Topological Data Analysis for Rossby Wave Estimation

by

Richard J Ross

(Under the Direction of Nicole A. Lazar and Lynne Seymour)

Abstract

Multiscale methods are a widespread topic of interest in statistical research. One particularly valuable aspect of these methods is to identify meaningful features in data across different scales, which allows practitioners to interact with their data in an interpretable manner. We present a multiscale method related to the growing topic of topological data analysis (TDA) which we call a sequential filtration. This method provides a multiscale application of TDA principles, which serves as a powerful tool in estimation of a response surface. We apply these methods to atmospheric data, specifically geopotential heights in the northern hemisphere. After describing a variety of applications of sequential filtrations, we show the performance of each in estimating the number of Rossby waves on a daily basis, leading to a discussion of the utility of multiscale persistence generally.

Index words: Persistent Homology, Morse Filtration, Sequential Morse Filtration, Rossby Waves, Climate Statistics

Topological Data Analysis for Rossby Wave Estimation

by

Richard J Ross

B.S., Brigham Young University-Idaho, 2014

M.S., University of Georgia, 2017

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2022

Topological Data Analysis for Rossby Wave Estimation

by

Richard J Ross

Major Professor 1:	Nicole A Lazar
Major Professor 2:	Lynne Seymour
Committee:	Dan Hall
	Thomas Mote
	Jaxk Reeves
	Paul Schliekelman

Electronic Version Approved:

Ron Walcott

Vice Provost for Graduate Education and

Dean of the Graduate School

The University of Georgia

May 2022

Acknowledgments

Throughout the process of writing this dissertation, I received support from many. I'd like to express gratitude to a few of them here.

First, I'd like to acknowledge and express gratitude to my advisors, Dr. Lynne Seymour and Dr. Nicole Lazar. Dr. Lazar's advice, pragmatism, and diligent encouragement helped me to dive into TDA. Your suggestions and ideas continually helped me to step forward instead of backward, and have impacted my approaches not just to statistical research, but other aspects of my professional life as well. Dr. Seymour has constantly been a source of creativity and practicality. Your counsel and good humor have helped me focus on the best goal of all: progress.

I'd like to express my appreciation for all four members of my committee, Dr. Dan Hall, Dr. Thomas Mote, Dr. Jaxk Reeves, and Dr. Paul Shliekelman. In addition to serving on the committee, each contributed very insightful ideas throughout the process and aided in my graduate studies. I am especially grateful to Dr. Mote helping me to better understand the science governing atmospheric circulation, including answering questions about small details.

I'd like to acknowledge the immeasurable support I received from my family. My wife, Laura, has been actively supportive through every step of this process; your enthusiasm for our family helped this work come to fruition. My children,

Zander and Evangeline, also supported me by sacrificing time together and I appreciate their patience which often belied their ages.

I'd also like to acknowledge the support of my parents, Dr. Brant Ross and Janna Ross. Your support has been unending, and your legacy serves as a powerful reminder of what we can accomplish with dedicated hard work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many different scientific endeavors, data are collected and researchers use them to seek answers to many different questions. Great science is performed in countless subject areas and there is no single way to describe all valid research questions. We have found that one of the more common questions could be described as a question of "What in these data is a true pattern with some cause, and what parts of these data are just noise"? In this dissertation, we will present one specific incarnation of that problem from atmospheric science. There's been a good variety of work done towards answering this question, and we hope to highlight many of the current approaches to providing an answer. We will then proceed to highlight some of the challenges that these approaches face. It turns out that this question is a hard one to answer!

We will then provide some information on a growing topic in statistical research named Topological Data Analysis (or TDA), which we feel is well-positioned to answer this question. We will summarize some of the foundations of TDA, moving towards a description of uses of TDA which are used on data somewhat similar to the real data at hand in the motivating scenario. We will

describe why work up to this point in TDA is not quite equipped to handle the data at hand. The solution to this question was not sitting ready-made for use.

In subsequent chapters, we will discuss how we're adapting TDA methods to directly answer the scientific question at hand. Namely, how many Rossby waves are present in the northern hemisphere at a given time? We'll go into more detail about Rossby waves and their importance later in this chapter. We'll describe a new use of TDA that allows us to estimate them at scale consistently with clear specifications, something that current answers to the question of counting these waves sometimes struggles to do.

1.1 Motivation for Analysis

Since we wish to develop methods that help us to model the structure of atmospheric circulation, it is first important to provide a description of how that circulation system functions, at least at a basic level. The two largest contributors to this system are the temperature gradient, which we explain below, and the Coriolis force, which is the frictional force on the air due to the rotation of the earth.

We'll go into how researchers currently estimate Rossby wave frequencies. We'll describe some of the challenges one must overcome in such estimation and will highlight a few key issues that arise from these current estimation methods, which seem to create and solidify disagreement among those who study the polar stream and related topics.

1.1.1 Rossby Waves and the Polar Jet Stream

Since we wish to develop methods that help us to model the structure of atmospheric circulation, it is first important to provide a description of how that

circulation system functions at least at a basic level. The two largest contributors to this system are the temperature gradient, explained below, and the Coriolis force, which is an apparent force on the rotating earth.

The earth's atmosphere is composed of a mixture of gases whose particles flow freely above the surface of the earth. Generally speaking, warm air (which often carries more moisture) is more bouyant and tends to rise, then spread poleward, which explains the Hadley circulation. Cool air which is often less moist tends to contract in volume. Since the equator is generally warmest, warm air tends to expand outward from the equator towards the poles. The atmosphere is generally warmer towards the equator where there is a surplus of net radiation and cooler towards the poles where there is a radiation deficit at a given elevation. The observed dynamics of this exchange of warm and cool air operate in specific ways in certain parts of the atmosphere, and atmospheric models help describe how this exchange functions de Silva and Carlsson [6].

The first layer of the atmosphere, which extends from the surface of the earth to an elevation of about 12 km is called the troposphere. In the troposphere, a first-level approximation for atmospheric circulation is known as a three-cell model (Figure 1.1). This model is not fully accurate, but is useful in considering circulation within each hemisphere. In this model, each hemisphere is divided into three regions: the tropical region ($\approx 0^\circ - 30^\circ$ N), the mid latitudes ($\approx 30^\circ - 60^\circ$), and the polar region ($\approx 60^\circ - 90^\circ$). In the tropical and polar cells, air close to the surface travels towards the equator until reaching the boundary between cells, at which point air tends to travel upwards, cooling and spreading poleward. This cooler air cycles towards the pole until it cools and starts to sink, getting closer and closer to the surface. As this occurs, the air starts to warm and travels poleward again. In the midlatitudes, warm air at the surface travels poleward, while cool air at higher altitudes travels towards the

equator. The polar, mid-latitude, and tropical cells are known as Polar, Hadley, and Ferrell cells respectively, and the exchange of air between the cells tends to spread heat and moisture in the atmosphere which helps to transport moisture across the surface of the earth.

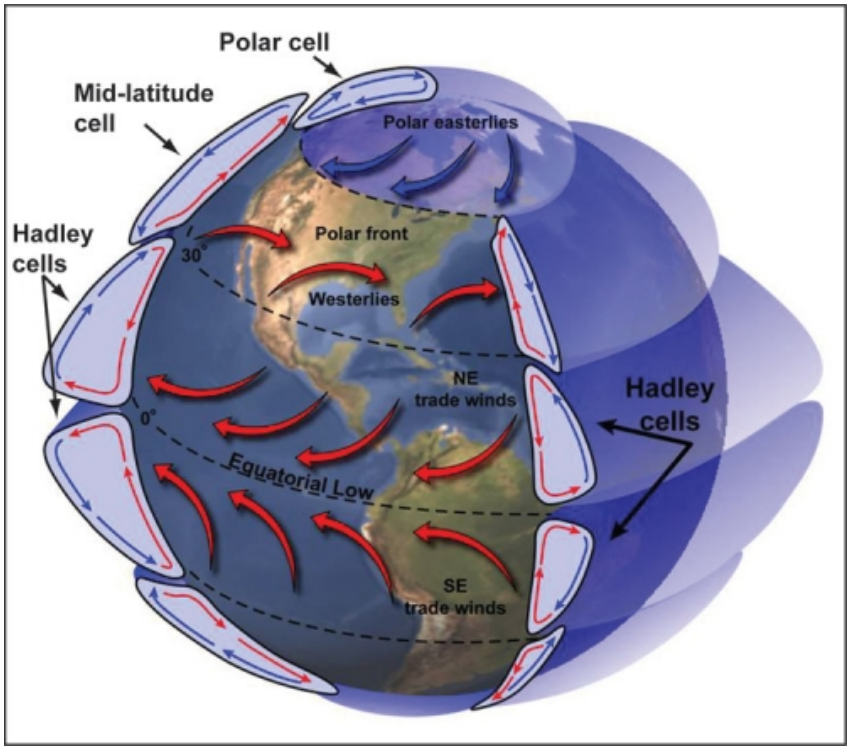


Figure 1.1: Three-Cell Model for atmospheric circulation. Image Credit Piana [20]

The three-cell model helps to describe how air circulates from north to south, but air certainly flows along latitudinal lines as well. As the earth spins daily on its axis, the planet itself rotates in an anticlockwise direction when viewed from space. The Coriolis force deflects motion to the right in the Northern Hemisphere and to the left in the Southern Hemisphere Winds traveling south from the equator deflect to the left. The opposite is true for winds traveling from the poles, further cementing the tendency for wind to flow from west to east. Wind can flow in any direction, but the Coriolis force deflects large scale motion relative to the flow that would be imposed by the pressure gradient.

The jet stream is a band of fast-flowing air traveling from west to east across the northern hemisphere. This band of air is focused in the midlatitudes ($\approx 30 - 60^\circ$ N). The actual north-south extent of the jet stream is not usually the entire range of latitudes of the cell, but it may deviate from its central latitude by as much as 15°. On a planetary scale, the flow in the atmosphere is approximately in geostrophic balance, which is a balance of the forces produced the pressure gradient and the Coriolis effect. The thermal wind is the vector difference between the lower- and upper-troposphere geostrophic wind. The increase in wind speeds with height in the mid-latitude troposphere, which leads to the jet stream, is produced by the temperature gradient and described by the thermal wind.

Atmospheric circulation tends to follow a model described in Palmen [19], which is a more accurate representation of ciculation than the three-cell model. The dominant circulation is the Hadley cell in the tropical belt, while the middle latitudes, circulation is more driven by thermal wind. This revised model is more accurate at explaining circulation in the northern hemisphere.

Large scale meanderings of the polar jet stream in the northern hemisphere are known as planetary waves, the most important form of which are quasi-stationary mid-latitude Rossby waves. The number of Rossby waves present in the jet stream, or wavenumber, is in almost constant flux. Waves travel through the midlatitudes, forming new waves or combining multiple waves together which increases and decreases the total number of waves present. This is especially evident seasonally.

In summer, the poles are warmer than they are in winter, while the equator receives the same amount of solar heat year round. Temperature gradients are accordingly greater in the winter and lower in the summer. Since higher temperature gradients correspond to greater differences in pressure and den-

sity between the polar and equatorial air masses, this causes a smaller number of high amplitude Rossby waves. In contrast, since summer generally experiences a lower temperature gradient, the decreased temperature gradient results in slower zonal wind speeds, which better support more waves, as seen in the Rossby wave equation. Wavenumber is typically a number between zero and ten. This number could be zero in the case of symmetric flow. Wavenumbers of one-three are sometimes called long waves, while wavenumbers of four-ten are sometimes called synoptic waves.

1.1.2 Understanding Rossby waves

Literature on the number of waves present on each day listed in a systematic manner is lacking, in part because the determination of the number of waves has been somewhat subjective. One of the goals of this dissertation is to provide a more rigorous and repeatable way to estimate or calculate the number of waves present in a given isobaric surface, which is a surface in the atmosphere where the pressure is equal everywhere along that surface. Although data for various isobaric surfaces is available, we utilize the 500 hectoPascals (hPa) surface, which is the set of locations in the atmosphere where the pressure has been measured at 500 hPa. Various work in the field sometimes uses other heights, but 500 hPa is a commonly-used pressure value for this type of work. This is because this surface is approximately in the middle of the troposphere and this is a level where we tend to see a minimum in divergence in windflow.

Recent work has pointed to the importance of knowing the number of waves, as an increased number of waves is associated with persistent episodes of extreme weather Mann et al. [15] because move waves tend to be slower moving, as mentioned earlier. This extreme weather which includes extreme precipitation and temperature events Screen and Simmonds [23], results from slower

moving waves (when there are more waves, their speed is slower than when there are fewer waves). The movement of Rossby waves has been associated with differences in North American and European weather patterns Kaspi and Schneider [13] as well as serving as one indicator for the extent of circulation between cells. Figure 1.2 shows two images of the jet stream which help to highlight the type of variation which can be expected in the number and amplitude of Rossby waves. One question which we'd like to focus on is how one can count these waves and how to identify which north south deviations in the polar jet stream constitute Rossby waves and which are small deviations due to other forces.

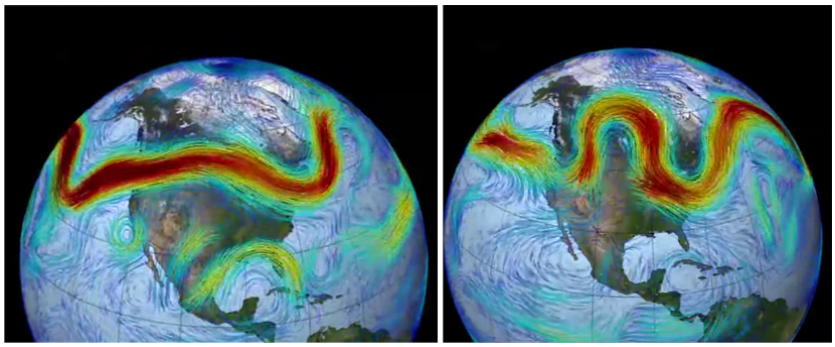


Figure 1.2: Two still images from NASA's Visualization Explorer Project titled "Aerial Superhighway"Viz [26]. Note that the same geographic region can exhibit highly variable behavior in Rossby wave amplitude, frequency, and the overall jet stream size.

One of the main tools used to understand the number of Rossby waves present at a given time is the following wave equation:

$$\omega = Uk - \beta \frac{k}{k^2 + l^2}, \tag{1.1}$$

where ω is the frequency of the waves, k and l are the zonal and meridional wavenumbers, U is the mean wind speed, and β is called the Rossby parameter and is calculated as a function of the angular speed of the Earth's rotation, the

latitude where the jet stream is centered, and radius of the earth. The zonal wavenumber denotes the total number of trough ridge pairs which can be observed around the globe at a given latitude. Rossby waves contribute to the zonal wavenumber but so does any other cause of local maxima minima in the wave form. Meridional wavenumber refers to the number of troughridge pairs which are observed around the globe at a given longitude. This number is typically small.

Using this equation, one can theoretically calculate the frequency of Rossby waves at a given time and latitude/longitude on the earth. However, this equation is bound to provide different values of ω at different locations on the earth, as the zonal and meridional wavenumbers vary from one location to the next, as well as the varying radius of the earth, which is larger at more equatorial locations and smaller at more polar locations. There are certainly some locations which one would never choose when calculating ω, like locations at the actual equator or at the north pole, both of which are not suitable locations to locate or count Rossby waves, but should one attempt to calculate that number at 30° N, 45° N, 60° N, or somewhere in between? There is not a single approach which provides a uniformly correct value for ω in the way that one might hope it would.

One of the specific obstacles in calculating wavenumber here is that as seasons change, the jet stream shifts in some predictable ways. One of these shifts which affects the choice of location when calculating the estimated wavenumber is that the center of the jet stream, which is sometimes close to 45° N, travels south in the winter and north in the summer by as much as 10°, as seen in Figure 1.3. Even if the jet stream exhibited the same wavenumber in winter and summer, using the same location and time of day during those two days may yield a very different frequency. That is not to say that the wave equation is incorrect , but

its use may produce results which count the frequency of non-Rossby waves during some portion of the year while counting the exact Rossby wavenumber at other times in the year. Essentially, the nature of this wave structure does not lend itself to the use of a single location, or even a single latitude.

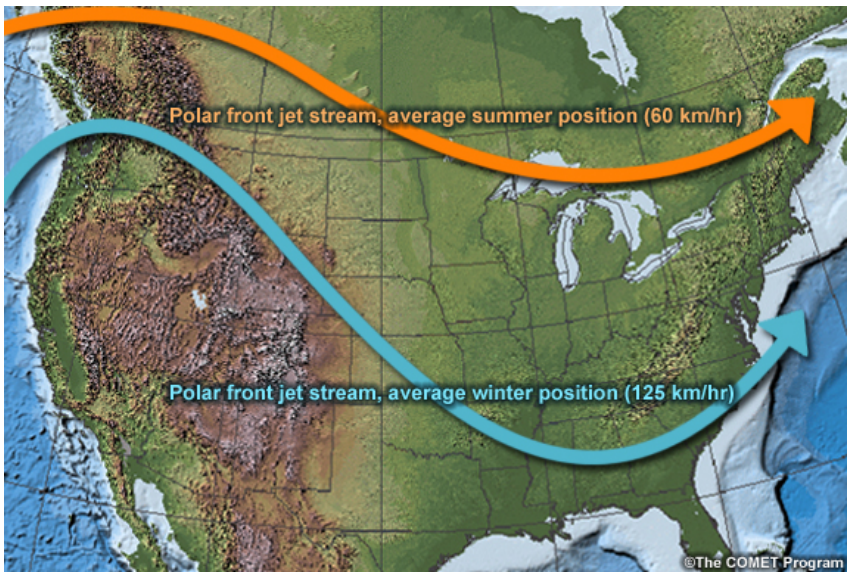


Figure 1.3: An illustration of seasonal trends in the center of the polar jet stream from North Carolina State University's climate office. In winter months when the temperature gradient is highest, the jet stream resides further south. In summer months when the gradient is smaller, the jet stream moves north.

