Water Air Soil Pollut (2012) 223:4841–4851
DOI 10.1007/s11270-012-1240-3

Temporal Synchronization Analysis for Improving Regression Modeling of Fecal Indicator Bacteria Levels

Michael Cyterski · Shuyan Zhang · Emily White · Marirosa Molina · Kurt Wolfe · Rajbir Parmar · Richard Zepp

Received: 9 September 2011 / Accepted: 12 June 2012 / Published online: 4 July 2012

Abstract Multiple linear regression models are often used to predict levels of fecal indicator bacteria (FIB) in recreational swimming waters based on independent variables (IVs) such as meteorologic, hydrodynamic, and water-quality measures. The IVs used for these analyses are traditionally measured at the same time as the water-quality sample. We investigated the improvement in empirical modeling performance by using IVs that had been *temporally synchronized* with the FIB response variable. We first examined the univariate relationship between multiple "aspects" of each IV and the response variable to find the single aspect of each IV most strongly related to the response. Aspects are defined by the temporal window and lag (relative to when the response is measured) over which the IV is averaged. Models were then formed using the "best" aspects of each IV. Employing iterative cross-validation, we examined the average improvement in the mean squared error of prediction, MSEP, for a testing dataset after using our temporal synchronization technique on the training data. We compared the MSEP values of three methodologies: predictions made using unsynchronized IVs (UNS), predictions made using synchronized IVs where aspects were chosen using a Pearson correlation coefficient (PCC), and predictions using IV aspects chosen using the PRESS statistic (PRS). Averaging over 500 randomly generated testing datasets, the MSEP values using the PRS technique were 50 % lower ($p<0.001$) than the MSEP values of the UNS technique. The average MSEP values of the PCC technique were 26 % lower ($p<0.001$) than the MSEP values of the UNS technique. We conclude that temporal synchronization is capable of significantly improving predictive models of FIB levels in recreational swimming waters.

Keywords Multiple linear regression · Fecal indicator bacteria modeling

M. Cyterski (✉) · S. Zhang · E. White · M. Molina · K. Wolfe · R. Parmar · R. Zepp
United States Environmental Protection Agency (USEPA), Office of Research and Development, National Exposure Research Laboratory, Ecosystems Research Division, 960 College Station Road, Athens, GA 30605, USA
e-mail: cyterski.mike@epa.gov

Present Address:
E. White
Department of Chemistry, The University of the South, Sewanee, TN, USA

1 Introduction

Protecting public health by limiting exposure to harmful pathogens at recreational swimming waters is the goal of a variety of local, state, and federal agencies. Unfortunately, it is difficult to definitively and quickly measure the levels of these pathogens to meet this objective. Concentrations of fecal indicator bacteria (FIB) are monitored at these sites instead of the actual illness-inducing pathogens, but the time required for current laboratory methods (culturing and counting colonies on agar plates) to measure FIB levels does not allow managers to make immediate decisions regarding beach closings on a daily basis. New techniques are emerging to allow for more rapid FIB assays (Griffith and Weisberg 2009; Lee et al. 2010), but it will take time to perfect and distribute these new technologies. Historically, empirical/statistical models have frequently been used to predict FIB levels in recreational waters (Borsuk et al. 2002; Olyphant 2005; Benham et al. 2006; Hou et al. 2006; Boehm et al. 2007; Nevers et al. 2007; Gronewold et al. 2009). These models use nearby environmental conditions to generate expected FIB concentrations for the day. If the predicted FIB concentration exceeds some accepted standard, the manager closes the beach in the interest of protecting public health. The accuracy and dependability of these empirical models thus receives much attention, and much published research seeks to improve techniques for modeling FIB concentrations (Olyphant et al. 2003; Olyphant and Whitman 2004; Nevers and Whitman 2005; Francy and Darner 2006; Ge and Frick 2007; Frick et al. 2008; Holtschlag et al. 2008; Telech et al. 2009). These efforts usually focus on variable/model selection techniques, and/or examining what independent variables (IVs) prove to be most useful at various sites.

A common statistical approach is to use multiple linear regression (MLR) to predict FIB levels. The IVs in MLR models are a combination of meteorological, hydrodynamic, and water-quality measures taken either near the beach or at the site itself. For most IVs, a value is recorded at the moment the water sample for FIB levels is taken, but rainfall is a noted exception. Cumulative precipitation over some period prior to the FIB water sample is often used as an IV in predictive beach bacteria modeling efforts (Hose et al. 2005; Neumann et al. 2006; Francy and Darner 2006; Nevers et al. 2009). Complicating efforts to develop useful MLR models is that FIB time series can exhibit complex temporal structure (Boehm et al. 2002). Investigating the characteristics of the temporal structure of both FIB measurements and environmental correlates has lead to improved empirical modeling. Ge and Frick (2009) present a wavelet decomposition analysis for non-stationary time series of FIB and candidate explanatory environmental data from Huntington Beach, OH. They found differences in scale, phase angles, and duration of temporal patterns in the time series they examined and used this information to reduce predictive error via the development of a wavelet-transformed MLR analysis. Here, we present an alternative approach to discover and adjust for differences in temporal patterns within FIB and environmental data, leading to better MLR models.

