Towards an Open Platform for Urban Simulation

STRC
2005

Paul Waddell
University of Washington

Outline

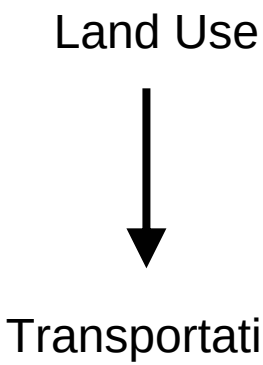
- Integrated Land Use - Transport Modeling
- UrbanSim
- Opus – Who, What, Why

Integrated Land Use - Transport Modeling

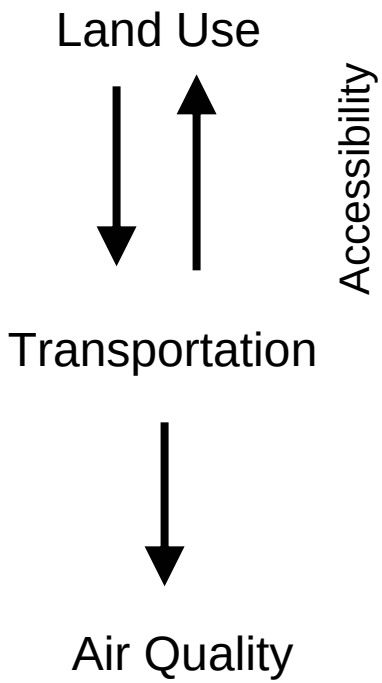
- Goal:
  - Improve Transport, Land Use and Environmental Policies and Investments – C*onsidering their Interactions*
- Means:
  - Models of Markets: Demand, Supply, Policies
    - Real Estate, Transport, Labor
  - Economic, Social and Environmental Assessment