This dynamic nature of where the polar jet stream resides brings up the issue of how to estimate waves in a scalable way. If a research group wanted to get an estimate for the wavenumber, it is not enough to simply count troughs and ridges at a single latitude, as we could perhaps try if the jet stream's center line were stable. Instead, it seems that researchers need to either pick general latitudes that seem to work well for certain time periods (such as the seasons), or to manually "find" the centerline for each isobaric surface or wind map and make estimations from there. While this approach may work quite well when examining a small number of surfaces, it becomes quite time consuming at scale and may not be preferable.

Wind maps can help to provide useful information about the prevailing behavior of the jet stream, but without a concrete set of criteria for identifying Rossby waves, this method can sometimes lead to the estimated wavenumber falling subject to the eye of the beholder, making the judgement of the person reading the wind map the only source on which to rely when finding the frequency. If a new person reads the same map, they may not agree on the number of waves.

One alternative to wind maps is known as an isobaric surface. This surface is obtained by first selecting a level of barometric pressure (such as 500 hPa), and then observing the elevation above sea level at which the given barometric pressure is observed. Isobaric surface data are publicly available through radiosonde observations that became routine in the late 1940s, as well as atmospheric reanalysis products, such as the NCEP/NCAR reanalysis project Kalnay et al. [12]. These data (which date from before 1955 through the present year) are provided with accompanying information on how they were obtained and estimated. Figure 1.4 shows the midlatitudes of one such surface. Isobaric surfaces illustrate the movement of the jet stream in addition to Rossby waves. These surfaces are also a scalar valued object which make analysis more direct than a vector-valued object like a wind map.

In this isobaric surface, 500 hPa heights are highest closer to the equator (the bottom of the figure). This makes sense as warmer temperatures correspond to higher pressure throughout the troposophere due to the hypsometric relationship. There is a prevailing wavelike formation which corresponds to the jet stream. This formation does not look the same as one might expect from wind maps. On an isobaric surface, the prevailing trend of values is a decrease when traveling from the equator to the north pole, with warps in the surface due to

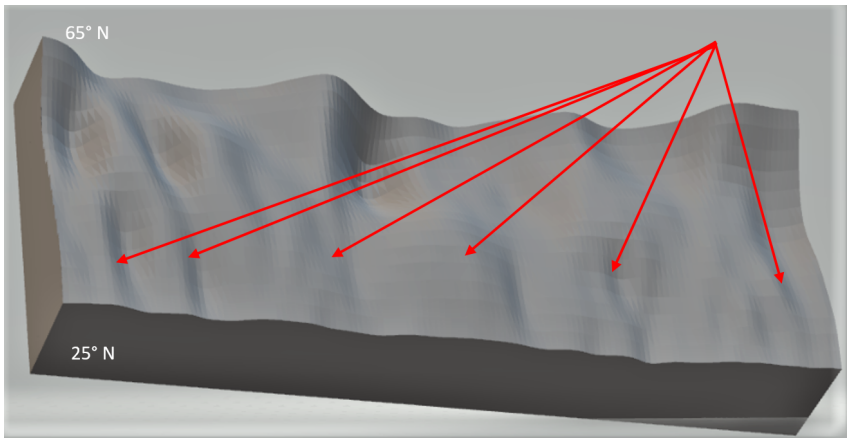


Figure 1.4: An image of the smoothed isobaric surface for Oct 11, 2016 between $25^\circ N$ and $65^\circ N$ based on NCEP/NCAR reanalysis data from NOAA [12]. The pattern of Rossby waves is not obvious, but some waves (red arrows) are noticeable as ripples along the horizontal midline of this surface (which is where the jet stream should be centered)

variation in pressure from the general temperature trend. This includes some long waves which correspond to Rossby waves and some other features.

The dynamic nature of atmospheric circulation makes finding or reporting a single elevation which always corresponds to the location of the jet stream impossible. The pattern of waves on the date shown in Figure 1.4 (October 11, 2016) is apparent; some waves wander further north than others, but there are five or six waves present, as seen by the vertical "ridges" along the isobaric surfaces which correspond to the north-south meanderings of the jet stream.

At this point, there are two very important issues to address. The first deals with the way that wavenumbers are currently estimated, and the second deals with the use of TDA in this application. We discuss the first issue in this chapter, while chapter 4 will address the second question in more detail.

If scientists have associated wavenumbers with extreme weather and other phenomena, then one might assume that a single method already exists for

calculating the wavenumber. Unfortunately, this is not the case. From what we have seen in the literature, wavenumber calculations are performed using a variety of methods. Anecdotally, climatologists may look at images such as those seen in Figure 1.2 and report the number of waves they can see using "the eye test" or some other rule which may or may not be the same as other researchers. This could include counting the number of wave/trough pairs at a given latitude such as $45°N$ or at varying latitudes as mentioned above. These methods may be prone to errors in observation and are not necessarily reproducible, as the rules for selecting a latitude are not well-defined and different groups may use different rules. Relying on judgement, no matter how grounded in experience, decreases the reproducibility of research by not providing a consistent method for estimating the true frequency ω.

The second method involves Equation 1.1. The problem here is the same as we noted before: there is too much chaotic behavior in the atmospheric system for this wave equation to be totally precise due to other forces which act on the jet stream in unpredictable ways. This method improves on the previous method in some ways. The most obvious improvement relates to reproducibility: If one used this equation to compute the Rossby wavenumber and reported the day, time, and location used in computing that wavenumber, anyone else could perform the same calculations and obtain the exact same value. The downside of this equation was mentioned before but is perhaps best illustrated by asking how one picks the location and time used in computation. Is there an automated way to pick a location? Will that location always work? These questions point out some room for improvement in finding Rossby wavenumber.

There are two additional ways which are currently employed for measuring wavenumber which we've found. The first is fairly straightforward: one can

define a specific contour, such as the 5400m contour, and a specific latitude, such as 50°N. Wavenumber is then computed by dividing the number of times that the countour crosses the given latitude, then dividing that number by 2. Thomas LeBlanc [25] shows one example of this method.

Another method is based on using spectral analysis. Researchers have examined data from projects such as the Poseidon data from the North Atlantic to find signal the frequency/wavenumber spectrum to calculate wavenumber. One example of this method is found in Osychny and Cornillon [17], where the authors found the highest amount of Rossby wave activity in the midlatitudes below 40°N.

Consequently, current estimates of wavenumbers are not necessarily consistent from one research group to the next especially when a north/south meander is on the border between being classified as a Rossby wave or not. Some using the eye test might agree or disagree on the number of waves, with no way to settle which individual's eyes did a better job at discerning the wavenumber. Others may use the wave equation, but disagree on the location or time to use when computing frequencies.

Currently used quantification methods open the door for researchers using the same data to report differing wavenumbers, which can lead to confusion. In Screen and Simmonds [22] the authors investigate the relationship between a phenomenon called Artic amplification, which describes the faster near-surface warming in Arctic latitudes as opposed to equatorial or mid-latitudes. Part of this work focuses on one way of looking at Rossby waves proposed earlier in [9] and compares it to a new manner described in [22]. In the conclusion of that paper, Screen and Simmonds note that "[these] two different measures of amplitude show notably different trends, and in many cases the trends are of opposing sign". While this work is a bit more focused on measuring the

amplitude of waves as opposed to wavenumber, it is clear that work in this field is still not quite sure of how best to measure various attributes of these waves.

Clearly, there is room for improvement in the way that wavenumbers are reported and computed. The inherent variability in the number and size of waves suggests that statistical techniques may help to quantify wavenumbers in a consistent and reproducible way. We will discuss how Morse theory and persistent homology can be applied to this question.

1.2 Foundations of Topological Data Analysis

Topological Data Analysis (TDA) is a growing field of research for both mathematicians and statisticians. TDA gives a framework for recognizing and summarizing the structure, or shape, of data, often considered from a geometric perspective. As the name implies, this area of research has at its roots the mathematical field of topology, which includes the study of homological features of an object. Homological features are critical defining characteristics of an object or dataset; if two objects have the same homological features, then one can be transformed into the other by expanding/compressing different portions of that item. The development of topology in relation to data analysis capitalizes on knowledge that some homological features present in data are not meaningful; some data may appear to have different topology even though the main features between two observations may be similar. Properly developed methods can filter out features which are not meaningful in discerning between objects, thus maximizing the ability to make correct comparisons.

TDA approaches have been used in many applications, such as 3D image recognition Singh et al. [24], asthma endotype classification Hinks et al. [11],

identification of the level of fragmentation in the U.S. Congress P. et al. [18], and predicting fluid flow through various materials Moon et al. [16].

Much of the literature developing TDA focuses on this framework's capabilities as a method of exploring data or extracting insight from data, but statistical inference via TDA has much less presence in the literature. We will proceed to discuss the theoretical framework for TDA, especially as outlined in Ghrist [10] and Carlsson [3]. This framework includes an overview of how to construct a simplicial complex (a mathematical structure constructed from the points in a dataset), using persistent homology to describe and understand each of the homological features of the data, and how to summarize that homology with a persistence diagram or barcode, two of the most well-known graphics used in TDA.

1.3 The Simplicial Complex

As a first step in an approach to analyzing data, one may think of data as points in a high dimensional cloud. Before examining the shape or specific features of this cloud, one may first want a consistent way to describe whether two points or a group of points are connected to one another. The mathematical simplex and related complex will help achieve this goal. A simplex is a collection of points in space as well as the line segments that attach these points. A simplicial complex is a set composed of any number of simplices. Any dataset can be divided into simplices based on some criterion such as a distance measure. Any number of the resulting simplices (up to and including all of them) can be considered together as a simplicial complex.

Many data can be considered as a point cloud. In this case, one can consider the Euclidean distance between points as a way to decide whether two points

are part of the same shape or structure in the cloud. For a given threshold, simplexes can be constructed using rules based on this distance, and then the collection of non-empty simplexes is considered as the simplicial complex for the given distance threshold and data. Note that one or many singleton simplexes may exist and are included in the complex even though they do not hold much insight for most applications. There are many different approaches which can be used to construct this complex, but we'll just describe two here, as explained in Ghrist [10].

One approach for constructing this complex (called the Cĕch complex) constructs each simplex by considering the pairwise distance between the points. A second method for generating the simplicial complex (called the Vietoris-Rips or Rips complex) depends on distance between each individual pair of pointsCarlsson [3]. For both of these approaches, we consider our data of dimension d as occupying points in a Euclidean space E^d, even though topology has applications in non-Euclidean space.

The Cĕch complex is constructed by selecting a filtration value ϵ, and then forming simplexes within the data based on the points in the cloud whose closed ball neighborhoods of radius $\epsilon/2$ have at least one point of intersection. This method uses the Euclidean distance measure and can be thought of as creating a sort of Venn diagram with balls around each point with radius ϵ. The simplex exists if and only if there is some point or collection of points of total overlap amongst the points in question. This means that several points sampled at even spacing from the circumference of a circle could potentially be in several simplexes without all being in the same simplex.

One desirable property of the Cĕch complex results from the Cĕch theorem (also called the nerve theorem) Bjorner [1], which states that the Cĕch complex and the original data cloud have the same homology . In essence, performing

data analysis which uses the Čech complex is possible because the complex is capable of capturing the shape of the original space.

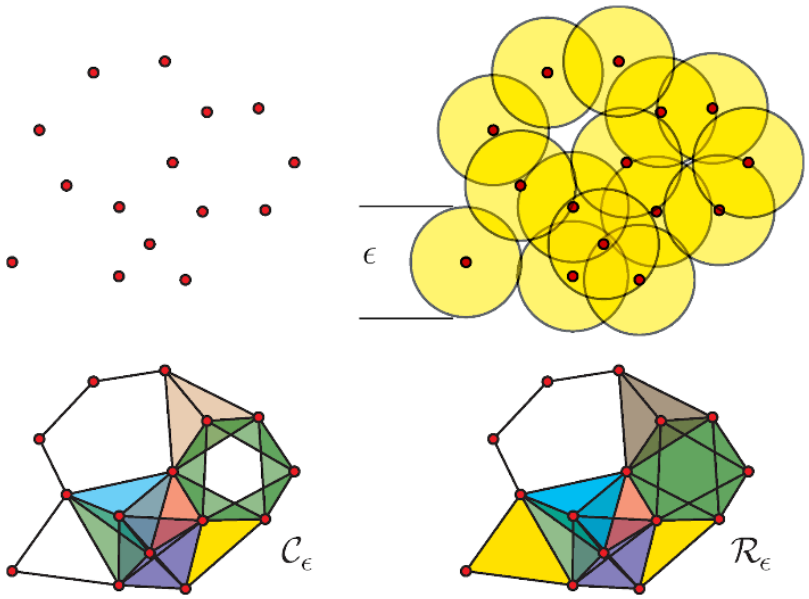


Figure 1.5: Figure 2 of Ghrist [10]. A fixed set of points [upper left] can be completed to a Čech complex $\mathcal{C}_\epsilon$ [lower left] or to a Rips complex $\mathcal{R}_\epsilon$ [lower right] based on a proximity parameter ϵ [upper right].

The Rips complex is constructed by first choosing a value for ϵ. Instead of considering points whose $\epsilon/2$ balls overlap, we consider as simplexes collections of points which are pairwise within distance ϵ. This complex can be thought of as inflating points, as if they were expanding droplets of water. Just as water droplets combine when they touch, forming one larger droplet, so too we consider two points to be part of the same complex when their $\epsilon/2$ balls overlap. This means that a circle with hundreds of points sampled at even spacing from its circumference would combine all at once and the entire circle would count as one simplex in the simplicial complex for a proper choice of ϵ.

Figure 1.5 helps to illustrate the differences between these two complexes for the same value for ϵ. The clearest illustration of this difference is shown in the "hexagonal" pattern of six points on the far right of the complex: in the

Cěch complex, the points are gathered into simplexes which take a triangular shape because not all the points have a shared intersection in their $\epsilon/2$ balls. In the Rips complex, the points form a single simplex because the points connect together from one to the next, 'chaining' into a a simplex of the six points in question.

Rips complexes do not follow the Cěch theorem, but are much less computationally expensive. Fortunately, it has been shown in de Silva and Carlsson [5] that for any $\epsilon > 0$, there is a chain of inclusion maps

$$\mathcal{R}_\epsilon \hookrightarrow \mathcal{C}_{\epsilon\sqrt{2}} \hookrightarrow \mathcal{R}_{\epsilon\sqrt{2}}.$$

This chain garauntees that the Rips complex can capture all of the topological features of the Cěch complex. This idea makes it possible to use the computationally efficient Rips complex while still leveraging the benefits of the Cěch theorem as a tool for dimension reduction from the original data to the chain of complexes.

Different values of $epsilon$ provide different perspectives: a very small value of ϵ will return a simplicial complex where many or most simplexes are singleton points (thus providing no information about meaningful topological features), while a very large ϵ value will return only a single simplex consisting of all points, which is an equally uninformative representation of the point cloud. Figure 1.6, which shows a set of points randomly sampled inside the selected region helps to illustrate this idea. The two smallest values of ϵ yield a simplex which does not capture the true homology (an annulus) of the point cloud, while the two largest values yield a cloud which obscures the homology in a different way.

However, in TDA we are not faced with the decision of selecting a single value for ϵ. These approaches look at the totality of filtration values. We are not attempting to treat ϵ as a value to optimize, but rather as spectrum of values

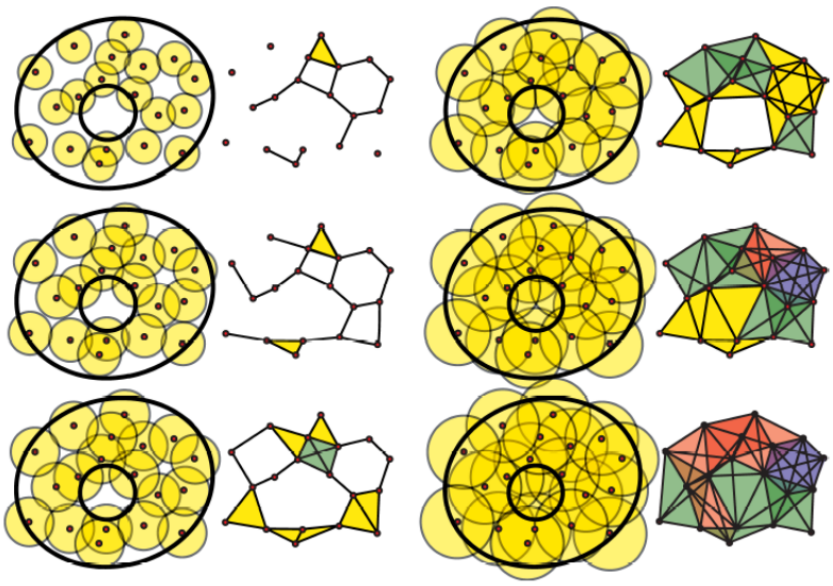


Figure 1.6: Figure 3 of Ghrist [10]. A sequence of Rips complexes for a point cloud dataset representing an annulus. Upon increasing ϵ, holes appear and disappear. Which holes are real and which are noise?

which can be informative about the persistent features in the data. The maximum homology dimension for this type of approach is $k - 1$, where k is the number of dimensions in the dataset.

