CROSS CORRELATION ANALYSIS OF RESIDENTIAL DEMAND IN THE CITY OF MILFORD, OHIO

Y.R. Filion[1], B.W. Karney[2], L. Moughton[3], S.G. Buchberger[4], B.J. Adams[5]

[1] Centre for Water Systems, University of Exeter, Exeter, UK, *yves.filion@utoronto.ca*.
[2] Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada, *karney@ecf.utoronto.ca*.
[3] Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio, *moughtlj@uc.edu*.
[4] Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio, *Steven.Buchberger@uc.edu*.
[5] Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada, *adams@ecf.utoronto.ca*.

Abstract

Estimating future demands with a high degree of accuracy in water distribution network design remains an elusive goal. The desired outcome is to match the design demands to the demands that are eventually "realized" in the built system. To this end, this paper explores the cross correlation between demands in an existing system in order to gain a better picture of the representative spatial and temporal patterns of design demands. The aim of the paper is to analyze the cross correlation in the residential demand data collected in the city of Milford, Ohio. More specifically, the paper begins to answer five important questions concerning the cross correlation of the Milford demand data: how strongly cross correlated are indoor, residential demands? How strongly correlated is the deterministic, diurnal component of residential demand? How strongly correlated is the random noise component of residential demand? To what extent does the choice of time step influence the strength of correlation between these 3 demand components? Does the correlation between these 3 demand components differ significantly between weekdays and weekends? To answer these questions, a periodic regression model was used to isolate the deterministic and the random noise components from the residential demand data collected in Milford. Correlation indices were formulated to measure the cross correlation of residential demand, its deterministic, diurnal component, and its random noise component. The Milford results pointed to a number of preliminary findings: (1) both residential demand and its deterministic, diurnal component had a positive and moderate to high correlation, while the random noise component of demand had a low level of correlation for the cases investigated; (2) increasing the time step length (from 600 s to 3,600 s) did increase the strength of the correlation in residential demand and its deterministic, diurnal component. This suggests that a longer time step increases both the coherence in diurnal demand patterns and their synchronicity. It is unclear whether time step length had any influence on the correlation of the random noise component of demand; (3) both residential demand and its deterministic, diurnal component were more strongly correlated during weekend periods than during weekday periods. This finding suggests that weekend periods may be characterized by less erratic water use patterns between customers leading to more coherent and synchronous diurnal patterns. It is unclear whether the random noise component was influenced by day-of-week effects. The implications of these preliminary results are discussed in the context of extended period simulation (EPS) and water quality modeling as they pertain to cost-effective design.

Keywords
Residential water demand, cross correlation, periodic regression.

1. INTRODUCTION

To gain an accurate picture of how a water distribution network will perform once it is built or improved, it is crucial that the demands estimated at the design stage be similar to those eventually observed in the built network. In the context of an extended period simulation (EPS), the crucial challenge is to find representative spatial patterns of demand across network nodes for peak and off-peak periods. Similarity between patterns of diurnal demand assumed in design (performed with EPS) and those observed in the field will ensure that the network is appropriately sized (at a reasonable cost) to provide acceptable pressures during peak and off-peak periods.

A number of researchers have begun to explore how the spatial correlation of demand affects the hydraulic performance and the cost of a network. Tolson *et al*. (2004) applied a chance-constrained optimization model to a simple network to show that increasing the strength of cross correlation between demands increases network cost at a fixed level of hydraulic reliability. The results of Tolson *et al*. pointed to the notion that a higher level of coordination between network demands produces larger fluctuations in nodal pressure, and more frequent low-pressure failures. Kapelan *et al*. (2005) and Babayan *et al*. (2005) developed multi-objective evaluation models to test the New York tunnels problem under correlated demands. Both studies confirmed the results of Tolson *et al*. and showed that correlated demands yield a higher network cost at a fixed level of hydraulic reliability or robustness. In a related effort, Filion *et al*. (2005, 2006) performed a systematic correlation analysis to better understand the relationship between correlated demands, network cost, and hydraulic performance. The studies showed that in systems without tank storage, the standard deviation of nodal pressures (and thus their variability) is sensitive to the strength of cross correlation between nodal demands. In other words, a stronger correlation between nodal demands leads to a greater variation in pressure. Furthermore, at a fixed level of hydraulic reliability, network cost is sensitive to the strength of cross correlation between demands. With more tightly correlated demands, larger pipe sizes—and a higher network cost—are required to maintain a desired hydraulic reliability in the face of larger pressure fluctuations.

The aim of this paper is to investigate the cross correlation of demand in a real system. To do this, indoor residential demand data from 20 single-family residences in the city of Milford, OH, are analyzed for cross correlation. The paper begins to answer five important research questions: how strongly cross correlated are indoor, residential demands? How strongly correlated is the deterministic, diurnal component of residential demand? How strongly correlated is the random noise component of residential demand? To what extent does time step selection influence the strength of correlation between these 3 demand components? Does the correlation between these 3 demand components differ significantly between weekdays and weekends? The results of the analysis begin to suggest where effort should be focused in future to better characterize design demands to achieve greater accuracy in network modeling and more cost-effective designs. Unfortunately, what cannot be answered at this stage is how representative the preliminary findings for Milford are for other systems.

