Urban Water Demand Forecasting: Review of Methods and Models

Emmanuel A. Donkor, S.M.ASCE[1]; Thomas A. Mazzuchi[2]; Refik Soyer[3]; and J. Alan Roberson, P.E.[4]

Abstract: This paper reviews the literature on urban water demand forecasting published from 2000 to 2010 to identify methods and models useful for specific water utility decision making problems. Results show that although a wide variety of methods and models have attracted attention, applications of these models differ, depending on the forecast variable, its periodicity and the forecast horizon. Whereas artificial neural networks are more likely to be used for short-term forecasting, econometric models, coupled with simulation or scenario-based forecasting, tend to be used for long-term strategic decisions. Much more attention needs to be given to probabilistic forecasting methods if utilities are to make decisions that reflect the level of uncertainty in future demand forecasts. **DOI: 10.1061/(ASCE)WR.1943-5452.0000314.** © *2014 American Society of Civil Engineers.*

Author keywords: Urban water demand; Forecasting; Time series; Uncertainty; Decision making.

Introduction

Reliable urban water demand forecasting provides the basis for making operational, tactical, and strategic decisions for drinking water utilities (Billings and Jones 2008; Gardiner and Herrington 1990) and is critical for several reasons. For instance, utilities need to know what the water demand for today and tomorrow will be to operate their treatment plants and wells appropriately to meet these demands. Utilities also need to predict the water demand 20–30 years in the future to develop new water sources and/or expand their treatment plants. Because decisions based on deterministic (single-point) forecasts do not accommodate possible deviations in demand, accounting for the uncertainty in these forecasts provides the basis for accurately quantifying the risk of water shortages and hence revenue risk, enabling utilities to optimize their operational and investment decisions (Cutore et al. 2008). Water utility examples of how this has been done in practice can be found in Hazen, Sawyer, and PMCL (2004) and Palisade Inc (n.d.).

Although forecasting is not a new discipline, its application in the water sector for demand estimation is fraught with many problems to the extent that it is known to be notoriously difficult, probably because of the nature and quality of available data, the numerous variables that are hypothesized to affect water demand (Arbués et al. 2003), and the multiplicity of forecast horizons and periodicities involved. These characteristics have engendered a plethora of studies in an attempt to improve forecast reliability. Despite this effort, water demand forecasting practice, undertaken by utilities and their consultants, currently differ extensively with respect to the methods and models used. Arbués et al. (2003) have reviewed the literature on residential water demand modeling, in which the focus was on cross-sectional data for pricing purposes. Although cross sectional models are important for identifying determinants of water demand for pricing and/or demand management purposes, they are inadequate for the kind of planning decisions that utilities make when future demand is uncertain. We therefore differentiate our study by focusing on methods and models for time series data.

As a contribution towards improving future forecasts, this paper surveys a sample of the recent literature on water demand forecasting as published from 2000 to 2010. The intended purpose is to provide a guide to the literature for practitioners looking for guidance on adopting methods and models to support forecast-based decision making, and for researchers seeking to extend the current state-of-the-art in the field. In concert with this purpose, we focus attention on four main objectives. Firstly, we provide a framework for demand forecasting as a background to understanding the literature. In this paper, we establish the basis for water demand forecasting by characterizing the water utility decision problems that create this need according to the level of planning, forecast horizon, and forecast periodicity. We also review the forecast variables, determinants used, and the predominant measures of forecast error.

Secondly, we create an inventory of water demand forecasting methods and models published from 2000 to 2010. We differentiate between forecast models, methods, and approaches for clarity. We will refer to forecast models when discussing specific mathematical formulations for predicting demand from time series data. Forecast method and technique will be used interchangeably and will refer to the class of qualitative and quantitative means by which forecast models are formulated and estimated. We use the term forecast approach to refer to a collection of methods, tools, and processes for estimating future water demand. By limiting the study period to 2000–2010, we focus attention on the recent literature, with the assumption that any useful method and model in the distant past would either have been extended or applied within this time frame. In such a study, it is impossible to list all published sources on the

[1]Doctoral Student, Engineering Management and Systems Engineering, School of Engineering and Applied Sciences, George Washington Univ., 1776 G St. NW, Suite 101, Washington, DC 20052 (corresponding author). E-mail: eadonkor@gwu.edu; eadonkor@gwmail.gwu.edu

[2]Professor, Engineering Management and Systems Engineering, School of Engineering and Applied Sciences, George Washington Univ., 1776 G St. NW, Suite 101, Washington, DC 20052.

[3]Professor, Decision Sciences, School of Business, George Washington Univ., 2201 G St. NW, Funger Hall, Suite 415C, Washington, DC 20052.

[4]Director of Federal Relations, American Water Works Association, 1300 Eye St. NW #701W, Washington, DC 20005.

Note. This manuscript was submitted on March 27, 2012; approved on September 10, 2012; published online on September 13, 2012. Discussion period open until July 1, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 140, No. 2, February 1, 2014. © ASCE, ISSN 0733-9496/2014/2-146-159/$25.00.

J. Water Resour. Plann. Manage. 2014.140:146-159.

subject, and hence some good sources could have been overlooked. However, we do not find this possibility a limitation on the general sense of the nature of water demand forecasting that we glean from the sample we use. Sometimes we cite literature outside the time frame and the water domain, if we find them illustrative of the issue under discussion.

Thirdly, we synthesize the literature, classifying it by method and forecast periodicity, to identify where researchers and practitioners have placed emphasis on water demand forecasting in recent times. We find it useful to emphasize the difference between forecast horizon, which refers to how far into the future demand is to be predicted, and forecast periodicity, which refers to the time span between consecutive forecasts, for example hourly or monthly, because a combination of forecast horizon and forecast periodicity determines the number of forecasts and may thus influence forecast accuracy.

Our final objective is to draw attention to what we see as weaknesses, problems, or gaps in urban water demand forecasting and to suggest how these can be addressed.

Framework for Water Demand Forecasting

Basis for Water Demand Forecasting

Planning for decision making forms the basis for forecasting in the water sector. To this end, a set of water demand forecasting literature differentiates forecast practice by the level of planning associated with the forecast (Gardiner and Herrington 1990), or in accordance with the forecast horizon (Billings and Jones 2008). In terms of planning level, all water demand forecasting exercises can be used for strategic, tactical, or operational decision making. These concern decisions for capacity expansion, investment planning and system operation, management, and optimization, respectively. In terms of forecast horizon, water demand forecasting can be categorized as either long-term, medium-term, or short-term, with these horizons being reflective of the planning levels, respectively (Alvisi et al. 2007; Ghiassi et al. 2008; Jain et al. 2001).

No generally accepted time frame exists for these horizons. Billings and Jones (2008) contain different time horizons on what constitutes long-term, medium-term, and short-term forecasts. One such definition classifies forecasts spanning more than two years as long-term, those from three months to less than two years as medium-term, and forecasts for one to three months as short-term. This contrasts with Gardiner and Herrington (1990), in which these categories are classified as annual forecasts for 10 years or more (long-term), annual forecasts for 1 to less than 10 years (medium-term), and hourly to monthly forecasts up to a year (short-term). In terms of reported research, Ghiassi et al. (2008) prepared monthly demand forecasts for two years, weekly demand forecasts for six months and daily demand forecasts for two weeks, and characterize these as long-term, medium-term, and short-term, respectively. In Table 1, we follow Gardiner and Herrington (1990) in categorizing the forecast horizon by planning level and summarize how these relate to the decision problem and the forecast periodicity.

Forecast Variables and Determinants

Apart from the various planning levels and horizons that tend to complicate urban water demand forecasting, the forecast variable of interest and the determinants of water demand (exogenous variables) are two features that add to the complexity. In decreasing order of frequency, results of a survey by the American Water Works Association (AWWA) portray the kind of variables that are of interest to water utilities in urban water demand forecasting: peak day (73.9%), daily total system demand (65.9%), monthly total system demand (65.6%), annual per capita demand (65.4%), annual demand by customer class (58.0%), and revenue (57.9%) (Billings and Jones 2008). These results clearly show that urban water demand forecasting can involve different variables measured at different periodicities. Our annotated reference in Table 2 provides additional evidence to support this fact. Appreciating this fact is important because the question regarding which method to use for urban water demand forecasting cannot be adequately answered without specifying the forecast variable, its periodicity, and the horizon. For example, interest in the annual variation of per capita demand makes the related variable a candidate for models amenable to medium to long-term forecasting. Forecasting this variable in the long-term (Burney et al. 2001; Polebitski and Palmer 2010) may require completely different determinants as compared with its short-term equivalent (Zhou et al. 2000) (see Table 2). Similarly, because total system demand can be measured on an hourly up to a yearly basis, different forecasting models may be required for its forecast, for a given horizon. Total system demand may sometimes include leakages and other system flows that cannot be considered as real demand. Whether it is more useful to model real demand or total system demand (including leakages) should depend on the purpose of the forecast. Whereas it may seem appropriate to model real demand for purposes of forecasting revenue, system operations and optimization may require a forecast of total system demand.

