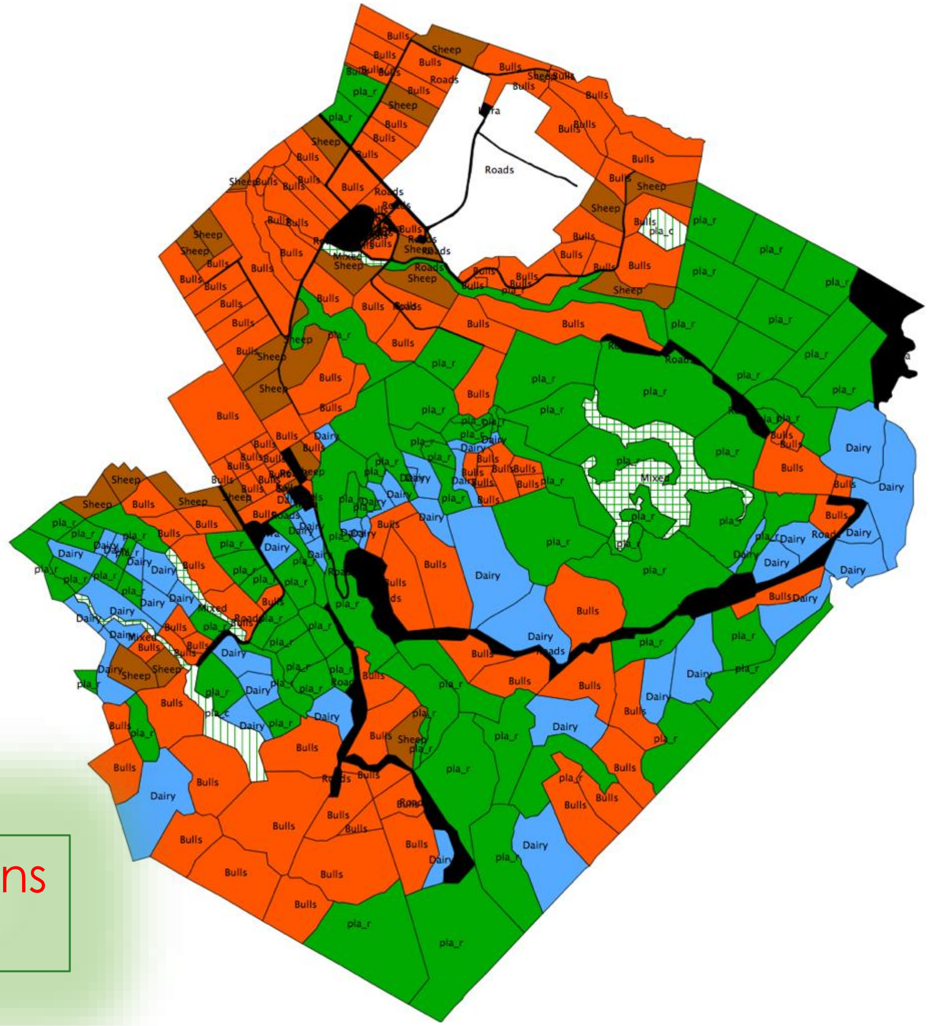
solution 20



Highly profitable solution



“Environmentally friendly” solution



Maori DMs

100 Intermediate Solutions Exist

• • •

Government DMs

Four Compromised Solutions Found

The Maori participants told the government that WISDOM is how they wanted to do future planning

Final Decision-Making with Four Solutions

- Four Stake-holder Groups:
  - Maori landowner, Rotorua district planner, Senior Waikato regional planner, and two senior forest planners
  - Each interview took 90-120 min
- **Analytical heierachy process** (AHP) to decide on one preferred solution
  - Pair-wise comparison by stake-holders
  - An analytical process, thereafter

Final AHP Process & Solu

- Pair-wise decision-making by a group
- One of Four solutions chosen

SCENARIOS	PRIORITIES	IDEALISED PRIORITIES
1	0.1663	0.7387
2	0.1774	0.7882
3	0.2251	1 WINNER
4	0.1345	0.5971

Astronomical Search Space

Pareto-optimal Solutions (10-100s)