The paper is structured as follows: the periodic regression model used to isolate the deterministic, diurnal and random noise components from the indoor residential demand data is described. The correlation indices used to measure the cross correlation of residential demand, the deterministic component of demand, and the random noise component of demand are presented. Following this, the Milford, OH demand data is analyzed for cross correlation. The paper concludes by placing the correlation results into the context of the accuracy in hydraulic and water quality modeling.

Water Distribution Systems Analysis Symposium 2006

2. METHODOLOGY

2.1 Milford, OH, Residential Water Demand Data

Buchberger and Wells (1996) collected instantaneous, residential water demand in the city of Milford, OH. Continuous data loggers were installed on the service line of 21 single-family homes along a dead-end loop to collect high-resolution data (1 s interval) on indoor, residential demand from February 1996 to November 1997. The single-family homes in the Milford data set were numbered from 1 through 21. The indoor residential demand data was converted into single, equivalent rectangular pulses (SERPs) that preserve the start time, duration, volume, and average intensity of the water demand (Buchberger and Wells 1996).

2.2 Time and Space Aggregation

In this paper, SERP data series for the days of Wednesday, October 15, 1997 and Sunday, October 19, 1997 was used in the analysis. The demand data for these 2 days was found to have a high number of daily demand counts (high demand data recovery). Home 16 was excluded from the analysis because it showed a low number of daily demand counts. A MATLAB script named Cumulative SERPS (Zhang *et al*. 2005) was used to aggregate the SERP data over groups of homes and over time steps of specified length. For example, Fig. 1 indicates indoor residential demand on Sunday, October 19, 1997 from 00:00:01 to 23:59:59 for SERP data aggregated over 10 homes for time step lengths of 600 s (10 min) and 3,600 s (60 min).

2.3 Periodic Regression of Aggregated Water Demand

The aggregated indoor, residential demand data $q_k(t)$ was assumed to be comprised of a deterministic, diurnal component, $f_k(t)$, and a random noise component, $\varepsilon_k(t)$ as follows

$$q_k(t) = f_k(t) + \varepsilon_k(t) \tag{1}$$

where k = index that refers to the group of homes across which demand was aggregated. Periodic regression was used to isolate the deterministic, diurnal component, $f_k(t)$, and the random noise component, $\varepsilon_k(t)$, from the aggregated indoor, residential demand, $q_k(t)$. The demand model in (1) is expanded to

$$q_k(t) = \mu_k + \sum_{i=1}^{m}\left\{a_{ik}\cos\left(\frac{2\pi t}{T}\ i\right) + b_{ik}\sin\left(\frac{2\pi t}{T}\ i\right)\right\} + \varepsilon_k(t) \tag{2}$$

where μ_k = mean of aggregated indoor, residential demand for group k; a_{ik}, b_{ik} = harmonic coefficients for harmonic i and group k; t = time step number; T = total number of time steps in 24 h period (fundamental period is therefore 24 h); $\varepsilon_k(t)$ = random departure from the periodic regression curve (random noise or residual component); m = number of significant harmonics.

