

Case Study: Improvements to a Real-Time Network Modelling Framework

S. M. Masud Rana¹, Paulo José Oliveira², Tian Qin³, Dominic L. Boccelli⁴

^{1,2,3,4} Department of Chemical and Environmental Engineering
University of Cincinnati, Cincinnati, OH, 45221-0012

¹ranasd@mail.uc.edu, ²oliveipa@mail.uc.edu, ³qintn@mail.uc.edu, ⁴dominic.boccelli@uc.edu

ABSTRACT

Short-term water demand forecasts are valuable for distribution system operators controlling the production, storage and delivery of drinking water. In certain problems, such as real-time pump scheduling, the cycle of data acquisition, model computation, and decision-making is time-sensitive, and requires an automatic procedure to handle the transfer of information between data source(s), forecasting model(s), and the operator. Recent development of a composite demand-hydraulic model integrates a demand time series model with a hydraulic network model to estimate and forecast demands using measurements typically available to water utilities. The application to a real-world network model with approximately 12,000 demand nodes and six flow measurements resulted in good representation of the observed flow rates. However, the performance of the demand-hydraulic algorithm, and subsequent analysis, has demonstrated limitations in two aspects of the demand estimation and forecasting framework: the temporal representation of the estimated demands, and the clustering approach needed to reduce the scale of the parameter estimation problem. The current research will present preliminary results associated with data-driven approaches for representing the temporal demands and application of alternative clustering algorithms to improve the overall demand estimation process.

Keywords: real-time modeling, time series, clustering

1 BACKGROUND

Real-time modeling applications are currently being developed to assist utilities in the operation and management of urban water infrastructure systems (e.g., [1] – [4]). In order to realize the potential benefits of these applications, short-term water demand forecasts and the associated uncertainty estimates are required to provide the information necessary to make real-time decisions. While a general framework has been presented for incorporating uncertainty within real-time network modeling [5], the extension of real-time demand estimation algorithms [6, 7] to short-term forecasting and uncertainty estimation remains a developing area.

Recent efforts have developed a composite demand-hydraulic modeling framework using a Bayesian Dynamic Network approach capable of estimating both water demands and the parameters of an underlying time series modeling structure [8]. While the application to a real-world network generated good representation of the measured flow rates in the system [9], there remain opportunities to improve the time series modeling approaches to better accommodate non-linear dynamics as well as the general clustering approach to ensure well-behaved demand estimation problems (recent research has demonstrated that measurement locations and clustering decisions can result in ill-conditioned estimation problems [10]). Thus, our current research has been focused on improved approaches to: 1) represent the time series of estimated demands, and 2) cluster demand nodes to improve demand estimation.

2 METHODS

The following presents a summary of the methods currently being evaluated to improve the overall real-time demand network modeling framework.

2.1 Time series modeling

The current implementation of the composite demand-hydraulic model includes the capabilities for representing the estimated demands from a family of seasonal autoregressive integrated moving average (SARIMA) models. The initial application and investigation of a real-world case study utilized a SARIMA model with 24-hour seasonality and three autoregressive parameters, i.e., SAR(3). To improve the representation of the non-linear dynamics within the demand estimation and forecasting, a k-nearest neighbor (knn) approach has been investigated.

The knn approach identifies the k-set of previous demand estimates most similar to the current estimates and generates forecasted demands based upon the k-set of historical data. Given a vector of observed demands, $\mathbf{X}(\mathbf{t}) = [x(t) \ x(t-1) \ x(t-2) \ \dots \ x(t-M)]'$, the knn algorithm identifies the k-sets, $\mathbf{X}^*(\mathbf{t}_k)$, most similar to $\mathbf{X}(\mathbf{t})$, where t_k represents some previous time $t_k < t$, and M represents the length of past time demand estimates used in the knn algorithm. Using the k-set of previous demands, the forecasted value n -steps ahead can be estimated by $x(t+n) = 1/k \sum x(t_k+n)$, $\forall k$ – an average of the n -step ahead values from the k-sets, with the standard deviation of these values representing the demand uncertainty. The forecasted uncertainty around $x(t+n)$ is assumed to be normally distributed. For this study, the knn demand estimation and forecasting was performed using the total system demand from a real-world distribution system.

2.2 Demand clustering

For many networks, the number of available hydraulic measurements is less than the number of nodes requiring demand estimates necessitating the clustering of demand nodes. While some systems have district metered areas that effectively separate the system into different zones, many systems do not have such physical separations requiring alternative approaches to provide some level of spatial aggregation. Previous research has focused on using a hydraulic path based approach for developing clusters [11]. However, the approach may result in ill-conditioned parameter estimation problems resulting in inaccurate or highly uncertain parameter estimates. The current approach being evaluated uses a singular value decomposition (SVD) method [12] to cluster demand nodes. The approach uses the SVD of the sensitivity matrix to identify the demand clusters. Preliminary results associated with the SVD approach have been applied to a real-world network model.

