
# Welcome to the Webinar

## *A Deep Dive into Advanced Analytics Model Review and Validation*

# Speaker Introduction

PRINCETON CONSULTANTS
*Information Technology and Management Consulting*

GUROBI OPTIMIZATION

- **Dr. Irv Lustig**

  - Optimization Principal, Princeton Consultants
    - Lead Optimization Consulting Sales and Optimization Project Implementations
  - Former Positions
    - CPLEX Optimization, Inc. Director of Numerical Optimization
    - ILOG, CPLEX Product Manager, eventually Vice President, ILOG Direct
    - IBM, Worldwide Optimization and Supply Chain Technical Sales Leader, then Manager of Optimization and Mathematical Software at IBM Research
  - Ph.D. in Operations Research, Stanford University
    - Ph.D. Advisor:  George Dantzig
  - INFORMS Certified Analytics Professional

# Speaker Introduction

- **Ugo Feunekes, M.Sc.F**
  - Co-Founder and CTO of Remsoft
    - Chief architect of Remsoft's suite of software
    - Leads Remsoft's research and development program and overall product development
  - Holds Bachelors and Masters of Science in Forestry, as well as a Bachelors of Mathematics and Computer Science
  - Experienced trainer and modeler with a depth of knowledge and experience in the realm of fire behavior, forest management, and asset lifecycle optimization.
  - Edelman Laureate
    - *Taking the Politics out of Paving: Achieving Transportation Asset Management through O.R.* (2010)

# About Princeton Consultants (www.princeton.com)

PRINCETON CONSULTANTS
*Information Technology and Management Consulting*

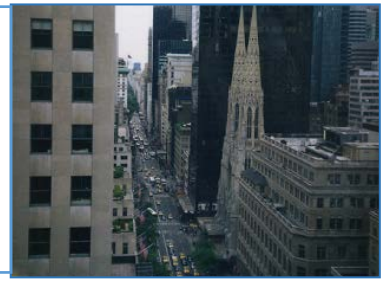
Princeton New Jersey

Information Technology & Management Consulting

*Advanced Analytics*
*Application Development*

*Strategy*
*Process Improvement*

Optimizing Performance



New York City

- Stability: over 35 years in business; over 1600 successful completed projects
- Clients: Industrial companies and their Logistics providers
- Analytics-focused Professional Staff: 85 full-time consultants, 2/3rds with graduate STEM degrees; plus network of top independent consultants and university professors. 50+ experienced data scientists/developers
- Experience: Senior Staff (top 28 consultants) average 15+ years experience; Firm Leadership (9 Directors) average 20+ years experience

# Princeton Clients: Diversified Industries

GUROBI OPTIMIZATION


PRINCETON CONSULTANTS
*Information Technology and Management Consulting*

GUROBI OPTIMIZATION

# Advanced Analytics Model Review & Validation

## A Quality Assurance Service for Analytics Leaders

PRINCETON CONSULTANTS
*Information Technology and Management Consulting*

# Benefits of the Service

- **Validate and improve optimization and predictive models with an expert third party review**
  - Uncover new ideas for improvement
  - Benchmark your models and group against best practices
  - Give business leaders more confidence in your solutions
  - Help your specialists improve their development skills

# Important Questions Addressed by Us

GUROBI OPTIMIZATION

- What is a correct model?
  - What makes a model "correct"?
- What data is being integrated and how?
- How are solutions published and used in the business?
- What validation was done to see that the model met the needs of the business?
- How sensitive are the answers to the inputs?
- How often is the quality of models assessed, considering the changing nature of the data and the business?
- Did the implemented model reflect the intentions of the practitioner?
- Is data captured to allow models to be evaluated and tested outside of an operational system?
- If algorithms use random number generation techniques, are the random number sequences reproducible?