1.4 Persistent Homology

Since the choice of ϵ determines how close two points must be for inclusion in the same simplex, it is not hard to imagine that for any given dataset, there may be two features which are present at non-overlapping choices of ϵ. There is no single value of ϵ which tells the entire story. Additionally, each data cloud will yield homological features at different values of ϵ, which requires a new investigation of which value(s) to select and explore. This variability of ϵ indicates that a systematic way to evaluate which value or range of values offer meaningful

information holds high value. Due to both a) the presence of features at non-overlapping ranges of ϵ and b) the variable nature of what values make sense when using a new set of data, we examine each set of data at many values of ϵ for the sake of discovering the full range of topological features.

The persistence complex is a series of the simplicial complexes, each at a different ϵ value. As the values increase from zero towards infinity, the shape of the data can be seen more clearly; some features are present only in certain ranges of ϵ, while perhaps a large prevailing feature (such as a large hole in the sample space) will be present in more of the individual simplicial complexes. As the value of ϵ shifts, the complex (Rips, Cěch, or otherwise) may change accordingly. If one were to choose an increasing sequence $(\epsilon_1 < \ldots < \epsilon_i < \epsilon_j < \ldots < \epsilon_{max})$, then the sequence of complexes $(\mathcal{R}_{\epsilon_1}, \mathcal{R}_{\epsilon_2}, \ldots)$ will show how the complex changes.

Given a point cloud X and a filtration value ϵ_k, increasing ϵ will cause simplexes to combine in a predictable and consistent way. Given two specific simplicial complex based on filtration value ϵ_i and another filtration value ϵ_j, there is an inclusion relationship of these complexes which preserves the homomorphism, $f_k^{i,j} : H_k(R(X;\epsilon_i)) \rightarrow H_k(R(X;\epsilon_j))$ Ghrist [10]. Pairing the existence of this inclusion relation with the sequence of filtration values, we can track and summarize how each topological feature persists with the filtration value. This information can be summarized in a number of ways, which we will discuss in section 1.5 and 1.6.

When we speak of persistent homology groups (topological features), we mean the k^{th} Betti number. The k^{th} Betti number is an approximate numerical measure of the homomorphism. Each Betti number denotes the number of k-dimensional holes in the space. We are primarily concerned with the first three Betti numbers. Dimension zero ($k = 0$) corresponds to the number

of connected components in the topological space, which correspond to zero-dimensional holes. This is the number of simplexes in a given simplicial complex. Dimension one denotes the number of one dimensional holes in the space. This would correspond to empty spaces surrounded by points in a single simplex. Dimension two features are two dimensional "holes", sometimes referred to as voids. The maximum homology dimension for this type of approach is $k - 1$, where k is the number of geometric dimensions in the dataset.

Imagine a single point in space. This point is a dimension zero feature because it is a single connected component (a simplex composed of one point). There are no dimension one or two features in this space. Now, consider an empty circle in space. This object is composed of a single connected component (dimension zero feature), and contains a dimension one feature (the hole in the middle of the circle) but no features of higher dimension. Contrast those objects with a hollow torus, which contains one dimension zero feature as the whole object is a single connected component. This torus has two dimension one features: the circle that goes around the hole in the torus, and the circle that acts as a ring around the torus without going around the hole (see Figure 1.7) . These features are present regardless of whether the torus is hollow or not. This hollow torus has a dimension two feature (the hole which comprises the interior of the torus). This dimension two feature would not be present if the torus were full instead of hollow.

For the sake of clarity, there are features at higher dimensions. We don't discuss them here because the interpretation of such features is difficult at best and our work generally focuses on dimension-zero and dimension-one features. The dimension of the data determines how many dimensions of persistent homology exist, and existing algorithms can generate persistence diagrams for

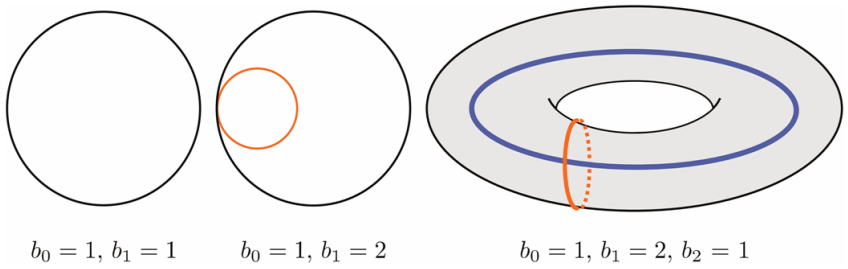


Figure 1.7: Homological feature counts for a hoop, two adjoined hoops, and an empty torus. Note that all three objects have one dimension zero feature because they are composed of one single connected component. The adjoined hoops and empty torus have two dimension one features because there are two loops which enclose an empty space. Only the empty torus has a dimension two feature because the shell of the torus encloses a two dimensional space.

dimension-zero, one, and two; additional dimensions are generally viewed as having very little interpretability.

1.5 Persistence Diagram

At each value of ϵ (called a filtration value), one can count the number of features present of each dimension by summarizing the simplicial complex. The first time that a feature appears, it is said to be "born". That feature is said to persist until it joins with another feature with smaller birth value, at which point we record that filtration value as the "death" of the feature. In Figure 1.8, there are six features born at $\epsilon = 0$, and deaths occur at $\epsilon = 5, 10$ (twice), 15, and $\sqrt{325}$, leaving only one feature for all $\epsilon > \sqrt{325}$. Since this last feature persists for all values of ϵ, we don't consider values beyond the last death, leaving only a final trivial feature, so called because this connected component contains no information other than the existence of a complex which contains a single simplex. When focusing on which homological complex holds true meaning and which parts of a complex can be classified as noise, the existence of such a com-

plex doesn't provide different information than knowing all previous birth and death times.

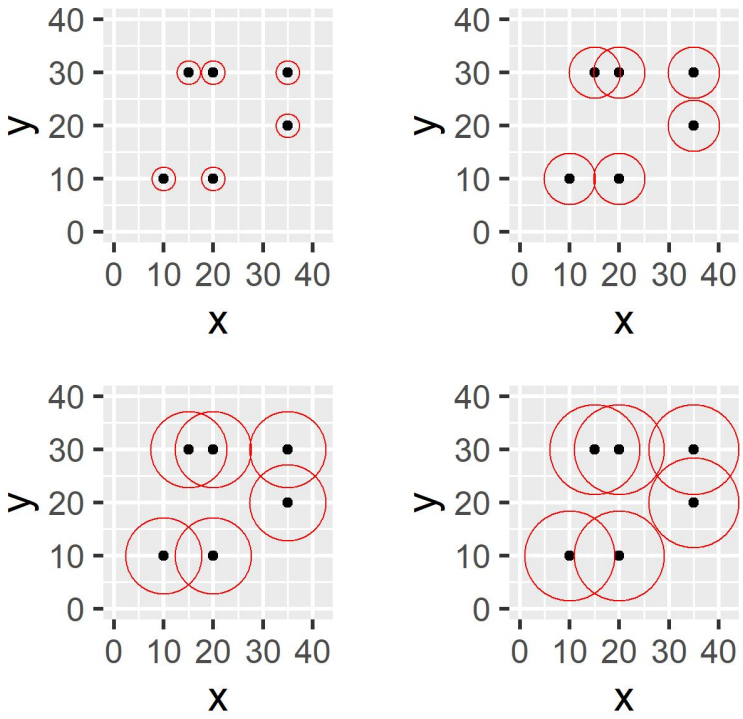


Figure 1.8: Rips filtrations for a six point dataset in space with four different values for ϵ (5, 10, 15, $\sqrt{325}$). With the Rips filtration, all dimension-zero features are born simultaneously, and die off as circles overlap.

With an adequate range of filtration values, we can thus record the birth and death times of each nontrivial feature and report a lifetime (death time - birth time) associated with each one. The summary of lifetimes can be displayed in a horizontal image known as a barcode. Features are displayed such that the feature born first is shown at the top of the barcode, with the last feature born shown at the bottom of the barcode. For each feature, the birth time and death time are marked and a filled in bar connects them, creating a staggered sequence of bars which mark the birth and death times for each feature. Another way to visualize this information is known as a persistence diagram.

In a persistence diagram such as that shown in Figure 1.9, the x-axis represents the filtration value where a feature first appears (called the "birth" of the feature), while the y-axis represents the largest filtration at which the feature persists ("death" of the feature). Since death cannot occur until after birth of a feature occurs, we see that all points will be on or above the identity line. The further away from the identity line that we see a point, the longer its lifetime is, indicating that the feature is present at a larger range of filtration values. Thus the persistence diagram helps extract true shape and structure from the data, as these features should persist over a larger range of filtration values, while avoiding noise and false features, which are not likely to persist for long. Since data can vary greatly, there cannot be a single lifetime length to point to for features of a given dimension, but often the true features of the data are immediately apparent from the persistence diagram.

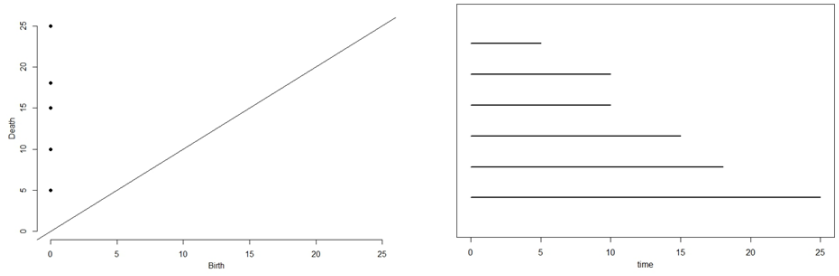


Figure 1.9: Persistence Diagram and Barcode for dimension-zero homological features for data used in Figure 1.8. All features are born at zero; deaths occur at $\epsilon = 5, 10$(twice),15, and $\sqrt{325}$.

Topological summaries of data, whether point clouds or some other type of data, can provide interesting information about the number, dimension, and persistence of each homological feature. One thing that TDA methods don't currently provide is information regarding location. In each case of constructing a simplicial complex, the information stored only includes that which is summarized in the barcode or persistence diagram; birth times, death times,

and the dimension of each feature. The advantage of this style of computation is that the homology of a cloud may not be sensitive or focused on location; storing and carrying location information through the entire process creates bulky algorithms and conflates homological information with raw values which may or may not be meaningful or interpretable.

As we continue in our discussion of TDA, it is valuable to keep in mind what information is and is not extracted and kept though the process. This introduction focused on TDA performed on data which takes the form of a point cloud, with no real response value. When data take a different form, as we will see in the next section, a different filtration proves more useful.

1.6 The Morse Filtration

When data are simple (such as a list of coordinates or a multivariate observation with no response), the Rips and Cěch complexes perform well in summarizing topological features of the point cloud. However, many data consist of some number of predictors as well as one or more response variables. In the case of a univariate response, TDA methods adopt the use of a tool called the Morse filtration Bubenik et al. [2], which embraces such a structure. We will define the Morse filtration as well as highlight its uses.

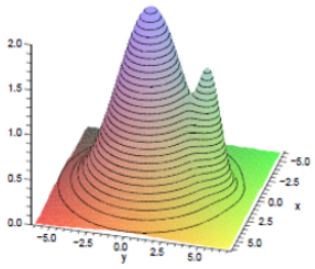
Consider a manifold $\mathbb{M}$ with a function $f : \mathbb{M} \to \mathbb{R}$. Then we define the level sets $\mathbb{M}_f$ as the subset of the manifold in the direction of f. As an example, if we consider $\mathbb{M}_{f>a}$, where a is called filtration value, then we will ignore the parts of the response surface with a value less than a. The smallest nontrivial level set is $\mathbb{M}_{f \geq a_{max}}$, which only returns the set of points where the response takes its absolute maximum value. This is analogous to a simplicial complex with $\epsilon = 0$ in the persistence complex.

As the level sets measure against a set of sequentially decreasing response values, the results will show when each local maximum of f appears and over which values of a that maximum remains as a distinct level set from the other local maxima of the function. Each level set found in this way comprises a zero-dimension homological feature of $\mathbb{M}$, since this set forms a single connected component over a specific range of values for a.

If all local maxima of $\mathbb{M}$ are absolute maxima, then the Morse filtration begins much the same as the Rips or Cěch complex: all features are "born" at the same value, in this case a_{max}. Features would then "die", or cease to persist as unique features, as a reaches each local minimum. Visually, this occurs as two components of the response surface join to form a single component. If some local maxima are not absolute maxima of f on $\mathbb{M}$, then each feature's birth occurs at f_{max_i}. In this case, the general pattern of the persistence diagram or barcode may take a different pattern or shape than it does in the case of the complexes discussed previously.

The process of starting at the maximum value and slowly decreasing the filtration value a is known as taking the superlevel sets of $\mathbb{M}$. Alternatively, one can start at the minimum value ($\mathbb{M}_{f \leq a_{min}}$) and slowly increase the filtration value (known as taking the sublevel sets of $\mathbb{M}$). Both methods rely on the differentiability of the surface, which guarantees adequate smoothness Bubenik et al. [2]. We can use this procedure to find the number and lifetime of features of other dimensions. Figure 1.10 helps to illustrate sublevel sets of a two-feature manifold. Just as in Figure 1.6, the two bumps in the function are identified as distinct or as a single feature, depending on the choice of ϵ.

Morse theory has many potential applications, especially for data with a geographical location. The spatial characteristics of an observation can serve as covariates and the response variable is a natural measurement at that location.



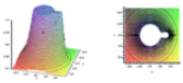


FIGURE 2.2. The sublevel set at $r = 1$ has one hole.

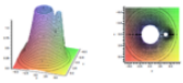


FIGURE 2.3. The sublevel set at $r = 1.2$ has two holes.

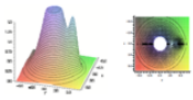


FIGURE 2.4. The sublevel set at $r = 1.5$ has one hole.

Figure 1.10: Figure 2 of Bubenik (2009). A mixture of two bump functions, along with contours which highlight sublevel sets for varying values of ϵ.

One example of this is that one can use sublevel sets to find and draw topographical maps, using the latitude and longitude of a point as the covariates and the ground elevation above sea level as the response. Thus, one should be able to identify features using Morse theory directly to create a map for a hiker or cartographer. One application of interest for our work centers on studying connected components in the Earth's atmosphere. In the next section, we will describe some of the basic mechanisms which guide that system in the northern hemisphere, focusing particularly on the jet stream's northward and southward motion. This is a case where the true homological structure is not known; we obtain a sample of data but do not have the "outline" of the true structure. It is our goal to describe a process which allows us to identify that structure.

In this chapter, we've described one topic of active study among atmospheric researchers. We have described some sources of disagreement in measuring Rossby waves as well as some of the complexities which may contribute to these disagreements. We've even shown examples where researchers in the field have clearly identified that the conclusions one reaches about changes in

the atmosphere are somewhat dependent on the way that one measures various attributes of the waves in question. We've described some of the background for using TDA to analyze the persistent homological features of data. This background lays the framework for the development of new tools which can help to address the question of estimating the number of Rossby waves.

In chapter 2, we develop and provide a new TDA tool that helps us consider multiscale persistence. There, we combine some of the ideas behind persistence diagrams and barcodes with the desire to look at persistence across multiple scales. The resulting sequential Morse filtration leads to the genesis of a new visualization, the seahorse plot.

Chapter 3 contains a description for how we can apply this new filtration to geopotential height data, especially as it relates to estimating Rossby waves for a given isobaric surface. Given such a surface, we discuss how the sequential Morse filtration and accompanying seahorse plots can aid in such estimation in a general sense.

Chapter 4 provides a simulation which can generate surfaces which mimic the behavior seen in isobaric surface data. Creating such a simulation enables us to use filtrations and seahorses to estimate Rossby waves and validate the quality of such estimation. We provide some interpretations of the results of our simulation and its usefulness in estimating Rossby waves and in assessing multiscale persistence more generally. We also discuss future work and describe additional ways to consider or summarize the information contained in a seahorse plot.

Chapter 2

The Sequential Morse Filtration

One can consider an isobaric surface as a manifold quite naturally: the latitude and longitude of a location on the surface of the earth are the covariates of the manifold M, while the elevation of the isobaric surface is the height of the manifold. Since these surfaces are relatively smooth, we can consider the function which creates the isobaric surface as differentiable and thus valid for Morse filtration TDA.

When using Morse filtrations, we would expect that at a given filtration height, the number of connected components above that height would help us to count the most relevant features in that surface. Since the focus of our current work revolves around finding and identifying Rossby waves, we are interested in knowing whether Rossby waves show up as distinct components when the Morse filtration is applied. However, Figure 1.4 shows that all five or six waves are part of the same connected component, which corresponds to higher-temperature regions closer to the equator. Consequently, using Morse filtrations yields results very much in line with our discussion of isobaric surfaces

from Section 1.1.1: geopotential heights reflect the temperature gradient and are higher at the equator.

Consider Figure 2.1, which shows superlevel sets of geopotential height data from Oct. 11, 2016 with a filtration height of 3100 meters. This filtration yields three separate connected components. The first component, which dominates the visible space, contains most of the 25 ° N line and includes wavelike portions of multiple Rossby waves. The other two connected components are composed of small sections at the 65° deg N line. These smaller features are part of an isolated area of high pressure and consequently higher geopotential heights in the upper latitudes. This is a feature of interest in climatology, but not strictly tied to the number of waves.