3 RESULTS

The following provide preliminary results into our investigation of improved time series modeling and clustering for demand estimation. These results illustrate the initial development and assessment of the individual techniques with the application to a full case study left as future research.

3.1 Time series modeling

Figure 1 presents the results of the SAR(3) model [top], the knn model with $M = 10$, knn(10), and the knn model applied to hourly differenced data with $M = 5$, knn_D(5). The solid line represents the observed demands, and the shaded regions represent the 95% confidence intervals for each of seven

24-hour forecasting periods – i.e., the shaded regions show the forecasting uncertainty from 1- to 24-hours ahead. The evaluation of the three methods were focused on the sharpness – the average size of a given prediction interval – and reliability – the percentage of forecasted estimates with in the 95% confidence interval range. In general, the sharpness (as well as the mean average percent error and root mean squared error), improved as the models moved from the SAR(3) [1.02] to the knn(10) [0.89] and knn_D(5) [0.68] models. However, the reliability was not significantly affected by model selection with the reliability ranging from 82 – 86%.

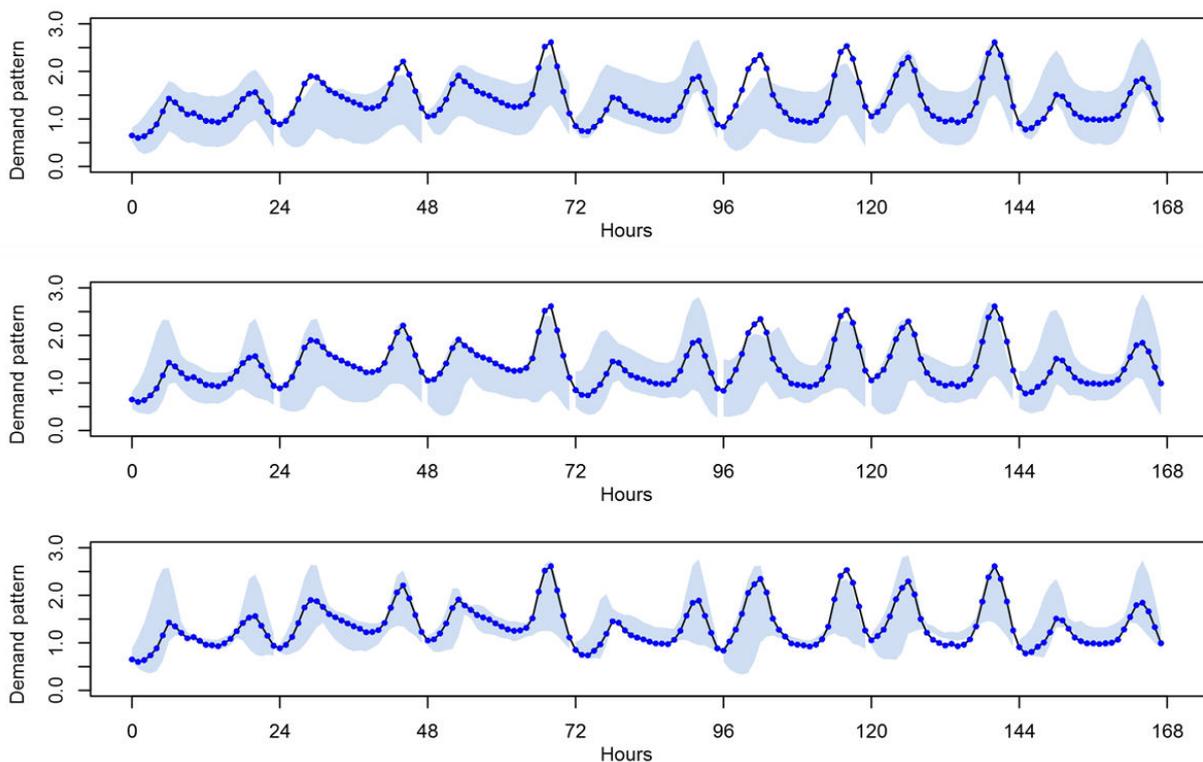


Figure 1. Observed total demand data [lines] and 95% confidence intervals of the forecasted demands for seven 24-hour periods [shaded region]; for each 24-hour interval, the uncertainties are shown for 1- to 24-hr ahead forecasts.

3.2 Demand clustering

Figure 2 shows the resulting set of six demand node clusters using the SVD approach for a network model with approximately 12,000 nodes and six flow measurements. The available flow measurements used within the demand estimation algorithm are shown on the maps as arrows. The bi-directional arrows represent two flow meters. In general, each cluster is effectively associated with one flow measurement, and the overall clustering based on the SVD approach generally represents the main transport dynamics within the system. The main treatment plant [center, right] provides water to both the north and the south portions of the system with the three uppermost clusters representing the northern region and the three lower clusters representing the southern region. The northeast cluster (orange) represents the portion of the system influenced primarily by

a secondary treatment plan. Future research will be performed to evaluate the impact of clustering on the demand estimation and forecasting as well as the conditioning of the overall estimation problem.

