Seattle Water Demand Forecast Model

Final Project
CEE 588A: Introduction to Urban Simulation
Professor Paul Waddell

Kathleen King, Austin Polebitski, Lee Traynham
June 6, 2006

With very special thanks to Liming Wang and Hana Ševčíková

Table of Contents

Introduction

Topic Selection

On June 6, 1889, the "Great Seattle Fire" destroyed much of Seattle, including the 64-acre business district. The city's response was largely hindered by a lack of water, exacerbating the extent of the damage. At the time, only wells, springs, and a patchwork of private companies supplied water. When the mayor proposed that the city develop its own municipal water system, the 1,875 to 51 vote showed resounding approval ("Water System History" 1).

Today, Seattle Public Utilities (SPU) is responsible for meeting the water demands of the majority of King County, not just in the extreme cases of fire and drought, but year-round. Expected demands and the supply system capacity must be regularly assessed in order to determine when system expansion will be both necessary and cost-effective. From financing to planning, design, and construction, expansion projects often require several years to implement. As such, proper planning requires that the agency be able to forecast Seattle's water demand well into the future.

Purpose

In assessing their most recent water demand model, Seattle Public Utilities was assisted by the University of Washington's Water Resources and Drought Management Planning Group. Under the leadership of Dr. Richard Palmer, the group evaluated SPU's water demand model when it became available in March, 2006. After a thorough review of the model, King County representatives posed the task of enhancing the model by altering format, methodology, and other relevant attributes. As such, a subset of the research group enrolled in Professor Paul Waddell's CEE 588A: Introduction to Urban Simulation course, with the intent of determining the suitability of urban simulation methods for forecasting water demand in Seattle.

Developed in the 1960's, urban simulation modeling was initially used as a tool for transportation planning and housing policy evaluation (Waddell, 2006). However, modern versions of the software are much more versatile and capable. For instance, "Opus is a new platform for urban and regional simulation, designed to support the development and integration of model components to simulate the effects of transportation, land use, and environmental policies on the evolving urban spatial system" ("Users Guide", 2006). Having taken advantage of advances in computing, modeling, data availability, and behavioral theory, modern urban simulation has become an analytical tool of choice to evaluate complex management policies and long-term goals (Waddell, 2006). Because of its ability to integrate large amounts of data describing separate, but connected systems, urban simulation lends itself to the task of forecasting future water demands in the face of ever-changing land use patterns, regulatory policies, climate and environmental conditions, and resource management strategies.

Scope

The work described herein was undertaken by three graduate student members of the Water Resources and Drought Management Planning Group as a term project for the CEE 588A course. The goal of this venture was to develop a simple forecast model using independent variable data produced by the UrbanSim 4 package. The practicality, transparency, flexibility, and accuracy of the UrbanSim model were to then be compared to SPU's model in order to determine the effectiveness of urban simulation as a water demand forecasting tool. A summary of tasks completed for the project is below, and each task is discussed in greater detail in subsequent sections of this report.

1.) Researched Water Demand Modeling Alternatives and Historical Practice
    a. Reviewed Literature on Water Demand Forecasting
    b. Researched the Various Historical Demand Models Developed and Utilized by SPU
2.) Gained Familiarity with Opus Platform, Python, and UrbanSim4
    a. Reviewed Literature on Urban Simulation Applications to Real Systems
    b. Studied Opus, UrbanSim4 and Python User Manuals
    c. Actively Participated in Hands-on UrbanSim 4 Tutorials
3.) Developed Forecast Model
    a. Collected and Processed Data
    b. Coded Models

c. Ran Models
d. Compared UrbanSim model w/ SPU Model

Long-Term Water Demand Forecasting Methods

Background

A forecast of future water demands is essential in water resources planning—not only for preparing water supplies to meet anticipated future demands, but also for understanding what factors drive water demand. An accurate forecast will help ensure that water is managed properly and that supply will be adequate for future decades; however, demand forecasting is an evolving art, and there is not one single method that is widely agreed upon as the best approach. Many recent forecasting methods have evolved to try and accurately project not only water demand, but also demography, economy, and social attitudes of a region.

Water conservation, defined by Baumann et al. (1980) as "the socially beneficial reduction of water use or water loss," has played an important role in reducing water demands in recent decades. However, water conservation effects are now reflected in recent (since the early 1990s) historic water-use data, adding another layer of complexity to water demand analysis. The nature of water conservation makes it difficult to predict; the timing and extent of conservation are a function of both policy and public education and attitude (political/social feasibility). In addition, the practical feasibility of conservation must be considered (i.e. the set amount of water that an individual needs at a minimum). Household water use is often tied to more durable appliances such as toilets, dishwashers, and laundry machines. Even if advances in technology, changes in plumbing codes, the price of water, or public awareness encourages increased water conservation, the delay before the installation of water-saving appliances may cause a time-lag, making the task of simulating the casual relationship more difficult.

Water demand forecasts have been evolving since mid-1960s (Gottlieb, 1963; Howe and Linaweaver, 1967). Prior to that, water demand was thought to be a rudimentary function of the number of residential and industrial water users. Since the inception of water demand forecasting, two distinct categories of water demand forecasting have formed: long-term and short-term. Short-term demand forecasts often project water demand within a single year, season, or even week. Long-term forecasts, however, commonly look at timeframes of ten to fifty years. Trying to extent forecasts beyond fifty years is often a fruitless endeavor, since the uncertainties in policy, lifestyle, and physical infrastructure are too broad to properly simulate. In this demand forecasting review, we will be focusing on water demand forecasting techniques appropriate for long-term (~30 yr.) scenarios. Long-term forecasts are often more complex because water demand is more elastic in the long range (due to conservation implementation delays, etc.) and longer time frames increase uncertainty about future conditions. As Brookshire et al. (2002) note, although demand elasticity increases as timeframe increases, there is not enough variability in water price nationally to fully understand the affect of water price on demand. Furthermore, water is at least moderately under priced. This adds another element of uncertainty when simulating consumer response to different pricing schemes that are outside the observed range of pricing techniques.

Forecasting Methods

Demand forecasting methods currently range from the relatively simple to the highly complex. The most common forecasting methods include:

(i) Trending

Trending, or trend analysis, is a simple method that no longer common in long-term water resources planning. It forecasts water demand by fitting a trend line to historic data and extrapolating into the future. This method may be appropriate for small utilities for short-term projections, but grossly overestimates water demands for long-term forecasts.

(ii) Per Capita

Per capita demand forecasting requires data on current per capita water usage and applies it to a population forecast. The change in demand is proportional to a change in population. This method does not take into account changes in public perception of water, economy, technology, etc.

(iii) Sectoral Disaggregation

Forecasting by sectoral disaggregation is similar to the per capita method, but per capita water use factors are broken down into smaller subsets of customers to make the forecast more detailed. This requires a population forecast for each customer subset, as well as past usage data for each subset. However, like the per capita forecasting method, sectoral disaggregation fails to account for changes in lifestyles, economy, technology, etc.

(iv) Econometric

Econometric analyses are a more common method of demand forecasting. These models establish a statistical relationship between water demand and several determinants (commonly including: price, income, number of household, weather, employment). Historic data for all the determinants, as well as past water usage, is required. In addition, future forecast values are needed for the determinants. The benefit of an econometric model is that forecasts can be independently created for any of the determinants (e.g. for weather, climate change forecasts), and these forecasts can be incorporated to reflect how the changes will alter demand.