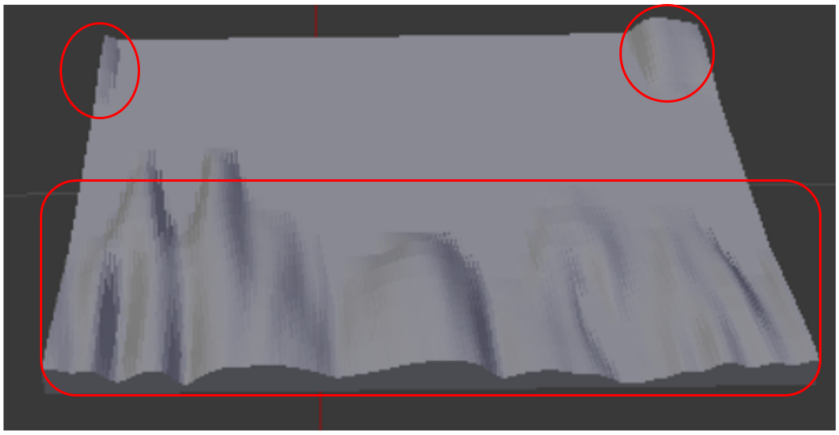


Figure 2.1: A superlevel set of the surface shown in Figure 1.4 with $\epsilon = 3100$. Two dimension zero features are present: one in the North end of the surface (which wraps around the vertical edges in this view) and one focused near the southern side of this surface. Rossby waves do not appear as individual features.

Analyzing isobaric surfaces using Morse filtrations does not provide an adequate framework for counting Rossby waves. The structure of these surfaces does not divide waves into separate connected components, instead identifying the entire equatorial to poleward feature as a single dominating connected component. One potential solution to this challenge involves looking at wind speed

surfaces instead of geopotential height surfaces which helps to isolate areas of high wind speed including the polar jet stream.

Using reanalysis data of wind speeds presents the same problems, but in a slightly different form. The jet stream is the prevailing wind feature in these latitudes, so the result of using Morse filtrations on a wind speed map would yield only one component which encompasses the entire jet stream, only providing additional features if wind speeds die out in one or more areas across the Hadley cell but then pick up again as one looks from east to west. This result would not help to quantify the number of waves present; it would only identify that the jet stream exists and provide information on how many times it fades out and surges back into relatively strong wind values.

However, data regarding both geopotential heights and wind speed show one clear trend. At each latitude close to the midline of the jet stream there is clear evidence of Rossby waves in the form of rising and falling heights on the isobaric surface. Figure 2.2 shows three such wave forms.

In each individual waveform there is clear evidence of peaks and troughs, likely corresponding to Rossby waves. Some peaks only appear at one latitude, as in the peak around $250°E$ in Figure 2.2(c) at the $35°N$ line, while other waves are clearly present at all three latitudes (see the dominating peak in all three waves around $0°E$). Combining information about the location and persistence of connected components at many successive latitude lines can provide information about the overall structure of waves, including the quantity of waves present and the persistence of each component. Connected components which persist over a large range of filtration heights and which appear over many different latitudes are meaningful ripples in the manifold, and likely indicative of the presence of waves.

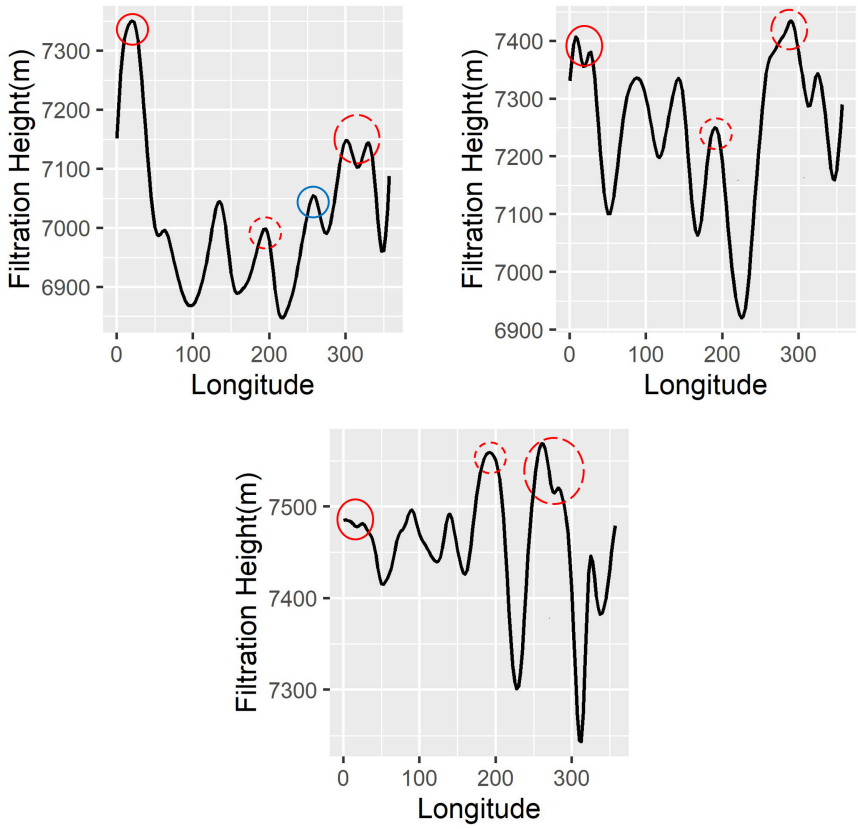


Figure 2.2: 500 hPa geopotential heights for $35°N$(a), $45°N$ (b), and $55°N$ (c) on Oct 11, 2016. Individual features can be seen as peaks in the curves. Some waves are present in all three curves (red circles), while others are present in only one curve (blue).

We propose a method of using Morse filtrations in a sequential manner which enables the isolation of Rossby waves from the dominating climatological trends in the data. This sequential Morse filtration provides an opportunity to gather information at each latitude and then combine this information across latitudes to help distinguish topological features from topological noise. The method for using a sequential Morse filtration on geopotential height data is as follows:

1. Select a number of connected components on which to focus.

2. Perform Morse filtrations at each lower-dimensional wave. In our case this is a two-dimensional wave form at each latitude available in the 25-65° N range.

3. Create a summary of this information which displays how the features persist over filtration heights (within a latitude) and adjacent waves (across latitudes).

4. Perform analysis on the summary data to answer statistical questions of interest.

More generally, we define the sequential Morse filtration mathematically. Consider a manifold M with at least two covariates, one of which we will call d and the other called a with a function $f_{\cdot}M \to \mathbb{R}$. Then we define the sequential Morse filtration of M along d by considering $M_{f>a|d=d_i}$, where a is the filtration value and d_i is one in a range of values for d. Clearly, d_i should take only values between d_{min} and d_{max}. These two values are determined by the nature of the data. In our case, d_{min} is equal to the equator where latitude is at its minimum value and d_{max} is equal to the $90°N$, the north pole. If one were to apply this method to different data, the range of the dimension of interest should dictate the choice of d_{min} and d_{max}.

By varying d across its range of values, we obtain information about how the number and persistence of connected components varies through d. Whereas typical filtrations help to classify the "shape" of data, sequential Morse filtrations can help to identify more about the "texture" of data, showing what each connected component or hole or other topological feature looks like. Sequential Morse filtrations can accordingly illuminate aspects of topological features which would otherwise remain unexplored.

Since NCEP/NCAR data are available at the $2.5°x2.5°$ resolution, we can look at "slices" of latitude every $2.5°$. The jet stream is known to occupy space limited between $25°N$ and $65°N$, which gives seventeen different "slices" to combine. For each slice, we perform Morse filtrations at a large number of filtration heights and count the number of zero-dimension homological features present. This value typically varies from zero to seven, depending on the season of the year as well as natural variation in the atmospheric system. Equipped with this information, we can display a summary graphically, as in Figure 2.3. Due to the temperature gradient and tendency to observe higher feature counts in the middle set of filtration values of the surface where the tallest waves are still present and the shortest waves are detected, these plots generally create shapes similar to that in the figure, causing us to refer to these images by the name "seahorse plots".

Since the seahorse plot provides a new way to summarize topological information from data, it is instructive to talk about how the literature uses the visualizations already discussed, namely persistence diagrams and barcodes. In Fasy et al. [8], a single value c_n is calculated based on distance metrics on the diagram. A confidence band is then defined as the region between the identity line $d = b$ to the line $d = b + c_n$, which forms a constant lifetime against which each observed feature's lifetime is compared. As shown in Figure 2.4, features whose deaths occur outside the band are deemed significant. A similar method can be applied to barcodes since they display the same information. In particular, for each topological feature with birth time b_i and death time d_i, a feature is found to be significant if $d_i - b_i > c_n$.

Some have attempted to present variations on this idea, incorporating the notion that perhaps the confidence band should not have a constant length for all birth times[8]. In particular, one method suggests that for larger values of

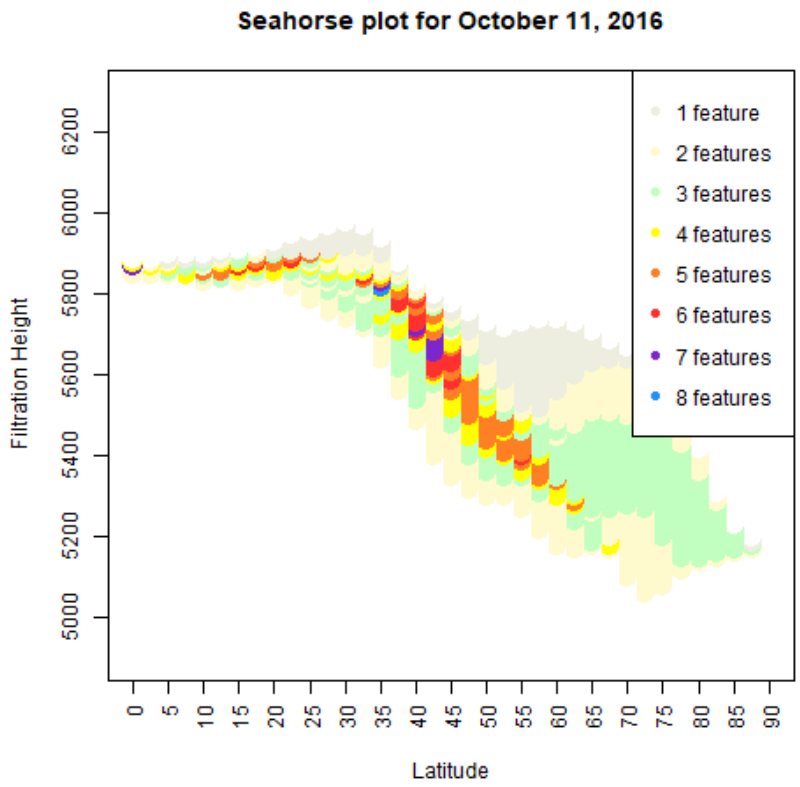


Figure 2.3: Seahorse plot of wave counts on Oct 11, 2016. For varying latitudes, the wavenumber is computed for many consecutive filtration heights. Red, purple, and blue pixels indicate combinations of latitude and filtration height where a Morse filtration finds five or more dimension zero features. If such a pattern persists across many latitudes and many filtration heights, the resulting collection of red and purple pixels indicates some strength of evidence in the presence of five or more Rossby waves.

b_i, the threshold which d_i must exceed is larger. There are some data for which this may make sense, while in other settings this adaptive confidence limit may not fit the type of shapes or features one wishes to detect.

One way to use visualizations like seahorse plots is to focus on a specific number of features. Due to the sequential nature of the seahorse plot, we can see persistence in two dimensions. Persistence over filtration values is seen when

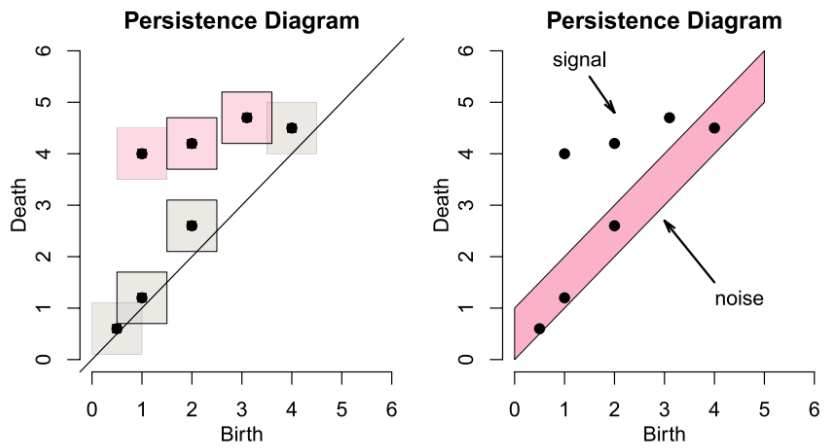


Figure 2.4: Figure 4 from Fasy et al. [8]. In this work, the authors consider features with a lifetime longer than a specific constant to be 'signal' and other features are considered 'noise'.

the seahorse plot yields the same value for many vertically consecutive values. Persistence over latitudes can be seen when the plot yields the same value over horizontally consecutive values. We can illustrate this by exploring the presence and persistence of five-wave structures, which appear as clusters of neighboring blue dots. In Figure 2.3, which is a day in mid-October 2016, one can see a very large cluster of connected blue dots, which is indicative that the wavenumber for this day is at least five.

Seahorse plots like this created for consecutive days can provide a measure for each day of a given year in the form of the size of the region where five or more features are detected and interpret it as one measure of relative evidence of a five-wave structure on the given day.

We record the size of the largest connected cluster for each day, where **connected** is defined by adjacency of points. Each day's cluster size can be considered as observations in a time series which show the strength of evidence

supporting the presence of a five-wave structure. Figure 2.5 shows this time series for 1950.

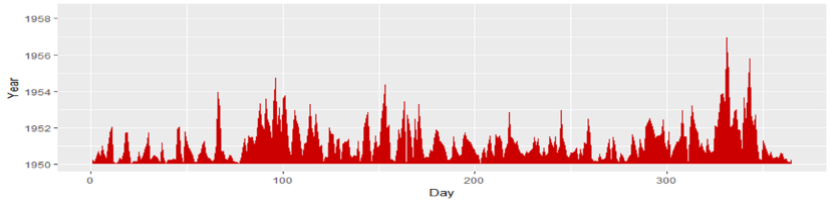


Figure 2.5: A time series of five-wave cluster size for 1950. Each day, we record the size of the largest connected region where $w \geq 6$, i.e. where five or more waves are detected.

We can recreate this plot year after year to get a good idea of how this measure looks as well as whether it changes over time. When the northern hemisphere has a decreased temperature gradient, we expect to see more Rossby waves. This lower gradient occurs seasonally as well as any other time when polar regions and equatorial regions are more similar to one another in termperature. Figure 2.6 shows this metric (the strength of evidence for five-wave structures) year after year from 1950 - 2000.

This visualization is called a "ridgeline" plot. As best we can tell, the plot used to be named a "joyplot", thus dubbed due to the cover art from an album by a band named 'Joy Division'. This title was suggested by Dr. Jenny Bryan on Twitter on April 24, 2017. Around that same time, Dr. Claus Wilke wrote an R package named "ggjoy" to assist in creating these plots. However, due to some of the background of the band's name, the name "joyplot" has since been abandoned and the ggjoy package has been deprecated. In a blog post on September 15, 2017, Wilke announced this deprecation which was paired with the posting of a new package named "ggridges". He additionally suggests the name "ridgeline plot", which we use throughout the this work.

In Figure 2.6, we see a slight but noticeable increase in the this evidence in the spring and fall seasons. In addition, the figure shows that fall formations

of wavenumber 5 or greater begin earlier in the year while lasting just as long into the winter, as seen from the "stretching" of the set of higher values in the fall season to a larger range of days. These two aspects of the plot are consistent with both a) an increase in the certainty of five wave formations as well as b) indicating that such formations are lasting for longer periods of time. Both of these phenomena are supported by the claim that under decreased temperature gradients, the atmosphere will feature more Rossby waves and subsequently we will observe high numbers of waves more often.

Temperature gradient is at its very lowest in summer. It naturally follows that we might expect that five-wave structures should then be highest in the summer, which is clearly not the case in Figure 2.6. The evidence for such a structure is lower in the summer for two main reasons. First, seahorses are smaller overall in the summer than the winter, so the total number of non-zero points is smaller. This fact makes it even more clear that winter, when the seahorse is largest, has the least amount of five-wave structures. The second reason: seahorses in summer months have smaller values for five-wave structures because they instead have greater evidence for structures with more than five waves, often supporting the idea of seven or more waves.

Sequential Morse filtrations provide a conclusion consistent with climatological theory as we would expect based on the construction of the plot. Numerical summaries of information contained in a seahorse plot can provide insight on both a) the number of features present on a given day and b) how those features persist both over a range of filtration heights and across varying latitudes. We've discussed the size of a region where a specific number of features are detected; sizes for regions where any given number of topological features are present may serve to identify characteristics on any geopotential height data.

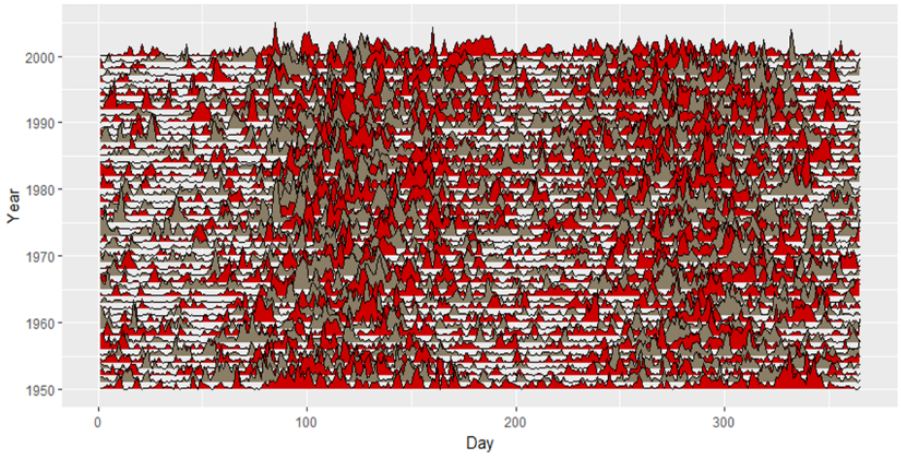


Figure 2.6: Ridgeline plot of five-wave cluster size annually, 1950 - 2000. Even numbered years are shown in red, odd-numbered years shown in tan. The evidence supporting the claim of five-wave formations is increasing through time (higher peaks for longer periods of time in the spring and fall) and the length of five-wave seasons is expanding (the periods of time with high heights are lasting longer).