Land Use - Transportation Interaction

Traditional Approach



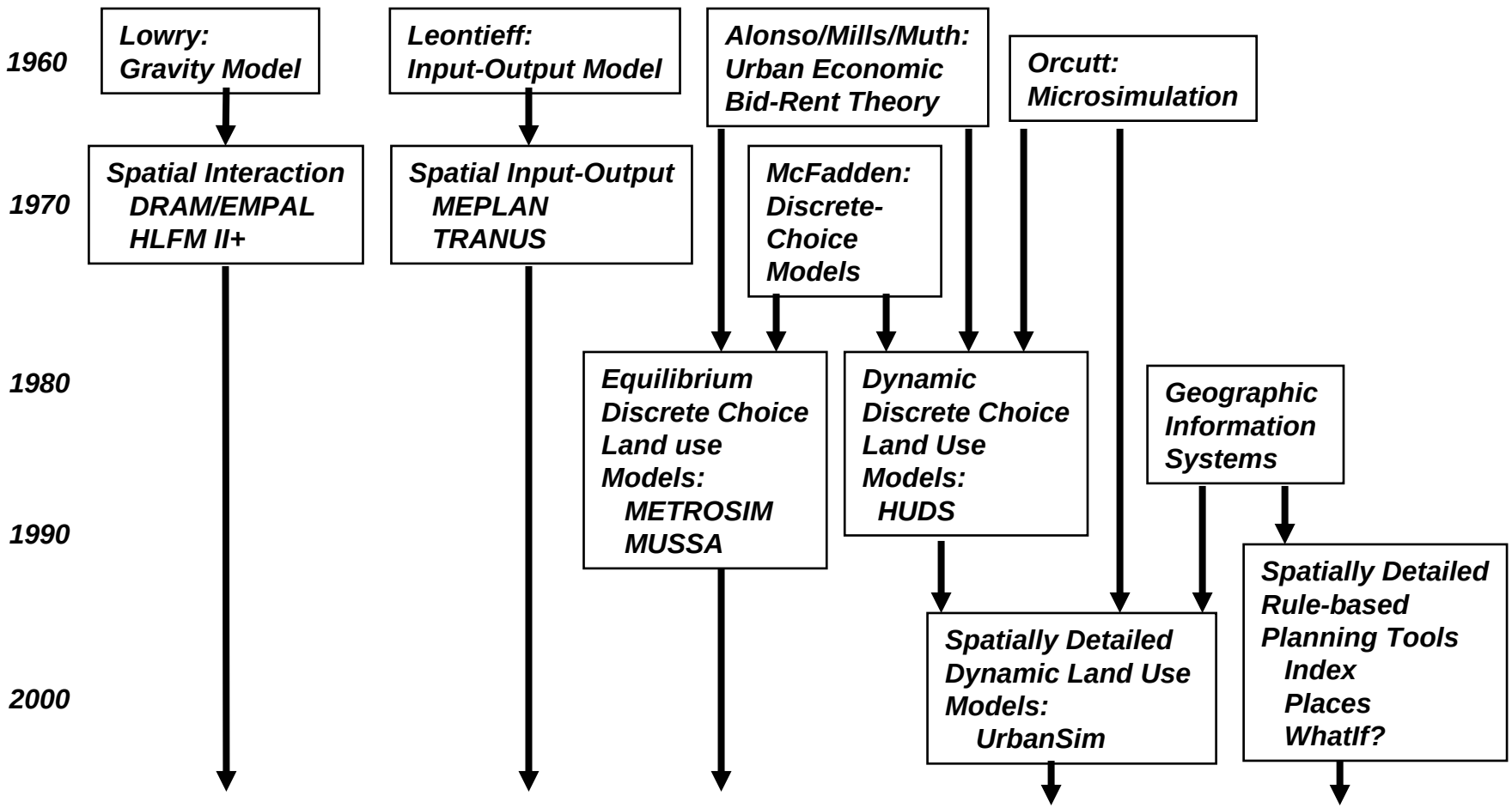
Integrated Modeling



Integrated Land Use – Transport Models

- From ~ 1970:
  - Dram/Empal – Putman
  - Meplan – Cambridge - Echenique
  - Tranus – Caracas - de la Barra
- From ~ 2000:
  - Delta – London – Simmonds
  - ILUMASS – Dortmund - Wegener
  - ILUTE – Toronto - Miller
  - Metrosim – Buffalo - Anas
  - MUSSA – Santiago - Martinez
  - Pecas – Calgary – Hunt
  - UrbanSim – Seattle – Waddell

