How Integrated Models Have Influenced Decision-making:

Lessons from the UrbanSim Experience

Paul Waddell
Center for Urban Simulation and Policy Analysis
University of Washington

Time for Assessment

- **Have** integrated models influenced decision-making?
- As much as expected?
- What worked well?
- What have been the biggest problems?
- How can these be addressed?

- We examine these questions in the context of UrbanSim

Support for UrbanSim Development

- NSF Urban Research Initiative
- NSF Biocomplexity *
- NSF Information Technology Research *
- NSF Digital Government *
- Environmental Protection Agency *
- Federal Highway Administration
- Oregon Department of Transportation
- Oahu Metropolitan Planning Organization
- Puget Sound Regional Council *

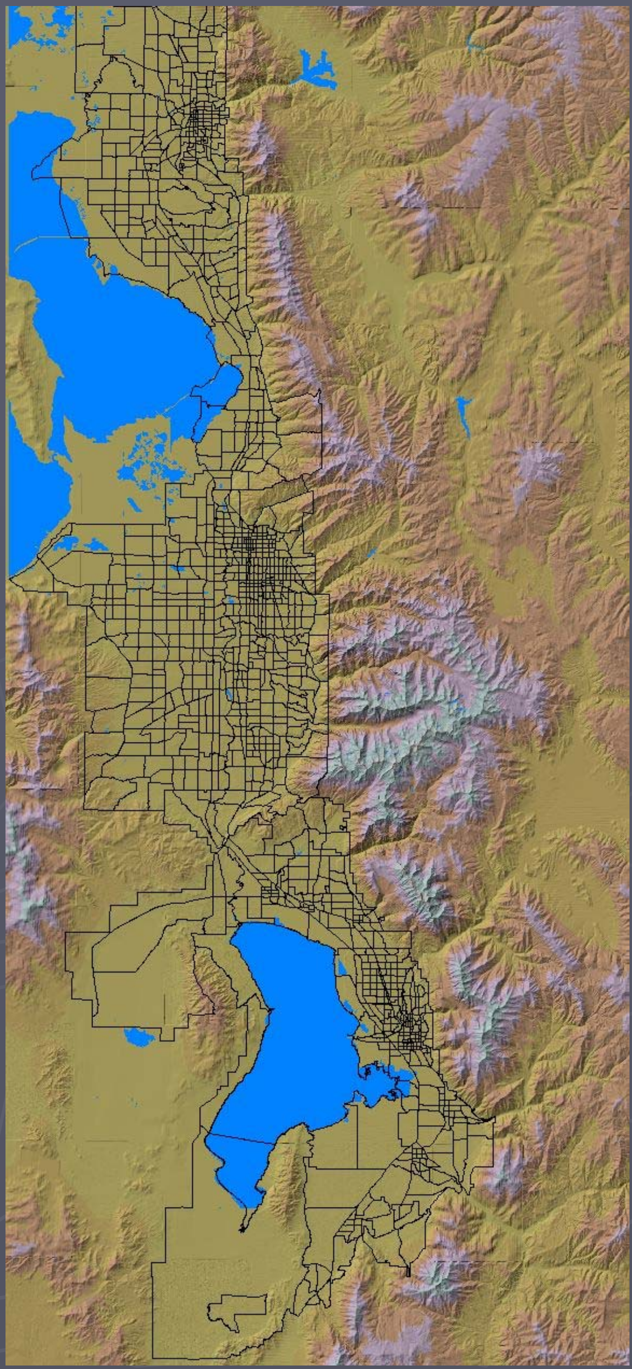
* Currently active grants

UrbanSim Development History

- Research and Development Phase (1996-2000):
  - Initial Design: Honolulu (OMPO), 1996
  - Initial Prototype: Eugene-Springfield (ODOT/LCOG), 1998
  - Initially released on the web as Open Source Software, 1998
- Development and Application Phase (post 2000):
  - Recent US Applications:
    - Houston (HGAC) - baseline land use adopted for RTP
    - Salt Lake City (WFRC) – used in settling highway lawsuit
    - Seattle (PSRC) – being used to support EIS for Vision 2020 update
    - Washtenaw County (SEMCOG) – pilot planning project
    - Phoenix (ASU) – research project
  - Recent European Applications
    - Amsterdam (RIVM) – EU study of impacts of economic deconcentration
    - Paris (IAURIF) – study of impact of rail projects
    - Zurich (ETH) – research project
  - Many other applications in various stages, not well documented

A Case Study:
Wasatch Front Region, Utah

(with John Lobb, formerly at WFRC
and Joel Franklin, UW)



A Case Study: Wasatch Front Region

- Existing Transportation System
  - Dominated by the automobile (~90% of all trips by auto)
  - 2 highly successful light rail lines
- Existing Land-usage
  - Low density
  - Subdivisions, retail centers and office parks
- Population:
  - 1.6 million in 2000
  - ~3.0 million by 2030
- Envision Utah
  - Highly successful visioning process
  - Intensive public outreach/involvement
  - However, the process mixed outcomes and regional goals

