Correlates of Depression in Bipolar Disorder

Paul J. Moore*, Max A. Little, Patrick E. McSharry, Guy M Goodwin, John R Geddes

***Abstract*—We analyze time series from 100 patients with bipolar disorder for correlates of depression symptoms. Since the sampling interval is non-uniform we quantify the extent of missing and irregular data using new measures of *compliance* and *continuity*. We find that uniformity of response is negatively correlated with the standard deviation of sleep ratings ($\rho = -0.26, p = 0.01$). To investigate the correlation structure of the time series themselves, we apply the Edelson-Krolik method for correlation estimation. We examine the correlation between depression symptoms for a subset of patients and find that *sleep* and *appetite/weight* show a lower average correlation than other symptoms. Using surrogate time series as a reference data set, we find no evidence of general correlations between time series, though we note a possible loss of information from sparse sampling. Finally, we suggest that some of the methods for handling non-uniform sampling might be applied more generally in health telemonitoring applications.**

***Keywords*—Bipolar disorder, Mood variability, Time series analysis, Public healthcare, Psychiatry,**

I. Introduction

HEALTH telemonitoring can benefit both patients and healthcare providers. A systematic review by Polisena *et al.* [1] found that home telehealth saved costs in 20 out of 22 studies, though it did note the poor quality of most of the economic evaluations. Another review by Paré *et al.* [2] examined 65 empirical studies of telemonitoring over four types of chronic illnesses: pulmonary conditions, diabetes, hypertension, and cardiovascular diseases. They drew no conclusion about economic viability but only because this was the subject of few studies most of which had no detailed analysis. However, they suggested that telemonitoring might have a positive effect on the patients' condition and that this would be a promising avenue for research. A more recent *BMJ* review [3] found evidence of fewer hospital admissions and lower mortality among patients allocated to receive telehealth interventions, though again there was no evidence of cost savings. However there are other benefits from both the patient's and clinician's point of view. The patients are monitored in their own environment, avoiding 'white coat syndrome' and they may have the freedom to manage their own reporting.

Most obvious from the researcher's point of view is the automated acquisition of data for analysis, sampled more often than an outpatient appointment would allow. Here, though, the freedom afforded to the patient has a potential disadvantage for time series analysis. If data can be returned at any time then the analyst cannot assume a regular reporting interval. Since most time series methods require uniform sampling, a common approach is to interpolate the data as a preprocessing step. In this study we apply methods that may be used directly on non-uniform data and introduce two new measures for quantifying non-uniformity. The structure of the paper is as follows. In §II we introduce time series analysis and the Edelson-Krolik method for estimating correlation. In §III we describe measures for quantifying non-uniformity in time series and in §IV show their application to telemonitored data. §V describes several different applications of the Edelson-Krolik correlation and correlation between time series using surrogate data. Finally, §VI summarizes the findings of the study.

II. Time series

Time series analysis involves the description, explanation and prediction of observations taken sequentially in time [4]. Description implies the use of numerical and graphical descriptive statistics such

P. J. Moore (**Corresponding author*) is with the Oxford Centre for Industrial and Applied Mathematics (OCIAM), Mathematical Institute, University of Oxford.

M. A. Little is with the MIT Media Lab and Aston University, Birmingham UK.

P. E. McSharry is with the Smith School of Enterprise and the Environment, University of Oxford.

G. M. Goodwin is with the Department of Psychiatry, University of Oxford.

J. R. Geddes is with the Department of Psychiatry, University of Oxford.

as time plots and the correlogram. Correlograms can reveal *seasonality*, which is the tendency to repeat a pattern of a certain periodicity such as a yearly cycle and *trend* or long-term variation up or down. Whereas description provides information about a given time series, inference induces a general form based on a finite number of observations. An example is time series regression which attempts to model an underlying relationship between dependent variables and time. Regression is often applied in the context of time series prediction because of its many practical applications. Linear approaches are popular because they are readily interpretable and convenient [5]. Stationary, linear time-invariant (LTI) Gaussian systems introduce several symmetries that have many conveniences including statistical stability, sufficiency of 1st and 2nd-order moments, and convex and analytic inference procedures [6]. Nonlinear models can represent regime switching behaviour and parsimonious nonlinear models have been shown to outperform linear methods in economic forecasting [7].

Correlation estimation

The autocorrelation function is an important measure of serial dependence in a time series, and is defined for a stationary random process $Y(t)$ as

$$\rho(s) = \frac{\gamma(s)}{\gamma(0)} \tag{1}$$

where s is the time lag and $\gamma(s)$ is the *autocovariance function* defined as the covariance between $Y(t)$ and $Y(t-s)$. An informative way of representing the serial dependence in a time series is by a graph of autocorrelation coefficients $\rho(k)$ against the integer lag k. This sequence represents a sample autocorrelation function (a.c.f.) and is called the *correlogram* [8]. Since natural time series often have missing or irregular data, it is often the applied sciences that have derived methods for their analysis. In astrophysics Edelson and Krolik [9] derived the discrete correlation function (DCF) for correlation estimation in non-uniform time series. It is defined for two discrete, centered time series a_i and b_j, first as a set of unbinned discrete correlation values

$$UDCF_{ij} = \frac{a_i b_j}{\sqrt{(\sigma_a^2 - e_a^2)(\sigma_b^2 - e_b^2)}} \tag{2}$$