Recognition that the predictive power of rainfall is improved by examining its sum over a temporal window, we hypothesized that regression models might be improved if all potential IVs were processed in this manner, rather than relying on the instantaneous value on an IV taken at the time of FIB sampling. In addition, the fact that previous research (Boehm et al. 2002; Ge and Frick 2009) found evidence of temporal phase shifts and incongruities between water-quality data series and environmental variables, we deduced that incorporating a temporal lag could increase the relationship between an IV and the FIB response. In the field of time series analysis and signal processing, there are a number of methods that address phase differences (either temporal or spatial) between multiple data series. The cross-correlation function (CCF) estimates the correspondence between two series when their relationship is subject to a temporal lag (Jenkins and Watts 1968). Spatial cross-correlation is also used in image processing to sharpen and clean distorted data (Yan 1993). Distributed-lag models incorporate a time lag when developing a regression relationship between independent and dependent data series, defining the dependent variable as a function of previous values of the independent variables (Cromwell et al. 1994). The Granger causality test seeks to determine the significance of one time series for predictions of another series, and relies on lagged values of the independent series to perform the test (Granger 1969).

To summarize, our proposed methodology combines the idea that the dependent and independent data series may be temporally phase-shifted, with the notion that some function (mean, sum, standard deviation, etc.) of the independent variables over a temporal window may produce a better correspondence to the dependent variable. We named this approach *temporal synchronization analysis* (TSA). Essentially, TSA seeks to maximize the correspondence between each IV and the FIB response by examining different combinations of temporal windows and lags over which the mean value of the IV is calculated. These windows and lags are defined relative to the time at which the FIB concentration is measured. We set out to test whether a TSA could improve both the ability of a multiple linear regression (MLR) model to fit a dataset, and more importantly, the ability of an MLR model to make predictions of new observations.

2 Methods

2.1 Study Site and Data Collection

South Shore Park in Milwaukee, Wisconsin is located within the residential area of Bayview on the west coast of Lake Michigan. The park includes South Shore Beach, a public beach area with 150 m of sandy shoreline. The beach is located adjacent to the South Shore Yacht Club and a paved parking area, which drains into the lake. A rock embankment juts into the lake, separating the sandy beach area from a cobble/pebble beach area with a high sloping shore. The entire beach and marina are partially enclosed by an offshore breakwall, which limits wave action, water circulation, and exchange with the outer harbor. The beach is about 4 km south of Milwaukee Harbor, where the Milwaukee Metropolitan Sewerage District Jones Island Water Reclamation Facility is located. The Milwaukee, Menomonee, and Kinnickinnic rivers also discharge to Lake Michigan inside the Milwaukee Harbor breakwall.

Historically, South Shore has poor water quality with 34 % of samples collected from 2003 to 2009 exceeding water-quality criteria standards. Potential sources of fecal contamination include combined sewer overflows (CSOs), urban/suburban and agricultural runoff from the Milwaukee River Basin, runoff from impervious surfaces including parking lots and the beach face, and gulls (Scopel et al. 2006). A detailed spatial assessment found that poor beach water quality was for the most part a local phenomenon with contamination originating at the shoreline (McLellan and Salmore 2003). The beach is routinely monitored for *Escherichia coli*, during the 14-week swimming season (late May to early September) by the Milwaukee County Health Department. Water quality (water temperature, specific conductance, pH, dissolved oxygen, chlorophyll, and turbidity) and meteorological conditions (air temperature, rainfall, wind speed, and wind direction) are continuously recorded by a sonde and weather station maintained by the University of Wisconsin-Milwaukee Great Lakes WATER Institute. Data are transmitted every hour (weather) or half hour (water quality) to a website via Ethernet communication. The sonde is deployed just northwest of the beach, off the east end of the South Shore Yacht Club dock and the weather station is installed on top of a post, at the beach near the boat launch. Additional meteorological data can be obtained from a National Climatic Data Center (NCDC) weather station at the General Mitchell International Airport in Milwaukee. Real-time and historical bacterial (since 2003), meteorological, and in-lake sonde data are publically available through the Wisconsin Beach Health website (http://www.wibeaches.us).