Lawsuits

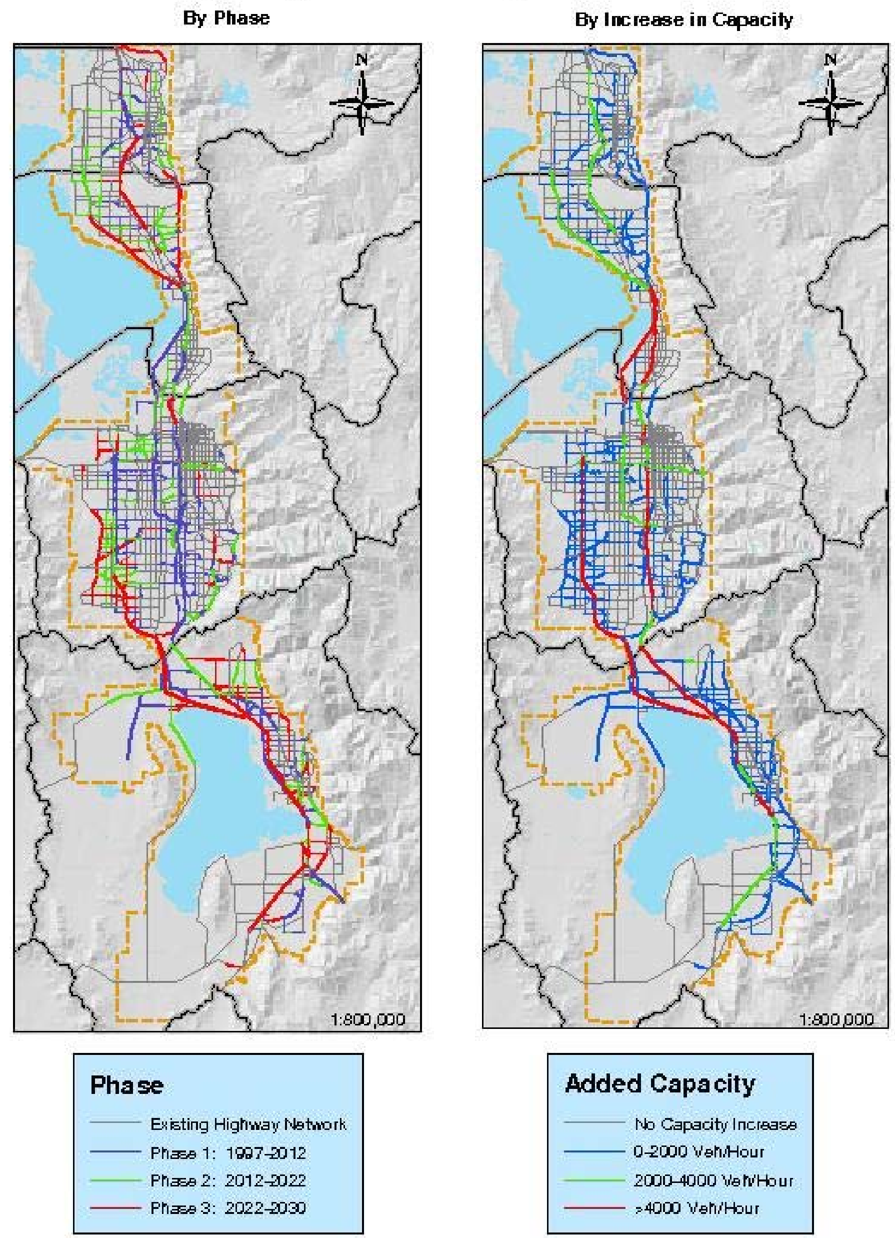
- Legacy Highway
  - North of Salt Lake City
  - Wetlands (adjacent to The Salt Lake)
  - Construction halted by court (Clean Water Act violations)
- Long range plan analysis
  - Technical analysis challenged
  - Lawsuit settled: Test UrbanSim for suitability for use, with peer review by 12/31/03

Legacy Highway

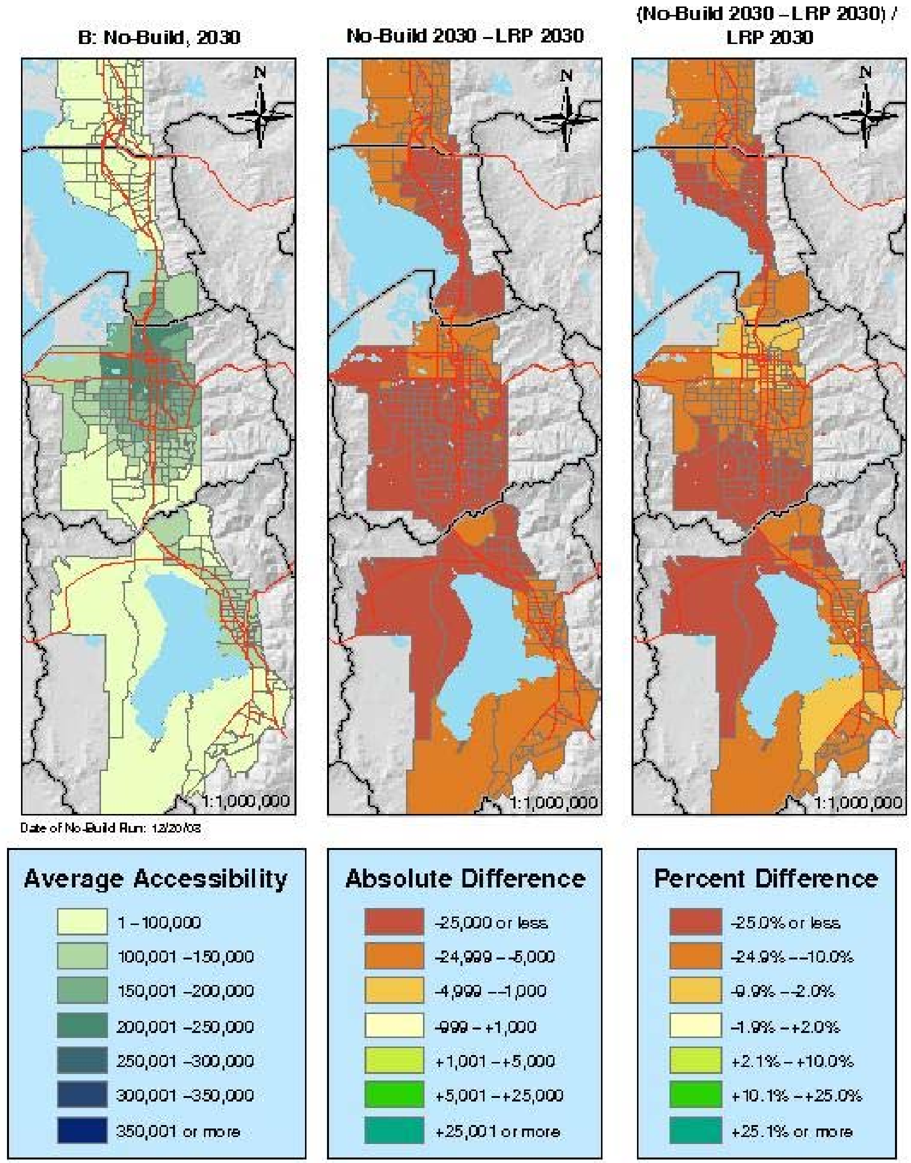
Sensitivity Testing of Integrated Land Use and Transportation Models

