

Observations on Surgical Demand Time Series

Detection and Resolution of Holiday Variance

Ian C. Moore, M.A.Sc.,* David P. Strum, M.D.,† Luis G. Vargas, Ph.D.,‡ David J. Thomson, Ph.D., P.Stat., C.Stat., P.Eng.§

Background: Surgical scheduling is complicated by both naturally occurring and human-induced variability in the demand for surgical services. Surgical demand time series are decomposed into periodic, lagged, and linear trends with frequent occurrences of nonconstant variations in mean and variance. The authors used time series methods to model surgical demand time series in order to improve the scheduling of scarce surgical resources.

Methods: With institutional approval, the authors studied 47,752 surgeries undertaken at a large academic medical center. They initially extracted periodic information from the time series using two frequency domain techniques: the harmonic F test and the multitaper test. They subsequently extracted lagged (correlated) behavior using a seasonal autoregressive integrated moving average model. Finally, they used moving variance filters on the residuals to identify variance in the time series that coincided with major US holidays.

Results: Linear terms such as periodic cycles, trends, and daily and weekly lags explained 80% of the variance in the raw time series. In the residuals, the authors used moving variance filters to detect nonlinear variance artifacts that correlated with surgical activities on specific US holidays.

Conclusions: After extracting linear terms, the remaining variance was attributable to a combination of nonlinear and unexplained random events. The authors used the term *holiday variance* to describe a specific nonlinear disturbance in surgical demand attributable to statutory US holidays. Resolving these holiday variances may assist in management and scheduling of scarce surgical personnel and resources.

TIME series problems are ubiquitous, with growing importance in recent years. Although time series analyses are in the domain of statisticians, these analyses are most

commonly practiced by nonstatisticians and managers. The development of times series analyses has been facilitated by evolution of reliable easy-to-use software packages and consequent on huge recent gains in computing power. Large time series data sets are continuously emerging not only in econometrics, but also in the biomedical sciences and hospital management.¹

Healthcare institutions operate in austere fiscal environments. They are stressed by overwhelming demands for surgical services simultaneously with the reality of insufficient resources caused by a combination of globalization and demographics of an aging population. Surgical scheduling is often complicated by variability inherent in the stochastic demand for surgical services. Surgical services might be better managed with improved methods to forecast and quantify alterations in surgical demand and thus allocate capacity more appropriately.

Time series may be analyzed in both the time and the frequency domain. The concept of data in the time domain is trivial to understand, whereas many people have difficulty visualizing the formal mapping between time and frequency domains. This mapping is termed the *Fourier transform*, where a mathematical transformation is made on data to make observations in the frequency domain. Analyses in the frequency domain usually provide new insights about the periodic behavior of the data under investigation. As an example, weekly and abnormal dependencies are well known, but detection of subtle periodicities requires frequency domain analysis. In this article, we used a combination of time and frequency domain methods.

Quantification and prediction of trends in surgical demand are important and technically challenging. Advanced time series methods^{2–8} are complicated combinations of modeling and filtering using a mixture of linear and nonlinear models that may be applied in both the frequency and time domains. Surgical demand time series are comprised of linear, autocorrelated (lagged), and periodic terms with frequent nonconstant variations in mean and variance.

Our broader goal was to identify, describe, and understand sources of variability inherent in time series originating from production and delivery of surgical services at a large academic medical center. Identifying these variances may assist in scheduling of important personnel and scarce surgical resources. Our specific objective was to identify and quantify artifacts in our time series attributable to holiday behavior.

* Ph.D. Candidate in Mathematics and Statistics, § Professor of Mathematics and Statistics and Canada Research Chair in Statistics and Signal Processing, Department of Mathematics and Statistics, † Associate Professor of Anesthesiology and Senior Fellow, Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania. ‡ Professor of Operations, Decision Sciences, and Artificial Intelligence, The Joseph M. Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania.

Received from the Departments of Anesthesiology and of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada. Submitted for publication January 24, 2008. Accepted for publication April 24, 2008. Assistance for this research was provided by Operating Grant No. 04-27 from Physician's Services Incorporated Foundation, Toronto, Ontario, Canada. Financial assistance for Dr. Strum was provided by the physicians of Ontario through the Arthur Bond Scholarship granted by Physician's Services Incorporated Foundation. Financial assistance for Mr. Moore was provided by grant No. RGPIN 250268-03 from the Natural Sciences and Engineering Research Council of Canada, Ottawa, Ontario, Canada. Presented in part at the Canadian Operations Research Society Meeting, Montreal, Canada, May 10, 2006; the Institute for Operations Research and Management Science Meeting, Pittsburgh, Pennsylvania, November 6, 2006; the Canadian Operations Research Society Meeting, London, Ontario, Canada, May 15, 2007; and the Annual Meeting of the American Society of Anesthesiologists, San Francisco, California, October 17, 2007.

Address correspondence to Dr. Strum: Department Anesthesiology and Critical Care, 680 Dulles, 3400 Spruce Street, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania 19104-4823. strumd@uphs.upenn.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

Materials and Methods

To study surgical demand time series, we analyzed all surgeries performed at a large academic health sciences center over a 6-yr period from 1989 to 1995.⁹⁻¹² Use of anonymous patient records was approved by the human subjects review committee of the institution that collected the data.

Variables in our data set included total time (defined as the time from entry into the operating suite until arrival in the recovery room), surgical time (defined as the time from incision to closure of the surgical wound), age, American Society of Anesthesiologists physical status, type of anesthesia, Current Procedural Terminology codes, and emergency status. Our data consisted of 47,752 individual surgeries collected over 2,191 days with an average of 21.8 surgeries per day, with surgical procedure times recorded in universal time with an accuracy of ± 5 min. Surgical schedules were parameterized using Allegro Common LISP (Franz Inc., Oakland, CA), and total times and surgical times were aggregated daily. We estimated apparent demand for surgeries as total times aggregated daily to produce a surgical demand time series.

