Wavelets for Atmospheric Gas Analysis

by

Nicholas Murdoch

A research paper
presented to the University of Waterloo
in partial fulfillment of the requirements for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2010
Supervised by John C. Lin, Ph.D.

© Nicholas Murdoch 2010

I hereby declare that I am the sole author of this research paper. This is a true copy of the research paper, including any required final revisions, as accepted by my examiners.

I understand that my research paper may be made electronically available to the public.

Abstract

Atmospheric CO_2, CH_4, and CO are measured in a variety of locations across Canada. The wavelet transform is used to find points of interest, such as the daily carbon cycle, as well as longer term oscillations. Wavelet coherence finds patterns linking the sequences around 7 to 10 day oscillations, mostly during the summer. The daily CO_2 cycle is stronger in more southern and urban locales, indicating that it is linked to vegetation and possibly human activity. CH_4 is also much more present in the urban environment. The qualities of the wavelet transform are discussed, primarily in reference to turbulent data. Selection of a proper mother wavelet is discussed.

Acknowledgements

I would like to thank my supervisor, John C. Lin, Ph.D., for providing a topic to me, and to Anthony L. Endres, Ph.D., for reading my paper.

Contents

List of Figures

Chapter 1

Introduction

In studying large datasets, one is often faced with the problem of extracting a small set of significant parameters controlling the generation of the data. Looking more specifically at a stochastic process, such as the mixing of gases in the Earth's atmosphere, the properties of the apparently random noise carry information. In order to render the data useful, these parameters should be related to known or measurable physical or numerical properties, so that their effects can be estimated for areas of limited knowledge, such as future forecasts, or locations where data acquisition is more limited.

Computers provide an avenue for studying these problems. Operations can be carried out on the large datasets, which transform them through various methods to give a quantitative look at some of the underlying patterns present. If multiple variables are available, correlations between them can be measured. Due to the speed of modern computers, these calculations can be carried out quickly and accurately, allowing the researcher to test hypotheses concerning these large datasets.

This paper will focus on one such technique - the wavelet transform - and its application to records of the presence of particular chemicals in the Earth's atmosphere.

1.1 The Atmosphere

The Earth's atmosphere is immeasurably complex - on the order of 10^{44} atoms reacting, colliding, and absorbing or emitting energy among one another endlessly. Chemical processes on the Earth's surface add or subtract molecules and energy from the atmosphere, in a vast number of ways. Radiative energy from the sun is added to the system, while energy is radiated from the atmosphere itself, back into space. Making sense of all this is challenging. A common simplification is to regard the atmosphere as a stochastic process -

by averaging over large numbers of molecules, the expected outcome becomes a probability distribution, and there are techniques for measuring and using these results.

On a human scale, the weather and climate affect where humans may live in comfort, how they perceive the world around them, and the decisions they make daily. In turn, human and other biological activities play a part in the evolution of the atmosphere. The human sense of smell, for example, is reliant on local atmospheric currents and mixing to convey the molecules representing different scents from their source to the discerning observer. The image of a rainbow depends on the scattering of light from microscopic water droplets suspended in the atmosphere. In turn, when people burn fuels, they add to the mix of the atmosphere.

As a result of the atmosphere's complexity, there are a wide variety of methods for obtaining data concerning the atmosphere. Tools measuring temperature, chemical composition, the passage or obstruction of electromagnetic radiation, and so forth, allow humans to peer into the properties of the atmosphere and the mechanisms governing it. The eventual goal is to be able to predict, and thus control, the way humans impact the atmosphere they live in.

This paper will focus on greenhouse gases, carbon dioxide and methane in particular. These are the gases capable of absorbing and emitting radiation in the infrared or thermal wavelengths. As a result, they contribute to the temperature of the atmosphere, making them of interest. The fractional presences of these gases, in parts per million, or parts per billion, are commonly measured, so large datasets are available. Carbon monoxide will also be discussed, though it is only indirectly related to the greenhouse effect.

1.2 Wavelet Analysis

The concept of integral transforms goes back to the 18th century, in work by Fourier and others. In general, an integral transform can be expressed as follows:

$$Tf(\omega) = \int_{t_1}^{t_2} K(t, \omega) f(t) dt \tag{1.1}$$

By multiplying the original function, f, by a kernel function, K, and integrating, a new function, Tf, is produced. Depending on the properties of the kernel function chosen, the output function may be a unique representation of the data within a new domain on the variable ω. This can lead to simplifications in expression of complicated functions, if the correct kernel is chosen.

The wavelet transform is a relatively new integral transform, having been developed by Morlet and Grossman in the early 1980s [2], and constitutes a set of criteria which the

kernel function must satisfy. The intention of the wavelet transform is to represent the function in both frequency and spatial domains, such as position or time, simultaneously. Where the original function is entirely in the spatial domain, and the Fourier transformed function is entirely in the frequency domain, a wavelet transform combines both. This makes it possible to see not only which frequencies dominate, but where in space or time they occur.

In discrete form, the integral becomes a sum of terms, and can be applied to measured data. Much like the Fourier transform, the structure of the coefficients makes a fast algorithm for the calculation of the discrete wavelet transform possible.

These properties make the wavelet transform useful in applications where oscillations may exist only temporarily, or have varying frequency over time - quite common in natural stochastic processes.

Once two or more wavelet transforms have been calculated for different sequences, it is also useful to be able to compare them quantitatively, to see where they show similar strong oscillations. In literature this is often called wavelet coherence, referring to the strength of the interference between the two sequences when they are treated as waves. Identifying common oscillations can indicate that a single process is governing those oscillations, or that they are otherwise linked.

Chapter 2

Wavelet Analysis

The intention of this chapter is to provide an introduction to wavelets, first in general, and then with a focus on the features relevant to this research. The next chapter further narrows the discussion to specific techniques, after the data has been introduced. A series of diagrams illustrating the terms and properties of the wavelet accompanies the text to assist in explanation. The majority of these diagrams have been prepared using R [10], a free statistical software environment, Wavethresh [8], a wavelet analysis package for R, and SOWAS [6] [7], a wavelet spectral analysis package for R.

2.1 The Wavelet Transform

The purpose of the wavelet transform is to move from the spatial domain into a combination spatial and frequency domain. Additionally, the number of non-zero coefficients should preferably be reduced. A wavelet is essentially a single wave, of a particular frequency, at a particular location in space. By changing the frequency and location, the set of basis functions comprising the wavelet transform are produced. The coefficient of a particular wavelet is higher if that wavelet more accurately represents the local properties of the original function - the goal is to represent the function accurately with a minimal number of wavelets.

Since there are two dimensions in the output, frequency and location, the function $K(t, \omega)$ from equation 1.1 can be rewritten using wavelet $\Psi_{j,k}(t)$, where j indicates the frequency and k indicates the location. Thus, equation 1.1 over a finite time becomes the wavelet transform:

$$\tilde{f}(j,k) = \int_0^1 \Psi_{j,k}(t) f(t) dt \tag{2.1}$$

When performing the wavelet transform, it is useful to refer to the value associated with each wavelet, $\tilde{f}(j,k)$, as wavelet coefficient d_{jk}, and to the set of such coefficients for all j,k as the wavelet spectrum.

In order to achieve this in a consistent and useful fashion, the wavelet should satisfy several properties. These properties restrict the space of all possible transform functions down to the ones conforming with the idea of a wavelet. The first three criteria are necessary, while the last two add useful features to the wavelet transform.

- *Admissibility* A valid wavelet function must have a mean of zero. This helps to ensure that only local frequency information is included. Together with *invertibility*, this property is needed to ensure the wavelet functions are orthogonal to one another.

- *Invertibility* It should be possible to reproduce the original function through some transform applied to the wavelet coefficients. Like *admissibility*, this ensures that all energy is preserved, and for operations to be performed on the wavelet spectrum, then transformed back to the original spatial domain.

- *Similarity* An important feature of the wavelet transform is that all the wavelets are similar. Aside from scaling and translation, each wavelet function used is identical. One "mother" wavelet can be used to generate every other wavelet function. Due to this, results can be compared across different frequencies and locations - the same value will hold the same meaning. Since higher frequency wavelets are narrower, the resolution of the results correspondingly becomes more local. The comparison between wavelets thus relates a number of wavelengths at a given frequency, rather than a similar distance in space. At longer wavelengths, the wave is being measured over a wider range.

- *Regularity* To be useful, the wavelet should be localized in the spatial as well as frequency domain. So when taking the Fourier transform, there are only a finite number of non-zero coefficients. This improves the utility of the transform. The Haar wavelet discussed later does not satisfy this property, and the result is a greater number of non-zero wavelet coefficients, making it a less useful transform.

- *Cancellation* In some cases, higher order moments of the wavelet should also be zero or near zero. This property is important in certain areas of turbulence study, in characterizing the nature of the stochastic process.

  - *Moment* The mathematical moment μ_k of a probability distribution is the expected value to some power: $\mu_k = E((X-\mu)^k)$. This can be applied to a set of wavelet coefficients to find moments at a particular frequency or location in space. The first moment is the mean, thus *admissibility*, but higher ones may also be useful.

2.1.1 The Mother Wavelet

Every process needs to start somewhere. The wavelet transform begins with the "mother" wavelet: a characteristic function from which every other wavelet is generated. Since this function will be convoluted with the data sequence, properties of both the original sequence and the mother wavelet can be seen in the results. Selection of an appropriate mother wavelet endows the results with the desired characteristics. A simple example is the Haar wavelet (Nason [8]). It contains discontinuities, so the regularity and cancellation properties are not possible, but it is easy to see that it can satisfy the admissibility, invertibility and similarity conditions.

$$\Psi(x) = \begin{cases} -0.5 & x \in [\text{-}1,\ 0) \\ 0.5 & x \in [0,\ 1) \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

Since the Haar wavelet is only non-zero on a finite domain, and has a basic square wave shape, it picks up local information about frequency. The Haar wavelet can be seen in Figure 2.2a.

The function to convert the "mother" wavelet into the daughter wavelets is as follows:

$$\Psi_{j,k}(x) = 2^{j/2}\Psi(2^j x - k) \tag{2.3}$$

For varying values of j, the wave is scaled to different lengths. Translation is handled by k. If j and k are limited to integral values, such as $j = 0, ..., J$, $k = 0, ..., 2^j - 1$ for each j, then an orthonormal basis is progressively formed with higher and higher accuracy as J increases. This progression can be seen in Figure 2.1. If the transform is being applied to a discrete sequence, the upper limit on J is reached with the Nyquist frequency - waves of length shorter than two data points are aliased away.

One commonly used mother wavelet, for example by Farge [2] and Furon et al. [4], is the Morlet wavelet, given by:

$$\Psi(x) = e^{ik_\Psi x} e^{-(|x|^2/2)} \tag{2.4}$$

The wavevector k_Ψ is used to adjust the wavelet to make it admissible - additionally, a small correction factor is necessary to make the average zero, but for $|k_\Psi| = 6$ it turns out this correction is smaller than computer rounding errors so it can be neglected. In form, the Morlet wavelet is a wave with wavenumber k_Ψ, windowed by a Gaussian of variance one, so it quickly decays to zero. The Morlet wavelet with $k_\Psi = 6$ is shown in Figure 2.2b. This wavelet matches oscillations well, and since it is complex, the phase and amplitude can be separated, letting us look at the energy present at any point.

Another fairly common complex wavelet, also from Farge [2], is the Paul mother wavelet:

$$\Psi_m(x) = \Gamma(m+1)\frac{i^m}{(1-ix)^{1+m}} \tag{2.5}$$

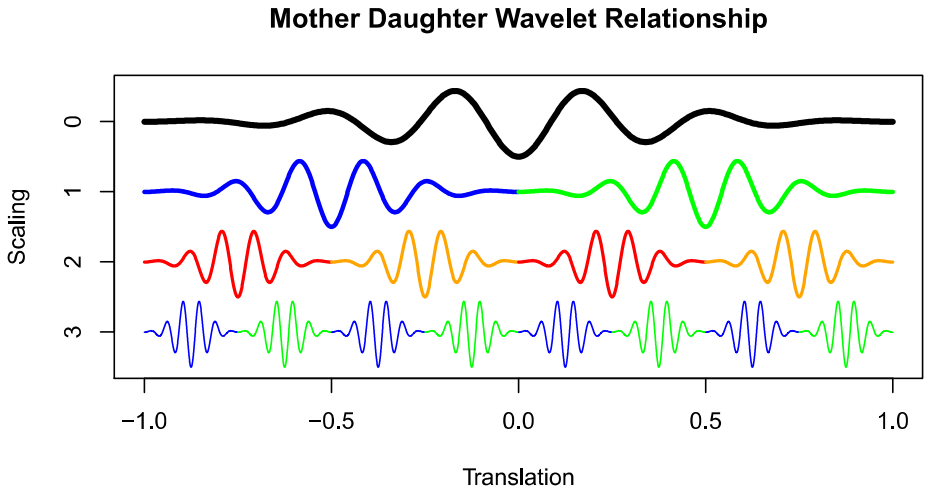

Figure 2.1: Cascade from mother wavelet to child wavelets. Mother wavelet is in solid black. Scale values of j lie along the vertical axis, while translations k are along the horizontal axis. This image uses the real component of the Morlet wavelet as an example.