for a measured pair of observations (a_i, b_j) whose time difference is Δt_{ij}. Here a_i and b_j are a concise notation for $a(t_i)$ and $b(t_j)$ respectively, σ_a, σ_b are the respective standard deviations and e_a, e_b are estimates of the measurement noise in each time series. The discrete correlation function is derived by averaging the set of M unbinned values

$$DCF(\tau) = \frac{1}{M} \sum_{|\Delta t_{ij} - \tau| < \frac{\Delta\tau}{2}} UDCF_{ij} \tag{3}$$

where τ is the bin centre and $\Delta\tau$ is the bin width. The standard error is given by

$$\sigma_{DCF}(\tau) = \frac{1}{M''} \left(\sum (UDCF_{ij} - DCF(\tau))^2 \right)^{1/2} \tag{4}$$

recalling that $UDCF_{ij}$ is a set and $DCF(\tau)$ is a scalar for given τ. The summation is over $|\Delta t_{ij} - \tau| < \frac{\Delta\tau}{2}$ as before and the normalising constant M'' is equal to $((M-1)(M'-1))^2$ with M' the number of unique measurement times for the series a_i.

The Edelson-Krolik method is closely related to the *variogram*, an approach that is well known in geostatistics where it is used to model spatial correlations [10]. It was until recently rarely mentioned in texts on time series or in the statistical literature as a whole [11] with the exception of Chatfield [4] and Diggle [8], who defines the variogram as follows:

$$V(k) = \frac{1}{2}\mathbb{E}[\{Y(t) - Y(t-k)\}^2] \tag{5}$$

$$= \gamma(0)\,(1 - \rho(k)) \tag{6}$$

where terms are defined as before. A plot of the quantities $v_{ij} = \frac{1}{2}\{y(t_i) - y(t_j)\}^2$ for all delays $k_{ij} = t_i - t_j$ is called the sample variogram. As with the DCF, random scatter in the plot may arise from small sample sizes used in calculating v_{ij}. This scatter can be reduced by averaging v_{ij} over binned time values to give $\bar{v}(k)$.

The binned variogram and discrete correlation function are examples of a *slotting* approach that uses a rectangular kernel to bin pairs of observations. They belong to one of four categories identified by Broerson *et al.* [12] for handling non-uniform data. The other categories are direct transform approaches, such as the Lomb-Scargle (LS) periodogram [13], model-based estimators (which presuppose a knowledge of the time series dynamics) and resampling through interpolation. The Lomb-Scargle approach, kernel methods (though

not slotting) and linear interpolation are compared in [14]. Since the data analyzed in this study has high relative noise and large gaps in the time indexes, we apply the Edelson-Krolik slotting approach. It provides a sample correlogram directly and avoids the assumptions necessary for interpolation or model-based estimators.

III. Measures of non-uniformity

We next introduce two measures for quantifying missing and non-uniform responses in time series. The first, which we call *compliance* measures the proportion of real observations in a time series which contains imputed values. The second measure, called *continuity* quantifies the sampling regularity among those real observations. Both measures are easily derived from a uniformly sampled series with missing data, but here we start from an irregular series and assume that a response is valid for an interval rather than a single point in time. This condition would apply, for example, to the answer from questionnaire where the relevant interval is the week prior to the response. We begin by considering the process of resampling the time series into a homogenized equivalent with uniform intervals.

Compliance

Figure 1 illustrates the resampling process assuming that sampling is approximately once per week and that responses are valid for the previous week. The optimal weekday w for the resampled time series is chosen to minimise the total deviation of the original responses from their corresponding resampled position on the X-axis or 'comb' of weekdays. The deviation in this case is the elapsed time to the first response within seven days.

The comb is then populated from the original series as follows. Starting from weekday w at the start, or the last instance of w before the start, of the time series we record any response within seven days. We repeat the search from weekday w in the following week and continue until the last response of the time series is reached. If no response is found within seven days, a missing value is imputed by random selection from the previous four responses. The imputed value itself is chosen for the purposes of illustration and does not affect the non-uniformity measures.

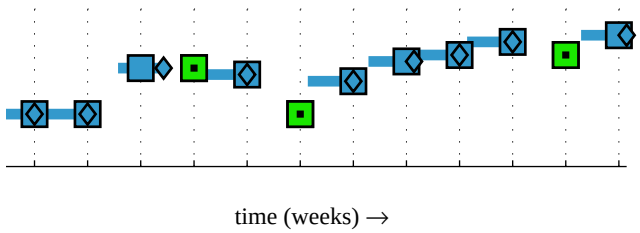


Fig. 1. Illustration of resampling. Diamond markers represent the original, non-uniform time series and the horizontal lines to the left of each marker show the period over which the response is valid. Square markers represent the resampled series and those with a square central dot are imputed values. The X-axis or 'comb' shows the optimal weekday which when aligned with the original series gives the minimum total distance (deviation) of the sample time from the response time.

Figure 2 shows the effect of resampling on two example series. Most of the original responses are not shifted while some are moved to an earlier time point and where this cannot be accomplished, an imputation is made.