Previous modeling efforts of South Shore Beach include hydrodynamic and water-quality models developed to describe the fate and transport of fecal coliform in Milwaukee Harbor and nearshore Lake Michigan. Modeling results indicated that the fecal coliform load from rivers and CSO/SSO events had only a slightly greater than marginal impact on the beach site, with local sources (e.g., stormwater runoff and birds) being more important (MMSD 2005). South Shore Beach was also one of the 55 beaches included in a regional forecast model for southern Lake Michigan, from Milwaukee, Wisconsin to Michigan City, Indiana (Whitman and Nevers 2005). Adoption of a predictive model is currently under consideration by the City of Milwaukee. In addition to issuing advisories based on the monitoring of *E. coli* levels (where <235 CFU/100 mL is acceptable, 235–1,000 CFU/100 mL results in a water-quality advisory, and >1,000 CFU/100 mL results in a closure advisory) a rainfall threshold of 2.5 cm in 24 h is also used at South Shore to predict poor water-quality and issue advisories.

In the summer of 2008, we collected water samples to measure FIB (*Enterococci* sp.) concentrations from three sites along the beach, along with gathering water-quality information from an additional sonde station as well as an ultraviolet radiation sensor (Fig. 1). For each sampling event, a beach-wide FIB concentration was calculated by taking the mean value across the three sampling sites. Concentrations of *Enterococci* were analyzed using standard membrane filtration methodologies (EPA Method 1600, USEPA 2006) and reported as Colony-Forming Units (CFU)/100 ml. For this study, we focused on modeling the FIB in waist deep water samples (taken from approximately 0.3 m underneath the surface of the water), although shin-deep samples were also gathered on occasion. Samples were taken 3 or 4 days each week, and on these days, samples were taken at three times: 9:00 am, 11:30 am, and 3:00 pm. Based on preliminary analyses showing low correlations, these intra-day samples were treated as independent observations. We used the natural logarithm of CFU measurements for modeling purposes. To avoid the effects of multi-collinearity on MLR results, the IVs were filtered prior to analysis to remove sets of highly correlated IVs (Pearson correlation coefficient >0.8). After data preparation, 44 FIB observations and 13 IV's were retained. The IV's were: water temperature, conductivity, water depth, pH, turbidity, chloride, NH_4, NO_3, dry-bulb air temperature, wet-bulb air temperature, relative humidity, wind speed, and wind direction. Wind speed and direction were combined and transformed into two spatial components: parallel to the shore (A component) and perpendicular to the shore (O component).

2.2 Cross-Validation

To begin the analysis, we randomly split the raw data into two parts: 75 % of the 44 observations (33) were used as training data, and the methods described in Sections 2.3 and 2.4 were applied. The remaining 25 % of the observations (11) were set aside as testing data. This randomization was done 500 times to create a population of 500 paired training/testing datasets that could subsequently be analyzed.

2.3 Temporal Synchronization

Each FIB measurement has a "time stamp" in the dataset. This is a real number that represents the day and time each observation was collected. For instance, an observation made on July 13, 2008, at 9:00 am would have a time stamp of 39,642.375. The integer 39,642 is the number of days since day 1 (January 1, 1900, in our convention, but this could be any day for calculation purposes). The decimal 0.375 represents the time of day (9/24 h). The time stamp for every IV measurement was also known. However, the IV's were measured much more frequently than the response variable, typically at intervals of every 5, 15, or 30 min. In creating a traditional empirical modeling dataset, one would match up the response variable to the corresponding value for each IV, i.e., the IV value whose time stamp equals (or is closest to) the time stamp of the FIB response. For example, if the FIB response were measured at 9 am on June 18th, you would match that up with the water temperature measured at 9 am on June 18th, the wind speed measured at 9 am on June 18th, and so on.

For this synchronization analysis, we wrote R code (R Development Core Team 2008) that would generate not merely a single IV value for each FIB measurement, but many different values for each IV. These various IV values were created by averaging the IV measurements over a temporal window and lag with respect to the time stamp of the response variable (Fig. 2). We used four windows (0, 0.5, 1.5, and 3 days) and four lags (0, −1, −1.5, and −3 days; the negative lags indicating that we used the IV values taken prior to the response being measured). Note that the 0-day window and 0-day lag corresponds to a traditional approach to FIB regression modeling mentioned in the previous paragraph. We called each window–lag combination an "aspect," and thus generated 16 aspects (4 windows×4 lags) for each independent variable.

Next we set out to choose a single aspect (out of the possible 16) for each IV for modeling the response variable. We considered two criteria to determine the aspect selection: a Pearson correlation coefficient between the response variable and each aspect, with the highest coefficient indicative of the best linear relationship; and the predicted residual sum of squares (PRESS, Draper and Smith 1981) statistic calculated from a univariate regression of the response on each potential IV aspect. PRESS is defined as:

$$\sum_i \widehat{\varepsilon}_i^2 = (\widetilde{y}_i - y_i)^2$$

where $\widetilde{y}_i$ is the predicted response based on a regression model formulated without using case i, and y_i is the ith observation. The Pearson coefficient emphasizes the best fit of a model to the data, while the PRESS statistic emphasizes prediction of observations not seen by the model. Lower PRESS values indicate greater predictive accuracy.

