The Devil is in the Details: From Zones to Parcels in Integrated Models

Paul Waddell
Professor and Chair
Department of City and Regional Planning
University of California Berkeley

1. **The Devil is in the Details**
2. Agent Details
3. Temporal Details
4. Geographic Details
5. A Strategy for Incremental Development

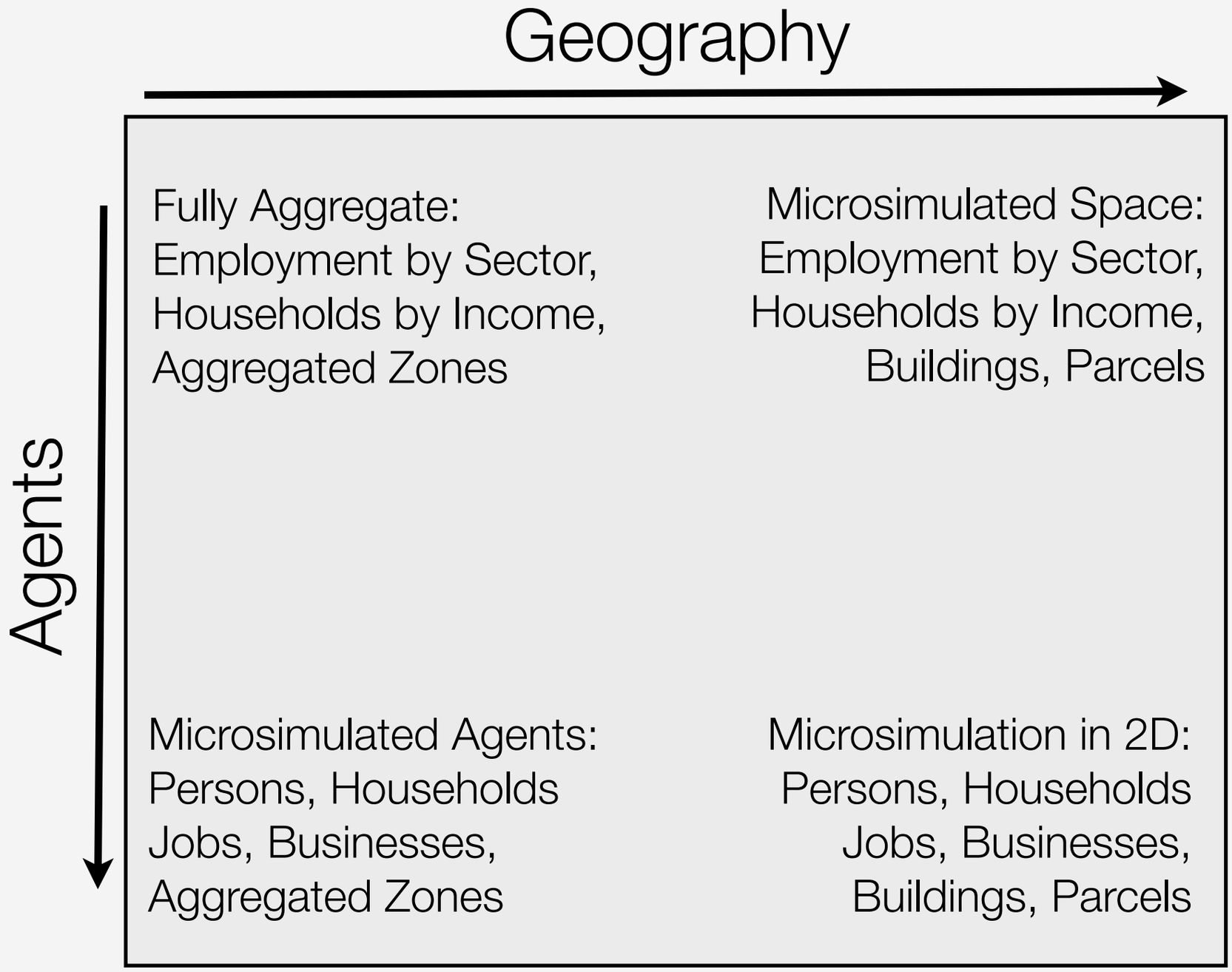
The idiom "the devil is in the details" derives from the earlier phrase, "God is in the detail;" expressing the idea that whatever one does should be done thoroughly; i.e. details are important.

Wikipedia

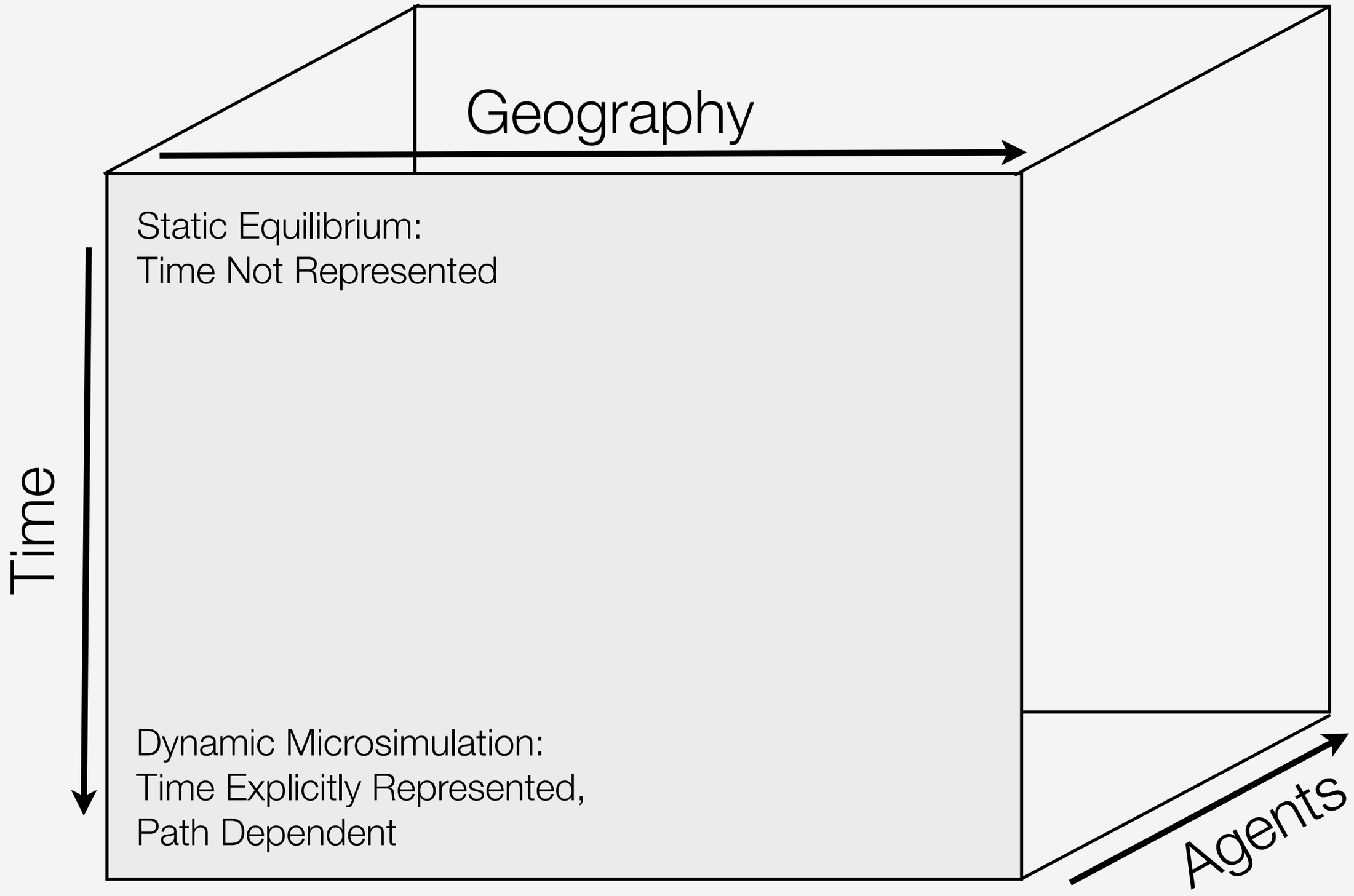
How Much Detail do we need in Integrated Modeling?

- **How much detail do we need in land use and transportation models in order to achieve models that are unbiased and allow us to assess relevant transportation and land use policies?**
  - How small should zones be?
  - Should we use zones at all? What about using parcels, or small gridcells?
  - How much detail do we need in terms of population and employment?
  - Should we use aggregations of households and jobs, or microsimulate?
- **Which details are important?**
- **What biases might our models have if we leave out details?**
- **What errors might we introduce if we use too much detail?**

Detail in 2 Dimensions



Detail in 3 Dimensions



1. The Devil is in the Details
2. **Agent Details**
3. Temporal Details
4. Geographic Details
5. A Strategy for Incremental Development

To Microsimulate, or Not to Microsimulate. That is the Question

- **Traditional urban models have ALL been aggregate**
- **Economic models use a representative agent to motivate model**
- **Urban economics is based on the Monocentric model**
  - Disaggregated at most by industry, households by high vs low income
  - Analytically based models are not tractable with much detail
- **Later, more applied models began disaggregating:**
  - DRAM/EMPAL: households by income quartile, employment by sector
  - TRANUS/MEPLAN/PECAS: use more sectors and household categories
- **Microsimulation models are fundamentally different:**
  - Originated with work of Guy Orcutt, initially for policy analyses like tax incidence
  - Represent individual agents (households, persons, jobs, businesses)
  - Maintain these as lists, and update them as the model progresses

To Microsimulate, or Not to Microsimulate. That is the Question

- **Arguments used in favor of aggregating individual agents:**
  - Data is more compact and easy to use (when small numbers of types used)
  - Leads to simpler models (depends on model logic: can be very complex / black box)
  - Models run faster (this is debatable and depends on complexity and implementation)
  - Less prediction error due to aggregation (this is an empirical question)
- **Arguments used in favor of microsimulating individual agents:**
  - Data are more natural to understand and work with since they represent real agents
  - Models can be more straightforward in their logic: agents make choices
  - Models can be designed to run at least as fast as aggregate models (especially aggregate models with many categories of agents and outcomes)
  - Avoids Ecological Fallacy: a common error in social science research, where individual behavior is inferred from aggregate data
  - Model parameters are less likely to be biased
- **Field is moving strongly in the direction of microsimulation of agents**