(v) Variable Flow

Variable flow models are becoming more common for forecasting water demands. The variable flow method utilizes water use factors (consumption per household or employee), like the sectoral aggregation method, but modifies the water use factors over time to account for changes in price and conservation measures. Water use factors can be obtained from past billing data; however the importance of weather-normalizing the data beforehand is noted. After the water use factors are derived from weather-normalized data, the factors are multiplied by demographic projections, adjusted for price and income effects, and integrated with conservation planning to create the water demand forecast.

(vi) End Use

End use models are a more detailed approach to water demand forecasting. End use models identify the quantity of water used for specific activities (e.g. flushing the toilet), and by applying this together with the population data and frequency of the water-consuming activity, a water demand forecast is formed. End use models are helpful for evaluating the impact of conservation programs, but are not widely used to forecast water demands because they are so data-intensive.

Discussion

Trending, per-capita, sectoral disaggregation, and end use methods are not widely used for regional long-range water demand forecasts. While end use models are typically avoided because they require data that may be too detailed and variable for a regionally-scoped project, the other methods are seen as markedly simplistic for modern day long-term demand forecasting. However, some of the more basic approaches are still useful for small utilities for short-range forecasts.

Econometric models are a common demand forecasting method, but the substantial data requirements have limited their use. Additionally, concerns about the handling of price variables and conservation in econometric models are another disincentive. As more recent data is made available to aid model forecasts, it becomes difficult to separate price effects and conservation effects. Similarly, trying to quantify the impacts of conservation programs throughout historic datasets is a complicated undertaking. Most econometrical models do not account for the social behavior of individual consumers and instead attempt to account for changes in social attitudes towards water with an aggregated approach. Another attribute of the econometric method is that the forecasts are more accurate when the demand equation is site-specific. This can be seen as an advantage because the method is flexible enough to be used in a variety of locations or a disadvantage because individual econometric models cannot be directly compared to each other. The formation of the model also requires more effort since a standard demand equation does not exist. There are some “off the shelf” models available (e.g. IWR-MAIN), but these now recommend that clients estimate their water use equations. Although the detail of econometric models may increase the accuracy of forecasts, the transparency of the models has been criticized in the past. Furthermore, the complexity of econometric models may imply more certainty than actually exists.

Variable flow models have become more popular recently as a result of some of the complications of econometric modeling. The Outlook (Central Puget Sound Supply Forum 2001) utilized the variable flow factor approach in its forecast of water demands for Seattle-Everett-Tacoma, which covered more than 158 water utilities. The cities of Tacoma and Everett both independently employ the variable flow factor approach in their water demand forecasts. Supporters of the variable flow method site that it requires less data but may obtain results comparable to those obtained by the econometric method. Although the variable flow factor methods have increased in popularity, there is criticism over its simplicity. Seattle has recently switched from an econometric model to a variable flow water demand forecast model. They are currently using @Risk (a Microsoft Excel add-in) to create the variable flow model. The U.S. Forest Service has also utilized the variable flow factor approach to projected future water demands at a national scale. Although variable flow factor techniques increase transparency, the loss of detailed data is also a source of criticism.

San Diego has created residential water demand models by regressing per-household water use (with household subdivided into single-family and multi-family). Determinants of water demand in the San Diego model are weather, price, median household income, units per acre, persons per household, and time of year, and different equations are used for each housing type.

It may seem most useful to look at forecasting in the Pacific Northwest because of the continuity in water resources issues. However, although many Europeans didn't begin publishing reports on water demand forecasting methods until the 1990's, several noteworthy studies from abroad have emerged over the past several years. DAWN, a hybrid demand forecasting model developed by a group in Greece, combines an econometric demand forecasting model with a social simulation layer (Athanasiadis et al., 2005). The agent-based social model attempts to reflect how conservation actually takes affect; this is done by simulating how "agents" react and interact for a given scenario. Agents represent water consumers ("consumer agents", CAs) and suppliers ("water supplier agents", WSAs), which correspond to social stakeholders. These agents represent a society, where social interaction is simulated through an *influence diffusion mechanism*, which "simulates the propagation of water conservation signals within the consumer community, as a function of individual social relationships" (Athanasiadis et al., 2005). Each simulation agent consumes water based on an econometric model (which accounts for changes in consumption based on water-pricing policies) while integrating the effects of the social interaction model. For more detail on the DAWN simulation model, please see the paper by Athanasiadis et al. (2005).

Another approach, one created in the United Kingdom, uses a decision support system, DFMS, to emphasize the dynamic relationship between demand forecasting and management (Froukh, 2001). This method attempts to reduce complications from the overlap of the different interests, and instead makes the forecast and management processes more robust. The DFMS consists of a database, GIS, an expert system, predictive models, a multi-objective decision component, hypertext files and a user-interface. The process begins with GIS, where the user selects the demand zone. Thereafter, the expert system is employed to use rule-based inference and qualitative reasoning (from user answers to specific prompter questions) to determine future domestic demand forecasts. Input of both qualitative and quantitative data, along with the multi-component approach, hope to make this model more responsive and adaptable. DFMS combines four different water demand forecasting techniques, in addition to the multi-objective decision component which is meant to help the user choose the best forecasting and conservation methods based on objectives with variable weights. The details of the model are to be hidden from the user, which allows for a complex model that is easy to operate; yet, this takes away from the transparency of the model and the ability.

"Best Practice" Techniques

When Western Water (Australia) formed there 2004 Demand Forecasting Report, they identified eight "best practice" techniques in water forecasting, which are shown below:

1) *Climate correction of historical bulk water production records*—to help develop an understanding of the impact of historical pricing changes, restrictions on water use and underlying trends in total production.

2) *Climate correction of customer database records*—allows customer database data (the sum of metered consumption in key consumer categories) to be analyzed for the influence of climate and underlying trends in the same manner as the analysis of water production records.
3) *Cross-sectional analysis of customer consumption records*—provides insight into the demand drivers, particularly in the residential customer categories, and information on how key factors (e.g. household income, lot sizes, land zoning, dwelling types, occupancy rates, household sizes and weather) have an impact on the level of consumption.
4) *Disaggregation of forecast into customer categories and forecasting on the basis on the number of accounts*—allows the impact of key drivers to be catered for, such as changes in the mix of dwelling types, in order to increase the accuracy of the forecasts.
5) *Disaggregation of forecast into internal and external use*—allows the factors influencing each area of use to be forecast separately in order to increase forecast accuracy. Internal use created sewage, so it is important to consider the two separately to aid integration of wastewater management with demand forecasting.
6) *Disaggregation of forecast into individual end-uses*—provides information needed to asses the impact of different demand management initiatives.
7) *Development of econometric time series regression model*—develops a greater understanding of how the business cycle has influenced historical demand and how growth in the economy and household income can be expected to influence demand in the future.
8) *Detailed water use survey*—helps identify trends in appliance ownership and attitudes to water use, which is essential for a greater understanding of the factors influencing water use.

When selecting a demand forecasting technique, it is important to consider the tradeoffs between performance, transparency, and bulkiness (e.g. data collection and analysis requirements) of the model. Many local utilities in the Puget Sound region have switched to a variable flow approach because they felt that, although this method is less data-intensive than the econometric analysis, the forecasting results are of comparable accuracy while increasing transparency and decreasing the data requirements.

