

---

# Status Report on UrbanSim and the Open Platform for Urban Simulation

---

**Paul Waddell**

Department of City and Regional Planning

UC Berkeley

[www.urbansim.org](http://www.urbansim.org)

May 17, 2010

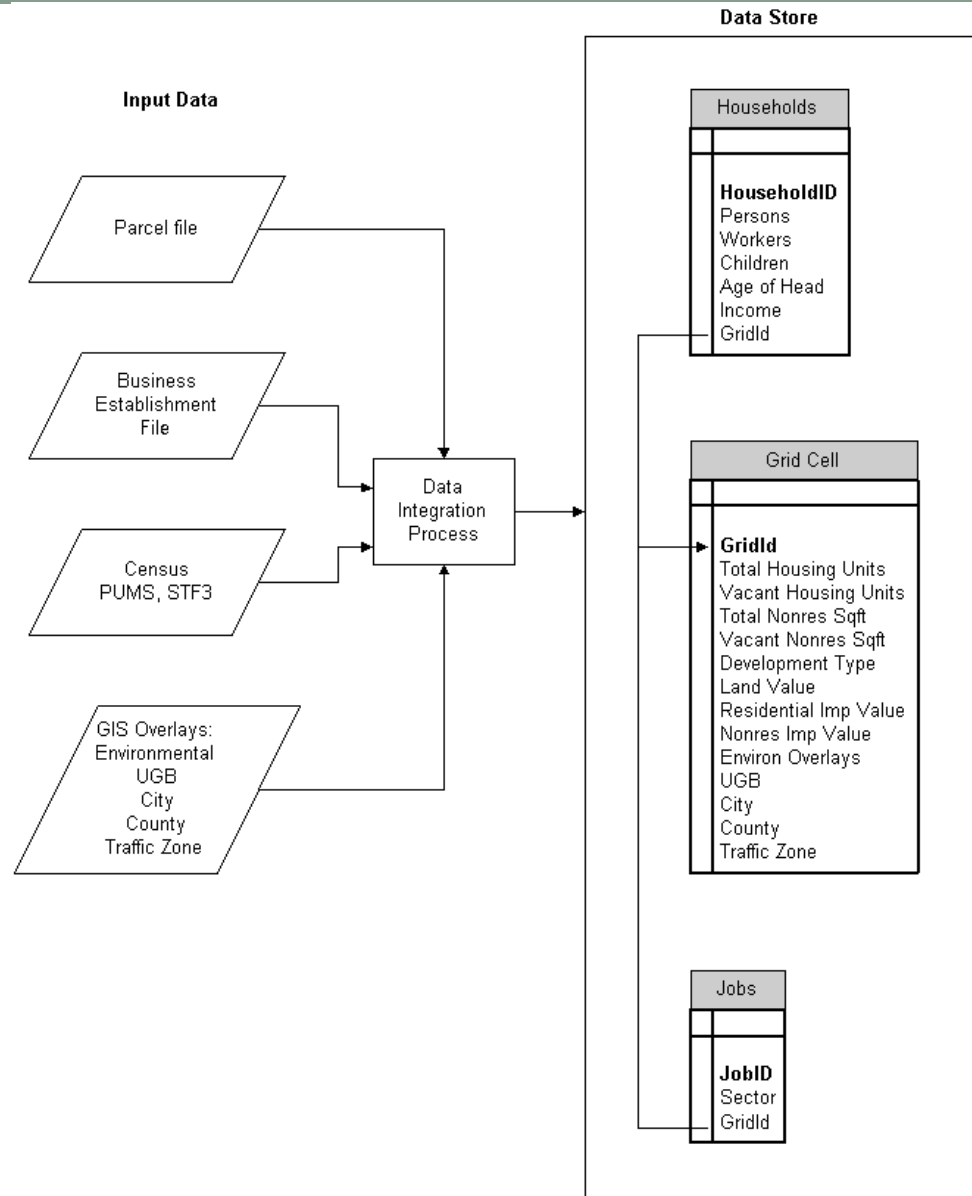
# Agenda

- Brief Updates on UrbanSim Development
  - Data Integration
  - Model Uncertainty
  - Zone version of UrbanSim
  - Interactive Database and Scenario Creation
- Integrated Model Application Case Study:  
Puget Sound

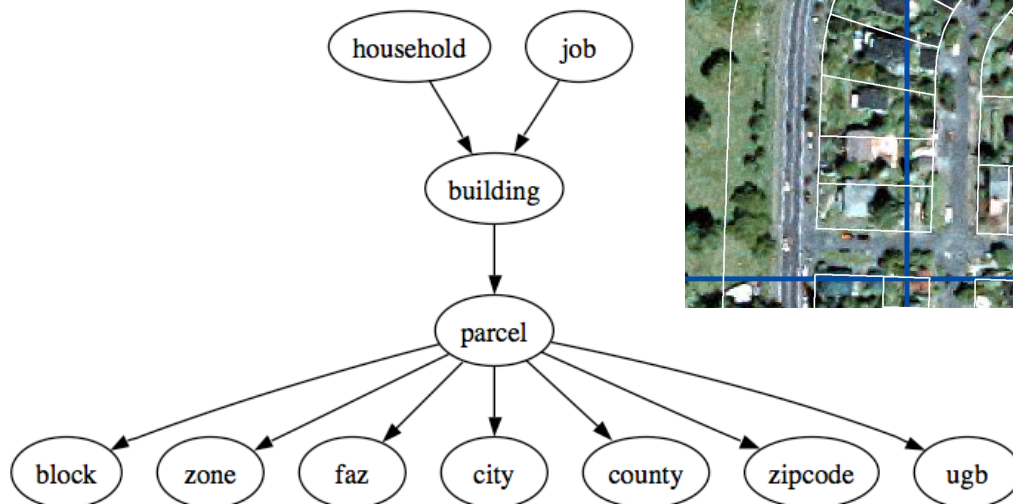
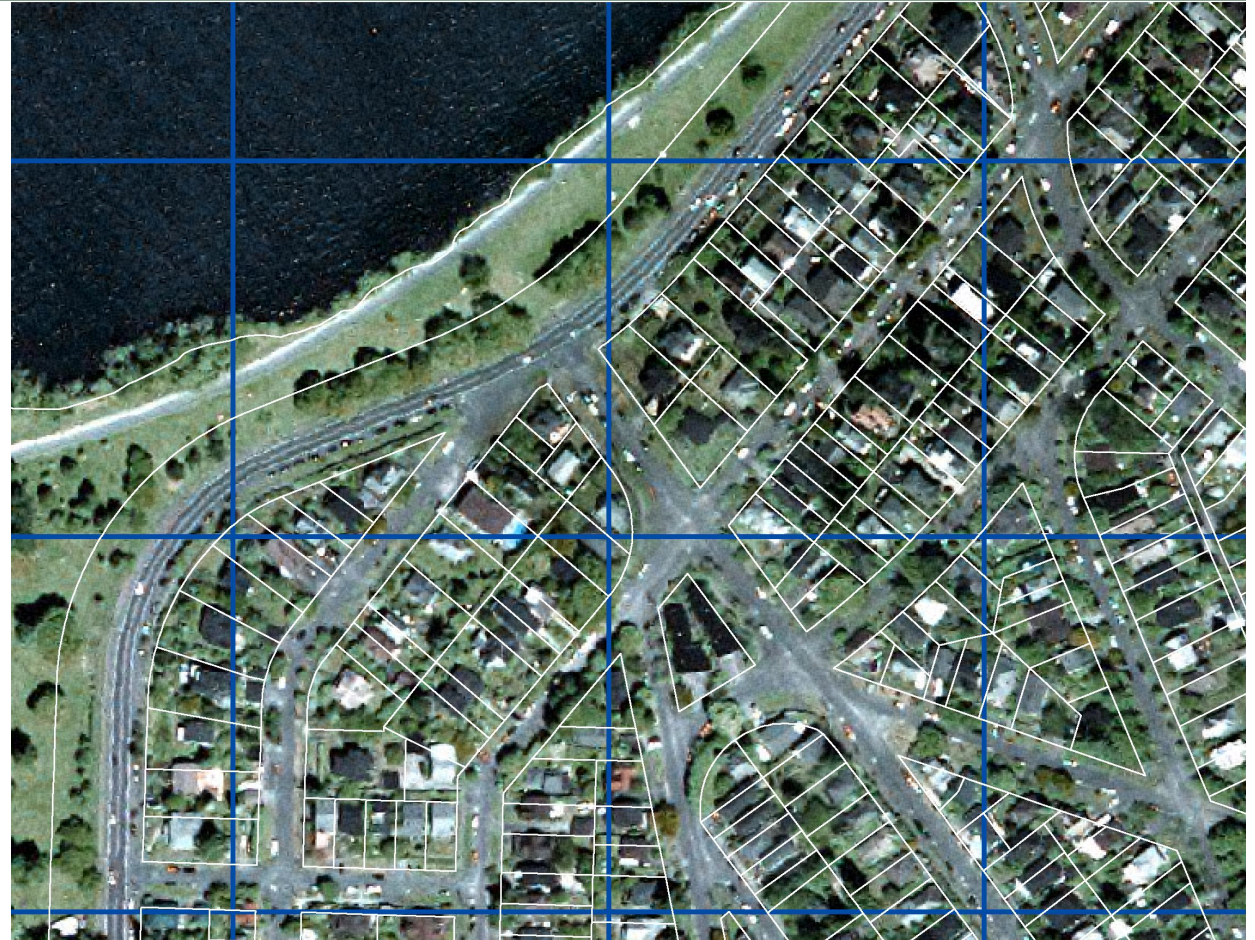


