Beyond Linear Methods of Data Analysis: Time Series Analysis and Its Applications in Renal Research

Ashwani K. Gupta¹ Andreea Udrea²

¹Department of Nephrology, University of Florida, Jacksonville, Fla., USA; ²Department of Automatic Control and Systems Engineering, Polytechnic University, Bucharest, Romania

Key Words
Analysis methods · Kidney disease · Time series models

Abstract
Analysis of temporal trends in medicine is needed to understand normal physiology and to study the evolution of disease processes. It is also useful for monitoring response to drugs and interventions, and for accountability and tracking of health care resources. In this review, we discuss what makes time series analysis unique for the purposes of renal research and its limitations. We also introduce nonlinear time series analysis methods and provide examples where these have advantages over linear methods. We review areas where these computational methods have found applications in nephrology ranging from basic physiology to health services research. Some examples include noninvasive assessment of autonomic function in patients with chronic kidney disease, dialysis-dependent renal failure and renal transplantation. Time series models and analysis methods have been utilized in the characterization of mechanisms of renal autoregulation and to identify the interaction between different rhythms of nephron pressure flow regulation. They have also been used in the study of trends in health care delivery. Time series are everywhere in nephrology and analyzing them can lead to valuable knowledge discovery. The study of time trends of vital signs, laboratory parameters and the health status of patients is inherent to our everyday clinical practice, yet formal models and methods for time series analysis are not fully utilized. With this review, we hope to familiarize the reader with these techniques in order to assist in their proper use where appropriate.

Introduction

Continuous developments of experimental techniques in cellular biology and physiology provide us with newer insights into biological processes. In order to fully understand these processes and their dynamic interplay, we need to develop tools that can help us identify, interpret and use this information. Time series models and analysis methods are techniques that can be useful in the characterization of simple and complex biological behaviors and derive predictions. Time series analysis can also be applied to the study of nonlinear systems, commonly encountered in biology and medicine [1]. This is an area of tremendous scope and applicability to renal research.

What Is a Time Series?

Experiments in a variety of experimental and physiological domains lead to time series of one (univariate time series) or more measurable parameters (multivariate time
series). Quite simply, a \((\text{discrete})\) time series \(Y\) is a set of observations arranged over time and spaced at often equal time intervals:

\[
Y = \{Y_i \mid i = 1, N\},
\]

where \(N\) is the length of the measurement.

Individual data points in a time series are not independent observations. Usually, biological plausibility can account for the present values as a function of the past values along with interventions performed. Several exploitable examples of time series can be found in our clinical experience. In a patient with chronic kidney disease, there can be time series of hemoglobin over time, iron stores over time, blood pressure (BP) over time and many others. Obtaining repeated measurements on the same subject over time is useful for the study of time-variant trends/patterns over cross-sectional measurements alone. Informal assessment of these time series (e.g. comparing BP measurement over the past several clinic visits to determine adequacy of control or comparing it to statistically derived population cutoffs for normal and disease states) is a part of our everyday clinical practice, yet formal analysis of time series provides several advantages.

Medical time series are often congested with noise. The noise may represent biological variability, effects of interventions performed or a variety of external factors that cannot be controlled for. Statistical methods can help us reliably interpret signals and highlight trends that may have causal implications. Nevertheless, while traditional statistics do not take into account the time-ordered nature of these observations, time series analysis provides us tools to do so. Quantitative analysis of time series data is often limited to calculating the mean and standard deviation. Although useful, these may not provide adequate characterization of periodicities, seasonal trends or correlations in the data, e.g. hemoglobin variability in dialysis patients treated with erythropoietin is a complex phenomenon (fig. 1). In a review, Arneson et al. [2] compared four different methodologies used to characterize hemoglobin variability. Using the mean, standard deviation or the absolute hemoglobin change [3] provides us with a summary measure that is easy to understand conceptually. It is, however, inadequate to distinguish between trends such as consistently increasing/decreasing, cyclical patterns or completely erratic patterns. A residual standard deviation methodology has been proposed previously [3]. It measures the deviation from a regression line and therefore controls for linear trends only in the data. Gilbertson et al. [4] have also used fluctuation across thresholds to characterize hemoglobin variability. Other biomarkers such as parathyroid hormone also fluctuate widely in patients with chronic kidney disease. Application of time series methods to these problems can provide avenues for new discoveries and research.

**Time Series Modeling and Analysis**

*Time series analysis* is a set of mathematical methods that help us understand the nature of the time series and what they represent. Time series models are helpful for characterization of a measured process but, more importantly, to derive inferences about the underlying physiology or pathophysiology and to forecast and predict outcomes.

**Types of Time Series Analysis**

Time series methods can be broadly classified as either linear or nonlinear methods.