Methodology

Selected Modeling Approach

After examining current water demand forecasting methods, we chose to employ a multiple regression model based on land use and demographic information available through UrbanSim and PSRC forecasts. In some way, his resembles econometric models, which commonly use regression models. However, the model's determinants also relate it to a variable flow model. Determinants we have chosen to consider include population, land use, percent residential, average income per housing unit, and lag demand.

A multiple regression model is a multivariate statistical technique, which examines the variable being forecast (e.g. water demand) and multiple other variables (e.g. population, lot size). Multiple regression models have the advantage of forecasting water demand while considering forecasts of its determinants (e.g. population forecast). The equation for a multiple regression resembles:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \varepsilon$$

Where:

y = dependent variable, e.g. water demand
x = predictor variable, e.g. population
β = regression coefficients
ε = residual term

When using a multiple regression model, it is important to minimize the correlation between predictor variables. If predictor variables are related to each other, a change in one variable will cause other variables to change, resulting in an exaggerated change in the dependent variable. This is a problem of collinearity. Although relationships between predictor variables may be unavoidable, an effort to minimize collinearity is important for maximizing the prediction power of the regression model.

Utilization of UrbanSim

Regression coefficients can be calculated inside UrbanSim by using population data for the present year. Additionally, UrbanSim has the capability to create "rolled-back" land use data, which removes housing, etc., based on when building were created. However, historic population information is not available on the gridcell level, which impacts the ability to perform the regression over a longer historic period. If gridcell population is used as a predictive variable, the regression in UrbanSim can only include information from 2000 on. By using past consumption data from Seattle Public Utilities (SPU) for the dependent variables (aggregated to the gridcell level) and information from UrbanSim for the predictive variables, regression coefficient values can be obtained.

UrbanSim can output forecasted information that includes: regional distribution of population, households by type, businesses by type, land use by type, units of housing by type, square footage of nonresidential space by type, densities of development by type of land use, prices of land and improvements by land use. UrbanSim forecasts for these categories can be fed into the previously developed (as described above) regression model as predictive variable forecasts. Therefore, to create a water demand forecast, the regression coefficients (derived in UrbanSim using current information for the predictive variables) are used alongside UrbanSim forecasts of the predictive variables.

Model Verification

Once the water demand forecast is created using UrbanSim outputs, the forecast is compared to other water demand forecasts for the City of Seattle. Results from SPU's current forecasting method, in addition to previous forecasting methods, can be weighed against outputs from the water demand module incorporated into UrbanSim. Although this comparison will provide an idea of how well the model is working, it should be noted that water demand forecasts are traditionally predict exaggerated values for future water demands.

Data Collection and Processing

To be able to create a model that represents actual water demand, observed data had to be collected and properly formatted so that regression coefficients could be determined. The City of Seattle has for many years kept a thorough database of bimonthly consumption data at the parcel level. It was from this database that we have been able to create the consumption record to be used in UrbanSim for water demand modeling.

Dependent Variable - Water Demand

To accurately forecast water demand, historic data is typically used to calibrate and validate the model. For calibration of our model we used bimonthly billing and consumption data from 1992 – 2001.

Access Databases of Historic Consumption Data by Parcel

Access databases of historic consumption data were located in house from previous research. The Access databases were not formatted properly for direct input into the SQL database and so some formatting had to occur. The database was first related to a parcel database because the SPU account information only had account id's. By relating the accounts to the parcel it was possible to later relate the parcels to the gridcells that are used in the UrbanSim simulation. After the accounts were related to the parcels, each parcel was broken down into sector demand. Four sectors were created from the original database; these were commercial single and multiple register parcels, and residential single and multiple register parcels. After the sectors were separated into their separate databases, the data was prepped and then read into the MYSQL database by sector.

Conversion to MYSQL Database with Consumption by Gridcell

After the data was read into the MYSQL database, the process of relating the parcel to the gridcell could occur. Through this relating process, a fraction of demand from each parcel was assigned to a gridcell, because not all the parcels lie exactly within the gridcell layout. After the parcel data was joined to the gridcell through this fractionation process, the data was ready to be read into UrbanSim. After initial runs for determining coefficients of pertinent variables, it was felt that the runtime was too slow, Liming Wang, the class TA was able to alleviate this by writing the appropriate data per regression run to a float file before reading into UrbanSim to allow faster access and therefore much quicker runtimes.

Independent Variables: Weather

Weather is a key predictor for peak consumption when forecasting demand on a daily timestep. As a factor it has been seen that weather can affect demand, causing peak demand to be upwards of double what it typically is for a given system. This is typically attributed to outdoor lawn watering, car washing, and other outdoor water use though some of the increase is also increased measured as in-house usage because of the temperature rise.

<u>Historic Weather Data</u>

Historic weather data from 1991-2005 was tabulated and formatted so that certain variables could be examined and used in the regression process. Maximum and mean bimonthly temperatures as well as total precipitation values were created from daily time-series records that were obtained from the NCDC database. To be able to use the weather data as a predictor variable, a timeseries of the UrbanSim data must be created. For coefficient determination, the timeseries created from the unrolled data would have to be sampled over the period available and coefficients determined from that. Because we are currently able to only use cross-sectional data weather will have to be semi-omitted from the study. This is because the weather variable will act as a constant, picking up variability that is not actually variability explained by the weather variables because we are not able to create a timeseries of data for all UrbanSim variables back to the 1990's.

<u>Forecasts</u>

If weather is joined onto the water demand module in the future, climate change impacts on water demand could be researched. With temperature forecasted to increase 2-5 degrees F in the next 100 years (NCDC, 2006), water demand patterns could become significantly different because of increased peak demand as days grow warmer for longer periods of time and shifted seasonal precipitation and temperature could potentially shift when water is demanded as well as amount demanded. Climate change impacts coupled with changed land-use patterns could also compound effects of one-another causing pattern shifts as well.

Gridcell Data

Historic (Unrolled)

To be able to use the weather data, historic parcel data, and lagged water demand as predictor variables, historic parcel data had to be generated through a process of "unrolling" UrbanSim back to 1990. The unrolling process involves eliminating parcels based on built year. By removing parcels of certain built year, the data for each parcel is closer to representing the past than if the parcels were left as is with the 2000 year data in place. After the data was unrolled, variables could be predicted by sampling over all the years so that inter-year variability could be explained by the coefficients. This also allows weather to act as a predictor variable because multi-year data is being sampled and therefore allowing a trend to be predicted. One thing the rollback does not do is generate population, income, or jobs into the past. Instead that would have to be a separate effort that is outside the scope of this project but would provide data so that a multi-year regression could be performed.

Forecasts

The gridcell data used for forecasting future years is all generated by UrbanSim. As UrbanSim moves ahead every year with its simulation, data is sent from UrbanSim to the water demand model for gridcell variables that are used. The coefficients that have been determined previously are used to predict water demand at each gridcell. The gridcell water demand is then summed across the Seattle region for each billing period to give bi-monthly and total annual demand.

Process

Because weather has been omitted from our final python script due to inability to incorporate historic census data into the unrolled parcels, only two scripts are needed to unroll the gridcells, prepare the gridcells for analysis, obtain regression coefficients, and forecast water demand. The two python files are located within the appendix in full code.