- Tested several scenarios:
  - Long Range Plan (Baseline)
  - No-build
  - Drop a highway project
  - Drop a light rail project
  - Add parking pricing
  - Impose Urban Growth Boundary
- Automated interaction of land use and travel models

Long Range Plan Roadway Improvements

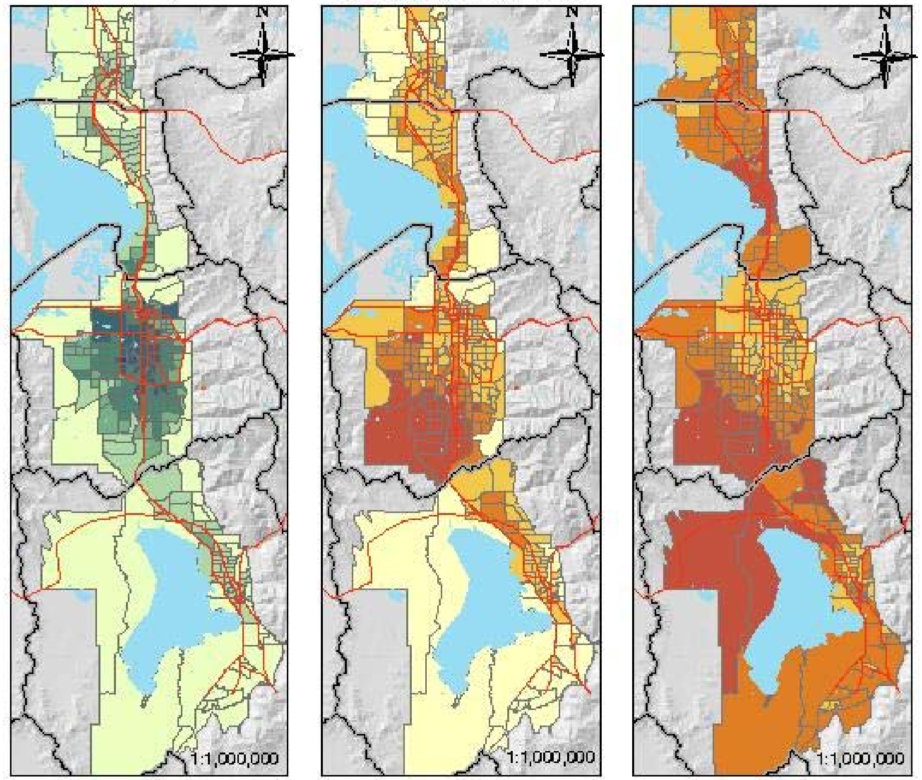


Scenario B: No-Build Comparison to Scenario A: LRP
Access to Employment (1-Car Households)



Scenario B: No-Build Comparison to Scenario A: LRP Land Price

B: No Build, 2030 | No Build 2030 – LRP 2030 | (No Build 2030 – LRP 2030) / LRP 2030



Date of No Build Run: 12/20/03

Dollars per Acre

- $0.00 – $50,000.00
- $50,000.01 – $100,000.00
- $100,000.01 – $150,000.00
- $150,000.01 – $200,000.00
- $200,000.01 – $300,000.00
- $300,000.01 – $500,000.00
- $500,000.01 or more

Absolute Difference

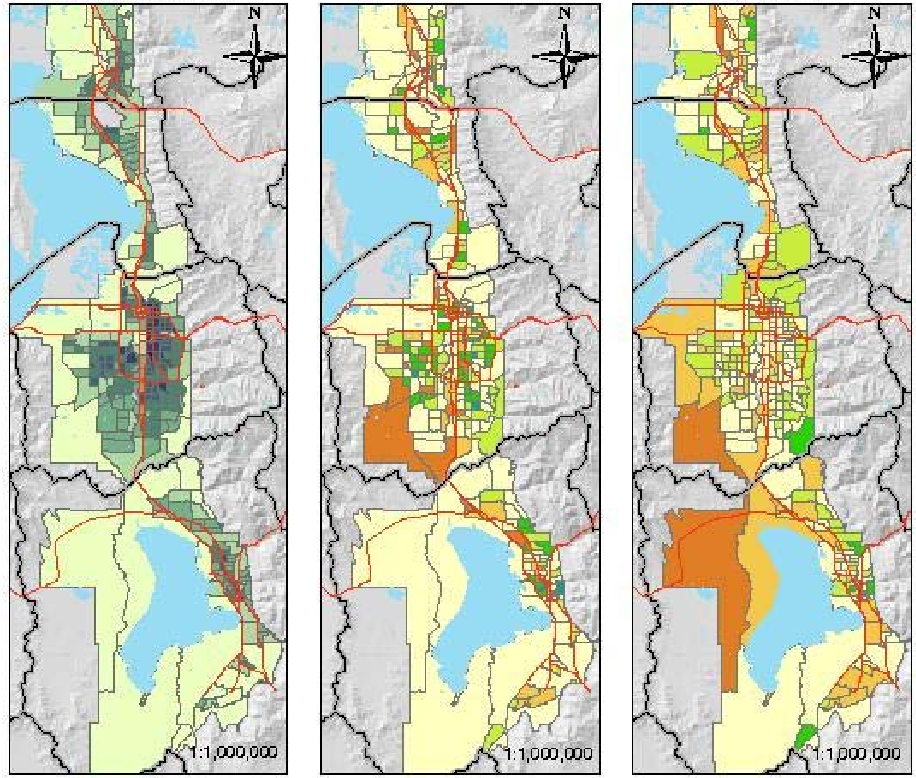
- -$40,000.00 or less
- -$39,999.99 – -$20,000.00
- -$19,999.99 – -$5,000.00
- -$4,999.99 – +$5,000.00
- +$5,000.01 – +$20,000.00
- +$20,000.01 – +$40,000.00
- +$40,000.01 or more