*Linear analysis of time series* can be performed either in the frequency domain (spectral analysis) or in the time domain (autocorrelation and cross-correlation analysis). A time domain graph shows how a signal changes over time. In many cases, the most distinguished information is hidden in the frequency content of the signal. A given time signal can be converted to its frequency domain using the Fourier transform:

\[
y_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} y_n e^{i\frac{2\pi}{N}kn},
\]
where the frequencies $f_k = k/N\Delta t$, $k = -N/2, \ldots, N/2$ and $\Delta t$ is the sampling time interval. The power spectrum of a discrete signal is defined as the square of the amplitude $|y_k|^2$ of frequency $f_k$.

An illustrative example is shown in figure 2. Figure 2b shows the tracing of nephron tubular pressure over time (a time series). Figure 2a shows a frequency domain graph from the same data. The characteristic pattern of peaks seen at varying frequencies makes it apparent how information hidden in the time domain graph is clearly visible on the frequency domain graph. The frequency domain graph is also called a power spectrum graph. It shows how much of a given signal lies within each given frequency band over a range of frequencies. If a signal changes rapidly, we say that it is of high frequency (HF). If the changes are spaced apart in time, we expect to see a low frequency (LF) signal. A power spectrum can also help interpret the nature of the change. If the signal slowly changes over time, i.e. smoothly transitions from one state to another, we expect to see a hump in the power spectrum rather than a peak. If this signal does not change at all, then we say it has zero frequency or no frequency. The frequency is measured in cycles per second or ‘Hertz’ (the more common name).

**Figure 2.** a Experimental time series graph. b Power spectrum showing characteristic peaks.

Frequency Domain Analysis – Applications in Nephrology

Power spectral analysis has been utilized extensively in the study of autonomic function (sympathovagal balance) or dysfunction in patients with chronic kidney disease [5], dialysis-dependent renal failure [6–16] and kidney transplants [17–21]. Classically, the study of autonomic function has involved the administration of either pressor (phenylephrine or angiotensinogen) or depressor stimuli (amyl nitrite or sodium nitroprusside) and assessment of changes in heart rate produced thereof. Power spectral analysis can be used for rapid assessment of the integrity of the cardiovascular autonomic reflexes at the bedside [22]. Beat-to-beat tracings of BP and pulse interval over time domain are used to generate the power spectrum. The characteristic peaks seen in the power spectrum graph can then be correlated with underlying physiological processes. By studying the number, frequency and amplitude of the oscillatory components one can detect the underlying rhythmic factors affecting the cardiovascular system, including the cardiac and respiratory cycle, and vasomotor activity [7]. This form of rhythmicity associated with the baroreceptor arc is a part of normal physiology and does not require the administration of pharmacological agents. The power spectrum of pulse interval variability has been shown to contain two distinct components – a HF and a LF component. The interpretation of these peaks begs circumspection from the reader. The coexistence of two peaks in the power spectrum of a physiological process can correspond to two essentially different situations. There could indeed be two independent modes or a single mode whose instantaneous frequency changes in time from one value to another. This brings up another important cautionary note for the reader. Confirmatory evidence is required to prove that the hypothesized physiologic process is indeed the causal mechanism resulting in the peak of the power spectrum.
The HF power (0.20–0.35 Hz) of the decomposed spectrum of pulse interval variability has been shown to be a marker of vagal activity. Physiological measures known to increase vagal tone, such as controlled respiration, cold facial stimulation and rotational stimuli, have been shown to increase the HF peak [23, 24]. Conversely, pharmacologic blockade of vagal drive using atropine abolishes the HF peak [23, 24]. A similar approach can be used when analyzing the power spectrum of any signal. A stepwise algorithm can include the identification of characteristic peaks in power spectra, confirming underlying physiologic mechanisms generating the peak and finally being able to attribute changes in the amplitude or frequency of the peaks to pathogenic disease processes. When power spectral analysis was applied to study the LF component (0.05–0.15 Hz) of pulse interval and BP variability, it was determined that while it is significantly enhanced by measures that stimulate the sympathetic nervous system such as active standing or tilt, it is not entirely abolished by β-blockade. Therefore, the interpretation of the LF component has remained controversial. The results of studies using frequency domain techniques to assess autonomic function in patients with chronic kidney disease (table 1) and kidney transplants (table 2) are summarized.

Our next focus is on an area of renal physiology which bears no similarity to the cardiovascular physiology described above, but for the identification of existing time series and the application of time series methods. Renal autoregulation refers to the ability of the kidney to maintain a relatively constant blood flow over a wide range of perfusion pressures. Classically, renal autoregulation has been studied utilizing micropuncture experiments to perfuse the distal tubules and measuring changes in the single nephron glomerular filtration rate in response to changes in tubular flow. The invasiveness and technical expertise required to conduct these experiments are apparent from the description. An alternative and noninvasive method for studying renal autoregulation is spectral analysis of the time series of single nephron blood flow measured by laser Doppler velocimetry [25]. The power spectrum (fig. 2) shows characteristic peaks seen at different frequencies. A HF peak seen at 0.2 Hz has been correlated with the myogenic response and the LF peak seen at 0.035 Hz represents the tubuloglomerular feedback (TGF). Administration of furosemide blocks TGF and has been shown to abolish the LF peak [26]. This confirms the correct attribution of TGF as the mechanism underlying the LF peak seen.