To initially better understand our time series, we examined graphical dependent variable and box plots. Assuming that the recording errors were independent, the mean of the raw aggregate time series was 58.8 h, with an SD of 27.0 h.

Surgical demand time series are aggregates of linear (*i.e.*, periodic, trend, lagged) and nonlinear (*i.e.*, nonconstant variance) terms. Beginning with the linear elements, we subtracted obvious periodic and lagged (correlated) elements sequentially leaving a residual. We studied these nonlinear effects of holiday variance and detail these analyses below in the order that they occurred.

The appendix describes terminology, statistics, and processes used in our time series analyses. We abstracted these descriptions from the body of the article to increase readability for clinicians while maintaining a level of detail more suitable for a technical audience.

Statistics

We used the following statistical tools in analyzing the surgical demand time series: graphical dependent variable plot, histogram, box plots, multitaper spectral estimates, harmonic F test, seasonal autoregressive integrated moving average (ARIMA) model, autocorrelation function, partial autocorrelation function, cumulative residual spectrum, quantile-quantile plots, and a moving sample variance filter. In the next section, we illustrate our results using these tools.

Linear Effects: Periodic Components

To model and detect periodic artifacts in the data, we computed a multitaper spectral estimate (appendix) of

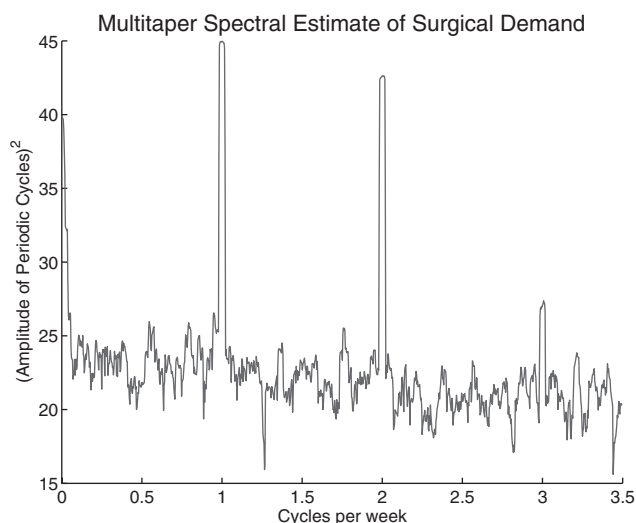


Fig. 1. Multitaper spectral estimate (bandwidth 6.5, 11 discrete prolate spheroidal sequence tapers) of the daily cumulative surgical demand over 6 yr starting July 1, 1989. The rectangular “bumps” and the typical responses to periodic effects are 1.0, 2.0, and 3.0 cycles/week and are explained by the periodicity from the week. However, the period near 1.4 cycles/week does not have such a simple explanation.

the data (fig. 1). For example, if we see a distinct spike at 0.5 cycles a week in a spectral estimate, we know that there is a 14-day cycle. With the aid of a harmonic F test (appendix), we determined the number of statistically significant period detections in the time series using three confidence level thresholds: 98, 99, and 99.5% (table 1). Of the periods detected, the three strongest, as seen in the multitaper spectrum, were derived from the weekly 7-day cycle and its first two harmonics.

These periods were removed from the data using the following procedure:

- The harmonic F test was computed from the data.
- All Fourier transform coefficients of the data set corresponding to those F statistics that exceeded the null hypothesis 99% confidence were retained.
- Periodic data elements from the coefficients were reconstructed using a nonperiodic Fourier series estimate of the time-varying mean:

$$\hat{u}_t = \sum_{j=0}^J \hat{a}_j \cos(2\pi f_j t) - \hat{b}_j \sin(2\pi f_j t)$$

Table 1. Sensitivity to Choice of Harmonic F Statistic Threshold ($NW = 6.5$, $K = 11$)

Nominal significance level, %	98	99	99.5
F statistic threshold	4.79	5.85	6.99
Number of detections	83	54	30
Expected	44	22	11

K = number of discrete prolate spheroidal sequences; NW = bandwidth of the discrete prolate spheroidal sequence.

as found in Chatfield,³ where \hat{u}_t represents the periodic reconstruction, J represents the number of significant terms identified by the F test, and f_j represents the corresponding frequencies. The \hat{a}_j , \hat{b}_j coefficients represent amplitude estimates of the cosine and sine terms, respectively; these were derived by the expected value calculations that aid in the calculation of the harmonic F test as found in Thomson.⁸

- \hat{u}_t was subtracted from the data used in the first step and repeated all the above until all F statistic readings were below the 99% significance threshold. The effect of choosing a low significance threshold (*i.e.*, 90%) and subtracting false detects creates large divots in the residual spectrum. On the other side, choosing a significance threshold that is too high (*i.e.*, 99.999%) will run the risk of failing to detect many important cycles in the data. Therefore, our 99% significance threshold was chosen based on the relative flatness of the residual spectrum.

Before removing these periodic effects, a histogram showed that the distribution of our raw data was bimodal, with the lower mode attributed to weekends and the upper mode attributed to weekdays. Interestingly, this bimodal effect was reduced to a single mode after the periodic background was subtracted. These periodic terms are important because, cumulatively, they account for approximately 80% of the total variance. Moreover, unlike the short-term autocorrelation effects to be described next, these periodic effects can be used to predict the series several months in advance. These longer-term predictions will be the subject of future research.

Linear Effects: Lagged (Correlated) Components

We say a time series is lagged or autocorrelated when demand on a given day is correlated with the demand from any previous day. We detected autocorrelation using lagged variables and a regression on the data from a given day on data from previous days. To detect lags, we used a partial autocorrelation function, autocorrelation function, cumulative residual spectrum, and quantile-quantile plot to identify the order of moving average, autoregressive and differences to construct an ARIMA(1,1,1) \times (1,0,1)₇ model.