Percent Difference

- -25.0% or less
- -24.9% – -10.0%
- -9.9% – -2.0%
- -1.9% – +2.0%
- +2.1% – +10.0%
- +10.1% – +25.0%
- +25.1% or more

Scenario B: No-Build Comparison to Scenario A: LRP Residential Units

B: No Build, 2030 | No Build 2030 – LRP 2030 | (No Build 2030 – LRP 2030) / LRP 2030



Date of No Build Run: 12/20/03

Units per Acre
- 0.0 – 1.0
- 1.1 – 2.0
- 2.1 – 3.0
- 3.1 – 4.0
- 4.1 – 5.0
- 5.1 – 10.0
- 10.1 or more

Absolute Difference
- -0.25 or less
- -0.24 – -0.10
- -0.09 – -0.05
- -0.04 – +0.05
- +0.06 – +0.10
- +0.11 – +0.25
- +0.26 or more

Percent Difference
- -50% or less
- -49.9% – -10%
- -9.9% – -2.5%
- -2.4% – +2.5%
- +2.6% – +10%
- +10.1% – +50%
- +50.1% or more

Some Lessons from this Case

- Tensions over impacts of highway projects
- Tensions between regional transportation planning and local land use control
- Credibility of models: high burden of proof
- Scale of project sensitivity
- Model and data diagnosis
- Small data errors, large time investment
- Need to make many things easier:
  - Data development, cleaning and synthesis
  - Model estimation
  - Model diagnosis and validation
  - Incremental improvements

Lessons from Toronto Workshop on Integrated Modeling

- Hosted by Eric Miller, Jan 2005
- Many integrated modeling projects in North America and Europe
- Most share substantial design aspects
- All require extensive software infrastructure
- None have fully reached their objectives

- Began Initiative to Collaborate in Building an Open Platform for Urban Simulation (OPUS)

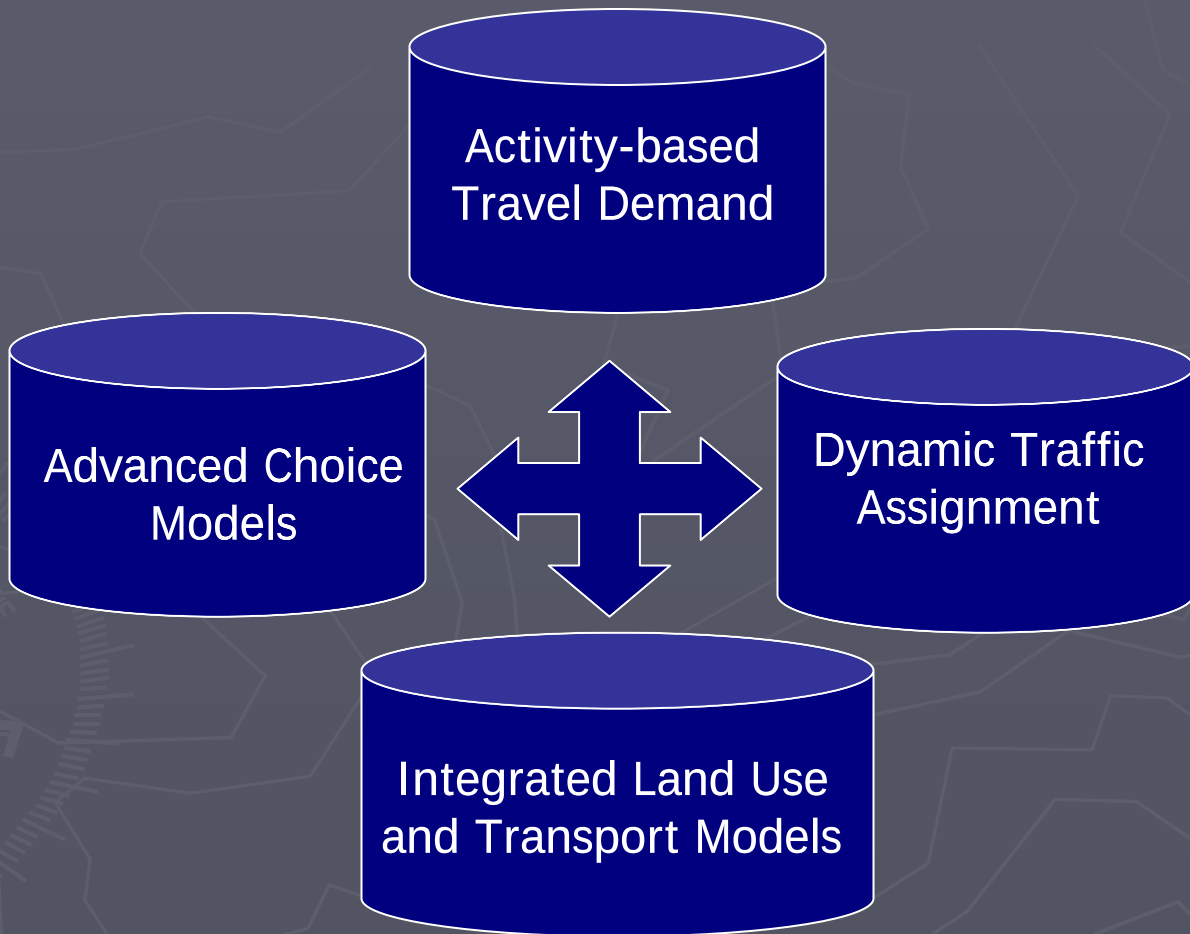
Open

Platform for

Urban

Simulation

Axes of Innovation in Land Use – Transportation Modeling



This initiative marks the beginning of an international collaboration across these research areas, to build an Open Platform for Urban Simulation (OPUS)