# Engagement Summary

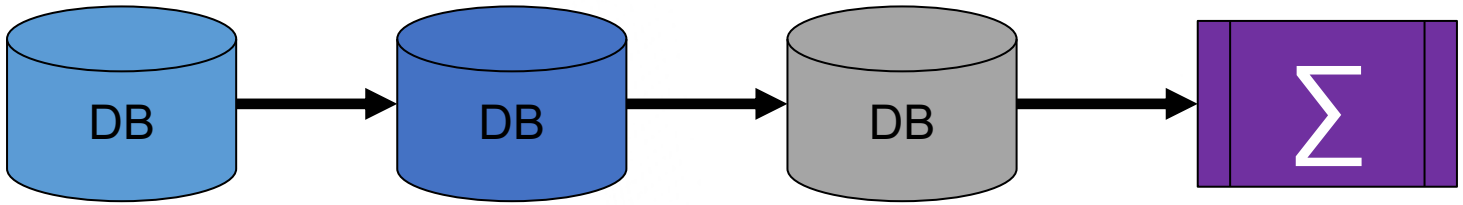
- Interview stakeholders from business and analytics development to understand business problem and context
- Review of existing models and procedures
- Review of data sources
- Implementation of models in alternative technologies to compare results
  - Languages, Solver, Analytics Engines
- Experiments with models to uncover issues
- Report of findings with suggestions for improvements and possible further investigation

# Sample Optimization Questions

- If the problem has multiple optimal solutions, how does the chosen algorithm affect the variability in the answers?
- Could data be presented that would make the model infeasible?
- Are best practices followed for generating values passed to the model?
  - Precision vs. Accuracy
- If the solver is integrated into an application, have all appropriate exceptions been trapped?
- Were best practices observed to facilitate future model updates?
- Will the model scale as the dimensions of input data change?
- Is the underlying solver being used in the best way possible?
- For mixed integer programming, is pure optimality really needed?
- Would an alternate model yield better performance?

# Understand how data feeds the system

- Two sources of data
  - Automatic feeds
    - Continually check the assumptions




  - User inputs
    - Create automatic alerts for data that falls outside control limits

| | |
|---|---|
| **Budget** | **$40,500,000** |
| | |
| **Budget** | **$50,400,300,200,000** |

# It's All About the Data!

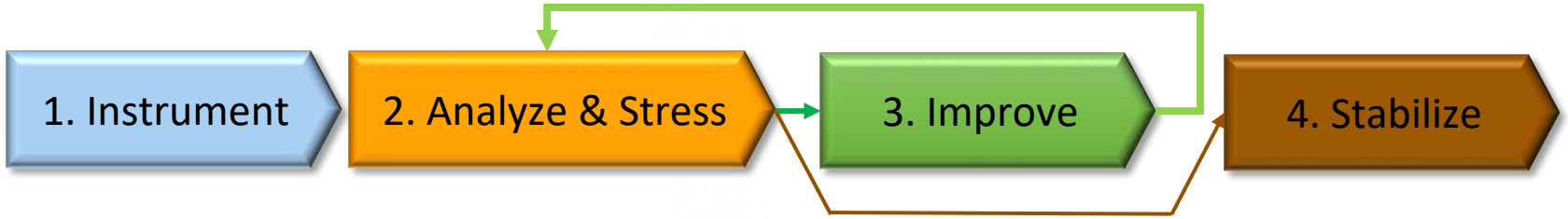
**Problem: Users do not distinguish between bad data and bugs**

- A bad recommendation means "it doesn't work"
- Result: can't verify; long validation cycles; frustration


**Solution: Data Cleaning is a core component of Optimization**

- Clarity of "why" includes the incoming data
- Allow users to easily change the data where possible

# Princeton Optimization Implementation Verification Methodology




- Create formal Optimization Algorithm Metrics (OAMs)

- Compute OAMs of current implementation
- Identify system failure and under-performance areas; Aggregate results; Compare to previous results
- Determine adversarial stress conditions

- Research and identify improvement strategies to algorithms
- Improve optimization solution quality, efficiency, robustness; Re-run