2.4 Model Selection

At this point, each of the 500 training datasets had been changed into three datasets: one where none of the IV's had been synchronized (the UNS dataset); one "synchronized" dataset that had IV aspects chosen by a Pearson coefficient (the PCC dataset); and another "synchronized" dataset where IV aspects were chosen using the PRESS statistic (the PRS dataset). We then

Fig. 1 South Shore Beach, Milwaukee, with indications for the locations of the three FIB sampling locations and the placement of monitoring equipment

used multiple linear regression methods to develop predictive models of FIB concentrations using these three datasets. In developing our MLR models, two different selection procedures were used:

1. For the UNS and PCC datasets, we used the Akaike Information Criterion (AIC, Akaike 1974) as a variable selection criterion.
2. For the UNS, PCC, and PRS datasets, we used the PRESS statistic to perform model selection.

For each of these analyses, we employed a backwards-stepwise variable selection algorithm. That is, we started with the full model of 13 IVs, and at each step, an IV may be removed or added back into the model, depending on changes in the PRESS and/or AIC criterion.

2.5 Measuring Model Performance

To measure how well a chosen model fit its corresponding training data, we calculated the mean squared error (MSE) using the 33 training observations. The MSE is the average of the sum of all 33 squared differences between each observation and its value estimated by the model. However, to measure predictive model performance (procedure 2 in Section 2.4), a model was selected using the 33 observations in the training dataset, and then a mean squared error of prediction (MSEP) was calculated for the 11 observations in the testing dataset. The MSEP is the average of the sum of all 11 squared differences between each observation in the testing dataset and its value predicted by the model.

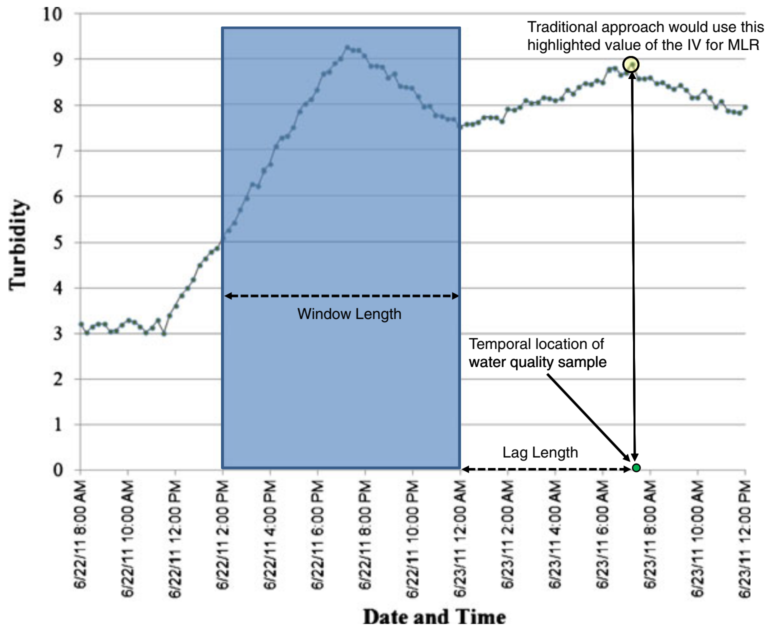

Fig. 2 Diagram showing how temporally synchronized IVs are calculated using windows and lags relative to time at which the response variable is collected

To summarize, the following four-step process was repeated 500 times:

1. Seventy-five percent to 25 % randomized splitting of the original raw data into training and testing sub-groups;
2. IV aspect selection (both Pearson coefficient and PRESS) using the training data;
3. Model section (PRESS or AIC) using the training data;
4. Model performance metrics calculated (MSE for training data, MSEP for testing data).

Afterwards, we made the following comparisons:

a. MSE of UNS vs. PCC (procedure 1, Section 2.4)
b. MSEP of UNS vs. PRS (procedure 2, Section 2.4)
c. MSEP of UNS vs. PCC (procedure 2, Section 2.4)
d. MSEP of PCC vs. PRS (procedure 2, Section 2.4)

These comparisons would allow us to measure the effectiveness of TSA for improving MLR fitting and predictive capabilities of the best-fit models.

3 Results and Discussion

3.1 MSE of UNS vs. PCC

Across the 500 datasets (Table 1), the mean MSE for models developed using temporally synchronized data (IV aspects chosen using correlation coefficients) was 35 % smaller than the mean MSE using unsynchronized data (0.513 vs. 0.795, $p<0.001$). When the statistical objective is fitting a set of training data, rather than prediction of new observations, selecting IV aspects using the maximum correlation coefficient should be sufficient.

3.2 MSEP of UNS vs. PRS

Figure 3 shows the 500 MSEP values for models using the PRS data (solid line) and the UNS data (dashed line). The mean MSEP for the PRS data was 1.85, and for the UNS data was 3.67, nearly twice as large ($p<0.001$). For many datasets, the UNS data produced an MSEP similar to that of the PRS data, but there were also many spikes seen, where the MSEP for the UNS data was much higher than that of the PRS data. We can conclude that temporal synchronization helped to improve predictions of new observations when fitting a multiple linear regression model.