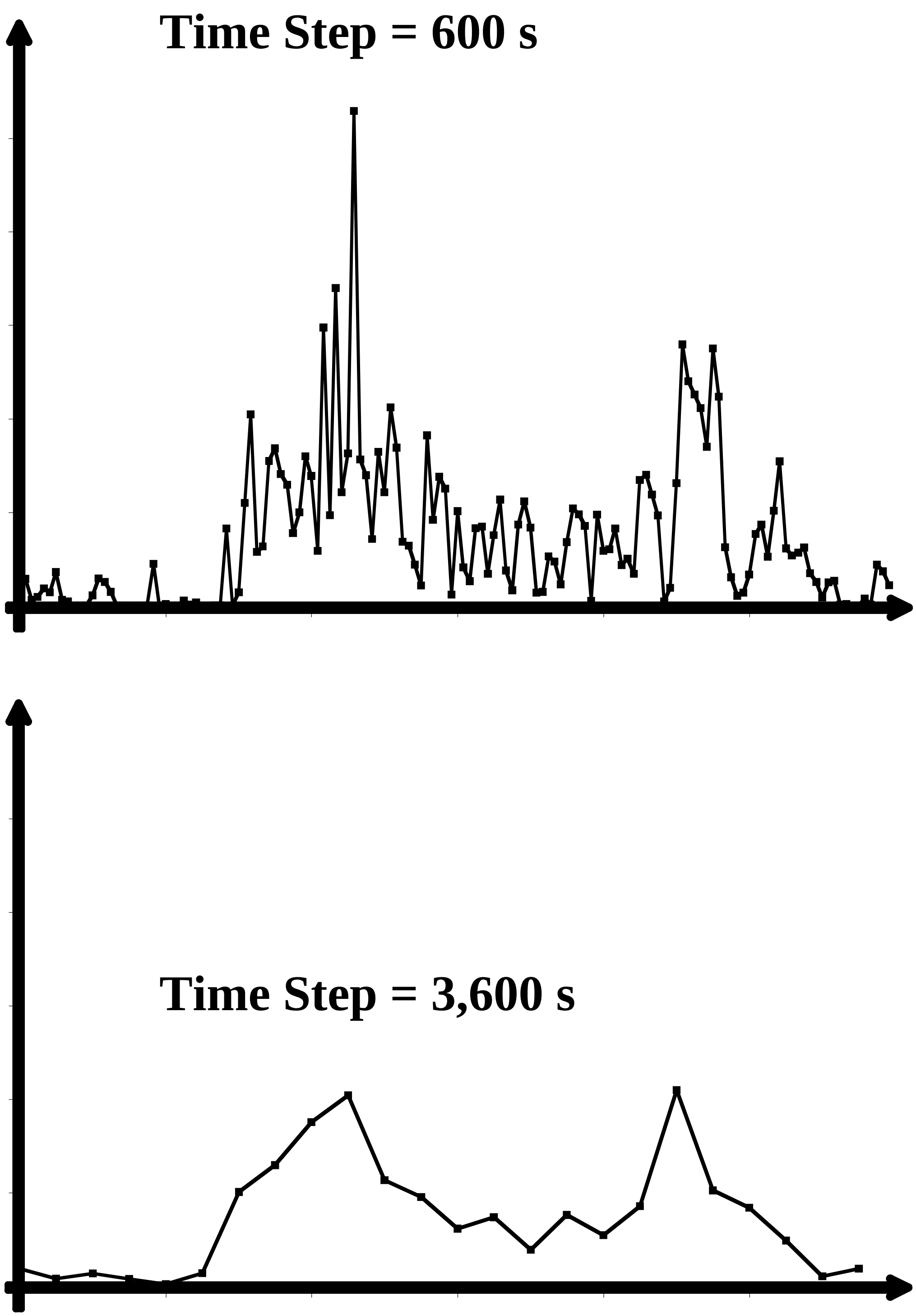

Figure 1. Indoor residential water demand on Sunday, October 19, 1997 from 00:00:01 to 23:59:59 aggregated across 10 homes and over time step lengths of 600 s (10 min) and 3,600 s (60 min).

Water Distribution Systems Analysis Symposium 2006

A periodic regression curve was fit to the demand data on the days of Wednesday, October 15, 1997 and Sunday, October 19, 1997 by the method of least squares to a series of i = 1, 2, 3,…, m significant harmonics. A typical regression of the demand data of October 19, 1997 from 00:00:01 to 23:59:59 is shown in Fig. 2. The figure indicates aggregated indoor, residential demand, $q_k(t)$, the deterministic, diurnal component, $f_k(t)$, and the random noise component, $\varepsilon_k(t)$ (residual).

Note that trigonometric functions only provide a convenient means to explain daily variations in demand. Yevjevich (1972) observed that there is no proof that the periodicity of hydrological phenomena is explained by trigonometric functions. The same is likely true in the case of residential water demand. Periodic regression is simply a mathematical fitting procedure and has no physical basis. Despite this, there is strong evidence to support the notion that residential water use and the social patterns of customers in a network exhibit a periodical trend that are heavily influenced by astronomical cycles such as the earth's rotation about its axis and the earth's rotation about the sun. In this light, periodic regression must therefore be seen as a tool to emulate residential water demand rather than to explain the physical and causal forces that are ultimately responsible for its occurrence.

2.4 Significant Harmonics

The significance of a regression harmonic is determined in part by the percentage of the variance in the residential water demand data that can be explained by that particular harmonic. The percentage of variance accounted for by the i^{th} harmonic is calculated as

$$V_i = \begin{cases} \dfrac{a_i^2 + b_i^2}{2\,s^2} \times 100\%, & i < \dfrac{T}{2} \\[2ex] \dfrac{a_i^2 + b_i^2}{s^2} \times 100\%, & i = \dfrac{T}{2} \end{cases} \tag{3}$$

where V_i = percentage of variance explained by the i^{th} harmonic; s^2 = variance of aggregated indoor residential demand. The criteria for determining significant harmonics in the aggregated residential water demand data is described in the section entitled "Data Analysis and Results".