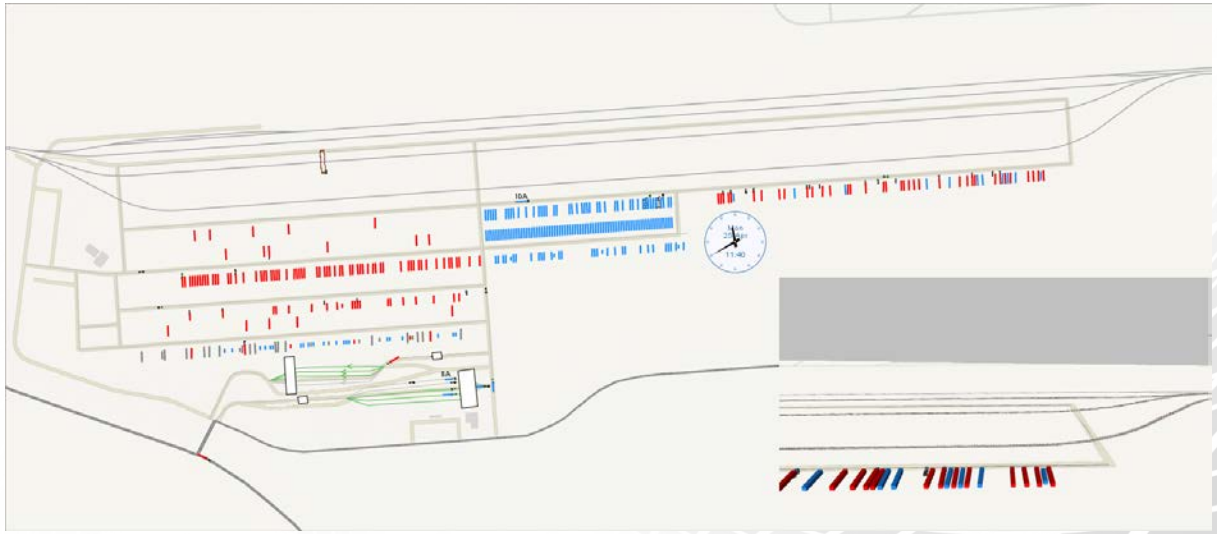
- Pass current implementation on for integration and testing
- Integrate implementation into application and processes for formal QA testing and production

PRINCETON CONSULTANTS
*Information Technology and Management Consulting*

# Simulation: A Tool to Help with Model Validation

- Build a simulator of data inputs
- Good simulation packages allow you to create graphical views of solutions that help end users validate the analytics
- Allows robust testing, especially with operational systems
  1. Validate data quality and completeness
  2. Iron out interface design kinks
  3. Fine tune the model
  4. Gain end user buy-in


Analytics Magazine, July/August, 2018



GUROBI OPTIMIZATION


Enabling clients to solve complex problems….. smarter

REMSOFT®

Recognized Worldwide as the INDUSTRY STANDARD in Forestry Planning and Scheduling.

More than half a billion acres of land managed globally. SUSTAINABILY.

150+ Clients in 15 Countries

Corporations
Consultants
Governments
Financial Institutions
Universities

Expert Team With Ongoing Investment in R&D

- Continued development of cloud platform and SaaS products
- Core IP & algorithms for modelling & optimization
- Systematic, Iterative Process to engage customers & proven implementation methodology