To study lagged behaviors in the surgical demand, we filtered the data to remove serially correlated artifacts by using a seasonal ARIMA(p,d,q) \times (P,D,Q)_s model (appendix) on the nonperiodic residuals; we determined to use $p = 1$, $d = 1$, and $q = 1$ parameters, seasonal parameters $P = 1$, $D = 0$, and $Q = 1$ with a seasonal differencing of $s = 7$. We used the seasonal ARIMA model to identify and remove sources of variation that we believed were due to lagging of surgical demand caused by day-to-day deferral of nonurgent surgeries.

Nonlinear Effects: Holiday Variance

After the extraction of linear artifacts, we used variance filters to identify and analyze nonlinear variance in the residuals attributable to statutory holidays. We used a nonweighted moving variance filter initially because it is simple. We compared results from the nonweighted filter with results using a discrete prolate spheroidal sequence (DPSS) weighted moving variance filter we believed would exhibit less bias in statistical estimates. Averages of these filtered results were taken over each of the 6 yr. We used this approach to identify variance attributable to specific statutory US holidays.

Results

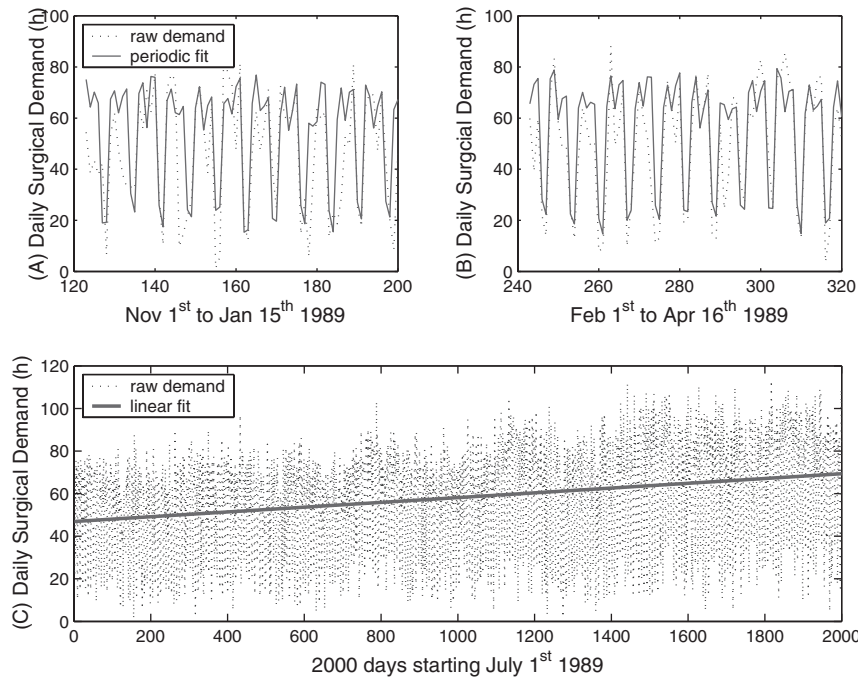
Examination of a dependent variable plot (figs. 2A-C) illustrates the nature of the time series in which the human eye has difficulties identifying and resolving trends in the raw data. We illustrated the raw data superimposed with the reconstructed periods over a 2-month stretch for both the least (fig. 2A) and most (fig. 2B) volatile extremes of surgical demand during the first year. Histograms of the weekdays and weekends revealed two almost nonoverlapping modes, the first attributed to weekends and the second attributed to weekdays.

Figure 3 illustrates heterogeneous means and variances of the surgical demand time series summarized by day of the week. The mean and SD of the work weekday time series were 72.7 and 17.2 h, respectively; the mean and SD of the weekend time series were 29.9 and 10.1 h, respectively. The outliers (low) on Mondays and Fridays exceeded similar outliers in the mid week and represent holiday behaviors associated with "long weekends."

Linear Effects: Periodic Components

With the aid of the harmonic F test, we determined there were 83 periods in the time series with a confidence level of 98%, 54 periods with a confidence level of 99%, and 30 periods with a confidence level of 99.5% (table 1). Of these detected periods, the three strongest, as clearly seen in the multitaper spectrum at 7, 3.5, and 2.3 days, respectively, are derived from the weekly 7-day cycle and its first two harmonics (fig. 1). Seventy-six percent of the variance in the raw data was attributed to the 54 periodic elements that were above the 99% threshold; for the most significant periodic elements, see table 2. Table 1 highlights that the expected number of detections reflects the fact that there are the same number of degrees of freedom in the time and frequency domains. Therefore, a test is made at each frequency, and for the 99% significance level we expected 22 detections and observed more than twice that number. Examination of the box plot in figure 3 illustrated heterogeneous variations in mean and variance in the raw

Fig. 2. Fitted estimates of surgical demand. (A) Periodic fit during the most volatile period in the initial year, November 1 to January 1, 1989. (B) Periodic fit during the least volatile period in 1990, February 1 to April 16. (C) Linear fit during the first 2,000 days starting July 1, 1989. Trend and periodic trends, as shown here, are important to modeling and predicting surgical demand.



data. In contrast, box plots for the raw data after extraction of the periodic and lagged linear artifacts (fig. 4) indicated a nearly stationary time series with relatively homogeneous mean and variance. Statistical analyses of the residuals after extraction of the linear artifacts (fig. 5) indicated the success with which these artifacts were removed.

Because much of the variance in our time series was explained by periodic components, the workweek and annual and semiannual cycles were expected. Excluding these, there remain many statistically highly significant components that do not have trivial explanations (table 2).

