

WAVELET ANALYSIS OF NORTHEASTERN UNITED  
STATES CLIMATE DATA

by

MICHAEL WOMACK

A THESIS

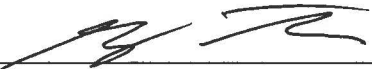
Presented to the Department of Physics  
and the Robert D. Clark Honors College  
in partial fulfillment of the requirements for the degree of  
Bachelor of Science

June 2017

## **An Abstract of the Thesis of**

Michael Womack for the degree of Bachelor of Arts  
in the Department of Physics to be taken June 2017

Title: Wavelet Analysis of Northeastern United States Climate Data

Approved: \_\_\_\_\_

Gregory Bothun

Climate change has been an intensely debated topic over the last decade, and ways of detecting and quantifying climate change are necessary for informing the development of policies related to responses to climate change. Recent research in climate science has used the technique of climate indexing to represent data. Climate indexing averages multiple metrics to provide a better overall measure of regional climate. We present the uses, limitations, and results of applying wavelet transforms to climate indexed data taken from the Northeastern United States. We analyzed three individual metrics (maximum and minimum temperature, precipitation) along with the index metric (average of the three metrics). Monte Carlo simulations suggest that wavelet transforms will return nearly identical results when applied to raw data versus z-scored data, given sufficiently large data sets. Our results suggest that wavelet transforms are a useful tool for characterizing time-dependent climate behavior, but it is likely more useful to consider raw data than transformed data to give statistically reliable results.

## **Acknowledgements**

I would like to thank Greg Bothun for his patience and input in advising this thesis. His classes have also been a highlight of my Clark Honors College experience, and he has challenged me to become a better scientist and thinker. I would also like to thank Casey Shoop and Dave Sutherland for serving on my thesis committee and for their helpful feedback. Additional thanks to Bishara Korkor for his help with programming while I learned the coding skills necessary to work on this project. The Robert D. Clark Honors College has been a great experience and I am very thankful for all the professors who have pushed me outside of my comfort zone and introduced me to ideas I would never have found myself. I would further like to thank the Math and Physics professors I have had for their support and willingness to indulge my endless questions. Finally, I would like to thank my family and friends, as they have been an incredible source of wisdom and support over the last four years.

## Table of Contents

Introduction	1
Research Questions	2
Northeast Regional Behavior	4
Description of the Data Set	7
Data Analysis Methods	9
Methods	16
Morlet Wavelet	16
Convolution Method	18
The Cone of Influence	21
Statistical Testing	22
Results and Discussion	25
Example Wavelet Analysis Plot	26
Wavelet Transforms of Z-scored Data versus Raw Data	28
Influence of Trends on Wavelet Analyses	32
Wavelet Analysis of the Northeastern United States Climate	36
Overall Northeastern Climate	37
Difference In Climate for West/East Appalachian Regions	42
Harr Wavelet Applied to Single Station Data	46
Conclusion and Future Work	48
Summary of Results	48
Future Work	49
Bibliography	52

## **List of Accompanying Materials**

The code used in this thesis can be found at:

<https://github.com/anTIMattErz/MWWavelets>

## List of Figures

Figure 1: Example of a stationary signal	10
Figure 2: Example of non-stationary signal	11
Figure 3	14
Figure 4: Our Harr wavelet function	15
Figure 5: Sampled Morlet wavelet	17
Figure 6: Example wavelet coefficient plot	26
Figure 7: Comparison of the wavelet coefficients of raw data to wavelet coefficients of z-scored data	30
Figure 8: Pure sinusoidal signal with wavelet transform	33
Figure 9: Same sinusoidal signal from Figure 8, but with a linear trend added	34
Figure 10: Similar linear trend as in Figure 9, but with a twice the slope and a red-noise time series added	35
Figure(s) 11, 12: Averaged monthly maximum and minimum temperature for all weather stations	38
Figure(s) 13, 14: Averaged monthly precipitation and NEI for all weather stations	38
Figure 15: Distribution of USHCN stations relative to the Appalachian Mountains	42
Figure(s) 16, 17: Maximum and minimum temperature wavelet analyses for stations East and West of the Appalachian Mountains	43
Figure(s) 17, 18: Precipitation and NEI wavelet analyses for stations East and West of the Appalachian Mountains	43
Figure(s) 19, 20: Harr wavelet applied to average monthly minimum temperature for one weather station, versus applied to a red noise sequence of same length and statistical properties	46

## Introduction

Over the last two decades, climate change has become one of the most controversial issues in science and politics. As researchers learned more about how Earth's climate systems interact and evolve, it became apparent that humans have become a significant force. In 1990, the Intergovernmental Panel on Climate Change (IPCC) produced their first report, stating, "There is concern that human activities may be inadvertently changing the climate of the globe through the enhanced greenhouse effect".<sup>1</sup> The IPCC's third report was more direct, "There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities."<sup>2</sup> General consensus among climate research reports is that Earth's climate has entered a period of distinctly accelerated change, with much attention being given to the issue of global average warming. However, while climate change is typically discussed as a global issue, it is ultimately a local phenomenon. Not only do local factors such as geography and ecology cause climate change to differ from region to region, but the inhabitants of the regions are affected uniquely as well. An increase in average summer temperatures affects someone in Los Angeles, California differently than someone in rural Maine. Differences in regional infrastructure, population distribution, and economic markets result in a need for local, rather than simply global, responses to climate change.<sup>3</sup> In order for policy makers to develop mitigation and response strategies, local and regional climate behaviors must first be understood.

---

<sup>1</sup> Intergovernmental Panel on Climate Change, *Climate Change: The IPCC Scientific Assessment*, 1990, viii.

<sup>2</sup> Intergovernmental Panel on Climate Change, *Climate Change: Third Assessment Report*, 2001, [http://www.grida.no/publications/other/ipcc\\_tar/](http://www.grida.no/publications/other/ipcc_tar/). Summary for Policymakers.

<sup>3</sup> Frumhoff et al., *An integrated climate change assessment for the Northeast United States*, 420.

Several studies have looked at climate change evidence and impacts for the Northeast United States. These studies typically examine temperature, precipitation, and snowfall trends with an emphasis on creating models to predict future behavior for policy development. Little attention has been given to analyzing how subregions of the Northeast have been evolving over the last century. A subregion of the Northeast is a region that is a subset of the larger Northeast region, likely chosen due to a geographic property that differentiates it from the larger region. As an example, we will divide the Northeastern U.S. into the subregions East and West of the Appalachian Mountains. This study seeks to examine how temperature and precipitation patterns in the Northeast are changing, and whether the subregions defined by the Appalachian Mountains are evolving differently from one another. The main tool used in this study is the wavelet transform, which had received a flurry of interest in the late 1990's and early 2000's but never became an integral data analysis technique in climate science. This section presents the questions this thesis addresses, along with background information about the Northeast climate relevant to the research. Afterwards a brief description of the data set is provided, along with some motivation for the usage of wavelet transforms on climate data.

### **Research Questions**

This study seeks to answer two primary questions about the climate in the Northeast United States:

1. Are climate cycles in the NE U.S. shifting, and if so, how are they shifting?
2. Is there a difference in climate patterns in urban vs. rural areas, coastal vs. non-coastal areas, and in different altitude ranges?



With respect to question one, it has been determined that over the last 100 years the Northeast has seen a significant shift in climate behavior.<sup>4</sup> We know that the region is becoming warmer and wetter, but there has not been research that quantitatively shows how cyclical behavior in the region has shifted. Regional climates are influenced by numerous factors, many poorly understood, but many climate related processes are known to exhibit cyclical behavior. Examples include daily cycles due to the rotation of the Earth on its axis, and yearly seasonal cycles due to the tilt of the Earth on its axis while orbiting the Sun. Longer cycles such as the Atlantic Multi-decadal Oscillation (AMO) and Pacific Decadal Oscillation (PDO) drive periodic behavior in climate on longer periods. The interactions of these many cycles cause fluctuations in various regional climate metrics (temperature, precipitation/rain, snowfall, etc.), which helps determine the character of the region's climate.<sup>5</sup> We would like to know, for each climate metric, if and when climate cycles have shifted, in addition to general trends in the region.

Question two is concerned with how climate change affects subclimates in the Northeast. In her University of Oregon Honors College thesis, Wilkie found that, "During the site comparisons to the NEI, it became evident that there may exist climatic sub-regions within the NE."<sup>6</sup> The NEI (Northeast Climate Index) is a climate index designed to measure the climate of the Northeastern U.S. region. Of particular interest is the difference between coastal and non-coastal regions, as the Northeast region is split by the Appalachian Mountains. Wilkie noticed that the region to the east of the

---

<sup>4</sup> NOAA, Regional Climate Trends and Scenarios for the U.S. National Climate Assessment.

<sup>5</sup> Wilkie, Investigating Regional Climate Change in Northeastern United States, 7-8.

<sup>6</sup> Ibid. 39.

Appalachians had experienced more dramatic changes in temperature and precipitation than the Northeast region as a whole. While it is fairly straightforward to see changes in the average behavior of a given metric as a function of time, we specifically would like to see if there were changes in the cycles present in the metric. This would provide a hint that the underlying mechanism that determines the region's climate had changed, or there was some new external source of variation that was not previously affecting the region.

### **Northeast Regional Behavior**

The Northeast United States is defined as the states of Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont.<sup>7</sup> As a whole the region contains a highly diverse climate due to a variety of geographic factors. The Atlantic Ocean serves as a regulating force for the coastal region, while the Great Lakes and Lake Champlain influence inland areas. Westerly winds bring air masses from the interior of North America across the Northeast, and the polar jet stream is often directly over the region during the winter. This causes winters to be characterized by harsh cold, abundant precipitation, and frequent storms. The Appalachian Mountains help shield the southern part of the region from the interior air masses, while preventing the warm and humid air masses of the Atlantic Ocean from reaching the western half.<sup>8</sup> Additionally, orographic effects due to the presence of the Appalachians cause increased precipitation near the mountain range.

---

<sup>7</sup> Frumhoff et al. Confronting Climate Change in the U.S. Northeast. 1.

<sup>8</sup> NOAA, Regional Climate Trends and Scenarios for the U.S. National Climate Assessment, 11.

Summers in the Northeast are typically warm and humid in the southern region, while the northern region is moderated by the presence of cooler air masses from Canada. Northeastern winters are notable for their frequent storms that produce extreme snowfall, flood-inducing rainfall, and dangerous cold temperatures.<sup>9</sup> Warmer temperatures generally correspond to higher levels of precipitation and lower levels of snowfall. This is in contrast to the Northwest region of the United States, where lower temperatures correspond to higher levels of precipitation and snowfall.<sup>10</sup> Average temperatures in the Northeast decrease towards the North, along with increasing elevation and increasing distance from the coast. Precipitation and snowfall vary widely with geography and even month to month.<sup>11</sup>

Temperatures in the Northeast have risen since 1900, with the majority of the 1.5°F increase occurring in the last few decades. Since 1970 the average regional temperature has risen by 0.5°F per decade, with a more dramatic increase in the winter temperatures (4°F between 1970 and 2000). This trend is expected to continue, with winters experiencing greater warming relative to the summer. Following this trend, annual precipitation has gradually increased by between 5 and 10 percent across the Northeast. In early 1960's the region experienced a severe drought. Before 1970, the increase in precipitation was evenly spread between the spring, summer, and fall seasons, while the winter season saw little change. However, beginning in the 1970's and 1980's, this trend reversed, and winter precipitation has increased slightly while spring, summer, and fall precipitation has remained relatively constant. The increased

---

<sup>9</sup> Ibid. 12.

<sup>10</sup> Wilkie. 47.

<sup>11</sup> Frumhoff et al. *Confronting Climate Change in the U.S. Northeast*. 8.

winter precipitation is likely due to increased winter temperatures causing snow to fall as rain. There is a short-term drought once every two years across most of the Northeast, and once every three years over northern Maine, northern New York, and western Pennsylvania. Medium-term droughts occur once every fifteen years in the inland region, but do not occur in the coastal region. Long-term droughts occur less than once every thirty years.<sup>12</sup>

Population growth is less directly related to a shift in local climate, but it is important to note that urbanization has intensified throughout the Northeastern urban corridor, which was already the most populated region in the United States.<sup>13</sup> This could cause an increase in the urban heat island effect, where urbanized regions experience higher average temperatures due to various effects of urbanization.

The effects of climate change on the Northeast are especially profound due to the region's economic reliance on natural resources for industry and recreation. Briefly, decreased snowpack will negatively affect skiing and other large recreational industries in all states except Maine. Higher temperatures and increased runoff from precipitation will negatively impact the Northeast's large fishing industry, while also stressing local ecosystems and possibly threatening human health. While rising temperatures will likely reduce winter-related stresses, pre-existent summer stresses such as heat waves and air quality concerns are likely to exacerbated. Critical infrastructure in the Northeast is aging, has received insufficient maintenance, and suffers from inadequate capacity to handle peak loads. Climate change poses an especially large risk to the power and transportation infrastructures, where higher temperatures will draw more

---

<sup>12</sup> Ibid. 9.

<sup>13</sup> Ibid. 12.

power for cooling and more frequent storms and heavy precipitation could damage or destroy essential infrastructure. Rising sea levels coupled with increased storm frequency is likely to destroy much of the beachfront property in the Northeast.<sup>14</sup>

### **Description of the Data Set**

The data set that I am analyzing was originally constructed by Adrien Ann Wilkie as part of her University of Oregon Honors College thesis, *Climate Change in the Northeastern United States*, in 2010. Greg Bothun obtained an updated version of the data set through personal correspondence. She created a set of climate indices for the Northeast based on various collections of weather station sites. The different collections of weather stations were made by applying filtering criteria to all the weather stations in the Northeastern U.S. A climate index is a measurement of the overall climate of a system by combining a set of metrics into a single metric, the index. For example, Wilkie combined monthly average maximum temperature, minimum temperature, precipitation, and snowfall data into a single monthly value for each site by averaging the monthly z-score data for each metric. The index can then be analyzed individually for each site, regionally, or for the Northeast region as a whole. Climate indexing was originally developed in the mid-1990's by C.C. Ebbesmeyer and R.M. Strickland to analyze oyster population in the Pacific Northwest. They developed a Pacific Northwest Index (PNI) to quantitatively characterize the behavior of the Pacific Northwest climate.<sup>15</sup> In 2004, Harvey Rogers determined that the PNI was a robust indicator of the climate in the Northwest.<sup>16</sup>

---

<sup>14</sup> National Assessment Synthesis Team, *Climate Change Impacts on the United States*, 41-45.

<sup>15</sup> Ebbesmeyer & Strickland, Oyster Condition and Climate: Evidence from Willapa Bay.

<sup>16</sup> Rogers, Assessing the robustness of the Pacific Northwest Climate Index....

Wilkie obtained data from the United States Historical Climatology Network (USHCN) that recorded maximum temperature, minimum temperature, and precipitation and snowfall levels for each month at high-quality data collection stations. The USHCN stations meet quality-control criteria that include “length of record, percent of missing data, number of station moves and other station changes that may affect the homogeneity, and resulting network spatial coverage”.<sup>17</sup> Data was obtained from 137 stations that have complete monthly records from January 1900 to December 2015. Four stations were missing data before 1904. We substituted zero values for all missing data. The missing data is negligible for the analysis of the time series due to its sparseness. Wilkie originally used snowfall records obtained after 1930 from the National Weather Service’s Cooperative Station Network (NWSCoop) for each station as well, but this thesis focuses only on the rainfall and temperature data due to the incompleteness of the snowfall data set. None of the stations in the data set were impacted by proximity to the Great Lakes, which cause a lake-effect snow. Lake-effect snow occurs when cold air moves over a large, warm lake, causing deposition of water vapor on the opposite side of the lake as snow. Additionally, the USHCN states that their stations are located in rural areas or small towns, and thus should not be significantly affected by urbanization. Wilkie determined that it was not necessary to filter any stations due to proximity to urban centers.<sup>18</sup>

A z-score was calculated for each monthly value of each metric relative to the stations entire data set. Z-scores measure how many standard deviations away from the population sample mean a value is. The final form of the data set is thus monthly z-

---

<sup>17</sup> Wilkie, 24.

<sup>18</sup> Ibid. 27.

scores for maximum temperature, minimum temperature, and precipitation at each of the 137 data stations from January 1900 to December 2015. A monthly index value for each site can be made by averaging the monthly z-scores of each metric. These index values can also be averaged across regions, or the entire Northeast, to study the overall behavior of a region of interest. One of the primary results we discuss in this thesis is that a Morlet wavelet transform applied to z-score transformed data gives nearly identical results as the same wavelet transform applied to the raw data. This makes it sensible for us to make claims about the behavior of climate cycles in the region by looking at the indexed data, as signals detected in the transformed data correspond to the same signal in the original data. Assuming that we are performing the wavelet transform on sufficiently large data sets, the z-score transform does not add artificial signals.

Because higher temperatures correspond to higher levels of precipitation in the Northeast, we can examine regional behavior by using this index:

$$NEI_{\text{month}} = (\text{MaxTempDev}_{\text{month}} + \text{MinTempDev}_{\text{month}} + \text{PrecipDev}_{\text{month}})/3$$

Here, dev refers to the z-score values per month for the given metric. This index tracks how warm and wet a given month is on the record. Summers should have large index values, while winters should have negative index values, due to the use of z-scores.

## **Data Analysis Methods**

The most common technique for determining oscillatory trends in data is the Fourier transform. When analyzing a time series, the Fourier transform allows us to decompose the data into its individual harmonic components (sines and cosines). The amplitudes of each component tell us the relative amplitude of the frequencies appear in

the data set. While Fourier transforms are extremely effective for analyzing the frequency content of a stationary signal, they run into issues with signals whose frequency content shifts over time. A stationary signal can be thought of a signal whose frequency content does not change over time, such as a simple sine wave. An example of a non-stationary signal would be a sine for the first half of the signal, and then a sine with twice the frequency for the second half of the signal.

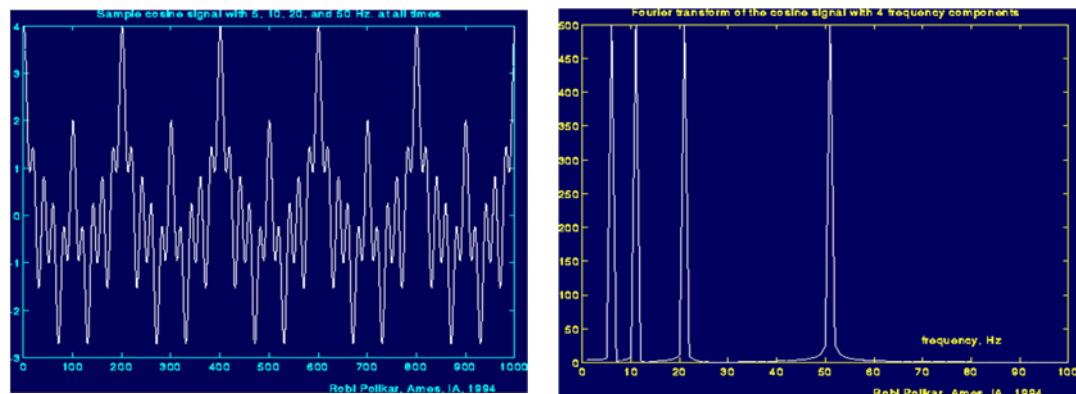
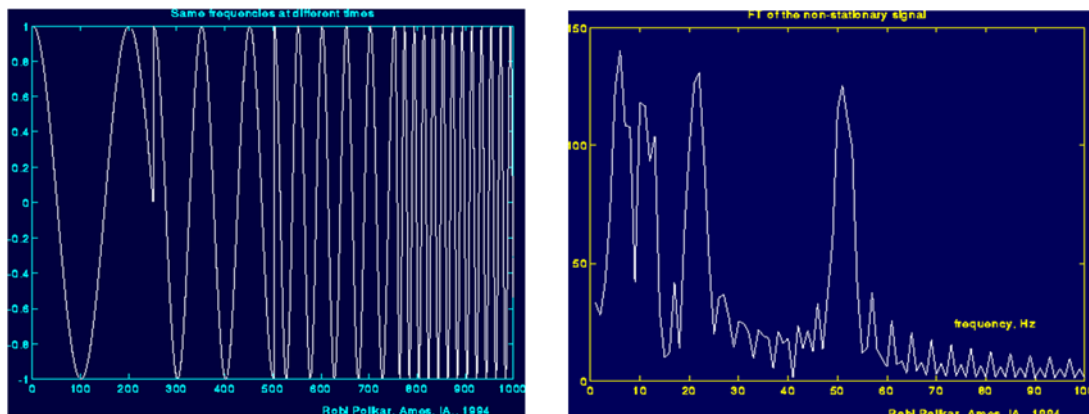


Figure 1: Example of a stationary signal, a sum of four cosines of equal amplitude with frequencies 5, 10, 20, and 50Hz present throughout the signal. Right figure is the Fourier transform representation of the data, showing power at 5, 10, 20, and 50Hz.<sup>19</sup>



<sup>19</sup> Polikar, “The Wavelet Tutorial”, <http://users.rowan.edu/~polikar/WAVELETS/WTpart2.html>. Used with permission from author.



Figure 2: Example of non-stationary signal. Signal has same frequency components as in Fig. 1, but occurring individually at different times. Right figure is the Fourier representation of the signal.<sup>20</sup>

Note that for the non-stationary signal in Figure 2 the Fourier transform has peaks at the same points as in the stationary signal of Figure 1, but with significantly more noise and less resolution. Note that one of the central purposes of this study is to identify when a signal's frequency content shifts. The Fourier transform in Fig. 2 shows that there are dominant periodicities at 5, 10, 20 and 50Hz, but without having seen the original signal we would have had no information as to when these frequencies were present. This comes from the “global support” of the sines and cosines used as basis functions in the Fourier transform. The support of a function refers to the domain of the function, i.e. a function with global support covers the entire time domain, while a function with compact support only covers a part of the time domain. The global support of sines and cosines cause the Fourier transform to consider the entire signal at once, and time dependent information is lost. An example of this loss of time dependence is that the Fourier transform of a reversed time series is the same as the original time series.

It is possible to introduce compact support into the Fourier transform by using the Windowed Fourier transform (WFT). In the WFT, a fixed length segment of the time series is considered, and the Fourier transform is calculated with the data in this segment. The width of the segment is varied to allow for varying frequency resolution, and the segment is moved along the time series to pick up local data. While this does allow for better time resolution of changes in the signal, frequency resolution is sacrificed. Other issues arise from the use of the fixed window, such as aliasing of high

---

<sup>20</sup> Ibid.

and low frequency components that are outside the frequency range of the window.<sup>21</sup> Additionally, frequency resolution in the low frequency bands is poor, and high frequency signals are often overrepresented.<sup>22</sup> These are major issues that would interfere with our ability to detect cyclical behavior in climate metrics (which are often in the low and high frequency ranges), and lead us to consider the wavelet transform.

Fundamentally, the wavelet transform is a generalization of the Fourier transform. A set of basis functions are matched to the data, and a good match gives a large amplitude coefficient for that function while a poor match gives a low amplitude coefficient. This is essentially the same idea as the Fourier transform. The primary difference between the two methods is that wavelet transforms utilize different sets of basis functions. These basis functions are derived from the scaling and translation of a “mother” wavelet  $\varphi_0(t)$ , with

$$\varphi_{ab}(t) = \frac{1}{a^{1/2}} \varphi_0\left(\frac{t-b}{a}\right)$$

where  $a (>0)$  and  $b$  are real. The parameter  $a$  is the scaling factor of the wavelet, and  $b$  translates the wavelet. The wavelet transform is defined as the convolution of the signal  $\tau(t)$  with the wavelets, similar to the Fourier transform (convolution of sines and cosines with the signal):

$$W(b, a) = \frac{1}{a^{1/2}} \int \varphi^* \left( \frac{t-b}{a} \right) \tau(t) dt$$

where  $*$  denotes the complex conjugate. At each scale and position (given by the parameters  $a$  and  $b$ ), a wavelet coefficient is calculated that describes how well the

---

<sup>21</sup> Torrence and Compo, A Practical Guide to Wavelet Analysis, 63.

<sup>22</sup> Lau and Weng, Climate Signal Detection Using Wavelet Transform..., 2392-2393.

wavelet matches the signal. While there is plenty of flexibility in choosing a function to act as a mother wavelet,  $\varphi(t)$  must satisfy two properties:

1. The function must be centered at zero and in the limit of  $|t| \rightarrow \infty$  the function must rapidly go to zero. This property is responsible for the local nature of the wavelet, which allows us to resolve time dependent features.
2. The function must have zero mean. This is called the admissibility condition, and implies the invertibility of the wavelet transform. By invertibility we mean that the original signal can be reconstructed from the wavelet coefficients obtained by the transform.<sup>23</sup>

There are additional properties that can be satisfied, but these two are essential. It is important to note that in this thesis we do not seek to use invertibility to reconstruct any signals, so the code implementation of the wavelet transform is **not** designed to preserve invertibility.

Two main types of wavelet transforms exist, each suited for different uses. In this study I used the continuous wavelet transform (CWT), which is useful for analyzing time series. The other common type of wavelet transform is the discrete wavelet transform (DWT), which is better suited for decomposition and reconstruction of data with minimal bases.<sup>24</sup> The DWT is widely used for data compression due to its ability to decompose signals into a sparse number of wavelet coefficients, but is less useful for analysis of time series for various reasons. In the DWT the time series is filtered and analyzed at various levels, rather than the direct convolution described above (which is how the CWT is implemented). My implementation of the CWT will be described later.

One of the most important parts of the wavelet transform is the choice of wavelet basis. Many wavelet bases have been developed for various purposes, and each

---

<sup>23</sup> Meyers, Kelly, and O'Brien, *An Introduction to Wavelet Analysis...*, 2859.

<sup>24</sup> Lau and Weng, 2394-2395.

basis is formed from either a mother wavelet that is either complex or real. The advantage of a complex wavelet is that it separates the phase and magnitude of the signal, while real wavelets only record the amplitude. Wavelets should be picked to match the expected signal in the data, as this implies that a large wavelet coefficient corresponds to a strong match between the data and the expected signal.<sup>25</sup> Climate scientists frequently employ the Morlet and Mexican hat wavelets, as they match the expected oscillatory nature of the time series.

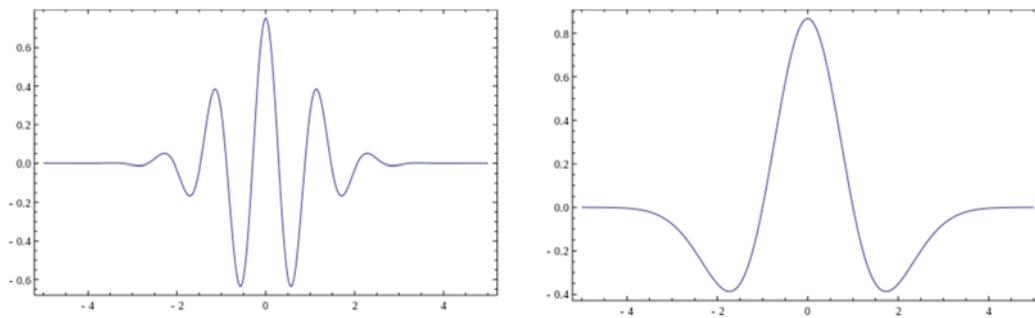


Figure 3: Left, the real part of the Morlet wavelet. Right, the Mexican hat wavelet.<sup>26, 27</sup>

Wavelet choice is particularly important, as the wavelet needs to be interpreted in the context of the data being analyzed. For the Morlet wavelet, it is fairly straightforward to see that there is a relation between the scale of the wavelet (how much it is dilated) and the corresponding Fourier period. For example, at a given scale we can say that the wavelet is able to detect a sinusoidal trend in the data of a certain period. In fact, we can typically give a physical period in units of time that correspond to any given scale. This conversion is detailed in the Convolution Method section of the thesis.

<sup>25</sup> Meyers, Kelly, and O'Brien, 2859-2860.

<sup>26</sup> Wikipedia. <https://upload.wikimedia.org/wikipedia/commons/0/0a/MorletWaveletMathematica.svg>

<sup>27</sup> Ibid. <https://upload.wikimedia.org/wikipedia/commons/0/08/MexicanHatMathematica.svg>

Another wavelet that we considered using was the Harr wavelet. This is simply a step function:

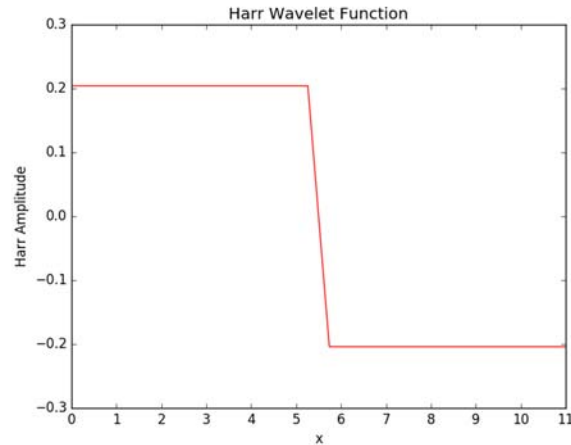


Figure 4: Our Harr wavelet function. Note that the function is zero outside of this plot.

The Harr wavelet is easily interpreted as the difference in averages over a given period. With respect to climate data, this allows us to detect differences in seasonal average behavior or perhaps a change in the baseline of a climate. If a set of data has a linear baseline that changes at some time (e.g. has a slight kink), the Harr wavelet should pick out this change in the baseline. The oscillatory nature of the Morlet wavelet make it insensitive to this kind of signal, so we see that performing wavelet transforms using a variety of wavelets may allow us to pick out different kinds of signals, yielding a more complete understanding of the features in the data.

Many wavelets have highly irregular forms that make them well suited for other types of data analysis, but may prevent them from having a clear physical interpretation in the context of analyzing a time series.<sup>28</sup> We decided to confine our attention to the Harr and Morlet wavelets, as these have waveforms that are physically relevant to the expected behavior of our data set.

---

<sup>28</sup> Meyers, Kelly, and O'Brien, 2865.

## Methods

We implemented a direct convolution algorithm for CWTs using Morlet and Harr wavelets. The code was written in Python, and was mostly a direct implementation of the methods described in Torrence and Compo's paper *A Practical Guide to Wavelet Analysis*. Part of the reason for following their paper is that they sought to establish a standardized procedure for implementing wavelet transforms so that results could be compared between studies. One of the issues that has hindered the acceptance of wavelet transforms is the difficulty in statistically testing results, a difficulty that partly stems from arbitrary implementations of the transform itself.<sup>29</sup> While we do not describe the details of the procedure, we do give a brief overview of our implementation as well as issues and considerations related to the code. A link to the code, hosted on GitHub, is provided in the Additional Materials section of this thesis.

### Morlet Wavelet

The Morlet wavelet is given by

$$\varphi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}$$

where  $\omega_0$  is the nondimensional frequency of the wavelet. This wavelet is simply a complex sinusoid windowed by a Gaussian envelope. We set  $\omega_0 = 2\pi$ , as this gives the easiest interpretation of the phase of the wavelet in terms of the time step  $\eta$ . Since the data is sampled at a frequency of one point per month, moving from point to point in the time series is equivalent to moving 1/12<sup>th</sup> a year. Thus, moving 12 data points is equivalent to one full period of the Morlet wavelet, and one year in the data. The

---

<sup>29</sup> Torrence and Compo, 61.

wavelet is given a scale, which is then converted into the corresponding Fourier period and used to sample the wavelet accordingly. For example, if we want the scale to be 1 year, we sample the wavelet so that it takes 12 points to reach the next local maximum from the center. Therefore, when we have traveled 12 points along the data, this corresponds to the wavelet oscillating once. Likewise, if we want the Fourier period of the wavelet to be 2 years, we sample the wavelet so that it takes 24 points to reach the next local maximum from the center.

To make sure that the wavelets at different scales have the same frequency response, we renormalized the wavelet by setting the total wavelet power to 1 in time space. This ensures that the wavelet has equal response to all frequencies.<sup>30</sup>

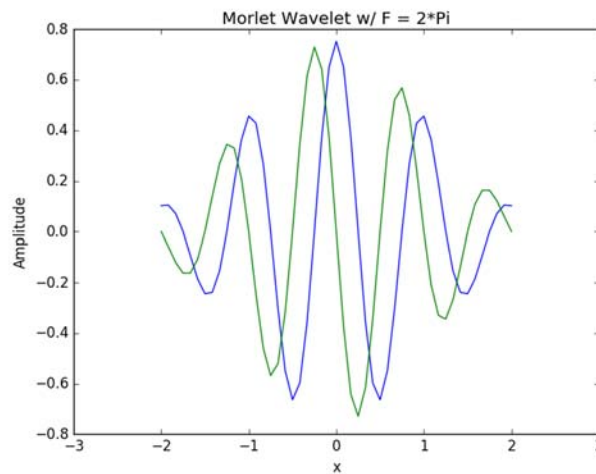


Figure 5: Sampled Morlet wavelet, real part in blue, imaginary part in green. The units of  $x$  are years, so we see that the peaks of the wavelet correspond to one year intervals. This matches the expected annual oscillatory behavior of a climate, but also gives an easy conversion of the scale of the wavelet to the period of oscillation it is detecting in the data.

---

<sup>30</sup> Ibid. 64.

We also implemented a Harr wavelet, but its implementation is essentially the same as the Morlet wavelet.

### **Convolution Method**

We decided to implement the CWT through a direct convolution algorithm following the definition of the wavelet transform. Most CWT implementations use a different method by applying the convolution theorem of Fourier analysis to do the calculations in frequency space, but this requires preprocessing of the data in order to mitigate the edge effects that come from the discretized Fourier transform.<sup>31</sup> The only issue with the direct convolution method is that it is computationally expensive, but the size of the data set we are using is small enough that this is not a severe hindrance. The wavelet coefficients are plotted using the Matplotlib library in Python.

Convolution refers to the process of multiplying two functions together, pointwise. For example, with a given wavelet, we can think of placing the center of the wavelet on top of the data at some point, and then multiplying the value of the wavelet at each point by the value of the data at each point. Summing these values gives us the value of the convolution, at the point where the center of the wavelet is. This value is the wavelet coefficient of the data at that point. By performing this convolution at every point in the time series, we determine the wavelet coefficients of the time series at every point. Recall that the wavelet coefficient is a measure of how well the given wavelet matches the data set at that point. We therefore are directly calculating how well the wavelet matches the data set, at every point along the data set. A large coefficient corresponds to a good match with the data.

---

<sup>31</sup> Meyers, Kelly, and O'Brien, 2860.



Here, the choice of wavelet becomes important. If the wavelet coefficient at a point on the time series is large, our wavelet is a good match with the data, but we need to know what this wavelet means physically to be able to say anything useful. In the case of the Morlet wavelet, when we know the wavelet scale, we can convert to the corresponding Fourier period of the wavelet. To convert from the wavelet scale to Fourier period (in years), we used the conversion formula:

$$\lambda = \left[ \frac{4\pi}{\omega_0 + (\omega_0^2 + 2)^{1/2}} \right] a$$

where  $\lambda$  is the Fourier period in years,  $a$  is the scale parameter, and  $\omega_0$  is the frequency of the wavelet (in this study, always set to  $2\pi$ ).<sup>32</sup> The Fourier period of the wavelet corresponds to the period of oscillation of the Morlet wavelet. For example, if we know that the Fourier period of the wavelet is 1 year, then the Morlet wavelet oscillates once over one year of the data set. Thus, a large wavelet coefficient at a given time in the data for the 1 year Morlet wavelet tells us that there is a strong 1 year oscillatory cycle in the data at that point. If the wavelet coefficient is positive, the phases of the cycles are the same, and if the wavelet coefficient is negative, the phases of the cycles are perfectly out of phase. Similarly, a large wavelet coefficient for a 10 year Morlet wavelet would tell us that at that point, there is a strong 10 year oscillatory signal in the data. A large wavelet coefficient for a Morlet wavelet of a given period at a given point corresponds to the presence of a signal with that period at that point in the original data set. This is why we are mostly considering the Morlet wavelet.

---

<sup>32</sup> Ibid. 2865.

For the CWT, we perform this convolution across the entire data set to get the wavelet coefficients for our wavelet at each point of the data set, but we also perform the convolution for a whole range of scales for the wavelet. This is why we use a mother wavelet, as dilating this mother wavelet allows us to detect different “periods” of the same signals. In the case of the Morlet wavelets, the period of the wavelet actually corresponds to the period of an oscillatory signal in the data, but the “period” (width of the wavelet) for other wavelets may correspond to signals that do not have a well-defined physical period.

The overall method for the CWT is summarized in the following steps:

1. Choose a wavelet and a range of scales (corresponding to different dilation factors of the mother wavelet) for the wavelet.
2. Convolve the wavelet with the lowest scale with the first data point of the time series. Record the wavelet coefficient at this point.
3. Slide the wavelet to the next point on the time series, and convolve again. Record the wavelet coefficient at this point.
4. Repeat 3. until we reach the end of the time series.
5. Scale the mother wavelet to the next highest scale in our range.
6. Repeat 2-4. for all the scales in our range.

This method of performing the CWT has the drawback that it is extremely redundant.

We do not need to calculate the wavelet coefficients at every single data point, but it is the easiest way to think about finding signals and it is the method whose results are the easiest to interpret. This comes at the loss of efficiency in our computation and an inability to accurately reconstruct the data from our wavelet coefficients. Fortunately, neither of these drawbacks are a concern for this study, as we are looking at a relatively small data set and we are not immediately interested in reconstructing the data.

Since the Morlet wavelet is a complex-valued function, the wavelet coefficients that we are calculating are also complex valued. Instead of looking at the actual wavelet coefficients, we typically use the wavelet power, which is the squared magnitude of the wavelet coefficient. This number will be real, but has the drawback that it is positive definite (always greater than or equal to zero, being zero only when the coefficient is zero as well). Thus, a negative coefficient will give a positive power, and we lose information about the relative phase of the signal. If we were interested in the phase of the signal picked up by the Morlet transform, we could easily decompose the wavelet power into the real and imaginary magnitudes, and detect the phase of the coefficient from those. We initially ignored the phase to simplify the plots given by the transform, but in the future work section of the thesis we describe reasons to switch to using the real part of the Morlet wavelet to detect the phase.

### *The Cone of Influence*

There is an issue with analyzing near the edges of the time series that involve the wavelet support going over the edge of the data set. A variety of methods can be used to help deal with this, but the most common is to simply pad the edge of the data set with zeros. This produces an edge effect commonly called the “cone of influence” (COI), where the zero padding serves to dampen the wavelet coefficients due to part of the wavelet picking up only zero values. The cone of influence will often be marked on the wavelet coefficient plots. Note that the COI’s effect increases with decreasing distance to the ends of the time series. The size of the cone of influence also tells us the decorrelation time for a spike in the data. The decorrelation time is the time it takes for a random spike in the time series to stop influencing the wavelet coefficients. This is

essentially the width of the wavelet function at a given scale, but is useful for determining if a set of large wavelet coefficients is due to noise or an actual signal. If the region of large wavelet coefficients is smaller than the COI's size, then it is likely to be the result of noise.

### *Statistical Testing*

It is also possible to develop statistical tests for determining a 95% confidence level for signal detection. Many geophysical processes can be approximated by red noise, a simple model of which is the univariate lag-1 autoregressive process:

$$x_n = \alpha x_{n-1} + z_n$$

where  $\alpha$  is the lag-1 autocorrelation,  $x_0 = 0$ , and  $z_n$  is taken from a Gaussian white noise series. We thus have the null hypothesis that the data is created by a stationary red noise process, and test our results against this assumption. Gilman et al. derived the expected Fourier power spectrum for such a red noise process,<sup>33</sup> while Torrence and Compo show that this power spectrum is also expected for the wavelet transform.<sup>34</sup> Assuming that the real and complex parts of the wavelet coefficient are normally distributed, the squared magnitude of the coefficient should follow a chi-squared distribution with two degrees of freedom. Multiplying the expected power spectrum by the corresponding chi-squared value gives an expected background spectrum for the wavelet coefficients.

We developed code to implement the statistical testing methods described in *A Practical Guide to Wavelet Analysis*, but found that while the method works on raw

---

<sup>33</sup> Gilman et al, On the power spectrum of "red noise."

<sup>34</sup> Torrence and Compo, 70.

data, z-scored data does not fit the hypotheses required to perform the tests. The z-score transform highly correlates the data, as every data point is normalized to the mean of the entire data set. This violates the required assumption that the data is a lag-1 autocorrelated process, and we cannot use the expected power spectrum that was developed by Gilman et al. In fact, it is uncertain whether we are able to perform any rigorous statistical tests on the z-scored data at all. This is a serious problem for applying the wavelet transform, as we would like to create confidence intervals to determine when detected signals are likely due to background noise. Therefore, the raw data set is essential for performing rigorous wavelet analyses, while the indexed data is only useful for rough analyses. Unfortunately, we were unable to obtain the raw data for this thesis.

While not statistically rigorous, a possible method for filtering out signals due to random noise would be to perform wavelet transforms on large sets of randomly generated data that have the approximate statistical properties we would expect the raw data to have. By z-scoring the randomly generated data, performing the wavelet transform on each trial, and then averaging the resulting wavelet coefficients, we can construct a reasonable approximate to the wavelet power spectra that random data would have given the assumed statistical properties. Setting this approximate power spectra as a baseline gives a way of testing if a signal is likely due to noise or is an actual event. Unfortunately, this is merely an approximation, and also requires prerequisite knowledge of the statistical properties of the raw data. Ultimately, if possible, it is much better to perform wavelet analyses on raw data rather than the indexed data.

Most time series are extremely noisy, which makes it difficult (if not impossible) to see any underlying trends in the data. This is frequently solved by smoothing the data. It is also possible to smooth the wavelet coefficients to reduce the influence of spikes in the time series. Note that smoothing the time series before applying the wavelet transform should always be avoided, as smoothing will remove any high frequency content in the data and bias the transform results.

## **Results and Discussion**

This thesis focused on three main groups of results:

1. Comparison of results of wavelet transforms of z-scored data to raw data.
2. How basic trends in data affect the wavelet coefficients.
3. Wavelet analysis of the Northeast climate metrics.

Note that all wavelet transform results discussed are for the Morlet wavelet unless otherwise specified. While we were interested in using the Harr wavelet, the data is simply too noisy to be able to extract meaningful results from the Harr wavelet transform here. We start with an example to show what information wavelet transform plots encode.

### Example Wavelet Analysis Plot

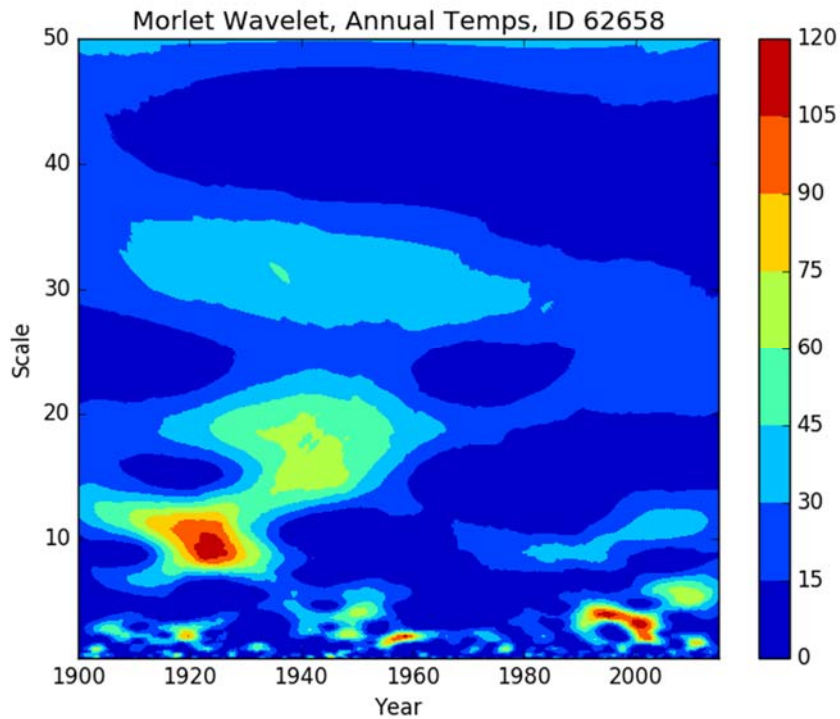


Figure 6: Example wavelet coefficient plot. These are the wavelet coefficients for a single weather station (USHCN ID 62658) for the stations monthly max temperature since 1900.

The x-axis for the wavelet coefficient plots corresponds to the year in the time series, while the y-axis corresponds to the scale of the wavelet. In many cases, we convert the scale directly to the Fourier period in years, as the Fourier period is physically relevant. In Fig. 6, we performed the CWT using a Morlet wavelet on a single weather station's average monthly max temperature records. Places where the wavelet coefficients are high mean that the wavelet was a strong match to the data at that point, and in the case of the Morlet wavelet, we know that an oscillatory signal of the Fourier period corresponding to the given scale is present in the data at that point.



The wavelet coefficients are not normalized, as we are mostly interested in their relative magnitude. A possible issue with only considering the relative magnitudes is that we suppress signals of small relative magnitude to the whole set of wavelet coefficients, when that signal is significant compared to the signals like it. For example, if we considered the Morlet wavelet with a period of 5 years and saw that for most of the time series it had a very small wavelet coefficient, but between 1960 and 1970 it had a wavelet coefficient that was one or two orders of magnitude larger, we would think that there was a strong signal in the data corresponding to the five year oscillation of the data between 1960 and 1970. However, if we also were looking at the wavelet coefficients for the Morlet wavelet with period of 20 years, and saw that it had a wavelet coefficient that was much larger than the maximum wavelet coefficient of the 5-year period Morlet wavelet at some point, this would cause the entire 5-year period Morlet wavelet signal to be suppressed in the data due to the relative normalization of all the wavelet coefficients. For our purposes this is not a problem, as the relative normalization tells us which scales are the largest influencers of the climate at any given time, but if we wanted to do a more detailed analysis, it would be important to keep this fact in mind. If we expected there to be a 5-year oscillatory signal in the data, restricting our wavelet scale range to be close to 5 years would give better results for determining if and how the 5-year oscillatory signal changes throughout the time series.

We should note that the cone of influence, which we have not marked in Fig. 6, affects the data near the edges of the time series. For wavelets with a Fourier period of approximately 50 years, the cone of influence affects the entire time series, so scales above approximately 45 are usually ignored in this thesis. This is also above the

Nyquist frequency of the data, so we are unable to reliably detect any signals with Fourier periods above 50 years. If we had a time series that extended further into the past, we would be able to detect longer period oscillations due to the decreased influence of the cone of influence on the data set. We generally do not include the cone of influence in our plots, though it is important to be aware of its presence.

Specifically considering Fig. 6, we see that there is a major 10-year oscillatory signal in the data around 1920, and possibly significant oscillatory signals of 15-20 year periods between 1920 and 1940 and 30 year periods between 1920 and 1980. There is also a noticeable 10-year period signal between 1980 and 2000. We could also look at the signals with periods below 10 years, but the amount of noise present in the data makes it hard to know whether a large wavelet coefficient for one of these signals is due to an actual signal or simply inherent random noise. This is the weakness of performing wavelet transforms of z-scored data as opposed to raw data: if we were considering the raw data, we could generate an expected power spectrum for the wavelet coefficients and have a rigorous way of determining if small-period oscillatory signals were due to an actual signal or noise. With the z-score data, the best we can do is provide an approximate test for filtering out false positives at any level. For the larger period oscillatory signals we expect that they will not be due to inherent noise in the data, as random noise should cancel out over large periods of time.

### **Wavelet Transforms of Z-scored Data versus Raw Data**

The database of climate data analyzed in this thesis uses the z-scored average monthly precipitation, maximum temperature, and minimum temperature. We used a

Monte Carlo method to determine the relationship between the wavelet coefficients of z-scored data versus the wavelet coefficients of raw data:

1. Simulate 100 sets of random time series data with the same statistical properties.
2. Perform the wavelet transform on each of these random time series individually.
3. Average the coefficients of all the randomly generated data sets to get the raw data wavelet coefficients.
4. Perform the z-score transform on each of the time series individually.
5. Perform the wavelet transform on each of the z-scored time series individually.
6. Average the wavelet coefficients of the z-scored time series to get the z-score wavelet coefficients.

This Monte Carlo simulation yielded the following plot:

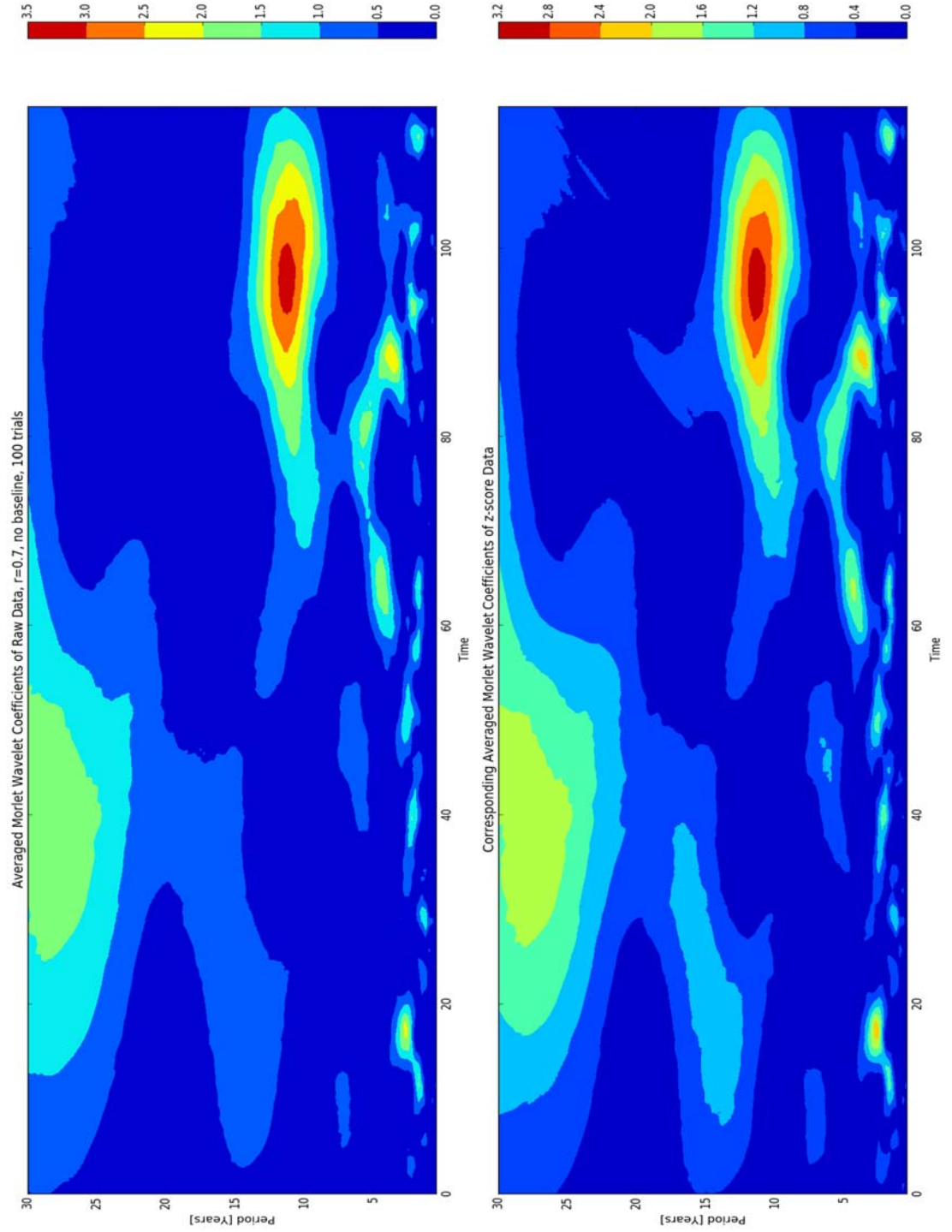


Figure 7: Comparison of the wavelet coefficients of raw data to wavelet coefficients of z-scored data.

This is the central result of this thesis. There is a direct correlation between the raw data wavelet coefficients and the z-scored wavelet coefficients. The minor differences that appear in the two plots are because the wavelet coefficient amplitudes for the raw data are slightly larger than for the z-scored data, but since we are not concerned with coefficient amplitudes in this thesis, this difference is irrelevant. From this simulation, we can be confident in assuming that the z-score transform does not add any artifacts to the wavelet analysis. Thus, a peak in the wavelet coefficients for the z-scored data corresponds to a peak in the wavelet coefficients for the raw data, and the z-score transform will not suppress any signals present in the raw data. This justifies making conclusions about the original data, and thus the physical climate system, when using results from a wavelet analysis of the z-scored data.

Note that in the Monte Carlo simulation, the wavelet transform was applied to the individual data sets, and then the wavelet coefficients were added afterwards. This would be like taking the wavelet transforms of the individual station data, then averaging the coefficients to get the overall wavelet coefficients for all of the stations. Our implementation of the wavelet transform takes the average of the station's data and then applies the wavelet transform to the averaged data, giving us the overall wavelet coefficients for all the stations. Since convolution is a bilinear process, these are equivalent procedures for calculating the wavelet coefficients. Though these two methods will give the same wavelet coefficients, it is much more efficient to average the data first, then apply the wavelet transform, as the wavelet transform algorithm only needs to run once. In the implementation used for the Monte Carlo method, the wavelet transform algorithm ran 100 times, and thus was incredibly slow.

## **Influence of Trends on Wavelet Analyses**

To assist in interpreting the results of wavelet analyses of climate data with the intent of detecting fundamental changes in the system (climate change), it is useful to look at basic signals and determine how they affect the wavelet coefficients. The most basic trend that could exist in the data that would signal a shift in the behavior of the climate is a linear trend. This could be caused by external forcing of the climate under consideration. An example of this external forcing could be intense urbanization in a region, causing a steady increase in the minimum and maximum temperatures due to the urban heat island effect.

The influence that a basic linear trend exerts on the Morlet wavelet transform is simple: linear trends bias the larger scale wavelets. This assumes that the linear trend is subtle relative to the data, otherwise the linear trend has a significant influence over all scales. The general rule of thumb is that the greater the slope, the smaller the scale that will have an increased overall wavelet power. Also note that this is considering a linear trend with zero overall mean. It is preferable to perform the wavelet transform on data that has had any linear baseline removed from it, as this baseline will artificially inflate the wavelet coefficients when it is not near zero. Here, near refers to the amplitude of the linear trend, relative to the signal's amplitude. Near the zero intercept, the influence of the linear trend is negligible due to the wavelet's zero mean, but away from the zero intercept the wavelet coefficients are enlarged. This generally suppresses signals that are near the points of intersection. Small slopes will not cause serious issues with biasing the wavelet coefficients, but large slopes should be removed from the data.

A few basic examples help demonstrate how a linear trend influences the wavelet coefficients of a signal. First, we consider a pure sinusoidal signal's wavelet transform:

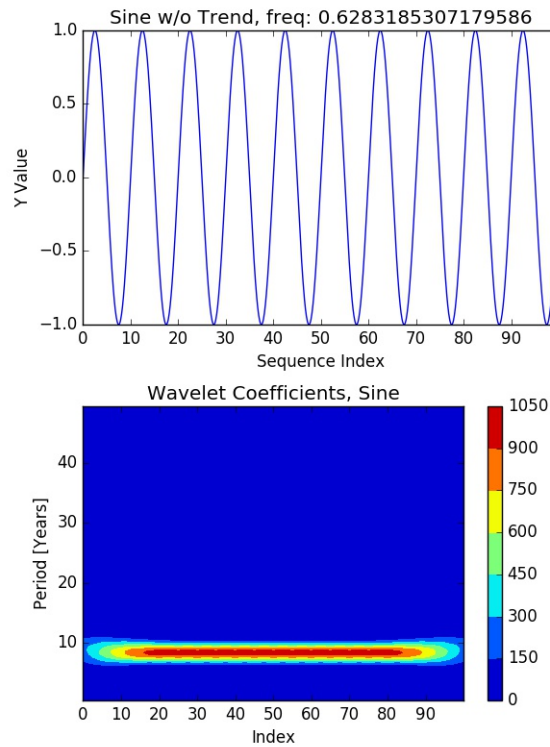


Figure 8: Pure sinusoidal signal with wavelet transform. Index refers to the number of the point in the time series. Note that the wavelet coefficients are not purely in the 10-year frequency range, but are localized around it. This is because of the redundancy of the CWT used. A 9.5-year period wavelet is also a good match to the data, as is a 10.5-year period wavelet.

If we add a linear trend with a significant slope compared to the amplitude of the data, a noticeable amount of power appears for the larger period wavelets. The spread around the 10 year period is due to the previously noted redundancy of the CWT. Wavelets that have a period close to 10 years will have a decent, but relatively weak, match to the signal, while the 10-year period wavelet will have the best match.

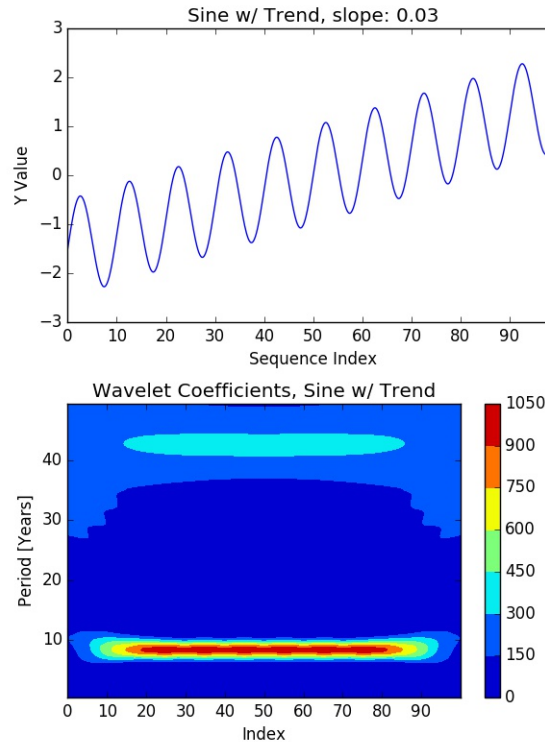


Figure 9: Same sinusoidal signal from Figure 8, but with a linear trend added.

The decreased amplitude of the wavelet coefficients in the 40-year period range near the edges of the plot are due to the cone of influence. The fringes at the edges of the plot for the 30-year period wavelet demonstrate the amplification of the wavelet coefficients as the linear trend moves away from zero. Because these fringes are in the cone of influence, they correspond to a better match than the coefficients imply, as their amplitude is significantly dampened by the zero-padding of the time series. This plot also demonstrates that the cone of influence is more severe for the larger period wavelets at a given time (index). The 40-year period wavelets would have greater power near the edges of the time series than near the center (index 50), but the cone of influence has a very strong dampening effect on the larger period wavelets at the edges.



Real climate data is not purely sinusoidal; in fact, it is frequently assumed to be well described by a random red-noise distribution. Our final example is simply a linear trend with red-noise added.

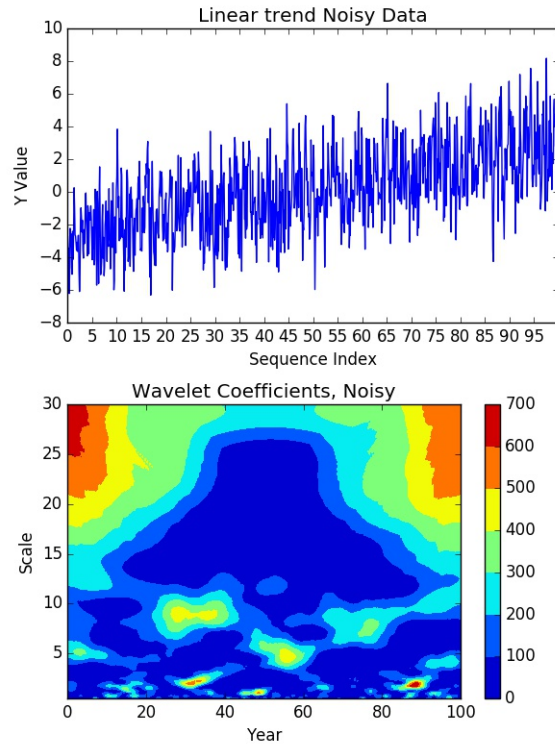


Figure 10: Similar linear trend as in Figure 9, but with a twice the slope and a red-noise time series added.

The same behavior from Figure 9 is present in Figure 10. By definition, a red-noise process has power at all scales, but there is more power in the larger scales than the smaller scales. This is clear for the 25+-year period wavelets, where the wavelet coefficients are significantly larger than zero everywhere. The edges also show the large bias introduced by the linear trend. For a pure noise time series, there should be equal power at all scales but there should be no obvious structure. In Figure 10, there is a clear increase in the power of the coefficients for the larger scale wavelets at the edges, despite being in the cone of influence. This corresponds to the parts of the time series

that are farthest from zero, confirming our claim that the wavelet power will be biased towards the parts of the linear trend that are away from the zero-intersection point.

Note that these examples have assumed that the linear trend in the data is present throughout the entire data set. In general, there could be a series of underlying linear trends with different slopes during different time intervals. Because wavelets have localized support, a change in the slope of a linear trend will not introduce any new behavior other than that illustrated in Figure 9 and Figure 10. A change in the slope of the linear trend will correspond to an increase or a decrease in the rate at which the wavelet power is biased at lower scales.

These results motivate a useful rule of thumb: wavelet power at large scales, if more present at the edges of the time series, likely correspond to the presence of a linear trend in the data.

### **Wavelet Analysis of the Northeastern United States Climate**

Wilkie's thesis created a climate index for the Northeastern United States using four metrics: average monthly maximum temperature, average monthly minimum temperature, average monthly precipitation, and average monthly snowfall.<sup>35</sup> Since the snowfall records are not complete, we decided to focus specifically on how the maximum and minimum temperature and precipitation metrics behaved. Typically, these metrics have a variety of oscillatory trends driven by various natural processes. We would expect to detect persistent signals at a few scales that perhaps shift over time, along with independent signals that correspond to anomalous events such as a droughts. Plots in the references listed at the end of this thesis clearly illustrate these expected

---

<sup>35</sup> Wilkie, *Climate Change in Northeastern United States*.

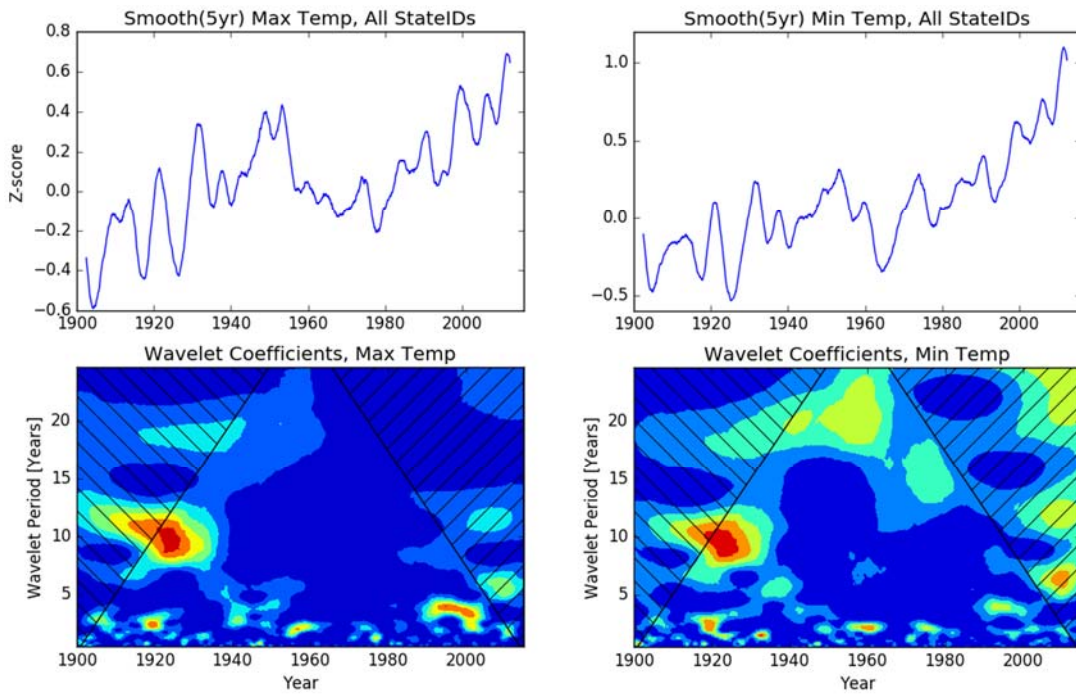
results.<sup>36</sup> Persistent signals can usually be explained by a known cycle, with an obvious example being the annual variation in temperatures due to the Earth's axial tilt and annual orbit about the sun. Longer period oscillations can be correlated to known natural processes that drive the climate's longer term behavior. We first consider the overall Northeastern United States climate behavior, and then divide the region by the Appalachian Mountains and consider these two subregions separately.

### *Overall Northeastern Climate*

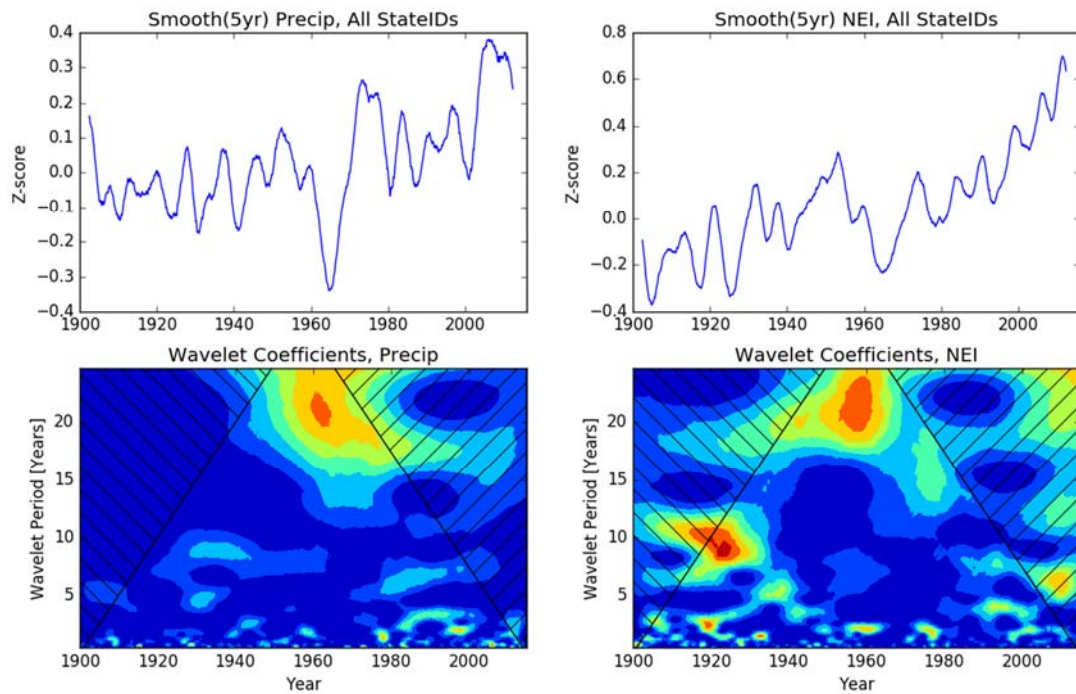
To determine the overall Northeastern climate, we averaged the monthly metrics for all 137 weather stations in the database developed by Wilkie and performed the wavelet transform on each metric. We also averaged the metrics together to make the Northeast Climate Index (NEI). The following plots show the z-scored data, smoothed with a Gaussian envelope of width of 5 years, and the Morlet wavelet coefficients of the unsmoothed data.

---

<sup>36</sup> In particular, see Lau & Weng, *Climate signal detection...* and Torrence & Compo, *A practical guide to wavelet analysis*.



Figure(s) 11, 12: Averaged monthly maximum and minimum temperature for all weather stations. The dashed regions denote the cone of influence.



Figure(s) 13, 14: Averaged monthly precipitation and NEI for all weather stations.

Figure 11 shows that the average monthly maximum temperature for the Northeast region as a whole has surprisingly little consistent variation. There is a very strong 10-year oscillation around 1920, but this cycle does not persist past 1940. Several events appear for wavelets below the 5-year period range, but these are usually just due to a combination of noise and the expected annual cycles. The large variation in the maximum temperatures from 1900-1930 and after 1990 likely obscure any relatively minor cyclical behavior in the time series. Note that while the smoothed time series appears to show a strong linear trend throughout, the linear trend is mostly only apparent from 1900-1945 (where it is very strong), and after 1980 (where it is also very strong). Thus, the power in the 20+-year period range from 1900-1940 is likely a product of the strong linear trend present in the data there. We do not observe a similar level of power in the 20+-year period wavelets for the trend starting in 2000 because the trend stays near zero until 2000, so the trend's influence on the coefficients is minimal during this time period. Ultimately, the wavelet transform does not reveal much about the overall maximum temperature behavior that the smoothed time series does not show in this case.

Figures 12 and 13 show significantly more oscillatory behavior. Both the precipitation and minimum temperature metrics show a strong 17-year period signal starting around 1930. This signal appears where the linear trend in the data is near zero, so it is likely that it is due to a real climate cycle rather than an artifact introduced by the linear trend. The significant drop in precipitation is explained by the fact that the Northeast faced a severe long term drought in the 1960's. Throughout the precipitation metric there also appears to be an approximately 7-year period cycle, which later

appears in the maximum and minimum temperature metrics after 1990. After the 1960 drought, this 7-year oscillation drops to a 5-year oscillation, but then gradually morphs back into a 7-year cycle.

Finally, the NEI wavelet analysis combine these trends together. The strong 17-year oscillatory signal is present, along with the weaker 7-year period cycle. The NEI suggests that the 17-year oscillatory signal we see in the minimum temperature and precipitation metrics is present throughout the entire time series. This cycle starts with a period of 17 years in 1900, then slowly morphs into a cycle with a period of 20 years by 1960. After 1960, two major 17-year and 25-year oscillations are present. A possible explanation is that there is only a 17-year oscillatory signal until 1960, where the drought causes a significant event that obscures the introduction of a new 25-year cycle not present before 1960. The NEI also shows that the 7-year oscillation is present throughout the entire time series, but is weak during the drought. A 13-year cycle, originally only apparent throughout the minimum temperature's time series, is reinforced by weak 13-year cycles that appear at different times in the precipitation and maximum temperature metrics. This suggests that the 13-year cycle is not only a part of the minimum temperature metric, but its absence in the second half of the maximum temperature metric and first half of the precipitation metric is strange. After 1970 the NEI has a strong linear trend that increases the power of the wavelet coefficients at all periods near the end of the time series, explaining the largely uniform increase in power after 2000. This strong linear trend is driven by the very strong trend in the minimum temperature metric and assisted by the weaker trend in the maximum temperature metric.

Combining these observations, we conclude that there is likely a strong 17-year oscillation throughout the entire time series, and a weaker 7-year oscillation that disappears during the 1960 drought. The drought may have obscured the introduction of a 25+-year oscillation, and a weak 13-year oscillation may be present throughout the data as well. After 1990, a strong positive linear trend increasingly biases the wavelet coefficients at all scales.

A potential problem with using the NEI to investigate the change in the Northeast's regional climate is that it obscures what specific metrics are changing. The NEI can be interpreted as a measure of the climate, in that positive values correspond to hotter and wetter periods, and negative values are colder and dryer periods, but we see in Figure 12 and Figure 14 that the minimum temperature metric has a larger influence of the resulting NEI wavelet coefficients than the other metrics. We need to be careful when we say that because the NEI has a 13-year oscillation throughout the data, as this doesn't necessarily correspond strictly to hotter and wetter periods followed by colder and dryer periods, since the individual metrics do not seem to directly reflect that statement. The precipitation metric does not have a 13-year oscillation before 1960, while the maximum temperature does not have a 13-year oscillation between 1940 and 1980. Generally, we should compare the NEI's results to the individual metric's results to make sure that signals are a product of all three metrics, rather than just one very strong signal in one metric.

### *Difference In Climate for West/East Appalachian Regions*

Wilkie noted that “there may exist climatic sub-regions within the NE.”<sup>37</sup> The Appalachian mountain range divides the Northeast region into two major geographic regions. To determine if the climate cycles present in the region west of the Appalachians differ from those present in the region East of the Appalachians, we filtered the climate stations into two groups based on their position relative to the Appalachian Mountains:

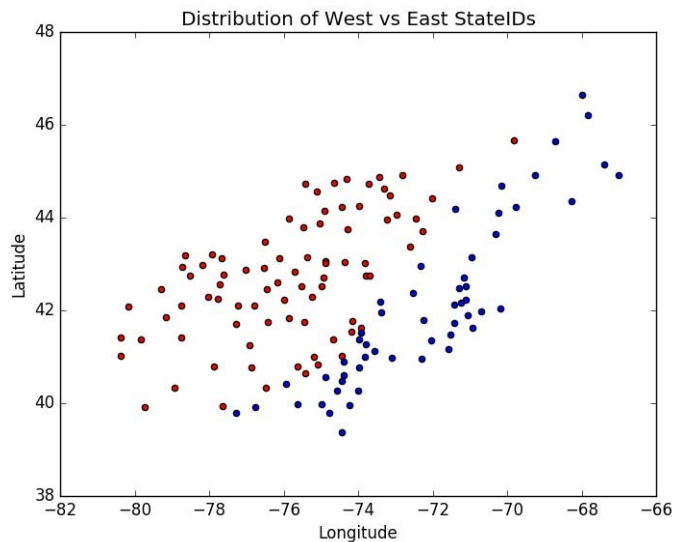


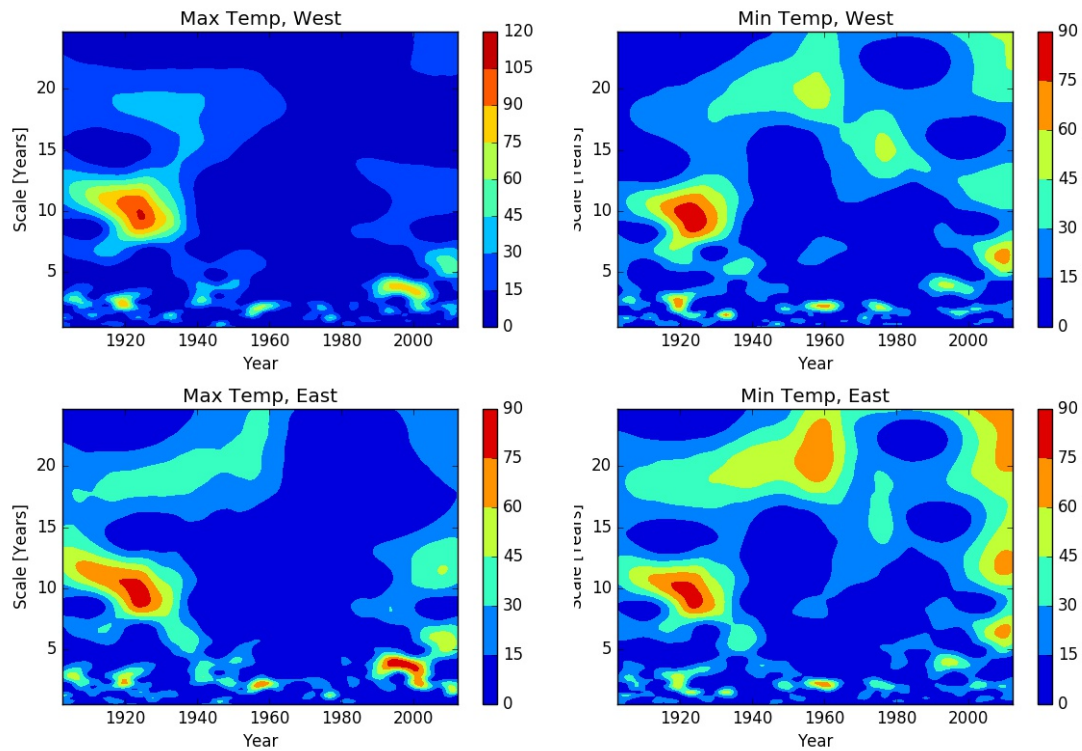
Figure 15: Distribution of USHCN stations relative to the Appalachian Mountains.

Using this sorting method, there were 83 stations West of the Appalachians, and 54 stations East of the Appalachians. These groups are sufficiently large that we can use our previous conclusion that the z-scored wavelet coefficients are approximately the same as the raw data wavelet coefficients.

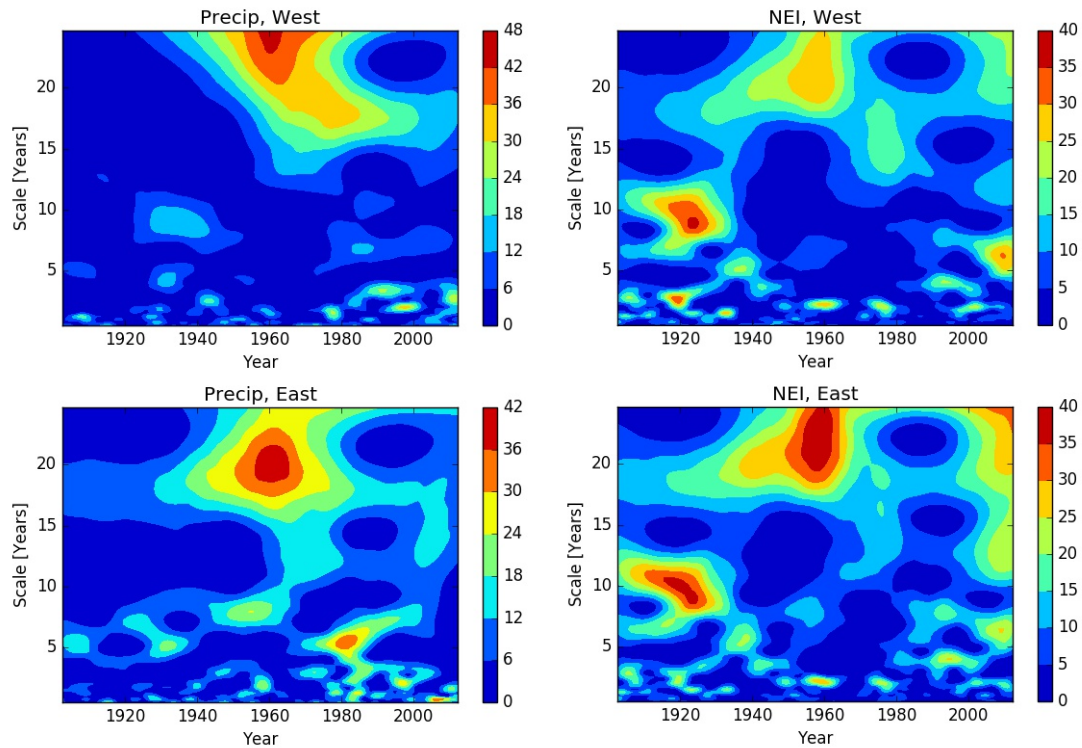
---

<sup>37</sup> Wilkie, *Climate Change in the Northeastern United States*, 38.





Figure(s) 16, 17: Maximum and minimum temperature wavelet analyses for stations East and West of the Appalachian Mountains.



Figure(s) 17, 18: Precipitation and NEI wavelet analyses for stations East and West of the Appalachian Mountains.

The cone of influence is not included on these plots, as they affect both plots identically and we are only interested in differences between the regions.

The maximum temperature wavelet analysis gave nearly identical results for each region, except in two places: between 1940 and 1960, the Eastern region had a moderately strong 25-year cycle that was not present in the Western region, and after 2000 the Eastern region had a significant 13-year cycle that was much stronger than in the Western region. Note that the 25-year cycle appears where the cone of influence ends, so it may have been present for longer in the Eastern region than implied by the plot.

In the minimum temperature wavelet analysis, there is a clear difference between the Eastern region and the Western region from 1900 to 1960 with respect to the 17-year cycle observed in the overall data. The Eastern region had a better match with the 17-year period Morlet wavelet until the drought, at which point the two regions behave nearly identically at all levels. After 2000, the Eastern region has greater wavelet power at all scales, which suggests that the linear trend in the minimum temperatures in the East was greater than the linear trend in the minimum temperatures in the West. Other than the difference in the strength of the 17-year oscillation from 1900-1960, the two regions behaved identically with respect to minimum temperature.

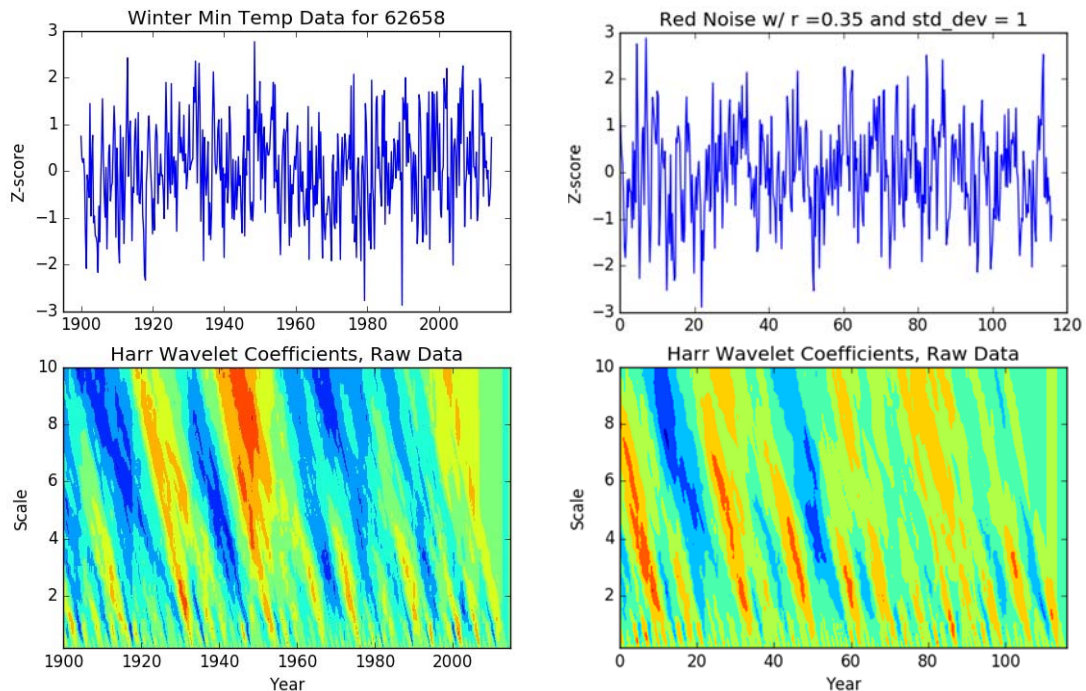
The precipitation wavelet analysis shows that the Western region's precipitation metric was dominated by the 1960 drought, with no other consistent behavior before 1960. After 1960, there is a strong 17-year cycle that is also present in the Eastern region. There is also a weak 7-year cycle that is much stronger in the Eastern region after 1970. The most noticeable difference between the regions is that the Eastern

region has a consistent 17-year cycle throughout the entire data set, but after the 1960 drought the Western region appears to pick up this oscillation as well. The Eastern region also has a consistent 7-year oscillation that is not consistent in the Western region. The wavelet power increases at the larger scales for both regions, with the East experiencing a slightly larger increase in power. This suggests that the Eastern region had a larger change in precipitation after 2000 relative to the Western region.

Finally, the NEI wavelet analysis gives slightly contradictory results. The large signal at 1960 appears to suggest that the Eastern region was more heavily affected by the 1960 drought than the Western region, while the individual metrics show the opposite behavior. This may be because there is a 25-year oscillation in the Eastern temperature metrics that are present throughout the data, but are suppressed by the cone of influence. When the metrics are combined to create the NEI, these cycles add to push the Eastern NEI wavelet power at 20-25-years above that of the Western NEI wavelet power. Note that the Western NEI wavelet analysis showed small signals in this scale range at 1960 for the temperature metrics, but a very large signal in the precipitation metric, which we would expect for a drought. Thus, the large signal in the 20-25-year range for the NEI in the Eastern region during 1960 is due to a combination of the weaker drought and 25-year oscillations in the temperature metrics, but the relatively smaller signal for the Western region is almost entirely due to the drought. However, other than the difference in the signal at 1960 the two regions behave almost identically with respect to the NEI until 2000. The NEI also shows that the Eastern region had a more severe linear trend than the Western region after 1990, with the difference driven by the difference in the linear trends in the minimum temperature metric.

### *Harr Wavelet Applied to Single Station Data*

In all the previous results, we used the Morlet wavelet to pick out oscillatory signals that appeared in the data at different times. We would also like to be able to determine if the difference in the averages, over a given length of time, changes throughout the data set. A physical example would be determining if the average seasonal temperature differences were different during the first few decades of the data set than the last few decades. The Harr wavelet detects these changes. With the Harr wavelet, scale directly translates to the period of time that the wavelet spans. A scale of two years means that the wavelet is comparing the average of the first year to the average of the second year.



Figure(s) 19, 20: Harr wavelet applied to average monthly minimum temperature for one weather station, versus applied to a red noise sequence of same length and statistical properties.

Unfortunately, the climate metrics used in this study are too noisy to return useful results for the Harr wavelet. The Harr wavelet applied to the climate data time series gives results that are very hard to distinguish from pure noise, as demonstrated in Figures 19 and 20. There is so much oscillatory behavior present in these time series that the Harr wavelet coefficients constantly detect these oscillations, which is exactly what the Morlet wavelet does. These oscillations in the Harr wavelet coefficients obscure any other type of signal that we may be interested in. It may work better to consider the Harr wavelet at specific scales to see how the metrics change with respect to that one scale, rather than performing a CWT across a whole range of scales. We leave this to be included as future work, as we are more interested in determining shifts in periodic behavior, a task better suited for the Morlet wavelet.

## **Conclusion and Future Work**

### **Summary of Results**

Quantitative tools for detecting changes in regional climates are in high demand to help policy makers develop plans for mitigating the local effects of climate change. Wavelet transforms, once a hot topic in math and physics research, never stuck as a central tool in climate science. This thesis sought to provide motivation for the application of wavelet transforms to climate time series with a focus towards detecting shifts in oscillatory trends in the data. Since climate data are being compiled into climate indices to better encapsulate the overall behavior of a climate, we showed that the wavelet transform applied to sufficiently large sets of climate indexed data will give results that can be interpreted in terms of the original raw data. However, wavelet transforms are significantly more useful when applied to raw data, as statistical testing methods can be applied that allow us to make rigorous conclusions about signal detection. For climate indices, we can only make approximate conclusions about the difference between a signal being due to noise or real physical behavior.

To help with the interpretation of wavelet analyses, we have shown that a linear trend biases wavelet power for longer period wavelets when the linear trend is not close to zero (relative to the data's original amplitude). As most of the climate metrics we investigated in this thesis do have a linear baseline, this helped us filter out some areas of increased wavelet power as due to a linear baseline, rather than the introduction of new oscillatory behavior.

In terms of analyzing the behavior of the Northeastern Climate region, our wavelet analyses showed that for the Northeast region as a whole, there is a strong 17-

year oscillation persistent throughout the NEI, a weaker 7-year oscillation, and a possibly strong 25-year oscillation that appears after the 1960 drought. While in this case the NEI is a decent descriptor of the overall Northeastern climate, it generally is a good idea to investigate the individual metrics that make up the index. Ultimately, the NEI wavelet coefficients are simply the averaged wavelet coefficients of the individual metrics, so only looking at the NEI can hide the source of a signal.

When considering the regions in the Northeast that are East and West of the Appalachian Mountains, we found that they exhibited relatively similar behavior except for the presence of a possibly stronger 25-year oscillation in the Eastern region. The Western region was more heavily affected by the 1960 drought, and after 1960 had a significant 17-year oscillation that was previously only present in the Eastern region. The Eastern region is also experiencing a more severe linear increase in its NEI and minimum temperature metrics.

In conclusion, this thesis has shown that wavelet transforms are a useful tool for investigating the behavior of climate indices, and could be a powerful way of rigorously detecting shifts in regional climates.

## **Future Work**

There are many possible directions for future work. New wavelets could be implemented to detect other behavior in the indices, and the current method for applying the Morlet wavelet could be better optimized to decrease the cone of influence. This would allow for the resolution of larger period signals. An algorithm could be implemented to correct for the cone of influence on wavelet power, as the power damping is linear with respect to the distance from the edge of the time series.

Unfortunately, this would introduce arbitrary choices into the data analysis, as the way that the data behaves outside of the recorded time series is not known, but such an algorithm could help with detecting larger period signals.

The most important next step would be to obtain the original data. This would allow for a comparison of the wavelet analyses for the z-scored data versus the raw data, and allow us to perform statistical testing to see if the conclusions we made using the z-scored data are supported by rigorous results from the raw data.

It would also be useful to explain the trends that we found in the data by known physical mechanisms. We could also attempt to extrapolate the current trends by taking the wavelets with the greatest power from the last few decades and reconstructing the major behavior of the climate metrics. This could be relevant for policy makers, as it gives them a rough short term model to inform mitigation policies. Unfortunately, this data set is sufficiently noisy that it might be difficult to distinguish relevant short term signals from noise without filtering the data or using the raw data.

Other wavelet studies have used the Morlet wavelet, but only looked at the real part and ignored the imaginary part. This has the advantage, over the power of the wavelet coefficient, of showing whether the wavelet is in phase or out of phase with the signal. This makes it more clear when the wavelet is oscillating many times with the data, or if it is just a strong match at a single point. Plotting the wavelet coefficients with finer resolution would also allow us to track shifts in cyclical behavior better. Lau and Weng detail a method that tracks the change in major oscillatory signals, and it could be useful to try and mathematically track the signals determined for each metric.



Finally, we could perform a similar wavelet analysis of other United States climate indices to see if the methods used in this study give comparable results for other data sets.

## Bibliography

- Baliunas, S., P. Frick, D. Sokoloff, and W. Soon, 1997: Time scales and trends in the central England temperature data (1659-1990): A wavelet analysis. *Geophys. Res. Lett.*, **24**, 1351-54.
- Birk. K., A.R. Lupo, P. Guinan, and C.E. Barbieri, 2010: The interannual variability of Midwestern temperatures and precipitation as related to the ENSO and PDO, *Atmòspha*, **23**, 95-128.
- Burrus, C. Sidney, Ramesh A. Gopinath, and Haitao Guo. 1998. Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall.
- Ebbesmeyer, C.C. and R.M. Strickland. 1995. Oyster Condition and Climate: Evidence from Willapa Bay. Publication WSG-MR 95-02, Washington Sea Grant Program, University of Washington, Seattle, WA.
- Frumhoff, P.C., J.J. McCarthy, J.M. Melillo, S.C. Moser, and D.J. Wuebbles. 2007. Confronting Climate Change in the U.S. Northeast: Science, Impacts, and Solutions. Synthesis report of the Northeast Climate Impacts Assessment (NECIA). Cambridge, MA: Union of Concerned Scientists (UCS).
- Frumhoff, P. C., J.J. McCarthy, J.M. Melillo, S. C. Moser, D. J. Wuebbles. C. Wake, and E. Spangler-Siegfried, 2008: An integrated climate change assessment for the Northeast United States. *Mitig. Adapt. Strat. Glob. Change.*, **13**, 419-423.
- Gilman, D. L., F. J. Fuglister, and J. M. Mitchell, Jr., 1963: On the Power Spectrum of "Red Noise". *Journal of the Atmospheric Sciences*, **20**, 182-184.
- Haberman, Richard. 2013. Applied Partial Differential Equations. Pearson.
- Kunkel, K. E., L. E. Stevens, S. E. Stevens, L. Sun, E. Janssen, D. Wuebbles, J. Rennels, A. DeGaetano, J. G. Dobson. 2013. Regional Climate Trends and Scenarios for the U.S. National Climate Assessment Part 1: Climate of the Northeast U.S. NOAA Technical Report NESDIS 142-1.
- IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp.
- IPCC, 1990: Climate Change: The IPCC Scientific Assessment. [Houghton, J. T., G. J. Jenkins, and J. J. Ephraums (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 412pp.

- Meyers, S.D., B. G. Kelly, and J. J. O'Brien, 1993: An Introduction to Wavelet Analysis in Oceanography and Meteorology: With Application to the Dispersion of Yanai Waves. *Monthly Weather Review*, **121**(10), 2858-2866.
- National Assessment Synthesis Team, *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change*, US Global Change Research Program, 400 Virginia Avenue, SW, Suite 750, Washington DC, 20024.
- Lau, K.-M. and H.-Y. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.*, **76**, 2391-2402.
- Ogden, Todd. December 1, 1996. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser.
- Percival, Donald B. and Walden, Andrew T. June, 2006. *Wavelet Methods for Time Series Analysis*. Cambridge University Press.
- Polikar, Robi. "The Wavelet Tutorial." Accessed Sept. 9, 2016. <http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>. Hosted by Rowan University. Images used with permission.
- Santos, C. A. G., C. O. Galvão, K. Suzuki, and R. M. Trigo, 2001: Matsuyama City Rainfall Data Analysis Using Wavelet Transform. *Ann. Jour. Hydr. Eng.*, **45**, 211-216.
- Rogers, Harvey, 2005: Assessing the robustness of the Pacific Northwest Climate Index and its potential to predict Northwest Climate variation [thesis]. [Eugene (OR)]: University of Oregon.
- Torrence, C. and G. P. Compo, 1998: A practical guide to wavelet analysis, *Bull. Amer. Meteor. Soc.*, **79**, 61-78.
- Walnut, David F. January 27, 2004. *An Introduction to Wavelet Analysis*. Birkhäuser.
- Wikipedia. *Real-valued Morlet wavelet*. 2016. [https://en.wikipedia.org/wiki/Morlet\\_wavelet#/media/File:MorletWaveletMathematica.svg](https://en.wikipedia.org/wiki/Morlet_wavelet#/media/File:MorletWaveletMathematica.svg)
- Wikipedia. *Mexican hat*. 2016. [https://en.wikipedia.org/wiki/Mexican\\_hat\\_wavelet#/media/File:MexicanHatMathematica.svg](https://en.wikipedia.org/wiki/Mexican_hat_wavelet#/media/File:MexicanHatMathematica.svg)
- Wilkie, Adrien Ann, 2010: Investigating Regional Climate Change in Northeastern United States [thesis]. [Eugene (OR)]: University of Oregon.