3.3 MSEP of UNS vs. PCC

The MSEP values produced by models fit to the PCC datasets are also smaller, on average, than the mean MSEP values for the UNS datasets (Table 1). The mean MSEP value for the PCC datasets was 2.71, versus 3.67 for the UNS datasets ($p<0.001$).

Table 1 Statistics for the 500 MSE and MSEP values obtained from models developed using temporally synchronized data (both PRESS-selected IV aspects, PRS, and correlation coefficient-selected IV aspects, PCC) and unsynchronized data (UNS)

Metric	Dataset	Mean	Standard deviation	95 % C.I. on the mean	
				Lower bound	Upper bound
MSE	UNS	0.795a	0.206	0.777	0.813
	PCC	0.513b	0.137	0.501	0.525
MSEP	UNS	3.67a	4.99	3.23	4.11
	PCC	2.71b	1.29	2.59	2.82
	PRS	1.85c	0.94	1.76	1.93

The natural logarithm of *Enterococci* CFU measurements were modeled in this study. A difference in the lower case letters associated with the mean MSE and MSEP values indicate a significant difference at α=0.05 confidence level

3.4 MSEP of PCC vs. PRS

The MSEP values for the PRS datasets were indeed smaller, on average, than the MSEP values for the PCC datasets (Table 1, 1.85 versus 2.71, $p<0.001$). Thus, choosing IV aspects using a method that emphasizes predictive modeling (PRESS), will result in better predictions than synchronizing data (choosing IV aspects) using a simple correlation coefficient that emphasizes data fitting.

We found a statistically significant difference between the mean MSE of the UNS and PCC datasets, as well as the mean MSEP values of the UNS, PCC, and PRS datasets. The MSEP values obtained from models developed using temporally synchronized data with PRESS-selected IV aspects were lower than from models developed using temporally synchronized data with correlation coefficient-selected IV aspects, and both were lower than models developed using unsynchronized data.

3.5 Effects of TSA on Variable Significance

These results show that the TSA is a useful technique for reducing the MSE for fitting data, as well as improving predictive performance (as measured by the MSEP). However, some additional questions concerning the consistency of the technique remain. Are the models selected in each of the 500 dataset randomizations similar to each other, i.e., are the same variables significant from randomization to randomization? To answer this, we examined the number of times each variable was chosen to be included in the final model of the 500 training datasets (Table 2). Nine variables were more significant contributors after synchronization, while four IVs were significant less often after temporal synchronization. Overall, synchronization reduced the variability in IV significance. The standard deviation of the

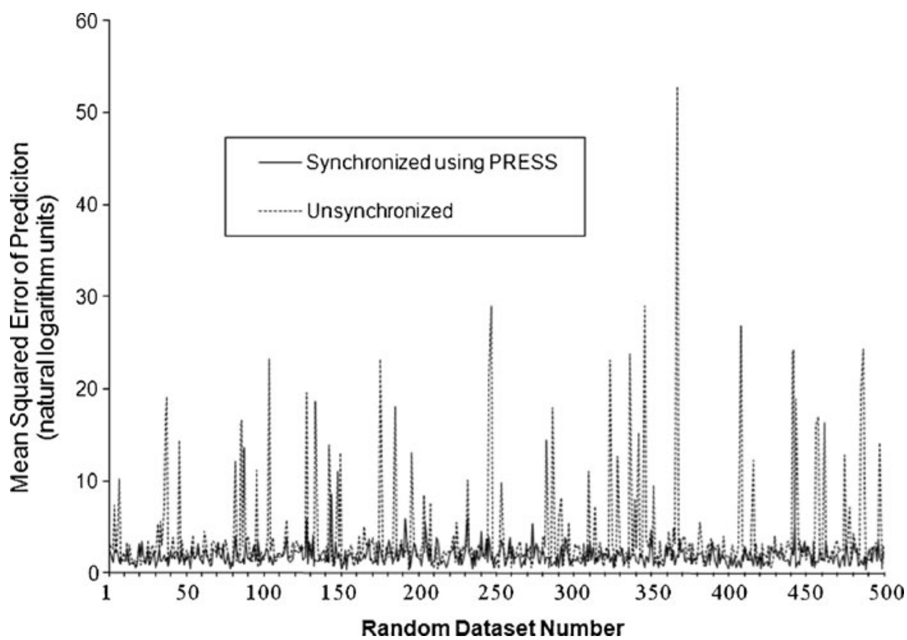

Fig. 3 A comparison between the MSEP values of models developed using temporally synchronized data (IV aspects selected using PRESS residuals) and unsynchronized data. Five hundred randomly generated training datasets (with 33 observations) were examined. Each testing dataset had 11 observations. Note that we modeled the natural logarithm of *Enterococci* CFU measurements in these models

UNS column in Table 2 is 119, while it is 89 for the PRS column. Temporal synchronization increased the significance of the seven least-significant UNS variables (the first seven IVs of Table 2), and reduced the overall significance of four of the six most significant UNS variables.