1. The Devil is in the Details
2. Agent Details
3. **Temporal Details**
4. Geographic Details
5. A Strategy for Incremental Development

Time is of the Essence

- **Urban models have generally ignored time**
  - Time-abstract models use a notion of static equilibrium from economics
  - We assume that a city-region is in equilibrium, perturb the equilibrium, and observe a new static equilibrium
  - Attractive for theoretical analysis since it follows from theory and provides consistent answers every time
- **Recent models have begun to represent time explicitly**
  - Empirical observation suggests that markets may often be in sustained disequilibrium: subprime mortgage crisis and subsequent global recession
  - Increasingly common to represent annual time steps reflecting differing response times in real estate supply, intra-year changes in household location, business location, prices
  - Path dependence is a feature of this kind of modeling: changes today have implications for later choices (developers go bankrupt because of imperfect foresight)
- **Field is moving strongly in the direction of explicit representation of time**

1. The Devil is in the Details
2. Agent Details
3. Temporal Details
4. **Geographic Details**
5. A Strategy for Incremental Development

Location, Location, Location

- **Most common arguments in favor of using zones vs parcels are:**
  - Less data requirements
  - Easier to develop
  - Easier to diagnose
- **Most common arguments against using zones:**
  - Modifiable Aerial Unit Problem (MAUP): model results depend heavily on configuration of zones; parameters sensitive to zone configuration
  - Ecological Fallacy: easy to fall into a classic mistake in social science research: inferring individual behavior from aggregate data
  - Walking scale is below the radar: may bias models with respect to intra-zonal trips, non-motorized trips, transit trips (with walk access); this is compounded by using travel networks that exclude local streets

Location, Location, Location

- **Most common arguments in favor of using parcels vs zones are:**
  - Data is becoming more readily available from tax assessors and commercial sources
  - Parcels are real: they are the unit of land that is owned, subdivided, and developed
  - Easier to interface local jurisdictions plans and zoning with regional plans
- **Most common arguments against using parcels:**
  - Messy data
  - Large data storage and processing
  - Messy data
  - Difficult to standardize across jurisdictions
  - Messy data
  - Can take 2+ years to make parcel-level data useful for modeling using generally available data techniques

1. The Devil is in the Details
2. Agent Details
3. Temporal Details
4. Geographic Details
5. **A Strategy for Incremental Development**

Start Simple, Then Add Details

- **Experience has demonstrated that an incremental development path may be most productive:**
  - Begin with a very simple zone model configuration
  - Begin using and assessing the model
  - Add detail to the model incrementally, targeting most pressing needs
- **Incremental development using a single platform**
  - Open Platform for Urban Simulation (OPUS) and UrbanSim
  - Use microsimulation of agents and explicit representation of time, just vary geographic detail
  - Modular configuration and estimation of models
  - Zonal model configurations
  - Synthesizing parcel details
  - Parcel model configurations
  - Exploiting parcels and local streets

UrbanSim: Start From the Simplest Zonal Configuration

Household Location Models

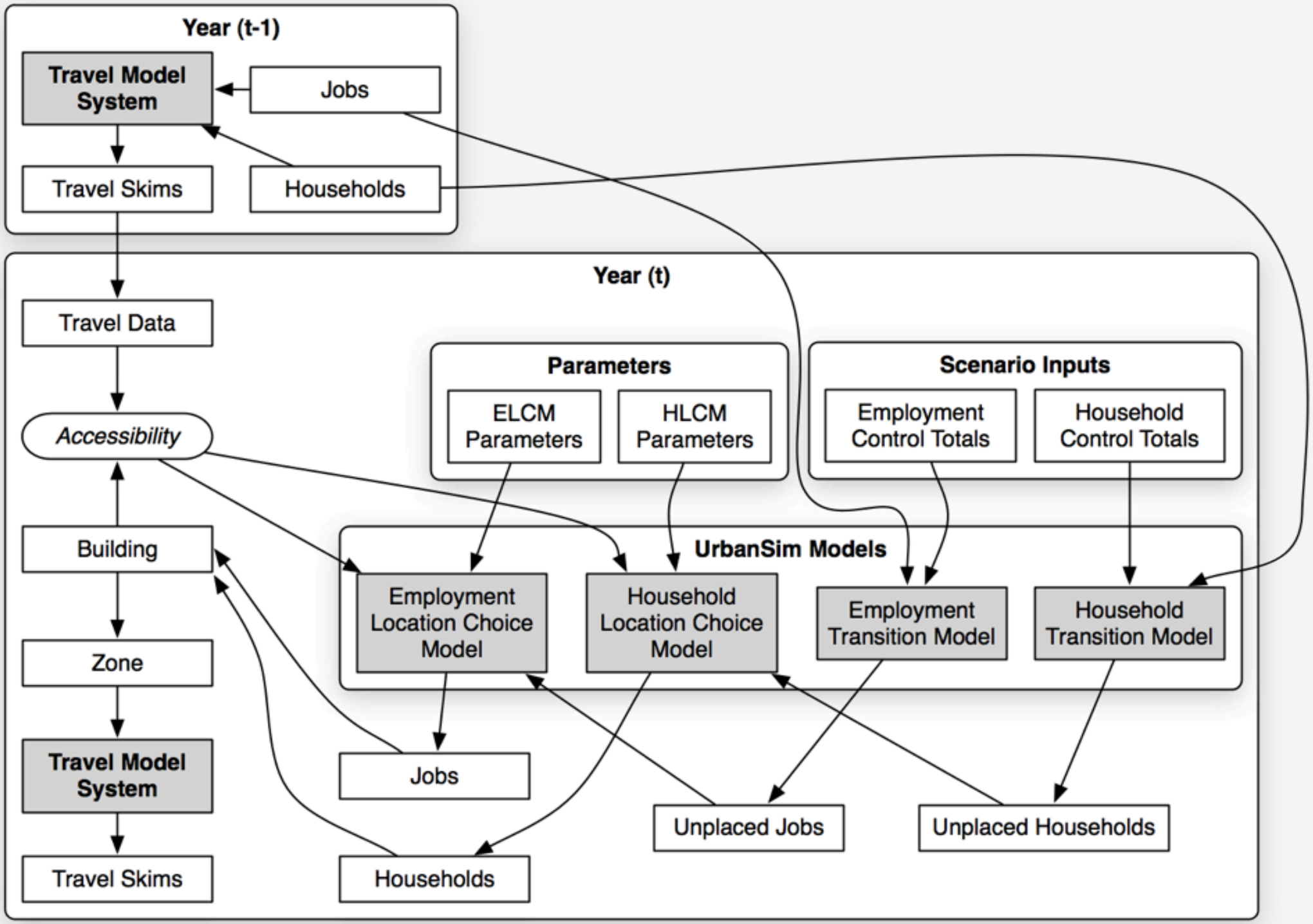
- **Household Transition Model**
- **Household Location Choice Model**

Employment Location Models

- **Employment Transition Model**
- **Employment Location Choice Model**

No representation of supply side of real estate market, or prices. No relocation of agents once placed. Becomes an 'incremental' model, allocating growth.

The Simple Zone Configuration of UrbanSim: In Detail



UrbanSim: Add Relocation Dynamics

Household Location Models

- Household Transition Model
- Household Relocation Model
- Household Location Choice Model

Employment Location Models

- Employment Transition Model
- Employment Relocation Model
- Employment Location Choice Model

Being used in Research Triangle Park, North Carolina. No representation of supply side of real estate market, or prices. Last resort when there is no data on supply.

UrbanSim: Add Real Estate Supply and Price

Land Development Models

- Real Estate Price Model
- Residential Development Project Location Choice Model
- Nonresidential Development Project Location Choice Model
- Building Construction Model

Household Location Models

- Household Transition Model
- Household Relocation Model
- Household Location Choice Model

Employment Location Models

- Employment Transition Model
- Employment Relocation Model
- Employment Location Choice Model

Being used in Maricopa County (MAG). Two different levels of geography being tested: travel model zones, and 'super-parcels' or aggregated parcels by land use/block.

UrbanSim: Add Labor Market & Workplace

Land Development Models

- Real Estate Price Model
- Residential Development Project Location Choice Model
- Nonresidential Development Project Location Choice Model
- Building Construction Model

Household Location Models

- Household Transition Model
- Household Relocation Model
- Household Location Choice Model

Employment Location Models

- Employment Transition Model
- Employment Relocation Model
- Employment Location Choice Model

Workplace Location Models

- Economic Transition Model
- Home-based Job Choice Model
- Workplace Location Choice Model
- Job Change Model

UrbanSim: Shift From Zones to Parcels as Locations

Land Development Models

- Real Estate Price Model
- Residential Development Project Location Choice Model
- Nonresidential Development Project Location Choice Model
- Building Construction Model

Household Location Models

- Household Transition Model
- Household Relocation Model
- Household Location Choice Model

Employment Location Models

- Employment Transition Model
- Employment Relocation Model
- Employment Location Choice Model