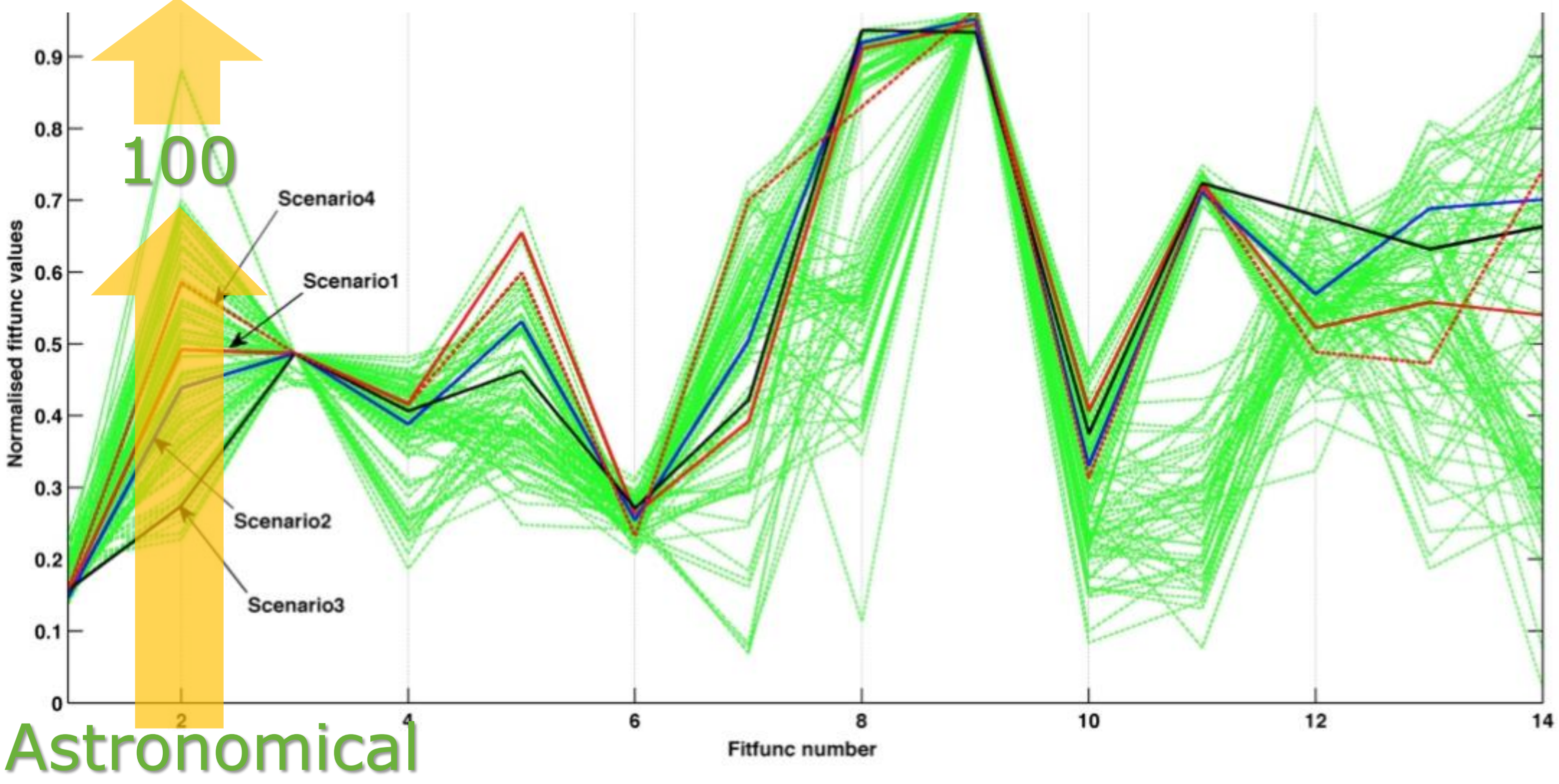
Handful of Trade-off Solutions (3-5)