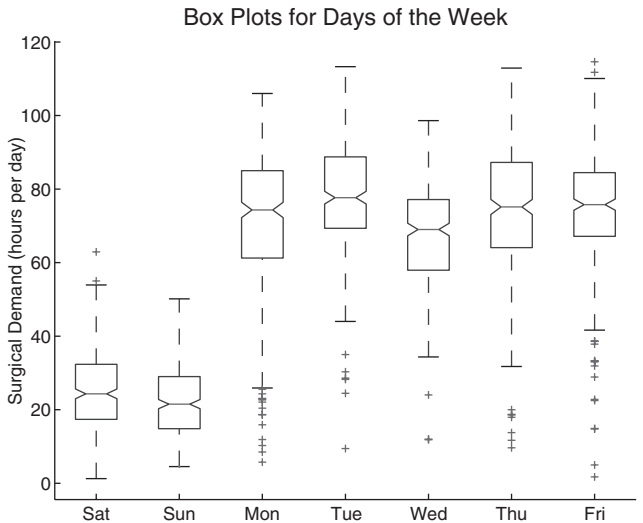


Fig. 3. Box plots for each day of the week of raw surgical demand (*waist* = median, *box* = 2 SD, *crosshatch* = 3 SD, outlier above 3 SD). Box plots indicate a nonstationary time series heterogeneous in both mean and average variance.

Linear Effects: Lagged (Autocorrelated) Components

We used a seasonal ARIMA model to remove lags and linear trend behavior from the post-periodic extracted residual. Nearly 4% percent of the variance in the raw data were attributable to linear trend and lagged elements. After seasonal ARIMA extraction, the statistics of the residuals fall within the confines of the SE thresholds for the autocorrelation function, partial autocorrelation function, and cumulative residual spectrum (fig. 5). A quantile-quantile plot indicated the post-seasonal ARIMA extracted residuals were normally distributed except for a heavy left tail.

Table 2. Highly Significant nontrivial Frequencies Derived from the Harmonic F Tests

	Frequency, cycles/wk	Period, days	Label	Significance Level	One-sided Amplitude, h
1	2.0000	3.50	W	1.0000000	9.48
2	1.0000	7.00	W	1.0000000	12.43
3	3.0000	2.33	W	0.9999982	1.38
4	0.0192	365.00	A	0.9999978	2.47
5	3.4201	2.05		0.9999115	0.70
6	0.8435	8.29		0.9998000	0.74
7	0.5351	13.08		0.9998000	0.82
8	1.3642	5.13		0.9997000	0.88
9	1.5799	4.43		0.9996000	0.78
10	1.3259	5.27		0.9995000	0.58
11	0.3451	20.29		0.9994000	0.69
12	0.8658	8.08		0.9993000	0.78
13	0.7268	9.63		0.9992000	0.70
14	1.0383	6.74		0.9991000	0.83
23	0.2492	28.09	L	0.9970000	0.68

A = annual periodic event; L = periodic event possibly due to lunar effects; W = weekly periodic event.

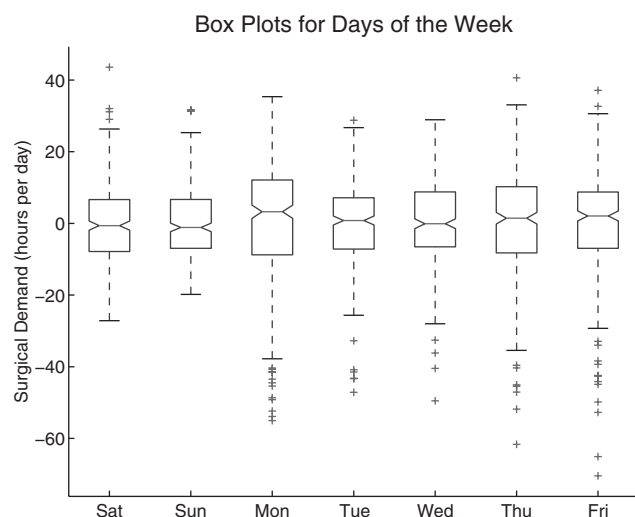


Fig. 4. Box plots for each day of the week after extraction of seasonal autoregressive integrated moving average $ARIMA(1,1,1) \times (1,0,1)$, fit, mean and periodic cycles. Box plots indicate a nearly stationary time series relatively homogeneous in mean and average variance.

Nonlinear Effects: Holiday Variance

To study the behavior of nonlinear variance clusters, a weighted moving variance filter was used on the data after extraction of the linear terms. The horizontal lines in figure 6 indicate the lower 5% and upper 95% thresholds of a gaussian random sample for data yielding the same SD as our post-seasonal ARIMA extracted residual. These results indicated volatile surges in surgical demand directly correlated with every major US holiday. We coined the term *holiday variance* to describe these repeated discrete disturbances in variance that coincided with statutory holidays. Using a nonweighted variance filter, we were able to resolve these disturbances in

only modest detail. In contrast, when using DPSSs¹³ as weighting coefficients, we observed increased resolution of holiday activity. This is well illustrated during the Christmas holiday period, as the results in figure 7 resolve holiday variance into Christmas and New Year's. Contrast this with the relatively poor statistical properties of the unweighted estimator used in figure 6 that biased holiday activity stretched across the entire Christmas week.

Discussion

Our analyses indicated significant variance in the time series attributable to periodic, linear trend, lagged, and holiday disturbances in surgical demand. Initial spectral estimates indicated that surgical demand was strongly periodic. The majority of these periodic behaviors were attributed to midweek highs and weekend lows. However, reality seems to be more complex than this limited view would indicate. We speculate two major separate processes inherent in our time series. The first is a stochastic process running throughout the entire 7-day week, consisting of random demand for emergencies that is plainly visible on the weekends. Superimposed on the underlying stochastic processes is a highly periodic elective weekday surgical demand, driven in large part by the accumulation of surgical specialty demand performed in reserved specialty block time. Within the detected periodic activity may be additional unexplained natural and socioeconomic phenomenon. Examples may include lunar cycles near 28 days or biweekly payroll disturbances in addition to other subtle periodic structures within the healthcare system that we have yet to identify. The heterogeneous means and variances in the