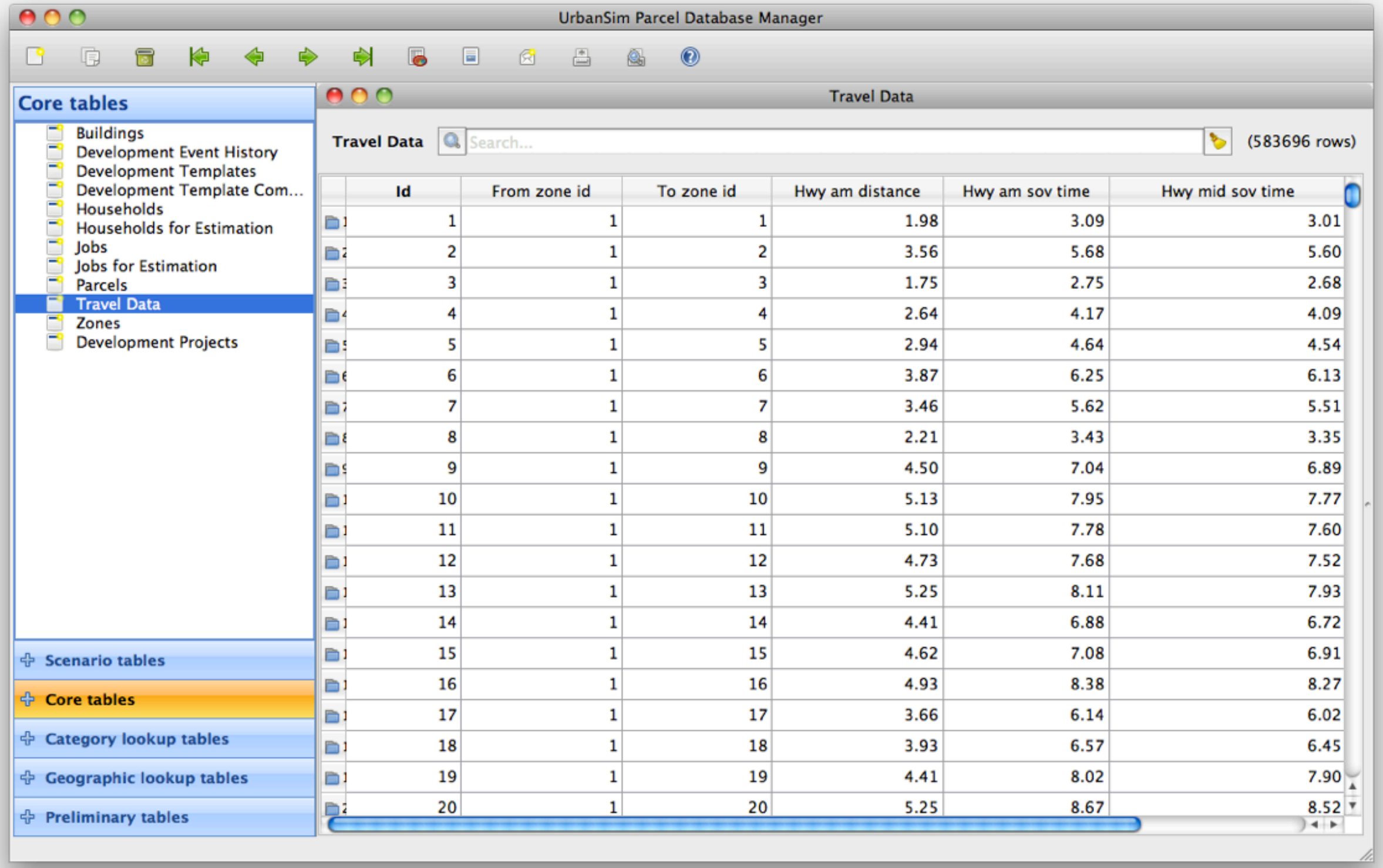
Workplace Location Models

- Economic Transition Model
- Home-based Job Choice Model
- Workplace Location Choice Model
- Job Change Model

Tools for Database and Model Development

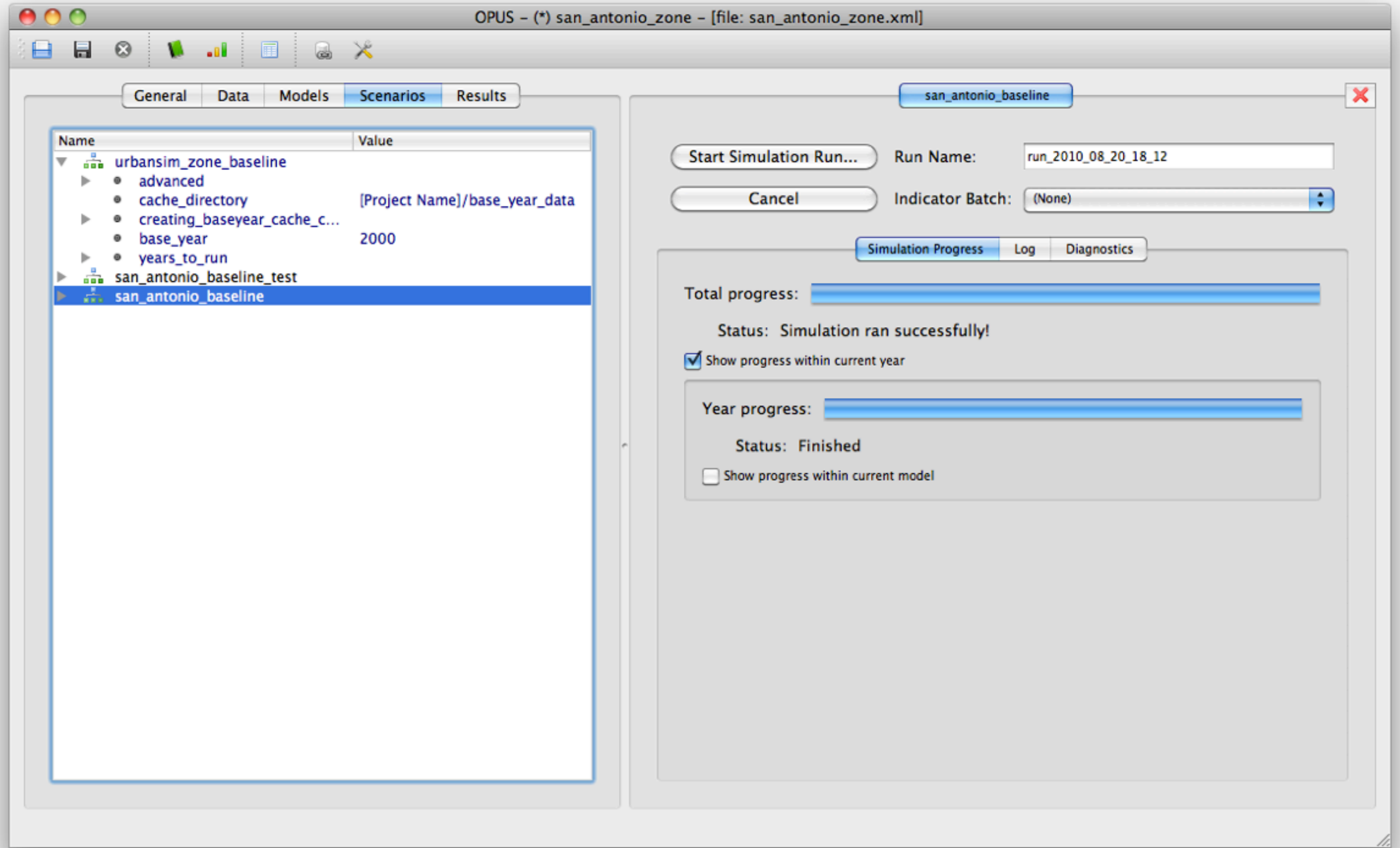
- **Database tools**
  - Schema generator for multiple platforms (MySQL, Postgres, SQLite...)
  - Python scripts for converting and loading data into the database
- **For zonal models, can start with aggregate zone-level data**
  - Zonal input file for Trip Generation
  - Travel skims
- **Zonal models can use traffic analysis zones, or other units of geography**
  - Paris uses Communes (~ 1,300 for Ile de France region)
  - Association of Bay Area Governments currently using traffic analysis zones (~1,450)
  - Maricopa Association of Governments experimenting with Super-parcels (~70,000)
- **Overcoming parcel data messiness**
  - Data synthesis and visualization
- **Multi-level modeling (county, district, zone, parcel)**

Database Schema, Data Loading, Browsing, Editing



	Id	From zone id	To zone id	Hwy am distance	Hwy am sov time	Hwy mid sov time
1	1	1	1	1.98	3.09	3.01
2	2	1	2	3.56	5.68	5.60
3	3	1	3	1.75	2.75	2.68
4	4	1	4	2.64	4.17	4.09
5	5	1	5	2.94	4.64	4.54
6	6	1	6	3.87	6.25	6.13
7	7	1	7	3.46	5.62	5.51
8	8	1	8	2.21	3.43	3.35
9	9	1	9	4.50	7.04	6.89
10	10	1	10	5.13	7.95	7.77
11	11	1	11	5.10	7.78	7.60
12	12	1	12	4.73	7.68	7.52
13	13	1	13	5.25	8.11	7.93
14	14	1	14	4.41	6.88	6.72
15	15	1	15	4.62	7.08	6.91
16	16	1	16	4.93	8.38	8.27
17	17	1	17	3.66	6.14	6.02
18	18	1	18	3.93	6.57	6.45
19	19	1	19	4.41	8.02	7.90
20	20	1	20	5.25	8.67	8.52