One way to capitalize on all the information contained in the seahorse plot as opposed to only the size of one specific region instead utilizes the feature counts throughout the entire nontrivial range of filtration values and latitudes. Everywhere in the seahorse plot where the Morse filtration detects at least one feature is a part of the surface, and we can use information about the size of the region with each number of features to summarize information about the dimension-zero homological features.

Table 2.1: Summary of feature count values for July 1 and Oct 11, 2016. As in most days, smaller feature numbers have high counts since there is almost always some fluctuation in values. In comparison with Oct 11, the total number of pixels with features in July is lower; the seahorse shape shrinks in the summer and expands in the winter due to the changes in temperature gradient.

Feature Count	1	2	3	4	5	6	7	8
Number of pixels (July 1, 2016)	342	656	372	172	117	35	3	0
Number of pixels (Oct 11 2016)	253	707	515	221	100	49	25	3

Tabulating information about feature counts serves as one numerical summary of the information contained in the seahorse plot. Some nuance included in the plot is no longer present, but the overall trend that low feature counts are most common, decreasing by a significant ratio from one feature count to the next. As seasons tend to exhibit different numbers of waves with varying shapes of seahorses, Table 2.1 shows the frequencies for two different days; one in summer and one in fall. One interesting difference between the summary for October and July data lies in the five feature counts: In October, there are less than half as many five-feature pixels as there are four-feature pixels($p_4 \approx 2.23 * p_5$), while the ratio is much smaller in July ($p_4 \approx 1.47 * p_5$).

Comparing two days by themselves can highlight some similarities and differences, but can these summaries, either graphical or numeric, provide clear information about the true number of features? Previously, we mentioned two methods currently used in calculating wavenumber, one using a location and time dependent formula and the other based on interpreting wind maps. In either case, some amount of subjective judgment is needed.

When choosing a location as a central point when using the formula for ω, the choice of location impacts the estimated frequency and the correct location should change at least some over time, seeing as the central latitude where the polar jet stream resides shifts north and south throughout the year. Using this formula is therefore prone to inconsistencies which are difficult to control; prudent science indicates that the best locations to detect the jet stream are moving, but how does one pick one latitude (or longitude) over another? There is not a clear and consistent way to do this that avoids potential confounding information. Using estimates from this formula results in inconsistencies with unknown bias; these inconsistencies contribute to some scientists uncertainty about what wavenumber actually means.

Using wind maps is subject to a different type of challenge. Since there is no set formula for computing frequency, finding the correct number of Rossby Waves is more or less dependent on the eyes and subjective training of the individual looking at the map. Some people may have very consistent judgment on these maps, but there is bound to be some degree of variability from one individual or group to the next when making such determinations.

Both methods currently used for calculating Rossby wave frequency are backed by sound theory and are usually exactly the same as that of other researchers or close to it. However, when the number of Rossby waves only varies from two to eight, being off by one is very meaningful. Superseding all these potential pitfalls when finding wavenumbers is the fact that there is no real training set of data. Individuals who read wind maps and estimate frequencies are rarely given a set of maps for days where the true wavenumber is known; how could such a number be provided when there are many different factors which contribute to wind speed, geopotential heights, and any other data which provides information about Rossby waves? Without such a set of days, its difficult to assess the quality of any method for computing the wavenumber. Our goal is now to create such a set of simulated isobaric surfaces and then use them to both a) create an honest representation of these surfaces which generates a surface with a given number of waves and b) to develop a classifier which associates a geopotential height surface to various wavenumbers with predicted probabilities.

With such a classifier, we will provide practitioners with a tool that is useful in ways which are not currently present. First, our classifier is transparent. Whichever classifier we use, we can clearly explain how we come to whatever estimates we reach. We value this transparency because it provides a clear vehicle for potential improvement: with a transparent mode of estimation, those

who work with this kind of data can iterate, suggest changes, and even see how a change in the classifier performs on both simulated and historical data. The second way that our classifier adds value is that it is objective. We don't think that anyone intentionally estimates incorrect wavenumbers, but currently available methods often leave a bit too much up to the literal eyes of the beholder of a wind map or isobaric surface. Some may feel more optimistic about whether a small northward peak in the polar jet stream is a Rossby wave, while others may dismiss that as not quite reaching their internal threshold for inclusion as such a feature. The same set of eyes might even make different conclusions if they looked at the same map at two different times! The consistency of a classifier provides clear benefits.

Another benefit of developing a classifier is speed. An expert takes some time to look at one of these maps and determine their best estimate of the number of Rossby waves; this can add up quite quickly if one is trying to estimate this number for many plots, such as estimating this number for one plot per day from 1980 and one plot per day from 2010. However, with our work, we can get these estimates quite quickly for a large number of surfaces. This would enable researchers not only to compare two individual years, but to compare decades or even longer periods of time while keeping the transparency and consistency mentioned above!

Chapter 3

Simulating Geopotential Heights

3.1 Basis of Simulation

As noted previously, we cannot use reanalysis data directly to build or validate a model because the true number of Rossby waves is an unknown quantity. Subsequently, there is no way to compare the number of waves between days; the response value of interest is not known for either day. This is part of the foundation for our work; since the true number of waves cannot be found empirically, it is impossible to validate any method of estimating the number of Rossby waves on a given day.

We propose one solution to this problem: applying trends seen in reanalysis data to create simulated isobaric surfaces which mimic real data. Using simulations allows us to create a surface with any given number of Rossby waves, so we can get a large sample of days with each wavenumber. Each of these simulated surfaces can include various structural components which are consistent with the type of variability seen in NOAA's reanalysis data which currently produces

so much inconsistency in estimation by traditional methods. The advantage here is that one facet of this imposed structure is the type and number of Rossby waves present in each simulated surface.

In this chapter, we discuss the different components that comprise each simulated surface. We begin the discussion by focusing on overall trends in geopotential height data throughout the year such as the effect of the temperature gradient seen in the decline in heights from equatorial to polar regions. We then progress to other components which contribute in a much more localized and/or seasonal way, such as the presence and effect of isolated high and low pressure zones as well as the presence and size of each Rossby wave.

In constructing simulated surfaces, we take an additive approach: each "layer" of the simulation is considered as though it exerts its effects on the surfaces in a manner that is agnostic to the size and location of the effects of any other layers. Using additive effects allows for a layer-based construction of each simulated surface, and helps to reduce complexity. Since there is no other known method for performing such simulations, this construction method has useful intuitive interpretations and produces surfaces that reasonably imitate the patterns we see in observed data, while providing a basis to obtain model-based predictions of wave numbers.

This additive simulation has some very clear benefits. The relatively intuitive nature of the simulation should improve the capacity for statisticians, atmospheric scientists, and others to discuss the simulation overall as well as to discuss a specific layer. Lets suppose, for example, that a simulated surface looked fairly odd to a practitioner. They could, with a few pieces of information, both a) recreate the surface as many times as they wished and b) decompose the simulated surface into its constituent parts, which enables them to see which layer or combination of layers in contributing to any odd behavior. A more

complicated model is not likely to have this level of intuition and is even less likely to have the same sort of decomposition as an option.

Another convenient aspect of this additive approach to the construction of the simulation is that it allows one to simulate some situations that don't currently happen with any regularity. What would a January day look like if it had a wavenumber more in line with a summer month, such as seven or even eight waves? Simulations which are only based on historical data and the dynamics of atmospheric circulation are not likely to produce such a day, but our simulation can create such a surface just as easily as creating a day in August with that number of Rossby waves. This is a valuable feature because some suspect that the temperature gradient may continue to decrease as time moves forward, so practitioners may want to take a look at what surfaces might look like in winter months with more waves. While it is perhaps a less likely future scenario, the simulation could generate a surface in summer months where there are very few (or even zero) Rossby waves. The customizability in this admittedly simple simulation provides a lot of flexibility in the types of surfaces that we can generate and analyze with a classifier.

The third beneficial aspect of this approach that we discuss here is that this simulation is that the modular nature of layers means that others can easily add layers, remove layers, or alter existing layers in fairly rapid succession. As one example, one could pretty quickly modify these layers to make a simulated 700 hPa surface instead of a 500 hPa surface by modifying each layer. If a researcher felt that one or more layers were missing, making the necessary addition would not be too difficult. This modular nature allows for adjustment and/or improvement of each layer without worrying to much about effects on any other layers.

The following sections describe each of the layers present in the simulation. These layers are: overall trend due to temperature gradient, signal due to Rossby waves. Isolated highs and lows, and geopotential noise.

3.2 Simulation layer 1: Overall Trends

The first layer we would like to add to each simulated surface is what we feel is the foundation for these data. Due to the nature of the temperature gradient in the northern hemisphere, we know that geopotential heights are higher near the equator and lower towards the poles. This is because warmer temperatures are generally associated with higher pressure, which means that isobaric surfaces will have higher heights in equatorial regions. It is true that in regions with cooler temperatures, the isobaric surface has lower heights, often seen in polar regions. This effect is seen throughout each year of data, and is more clearly present in the winter as the temperature gradient is highest during winter months. Our simulation needs to incorporate this large-scale component in the structure of each simulated surface. However, simply accounting for the change in heights based on distance to the equator or poles is not sufficient. There are other structural components which contribute meaningfully to the overall structure of the data, regardless of the number of waves.

Geopotential heights are affected by surface features in the northern hemisphere, such as the shift from land to ocean, effects due to mountain ranges which serve as barriers to the free flow of air along the surface, and the presence of various currents which tend to push air into specific regions, affecting pressure and therefore geopotential heights.

Each of these effects are stationary in location, so their effect on a given isobar should be consistent over time. We can capture all of these trends by

taking the mean isobar over time. Rossby waves have peaks as well as troughs, which means averaging out over a longer time period yields a reasonable estimate of geopotential heights with all short-term features removed. Figure 3.1 shows a simulated surface with just this layer. It turns out that for this year (and many years), the averages are quite smooth from east to west, so the height of the surface is almost entirely determined by latitude. We do not consistently see much longitudinal effect.

Rossby waves certainly persist as a feature in the northern hemisphere, but their presence travels along the hemisphere with the jet stream, which is what makes this averaging fulfill the intended role. By starting with the mean geopotential height when generating a simulated surface, we ensure that we are starting not with a blank canvas, but a representative starting point for isobaric surface data. These average surfaces will have some high and low heights along a given latitude, but these highs and lows are generally much smaller in magnitude than one typically sees from a Rossby wave. We can use these averaged values as the first layer in each simulated geopotential height surface.

In constructing each simulated surface, we use the mean annual surface from whatever year is randomly selected for a given surface. This helps to provide a baseline layer for the simulation which is based on that particular year's average geopotential height data.

We have not included a seahorse plot for this simulation layer because the graphic is uninformative. Since geopotential heights are constant for any given latitude, that means that there are no dimension-zero features with non-zero lifetimes. The seahorse plot would look much like 3.3 but with no non-zero pixels. For this reason, we've not included the seahorse for this layer by itself.

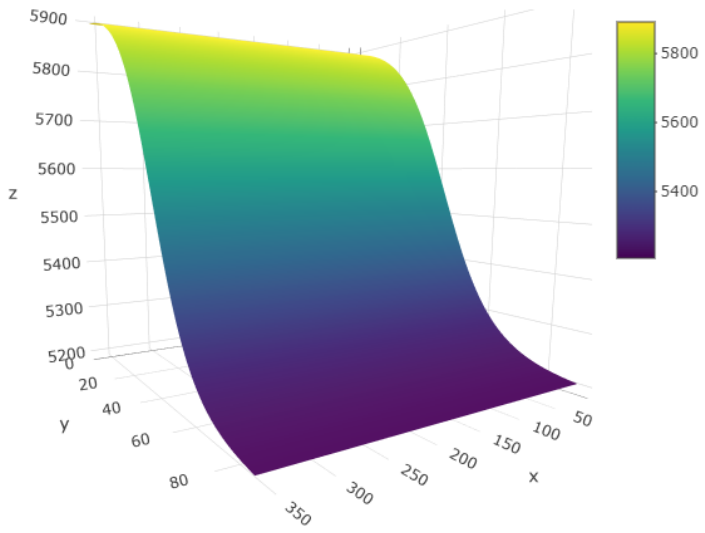


Figure 3.1: A simulated surface which includes only layer 1 from 1980. The x-axis represents longitude, the y-axis represents latitude, and the z-axis represents the average height of the isobaric surface at that latitude for the given year.

3.3 Layer 2: Rossby Waves

When the jet stream wanders poleward, it brings with it warm moist air from the equator. We see the effect of this circulation on geopotential height data: the peaks of Rossby waves cause geopotential heights to remain higher when traveling from north to south, sometimes creating new peaks but often just causing the region of highest geopotential heights to jut up to higher latitudes than seen elsewhere. This added height shows up in reanalysis data almost as if one were looking at one side of a mountain with a large plateau; the waves create various ripples which extend out and down from the plateau.

Rossby waves in the data show up as northward bulges and southward canyons regions of latitude and longitude with higher overall heights at the peaks and lower heights at the troughs. These elliptical regions have their major

axis more or less along a line of longitude with their minor axis along a line of latitude. These are not strict requirements though, sometimes the wave can be somewhat tilted off of the "vertical" appearance it typically maintains.

The number of waves plays a role on how large an effect the wave exerts on geopotential heights. At times with fewer waves, Rossby waves tend to have larger amplitudes (higher heights) and extend further in both the north-south and east-west directions. Conversely, heights and axis size decrease in the presence of a higher wavenumber. Regardless of the number of waves, the effects of these waves are heavily concentrated in the region from 25 ° N - 65 ° N and have little to no effect outside of this region.

To simulate the effects of Rossby waves on geopotential heights, we first select the number of up-swells that we intend to include. This value should be somewhere between two and ten, as the number of Rossby waves in the simulation will be no larger than the number of up-swells. We then randomly choose the longitude where each up-swell is centered, with some loose constraints to ensure spacing between waves which accurately shows the types of spacing which occur naturally in empirical data. Due to the ways that Rossby waves can clash and combine with each other though, the spacing between up-swells is not explicitly guaranteed; this allows for the possibility of two up-swells comprising part of the same wave.

The number of up-swells is an upper bound on the total number of waves in the simulated surface, but there may be fewer waves than individual up-swells. This phenomenon occurs more often when the wave number is larger, just as would be expected. The more waves there are, the more often they clash into one another; with smaller amplitudes these waves are more easily impacted, interrupted, and shifted by lower frequency waves which can contribute to collapsing one Rossby wave into another.

We mentioned previously that each Rossby wave extends more in the north-south direction than the east-west direction, but these waves are not perfectly vertical. It is not unusual for a wave to "lean" eastward or westward against the overall flow of the jet stream. Each up-swell in our simulation incorporates this idea of lean, resulting in waves that are not always perfectly vertical. This helps each simulates surface to more closely emulate empirical and reanalysis data, but it should have little to no effect on the homology of the surface since our filtrations are agnostic to translation and rotation of features.

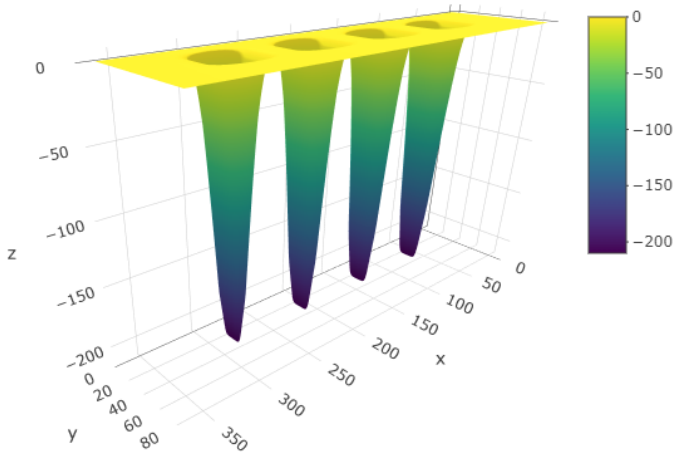


Figure 3.2: layer 2 only from a simulated surface.

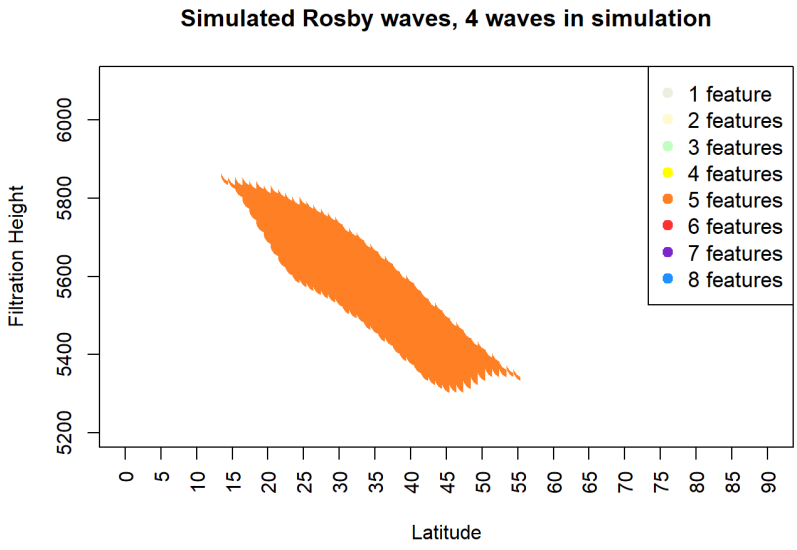


Figure 3.3: A seahorse plot of a simulated surface using the first two layers described in the simulation. The heart of this seahorse is large and uniformly shows evidence of five features, which corresponds to four waves.