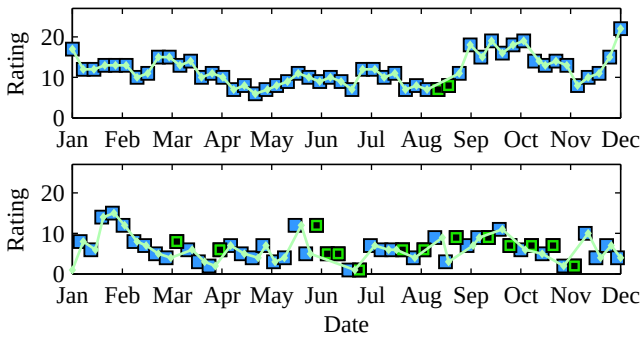


Fig. 2. Effect of resampling on high and low compliance time series. The original responses are denoted by small diamond markers and the resampled series by the larger square border. Imputed values are shown with a central square dot. The top plot represents an approximately uniform original time series in which resampling preserves the time stamps of the original responses: most diamond markers are centered in the squares. The lower plot illustrates a non-uniform series where many responses are late and some are missing. The late responses are shown by a diamond marker located to the right of center of the square border.

We define compliance as the proportion of non-imputed values in the resampled time series. Imputations occur when a response is later than the sample period τ which in this application is equal to 7 days. Formally,

$$C_m = \frac{1}{N}\sum_{k=0}^{N-1}\Theta\left[\sum_{i=1}^{N'}\mathbf{1}[k\tau \le t_i < (k+1)\tau]\right] \quad (7)$$

where C_m is compliance, τ is the uniform sample period and t_i is the i^{th} element of the time vector for the original series, which has N' points. N is the number of points in the resampled series and is equal to the number of weeks spanned by the original time series, allowing for the period of validity. The function Θ is equal to 0 if its argument is 0 otherwise it is equal to 1, and the indicator operator $\mathbf{1}$ has value 1 for a boolean argument of `true` and 0 for `false`. The value of C_m lies between 0 and 1.

As long as the original series covers all the new sample time points, there will be no imputations and the compliance is 100%. For example if responses are returned more often than every week, a uniform series may be derived by discarding some responses and without loss of compliance. A non-uniform series may also exhibit full compliance as long as no response is more than six (more generally, $\tau - 1$) days late. However, longer gaps result in an imputed value being added to the uniform series and compliance being reduced. The measure thus penalizes missing data but not additions or late returns.

Continuity

A low compliance implies that there is a large proportion of imputed points in the resampled series but gives no information about their distribution throughout the observed responses. A second measure which we call *continuity* measures the connectedness of non-imputed responses in the resampled time series. To develop the measure, we examine the sequence of points in the resampled series and label them with a state indicator of P for imputed and R for not imputed. The number of sequential state changes R $\rightarrow$ P is a count of the discontinuity and we use the ratio of this count to $N_r - 1$, where N_r is the number of R states. A simple example is the sequence RRR PPP R PPP R. Here there are 2 sequential changes of state from R to P out of a total of five R states giving a continuity of $2/4$. The sequence RRRRR then has a continuity of 1, and the sequence RPRPR has a continuity of 0. In general we then have

$$C_t = 1 - \frac{1}{N_r - 1}\left(\sum_{k=1}^{N-1} \mathbf{1}[\,(w_k, w_{k+1}) = (\mathsf{R}, \mathsf{P})]\right) \quad (8)$$

where C_t is continuity, N is the length of the resampled series and $w_k \in \{\mathsf{R},\mathsf{P}\}$ is the state of the k^{th} data point. The minimum possible continuity occurs when the P states are distributed singly throughout the time series. In this case,

$$C_{t\,(min)} = 1 - \frac{N_p}{N - N_p - 1} \quad (9)$$

$$\approx \begin{cases} \frac{2C_m - 1}{C_m} & \text{if } C_m \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

for $N \gg 1$ where N_p is the number of P states. It can be seen from (10) that as the compliance approaches 1, the minimum possible continuity approaches the compliance.

So compliance is the proportion of non-imputed responses and continuity is the proportion of correct intervals among them. Continuity summarizes the interval distribution using the probability density located only at the desired interval. The location of the remaining mass, corresponding to the distribution shape, does not influence its value.

This approach gives an advantage over standard dispersion measures (of either the raw or the homogenized series) because all intervals longer than the sampling period are classed together. Long gaps in the time series, when the patient fails to respond for a period, do not greatly influence the continuity value, although they are reflected in the compliance. The property is also relevant to the autocorrelation calculation because time series with high continuity can be treated as uniform for this purpose. Both compliance and continuity can be useful in both selection of near-uniform series for the application of standard methods and for exploring non-uniformity as an informative property in itself.

IV. Application of measures

We apply the measures to time series from 153 patients with bipolar disorder who were monitored between 2006 and 2011. Data were collected as part of the OXTEXT programme funded by the National Institute for Health Research which investigates the potential benefits of self monitoring of mood for people with bipolar disorder. The sub-sample of participants in this study was selected from the OXTEXT cohort, and includes those patients who had used the mood monitoring prior to recruitment

into OXTEXT and who had given consent for the use of anonymised retrospective data for exploratory time series analysis.

The mood data is returned approximately each week and comprises answers to standard self-rating questionnaires for both depression and mania. The rating scale used for depression is the *Quick Inventory of Depressive Symptomatology - Self Report (QIDS-SR$_{16}$)* [15] which has 16 questions covering nine symptom domains for depression (Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision) [16]. This self-rated instrument has highly acceptable psychometric properties, including high validity [17]. Each domain can contribute up to 3 points giving a total possible score of 27 on the scale. The severity of mania is quantified using the *Altman Self-Rating Mania Scale (ASRM)* [18] which has 5 questions, each of which can contribute up to 4 points, giving a total possible score of 20.