REMSOFT®

GUROBI OPTIMIZATION

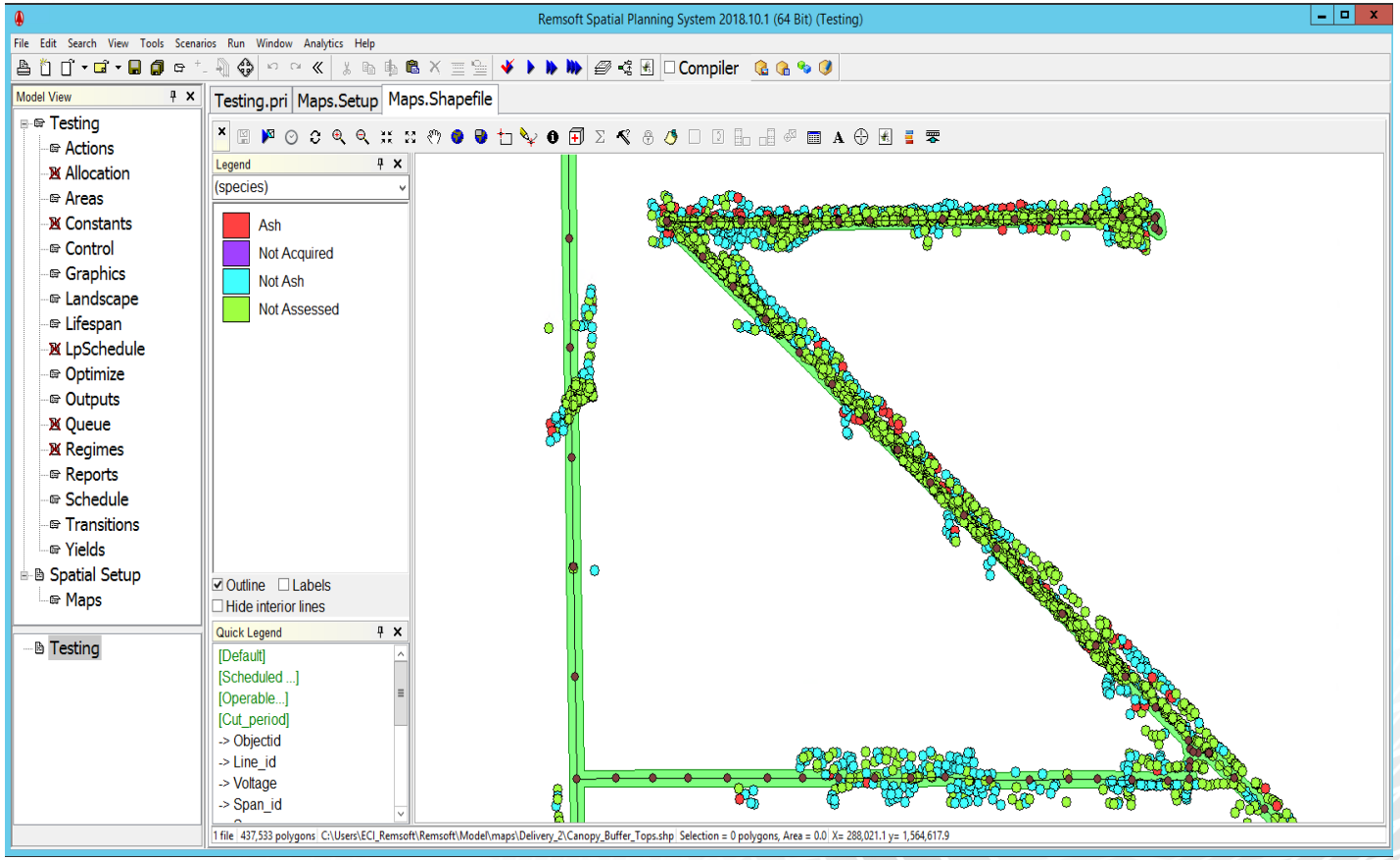
# Remsoft Technology

GUROBI OPTIMIZATION

## Two advanced platforms for optimized planning and decision support

- Woodstock Modeling Platform for the desktop
- Remsoft Rise Platform for the cloud

Together they provide the foundation for all of our planning solutions.