3.4 Layer 3: Atmospheric Noise

With the overarching trends present in the data as well as a simulated set of up-swells which make up the Rossby waves, the simulation captures the overall signal structure we expect to see in geopotential height data. However, the chaotic nature of atmospheric circulation includes many sources of variation which are not captured by the trends described in the previous sections. Some of the other variation can be described as 'noise'; random variability which increases or decreases geopotential heights by varying amounts in a mostly random way. Even though this source of variation is unlikely to produce changes in the persistent homology and hence the seahorse plot for a given simulated surface, we include this layer to create a more realistic surface.

We add this random noise to a surface by adding normally distributed heights at each 2.5° x 2.5° point on the northern hemisphere. These heights follow a $N(0, s_{residuals})$ distribution, where s is the standard deviation of a given day's residuals compared to the annual mean. These values are smoothed over the surface using a kernel smoother and scaled to ensure the proper amount of variability to match with real data. This layer plays the role of random noise present in the atmosphere. One realization of such noise is shown in Figure 3.4.

As mentioned at the start of this section and throughout this work, meaningful dimension zero features should exhibit multiscale persistence. When we consider the potential effect of this noise layer, we expect to see additional deformations on the simulated surface. Areas with positive noise may look like small bulges in the surface, while negative noise will be seen as slight puckers in the same surface. Overall, this creates a surface that could be considered as more bumpy and varied. However, a feature which is born and dies at two heights which are meaningfully different will be minimally impacted. When we con-

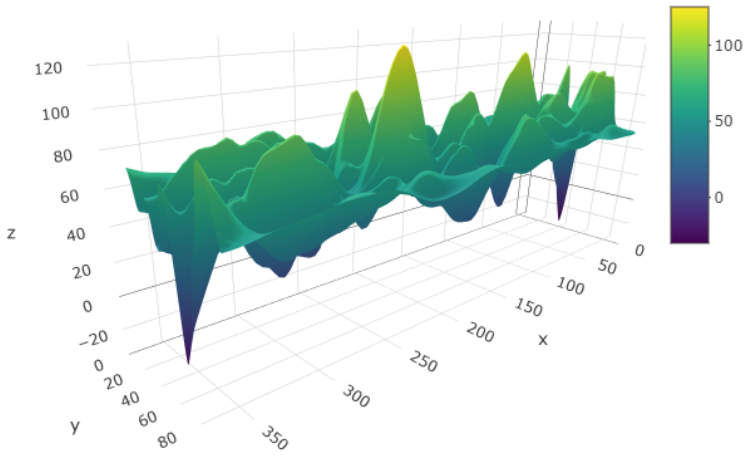


Figure 3.4: Layer 3 from a simulated surface, which fills the role of atmospheric noise. This noise is generated via smoothed random normal noise across the northern hemisphere.

sider Figure 3.3, we can see that at a latitude of 35°N, the seahorse reports four waves from a height of about 2850 meters through 2550 m. The effect of a small amount of noise, such as the noise added in this layer, would at worst alter these birth and death values by a little.

This noise may serve to increase the lifetime of one or more features. In theory, noise could push the peaks of these waves even higher and the troughs even lower. In such a case, the lifetime of any such features would be longer than would be seen without this layer. Conversely, if the noise pushes a peak downward and pulls a tough upwards, then we might see a lifetime that is a little smaller than would be seen without this noise layer. This is one of the best features of seahorse plots and their use in examining multiscale persistence.

The effect of a localized source of random noise might have a slight local effect on the lifetime of a particular feature, but since we look at the persistence

of dimension zero features over a range of filtration heights as well as looking at persistence over a range of latitudes, these local effects should have a relatively small impact on our ability to detect the true number of rossby waves we added to a simulated surface in the second layer of the simulation, as seen in Figure 3.5.

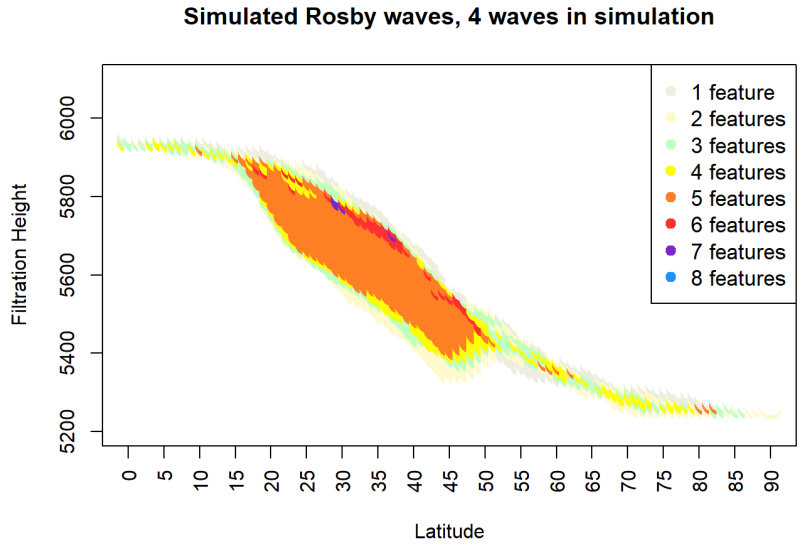


Figure 3.5: A seahorse plot of a simulated surface using the first three layers described in the simulation. The heart of this seahorse is large and uniformly shows evidence of five features (which corresponds to four waves).

The two types of local effects we discussed are visible in Figure 3.5: there are some areas (especially around $35°N$ and 2900m height) where the noise is "pushing" one of the peaks down, which leads to an inflated feature number; the seahorse reports six features when there are only five waves. In regions near the bottom of the surface, like the $50°N$ and 2600 m height area, the noise instead sometimes "pushes" up on a trough. The net effect of this is that the trough is now above the filtration height, so that feature is not detected. The results is that the seahorse shows an undercount of the number of features in the surface as seen by a local density of 3 feature and 4 feature counts.

Each of these local effects are expected and support the importance of the idea of multiscale persistence. In each of these local effects, we do see a trend in the plot that has some persistence in just one dimension (in this case, across latitudes) but that trend does not exhibit persistence across both scales. Additionally, this persistence is nowhere near the strength of persistence seen in the heart (where the waves dominate the overall behavior of dimension zero features.

As we would expect, the seahorse in Figure 3.5 has largely the same appearance as Figure 3.3: it has a large "heart" region with five features (what we would expect with four waves) that exhibits persistence in both latitudes and in filtration heights. The main difference between these two seahorses is that the addition of a third layer adds a fair amount of noise, which leads to locally high feature numbers even in the presence of no true features. This noise does not meaningfully alter the main message of the seahorse; there is a large range of latitudes and filtration heights with five dimension zero features, which is present because of the four waves we introduced in layer 2.

3.5 Layer 4: Non-Rossby signal

In addition to this random noise, each simulated surface includes a small number of up-swells in geopotential height, which are not as large in magnitude as each Rossby wave. This sort of isolated cell of high pressure occasionally appears in the northern hemisphere, especially in the polar regions. Many times these cells are separate and disconnected from the jet stream and all Rossby waves, but occasionally, these cells can be close to or even on in the same latitude and longitude as the peak of a high amplitude Rossby wave. In some cases, higher amplitude Rossby waves lead to the genesis of these high pressure cells. Each sim-

ulation run includes between two and four such cells, with magnitudes equal to the 70^{th} percentile of residual values for that day versus the annual average.

While the smoothed noise throughout the hemisphere which we mentioned already is not likely to have any effect on our ability to detect Rossby waves, isolated high pressure cells could potentially exert a different effect on feature counts in a seahorse plot, especially across several filtration heights. We are interested in assessing whether this is likely to create difficulties in detecting the correct number of Rossby waves.

The easiest way to encounter such difficulty is in a theoretical case where two Rossby waves are fairly close to each other longitudinally with one of these upswells between them. In this case, the persistence of each wave will be reduced since those two features join together to become a single connected component at a higher filtration height than might be expected, which could lead to underestimates of the number of waves. This type of underestimation is most likely to occur when the true number of waves is large. The reason for this is that Rossby waves are often well-spaced, so in surfaces with fewer Rossby waves, each wave is generally not close enough the others for this type of effect to occur.

Question for Lynne and Nicole: Do I need a figure that shows the thing in the above paragraph?

Due to the nature of where these cells are positioned, it is possible that an isolated cell may be generated between two Rossby waves which are far apart longitudinally, which is most likely when the true number of waves is at its lowest observed numbers, such as two or three. In this case, if the isolated cell is not so much isolated as it is positioned rather close to the jet stream, it is possible that sequential Morse filtrations will detect the cell and the Rossby waves at a meaningful number of filtration heights and enough similar latitudes that the

corresponding seahorse plot may lead to the conclusion that there are more features than expected based on the true wave number.

The two previous examples help to illustrate that we do expect some noise and imperfect seahorses to result from simulating isobaric surfaces from this simulation. Using feature and cell placement which are bound to occur at random in some simulation runs, it would not surprise us to encounter some simulation runs in which even sequential Morse filtrations may miscount the true number of waves. However, in most cases, we should be able to identify which parts of the surface are true signal while filtering out the parts of the isobaric surface which are not associated with the waves we care about most. Another interesting thing to note is that based on the context of the data, we know what types of prediction error are most likely. When the simulation generates a large number of waves, there is a higher chance that the model will under-predict versus over-predict. When the simulation instead generates a small number of waves, we are more likely to over-predict the number of waves.

There is a second type of non-Rossby signal that we find in the data. Very similar to the up-swells we've noted in this section, we also observe down-swells in the data from time to time that are larger than the random noise identified in layer four. Each simulated surface incorporates between four and twenty of these smaller down-swells, and we combine both of these sources of variation together in a fourth layer of the simulation. Figure 3.6 shows one example of this layer.

Layer three and layer four are designed to replicate separate aspects of the isobaric surface. While layer three adds in the type of random noise which is quite unlikely to alter the heart of the seahorse, it does have the potential to change the number of dimension zero features in a small area almost anywhere without otherwise dominant dimension zero features. Layer four's impact is limited

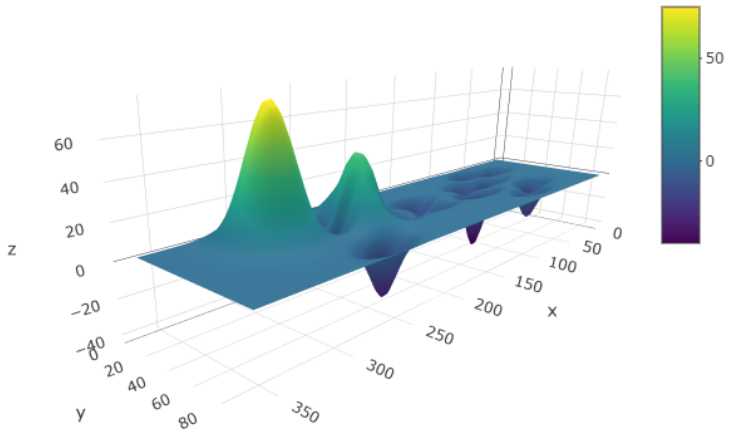


Figure 3.6: Layer four from a simulated surface, a combination of upward and downward swells in the simulated surface.

to a smaller number of regions, but has more potential to alter the number of dimension zero features at a certain latitude over a larger number of filtration heights. As is turns out, both layer three and four have similar effects for our purposes: when attempting to identify the number of Rossby waves present in the data, all non-wave contributions serve to essentially 'distract' the seahorse plot from clearly displaying the true number of waves.

We note here that each layer of the simulation serves a purpose, and layers which may feel much like noise when using this work to estimate Rossby waves may be absolutely essential if one were using this approach to multiscale persistence for another purpose, such as attempting to counting isolated cells in the polar region. Such work may require practitioners to consider the effects of Rossby waves to be noise or otherwise unimportant. We have discussed how some portion of this noise (atmospheric variability) is unlikely to affect our

model's predictions while other noise from isolated cells has a stronger likelihood of altering our model's performance.

In addition, there are signals in geopotential height data which do not relate to Rossby waves. Although we are not particularly interested in these types of signals, we need to include them in our simulation so that the performance of our model can be tested in as realistic a scenario as possible.

Now that our simulated surface has all four layers, we should check the visual of the seahorse plot. Figure 3.7 shows this seahorse, and the change we see is largely expected. In this surface, there are two postive upswells in the $20° - 40°N$ range (see Figure 3.6), which makes the trough two of the waves less deep (one more than the other), this results in regions in that range at lower filtration heights where three or four features are detected, even when four waves would typically indicate the presence of five features. The various downswells (there are thirteen distinct ones in Figure 3.6) exert very mild variation in feature numbers towards the heart of the seahorse, and translate to a bit more variation in the tails as well.

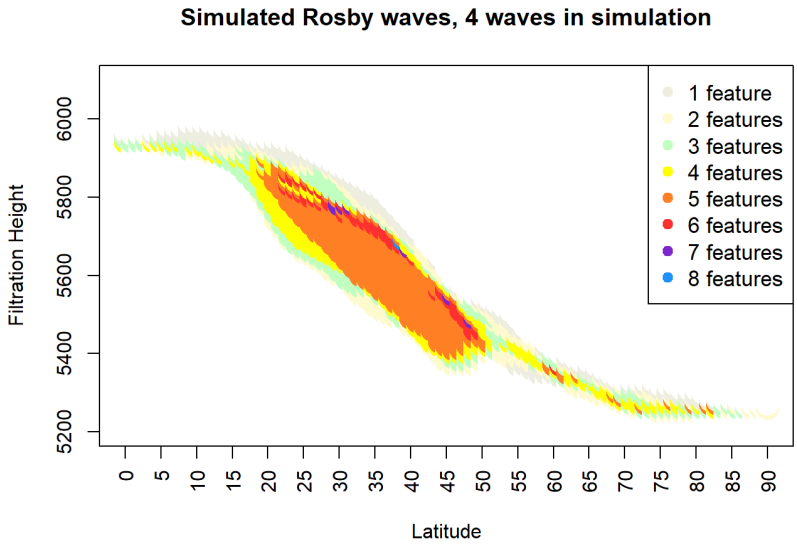


Figure 3.7: Seahorse plot of a simulated surface with four layers

3.6 Simulated Surfaces

Combining each of these layers yields a simulated geopotential height surface with underlying geopotential heights based on a specific day and year. These values informed by empirical data help guarantee that the surface takes the correct overall shape. This first layer is based only on reanalysis data and removes any transient features, including random variability and wave structures including the effects of Rossby waves.

The key to each simulated surface is that the second layer of the simulation contains information about the Rossby waves, which manifest as increases and decreases in the geopotential height values due to the peaks and troughs of these waves. If one were to perform sequential filtrations on this layer alone, the result would always show evidence exactly in favor of the true wave number. Combining wave-based fluctuations to date-based geopotential height values, we obtain a surface which provides an accurate representation of these data while providing a crucial piece of information: the number of Rossby waves are known for each surface, which means that its possible to see if an estimated or calculated wavenumber is correct or not.

Each surface contains noise in two forms; the first includes smoothed random normal noise and the second contains non-Rossby fluctuations. These two sources of noise contribute the type of variability which the overall trend values have filtered out, but which typically arise in this type of data.

Each surface then takes the same overall form as a real day. Although each date in time exhibits behavior for only a single wavenumber or evidence of a transitioning wavenumber, these simulated surfaces allow us to examine any day and year from 1955 onward while imposing a specified wave number, allowing for the possibility of examining the surfaces on the same date with the same wave

number but a different realization of the position, amplitude, and orientation of each Rossby wave, as well as variation on each type of noise.

When we take all of these layers together, we get a simulated isobaric surface like the one in 3.8. In this particular surface, the Rossby waves are particularly clear since their amplitude is large. Some simulation runs yield more subtle Rossby waves, and variation from each other layer varies some from one simulated surface to the next. The overall trend from layer one is clear, as is the set of Rossby waves, but layer three and layer four are a bit more subtle.

Layer three is the most visible in Figure 3.8 at high latitudes. While the entire surface resides at relatively low heights, smaller localized deformations are present in several places. We can see some of this small-scale deformation in the middle latitudes.

Layer four is at its most visible around $20°N$ and about $300°W\,(60°E)$ longitude; the large upswell contributes to the region on the surface with the highest geopotential height on the entire surface. Some of the down-swells, which are a bit smaller than the upswells in amplitude, are visible at high latitudes and in a few other locations in the figure.

3.7 Goals of Simulation

One of the main reasons to use a simulation to create surfaces which replicate the behavior of the 500 hPa isobar is that these is that we can create a surface with a "known" true number of Rossby waves. This will allow us to assess the accuracy of any potential method for estimating the number of Rossby waves for a given surface. Before doing so, we would like to outline how we plan to use the simulation to accomplish this goal.