A. Data selection

The initial set of 153 patients is first cleaned by removing repeated response values, that is those which share the same time stamp. These repeats arise when a patient resubmits a rating score either by mistake or in order to correct an earlier response. Assuming that earlier values are being corrected, we remove repeated responses by taking the most recent in the sequence. We then create Set A (n=93) with members whose time series have at least 25 data points, or approximately six months duration. Figure 3 illustrates the data selection process.

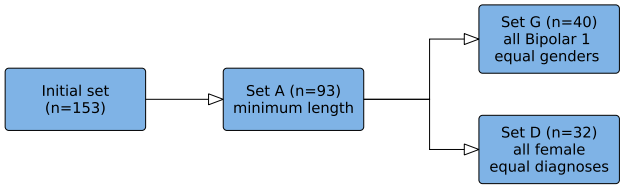


Fig. 3. Flow chart for data selection. From the initial data set, Set A (n=93) of time series having a minimum length of 25 data points is selected. Two further subsets are then selected from Set A. Set G (n=40) has equal numbers of each gender, all with a diagnosis of Bipolar I disorder. Set D (n=32) has equal numbers of patients having Bipolar I and Bipolar II diagnoses, all of whom are female. The selection algorithm matches patients by time series length. Where no patient of matching length can be found, the range is progressively widened until one or more matches is found.

Two further subsets are then created from Set A, one having equal numbers of male and female patients and a second with equal number of Bipolar I (BPI) and Bipolar II (BPII) diagnoses. The first subset is labelled as Set G (n=40) and contains patients all of whom have a diagnosis of BPI disorder. It is created by selecting all the patients with BPI from Set A and removing the female patient with the shortest time series length. The second subset, labelled Set D (n=32), has equal numbers of patients diagnosed with BPI and BPII disorder, all of whom are female. Set D is created by retaining the 16 female BPII patients from Set A and selecting 16 BPI female patients to match for time series length. The selection algorithm attempts to match the length for each individual patient by progressively widening the search range until a suitable match is found. Descriptive statistics of the subsets are given in the electronic supplementary material §I.

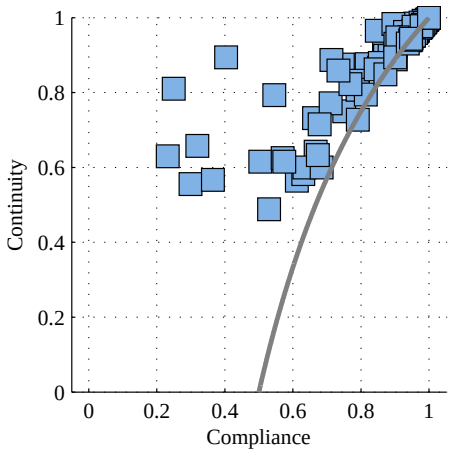


Fig. 4. Scatter plot of continuity against compliance for patients having at least 25 points in their time series (n=93). The approximate minimum continuity limit $2-C_m^{-1}$ from (10) is shown as a line. There are some short time series which have continuity values slightly lower than this limit. As compliance tends towards 1, the minimum possible continuity tends towards compliance. Those series in the upper left of the plot with high continuity and low compliance have large gaps where there is a long sequence of imputed points.

Non-uniformity

Using the subset of data labelled Set A, we derive the compliance and continuity measures for each patient. A scatter plot is shown in Figure 4. From (10) we see that the minimum continuity tends towards the compliance as the compliance approaches 1. For lower compliance where there is a higher proportion of imputations, the continuity is more dispersed.

For the next analysis, we assume that any text message latency of more than a day is small in comparison with the patient's delay in responding to the prompt. The ideal would be to find daily network latency patterns and show these are not significant, but such data is hard to obtain and would need to relate to all network providers and any interactions between them. However, some evidence can be found from the data itself.

A prior expectation would be that a significant number of patients would not respond immediately, although the delay cannot be distinguished from any network latency. However, if we assume that there is significant network latency, we would predict that no patients could return messages consistently every week. An analysis of the 2009 data shows that two patients return responses with near perfect regularity: one with 98% intervals being exactly of one week, the second with 96%. These two cases do not show significant network delay either in message sending or its reception by the system. This evidence is not definitive because the individuals may be responding at will rather than in response to the prompt, and may have highly reliable network providers along with reliable reception by the system. A quantitative latency test would provide a firmer basis for this assumption and may be a useful study, especially if ratings are to be sampled more frequently than every week.

Demographic and mood data

We examine the correlation between continuity and both demographic and mood data over the set of patients using Set G (n=40) which has equal numbers of male and female patients and Set D (n=32) with equal numbers of Bipolar I (BPI) and Bipolar II (BPII) diagnoses. No pattern emerges in either case, and a two-sample Kolmogorov-Smirnov test does not distinguish the distribution of male vs female or BPI vs. BPII non-uniformity measures. Further details can be found in the electronic supplementary material §IV.