We have seen that spectral analysis gives us useful insights into renal autoregulation; however, it is not the best-suited mathematical tool available today for its study. Spectral analysis was developed for the analysis of invariant (stationary) time series. A stationary time series has a constant mean and a constant variance over time. However, both the myogenic and TGF mode give rise to non-stationary time series [27]. Interactions between such oscillatory systems are known to lead to nonlinear outcomes, including synchronization, amplitude and frequency modulation, and chaos [28]. To cope with the nonstationarity and inhomogeneity of biological time series, time frequency analysis using wavelet transforms offers a more suitable investigation approach. A spectral graph only provides information about the periodicities in a time series by summarizing the distribution and density of a particular frequency range in a signal, whereas wavelet analysis describes how the frequency of a signal changes over time and allows us to study the temporal dynamics of the rhythmic components. Marsh et al. [27] used wavelet analysis to study the instantaneous amplitude and frequencies of the myogenic response. By using a second wavelet transform, they derived the amplitudes and frequency of the modulation signal. They observed that the modulation signal operated in the same frequency ranges as TGF suggesting that TGF was responsible for modulation of the myogenic response. The frequency and amplitude of the modulation parameters have been found to significantly differ between normotensive and hypertensive rats [29], providing insights into the pathophysiology of disease states. Sosnovtseva et al. [30] applied wavelet analysis to demonstrate synchronization among nephrons that arise from the same cortical radial artery allowing them to function as a cohesive unit. At the same time, hypertensive rats were found to have reduced synchronization [30] or chaotic fluctuations [31, 32]. Another advantage of using wavelet analysis over spectral analysis is related to improved resolution over the LF scales. It is not surprising then that wavelet analysis has allowed the discovery of a very slow oscillatory mode, the origin of which is still not certain. This very slow oscillatory mode is able to modulate both the myogenic and TGF responses [29]. The wavelet transformation allows utmost flexibility to extract important information about the signal in time and frequency domain simultaneously, serving as a more powerful tool to detect specific regions within the signal, which may not appear significant in the frequency domain alone. Wavelets also allow us to properly analyze nonstationary time series which are frequently encountered in physiology and medicine. Application of wavelet analysis to data from existing experiments [27] conducted at a time such analysis was not available has allowed the discovery of the modulation characteristics of the myogenic re-