2.5 Cross Correlation of Water Demand

The last step of the analysis was to compute the cross correlation of residential water demand, $q_k(t)$, aggregated across groups of homes in the city of Milford on the days of October 15, 1997 and October 19, 1997. The cross correlation of residential water demand aggregated across home groupings k and l is computed as

$$r = \frac{1}{T}\sum_{t=1}^{T}(q_k(t)-\mu_k)\cdot(q_l(t)-\mu_l)/\,s_k \cdot s_l \tag{4}$$

where r = cross correlation coefficient of demand between home groupings k and l; s_k, s_l = standard deviation of residential demand for home groupings k and l.

Water Distribution Systems Analysis Symposium 2006

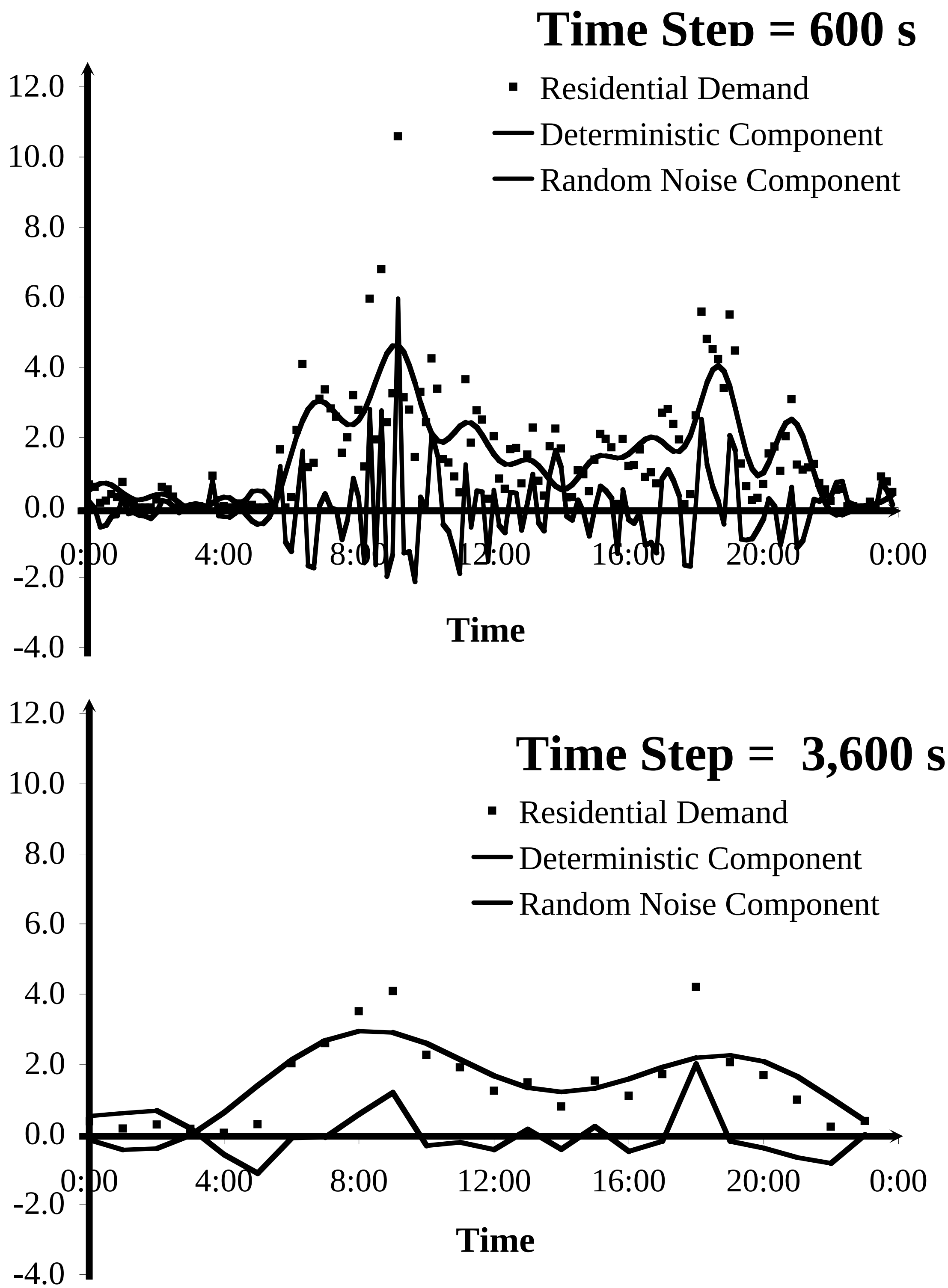

Figure 2. Periodic regression of indoor residential demand data aggregated across 10 homes and time step lengths 600 s (10 min) and 3,600 s (60 min) for October 19, 1997 from 00:00:01 to 23:59:59.

Water Distribution Systems Analysis Symposium 2006

The cross correlation of the deterministic, diurnal component of demand, $f_k(t)$, (explained with periodic regression) between home groupings k and l is computed as

$$r_D = \frac{1}{T}\sum_{t=1}^{T}(f_k(t)-\eta_k)\cdot(f_l(t)-\eta_l)\,/\,s_{k,D}\cdot s_{l,D} \tag{5}$$

where η_k, η_l = mean of deterministic, diurnal component of demand for home groupings k and l; $s_{k,D}$, $s_{l,D}$ = standard deviation of deterministic, diurnal component of demand for home groupings k and l.

