Received Total Wideband Power Data Analysis

Multiscale wavelet analysis of RTWP data in a 3G network

John Garrigan	Martin Crane	Marija Bezbradica
ARC-SYM Research Group	ARC-SYM Research Group	ARC-SYM Research Group
Dublin City University	Dublin City University	Dublin City University
Glasnevin, Dublin 9, Ireland	Glasnevin, Dublin 9, Ireland	Glasnevin, Dublin 9, Ireland
john.garrigan4@mail.dcu.ie	martin.crane@dcu.ie	marija.bezbradica@dcu.ie

ABSTRACT

Received total wideband power (RTWP) data is a measurement of the wanted and unwanted power levels received by a 3G radio base station (RBS) and is a concise indicator of uplink network performance. Using a statistical physics approach, we aim to detect periods of unusual activity between cells by assessing a sample of RTWP measurement data from a live network. Using wavelet correlation and cross-correlation techniques we analyse multivariate non-stationary time series for statistical relationships at different time scales. We analyse the seasonal component of the dataset as well as examining the autocorrelation and partial autocorrelation methods. We then explore the Hurst exponent of the dataset and inspect the intraday correlations for patterns of events. Next, we examine the eigenvalue spectrum using different sized sliding windows. Finally, we compare approaches for assessing multiscale relationships among several variables using the wavelet multiple correlation and wavelet zero-lag cross-correlation on non-stationary RTWP time series data.

CCS CONCEPTS

• Networks~Wireless access points, base stations and infrastructure
• Networks~Network performance analysis

KEYWORDS

3G; uplink; maximum overlap discrete wavelet transform; multiscale analysis; multivariate time series; non-stationary time series; received total wideband power

ACM Reference Format:

1. Introduction

Wavelet techniques have been used extensively in a broad range of research areas e.g. engineering [1], medicine [2], fractals [3] and geophysics [4]. Robertson et al in [1], first used wavelets for power engineering to analyse electromagnetic transients from power system faults and switching. Since then there has been a significant increase in the application of wavelet transforms to power systems including power system protection, power quality and load forecasting. Biomedical engineering research has also used wavelet transforms, specifically in seizure prevention techniques and pre-surgical evaluations [2]. Identifying correlations between non-stationary signals is a common approach particularly in finance in measuring the relationship between two or more signals over time. In portfolio management, the maximum overlap discrete wavelet transform (MODWT) is used as part of an investment portfolio optimisation process [5]. The zero-lag cross-correlation matrix and the MODWT have also been used to analyse SenseCam images [6] to strengthen the wearers memory.

The typical observation of recorded RTWP levels in a 3G network is that levels vary during the day [7], with levels close to the noise floor in low traffic periods and levels rising during busy periods. Generally, noise levels for a cell in a network exhibit seasonality or periodic fluctuations every 24-hour cycle. 3G networks are noise limited systems, therefore increases to RTWP above normal operating levels could mean a loss of coverage for users at cell edge radio conditions and with an undesirable impact on network capacity.

When elevated RTWP levels are observed in a network, several factors can be responsible. Firstly, passive intermodulation (PIM)[1] within faulty hardware can result in intermodulation products in a cellular operator's uplink (UL) band; another possible cause is external interference; this can appear as spurious emissions from another party's transmitter; thirdly RTWP rises with user

---

[1] Passive Intermodulation occurs when two signals mix in a non-linear device such as a mechanical connector and generate a third frequency which falls within the operators own band resulting in interference.

traffic; large volumes of high-speed packet access (HSPA) traffic correlate with increased RTWP levels. Correlations and wavelet techniques have been used frequently in wired communication systems [8], [9], [10], [11] and have a rich history. In 3G networks, orthogonal spreading codes are combined with users' data packets to spread the data across the full 5MHz channel. At the radio base station receiver, the same spreading code sequence can extract the original data using signal correlation. In Fig. 1 below, a user's data is encoded using an orthogonal spreading code, only the intended recipient knows the code. This use of orthogonal codes allows concurrent use of the RF physical channel by multiple users.
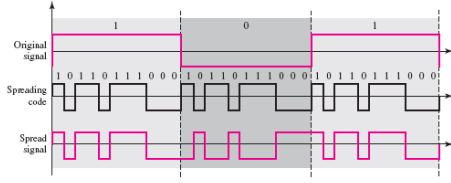

Figure 1: How direct sequence spreading codes works. As each user is given a unique spreading code, each user's uplink signal looks like noise to one another due to the orthogonality of the codes used.

Using wavelets, RTWP data are decomposed into their component scales in short time windows, enabling us to study the correlation at various scales. This gives us a more convenient way to establish overall multiple relationships between cells and to minimise the time to fault find such issues. Using such techniques, it would allow a network optimisation engineer to quickly evaluate whether the interference patterns seen in the RTWP reports are correlated with the noise profiles of other cells. This approach would also improve the turnaround time for fault detection as a large number of cells can be quantitively analysed together rather than a traditional approach of assessing each neighbouring cell individually. This paper is organised as follows: In Section 2 we review the methods used, Section 3 describes the RTWP dataset, meanwhile Section 4 details the results obtained and finally in Section 5 conclusions are provided.

2. Methods

While telecom networks have a rich store of data ready for interpretation, little research into RTWP datasets has taken place [12]. In this section we introduce some aspects of the datasets using typical time series analysis techniques. We calculate the zero-lag cross-correlation matrices of the multivariate raw time series data to characterise dynamical changes. From this we look at the eigenspectrum for noise level patterns at different time scales. Finally, we measure the overall relationships at different scales among observations in a multivariate random variable with multiple wavelet correlation/cross-correlation approaches.