Evolution of Land Use Models



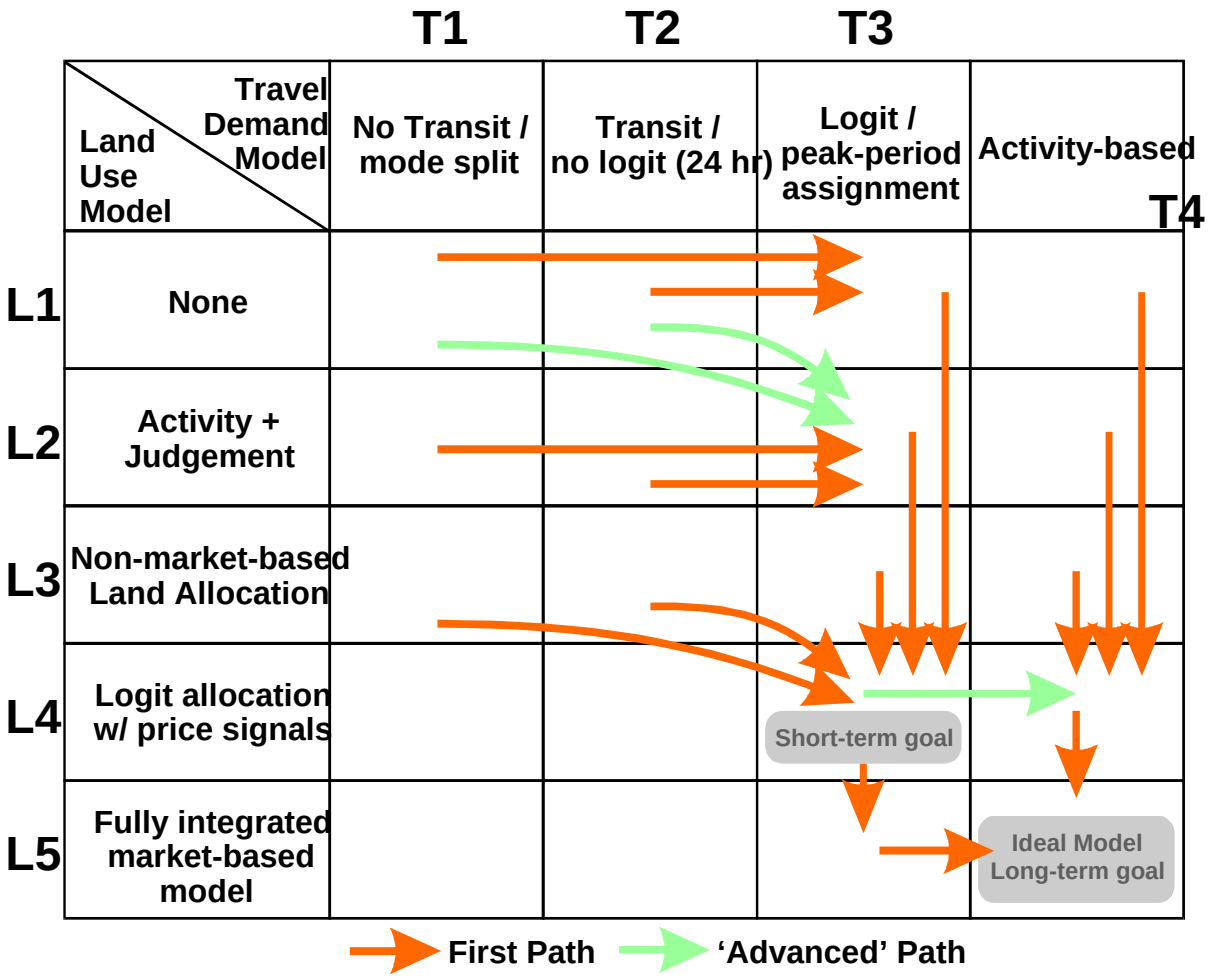
Emerging Common Design Elements

- Agent-level Microsimulation
  - Households – persons
  - Firms – jobs
  - Parcels – grid cells
- Dynamic Path-dependence
  - Real estate demand changes rapidly
  - Real estate development takes time to respond
  - Disequilibrium is the norm: “Boom and Bust”
  - History affects current choices

Emerging Common Design Elements

- Many Key Processes are Choice Problems
  - Household Residential Location
  - Household Vehicle Ownership
  - Individual Workplace
  - Firm Location
  - Real Estate Development
- Increasing Use of Discrete Choice Models
- Convergence with Activity-based Travel Modeling

Evolution of Land Use – Transport Models



Source: Miller, Kriger and Hunt 1998

UrbanSim

- Designed as a Land Use Model to Couple to Existing 4-Step Travel Models
- Applications in Multiple Metropolitan Areas
- Funded Mainly by National Science Foundation
  - Urban Research Initiative
  - Information Technology Research
  - Digital Government
  - Biocomplexity
- Collaborators from Multiple Disciplines
- Open Source Software (www.urbansim.org)