The cross correlation of the random noise component of demand, $\varepsilon_k(t)$, between home groupings k and l is computed as

$$r_N = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_k(t)\cdot\varepsilon_l(t)\,/\,s_{k,N}\cdot s_{l,N} \tag{6}$$

where $s_{k,N}$, $s_{l,N}$ = standard deviation of random noise component of demand for home groupings k and l. The expression in (6) assumes that the mean of the random noise component is zero.

3. DATA ANALYSIS AND RESULTS

The cross correlation analysis of the indoor residential demand data was conducted over a number of random trials. In each trial, homes were randomly selected from the pool of 20 homes to form 2 groupings of 10 homes. Throughout this paper, these 2-home groupings will be called Group A and Group B. A total of 10 random trials were generated, as indicated in Table 1.

Table 1. Homes randomly selected to form Groups A and B in 10 trials.

	Home Identifiers — Group A	**Home Identifiers — Group B**
Trial 1	9, 17, 3, 1, 10, 2, 13, 20, 8, 5	15, 12, 11, 14, 19, 6, 7, 21, 4, 18
Trial 2	14, 4, 21, 3, 12, 10, 6, 5, 7, 20	2, 13, 17, 9, 8, 18, 1, 15, 11, 19
Trial 3	17, 19, 13, 21, 12, 6, 1, 2, 14, 20	10, 4, 3, 8, 9, 5, 7, 11, 18, 15
Trial 4	20, 1, 21, 13, 15, 5, 11, 9, 12, 17	6, 10, 4, 3, 7, 8, 2, 18, 14, 19
Trial 5	21, 15, 10, 4, 9, 17, 12, 3, 5, 8	13, 11, 18, 1, 2, 6, 20, 7, 14, 19
Trial 6	3, 5, 20, 10, 6, 11, 8, 2, 1, 7	19, 17, 13, 18, 14, 9, 12, 15, 21, 4
Trial 7	6, 12, 3, 8, 21, 5, 19, 14, 4, 17	7, 18, 13, 11, 1, 9, 20, 2, 15, 10
Trial 8	6, 8, 10, 3, 4, 12, 13, 7, 1, 20	5, 15, 9, 11, 14, 2, 18, 19, 21, 17
Trial 9	13, 11, 14, 3, 5, 12, 19, 10, 9, 15	17, 8, 18, 1, 4, 7, 21, 20, 2, 6
Trial 10	4, 20, 10, 6, 19, 2, 15, 14, 17, 18	12, 1, 5, 7, 13, 3, 9, 11, 21, 8

In each trial, indoor residential water demand was aggregated across the 10 homes selected in Group A and Group B (Table 1). Periodic regression was performed on the aggregated residential water demand of Group A and Group B. The periodic regression produced a deterministic, diurnal demand, $f(t)$, and a random noise (or residual) demand component, $\varepsilon(t)$, for Groups A and B. The time series data of residential demand, $q(t)$, diurnal demand component, $f(t)$, and random noise component, $\varepsilon(t)$, in Groups A and B were used to compute the cross correlation coefficients r, r_D, and r_N. This procedure was repeated in the 10 trials for the 600 s and 3,600 s time step data on October 15, 1997 and October 19, 1997. The results of the analysis are indicated in Tables 2 and 3. These tables include summary statistics on the minimum, maximum, and mean values of cross correlation coefficients. By the central limit theorem, the mean correlation is assumed to follow a normal distribution. The 95% confidence limits on the mean are also reported in these tables.

The criteria followed to determine significant harmonics in periodic regression deserve further discussion. Initially, Fisher's ratio test (Yevjevich 1972) was applied to identify the significant harmonics of regression. Unfortunately, the ratio test yielded inconsistent results and was abandoned. Thus, an *ad hoc* procedure was adopted in its place. It was found that aggregating indoor residential water demand over a 600 s time step produced a diurnal pattern with significant fluctuation (see Fig. 1). For this time step length, indoor residential water demand was regressed with 8 to 12 of the most significant harmonics (not necessarily in sequential order) that explained at least 95% of the variance in indoor residential demand as in (3) (see Fig. 2). The aggregation of indoor residential water demand over a time step of 3,600 s revealed a clear diurnal pattern with well-defined peaks in the early morning and early evening (see Fig. 1). Given this observation, only the first 2 harmonics were used in the regression of indoor residential water demand to describe the 24 h periodicity and the 2 characteristic peaks (see Fig. 2).