Motivation

- Many research projects want to do some or all of these:
  - Modeling urban spatial development
  - Modeling activity patterns
  - Modeling dynamic traffic conditions
  - Modeling environmental conditions (air, water, habitat)
  - Microsimulation using individual agents
  - Sophisticated models of agent choices
  - Robust model estimation and validation
  - Analysis of uncertainty
  - Capacity to evaluate land use and transport policies
  - Integrated spatial analysis and visualization
- The effort to do all of this well, independently, is massive
- Linking models is currently very problematic
- Writing/maintaining software takes too much effort
- Making incremental improvement is hard

Experience from UrbanSim Project

- Open Source with regular releases since 1998
- Applications in numerous metropolitan areas
- Intensive testing and building procedures
- Advanced software engineering and performance tuning
- Extensive documentation
- Elaborate tools to facilitate database development
- Sophisticated run management system
- But...

Experience from UrbanSim Project

- Open Source with regular releases since 1998
- Applications in numerous metropolitan areas
- Intensive testing and building procedures
- Advanced software engineering and performance tuning
- Extensive documentation
- Elaborate tools to facilitate database development
- Sophisticated run management system
- But…it is still too hard to do important, basic things:
  - Create and clean input data
  - Add variables
  - Modify specifications
  - Estimate models
  - Diagnose models
  - Interface with travel models

Design Requirements for Opus

- Must be Open Source
- Must be Very Highly Modular, Below Model Level
- Must have Extensive Data Analysis, Synthesis, Visualization
- Must be Extensible by User-Contributed Packages
- Must Provide Scripting Capacity
- Must Provide High Performance for Production Use
- Must Integrate Model Estimation and Application
- Must Integrate Spatial Analysis and Visualization
- Must Represent Multiple Levels of Geography
- Must Allow Integration of Heterogeneous Components
  - Computing languages
  - Model scopes (e.g. land use, traffic assignment)
  - Modeling approaches (discrete choice, ABM, rules)

To our knowledge, no single platform currently does all this.

Why Open Source?

- Facilitates Academic and Public Agency Accountability
  - Transparency
  - Verifiability
  - Reproducibility
- Precondition for Open Collaboration
  - Symmetric incentives to share information
- Productivity: Leverage Existing Components
- Protection:
  - Users Access to Software
  - Developers Access to Software
  - From Liability
- Robustness
  - Increases Number of People Checking Code

Why Modular? Innovation

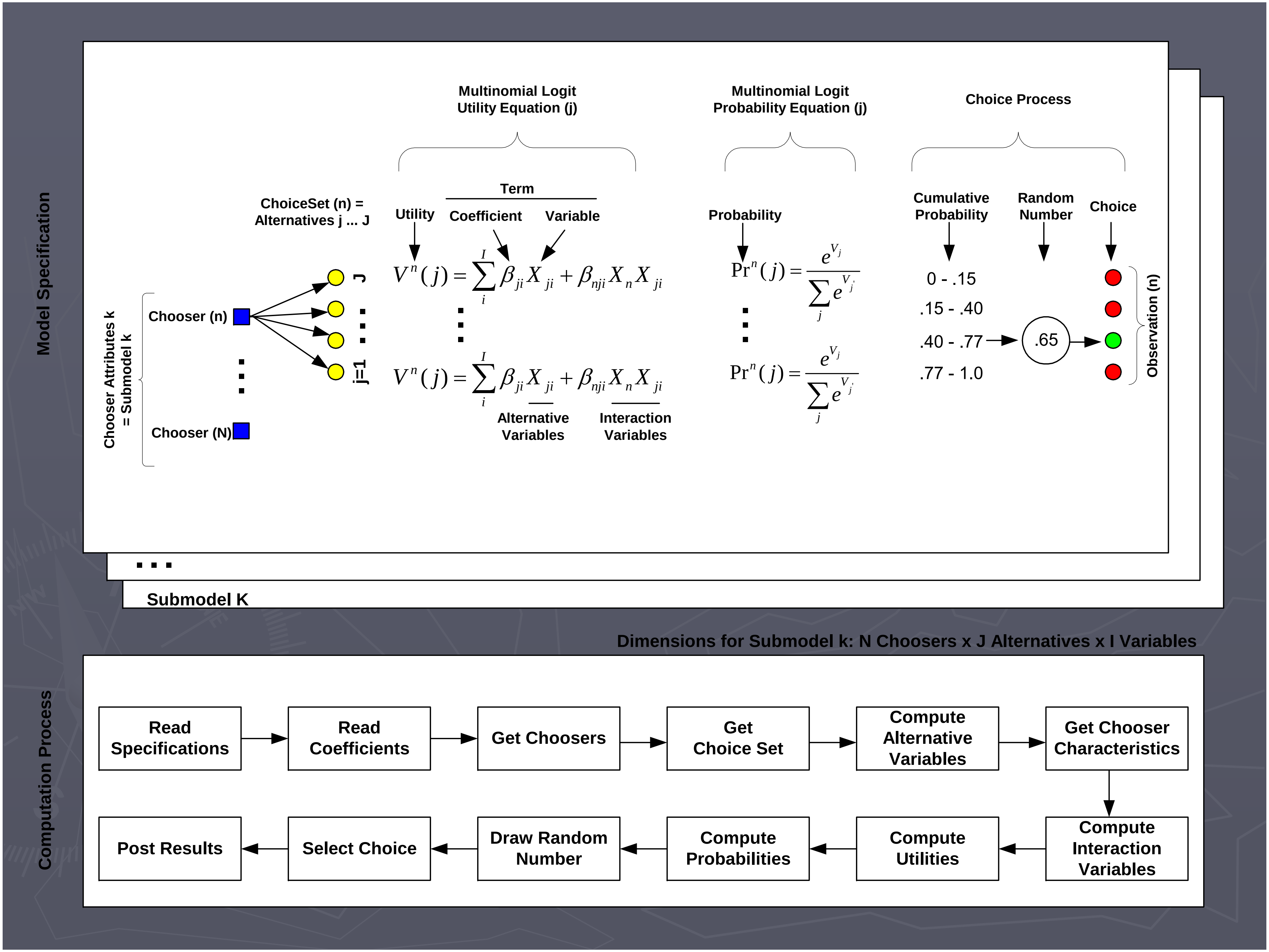
- Agents: individual, sample, aggregate segments?
- Space: parcel, cell, traffic zone, larger districts?
- Time:
  - Time-abstract equilibrium or explicit temporal dynamic?
  - Size of time steps?
  - Event-driven or chronological time?
- Behavior:
  - Independent choices, or interdependent?
  - How much detail?
    - Demographic: e.g. kids leaving home, marriage markets?
    - Intra-household negotiation?
    - Social networks?