A Single Compromise Solution (1)

Multi-Objective Optimization (EMO)

Group Decision-Making

Final Choic e (AHP)

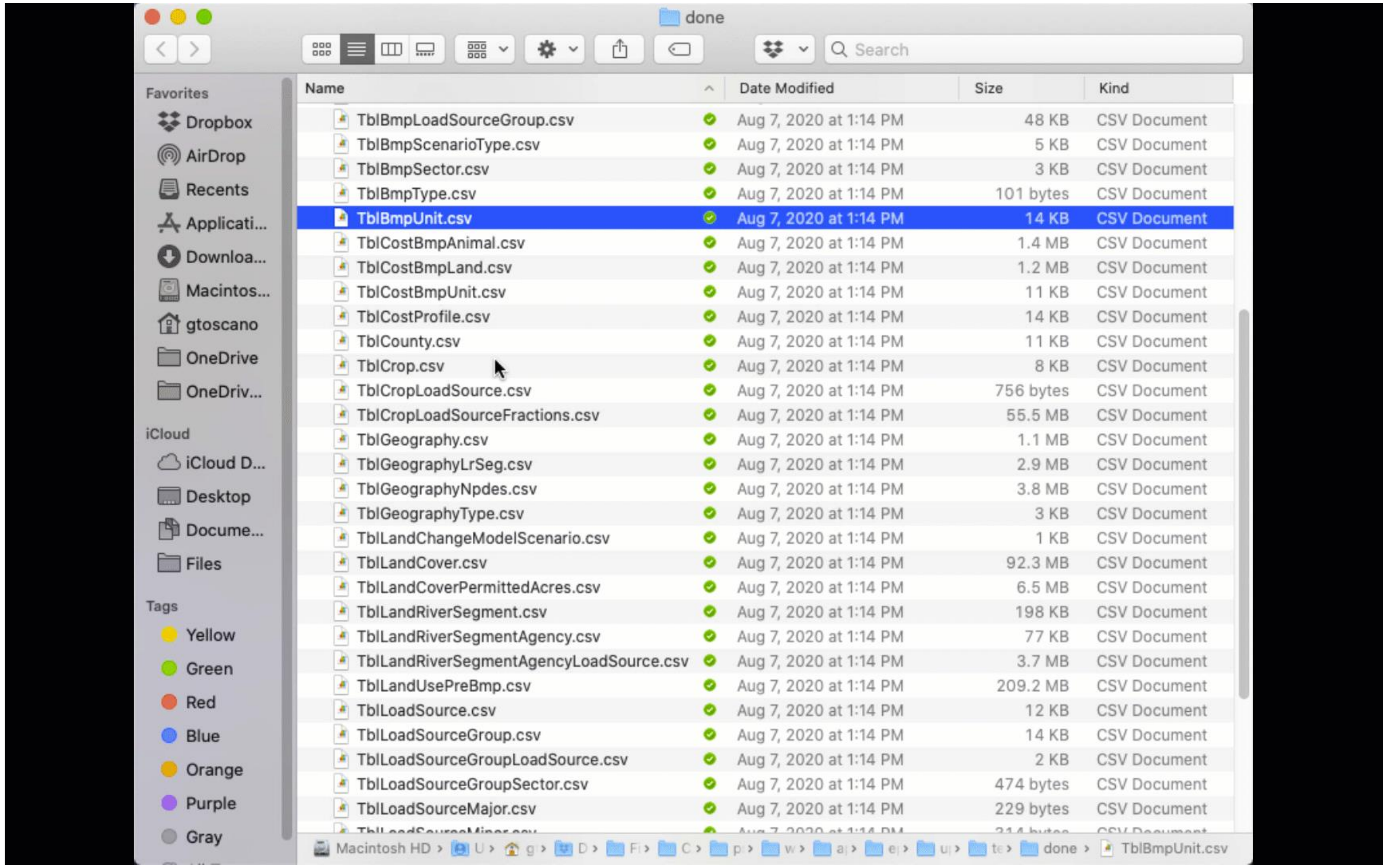


CAST System and Optimization (Current Study)