Regression Model Code Description

The regression model code uses data generated by the prepare_data.py script. The prepare data script unrolls parcel data back to 1990. This data would be used in weather included version of the model. The prepare data script also caches gridcell data to the hard disk so that the regression model can run faster. The regression model, after taking in the cached gridcell data and the provided household dataset for the year 2000, performs a linear regression to generate coefficients for pertinent variables to be used in the forecasting. Data passing was originally done from the MYSQL database to the computer over the network, this ended up also slowing done the runtime significantly. To combat this, the water demand data for each sector was cached to the hard disk during the prepare data script run.

The regression script reads the integer or floating point files in and regresses the chosen variables against the water demand consumption data for each bill period, generating specific coefficients for that period that are later used in the forecast. The bill periods are defined by bi-monthly timesteps starting in January, so bill period 1 would be combined consumption from January and February. After the regression is performed and coefficients generated, the forecasting begins. This is all performed with the consumption_weather_single_year.py script.

Forecast Model Code Description

The forecasting uses pre-run gridcell data for the years 2001-2005. The gridcell data is passed to the water demand model, which has been populated with the appropriate coefficients for each bill period, and is used to generate future water demands for each bill period. The script was written in such a way that each bill period had to be forecasted as separate runs, i.e. when the script is run it generated 2001-2005 water demand for a selected period such as January-February.

Progress

Regression

Throughout the course of the quarter, our team has made important progress despite some problems we have run into regarding data limitations and computational constraints. We have integrated historic water consumption data into a MYSQL database, converted parcel-scaled data to gridcell level detail, and regressed water demand based on different combinations of predictor variables until we achieved satisfactory results.

After we ran the regression model using several different combinations of predictor variables, we found that out R^2 value plateaued at around 0.3. We tried to minimize the number of predictor variables and correlation by keeping predictors that increased the R^2 value by more than a percent, with a significant t-value.

The following gridcell variables proved to be useful for increasing the power of our regression equation:

- Constant value
- Average income per housing unit
- Natural logarithm of the total land value
- Natural logarithm of the total population within walking distance of a gridcell
- Percent residential
- Demand lag (the gridcell's demand during the same bimonthly period of the previous year)

We obtained the following results:

Coefficient	Estimate	Standard Error	t-values
constant	-7.882	0.265	-29.754
average_income_per_unit	6.870E-06	9.790E-07	7.010
land_value	0.456	0.018	25.364
population	0.346	0.020	16.953
percent_residential	0.013	0.001	20.532
demand_lag1	0.005	1.116E-04	45.465

Table 1 - Regression Coefficients (Period 4, With Lag)

Number of observations: 16361
R-Squared: 0.2953

Adjusted R-Squared: 0.2951

All of the variables we used were highly significant; however, we were surprised that our R^2 value remained so low. Water consumption is subject to the dynamics of human behavior, which is difficult to predict in a model. Additionally, modeling water consumption on such a small (150 meters by 150 meters) grid might overstate the difficulties of modeling water demand on a regional scale. In other words, using the regression equation on regionally-scaled data might provide better results.

Forecasting

Our initial forecasts cover the years 2001 to 2005, enabling us to compare our results to actual consumption values and to forecasts from SPU. After creating forecasts which used demand lag as a predictor variable, we realized that water consumption was increasing significantly every year. Therefore, we also decided to run forecasts without demand lag in order to see if we could get more realistic results.

The results of our forecasts are included in the Appendix. Although the prediction power and accuracy of the forecasts are not sufficiently accurate for water resources planning based on the five year test forecasts, the progress we have made thus far gives us a strong foundation for creating a more robust model. Please refer to Recommendations for Future Work below for some of our thoughts on how forecasting water demand in UrbanSim can be improved.

Results and Conclusions

From the initial results of our model runs we see that UrbanSim does a good job of catching the seasonality of water demand without having a weather variable for prediction. This is because we have broken the water demand model into bi-monthly periods for the regressing of the coefficients. This allows the coefficients being used in the model to have varying magnitude for each season and so the model picks up the changing trends for each season. We see that the demand lag variable is a bad predictor because it forces the use of last years demand forecast to effect the next years. This seems to be causing the demand to “blowout” and increase consumption by exorbitant amounts for each year forecasted. The model with the lag demand variable was increasing consumption by about 10 MGD a year, whereas the model without the lagged demand variable was increasing by only 2 MGD over 5 years, which is much more realistic.

What we can't say is how this model compares to the actual demand data. This is because we realized that the consumption data in the database had a larger area covered than originally thought, and so more time is needed to finish that evaluation. What can be said is that if the database used covered the Seattle area and purveyor residential demand, than the UrbanSim water demand model predicted that demand very well. This is despite the fact that weather, conservation, and yard size were not included as variables. With the addition of these we might see a much better and realistic prediction of actual values. Currently water demand is decreasing for the Seattle water system, which was not picked up by our UrbanSim model, mainly because conservation was not included.

Challenges

In attempting to develop a functional water demand model in the Opus environment, several problems were encountered. It has been noted that, with respect to urban simulation, "Early simulation efforts were hampered by limitations in data, computational capacity, modeling techniques, behavioral theory, and did not meet the optimistic expectations set out for them" (Waddell, 2006). Despite advances in several of these areas, significant limitations still abound.

Data Limitations

<u>Historic Data</u>

The first significant hindrance to the project was a lack of sufficient data. In order to develop coefficients for a regression model, it is necessary to have accurate historical data for all variables. However, historic data prior to the year 2000 was not readily available for the proposed independent variables. As described above, the historic values had to be "unrolled" from more recent data, and the accuracy of the results of this procedure is highly uncertain. As such, coefficients were derived using the five years worth of available data (2000 – 2005). However, doing so eliminated the opportunity to calibrate the model by comparing predicted water demands to actual values during these years.

Employee Data

Because water is consumed on a per-person basis, the anticipated number of people consuming is often considered a key variable for demand prediction. Unfortunately, historic values for the number of employees in a gridcell were not available for this study. A regression model developed for commercial water demand would have lacked information for a significant explanatory variable, severely limiting the model's explanatory power. Because of this, resources were focused on generating a model to predict residential demands.

Database Size

The large size of the input databases quickly became an additional concern. Containing ten years worth of consumption data distributed by parcels, some of the databases consisted of millions of records. This led to difficulties when the data was prepared and processed in order to develop MYSQL databases. These issues were eventually overcome by the course teaching assistant, Liming Wang, however, portions of the model continue to suffer from relatively slow run times.

Complex Coding

Python Experience

One of the largest project constraints was the general inability of group members to code in the Python language. This problem was most persistent, as it prevented the members from readily implementing ideas in a timely and efficient manner. The group became heavily dependent on the teaching assistants, who, though extraordinarily helpful, were not available twenty-four hours a day.

Time Series Data

The need to perform a regression on time series data presented further difficulties. Unlike the sample models presented in class, the water demand model was to be developed using ten years worth of historical data, instead of just one. In addition, because of the seasonal nature of water demand, it was necessary that the model be run at a smaller, bimonthly time-step, as opposed to the one year time-step. It was not immediately clear how these deviations might be incorporated into the model code. Liming Wang addressed these concerns to the best of his ability. While the model currently runs at the bimonthly time-step, regression coefficients are generated using one year's worth of data only. Unfortunately, this precludes the use of weather data, which is constant across all gridcells for a given bimonthly period and year.

