Statistical Learning Techniques for Spectrum Based Signal Synthesis

D. Riegert & J.W.W. Pohlkamp-Hartt

May 27, 2015

1 Introduction

Many times in the study of time series data we are presented with the problem of synthesizing data. This synthesis may be required to interpolate missing data or to predict future values and trends. In either case there are a variety of approaches that a statistician can employ and achieve reasonable results. The focus of this report is to spotlight a lesser known synthesis method using Multitaper Spectrum Estimation and provide a framework and tools for use on real data problems.

The technique we will employ is to perform an inverse fourier transform on the complex regression coefficients that are produced when computing the F-test for line components. This method is attributed to Dr. David Thomson, who has not published a paper outlining this method but has utilized it within analysis and lectured on the topic. After outlining the details of how the method works we will provide some insight and advancements to the basic procedure using techniques borrowed from statistical learning theory. Analysis of well weather data is provided to demonstrate these techniques in a real world context.

2 Spectrum Estimation

Spectrum estimation is the process of transforming time correlated data from the time domain to the frequency domain. We perform this transformation in an effort to study periodic structure that exists within the data. The study of this periodic structure can allow us model the deterministic elements of the time series and synthesize estimates.

Many methods of spectral estimation exist with the majority of research being done under the assumptions of Gaussian noise and stationary signals. Following these two assumptions we approach estimating the spectrum in this basic way; take a Fourier transform of the time series multiplied by a sequence of weights (a window or taper). The choice of window is where most methods differ. For most methods, depending on the choice of windows, there is a trade off between the bias and the variance of the estimate[1].

3 Multitaper Method

This trade off is controlled when using the Multitaper Spectral Estimation method (MTM) [4]. The MTM was introduced by David J. Thomson and allows for bias control without a significant corresponding increase in variance. The MTM is similar to other methods in that it uses windowed Fourier transforms of the data series to produce spectral estimates. However, it differs through use its of an orthogonal family of windows instead of a single choice. This orthogonal family consists of a group of discrete prolate spheroidal sequences (DPSS, or Slepians) [3]. By using any orthogonal family of functions they will have maximal energy within a given frequency band. An attractive property of the Slepians is that the Fourier transformations of the windows minimize the weight given to out-of-band frequencies.

The method begins by defining a time-bandwidth product NW, with N the number of data points, and W the bandwidth parameter. Given NW, we choose to compute between NW and $2NW$ Slepian sequences of length N, $\nu_t^{(k)}$, where $t = 0, ..., N-1,\ k = 0, ..., K-1$, the number of windows denoted by K. We use these as windows for K Fourier transformations, called the eigenspectra of the data.

$$Y_k(f) = \sum_{t=0}^{N-1} \nu_t^{(k)} e^{-2i\pi ft} x_t, \tag{1}$$

The initial naive spectral estimate is then formed as

$$\overline{S} = \frac{1}{K} \sum_{k=0}^{K-1} |Y_k(f)|^2, \tag{2}$$

the average of the K eigenspectra.

The choice of NW and K are important to the shape of the spectrum and, by extension, the synthesized estimates for the time series. For smaller values of W, we get a higher frequency resolution but increased variance in our estimates. The opposite hold for larger values. After setting W, the choice of K works as a bias-variance trade-off. For K closer to $2NW$ we get more eigenspectra providing less variance to the estimates but higher out-of-band power, which increases the bias of the estimate. Lower values of K do not suffer as poorly with out-of-band bias but have increased variance. The choice of parameters is important to further evaluation of the data in the frequency domain. Supervised selection based on known characteristics of the data is the common practice, although this can lead to selection bias in your research. We employ an unsupervised method motivated by the normality assumption for the noise of the time series. The details of this method can be found here [2].

4 The F-test

The use of Thomson's synthesis method relies on the signal detection and complex mean values,$\hat{\mu}(f)$ resulting from the computation of the F-test for detection of line components. We think of the F-test as a regression problem, where we are regressing the frequency domain Slepian sequences taken at the base frequency, $V_k(0)$, onto the eigenspectra for each frequency, $Y_k(f)$.

$$Y_k(f) = \hat{\mu}(f)V_k(0) + e(f), \tag{3}$$

where $e(f) \sim CN(0, \sigma^2)$ is Complex Gaussian distributed with variance equal to the background noise of the environment, $\sigma^2 = S_N(f)$.