1. CAST database (built from CSV provided files)
2. Queries and extract information from the built database
3. Installation and testing of Bayota/VICO software
4. Development of a base evolutionary algorithm for testing purpose
5. Hybrid method development
6. CAST system integration: Immediate plan

CAST System and Optimization: CAST Database

- 56 CSV files
- Easy but slow access to data
- Difficult to analyze information
- To have a “working-in-situ” experience, we have **built a local database** using the provided CSVs



CAST System and Optimization: Queries and extra[redacted]

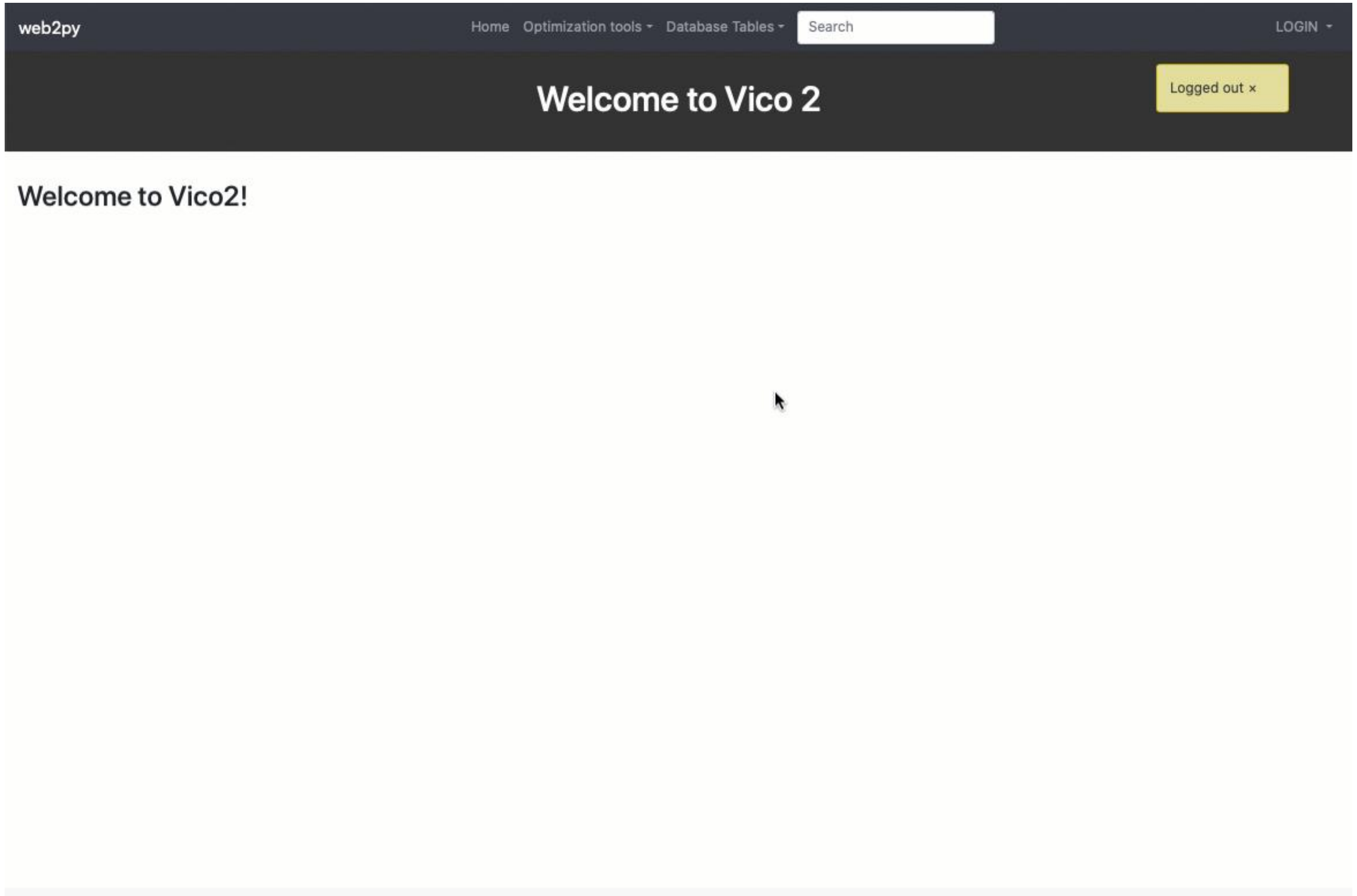
Using the Built-Database:

- Fast access to data
- Relations in a glimpse
- Database structure
- Standard queries
- Complex queries
    - select * from load_source where land_use_acres == 'T' and id in (select DISTINCT (load_source_id) from bmp_efficiency where land_river_segment_id in (select id from land_river_segment where county_id == (SELECT id from county where name == 'Bradford')));

The Built-Database is easily expandable

CAST System and Optimization: Web Interface

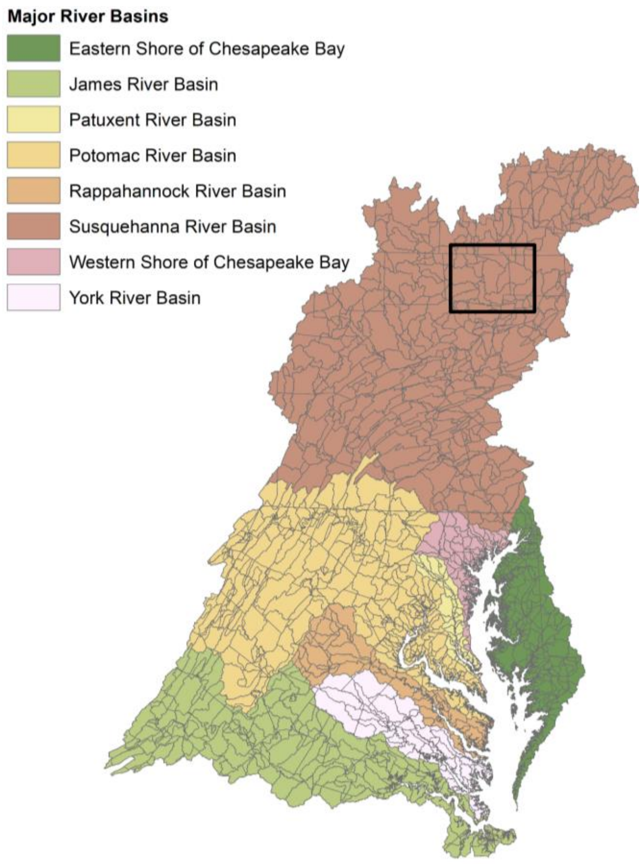
- **Built a Web interface**
- Process data and call optimization software easily and quickly
- Help to share and analyze data and results for optimization purpose and also for EPA users