Behavioral Theory

Despite advances in behavioral theory, very little is known about the way in which individuals make water consumption choices. For example, demand elasticities are a measure of how much demand varies with a given change in price. However, there is no agreed upon value for the demand elasticity of water, and various studies have produced a wide range of possible values. Because individuals tend to behave differently across time, geographies, social classes, etc., a great deal of uncertainty persists when forecasting water demand.

Further Investigation

Address Challenges

In order to develop a fully functional water demand forecasting model for the city of Seattle, the challenges described above should be addressed to the extent possible. Datasets should contain at least 15 years worth of historic values and information for all demographic, economic, and climate variables so that a reliable regression model can be produced. This data may be available from the PSRC and/or other local agencies, but it may require significant processing before it can be incorporated into the model. Next, group members must gain proficiency with Python. Unfortunately, this is likely to require excessive amounts of time. An alternative may be to temporarily engage the services of an experienced programmer, as was done for the purposes of this class. When code writing is no longer an obstacle, it will be less difficult to overcome the obstacles presented by the time series data.

Enhance the Model

With capable coders and sufficient data, the model can be greatly improved upon such that it may meet the goals initially laid out for this project. Regression coefficients can be calculated using data from multiple years, weather data can be incorporated, and a commercial model can be developed.

In the future, the demand model can be further enhanced in several ways. First, additional explanatory variables can be examined and potentially integrated into the current code. Variables of interest that are not currently available might describe water price, conservation efforts, income distribution, lot sizes, drought events, economic recession, and more. Some of these variables can be created within the Opus environment using available UrbanSim data. Others can be generated outside of the model and then incorporated as input tables.

SPU supplies water not just to inner-city communities, but to surrounding areas as well. The model can be adapted to predict demands in these areas as well. This task will likely prove challenging. Water delivered to outlying regions is generally wholesale, and, in some cases, historic data for both dependent and independent variables may not be available at the gridcell scale.

Finally, conducting a more extensive statistical analysis of the data may prove to be advantageous. Input data can be examined for outliers, correlated variables, and the need for dataset transformations by performing exploratory data analyses. Accounting for the findings of these exercises may result in more consistent and accurate results.

Sources

Arbués, F., M.A. Garcia-Valiñas and R. Martinez-Espiñeria. 2003. Estimation of Residential Water Demand: A State of the Art Review. *Journal of Social-Economics*, 32. 81-102.

Athanasiadis, I.N., A.K. Mentes, P.A. Mitkas and Y.A. Mylopoulos. 2005. A Hybrid Agent-Based Model for Estimating ResidentialWater Demand. *SIMULATION*, 81 (3): 175-187.

Brookshire, D., H.S. Burness, J.M. Chermak and K. Krause. 2002. Western Urban Water Demand. University of New Mexico, Department of Economics.

Brown, T. C. 2000. Projecting U.S. Freshwater Withdrawals. *Journal of Water Resources Research*, 36 (3): 769–780.

Dandy, G., T. Nguyen and C. Davies. 1997. Estimating residential water demand in the presence of free allowances. *Land Economics*, 73 (1): 125-139.

Foster, H.S., and B.R. Beattie. 1979. Urban Residential Demand for Water in the United States. *Land Economics*, 55 (1): 43-58.

Froukh, M.L. 2001. Decision-Support System for Domestic Water Demand Forecasting and Management. *Water Resources Management*, 15: 363–382.

Gottlieb, M. 1963. Urban domestic demand or water in the United States. *Land Economics*, 39 (2): 204–210.

Howe, C.W. and F.P. Linaweaver. 1967. The impact of price on residential water demand and its relationship to system design and price structure. *Water Resources Research*, 3 (1): 13–32.

IWR-MAIN. http://www.iwrmain.com/

Kiefer, J.C. and G.A. Porter. 2000. Development of Probabilistic Water Demand Forecast for the San Diego County Water Authority. Planning and Management Consultants, Ltd.

Ludlum, M. 2004. Seattle Public Utilities: 2004 Water Demand Forecast Update, Phase 1, Literature Review.

Nauges, C., A. Thomas. 2003. Long-run Study of Residential Water Consumption. *Environmental and Resource Economics*, 26 (1): 25 - 43.

NCDC. http://www.ncdc.noaa.gov/oa/climate/climatedata.html#daily. SEA-TAC Meteorological Station. Last Accessed: Jun 2006. Last Updated: Jun 1, 2006 by Mark.Lackey@noaa.gov.

"Opus: The Open Platform for Urban Simulation and UrbanSim Version 4, Reference Manual and Users Guide." Center for Urban Simulation and Policy Analysis, University of Washington. Jun 2006.

Palmer, R. N. and K. V. Lundberg. "Integrated Water Resource Planning."

Waddell, Paul. CEE 588A: Introduction to Urban Simulation. Jun 2006. http://faculty.washington.edu/pwaddell/UrbanSim/.

"Water System History." Seattle Public Utilities. 2006. http://www.seattle.gov/util/About_SPU/Water_System/History_&_Overview/WATERSYST_200312020908156.asp

Western Water. 2004. Demand Forecasting Report: Water Demands and Wastewater Flows.

Appendix A: Code

Prepare_data.py

```
print "Create MySQL connection"
from opus.urbansim.store.scenario_database import DbConnection
from opus.core.dataset import DataSubset
from numarray import where
import os

#step 1. unroll gridcell from 2000 to 1990
from opus.washtenaw.estimation.estimation_config import run_configuration
from opus.urbansim.model_coordinators.cache_mysql_data import CacheMysqlData
from opus.core.configuration import Configuration

def rm_rf(path):
    """Recursively deletes a directory and all its children (directories or files).
    Can use a relative or absolute path"""
    if not os.path.exists(path):
        return
    if os.path.isdir(path):
        contents = os.listdir(path)
        for file in contents:
            absolute_path = os.path.join(path, file)
            rm_rf(absolute_path)
        os.rmdir(path)
    else:
        os.remove(path)

if __name__ == '__main__':
    cache_directory = "D:/urbansim_cache/water_demand"

    gridcell_config = Configuration({
        'in_storage_type':'mysql',
        'db_host_name':'artemis.ce.washington.edu',
        'db_user_name':'urbansim',
        'db_password':'cee530',
        'db_input_database':'psrc',
        'db_output_database':None,
        'cache_directory':cache_directory,
        'base_year':2000,
        'tables_to_cache':[
            'gridcells',
    #       'households',
    #       'jobs',
        ]})

    #CacheMysqlData().run(gridcell_config)

    # step 2 cache water demand data by
    dbcon = DbConnection(db="water_demand_seattle",
                    hostname="trondheim.cs.washington.edu",
                    username="waterdemand",
```

```
                    password="wewantH2O")

    print "Create Storage object."
    from opus.urbansim.storage_creator import StorageCreator
    storage = StorageCreator().build_storage(type="mysql", location=dbcon)

    from opus.water.datasets.consumption import ConsumptionSet
    consumption_types = ['WRSR', 'WRMR', 'WCSR', ] #'WCMR'
    for consumption_type in consumption_types:

        consumption = ConsumptionSet(in_storage = storage, in_table_name=consumption_type+'_grid')

        for year in range(1990, 2001):
            print "%s %s" % (consumption_type, year)
            year_index = where(consumption.get_attribute("billyear") == year)
            out_storage = StorageCreator().build_storage(type="flt", location=os.path.join(cache_directory,
str(year)))
            consumption_subset = DataSubset(consumption, year_index)
            consumption_subset.write_dataset(out_storage=out_storage,
out_table_name=consumption_type.lower())
```