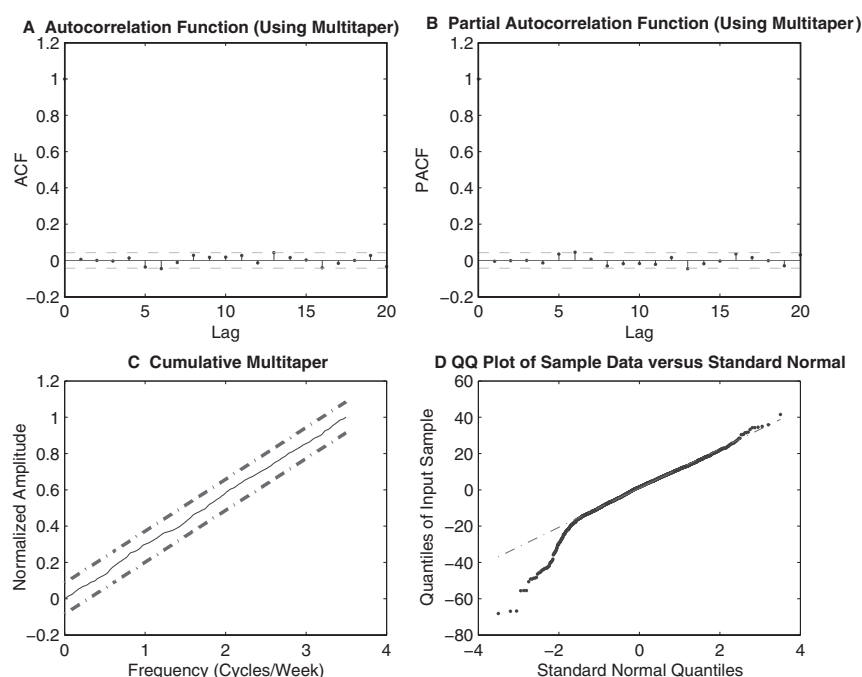


Fig. 5. Analysis of residuals after extraction of seasonal autoregressive integrated moving average $ARIMA(1,1,1) \times (1,0,1)$, fit, mean and periodic artifacts. (A) Autocorrelation function (ACF). (B) Partial autocorrelation function (PACF). (C) Cumulative residual spectrum. (D) Normal quantile-quantile (QQ) plot. Note that the nonlinear behavior of the lower left tail (D) is almost entirely due to holiday variance.

Non-weighted Variance Filter Estimates

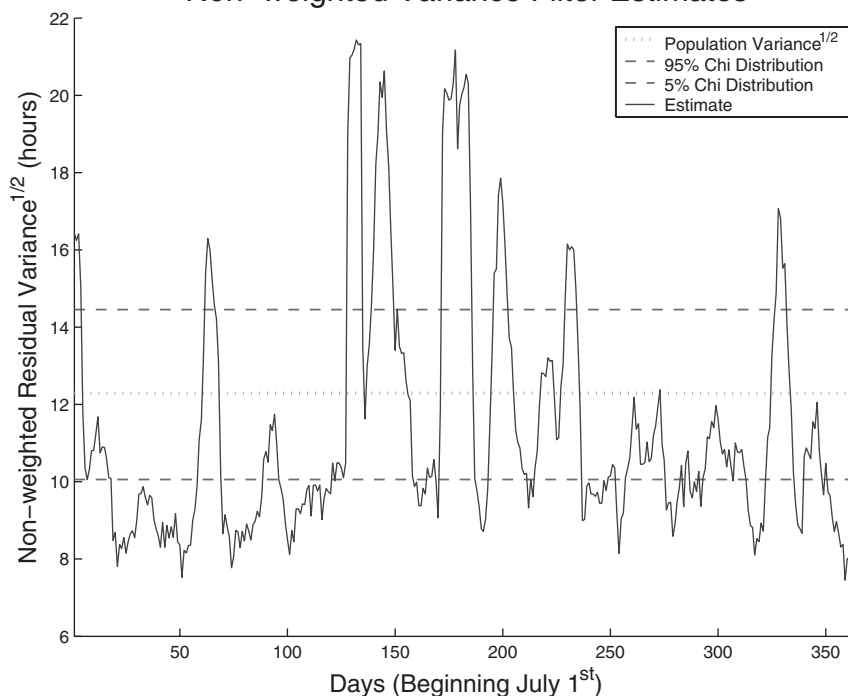


Fig. 6. Mean of 7-day nonweighted moving variance filter estimates of surgical demand residuals over a 1-yr cycle. The horizontal dashed lines are set at the 5% (lower) and 95% (upper) thresholds of the chi-square distribution with 42 degrees of freedom. Note that the overall average variance is a poor fit almost everywhere.

time series were highly nonlinear and therefore difficult to analyze using more traditional statistical methods without encountering potentially serious statistical bias.

The seasonal ARIMA model confirmed a gradual linear upward trend indicating a slow overall growth in surgeries throughout the studied period. In addition, significant lag components were identified at 1 and 7 days. A significant lag 7 indicates that surgical demand on any given day is strongly influenced by activity that occurred on the same weekday of the preceding week. An impor-

tant event such as a traffic accident might be an example of behaviors leading to a lag 1. The unexpected casualties from the accident could overwhelm the emergency capacities of local hospitals, and surgeries would have to be reordered according to urgency. This reordering of surgeries would delay less urgent surgeries to the following day. Our analyses indicate that lagging of surgical cases happens daily and weekly; whatever is not done on time is pushed to the following day or, alternately, the following week. Weekly lagging may be attributable to

DPSS Weighted Variance Filter Estimates (NW = 2)