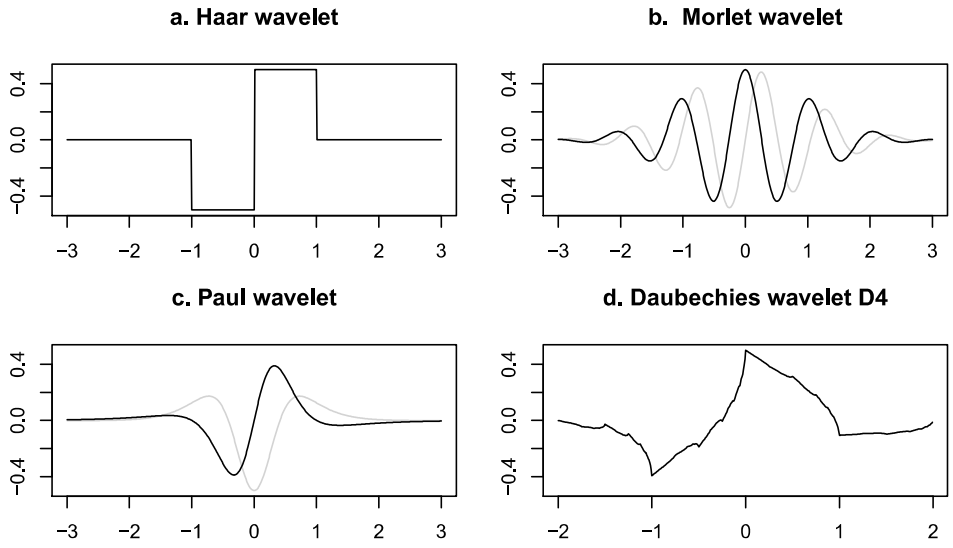

Figure 2.2: Four different mother wavelets. Black indicates real component, grey is imaginary. Morlet wavelet has $k_\Psi = 6$, Paul wavelet has $m = 4$.

This transform is mostly used in quantum mechanics, where certain operations can be performed in wavelet space. The value of m affects the number of vanishing moments, controlling the *cancellation* criterion. This wavelet is also used for analysis of music, since the wavelets are strictly forward in time and causal relationships are preserved [2]. An example of a Paul mother wavelet with $m = 4$ appears in Figure 2.2c. Because the Paul wavelet has fewer oscillations than the Morlet wavelet, it is less narrowly banded in the frequency domain - but sometimes the extra moments are useful.

An alternative method of selecting the mother wavelet, which only applies to discrete signals, is to start at the highest frequency with a set of coefficients, and then progress downward to lower frequencies iteratively. This leads to the Fast Wavelet Transform - which executes in time on the order of the number of coefficients being calculated, each coefficient only taking a small number of operations. One such set of mother wavelets is called the Daubechies family of wavelets, which are derived by maximizing the number of zero moments for a given number of initial points. So D2 is the Daubechies mother wavelet derived from 2 points, D4 uses 4 points, and so on. In practice, every increase of 2 points leads to 1 extra zero moment. An example, D4, is shown in Figure 2.2d. Unlike the Morlet and Paul wavelets, it is not smooth - but the practicality of fast calculations makes it useful in some areas.

For the remainder of this paper, the Morlet wavelet is used. The ability to isolate energy and phase, and the improved frequency resolution, makes it more useful for investigating data sets for objects of interest. Since continuous transforms will be performed, it is adequate in execution time.

2.1.2 Continuous vs. Discrete Transform

There are two distinct ways of selecting the scale and translation factors when generating wavelets. In the continuous case, all values, or at least closely spaced values, are used. In the discrete case, only integer scales and translations are applied.

These methods serve different purposes - the continuous transform provides information at every scale and frequency on the domain, so it is possible to estimate more accurately the exact location of any anomaly. As a result, the continuous transform is more suitable for exploratory work - locating and investigating interesting frequencies. In contrast, the discrete transform only includes frequencies that are powers of two - but since fewer coefficients are being calculated, it is much quicker. The discrete coefficients are also orthogonal with an appropriate mother wavelet - meaning data corresponds exactly to the wavelet coefficients. Since the continuous wavelets overlap, they are not orthogonal - making an inversion, or calculation of the total energy present, more difficult. Thus the discrete transform is more suited to problems where the data should be preserved, and an inverse

transform is desirable. For example, in compressing a signal, it is often possible to apply the wavelet transform and then use a threshold, recording only the largest coefficients. If the majority of the signal is represented by a few coefficients, this can be quite successful. On the other hand, when Gaussian noise is present in the signal, this adds equally to all the coefficients - so this advantage is somewhat lessened, because energy is distributed more evenly.

For the purposes of this paper, the continuous transform is used to explore the data available. The main intention is to find processes acting on the atmosphere in an oscillatory fashion, and the continuous transform fits this concept well. When new avenues of research open, though, it is important to weigh these benefits against each other and select the appropriate method.

2.2 Cross-Spectrum Analysis

Moving beyond single variable sequences, sometimes a method for directly comparing two different variables is wanted. The goal is to show areas where both signals have large coefficients. If done properly, this should reduce the amount of guesswork in identifying potential correlated frequencies and ranges of points.

The easiest way to perform this is to simply multiply together the absolute values of the coefficients at a particular location - so if both are large, the result is large, and can be spotted. Implementation is simple - multiply pointwise the matrices containing the coefficients. For two wavelet spectra $\tilde{f}^x$, $\tilde{f}^y$, with coefficients d^x_{jk}, d^y_{jk}, Furon et al. [4] construct the cross-wavelet transform $\tilde{f}^{xy}$, with coefficients:

$$d^{xy}_{jk} = d^x_{jk}\overline{d^y_{jk}} \tag{2.6}$$

Here the complex conjugate of the second term is used so the product is power. Unfortunately, this method runs into trouble - very high spikes in a single time sequence can still show up in the end result, even if the other sequence is smooth in that region. To prevent this, some sort of normalization needs to be introduced.

2.3 Wavelet Coherence

Wavelet coherence provides this normalization. After the cross-spectrum analysis has been completed, smoothed sequences are calculated - averaging over local values in space and frequency for both the original wavelet spectra and the cross-wavelet spectrum. Once the smoothed coefficients have been found, the coherence can be calculated. Thus the

significant points will have to rise above the background level in both sequences. For a smoothing operator S, the wavelet coherence, also from Furon [4], $R(j,k)$ can be used to identify points of interest.

$$R^2(j,k) = \frac{|S(j^{-1}d_{jk}^{xy})|^2}{S(j^{-1}|d_{jk}^{x}|^2)S(j^{-1}|d_{jk}^{y}|^2)} \tag{2.7}$$

Multiplying by the inverse of the scale, j^{-1}, allows the smoothing operator to be used equivalently at all scales - since it is smoothing in both location and scale. The result of this formula is a value between 0 and 1, which shows the relative local strength of the coherence.

The phase difference between the sequences can also be calculated, using $\theta(j,k) = \tan^{-1}(\mathrm{Im}(S(j^{-1}d_{jk}^{xy}))/\mathrm{Re}(S(j^{-1}d_{jk}^{xy})))$, working from the smoothed cross-wavelet value. The phase is useful in determining how the particular oscillations are related - whether they are in phase, or opposite phase, or one is lagged behind the other by some amount of time.

By setting a rigid structure on the coherence calculation, it is also possible to do a significance test - letting the computer automatically mark areas of potential interest. The SOWAS package is able to perform these tests and indicate the significant signal.

2.4 Identifying Features of Interest

Larger coefficients can indicate some oscillatory process - but they can also be artifacts of random noise, since stochastic processes are a common target of the wavelet transform. The more a coefficient stands out, and similar values in the surrounding coefficients, can help to determine which it is. The longer the duration of a high energy feature, the more likely it is to be an actual oscillation, and if it is an ongoing process, it will likely also appear at a similar frequency several times in the time series. Certain singular peaks, characteristic of turbulent motion, may also show up in particular ways in the wavelet spectrum.

An important consideration is the envelope of significant values. An individual point in the original sequence affects every wavelet it is a part of. This builds up to a wider cone of influence as the scale increases, and the wavelets being used get wider. The ends of the sequence take consideration - no values are present past some point, so the ends of the usable sequence curve inward at larger scales, eventually reaching the point where only a single long wavelet covers the entire domain. This is typically the scale the mother wavelet is at. It is also possible to perform the wavelet transform as if the data was cyclical - in this case the wavelet coefficients at the ends are being calculated using some of the beginning of the sequence and some of the end.

2.4.1 SOWAS Wavelet Power Spectrum

In the plots produced by SOWAS, the original horizontal axis of a time series is preserved. The vertical axis becomes the axis representing scale - with longer wavelengths and lower frequencies at the top, shorter scales at the bottom. The units displayed are time - so longer times mean lower frequencies. The magnitude of the wavelet coefficients is indicated by the colour of each pixel, following a legend appearing to the right of the plot. The cone of influence where no data repetition is needed is marked by a black line on the plot that starts at the bottom corners and curves up and inward at higher scales. Outside this the data is wrapped around as if the sequence were periodic.

A simple demonstrative example is the chirp - an oscillation that increases and subsequently decreases in frequency. One such signal can be seen in Figure 2.3. Since this signal is composed of a single wave of varying frequency, it shows up very clearly in the wavelet power spectrum seen in Figure 2.4. The curve of significant values in the diagram does a good job of describing the original signal - it is easy to read the low frequency at either end, rising to a high frequency in the middle and then dropping down again at the other end. With no noise present, this is an ideal representation of the chirp.

Adding in noise demonstrates the usefulness of the wavelet transform. The time sequence in Figure 2.5 has two sin waves masked by additive Gaussian noise - the Gaussian noise has a variance of 1.0, and the sin waves are scaled to magnitude 0.5 and 0.25. While it is possible to get an idea of where the higher magnitude sin wave appears in the time sequence, as some oscillations are visible around -200, the lower magnitude one is almost impossible to detect. But when the wavelet transform is applied, in Figure 2.6, not only is it possible to see both sin waves - but an estimate can be made to their wavelengths. One is around 128, the other is around 64. The spatial limits are more difficult to detect, due to noise obscuring particularly the lower energy oscillation, but it could be estimated that the longer wavelength signal is applied over the duration $[-300, 0]$ and the shorter wavelength appears around $-600, [-100, 200]$, and 700. This is located visually by finding the coefficients with largest magnitude on the plot. Knowing the general structure of the signal, that is, two different sin waves of constant frequency, helps with this. So it is useful to have some idea of what signals will be present, when the wavelet transform is used.

The underlying sequence used here was $y = N(0, 1) + 0.25 \cdot \sin(2\pi x/63)$, where $N(\cdot, \cdot)$ is the normal distribution. An additional $0.5 \cdot \sin(2\pi x/126)$ was added on the interval $[-400, 0]$. So the actual wavelengths are 63 and 126 - which the wavelet transform fairly accurately detects. The domains are more difficult to measure - the lower magnitude oscillation is obscured by the random noise. This is characteristic of the Morlet wavelet being used for the transform - frequency data is preserved more accurately than spatial data.

This tradeoff between frequency and spatial accuracy can be accounted for by the

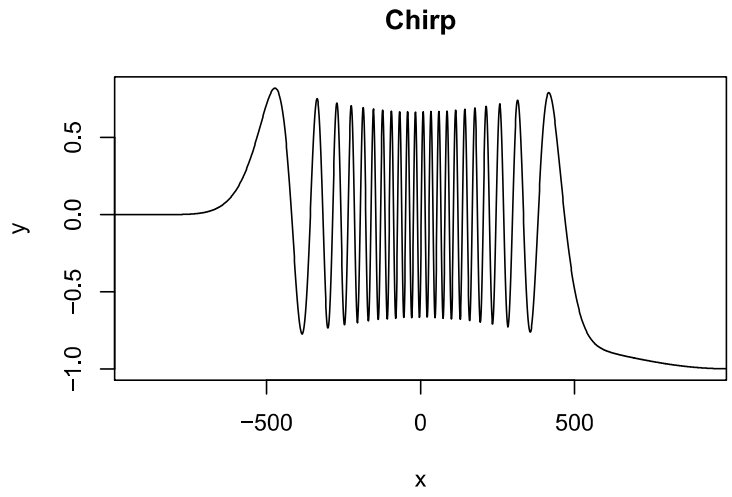

Figure 2.3: Chirp signal - oscillations at higher frequencies at the centre.

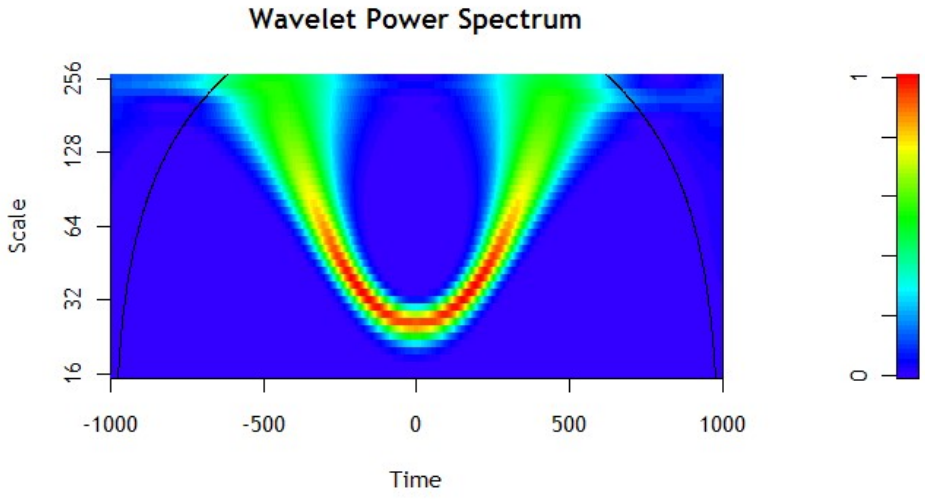

Figure 2.4: Wavelet Power Spectrum of the chirp in Figure 2.3

Heisenberg uncertainty principle - the narrower the spatial domain, and thus the more accurate the coefficient in space, the less well the frequency can be determined, since only a small fraction of the wave can be seen. The Morlet wavelet is wide enough to oscillate several times - making it good for isolating frequency information, but tending to blur more in the spatial domain.