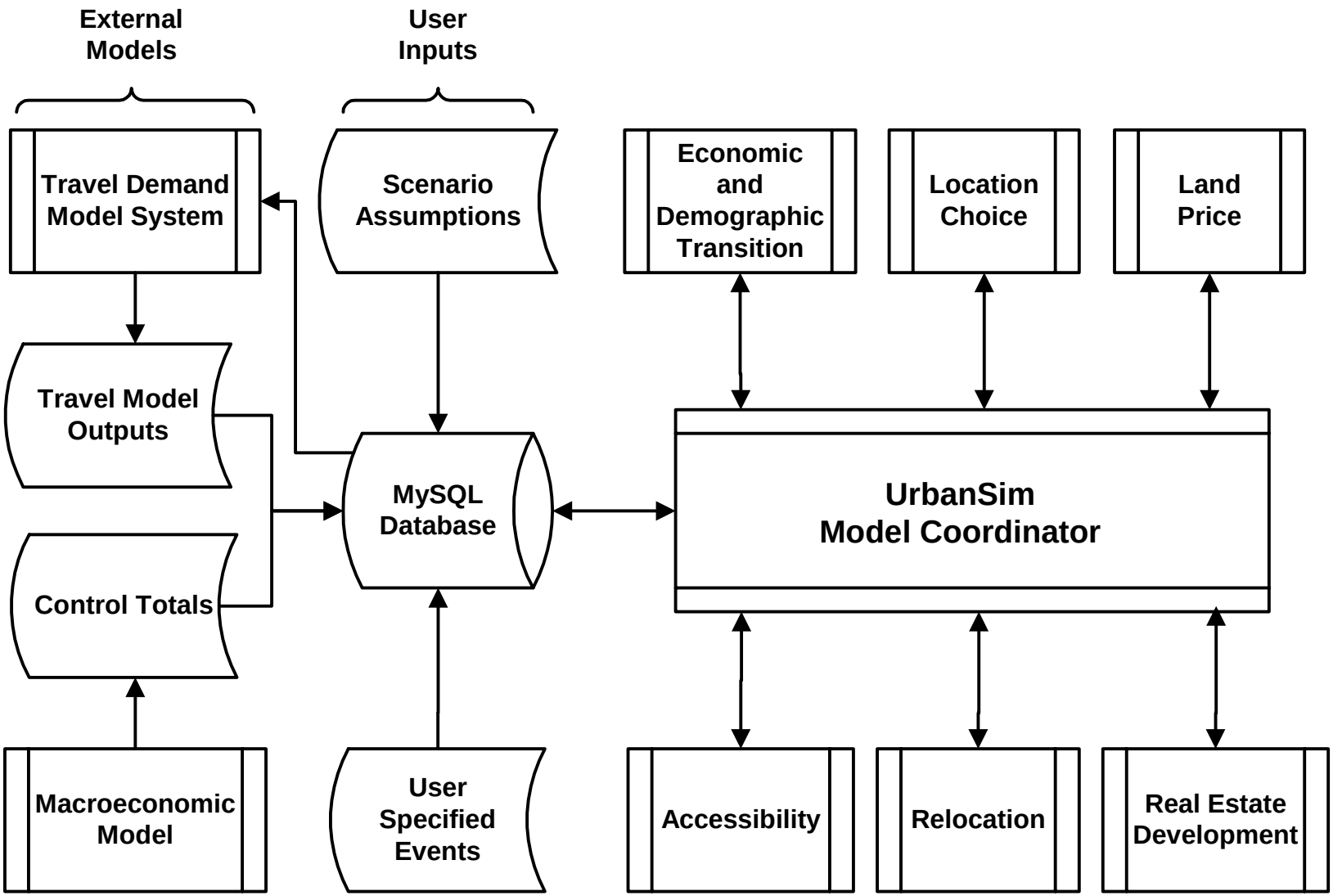
UrbanSim: Data Inputs/Outputs

- Inputs
  - Regional Control Totals
  - Parcel Data
  - Business Establishments
  - Household Data (Census, Travel Survey)
  - Land Use Plan
  - Environmental Constraints
- Outputs by traffic analysis zone and grid cell
  - Households by Income, Age, Size, Children, Workers
  - Employment by Industry Sector
  - Land Use and Density
  - Housing Units, Commercial SQFT, Prices by Type

UrbanSim Model Components

- Household and Employment Mobility and Location
  - Multinomial logit
  - Grid cell is choice outcome
  - Individual household or job is unit of analysis
- Real Estate Development
  - Multinomial logit
  - 24 development type outcomes
  - Grid cell is unit of analysis
- Real Estate Price Estimation
  - Hedonic regression
  - Grid cell is unit of analysis