The first three research questions posed at the beginning of this paper pertain to the strength of the correlation of the aggregated indoor residential demand, $q(t)$, its deterministic, diurnal component, $f(t)$, and its random noise component, $\varepsilon(t)$. The data in Tables 2 and 3 show that the mean correlation of residential demand is positive and moderate to high (+0.106 to +0.670) on October 15 and 19 for time steps of 600 s and 3,600 s. (Note that in this paper a low correlation spans the range ±0.0–0.3, a moderate correlation spans the range ±0.3–0.6 and a high correlation spans the range ±0.6–1.0). The mean correlation of the deterministic, diurnal component is also positive and moderate to high (+0.299 to +0.840) on October 15 and 19 for time steps of 600 s and 3,600 s. The mean correlation of the random noise component is negative and low (-0.088) on Oct 15 for the 600 s time step data. Otherwise, the mean correlation is positive and low in all other experiments (+0.051 to +0.256). Since aggregated residential demand, $q(t)$, shows an elevated correlation in Tables 2 and 3, it follows that the deterministic, diurnal component, $f(t)$, would show a similarly high level of correlation since it explains a large part of the variation of residential demand, $q(t)$.

The fourth research question asks whether the time step length heavily influences the strength of correlation between the three demand components? The data in Tables 2 and 3 hint that increasing the time step length strengthens the correlation between demands. Indeed, the mean correlation of aggregated residential demand, $q(t)$, increases from +0.106 to +0.358 on Oct 15, 1997 and from +0.412 to +0.670 on Oct 19, 1997 when the time step is increased from 600 s to 3,600 s. Similarly, the mean correlation of the deterministic, diurnal component, $f(t)$, increases from +0.299 to +0.541 on October 15, 1997 and from +0.653 to +0.840 on October 19, 1997 when the time step is increased from 600 s to 3,600 s. It is unclear to what extent the mean correlation in the random noise component, $\varepsilon(t)$, is influenced by time step length. These preliminary results seem to support two propositions: (1) that aggregating demand data over longer time steps tends to enforce a coherent and well-defined diurnal pattern in a grouping of homes, and; (2) that longer time steps tend to increase the similarity and synchronicity of diurnal patterns between the 2 grouping of homes, leading to stronger correlations in demand between these groupings.

Water Distribution Systems Analysis Symposium 2006

Table 2. Cross correlation values for the aggregated residential water demand, the deterministic, diurnal component of demand, and the random noise component of demand on October 15, 1997 for time step lengths 600 s (10 min) and 3,600 s (60 min).

Date: Wednesday, October 15, 1997 **Period: 00:00:01 – 23:59:59**						
	Time Step = 600 s			**Time Step = 3,600 s**		
	$\boldsymbol{r}$	$\boldsymbol{r_D}$	$\boldsymbol{r_N}$	$\boldsymbol{r}$	$\boldsymbol{r_D}$	$\boldsymbol{r_N}$
Trial 1	-0.010	+0.110	-0.149	+0.277	+0.498	-0.087
Trial 2	+0.072	+0.304	-0.163	+0.331	+0.649	-0.084
Trial 3	+0.081	+0.227	-0.085	+0.277	+0.366	+0.069
Trial 4	+0.142	+0.403	-0.098	+0.461	+0.768	+0.001
Trial 5	+0.227	+0.563	-0.079	+0.571	+0.699	+0.257
Trial 6	+0.094	+0.370	-0.101	+0.352	+0.282	+0.324
Trial 7	+0.083	+0.186	-0.059	+0.275	+0.311	+0.125
Trial 8	+0.170	+0.337	-0.016	+0.286	+0.708	-0.168
Trial 9	+0.125	+0.185	+0.015	+0.328	+0.585	-0.073
Trial 10	+0.078	+0.309	-0.145	+0.424	+0.540	+0.141
Min.	-0.010	+0.110	-0.163	+0.275	+0.282	-0.168
Max.	+0.227	+0.563	+0.015	+0.571	+0.768	+0.324
Mean	**+0.106**	**+0.299**	**-0.088**	**+0.358**	**+0.541**	**+0.051**
Std. Dev. of Trials	0.064	0.131	0.057	0.098	0.173	0.161
Lower 95% C.I. on Mean	+0.066	+0.218	-0.124	+0.297	+0.433	-0.049
Upper 95% C.I. on Mean	+0.146	+0.380	-0.052	+0.419	+0.648	+0.150

The fifth research question asks whether day-of-week effects have any bearing on the strength of correlation between the three demand components? Again, the data in Tables 2 and 3 seem to indicate that demands are more strongly correlated on weekends than on weekdays. Specifically, the mean correlation of aggregated residential water demand is +0.106 (600 s time step) and +0.358 (3,600 s time step) on Wednesday, October 15, 1997, while on Sunday, October 19, 1997, the mean correlation increases to +0.412 (600 s time step) and +0.670 (3,600 s time step). Similar increases are observed for the deterministic, diurnal component, f(t), in Tables 2 and 3. The data in these tables make it unclear whether the mean correlation in the random noise component, $\varepsilon(t)$, is influenced to any degree by day-of-week effects. The higher correlation on weekends might be due to the possibility that more people are at home (and thus not at work) and maintain more predictable schedules during weekends. A great number of users on weekends might produce more coherent diurnal patterns in residential water demand and lead to more synchronous water use during weekends. To reach a definitive conclusion on this issue would require extensive end-use modeling (Blokker and Vreeburg 2005) of the 20 homes in order to properly establish instantaneous temporal water use on a fixture-by-fixture basis on weekdays and weekends. An analysis of this kind is beyond the scope of this paper.