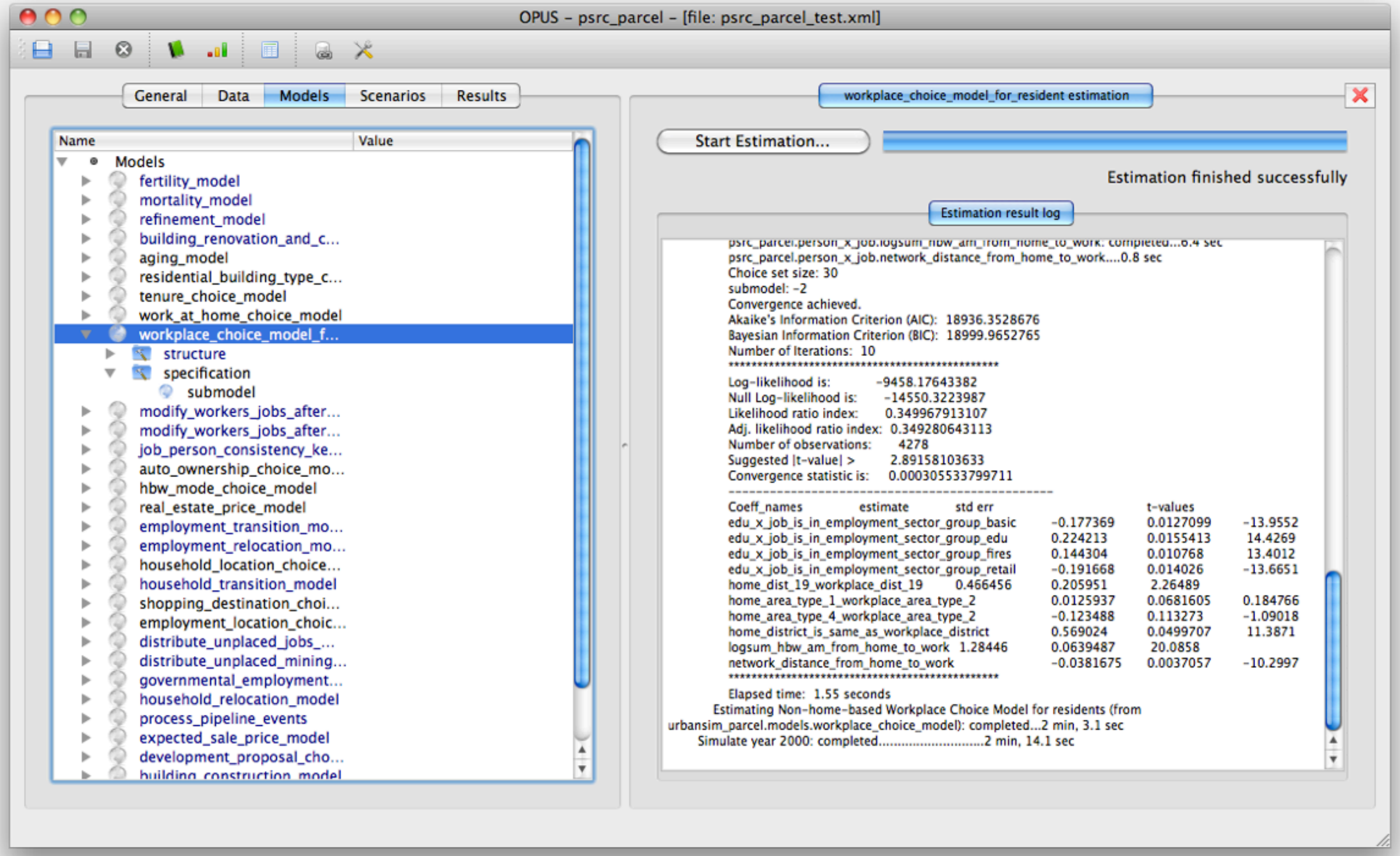
Interface for Developing, Estimating, Running Models



Measuring Progress: Benefits and Costs

- **Incremental model development ideally will monitor an evaluate progress:**
  - How much better is a new innovation compared to the best available version?
  - How much more costly is it in terms of computational expense or data effort?
- **How to measure progress**
  - Model estimation results
  - Model sensitivities
  - Calibration of model uncertainty over time

Estimation of Workplace Choice Model in GUI



Sensitivity Analysis: Relative Influence of Variables

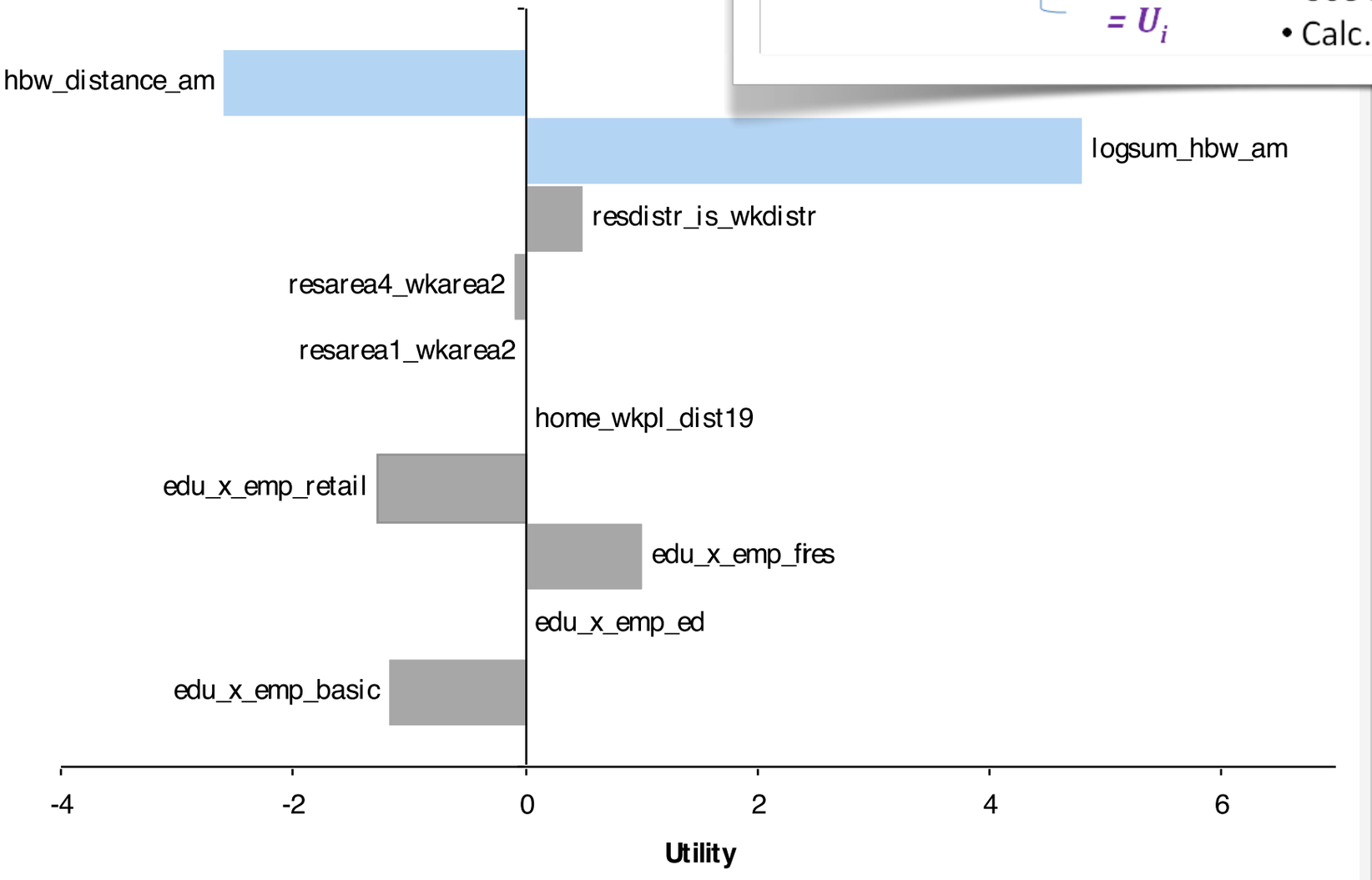
Estimated Parameters (β)

$\beta_{i1} x_{i1} +$
$\beta_{i2} x_{i2} +$
...
$\beta_{ik} x_{ik} +$
...
$= U_i$

All variables (x) except one held at median value

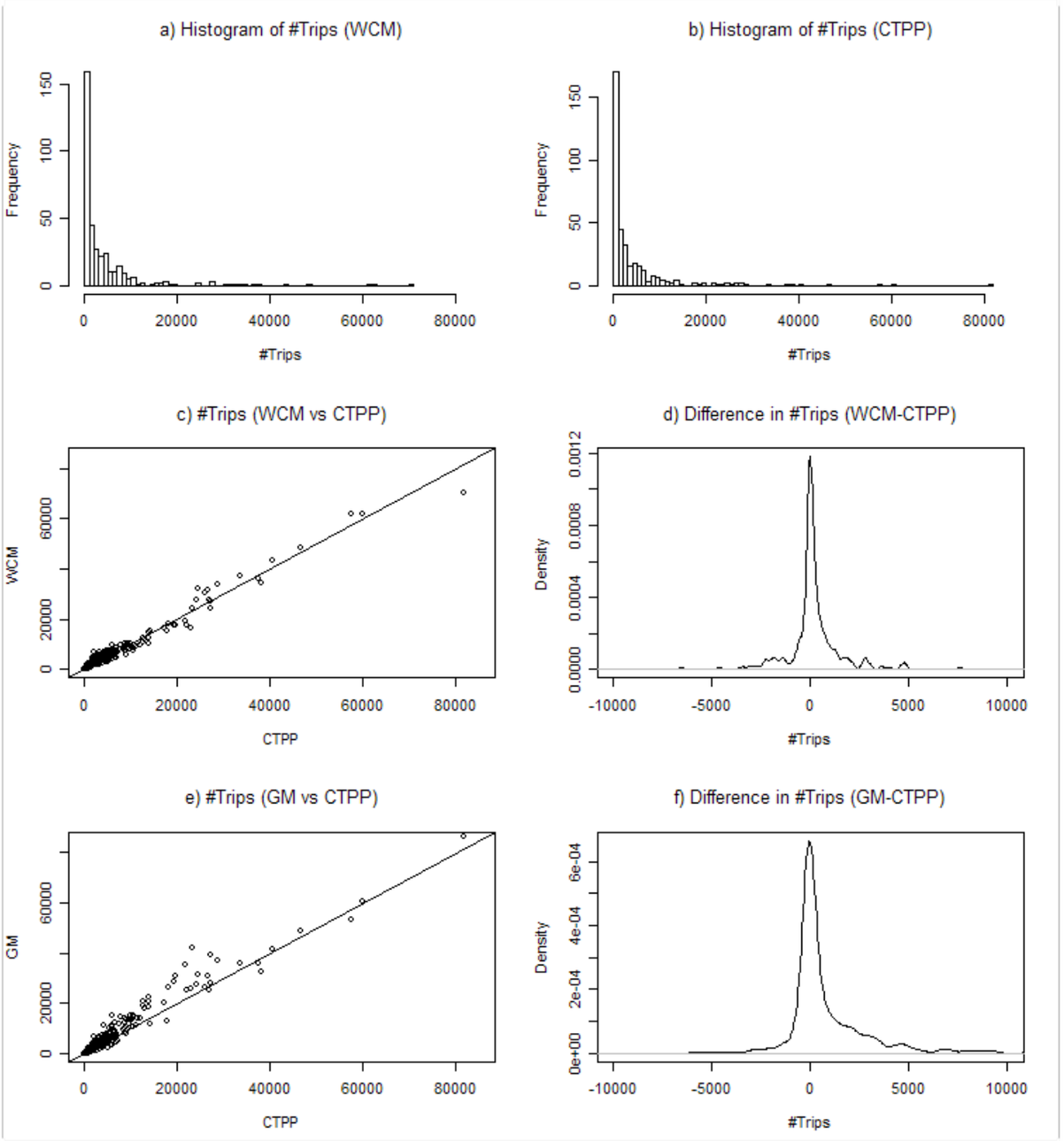
One variable:
- Use 5th & 95th percentile values
- Calc. Δ U for indication of influence