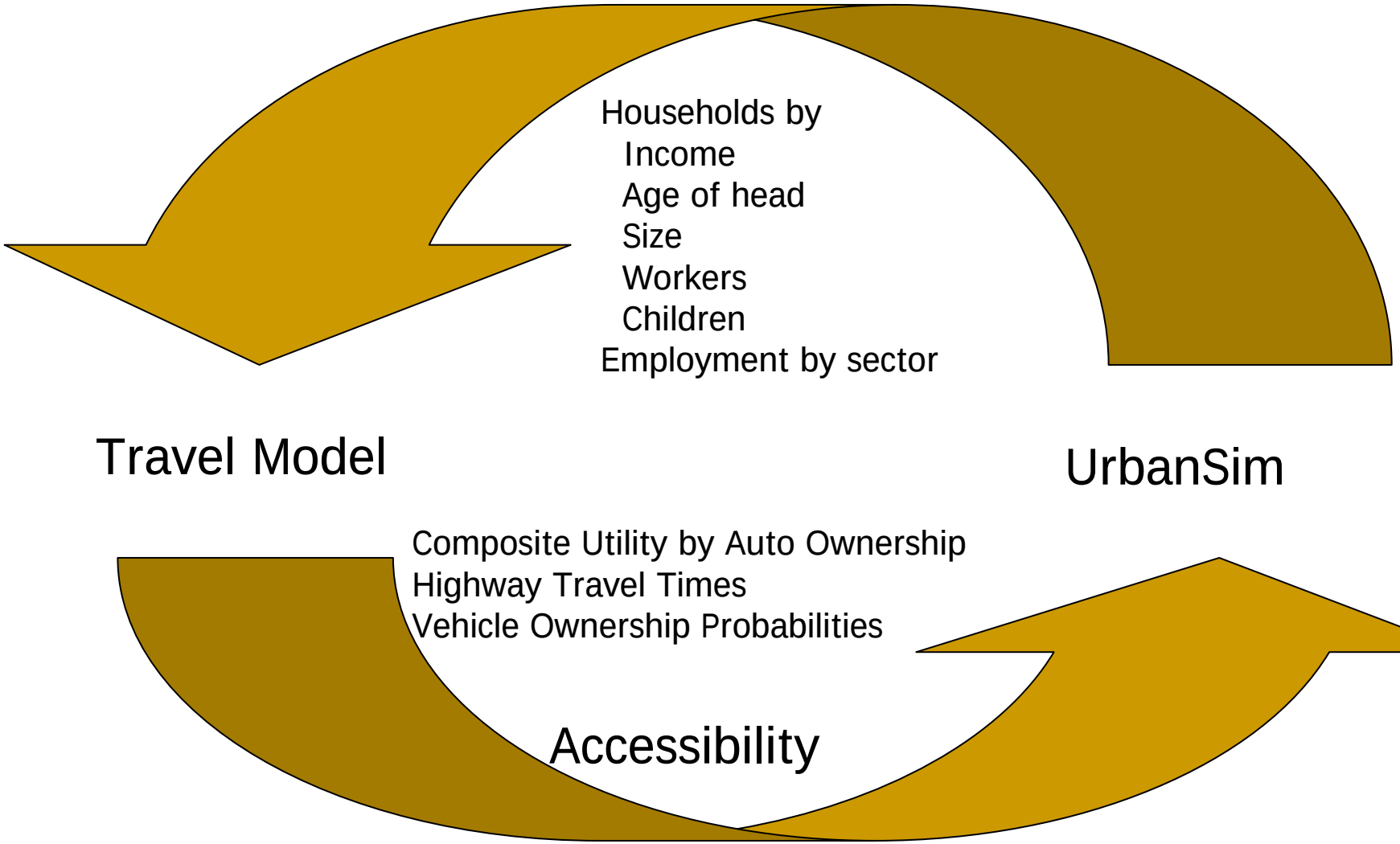
UrbanSim Model Structure



UrbanSim Software Architecture

- UrbanSim uses a set of *component models* representing actors and processes in the urban environment:
  - Land price
  - Real estate developer
  - Residential location
  - Employment location
  - Demographic transition
  - Economic transition
  - travel
- Conflicting demands - component models
  - should be as independent as possible to facilitate evolution and change
  - need to be able to interact efficiently
- Approach:
  - “pluggable models” – these communicate only via a shared database of grid cell and other data, not directly with each other

UrbanSim – Travel Model Interactions



UrbanSim Accessibility Measures

- Logsum-based Measures of Accessibility
  - Accessibility to employment

$$Access_i = \sum_j Jobs_j e^{Logsum_{ij}}$$

Logsum from Mode Choice

UrbanSim Accessibility Measures

- Trip-Weighted Measures of Accessibility
  - Accessibility for Home-based Work Purpose

$$Access_i = \frac{\sum_j (Trips_{ij} Logsum_{ij})}{\sum_j Trips_{ij}}$$

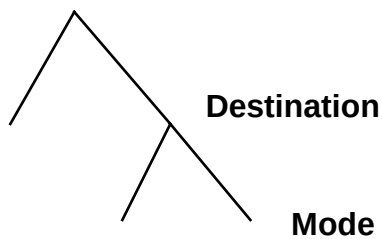
Logsum from Mode Choice

UrbanSim Accessibility Measures

- Accessibility from Combined Destination and Mode Choice
  - Accessibility for Home-based Work Purpose

$$Access_i = Logsum_i$$

Logsum from Destination & Mode Choice



UrbanSim Accessibility Measures

- Activity-based modeling opens new opportunities:
  - Long-term choices given:
  - Residence location
  - Workplace
  - School location
- Accessibility is an **individual** expectation
  - Time
  - Cost
  - Reliability

UrbanSim Walking-Scale Accessibility

- Most Travel Models Limited in Measuring Pedestrian-scale Accessibility
  - Large Zones
  - Coarse Networks
  - Difficulties Predicting Intra-zonal Travel
  - Difficulties Predicting Walking Trips
- UrbanSim Augments Travel Models by
  - Measuring Walking-scale Accessibility
  - Walking-distance (e.g. 600 meters) Radius
  - Measure Access to Shopping and Other Activities