REMSOFT®

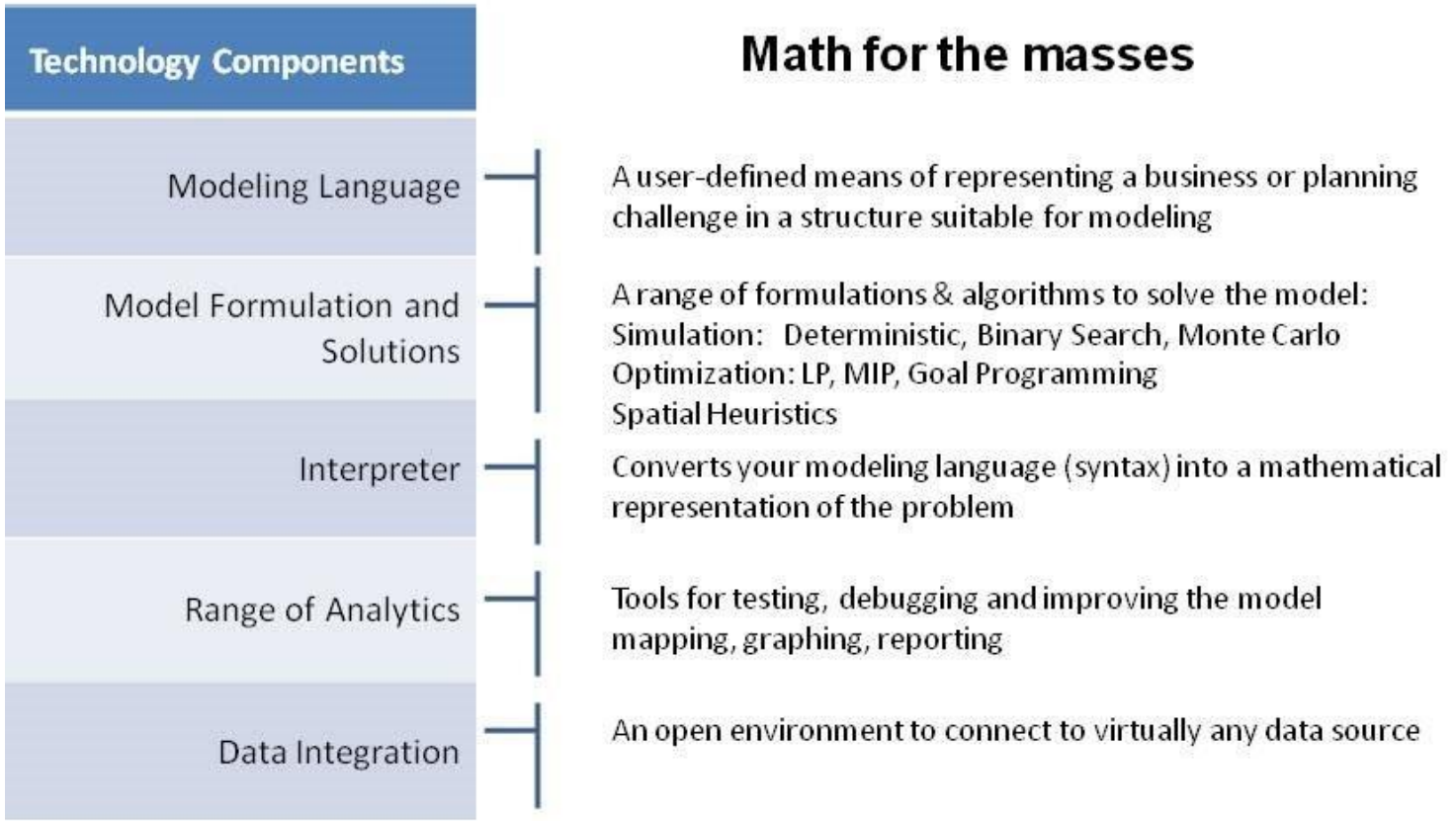
Software to schedule, manage and visualize high-value decisions

# Woodstock Modeling Platform

GUROBI OPTIMIZATION

## Technology Components

- Modelling engine: syntax, editor, algorithms
- Allocation Optimizer: products to destinations
- Spatial Optimizer: spatial scheduling
- Regimes: Multi-period scheduling
- Solver (OEM)




REMSOFT®

# Remsoft's Engagement with Princeton

## Background

- Highly complex models, primarily MIP
  - Optimize a model with 944,490 rows, 4,765,013 columns and 24,439,033 nonzeros
  - Variable types: 4,578,017 continuous, 186,996 integer (0 binary)
- Wide ranging planning challenges
  - Crew scheduling
  - Transportation / Harvesting
  - Long term (100 year) sustainable forest/habitat management plan
- Frequent data refresh – even while building and testing models

# Remsoft's Engagement with Princeton

## The Issue

- Unreasonably long solve times
  - Days or longer when short term decisions are required
  - Frequent infeasibilities
- Unpredictable solve times
  - Different Gurobi tuning parameters required for different models – what works for one model does not work for another
  - Adding a single additional constraint makes solving impossible…. sometimes
- Client reluctant to simplify model or reduce constraints

# Our question for Princeton Consultants:

## What's happening? What is causing these issues?

- Are there things we could tweak in our modeling platform to improve solver performance?
- Are there better ways to structure some of these models?
- Can we improve Gurobi performance by changing parameters?
- Are there best practices you can share?