Workplace Location Choice Model



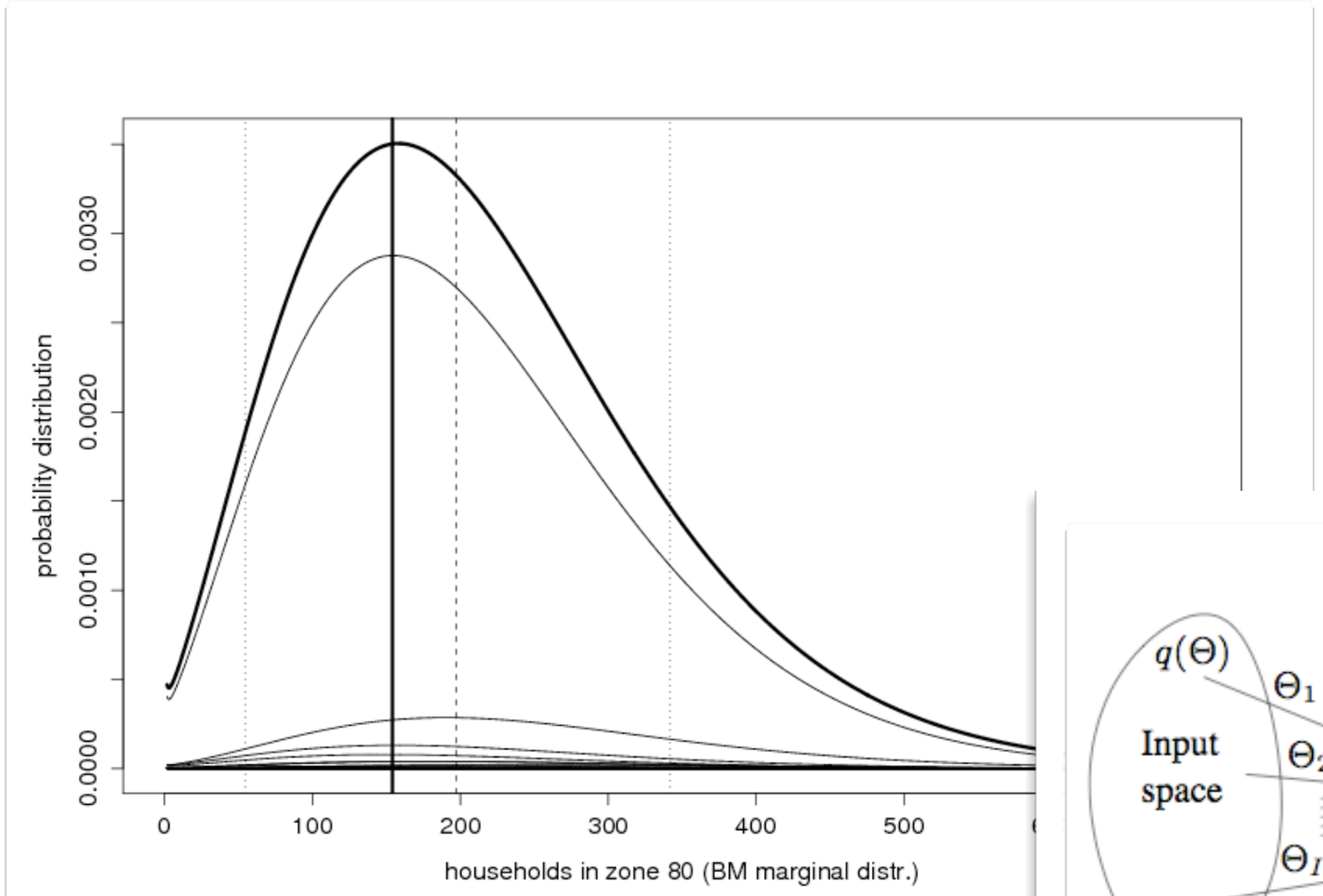
Calibration and Validation of Workplace Choice Model

- Model predicts individual workplace (attaches a job to a person) at parcel and building level
- Comparison of predicted values (commute trips) to observed values from CTPP to compute errors
- Calibration: Adding variables to specification to reduce errors.
- Errors compared to previous HBW Trip Distribution Model (gravity model)
  - RMSE Gravity Model = 2558.65
  - RMSE New Model = 1440.01



Application: Puget Sound Regional Council

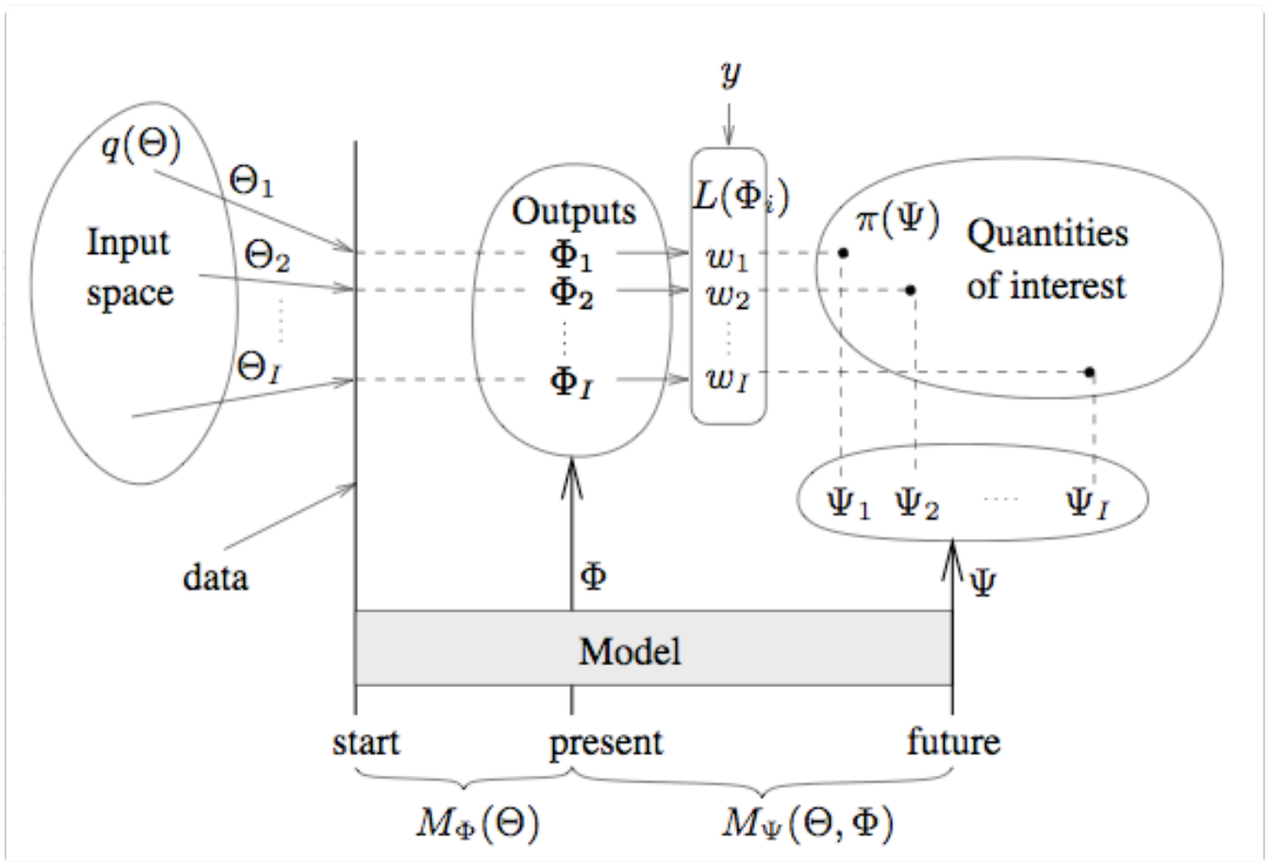
Calibrating Uncertainty in UrbanSim Model Application



Method	# Cases Missed by 90% Confidence Interval	Percent Covered Cases
Bayesian Melding	31	88%
Multiple Runs	163	38%

Bayesian Melding

Results from Eugene-springfield in Transportation Research B, 2007; Seattle Viaduct application in Transportation Research A, 2011.



Problem: if we are going to use parcel and building level detail to improve walking-scale analysis, we have to update streets, parcels and buildings as part of the simulation...

Creating a new open source platform for this: UrbanVision.

3 months into project for Metropolitan Transportation Commission. Scheduled launching for Sustainable Communities Strategies Public Engagement Process in Fall 2011.

Synthesizing Detail with Procedural Geometric Models

We created Monterey Peninsula streets, parcels, buildings in 3D using only 5 minutes of user input of the highways, coastline, undevelopable areas, center of jobs, total population and jobs.