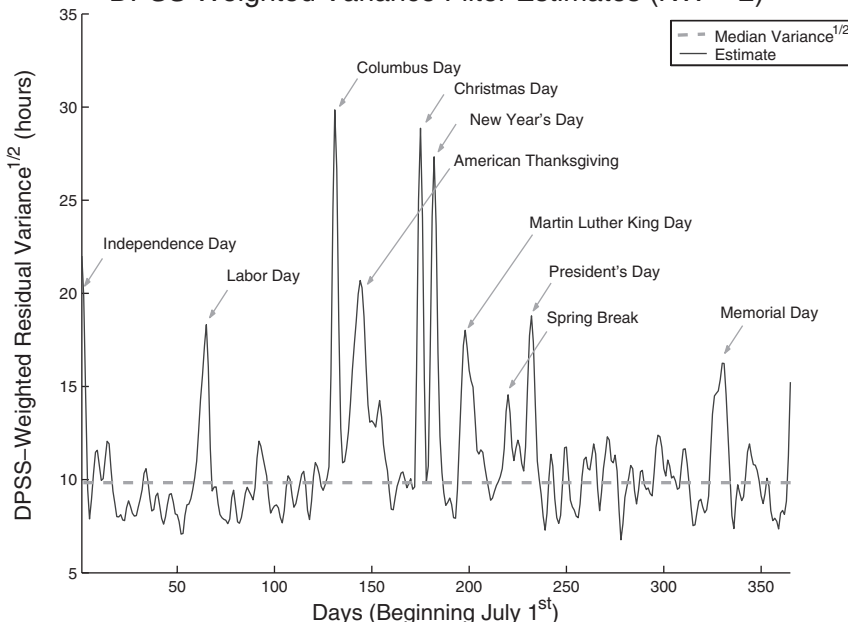


Fig. 7. Mean of 7-day discrete prolate spheroidal sequence (DPSS) weighted variance filter estimates of surgical demand residuals over a 1-yr cycle. Note that the background variance excluding holidays is approximately 9.8, which is smaller than the nonweighted variance filter estimates.

delaying surgeries from repeating weekly blocks of time reserved for surgeons to the next available surgeon-specific opportunity. Major events capture attention; however, our results illustrate that mundane but significant lagging of surgeries occurs on a routine daily and weekly basis.

The DPSS weighted moving variance estimates (fig. 7) indicated that every peak, in the filtered squared residual, represents a specific US holiday; we termed this holiday variance. We can observe that these variance estimates represent a substantial improvement in resolution relative to using a nonweighted moving sample variance filter estimate, illustrated in figure 6.

Previous studies of similar surgical demand time series used different analyses methods. Williams *et al.*¹⁴ analyzed times for a single surgical procedure (knee arthroscopy) using interrupted time series methods (one-way analysis of variance or linear regression) and showed shorter surgical times with regional anesthesia. However, these authors did not analyze continuous diverse surgical demand, did not address issues of statistical bias associated with complex time series and linear methods, and explained less than 30% of variance in the dependent variables, all factors that might affect the potential of these methods to predict the future.

Dexter *et al.*¹⁵⁻¹⁷ analyzed surgical demand time series similar to our own to predict staffing requirements. These authors used boxcar averaging methods and predicted staffing requirements for weekdays, weekends, and holidays.¹⁵⁻¹⁷ The boxcar averaging periods ranged from 4 weeks to as short as 12 or 24 h. These investigators assigned confidence intervals for their estimates and related the number of surgical cases started to the hours of surgical service provided using discontinuous time series. However, these authors did not report estimates of variance explained by the dependent variables and did not study the periodic elements that, in our studies, explained much of the variance.

In contrast, we studied a continuous surgical demand time series heterogeneous in mean and variance using methods that identified linear trends, lags, holiday variance, and periodic elements. We identified periodic elements not previously recognized in surgical demand time series, including short- and long-term periodic elements. Using these methods, we detected 54 periodic terms at the 99% detection threshold that explained a large portion of variance (76%) in the dependent variable. Most of the first dozen of these periodic terms have important amplitudes greater than 0.7 h. Boxcar averaging is undersampled in "4-week" blocks and would hide many of the periodic elements we detected with periods less than 1 month and might alias longer periods. We described a method of time series analyses capable of fine-grained studies and improved fidelity relative to previous studies using similar surgical time series. Further development of these techniques should improve predictions of surgical demand with a potential for improved capacity planning and staffing models.

Limitations

To remove the linear trend in our model, we used first differencing within an ARIMA model ($d = 1$). Had we instead used a least squares regression model to remove the linear trend, we might have improved even further the accuracy of estimate of the trend. We tried dividing the raw data by trend, but this resulted in a decreasing variance.

We made empirical comparisons only between non-weighted and DPSS weighted moving variance filter estimates. Detailed analytical comparisons were beyond the scope of current article and will be the subject of future research.

It will be apparent to many readers that apparent demand is not the actual daily demand for surgical services; rather, it represents true demand censored by scheduling policies in effect at the institution that produced the services. Predictions based on true surgical demand would likely improve prediction accuracy; nonetheless, apparent demand is the only demand metric most institutions record, and we assume that over time, apparent demand represents a lagged but close approximation of true demand. We believe, however, that our methods would also be applicable to data series representing uncensored true demand. In this article, whenever we refer to "demand," it is apparent demand we actually refer to.

Although we have removed the periodic and ARIMA artifacts from the data, we propose in future research to also remove the holiday variance in a statistically justified manner. Pending further research, we propose to incorporate these artifacts into a prediction model to better predict surgical demand in the operating rooms.

Conclusions

Our results suggest that demand for surgical services is highly periodic. The main cycles are the obvious one, two, and three cycles per week; however, there are several others in addition. The data had strong daily and weekly lags, suggesting that surgeries are routinely deferred to the next day or the next surgeon-specific block time, which is often in the same weekday of the subsequent week. There was also a gradual upward linear trend that indicated that demand for surgical services at the hospital increased steadily over the study period.

Linear effects, including periodic cycles, trends, and daily and weekly correlations, explained 80% of the variance in the raw data. The remaining variance was attributable to a combination of nonlinear and unexplained random events. We were able to identify surges of volatile activity correlated to statutory US holidays. We coined the term holiday variance to describe these disturbances in surgical demand. Resolving these variances may ultimately improve scheduling of personnel

and scarce surgical resources. Poor scheduling or capacity allocation can lead to suboptimal utilization and staffing of costly operating rooms.