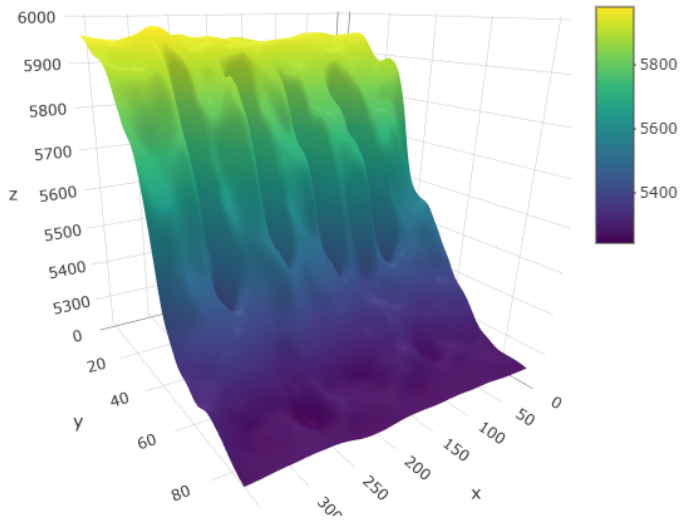


Figure 3.8: A simulated surface using all four layers,including trend from temperature gradient, four Rossby waves, random noise, and a combination of upward and downward swells.

We first use the simulation to build a "training" set of surfaces which is composed of one thousand surfaces for each wavenumber from two to nine for a total of nine thousand surfaces in this training set. In Chapter 4, we will describe each of three metrics we will use to make estimations. Cross-validation will be used on the training set to enable us to tune how to best make estimates for a given metric, but accuracy will be measured by looking at the percent of predictions that are correct using a different "test" dataset. We will consider a prediction to be accurate if the estimated wavenumber matches exactly with the true wavenumber.

Table 3.1: Frequencies of each wavenumber in the "test" dataset. These frequencies are weighted towards more common wavenumbers in current data as well as anticipated future trends of increasing wavenumbers.

Wavenumber	2	3	4	5	6	7	8	9
Frequency	10	20	40	60	60	60	40	40

In building a "test" set of simulated surfaces, we chose a number of surfaces to try to mimic the general frequency of each wavenumber that we tend to see in real data, with perhaps a bit more emphasis on high wavenumbers, since data points to an decreasing temperature gradient which should lead to an increase in Rossby waves over time. The set of simulated surfaces with each wavenumber is given in Table 3.1. It may be possible that we have overrepresented the number of nine wave surfaces we include in the test set; prediction results will be shown for each wavenumber separately for the convenience of the viewer.

Chapter 4

Interpreting Multiscale Persistence

4.1 A Return to the Seahorse Plot

Each seahorse plot is a visualization of a sequential filtration, which shows the number of dimension 0 features across all latitudes at a given filtration height and a given latitude. Recall from the initial description that this visualization is a multiscale approach for understanding both how many features are present at a certain filtration height and captures some big-picture information about where those features occur. When used to describe an isobaric surface, the seahorse plot generally includes the two features: shape and structure.

First, we will address the shape of a seahorse plot. The non-trivial region of the plot is tapered at the high and low ends of latitude. These tapers follow naturally from the nature of the surface. Closer to the north pole, the circumference of a given latitude line decreases. This indicates that for northern latitudes, the distance from one longitude to the next (in meters) becomes smaller and smaller. Additionally, once one reaches latitudes above the jet stream, the plot

should not show additional signal due to Rossby waves, and other sources of topological features are not very persistent in height or across latitudes. At the "bottom" end of the seahorse plot we can see indications of the presence of topological features around the equator, which is south of the jet stream. Our understanding of atmospheric science and empirical data indicate that a seahorse will not detect dimension 0 features related to Rossby waves at equatorial latitudes. and there are not any or many features at any given height at these central latitudes. This feature is one of the main reasons for the name of the plot, as the shape resembles seahorse with an extended tail, as in 4.1.

Figure 4.1: A a seahorse, the inspiration for the name of the sequential filtration plots used throughout this work. Photo Credit: Charleston Animal Society

The second characteristic that shows up in geopotential height data is the structure of the shape that emerges. We notice this behavior in a couple key ways. At the tapered edges (which we sometimes call the "head" and "tail" of the plot), we notice a scattered number of pixels with three or more features. These are usually fairly isolated and show instances where a feature may have a particularly short lifetime if it persists for very few filtration heights or a may

not persist over many latitudes. In many cases, small sources of noise are subject to both of these, leading to pixels that reside in relative isolation in the plot.

Another way that layers come into play is found in the central region of the plot. Since Rossby waves persist over a larger range of both filtration heights and latitudes, the plot displays concentric regions of pixels with a given feature number, with higher numbers in the middle. This trend creates a region that we sometimes call the "heart" of the seahorse plot. Generally speaking, we would expect that the number of Rossby waves in an isobaric surface would be indicated by the highest feature number that is prominently present in the heart, as such a structure indicates that a given number of features are present over a variety of heights and latitudes; this indicates persistence across both scales.

4.2 Identifying Trends in Seahorse plots

Visualizing a sequential filtration can be a useful tool, but understanding which parts of a seahorse are the most important is critical in interpreting these plots. In this section, we will discuss which features are most meaningful, which are less meaningful, and how to best interpret a set of seahorse plots. These interpretations are useful in many scenarios, including ones where we want to understand how this information changes from one day to the next.

We begin our explanation of how to understand and interpret these plots by considering all the possible range of feature numbers which we may expect to see in these images. Although seahorse plots display the number of points whose sequential filtration yields a specific number of dimension zero features, (one, two, etc) we recall that we are using 'superlevel' sets in these filtrations. Subsequently, a surface with w Rossby waves, each forming a trough shape in the isobaric surface, should detect $w + 1$ dimension zero features. This

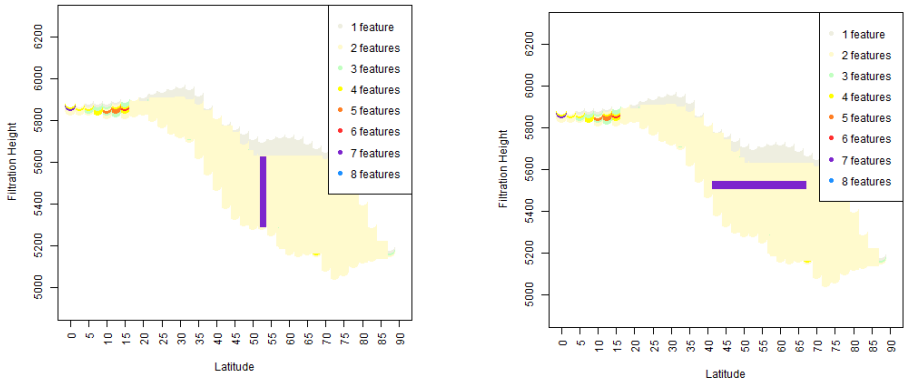
means that anywhere where we detect zero, one, or even two features is not a combination of latitude and filtration height where we believe there is any strong indication of Rossby waves. This is because we typically expect to see at least two Rossby waves in the jet stream on any given day. Accordingly, the coloring used to create seahorse plots for an isobaric surface uses more muted tones for $w \in 0, 1, 2$.

With this increased understanding of some of the lower-count pixels in the plot, we turn our focus on the presence and size of clusters of points with like feature counts. In Chapter 2 we highlighted a key motivation for this work: Rossby waves should be visibly persistent across many adjacent combinations of filtration height (an indication of depth of the waves) and latitudes (an indication of the north-south amplitude of the waves). This means that although we can see many small details in a seahorse plot, we may want to direct much of our focus on clusters of points with similar feature counts. A single isolated pixel whose feature count is eight is almost certain to be the result of random fluctuations in the isobaric surface since it demonstrates no persistence in either dimension: height or latitude. However, a large cluster where the seahorse sees mostly five and six as feature counts (such as the 'heart' of a seahorse) is likely an indication of four to five persistent waves, which may all be Rossby waves.

It is certainly possible for a seahorse plot to include more than one cluster, but we find that seahorse plots don't typically feature many clusters in the 'head' or 'tail' and typically only feature one large cluster in the central region, typically $25 - 65°$ N. This central cluster features combinations of filtration height and latitudes that we would typically associate with Rossby waves. In many or most seahorses based on daily average heights, we find that there is a single such cluster. In the winter months, the heart includes many points with three to five features, indicating evidence for the presence of two to four Rossby waves.

In summer months, this cluster often features many points with five to eight features, indicating evidence of four to seven Rossby waves.

We reiterate here the importance of clusters in these plots, especially in the heart of the plot. Clusters of connected points with identical or very similar feature counts show persistence across multiple scales. We introduce two hypothetical scenarios to solidify how persistence is connected to these plots.



(a) (b)

Figure 4.2: Hypothetical examples of persistence along each scale. (a) shows persistence across filtration heights alone, while (b) shows persistence across only latitudes.

First, consider a seahorse plot that looks much like Figure 2.3, except that the entire central region consists of only points with four features detected, so the central region is almost completely yellow, with the following exception: At $50°N$, there is a column of purple points (indicating seven topological features of dimension zero) which starts at 5250m above sea level and ends at 5650m.Figure 4.2(a) displays such a configuration.

This vertical cluster shows strong evidence of persistence across many filtration heights, as the feature number does not change meaningfully over that entire 200 meter range.

As a related but differing example, consider the same plot with homogeneous feature counts of '4' across its central region with a different twist: in this

plot, suppose that the feature count was seven at the filtration height of 5500m from $40°N$ continuing until $70°N$. Figure 4.2(b) displays such a configuration. This connected region would be an indication of persistence across that entire range of latitudes, even though it is not an indication of persistence across many filtration heights.

The two previous examples are unrealistic; it is very rare to persist along many filtration heights but no latitudes or vice versa. However, we can see from these examples that a cluster may persist more along one scale than the other, and a roughly 'square' cluster is an indication that a cluster persists across both scales. Thus, clusters in a seahorse plot provide information about multiscale persistence. We are not aware of other work that attempts to capture persistence in this manner. While this particular application captures persistence across filtration heights and latitudes, one could certainly apply the same framework to persistence across longitudes. More generally, one could examine persistence across any two scales, although the use of sequential Morse filtrations to analyze multiscale persistence would suggest that at least one of those scales be reasonably considered as a response variable, as is the geopotential height in our data.

This work helps to highlight one quality of analyzing multiscale persistence. In a low-dimensional dataset like ours (where $k = 3$), the choice of scales is fairly straightforward. However, in the case where statisticians wish to perform similar analysis on more high-dimensional data, this choice might not be as obvious.

Another point to consider is what a seahorse plot does and does not communicate. We will compare a seahorse plot like Figure2.3 with a barcode like the one in Figure 1.9. The dimension and type of data is a bit different between the two, but both give a sense of both the number of features present at various

filtration values as well as the birth and death of features. In the barcode, we track this information by showing the lifetime of each feature distinctly. The seahorse instead shows this information as a summary of all the features, as the color of any given pixel displays the number of features which are still 'alive' at a given filtration value and latitude. Both figures convey information about which filtration values are the floor and ceiling of reasonable filtration values to consider; the barcode by showing a minimum and maximum along the x axis and the seahorse showing the general shape for which it is named (which may look very different if used for data other than geopotential height values).

While these two summaries have some similarities and some differences, it is clear that each one is designed to properly summarize the persistence of dimension-zero features for the corresponding data. We recommend the use of seahorse plots for data that have a similar overall structure to three-dimensional data with a single response value as one of the dimensions. For practitioners interested in summarizing data with a higher dimension, we suggest looking at barcodes and seahorse plots for suggestions of how to develop additional tools for visualizing and summarizing those data.

If seahorses are a useful display of multiscale persistence, what can they tell us about how Rossby waves change over time? Before we move ahead too far, its important to be clear about what we do and do not know from looking at feature numbers in a visual like the seahorse plot. First, these plots endeavor to provide information about the number of sublevel sets or troughs on our response surface. A feature number of four would indicate that there are four troughs on the surface at a given combination of latitude and filtration height. The Morse filtration does not have any way to designate the source of any given trough (or peak), and we do not claim to interpret such a designation. However, Rossby waves are known to persist across a moderate to large range of filtration

heights and latitudes. Based on this knowledge of how these waves manifest in our data, we can use information about feature numbers to understand and infer meaningful information about the number of Rossby waves. One of the key ideas we leverage here is that Rossby waves are among the features with the greatest degree of persistence in these geopotential height surfaces.

We should view the feature numbers on a seahorse plot not as an exact value of the number of Rossby waves for a given day, but rather as a good estimate for the upper bound of the number of Rossby waves present in the data. For any given combination of filtration height and latitude, the number of dimension-0 features is not a declaration that "there are w waves present on this day", but rather as a general indication of the maximum number of waves which may be present. We'll discuss this more, but we are particularly focused on the heart of each seahorse plot.

While there are multiple ways to approach the interpretation of this multiscale version of persistence, we would like to focus on two specific ways that we can discuss changes in Rossby waves over time. First, how do wavenumbers change throughout a single year? In Figure 4.3, we can see a seahorse for day 90, 180, 270, and 360 of 1980. These days were chosen to give a spaced out view of seahorses through all four seasons within a given year.

As seen in Figure4.3, we can see that the overall structure of the seahorse plot remains largely the same through the year, as mentioned above. In spring and summer like days 90 and 180, we can see that there are varying strength of evidence of a cluster of six or seven features ,as in Figure 4.3(a) and (b). However, we can see that in fall and winter, such as in Figure 4.3(c) and (d) , there is little evidence of a six or seven feature cluster. This is expected, as scientists generally agree that there are more Rossby waves in warmer seasons, and fewer of these waves in colder months. While Figure 4.3 shows four days that were evenly

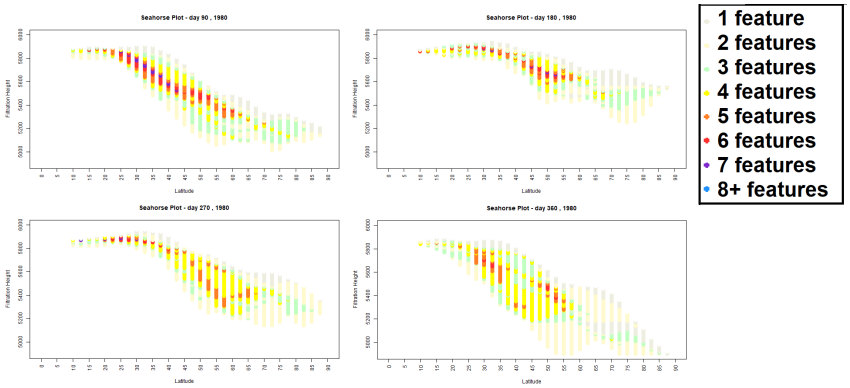


Figure 4.3: Four seahorses from 1980, spaced across seasons. From left to right and top to bottom, March 29, June 27, Sept 25, and Dec 25. In the winter (bottom right), the heart has lower feature counts, while summer and fall yield hearts with higher feature counts. We note the seasonal changes in the overall size of the plot: summer has the smallest plot overall, while winter and spring have the largest amount of nonzero pixels.

spaced, the results are largely the same across the span of one year. comparing seasons.

We have gathered that seahorses do display change in how many features are detected using sequential filtrations within a given year, and note that this is in line with general expectations about the ebb and flow of wavenumbers. A related question focuses on looking at wavenumbers across years. We'd like to use seahorse plots to investigate if there is any evidence pointing to a change in wavenumbers over time. While we do expect these numbers to shift within a year, it would be noteworthy if, for example, seven wave formations appear more often in the 2000s as opposed to the 1980s. We'll examine a few plots to look into this question.

In figure 4.4, we see seahorses for approximately the same days (90, 180, 270, and 359), this time in 2010. While we would expect to see some amount of random variation in seahorses from the same day across years, one thing that

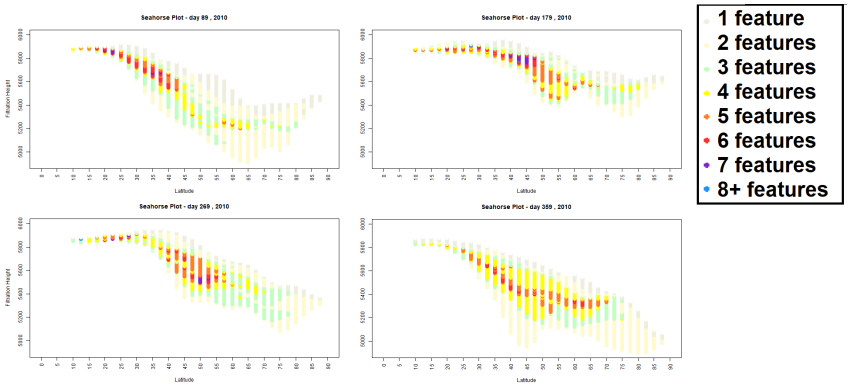


Figure 4.4: Four seahorses from 2010, spaced across seasons. From left to right and top to bottom, March 29, June 27, Sept 25, and Dec 25. In the winter (bottom right), the heart has lower feature counts, while summer and fall yield hearts with higher feature counts. We note the seasonal changes in the overall size of the plot: summer has the smallest plot overall, while winter and spring have the largest amount of nonzero pixels.

we notice is that this more recent set of seahorses seem to indicate an increase in both a) the highest wavenumber seen in the heart of the seahorse plot and b) the size of the central region of that plot. These changes over time are consistent with the hypothesis that the polar jet stream experiences more Rossby waves now than it did in 1980, and could provide some evidence to support the notion that wavenumbers have increased over time.

To pick one particular point of reference, let's examine day 270 from 1980 and 2010. In 1980, the central region of the seahorse includes a combination of filtration heights and latitudes that range from about 5300 to 5600m and $32 - 40°N$, respectively. In this cluster, we can see that the largest wavenumber in the cluster is four (remember that the wavenumber is one less than the number of features found in the seahorse). If we then turn our attention to day 270 in 2010, we now see that the location of the central region has shifted somewhat, and is located in the range of 5400-5700m for filtration heights and $35 - 45°N$.