# Data Integration in UrbanSim

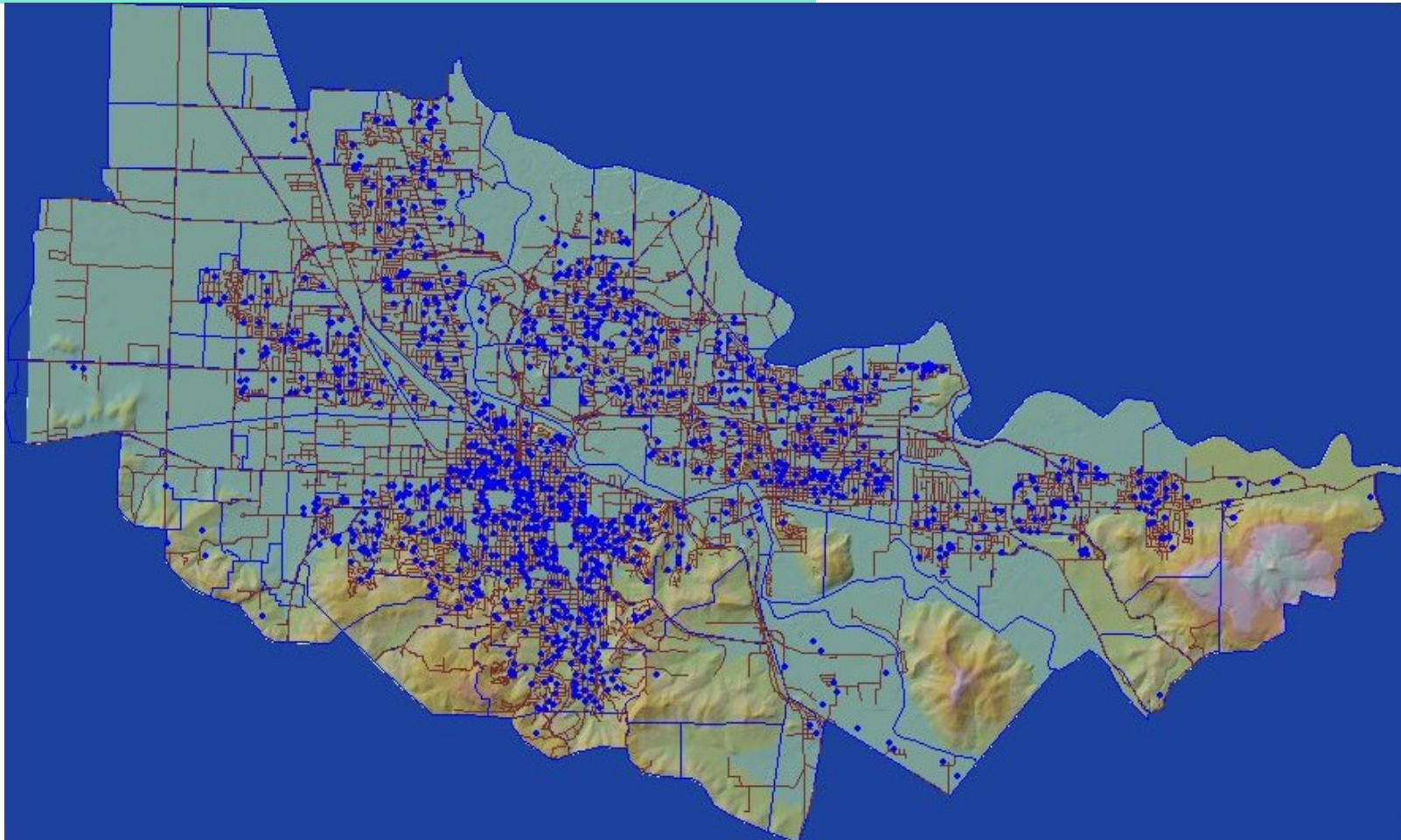
# UrbanSim Data Integration Process



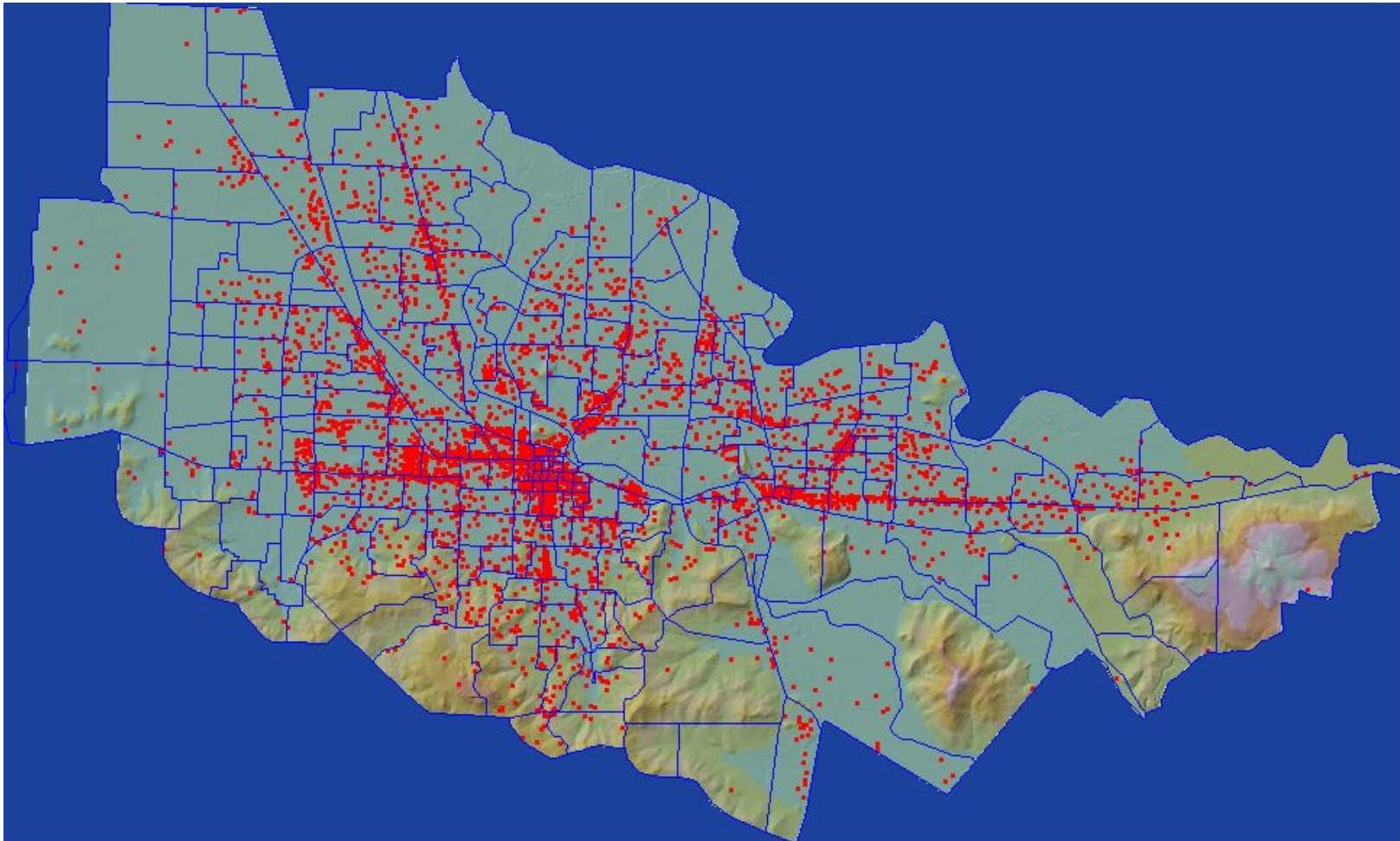
# New Model System Based on Parcels and Buildings