2.4.2 SOWAS Wavelet Coherence

The wavelet coherence plots are presented in a similar fashion to the power spectra. As well as the linear scale, it is possible to use a logarithmic scale for the magnitude of coefficients. Due to the nature of turbulent energy at different frequencies, the logarithmic scale is more useful in practice, since turbulence follows a power law which becomes linear on the logarithmic scale. However, for the example, the regular scale is used because it functions adequately.

The coherence is calculated between two previous examples in Figure 2.7 - the chirp, and the noise overlaid with sin waves. The expected behaviour would be to see peaks where the sin waves cross similar frequencies in the chirp. Since the larger magnitude sin wave ends up inside the arc of the chirp, it barely registers. The lower magnitude sin wave crosses the entire domain, and its intersection can be seen at (-300, 64) and (300, 64) in the coherence plot.

This plot also shows a potential pitfall with coherence - in a completely noisy time sequence, after the coefficients have been smoothed out, there will still be some peaks. If the other time series has a local peak that is part of a signal, the two can combine to produce false positives. These are visible particularly around wavelengths 30 and 250. Since the chirp spectrum has higher values here, and the noisy spectrum is randomly large in the same spots, points with relatively good coherence appear.

The phase plot can be read in conjunction with the coherence to find the relative phases of the two signals. For example, at a wavelength of 256, the wavelets are in opposite phase. But as the wavelength decreases, the phases adjust and the two signals are in phase by a wavelength a little above 64. In the lower part of the plot, the phase is entirely random - the branching structure is an artifact of the continuous variation in frequency.

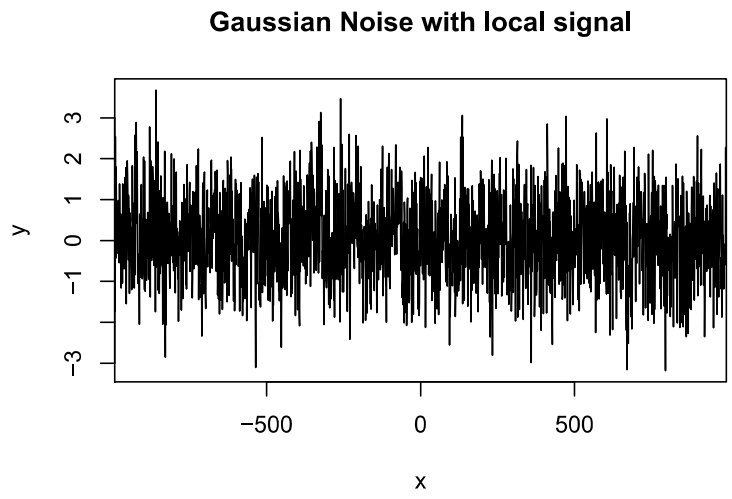

Figure 2.5: Noisy signal: Gaussian noise applied with variance 1 across the entire domain. Two sin functions added. The first spans -400 to 0, with a magnitude of 0.5. The second spans the entire domain, with a magnitude of 0.25. Neither is easy to detect by visual inspection.

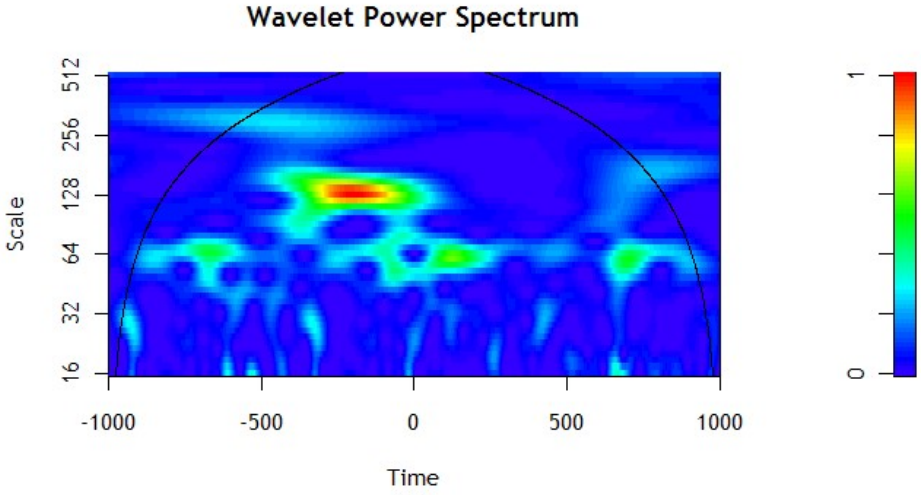

Figure 2.6: Wavelet Power Spectrum of the noisy signal in Figure 2.5. The two sin waves show up as higher values, near their respective wavelengths of 63 and 126. The Gaussian noise also adds random power spikes on the rest of the image.

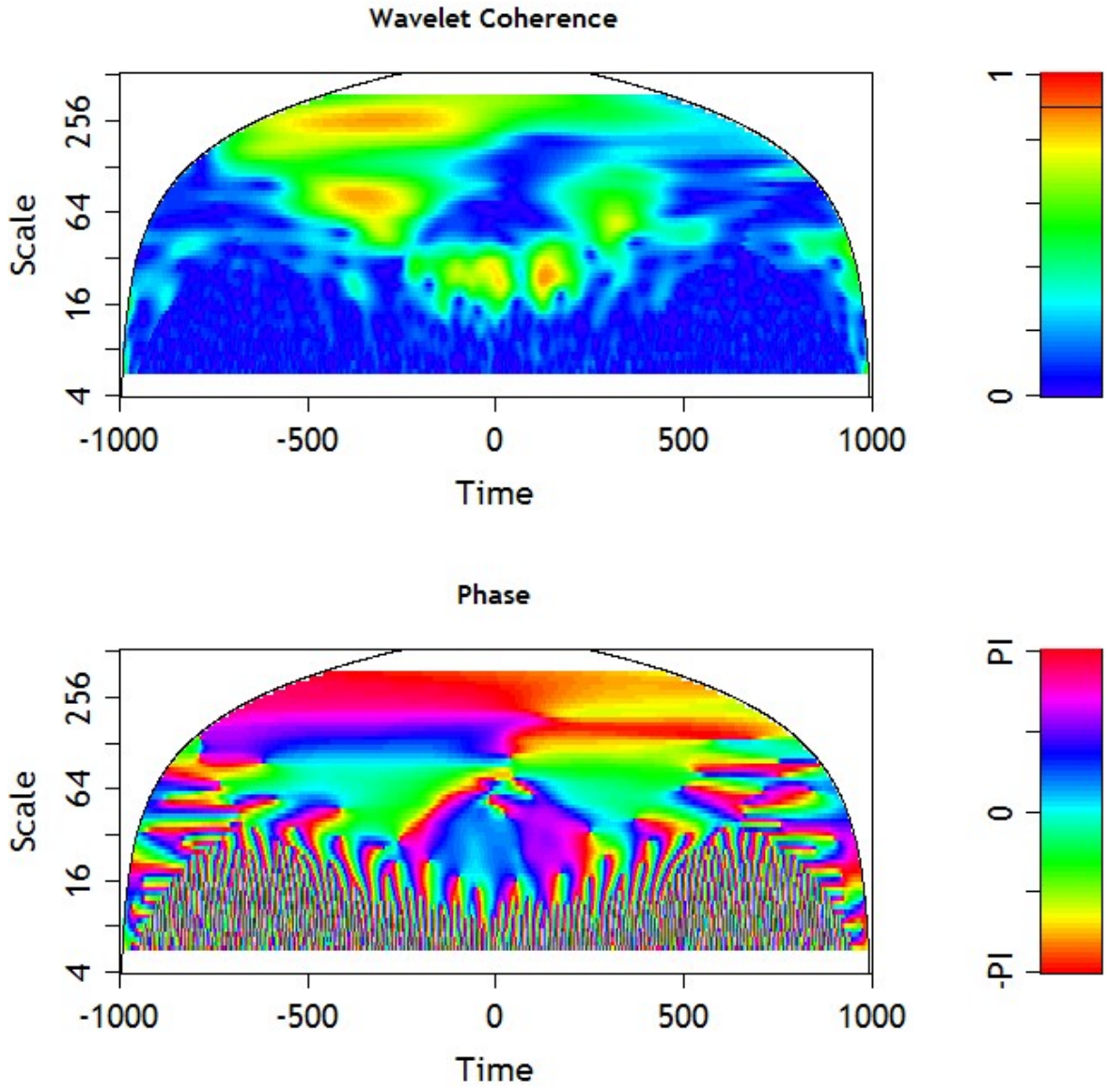

Figure 2.7: Wavelet Coherence between chirp (Figure 2.4) and noisy (Figure 2.6) wavelet spectra.

Chapter 3

Studying The Atmosphere

Now that the stage has been set, with the wavelet method introduced and explained, the data of interest can be brought into the picture. The atmosphere provides plenty of opportunity for study and numerical analysis - as the sheer number of processes and molecules interacting allows for new avenues of research. This chapter will introduce the species being studied, how the data was acquired, and then present the results of applying the wavelet transform.

A problem when studying the atmosphere is that there is very little control. It is nearly impossible to run the same experiment twice with only one variable changed - all the various properties are constantly interacting and reacting, and the best that can be expected is similar settings. This limits the ways processes can be studied - it is difficult to isolate them and get specific results. Looking at long-running datasets collected in a variety of distinct locations is one way to approach this problem. The long term average behaviours can be determined, and comparison between locations showing the same features can help find the source, or at least other measurements correlated with those features.

3.1 Turbulence

A major driving force in the complexity of the atmosphere is turbulence. In fluids possessing a large Reynolds number - comparing inertial and viscous forces acting on the fluid - the flows do not progress smoothly, and break down into chaotic whorls and eddies. This prevents deterministic flow calculations, and allows mixing between the particles in the fluid, which in turn allows chemical reactions to take place. In turbulence, energy is introduced at a large scale, where advection occurs, then cascades down through vortices of decreasing size until it reaches a scale where it can be dissipated by viscosity.

One of the first approaches that made it possible to study turbulence was by A.N. Kolmogorov. Kolmogorov [5] suggested that the energy present at frequencies between two length scales - external energy, and dissipation - should also become constant, cascading at a regular rate from one to the other according to a power law. He derived this power law from the Navier-Stokes equations and some assumptions about the way energy is transfered between different wavelengths. Kolmogorov's first assumption was that energy is transfered from wavenumbers k_1, k_2 to a third wavenumber k_3 if and only if $k_3 = k_1 + k_2$. As long as $k_1 \approx k_2$, then $k_3 \approx 2k_1$. If this applies at certain scales, then the process is self similar on those scales. At low wavenumbers, the process is affected by external energy - the lowest wavenumber at which this is no longer true, and energy is entirely moving within the system, is called k_{min}. At high wavenumbers, the energy is dissipated through viscosity - this higher wavenumber is called k_v. So there exists a range $k_{min} < k < k_v$ where the process should be entirely self similar, and thus can be governed by a power law.

The presence of this range depends on the Reynolds number being sufficiently large, and the boundary conditions of the system being appropriate such that energy is not moving in or out due to other mechanisms within this range of wavenumbers. This range is often called the inertial range.

The other assumption Kolmogorov made was that a single constant governed this process in all systems, leaving only the energy dissipation and wavenumber as contributors. In practice this assumption can be demonstrated to be incorrect, with other factors affecting the energy in the system, but it turns out that for a given single system the method is fine. Kolmogorov used dimensional analysis on the wavenumber k with units L^{-1}, and energy dissipation ϵ with units L^2T^{-3}, along with the overall energy density, with units L^3T^{-2}, to derive a relationship between the three numbers: $E \propto k^{-5/3}\epsilon^{2/3}$.

Since this energy cascade happens over frequencies, the natural response is to study it using the Fourier transform. This is where confirmation of Kolmogorov's power law comes from - observed power spectra fall away at the expected $k^{-5/3}$ rate. Because the wavelet transform also presents frequency information, it should verify the law. Unfortunately, after performing the wavelet transform, several features consistently appear in turbulent data which cannot be predicted by Kolmogorov's law - coherent structures that are larger and last longer than should be possible if the energy cascade is entirely homogeneous. According to Farge et al. [3], the reason these sort of features do not appear in the Fourier transform is that the spatial insensitivity of the Fourier transform produces ensemble averages, eliminating these coherent structures by blending them with the general energy in the signal. Since the wavelet transform is sensitive to spatial changes in the signal, these coherent structures become visible. These general properties of turbulence give us some idea what to expect when applying the wavelet transform.

3.2 Environment Canada Greenhouse Gas Network

Internationally, data is collated by the World Data Centre for Greenhouse Gases [1], concerning a variety of measurements related to greenhouse gas. These measurements are taken at stations located globally, with a variety of measurement devices for different purposes. A dozen of these stations are run by Environment Canada. The majority of these sites measure CO, CO_2, CH_4, N_2O, and SF_6 quantities, as well as isotopes of H_2 and CO_2. In particular, the CO_2, CO, and CH_4 have been measured hourly in Canada, at some stations for over two decades, so this paper will focus on these three gases.

3.2.1 Carbon Dioxide, Carbon Monoxide, and Methane

The most abundant greenhouse gases in the atmosphere, ordered by the magnitude of their radiative forcing, are water vapour, carbon dioxide, and methane. While the IPCC [11] lists carbon dioxide and methane in their charts of greenhouse gases, they ignore water vapour because it is not well mixed or present in the atmosphere over long periods. Water vapour condenses at varying rates due to atmospheric temperatures, leading to a range from about zero to ten percent of the atmosphere being water vapour by volume. Carbon monoxide reacts with other chemicals in the atmosphere to form CO_2 or CH_4, so its presence indirectly leads to greenhouse gases as well. Carbon monoxide is generally present for shorter times, due to these reactions, so its long term greenhouse effect is not often measured.

Different techniques are used to measure the partial volumes of the three different species. Gas chromatography involves feeding the sample through a tube to separate the species, then reaching a detector. A flame ionization detector (FID) is used to measure the presence of methane, as the FID is most sensitive to hydrocarbons, and insensitive to water, CO_2, and other gases in the atmosphere which do not contain carbon. Carbon monoxide can be catalyzed to produce methane, after which it can also be measured by the FID. In order to measure carbon dioxide, a nondispersive infrared sensor (NDIR) is used. This checks the obsorption at particular wavelengths in the infrared spectrum to determine what quantity of carbon dioxide is present.

3.2.2 Environment Canada Station Locations

Environment Canada's measurement stations for greenhouse gases are scattered across Canada. Their geographical locations can be seen in Figure 3.1. Additional information about the stations and measurement devices is found at GAWSIS [12].

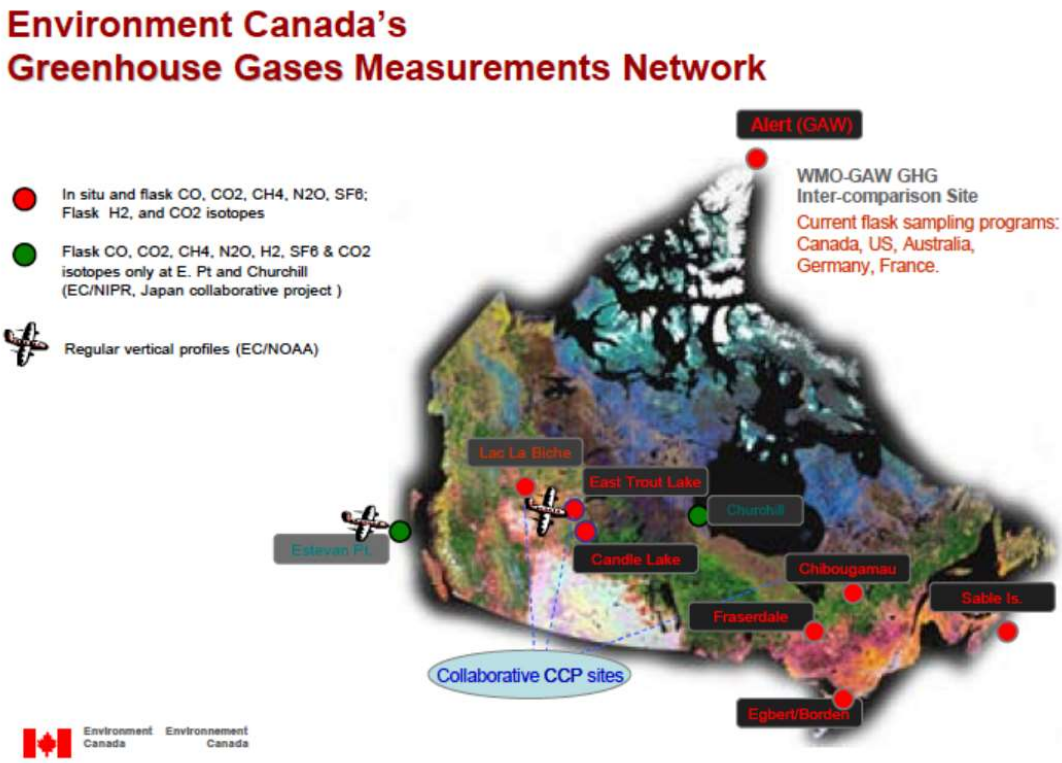

Figure 3.1: Environment Canada Greenhouse Gas Network station locations.

One concern with this data is that the sequences are not in general continuous - for various reasons there are frequent gaps in the data, when the acquisition devices were not running. This presents a problem for the wavelet transform which expects a continuous signal. These missing sections need to be handled appropriately so they do not affect the other measurements too much.

The Alert Background Air Pollution Monitoring Network Observatory has been operational the longest, with flask measurements dating back to 1975. CO_2 and CH_4 measurements have been made hourly since 1988, and CO measurements since 1998. Alert is located at the northeastern tip of Ellesmere Island, in Nunavut. Due to its location, there is very little vegetation or human activity nearby - so most gas presence has been transported from other areas. It is possible that in some seasons, this is traveling across the North Pole from Europe and Russia.

The Candle Lake site is located in the northern boreal forests of Saskatchewan. Measurements of gases have been made since 2002. Much like the Candle Lake station, the East Trout Lake observatory is located in northern Saskatchewan, at the southern edge of a boreal forest. Measurements here have been made since late 2005. In the summer western Canada is affected by forest fires, which may be visible in the results.

The Chibougamau station was established in 2007. It is located in the boreal forest in Quebec, with the intention of studying the effects of ongoing forestry activities in the region.

Egbert is in southern Ontario, roughly 80km north of Toronto. Measurements here have been made since early 2005. The proximity to a large city, and cultivated land, means this is likely the station with the largest human influence visible over a long period. Another station is located in Toronto, on the roof of the National Headquarters of the Atmospheric and Climate Science. Unfortunately regular hourly measurements of methane and carbon monoxide have only been made from late 2003 to early 2005.

The Fraserdale observatory is located in northern Ontario, on the edge of the Hudson Bay Lowland and a boreal forest. Wetlands are common in the region, and the site is situated near a reservoir formed by a hydroelectric dam. A more in depth study of the processes controlling CH_4 levels was performed by Worthy et al. [13]. They noted the diurnal cycle, and certain seasonal highs in methane levels.

Lac La Biche is in northern Alberta. It is situated close to peatlands and wetlands, which generally act as a carbon sink, absorbing CO_2, but also release fairly large amounts of methane. The measurements made here should be expected to be representative of wetlands in general, which cover about 10% of Canada's land area. Measurements here have been made since the first quarter of 2007.

Sable Island is a small island off the coast of Nova Scotia, in the Atlantic Ocean. Its remote location, much like Alert, means it is primarily measuring longer term trends in

airflow arriving from other areas, since there is very little locally. The Sable Island station has been operational since late 2002.

3.3 Application of Transform

3.3.1 Wavelet Analysis of Environment Canada Data

The wavelet transform was applied to the longest possible time sequence at each location such that all 3 chemical species were being recorded. Unfortunately there are gaps in this data. Because the wavelet transform depends on accurate lengths to produce a correct image, it was necessary to fill these gaps so that the coefficients could be estimated. As can be seen in the time sequences, the method chosen was to repeat the previous measurement to cross the gaps. The result is a constant value, which does not add to the magnitude of the wavelet coefficients, though it does have an influence on the nearby values which may lead to ringing. Once the wavelet transform had been calculated, the times without original measurements were marked with white gaps. Depending on the size of the gap and the wavelength in question, wavelengths near the gaps may be affected - so care should be taken when reading the plots not to rest excessive significance on features located close to these gaps.

The plots are presented with Alert first, then from West to East across Canada by geographical location. First carbon dioxide, then carbon monoxide, then methane are displayed. Due to the number of plots generated, only those which are representative of the plots in general, or have special features visible, are included. Some of the locations also had shorter time sequences available - Chibougamau, Lac La Biche, and Toronto. So these were not included unless they had particularly significant features, so that longer term features could be seen.

Each plot uses an annual scale along the horizontal axis, and a daily scale along the vertical axis. Due to the turbulent processes involved, the colour scale is logarithmic, with base ten - so the brightest signals are up to 13 orders of magnitude larger than the background level of noise. The minimum scale is adjusted to be zero based by subtracting the smallest coefficient, corresponding to the smallest wave measured. Depending on the station, this minimum level varies somewhat, so colours should not be compared directly between plots.

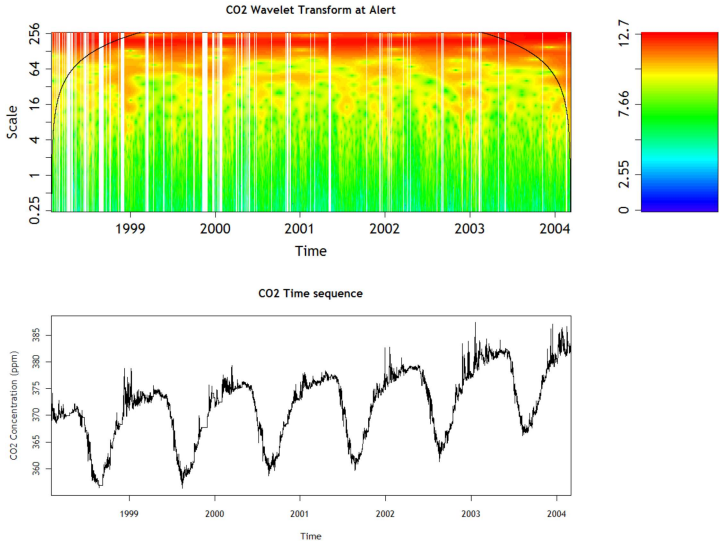

Figure 3.2: Wavelet transform power spectrum and time sequence of carbon dioxide at Alert station.

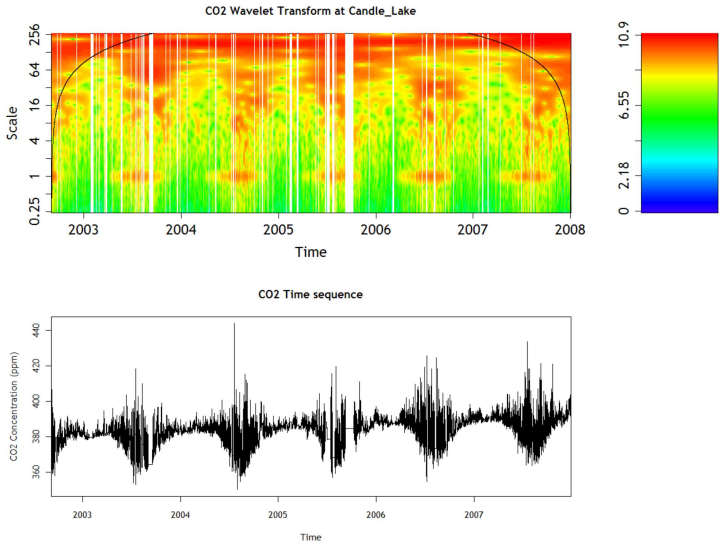

Figure 3.3: Wavelet transform power spectrum and time sequence of carbon dioxide at Candle Lake station.

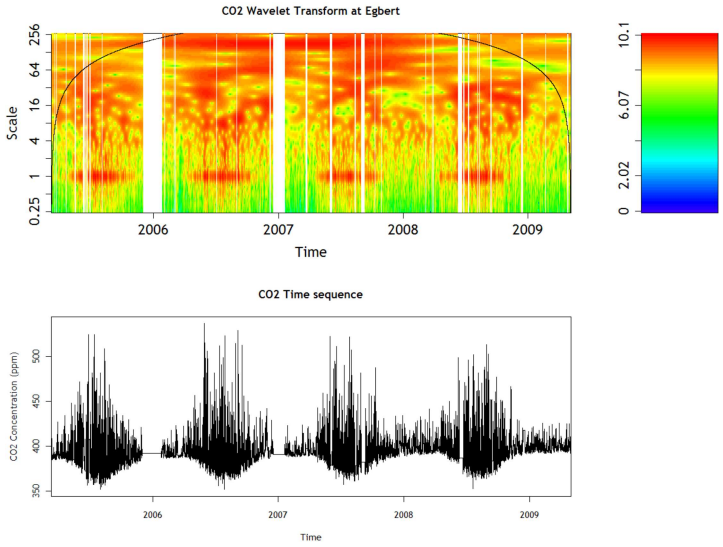

Figure 3.4: Wavelet transform power spectrum and time sequence of carbon dioxide at Egbert station.

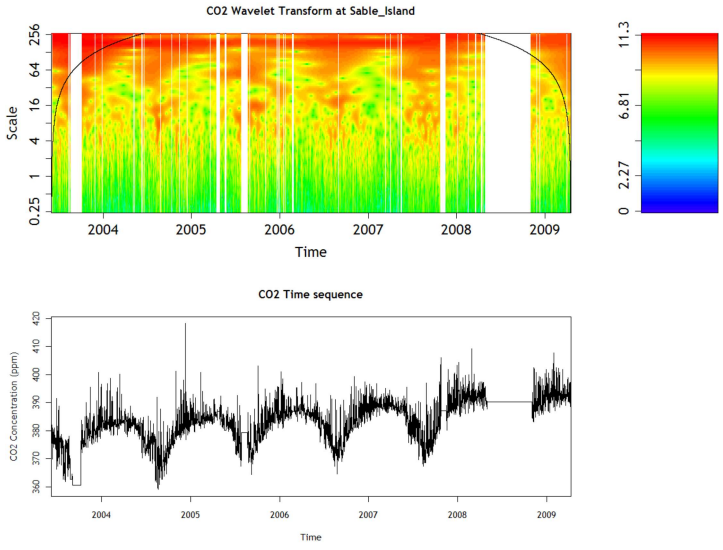

Figure 3.5: Wavelet transform power spectrum and time sequence of carbon dioxide at Sable Island station.

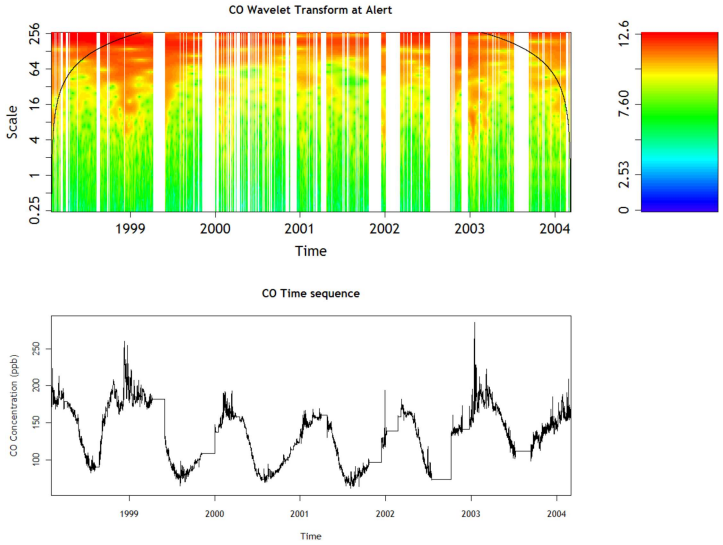

Figure 3.6: Wavelet transform power spectrum and time sequence of carbon monoxide at Alert station.

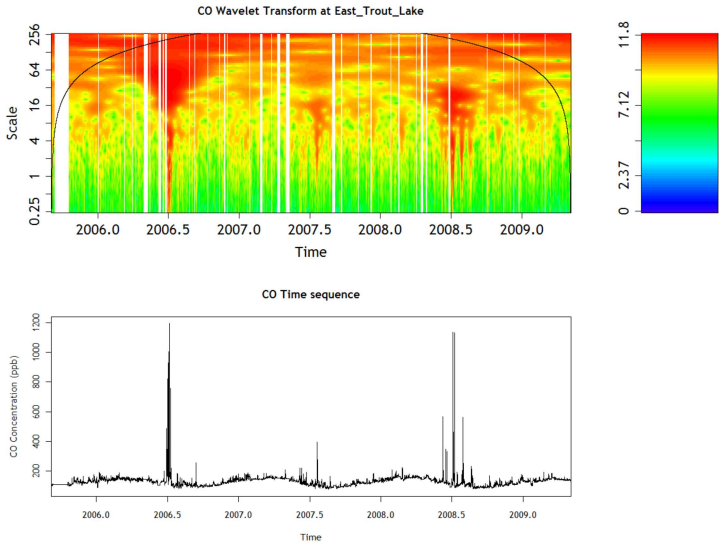

Figure 3.7: Wavelet transform power spectrum and time sequence of carbon monoxide at East Trout Lake station.

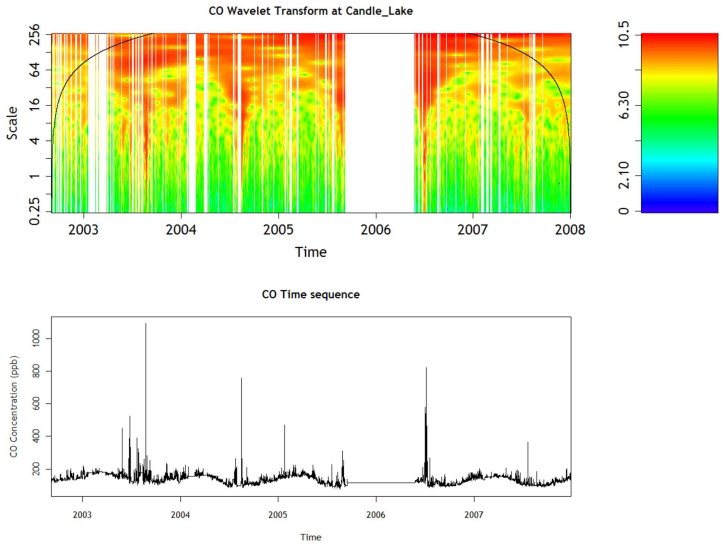

Figure 3.8: Wavelet transform power spectrum and time sequence of carbon monoxide at Candle Lake station.

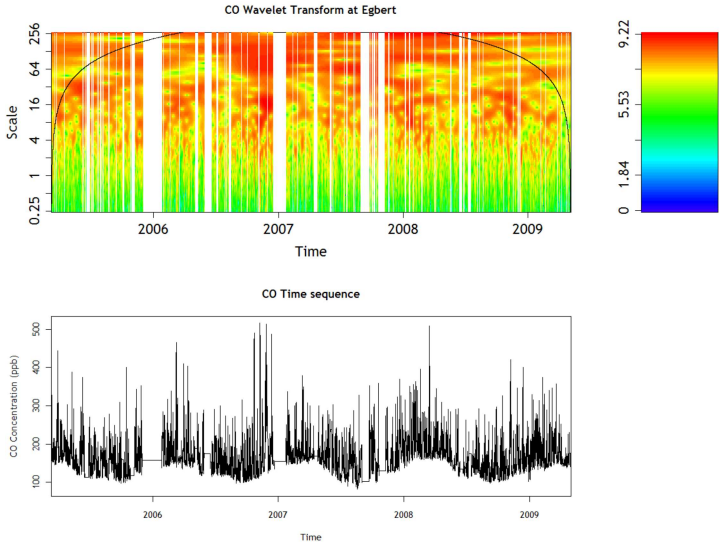

Figure 3.9: Wavelet transform power spectrum and time sequence of carbon monoxide at Egbert station.

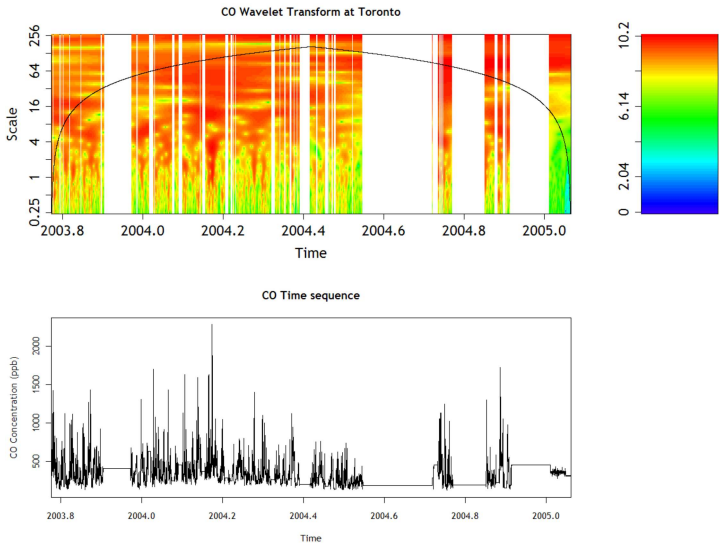

Figure 3.10: Wavelet transform power spectrum and time sequence of carbon monoxide at Toronto station.

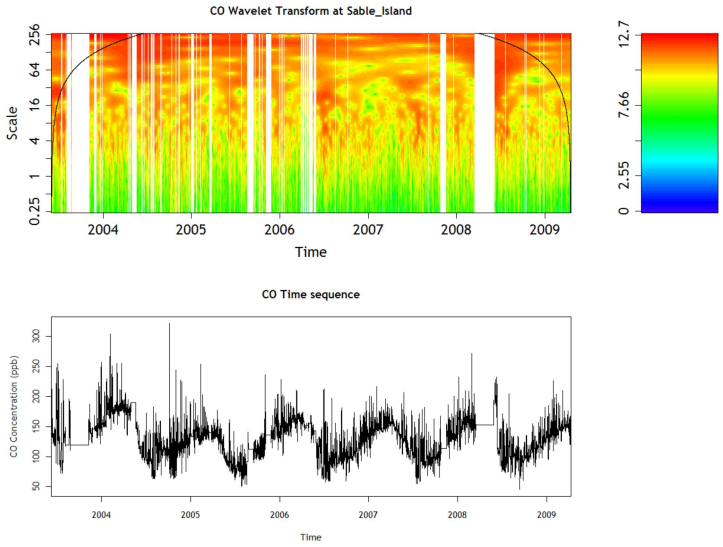

Figure 3.11: Wavelet transform power spectrum and time sequence of carbon monoxide at Sable Island station.

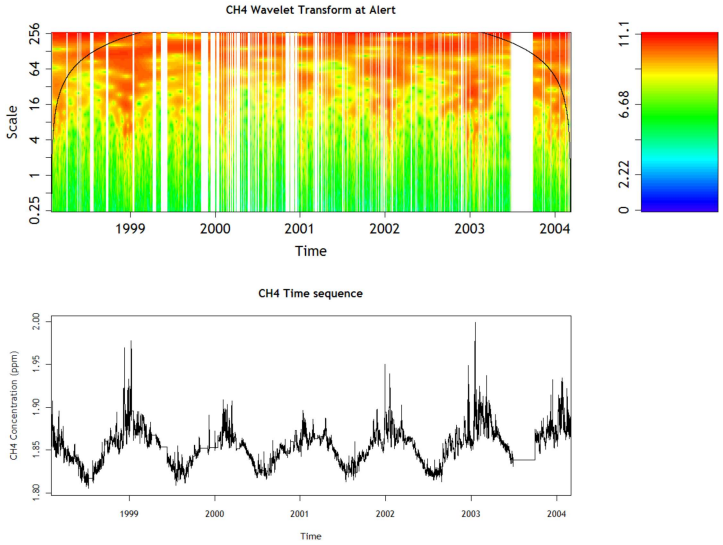

Figure 3.12: Wavelet transform power spectrum and time sequence of methane at Alert station.

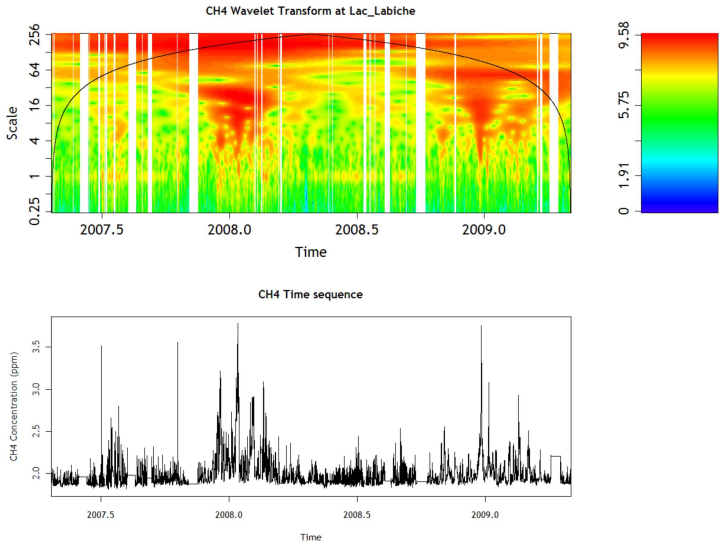

Figure 3.13: Wavelet transform power spectrum and time sequence of methane at Lac La Biche station.

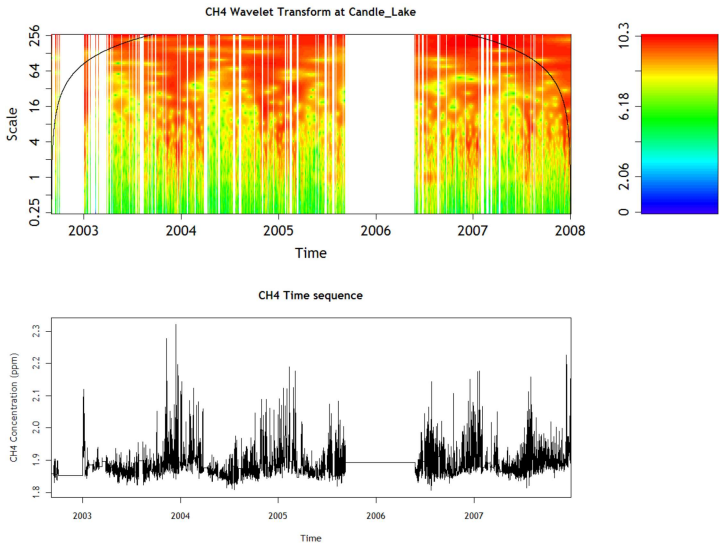

Figure 3.14: Wavelet transform power spectrum and time sequence of methane at Candle Lake station.

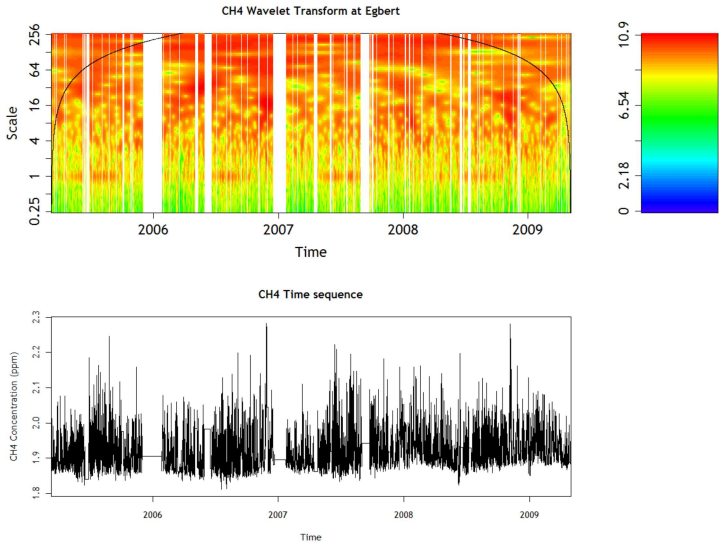

Figure 3.15: Wavelet transform power spectrum and time sequence of methane at Egbert station.

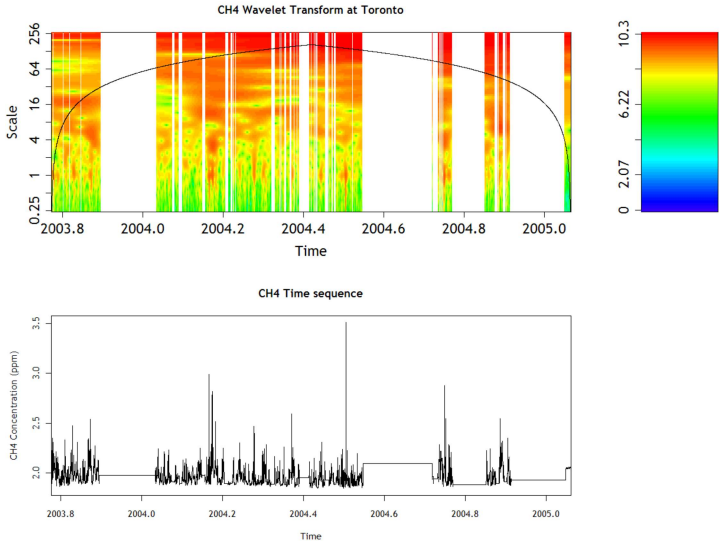

Figure 3.16: Wavelet transform power spectrum and time sequence of methane at Toronto station.

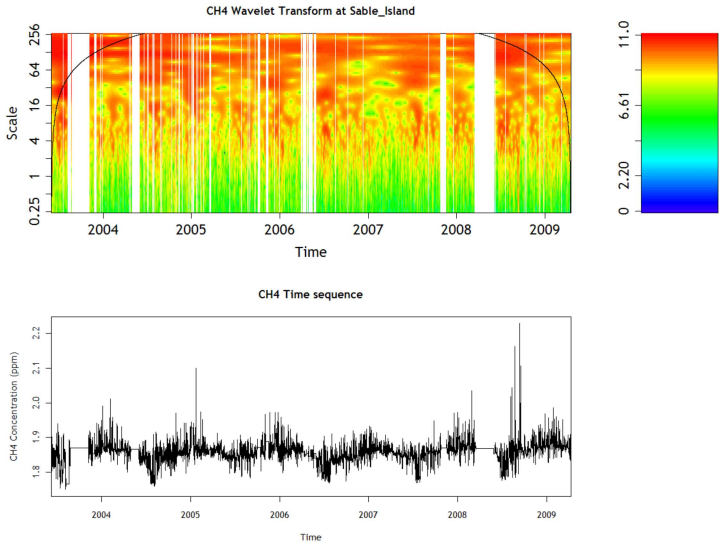

Figure 3.17: Wavelet transform power spectrum and time sequence of methane at Sable Island station.

3.3.2 Wavelet Coherence study of Environment Canada Data

Wavelet coherence was calculated between each pair of chemicals at the ten sites. In order to reduce the number of images presented, only those of particular interest are reproduced here. Some of the time sequences had fairly large gaps in multiple species, resulting in very little usable information in the coherence plots.

The wavelet coherence plots include black outlines to indicate statistical significance. Two weights of lines are used here - thick and thin. The thin lines simply show the 95% confidence level - coefficients that have a 95% or better chance of being non-random. In order to improve this, a second confidence measure is performed, generating the thick lines. This measure uses knowledge of the frequency and time scales to determine what is actually significant - if a signal appears for a smaller duration, or a narrow band of frequencies, it is more likely to be due to turbulence and should be ignored. So the thick lines provide the strongest indication that a particular oscillation is significant. These lines also make it easier to refer to the phase diagram, as the outlines correspond directly.

3.4 Results

3.4.1 Wavelet Transform

Comparison of the wavelet transform plots allows features particular to specific locations to be detected. Patterns appearing at certain wavelengths can also be interpreted. By linking these to physical phenomena or historical events, hypotheses can be formed as to their source.

As expected, the carbon dioxide power spectra, Figures 3.2 to 3.5, show a strong daily cycle, due to plant photosynthesis during the daytime which consumes carbon dioxide. This oscillation is strongest at Egbert, likely due to the more southern and agricultural location. Candle Lake is fairly representative of the other boreal forest locations in this regard. At Alert, there are few oscillations of less than half a year in length - what little does show up may be almost entirely due to the local coherent structures expected due to turbulence. The Sable Island station, while still isolated from the daily cycle, shows more weekly to monthly coherence, the majority of which appears around August to October. This may be due to the contrast between marine and continental air, as the station is on the coast.

The carbon monoxide spectra, Figures 3.6 to 3.11, have far more interannual variation - though the annual trends are fairly consistent, the shorter scale cycles vary quite a bit from year to year. Of note is the large spike in carbon monoxide at Candle Lake and East

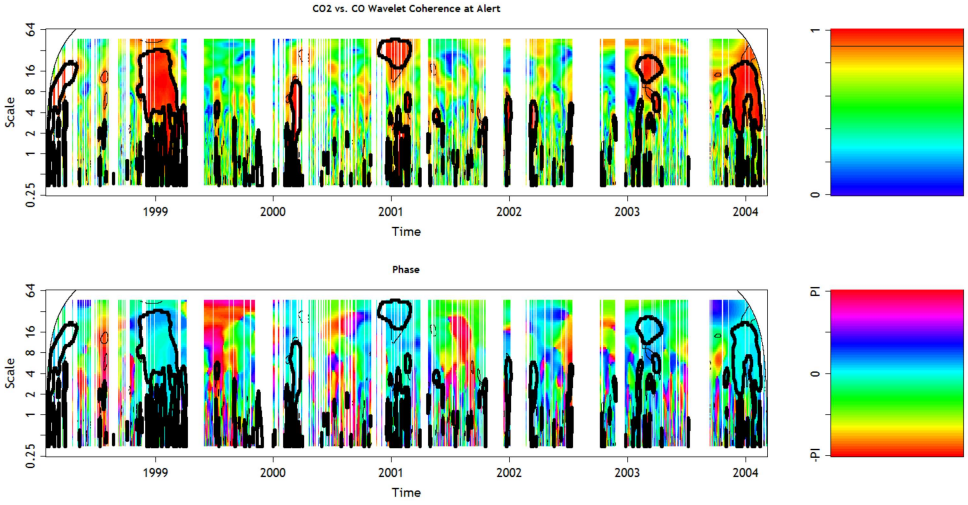

Figure 3.18: Wavelet coherence and phase of carbon dioxide and monoxide at Alert station.

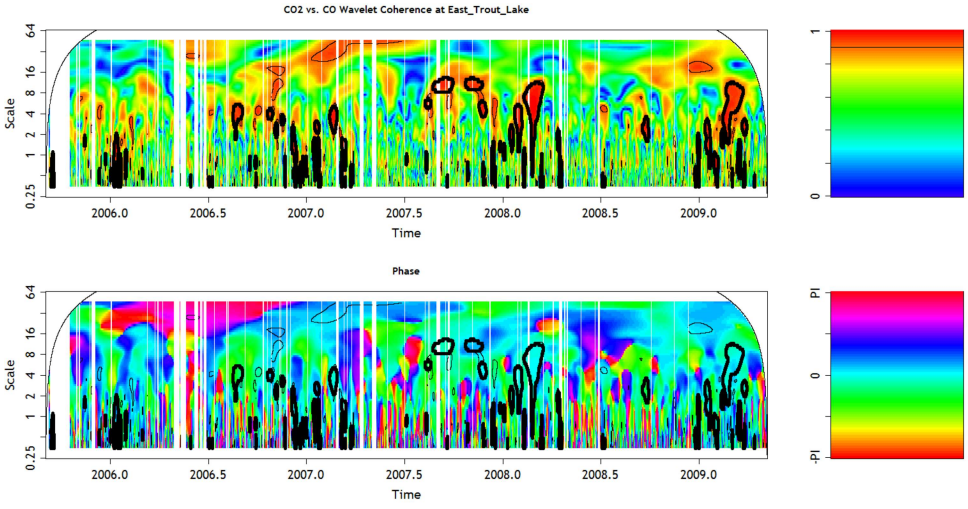

Figure 3.19: Wavelet coherence and phase of carbon dioxide and monoxide at East Trout Lake station.

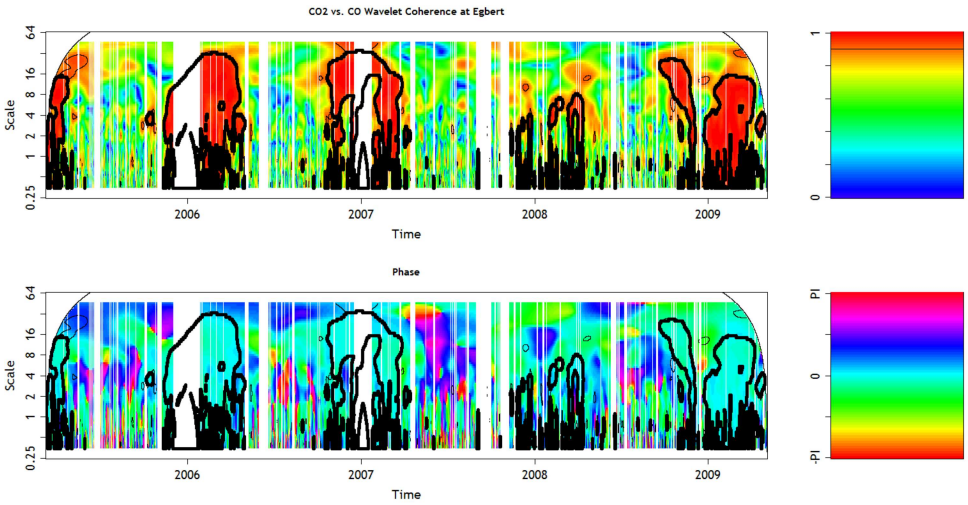

Figure 3.20: Wavelet coherence and phase of carbon dioxide and monoxide at Egbert station.

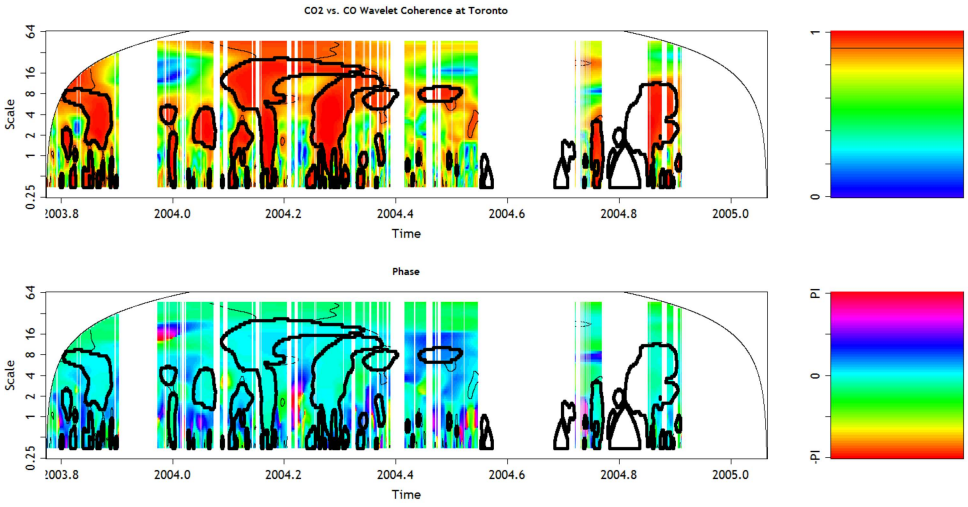

Figure 3.21: Wavelet coherence and phase of carbon dioxide and monoxide at Toronto station.

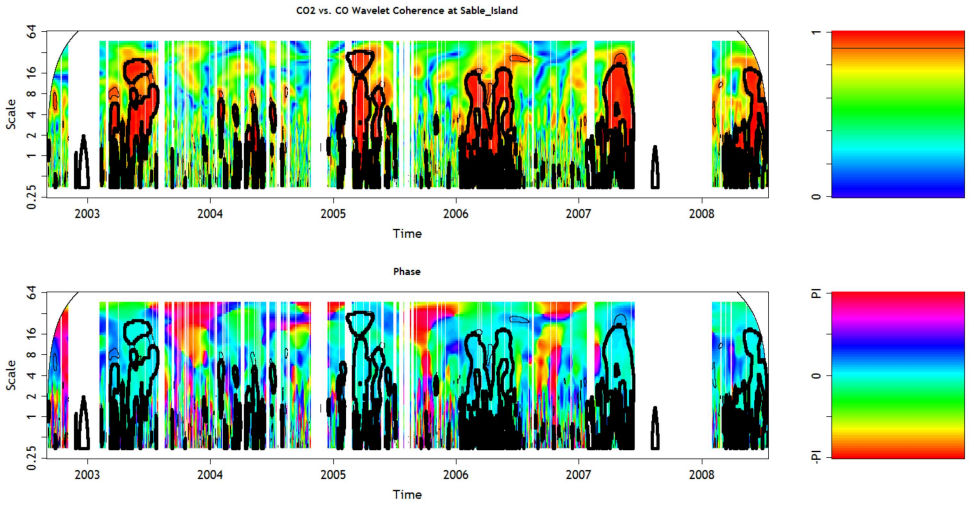

Figure 3.22: Wavelet coherence and phase of carbon dioxide and monoxide at Sable Island station.

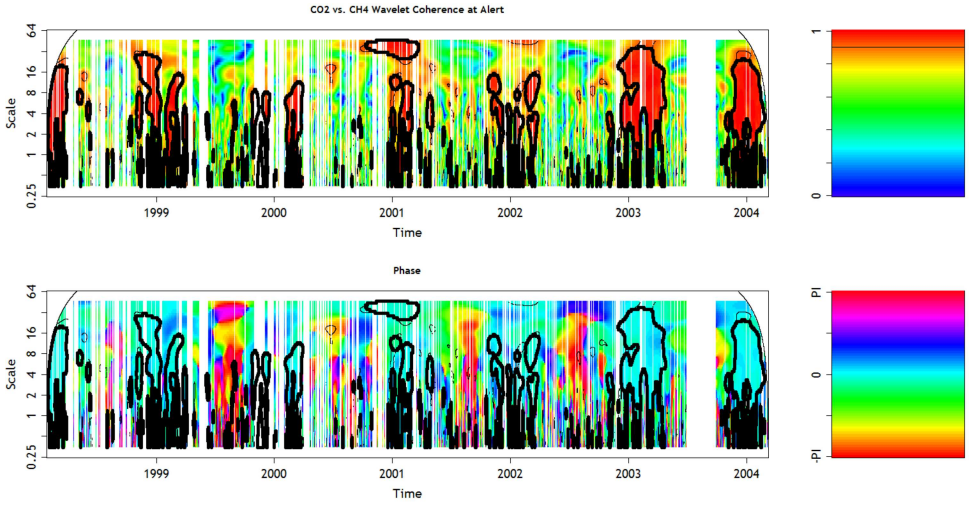

Figure 3.23: Wavelet coherence and phase of carbon dioxide and methane at Alert station.

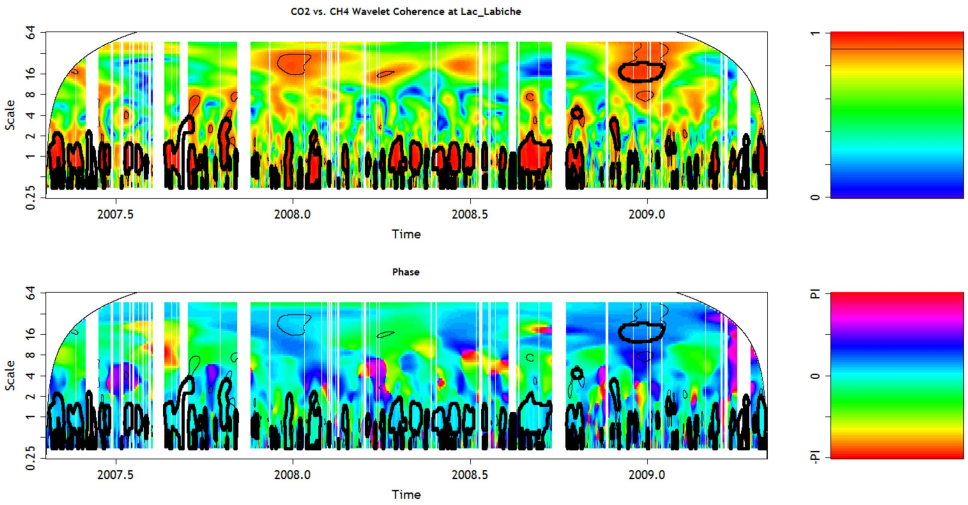

Figure 3.24: Wavelet coherence and phase of carbon dioxide and methane at Lac La Biche station.

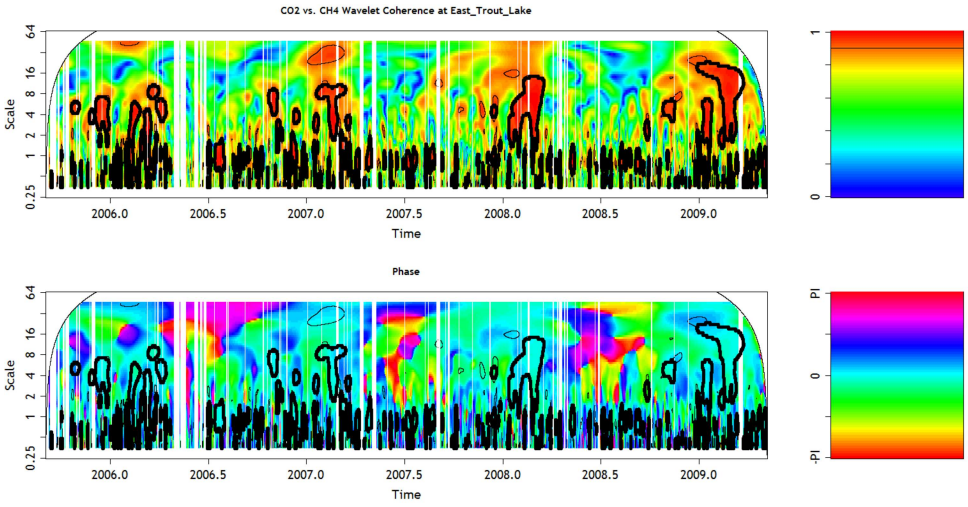

Figure 3.25: Wavelet coherence and phase of carbon dioxide and methane at East Trout Lake station.

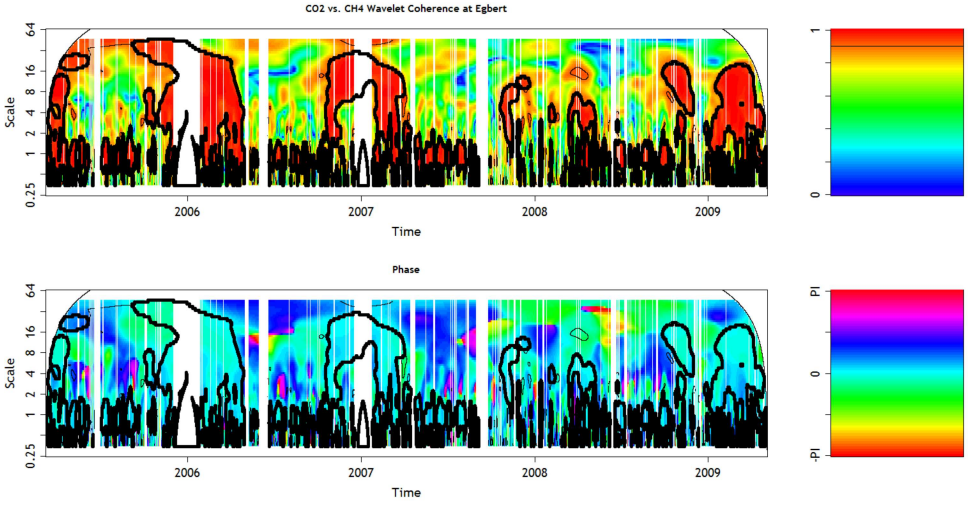

Figure 3.26: Wavelet coherence and phase of carbon dioxide and methane at Egbert station.

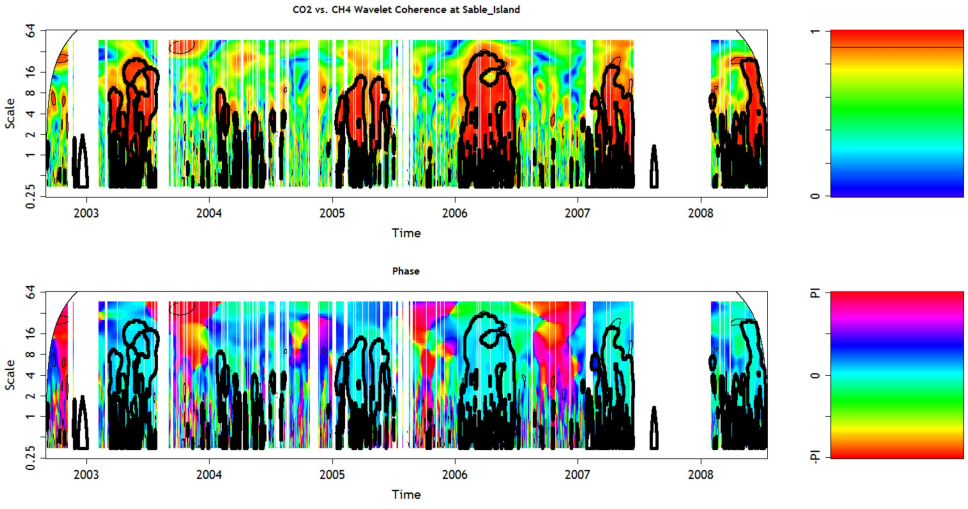

Figure 3.27: Wavelet coherence and phase of carbon dioxide and methane at Sable Island station.

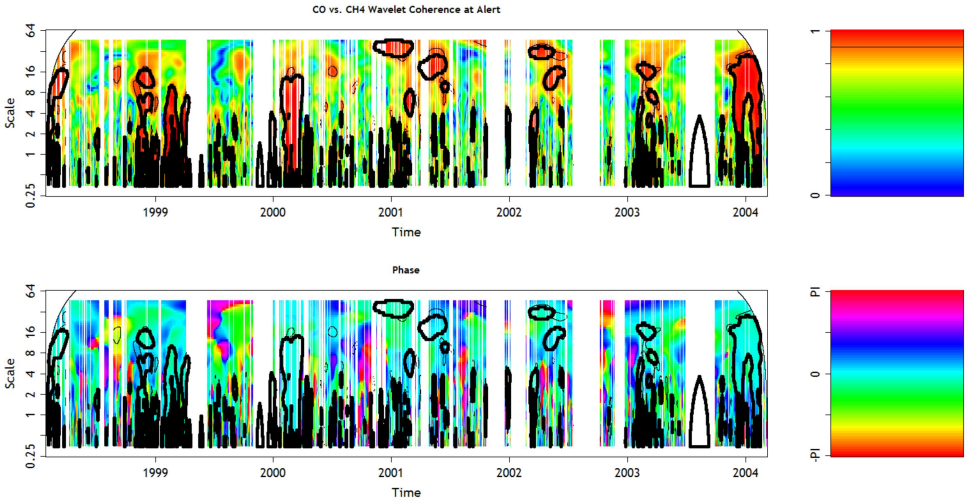

Figure 3.28: Wavelet coherence and phase of carbon monoxide and methane at Alert station.

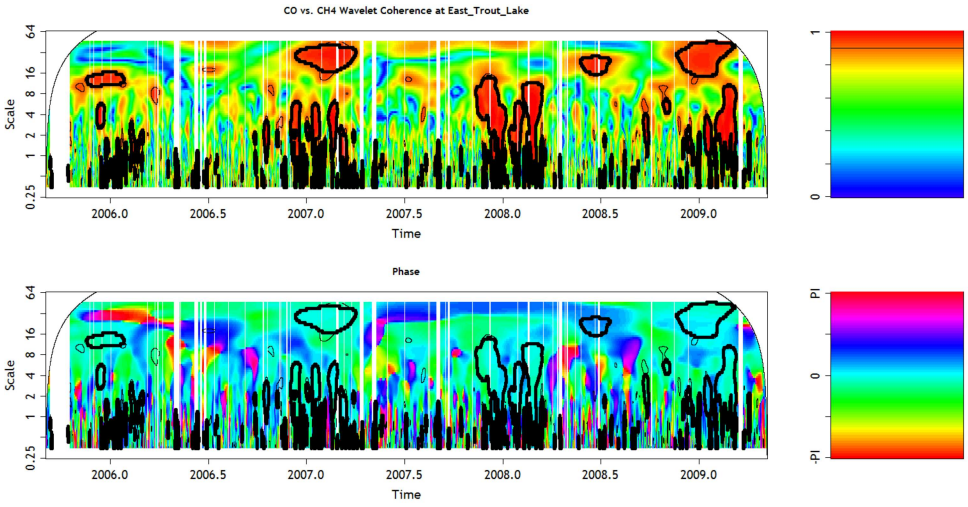

Figure 3.29: Wavelet coherence and phase of carbon monoxide and methane at East Trout Lake station.

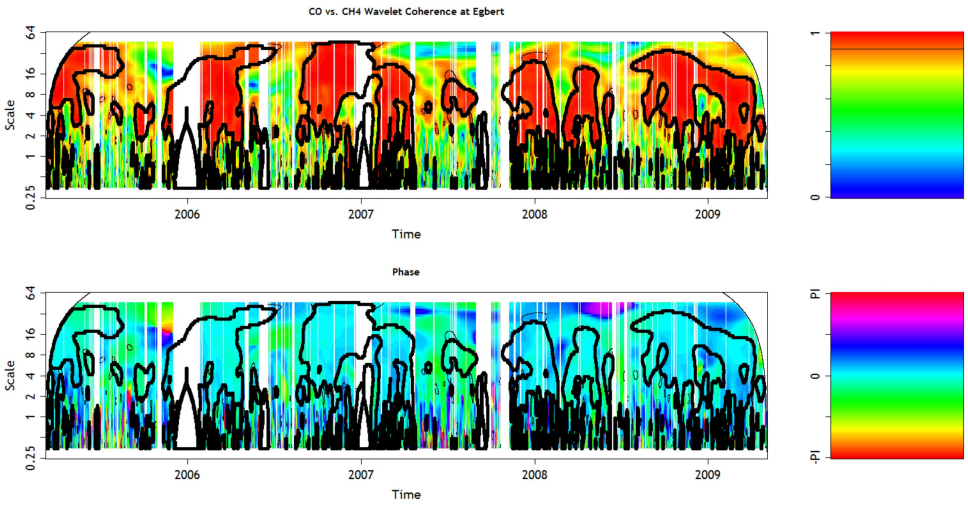

Figure 3.30: Wavelet coherence and phase of carbon monoxide and methane at Egbert station.

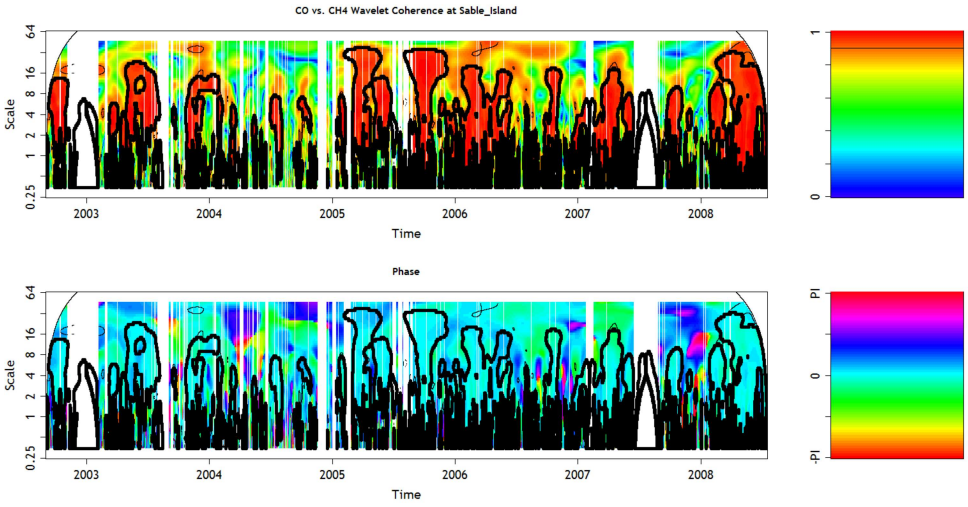

Figure 3.31: Wavelet coherence and phase of carbon monoxide and methane at Sable Island station.

Trout lake around the middle of 2006. This seems to be the result of a large forest fire that burned in northern Saskatchewan in the last week of June. The NASA Earth Observatory [9] published an image on June 29th, 2006, seen in Figure 3.32 showing the extent of the smoke caused by fires, which were unusually large. Other spikes in carbon monoxide levels are most likely due to other large fires - this set of fires is notable because it affected several stations at once, during the overlapping period where multiple stations were operational. Though thousands of fires occur each summer, only a few seem to cause peaks this large.

The methane plots, Figures 3.12 to 3.17, show stronger oscillations in the summer at the daily scale, and in the winter at the weekly to monthly scale - particularly at Lac La Biche and Candle Lake. As before, Alert only shows the oscillations on longer wavelengths associated with annual cycles. At Egbert and Sable Island, the weekly to monthly oscillations seem to exist over the entire year, peaking around September and January. Like the carbon monoxide plots, several show large spikes. However, these are of smaller magnitude, and do not generally last as long. Lac La Biche shows the strongest of these spikes - possibly due to the wetlands nearby.

3.4.2 Wavelet Coherence

Alert, shown in Figures 3.18, 3.23, and 3.28, shows relatively little coherence. It mostly appears around December to March, when presumably the temperature stays consistently below zero and it is dark. Sable Island, by contrast, in Figures 3.22, 3.27, and 3.31, shows strong coherence from 1 month down to less than a day - primarily in February to June. These coherences are present over more wavelengths and a larger part of the year than most of the other stations - the variation between marine and continental air is possibly providing these oscillations, due to regular changes in wind pattern.

Lac La Biche, seen in Figure 3.24, shows strong coherence between carbon dioxide and methane at the daily level. Phase information indicates that the two processes are in phase - CO_2 and CH_4 are being taken up or released together at that daily scale. The winter months show an out of phase coherence - CO_2 is leading by around $\pi/6$, which at the 20 day wavelength means roughly two or three days.

East Trout Lake, in Figure 3.25, shows a weaker daily coherence between carbon dioxide and methane than Lac La Biche. The oscillations around 3 to 7 days in spring are also in phase, or possibly even CH_4 leading in places. The regular annual nature of these oscillations makes it unlikely that they occur randomly, but I cannot say what significance they hold. Looking at Figure 3.19 to compare carbon monoxide and methane, the roughly weekly coherence appears in the Spring, but the daily cycle does not - so some process is likely linking the three gases at the weekly scale. The coherence is less well defined here, so carbon monoxide is likely the odd one out, matching less well with the other gases. In

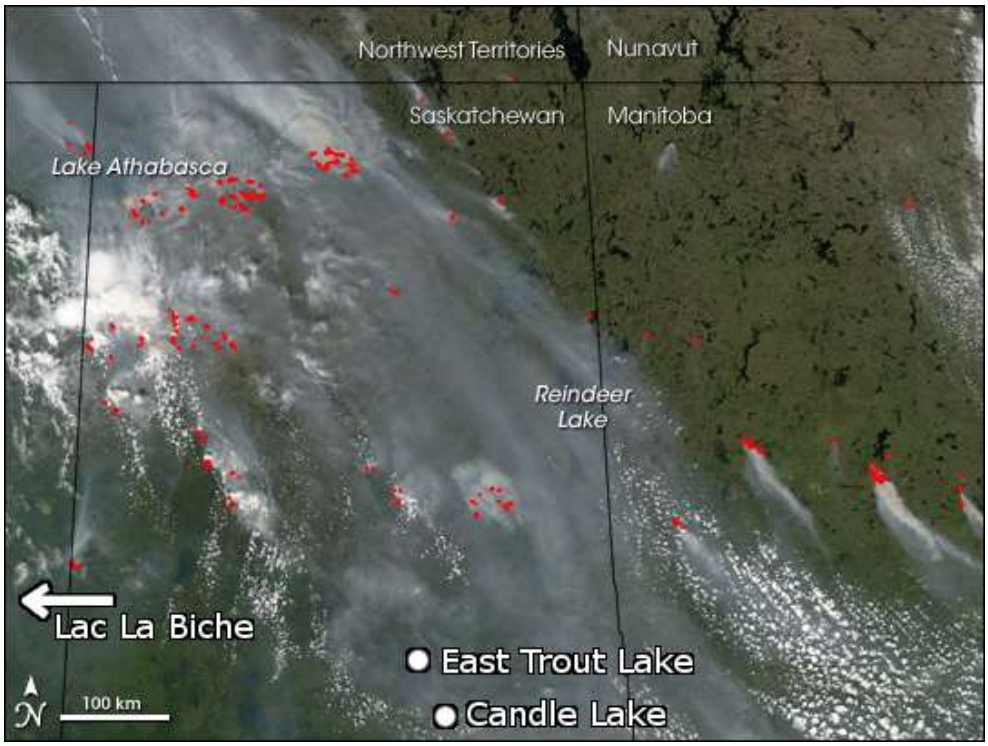

Figure 3.32: NASA Earth Observatory Image Of The Day - June 29, 2006. Forest fires are indicated in red, and large amounts of smoke are visible across Saskatchewan in this satellite image. The locations of the three nearest stations are indicated approximately.

Figure 3.29, the carbon monoxide and methane levels at East Trout Lake are compared. Again, the roughly 4 to 8 day oscillations are present in the spring, but there is a fairly strong monthly component as well, again with the two processes in phase.

Methane and carbon monoxide, compared at the Egbert station in Figure 3.30 show very good coherence from around 4 to 20 day wavelengths. These appear mostly in the months of September to April, while in the summer the two are less well correlated. Carbon monoxide leads these cycles slightly in phase, indicating that possibly it is due to the conversion of carbon monoxide to methane.

Chapter 4

Conclusions

In approaching data with unknown properties, or underlying processes, especially where the data set is large, it is important to have a set of diagnostic tools which can be used to study the data. The wavelet transform, particularly in continuous form, can be readily adapted to this purpose. If oscillatory energy exists in a sequence, then the wavelet transform locates it. If, as is the case with the atmospheric gases, certain processes only happen at certain times of year, the wavelet transform helps to identify when these processes occur.

The main hurdle with the wavelet transform is understanding which parameters to use, and how to interpret the results it produces. It is hoped that this paper can help to inform these decisions, and give an idea of how its application can extract useful information. The wavelet coherence method is less immediately useful - it requires two time series spanning the same domain, both having oscillatory properties. Once such sequences have been found, though, it is easy to identify where they coincide.

Applying the wavelet transform to greenhouse gases allows a new view of the ways these gases interact and vary over time. Between speculation and unidentified processes, a great deal of information has yet to be revealed. The daily cycle of CO_2 could be studied to estimate the growing season and general presence of vegetation in an area. Weekly monthly oscillations appearing in the wavelet coherence plots indicate other correlations.

Future research into quantifying these unknown processes, and looking into literature for potential explanations would be one potential way to continue this work. Another would be to move to atmospheric simulations, and see how the simulation appears in wavelet space during or after the run to improve the quality of results.

This research paper was produced thanks to the guidance of Dr. John C. Lin, particularly in providing the datasets and suggesting interpretations of the information gained. I hope to continue this work in the future, applying the knowledge gained to make contributions in this area of research.

Bibliography

[1] Japan Meteorological Agency. World data centre for greenhouse gases (WDCGG). `http://gaw.kishou.go.jp/wdcgg/`, 2010. 18

[2] M. Farge. Wavelet transforms and their applications to turbulence. *Annu. rev. fluid mech.*, 24:395–457, 1992. 2, 6, 8

[3] Marie Farge, Nicholas Kevlahan, Valérie Perrier, and Éric Goirand. Wavelets and turbulence. *Proceedings of the IEEE*, 84(4):639–669, 1996. 17

[4] Adriana C. Furon, Claudia Wagner-Riddle, C. Ryan Smith, and Jon S. Warland. Wavelet analysis of wintertime and spring thaw CO_2 and N_2O fluxes from agricultural fields. *Agricultural and Forest Meteorology*, 148:1305–1317, 2008. 6, 9, 10

[5] A. N. Kolmogorov. Dissipation of energy in the locally isotropic turbulence. *C. R. Acad. Sci. U.S.S.R.*, 30:301–305, 1941. 17

[6] D. Maraun and J. Kurths. Cross wavelet analysis. significance testing and pitfalls. *Nonlin. Proc. Geoph.*, 11(4):505–514, 2004. 4

[7] D. Maraun, J. Kurths, and M. Holschneider. Nonstationary gaussian processes in wavelet domain: Synthesis, estimation and significance testing. *Phys. Rev.*, 75:016707, 2007. 4

[8] Guy P. Nason. *Wavelet Methods in Statistics with R.* New York : Springer, 2008. 4, 6

[9] NASA Earth Observatory. Image of the day. `http://earthobservatory.nasa.gov/IOTD/`, 2006. 38

[10] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0. 4

[11] V. Ramaswamy, O. Boucher, J. Haigh, D. Hauglustaine, J. Haywood, G. Myhre, T. Nakajima, G. Y. Shi, and S. Solomon. Radiative forcing of climate change. In *Climate Change 2001: The Scientific Basis*. IPCC, 2001. 18

[12] QA/SAC Switzerland, Keith Puckett, and Doug Worthy. GAWSIS station information system. `http://gaw.empa.ch/gawsis/reports.asp`, 2010. 18

[13] D. E. J. Worthy, I. Levin, N. B. A. Trivett, A. J. Kuhlmann, J. F. Hopper, and M. K. Ernst. Seven years of continuous methane observations at a remote boreal site in Ontario, Canada. *Journal of Geophysical Research*, 103(D13):15995–16007, 1998. 20