In the drinking water industry, many variables are considered influential in determining water demand. These range from socioeconomic to various derivatives of weather-related variables. Examples of these weather-related variables and how they are used can be found in Coomes et al. (2010) and Brekke et al. (2002). Although "A good understanding of the factors influencing demand and reliable estimates of the parameters describing demand behavior and consumption patterns are prerequisites [to a good forecast]" (Burney et al. 2001), the enormity of these variables can create frustration for water utility mangers. As an illustration of the size and variability of the variables that can be considered, we refer to a report prepared for the Water Research Foundation by Coomes et al. (2010), in which the authors tested the effect of 26 variables on average daily water use for 293 residential customers of the Louisville Water Company. The availability and choice of these independent variables can also influence the forecasting models used. For instance, whereas population projections and per capita demand are the drivers for unit rate models, these have no consideration when exponential smoothing or Box-Jenkins models are formulated.

Measures of Forecast Error

Because it provides a means of measuring forecast accuracy, forecast errors play a significant role in the selection of appropriate

Table 1. Relationship among Planning Level, Water Utility Decision Problems, and Forecast Attributes

Planning level	Decision problem	Forecast horizon	Forecast periodicity
Operational	System operation management and optimization	Short-term (less than 1 year)	Hourly daily weekly monthly
Tactical	Revenue forecast; investment planning; staging system improvement	Medium-term (1–10 years)	Monthly annual
Strategic	Capacity expansion	Long-term (more than 10 years)	Annual

J. Water Resour. Plann. Manage. 2014.140:146-159.

Table 2. Annotated Reference of Water Demand Forecasting Methods of Selected Journal Papers from 2000 to 2010

Reference	Purpose	Location	Variable/period	Determinants	Horizon
Univariate time-series-only papers					
Alhumoud (2008)	Uses correlation to assess the relationship between water consumption and its determinants and provides a descriptive model of annual water consumption using AR(1) model. To assist Govt. Reform subsidy policy and capacity planning	Kuwait	Total annual residential water consumption	Y_{t-1}	Long-term: 22 annual forecasts
Regression-only papers					
Brekke et al. (2002)	To understand the determinants of demand to enable demand management and supply planning	Lake-haven Utility District	Monthly demand (single family residential)	Time; seasonal dummies; derivatives of weather, price	None
Lee et al. (2010)	Uses Bayesian moment entropy to estimate the regression relationship between water demand and population density as a means of informing conservation policy and infrastructure planning	Phoenix, Arizona	Log water demand by census tract	Population density	Long-term: 26 annual forecasts
Polebitski and Palmer [(2010) see reference for complete list of variables]	Used three regression models to explain observed temporal and spatial variation in residential water demand for regional infrastructure expansion, water resources management and an understanding of the determinants of water demand	Puget Sound area, Seattle, Washington	Bimonthly single family per capita consumption	Density, building size, lot size, household size, income, price, temp, rain, drought dummies	Long-term: 12 bimonthly forecasts
Scenario-based papers					
Burney et al. (2001)	Econometric modeling of total annual water demand under eight scenarios to determine capacity expansion and associated investment costs. Considers the concept of co-integration	Kuwait	Total annual water demand, per capita demand	Population; per capita demand; price	Long-term: 26 annual forecasts
Goodchild (2003)	Models per capita demand as a function of weather variables to assess impact of climate change scenarios on water demand	Essex Area, U.K.	Daily per capita demand	Max temp; days since 2 mm rainfall; evapotranspiration; temp dummy	Long-term: forecast for 2020 from a base year of 2000
Wei et al. (2010)	Uses econometric models to forecast annual industrial water demand under different scenarios	Beijing, China	Annual demand	Employment; inventory cost; time; industrial value	Long-term: 2008 and 2015 base year: 2007
Williamson et al. (2002)	Forecast household water demand under 10 scenarios for purposes of informing conservation policy	Yorkshire Water Region, U.K.	Household water demand	Residents; bedrooms; appliance ownership; property type	Long-term: forecast for 1991–2025
DSS-based papers					
Feng et al. (2007)	For purposes of assessing impact of future demand on available water resources, uses a DSS to prepare forecasts of water use under various scenarios for 2010 and 2020	China	Total annual water demand	Not stated	Long-term: forecasts for 2010 and 2020 from 1997
Froukh [(2001) see reference for complete list of variables]	Developed a DSS which determines which of four forecasting approaches: simple linear regression; econometric model; microdemand and classified consumption. Incorporates future demand management strategies (metering, rationing, pressure reduction, and pricing)	Applied to Swindon Area of Thames Water Utility, U.K.	Daily demand	Econometric model: income; occupancy rate; price; rainfall; temperature	Long-term: specific horizon not stated

J. Water Resour. Plann. Manage. 2014.140:146-159.

Table 2. (*Continued.*)

Reference	Purpose	Location	Variable/period	Determinants	Horizon
Jain and Ormsbee (2001)	Describes a draught management DSS with a water demand forecasting module that enables short-term forecasts (5 days ahead) to be made using regression, time series analysis or artificial neural networks	Lexington, Kentucky	Daily demand	Water demand of previous days, total rainfall, max air temperature, sunshine hours	Short-term 5 days
Levin et al. (2006)	Demonstrates the use of a demand-side management least-cost planning decision support system (DSS), based on an end-use model, for estimating long-term water demand incorporating water demand management measures	Areas under San Francisco Public Utilities Commission	Daily water use per customer category. Includes population and employment forecasts	Water use per category population conservation measures	Long-term: forecasts for 2030 from a base year of 2001
Mohamed and Al-Mualla (2010b)	Uses the constant rate model in the IWR-MAIN DSS to forecast water demand for 20 years and 30 years horizon to inform capacity planning decisions	United Arab Emirates	Monthly per capita water use per subsector	Population, temperature, rainfall, seb-sector size	Long-term: forecasts for 2020 and 2030 base year: 2002
Mohamed and Al-Mualla (2010a)	Uses the specify forecasting model option in the IWR-MAIN DSS to forecast water demand for 25 years horizon. Prepares forecasts under different scenarios incorporating different mixes of dependent variables	United Arab Emirates	Monthly per capita use for each subsector	Population, temperature, rainfall, seb-sector size	Long-term: forecasts for 2020 and 2030 base year: 2002
Neural nets and regression					
Adamowski and Karapataki (2010)	Motivated by the absence of conjugate gradient Powell-Beale (CGPB) and resilient back propagation (RP) ANN models, the authors set out to compare the forecasting ability of these models as compared with the Levenberg-Marquardt(LM) ANN method and a multiple linear regression model. Concludes that the LM model is best	Athalassa and Public Garden, Cyrus	Peak weekly demand	Peak demand for previous week; Max. Temp. For current and previous weeks	Short-term: horizon not stated
Cutore et al. (2008)	In an attempt to account for forecast uncertainty, the paper uses a shuffled complex evolution metropolis (SCEM-UA) algorithm to calibrate the parameters of an artificial neural network (ANN) for daily water consumption, and compared fore cast results with those obtained from a Bayesian ANN, regression and adaptive neuro-fuzzy inference system (ANFIS). Obtained similar performance among models	Catania, Italy	Daily water demand	Demand of previous day; working day dummy; day of week	Short-term: forecast horizon not stated
Firat et al. (2009)	Compares the performance of generalized regression neural networks (GRNN); feed forward neural networks (FF); radial basis neural networks (RB); multiple linear regression (MLR). Concludes that GRNN is the best	Izmir, Turkey	Monthly water consumption	Monthly bill; population; households; GNP; inflation; temp; rainfall; humidity	Not stated. Used 24 records for validation
Herrera et al. (2010)	Compares four methods: ANN, projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), random forest (RF) and support vector regression (SVR). Concludes that SVR was best	Spain	Hourly demand	Up to 2 lags of hrly demand; demand for previous week; temp; wind velocity; pressure; rain	Short-term: uses 1 week of data for validation.