3.6 Analysis of Window–Lag Characteristics

We also asked if, in terms of the IV aspect selections, the same windows and lags were typically selected as "best" for each variable. If so, interpretation of these aspects becomes more reasonable. When an IV appeared in the final model of any of these randomly created datasets, we recorded its aspect (i.e., its window–lag combination, Table 3). Seven IVs (conductivity, chlorophyll, NH4, NO3, dry-bulb air temperature, humidity, and wind U) have one strong peak corresponding to a single aspect, while four IVs (water temperature, water depth, wet-bulb air temperature, and wind V) have two spikes. Only pH and turbidity show three or more distinct peaks. These results indicate that window–lag selections have some consistency for most of the IVs across the dataset randomizations. Watershed source-tracking models based on hydrodynamic and hydrologic principles would have to be examined to see if there is any physical explanation for the predictive capabilities of certain window–lag combination for each IV. Only three IVs (water depth, NH_4, and NO_3) had L_0W_0 (the traditional instantaneous measurement of the IV) as the strongest aspect for predictive modeling purposes, indicating that temporal synchronization improved the predictive power of 77 % (10/13) of the IVs.

Table 2 Number of times each independent variable was included in the final model using the 500 training datasets

Independent variable	UNS	PRS	Sync effect
Conductivity	67	215	+
Wind O	90	147	+
pH	93	230	+
NH_4	96	305	+
NO_3	109	160	+
Turbidity	124	185	+
Dry-bulb air temp	204	221	+
Wet-bulb air temp	210	136	−
Chlorophyll	228	169	−
Wind A	234	223	−
Water temperature	242	376	+
Humidity	337	397	+
Water depth	486	349	−

The sync effect column indicates if synchronization increased the importance of the given IV (+) or decreased that IV's significance (−)

3.7 Methodological Comments

In this TSA of South Shore Beach data, for simplicity and clarity, we investigated a limited number of windows and lags, the longest of either being 72 h. Expanding this to a week or more, if the data allow, could produce additional predictive improvements. However, TSA is a brute force method without any elegant analytical solution for the "best" window and lag for each IV. There are an infinite number of window/lag combinations that could be explicitly tested, and there is no a priori means of determining which may be the best. Computational time may increase substantially as the number of windows and lags investigated increases.

We computed a mean value of the independent variables over each window, but other statistics, such as the median, standard deviation, maximum or minimum value of the IVs might be more potent correlates to the response variable. Maximizing the Fisher Information (Lehmann and Casella 1998) of the IV over various temporal windows could also possibly lead to better predictive models.

The TSA can be likened to a transformation of each IV, but instead of using a logarithmic or square root transformation to linearize the relationship between the IV and the response, we are temporally transforming the IV to discover if windows and lags can improve its predictive power. These most useful windows and lags would most certainly be dependent on site-specific relationships of pathogen sources to each individual beach site. The best aspect of each IV might even vary annually or seasonally, again depending on site-specific dynamics of pathogen fate and transport. Because the TSA allows the analyst to choose different aspects for each IV (i.e., independent of the other IVs), the technique provides a great deal of flexibility in fitting the FIB data. The TSA basically pre-processes the IV data to find patterns that correlate most highly with the dependent

Table 3 The frequency that each window–lag combination was seen as the "best" aspect for each IV in the selected models for the 500 PRS datasets

Lag–window combination	Water temperature	Conductivity	Water depth	pH	Turbidity	Chlorophyll	NH_4	NO_3	Dry-bulb air temperature	Wet-bulb air temperature	Humidity	Wind O	Wind A
L_0W_0	7	4	119	18	0	1	260	103	169	6	0	2	0
$L_0W_{0.5}$	3	21	2	4	10	0	1	0	29	0	16	45	0
$L_0W_{1.5}$	1	0	222	0	55	0	0	0	0	0	277	5	0
L_0W_3	1	1	6	0	2	0	0	0	0	0	80	0	0
$L_{0.5}W_0$	6	0	0	23	1	0	0	4	3	0	21	0	0
$L_{0.5}W_{0.5}$	4	0	0	1	0	0	0	0	0	0	0	0	0
$L_{0.5}W_{1.5}$	0	0	0	3	0	0	0	0	0	0	0	0	0
$L_{0.5}W_3$	13	4	0	11	0	0	0	0	4	22	0	1	2
$L_{1.5}W_0$	22	1	0	0	26	0	0	4	10	56	2	1	0
$L_{1.5}W_{0.5}$	2	0	0	7	2	0	0	6	0	5	0	4	0
$L_{1.5}W_{1.5}$	0	0	0	11	0	0	0	3	2	36	0	0	0
$L_{1.5}W_3$	133	1	0	30	0	5	0	1	0	2	0	7	188
L_3W_0	20	151	0	19	67	0	4	29	0	1	1	0	0
$L_3W_{0.5}$	4	6	0	71	10	8	0	0	2	3	0	0	0
$L_3W_{1.5}$	148	1	0	29	5	0	2	0	0	0	0	0	33
L_3W_3	12	25	0	3	7	155	38	10	2	5	0	82	0