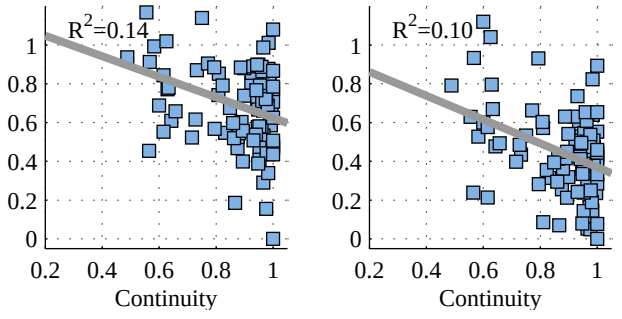
Next we look for correlates of non-uniformity with mood. There are 9 variables for depression corresponding to symptoms of sleep, appetite etc., and 5 variables for mania which we summarize for each patient by mean, standard deviation and mean absolute difference. We take the rank correlation for each symptom with continuity over the set of 93 patients in Set A. The results are shown in Table I. No correlations were found between mean symptom levels and continuity. For the dispersion statistics only sleep in the depression questionnaire was found to have a correlation significant at the 1% level.

		Variability measure	
Domain	*Mean*	*Std. dev.*	*Mean abs. diff.*
Sleep	+0.14 (0.18)	-0.26 (0.01)	-0.25 (0.02)
Feeling sad	-0.13 (0.21)	-0.17 (0.10)	-0.09 (0.39)
Appetite/wt	-0.06 (0.59)	-0.04 (0.75)	-0.02 (0.88)
Concentration	-0.12 (0.24)	+0.01 (0.94)	-0.00 (0.96)
Self-view	-0.13 (0.22)	-0.15 (0.14)	-0.13 (0.23)
Death/suicide	-0.11 (0.27)	-0.15 (0.16)	-0.19 (0.06)
Gen. interest	-0.11 (0.29)	-0.16 (0.12)	-0.19 (0.07)
Energy level	-0.08 (0.43)	-0.14 (0.19)	-0.05 (0.61)
Slowed down	-0.09 (0.39)	-0.08 (0.45)	-0.01 (0.91)

TABLE I
RANK CORRELATION (p-VALUES) BETWEEN DEPRESSION SYMPTOMS AND CONTINUITY FOR SET A.

Variability of sleep correlates negatively with continuity when measured by standard deviation ($\rho = -0.26$, $p = 0.01$) and mean absolute difference between sequential values ($\rho = -0.25$, $p = 0.02$). A similar result was found when using compliance as the non-uniformity measure. The scatter plots for both statistics are shown in Figure 5.

We note that there will be a sampling distribution for both the mean and variability measures arising from the limited sample sizes, which would manifest in Figure 5 as a range for each point.



(a) Standard deviation (b) Mean absolute difference

Fig. 5. Scatter plots for sleep against continuity. In (a) the standard deviation of all the resampled sleep values (excluding imputed points) for each patient are plotted against the continuity score for that patient. In (b) the mean of the absolute difference between sequential resampled values, again ignoring imputed points, is used. For both cases, patients with lower continuity show a higher variability in sleep responses on average. The linear least-squares fit is marked as a line.

For some symptoms, any correlation with non-uniformity might be hidden by this effect. However, since the same sampling limits apply to all symptoms we can distinguish sleep variability as having a relatively strong association with non-uniformity of response.

The relation of non-uniformity of response with sleep variability is an important finding from this analysis. The association is also interesting if response uniformity is taken as an indicator of general functioning. We would expect that delays in responding are caused by holidays, work commitments, physical illness, forgetting to reply, a low priority for replying or chaotic behaviour. Psychological factors may have an influence and several of the symptoms explicitly measured on the QIDS scale are relevant, in particular severe lassitude or lack of energy, a lack of interest, poor concentration, and thoughts of death/suicide. As pointed out, it is quite possible that correlations with these variables exist but that they are below the noise threshold. The relatively stronger effect of sleep points to a number of possibilities. Firstly, a strong association between sleep and mental illness is well established, if not well understood [19]. So one possibility is that sleep is simply the strongest indicator of an underlying disorder which causes irregularity through the behavioural issues listed above. The causation might be more direct, for example sleep causing problems with memory or other functioning, leading to lost or delayed ratings. However, it is a high variability of sleep ratings rather than a high mean rating that predicts non-uniformity of response. It may be that there is some adaptation to poor sleep, whereas inconsistent sleep leads to inconsistent behaviour. The data is too noisy and does not provide a strong enough effect to distinguish these scenarios.

V. Application of Methods

We now apply the Edelson-Krolik method to calculate autocorrelation and correlation using the time series for depression. We first examine evidence of seasonality from the correlogram for individual patients. We then look at the correlation between symptoms of depression and finally apply a surrogate data method to detect correlations among the set of time series themselves.

Seasonality

We examine the autocorrelation function of the depression time series using the Edelson-Krolik method to determine the autocorrelation at successive lags. Four examples of correlograms are shown in Figure 6 in comparison with a standard correlogram (lighter line) which has not been adjusted for non-uniform response times. The third plot from the top shows a yearly seasonality for both the Edelson-Krolik method and the unadjusted correlogram with the latter having a peak correlation at less than 50 weeks and less seasonal variation.

Figure 7 is the Lomb-Scargle periodogram corresponding to this time plot. It shows a peak of spectral power at 370 days indicating a yearly seasonality. The depression time series do not in general show clear evidence of yearly periodicity, though some have a peak at or near this period. Most exhibit a rapid decrease in correlation with lag and some show evidence of a trend, indicated by the correlogram not tending to zero as the lag increases.