150 Meter Grid Cells
And Parcels

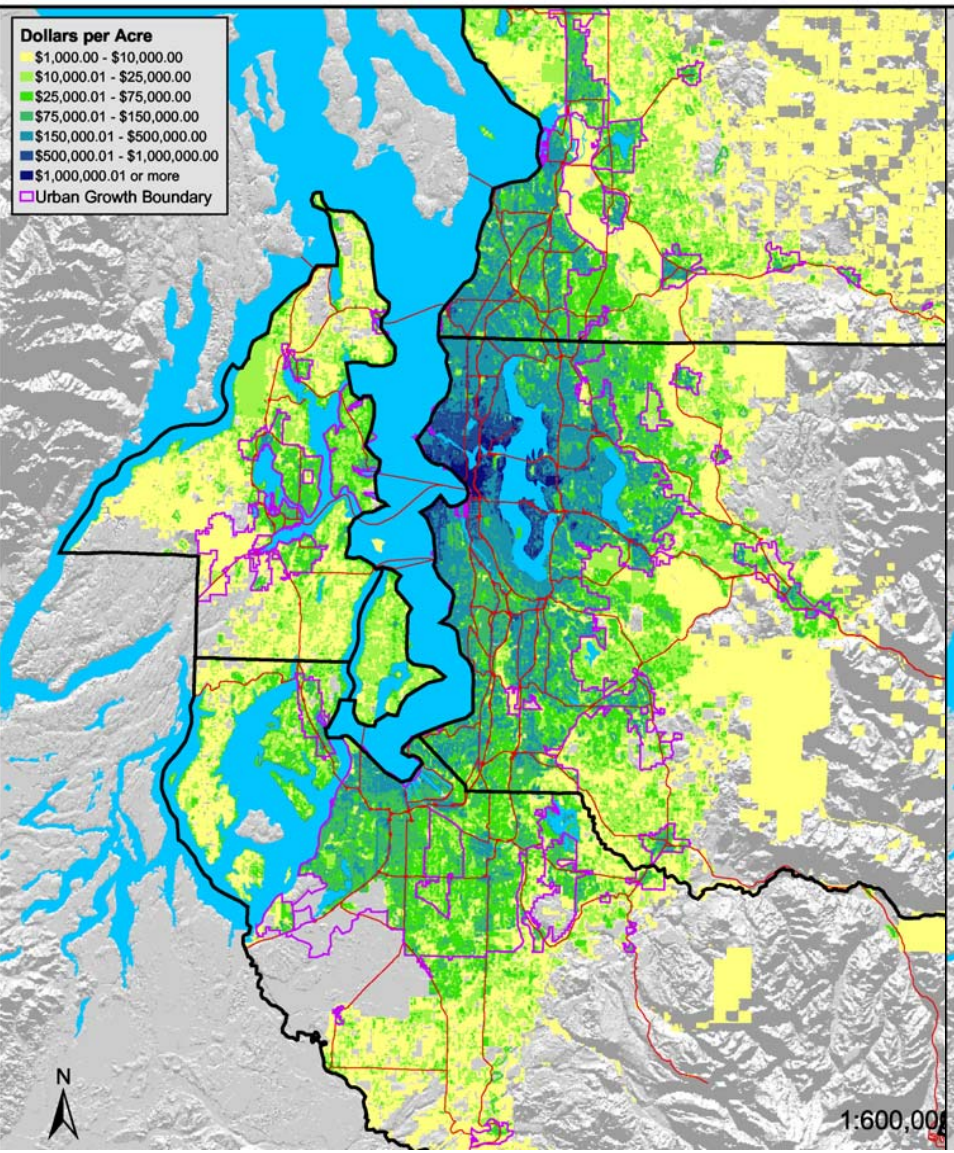
Land Price

$$\ln(P_i) = \alpha + \beta S_i + \delta N_i + \gamma A_i + \varepsilon_i$$

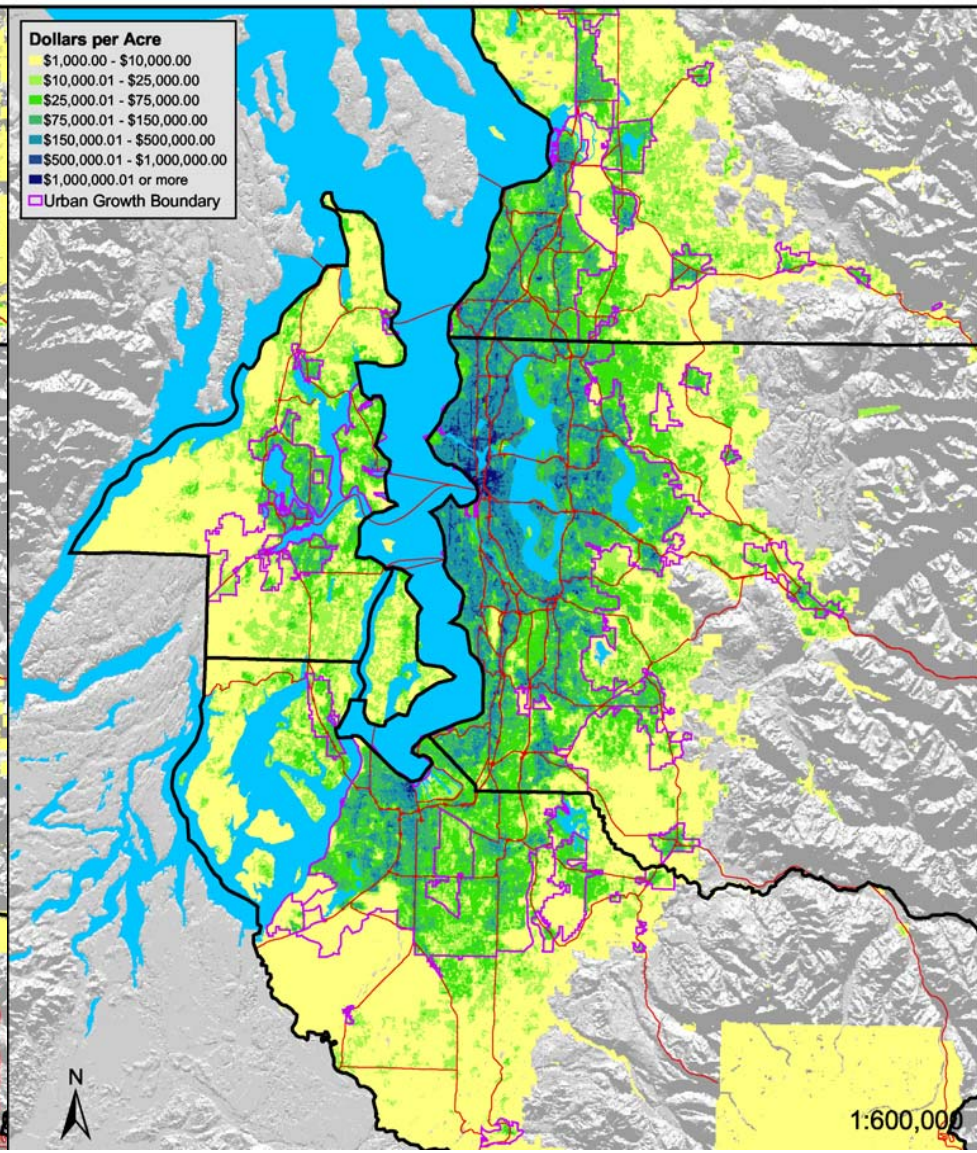
S is a set of site attributes
N is a set of neighborhood attributes
A is a set of accessibility attributes

- Predicts grid cell land price
- Multiple regression (hedonic) specification
- Used sample of 100,000 cells
- Explains 80% of variation
- Example of variables used:
  - Type of development in cell
    - Residential, Mixed Use, Commercial, Industrial, Government
  - Travel time to Seattle CBD
  - Highway
    - Adjacency; distance from
  - Housing units
    - In cell; within walking distance
  - Commercial sqft
    - In cell; within walking distance
  - Employment within walking distance
  - Percent: floodplain; water; wetland; open space; public land
  - Land Use Plan designation

PSRC Region
2000 Total Land Value per Acre by Gridcell



PSRC Region
Predicted 2000 Total Land Value per Acre by Gridcell



Household Location Choice

- Predicts grid cell location choice
- Applies to new and moving households
- Multinomial Logit specification
- Used 1999 Household Travel Survey
  - 2,364 households who moved within 5 yrs
- Example of Variables used:
  - Housing cost to income ratio
  - Income * improvement value/unit
  - Trip-weighted utility for HBW by SOV; Transit
  - Near arterial road
  - Housing density within walking distance
  - Housing density if HH has children
  - High density if HH is young
  - Mixed use development if HH is young
  - Percent High Income if HH is high income
  - Percent Mid Income if HH is mid income
  - Percent Low Income if HH is low income
  - Percent Minority if HH is minority
  - Percent Minority if HH is not minority

$$U_i = u_i + \varepsilon_i$$

$$P_i = \frac{e^{\beta_i \cdot x}}{\sum_{\forall i'} e^{\beta_{i'} \cdot x}}$$

Employment Location Choice

$$U_i = u_i + \varepsilon_i$$

$$P_i = \frac{e^{\beta_i \cdot x}}{\sum_{\forall i'} e^{\beta_{i'} \cdot x}}$$

- Predicts grid cell location choice
- Applies to new and moving jobs
- Multinomial Logit specification
- 1995-2000 TAZ employment change
- Example of Variables used:
  - Land value in area
  - Total value of land and improvements
  - Trip-weighted (destination) utility for HBW by SOV
  - Travel time to Seattle CBD
  - Employment by sector in area
  - Industrial; commercial sqft
  - Near arterial road
  - Near highway
  - Housing density in area
  - Percent low income in area; mid-income
  - Building age