Table 2. (*Continued.*)

Reference	Purpose	Location	Variable/period	Determinants	Horizon
Jentgen et al. (2007)	Identifies the best forecasting methods that can be used to optimize pumping schedules for minimizing energy costs. Compares heuristics, regression and ANN methods. Concludes that performance is about equal among methods, although ANN is better for hourly forecast and Regression is better for daily forecasts	Several utilities in the US	Hourly demand Daily demand	Temperature, rainfall and Lag 1 of demand	Short-term 24 h for hourly demand 7 days for daily demand
Neural nets and time series					
Ghiassi et al. (2008)	Compares the performance of dynamic artificial neural network(DAN2); ARIMA and feed forward back propagation neural network (FFBP), to inform schedule pumping to minimize electric cost, maintenance scheduling, system expansion, maintenance and decisions and strategies for purchasing. Concludes that DAN2 outperforms the others	San Jose Water Company, California	Multiple periodicities: monthly; weekly; daily and hourly demand	Univariate demand series; multivariate hourly model includes temp	24 months 26 weeks 14 days 48 h
Neural nets, regression, and time series					
Bougadis et al. (2005)	Compares relative performance of models selected from simple linear regression (SLR) multiple linear regression (MLR), univariate time series and 3 artificial neural networks (ANN), for purposes of identifying a usable model for sizing and staging facilities for development Concludes that ANN models were the best	Ottawa, Ontario, Canada	Peak weekly demand	Various lags of peak demand, temp and rainfall	Short-term
Jain and Ormsbee (2002)	Compared the performance of regression, univarite time series and artificial intelligence(AI) models, on daily water consumption, to inform optimal operating policy. Concludes that AI models were the best	Lexington, Kentucky	Daily demand	Various lags of demand; max temp; rainfall; rainfall occurrence	Short-term
Jain et al. (2001)	Generally compares 5 regression models, 2 univarite time series models and 6 artificial neural networks (ANN) models as methods for forecasting weekly water demand. Concludes that the complex ANN models out-performed the others	Indian Institute of Technology, Kanpur	Weekly demand	Various lags of demand; max temp; rainfall; rainfall occurrence	Short term
Composite forecasts					
Altunkaynak et al. (2005)	Decomposes monthly water demand series into trend and stochastic components and uses Fuzzy logic and AR(3) to forecasts the latter. Claims that Fuzzy logic is better than AR	Istanbul, Turkey	Monthly water demand	3 lads of monthly water demand	Not stated
Alvisi et al. (2007)	Motivated by hourly pumping requirements, develops an adaptive model for real-time, near optimal control of distribution networks, by decomposing daily demand into season, trend and residual components. Uses fourier series for seasonal component and AR(1) for residuals	Castelfranco Emilia, Italy	Daily demand hourly demand	Historical data	Short-term: 1–24 h forecast horizon

Table 2. (*Continued.*)

Reference	Purpose	Location	Variable/period	Determinants	Horizon
Aly and Wanakule (2004)	For operational purposes, prepared daily forecasts using regression by adjusting monthly forecasts obtained from a Holt-Winters multiplicative exponential smoothing model	Tampa Bay water, Florida	Daily and monthly demand	Prep, temp, humidity, lagged demand	Short-term: 6 days ahead
Caiado (2010)	Investigates if combining forecasts from Holt-Winters, ARIMA, and GARCH models improves forecast accuracy. Uses simple to weighted averages to combine forecasts	Spain	Daily water cons	Univariate daily cons	Short-term: 1–7 days ahead
Gato et al. (2007a)	Models daily use a function of base use and seasonal where base use is dependent on climatic factors	East Doncaster, Victoria, Australia	Daily demand	Thresholds of temp and rainfall; day of the week	Short-term: forecast horizon not stated
Gato et al. (2007b)	Models daily use a function of base use and seasonal use where base use is independent of climatic factors. Uses regression for base use and fourier series for seasonal use	East Doncaster, Victoria, Australia	Daily demand	Thresholds of temp and rainfall; day of the week	Short-term: forecast horizon not stated
Wang et al. (2009)	Compares the performance of forecast obtained from a weighted average of a regression and a back-propagation ANN model, and those from these models separately. Concludes that the combined forecast was better	China	Annual domestic and industrial water demand	Historical demand, time	Long-term: forecasts for 2005, 2010 and 2030
Wu and Zhou (2010)	Develops a forecast model to assist in long-term planning, by decomposing annual demand into trend and stochastic components. Compares results for four models: Regression for both trend and stochastic components, ANN for both trend and stochastic components, and regression and ANN interchanged for the two components, respectively. Concludes that combining forecasts from disaggregated models yields better results	Dalian, China	Annual demand	Population, GDP; Avg. Annual temp; Greenery coverage; Lag 1 of demand	Long-term: forecasts for 2010, 2020
Zhou et al. (2000)	Decomposes per capita water consumption into base, seasonal, climate and residual components. Models base and climate components with regression, seasonal with fourier, and residual with AR (p), where p takes values of 1, 2, and 7	Melbourne, Australia	Daily per capita cons	Max temp; Precip; Evapotration; prep index; Rain; Post precipitation. days	Short-term
Zhou et al. (2002)	Models hourly demand by adjusting daily demand obtained from decomposition and modeling with separate methods	Melbourne, Australia	Daily demand hourly demand		Short-term: 24 h ahead

J. Water Resour. Plann. Manage. 2014.140:146-159.

models and in providing insights in recommending changes to existing models to reduce deviations in future forecasts. The accuracy of forecasts is evaluated by comparing them with observed demand. The general approach to model selection is to consider competing forecasting models in a sequence of steps: (1) divide the data set into an estimation period and a hold-out period; (2) use the estimation period to model demand; (3) evaluate the accuracy of the models by comparing the forecasts with the observed values for both the estimation period and the hold-out period; and (4) select the best model on the basis of its performance, as measured by any of the popular error measures specified in Eqs. (1)–(4). For N time periods, Y_t and $\hat{Y}_t$ represent the actual observation and the forecast value at time t, respectively, and $Y_t - \hat{Y}_t$ measure the forecast error:

$$\text{Mean Absolute Error (MAE)} = \frac{1}{N}\sum_{t=1}^{N}\left|Y_t - \hat{Y}_t\right| \tag{1}$$

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{2}$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{N}\sum_{t=1}^{N}(Y_t - \hat{Y}_t)^2 \tag{3}$$

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\text{MSE}} \tag{4}$$

When comparing forecast performance for a given sample data, the model with the least value of a chosen error measure is deemed the most accurate. These functions measure different characteristics of forecast error and hence sometimes model ranking could be different, depending on which error measure is used. In Cutore et al. (2008), for example, root mean squared error (RMSE) and a less known error measure in demand forecasting, the Nash-Sutcliffe index of efficiency (EI), ranked four alternative models differently. The mean absolute error (MAE) prefers models with the least deviations on average. The mean absolute percentage error (MAPE) is similar but has no units. The mean squared error (MSE) and its associated RMSE penalizes models that have large deviations and hence is used to select models that fit the data well. For an example of how both MAPE and RMSE were used in water demand forecasting, see Altunkaynak et al. (2005).

Point forecasts are not accurate most of the time and hence it is better to speak of their reliability which can be (theoretically) determined from predicted forecast error once a particular forecast model is selected. However, this prediction assumes that the selected model remains true and stable over the forecast horizon. Defining forecast reliability calls for establishing threshold values on the error measures to judge how reliable forecasting has been for a given utility. The nonexistence of such threshold values makes it problematic for comparing the performance of models across utilities with different capacities. Of the various metrics, MAPE might be the only one for which a threshold value can be used to compare forecasting performance among utilities because it is independent of system capacity.

An Inventory of Water Demand Forecasting Methods and Models

In what follows, we present an inventory of water demand forecasting methods and models. An annotated reference list of these methods and models, as appearing in the water demand forecasting literature in our sample is presented in Table 2, in which the purpose of the study, forecast variable and periodicity, determinants used, and forecast horizon are identified.

Forecasting with Qualitative Methods, and Unit Water Demand Analysis

Judgmental or qualitative methods include the use of heuristics or rule-based methods to forecast the value of a variable of interest. According to Gardiner and Herrington (1990), "... [these] approaches rely on the experience of an individual or, less likely, a group, and may be either entirely subjective in nature or a modification of more objective results derived from other approaches." In practice, the use of qualitative methods is necessitated by a desire for rudimentary forecasts for purposes of simplicity. Billings and Jones (2008) describe the method as applied by water utilities, whereas Jentgen et al. (2007) report specific cases where utilities in the United States (Jacksonville Electric Authority, San Diego Water Department, Colorado Springs Utilities, and Las Vegas Valley Water District) use heuristics, regression, and neural networks to prepare short-term forecasts for optimizing pumping schedules.