Time Series in Nephrology

Nephron Physiol 2013;124:14–27
DOI 10.1159/000356382

Downloaded by: 54.70.40.11 - 11/25/2017 4:27:58 AM
<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Population</th>
<th>Controls, n</th>
<th>Findings</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axelrod et al. [5]</td>
<td>To use a quantitative, noninvasive method of assessing autonomic control,</td>
<td>10 CKD 8 HD 7 PD</td>
<td>30</td>
<td>Reduced LF power in all patients vs. controls. Chronic renal failure &gt; PD &gt; HD (nonsignificant trend)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>based on the spectral analysis of beat-to-beat fluctuations in HR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloarec-Blanchard et</td>
<td>To investigate autonomic mechanisms underlying persistent hypotension in</td>
<td>6 HD</td>
<td>6</td>
<td>Reduced LF power in HD vs. controls when standing but no differences in supine position</td>
<td>These results indicate that an impairment in both parasympathetic and sympathetic functions exists in HD patients with persistent hypotension and that the apparent sympathetic dysfunction could result from a reduction in cardiovascular responsiveness to sympathetic stimulation</td>
</tr>
<tr>
<td>al. [6]</td>
<td>long-term HD patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takahashi et al. [7]</td>
<td>To investigate autonomic mechanisms underlying persistent hypotension in</td>
<td>11 HD patients prone to IDH (group H) 11 normotensive patients (group N) and 10 healthy controls (group C)</td>
<td></td>
<td>The HF amplitude, and LF/HF ratio, an index of sympathetic predominance, were in the order of groups C &gt; N &gt; H (p &lt; 0.01). While the HF amplitude decreased with standing in all three groups (p &lt; 0.05 for all), the LF/HF ratio increased only in groups N and C (p &lt; 0.05 for both) but not in group H. Conversely, plasma noradrenaline level was in the order of groups C &lt; N &lt; H (p &lt; 0.001). Furthermore, while the LF/HF ratio correlated positively with the plasma noradrenaline level in group C (r = 0.73, p &lt; 0.01), it correlated negatively in group H (r = 0.69, p &lt; 0.05)</td>
<td>These changes support compensatory baroreflex-mediated activation of the sympathetic nervous system (and suppressed parasympathetic activity) during ultrafiltration-induced intravascular volume depletion. Also, activation of the sympahto-inhibitory cardiodpressor reflex (Bezdj-Jarisch reflex), which is a physiologic response to a critical reduction in intravascular volume and cardiac filling, is the cause of sudden IDH</td>
</tr>
<tr>
<td></td>
<td>chronic kidney disease (CKD) patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Studies in chronic kidney disease (CKD) patients
<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Population</th>
<th>Controls, n</th>
<th>Findings</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giordano et al. [11]</td>
<td>To study the effects of the HD procedure on autonomic function</td>
<td>11 diabetic HD patients; 9 nondiabetic HD patients</td>
<td>10 healthy subjects; 10 type II diabetic patients</td>
<td>ESRD patients had increased baseline sympathetic activity which increased further during dialysis when compared to the predialytic period in both diabetic and nondiabetic patients. Sympathetic activity remained elevated in diabetic HD patients even in the postdialytic period while the nondiabetic HD patients shifted to a parasympathetic predominance</td>
<td>Nondiabetic HD patients have reversible changes in HRV (which is correlated to a change in BUN). Diabetic HD patients on the other hand have autonomic dysfunction related to both diabetes and uremia</td>
</tr>
<tr>
<td>Karayaylali et al. [12]</td>
<td>To study the role of autonomic dysfunction in IDH</td>
<td>22 HD patients</td>
<td>20 age and sex matched</td>
<td>LF and HF were decreased in HD patients. HF spectral power was significantly correlated with dialysis duration, end systolic volume, end diastolic volume, ejection fraction and fractional shortening</td>
<td>Cardiac autonomic neuropathy is dependent on ESRD duration</td>
</tr>
<tr>
<td>Lee et al. [13]</td>
<td>To study the role of autonomic dysfunction in IDH</td>
<td>12 IDH-prone HD patients</td>
<td>12 age- and sex-matched HD patients without IDH</td>
<td>In the control group, the LF/HF ratio power index rose progressively and reached significantly high levels at 4 h of HD compared to the basal values. In the group prone to hypotension, the power index remained unchanged and systemic vascular resistance was lower than that in the control group</td>
<td>Autonomic dysfunction contributes significantly to IDH</td>
</tr>
<tr>
<td>Polak et al. [14]</td>
<td>To study the relationship between high PTH and autonomic neuropathy</td>
<td>21 HD patients with PTH &gt;275</td>
<td>19 HD patients with PTH &lt;275</td>
<td>The values of lnTP and lnLF were lower in patients with higher PTH levels. lnPTH was negatively correlated and independently related to lnTP, lnLF and lnHF</td>
<td>PTH affects the activity of both parts of the autonomic nervous system: sympathetic and parasympathetic. A high PTH level deteriorates total autonomic activity</td>
</tr>
<tr>
<td>Di Leo et al. [15]</td>
<td>To study the effect of age on autonomic function in ESRD patients</td>
<td>30 HD patients</td>
<td>32 age- and gender-matched normal individuals</td>
<td>In uremic patients, LF band on lying was significantly lower in elderly uremics when compared with elderly controls, whereas LF band on standing was significantly lower in elderly than in younger uremics</td>
<td>There is a significant inverse relationship between aging and LF band values, especially in uremics. A sympathetic autonomic derangement is greatly present in older uremics, in particular after 50 years of age</td>
</tr>
<tr>
<td>Vita et al. [16]</td>
<td>To study uremic autonomic neuropathy by spectral analysis of HR</td>
<td>30 HD patients</td>
<td></td>
<td>A significant reduction in the LF value on supine uremic patients who were not found to have autonomic neuropathy by conventional autonomic function testing suggests that an early sympathetic involvement exists that traditional autonomic tests were unable to detect</td>
<td>Autonomic dysfunction in chronic uremic patients involves both the sympathetic and parasympathetic pathways. It is not limited to the parasympathetic system</td>
</tr>
</tbody>
</table>

BUN = Blood urea nitrogen; ESRD = end-stage renal disease; HD = hemodialysis; HR = heart rate; HRV = heart rate variability; IDH = intradialytic hypotension; MAP = mean arterial pressure; PD = peritoneal dialysis; PTH = parathormone.
### Table 2. Studies in kidney transplant recipients