The authors thank Gerard Bashein, M.D., Ph.D. (Professor of Anesthesiology, University of Washington, Seattle, Washington), for his assistance with the manuscript.

Appendix: Technical Details

This appendix describes terminology, statistics, and processes used in our time series analysis. We abstracted these descriptions from the text of the article to increase readability for clinicians while maintaining a level of detail more suitable for a technical audience.

Chi-square Distribution

The chi-square distribution,^{18,19} denoted by χ^2 , is one of the most widely used distributions in statistical significance tests and plays an important role in making statistical inferences; an example of its use is in figure 6. If x_i are n independent, gaussian-distributed random variables of mean μ and variance σ^2 , then

$$z = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2$$

is chi-square distributed, and we write $z \sim \chi_n^2$, where n is the degrees of freedom. We used this distribution to construct the thresholds in the harmonic F test. We used it to test whether the residuals were normally distributed, after removal of time series artifacts (fig. 6). The square of a normally distributed data set is, by definition, a chi-square distribution.

Discrete Prolate Spheroidal Sequences

Discrete prolate spheroidal sequences¹³ (DPSSs) are a family of sequences discovered by David Slepian, Ph.D. (1923–2007), in 1978 at Bell Laboratories (Murray Hill, NJ) which, using his notation, are the eigenvectors $v_n^{(k)}$ of order k , dimension N , and bandwidth W of the following algebraic eigenvalue equation:

$$\sum_{m=0}^{N-1} \frac{\sin[2\pi W(n-m)]}{n-m} v_m^{(k)}(N, K) = \lambda_k(N, W) v_m^{(k)}(N, K).$$

The importance of DPSSs lies in the fact that they have maximum energy concentration in frequency given any sequence defined on a finite time interval; i.e., they define the famous Heisenberg uncertainty principle for finite amounts of sampled data. DPSSs play an important role in the statistical robustness of multitaper spectral estimates (fig. 1) and as weighting coefficients in our variance filter estimates (fig. 7).

Multitaper Spectral Analysis

Although there is admittedly a conceptual threshold to understanding the frequency domain quantities, many problems are significantly simpler in the frequency domain than they are in the time domain. Spectral analysis allows us to study the distribution of power (including the amplitude of periodic components) of a data set in the frequency domain. A band-limited function can not be time-limited, as Claude Shannon indicated in a famous landmark article²⁰; therefore, we are left with the task of estimating a band-limited function in the frequency domain given a finite time series sample set. The Fourier transform⁶ is an integral part of estimating the spectrum of a data set that is divided into magnitude and phase responses. Spectrum estimation relates to the magnitude response of our time series in the frequency domain.

When selecting a spectral estimator for a given data set, we factor in considerations of variance, resolution, spectral leakage, and bias. The simplest of all spectral estimators is the periodogram^{3,6,21}

$$S_p(w) = \sum_{\ell=-(N-1)}^{N-1} r_{(p)}(\ell) e^{-i w \ell},$$

where $r_{(p)}(l)$ is an autocorrelation sequence conventionally estimated as

$$r_{(p)}(\ell) = \frac{1}{N} \sum_{n=1}^{N-|\ell|} x_n x_{n+|\ell|}.$$

Unfortunately, periodogram estimates are one of the poorest of spectral estimators due to both statistical bias and variance^{8,21}; the corresponding autocovariance estimate inherits these problems. To minimize this, a multitaper spectral estimate is an excellent choice because of its statistical robustness. Technical explanations of the multitaper technique are beyond the scope of this article and are detailed elsewhere by Thomson,⁸ section V-6 and figure 6 of Thomson *et al.*²²

Harmonic F Test

We used a harmonic F test to aid in removal of the periodic components from our data set. The harmonic F test⁸ is the ratio of variance explained by the assumed periodic component at frequency f to that variance remaining in the residuals. We applied the harmonic F test to our time series while using the multitaper simultaneously to confirm periodic trends. We used the F test to identify periods, connected these to a nonperiodic Fourier series (done with a fast Fourier transform⁶) to form a periodic reconstruction, and subtracted this from the data. Strong periodic components tend to mask weaker ones; therefore, this process was repeated recursively until all F statistics tested below the 99% significance level.

Seasonal Arima Model

We used a seasonal ARIMA model to study stationary behaviors in residuals after extraction of periodic artifacts. Seasonal ARIMA models are covered in numerous texts,^{2-5,7} so we give only a summary here. We assume z_t is a purely random process sampled from a gaussian distribution with mean zero and variance $\sigma_{(z)}^2$. Under this assumption, an ARMA model can be summarized as a moving average (MA) process x_t of order q , MA(q) of the form,

$$x_t = \beta_0 z_t + \beta_1 z_{t-1} + \dots + \beta_q z_{t-q}$$

where β_i are constants and the Z s are scaled so that $\beta_0 = 1$. Using the backshift operator $Bz_t = z_{t-1}$, we can rewrite the above expression as

$$x_t = (\beta_0 + \beta_1 B + \dots + \beta_q B^q) z_t = \theta(B) z_t.$$

Then process x_t is said to be an autoregressive (AR) process of order p if

$$z_t = (1 - \alpha_1 B - \dots - \alpha_p B^p) x_t = \phi(B) x_t$$

We can express an AR process of finite order as an MA process of infinite order. We can combine the above MA and AR processes into a mixed model of the following form:

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + z_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q},$$

where the above can be rewritten as;

$$\phi(B) x_t = \theta(B) z_t$$

thus, the above is known as the ARMA model. Because most time series are nonstationary, the ARMA model described above can be extended to include simple nonstationary artifacts such as trends and

periods. This is achieved through the process of differencing the series and when applied to the above equation is called an ARIMA model. This is done by the introduction of a differencing operator, $\nabla^d x_t$, by applying the following:

$$w_t = \nabla^d x_t = (1 - B)^d x_t.$$

The equation above can be extended to the following:

$$\phi(B)w_t = \theta(B)z_t.$$

The above equation is said to be an ARIMA(p, d, q) model.