Unit water consumption is a variable that is extensively forecasted using rudimentary methods. Demand forecasting based on what Brekke et al. (2002) terms "unit water demand analysis" employs the consumption per unit of a customer category, for example, per capita water demand, and the number of units of that category, such as population/size of domestic customers, to forecast water demand. In this case, the demand for a given future time period t (Q_t) is computed by taking the product of the unit consumption (q_t) and the number of units (N_t). Where a utility's customer mix includes other categories, the method requires disaggregating demand by customer segment, preparing forecasts for each, and then adding these forecasts to generate the total [see Hazen, Sawyer, and PMCL (2004) for a comprehensive case]. The mathematical expression for this sectoral forecasting model is presented in Eq. (5), where for C customer categories indexed by i, the demand forecast at a future time t is given by

$$\sum_{i=1}^{C} Q_{i,t} = \sum_{i=1}^{C} q_{i,t} * N_{i,t} \tag{5}$$

Although forecasting unit rates by rudimentary methods is not feature extensively in the literature, Billings and Jones (2008) and Jentgen et al. (2007) have noted that in practice, it is the simplest approach used by most utilities. All that is required is to estimate $q_{i,t}$ and $N_{i,t}$ to obtain a forecast. As an example, the Washington metropolitan area (WMA) water studies have consistently adopted a "unit use coefficient approach ... [since] 1990 ... [because] ... it is a transparent and easily understandable method ... and was judged to provide the right balance between data needs and accuracy" (Hagen et al. 2005). It is therefore not surprising to observe that 65.4% of the utilities surveyed in Billings and Jones (2008) spend resources in forecasting per capita water demand. The reliability of these forecasts is questionable when simple rules of thumb or expert judgment, instead of empirical analysis, form the basis for estimating $q_{i,t}$ and $N_{i,t}$ for each customer category. In annual forecasts, this results in trend lines with limited variability, as compared with the historical profile of the series.

Forecasting through Univariate Time Series Analysis

Time series models, or what is technically known as extrapolation forecasts in the water industry (Billings and Jones 2008; Gardiner

J. Water Resour. Plann. Manage. 2014.140:146-159.

and Herrington 1990), forecast future water demand on the basis of past observations and associated error terms. This class of models does not account for the effect of exogenous variables such as weather or price, but rather rely on the assumption that past trends will be repeated in the future. Their failure to take into consideration the effects of changes in demographic, economic and technological variables, and water demand management strategies (such as public awareness campaigns and/or price adjustments) in influencing future water demand is the main criticism.

Moving Average and Exponential Smoothing Models

Deterministic moving averages and exponential smoothing models are simple time series models whose identification can be facilitated by understanding the eight generic time series profiles suggested in Mun (2010). Each of these profiles, depicted in Fig. 1, can have one or more of three main components: a level component (L_t), a trend component (T_t), and a seasonal component (S_t), and the underlying mathematical model for each depends on which components are present and the nature of the variation (constant or changing), observable in the series. Related models are formulated as Eqs. (6)–(12), where m = number of periods in the forecast lead-time; Y_t = observed value of demand at time t; $\hat{Y}_t$ = forecast of demand for period t; and $\hat{Y}_{t+m}$ = forecast of demand for m periods ahead from period t. Besides these eight profiles described in this paper, Gardner (2006) presents an extensive list, in a state-of-the art review of exponential smoothing models, where linkages with discrete-time stochastic process models have been identified. In addition, stochastic versions of these models are possible, through the inclusion of a random error term, in what is known as state-space models (Gardner 2006; Hyndman et al. 2005). Although we do not specify any specific method for initializing the models formulated in Eqs. (6)–(13), all exponential smoothing models require an initial forecast. An extensive coverage of alternative methods for doing so can be found in Taylor (1981).

There are four profiles that exhibit no seasonal variation, as depicted in Figs. 1(a–d). Where the variation is constant and the series has no trend [Fig. 1(a)], it is best modeled by a single moving average as formulated in Eq. (6), where k = number of historical periods used in calculating the moving average:

$$\hat{Y}_t = (Y_{t-1} + Y_{t-2} + \cdots + Y_{t-k})/k \tag{6}$$

In the case where the variation increases or decreases with time as in Fig. 1(b), the m period ahead forecast is modeled with a single exponential smoothing function as in Eq. (7):

$$\hat{Y}_{t+m} = L_t \tag{7}$$

where $L_t = \alpha Y_t + (1-\alpha)L_{t-1}$; and $\alpha \in (0,1)$ = smoothing parameter for the level of the series.

In Fig. 1(c), the series has a trend and a constant but nonseasonal variation around the trend. This is modeled as a double moving average for which the m period ahead forecast is given by Eq. (8):

$$\hat{Y}_{t+m} = L_t + mT_t \tag{8}$$

where $L_t = 2M_t - D_t$; $T_t = 2(M_t - D_t)/(k-1)$; $M_t = (Y_t + Y_{t-1} + \cdots + Y_{t-k+1})/k$; $D_t = (M_t + M_{t-1} + \cdots + M_{t-k+1})/k$; M_t = 1st or single moving average; and D_t = 2nd or double moving average.

If the nonseasonal profile is characterized by a trend and a changing variation as depicted in Fig. 1(d), the forecast is also obtained by using Eq. (8), in what is termed a double exponential smoothing model. In this case

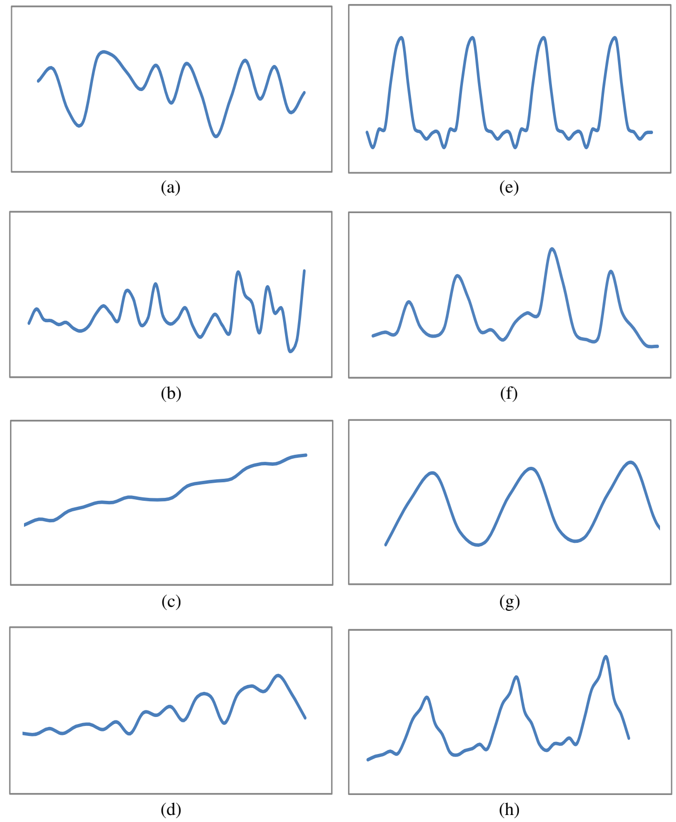

Fig. 1. Generic time series profiles: (a) no trend; no seasonality; constant variation; (b) no trend; no seasonality; changing variation; (c) trend; no seasonality; constant variation; (d) trend; no seasonality; changing variation; (e) seasonality; no trend; constant seasonal variation; (f) seasonality; no trend; changing seasonal variation; (g) seasonality; trend; constant seasonal variation; (h) seasonality; trend; changing seasonal variation

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$
$$\gamma \in [0,1] = \text{smoothing parameter for the trend}$$

The remaining four profiles exhibit seasonal variation: a constant seasonal variation without a trend [Fig. 1(e)], an increasing seasonal variation without a trend [Fig. 1(f)], a constant seasonal variation with a trend [Fig. 1(g)], and an increasing seasonal variation with a linear trend [Fig. 1(h)]. The mathematical formulations of the underlying exponential smoothing models are termed as follows:

Seasonal additive [Fig. 1(e)], modeled in Eq. (9)

$$\hat{Y}_{t+m} = L_t + S_{t+m-p} \tag{9}$$

where $L_t = \alpha(Y_t - S_{t-p}) + (1-\alpha)L_{t-1}$; and $S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-p}$.

Seasonal multiplicative [Fig. 1(f)], modeled in Eq. (10)

$$\hat{Y}_{t+m} = L_t S_{t+m-p} \tag{10}$$

where $L_t = \alpha(Y_t/S_{t-p}) + (1-\alpha)L_{t-1}$; and $S_t = \delta(Y_t/L_t) + (1-\delta)S_{t-p}$.

Holt-Winters additive [Fig. 1(g)], modeled in Eq. (11)

$$\hat{Y}_{t+m} = L_t + mT_t + S_{t+m-p} \tag{11}$$

where $L_t = \alpha(Y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$; $T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)L_{t-1}$; and $S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-p}$.

Holt-Winters multiplicative [Fig. 1(h)], modeled in Eq. (12)

$$\hat{Y}_{t+m} = (L_t + mT_t)S_{t+m-p} \tag{12}$$

where $L_t = \alpha(Y_t/S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$; $T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$; and $S_t = \delta(Y_t/L_t) + (1-\delta)S_{t-p}$.