# Precision vs. Accuracy

GUROBI OPTIMIZATION

Standard Recommendation

- Use double precision in representing values for an application

Issue

- The double precision value may be too precise with respect to the business problem!

Example from Remsoft Application:

- Estimate amount of wood boards to 2 digits of accuracy after decimal point
- Multiply and divide by other data, such as rates and percentages to get hectares of forest
  - Could get a value such as 45.123456
- Constraint on hectares could be asking for accuracy down to the square micrometer!

Carefully consider the accuracy required in satisfying constraints, and possibly use *less* precision

- E.g., round numbers to 2 digits after decimal

Important to Consider when computing data and representing expressions related to money

- Do the constraints/objectives need accuracy to the pennies when values are in the millions of dollars?

# Multiple Objectives

- Remsoft
  - Has the ability to specify “goals”
  - Customers typically used blended objectives
  - Blended objectives could be mixing different unrelated quantities
    - E.g. Costs in Dollars and Area in Hectares
    - Blended objectives can create numerical difficulties for solvers
  - Princeton recommended a hierarchical approach

- Gurobi supports multiple objectives
  - Two approaches
    - Blended
    - Hierarchical

- In general, Princeton finds that hierarchical approaches can more easily be mapped to business requirements

# Constraint Violations and Normalizing Units

- Consider the following two constraints:
  - X ≥ 1500
    Y ≥ 40
- Remsoft system allows user to say these are soft constraints and solve by creating an objective that measures the total constraint violation
- If weighted equally, values of X and Y of 1490 and 30, respectively, are considered equal violations of the two constraints
- Might be better to normalize the constraints, and use a percent violation:
  - X/1500 ≥ 1
  - Y/40 ≥ 1
- Or appropriately weight the constraints in the penalty function

# Modeling Recommendations: Add Extra Constraints

- For one MIP model, variables that were implicitly binary due to a combination of constraints were not declared as binary
- Gurobi Presolve did not figure out that those variables were supposed to be binary with the constraints as the model was formulated
- By specifying those variables to be binary, and creating additional constraints to link those binaries with other variables, Gurobi could branch on those variables and create stronger cuts
- Gurobi Presolve created a smaller problem, and solution times were significantly decreased (from hours to minutes)

# Diagnose Performance Issues Via Relaxation

- One Remsoft model had the following characteristics:
  - Binary decision variable $V(m,p)$ represents whether a machine *m* is used for work in period *p*.
  - Requirement that if machine *m* is used for work in period *p*, then at least *n* machines are used in that period, effectively:

$$Machines(p) = 0 \text{ OR } n \leq Machines(p) \leq n+1$$

  - Equivalent representation is

$$n\,V(m,p) \leq Machines(p)$$
$$Machines(p) \leq (n+1)V(m,p)$$

  - The variables $V(m,p)$ appear elsewhere in the model

- Model took many hours (overnight) to get any reasonable solution

- If $V(m,p)$ was relaxed to no longer be integer, Gurobi solved model to 1% gap in 39 minutes

- One can conclude that these variables are what make the problem difficult
- Leads to an alternate formulation of the model that introduces new variables and constraints that interact with the $V(m,p)$ variables that should generate stronger cuts for Gurobi

## Following the engagement

Remsoft

- Made programming changes to address some of the rounding issues identified. Users now have the option to “turn on” this feature
- Increased model debugging tools to manipulate/analyse matrices
- Employed best practices for modeling both within our company and in the work we do with our clients
- Added Princeton recommendations to our model audit service

# Conclusion

- Deployment of Optimization Models Requires
  - Optimization Modeling and Algorithmic expertise
  - Testing expertise
  - Software engineering expertise
- An independent review can assure that your models are delivering the value you expect
  - Deployment in real-time systems can be a challenge to get right
- What percentage of your models have errors??