The sum of each column equals the numbers in the "PRS" column of Table 2. The terms in the leftmost column indicate the lag-window combination of each IV. For example, $L_{1.5}W_{0.5}$ means the IV was averaged over a window of 0.5 days (12 h) using a lag of 1.5 days (36 h prior to water-quality sample)

data series. There is no reason the analyst must follow up the TSA with a multiple linear regression model. Other statistical techniques, such as Partial Least Squares regression, Gradient Boosting, or even neural networks, could then be used to quantify the relationship between the dependent data and the IVs.

The TSA is data intensive and will not be feasible in all circumstances. It relies on temporally dense IV data relative to the FIB water-quality measurements, but we note that meteorological data from monitoring stations such as National Climatic Data Centers are readily available over the internet and are typically collected every 10 or 15 min. Water-quality measurements on such a fine temporal scale are not usually available unless additional instrumentation has been deployed at the site, which could limit the usefulness of the TSA technique for small municipalities that lack the funding for intensive beach monitoring programs.

Although this analysis was only done for a single beach site, enclosed beaches like South Shore are known to be especially susceptible to contamination (Grant and Sanders 2010). Understanding the factors driving contamination and developing methods that can successfully model FIB levels is most critical for the protection of public health at these recreational water use sites.

4 Conclusions

There is recognition that an environmental variable such as rainfall should be examined over some temporal window for effects on water-quality measurements to become evident, and the existence of complex temporal structures within both FIB and IV time series has been documented. This study shows that manipulating all candidate IVs using the TSA can lead to better predictions for FIB empirical modeling applications. Only three out of thirteen IVs (water depth, NH_4, and NO_3) had L_0W_0 (corresponding to the traditional instantaneous measurement of the IV at the time a FIB water sample is collected) as the strongest aspect for predictive modeling purposes. Ideally, empirical modeling results from analyses such as these can provide insight for management of watershed pathogen sources to limit exposure to contamination at beaches.

Acknowledgments The authors wish to thank Michael Tryby (USEPA) and anonymous referees for insightful review comments. We acknowledge the support of the City of Milwaukee Health Department, South Shore Yacht Club, and Daniel Feinstein. This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Benham, B. L., Baffaut, C., Zeckoski, R. W., Mankin, K. R., Pachepsky, Y. A., Sadeghi, A. M., Brannan, K. M., Soupir, M. L., & Habersack, M. J. (2006). Modeling bacteria fate and transport in watersheds to support TMDLs. *Transactions of the ASABE, 49*, 987–1002.

Boehm, A. B., Grant, S. B., Kim, J. H., Mowbay, S. L., McGee, C. D., Clark, C. D., Foley, D. M., & Wellman, D. E. (2002). Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environmental Science and Technology, 36*, 3885–3892.

Boehm, A. B., Whitman, R. L., Nevers, M. B., Hou, D., & Weisberg, S. B. (2007). Now-casting recreational water quality. In L. Wymer & A. Dufour (Eds.), *Statistical Framework for Water Quality Criteria and Monitoring*. West Sussex: Wiley and Sons.

Borsuk, M. E., Stow, C. A., & Reckhow, K. H. (2002). Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. *Environmental Science and Technology, 36*, 2109–2115.

Cromwell, J., Hannan, M., Labys, W., & Terraza, M. (1994). *Multivariate tests for time series models. Quantitative Applications in the Social Sciences Series, No. 100*. Thousand Oaks: Sage Publications, Inc.

Draper, N., & Smith, H. (1981). *Applied regression analysis*. New York: Wiley.

Francy, D. S. & Darner, R. A. (2006). *Procedures for Developing Models to Predict Exceedances of Recreational Water-Quality Standards at Coastal Beaches. USGS Techniques and Methods Report 6-B5*.

Frick, W. E., Ge, Z., & Zepp, R. G. (2008). Nowcasting and forecasting concentrations of biological contaminants at beaches: a feasibility and case study. *Environmental Science and Technology, 42*, 4818–4824.

Ge, Z., & Frick, W. E. (2007). Some statistical issues related to multiple linear regression modeling of beach bacteria concentrations. *Environmental Research, 103*, 358–364.

Ge, Z., & Frick, W. E. (2009). Time-frequency analysis of beach bacteria variations and its implication for recreational water quality modeling. *Environmental Science and Technology, 43*, 1128–1133.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424–438.

Grant, S., & Sanders, B. (2010). Beach boundary layer: a framework for addressing recreational water quality impairment at enclosed beaches. *Environmental Science and Technology, 44*(23), 8804–8813.