Consumption_weather_single_year.py

```
import os
from opus.core.simulation_state import SimulationState
from opus.urbansim.session_configuration import SessionConfiguration
from opus.core.attribute_cache import AttributeCache
from opus.urbansim.model_coordinators.misc import create_datasets_from_flt
from opus.urbansim.store.scenario_database import DbConnection
from numarray import where, logical_or, ones, zeros, exp
from opus.urbansim.storage_creator import StorageCreator
from opus.water.prepare_data import rm_rf

print "Create gridcell and water conpsumption dataset from cache directory"
cache_directory = "C:\Documents and Settings\student\My Documents\urbansim_cache"

SimulationState().set_cache_directory(cache_directory)
SimulationState().set_current_time(2000)

# have session_configuration points to dbcon, in case we may need some tables when computing variables
#dbcon = DbConnection(db="water_demand_seattle",
#                   hostname="trondheim.cs.washington.edu",
#                   username="waterdemand",
#                   password="wewantH2O")

dbcon2 = DbConnection(db="psrc",
                   hostname="artemis.ce.washington.edu",
                   username="urbansim",
                   password="cee530")

from opus.urbansim.session_configuration import SessionConfiguration
sc = SessionConfiguration(data={"in_storage_location":dbcon2}).get_session_configuration()

# what type of water consumption
```

```
consumption_type = 'wrsr'
required_datasets = {
#                   'job':{'package':'urbansim'},
                    'household':{'package':'urbansim'},
                    'gridcell':{'package_name':'urbansim'},
                    consumption_type:{'package_name':'water'}
                    }

datasets = create_datasets_from_flt(required_datasets, None, {'in_storage':AttributeCache()})
consumption = datasets[consumption_type]

print "Create DataSet object for weather"
import opus.water
us_path = opus.water.__path__[0]
from opus.water.datasets.weather import WeatherSet
weather = WeatherSet(in_storage_type = "tab",
                    in_storage_location = us_path,
in_table_name = "weather.tab", id_name="year_id")

##print "Create ZoneSet"
##from opus.urbansim.datasets.zones import ZoneSet
##zones = ZoneSet(in_storage=storage)


# Group records by billmonth into sub model...
# categorize months into group, and assign a sub_model id for each group
# we will estimate a set of coefficient for each sub model group
#category_bins = [2,4,6,8,10,12]
#sub_model_ids = consumption.categorize("billmonth", category_bins) + 1
#consumption.add_attribute(sub_model_ids, "sub_model_id")

#or
#"""Arranging consumption data by 2-month bill period"""
print "Selecting data for bimonthly period"
months = [11, 12] # which two months to estimate and simulate
weather_attributes = ["t_max4"] # what are (is) the correspondent weather attribute name(s)
month_index = where(logical_or(consumption.get_attribute("billmonth") == months[0],
                               consumption.get_attribute("billmonth") == months[1]))[0]

index_est = month_index

# records consumption for a single year in given 2 months are less than 50000, so don't do sample
# but use full set by set index_set to None
#index_est = array(sample(month_index,50000))

#attache weather attribute to consumption data, as NONDERIVED attribute
from opus.core.attribute_meta_data import AttributeType
consumption.join(weather, name=weather_attributes, join_attribute="billyear",
                metadata=AttributeType.NONDERIVED)

print "Create Specification"
from opus.core.equation_specification import EquationSpecification
specification = EquationSpecification(
                 variables =
                  ("constant",
                   "consumption:opus.core.func.disaggregate(opus.urbansim.gridcell.percent_residential)",
```

```
"consumption:opus.core.func.disaggregate(opus.urbansim.gridcell.ln_total_population_within_walking_dis
tance) as population",

"consumption:opus.core.func.disaggregate(opus.urbansim.gridcell.average_income_per_housing_unit)",
                "consumption:opus.core.func.disaggregate(opus.urbansim.gridcell.ln_total_land_value)",
#               "t_max4",
                "sum_demand_lag1"  # we can add variables of last year (in this case, 1999) by using variable
name + "_lag1"
               ),
              coefficients =
               ("constant",
                "percent_residential",
                "population",
               "average_income",
#               "t_max",
               "total_land_value",
               "demand_lag1"
               )
             )

print "Create a model object"
from opus.core.regression_model import RegressionModel
model = RegressionModel()

print "Estimate coefficients"
coefficients, other_est_results = model.estimate(specification, consumption,
                outcome_attribute="opus.water.wrsr.sum_demand",  # if outcome_attribute is
opus.core.func.ln(), the simulation results need to take exp()
                index=index_est,
                procedure="opus.core.estimate_linear_regression",
                data_objects=datasets)


#prepare for simulation using specification and coefficents just estimated
from opus.core.chunk_specification import ChunkSpecification
from opus.core.resources import Resources
from opus.core.dataset_factory import DatasetFactory

for year in range(2001, 2006):  #simulation from 2001 to 2005
   print "\nSimulate water demand %s" % year
   #load datasets from correspondent year
   SimulationState().set_current_time(year)
   datasets = create_datasets_from_flt(required_datasets, None, {'in_storage':AttributeCache()})
   gridcells = datasets["gridcell"]

   #create a ConsumptionSet instance out of gridcells - simulate water demand for every gridcell
   resources = Resources({'data':{
         "grid_id":gridcells.get_id_attribute(),
         "billyear":year * ones(gridcells.size()),
         "billmonth":months[0] * ones(gridcells.size()),
         "sum_demand":zeros(gridcells.size())
         }})
   this_consumption = DatasetFactory().get_dataset(
                              entity_name = consumption_type, package='water',
                              arguments = {'in_storage_type':'ram',
```

```
                                            'resources':resources},
                                    )
    #join consumption set with weather data
    this_consumption.join(weather, name=weather_attributes, join_attribute="billyear",
                metadata=AttributeType.NONDERIVED)
    #run simulation
    result = model.run(specification, coefficients, this_consumption, index=None,
                 chunk_specification=ChunkSpecification(nchunks=3),
                 data_objects=datasets)

    #result = exp(result)
    this_consumption.modify_attribute("sum_demand", result)

    #keep only those with meanful water demand pridiction, e.g. residential_units > 0
    keep_index = where(result>0)[0]
    this_consumption.subset_by_index(keep_index)

    #write the output into a tab file
    out_storage = StorageCreator().build_storage(type="tab", location=os.path.join(cache_directory,
str(year)))
    this_consumption.write_dataset(attributes=["grid_id", "billyear", "billmonth", "sum_demand"],
                        out_storage = out_storage, out_table_name=consumption_type + '.tab'
                        )
    if os.path.exists(os.path.join(cache_directory, str(year), consumption_type)):
        rm_rf(os.path.join(cache_directory, str(year), consumption_type))
    this_consumption.flush_dataset()
    #print result
```

Appendix B: Regression Results

Lag Demand Scenario

period 1
January - February
Estimate regression for submodel -2
Number of observations: 16005
R-Squared: 0.405811186433
Adjusted R-Squared: 0.405625490823

