Analysis of Minute and Hour Data for Weather Stations on Southern Vancouver Island.

Matthew Wine
University of Victoria
Physics 411

Abstract— **Using methods learned in Phys 411: Time Series Analysis, weather data from the Greater Victoria region was analyzed to discern different patterns contributing to the overall weather and climate of the region.**

I. INTRODUCTION

Greater Victoria is located on the southern end of Vancouver Island, on the west coast of Canada. To the south is the Strait of Juan de Fuca, which is the main water way connecting up to the Pacific Ocean. To the east is the Haro Strait. The Gulf islands are located to the north and to the east. Most of the region is relatively flat, however once west of Metchosin and Colwood, the terrain becomes much more mountainous. This geography of the region plays an important roll in shaping the local climate. Global weather patterns also have major affects on the local climate. The Westerlie Prevailing winds blowing off the Pacific Ocean interact with the geography of the region heavily influencing the climate.

Temperature data was collected from a series of weather stations throughout the Greater Victoria region. There were two data collections: minute data collected from four different locations, and hour data collected from 34 different locations. By analyzing this data with different time series techniques, how different effects of on the climate could be seen. The minute and hour data will be discussed equally, however some results will be seen to emerge for both sets of data, while others can only be seen from certain techniques used on one of the data sets.

Fig. 1. Greater Victoria Municipalities

TABLE I
LOCATIONS OF MINUTE DATA STATIONS

	Latitude	Longitude
Deep Cove	48.6801	-123.4591
James Bay	48.4128	-123.3814
John Muir	48.3707	-123.7516
UVic	48.4629	-123.3138

II. MINUTE DATA

For the minute data, temperatures were collected from four different schools: Deep Cove (DC), James Bay (JB), John Muir (JM), and UVic (UV). Deep Cove is located at the northern tip of the Saanich peninsula. James Bay is located at the south west end o Victoria, near the entrance to Victoria harbour. John Muir is located in Sooke on near the south west coast of the island. UVic is from The University of Victoria on the east coast of the island in Saanich East.

A. Removing Non-Values

Before the data could be analyzed, the non-values (nans) had to be accounted for. It was not simply enough to remove them from the time series, as this would distort the frequencies present in the data. Because there were relatively small sections of data missing compared to the total length, a linear interpolation was used to fill in the missing values. For the hour data which was missing more data a different method had to be used, but for the minute data this worked well to fill in the missing points. The time series for the different stations can be seen in figure 2. At first glance they appear very similar, and appear to follow the same trend over time, as well as having similar spikes. In order to understand these features as well as unveil more information, many different types of analysis had to be performed.

B. Basic Statistics

The simplest was to look at the data, is to look at the mean, standard deviation, skew, and kurtosis. The values of these for each station can be seen in table II. In order to help visualize the data, it was helpful to plot a PDF of the temperatures as seen in figure 3. Looking at the PDF it is clear that John Muir stands out from the others. It has the

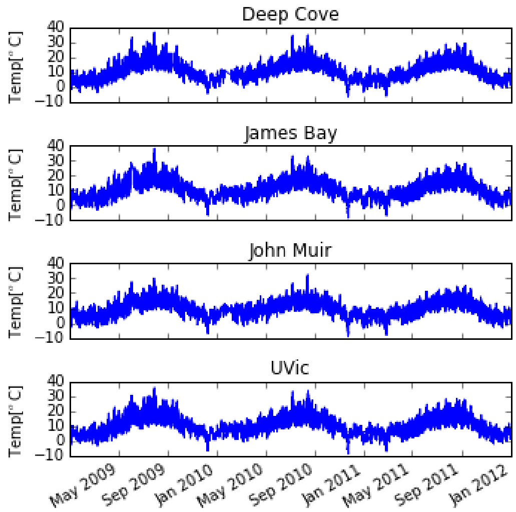

Fig. 2. Minute Data Time Series

lowest mean value, the lowest standard deviation, and unlike the other three has very little skew to it. The main cause for this difference most likely derives from the fact that John Muir, located in Sooke, is on the east coast of the island and more exposed to the weather coming off of the Pacific Ocean due to the Westerly Prevailing Winds. The res of the stations are shielded from these winds by the mountainous terrain to the west. These winds help moderate the temperatures keeping down larger variations seen in the other stations.

TABLE II
STATISTICS OF DIFFERENT SITES

	Mean	Std. Dev.	Skew	Kurtosis
DC	10.302	5.981	0.444	0.133
JB	10.311	5.837	0.480	0.229
JM	9.1120	4.836	0.064	0.141
UV	10.237	5.701	0.491	0.288

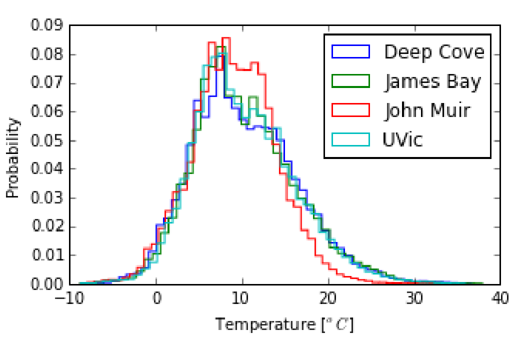

Fig. 3. PDf of Temperatures

C. Frequency Space

In order to determine the different frequencies composing the different time series, the discrete power spectra were calculated and plotted. This was done in two different ways. First this was done with no hanning window to observe the lower frequencies, then with a 50% hanning window to observe the high frequencies. These revealed multiple different peaks observable in all the stations. The obvious peaks showed up at one year and 24 hours, however, there were also very strong peaks at 12, 8, and 6 hours corresponding to integer multiples of the 24 hour frequency. These peaks can be see in figure 4.

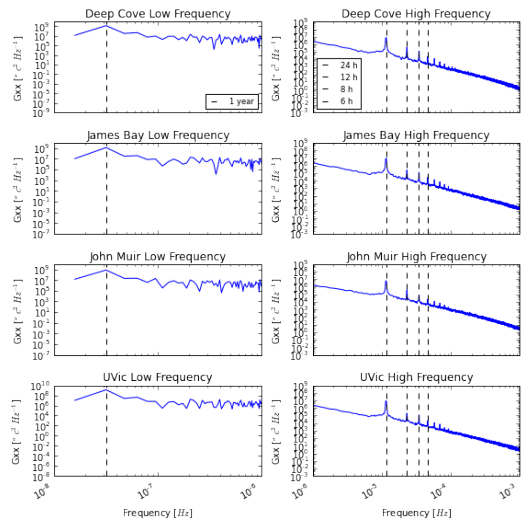

Fig. 4. Power Spectra of Minute Data Stations

D. Seasonal Variance and Filtering

The next thing that was to be investigated was how the temperature variance changes from summer to winter. To do this a band-pass filter was used to pick out the daily frequency and suppress everything else. When plotting the new time series of only the daily frequency, it can easily be seen that the temperatures have a larger variance during the summer than during the winter (see figure 5). This is because summer days offer more heating than winter so the change from day to night is more pronounced during the summer. We will see this again when looking at the hourly data later on.

E. Co-Variance and Lag Correlation

Next the co-variance between the different stations was investigated. This was plotted as seen in figure 6. It can be seen that at very low frequencies and the 12 hour frequency all stations have a high co-variance. However, at all other frequencies, the co-variance drops significantly, with the exception of between James Bay and UVic. The cause the the higher co-variance between these two stations has to do with there closer location to each other and the fact that they are both positioned close to Haro Strait which has similar effects on their climate. John Muir is much farther west, located net to the Strait of Juan de Fuca and Deep Cove is farther north and located next to the Saanich Inlet. While James Bay and UVic are roughly he same distance away from Deep Cove as John Muir, they both show higher co-variance with Deep

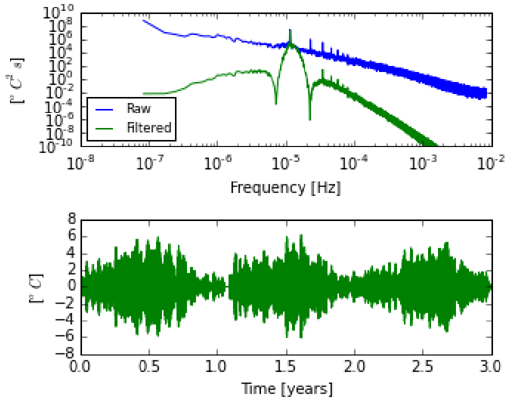

Fig. 5. (a) Filtering of the 24 hour frequency (b) Time Series of the 24 hour frequency

Cove than John Muir. This would be due to John Muir being on the west coast while Deep Cove is in a similar location to JB and UV on the east coast of the island, where they are protected from winds b the mountains. It can also be seen that the lowest co-variance is between John Muir and Deep Cove, which makes sense as they are the farthest apart and in very different regions.

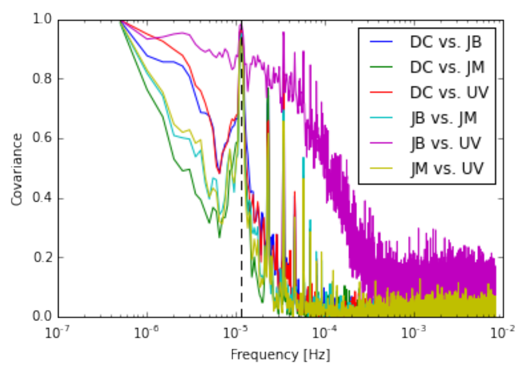

Fig. 6. Co-Variance Between Minute Data Stations

In order to determine the direction of the moving weather front, the lag correlation between different stations was calculated(figure 7). This was done between James Bay and Deep Cove to look at the North-South direction, and between James Bay and John Muir to look at the East-West direction. It was also done at different periods of the year to see how it changed with the seasons. Looking at the different plots, it can be seen that there is little variation between the different times of year. this appears to indicate that the weather front moves at about the same rate through out the year. When looking at the lag correlation between James Bay and Deep Cove, it can be seen that there is roughly a 30 min lag between the two. The distance between the two stations being $30.3\ km$. This indicates that there is motion of the weather front moves very slowly south to north, at about $1\ \frac{km}{min}$. When looking at the lag correlation between James bay and John Muir, the lag is much smaller, about 8 minute in summer and only 1 minute in winter. The distance between the two stations is $27.7\ km$. From this it was determined that the weather front moved west to east at a rate of $3.5\ \frac{km}{min}$ in the summer and $28\ \frac{km}{min}$ in the winter. The distances between the stations were calculated using the Haversine formula:

$$h = hav\left(\frac{d}{R}\right) = hav(\phi_2 - \phi_1) + cos(\phi_1) \cdot cos(\phi_2) \cdot hav(\lambda_2 - \lambda_1)$$

$$d = R \cdot hav^{-1}(h)$$

where R is the radius of the earth, and d is the distance between the points, ϕ and λ are the latitude and longitude respectively, measured in radians, and hav is the haversine function defined as:

$$hav(\theta) = sin^2\left(\frac{\theta}{2}\right) = \frac{1 - cos(\theta)}{2}$$

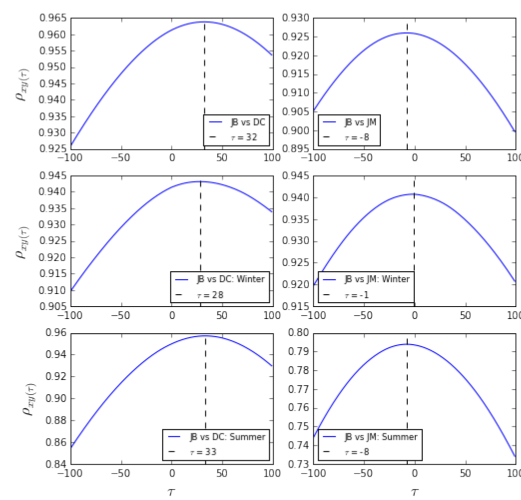

Fig. 7. Lag Correlation Between Minute Data Stations

F. *Minute Data Summery*

A multitude of different methods were used to analyze the data collected from the 4 different weather stations located around Greater Victoria.

Simply looking at the mean, standard deviation, skew and kurtosis showed how different the the eastern side of the island had a wider variation in climate, while the eastern side was more mild and consistent. Plotting the power spectrum for the different stations showed the main frequencies at the one year and 24 hour mark, but there were also peaks seen at integer values of the 24 hour frequency. Filtering the signal to only show the 24 hour frequency showed that the summer months had a higher range of temperature changes over the course of a day. Looking at Co-Variance showed a higher correlation between James Bay and UVic, while the rest only showed a strong correlation at the daily frequency. The Lag correlation showed the weather front moved from west to east, and moved significantly faster during the winter.

III. HOUR DATA

The next section used hour data collected for 34 locations located around southern Vancouver Island. Most of the stations were missing sections of data,and some had to be discarded due to missing too much such as the station in figure 8. The locations of the stations as well as which ones were used can be seen in figure 9.

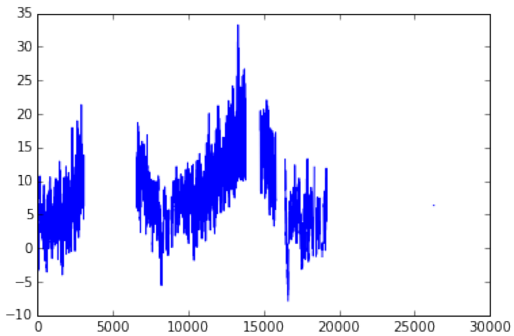

Fig. 8. Example of a Removed Time Series

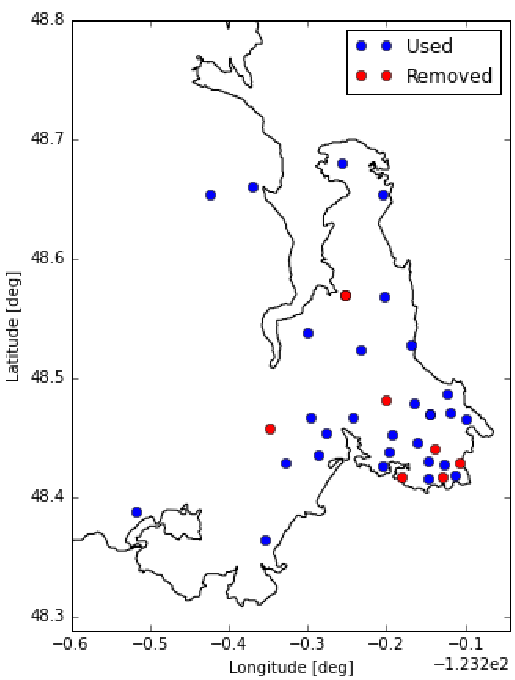

Fig. 9. Map of Stations on Vancouver Island

A. Removing Non-Values

The removal of the nans from the hour data was not as simple as for the minute data, due to larger sections of data missing from some of the stations. to remove these nans, a new time series was made of the average temperatures over all available stations. However after doing this there were still around 140 hours where all the stations were down. To fill in these last few point a linear interpolation was used. Any nan values in the different series just used the values from the average temperature time series. The figures 10 and 11 below show one of the time series before and after replacing the nans.

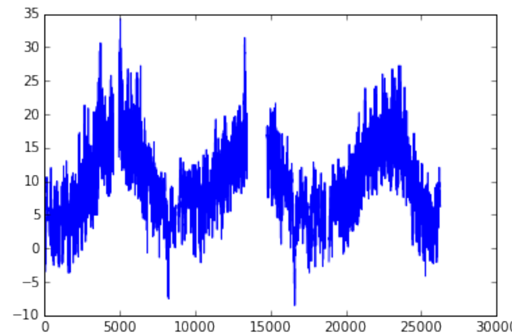

Fig. 10. A Time Series before replacement of nan values

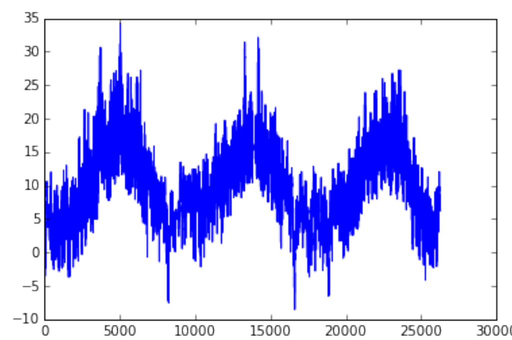

Fig. 11. A Time Series after replacement of nan values

B. Temperature Maps

The time series at each station were then used to extrapolate to the rest of the island. This was done using a overlaying a 1000x1000 grid over the map and interpolating for each point on the grid using a cubic interpolation to fill in the values. Figure 12 below shows the average temperature over the entire time series. While this particular plot is not the most informative, there are many important things that can be learned from it. First it can be seen that the highest annual average temperature is around Saanich West, and the lowest are appear located in the Highlands municipality, around the Saanich Inlet. Another important observation is the lack of data points between Sooke, Colewood, and the Malahat make analysis of this region difficult, as these more inland regions temperatures are being extrapolated from far away stations located much closer to the coast. In order to obtain more information, the standard deviation was also plotted as seen in fig 13. From this plot an important feature can be seen. The stations along the coast have a much lower standard deviation than the inland stations. From this it can be determined that along the coast there is less variation in the yearly temperatures. These plots only give a very broad idea of the weather. In order to obtain more detail, different seasons had to be looked at. For this the winter and summer seasons were looked at separately.

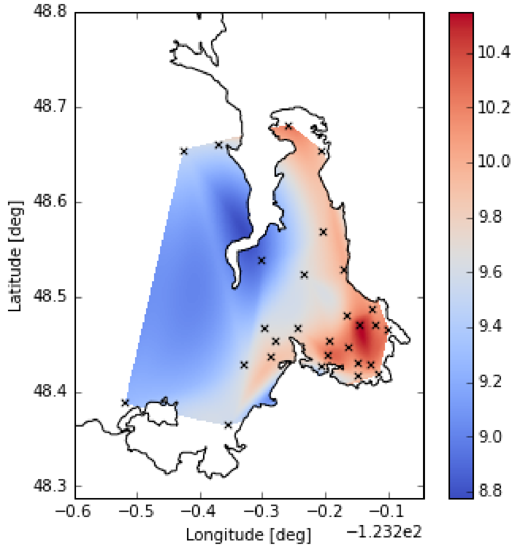

Fig. 12. Mean Temperature: Full Time Series

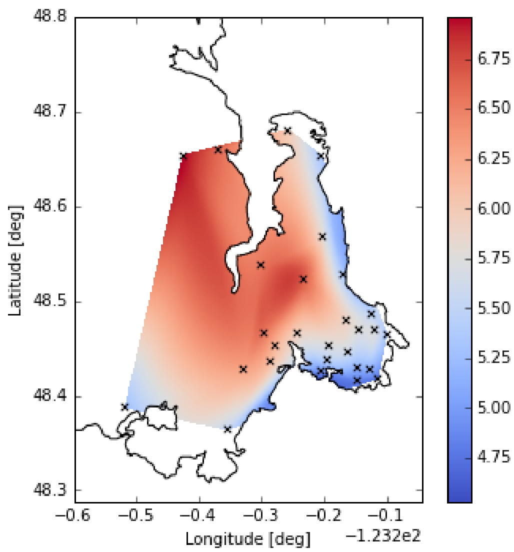

Fig. 13. Standard Deviation: Full Time Series

Looking at the winter mean temperatures it can be seen that along the coast the temperature is higher than farther inland, the highest and lowest temperatures being seen in similar locations as in the annual mean. When looking at the summer data it looks very different. The inland regions have a higher mean than the coastal regions. The lowest summer temperatures appear in Sooke and the southern end of Victoria West. The differences between summer and winter become very apparent when looking at these plots. During summer the inland regions are higher than the coastal regions, while the opposite is true during the winter.

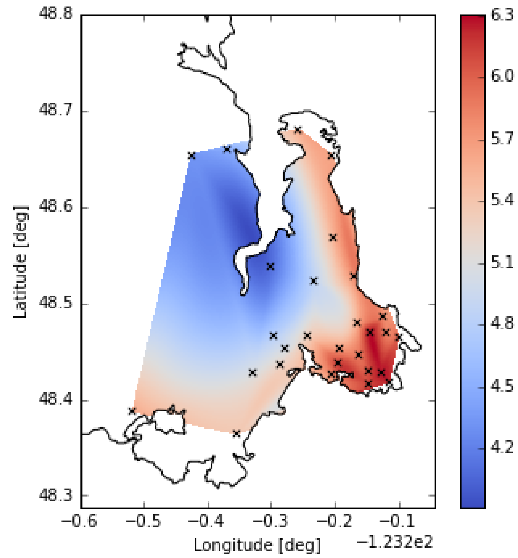

Fig. 14. Mean Temperature: Winter

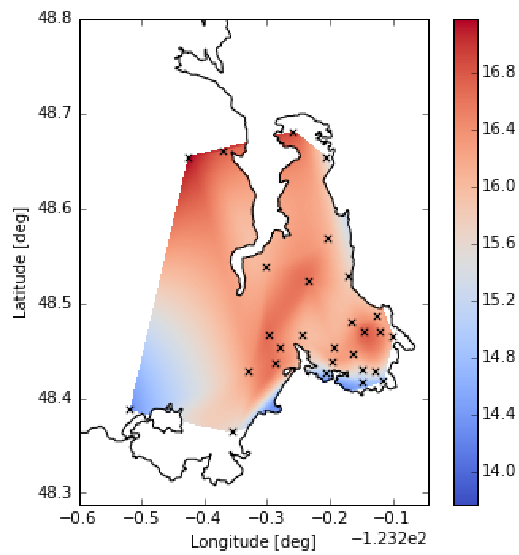

Fig. 15. Mean Temperature: Summer

When looking at the standard deviation for summer and winter, a few more important features become apparent. First it is clear both plots agree this the annual standard deviation plot in showing the standard deviation is lower on the coast than farther inland. However it is also apparent that during winter the standard deviation is much lower than in the summer. This agrees with what was seen in the minute data when lo0king at the 24 hour frequency; that the winter and less variance in day night temperature cycle. This is again the same effect being seen as in the minute data. The increased heating during the summer heats the ground during the day increasing daily temperatures. The low specific heat capacity of the ground means it cannot hold on to that heat very well and it cools off quickly during the night resulting in large changes in day night temperatures. During winter, when the ground receives less heating during the day, there is less change from day to night. The reason for the more mild climates near the coast is due to the higher specific heat capacity of water. The oceans act as giant heat sinks, absorbing large amounts of incoming solar radiation while not changing drastically in temperature like the ground. This helps keep the temperatures more mild during the day. during the night, all the energy the water has absorbed can be released, keeping the temperature from dropping as drastically as it would inland away from the water.

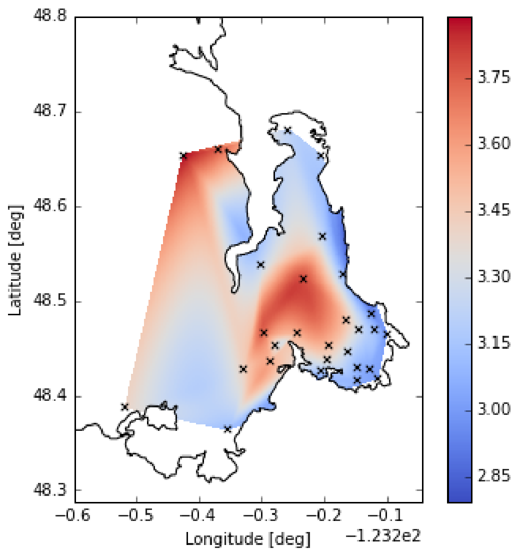

Fig. 16. Standard Deviation: Winter

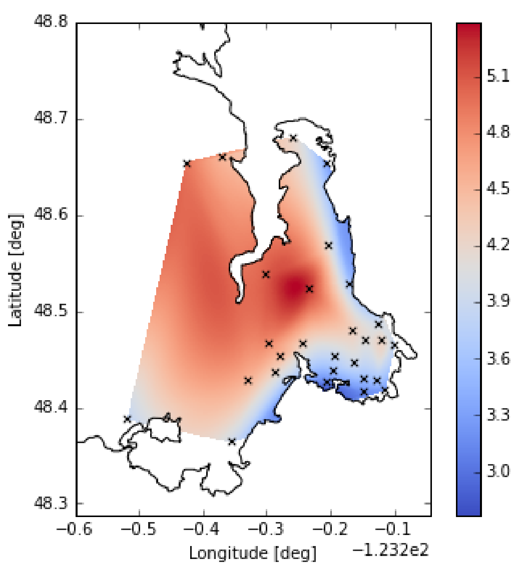

Fig. 17. Standard Deviation: Summer

C. Spectrogram

In the previous section it was discovered that the temperature time series has specific features that change though out the year. This means they are non-stationary time series. This means that through out the year there will be different frequencies present in the signal. In order to see what frequencies are present at different times in the year a spectorgram of the time series had to be made. To do this the time series was divided into a few hundred sections, each partially overlapping. The power spectrum of all these was then calculated, as seen in figure 19. It is noticeable from this plot that all sections have a strong correlation at he 24 hour frequency, but everywhere else are in far less agreement. To see the differences between the sections the spectrogram in figure 18 had to be examined. The spectorgram had many different lines which appear more prominent at different times. these different lines correspond to the frequencies of integer multiples of the 24 hour peak. The first peak corresponding to the 2 hour peak can be seen to be strong across the entire time series, which makes sense as the day night cycle is prevalent year-round. The 12 hour frequency is the next line up. This frequency was seen to be more prominent during the winter months than during the summer. The rest of the frequencies are more prevalent during the summer and weaker during winter.

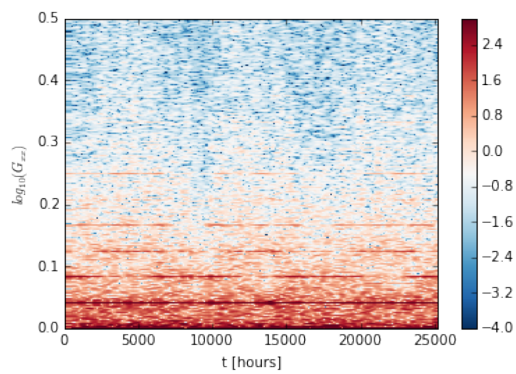

Fig. 18. Spectrogram of Hour Data Time Series

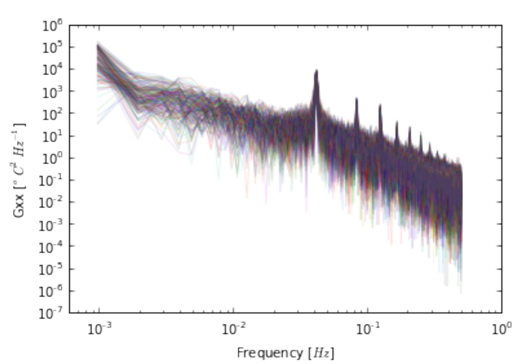

Fig. 19. Spectrogram Frequencies

D. Hour data Summery

The methods used to analyze the hour data revealed many different findings and helped support those found from the minute data. There was a lot of information found from the temperature maps. They showed how inland regions have higher summer temperatures and lower winter temperatures than coastal areas. It also showed that inland regions have a higher variance from day to night than coastal areas. Looking at the spectorgram for the hour data it was seen that the 24 hour frequency was prominent throughout the entire time series, the 12 hour frequency was only prominent during winters, and any subsequent peaks were only prominent during summer.

IV. CONCLUSIONS

The analysis of the both the minute and hour data revealed a lot of information about the local climate of the greater Victoria region. It was seen that the coastal regions had the most moderate climates, both in terms of daily and annual periods. However the west coast was seen to have the most mild climate, due to its exposure to the Westerlie winds. It was seen that daily variations in temperature were greater in summer than in winter, due to a larger variation in heating from day to night. It was seen that John Muir had the least co-variance with any of the other stations while, James Bay and UVic had the highest co-variance. By looking at the lag correlation between stations it was seen that the weather front moved west to east, and the speed at which it moved was greatly increased during the winter compared to the summer. Mapping the weather data for different seasons showed the difference between coastal and inland regions. the inland areas tended to experience more extreme temperatures in winter and summer, as well as having greater variation from day to night. Using the data to create a spectrogram, it was seen that different frequencies were present in different seasons. the 12 hour frequency being more prominent in winter, while the 8 and 6 hour frequencies were more prominent in summer.