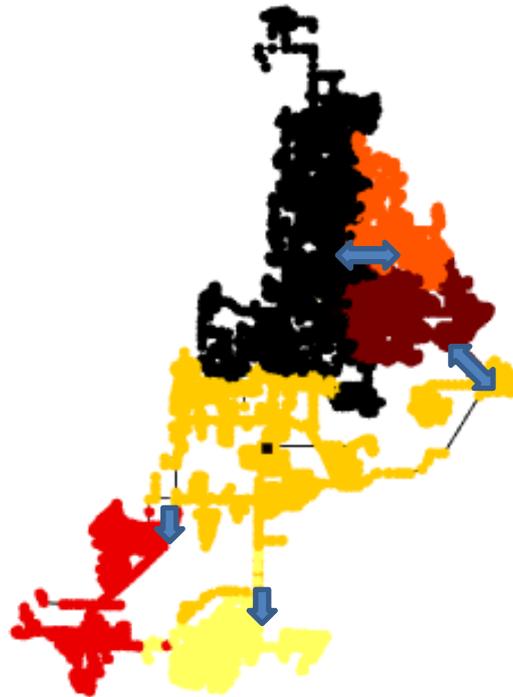


Figure 2. Illustration of six clusters within a large network model using SVD and six available flow measurements.

4 SUMMARY

Real-time demand estimation and forecasting are necessary for the development of real-time operation and management applications. While demand estimation algorithms have long been developed, challenges are still present in ensuring adequate demand estimates in real systems and translating those results into adequate demand forecasts. Previous research associated with the development of a composite demand-hydraulic modeling framework has shown promise for both demand estimation and forecasting with respect to a real-world system. However, there is still room for improvement in many aspects, including the representation of the estimated demand time series and the approach for clustering networks for given a set of observation locations to ensure adequate parameter estimation. Currently, efforts are on-going to evaluate data-driven demand modeling through k-nearest neighbors and the clustering through singular value decomposition to improve the overall demand estimation and forecasting framework. These efforts are part of a continuum in research activities associated with improving real-time modeling capabilities.

References

- [1] D. Jung, K. E. Lansey, Y. Choi and J. H. Kim, “Robustness-based optimal pump design and scheduling for water distribution systems,” *J. Hydroinf.*, vol 18, no. 3, pp. 500-513, 2016.
- [2] L. Murphy, D. McIver, G. Dandy, C. Hewitson, J. Frey, L. Jasobsen and M. Fang, “GA optimization for Las Vegas Valley water distribution system operations and water quality,” *World Water and Environmental Resources Congress*, ASCE, 2007.
- [3] X. Yang and D. L. Boccelli, “Bayesian approach for real-time probabilistic contamination source identification,” *J. Water Resour. Plann. Manag.*, vol 140, no. 8, 2014.
- [4] S. M. M. Rana and D. L. Boccelli, “Contaminant spread forecasting and confirmatory sampling location identification in a water distribution system network,” *J. Water Resour. Plann. Manag.*, vol, 142, no. 12, 2016
- [5] C. J. Hutton, Z. Kapelan, L. Vamvakieridou-Lyroudia and D. A. Savic, “Dealing with uncertainty in water distribution system models: A framework for real-time modelling and data assimilation,” *J. Water Resour. Plann. Manag.*, vol 140, no. 2, pp. 169-183, 2014.
- [6] F. Shang, J. G. Uber, B. G. van Bloemen Waanders, D. Boccelli and R. Janke, “Real time water demand estimation in water distribution system,” *8th Annual Water Distribution Systems Analysis Conference*, 2006.
- [7] D. Kang and K. Lansey, “Real-time demand estimation and confidence limit analysis for water distribution systems,” *J. Hydraul. Eng.*, vol 135, no. 10, pp. 825-837, 2009.
- [8] J. Chen, Stochastic Demand-Hydraulic Model of Water Distribution Systems, Ph.D. Dissertation, University of Cincinnati, 2015.
- [9] P. J. Oliveira, S. M. M. Rana, T. Qin, H. Woo, J. Chen and D. L. Boccelli, “Case Study: Evaluation of a Composite Demand-Hydraulic Modelling Framework,” *Water Distribution System Analysis Conference*, 2016
- [10] S. M. M. Rana, D. L. Boccelli, A. Marchi and G. Dandy “Impacts of measurement location and spatial aggregation on demand estimation in water distribution systems,” *World Water and Environmental Resources Congress*, 2017
- [11] T. Qin and D. L. Boccelli, “Grouping water demand nodes by similarity among flow paths in water distribution systems,” *J. Water Resour. Plann. Manag.*, vol 143, no. 8, 2017
- [12] G. Sanz and R. Pérez, “Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition,” *J. Water Resour. Plann. Manag.*, vol 141, no. 10, 2015