Water Distribution Systems Analysis Symposium 2006

Table 3. Cross correlation values for the aggregated residential water demand, the deterministic, diurnal component of demand, and the random noise component of demand on October 19, 1997 for time step lengths 600 s (10 min) and 3,600 s (60 min).

Date: Sunday, October 19, 1997 **Period: 00:00:01 – 23:59:59**						
	Time Step = 600 s			**Time Step = 3,600 s**		
	$\boldsymbol{r}$	$\boldsymbol{r_D}$	$\boldsymbol{r_N}$	$\boldsymbol{r}$	$\boldsymbol{r_D}$	$\boldsymbol{r_N}$
Trial 1	+0.457	+0.704	+0.047	+0.691	+0.832	+0.296
Trial 2	+0.415	+0.771	-0.062	+0.724	+0.897	+0.292
Trial 3	+0.433	+0.614	+0.078	+0.660	+0.734	+0.397
Trial 4	+0.432	+0.690	+0.033	+0.711	+0.755	+0.505
Trial 5	+0.371	+0.609	-0.043	+0.677	+0.873	+0.220
Trial 6	+0.522	+0.732	+0.118	+0.682	+0.902	+0.184
Trial 7	+0.311	+0.436	+0.025	+0.528	+0.760	+0.045
Trial 8	+0.408	+0.687	-0.045	+0.704	+0.925	+0.210
Trial 9	+0.352	+0.564	+0.037	+0.631	+0.805	+0.198
Trial 10	+0.419	+0.727	-0.028	+0.694	+0.918	+0.213
Min.	+0.311	+0.436	-0.062	+0.528	+0.734	+0.045
Max.	+0.522	+0.771	+0.118	+0.724	+0.925	+0.505
Mean	**+0.412**	**+0.653**	**+0.016**	**+0.670**	**+0.840**	**+0.256**
Std. Dev. of Trials	0.058	0.100	0.059	0.057	0.073	0.126
Lower 95% C.I. on Mean	+0.376	+0.592	-0.021	+0.635	+0.795	+0.178
Upper 95% C.I. on Mean	+0.448	+0.715	+0.053	+0.705	+0.885	+0.334

4. DISCUSSION

While it is difficult to make sweeping conclusions on the strength of the data presented in this paper, some general remarks are in order. The limited data collected tentatively suggest that the cross correlation coefficients of aggregated residential demand, $q(t)$, and the deterministic, diurnal component of demand, $f(t)$, were below +1.0 in Milford, OH, on October 15 and 19, 1997. A correlation coefficient of +1.0 signifies a perfect correlation or synchronicity between demands in a network. This result has two possible implications. First, this suggests the need to re-evaluate the oft-used assumption that synchronicity exists between network demands in extended period simulation and that a single peaking factor can be applied to all network demands (or all demands in a single pressure zone). More analysis is needed to reach a firm conclusion on this matter. Second, supposing that in most systems demands are not strongly cross correlated (cross correlation coefficient is well away from the negative and positive extremes of ±1.0), it may be worthwhile for utilities to perform a cross correlation analysis on their demand data to better establish demand patterns for design and rehabilitation work. This could help reduce the difference between demands estimated in design and demands observed in the system once it has been built and thus help utilities avoid over-design costs (building larger pipes and components than needed) or avoid low pressures due to under-design.

While far from comprehensive, the data suggests that cross correlation in demand might be weaker when using a shorter time step. This finding might have implications for water quality modeling where the time step is typically of short duration. Under these circumstances, the assumption of synchronicity in demand is likely specious and can possibly lead to large inaccuracies in water quality predictions since the movement and distribution of chemical constituents in a network depend heavily on the accurate determination of temporal and spatial patterns of demands. Thus, properly establishing the cross correlation between network demands is perhaps important in water quality modeling.

5. FUTURE WORK

The correlation values in Tables 2 and 3 were derived with only 10 random trials. Such a small number of trials make it impossible to reach definitive conclusions about demand correlation in the Milford system. In future work the aim will be to deploy Monte Carlo Simulation to generate a much larger number of trials to reduce the variability in the minimum, maximum, and the mean of the cross correlation coefficients. Working with a large number of trials will also make it possible to conduct hypothesis testing in order to test the statistical significance of cross correlation.