2.1 Correlation Dynamics

Using the zero-lag cross-correlation matrix (henceforth correlation matrix), dynamical changes in non-stationary multivariate time series can be characterised. To analyse the impact of abnormal noise rises in 3G networks, a cell with a known external interference source was included as per the analysis. To see how this source impacts on geographically neighboured cells and if any lead or lag pattern exists between such cells, we used a correlation matrix consisting of 10 geographically clustered cells. Of these, one has the source and we analyse its impact on the other 9. The correlation matrix is calculated using a sliding window of size less than the number of cells. Similar windowing techniques have been looked at in OLS hedging models [13].

Given a time-series of RTWP measurements $R_i(t), i = 1, \dots, N$, the series within each window is normalised using $r_i(t) = [R_i(t) - \widehat{R_i(t)}]/\sigma_i$ where σ_i is the standard deviation of R_i and $\langle \dots \rangle$ denotes a time average over the period and is given by $\sigma_i = \sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$. The correlation matrix may be expressed in terms of $r_i(t)$ as follows: $C_{ij} = \langle r_i(t) r_j(t) \rangle$. C_{ij} has values $-1 \leq C_{ij} \leq 1$, where $C_{ij} = 1$ corresponds to perfect correlation, $C_{ij} = -1$ to perfect anti-correlation, and $C_{ij} = 0$, to uncorrelated pairs of cells. The correlation matrix can be expressed as $C = [RR^T]/T$ where T is the transpose of a matrix and R is an $N \times T$ matrix with elements r_{it} [14].

The eigenvalues λ_i and eigenvectors $\underline{v}_i$ of C come from the characteristic equation $C\underline{v}_i = \lambda_i \underline{v}_i$. The eigenvalues of C are ordered by size, such that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. Given that the sum of the matrix diagonal entries (the Trace, T_r)) remains constant under linear transformation [15], $\sum_i \lambda_i = T_r$ for C. Hence, if some eigenvalues increase then others must decrease, and vice versa, (Eigenvalue Repulsion) [16].

Two limiting cases [15],[17] exist for the distribution of the eigenvalues: (i) perfect correlation, $C_i \approx 1$, when the largest is maximised with value N, (all others being zero). (ii) when each time series compromises random numbers with average correlation $C_i \approx 0$ and the corresponding eigenvalues are distributed around 1, (where any deviation is due to spurious random correlations). For C_i between 0 and 1, the eigenvalues furthest away from λ_{max} can be much smaller. To investigate the dynamical changes in eigenvalue distribution we use sliding windows with eigenvalues normalised thus $\tilde{\lambda}_i(t) = [\lambda_i - \underline{\lambda}]/\sigma_\lambda$ where $\underline{\lambda}, \sigma_\lambda$ are mean, standard deviation of the eigenvalues from a subsection of the eigenspectrum of C. By normalising the eigenvalues, we can compare eigenvalues at both ends of the spectrum, though their magnitudes differ. To calculate $\underline{\lambda}$ and σ_λ above a low volatility part of the eigenspectrum is used to enhance the visibility of high periods (also the full time period can be used) [5].

2.1.1 Maximum Overlap Discrete Wavelet Transform.

The Maximum Overlap Discrete Wavelet Transform, (MODWT) [19], transforms a series into coefficients related to the variations over a set of scales. Like the DWT, the MODWT outputs a set of time-dependent wavelet and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j = 2^{j-1}$ for each decomposition level $j = 1, \dots, J_0$ [14]. However, unlike the DWT, the MODWT, has a high level of redundancy. The advantages of the MODWT over DWT are its non-orthogonality and ability to handle any sample size $N \neq 2^j$ [14]. With MODWT, a signal can be broken into J levels by applying J pairs of filters. The coefficients at the Jth level are found by applying a rescaled *father* wavelet:

$$\tilde{D}_{j,t} = \sum_{l=0}^{L_j - 1} \tilde{\varphi}_{j,l} f_{t-l} \qquad (1)$$

for all $t = \dots, -1, 0, 1, \dots$, where f is the function to be decomposed [14]. The rescaled *mother*, $\tilde{\varphi}_{j,t} = \frac{\varphi_{j,t}}{2^j}$, and father, $\tilde{\phi}_{j,t} = \frac{\varphi_{j,t}}{2^j}$, wavelets for the jth level are a set of scale-dependent localised differencing and averaging operators and can be seen as rescaled versions of the originals. The jth level equivalent filter coefficients have a width $L_j = (2^j - 1)(L - 1) + 1$, ($L$ is the width of the $j = 1$ base filter [14]). The filters for levels $j > 1$ are not explicitly constructed as the detail and scaling coefficients can be found, using an algorithm involving the $j = 1$ filters operating recurrently on the jth level scaling coefficients, to get the $j + 1$ level scaling and detail coefficients [14].

2.1.2 Wavelet Variance.

The wavelet variance $v_f^2(\tau_j)$ is defined as the expected value of $\tilde{D}_{j,t}^2$ considering only non-boundary coefficients[2]. The unbiased estimator of the wavelet variance is achieved by omitting the coefficients impacted by boundary conditions and is calculated as follows:

$$v_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,l}^2 \qquad (2)$$

where $M_j = N - L_j + 1$ is number of non-boundary coefficients at j^{th} level [14]. The series behaviour over different horizons on a scale-by-scale basis is shown by the wavelet variance.

2.1.3 Wavelet Covariance and Correlation

Like the wavelet variance above, the wavelet covariance between functions $f(t)$, $g(t)$ is defined as the covariance of wavelet coefficients at scale j. The unbiased estimator of the wavelet covariance at the jth scale is:

$$v_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,l}^{f(t)} \tilde{D}_{j,l}^{g(t)} \qquad (3)$$

Again, all wavelet coefficients affected by the boundary are removed [14], and $M_j = N - L_j + 1$. The MODWT estimate of the wavelet correlation between functions $f(t)$ and $g(t)$ is found with the wavelet covariance and square root of the wavelet variance of

the functions at each scale j [14]. The MODWT estimator, of the wavelet correlation is given by:

$$\rho_{fg}(\tau_j) = \frac{v_{fg}(\tau_j)}{v_f(\tau_j)v_g(\tau_j)} \qquad (4)$$

where, at scale j, $v_{fg}(\tau_j)$ is the covariance between $f(t)$ and $g(t)$, $v_f(\tau_j)$ is the variance of $f(t)$ and $v_g(\tau_j)$ is the variance of $g(t)$ [14].

2.1.4 Wavelet Multiple Correlation and Cross-Correlation

The wavelet multiple correlation and cross-correlation give the overall statistical relationship at different time scales among a set of multivariate random data. The wavelet multiple correlations (WMC) $\varphi_X(\lambda_j)$ are defined as one single set of multiscale correlations calculated from X_t where $X_t = (x_{1t}, x_{2t} \dots, x_{nt})$ is a multivariate stochastic process and $W_{jt} = w_{1jt}, w_{2jt}, \dots, w_{njt}$ the respective scale λ_j wavelet coefficients from application of the maximum overlap discrete wavelet transform to each x_{it} process [20]. The wavelet multiple correlation is:

$$\varphi_X(\lambda_j) = Corr(w_{ijt}, \hat{w}_{ijt}) = \frac{Cov(w_{ijt}, \hat{w}_{ijt})}{\sqrt{Var(w_{ijt})Var(\hat{w}_{ijt})}}, \quad (5)$$

where w_{ij} is chosen to maximise $\varphi_X(\lambda_j)$ and $\hat{w}_{ij}$ are the fitted values in the regression of w_{ij} on the rest of the wavelet coefficients as scale λ_j.

Allowing for a lag of τ between observed and fitted values of the variable selected [20] as the criterion variable at each scale λ_j the wavelet multiple cross-correlation (WMCC) is defined as:

$$Corr(w_{ijt}, \hat{w}_{ijt+\tau}) = \frac{Cov(w_{ijt}, \hat{w}_{ijt+\tau})}{\sqrt{Var(w_{ijt})Var(\hat{w}_{ijt+\tau})}} \qquad (6)$$

3 The Dataset

The RTWP dataset (RNC20_30 data) has uplink receive level values at a radio base station (RBS) taken from a live 3G network. The dataset is extracted from the operational support database in CSV format and is post-processed in R. The analysis was performed offline due to the unavailability of dedicated hardware to perform real time analysis. In order to productionise this analysis a real time streaming engine would be configured to ingest the RTWP reports. From there an R or Python scripts would parse and manipulate the data before the wavelet algorithms analyse the data for correlations between cells to identify the presence or absence of external interference. RTWP data is available at granularities raw data (15 min), hourly and daily. The analysis here concentrates on a subsample of the RNC20_30 data dataset using the raw format for representation purposes. A site with a known external interference source was identified and its neighbouring cells analysed to see if a lead-lag relationship exists between these proximal sites.

---

[2] The MODWT treats the time series as if they are periodic using "circular boundary conditions". There are Lj wavelet and scaling coefficients that are influenced by the extension, which are referred to as the boundary coefficients.

3.1 Data Visualisation

In Fig. 2 we see RTWP levels in decibel-milliwatts (dBm) plotted against time for site LX0088. Sector is the industry term for an antenna on a mast radiating a specific frequency. Typically a site has 3 sectors, each covering a unique geographic area so one expects different noise profiles based upon area subscriber density and mobility.
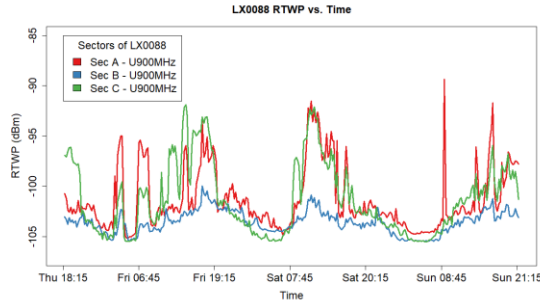

Figure 2: RTWP levels vs. Time showing the presence of external interference particularily on Sec A and C.

In Fig. 2 above, Sector B (Sec B) shows a typical noise profile for a cell operating as normal. As usage decreases after midnight, the noise level is close to the noise floor of ≈-105dBm. Network activity increases into the day and again decreases at night. Sectors A and C show very different profiles to Sector B. Sudden bursts of noise such as on Sunday at ≈08:45am appear and vanish once again.

Early Friday morning there are two very obvious sudden bursts of noise in Sectors A and C and to a lesser extent in Sector B. From above, the chance of all three sectors exhibiting similar noise signatures such as that seen on Friday at 06:45 is very small as they cover different geographic areas, this strongly points to the presence of local external interference.

Table 1: Summary statistics for RNC20_30

Freq. Band	Mean (dBm)	Median (dBm)	S.D (dBm)	Max (dBm)	Min (dBm)	Count
U900 F0	-104.44	-105.10	1.14	-65	-110	1463
U21 F1	-104.98	-105.35	0.79	-73.63	-110	1527
U21 F2	-104.84	-105.24	0.87	-73.93	-110	1488
U21 F3	-105.04	-105.38	0.82	-75.03	-110	1509

Table 1 shows typical values for the 4 different frequencies in the network based on a sample of 5987 cells. The table shows that the U900MHz cells mean exceeds that of the U2100MHz cells. The table below indicates that the mean of the U900MHz cells is higher than that of the U2100MHz cells. Since the mean value for all the frequency bands is higher than the median, this is indicative of a right skewed distribution of values. The maximum values indicate the strength of interference sources which average 30dB of noise

for U2100 and 40dB of noise for the U900 cells. Minimum values of -110dBm indicate that some "deaf" cells in the network should be investigated for physical build issues.
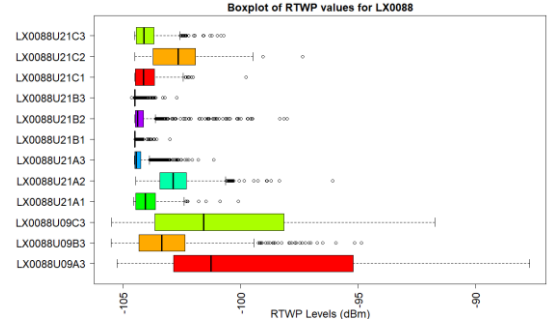

Figure 3: Boxplot of RTWP values for LX0088

The box plot in Fig. 3 shows RTWP values for all cells across all frequencies belonging to site LX0088. From this representation, the range of values for all three U900MHz cells is obviously much larger than that of the U2100MHz cells. The median values for U900MHz Sector A and C are higher than that of Sec B which is also evident on the U2100MHz cell particularly U21A2 and U21C2.

4 Results

4.1 Time Series Decomposition

Time series decomposition using Loess [21] splits the RTWP data into its four main components, the trend, cyclical, seasonal and irregular parts. Fig. 4 in the Appendix shows the seasonal decomposition for LX0088U09A3, the cell showing external interference as discussed in Fig. 2. Fig. 4 shows from top to bottom (a) the original time series, (b) the seasonal, (c) the trend and (d) the remainder components.
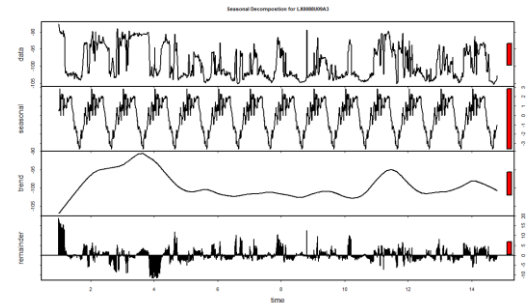

Figure 4: Seasonal decomposition for LX0099U09A3

The seasonal component indicates a daily periodicity and the vertical red bars on the right-hand side of the image show that the seasonal signal is large relative to the data *variation*. This component clearly shows the early morning and afternoon spike in noise levels typically associated with waking and lunch time heavy

usage patterns. During the consistently high noise periods during day 2 and 3 in the original data, the trend also increases before falling with the noise after day 4.

4.2 Autocorrelation and Partial Autocorrelation Analysis

Other statistical properties of interest for patterns include the autocorrelation (ACF) and partial autocorrelation (PACF) functions. For a normal network cell, we should see a strong internal link between the signals at regular intervals given the strong seasonality from above. Before we assess the ACF and PACF, differencing was required in order to ensure data stationarity. First order differencing was used to stabilise the mean and ensured a constant variance around the mean.
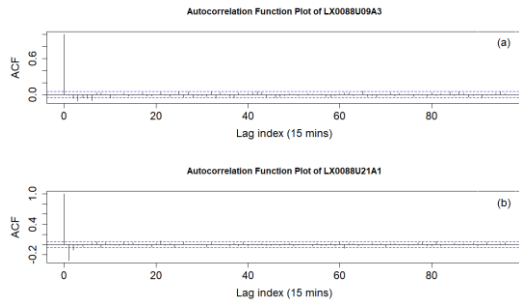

Figure 5: (a) Autocorrelation plot for LX0088U09A3, (b) autocorrelation for LX0088U21A1

Fig. 5(a) shows the ACF shut off immediately after lag 0 with a few significant negative lags at 2, 3 and 6. The corresponding PACF plot in Fig. 6(a) in the Appendix shows a slowly decaying negative PACF up to lag 7. These 2 plots, we use to reliably forecast future values for LX0088U09A3 an ARIMA(3,1,5) model [22]. Similarly, Fig. 5(b) and 6(b) also in the Appendix show a comparable behaviour for LX0088U21A1 which covers the same geographic area.
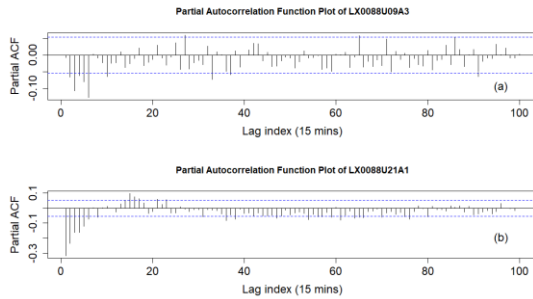

Figure 6: (a) Partial autocorrelation plot for LX0088U09A3, (b) partial autocorrelation plot for LX0088U21A1

In Fig. 5(b) in the Appendix we see a sudden shut off in ACF with large negative lags for values 1 & 2 while the corresponding PACF

plot in Fig. 6(b) shows a gradual decrease in the PACF with significant lags for values up to 6. Again the process suggests an ARIMA(2,1,6) model. In Fig. 5(b) the first lag shows a negative autocorrelation at lag 1, implying that if a value is above average then subsequent values will be below average. The ACF plot for LX0088U09A3, the cell with a verified external interferer shows that the first significant value occurs at lag 2 and this is also negative. This implies that if an RTWP measurement is above average, the value is expected to remain above average until 2 lags have occurred which equates to 30 min. Comparing this to the equivalent ACF plot for LX0088U21A1, the U900MHz cell will remain above average for up to 30 min while the next value for the U2100MHz will be below average.

4.3 Hurst Exponent

The Hurst exponent is a statistical measure of the long-term memory of a time series and describes the rate of decrease of autocorrelations with lag. The Hurst quantifies the relative tendency of a time series either to regress strongly to the mean or to cluster in a certain direction [23]. It varies between 0 and 1: $H = 0.5$ implies a random walk or independent process; $0 \leq H < 0.5$ then the time series is *anti-persistent* meaning that a time series with decreasing trend is more likely to show an increasing trend next. If $0.5 < H \leq 1$ the process is persistent, (*i.e. if we have an increasing time series then it is more probable that it will continue to show an increasing trend* [24]).
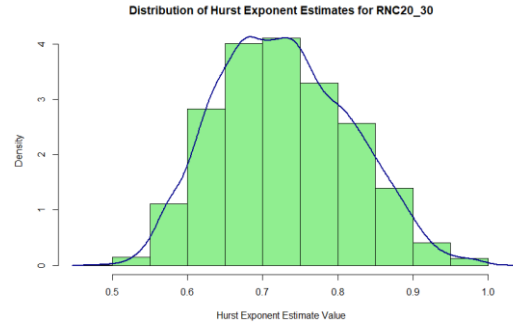

Figure 7: Distribution of Hurst exponent estimates for RNC20_30

Fig. 7 in the Appendix section shows the Hurst distribution for the RNC20_30 dataset where $0.5 \leq H < 1$ indicates long term positive autocorrelation such that high noise levels will likely be followed by other high values. Further, as the Hurst seems normally distributed, we assume that the original data sampled from have similar properties.
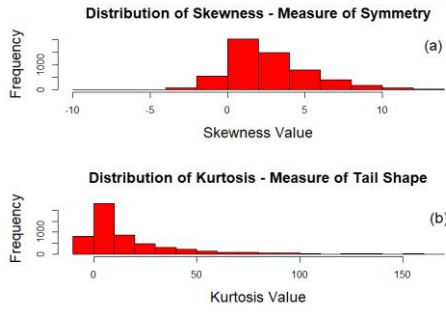
Figure 8: RNC20_30 Distribution of (a) Skewness and (b) Kurtosis

The skewness distribution in Fig. 8(a) in the Appendix indicates that many of the variables in the RNC20_30 dataset are clearly right skewed indicated by the positive values while the kurtosis distribution in Fig. 8(b) indicates that most of the variables in the same dataset have fat-tailed or *leptokurtic* distributions.

4.4 Intraday Correlations

The intraday correlations show the co-movements of RTWP values between days for a small sample of cells and describe the intraday volatility for a cell with a known interference source and one without. Fig. 9(a) and 9(b) show the intraday correlations for both LX0088U21A1 and LX0088U09A3 for a two-week period. LX0088U21A1 provides coverage to the same geographic area as LX0088U09A3 but uses a different frequency band which is less susceptible to *uplink* interference. The first noticeable observation in Fig. 9(a) is the strong positive correlation for cell LX0088U21A1 for all days with a minimum correlation of 0.4 between Day 7 and Day 8 while the strongest correlation occurred between Day 3 and Day 4 as well as between Day 11 and Day 3 with values of 0.9.
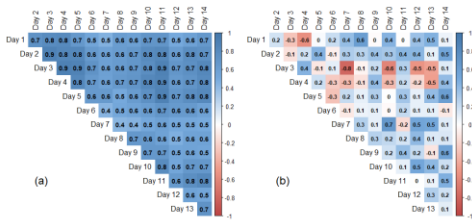

Figure 9: (a) Intraday correlation for LX0088U09A3 and (b) LX0088U21A1

Fig. 9(b) also details the intraday correlations for cell LX0088U09A3 from a verified external interference source. LX0088U21A1 and LX0088U09A3 differ greatly: In Fig. 9(a) we see contrasting values between days, between Day 10 and Day 7

the correlation is positive and strong, contrasting with strong negative correlations as between Day 7 and Day 3. On average, the intraday correlations are typically weak with values between 0.2 and -0.2. Fig. 10 also in the Appendix is a plot of RTWP for both cells over the 14 days. We see the contrast in RTWP values for both cells and the impact an external interference source has on LX0088U09A3.
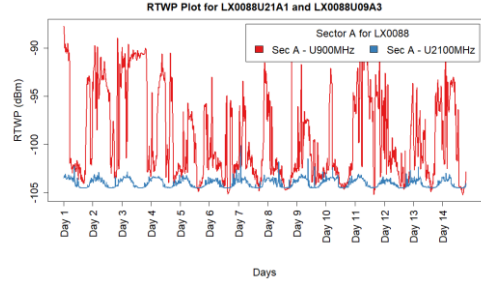

Figure 10: RTWP of LX0088U09A3 and LX0088U21A1

In Fig. 10, we see that on Day 3 the magnitude of the interference levels for LX0088U09A3 is consistently high throughout the entire day and rolls over into day 4 which explains the strong negative interday correlations between Day 3 and Day 7, Day 10, Day 12 and Day 13. Since the interference affected Day 4 also, we expect to see strong negative correlations for this day also.

4.5 Eigenvalue Dynamics

The first eigenvalue (λ_1) of a correlation matrix shows the maximal variance of the variables which can be accounted for with a linear model by a single underlying factor [25]. For all positive correlations, this first eigenvalue is roughly a linear function of the average correlation among the variables [26]. As per Fig. 4 in the Appendix, the RTWP data has a seasonality of 24 hours, suggesting that the correlation between variables would also exhibit a periodicity across the same scale and likewise the eigenvalues of the correlation matrix.

With 10 cells from the RTWP dataset, one having a known interference source, we look at the eigenvalue dynamics for unusual patterns: Fig. 11 gives the largest eigenvalue for four different window sizes: (a) 90 mins, (b) 300 mins, (c) 750 mins and (d) 1500 mins. Using normalised eigenvalues as described above (1), the dynamics of (λ_{max}) should show the presence of unique events.

Obviously as the window size increases the high frequency information visible in Fig. 11(a) in the Appendix section is lost due to the smoothing effect of the larger window size. The changes in magnitude of λ_1 for the different window sizes indicate substantial changes in the noise levels between cells at these time periods. In Fig. 11(c) we clearly see 14 distinct double peaks, (some of these double peak events are identified by a red asterisks) a sudden rise followed by a small decline then a rise and then another drop. Each double peak represents a day's information. These troughs on the 14 peaks coincide with early afternoon noise levels and seem to

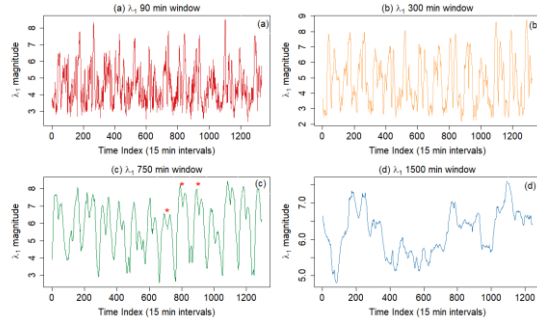signify low volatility periods as all cells are expected to have peak RTWP values at the busy times of day.

Figure 11: Largest Eigenvalue for (a) 90min (b) 300min (c) 750min (d) 1500min windows

In Fig. 11(d) in the Appendix we see a smoothly varying function with eigenvalue magnitudes varying between 5.0 and 7.0. This plot captures the daily noise variations between cells as a moving average and therefore any low frequency variations are smoothed out. As stated in Section 2.1 above, the sum of all eigenvalues must equal $Tr(C)$.

As C is a 10x10 matrix, from Fig. 11(d) we can say that λ_1 explains between 50-70% of the system noise and fluctuations between these values, evident between time index 400-700 are due to other sources of noise which cause variation in eigenvalue magnitudes at these times.

4.6 Wavelet Multiple Correlation

This section introduces the wavelet multiple correlation method to assess the overall statistical relationship between many variables. Macho [20] introduced this approach when studying the wavelet multiple correlation and cross-correlations between the Eurozone stock markets. The wavelet multiple correlation method produces a single statistical measure of the multivariate sample on a scale-by-scale basis. Using this approach, the wavelet correlation between pairs of variables can be assessed in a single visualisation rather than having to compare multiple wavelet statistics for all the variables under analysis.

We decomposed the RTWP dataset using the Daubechies least asymmetric (*LA8*) wavelet filter which has a filter of length *L*=8. This filter is used extensively in finance [13] [5] as it provides a reliable estimate of correlation between long memory time series [27]. Based on the findings in Section 4.3 this was an appropriate choice filter given an average H value of 0.7. The maximum level of decomposition using the LA8 filter is given by $log_2(T)$ where T is the time series length, in our case this is 1344 equating to a maximum decomposition level of 10. For this study, we use $J = 7$ resulting in 7 wavelet coefficients and one scaling coefficient for each interval in the RTWP dataset i.e. $\widetilde{\omega}_{i1}, ..., \widetilde{\omega}_{i8}$ and $\widetilde{v}_{i8}$ respectively. Since a MODWT approximates an ideal band-pass filter (with bandpass given by the frequency interval

$[2^{-(j+1)}, 2^{-j}]$) for $J = 1, ..., J$. Inverting the frequency range, the corresponding periods are within $(2^j, 2^{j+1})$ time unit intervals [27].
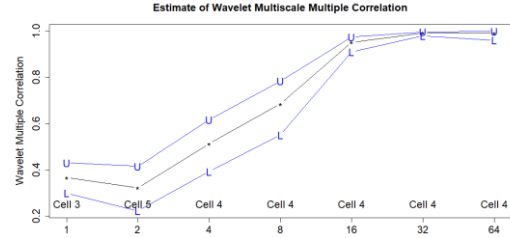

Figure 12: Estimate of wavelet multiscale correlation

Since the RTWP dataset is sampled at 15 min intervals, the wavelet coefficients represent the following intervals, 30-60 min, 1-2 hours, 2-4 hours, 4-8 hours, 8-16 hours, 16-32 hours (daily scale) and 32-64 hours (2-day scale).

Fig. 12 in the Appendix shows the wavelet multiple correlation for a sample of 10 cells from the RTWP dataset. It gives the strength of the wavelet correlations between the sample of 10 cells across the 7 wavelet scales. The straight blue lines represent the upper and lower bounds of the 95% confidence intervals. For each wavelet level, the variable which maximises the multiple correlation against a linear combination of the rest of the variables is also plotted. From the wavelet multiple correlation for a sample of 10 cells, we see that all the correlations are positive and indicate a strong positive relationship. The scale rises with the correlation between cells up to perfect cell correlation at the longest scale. It is also evident that as the scales increase the confidence interval narrows showing increasing certainty of the estimate. Since near perfect correlation exists at wavelet scales greater than 16, an exact linear relationship between the RTWP values of the 10 cells cannot be ruled out. The presence of such a relationship means that noise levels for any cell can be estimated by the overall noise levels of the other cells in the sample. Another interesting point is the brief decrease in correlation between scales 1 and 2 before the correlation almost linearly increases to perfect correlation at longer scales. This shows that, at smaller time scales, clear discrepancies exist between the RTWP levels but that over longer time periods the RTWP values follow the same overall trend.

4.7 Wavelet Multiple Cross-Correlation

Fig. 13 also in the Appendix shows the wavelet multiple cross-correlation for different wavelet scales with leads and lags up to 25 hours. Like the wavelet multiple correlation above, the multiple cross-correlation decomposes the usual cross-correlation and produces patterns which show the relationship between multiple variables across various physical time scales [27]. Each wavelet scale plot shows in its upper left-hand corner the network cell maximising the multiple correlation against a linear combination of

the other variables and, thus, signals a potential leader or follower for the whole system [20]. The wavelet multiple cross-correlation analysis was performed for several different lead/lag values and for brevity Fig. 13 shows the positive and negative lag up to 25 hours.
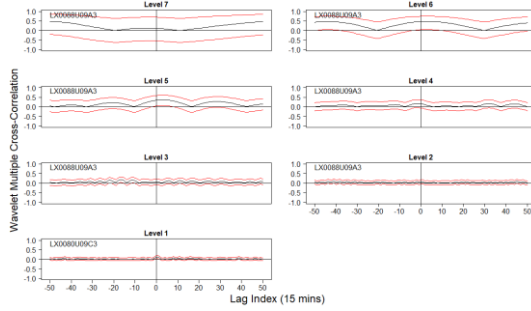

Figure 13: Wavelet multiple cross-correlation for a sample of 10 cells at different wavelet scales. The continuous red line corresponds to the upper and lower bounds of the 95% confidence interval.

Like Fig. 12 from the Appendix, Fig. 13 also shows consistent positive correlations across all lags for all lead/lag values. Upper and lower bounds of the 95% confidence interval are shown as continuous red lines and again the variable maximising the multiple correlation against a linear combination of the rest of the variables is shown in the upper left-hand corner for each scale. When both confidence intervals are above the horizontal axis on the right-hand side of the graph, this indicates a positive statistically significant lagging wavelet cross-correlation and inversely, a positive confidence interval on the left-hand side of the chart indicates a positive statistically significant leading correlation. From the results in Fig. 13, we see that for the majority of lag values, the cross-correlation between the multivariates is statistically insignificant since neither of the confidence intervals are greater than zero. Clearly there are brief periods around the zero-lag mark across a number of wavelet scales indicating a statistically significant cross-correlation. Fig. 13 shows that there are statistically significant events in Levels 1, 5 & 6 for lags of 1, 4 and 4 respectively where levels 1, 5 & 6 represent 30-60 mins, 8-16 hours and 16-32 hours respectively. In Fig. 13, the Level 1 plot identifies the presence of a statistical lag relationship for LX0080U09C3 at lag 1 at the 30-60 min time scale. The Level 3 plot in Fig. 13 indicates that RTWP values for LX0088U09A3 tends to statistically lag the other cells for time scales of 4-8 hours. We see in the Level 3 (2-4 hours) plot statistically significant periods for LX0088U09A3 at lag values of -11 and -17 indicating possible RTWP levels for LX0088U09A3 leading the other cells in the analysis for time scales of 2-4 hours. The Level 6 plot also shows a statistically significant lag for LX0088U09A3 at lag values of 4 for time scales 16-32 hours.

4.8 Conclusion

High RTWP levels have serious consequences in modern wireless networks due to their adverse effect on coverage of a wireless transceiver site and negative impact on customer satisfaction. Quickly classifying the cause and identifying the source of the problem is key to resolving issues as efficiently as possible while at the same time minimising operational expenditure in the fault-finding process. With the explosion in usage of noisy packet switched data sessions, monitoring and protecting *uplink* noise levels is key for a wireless operator hence alternative data mining techniques are needed.

By examining the eigenvalue spectrum of the correlation matrix for a small sample of cells we looked at dynamical changes in the RTWP values using different window sizes which reflect large changes in the RTWP values of the sample. Using this approach an engineer could quickly identify major changes in RTWP values for a group of cells rather than assessing individual graphs of RTWP values and quickly quantify the magnitude of the change in noise levels for a group of cells rather than assessing the impact at an individual cell level.

The maximum overlap discrete wavelet transform enabled the multilevel decomposition of the raw RTWP time series into their respective coefficients for different time horizons. This technique enables RTWP measurements to be investigated for correlations and cross-correlations over several different time horizons. The multiscale wavelet correlation showed how it can be used to pinpoint short term deviations in RTWP values when assessed over numerous scales. Similarly, the multiscale wavelet cross-correlation identified a number of different lead and lag relationships not evident in standard approaches. These lead/lag relationships need further investigation to fully understand the complex interplay that exists between each network element.

Future work includes investigating the eigenvalue spectrum of the wavelet correlation matrix using a heatmap diagram for evidence of significant events across various wavelet scales. In this way, an engineer could easily identify periods of high RTWP values across different time horizons. To quantify the significance and meaning of the elements of the cross-correlation matrix C, it would be advantageous to quantify the correlations between such cells by comparing the statistics of the cross-correlation matrix to the null hypothesis of a random matrix using RMT as per [28], if the properties of C conform to those of a random correlation matrix, it signifies random correlations in C. Deviations of the properties of C from those of the random matrix convey information about genuine correlations requiring further investigation.

Another possible approach to be considered could involve time series clustering based on similarity or distance measurements, the discrete wavelet transform can be used as a metric of similarity. In doing so, time series that are similar are clustered together and may assist in identifying cells which exhibit similar patterns of interference. These clusters could then be further analysed to identify cells which are spatially grouped which may signal the presence of an external interference source.

References

[1] D. C. Robertson, O. I. Camps, J. S. Mayer and W. B. Gish, "Wavelets and electromagnetic power system transients," IEEE T Power Deliver, vol. 11, no. 2, pp. 1050-1058, 1996.

[2] T. Conlon, H. J. Ruskin and M. Crane, "Seizure characterisation using frequency-dependent multivariate dynamics," Comput. Biol. Med, vol. 39, no. 9, pp. 760-767, 2009.

[3] J. F. Muzy, E. Bacry and A. Arneodo, "Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method," Phys Rev E, pp. 875-884, 1993.

[4] A. Grinsted, J. C. Moore and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," Nonlinear Process. Geophys., vol. 11, no. 5/6, pp. 561-566, 2004.

[5] T. Conlon, H. J. Ruskin and M. Crane, "Multiscales cross-correlation dynamics in finacal time-series," Advances in Complex Systems, vol. 12, no. 04n05, pp. 439-454, 2009.

[6] N. Li, M. Crane and H. J. Ruskin, "Automatically Detecting "Significant Events" on SenseCam," Int J Wavelets, Multi, vol. 11, no. 06, p. 1350050, 2013.

[7] H. Holma and A. Toskala, WCDMA for UMTS, Fifth Edition ed., Chichester: Wiley, 2005, p. 48.

[8] S. P. Girija and K. D. Rao, "Smoothing term based noise correlation matrix construction for MIMO-OFDM wireless networks for impulse noise mitigation," Macao, 2015.

[9] P. K. Gkonis, G. V. Tsoulos and D. I. Kaklamani, "Dual code Tx diversity with antenna selection for spatial multiplexing in MIMO-WCDMA networks," *IEEE Commun Lett,* pp. 570-572, 2009.

[10] B. A. Bjerke, Z. Zvonar and J. G. Proakis, "Antenna Diversity Combining Schemes for WCDMA Systems in Fading Multipath Channels," *IEEE T Wirel Commun Le,* pp. 97-106, 2004.

[11] M. Keskinoz, S. Olcer and H. Sadjadpour, "Advances in signal processing for wireless and wired communications [Guest Editorial]," IEEE Commun Mag, vol. 47, no. 1, pp. 30-31, 2009.

[12] S. Hu, Y. Ouyang, Y. D. Yao, M. H. Fallah and W. Lu, "A study of LTE network performance based on data analytics and statistical modeling," Newark, 2014.

[13] T. Conlon and J. Cotter, "An empirical analysis of dynamic multiscale hedging using wavelet decomposition," J Futures Markets, vol. 32, no. 3, pp. 272-299, 2011.

[14] D. B. Percival and A. T. Walden, Wavelet methods for time series analysis, Cambridge, United Kingdom: Cambridge Univ. Press, 2008.

[15] K. Schindler, H. Leung, C. E. Elger and K. Lehnertz, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG," Brain, vol. 130, no. 1, pp. 65-77, 2006.

[16] G. Oas, "Universal cubic eigenvalue repulsion for random normal matrices," *Phys Rev E,* vol. 55, no. 1, pp. 205-211, 1997.

[17] M. Muller, G. Baier, A. Galka, U. Stephani and H. Muhle, "Algorithm for the Detection of Changes of the Correlation Structure in Multivariate Time Series," *Electronics and Electrical Engineering,* 2012.

[18] J. P. Bouchaud and M. Potters, Theory of Financial Risk and Derivative, Cambridge University Press, 2003.

[19] S. C. Burrus, R. A. Gopinath and H. Guo, An introduction to wavelets and wavelet transforms: A Primer, 1st ed., Prentice-Hall, 1995.

[20] J. Fernandez-Macho, "Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets," Physica A, vol. 391, no. 4, pp. 1097-1104, 2012.

[21] R. B. Cleveland, W. S. Cleveland, J. E. McRae and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess.," J Off Stat, vol. 6, no. 1, p. 3, 1990.

[22] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," Otexts.org, [Online]. Available: https://www.otexts.org/fpp.

[23] T. Kleinow, "Testing continuous time models in financial markets," Doctoral dissertation, Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, 2002.

[24] M. Kale and F. Butar Butar, "Fractal analysis of time series and distribution properties of Hurst exponent. Diss," Sam Houston State University, 2005.

[25] G. H. Dunteman, Principal components analysis, Newbury Park [etc.]: Sage Publications, 2016.

[26] S. Friedman and H. F. Weisberg, "Interpreting the First Eigenvalue of a Correlation Matrix," Educ. Psychol. Meas, pp. 11-21, 1981.

[27] B. Whitcher, P. Guttorp and D. B. Percival, "Wavelet analysis of covariance with application to atmospheric time series.," J. Geophys. Res., vol. 105, no. D11, pp. 14941-14962, 2000.

[28] T. Conlon , H. J. Ruskin and M. Crane, "Random matrix theory and fund of funds portfolio optimisation," Physica A, vol. 382, no. 2, pp. 565-576, 2007.

[29] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," J R Stat Soc, vol. 39, no. 1, pp. 1-38, 1977.