**Using Adaptive Frequency Granger Causality To Infer Cell Topology in the Circadian Clock**

V. Esli Tovar1, Iavor V. Dekov1 and Stephanie R. Taylor1

1Department of Computer Science

Colby College

Waterville, ME 04901

*Abstract*

The suprachiasmatic nucleus (SCN) of the mammalian brain controls sleep cycles, digestive activity, physical activity, hormone levels and more. The SCN has about 20,000 neurons and scientists have tried to map out how they are connected using different methods. The Granger Causality method has been traditionally used to infer connections in other systems. However, Granger Causality is highly reliant on the assumption that time series are stationary, thus it is not suitable for inferring connections in the SCN because gene expression data from the circadian system has a highly non-stationary nature. Recently, computational biologists have adapted the Granger Causality method to account for this property of gene expression data, creating the Adaptive Frequency Granger Causality (AFGC) method. In the original publication, AFGC was tested on synthetic data only. We have implemented AFGC and tested it on 649 SCN cells to determine how well this new method performs on real data. Our results suggest that AFGC may not be suitable for these data.

**Introduction**

Inferring the topology of cell networks is a growing field in the computational biology community.

Our goal was to use Adaptive Frequency Granger Causality (AFGC) to infer one-way direct connections using time series data from the SCN.

This was a difficult task considering that there are no well established methods to infer directional connections in non-stationary oscillatory time series. Although, there are techniques that are commonly used for inferring networks from time series, like Granger Causality. AFGC is an attempt to adapt this technique to infer directional connections in non-stationary oscillatory time series data. AFGC determines whether or not there is a correlation between two cells using an F-value. When an F-value is higher than a predetermined threshold it indicates that one of the cells in question is influencing the other.

AFGC, to our knowledge, has only been used on synthesized data.

For our project, we implemented the method and tested it on real data from the SCN.

**Granger Causality**

Granger Causality is a statistical hypothesis test used to find the correlation between time series to assess how much they affect each other. For example, given two sets of time series data, Granger Causality attempts to determine whether one time series is likely to influence another. When granger causality is high, it means that one cell heavily influences the other.

Macintosh HD:Users:ivdekov:Downloads:documents-export-2016-05-03:highGrangerCausality.eps*Figure 1: Example of high Granger Causality in simulated periodic time series data.*

Macintosh HD:Users:ivdekov:Downloads:documents-export-2016-05-03:lowGrangerCausality.eps

*Figure 2: Example of low Granger Causality in simulated periodic time series data.*

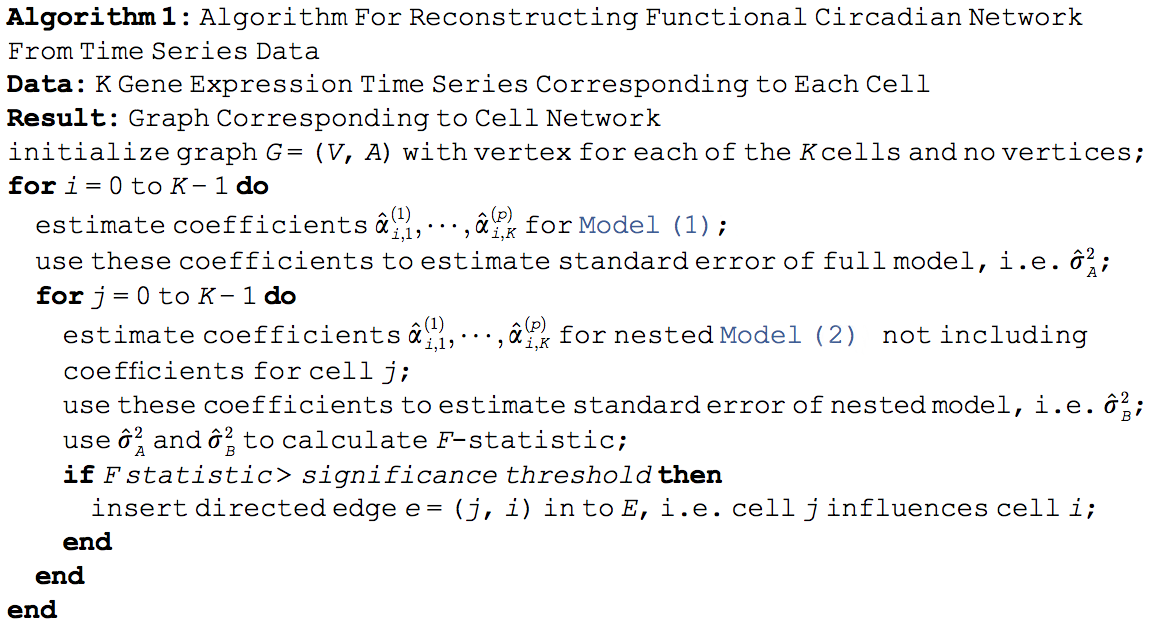
The limitation of Granger Causality is that it is highly reliant on the assumption that time series are stationary (that the data has a constant period), but gene expression data from the circadian system has a highly non-stationary nature. Therefore, AFGC was created.

**Adaptive Frequency Granger Causality**

The goal of the AFGC is to reconstruct the connections of an entire network of cells or in other words figure out which cells influence each cell in the network.

In the outer for loop, the coefficients of model 1 are estimated for each cell in the data. Then, the coefficients are used to estimate the standard error for the model 1. Model 1 is the full model which tests how well the i-th time series is predicted from the whole set of series.

In the inner for loop, the coefficients are estimated once more and the standard error of the model 2 is calculated for each cell in the data except cell j. Model 2 is the nested model that tests how well the i-th time series is predicted from the set series leaving out the j-th time series. These two standard error values are used to calculate the F-statistic, which is the standard error of the nested model divided by the standard error of the full model. If the F-statistic is higher than the preset threshold, then cell j influences cell i.

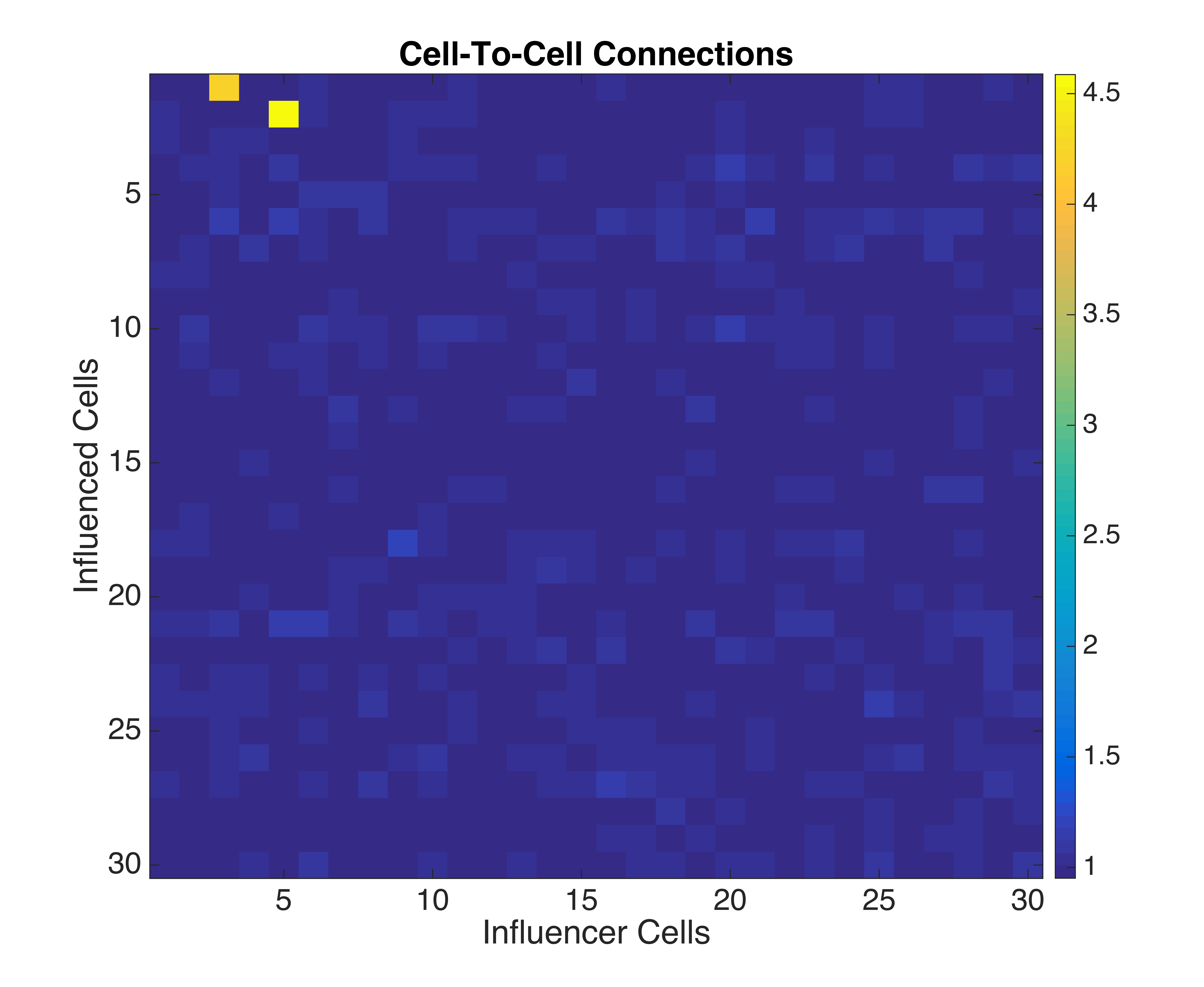


*Figure 3: Adaptive Frequency Granger Causality Algorithm*

What makes AFGC work for non-stationary data is that the method does not make any assumptions about the time series (ie. it does not pre-suppose that the time series are sinusoidal or even that they are periodic).

**Results**

We tested a MATLAB implementation of the algorithm above on synthesized data. This data was a collection of 30 sine wave time series for 200 steps with added noise. In two pairs of these sine waves, one time series was made to be a copy of the other but shifted to the right a few steps and with added noise so that the two time series were not exactly the same (see figure 1 for a demonstration of this). In this way, a simulated connection between these two time series (which were used to represent oscillating cells) was created. After running the AFGC algorithm on the synthesized data described above, we plotted an image of the F-values for each influencer cell that was removed in the inner for-loop of the algorithm. The image below displays the F-values as scaled colors. A high F-value is represented by a more yellow color.



*Figure 4: Results of AFGC on synthesized data. 2 high F-values (representing connections) produced.*

After running AFGC on the base data, we created the graph below, which illustrates the F-values through color.

**Macintosh HD:Users:ivdekov:Downloads:documents-export-2016-05-03:CellConnections.eps**

*Figure 5 : A color map of the F-values obtained by running AFGC on the base data collected from the SCN. Color represents F-value.*

In order to determine whether the algorithm worked effectively on the data, we plotted 49 graphs with time series data for the most influential cells, cells with the highest F-values, and those they influenced the most.

**Conclusions**

One can tell that the algorithm did not work effectively on the real data by looking at the graphs below: specifically the graph in the 1st row and 1st column as well as the graph in the 5th row and 2nd column.

Macintosh HD:Users:ivdekov:Downloads:documents-export-2016-05-03:mostInfluentialCells.eps

*Figure 5: Time series data of the 49 pairs of cells that generated the highest F-values.*

After inspecting the data we concluded that AFGC is not the best method for the data we used. There is no consistency among the 49 graphs above. We believe this is because AFGC is reliant on data that is sampled often, which documents the small changes over time. The data we used was sampled every hour versus every few minutes. As a result, we cannot assess how good AFGC is at predicting directional connections in non-stationary oscillatory time series like that of the SCN.

**Acknowledgements**

We would like to express our gratitude to Stephanie Taylor, who provided constant support and guidance regarding this project throughout the semester**.**

**References**

Pourzanjani A, Herzog ED, Petzold LR (2015)

On the Inference of Functional Circadian Networks Using Granger Causality. *PLoS ONE* 10(9): e0137540. doi:10.1371/journal.pone.0137540