<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Population</th>
<th>Controls, n</th>
<th>Findings</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yildiz et al. [17]</td>
<td>To study the effect of renal transplantation on uremic autonomic dysfunction with time and frequency domain HRV</td>
<td>14 posttransplant patients</td>
<td>14 chronic HD patients before transplantation</td>
<td>LF and HF spectral power were found to be significantly increased after renal transplantation, but the LF/HF ratio was not different from the pretransplant period</td>
<td>Autonomic dysfunction in HD patients is reversible and renal transplantation reverses the sympathetic and parasympathetic autonomic dysfunction simultaneously and at a relatively early stage</td>
</tr>
<tr>
<td>Gerhardt et al. [18]</td>
<td>To evaluate the effect of transplantation on baroreceptor sensitivity</td>
<td>20 kidney transplant recipients; 20 normotensive HD</td>
<td>20 age-matched healthy volunteers</td>
<td>The 66- to 129-mHz component (Mayer waves) of HR and BP spectrum in normals significantly exceeded its equivalent in kidney transplant recipients and patients on HD</td>
<td>Both in kidney transplant and dialysis patients a decreased activity of the generating compounds of the sympathetic nervous system can be detected compared to healthy volunteers</td>
</tr>
<tr>
<td>Rubinger et al. [19]</td>
<td>To evaluate the effect of SPK on autonomic function</td>
<td>HD patients</td>
<td>Age-matched healthy adults; transplant recipients</td>
<td>Renal transplantation normalized HRV in most patients</td>
<td></td>
</tr>
<tr>
<td>Boucek et al. [20]</td>
<td>To evaluate the effect of SPK on autonomic function</td>
<td>82 patients with DM I who were insulin and dialysis free following SPK</td>
<td>32 age-matched healthy controls and 13 patients with kidney transplant alone matched for age and duration of DM I at a comparable time point after transplantation</td>
<td>No significant improvement was seen in SPK patients in the posttransplant period and no differences were found compared with patients with kidney transplant alone</td>
<td>SPK does not lead to significant improvement in autonomic function at 25 months (±15) after SPK</td>
</tr>
<tr>
<td>Rubinger et al. [21]</td>
<td>To evaluate the effect of SPK on autonomic function</td>
<td>52 nondiabetic HD patients, 44 transplanted patients, 24 in the 1st year after renal transplantation and 20 &gt;1 year after renal transplantation</td>
<td>41 age-matched controls with normal renal function</td>
<td>In HD patients, BPV was increased, while HRV, BRS slope and LF and HF coefficients were markedly decreased compared with controls. Renal transplantation was associated with normalization of BPV at short term (≤1 year). In patients with long-standing functioning grafts (&gt;1 year), baroreceptor indices were significantly increased and returned to values similar to those of the control subjects</td>
<td>Renal transplantation improves BP and HRV and restores baroreflex function to near normal range on the long-term follow-up</td>
</tr>
</tbody>
</table>

BPV = BP variability; DM = diabetes mellitus; HD = hemodialysis; HR = heart rate; HRV = heart rate variability; SPK = simultaneous pancreas-kidney transplantation; BRS = Baroreceptor sensitivity.
**Fig. 3.** Interdialytic weight gain (IDWG) time series.

**Fig. 4.** Patterns of predialysis weights (PREWT) in dialysis patients. **a** Pattern 1: minimal fluctuation of predialysis weights. **b** Pattern 2: rising and falling predialysis weights.
response by TGF not known priorly. Important differences in this modulatory response have been observed in hypertension and hence allowed new knowledge discovery.

**Time Domain Analysis – Applications in Nephrology**

The simplest form of a time domain graph is to plot a signal (y-axis) over time (x-axis) and study the resulting graph. However, in order to mathematically describe a time series, an autocorrelation function (ACF) and a partial ACF (PACF) are needed. The ACF represents the covariance of an observation with its own time-lagged values. In biological time series, we do not expect serial measurements over time to yield independent observations. The ACF tells us how a variable is correlated to its own values in the past and future.

\[
RN(\tau) = \frac{\sum_{n=1}^{N} (Y_{n+\tau} - \bar{Y})(Y_n - \bar{Y})}{\sum_{n=1}^{N} (Y_n - \bar{Y})^2}; \bar{Y} = \frac{1}{N} \sum_{n=1}^{N} Y_n. \tag{3}
\]

The PACF measures correlation between observations that are \( k \) time periods apart after controlling for correlations at intermediate lags (i.e., lags less than \( k \)). In other words, it is the correlation between \( Y_t \) and \( Y_{t-k} \) after removing the effects of intermediate \( Y' \)'s. At lag \( \tau = 1 \), the ACF and PACF values are equal, since there are no intermediate lag effects to eliminate. Let us consider an example: the time series in figure 3 represents the interdialytic weight gain for a patient over a period of 3 years. While it is easy to see that the mean interdialytic weight gain for this patient has been 3.5 kg over the past 3 years, there is very little information in this regarding the trend of interdialytic weight gain over time? How do we detect the presence or absence of any systematic trends in this series? Would it be useful to systematically study patients whose nutritional status/weight does/does not fluctuate (fig. 4)? We begin our exploratory analysis with a plot of ACF and PACF. The most striking feature of ACF (fig. 5) is the strong correlation observed between every 3rd observation. This is not a surprising finding since a typical thrice-weekly dialysis patient often comes in more volume overloaded at the beginning of the week. While this is not new information, the elegance of time series methods in detecting this is profound. The PACF demonstrates significant spikes till a lag of 3, and this has implications to model fitting. The pattern of ACF and PACF is compared with known theoretical patterns for satisfactory model selection (table 3).