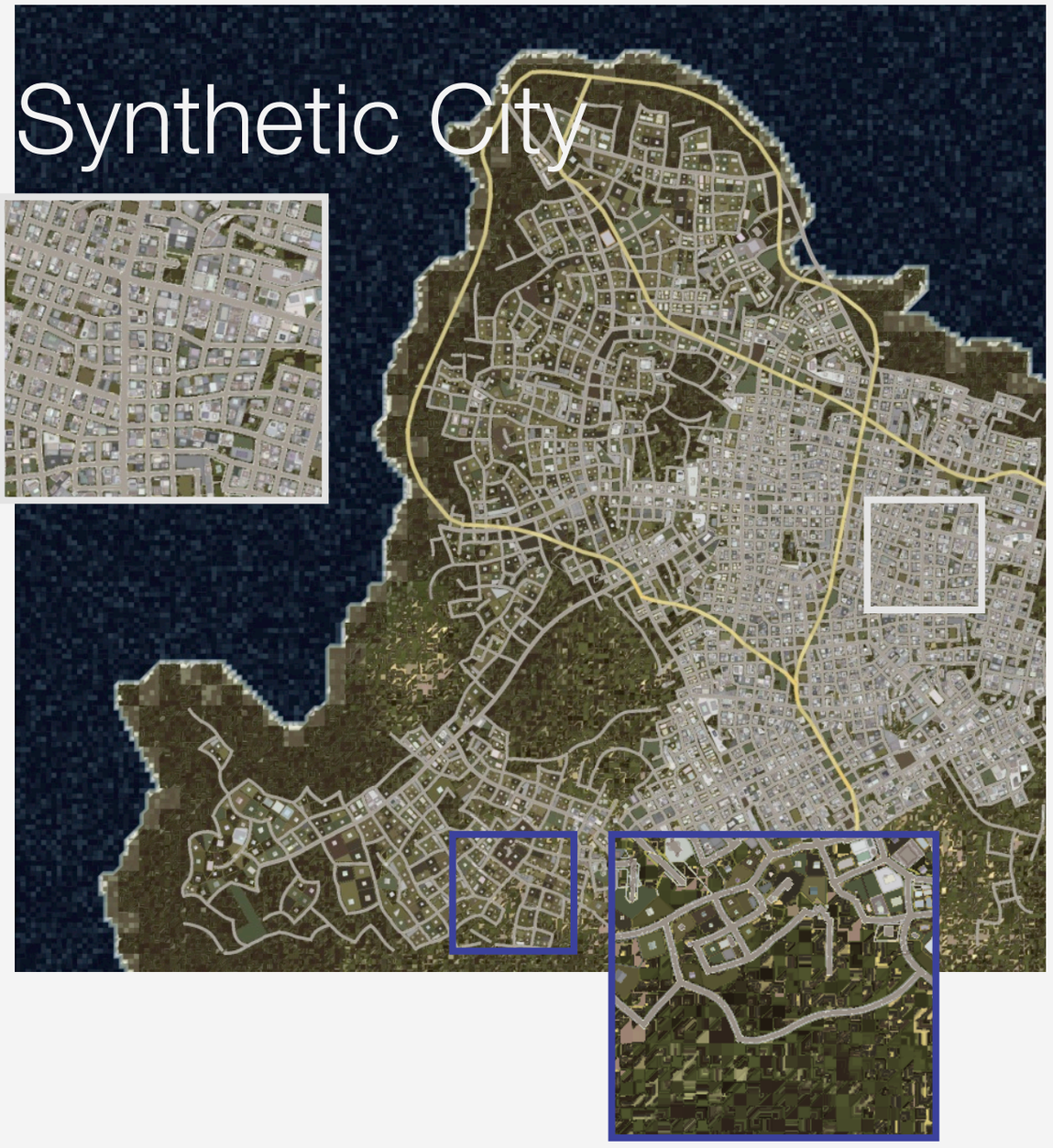
Synthesizing Detail with Procedural Geometric Models

We created Monterey Peninsula streets, parcels, buildings in 3D using only 5 minutes of user input of the highways, coastline, undevelopable areas, center of jobs, total population and jobs.

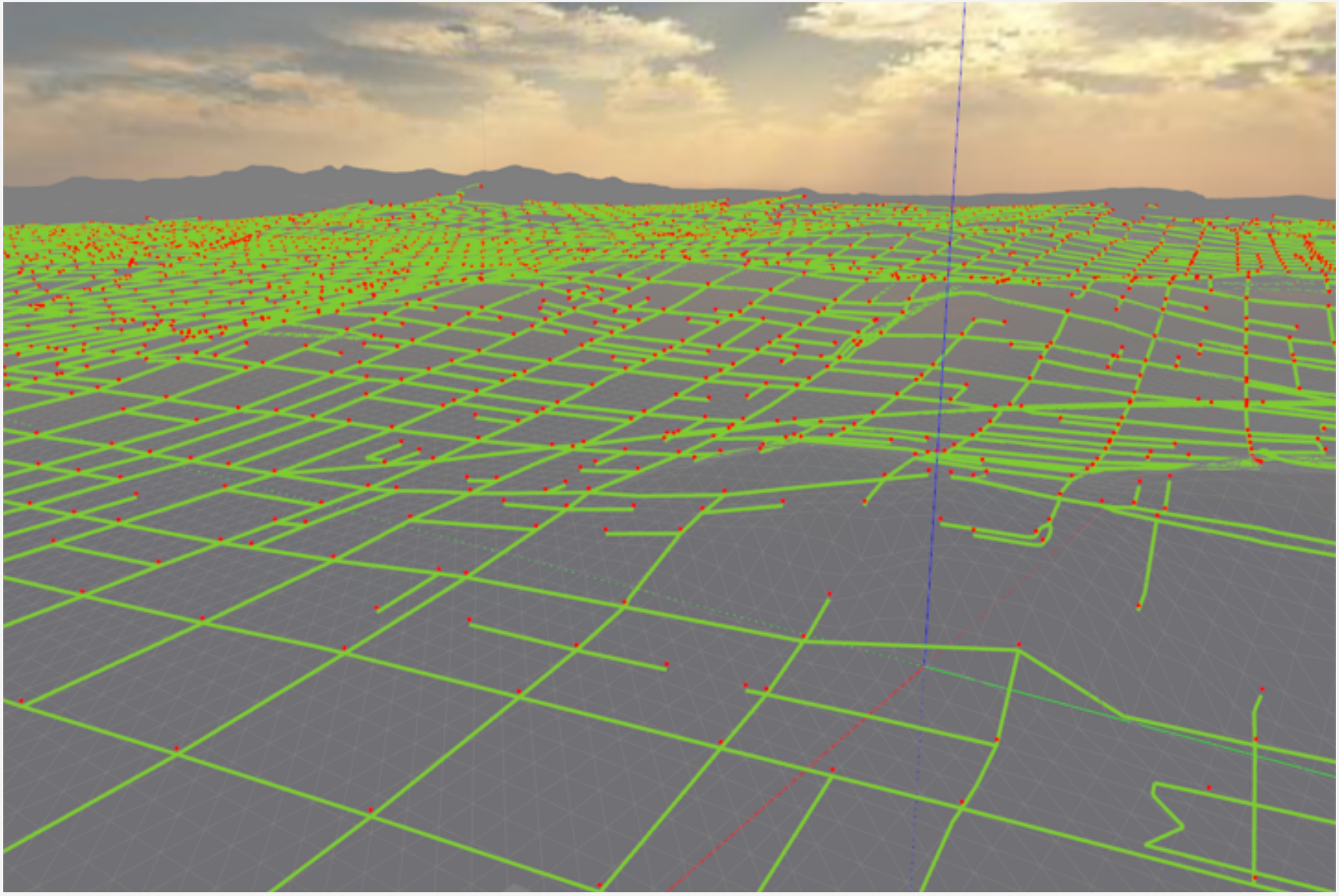
Synthesizing Detail with Procedural Geometric Models

We created Monterey Peninsula streets, parcels, buildings in 3D using only 5 minutes of user input of the highways, coastline, undevelopable areas, center of jobs, total population and jobs.

Synthesizing Detail with Procedural Geometric Models

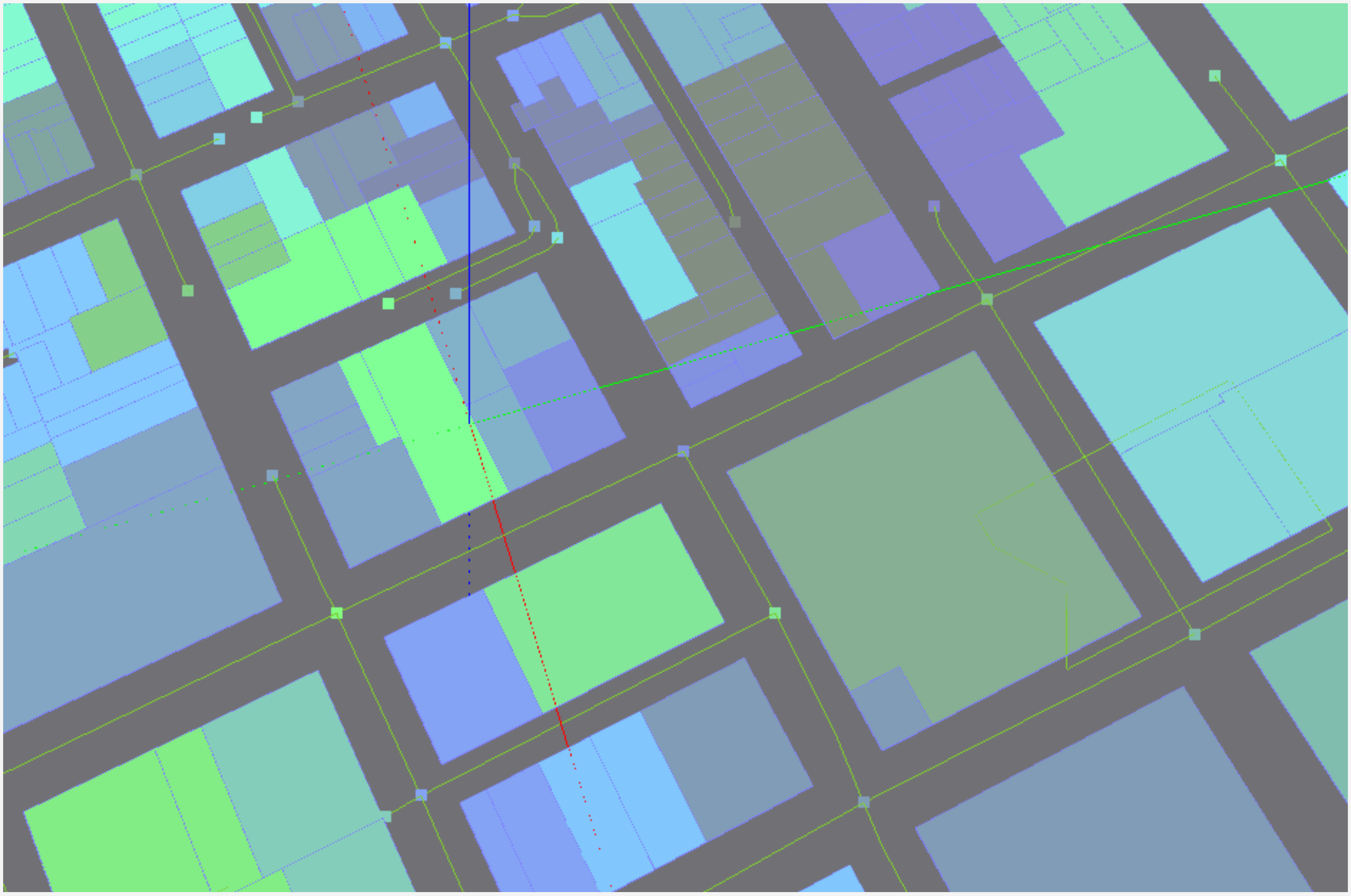


We created Monterey Peninsula streets, parcels, buildings in 3D using only 5 minutes of user input of the highways, coastline, undevelopable areas, center of jobs, total population and jobs.



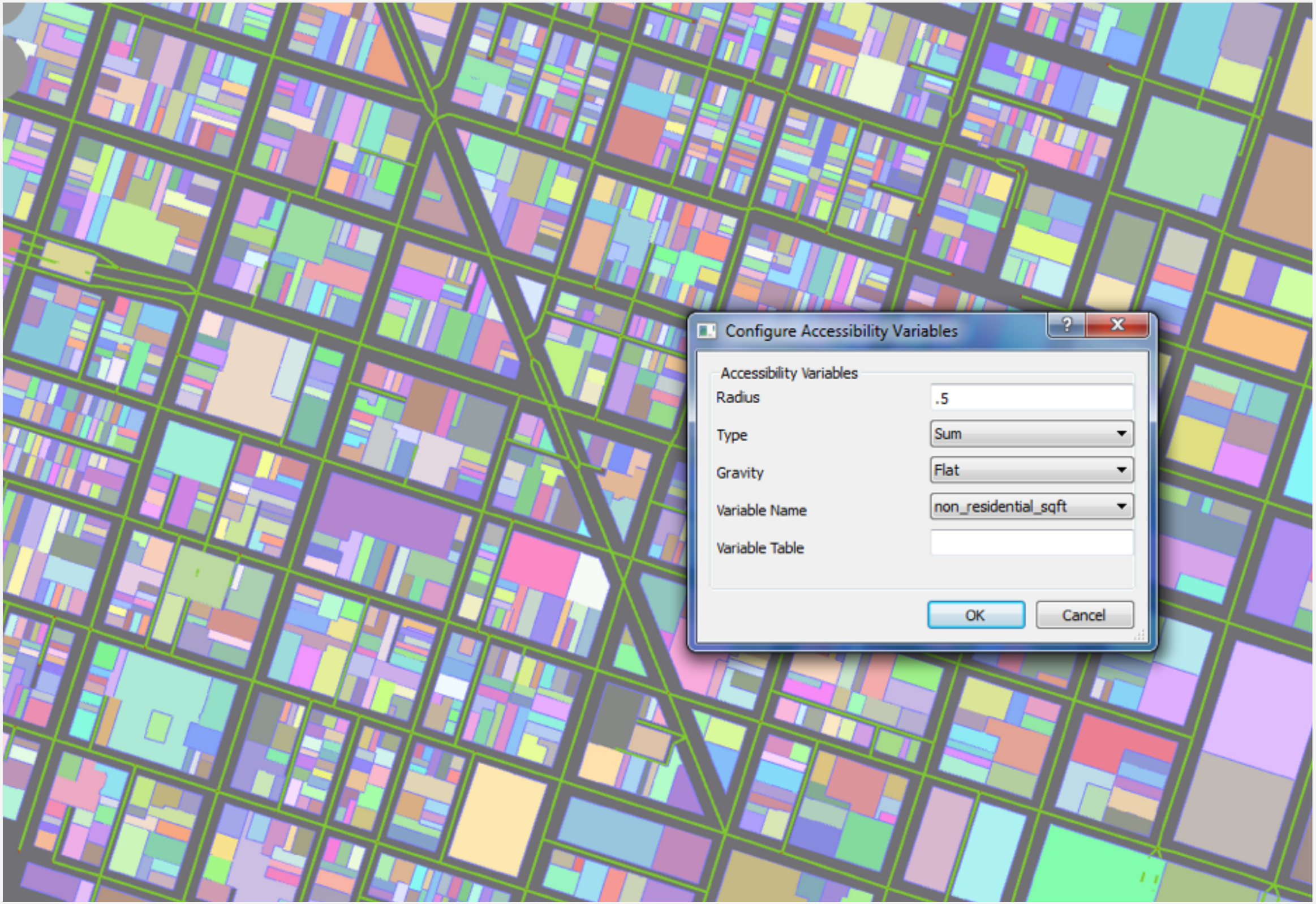
Synthesizing Constrained with More Real Data: Loading Local Streets and Terrain

UrbanVision



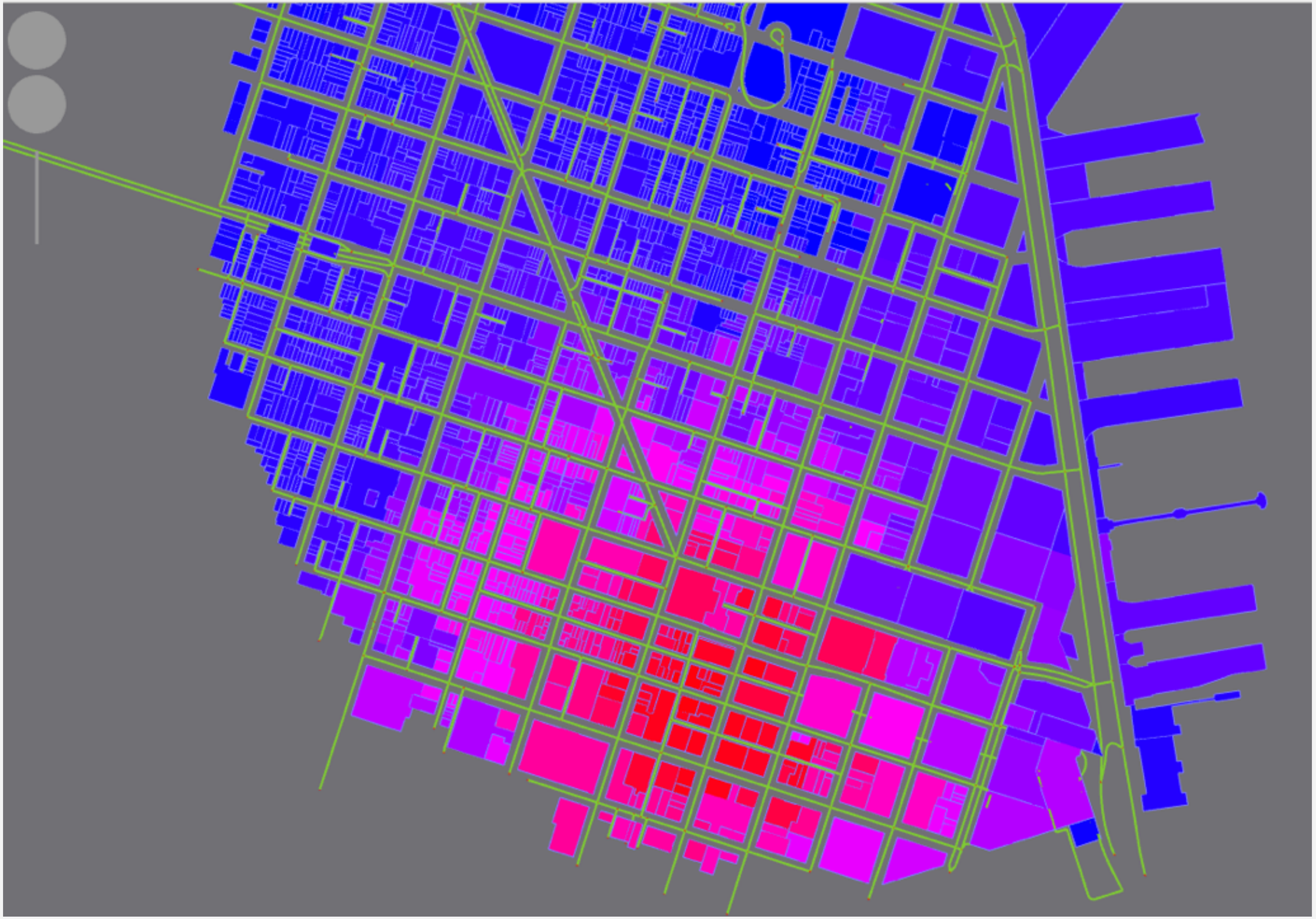
Calculating Nearest Local Street Node for Every Parcel

UrbanVision



Flexible walking - scale accessibility calculator using local streets, buildings, parcels