Why Modular? Adapts to Applications

- Focused research effort
  - High detail in some aspects of interest
  - Less detail needed in others
- Planning application
  - Corridor or Community Study
    - High detail within focus area, less needed outside
  - Regional Visioning and Scenario Analysis
    - Coarse, regionwide analysis needed
  - Regional Transportation Plan Final Analysis
    - Detailed, region-wide analysis needed

Modular Choice Models

- Selection of Choosers
- Sampling of Choice Sets
  - Random, stratified, weighted
- Variables
  - Modular variable computation with dependencies implemented
- Utilities
  - Linear implemented, non-linear planned
- Probabilities
  - Multinomial logit implemented, others planned
- Estimation
  - MNL implemented, others planned
- Choice selection
  - Random, Lottery, Constrained



Model Specification
Multinomial Logit Utility Equation (j)
Multinomial Logit Probability Equation (j)
Choice Process
ChoiceSet (n) = Alternatives j ... J
Term
Utility
Coefficient
Variable
Probability
Cumulative Probability
Random Number
Choice
Chooser Attributes k = Submodel k
Chooser (n)
Chooser (N)
J
j=1
$V^n(j) = \sum_i^I \beta_{ji} X_{ji} + \beta_{nji} X_n X_{ji}$
$\Pr^n(j) = \frac{e^{V_j}}{\sum_j e^{V_j}}$
Alternative Variables
Interaction Variables
0 - .15
.15 - .40
.40 - .77
.77 - 1.0
.65
Observation (n)
Submodel K
Dimensions for Submodel k: N Choosers x J Alternatives x I Variables
Computation Process
Read Specifications
Read Coefficients
Get Choosers
Get Choice Set
Compute Alternative Variables
Get Chooser Characteristics
Compute Interaction Variables
Compute Utilities
Compute Probabilities
Draw Random Number
Select Choice
Post Results

Extensible by User Packages

- The R-Project is an inspiration for Opus
- R is an Open Source Statistics package
  - Minimal core
  - User-contributed packages
  - Now standard to publish paper and R package
  - Writing packages is standardized
  - Installing user packages is trivial
  - Interactive and batch use modes
- Could do integrated modeling community do this?

Scripting Support

- Many processes need to be scripted
  - Long runs of land use and travel model chains over 30 simulation years
  - Processing a set of queued scenarios
  - Uncertainty analysis with many runs
  - Generating indicators from many completed scenarios
- In addition, scripting is excellent for developing prototypes
- But… computing speed is usually a problem with scripting languages

High Computing Performance

- Need to run models for large metropolitan areas – e.g. Paris, 11 million population
- Need to be able to run many scenarios
- Scripting languages generally interpreted (slow)
  - **Except** – when they are hybridized with low-level languages
  - Python + numarray
    - Low-level LAPLACK library in C
    - Scripting in Python
    - Computing performance close to native C, C++
  - Example: evaluates MNL logit on 1,000,000 agents with 25 choices each in about 6 seconds on this laptop.

Integrated Estimation & Application

- Too many problems arise from loosely coupled model estimation and application:
  - Redundant specifications lead to errors
  - Data setup is tedious and inefficient
  - Getting estimation results into usable form hard
  - Experimentation and iteration is very costly and error prone
- The solution:
  - One repository for the model specifications
  - Integrate model estimation into modular system
    - Shares application code, adds estimation step