**Time Series Models**

Autoregressive (AR), moving average (MA) or the combination of both: ARMA and ARIMA models can be fit to the data. ACF and PACF are used in the appropriate selection of the model type (table 3). The model residuals obtained are then checked for normality and randomness. By confirming that the residuals are truly random, we can conclude the lack of other systematic trends than those captured in the model. Once a satisfactory model has been obtained, it not only characterizes the time series, it can also be used for forecasting. There is one important caveat here. Before using any time series for this type of modeling and forecasting, we need to make sure that the series is stationary. If not, we often have to transform it to achieve stationarity. A stationary time series has a constant mean and a constant variance over time. Further, the covariance between any two time periods only depends on the distance between the two time periods. Therefore, if we know the separation between any two
points in a stationary time series, we know how they are related to each other. In simple words, a stationary time series repeats itself after a known time interval (fig. 6). This is not uncommon in physiologic domains. A non-stationary time series can only be used to draw inference about the time period that it represents while a stationary time series can be used to make generalizations about the future. When one encounters a nonstationary time series it must be transformed into a stationary time series. This can be done by a variety of mathematical operations. One of the commonly employed methods of removing non-stationarity is differencing (each observation is subtracted from the immediately previous observation). One may find that, indeed, the resulting time series is stationary, which can now be used for creating models. As an example, consider a process that increases or decreases by a constant value over each successive time period. This will result in a nonstationary time series that is amplified or dampened over time. By differencing, the series can become stationary. Using the correct methods of analysis as described above, we can now arrive at a logical deduction of the nature of the underlying process.

ARIMA models have been used in renal-related health services research. Examples can be found in a study of population-based interventions on laboratory utilization published in *JAMA* [33]. The authors used time series analysis to study the effect of a guideline disseminated in August 1994 in Ontario, Canada, stressing the futility of
urine microscopy in asymptomatic patients. Time series models were used to forecast the trends of laboratory utilization in the absence of intervention and were compared to actual trends observed. The differences in the actual and predicted rates were then used to calculate the numbers of tests avoided and cost savings thereof. The authors were also able to study the effects of physician education, changes to laboratory requisition forms and/or policy to change financial payments independently and when introduced together. In another paper published in JAMA in 2011, Hemmelgarn et al. [34] examined the trends in nephrologist visits and health care resource use before and after reporting of estimated glomerular filtration rate was implemented by the laboratory. The authors listed the use of time series analysis as a strength of the study design. In a nonrandomized setting, the interrupted time series design with a concomitant comparison group is the strongest quasi-experimental approach for evaluating longitudinal effects of interventions. It allows us to study the effect of an intervention while accounting for underlying trend or seasonal influences. Lam et al. [35] examined the effects of the publication of the 4D (Die Deutsche Diabetes Dialyse) study on statin prescription. The authors appropriately considered an ARIMA model for their analysis but chose a linear model instead, since statistical assumptions for the linear model were met. Specifically, the Durbin-Watson statistic for autocorrelation was reviewed but chose a linear model instead, since statistical assumptions for the linear model were met. Specifically, the Durbin-Watson statistic for autocorrelation was reviewed and autocorrelations and inverse and partial autocorrelations were visually inspected. An AR model has also been used in an attempt to forecast the number of patients with end-stage renal disease in the United States [36].

Cross-Correlation Analysis

One of the most challenging tasks in medicine is establishing causality. As physicians, we are interested in determining if a given time series is being driven or driving another time series. If yes, what is the nature of the driving force? Is there a lag between the intervention and the outcome? For example the time series of hemoglobin over time is being driven by the time series of erythropoietin over time. For this purpose, the two time series can be cross-correlated. Cross-correlations are computed after removing the autocorrelations present within the series, a procedure known as pre-whitening. In other words, we first de-trend the series to remove systematic trends within the series and then cross-correlate them to obtain a true relationship between two series. Pre-whitened cross-correlations have several applications in nephrology. They can be used to study the delay between BP reductions in response to a decrease in the dry weights of the patients. Graphic representation of time trends in dialysis patients can help avoid information overload and improve many patient outcomes. While these will need to be formally tested in randomized controlled trials, the potential applications of using time series models in nephrology appear very promising.

Nonlinear Time Series Analysis

Chaos theory developed in the later part of the last century, when biologists, physicists and mathematicians observed phenomenon that could not be explained by linear theory alone. What has often been called noise or error in linear models has been found to be an essential component of some systems studied under the chaos theory. Very small changes in initial conditions can lead to very different outcomes, making long-term predictions for a class of systems impossible (e.g. meteorology). Chaos theory has found several applications in medicine [37]. Attempts at the application of these concepts to research related to kidney diseases are just getting started.