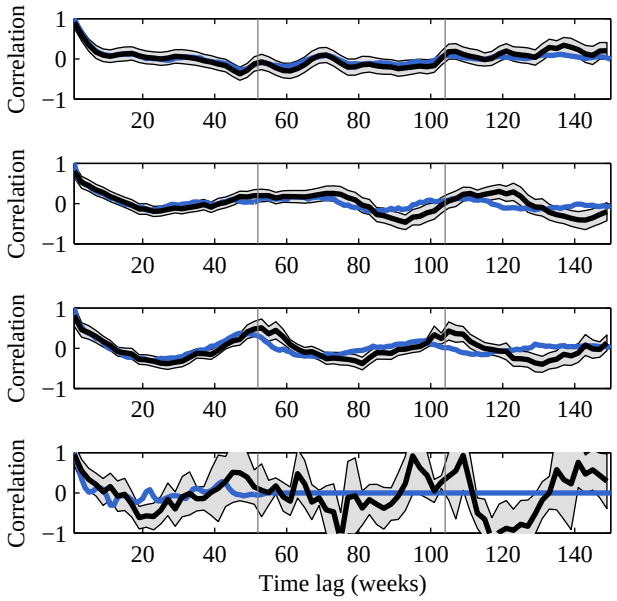


Fig. 6. Correlograms for the depression time series from four patients. In each plot, the dark line is the correlogram estimated using the Edelson-Krolik method with a bin width of 2 weeks and showing two standard errors each side as a filled region. The lighter line is the autocorrelation calculated under the assumption of a uniform series. Imputed points are not used in either calculation. In the time plot third from the top there is clear evidence of yearly seasonality of depression. The continuity values for the time series are, from top to bottom: 0.99, 0.92, 0.87 and 0.30. Vertical lines are year markers corresponding to 52 and 104 weeks. Note that correlograms are defined only at integer lags or bin centres but are shown as continuous lines for clarity.

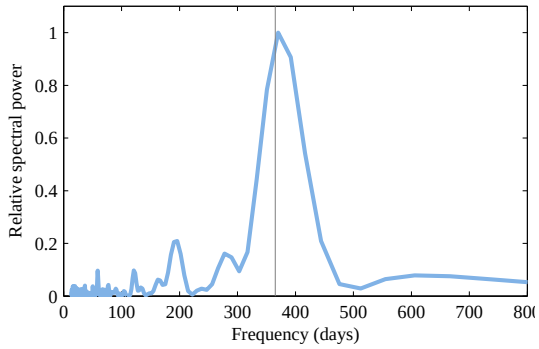


Fig. 7. Lomb periodogram for a patient exhibiting seasonality of depression. The corresponding correlogram in Fig. 6 is third from the top. The spectral power is normalised by the peak power and the periodicity of 365 days is marked as a vertical line. The peak is at a period of 370 days and a second much smaller peak occurs at 196 days. In general the depression time series do not show such clear evidence of yearly periodicity, although some patients have a peak at or near this period.

Correlation between depression symptoms

The correlation between depression symptoms is examined for patients who have at least 100 data points in their homogenized time series. The first 100 responses are taken, the imputed values removed and the means subtracted from the individual domain scores. Correlation between domains is then calculated using the Edelson-Krolik method (3) and the scores averaged over the set of patients. In order to provide a comparison between symptoms, only those patients with non-zero symptom series and positive correlations are selected. There are 6 patients showing some pairs of negative correlations but these did not show any common relationship. The subset of patients fulfilling these criteria is denoted Set E and its statistical properties are summarized in the electronic supplementary material §II, with further details about selection. The selection of Set E is illustrated in Figure 8.

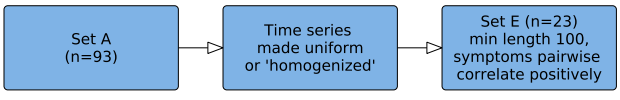


Fig. 8. Flow chart for data selection. From Set A (n=93), a homogenized set of time series is created and from this Set E (n = 23) is selected. It has at least 100 data points in the homogenized time series, and all the symptom time series for a patient have positive pairwise correlations.

A heat map showing the relationship between symptom domains is shown in Figure 9.

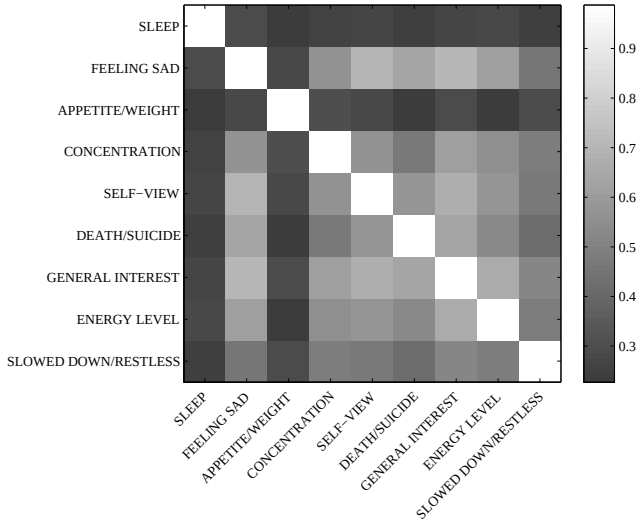


Fig. 9. Matrix of mean correlation between pairs of depression symptoms. For each patient in a set of 23, we find the correlation between pairs of symptoms and present the average over whole set. Only positive correlations greater than two standard errors from zero are used and patients with negative correlations or non-significant autocorrelations are excluded. The white diagonal represents the zero lag autocorrelations of individual domain time series.

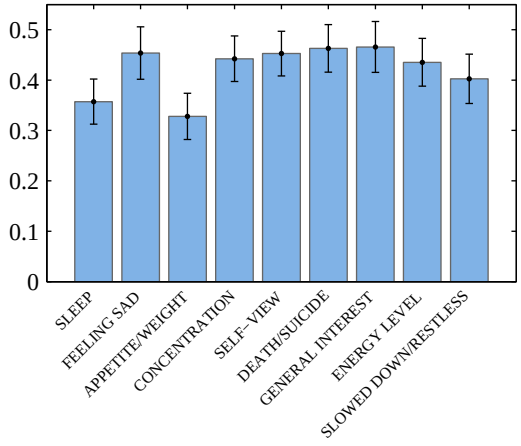


Fig. 10. Autocorrelation for symptom time series. The chart represents the mean first order autocorrelation of a set of 23 patients, with error bars showing the standard error. The symptoms *Sleep* and *Appetite/weight* have a lower autocorrelation than the rest implying a low relative correlation with symptoms which have a different autocorrelation structure.

On average, the symptom domains *Sleep* and *Appetite/weight* correlate less than other domains. By contrast *Feeling sad* correlates strongly with other domains while *Slowed down/restless* shows less correlation with others.

An analysis of the autocorrelation structure for symptom time series explains why the symptoms of *Sleep* and *Appetite/weight* tend to correlate less when paired with other domains. We take the 23

time series used above and find the autocorrelation at using the Edelson-Krolik method on the homogenized time series with imputed points removed. The results are shown in Figure 10. The symptoms *Sleep* and *Appetite/weight* have a lower autocorrelation than the other symptoms which explains their relatively low pairwise correlation in Figure 9. Although *Sleep* and *Appetite/weight* have a similar first order autocorrelation, Figure 9 shows that they do not themselves correlate as a pair, the reason being that their autocorrelation structure is somewhat different: the autocorrelation for *Sleep* remains higher than *Appetite/weight* as the lag increases. Autocorrelation coefficients up to a lag of 4 are shown in the electronic supplementary material §III.

We note that these two symptoms are the most amenable to objective measurement out of the nine symptoms in the QIDS rating scale and that *Slowed down/Restless*, which might also fall into this category also correlates less than the others. It may be that the other symptoms: *Feeling sad*, *Concentration, Self-view*, *Thoughts of death/suicide*, *Interest* and *Energy level* have a common factor which influences them more than it does the other three symptoms. This finding is similar to that in [20] which identified three factors in the IDS instrument: cognitive/mood, anxiety/arousal and sleep (or sleep/appetite for the self-rated instrument).

Time series correlation

In this section, we look for similar mood changes in patients by examining pairwise correlations between their time series of depression ratings. We take a set of 28 patients who have complete depression series during the years 2009 and 2010 which we denote as Set F.

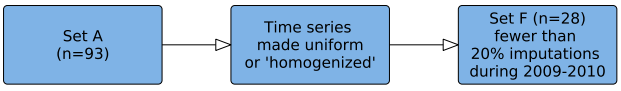


Fig. 11. Flow chart for data selection. From Set A (n=93), a homogenized set of time series is created and from this Set F (n = 28) is selected. It is comprised of time series which span the years 2009-2010 and have fewer than 20% of imputed points over that period.

The selection process is illustrated in Figure 11 and descriptive statistics are given in the electronic supplementary material §II. We create a reference set of surrogate time series by shuffling the time order of existing series while maintaining their mean, variance and autocorrelation function. The algorithm used for this process is described in [21] and is implemented using the *TISEAN* function `surrogates` [22]. The distribution of the pairwise correlations for both the original and surrogate data sets is shown in Figure 12.

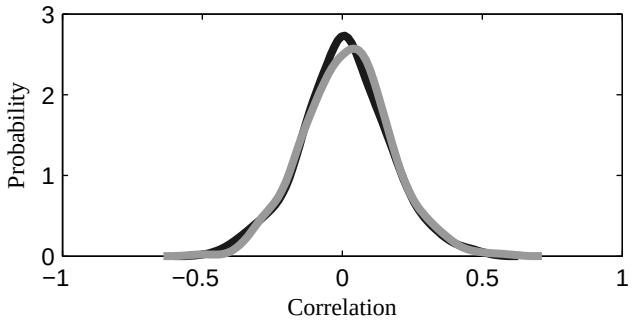


Fig. 12. Kernel density estimate of pairwise correlations between time series. The dark line is the density estimate for the original set of time series and the light line for the surrogate data. Each surrogate time series is derived from its original counterpart by taking the Fourier transform and randomizing the phases to obtain a time series with the same power spectrum. The method removes any correlation between pairs of time series that arises from a common source rather than by chance. The similarity of the distributions shows that in general there is no correlation present among pairs of the original time series.

The correlations between time series for original and surrogate data sets appear to have the same distribution and a two sample Kolmogorov-Smirnov test returns a value of p=0.53. Although external factors do not appear to have a strong influence on depression over the set of patients, this does not preclude the possibility that there may be strong environmental effects in individual cases.

VI. Conclusion

We have addressed the problem of describing and modelling time series with missing or irregularly spaced values. Two new measures for quantifying missing and non-uniform data were introduced and applied to a database of telemonitored mood data. The quantification of non-uniformity can be useful in 1) investigation of non-uniformity as correlate of other variables 2) selecting subsets of data where uniformity is a requirement 3) use as supplementary information for a clinician. We found that time series uniformity does not correlate with either gender or diagnostic subtype. However, variability of sleep correlates with continuity. This finding has

implications for selecting time series according to their uniformity since it may exclude patients with more variable sleep ratings.

The Edelson-Krolik method uses relative distances rather than fixed lags to determine time series correlation and so it is robust to non-uniform sampling intervals. We used the method to generate correlograms of depression ratings and showed that one patient exhibited mood with yearly seasonality. Most patients do not show evidence of seasonality, but rather a short term autocorrelation structure.

We examined correlations between depression symptoms and found that *sleep* and *appetite/weight* show a lower average correlation than other symptoms. We found evidence that the autocorrelation structure for these domains is different from that of the others. Finally, we examined correlations between patients' depression time series but found no evidence of correlation in general. We note that for some patients, the weekly sampling will be below the Nyquist frequency for depression, so information will be lost. A study identifying the range of frequencies in depression in bipolar would help in choosing an optimal sample rate, consistent with practical considerations. While all these findings relate to mood in bipolar disorder we hope that the methods applied may find more general application in health telemonitoring.

ACKNOWLEDGMENTS

The authors would like to acknowledge financial support from the John Fell Fund, Oxford University.

This article presents independent research funded by the NIHR under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0108-10087). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

[1] J. Polisena, D. Coyle, K. Coyle, S. McGill *et al.*, "Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations," *International journal of technology assessment in health care*, vol. 25, no. 3, pp. 339–349, 2009.

[2] G. Paré, M. Jaana, and C. Sicotte, "Systematic review of home telemonitoring for chronic diseases: the evidence base," *Journal of the American Medical Informatics Association*, vol. 14, no. 3, pp. 269–277, 2007.

[3] A. Steventon, M. Bardsley, J. Billings, J. Dixon, H. Doll, S. Hirani, M. Cartwright, L. Rixon, M. Knapp, C. Henderson *et al.*, "Effect of telehealth on use of secondary care and mortality: findings from the whole system demonstrator cluster randomised trial," *BMJ: British Medical Journal*, vol. 344, 2012.

[4] *The Analysis of Time Series: An Introduction*, 6th ed. Chapman & Hall/CRC, Jul. 2003.

[5] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, 6th ed. Springer Series in Statistics, 2009.

[6] M. Little, "Mathematical foundations of nonlinear, non-gaussian, and time-varying digital speech signal processing," *Advances in Nonlinear Speech Processing*, pp. 9–16, 2011.

[7] S. Arora, M. Little, and P. McSharry, "Nonlinear and nonparametric modeling approaches for probabilistic forecasting of the us gross national product," *Studies in Nonlinear Dynamics and Econometrics (in press)*, 2012.

[8] P. Diggle, *Time series: a biostatistical introduction*. Oxford University Press, 1990, vol. 5.

[9] R. Edelson and J. Krolik, "The discrete correlation function-a new method for analyzing unevenly sampled variability data," *The Astrophysical Journal*, vol. 333, pp. 646–659, 1988.

[10] N. Cressie, "Statistics for spatial data," *Terra Nova*, vol. 4, no. 5, pp. 613–617, 1992.

[11] J. Haslett, "On the sample variogram and the sample autocovariance for non-stationary time series," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 46, no. 4, pp. 475–484, 1997.

[12] P. Broersen, S. de Waele, and R. Bos, "The accuracy of time series analysis for laser doppler velocimetry," in *Proc. 10th Int. Symp. Laser Techn. Fluid Mech*, 2000.

[13] J. Scargle, "Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[14] K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, "Comparison of correlation analysis techniques for irregularly sampled time series," *Nonlinear Processes in Geophysics*, vol. 18, no. 3, pp. 389–404, 2011.

[15] A. J. Rush, T. Carmody, and P.-E. Reimitz, "The inventory of depressive symptomatology (ids): Clinician (ids-c) and self-report (ids-sr) ratings of depressive symptoms," *International Journal of Methods in Psychiatric Research*, vol. 9, no. 2, pp. 45–59, 2000.

[16] *Diagnostic and Statistical Manual of Mental Disorders: Text Revision, 4th edition.*, American Psychiatric Association, Washington, DC, 2000.

[17] A. Rush, "The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression." *Biological Psychiatry*, vol. 54, no. 5, 2003.

[18] E. G. Altman, D. Hedeker, J. L. Peterson, and J. M. Davis, "The altman self-rating mania scale," *Biological Psychiatry*, vol. 42, no. 10, pp. 948–955, Nov 1997.

[19] S. R. Pandi-Perumal and M. Kramer, *Sleep and mental illness*. Cambridge University Press, 2010.

[20] A. Rush, C. M. Gullion, M. R. Basco, R. B. Jarrett, M. H. Trivedi *et al.*, "The inventory of depressive symptomatology (ids): psychometric properties," *Psychological medicine*, vol. 26, no. 3, pp. 477–486, 1996.

[21] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 1st ed. Cambridge University Press, 2003.

[22] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The tisean package," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 9, no. 2, pp. 413–435, 1999.