In the system of equations specified in Eq. (9)–(12), $\delta \in (0, 1)$ = smoothing parameter for seasonal indices and p = number of periods in the seasonal cycle ($p = 4$ for quarterly data and 12 for monthly data). It is required that $p \geq m$. Water demand forecasting applications of the Holt-Winters seasonal multiplicative model can be found in Aly and Wanakule (2004) and Caiado (2010).

In practice, time-dependent water demand data do not have, although may contain semblances of, the nice profiles depicted in Fig. 1. A typical example can be found in Brekke et al. (2002) and in Fig. 2(a). Under such circumstances, the time series can be decomposed into its components, level (random/persistence), trend, and seasonal, as shown in Fig. 2, and then composite forecasting methods, discussed latter, can be used.

Stochastic Process Models

Sometimes, time series data can exhibit more complex profiles for which the eight generic exponential smoothing models cease to be adequate. Stochastic process models, which can be formulated in discrete-time or continuous-time, are more advanced alternatives that can be used to model these complex profiles. These models are mathematical formulations of processes that obey specific probabilistic and statistical laws and thus, their simulated forecasts result in a series of outcomes for each period over a given time span (Billings and Jones 2008). The value of using stochastic process models to forecast demand lies in the ability to quantify estimates of the level of uncertainty associated with forecast values. It is different from scenario-based approaches in that a large number of possible paths can result from the underlying process. Billings and Jones (2008) refer to this method as risk simulation and presents an example on five possible paths of an AR(1) process for population, with a mean growth rate of 2.4% and a range of 0.5–5%. Thousands of these paths can be generated, allowing for a probability distribution of the demand variable to be assessed at each point in time, and hence, an assessment of the level of risk associated with decisions based on the forecast.

Discrete-time models, typically known as Box-Jenkins models, are formulated in Eqs. (13)–(16), where ϕ = autoregressive or damping parameter; θ = moving average parameter; μ = mean value of the process; and ϵ_t = forecast error at time t. In this case, the error term (ϵ_t) is generally assumed to be independent and identically distributed (i.i.d) Normal $(0, \sigma)$ random variables (white noise), and the coefficients ϕ and θ take values with defined restrictions for stationary time series. Eqs. (13)–(15) model processes that follow an autoregressive process of order p, AR(p); a moving-average process of order q, MA(q); and an autoregressive moving-average process of order (p,q) [ARMA(p,q)], respectively. Where the time series has to be differenced by order d to make it stationary, a generalized formulation, using the stationary AR backshift operator $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and the invertible MA backshift operator $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_p B^q)$, can be used to represent an autoregressive integrated moving average process of order (p,d,q), ARIMA (p,d,q), as shown in Eq. (16):

$$Y_t = \mu + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t \tag{13}$$

$$Y_t = \mu + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \tag{14}$$

$$Y_t = \mu + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} \tag{15}$$

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\epsilon_t \tag{16}$$

Model formulation requires the creation and understanding of what is known as the autocorrelation function (ACF) and partial autocorrelation function (PACF) to specify the number of historical/lagged components of demand. This is amply verified by using the

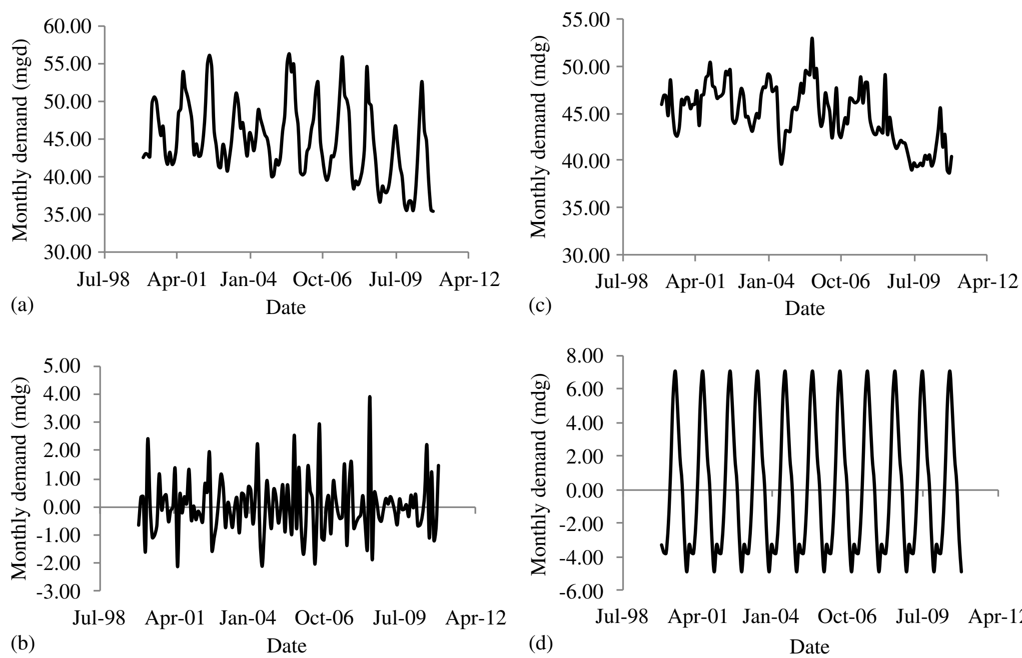

Fig. 2. Time series decomposition: (a) original demand profile; (b) base, level, or random component; (c) trend and cycle; (d) seasonal component

J. Water Resour. Plann. Manage. 2014.140:146-159.

Ljung-Box white noise test as a diagnostic tool for model adequacy. In our sample of papers, Alhumoud (2008), where the author uses an AR (1) process to model freshwater demand in Kuwait, is the only paper that considered a univariate time series model singularly. Other examples of Box-Jenkins models can be found, in comparison with other methods, in Bougadis et al. (2005), Caiado (2010), Ghiassi et al. (2008), Jain and Ormsbee (2002), and Jain et al. (2001), although their treatment is not as comprehensive as in Miaou (1990).

Forecasting with Time-Series Regression Models

The class of model we define as time-series regression models are those that produce forecasts on the basis of the relationship between water demand and its determinants Polebitski and Palmer (2010). This feature of accounting for the effect of exogenous variables differentiates this class from models obtained through univariate time series analysis. A general model formulation is presented in Eq. (17), where Y_t is water demand at time t:

$$Y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{j=1}^{v} \vartheta_j t^j + \sum_{m=1}^{s} \varphi_m s_{m,t} + \epsilon_t \qquad (17)$$

In Eq. (17) β_o is the regression intercept and i is an index of the ith independent/exogenous variable, for a total of n such variables. β_i and $x_{i,t}$ represent the coefficient and observed value of the ith independent variable, respectively. Trend (both linear and nonliner) and seasonality are accounted for by t and s, respectively, with the associated coefficients being ϑ_j and φ_m. The parameter j accounts for polynomial terms in the trend component and the seasonal index m takes maximum values of 6, 3, and 11 for daily, quarterly, and monthly seasonal variations, respectively.

For $n = 1$, $\vartheta_j = 0$, and $\varphi_m = 0$, Eq. (17) becomes a simple linear regression. When $\vartheta_j \neq 0$, time and its polynomial derivatives enter the regression equation, while constant seasonal variation is accommodated when $\varphi_m \neq 0$. Nonlinearities in the water demand time series and its determinants can be modeled by transforming Y_t and the $x_{i,t}$ variables using natural logs (Billings and Jones 2008, p. 258). This is useful for two reasons: (1) to stabilize the variance (because water consumption tends to be more variable when its level is high); and (2) to prevent the model from giving negative predicted values. Brekke et al. (2002) exemplifies the case where a mix of exogenous variables, time and its polynomials, and indicator seasonal variables, are used in a mixed time-series regression models.

In Eq. (17), the residuals (ϵ_t) are assumed to be free of autocorrelation. Regression models based on time-series data are, however, known for the serial correlation of the error terms. When this happens, the appropriate ARMA model must be used to model the structure of the error terms for subsequent integration with the parent model. For example, Burney et al. (2001) uses log transformation of Eq. (17) and models the error term with an AR (1) process. However, serial correlation of the error terms can result in correlation structures more complicated than the usual 1st order autocorrelation assumed, and in some cases the variance of the error terms are not constant. In this case of heteroscedasticity, the appropriate autoregressive conditional heteroscedasticity and generalized autoregressive conditional heteroscedasticity (ARCH/GARCH) model should be used, as demonstrated in Caiado (2010). The key lies in examining the ACF and PACF, and the Ljung-Box Q statistic, to determine an appropriate model for the residuals.