The Poincaré plot is a geometric method which allows us to study nonlinear dynamics of a complex system. Lerman et al. [38] have utilized the Poincaré methodology to elucidate nonlinear information contained within heart rate variability data obtained before and after hemodialysis that cannot be detected by power spectral techniques. The plot allows us to represent a time series in a two-dimensional plane (or phase space). Every pair of points \((Y_i, Y_{i+\tau})\) that can be obtained using a time series and a fixed lag \(\tau\) \(\{(Y_i, Y_{i+\tau})| i = 1, N - \tau\}\) represent a point on the plot. The plot can be interpreted by simple visual inspection of its shape. A common way to objectively analyze the geometry of the plot is to fit an ellipse to the plot. The ellipse is fitted onto the so-called line of identity at 45° to the normal axis. The standard deviation of the points perpendicular to the line of identity denoted by SD1 describes short-term variability. The standard deviation along the line of identity denoted by SD2 describes long-term variability. The ratio SD1/SD2 has also been used to characterize the shape of the Poincaré plot.

The term fractal was coined by Mandelbrot in 1975. Fractals are self-similar patterns that repeat themselves at different scales of representation giving rise to complex geometric forms. Biological time series can often be decomposed into fragments that repeat themselves over time. Biological images also exhibit similar behavior.
Fractal analysis methods provide a way to describe and quantify the irregular forms and patterns in 1D, 2D and 3D signals. The fractal dimension \( d_f \) provides information on the self-similarities of a waveform, contour or texture and has the following expression:

\[
d_f = \lim_{s \to 0} \frac{\log N(s)}{\log(1/s)},
\]

where \( N(s) \) is the number of squares with side length that contain a part of the analyzed object. If this limit exists, it is called the 'box-counting' dimension of the measured object. In practice, this limit converges slowly, that is why the following expression is used:

\[
\log(N(s)) = d_f \cdot \log \left( \frac{1}{s} \right).
\]

This represents the equation of a straight line of slope \( d_f \), the 'log-log' curve described by the points of \((\log(N(s)), \log(1/s))\) for different values of the side of the square \(-s\). The slope of the line that approximates the distribution of the points – the fractal dimension – can be determined by linear regression (least square method).

Fractal dimension can be used for signal and image analysis; for the characterization and classification of different waveforms and patterns; for delimitation between normal and affected tissue, lesions, and tumors. The simplest medical application is the morphological analysis of a structure, e.g. the renal arterial tree [39]. The renal microvasculature has been found to have the largest fractal dimension compared to microvasculature networks in the cortex, skin and thigh muscle [40]. The shapes of the renal proximal tubular cells have also been determined to be governed by fractal rules [41]. The box-counting method estimates the fractal dimension \( d_f \) of a texture, contour or waveform. The box-counting algorithm is easy to implement and has been used extensively, with several applications including analysis of complex images.

It has generally become accepted that the heart rate is of chaotic (highly nonlinear) nature. Therefore, the acquired time series of heart rate has a fractal-like structure. In general, any measure (entropy or unpredictability, fractal dimension or scaling exponent) that shows a decrease in complexity of the heart rate signals indicates an increased risk of heart disease. In hemodialysis patients, decreased approximate entropy has been found to predict mortality independent of the triangular index and ultra-LF power [42].

A decrease in the natural complexity of heart rate can also be demonstrated by calculating the scaling exponent (similar to \( d_f \)) in the long-term dynamics (>11 heart beats) and has also been associated with depression in hemodialysis patients [43].

Nonlinear time series analysis derives form chaos theory and encompasses the methods described above along with others such as Lyapunov exponents and approximate entropy. In general, one can reconstruct the attractor of a dynamical system from the time series generated by the system using Takens' embedding theorem and compute the correlation dimension of the attractor in order to geometrically characterize it.

### Problem Solving Using Time Series

Time series analysis is useful to characterize data trends where classical statistical methods alone may not be sufficient. Let us consider another example to illustrate this. If we have to compare BP control in 2 subjects over a 6-month period, an approach can be to calculate the mean BP for each subject and use a Student’s t test for comparing the means. While this approach is certainly valid, it provides us with no information regarding the time trend of BP. Essentially by summarizing the BP as a mean, we are losing information about trend over time which may be of interest. A paired t test perhaps provides a slightly better way of doing this, but can only be used to compare paired observations over time in the same subject. Logistic regression is built on the assumption of independence of individual observations. In a time series, individual observations are inherently dependent on one another since they are derived from the same subject. Therefore, logistic regression does not present us with a way of analyzing time series data. Repeated-measure ANOVA is perhaps the best suited of all methods, but it is limited by the assumption of equal variance over time. Moreover, in order to use ANOVA, one needs to design experiments to fit with comparison structures for which ANOVA can be used.

When faced with a problem involving time series data, the first step should include data visualization by simply plotting the observations over time. This is analogous to making a scatter plot or bar graph to characterize the data. Before proceeding further, the quality of the data should be verified. Any gaps in the time series should be identified. Missing observations can be dealt with by interpolation between the gaps. This can be achieved either via linear interpolation or via the application of smoothing filters (which are available in several statistical packages with time series capabilities). Smoothing of a time series can also reduce noise and eliminate outliers. One needs...
to be careful to avoid oversmoothing, which will make the data lose its original characteristics. If there are too many missing observations, smoothing is not recommended. Frequency analysis can be carried out at this stage. The next step involves characterization of the time series by plotting the ACF and the PACF described above. The series must also be tested for stationarity. Several methods are available, one is using ACF (if the ACF cuts off or dies quickly than the series is stationary). One important caveat here is that the time series needs to be of sufficient length (enough data points) for this to be done successfully. If there is nonstationarity, it needs to be corrected, for example by differentiating the time series. Once a stationary data series is obtained, appropriate models can be fit by analyzing ACF and PACF. One of the tenets of model building is parsimony. One should try to achieve the best model fit using the least number of parameters in the model. Conclusions from models built and any predictions made are only as good as the model itself. Therefore, the model needs to be evaluated for adequacy. The time series of residuals can be used to plot an ACF and a PACF of the residuals. If the residuals are random, one can conclude that the best model has been fit to the data. These models can be used to make predictions about time trends being studied. Popular statistical packages such as SAS and SPSS both offer an impressive suite of application to deal with time series data. Several textbooks [44–47] are available to help the inclined reader gain further expertise using these statistical packages for time series analysis.

Conclusions

This introductory review serves to provide an overview of time series models and analysis methods to investigators and clinicians who are looking for a solution to a concrete problem. In order to completely understand a phenomenon that contains temporal information, the data should be viewed as a sequence of events. Classical data analysis techniques tend to treat temporal data as unordered sequence of events ignoring the temporal sequence. A great deal of new knowledge can be gained by using appropriate methods of temporal data mining. Common statistical software packages such as SAS and SPSS provide impressive time series capabilities. MATLAB also provides a powerful suite of time series analysis methods and programming capabilities. Using these tools, it is possible to develop computer models and deploy them in an interactive user-friendly environment. The reader is referred to several textbooks [44, 45] dealing with the theory and application of time series analysis.

Time series methods have been used successfully in several areas of medicine, including the characterization of cardiac arrhythmias and the analysis of electrocardiograms and electroencephalograms.

Time series modeling and analysis offer avenues for new knowledge discovery and they have several advantages. They facilitate the understanding of dynamic phenomena. A deeper understanding of processes both at the micro (cellular) and macro (individual) level allows us to design better experiments. By studying phenomena at an individual scale rather than a population scale, they offer a way to personalize treatments. They allow indirect insights into a condition using data obtained by minimally invasive or noninvasive procedures.

At the same time, the reader should also learn to recognize some limitations of these techniques. The quality of the analyzed or modeled time series can affect the results (too short, nonstationary and too noisy). A time series with a minimum of 50 data points is recommended to carry out the analysis. If the investigator wishes to study seasonal or cyclic trends, at least 4–5 cycles should be observed in the series. Obtaining repeated measurements over extended periods of time can add to the costs of experiments. Missing values are often encountered in biologic time series. Although the use of interpolation and smoothing filters can rectify these to some extent, one must be cautious regarding introducing distortion or bias in the data when performing these preprocessing operations. Drawing predictions form a nonstationary time series will give misleading results if caution is not exercised in choosing the correct methods of data analysis. Readers involved in the planning and execution of research projects should carefully consider utilizing time series models and methods in the appropriate circumstance to their advantage.

Disclosure Statement

There are no conflicts of interest.

References


5 Takahashi H, Matsuo S, Toriyama T, Kawa-


8 Gerhardt U, Riedsch M, Steinmetz M, Ho-

9 Rubinger D, Sapoznikov D, Pollak A, Pop-


19 Grabczewska Z, Przybyl R: Effect of parathor-

20 Robinson TG, Carr SJ: Cardiovascular auto-

21 Sandgaard NC, Andersen JL, Holstein-Rath-

22 Robinson TG, Carr SJ: Cardiovascular auto-

23 Sandgaard NC, Andersen JL, Holstein-Rath-


28 Pikovskiy A, Rosenblum M, Kurths J: Syn-
chronization: A Universal Concept in Non-
linear Sciences. Cambridge, Cambridge Uni-


31 Yip KP, Holstein-Rathlou NH, Marshall DJ: Chaos in blood flow control in genetic and re-

32 Yip KP, Holstein-Rathlou NH: Chaos and non-linear phenomena in renal vascular con-

33 van Walraven C, Goel V, Chan B: Effect of population-based interventions on labora-

34 Hemmelgarn BR, Zhang J, Manns BJ, James MT, Quinn RR, Ravani P, et al: Nephrology visits and health care resource use before and after reporting estimated glomerular filter-


36 Nishinoya M, Goer WH, Koomans HA: Hemodynamic patterns and spectral analysis of heart rate variability during dialysis hypoten-


38 Giordano M, Manzella D, Paolissio G, Calien-

39 Kojima M, Hayano J, Fukuta H, Sakata S, Mu-

40 Lam NN, Jain AK, Hackam DG, Cuerden MS, Suri RS, Huo CY, et al: Results of a random-


46 Box GE, Jenkins GM, Reinsel GC: Time Se-

47 Pankratz A: Forecasting with Univariate Box-