We have shown how to account for periodic components in a nonstationary time series; however, in practice it is necessary to account for seasonally periodic components which repeat every s observations. We can further generalize the ARIMA(p, d, q) model to what is popularly known as the seasonal ARIMA model, denoted using the notation ARIMA(p, d, q) \times (P, D, Q) $_s$, where the seasonally periodic components are repeated every s observations and P, D, Q are the seasonal versions of p, d , and q , respectively.

We have made one significant change to the usual ARIMA methodology, namely that the usual estimates of the autocorrelation have been replaced by the Fourier transform of the multitaper spectrum estimate. Similarly, we replaced the cumulative periodogram with a cumulative multitaper spectrum.

Moving Variance Filter

We used weighted and nonweighted moving variance filters to study changes in variance of the residual after extraction of the periodic and lagged components. For simplicity's sake in the analysis to be presented in future work, we make the assumption that our process x_t is zero mean. To reduce estimator variance, we took the mean over six samples each corresponding to the same day of the week and month of each year (excluding the leap year), and the sample variance filter estimate $\hat{\zeta}_{(nw),t}$ of any given 7 days is:

$$\hat{\zeta}_{(nw),t} = \frac{1}{YP} \sum_{k=0}^{Y-1} \sum_{n=0}^{P-1} (x_{t-n+365k})^2,$$

where P corresponds to the 7 days of the week and Y corresponds to the 6 yr of the study period. The variance estimate $\hat{\zeta}_{(nw),t}$ is that of a chi-square distribution¹⁹ with $42 (P * Y)$ degrees of freedom where we determined the 5% and 95% thresholds.

To reduce estimator bias, we applied DPSS filter coefficients $v_n^{(k)}$ of bandwidth $NW = 2$. Hence, we denote our DPSS weighted variance filter estimate $\hat{\zeta}_{(w),t}$ of any given 7 days to be

$$\hat{\zeta}_{(w),t} = \frac{\sum_{k=0}^{Y-1} \sum_{n=0}^{P-1} (v_n x_{t-n+365k})^2}{Y \sum_{n=0}^{P-1} v_n^2}$$

One notable property of DPSSs is that they are orthonormal; hence,

$$1 = \sum_{n=0}^{P-1} v_n^2$$

This reduces the denominator of $\hat{\zeta}_{(w),t}$ to Y , which we have left in the notation in the event that v_n is arbitrary.

References

1. Solo V: The end of time series. *J Am Stat Assoc* 2000; 95:1346-9
2. Box GEP, Jenkins GM: Time Series Analysis: Forecasting and Control. London, Holden-Day, 1970, pp 1-125, 173-459
3. Chatfield C: The Analysis of Time Series: An Introduction, 6th edition. New York, Chapman & Hall/CRC, 2004, pp 1-276
4. Harvey A: Forecasting, Structural Time Series Models and the Kalman Filter. New York, Cambridge University Press, 1990, pp 1-99
5. Jenkins GM, Watts DG: Analysis and Its Applications. London, Holden-Day, 1968, pp 1-525
6. Oppenheim AV: Discrete-Time Signal Processing, 2nd edition. New Jersey, Prentice Hall, 1999, pp 8-93, 541-774
7. Priestly MB: Spectral Analysis and Time Series. New York, Academic Press, 1981
8. Thomson DJ: Spectrum estimation and harmonic analysis. *Proc IEEE* 1982; 70:1055-96
9. Bashein G, Barna C: A comprehensive computer system for anesthetic record retrieval. *Anesth Analg* 1985; 64:425-31
10. Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: Comparison of lognormal and normal models. *ANESTHESIOLOGY* 2000; 92:1160-7
11. Strum DP, Sampson AR, May JH, Vargas LG: Surgeon and type anesthesia predict variability in surgical procedure times. *ANESTHESIOLOGY* 2000; 92:1454-66
12. Strum DP, Vargas LG, May JH: Surgical subspecialty block utilization and capacity planning: A minimal cost analysis model. *ANESTHESIOLOGY* 1999; 90:1176-85
13. Slepian D: Prolate spheroidal wave functions, Fourier analysis, and uncertainty, V: The discrete case. *Bell Syst Tech J* 1978; 57:1371-430
14. Williams BA, Kentor ML, Williams JP, Figallo CM, Sigl JC, Anders JW, Bear TC, Tullock WC, Bennett CH, Harner CD, Fu FH: Process analysis in outpatient knee surgery: Effects of regional and general anesthesia on anesthesia-controlled time. *ANESTHESIOLOGY* 2000; 93:529-38
15. Dexter F, Macario A, Qian F, Traub RD: Forecasting surgical groups' total hours of elective cases for allocation of block time: Application of time series analysis to operating room management. *ANESTHESIOLOGY* 1999; 91:1501-8
16. Dexter F, Epstein RH: Holiday and weekend operating room on-call staffing requirements. *Anesth Analg* 2006; 103:1494-8
17. Dexter F, O'Neill L: Weekend operating room on call staffing requirements. *AORN J* 2001; 74:664-5, 668-71
18. Johnson NL, Kotz S, Balakrishnan N: Chi Square Distributions Including Chi and Rayleigh, Continuous Univariate Distributions. New York, John Wiley and Sons, 1994, pp 415-93
19. Rice JA: Distributions derived from the normal distribution, *Mathematical Statistics and Data Analysis*. Pacific Grove, California, Wadsworth and Brooks, 1988, pp 177-84
20. Shannon CE: A mathematical theory of communication. *Bell Syst Tech J* 1948; 27:379-423, 623-56
21. Percival D, Walden A: Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge, Cambridge University Press, 1993
22. Thomson DJ, Lanzerotti LJ, Vernon FL, Lessard MR, Smith LTP: Solar modal structure of the engineering environment. *Proc IEEE* 2007; 95:1085-132