Contrary to the assumptions informing the generalized model described in Eq. (17), demand at time t can be influenced by moving average and autoregressive terms, in addition to exogenous variables and their autoregressive terms. Such models are term ARMAX(p,q,b), and can be formulated as in Eq. (18), where a single exogenous variable is assumed. These are estimated as transfer function models, where the model structure is obtained by examining what is known as the cross-correlation function (Wei 2005)

$$Y_t = \mu + \sum_{k=1}^{p} \varphi_k Y_{t-k} + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} + \sum_{k=0}^{b} \beta_k x_{t-k} \qquad (18)$$

Identification of the structure of ARMAX(p,q,b) models in the water demand literature is currently ad hoc. The literature is replete with cases where water demand at time t is modeled as a function of a selection of its previous values and those of its exogenous inputs variables. Examples of this practice can be found in Jain and Ormsbee (2001), in which the use of independent variables such as water demand of previous days, total rainfall, and daily maximum air temperature were endorsed on the grounds of simplicity, and in Adamowski and Karapataki (2010), in which the authors based the choice of model determinants on correlations coefficients, leading to the selection of variables such as peak water demand for the previous week, maximum temperature for the current and previous weeks, maximum temperature from two weeks ago. Such simple approaches to model identification calls to questions whether model performance would have been improved with a more rigorous identification and estimation procedure, as in SAS, R or Matlab.

Forecasting with Scenario-Based Approaches and Decision Support Systems

Scenario-based approaches are basically regression models that determine the level of demand for long-term forecasts given specific scenarios. They are used when there is a need to account for uncertainty in demand forecasts brought on by a limited number of discrete combinations of the independent variables. The idea is to determine the effect on water demand of various future scenarios of the determinants. Examples of this approach for urban water demand forecasting can be found in Burney et al. (2001), Polebitski et al. (2011), Wei et al. (2010), and Williamson et al. (2002). Strzepek et al. (1999) presents a similar approach for linking climate change scenarios with water planning for agriculture.

Sometimes, the need for automating scenario-based approaches results in the development of custom-made decision support programs. The forecasts obtained from such programs are normally driven by system-defined models derived from different methods (Feng et al. 2007; Froukh 2001; Jain and Ormsbee 2001), allowing the decision maker to select the most appropriate combination of methods that satisfies assumptions concerning future scenarios. Currently, IWR-MAIN, developed by the U.S. Army Corps of Engineers' Institute for Water Resources, and Demand-Side Management Least-Cost Planning Decision Support System (DSS), created by Maddaus Water Management, Alamo, California, are two decision support packages for water demand forecasting. The use of these software packages have been demonstrated by Mohamed and Al-Mualla (2010a, b) and Levin et al. (2006), respectively.

Forecasting with Artificial Neural Networks

Artificial neural networks (ANN) and fuzzy logic techniques of forecasting water demand are advanced methods classified as nonparametric in Billings and Jones (2008). They can be used for both regression and time series models yet they do not require adherence to the assumptions that form the basis for these conventional methods. Nonetheless, identifying the optimal architecture first

J. Water Resour. Plann. Manage. 2014.140:146-159.

requires the determination of the structure of a univariate time series or regression model. The structural components in the data set are determined using a training data set (estimation period), whereas forecasts are produced and compared with a hold-out data set (hold-out period).

The appearance of neural network models for water demand forecasting in the literature normally involves a comparative assessment of the performance between different neural network models and conventional regression models (Adamowski and Karapataki 2010; Cutore et al. 2008; Firat et al. 2009; Herrera et al. 2010; Jentgen et al. 2007), with time series models (Ghiassi et al. 2008), or with both (Bougadis et al. 2005; Jain and Ormsbee 2002; Jain et al. 2001). In general, authors find ANN to perform better than the conventional methods, although results in Jentgen et al. (2007) and Cutore et al. (2008) were inconclusive.

Forecasting with Composite Models

Finally, an approach that has found wide application in water demand forecasting is classified as hybrid models (Jentgen et al. 2007). These models use more than one method and/or model to arrive at a composite forecast and usually involve some form of combination of forecasts from models through simple or weighted averages (Caiado 2010; Wang et al. 2009) or by applying a mix of methods and models to forecast the decomposed components of a time series (Gato et al. 2007a, b; Zhou et al. 2000). Within this group, a third class prepares forecasts at lower periodicities by adjusting forecasts obtained for higher periodicities (Alvisi et al. 2007; Aly and Wanakule 2004).

In the case of combining forecasts from different models to obtain a composite forecast, Eq. (19) is used, where $\hat{Y}_{i,t}$ is the predicted value of the time series at time t using the ith model. The β_i coefficients are determined by optimization or least squares regression to minimize the mean squared error [see Eq. (3)] between the composite forecast $\hat{Y}_t$ and the actual data (Clemen 1989):

$$\hat{Y}_t = \beta_0 + \sum_{i=1}^{n} \beta_i \hat{Y}_{i,t} \tag{19}$$

The urban water demand forecasting literature tends to favor composite forecasts developed by decomposition. An example of this practice is found in Wu and Zhou (2010), in which the authors used linear regression to model the deterministic component of demand and ANN to model the cyclical component. They subsequently compared their results with forecasts obtained from separately using conventional regression and ANN to model the deterministic and cyclic components, respectively.

In the paper "Combining forecasts: A review and annotated bibliography," Clemen (1989) provides the rationale for the use of these hybrid models: "... [the] idea of combining forecasts implicitly assumed that one could not identify the underlying process, but that different forecasting models were able to capture different aspects of the information available for prediction." In most instances, these composite forecasts are reported to have led to better forecasting performance for water demand (Wang et al. 2009).

Empirical Studies on Water Demand Forecasting: Current Emphasis

In this section, we present and discuss results on what appears to be the emphasis of water demand forecasting research, in terms of the methods used and the periodicity of the demand variable. In Table 3, we cross-classify our sample papers, as inventoried in Table 2, by periodicity of demand variable and by method.

The table shows a considerable variation in the occurrence of methods in the literature. Note the coupling of neural network and conventional methods, and the paucity of stand-alone univariate time series analysis papers. Very little focus has been placed on the latter and much less considered are judgmental methods. Two less known methods, microsimulation as examined in Williamson et al. (2002) and space-time forecasting with Bayesian maximum entropy (MBE) as in Lee et al. (2010), have been used for scenario-based approaches.

The summary statistics presented in Table 3 indicate that there is a considerable emphasis on three approaches, instead of using pure conventional methods: (1) scenario-based and DSS models, approaches which accommodate some amount of uncertainty in demand forecasting; (2) comparative assessment of performance between neural nets and conventional methods; and (3) recognition of the need to improve forecast accuracy by using hybrid models.

Over the analysis period examined in this paper, time-series regression models have been used extensively. In most instances they are not stand-alone but compared with neural networks and/or in combination with univariate time series models. The general conclusion, by researchers, from this line of work is that models developed from neural networks perform better than those developed by time-series regression or univariate time series models. This superiority in performance is attributed to the ability of neural networks to efficiently capture nonlinearities that may exist in the structure of time-series regression and univariate time series models. However, the much touted better performance of ANN over these conventional methods refers primarily to short-term forecasts with very little research conducted on how they compare over medium-to-long-term forecasts.

Table 3 shows that in terms of periodicity of the demand variable, the literature is skewed towards daily and annual variation of water demand, reflecting an emphasis on satisfying the mandate of water utilities, which is to maintain a reliable supply of potable water to consumers and to ensure that this level of reliability is maintained in future years. Thus, the recent literature analyzed in this paper seems to have focused on forecasting to address operations and strategic planning problems, such as system optimization and capacity planning (see Table 1).

A few additional characteristic features of Table 3 are worth noting:

1. Stand-alone regression models have focused on monthly water demand.
2. Annual variation in demand variables have attracted research employing scenario-based and DSS models, ostensibly to model uncertainties.
3. Long-term demand forecasting has not benefited much from the comparative performance assessment between neural networks and conventional methods.

The table also seems to suggest that less emphasis has been placed on hourly variation of demand. Contrary to this perception, hourly variation of demand has been accommodated in the papers that examined multiple periodicities.

Overall, improving forecast accuracy, accounting for uncertainty in long-term forecasts and maintaining system reliability now and in the future seem to have provided the impetus for current research in urban water demand forecasting. As evident in Table 3, it is difficult to answer the question "Which model is best for water demand forecasting?" without specifying the periodicity of the demand variable.