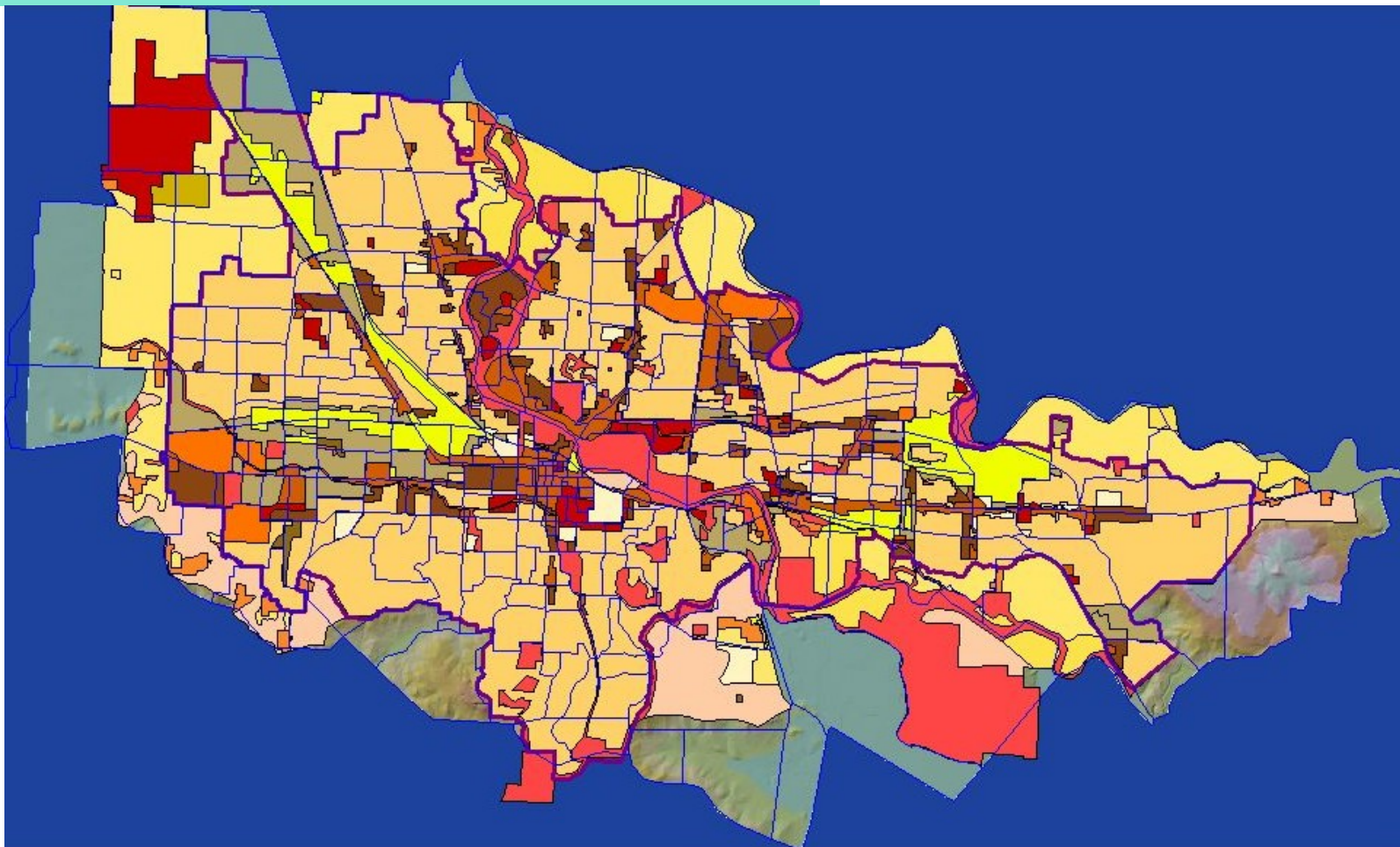
# Input Data: Household Survey



# Input Data: Employment

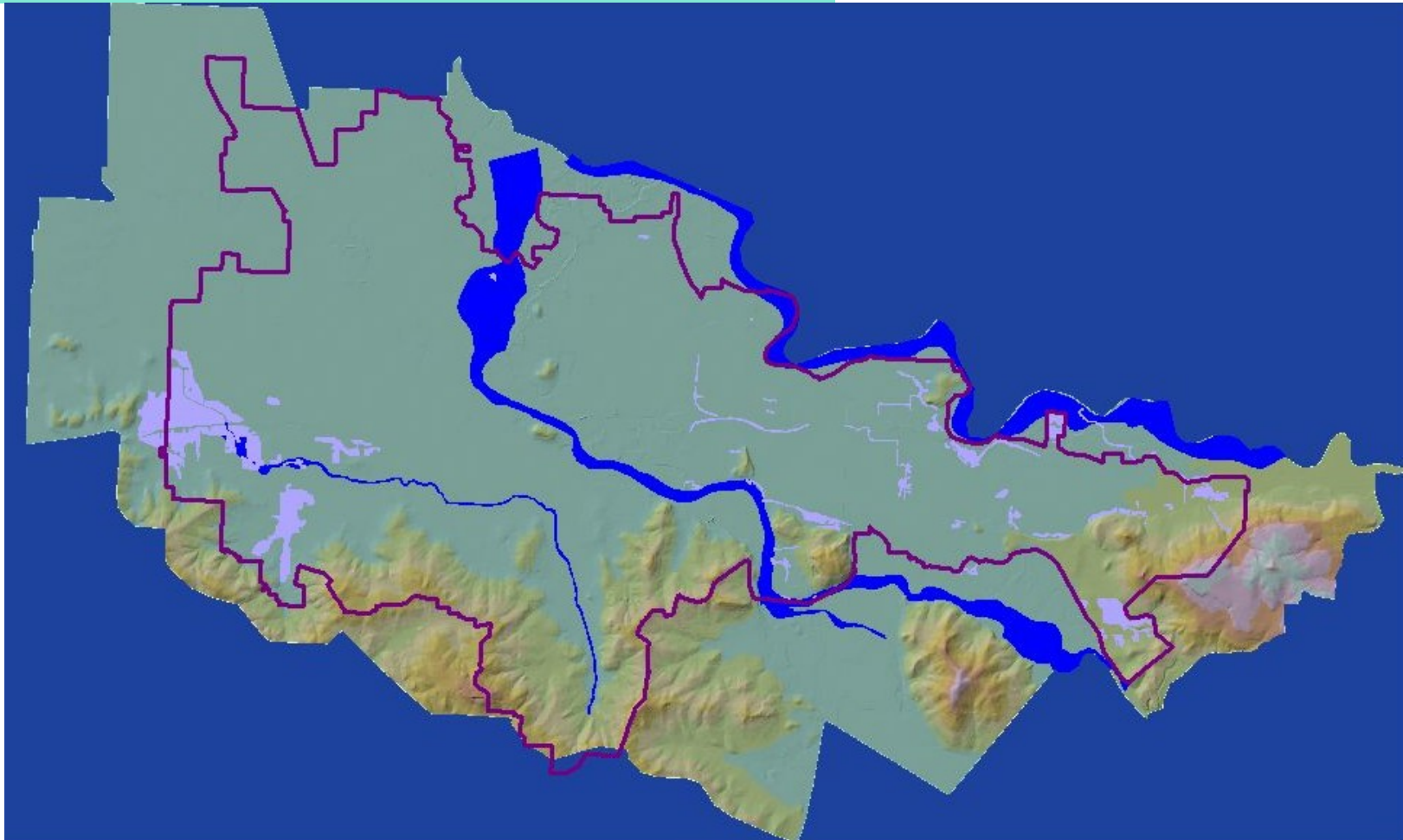


# Input Data: Land Use Plan





# Input Data: UGB and Environmental



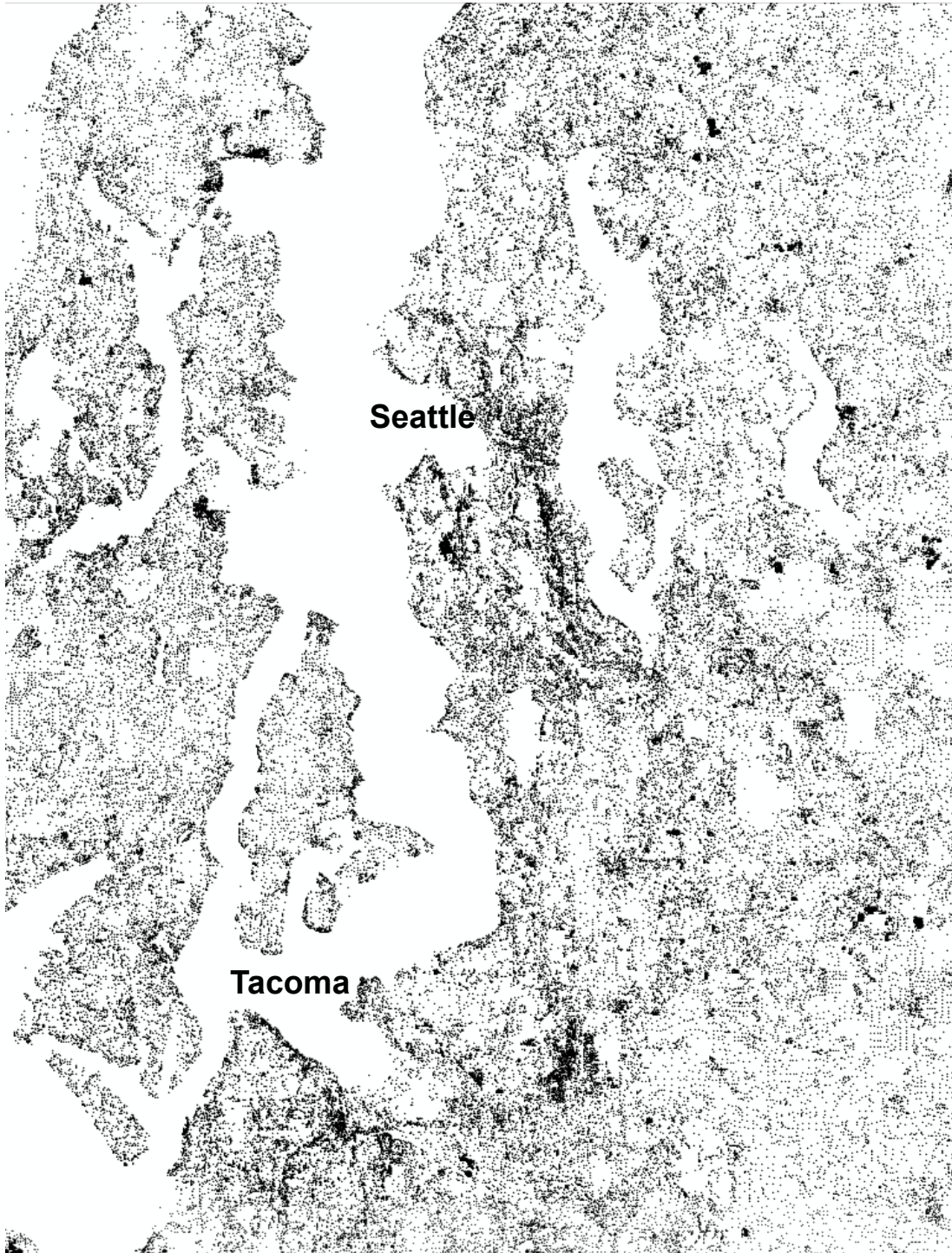
# Data Integration Challenges

- **Messy Data**
  - Many outliers, errors, and missing data
  - Inconsistent coding schemes among data sources
- **Difficult to integrate with other data sources**
  - Building-level data
  - Business establishment data
  - Market information (vacancies, prices, rents)
- **Volume of data too massive to manually correct**
  - 2+ million parcels in Bay Area
- **Problems hard to diagnose**
  - Which data is wrong? (which attributes/sources are incorrect? May have systematic patterns of omissions – e.g. tax-exempt properties)
  - Misgeocoding: some businesses are geocoded to the wrong place. Complicates the diagnosis.

## The Magnitude of the Problem

This map shows only buildings with missing values for “Building Type ID”, a description variable.

195,501 out of ~1,200,000  
Building Type ID = Null  
King, Kitsap, Pierce & Snohomish Co.

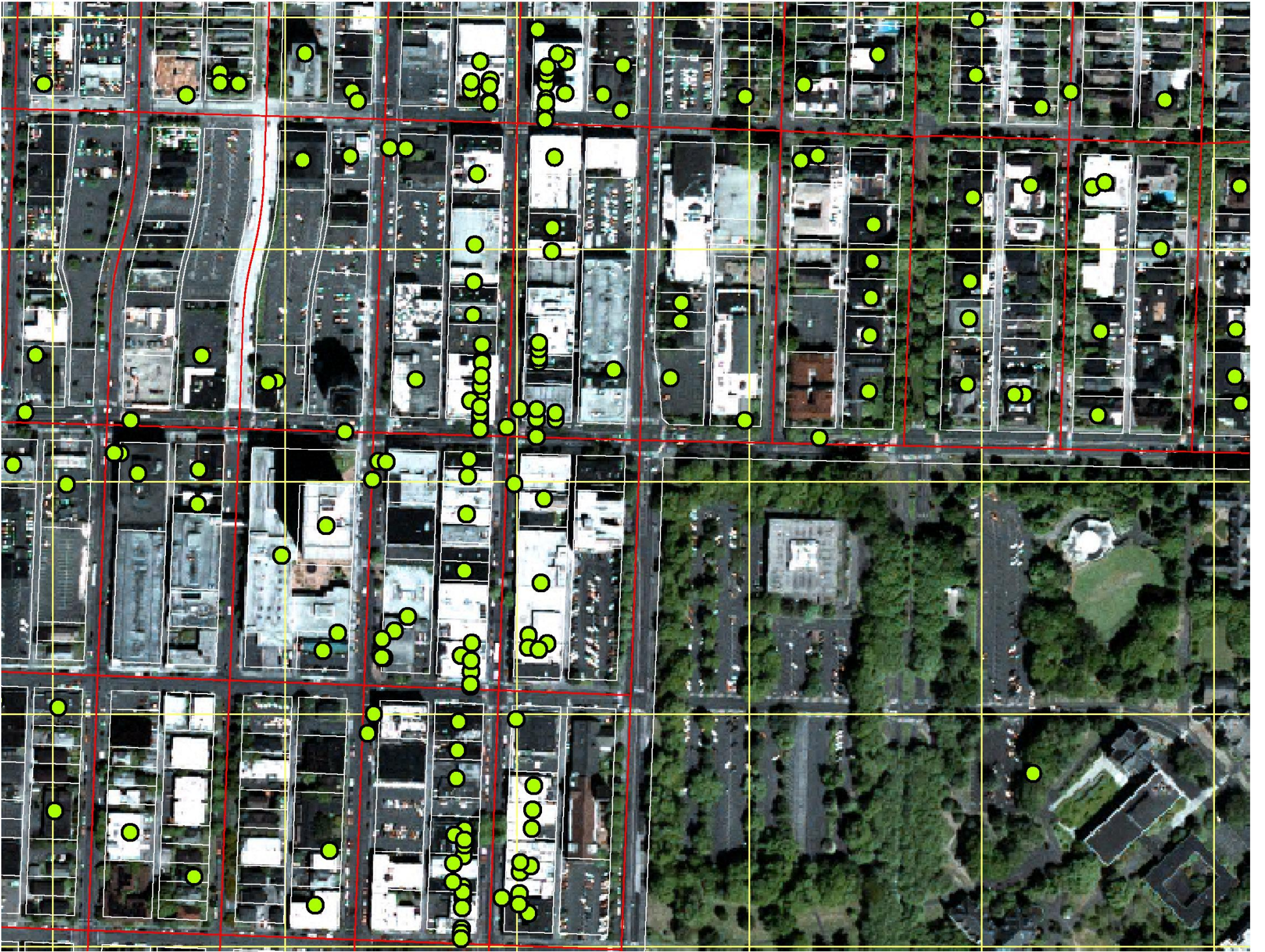


# Data Imputation Tool



Data Imputation Tool





# Ways Forward on Data Integration

## **Option 1:**

- Machine Learning/Data Mining
- Model patterns in the observed data
- Use the models to detect outliers, impute data
- Preliminary work on this now implemented using WEKA library

## **Option 2:**

- In some cases, the missingness level is very high
- Developing countries (e.g. Ghana and South Africa)
- Potential to Synthesize much of the data, subject to constraints, using procedural modeling

## **Option 3:**

- Potential to hybridize statistical/machine learning and procedural modeling, to synthesize from disparate sources?

# Data Imputation

## K-Nearest Neighbors for Continuous Attributes

Attributes: Stories, Bldg SF,  
Improvement Value, etc.

KNN Basics:

Finds k closest neighbors in n  
dimensional space.

Uses k neighbors target values  
to make prediction.

$$L^n(\mathbf{x}_1, \mathbf{x}_2) =$$

$$\sqrt[n]{\sum_{i=1}^{\#dim} |\mathbf{x}_{1,i} - \mathbf{x}_{2,i}|^n}$$

$$\hat{f}(x_q) \leftarrow \frac{1}{k} \sum_{i=1}^k f(x_i)$$



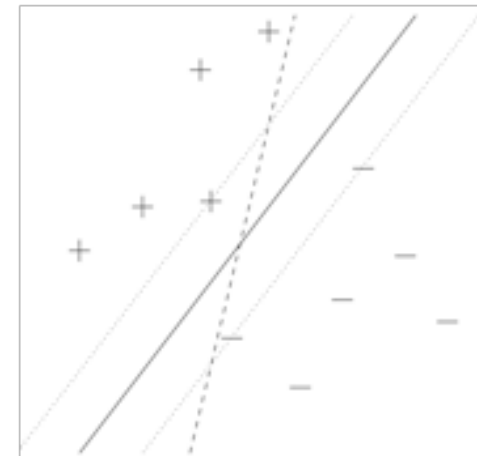
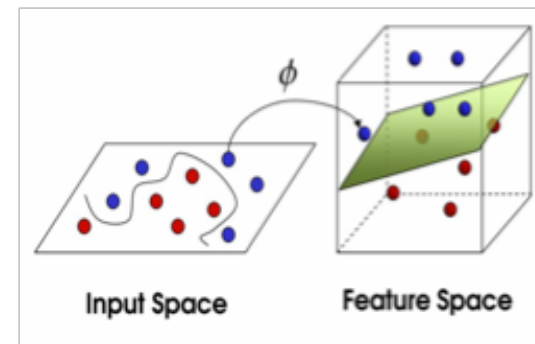
# Data Imputation

## Support Vector Machines for Categorical Attributes

Attributes: Building Use Code,  
Land Use Code, etc.

SVM maps training instances  
into higher dimensional  
space.

Creates hyper planes that have  
maximum distances from  
instances as category  
boundaries.



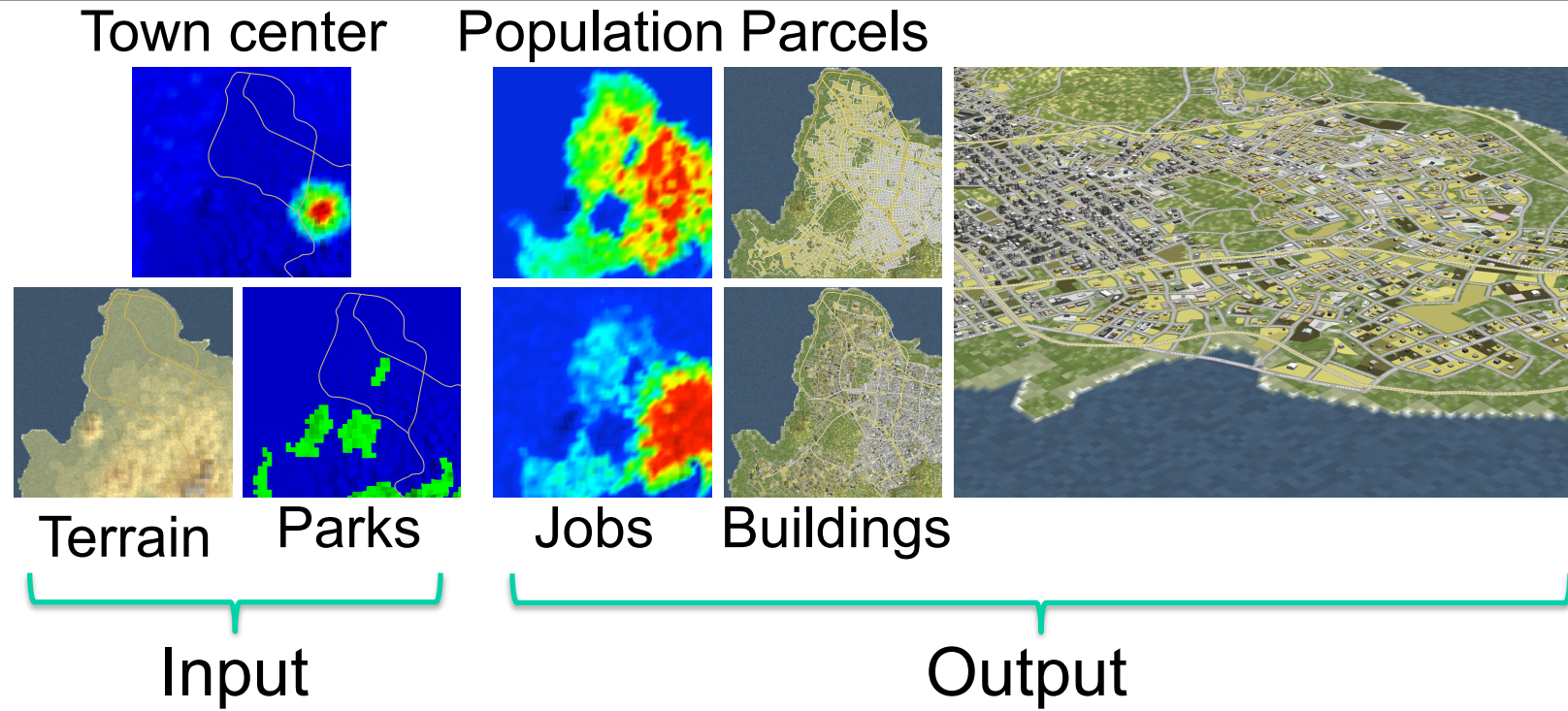
# Machine Learning/Data Mining

So far, only applied to single tables, single output

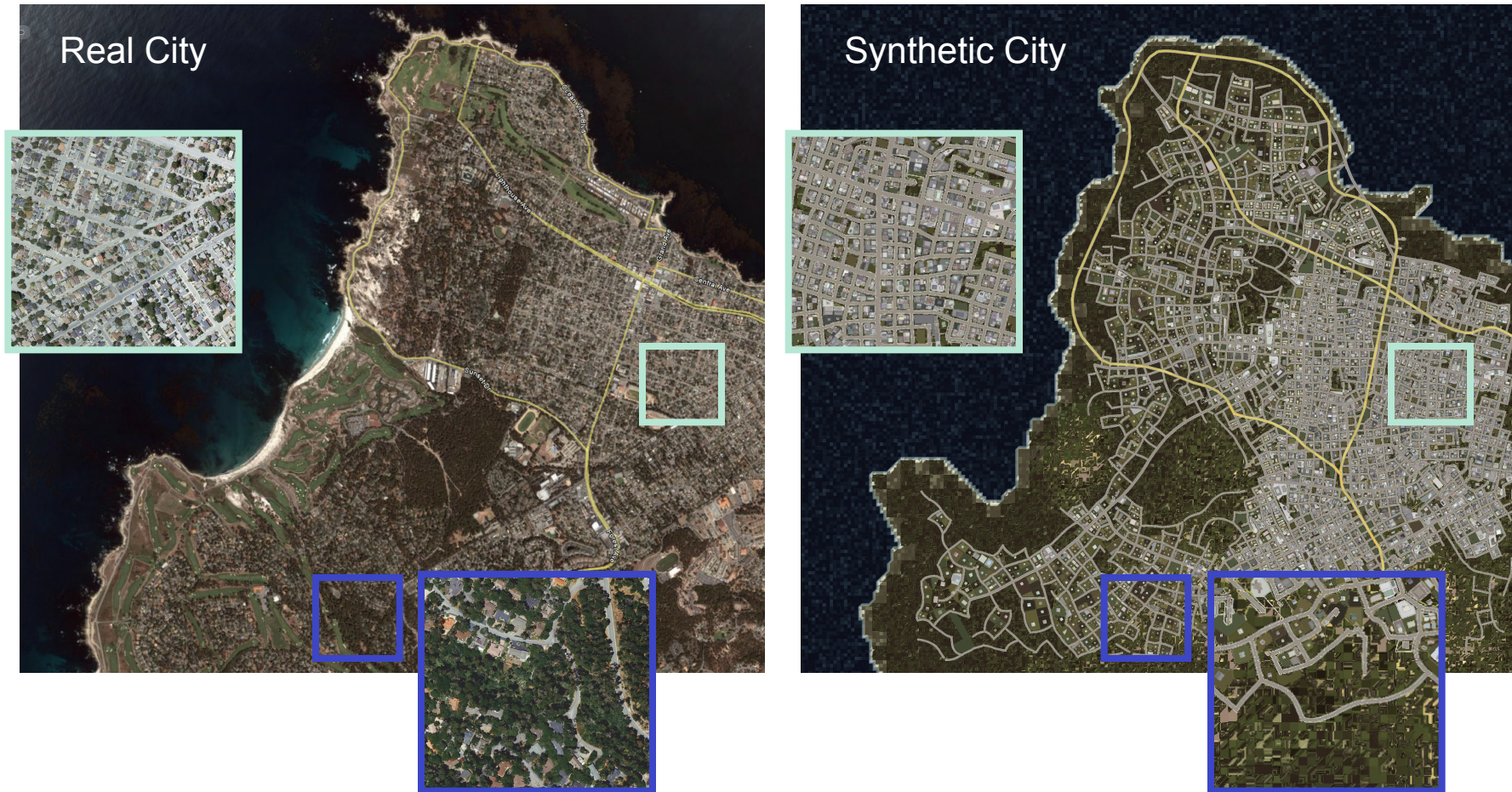
Need to develop analysis for:

- multivariate outcomes,
- across tables,
- some of which have poorly-defined (spatial) cross-references
- and mixtures of continuous, categorical and ordered outcomes

# Option 2: Procedural Modeling



# Results: Completion and Validation



Vanegas, et al, 2009



# Model Uncertainty

# Assessing Uncertainty: Bayesian Melding

- Developed rigorous methodology for assessment of uncertainty in integrated land use and transport models based on Bayesian Melding (published in Transportation Research A, 2007)
- Currently testing an application to the question: what would happen if the Alaskan Way Viaduct adjacent to the waterfront in the Seattle CBD were demolished? It is at risk of collapse in the next earthquake.

# Alaskan Way Viaduct Scheduled for Demolition in...

The next earthquake?

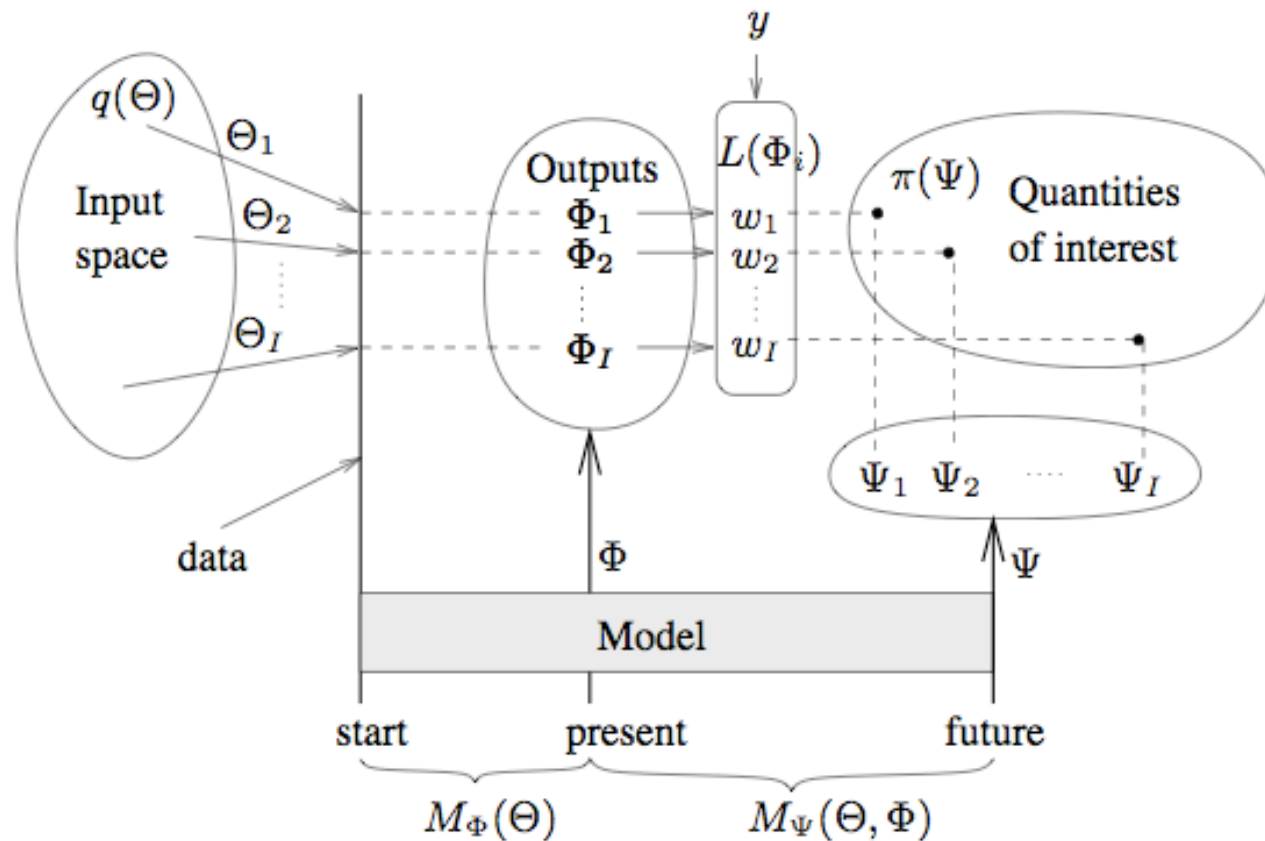
Some claim that alternatives which do not have comparable Traffic capacity will Cause massive failure of traffic in CBD and on I5.



Others claim that we should replace it with surface street and transit, and reclaim the waterfront. It won't cause much traffic impact because people adapt.

How much would a low-capacity alternative affect travel times over 10 years?

# Assessing Uncertainty with Bayesian Melding





# Assessing Uncertainty with Bayesian Melding

## Likelihood and posterior distribution

$y_k$  is sqrt of observed quantity in zone  $k$

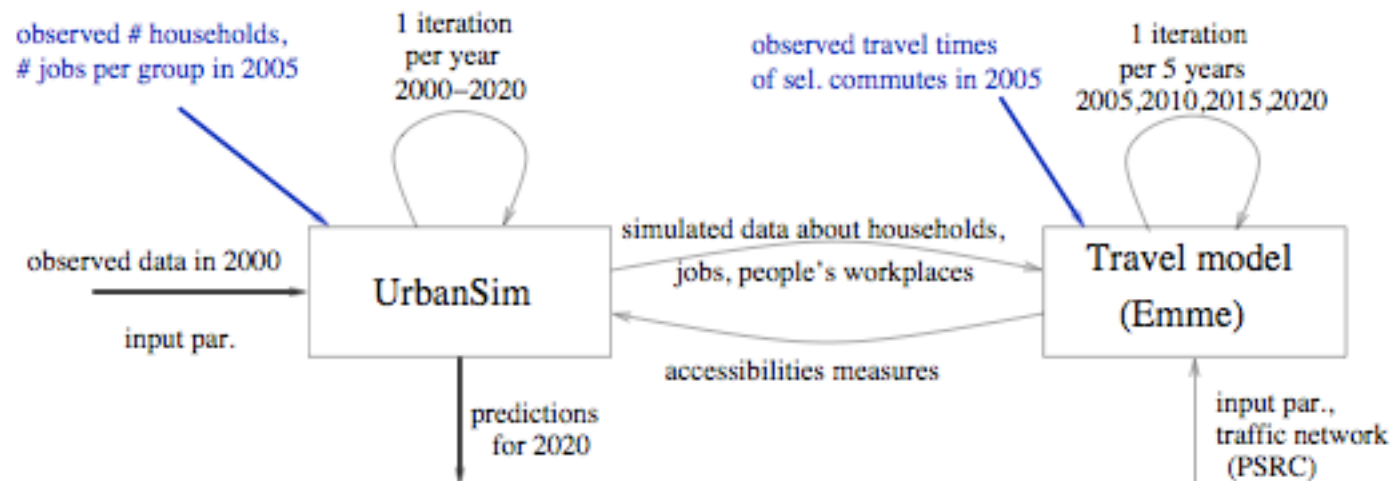
$$y_k | \Theta_i \sim N(\hat{a} + \hat{\mu}_{ik}, \hat{\sigma}_i^2)$$

$$w_i \propto p(y | \Theta_i) = \prod_{k=1}^K \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} \exp \left[ -\frac{1/2(y_k - \hat{a} - \hat{\mu}_{ik})^2}{\hat{\sigma}_i^2} \right]$$

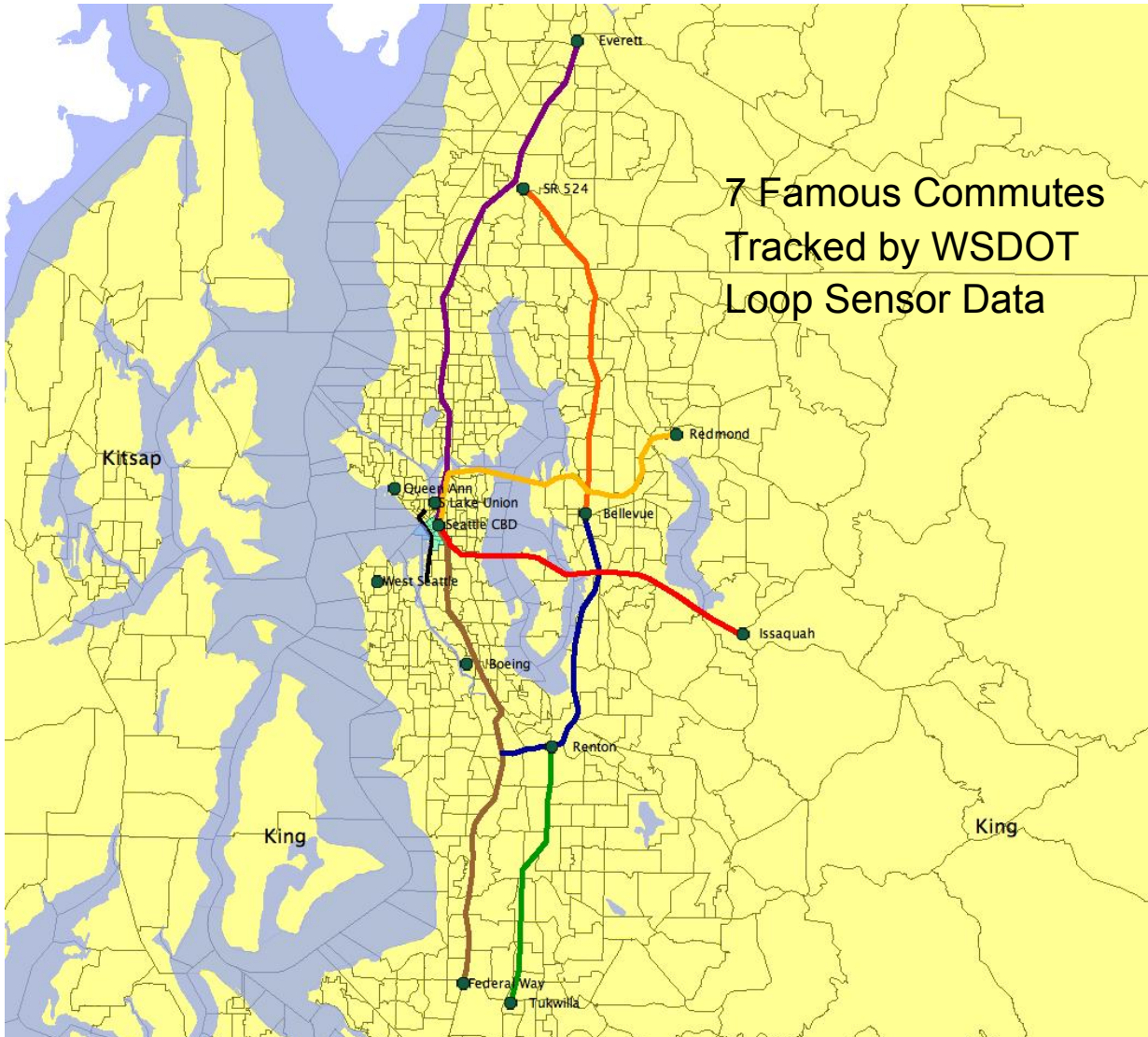
$$p(\Psi_k) = \sum_{i=1}^I w_i N(\hat{a}b_a + \Psi_{ik}, \hat{\sigma}_i^2), \quad k = 1, \dots, K$$

# Assessing Uncertainty with Bayesian Melding

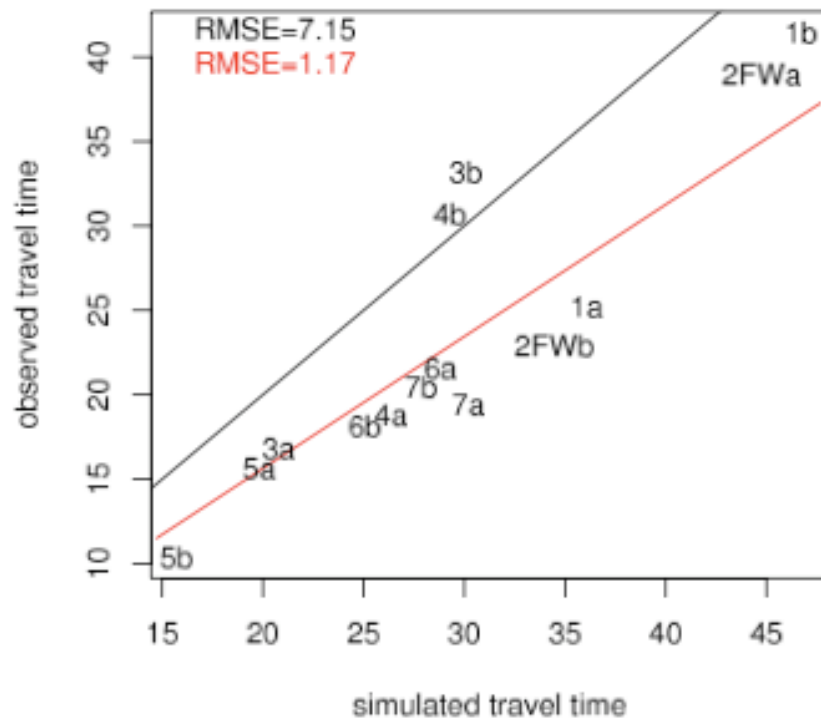
Quantity of interest: Travel times on selected routes



# Assessing Uncertainty with Bayesian Melding



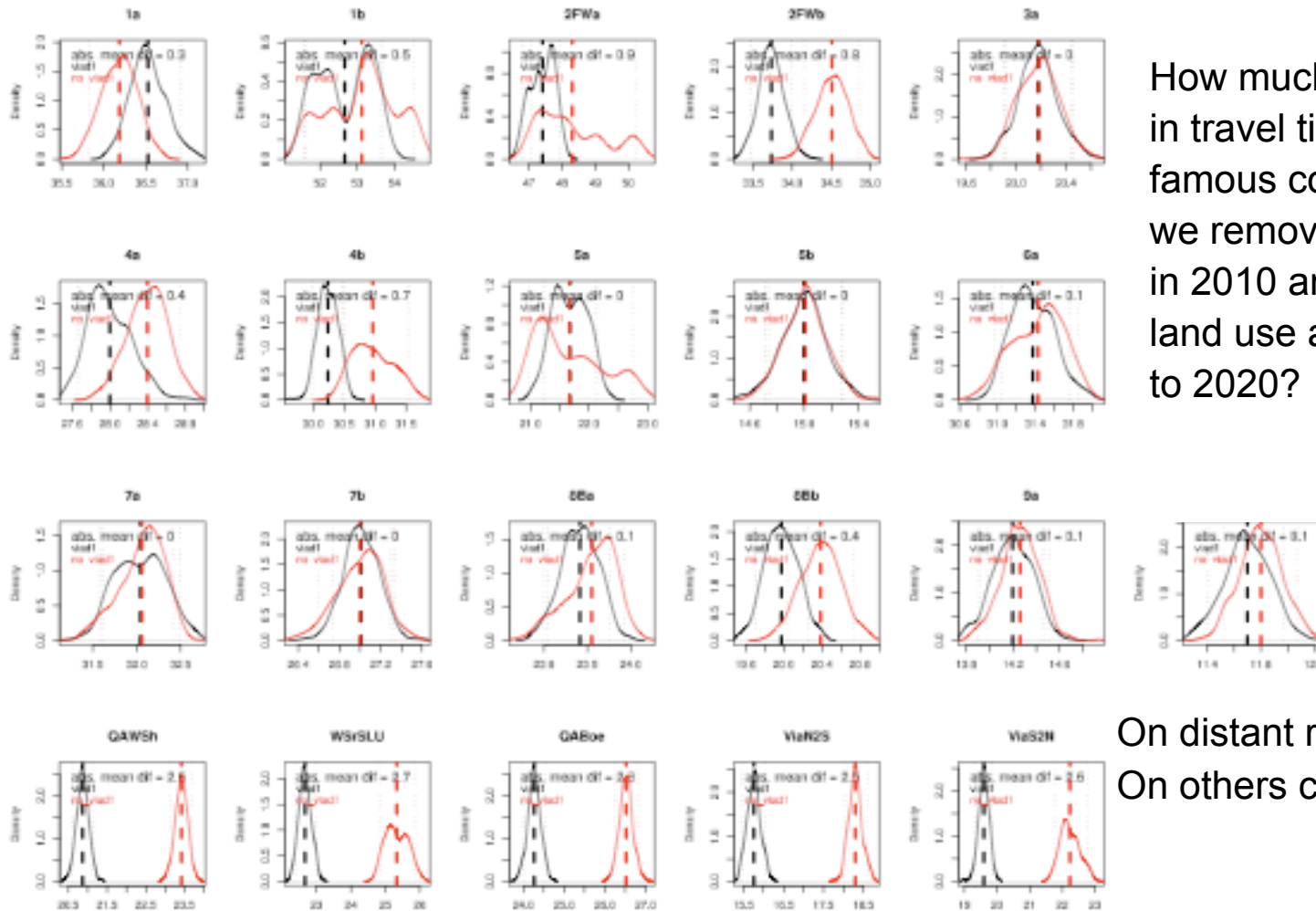
# Assessing Uncertainty with Bayesian Melding



Systematic bias in travel times predicted by travel model was corrected

$$\log(T) \sim N(\log(T_{sim}) - 0.25, 0.16^2)$$

# Assessing Uncertainty with Bayesian Melding



How much difference in travel time on those famous commutes if we remove the Viaduct in 2010 and simulate land use and transport to 2020?

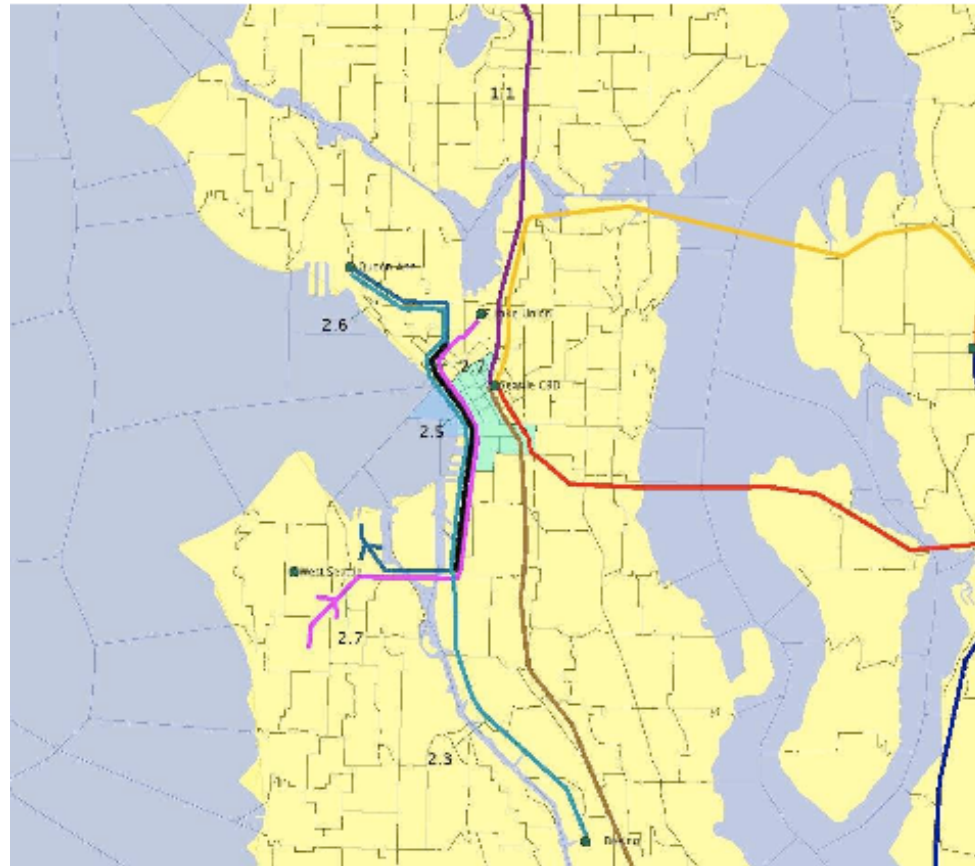
On distant routes < 1 minute,  
On others closer to 2-3 minutes

# Assessing Uncertainty with Bayesian Melding

route	scen.	2000	2010	inc.	2020	inc.
QAWSh (7.4mi)	viad	21.1	21.3		21.1	
	no v.		23.4	10%	23.7	12%
WSrSLU (7.4mi)	viad	20.9	22.1		22.8	
	no v.		24.7	12%	25.5	12%
QABoe (9.9mi)	viad	24.5	24.6		24.5	
	no v.		26.7	9%	26.8	9%
ViaN2S (5.2mi)	viad	16.0	16.1		16.0	
	no v.		18.4	14%	18.5	16%
ViaS2N (5.2mi)	viad	18.1	19.1		19.8	
	no v.		21.5	13%	22.4	13%

On shorter routes close to Viaduct, these translate to 9 – 16% increases in travel time

# Assessing Uncertainty with Bayesian Melding





# Zone Version of UrbanSim



# Zone Model System Documentation On-line

The screenshot shows the UrbanSim website interface. At the top, there is a blue header with the UrbanSim logo in green. Below the logo is a navigation menu with buttons for Home, Documentation, Download, Community, Research, and Forums. A search box with the text 'Jump' is located in the top right corner. Below the navigation menu, the breadcrumb trail reads 'You are here: Documentation/Zone Web > WebHome (16 Apr 2010)'. The main heading is 'A Zone-Based Version of UrbanSim' in green. The text below explains that UrbanSim can be configured to run at a Zone Level, Parcel Level, or Gridcell Level. It describes the zone level model configuration, noting that a zone can be any irregular polygon, from Traffic Analysis Zones to large administrative units like cities or census blocks. It mentions that the description is based on the development and initial testing of a zone-level UrbanSim configuration in San Antonio, Texas, Durham, North Carolina, and the Puget Sound, Washington. The text also discusses the flexibility of the OPUS platform and how the UrbanSim model configurations have been adapted from the parcel version to work at a zone level. It notes that in the zone version, buildings are aggregated data representing the composite of individual buildings of the same type in the same zone. Finally, it lists the models used in the zone version of UrbanSim, with a table providing details for each.

**A Zone-Based Version of UrbanSim**

UrbanSim can be configured to run at a [Zone Level](#), [Parcel Level](#) or [Gridcell Level](#). This section of the documentation describes the zone level model configuration. By zone, we mean any irregular polygons, from Traffic Analysis Zones used in travel models, to large administrative units such as cities, or small Census Blocks or Neighborhoods. It is up to the user to determine what level of geography fits their needs and data constraints best. The description below is based on the development and initial testing of a zone-level UrbanSim configuration that is being tested, with some variations, in San Antonio, Texas, Durham, North Carolina, and the Puget Sound, Washington.

Given the flexibility of the OPUS platform, it has been relatively easy to adapt the UrbanSim model configurations from the parcel version to work at a zone level. The `location_set` for the household location choice and employment location choice models are set to building, in both models. But in the zone version of the model, buildings are actually aggregated data, representing the composite of individual buildings that are of the same `building_type` and in the same zone. So if one defines two housing types, `single_family` and `multi_family`, then there will be two residential buildings per zone.

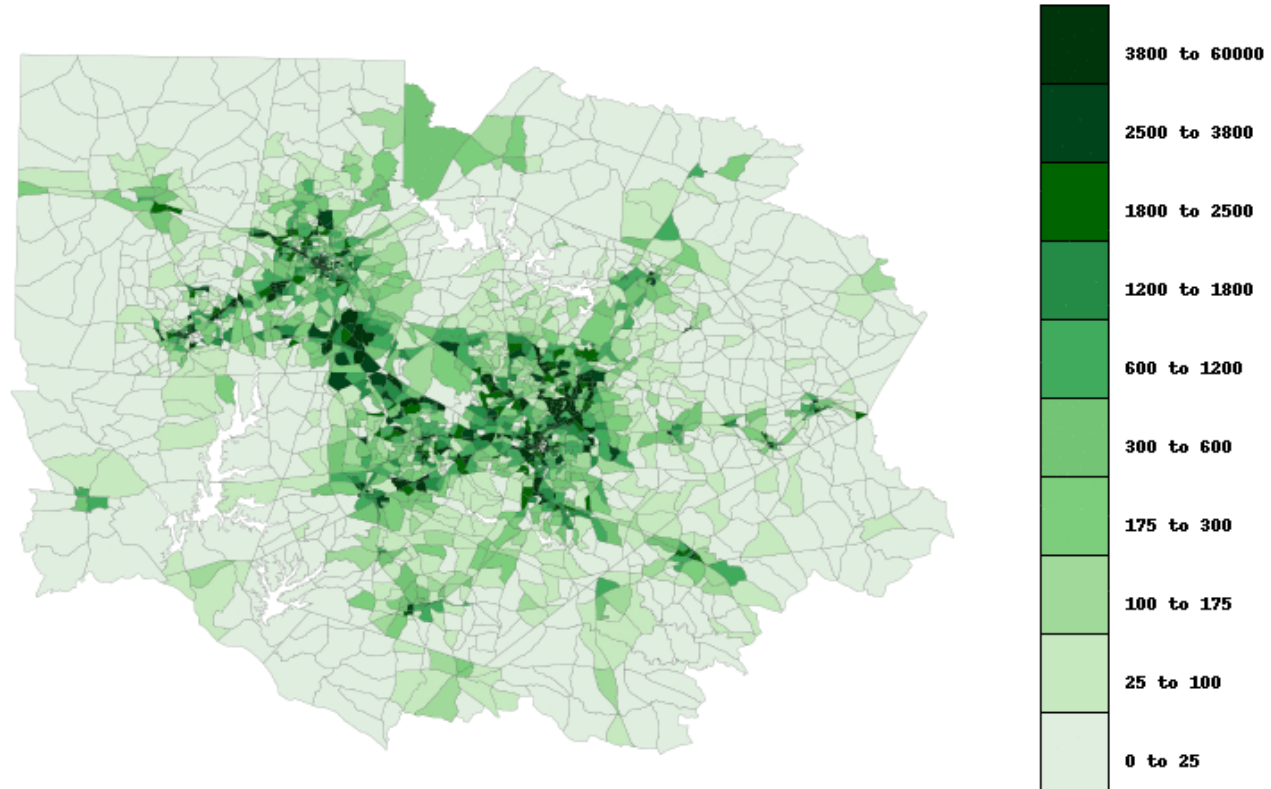
**Models in the Zone Version of UrbanSim**

Below are the models used in the zone version of UrbanSim. Each model name links to a more detailed description of its objective, algorithm, configuration and data.


Model Name	Abbreviation	Description
<a href="#">Scheduled Development Events Model</a>	SDEM	Handles Scheduled Development Events
<a href="#">Scheduled Employment Events Model</a>	SEEM	Handles Scheduled Employment Events

# Animated Map Indicators

2000



Automated Animated GIF images produced from Mapnik maps, using ImageMagick



# Database and Scenario Creation Tools and Graphical Interface – A Work in Progress

# Database Schema Management

- SQLAlchemy is an Object-Relational Mapper and database abstraction library, allowing use of standard Python syntax to create and query databases in multiple back-end database servers:
  - Postres, MySQL, SQLite, MS SQL Server, Access
- Elixir is a Declarative Layer on SQLAlchemy
  - Makes it easy to define database schema
  - Has a GUI builder, called Camelot, for browsing/editing tables

# Database Schema Definition using Elixir

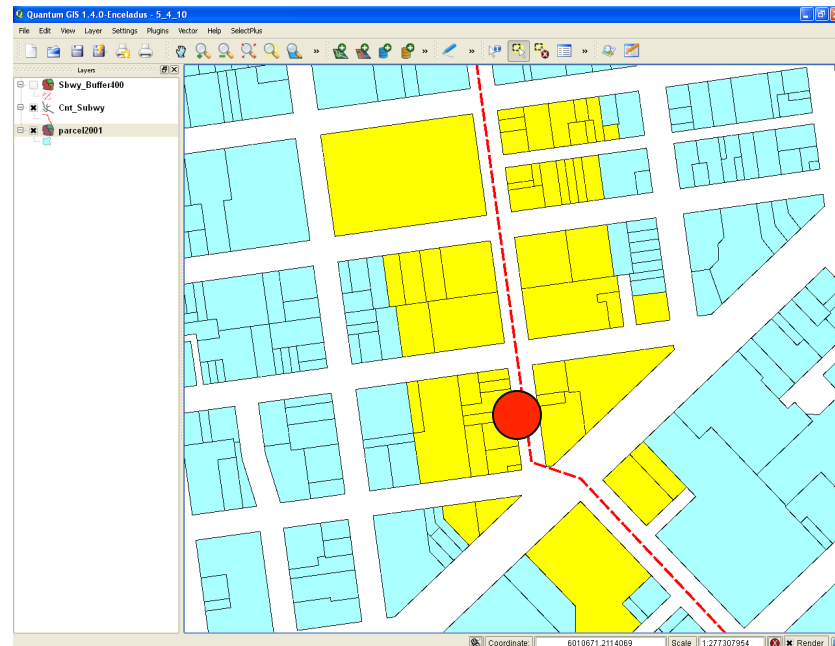
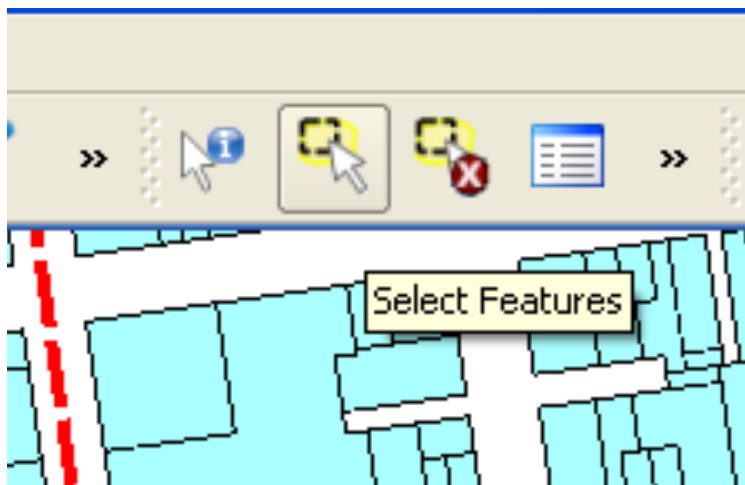
```
class Building(Entity):
  using_options(tablename='buildings')
  building_id = Field(Integer, primary_key=True)
  building_quality_id = Field(Integer)
  building_type = ManyToOne('BuildingType', colname='building_type_id')
  improvement_value = Field(Integer)
  land_area = Field(Integer)
  non_residential_sqft = Field(Integer)
  residential_units = Field(Integer)
  sqft_per_unit = Field(Integer)
  year_built = Field(Integer)
  stories = Field(Integer)
  tax_exempt = Field(Integer)
  parcel = ManyToOne('Parcel', colname='parcel_id')
```

# Scenario Creation Graphical Interface

- Requirements
  - Easy means to select parcels or other geographies
  - Ability to edit attributes
  - Ability to edit geometry (where appropriate)
  - Open source, multi-platform, easy to integrate with OPUS
- Strategy:
  - Quantum GIS for spatial data interaction, with customization
  - SQLAlchemy/Elixir/Camelot for tabular data management

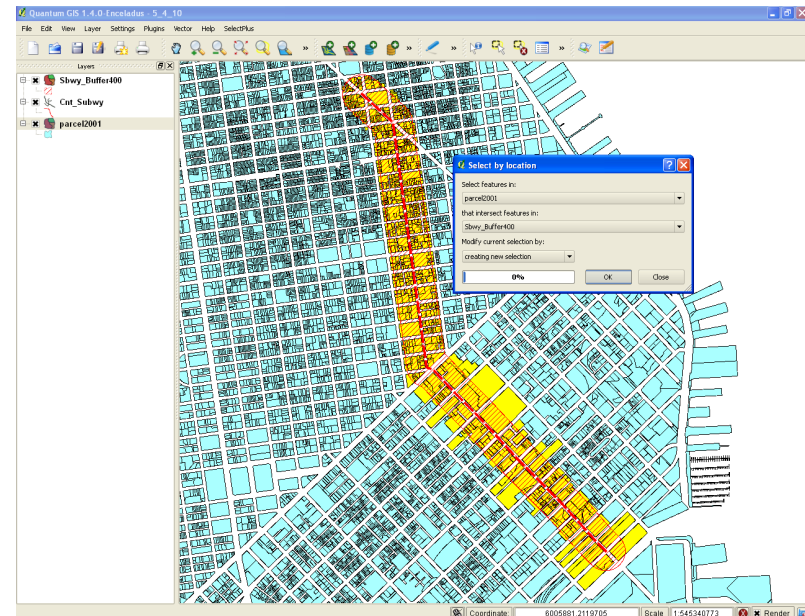
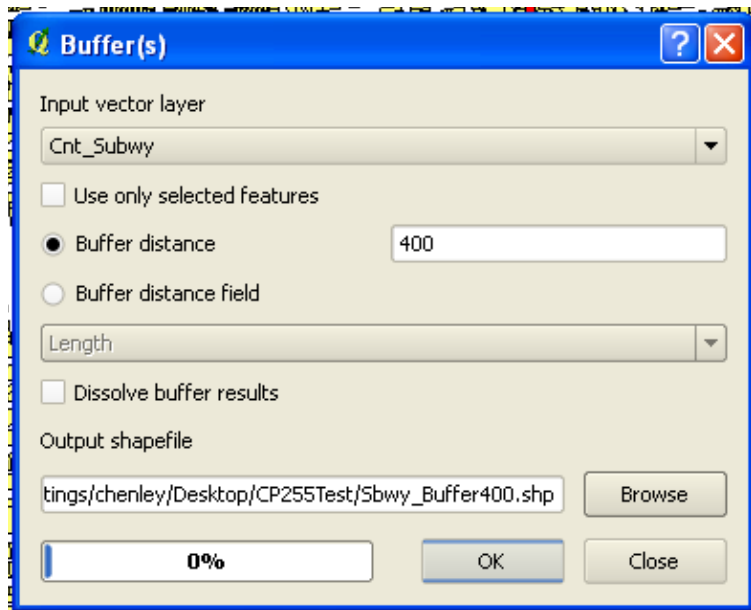
# Geographic Selection

- Basic spatial selection → Point and Click
- e.g. parcels near a planned subway stop



# Geographic Selection

- Location or proximity selection
- e.g. parcels within buffer distance of subway alignment





# Geographic Selection

- Selection by attributes
- e.g. parcels within buffer with FAR  $\leq 2$

Search query builder

parcel2001

Fields

- incomplete
- land\_val
- struc\_val
- dbiustype
- stories
- bldgsqft
- lidarsqft
- high\_sqft
- yrbuilt
- Far
- totaluses
- tractid
- taz\_1
- shape\_leng
- shape\_area
- planning\_2

Values

- 0
- 1
- 1.8
- 2
- 2.5
- 2.8
- 3
- 3.6
- 4
- 4.8
- 5
- 6
- 9

Operators

= < > LIKE % IN NOT IN

<= >= != ~ AND OR NOT

SQL where clause

Far <= 2

OK Test Clear Cancel Help

Attribute table - parcel2001

	far	totaluses	tractid	taz_1	shape_leng
5906	1.8	1000	010400	381	190.000063055
6219	1.8	1790	010700	376	371.8116630605
6220	1.8	0	010700	376	165.00213608
6222	1.8	3900	010400	382	236.003135073
6584	1.8	15900	010400	381	269.999312451
6585	1.8	0	010400	382	225.007346589
6587	1.8	0	010400	381	382.888683508
6590	1.8	0	010400	382	120.003856662
6693	1.8	1700	010400	381	235.000294048
6697	1.8	0	010700	376	165.001234232
6699	1.8	0	010700	376	215.002323677
6805	1.8	0	010700	376	168.097727627

Show selected records only Search selected records only

Advanced search Help

# Simple Table Editing

- Edit Development Constraints for Selected Parcels
- Constraint changes are made to a copy of the parcel table
- Changed values override original values in model runs

The screenshot shows a software dialog box titled "Dialog - TableEditor.ui". It contains two main sections: "Building Constraints" and "Permitted Building Types".

**Building Constraints:** This section features four horizontal sliders, each with a green handle. The sliders are labeled as follows:

- Height Limit:** The range is from 40' to 350'. The handle is positioned at approximately 100'.
- Floor-Area Ratio:** The range is from 1 to 9. The handle is positioned at approximately 1.5.
- Minimum sq. ft per unit (residential only):** The range is from 400 to 1200. The handle is positioned at approximately 400.
- Maximum Building Footprint (% of parcel size):** The range is from 50 to 100. The handle is positioned at approximately 50.

**Permitted Building Types (check all that apply):** This section contains a grid of 15 checkboxes, all of which are currently unchecked:

<input type="checkbox"/> Apartment	<input type="checkbox"/> Timeshare	<input type="checkbox"/> Hotel
<input type="checkbox"/> Condominium	<input type="checkbox"/> Church	<input type="checkbox"/> Industrial
<input type="checkbox"/> Flat	<input type="checkbox"/> Commercial	<input type="checkbox"/> Mixed Use
<input type="checkbox"/> Live/Work	<input type="checkbox"/> Entertainment	<input type="checkbox"/> Office
<input type="checkbox"/> Nursing home	<input type="checkbox"/> Gas Station	<input type="checkbox"/> Other
<input type="checkbox"/> SFH	<input type="checkbox"/> Hospital	<input type="checkbox"/> Parking
<input type="checkbox"/> Port	<input type="checkbox"/> Public	<input type="checkbox"/> School

At the bottom right of the dialog, there are two buttons: "OK" and "Cancel".

# Simple Interactive Table Filtering/Editing

For example, to edit control totals tables

The screenshot displays the Camelot software interface for editing control totals tables. The window title is "Camelot - [Households Control Totals]". The menu bar includes File, Edit, View, Window, and Help. The toolbar contains various icons for navigation and editing. On the left, a tree view shows the project structure under "Control\_totals", including folders for "Households control totals", "Business control totals", "Household relocation rates", "Job relocation rates", "Development constraintss", and "Development event historys". The main area shows a table titled "Households control totals" with 180 rows. The table has three columns: "Year", "Persons", and "Total number of households". The data is organized by year and then by the number of persons. On the right, there are interactive filters for "Persons" and "Year". The "Persons" filter is currently set to "all", and the "Year" filter is also set to "all".

Year	Persons	Total number of households
2000	0	133204
2000	1	100433
2000	2	67934
2000	3	20654
2000	4	7475
2001	0	133941
2001	1	100989
2001	2	68310
2001	3	20768
2001	4	7516
2002	0	134682
2002	1	101547
2002	2	68688
2002	3	20883
2002	4	7558
2003	0	135427
2003	1	102109
.....	-	.....