CAST System and Optimization: Case Study: Bradford, PA

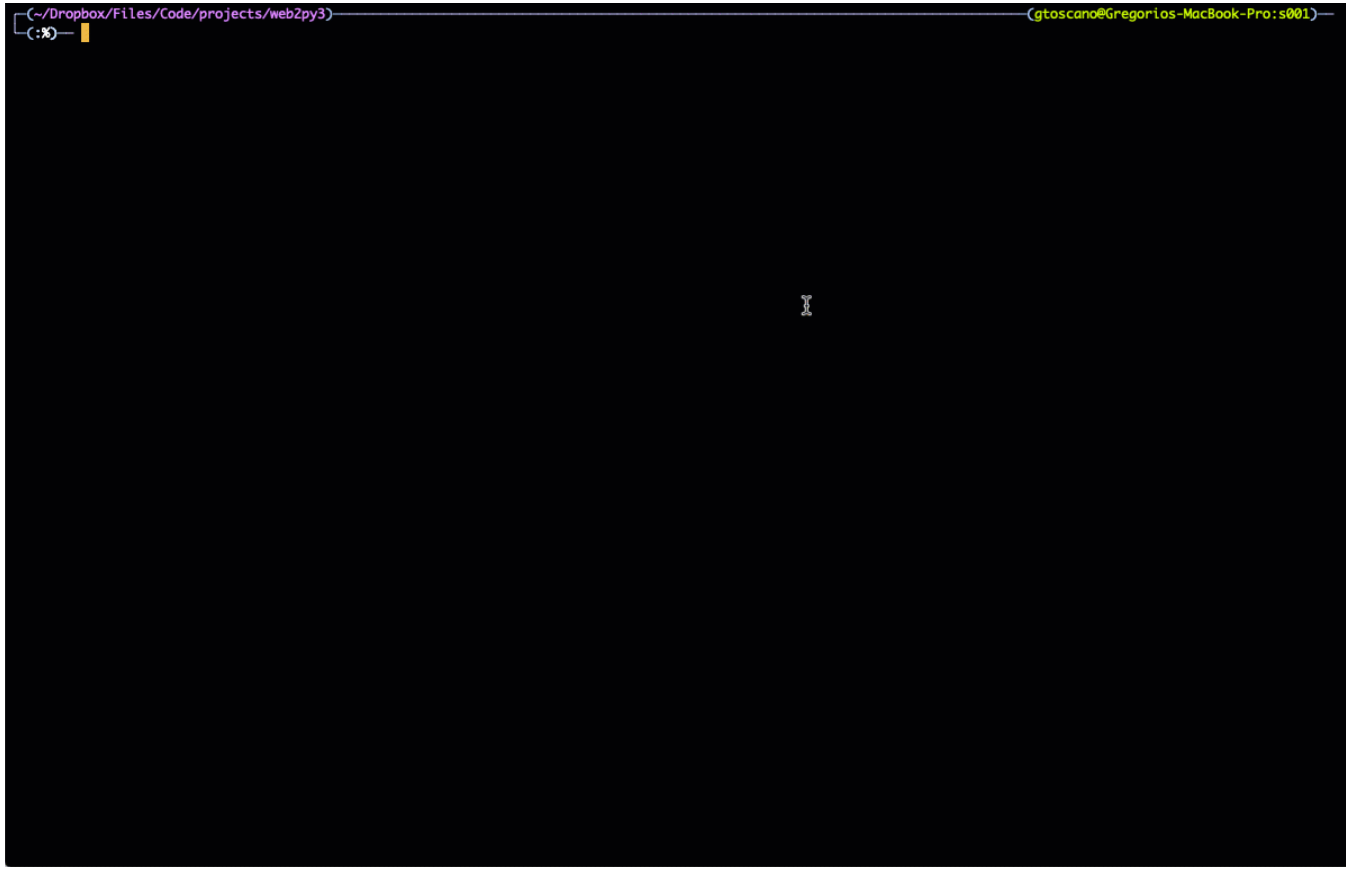
Problem Details:

- Land river segments: 22
- Load sources: 41
- BMPs: 190
- Parcels: 902
- Overlapping BMP/Parcel combination: 4,053
- Overall BMP/Parcel Combinations: 38,170



CAST System and Optimization: Bayota/VICO software

- Python-based (fast implementation)
- Requires several libraries
- **Pyomo** optimization framework
- Uses Ipopt as the optimization engine
- **Ipopt** is a well-known library for large scale nonlinear optimization that uses an **interior point** approach
- Ideal for **convex** problems (it locates the optimum)
- It may find a **local minimum/any non-opt. point** in **non-convex** problems



CAST System and Optimization: Bayota/VICO Results

- Objective function: **$155,256** (total cost of implementing all BMPs) Is it really the optimum?
- Number of objective function evaluations = 511
- Number of objective gradient evaluations = 496
- Number of equality constraint evaluations = 0
- Number of inequality constraint evaluations = 511
- Number of equality constraint Jacobian evaluations = 0
- Number of inequality constraint Jacobian evaluations = 496
- Number of Lagrangian Hessian evaluations = 495
- Total CPU secs in IPOPT (w/o function evaluations (FEs)) = 61.048
- Total CPU secs in NLP function evaluations = 6.014
- Time of optimization approach: 75.74 s.
- Total time: 1:49.35 min (109.35 sec).