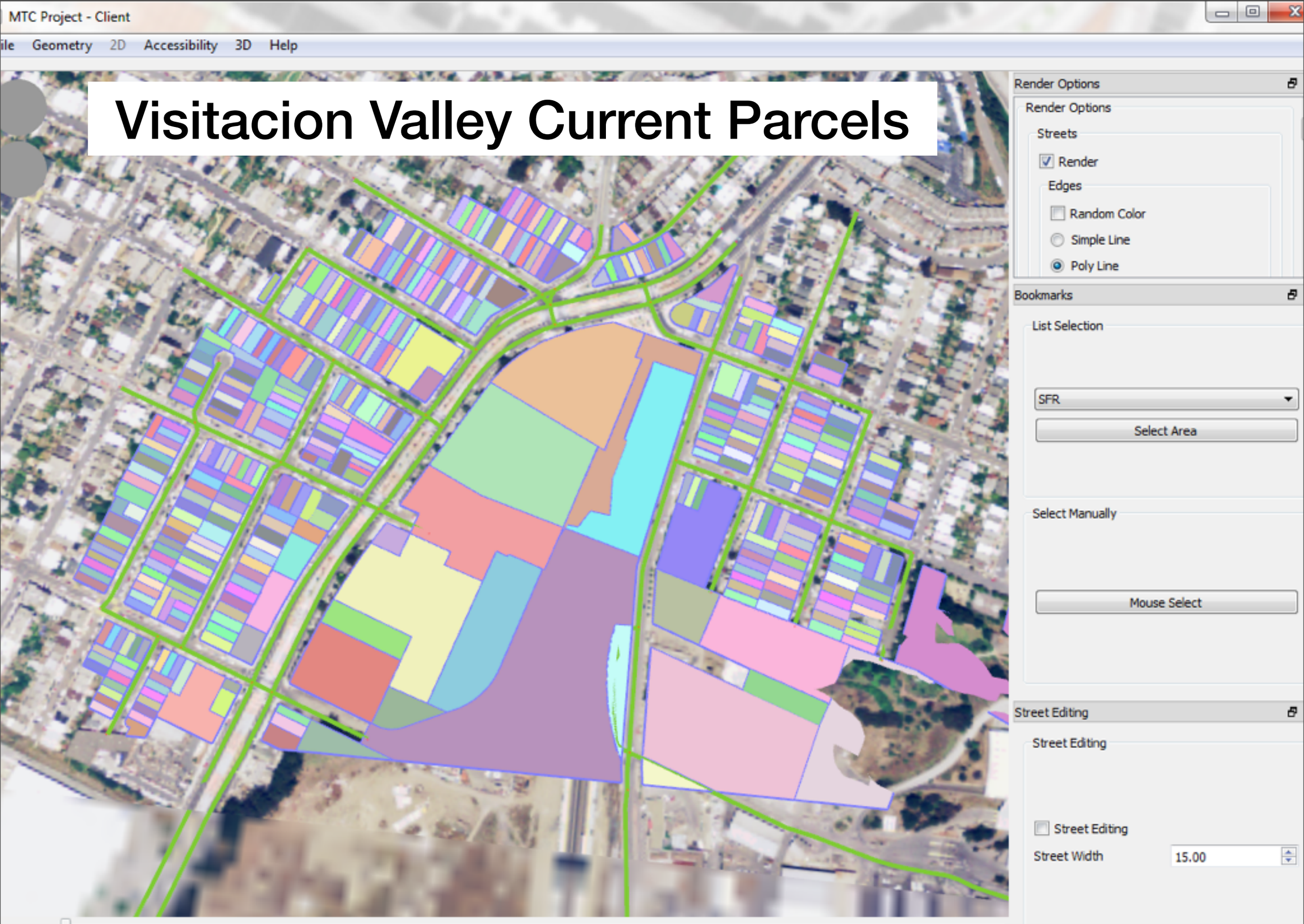
UrbanVision



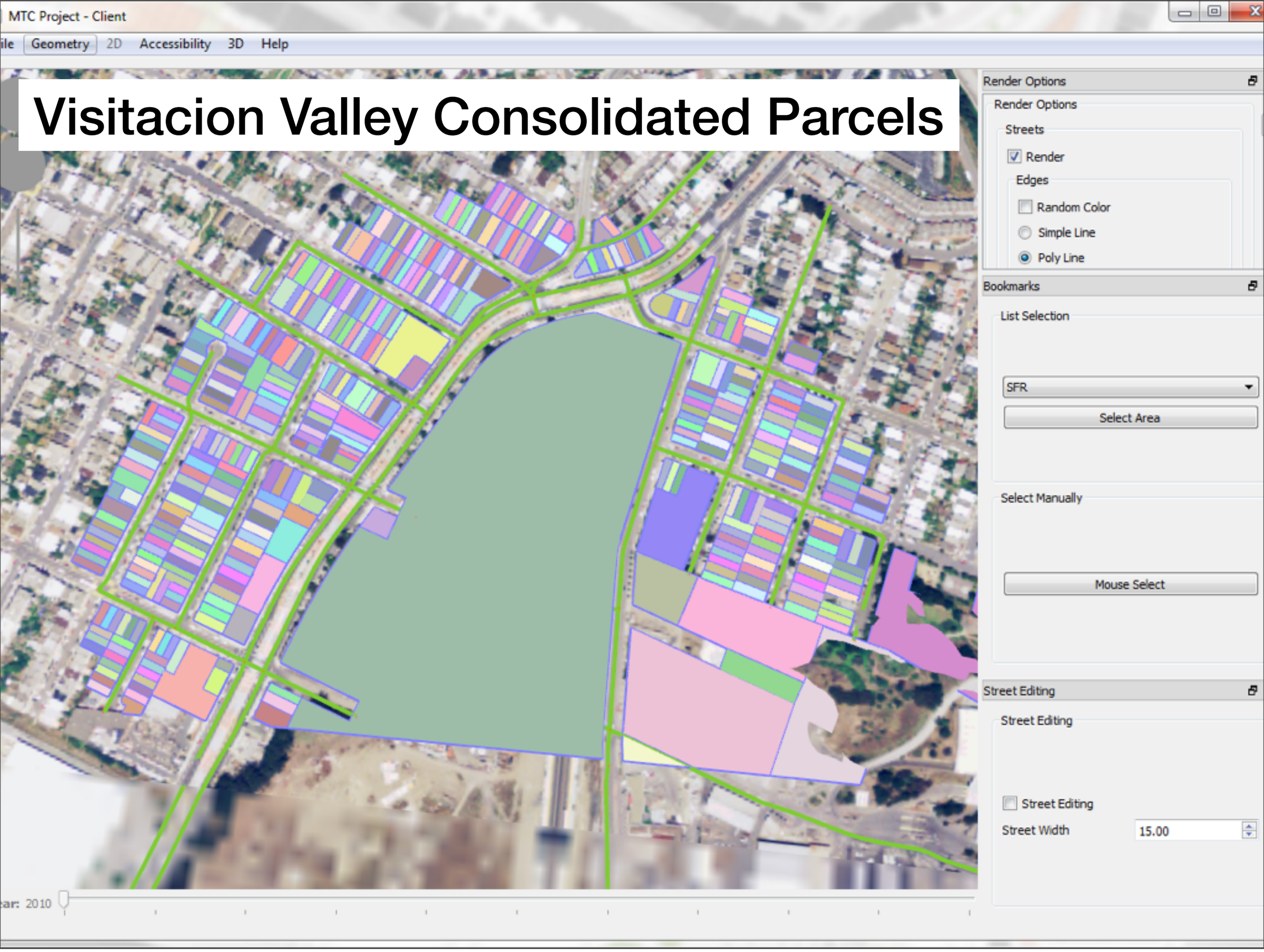
Heat map of non-residential building square footage within 0.5 kilometers on local streets

UrbanVision

Visitacion Valley Current Parcels



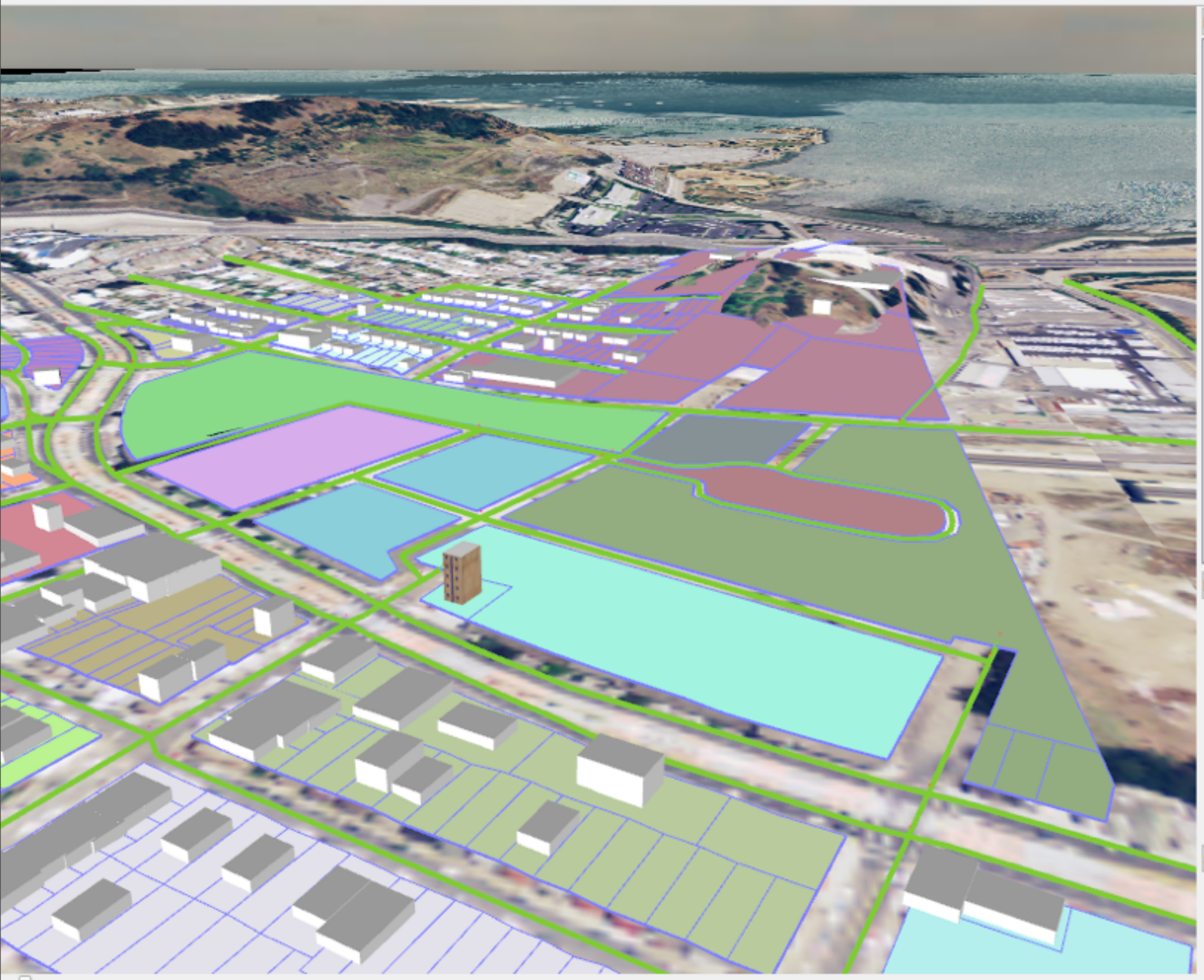
Visitacion Valley Consolidated Parcels



Sunday, May 8, 2011

roject - Client

eometry 2D Accessibility 3D Help



Render Options

Accessibility

Ortho

Aerial Image

Opacity

Terrain

View Terrain

Adapt geometry

Bookmarks

List Selection

SFR

Select Area

Select Manually

Mouse Select

Street Editing

Street Editing

Street Editing

Street Width 15.00

0

Sunday, May 8, 2011

MTC Project - Client

Geometry 2D Accessibility 3D Help

Landmarks + Procedurally Generated Buildings

Questions and Discussion

UrbanSim Links:

http://www.urbansim.org

Contact Information:

Paul Waddell
Department of City and Regional Planning
University of California Berkeley
email: waddell@berleley.edu