In the present work, an *ad hoc* criteria was used to choose significant regression harmonics that explain the variation in aggregated residential demand. This *ad hoc* procedure is not statistically objective and instead relies on a subjective judgment vis-à-vis what constitutes an appropriate fit. This is important since the cross correlation coefficients computed in (4)-(6) hinge on what harmonics are included in the regression. Thus in future work, hypothesis testing could be used to reach a statistically objective determination as to what harmonics should be included in the regression.

6. SUMMARY AND CONCLUSIONS

The focus of this paper has been on analyzing the cross correlation in indoor residential demand data collected in the city of Milford, Ohio. The paper has sought to answer five research questions: how strongly cross correlated are indoor, residential demands in the Milford system? How strongly correlated is the deterministic, diurnal component of residential demand in the Milford system? How strongly correlated is the random noise component of residential demand in the Milford system? To what extent does time step length influence the strength of correlation between these 3 demand components? Does the correlation between these 3 demand components differ significantly between weekdays and weekends? To begin to answer these questions, a periodic regression model was used to isolate the deterministic, diurnal and the random noise components of the residential demand data collected in Milford, Ohio. Correlation indices were formulated to measure the cross correlation of residential demand, the deterministic, diurnal component of demand, and the random noise component of demand. The Milford data was analyzed for cross correlation and the results pointed to a number of preliminary findings:

1. Residential demand and its deterministic, diurnal component showed a positive and moderate to high correlation, while the random noise component of demand showed a low level of correlation.

2. Increasing the time step length (from 600 s to 3,600 s) seemed to increase the strength of the correlation in residential demand and its deterministic, diurnal component. This suggested that a

Water Distribution Systems Analysis Symposium 2006

longer time step may increase the coherence in diurnal demand patterns as well as their synchronicity. It was unclear whether time step length had any influence on the strength of correlation in the random noise component of demand.

3. Residential demand and its deterministic, diurnal component were more strongly correlated during weekends than during weekdays. This finding suggested that weekend periods may be characterized by less erratic water use patterns that produce more coherent and more synchronous diurnal patterns. It was unclear whether the random noise component was influenced by day-of-week effects.

4. Further analysis is required with a larger number of trials to determine the statistical significance of the results.

The first finding suggests that the oft-used assumption of synchronicity between diurnal demands in extended period simulation may need to be reconsidered. By performing a cross correlation analysis of demand data, utilities may be able to obtain a more accurate picture of representative spatial and temporal demand patterns for use in design and rehabilitation work. The second finding suggests that diurnal demands are only weakly correlated when a short time step is used. In water quality studies where a short time step is typically used, it is even more critical to understand to what extent diurnal demands are correlated in space since the movement of chemical constituents depends heavily on the temporal and spatial patterns of demand.

7. REFERENCES

Babayan, A.V., Savic, D.A., Walters, G.A. (2005). “Multiobjective Optimization for the Least-Cost Design of Water Distribution System Under Correlated Uncertain Parameters.” *EWRI 2005: Impacts of Global Climate Change*, EWRI, Anchorage, Alaska, May 15-19, 2005.

Blokker, E.J.M., Vreeburg, J.H.G. (2005). “Monte Carlo Simulation of Residential Water Demand: A Stochastic End-Use Model.” *EWRI 2005: Impacts of Global Climate Change*, EWRI, Anchorage, Alaska, May 15-19, 2005.

Buchberger, S.G., Wells, G.J. (1996). “Intensity, Duration, and Frequency of Residential Water Demands.” *Journal of Water Resources Planning and Management*, 122(1), 11-19.

Filion, Y.R., Karney, B.W., Adams, B.J. (2005). “Stochasticity of Demand and Probabilistic Performance of Water Networks.” *EWRI 2005: Impacts of Global Climate Change*, EWRI, Anchorage, Alaska, May 15-19, 2005.

Filion, Y.R., Adams, B.J., Karney, B.W. (2006). “Cross Correlation of Demands in Water Distribution Network Design.” *Journal of Water Resources Planning and Management*, ASCE, In Press.

Kapelan, Z., Savic, D.A., Walters, G.A. (2005). “An Efficient Sampling-Based Approach for the Robust Rehabilitation of Water Distribution Systems Under Correlated Nodal Demands.” *EWRI 2005: Impacts of Global Climate Change*, EWRI, Anchorage, Alaska, May 15-19, 2005.

Tolson, B.A., Maier, H.R., Simpson, A.R., Lence, B.J. (2004). “Genetic Algorithms for Reliability-Based Optimization of Water Distribution Systems.” *Journal of Water Resources Planning and Management*, 130(1), 63-72.

Water Distribution Systems Analysis Symposium 2006

Yevjevich, V. (1972). *Stochastic Processes in Hydrology*. Water Resources Publications, Fort Collins, Colorado.

Zhang, X., Nadimpalli, G., Buchberger, S. (2005). *Cumulative SERPS User's Manual*. Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio.