Published in final edited form as:

Chaos Solitons Fractals. 2012 June 1; 45(6): 853–860. doi:10.1016/j.chaos.2012.03.003.

Estimation of time-delayed mutual information and bias for irregularly and sparsely sampled time-series

D. J. Albers^{1,*} and George Hripcsak^{1,†}

¹Department of Biomedical Informatics, Columbia University, 622 W 168th St. VC-5, New York, NY 10032

Abstract

A method to estimate the time-dependent correlation via an empirical bias estimate of the time-delayed mutual information for a time-series is proposed. In particular, the bias of the time-delayed mutual information is shown to often be equivalent to the mutual information between two distributions of points from the same system separated by infinite time. Thus intuitively, estimation of the bias is reduced to estimation of the mutual information between distributions of data points separated by large time intervals. The proposed bias estimation techniques are shown to work for Lorenz equations data and glucose time series data of three patients from the Columbia University Medical Center database.

Keywords

time-delay mutual information; nonlinear time-series analysis; information theory; high dimensional data; chaos; time-delay dynamical systems; Lorenz equations; physiology; non-uniform sampling

I. INTRODUCTION

In many experimental and computational contexts, experiments are designed in such a way that the data that are collected are uniformly and densely sampled in time, stationary (statistically), and of sufficient length such that various time-series analysis can be directly applied. Nevertheless, there is an increasing number of scientific situations where data collection is expensive, dangerous, difficult and for which there is little that can be done to control the sampling rates and lengths of the series.

This paper primarily focuses on establishing a methodology for confirming the existence of *time-based* correlation for data sets that may include highly non-uniform sampling rates and small sample sets. To achieve this, we will use an information-theoretic framework, specifically time-delayed mutual information [1] [2] [3] (TDMI), which does not rely on sampling rates, but rather on an overall number of data points used to estimate the TDMI for a given time-separation. Estimating the TDMI requires estimating probability densities

© 2012 Elsevier Ltd. All rights reserved

*Electronic address: david.albers@dbmi.columbia.edu

†Electronic address: hripcsak@columbia.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(PDF), and PDF estimates fundamentally have two components, the PDF estimate and the *bias* (or noise floor) associated with the estimation of the PDF.

The primary motivation for this work originates in work with physiologic time series extracted from electronic health records (EHR) [4, 5]. EHR data are never collected in a controlled fashion, and thus are pathologically irregularly sampled in time. Nevertheless, EHR data will likely be the only medium- to long-term source of population scale human health data that will ever exist because human data can be expensive, dangerous, and intrusive to collect. Therefore, if we ever wish to understand long-term physiologic trends in human beings, we must find a way to use EHR data, and that means finding a way to cope with time series whose values are mostly missing yet not missing at random. Here we wish to establish a method to begin to quantify predictability of a patient based on their physiologic time series. We hope to apply predictability in at least three ways. First, in testing physiological models on retrospective EHR, such as in glucose metabolism, we have found that while it is sometimes impossible to distinguish two models based comparing their outputs to raw data, sometimes models can be distinguished by comparing their estimates of derived properties like predictability [6]. Second, in carrying out retrospective research on EHR data, we often want to include only those patients whose records are sufficiently complete, yet there is no simple definition of completeness. We propose to use predictability as a feature that correlates with completeness: a record that is unpredictable across a wide range of variables may be so because it is incomplete. And third, the predictability of a set of clinical or physiological variables may be used as a feature to define a disease or syndrome (or, for example, to subdivide the disease or syndrome). Such a feature may add information beyond simple means, variances, and simple characterizations of joint distributions.

II. NONLINEAR CORRELATION IN TIME: CALCULATING THE TIME-DELAYED MUTUAL INFORMATION

A. Non-uniformly sampled time-series: basic notation

Begin by defining a scalar time-series as a sequence of measurements x taken at time t by $x_{t(i)}$ noting that t denotes the *real time* that the measurement was taken, and i denotes the *sequence time*, or sequential numeric index that denotes the measurement's sequential location in the time series. Denote differences in real time by $\delta t = t(i) - t(j)$ where $i \geq j$. Similarly, denote differences in sequence time by $\tau = i - j$ where $i \geq j$.

B. Time-delayed mutual information

Given two *distributions* whose entropy [7] can be defined, the amount that one distribution's information or differential entropy is *decreased* by knowing the other's distribution is known as the *mutual information* (MI) between the two distributions [8] [7]. We denote the joint probability distribution between a single variable measured at a time t and that same variable measured at a time $t - \delta t$ by $p(x_t, x_{t-\delta t})$; similarly, the marginal distributions are denoted $p(x_t)$ and $p(x_{t-\delta t})$. The *time-delayed mutual information* (TDMI) between a variable measured at time t and time $t - \delta t$ is then given by:

$$I(X_t, X_{t-\delta t}) = \int p(x_t, x_{t-\delta t}) \log \frac{p(x_t, x_{t-\delta t})}{p(x_t)p(x_{t-\delta t})} dx_t dx_{t-\delta t} \quad (1)$$

The maximum TDMI occurs for $\delta t = 0$ and is equal to the entropy of the time series. Because establishing the existence of non-null MI depends on the PDF/PMF estimation, establishing the existence of MI is primarily a problem of establishing a rough upper bound on the bias or noise floor.

C. Explicit estimation of probability mass and probability density functions

There are roughly four different ways to estimate the PDF or PMF for the mutual information calculation: (i) a standard histogram style estimate [9]; (ii) an adaptive bin size histogram style estimate [10]; (iii) a correlation function style estimate [11]; and (iv) a kernel density style estimate (KDE) [12]. In this paper we primarily use a KDE style estimate written for MATLAB [13].

D. Time-delayed mutual information for irregularly sampled or sparse data sets

In the context of the TDMI, the data set used to generate the PDFs for estimating the TDMI is generated by stepping through a time series and collecting all pairs of points that are separated by some *fixed* time or time window. If a point does not have a corresponding point δt in the past, it is not included in the data set; this means that, for a stationary process, non-uniformity in the sampling rate effectively decreases the sample size as it excludes points.

III. BIAS ESTIMATION THEORY FOR INFORMATION THEORETIC CALCULATIONS

Because the primary machines utilized for the TDMI calculation are PDF (PMF) estimators, the primary sources of error or bias of the TDMI lie with errors related to the PDF estimation. In general, there are at least three sources of bias related to PDF estimation techniques: (i) bias due to the number of points present in the sample; (ii) bias due to the resolution of the estimator (i.e., the bandwidth and the placement of the bin boundaries); and (iii) bias due to the particulars of the data set, such as non-stationarity or mixtures of statistically differing sources.

A. Estimator bias calculations for entropy and mutual information

Begin by noting that the TDMI between two *distributions* takes the following parameters as

arguments: the number of points, N ; the number of bins, b , (or bandwidth, $\frac{1}{b}$); and the real-time separation, δt , yielding $I(N, b, \delta t)$. Assuming, unrealistically, that for a given δt , $N \rightarrow \infty$ and $b \rightarrow \infty$ fast enough, then I converges to $I(\infty, \infty, \delta t) = I_\infty(\delta t)$ that has no bias and

no variance. In lieu of having $N = \infty$ and $\frac{1}{b} = \epsilon = 0$, there will exist bias. In this situation, the result of calculating I will be:

$$I(N, b, \delta t) = I_\infty(\delta t) + B(N, b, \delta t) \quad (2)$$

where $B(N, b, \delta t)$ is the bias of I at δt for a given N and b . The key problem remaining is the estimation of $B(N, b, \delta t)$.

In the case where the statistical estimator used to estimate the PDFs is the standard histogram, then the estimation of $B(N, b, \delta t)$ problem has been solved to second order in essence by Basharin [14], who calculated the bias of the entropy calculation, and in detail by Roulston (cf. section 6, on page 293 of [15]). (Higher order bias estimates of entropies have been made for some algorithms in some circumstances, c.f., [16].) The key qualitative result of Ref. [14] is, for a given $b, \delta t$, and sufficiently large N , B scales roughly as follows $B(N, b,$

$\delta t) \sim \frac{1}{N}$. Nevertheless, when the sample size is very small, the bias can be influenced by the abundance of empty bins. Translating this to mathematics, the bias can be expressed as $B(N, b, \delta t) \sim \frac{A}{N}$, where A is proportional to the amount of the support with no measurements (e.g., the number of empty bins [15]). Thus, when N is small, for histogram estimators, A can be

on the order of N or larger; for KDE estimators, because of their smoothing properties, the effects of the empty support are translated into an over-weighting of empty bins. For small data sets, histogram estimators tend to yield distributions with sharp peaks, while KDE estimators tend to yield distributions that are closer to uniform distributions.

IV. FIXED-POINT METHOD FOR ESTIMATING TDMI BIAS

We want the bias estimator to: (i) be fast, reliable and robust; (ii) be applicable to a wide variety of different estimators with different bias properties; and (iii), have the potential to depend on the data set in the presence of non-estimator bias. The four methods that follow aim at satisfying these goals.

The overall framework begins with the insight that very often the infinite-time correlation will be the same as the bias of the estimator, or $I(N, b, \infty) = B(N, b, \delta t)$, for all δt . This claim exploits the fact that we are interested in *time-dependent* correlations, and thus we are claiming that one can approximate $B(N, b, \delta t)$ with $I(N, b, \infty)$ *at least* for systems whose correlations decay to zero in time. We further propose that $I(N, b, \infty)$ can be approximated by removing or destroying the time-dependent information within the time-series and then estimating $I(N, b, \delta t)$ repeatedly and taking the average. Because the methods we propose rely, intuitively, on approximating bias with the infinite-time correlation, we call this bias method the fixed point bias, denoted as follows:

$$B_{FP}(N, b, X) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M I(N, b, X, \delta t=i) \quad (3)$$

where X is the time-series (to be defined later) whose time-dependent information has been removed. While it is possible to construct examples where infinite-time probabilistic correlations persist, in situations where the system is stationary, even when nonlinear correlations in time persist, $B(N, b, \delta t)$ can still be approximated when the time-dependent information has been removed. In contrast, when the system is non-stationary, or when multiple, possibly differing sources are aggregated, $B(N, b, \delta t)$ will likely represent both estimator and non-estimator bias.

Randomly permuted-data method

Given the two-dimensional data set used to estimate the TDMI for a given δt , randomly permute or shuffle the data in one of the marginals *without replacement* (meaning, the permutation does not replace or change any of the data points, it only changes the order of the data points).

To quantify an upper bound on the amount of time-based correlation retained by randomly permuting the data, consider the *mean distance* between the ordered and randomly permuted data sets. The mean distance *in time* between an ordered and randomly permuted *discrete-time* data set is given by:

$$\bar{\delta t} = \frac{2}{N^2} \sum_{i=1}^{N-1} \sum_{j=1}^{j < i} j \quad (4)$$

$$= \frac{N}{3} - \frac{1}{3N} \quad (5)$$

which, by the time $N = 5$, can be approximated by $N/3$. Note that when time is continuous, and integrals replace sums, one arrives at $N/3$ explicitly. Thus, the *average* term by term time correlation will be on the order of $N/3$. Nevertheless, it is important to keep in mind that while the *average* time correlation is $N/3$, the *time-based pairings have been randomized*, and thus when integrating, considerably more time-based correlation information has been lost.

Recall again that we want the bias estimation method to be applicable in circumstances where N is small. However, for small N , the permuted data method can have two issues that require quantification: (i) the effect of having a relatively short time-series which can lead to the inability to approximate the $\delta t = \infty$ fixed point as discussed in the previous paragraph (note that bootstrapping won't help when all the points are too correlated); and (ii), the effect of small sample size as related to the bandwidth/bin-width of the estimator. Thus, the following bias estimation techniques are aimed at addressing and quantifying both of these potential issues.

Uniform-mixed method

Given the two-dimensional data set used to estimate the TDMI for a given δt , replace one of the marginals with uniformly distributed random numbers.

Gaussian-mixed method

Given the two-dimensional dataset set used to estimate the TDMI for given δt , replace one of the marginals with normally distributed random numbers with a given mean and variance.

Accept-Reject generated random variable mixed method

Given the two-dimensional data set used to estimate the TDMI for a given δt , replace one of the marginals with random numbers generated by a distribution that approximates (using the accept-reject or like method) the non-ordered *distribution* of the original data-set. We employ the standard accept-reject PDF fitting algorithm with two different criteria, a strong fitting criterion (denoted AR-mixed strong) and a weak fitting criterion (denoted AR-mixed weak).

KDE-Histogram bin effect method

Given the two-dimensional data set used to estimate the TDMI for a given δt and any of the above methods for estimating bias, compare the TDMI fixed point for a KDE estimator and a histogram estimator with roughly equivalent bandwidth/bin-sizes. Because the smoothing property of the KDE, some probability mass is assigned to empty portions of the support (relative to N), thus leading to an overestimate of p for some portions of the support when N is small. In contrast the histogram estimator assigns zero probability mass for all bins with no data points, thus leading to an underestimate of the probability mass of a given bin. Thus in the circumstance where N is small, differences in the KDE and histogram estimates of the TDMI fixed point identify the existence of small sample size effects on the bias estimate.

V. IMPLEMENTATION OF THE FIXED-POINT METHOD FOR ESTIMATING TDMI BIAS

A. Data sets

The first data set we utilize is generated by the Lorenz equations [17] with the now standard chaotic parameter settings (i.e., $\sigma = 10$, $\rho = -28$, and $\beta = \frac{8}{3}$). We integrate the Lorenz

equations with a standard Runge-Kutta fourth order integration scheme with a step-size of 10^{-3} ; the resulting time-series is then sampled once per 100 steps.

The second data set we utilize includes three patients from the Columbia University Medical Center (CUMC) patient database. We selected these patients by first narrowing the set of patients to those who had glucose time series with at least 1,000 points so that we could test the bias estimate techniques on the same patient over a wide range of time series lengths. After examining a few hundred of such cases, we selected three patients who were particularly representative of the population. The three patients' time series of glucose values are shown in Fig. 1, thus displaying the sampling rates, proportion of missing values (which are interpolated as straight lines), and overall clustering and irregularity of measurement times. In particular, notice that: patient one has dense and somewhat uniform sampling over ~ 2 years; patient two is more irregularly sampled than patient one, but remains relatively uniformly sampled over a longer time period (~ 8.5 years); and patient three has bursts of measurements followed by large gaps between measurements over ~ 6.5 years.

B. TDMI estimation results

Figure 2 details the TDMI and bias estimates for the Lorenz equations. As can be seen in the long-time plot, nonlinear correlations do not completely dissipate until measurements are separated by at least 1,500 data points at the given measurement frequency. The raw TDMI estimates for both the 1,000 and 10,000 point strings are identical, certainly up to the bias and error. The TDMI results for the 100 point string, while a qualitatively different raw TDMI graph from the 1,000 and 10,000 point estimates, are not wildly different; moreover, the raw TDMI estimated on the 100 point data set is not on the same order of magnitude as any of the bias estimates. Thus, while 100 points may not be enough for an extremely accurate TDMI estimate, it is enough to establish a difference between persistent, existent correlation and bias. The variance of the various bias estimates again appears to follow a power-law in the number of points in a data string — implying that small sample size effects do not dominate the estimate of the estimator bias. And finally, the different bias estimation techniques do not differ enough to cause any difference in interpretation of the TDMI signal.

Figure 3, where the TDMI and random permutation bias is plotted for the three patients, demonstrates that the TDMI and our bias estimates are not sensitive to sampling frequency; in all cases the bias is below the TDMI signal as expected. We will now focus Fig. 4 which details the TDMI for patient one; we focus patient one because patient one has the weakest TDMI signal. First notice that the $\sim 4,000$ and 1,000 point TDMI results are very similar, including the correlation peak at 24 hours; however, the $\sim 4,000$ point TDMI appears to resolve a 48 hour peak as well, implying that the number of points does effect the detail of resolution of a given signal. Note that there is an important difference between the glucose results and the Lorenz-based results: for the glucose-based results, in general the difference between the TDMI and the bias is smaller when N is larger ($I(4000) - B(4000) < I(1000) - B(1000)$) whereas for the Lorenz-based results the opposite is true. The Lorenz-based result produces the expected outcome, whereas the glucose-based result produces a surprise outcome. This discrepancy warrants two comments. First, while the difference between the TDMI and the bias is smaller when N is larger, this effect is relatively small and does not change any of the conclusions that can be drawn from either TDMI signal. Second, it is unlikely that the glucose data source is stationary; in particular, this patient had a terminal disease and was slowly failing, and thus we hypothesize that the amount of predictive information in this patient's signal was likely decreasing (we will not attempt to further justify this hypothesis here). Refocusing on the bias estimates for the 1,000 and 4,000 point cases, overall, the various bias estimates yield very similar results (all bias estimates are on the same order of magnitude).

In contrast to the higher- N cases, the TDMI estimate for the 100 point data string rises after $\delta t \approx 48$ or equivalently after two days, and moreover differs qualitatively from the 1, 000 and $\sim 4,000$ point TDMI estimates. Thus precise TDMI estimates beyond two days with a resolution of quarter days are unlikely with only 100 points even if the points are sampled relatively evenly in time. Nevertheless, the TDMI estimated on the 100 point data set shows a sharp decay in correlation over the first 48 hours. So, while the TDMI estimated with only 100 points will not resolve perfectly, it will nevertheless retain some gross qualitative features of the higher point analogs. Finally, TDMI estimated using a 20 point data set can resolve nothing; even the raw TDMI is buried in the bias estimates and correlation decay over even the first 12 hours is not present.

C. Comparing the fixed point bias estimation methods

The features of a time series that control the number of points in a given δt bin are the length of the time series and the sampling frequency. Therefore we will not change the sampling rate and merely change the length of the time series (they have the identical effect, reducing the number of points in given δt bins).

We demonstrate how the different methods perform in three steps. First, Fig. 5(a) represents the variation of the bias estimates on a uniformly sampled time series (Lorenz data) of length N ; note that such a time-series will have $N - \delta t$ pairs of points in each δt bin. This figure demonstrates how the different bias estimates behave as the absolute number of points used for the PDF estimate varies. In practice, no one characterizes a time series by the number of points separated by a fixed time. Therefore, second, Fig. 5(b) demonstrates the same calculation for the non-uniformly sampled patient one (top plot, Fig. 1). Here again, we do not alter the sampling rate which is already irregular, but merely take the first 20, 100, 1, 000, and 4, 000 points in the time series. This figure demonstrates that the bias estimate methods apply well on real, irregularly time series. Finally, because statistical estimates are made with concrete statistical estimators, Fig. 5(c) demonstrates how the bias estimation methods compare for histogram and KDE PDF estimators.

1. Summary of conclusions

The broad observations relevant to the KDE-based TDMI bias estimate comparisons include: (a) uniform-mixed bias estimate is usually the lower bound or nearly the lower bound on bias estimates; (b) the Gaussian-mixed bias estimate converges to the uniform-mixed bias estimate in the limit of short data strings and is often an upper bound on all bias estimates for long data strings; (c) the Gaussian-mixed and uniform-mixed bias estimates have the least variance over data string lengths with Gaussian-mixed bias estimates being an upper bound on the uniform-mixed bias estimates; (d) the AR-mixed bias estimates converge to the uniform-mixed bias estimates in the limit of short data strings and often in the limit of a long data strings where the AR-mixed weak bias estimate is always an upper bound on the AR-mixed strong bias estimate; (e) the random-permutation bias estimate progresses from a lower bound bias estimate for long data strings, usually coinciding with the uniform-mixed bias estimate, to an upper bound bias estimate on short data strings; (f)

the fixed point bias estimates in all cases roughly obey the $\frac{1}{N}$ dependence on the number of points shown analytically for the histogram estimator; and (g) all of the fixed point bias estimation techniques yield estimates on the same order of magnitude.

VI. SUMMARY AND DISCUSSION

Robust and consistent with previous bias estimate techniques

We claim that the *fixed-point bias estimation method* for estimating the TDMI bias is accurate, easy to use, robust with respect to differing methodologies, and works for a variety of estimators. In particular, all the time-dependency removal schemes we employ reproduce the $1/N$ bias dependency previously known for histogram estimators and observed in the KDE estimators. Moreover, all the bias estimate schemes we employed are easy to use and compute as fast as the standard TDMI computation. Finally, because the bias estimation technique relies on re-ordering or replacing the data *before the PDF estimates are carried out*, this estimation scheme will work for all PDF estimators.

Contrasting the bias estimate methods

Given data strings that have a reasonable density and length, the random permutation bias estimate will likely yield both the best estimate of the bias and the fastest implementation of a bias estimate. Nevertheless, for short data strings, the random-permutation bias estimate is likely to overestimate the bias. This overestimate occurs primarily because randomly permuting a few time-correlated points will, on average, force an average maximum time separation of one-third the length of time represented in the data set. For short data-sets, this can imply a significant amount of time-based correlation. In the context of short data strings, the uniform-mixed bias estimates do work reasonably well. Moreover, in all data-string length contexts, using more sophisticated means of fabricating data such as accept-reject methods for generating distributions do not, in general, outperform either the uniform-mixed or Gaussian-mixed, or random permutation bias estimates.

Data-based constraints on the TDMI calculation

For data sets with fewer than 50 points, very little TDMI related information can be gained. However, for data sets with as few as 100 points, very often qualitative time-based correlation information such as a simple decay in correlation can be determined. As the number of points increase from 100 up to 1,000 and beyond, the TDMI and TDMI bias calculations continue to improve. For the data sets we considered in this paper, beyond 1,000 points all of the qualitative and some quantitative conclusions drawn from the TDMI signal did not change when more data points were added. Nevertheless, these statements depend on the density of the data in time, the time-resolution desired, and the uniformity of the measurements in time. Finally, when there is concern about the presence of small sample size effects, a comparison (via the difference) between histogram and KDE-based estimates of the TDMI fixed point bias can be used to detect and quantify the existence of the small sample size effects.

DJA and GH would like to acknowledge the Columbia University Department of Biomedical Informatics data-mining group for helpful discussions; D. Varn and J. Dias for a careful reading of the manuscript; and financial support provided by NLM grant RO1 LM06910, an award from Microsoft Research for the Phenotypic Pipeline for Genome-wide Association Studies, and a grant from The Smart Family Foundation.

References

- [1]. Kantz, H.; Schreiber, T. Nonlinear Time Series Analysis. 2nd edition. Cambridge University Press; 2003.
- [2]. Sprott, JC. Chaos and Time-series Analysis. Oxford University Press; 2003.
- [3]. Albers DJ, Hripcsak G. Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations. CHAOS. 2012; 22:013111. [PubMed: 22462987]

- [4]. Friedman C, Wong A, Blumenthal D. Achieving a nationwide learning health system. *Sci. Transl Med.* 2010; 2:57.
- [5]. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med.* 2010; 363:501–504. [PubMed: 20647183]
- [6]. Albers DJ, Hripcsak G. Using population scale EHR data to understand and test human physiological dynamics. 2011; arXiv:1110.3317.
- [7]. Cover, TM.; Thomas, JA. *Elements of information theory.* Wiley; 1991.
- [8]. Shannon, CE.; Weaver, W. *The mathematical theory of communication.* University of Illinois Press; 1963.
- [9]. Pompe B. Measuring statistical dependencies in a time-series. *J. Stat. Phys.* 1993; 73:587–610.
- [10]. Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A.* 1986; 33:1134–1140. [PubMed: 9896728]
- [11]. Kraskov A, Stgbauer H, Grassberger P. Estimating mutual information. *Phys. Rev. E.* 2004
- [12]. Moon Y-I, Rajagopalan B, Lall U. Estimation of mutual information using kernel density estimators. *Phys. Rev. E.* 1995; 52:2318–2321.
- [13]. Gray; Moore. *Joint Stat. Meeting.* 2003. Very fast multivariate kernel density estimation using via computational geometry.
- [14]. Basharin GP. On a statistical estimate for entropy of a sequences of independent random variables. *Theory Prop. App.* 1959; 4:333–338.
- [15]. Roulston MS. Estimating the errors on measured entropy and mutual information. *Physica D.* 1999; 125:285–294.
- [16]. Grassberger P. Finite sample corrections to entropy and dimension estimates. *Phys. Lett. A.* 1988; 128:369–373.
- [17]. Lorenz EN. Deterministic nonperiodic flow. *J. Atmosph. Sci.* 1963; 20:130–141.

Time-delayed mutual information for irregularly sampled time-series.

Estimation bias for the time-delayed mutual information calculation.

Fast, simple, PDF estimator independent, time-delayed mutual information bias estimate.

Quantification of data-set-size limits of the time-delayed mutual calculation.

FIG. 1. Time series for three well sampled patients (from top to bottom) with ~ 2 , ~ 8.5 and ~ 6.5 year records respectively. Note that the missing measurements have been interpolated by straight lines. The x-axis units are in hours; the y-axis units are in milligrams per decilitre (mg/dL).

FIG. 2.

Time-delay mutual information estimates and bias for the Lorenz equation data strings of various lengths.

FIG. 3.

Time-delay mutual information and bias for delays of 0 to 120 hours (5 days) for three well sampled patients. Notices: the bias has low variance and is below the raw TDMI curve; the diurnal peaks in predictability; the variation between heights of the respective patients' TDMI peaks.

FIG. 4.

Time-delay mutual information estimates and bias for patient one's glucose time series of various lengths.

(a) Mean and standard deviation for each fixed point bias estimate method for Lorenz equation data.

(b) Mean and standard deviation for each fixed point bias estimate method for glucose data.

(c) Mean and standard deviation for each fixed point bias estimate method for glucose data for *histogram* and *KDE* estimators; note the convergence of the methods as the number of points increases.

FIG. 5.

Comparing the bias estimation methods via both the mean and standard deviation of each bias estimate method (Figs. 4(b) and 4(c)) and the estimator type (Fig. 4(d)) as a function of the number of points in the glucose data set.