To detect a signal at a frequency, we test the null hypothesis $H_0 : \mu(f) = 0$. To do this, we obtain estimates of the $\mu(f)$ from linear regression and then use an F-test to determine if there is evidence that $\mu(f)$ is non-zero. The statistic for the F-test follows an $F(2, 2K-2, p)$ distribution and is computed by:

$$F(f) = (K-1)\frac{|\hat{\mu}(f)|^2 \sum_{k=0}^{K-1} |V_k(0)|^2}{\sum_{k=0}^{K-1} |\hat{r}_k(f)|^2}, \tag{4}$$

$$\hat{r}_k(f) = Y_k(f) - \hat{\mu}(f)V_k(0), \tag{5}$$

$$\hat{\mu}(f) = \frac{\sum_{k=0}^{K-1} V_k^\star(0)Y_k(f)}{\sum_{k=0}^{K-1} |V_k(0)|^2}, \tag{6}$$

$$V_k(f) = \sum_{t=0}^{N-1} \nu_t^{(k)} e^{-2i\pi ft}, \tag{7}$$

where Y_k are the eigenspectra for our time series x_t, and $\nu_t^{(k)}$ are the Slepian sequences in the time domain.

We are also able to get an estimate for the residuals, $\hat{r}_k(f) = Y_k(f) - \hat{\mu}(f)V_k(0)$. These residual are used to evaluate the validity of the parameter choices used in the MTM [2]. The choice of significance level, p, for the F-test will greatly effect the estimates returned from Thomson's method so care must be taken when making this choice. We will discuss later an unsupervised method for finding p and the effect on your estimates a choice of p has.

5 Inverse Fourier Transform Signal Synthesis

Following the identification of significant frequencies within the spectrum we want to synthesize these deterministic periodic trends to form an estimate of the time series without noise. The obvious way we may attempt to do this is by attempting to determine the phase and amplitude of the significant periodic components and modeling the time series as the sum of sinusoids with these properties. This has been shown to have be marginally successful (ref bpa work) but this can become a cumbersome set of computations if your set of significant frequencies is large.

Thomson has proposed an alternative method that follows directly from the F-test, by performing an inverse Fourier transform on the regression coefficients from the F-test we are able to get a reasonable approximation of our original time series. That is,

$$\mathcal{F}^{-1}(\hat{\mu})(t) = \sum_{f=-f_n}^{f_n} e^{i2\pi ft} \frac{\sum_{k=0}^{K-1} V_k^\star(0)Y_k(f)}{\sum_{k=0}^{K-1} |V_k(0)|^2}. \tag{8}$$

Knowing the denominator is a constant with respect to f we can move it outside the sum and since $e^{i2\pi ft}$ is constant with respect to k we will move it inside the inner summation,

$$\mathcal{F}^{-1}(\hat{\mu})(t) = \frac{1}{\sum_{k=0}^{K-1} |V_k(0)|^2} \sum_{f=-f_n}^{f_n} \sum_{k=0}^{K-1} e^{i2\pi ft} V_k^\star(0) Y_k(f). \tag{9}$$

Now using Fubini's Theorem as the interior functions are integrable we can change the orders of summation and move the $V_k^\star(0)$ outside the inner summation now as it is constant with respect to f.

$$\mathcal{F}^{-1}(\hat{\mu})(t) = \frac{1}{\sum_{k=0}^{K-1} |V_k(0)|^2} \sum_{k=0}^{K-1} V_k^\star(0) \sum_{f=-f_n}^{f_n} e^{i2\pi ft} Y_k(f). \tag{10}$$

Noting that the inner summation is now the discrete inverse Fourier transform of $Y_k(f)$ and $Y_k(f)$ is defined as the Fourier transform of $\nu_t^{(k)} x_t$, we can the inversion theorem for Fourier transforms to replace the inner sum,

$$\mathcal{F}^{-1}(\hat{\mu})(t) = \frac{1}{\sum_{k=0}^{K-1} |V_k(0)|^2} \sum_{k=0}^{K-1} V_k^\star(0) \nu_t^{(k)} x_t. \tag{11}$$

Next as x_t is constant under k we can move it outside the summation and we are left with a time weighted multiple of x_t.

$$\mathcal{F}^{-1}(\hat{\mu})(t) = \frac{\sum_{k=0}^{K-1} V_k^\star(0) \nu_t^{(k)}}{\sum_{k=0}^{K-1} |V_k(0)|^2} x_t. \tag{12}$$

Inverting this weighing we can get back an estimate of our time series,

$$\hat{x}_t = \frac{\sum_{k=0}^{K-1} |V_k(0)|^2}{\sum_{k=0}^{K-1} V_k^\star(0) \nu_t^{(k)}} \mathcal{F}^{-1}(\hat{\mu})(t). \tag{13}$$

From this computation we can return an estimate of our time series based on the regression coefficients found within F-test but it is unrealistic the expect that a periodic component will be found at each frequency. Instead of using the full set of coefficients we will use a subset that is significant under the F-test,

$$\hat{\mu}_\alpha(f) = \begin{cases} \hat{\mu}(f), & \hat{F}(f) > F_{(2nw-1,2,\alpha)}. \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Now we can synthesize an estimate of the significant periodic parts within our time series by replacing $\hat{\mu}(f)$ with $\hat{\mu}_\alpha(f)$ in equation 13.

6 Interpolation & Prediction

The preceding method allows us to reproduce the periodic components with a time series for times where we have observations. This may be useful as a de-noising process for data clean up but many problems require interpolation or prediction to be made from time series data. Luckily we can still use the same method with a small variation. The introduction of some proxy data in the time series will allow us to return an estimate of the periodic components operating at these unknown intervals.

For prediction of future data, we zero pad the data the desired number of data points and when we return the estimate of the significant components they will persist through the extra data points giving us a prediction estimate. The added bonus from zero padding is that we improve the frequency resolution which will allow for better estimation of the significant frequencies. It is common practice to zero pad to improve the resolution of the frequencies for analysis therefore in many cases there is negligible extra cost to achieve a prediction from the period components for most.

To interpolate the data we will want some approximation of the missing data first and then perform the same method as before. The most common method is to linearly interpolate the gap in data as this will not introduce any new periodic trends like is possible when using a spline based approximation. A drawback of using a linear interpolation is the reduced high frequency power that will be found in the spectrum. This will bias the interpolation by reducing the significance of the higher frequency components. As this bias will reduce propensity for spikes in the interpolation, this choice is reasonable for most scenarios. Another potential issue is when the gap edge points are extreme values of the process. In this situation you will find the interpolation will have either an incorrect central tendency or the have a distinct linear trend that is not apparent in the actual process. This issues is magnified for larger gaps where the initial interpolation will have more effect of the resulting synthesis. In situations where the gap edge values are dubious or the gap is large, we recommend using the mean of the series instead of a linear interpolation.

7 Cutoff Determination

When synthesizing data from a time series we need to to make a choice for the significance level of the periodic components to be used. This choice can drastically change the estimate we produce from the data. If we set the level low we will accept more frequencies producing a more turbulent estimate which can cause undesirable spikes in the data. Likewise if we set the significance level too high, the estimate will miss much of the period structure in the time series.

Finding the sweet spot for the significance level is dependent on the data as well as the problem at hand. As interpolation and prediction are two separate problems with differing methods and end goals, it is reasonable to assume the optimal significance level will not be always the same. To identify the best choice for both cases we will approach them separately.

For an interpolation problem, we are aiming to fill in a section of missing data with the information contained in the data on either side. It would therefore make sense to choose the optimal significance level for the set surrounding data. It also makes sense that the best choice would be chosen with respect to the gap size we plan to interpolate. To meet these aims we propose a cross validation based method, where you divide the data on each side of the gap into bins the size of the gap and interpolate the now missing data.

We denote the data on either side of the gap $X_r(t)\&X_l(t)$ and the size of the gap as n. For a set significance level α we do the following:

1. Starting with $X_r(t)$, we divide the data into bins of size n. In the likely event that the length of $X_r(t)$ does not evenly divide into bins of size n, truncate $X_r(t)$.
2. Replace one bin, $X_r((n-1)i+1)\dots X_r(ni)$ with a linear interpolation.
3. Compute an F-test and determine the significant frequencies.
4. Use Thomson's Method to produce an estimate of the data, $\hat{X}_r(t)$
5. Calculate the mean squared error of the estimate to the removed bin, $MSE_{r,i}(\alpha) = \frac{\sum_{t=(n-1)i+1}^{ni}(X_r(t)-\hat{X}_r(t))^2}{n}$.
6. Repeat steps 2 through 5 for each bin.
7. Find the mean across all bins, $MSE_r(\alpha)$.

We now repeat this process for $X_l(t)$ and take the average across both to obtain an overall measurement of interpolation error, $MSE(\alpha)$ This process is computed for a range of values of α with the minimum error producing level chosen, $\alpha_{opt} = \underset{\alpha}{\operatorname{argmin}}\ MSE(\alpha)$.

In the event that the gap in the data is larger that the size of the two adjacent portions, this method will not work. We recommend using either the prediction method below. If there are multiple gaps start with he smallest gap and then use this interpolated data as real data for larger gaps, this will ensure that you have the largest data series possible for the larger gaps.

For prediction problems, our main goal is to ensure the optimal set of periodic components for predicting an interval of new data. The two important parameters that will affect the significance level chosen are the length of the interval we want to predict, n_p, and the length of the data we intend to use to create this prediction, n_t. We assume $n_p < n_t$ as it considered unwise to attempt to predict a larger time series than the sample used for modeling.

To develop a value of the prediction error associated with each significance level, we split all the available data into overlapping bins of size $n = n_p + n_t$. The amount of overlap required is dependent on the amount of data we have, ideally we would like a minimum of 10 bins. The process now follows similarly to the one for interpolation, with the main difference being that for each bin we replace the last n_p data points with zeros instead of lineary interpolating.

For a data set $x_(t)$ and a set significance level α we do the following:

1. Split the data into overlapping bins of size $n = n_p + n_t$.
2. For bin i, $X_i(t)$, we replace the last n_p data points of $x_i(t)$ with zeros.
3. Compute an F-test and determine the significant frequencies.
4. Use Thomson's Method to produce an estimate of the data, $\hat{X}_i(t)$
5. Calculate the mean squared error of the estimate to the predicted times, $PSE_{l,i}(\alpha) = \frac{\sum_{t=n_t+1}^{n}(X_i(t)-\hat{X}_i(t))^2}{n_p}$.
6. Repeat steps 2 through 4 for but now replace the first n_p data points with zeros and calculate the mean squared error of the estimate to the predicted times, $PSE_{r,i}(\alpha) = \frac{\sum_{t=1}^{n_p}(X_i(t)-\hat{X}_i(t))^2}{n_p}$.

7. Find the mean error across both predictions for bin i, $PSE_i(\alpha) = \frac{PSE_{r,i}(\alpha)+PSE_{/,i}(\alpha)}{2}$.

8. Find the mean across all bins, $PSE(\alpha)$.

This will give you a metric for evaluating the performance of the prediction model for a given prediction size and training interval. The assumption of stationarity is vital to the use of multiple bins with the expectation that the prediction error of times far from the times to be as useful as those close to the latest times. If there is some concern about the stationarity of the process then the use of a weighted mean with greater weights for bins with more current times is recommended. In the event you intend to use all of the data for your prediction, $n_t = n$, it is advantageous to use as large a size n_t you can while maintaining 10 bins with no more that 50% overlap to ensure there is not over training in the significance level chosen.

[add section on bootstrap f-test option]

8 Boosting Residual Signals

In some situations all of the significant periodic components will not be utilized in the estimated data. This will result in temporally correlated residuals. If we were able to model the residuals similar to the original model we may be able to improve our synthesized data. To encapsulate this extra periodic information we use a gradient boosting method. By fitting periodic components to the residuals using Thomson method we can identify missing periodic components of our process. We then optimize our model with the new components by finding the minimum squared error for the linear combination of the old model and residuals $M_{i-1}(t) + \gamma_i F_i(t)$, where $M_i(t) = \sum_{j=1}^{i} \gamma_j F_j(t)$ and $F_j(t)$ is the model on the $(j-1)^{th}$ residual series. $F_1(t)$ is the original model found on the data. When concerned with over-fitting we can add a cross validation step in here and choose the gamma that minimizes the mean of the cross validated squared errors.

We then check whether this new model is significant compared to our old model. To do so we perform an F-test to compare the two models. Under the null hypothesis that there is no significant improvement in the fit of the boosted model to the data we would have our statistic,

$$F = \frac{\frac{SSE_{old} - SSEnew}{\#NewFrequencies}}{\frac{SSEnew}{GapSize - \#TotalFrequencies}}, \tag{15}$$

which should follow an $F_{\#NewFrequencies, GapSize-\#TotalFrequencies}$ distribution.
The algorithm follows,

1. For a time series, y_t, we find the optimum Thomson estimate, $\hat{y}_t$, as described previously.

2. We next find the residuals for these estimates, $r_t = y_t - \hat{y}_t$.

3. With these residuals we test to see if the model is significant using the F-statistic described above. If significant we accept these estimates as our non-boosted model for the time series.

4. Next, treat the residuals, r_t, as a new time series and find the optimum Thomson estimate, $\hat{r}_t$.

5. With this new estimate we now make a "greedy" model by finding $\underset{\gamma}{\operatorname{argmin}} \sum_{t \in T}(y_t - (\hat{y}_t + \gamma \hat{r}_t))^2$, where T is the set of times we are using to create the interpolation or prediction with. If there is concern of over-fitting, in this step use cross validation to identify the optimal γ.

6. We now define our new estimates of this greedy model, $\hat{y}'_t = \hat{y}_t + \gamma \hat{r}_t$ and test for significance with an F-test.

7. If the model is significant we repeat steps 4 and 5 with the new residuals, $r'_t = y_t - \hat{y}'_t$, as the time series.

8. We continue repeating steps 4 and 5 with the new residuals until the model is not considered significant under the F-test. At this point we will consider the model from the previous iteration as our final model.

9 Bootstrapped Signal Synthesis

So far in our data synthesis we have negated the effects of randomness on our estimates. For each data point we assume that there is a set of periodic components and a noise term. This noise term will bias our estimates of the periodic components. There are several ways to attempt the deal with this issue, we can use noise removal procedures by filtering the data or using data transformations like Principal Components Analysis [**?**] but these too can be detrimental to our estimates. In an effort to produce a more meaningful estimate of the missing data we will use a non-parametric bootstrapping method to estimate the role noise plays.

We start by assuming our time series data is sampled from a process made of periodic components and a noise term, $x_t = \sum_{i=1}^{n} P_i(t) + \zeta(t)$. As well we assume the samples are uncorrelated. If the samples are correlated but no new periodic components can removed by using boosting on the residuals, it may be best to model the data with the periodic components, an ARMA process and noise component. To gain an estimate of the distribution the synthesized data will follow we need to model the noise process, $\zeta(t)$. To do so, we estimate the periodic components using of Thomson's method and re-sample the residuals, $r_t = x_t - \hat{x}_t$.

We now introduce noise into our initial estimates of the missing data by drawing samples from the residuals for the data points we wish to synthesize, interpolate or predicted, we emulate the noise contribution to the periodic estimates for those times. For interpolation, we would add noise samples to the linear interpolation and for prediction we would replace the zero padding with noise estimates.

By repeating this process of drawing noise samples and obtaining Thomson Estimates, we produce a set of random samples for each estimated time. From these samples we can derive properties of the distribution for the synthesized data. This gives us an estimate of the mean value for each time as well as confidence intervals.

Along with the confidence intervals we get for our estimates from the bootstrapping, we can also estimate the overall confidence intervals for the synthetic data by modeling the residuals by a normal distribution and adding the confidence intervals for the noise and synthetic terms together. With this we can give a distinct interval at each time that we expect the missing data to have been found.

10 Data Analysis and Comparison

We decided to check the merit of the these techniques we should first evaluate their performance on fabricated data made of sinusoids in noise and then test the methods on common data sets from physical sciences. We are interested in knowing how well the optimized Thomson inverse estimates are at synthesizing unknown data and under what conditions do these methods perform best.

First we tested the methods with an artificial time series containing 7 sinusoids in Gaussian noise with signal to noise varying from .15 to .3,

$$y_t = .2\sin(2\pi.2t) + .3\sin(2\pi.35t) + .25\sin(2\pi.135t) + .3\sin(2\pi.305t) + .28\sin(2\pi.25t) + .15\sin(2\pi.05t) + .27\sin(2\pi.1t) + N(0,1). \tag{16}$$

Wanting to see how the Thomson inverse method works on a well behaved stationary series, we ran through the full set of procedures described here.

Beginning by attempting to interpolate 100 points in the middle of 1200 points from this series, we start by determining the optimal cutoff value by finding the minimum mean squared error across the range from .5 to 1. The minimum value was found to be .986. We notice the mean squared error behaves as we would expect for this series, with low cutoffs, between .5 and .9 performing poorly due to the inclusion of many falsely detected periodic components. The performance of the estimates improves as we continue to remove needless frequencies, minimizing on the optimal set and as we continue to remove more frequencies the performance diminishes due to the removed significant periodic components.

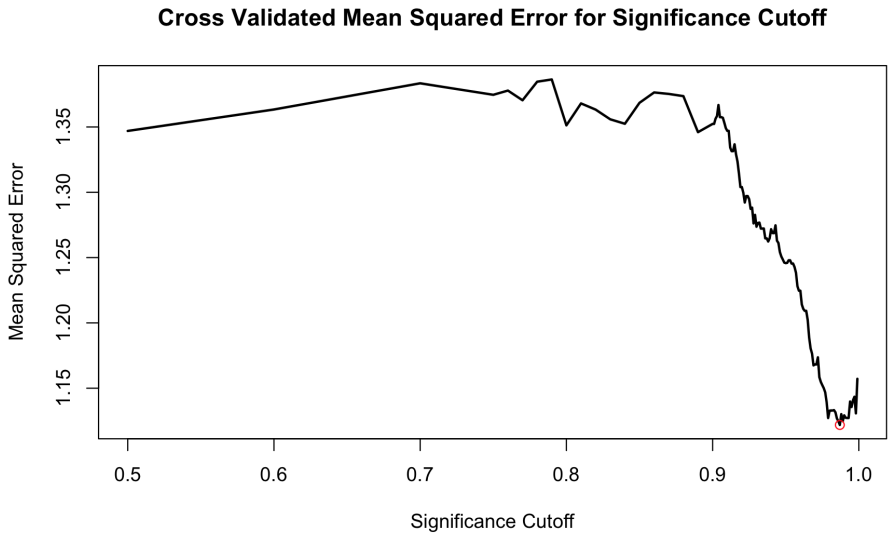

Figure 1: Cross-Validated Mean Squared Error of Varying Significance Levels for Interpolating 100 points of Sinusoidal Data.

Now with the optimal cutoff found, we perform our interpolation. The interpolation looks to perform well in the gap, with less variance than the true data, as we would expect with the noise levels present. Attempting to boost the residuals we found that no significant improvement could be made to the estimates by adding more periodic terms. This made logical sense to us with the simplicity of the structure of the data.

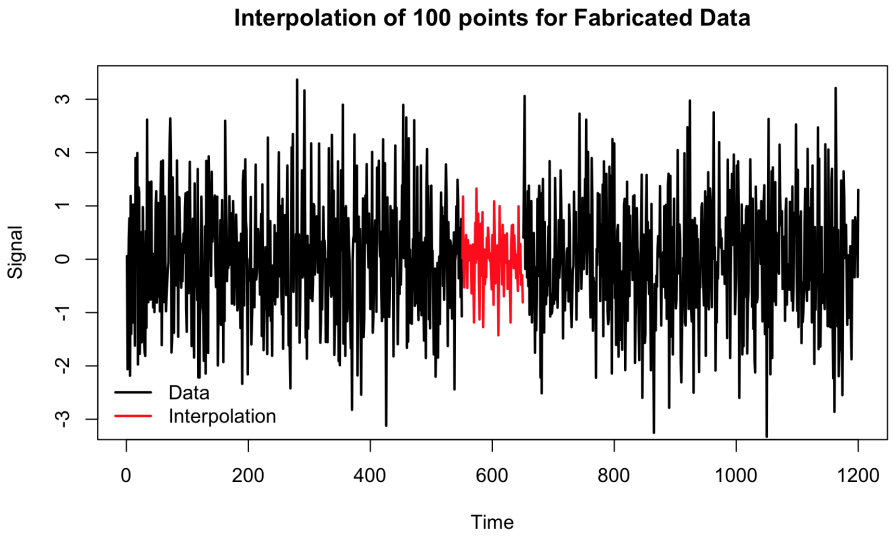

Figure 2: Interpolation of 100 points of Sinusoidal Data.

Finally sampling the residuals from the estimated series to add to the mean interpolation used gave us an estimate of the confidence intervals for the periodic error that may exist in the interpolation. Using the residuals to model the assumed Gaussian noise for our series we also obtain overall confidence intervals for the interpolated section. For this series the bootstrapped confidence intervals on the periodic interpolation are extremely small as this data is strongly periodic with little residual structure or noise interference. The overall confidence bounds appear appropriate, using $1 - \frac{1}{gapsize}$ as our significance level, we see the range to match that of the surrounding data.

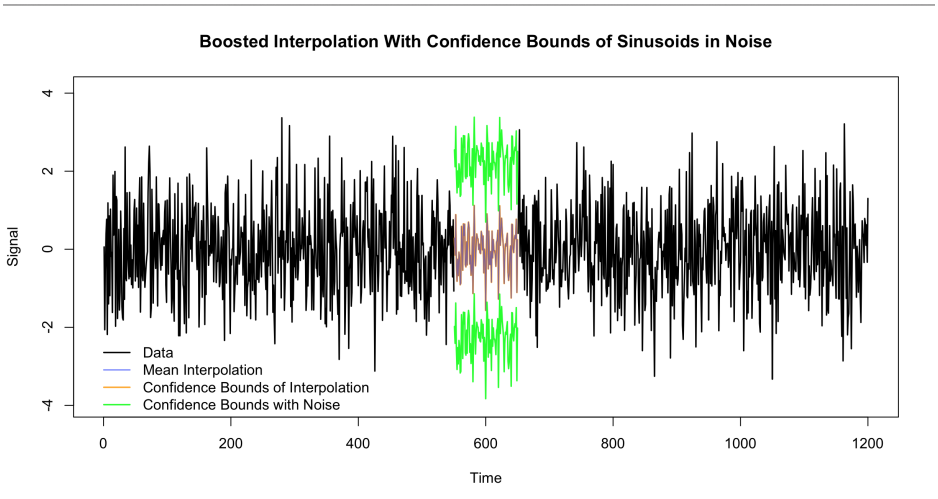

Figure 3: Interpolation of 100 points of Sinusoidal Data.

Comparing the Confidence region of the interpolation to the data removed from the series originally we see that only 4 points exceed our bounds and barely in all cases. For this fabricated data these methods perform quite well and give a useful representation of what could be found in the missing gap.

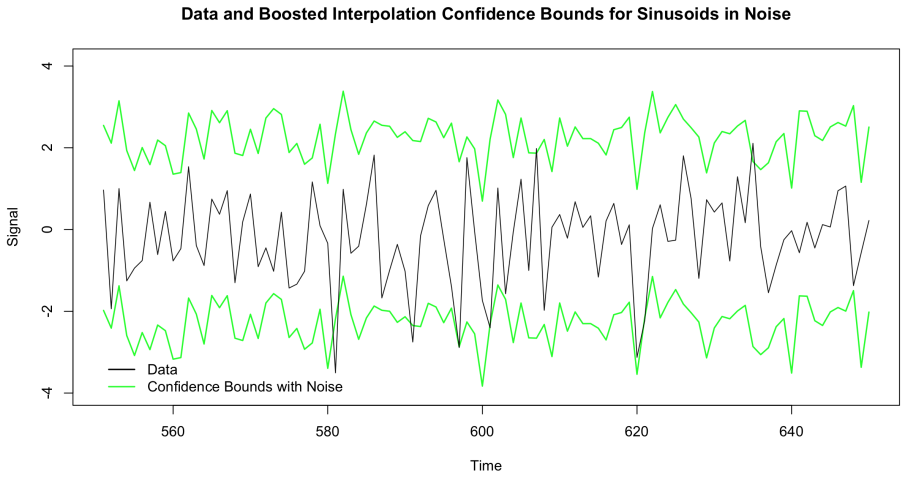

Figure 4: Interpolation of 100 points of Sinusoidal Data.

11 Conclusions on Techniques

This is awesome/the end

References

[1] M. Bayram and R. Baraniuk. Multiple window time-frequency analysis. In *Time-Frequency and Time-Scale Analysis, 1996., Proceedings of the IEEE-SP International Symposium on*, pages 173–176. IEEE, 1996.

[2] J. Pohlkamp-Hartt. The sphericity test for parameter selection for multitaper spectral estimation. 2014.

[3] D. Slepian and H.O. Pollack. Prolate spheroidal wave functions, fourier analysis and uncertainty - I. *Bell System Technical Journal*, 40(1):43–64, 1961.

[4] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(09):1055–1096, 1982.