Integrated Estimation

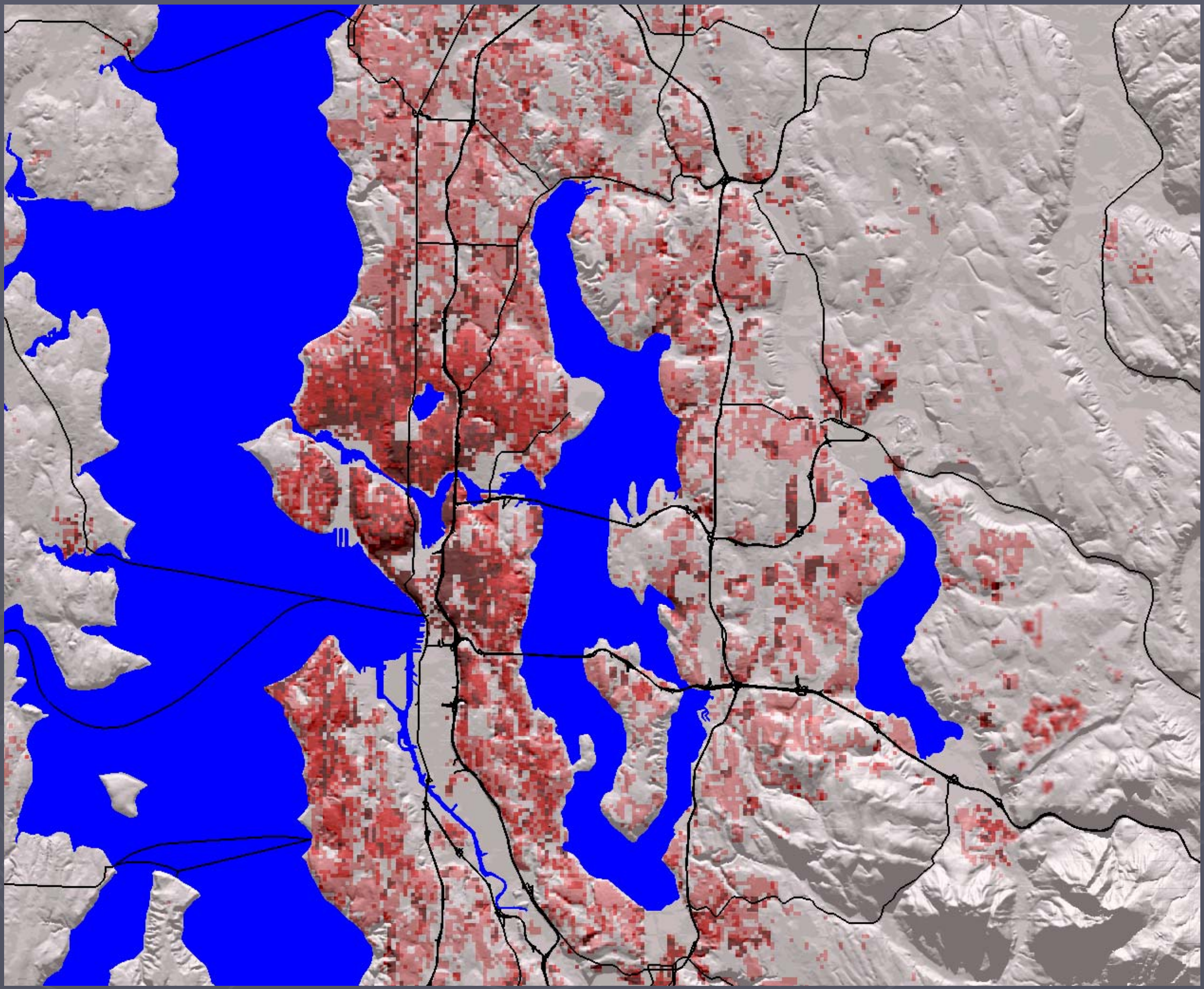
- Example: Single-housing unit development

```
********************************************************
Convergence achieved.
Estimates are:
[-0.2645337   0.17368394 -0.84682736 -0.14049767]
Standard errors are:
[ 0.08406553  0.01636794  0.01304732  0.01127312]
Convergence statistic is: 0.000421028108308
Log-likelihood is:        -30245.3885848
Log-likelihood ratio is:  0.211086788925
Number of observations:   16650
********************************************************
Elapsed time:  0.833183267706 seconds
```

Integrated Visualization

- Loosely coupled GIS is too inefficient
  - Can export data for making pretty maps
  - But too time consuming for exploratory work
  - And what about dynamic maps – animations?
- Need Integrated Spatial Analysis
  - Solution: Python Numeric packages for image processing – fast spatial queries
- Need Tightly Coupled Visualization
  - Must be able to display data in memory on map
  - E.g. Python with OpenEV

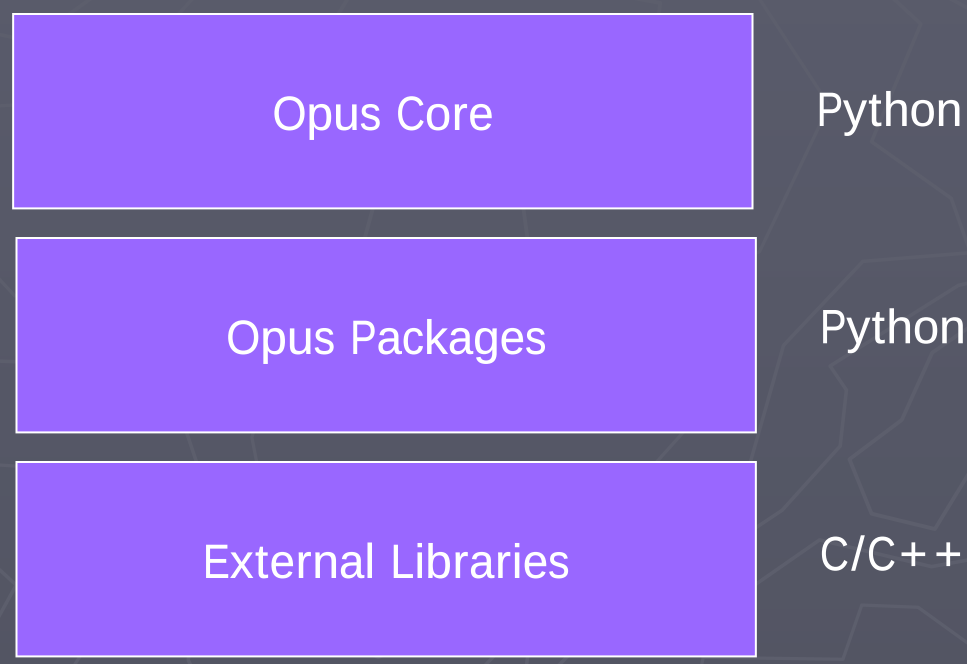
Integrated Visualization



Housing unit density displayed from data in memory. Used OpenEV with Python.