Coeff_names	estimate	SE	t-values
constant	-320.098	13.5376	-23.6452
average_income	- 0.000318945	4.97E-05	-6.41707
total_land_value	17.1724	0.914531	18.7773
population	11.6039	1.04779	11.0746
percent_residential	0.715083	0.0318516	22.4504
demand_lag1	0.466353	0.00658234	70.8491

==

period 2
March-April
Number of observations: 16170
R-Squared: 0.300663272835
Adjusted R-Squared: 0.300446947443

Coeff_names	estimate	SE	t-values
constant	-379.666	13.09	-29.0044
average_income	- 0.000356485	4.86E-05	-7.33159
total_land_value	20.4492	0.888184	23.0236
population	13.6119	1.01935	13.3535
percent_residential	0.862376	0.0306272	28.1572
demand_lag1	0.243222	0.00532198	45.7014

==

period 3
May-June
Number of observations: 16309
R-Squared: 0.353786518454
Adjusted R-Squared: 0.353588329936

Coeff_names	estimate	SE	t-values
constant	-351.283	14.1415	-24.8406
average_income	-0.00027265	5.20E-05	-5.2384
total_land_value	19.847	0.955422	20.773
population	10.5313	1.0934	9.63168

percent_residential	0.792112	0.0332248	23.841
demand_lag1	0.397596	0.00662731	59.9936

==

period 4
July-August

Number of observations: 16361

R-Squared: 0.362551192906

Adjusted R-Squared: 0.362356314029

Coeff_names	estimate	SE	t-values
constant	-550.042	20.8499	-26.381
average_income	-0.000120015	7.71E-05	-1.55712
total_land_value	36.3953	1.41568	25.7086
population	4.30916	1.60766	2.6804
percent_residential	1.27293	0.0493142	25.8127
demand_lag1	0.547054	0.0087865	62.2608

==

period 5
September-October

Number of observations: 16196

R-Squared: 0.396941257006

Adjusted R-Squared: 0.396755012799

Coeff_names	estimate	SE	t-values
constant	-445.382	18.7686	-23.7302
average_income	-0.00011329	6.90E-05	-1.64284
total_land_value	29.6867	1.27186	23.3413
population	2.8763	1.4527	1.97998
percent_residential	1.13607	0.0442999	25.6451
demand_lag1	0.553756	0.0078735	70.3315

==

period 6
November-December

Number of observations: 15973

R-Squared: 0.360944721055

Adjusted R-Squared: 0.360744603538

Coeff_names	estimate	SE	t-values
constant	-286.195	14.0161	-20.419
average_income	-0.000222646	5.13E-05	-4.34152
total_land_value	16.1864	0.947438	17.0843
population	8.46029	1.079	7.84085
percent_residential	0.693822	0.032818	21.1415
demand_lag1	0.421792	0.00652425	64.65

==

No Lag Demand Scenario

period 1

January-February

Adjusted R-Squared: 0.219192752919

Coeff_names	estimate	SE	t-values
constant	-602.526	14.8282	-40.6339
average_income_per_unit	-0.000504585	5.69E-05	-8.86989
land_value	33.6819	1.01359	33.2301
population	18.8016	1.19527	15.7299
percent_residential	1.37066	0.0349323	39.2377

==

period 2

March-April

Adjusted R-Squared: 0.210103599437

Coeff_names	estimate	SE	t-values
constant	-528.922	13.4697	-39.2677
average_income_per_unit	-0.000456085	5.16E-05	-8.83617
land_value	29.2216	0.92149	31.7112
population	17.3064	1.07976	16.028
percent_residential	1.20999	0.0315252	38.3816

==

period 3

May-June

Adjusted R-Squared: 0.210927758689

Coeff_names	estimate	SE	t-values
constant	-605.495	14.9064	-40.6199
average_income_per_unit	-0.000425168	5.74E-05	-7.40232
land_value	34.8125	1.01899	34.1639
population	16.7091	1.20268	13.8932
percent_residential	1.3777	0.0350885	39.2637

==

period 4

July - August

Adjusted R-Squared: 0.211272388396

Coeff_names	estimate	SE	t-values

constant	-927.665	22.186	-41.813
average_income_per_unit	-0.000336065	8.56E-05	-3.92441
land_value	59.3645	1.52009	39.0532
population	11.7966	1.78299	6.6162
percent_residential	2.19035	0.0523406	41.8481

==

period 5
September-October
Adjusted R-Squared: 0.212494501686

Coeff_names	estimate	SE	t-values
constant	-834.42	20.4918	-40.7198
average_income_per_unit	-0.000344739	7.87E-05	-4.38034
land_value	53.1369	1.40235	37.8912
population	10.9892	1.65455	6.64177
percent_residential	2.07026	0.0482867	42.8742

==

period 6
November-December
Adjusted R-Squared: 0.193459817956

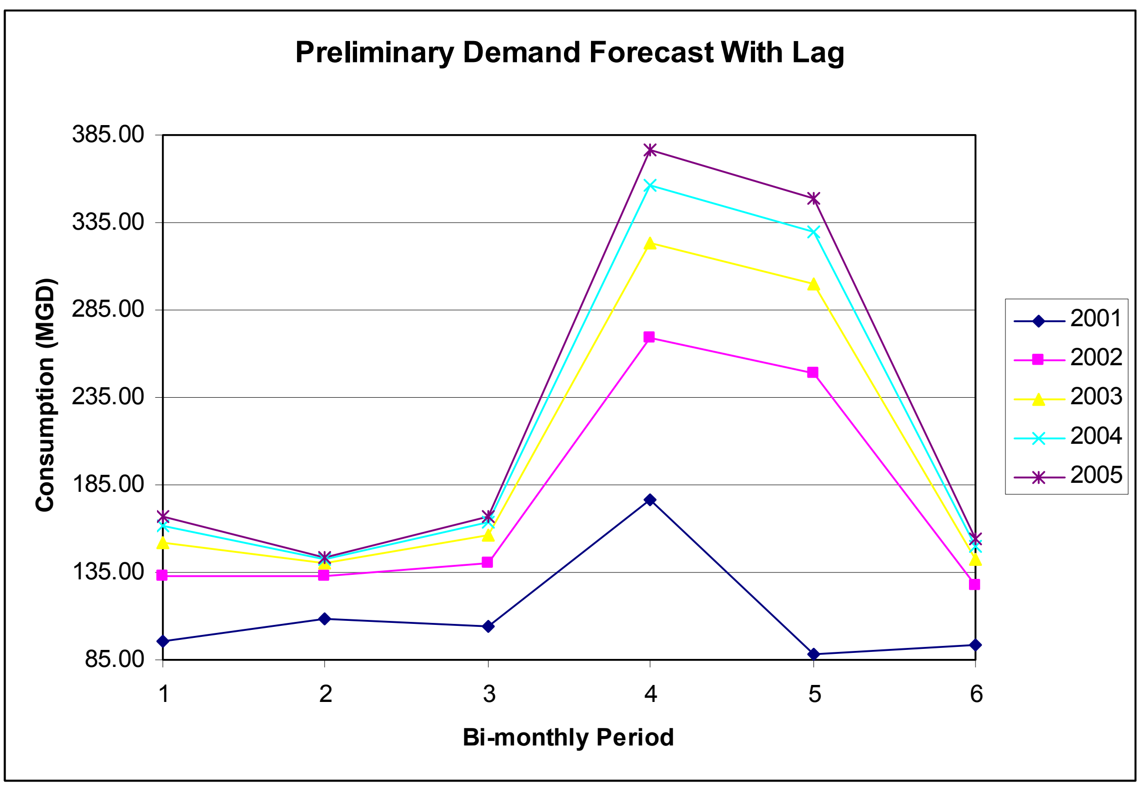
Coeff_names	estimate	SE	t-values
constant	-566.436	14.9717	-37.8337
average_income_per_unit	-0.000391932	5.75E-05	-6.81286
land_value	32.7758	1.02443	31.994
population	15.1721	1.20637	12.5767
percent_residential	1.31365	0.0352547	37.2617

==

Appendix C: Water Demand Forecasts

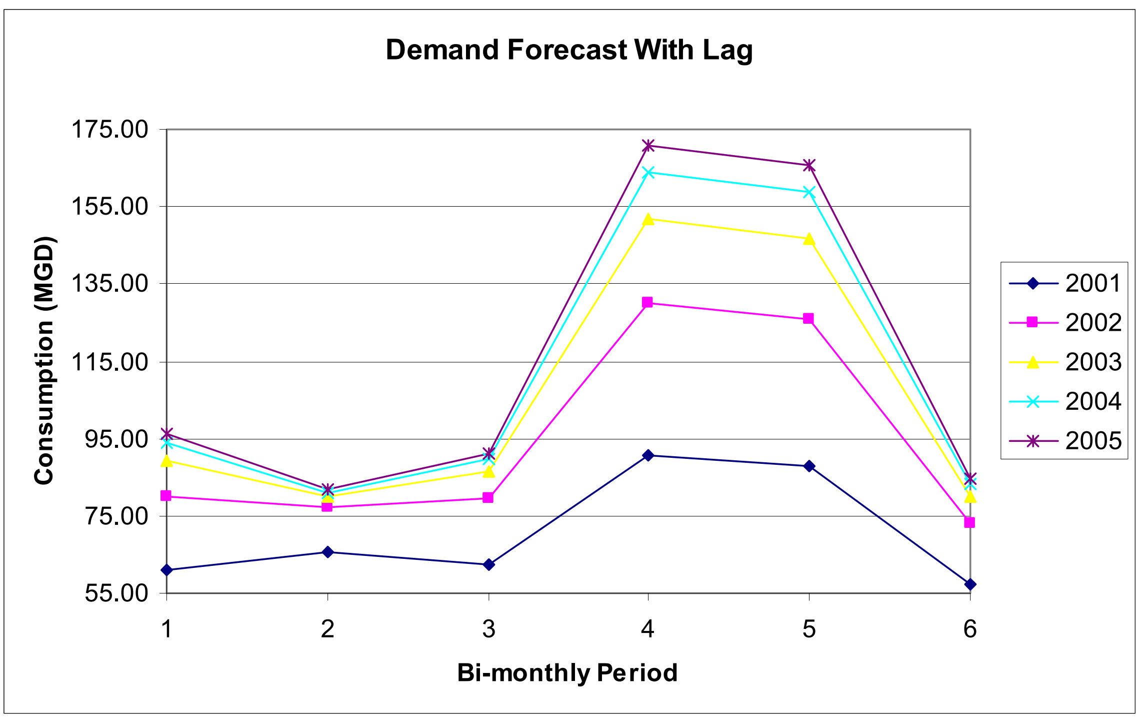
Preliminary Forecast Results With Demand Lag

Year	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
			Hundreds of Cubic Feet			
2001	7,551,000	8,879,000	8,475,000	14,652,000	31,230,000	7,656,400
2002	10,517,000	10,863,000	11,433,000	22,283,000	20,333,000	10,381,000
2003	12,008,000	11,420,000	12,729,000	26,817,000	24,448,000	11,628,000
2004	12,781,000	11,611,000	13,324,000	29,545,000	26,926,000	12,223,000
2005	13,204,000	11,713,000	13,626,000	31,230,000	28,459,000	12,533,000
			Millions of Gallons			
2001	5648.15	6641.49	6339.30	10959.70	23360.04	5726.99
2002	7866.72	8125.52	8551.88	16667.68	15209.08	7764.99
2003	8981.98	8542.16	9521.29	20059.12	18287.10	8697.74
2004	9560.19	8685.03	9966.35	22099.66	20140.65	9142.80
2005	9876.59	8761.32	10192.25	23360.04	21287.33	9374.68
			MGD			
2001	95.73	108.88	103.92	176.77	382.95	93.89
2002	133.33	133.21	140.19	268.83	249.33	127.29
2003	152.24	140.04	156.09	323.53	299.79	142.59
2004	162.04	142.38	163.38	356.45	330.17	149.88
2005	167.40	143.63	167.09	376.77	348.97	153.68



Forecast Results With Demand Lag

Year	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
			Hundreds of Cubic Feet			
2001	4,822,100	5,355,600	5,095,400	7,534,400	7,175,400	4,670,200
2002	6,326,200	6,287,000	6,490,600	10,775,000	10,253,000	5,968,200
2003	7,041,300	6,530,200	7,059,600	12,566,000	11,977,000	6,527,300
2004	7,393,800	6,612,300	7,305,600	13,569,000	12,952,000	6,780,400
2005	7,586,200	6,666,800	7,433,100	14,153,000	13,523,000	6,912,400
			Millions of Gallons			
2001	3606.93	4005.99	3811.36	5635.73	5367.20	3493.31
2002	4732.00	4702.68	4854.97	8059.70	7669.24	4464.21
2003	5266.89	4884.59	5280.58	9399.37	8958.80	4882.42
2004	5530.56	4946.00	5464.59	10149.61	9688.10	5071.74
2005	5674.48	4986.77	5559.96	10586.44	10115.20	5170.48
			MGD			
2001	61.13	65.67	62.48	90.90	87.99	57.27
2002	80.20	77.09	79.59	130.00	125.73	73.18
2003	89.27	80.08	86.57	151.60	146.87	80.04
2004	93.74	81.08	89.58	163.70	158.82	83.14
2005	96.18	81.75	91.15	170.75	165.82	84.76



Forecast Results Without Demand Lag

Year	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
Hundreds of Cubic Feet						
2001	7,696,500	6,878,600	7,648,900	11,663,000	11,261,000	7,342,900
2002	7,741,900	6,920,100	7,691,400	11,711,000	11,311,000	7,384,500
2003	7,767,300	6,942,900	7,715,900	11,745,000	11,345,000	7,407,500
2004	7,801,400	6,973,600	7,748,700	11,788,000	11,386,000	7,438,600
2005	7,854,500	7,021,200	7,800,000	11,856,000	11,451,000	7,486,900
Millions of Gallons						
2001	5756.98	5145.19	5721.38	8723.92	8423.23	5492.49
2002	5790.94	5176.23	5753.17	8759.83	8460.63	5523.61
2003	5809.94	5193.29	5771.49	8785.26	8486.06	5540.81
2004	5835.45	5216.25	5796.03	8817.42	8516.73	5564.07
2005	5875.17	5251.86	5834.40	8868.29	8565.35	5600.20
MGD						
2001	97.58	84.35	93.79	140.71	138.09	90.04
2002	98.15	84.86	94.31	141.29	138.70	90.55
2003	98.47	85.14	94.61	141.70	139.12	90.83
2004	98.91	85.51	95.02	142.22	139.62	91.21
2005	99.58	86.10	95.65	143.04	140.42	91.81