The literature analyzed in this study, as presented in Table 3, also reveals that neural networks and hybrid models are more appropriate for short-term forecasts. But for extended ones, in which incorporating future scenarios of a variable might be important,

J. Water Resour. Plann. Manage. 2014.140:146-159.

Table 3. A Cross-Tabulation of 2000–2010 Water Demand Forecasting Literature by Forecast Periodicity and Method

Periodicity of demand variable	Water demand forecasting approach: Univariate time series analysis	Times series regression	Scenario-based	DSS	Neural networks with conventional methods	Composite/hybrid	Total
Hourly	—	—	—	—	Herrera et al. (2010)	Zhou et al. (2002)	2
Daily	—	—	Goodchild (2003)	Froukh (2001), Jain and Ormsbee (2001) and Levin et al. (2006)	Cutore et al. (2008) and Jain and Ormsbee (2002)	Caiado (2010), Gato et al. (2007a), Gato et al. (2007b), and Zhou et al. (2000)	10
Weekly	—	—	—	—	Adamowski and Karapataki (2010), Bougadis et al. (2005), and Jain et al. (2001)	—	3
Monthly	—	Brekke et al. (2002) and Polebitski and Palmer (2010)	—	Mohammed and Al-Mualla (2010a, b)	Firat et al. (2009)	Altunkaynak et al. (2005)	6
Annual	Alhumoud (2008)	Lee et al. (2010)	Burney et al. (2001), Wei et al. (2010) and Williamson et al. (2002)	Feng et al. (2007)	—	Wang et al. (2009) and Wu et al. (2010)	8
Multiple periodicities	—	—	—	—	Ghiassi et al. (2008) and Jentgen et al. (2007)	Alvisi et al. (2007) and Aly and Wanakule (2004)	4
Total	1	3	4	6	9	10	33

scenario-based and DSS models are more suitable. However, the use of regression in modeling monthly demand follows the generally held view that short-to-medium-term demand is typically influenced by weather variables whereas long-term forecasts are more determined by socioeconomic factors (Adamowski and Karapataki 2010; Ghiassi et al. 2008). Ghiassi et al. (2008) emphasizes this view by noting that "...when analyzing or forecasting water demand over a longer time horizon such as decades, economic or demographic factors may be more effectively included in models." This view makes it imperative to use econometric models for long-term demand forecasting and apparently influenced the model developed by Burney et al. (2001). The economic and demographic factors, and the influences of demand management strategies, technological change and climate change on future demand, do not change quickly in the short-term and their long-term estimates can take one of several values. Thus, although weather variables are not predominantly included in long-term demand forecasting, the realities of uncertainty in climate change have resulted in papers that include various climate scenarios in long-term forecasts. The procedure may follow either Burney et al. (2001), Goodchild (2003), Polebitski et al. (2011), or Wei et al. (2010), or may be incorporated in a DSS model similar to Froukh (2001), Levin et al. (2006), and Mohamed and Al-Mualla (2010a, b).

Water Demand Forecasting: Problems and Recommendations

From the foregoing, the need to improve the practice of urban water demand forecasting calls for paying attention to some modeling issues observed in the literature. The first concerns the practicality of some models proposed in the literature. In all its forms, the search for models for practical application should lead to models whose input variables can be collected, monitored and used by the utility. Models that contain many variables, as is Coomes et al. (2010), and those that utilize derivatives such as "days since 2 mm of rainfall" (Goodchild 2003) pose the greatest challenge to practice in terms of collecting and keeping track of the data. Operationalizing such models will be practically difficult when regression is used for short-term forecasting. For determinants that are not under the control of water utilities, considerable difficulty in acquiring reliable forecasts of such variables will preclude the use of econometric models in favor of time series analysis models. Researchers should therefore take into consideration the ability of utilities to acquire and monitor predictors if the models they propose are to be used for forecasting. Models should be as parsimonious as possible without compromising on structural integrity and forecast quality.

Related to the problem of numerous determinants is what seems to be a naïve and baseless selection of autoregressive terms to model time series data as in Bougadis et al. (2005) and Jain et al. (2001), and the inclusion of lagged variables of both demand and weather derivatives in regression models [see, for example, Jain and Ormsbee (2002); Jentgen et al. (2007); Wei et al. (2010)]. The arbitrary use of various lags of the demand variable provides practicing water utilities with a less rational basis for their inclusion in forecasting models. Contrary to such practices, the use of autoregressive terms should be informed by the structure of the autocorrelation and partial autocorrelation functions to account for serial autocorrelation in the data, if any. This will require not a cursory exposure to, but a clear understanding of the Box-Jenkins methodology. Wherever lagged values of other variables are included, the proper unit root tests must be conducted to justify their inclusion to

J. Water Resour. Plann. Manage. 2014.140:146-159.

avoid the problem of spurious correlations that can exist in time series regression models. In this regard, we make reference to Burney et al. (2001) and Martinez-Espineira (2007), in which the authors handled the concept of cointegration and provided the right approach to including indicator variables in econometric forecasting of water demand. ARMAX(p,q,b) modeling in the drinking water industry is a fertile ground, given the flaws in current approaches to model identification when exogenous variables are incorporated in water demand forecasting.

An opportunity exists for conducting post evaluation of the accuracy of implemented models. None of the papers reviewed for this study report the results of such research, although it is one of the key recommendations made in Billings and Jones (2008). From the literature, it is difficult to evaluate if past forecasts turned out to be accurate and reliable. It will be useful to have a retrospective evaluation of selected methods, by comparing how forecast values compared with actual demand. For future research, we propose post evaluation of forecast methods for selected existing utilities. This could be done along the lines proposed by Fischer et al. (2009) and Shlyakhter et al. (1994), in which the authors evaluated the errors in energy demand projections in the United States. The value of this line of study is to contribute in improving the accuracy of future probabilistic forecasts as exemplified in the U.S. energy sector. A good source of data for starting such a research in the water sector is the Interstate Commission on the Potomac River Basin's (ICPRB) initiative, where every five years, the institution prepares annual water demand forecasts for a 20-year horizon (Hagen et al. 2005). Similar exercises could be undertaken for other forecast horizons other than the long-term.

To date, the quest for a generalized forecasting model that can be used by all utilities has eluded researchers. For each conventional method, there is currently no acceptable model for forecasting water demand, irrespective of the planning level involved. For instance, for univariate time series models, the probabilistic structure which generates total monthly demand is not known for certain. This contrasts the airline model described by Box et al. (1987) and models used for forecasting stock price returns. Such industry models are known to follow specific stochastic processes, making it possible for analysts to focus on determining the parameters of these models for a given data set. The existence of these stochastic industry models creates a need for similar models to be developed for the drinking water industry. This line of research can concentrate on identifying the probabilistic structures that generate the series for short-term, medium-term, and long-term water demand forecasting. The validation of such models will help utilities estimate and better manage the uncertainty in demand forecasts and will result in optimized operations and investment decisions.

Conclusion

This review has presented an overview of the water demand forecasting literature appearing from 2000 to 2010, in an attempt to provide some guidance, first for practicing professionals seeking to adopt methods and models suitable for addressing planning-related decisions that are dependent on future levels of demand, and secondly for researchers intent on extending and/or improving models. In conclusion, we observe from the literature analyzed in this study that:

1. The basis for urban water demand forecasting is enshrined in utility management decision problems that are dependent on uncertain /stochastic future levels of demand, and that the forecast horizon and periodicity are key drivers to method and model selection.
2. Although in practice, rudimentary approaches to unit rate modeling is predominantly preferred, this has not attracted research attention in recent times.
3. Rather, there's been an emphasis on scenario-based forecasting and approaches that use decision support systems, a comparison of the performance of neural networks against conventional methods, and the use of hybrid models, all in an attempt to either account for uncertainty and/or to improve forecast accuracy.
4. If utilities are to account for uncertainty and make decisions that incorporate this uncertainty in their forecasts, then the neglected probabilistic forecasting methods will require greater attention than currently received, as an advanced step beyond scenario-based forecasting.

Acknowledgments

This study was funded by the National Oceanic and Atmospheric Administration's (NOAA) Climate Program Office through the American Water Works Association, as part of the project "Decreasing climate induced risk for municipal water demand forecasting." We appreciate the feedback received from participants of two AWWA workshops designed to discuss outputs of this study and from three anonymous reviewers of this paper. Our citing of commercial software products and consulting entities is for illustrative purposes only and not an endorsement of their preference over other equally good alternatives.

References

Adamowski, J., and Karapataki, C. (2010). "Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000245, 729–743.

Alhumoud, J. (2008). "Freshwater consumption in Kuwait: Analysis and forecasting." *J. Water Supply: Res. Technol.—AQUA*, 57(4), 279–288.

Altunkaynak, A., Özger, M., and Çakmakc, M. (2005). "Water consumption prediction of Istanbul city by using fuzzy logic approach." *Water Resour. Manage.*, 19(5), 641–654.

Alvisi, S., Franchini, M., and Marinelli, A. (2007). "A short-term, pattern-based model for water-demand forecasting." *J. Hydroinf.*, 9(1), 39–50.

Aly, A., and Wanakule, N. (2004). "Short-term forecasting for urban water consumption." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2004)130:5(405), 405–410.

Arbués, F., Garcıa-Valiñas, M. Á., and Martınez-Espiñeira, R. (2003). "Estimation of residential water demand: A state-of-the-art review." *J. Soc. Econ.*, 32(1), 81–102.

Billings, B., and Jones, C. (2008). *Forecasting urban water demand*, 2nd Ed., American Waterworks Association, Denver, CO.

Bougadis, J., Adamowski, K., and Diduch, R. (2005). "Short-term municipal water demand forecasting." *Hydrolog. Process.*, 19(1), 137–148.

Box, G. E. P., Pierce, D. A., and Newbold, P. (1987). "Estimating trend and growth rates in seasonal time series." *J. Am. Stat. Assoc.*, 82(397), 276–282.

Brekke, L., Larsen, M., Ausburn, M., and Takaichi, L. (2002). "Suburban water demand modeling using stepwise regression." *J. Am. Water Works Assoc.*, 94(10), 65–75.

Burney, N., Mukhopadhyay, A., Al-Mussallam, N., Akber, A., and Al-Awadi, E. (2001). "Forecasting of freshwater demand in Kuwait." *Arabian J. Sci. Eng.*, 26(2B), 99–113.

Caiado, J. (2010). "Performance of combined double seasonal univariate time series models for forecasting water demand." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000182, 215–222.

Clemen, R. (1989). "Combining forecasts: A review and annotated bibliography." *Int. J. Forecasting*, 5(4), 559–583.

Coomes, P., Rockaway, T., Rivard, J., and Kornstein, B. (2010). *North America residential water usage trends since 1992*, Water Research Foundation, Denver.
Cutore, P., Campisano, A., Kapelan, Z., Modica, C., and Savic, D. (2008). "Probabilistic prediction of urban water consumption using the SCEM-UA algorithm." *Urban Water J.*, 5(2), 125–132.
Feng, S., Li, L., Duan, Z., and Zhang, J. (2007). "Assessing the impacts of south-to-north water transfer project with decision support systems." *Decis. Support Syst.*, 42(4), 1989–2003.
Firat, M., Yurdusev, M., and Turan, M. (2009). "Evaluation of artificial neural network techniques for municipal water consumption modeling." *Water Resour. Manage.*, 23(4), 617–632.
Fischer, C., Herrnstadt, E., and Morgenstern, R. (2009). "Understanding errors in EIA projections of energy demand." *Resour. Energy Econ.*, 31(3), 198–209.
Froukh, M. (2001). "Decision-support system for domestic water demand forecasting and management." *Water Resour. Manage.*, 15(6), 363–382.
Gardiner, V., and Herrington, P. (1990). *Water demand forecasting*, 1st Ed., Spon Press, Norwich, U.K.
Gardner, E. (2006). "Exponential smoothing: The state of the art-part II." *Int. J. Forecasting*, 22(4), 637–666.
Gato, S., Jayasuriya, N., and Roberts, P. (2007a). "Forecasting residential water demand: Case study." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2007)133:4(309), 309–319.
Gato, S., Jayasuriya, N., and Roberts, P. (2007b). "Temperature and rainfall thresholds for base use urban water demand modelling." *J. Hydrol.*, 337(3–4), 364–376.
Ghiassi, M., Zimbra, D., and Saidane, H. (2008). "Urban water demand forecasting with a dynamic artificial neural network model." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2008)134:2(138), 138–146.
Goodchild, C. (2003). "Modelling the impact of climate change on domestic water demand." *Water Environ. J.*, 17(1), 8–12.
Hagen, E., Holmes, K., Kiang, J., and Steiner, R. (2005). "Benefits of iterative water supply forecasting in the Washington, DC, metropolitan area." *J. Am. Water Resour. Assoc.*, 41(6), 1417–1430.
Hazen, Sawyer and PMCL. (2004). "The Tampa Bay water long-term demand forecasting model." ⟨http://www.tampabaywater.org/documents/conservation/LongtermDemandForecastingModelJune2003.pdf⟩ (Jun. 04, 2012).
Herrera, M., Torgo, L., Izquierdo, J., and Perez-Garcia, R. (2010). "Predictive models for forecasting hourly urban water demand." *J. Hydrol.*, 387(1–2), 141–150.
Hyndman, R., Koehler, A., Ord, J., and Snyder, R. (2005). "Prediction intervals for exponential smoothing using two new classes of state-space models." *J. Forecasting*, 24(1), 17–37.
Jain, A., and Ormsbee, L. (2001). "A decision support system for drought characterization and management." *Civ. Eng. Environ. Syst.*, 18(2), 105–140.
Jain, A., and Ormsbee, L. (2002). "Short-term water demand forecast modeling techniques—conventional methods versus AI." *J. Am. Water Works Assoc.*, 94(7), 64–72.
Jain, A., Varshney, A., and Joshi, U. (2001). "Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks." *Water Resour. Manage.*, 15(5), 299–321.
Jentgen, L., Kidder, H., Hill, R., and Conrad, S. (2007). "Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations." *J. Am. Water Works Assoc.*, 99(6), 86–94.
Lee, S., Wentz, E., and Gober, P. (2010). "Space-time forecasting using soft geostatistics: A case study in forecasting municipal water demand for Phoenix, Arizona." *Stochastic Environ. Res. Risk Assess.*, 24(2), 283–295.
Levin, E., Maddaus, W., Sandkulla, N., and Pohl, H. (2006). "Forecasting wholesale demand and conservation savings." *J. Am. Water Works Assoc.*, 98(2), 102–111.
Martinez-Espineira, R. (2007). "An estimation of residential water demand using co-integration and error correction techniques." *J. Appl. Econ.*, 10(1), 161–184.
Miaou, S. P. (1990). "A stepwise time series regression procedure for water demand model identification." *Water Resour. Res.*, 26(9), 1887–1897.
Mohamed, M., and Al-Mualla, A. (2010a). "Water demand forecasting in Umm Al-Quwain (UAE) using the IWR-MAIN specify forecasting model." *Water Resour. Manage.*, 24(14), 4093–4120.
Mohamed, M., and Al-Mualla, A. (2010b). "Water demand forecasting in Umm Al-Quwain using the constant rate model." *Desalination*, 259(1–3), 161–168.
Mun, J. (2010). *Modeling risk: Applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, and portfolio optimization*, 2nd Ed., Wiley, Hoboken, NJ.
Palisade Inc. (n.d). "Case study: ADWEC uses @RISK from Palisade to forecast water demand in Abu Dhabi—Palisade." ⟨http://www.palisade.com/cases/adwec.asp⟩ (Jun. 4, 2012).
Polebitski, A., and Palmer, R. (2010). "Seasonal residential water demand forecasting for census tracts." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000003, 27–36.
Polebitski, A., Palmer, R., and Waddell, P. (2011). "Evaluating water demands under climate change and transitions in the urban environment." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000112, 249–257.
Shlyakhter, A., Kammen, D., Broido, C., and Wilson, R. (1994). "Quantifying the credibility of energy projections from trends in past data—The United-States energy sector." *Energy Policy*, 22(2), 119–130.
Strzepek, K., et al. (1999). "New methods of modeling water availability for agriculture under climate change: The US cornbelt." *J. Am. Water Resour. Assoc.*, 35(6), 1639–1655.
Taylor, S. (1981). "Initialization of exponential smoothing forecasts." *AIIE Trans.*, 13(3), 199–205.
Wang, X., Sun, Y., Song, L., and Mei, C. (2009). "An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. II: Model application and results." *J. Environ. Manage.*, 90(8), 2612–2619.
Wei, S., Lei, A., and Islam, S. (2010). "Modeling and simulation of industrial water demand of Beijing municipality in China." *Frontiers of environmental science and engineering in China*, 4(1), 91–101.
Wei, W. (2005). *Time series analysis: Univariate and multivariate methods*, 2nd Ed., Addison Wesley, Boston, MA.
Williamson, P., Mitchell, G., and McDonald, A. (2002). "Domestic water demand forecasting: A static microsimulation approach." *Water Environ. J.*, 16(4), 243–248.
Wu, L., and Zhou, H. (2010). "Urban water demand forecasting based on HP filter and fuzzy neural network." *J. Hydroinf.*, 12(2), 172–184.
Zhou, S., McMahon, T., Walton, A., and Lewis, J. (2000). "Forecasting daily urban water demand: A case study of Melbourne." *J. Hydrol.*, 236(3–4), 153–164.
Zhou, S., McMahon, T., Walton, A., and Lewis, J. (2002). "Forecasting operational demand for an urban water supply zone." *J. Hydrol.*, 259(1–4), 189–202.

J. Water Resour. Plann. Manage. 2014.140:146-159.