Opus will Support Multiple Geographic Units of Analysis
Currently: Grid cells, Zones
Soon: Parcels, Neighborhoods

Integration of Heterogeneous Components: A Tiered Opus Architecture



Opus Core (Python)

Model Definition

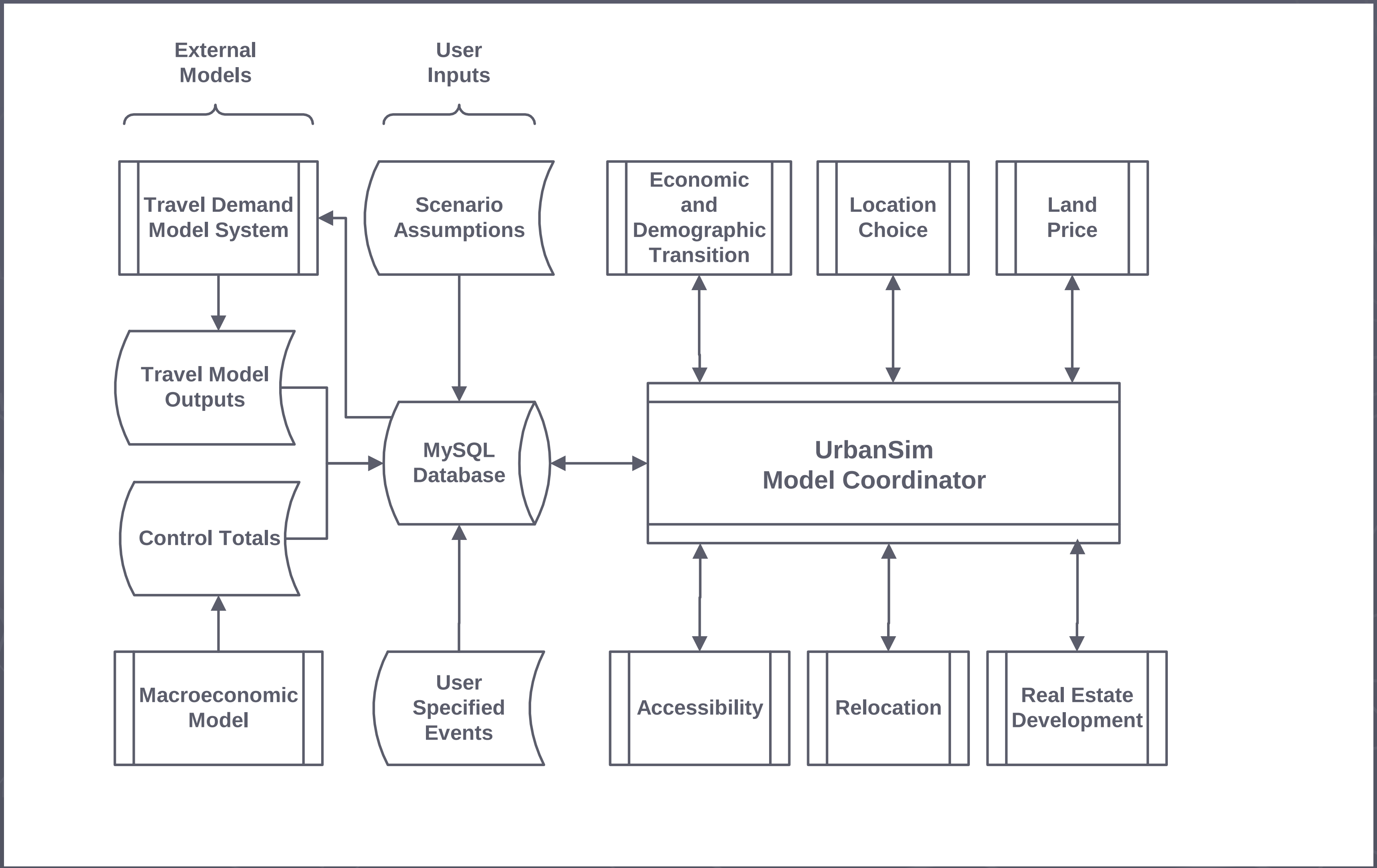
- Model Specification
  - Coefficients
  - Variables

- Discrete Choice Model
  - Sample Choosers
  - Sample Choices
  - Compute Utilities
  - Compute Probabilities
  - Select Choice

Data Management

- Storage
  - MySQL
  - XML
  - Binary
  - ASCII

Opus Packages (Python)
e.g. UrbanSim



Opus External Libraries (C/C++ with Python Wrappers)

Statistical and Numeric
- **Laplack**
- **BLAS**
- **Numarray**
- **R**
- Biogeme
- Amlet

Data Management and GIS
- **MySQL**
- **OpenEV**
- **PROJ4**
- **GDAL**
- Postgres
- SAGA

Travel Models
- PCATS
- DEBNetS
- ATESAME
- PACSIM
- MATSIM

Packages in Bold have already been interfaced to Opus

Status of Opus

- Core Architecture Implemented
- Generalized Storage Classes
- Discrete Choice Model Classes Implemented
- UrbanSim Models Converted to Opus
  - Faster performance
  - Much simpler and more readable code
- Multinomial Logit Estimator Implemented
- Connections to Biogeme Package in Progress
- Connections to Travel Model Systems Planned
- Release Planned for December 2005

Emerging Collaboration on Opus

- Eric Miller (U Toronto)
  - Modular travel model
- Ram Pendyala (USF), Ryuichi Kitamura (Japan)
  - Household Synthesizer
  - Activity-based Travel Model
  - Dynamic Mesoscopic Assignment Model
- Maren Outwater (CSI)
  - Activity-based Models, Assignment
- Chandra Bhat (UT)
  - Econometrics of Interdependent Choices
- Michel Bierlaire –EPFL, Lausanne
  - Biogeme – Discrete Choice Model Estimation
- Yan Song – UNC Chapel Hill
  - Development Templates

A Short Demo of Opus...?

- Seeing is believing
- Let’s create and run a zone-level household location choice model for King County, Washington
  - Instantiate household dataset
  - Instantiate locations
  - Specify the model
  - Compute the necessary variables
  - Sample alternative locations
  - Estimate the model coefficients
  - Simulate from the fitted model
  - Generate map of results
- If this doesn’t work...

Open
Platform for
Urban
Simulation

$$P(i) = \frac{e^{\alpha+\beta X_i+\delta Y_i+\theta Z_i}}{\sum_j e^{\alpha+\beta X_j+\delta Y_j+\theta Z_j}}$$

For more information or to collaborate
we're setting up a web portal soon at:
www.opus-network.org