Griffith, J. F. & Weisberg, S. B. (2009). *Evaluation of Rapid Microbiological Methods for Measuring Recreational Water Quality. Southern California Coastal Water Research Project Technical Report 485*.

Gronewold, A. D., Qian, S. S., Wolpert, R. L., & Reckhow, K. H. (2009). Calibrating and validating bacterial water quality models: a Bayesian approach. *Water Research, 43*, 2688–2698.

Holtschlag, D. J., Shively, D., Whitman, R. L., Haack, S. K. & Fogarty, L.R. (2008). *Environmental factors and flow paths related to Escherichia coli concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005. USGS Scientific Investigations Report 2008–5028*. Reston, VA.

Hose, G., Gordon, G., McCullough, F., Pulver, N., & Murray, B. (2005). Spatial and rainfall related patterns of bacterial contamination in Sydney Harbour estuary. *Journal of Water and Health, 3*, 349–358.

Hou, D., Rabinovici, S., & Boehm, A. B. (2006). *Enterococci* predictions from a partial least squares regression model can improve the efficacy of beach management advisories. *Environmental Science and Technology, 40*, 1737–1743.

Jenkins, G., & Watts, D. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day.

Lee, C., Griffith, J., Kaiser, W., & Jay, J. (2010). Covalently linked immunomagnetic separation/adenosine triphosphate technique (Cov-IMS/ATP) enables rapid, in-field detection and quantification of *Escherichia coli* and *Enterococcus* spp. in freshwater and marine environments. *Journal of Applied Microbiology, 109*, 324–333.

Lehmann, E., & Casella, G. (1998). *Theory of point estimation*. New York: Springer.

McLellan, S. L., & Salmore, A. K. (2003). Evidence for localized bacterial loading as the cause of chronic beach closings in a freshwater marina. *Water Research, 37*(11), 2700–2708.

MMSD (Milwaukee Metropolitan Sewerage District). (2005). *Bacteria source, transport and fate study—phase 1, Milwaukee Harbor Estuary hydrodynamic and bacteria modeling*. Milwaukee: Milwaukee Metropolitan Sewerage District.

Neumann, C. M., Harding, A. K., & Sherman, J. M. (2006). Oregon beach monitoring program: bacterial exceedances in marine and freshwater creeks/outfall samples, October 2002–April 2005. *Marine Pollution Bulletin, 52*, 1270–1277.

Nevers, M. B., & Whitman, R. L. (2005). Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan. *Water Research, 39*, 5250–5650.

Nevers, M. B., Whitman, R. L., Frick, W. E., & Ge, Z. (2007). Interaction and influence of two creeks on *Escherichia coli* concentrations of nearby beaches: Exploration of predictability and mechanisms. *Journal of Environmental Quality, 36*, 1338–1345.

Nevers, M. B., Shively, D., Kleinheinz, G., McDermott, C., Schuster, W., Chomeau, V., & Whitman, R. (2009). Geographic relatedness and predictability of *Escherichia* *coli* along a peninsular beach complex of Lake Michigan. *Journal of Environmental Quality, 38*, 2357–2364.

Olyphant, G. A. (2005). Statistical basis for predicting the need for bacterially induced beach closures: emergence of a paradigm? *Water Research, 39*, 4953–4960.

Olyphant, G. A., & Whitman, R. (2004). Elements of a predictive model for determining beach closures in a real-time basis: the case of 63rd Street Beach Chicago. *Environmental Monitoring and Assessment, 98*, 175–190.

Olyphant, G. A., Thomas, J., Whitman, R. L., & Harper, D. (2003). Characterization and statistical modeling of bacterial (*Escherichia coli*) outflows from watersheds that discharge into southern Lake Michigan. *Environmental Monitoring and Assessment, 81*, 289–300.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Scopel, C. O., Harris, J., & McLellan, S. L. (2006). Influence of nearshore water dynamics and pollution sources on beach monitoring outcomes at two adjacent Lake Michigan beaches. *Journal of Great Lakes Research, 32*(3), 543–552.

Telech, J. W., Brenner, K. P., Haugland, R. A., Sams, E. A., Dufour, A. P., Wymer, L. J., & Wade, T. J. (2009). Modeling *Enterococcus* densities measured by quantitative polymerase chain reaction and membrane filtration using environmental conditions at four Great Lakes beaches. *Water Research, 43*, 4947–4955.

USEPA (Unites States Environmental Protection Agency). (2006). *Method 1600: Enterococci in water by membrane filtration using membrane-enterococcus indoxyl-β-D-glucoside agar (mEI). EPA-821-R-06-009*. Washington: USEPA.

Whitman, R., & Nevers, M. B. (2005). *Regional and local factors affecting patterns of E. coli distribution in southern Lake Michigan*. Porter: U.S. Geological Survey.

Yan, H. (1993). Skew correction of document images using interline cross-correlation. *Computer Vision, Graphics, and Image Processing, 55*(6), 538–543.