Real Estate Development Model

$$U_i = u_i + \varepsilon_i$$

$$P_i = \frac{e^{\beta_i \cdot x}}{\sum_{\forall i'} e^{\beta_{i'} \cdot x}}$$

- Predicts grid cell development events
- Multinomial Logit specification
- One equation per starting development type
- Used 1995-2000 development events
- Example of Variables used:
  - Value of land and improvements
  - Land value per acre in area
  - Employment by sector in area
  - Housing units in area
  - Proximity to existing development
  - Development composition in area
  - Recent development events in area
  - Travel time to Seattle CBD
  - Trip-weighted travel utility
  - Highway adjacency and distance from
  - Percent: floodplain; water; wetland; stream buffer; steep slope

Problems (Research Opportunities)

- Coupling with 4-Step Travel Models
  - Slow and Theoretically Limiting
- Coupling of Model Estimation and Prediction
  - Separate Processes – Slow and Error Prone
  - Need to Explore More Complex Model Structures
- Software Implementation
  - + Robust and efficient Java code
  - + Good Development Methodology (Test/Build)
  - - But Too Complex: difficult to diagnose, extend
- Individual Model Components Need Refinement
- Need Improved System Validation Techniques

UrbanSim – Agenda

- Reimplement System
  - Python + numarray: Faster, Less Code, Clearer
  - Modular “Plug-and-play” Model Components
- Integrate Model Estimation and Application
- Integrate with Activity-Based Travel Model
- Integrate with Assignment
- Integrate with Air Quality Models

- There is too much to do without collaboration!

A Vision: International Collaboration

- We are not the Only Project with these Aims..
  - ILUTE, ILUMASS, Albatross, etc.
  - Similar in many design respects
  - Share many implementation requirements
- Why not Cooperate?
  - Create Common Core Architecture
  - Develop Interoperable Components
  - Reduce Resource Requirements
  - Focus Effort
  - Publish more Papers

Opus: An Infrastructute for Collaboration

- Open Platform for Urban Simulation
  - Must be Open Source
  - Must be Very Modular
  - Must Have Clear API
  - Must Accommodate Components Written in
    - C++, C, Python, probably also Java
  - Must Support Incremental Development
    - Interface Existing Software Components
    - Refine Components and Interfaces Incrementally
  - Must Support Remote Collaboration

Why Open Source?

- Consistent With Academic Research Standards
  - 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

A Complete Open Source System???

- Model Components –
  - From UrbanSim,ILUTE,ILUMASS…
    - Household and Firm Evolution (Demography)
    - Household and Firm Relocation/Location Choice
    - Real Estate Development
    - Workplace Choice
    - Vehicle Ownership
  - From CEMDAP, FAMOS, Albatross…
    - Activity Schedule, Destination, Mode Choices
  - From MATSIM, SUMO, Ciudadsim…
    - Traffic Assignment

A Complete Open Source System???

- Software Components
  - From Biogeme, R…
    - Model Estimation
  - From MySQL, Postgress…
    - Database Management
  - From Saga, TerraLib, Thuban…
    - GIS functionality
  - From wxWidgets, OpenGL…
    - Graphical User Interface
  - From Python…
    - Scripting Language

It Has Been Done in Another Domain

- Computational Statistics
- Commercial Product: S+ -> Open Source Version: R
  - R now widely recognized as a standard
- Core System
  - Shell with Basic GUI
  - Scripting Statistical Language
  - User-Contributed Packages
    - Standardized API
    - Standardized Documentation
    - 30 Seconds to Install a Package from Web
- Emerging Publishing Standard:
  - Papers publishing a new statistical method now commonly release an R Package implementing it, with test data.

Will this Be Difficult to Do for Land Use – Transport Modeling?

- Of course!
  - Coordination of Diverse Objectives
  - Globally Dispersed Collaborators
  - Theoretical Challenges
  - Technical Integration Challenges
  - Disagreement on Design Choices
  - Resources
- But the statistical computing community has overcome similar challenges…

Is it Worth Trying?

?