Note: 6.014/511 = 11 ms /FFE
required by Ipopt

CAST System and Optimization: An Alternate Approach
Base Genetic Algorithm (BGA) C++ (Proof-of-Concept)

- **Base Genetic Algorithm** (without any customization) **implemented** from scratch in C++
- Binary representation (naïve representation: 1 for a BMP means it is used in the whole parcel), 38,170 Boolean variables
- One-point crossover operator
- Flip-bit mutation operator
- Binary Tournament selection
- 0.8 crossover probability
- 1/L mutation probability (L: #variables)
- 30 executions

BGA requires 0.12 ms /FFE

CAST System and Optimization: Base Genetic Algorithm C++ (Cont.)

- BGA in C++ is 95 times faster than the Python's approach per FFE
  - Probably due to Python and C++ number crunching methods
- We run two experiments:
  - 1,000 generations, 500 population size (**500,000 FFEs**)
  - 1,000 generations, 1,000 population size (**1,000,000 FFEs**)

500,000 FEs

	Best initial random solution	BGA best solution	Time (sec)
AVG	$19,073,271	$132,356	88.35
STD	$1,010,853	$11,188	19.27
Best	$16,666,185	$116,978	71.22
Worst	$20,821,176	$159,847	144.82

1,000,000 FEs

	Best initial random solution	BGA best solution	Time (sec)
AVG	$18,187,730	$66,627	158.64
STD	$981,967	$5,491	16.13
Best	$15,909,309	$54,714	133.55
Worst	$19,791,377	$79,468	197.21

Comparison of BGA with Ipopt

Algorithm	Cost in $	Time in sec
Ipopt	155,256	109.35
BGA (FFE 500k)	116,978	71.22
BGA (FFE 1000k)	54,714	133.55

- BGA (FFE 500k) finds 25% cheaper solution in 35% less time
- BGA (FFE 1000k) finds 65% cheaper solution in 22% more time
- Importantly, Ipopt did not produce the optimum solution!

Comparison of BGA with Ipopt in Four Counties

County	Ipopt in $	Time in sec	BGA in $	Time in sec
Bradford, PA	155,256	109.35	**54,714**	**133.55**
Tioga, PA	161,864	102.97	**23,923**	**150.64**
Lancaster, PA	**20,981**	**51.59**	64,972	171.45
Howard, MD	**1,010**	**45.84**	3,914	69.78

- Complexity of the problem dictates which algorithm will work well
  - Usually, BGA works better on more complex problems
- Mixed results suggest a hybrid approach involving both
  - Ipopt as a mutation operator applied to best population member

CAST System and Optimization: Summary of Current Studies

- The preliminary study suggested that both **BGA** and **Ipopt** can produce good quality results and are potentially good candidates

- **Ipopt** approach can be used in a hybrid manner (for local search) within a GA
  - Taking best aspects of both optimization algorithms

CAST System Integration Plan

- Getting ready to test and integrate our algorithms within the CAST system
  - COVID’19 causing some challenges
- Met with members of different CAST teams to devise a plan to gain data access and evaluate security considerations:
  - CBP's application development team
  - CBP network team
  - CBP data center
  - CAST Experts

CAST System Integration: Immediate Plan

- MSU will obtain a copy of the CAST database with a secured access
- Our immediate Tasks:
  - Verify the performance of our current optimization methods
  - Gain further knowledge about the problem
  - Develop customized components to improve our methods
  - Integrate our algorithms with the CAST database

CAST System and Optimization: Overall Project Plan (2020 – 2026)

- Objective 1 (First 18 mths.): Single-objective, customized, hybrid optimization methods and scalability study
- Objective 2 (Next 18 mths.): Multi-objective, customized, hybrid optimization methods and multi-criterion decision-making aids
- Objective 3 (Next 18 mths.): Scalability study to state and watershed level with Distributed computing and AI-based learning from optimization (Innovization)
- Objective 4 (Final 18 mths.): Handling practicalities using robust, surrogate-assisted, and sustainable optimization