This small shift in location is not particularly noteworthy, but something else about this central cluster is quite meaningful: the maximum wavenumber in this cluster is six (up from four in 1980) and most of the filtration values indicate the presence of at least five waves.

Comparing just one pair of days is not intended to serve as proof of changing wavenumbers over time, but this evidence is a meaningful display of some of the changes consistently present in such comparisons; seahorses in 2010 show larger clusters with higher wavenumbers. We are now motivated to take a more systematic look at changes in wavenumbers, cluster size, and other differences in the persistent homology of geopotential height data.

4.3 Summarizing Trends in Seahorse Plots

We've discussed each of the kinds of trends we see in seahorse plots: within a single year, such as a seasonal trend, and across years at the same season to see/show changes over time within a season. While each plot does show some meaningful information about the number of waves or other features on a given day, we'd prefer to have a way of summarizing this information in a more condensed manner, such as numerically or with a single graph to show trends as opposed to a separate plot for every day of data. In this section, we'll discuss a several potential ways to create this type of summary and evaluate the usefulness of each one.

In discussing this usefulness, we need to reiterate that reanalysis data do not have a "true" wavenumber, so it is not possible to directly measure how well a summary displays this information. However, we can create many simulated isobaric surfaces which are created to have a specific number of waves. We can

evaluate each candidate summary for seahorse plots by counting how many plots are correctly identified using that summary.

Since seahorse plots consistently show trends in shape and structure, it may be useful to consider a tabular summary of feature numbers in the plot. Such a summary records the frequency of pixels that where the Morse filtration detects each number of features. Taken in this way, we can display these frequencies as in Table 4.1.

Table 4.1: The general form of a frequency table of feature numbers

Feature Number	1	2	...	8	9+
Frequency	n_1	n_2	...	n_8	n_{9+}

We'll use these tables to discuss how to summarize the results of a seahorse plot.

Since the goal here is to identify the number of waves present in an isobaric surface and we are accomplishing this by comparing real days to simulated ones, it is important for us to describe how to make these comparisons. Due to the amount of variation present in seahorse plots, it is unlikely that any two seahorses will be exactly identical in their feature number tabulations. Indeed, even for two simulated surfaces with the same starting conditions, we would expect mild to moderate variation due to randomness alone. However, the feature counts should follow some sort of pattern based on the wave number. For each metric, part of the discussion will focus on how that metric will be compared to simulated surfaces to make an estimate at the true wave number. We will include information based on simulated surfaces. The advantage here is that we can compare the estimate against a known wavenumber, which is one of the initial conditions in the simulation process.

The first candidate method we examine is to look at what proportion of the seahorse plot has a feature number greater than some given number i. One can

think about this metric as the proportional prevalence of high feature counts. This proportional prevalence measure for a seahorse could be measured by computing

$$p_{prev,i} = \frac{n_i + n_{(i+1)} + \cdots + n_{9+}}{n_1 + n_2 + ... + n_{9+}}.$$

Based on this measure, we can get several different values for each surface ($p_{prev,2}, p_{prev,3}, p_{prev,4}$ etc.). This set of proportional prevalence values can be viewed sequentially as in a scree plot. Using cross validation of 8,000 simulated surfaces (one for each wavenumber from two to nine), we found that the most accurate way to use $p_{prev,i}$ is to predict that there are w waves in a surface, where w is the smallest value of i with $p_{prev,i-i} < .51 * p_{prev,i}$. This choice means that we are not looking for a static threshold value for $p_{prev,i}$; predictions are all relative based on the overall set of $p_{prev,i}$ values.

This prediction method has a few bits of interesting behavior and overall it does not perform at a particularly high level. Table4.2 shows the outcome of using this model for predictions using our set of 330 simulated surfaces. Using $p_{prev,i}$ tends to yield correct predictions a little under half of the time. One big issue we face with this method is that we are especially prone to underestimating, especially for large wavenumbers.

One reason for this poor performance is that pixels that are far from the central cluster in the seahorse are counted as being informative about the number of Rossby waves, which we would typically expect to see only in the heart of the seahorse. This means that n_4, n_5, and similar counts are inflated due to sources that are not part of the true pattern.

A second candidate method to examine is based on the first. Since $p_{prev,i}$ is sometimes prone to estimation inaccuracies in Rossby wave estimation due to behavior in the tails, we propose a metric which operates in a similar manner, conditioned on the latitude of the points. In particular, we propose measuring

Table 4.2: A confusion matrix of predictions from 330 simulated surfaces using $p_{prev,i}$. We note that this method almost never over-predicts the true number of waves, but is sometimes prone to underpredicting. The overall accuracy of $p_{prev,i}$ is about 48.4%.

Predicted waves	2	3	4	5	6	7	8	9
2	8	1	0	0	0	0	0	0
3	1	19	11	2	0	0	0	0
4	1	0	29	21	1	0	0	0
5	0	0	0	37	28	4	0	0
6	0	0	0	0	31	33	5	0
7	0	0	0	0	0	23	28	6
8	0	0	0	0	0	0	7	28
9	0	0	0	0	0	0	0	6

this by quantifying the proportion of all pixels with a feature count above four subject to a condition: we will count only those pixels with such a feature count that lie between $25 - 65°N$ in the numerator of our fraction. In this manner, we will capture all the information from the heart where Rossby waves tend to show up without inflating our sense of the prevalence via the additional noise from the tails of the seahorse plot. Such a measure has the added benefit of ignoring features out in the tails of the plot, which are certainly not evidence related to the feature of interest. The formula for this metric is

$$p_{Tprev,i} = \frac{n_{iT} + n_{(i+1)T} + \cdots + n_{9+T}}{n_1 + n_2 + ... + n_{9+}}$$

, where T is used to denote the trimmed region of interest used in estimating prevalence.

This approach gives us eight different values, one for each value of i from two to nine. Just as we did with the set of $p_{prev,i}$ values, we used cross validation to determine a method for selecting the "best" wavenumber to predict. In this

case, we found that using $p_{Tprev,i-i} < .65 * p_{Tprev,i}$ gave the best results on the 9,000 simulated surfaces in the training dataset.

Compared to p_{prev}, p_{Tprev} performs better, as shown in Table 4.3. Overall, we feel that this metric is less than ideal because the prediction is based on comparisons between consecutive pairs of values instead of using as much information from the seahorse plot and accompanying set of frequencies as possible. The loss of resolution is expected with this method, and the prediction accuracy of this metric is not as high as we might hope, even for the first work on this method of estimation. Overall, we do find that this method has a prediction accuracy over 67%, which is encouraging and reinforces the idea that the tails can be a bit distracting overall.

Table 4.3: A confusion matrix of predictions from 330 simulated surfaces using $p_{Tprev,i}$. We note that this method almost never over-predicts the true number of waves, but is sometimes prone to underpredicting. The overall accuracy of $p_{Tprev,i}$ is about 67.5%.

Predicted Waves	2	3	4	5	6	7	8	9
2	5	0	0	0	0	0	0	0
3	1	16	3	0	0	0	0	0
4	3	3	36	7	0	0	0	0
5	1	1	1	53	15	1	0	0
6	0	0	0	0	45	22	2	0
7	0	0	0	0	0	37	17	3
8	0	0	0	0	0	0	21	27
9	0	0	0	0	0	0	0	10

In a slightly different approach, we propose a metric which simultaneously uses all the information in the original table. We can directly compare the set of observed frequencies of each feature number, starting at 1, and compare them to the set of counts recorded for the simulated surfaces in our training dataset. The distance measure is a familiar formula used in comparing expected and

Table 4.4: A confusion matrix of prediction results using χ^2. This metric never predicts a two-wave structure, and is more prone to over-prediction than either of the other two. Overall, this metric performs rather well, with a prediction accuracy over 86%.

Predicted Waves	2	3	4	5	6	7	8	9
2	0	0	0	0	0	0	0	0
3	9	18	5	0	0	0	0	0
4	0	2	35	4	0	0	0	0
5	1	0	0	54	0	0	0	0
6	0	0	0	2	59	1	0	0
7	0	0	0	0	1	59	0	0
8	0	0	0	0	0	0	31	9
9	0	0	0	0	0	0	9	31

observed counts, and is calculated as

$$\chi^2 = \sum_{i=1}^{9} \frac{(observed_i - simulated_i)^2}{simulated_i}.$$

This metric is not a test statistic, as it is not used in traditional inference, but uses the same name since the formula is reminiscent of the familiar statistic.

Using the χ^2 metric one would obtain a different value for each candidate wave number due to using a different set of simulated counts. We propose that the estimated wave number be whichever set of simulated values yields the smallest value of χ^2. Using this strategy is fairly promising. A confusion matrix of estimates is shown in Table 4.4.

We've described three different measures here for estimating Rossby wavenumbers. Each of the three has a natural explanation for why it might be used, and we see varying levels of performance for the three metrics. Unsurprisingly, χ^2 performs the best, likely due to the fact that it considers all the different feature counts simultaneously instead of combining those counts into a smaller sum-

mary. The prediction accuracy with this metric of over 86% serves as a strong indication that this metric can be used to estimate Rossby waves more broadly for data similar to the test set, which certainly includes all NCEP/NCAR data in the 1980-2010 era.

4.4 Alternative summaries for seahorse plots.

4.4.1 Movies

We've discussed several different ways to summarize information from seahorse plots, especially as it relates to using these summaries to estimate the number of Rossby waves in an isobaric surface. During our work, we encountered some other tools for summarizing these plots. While they have not yet proven useful for our goals, we feel that it's worth mentioning them here. These visualizations are interesting and show some meaningful information, and future work in this direction may wish to employ some or all of these tools in gaining more understanding from seahorse plots.

One way that may not feel like a "summary" per se is to view a video or animation of seahorses for a specific time period, such as a single year. We've created several animations of these plots, with a timing of around 1-3 seconds per plot. By watching these videos, one can get a better sense of how the seahorse and hence the isobaric surface changes over time. Among the behaviors that we have noticed in most or many years, it is quite typical to see the trends mentioned above: namely, the shift in shape, size, and location of the plot seasonally and the changes in the behavior of the central cluster or "heart" of the plot. Here is a link to one such video: **Seahorse Plots - 2000**

While these animations can help one to conceptualize the changes in Rossby wavennumbers through time, it is true that there is not an especially clear path from these videos to inference. While this is true, there is definite precedent for the use of movies as a dynamic representation used in assessing a summary Lazar and Kadane [14]. We feel that this movie may serve as a similarly useful tool. As noted by Lazar and Kadane, movies are flexible in the period that you

choose to display, they can help to display characteristics that might be difficult to observe otherwise, and movies can be used for pragmatic purposes such as identifying an outlying day which may feature an especially strange plot.

We've actually used one video-adjacent example already in this chapter: when we investigated four distinct seahorse plots from each of 1980 and 2010, we were basically analyzing a very granular version of the type of video we've discussed here. Instead of considering general trends in a seahorse plot based on a subset of days, videos allow us to see the entire time course we can get from the original data. Throughout this work we've used daily mean geopotential heights, but videos could display this data at any resolution, which could include NOAA's four times daily data for geopotential heights.

We used movies like these multiple times when looking at trends in these plots, but we were unable to develop any hard and fast rules for how to make estimates of wavenumbers based on the videos directly. While investigating these plots, it seems clear that there may be more work to do with these videos.

4.4.2 Ridgeline plots

We've just discussed the concept of using a movie or animation to view how seahorses change over time, especially over several or many consecutive days. In section 4.3, we discussed the idea of measuring quantities like p_{prev}, p_{Tprev}, $p_{center,i}$ and χ^2. Our goal there was focused on looking at these metrics for one surface. Doing so allowed us to estimate the number of Rossby waves present on a certain day. Here, we call back an alternate summary which may allow us to look at similar information over time.

While the traditional series or time-series plot allows one to examine how the value of some measure changes over a given time course, we showed a visualization in Figure 2.6 which varies somewhat from a traditional time-series

plot. We propose the use of ridgeline plots, as in Figure 4.5 to visualize any of the following over time:

1. p_{prev}
2. p_{TPrev}
3. $p_{center,i}$
4. $n_{4+}, n_{5+}, n_{6+}, n_{7+}$ or similar metrics
5. χ^2

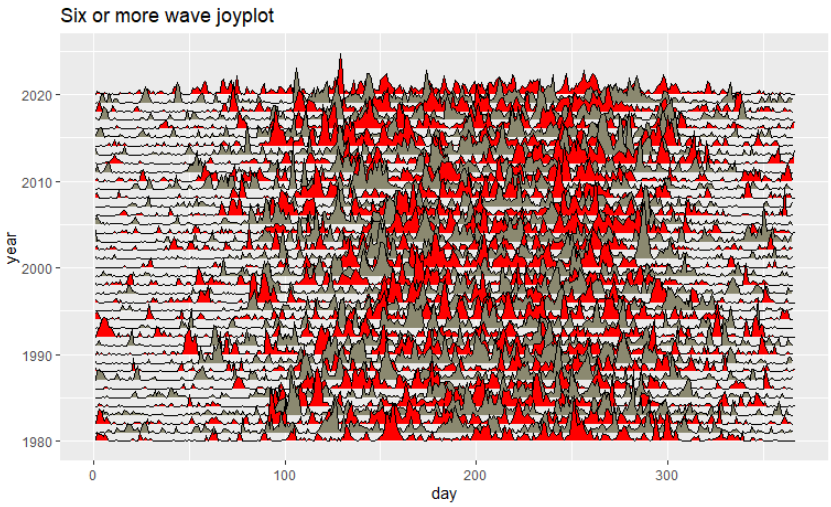


Figure 4.5: An example Ridgeline plot of n_{6+}, 1980-2010. While the trends are a bit difficult to parse in this plot, we do see the expected behavior of much higher values in the middle of the year in summer and fall months, and lower values in winter and spring months. There is a slight increase in the average value of n_{6+} over time

The reasoning behind the first three proposed plots should be fairly clear: the same reasoning to look at any of these measures for an individual day is the very reason why it would be beneficial to look at that measure over time. Each of these three measures are clearly defined for a specific surface, so creating a

joy plot for these surfaces would be a way to see how the values of that measure change over the entire time course, both within and across years.

The rationale behind 4. is not entirely different, but we note that these are no longer displayed as proportions. This means that if the entire seahorse grows in size, proportional measures will not display any change, while measures like n_{4+} will increase. A natural consequence here is that a measure based on frequencies may do a better job in a circumstance where one believes that the entire seahorse is growing, as opposed to only the central region growing. One potential expected result from these metrics comes from the higher indices on n. Since the number of Rossby waves is subject to some amount of change over time, let us consider n_{8+} as an example. In a typical year, we would expect that this measure has a frequency of zero or some very small number for the winter, spring, and fall seasons for the most part. On days where we see five or six Rossby waves, we may see a nonzero value for n_{8+} which is still small. If there is a day with seven or more waves, then we would expect to see a higher count for n_{8+}. Thus, if we create a joy plot, we expect to find mostly a flat series outside of the summer months. If the plot yields higher values of n_{8+} as you move forward in time or backward in time, this could point to a transition towards or away from seven-wave days.

Preliminary investigation did not yield much indication of a large shift of this nature, but would certainly be a measure to keep in mind in looking for shifts in the number of waves. One could spend some more time investigating the presence or absence of patterns in these ridgeline plots, and we believe that it is a very reasonable place to point future efforts on this type of research.

While we will not spend an immense amount of time discussing proposal 5., we'd like to explain our overall approach and the reason why it may not be ideal to create ridgeline plots in this way. First, there is not only one χ^2 measure

for a given surface on a given day; there is a different set of χ^2 values for each set of simulated counts (which are based on the average of multiple simulated surfaces). If one has a target number of waves, say five, then it is possible to look at the χ^2 values using that set of simulated counts, but the issue would then be that there is no way to tell from the ridgeline plot whether χ^2 is lower for six features than other feature counts. There are two other ways that one could consider ridgeline plots as it pertains to the χ^2 measure.

One somewhat novel idea would be to use the χ^2 measure to estimate the most likely wavenumber on each day through the year and then plot that as a series plot or ridgeline plot. This would certainly not serve any inferential purpose or operate as a way to estimate values, but could be one way of displaying the results of said estimation.

Another way of considering the χ^2 measure is to pick a short time course (such as one month or one year) and then to create a ridgeline plot where each series plot is one set of χ^2 measures instead of a different year. This may help investigators to get a sense of which χ^2 values are more likely or less likely at various points in that time course. We would generally expect that lower wavenumbers have the smallest χ^2 values in the winter months, while larger wavenumbers would generally have higher χ^2 values in the summer months. These ridgeline plots would need to employ some sort of transparency, and tracking the smallest value for each day may be tricky due to the nature of the ridgeline plot (with vertical offsets for each separate series in the plot). Further investigation of this style of ridgeline plot may yield additional information, and we encourage interested parties to take a look for any other patterns we did not notice in such visualizations.

Creating these ridgeline plots enables us to see how each of these metrics change in the two manners we've described before: change within a year and

across years. We observe the change within a year by looking at how values of the plot change within a single horizontal series plot. as well as looking at change across years. We observe the change across years by looking at how values of the plot change from one horizontal series to the next, hence showing the change in the given metric year-over-year, instead of day-after-day.

4.5 Future Work

We've discussed the use of sequential Morse filtrations for estimating Rossby wavenumber, in part by using simulated isobars to evaluate various methods of estimation. While this application serves as the primary motivation for the development of sequential Morse filtrations, our work has pointed to several directions for future work. In this section, we describe some potential directions for future work. Some of these directions discuss future developments related to Rossby waves in the polar jet streams while others point to uses of sequential Morse filtrations elsewhere.

4.5.1 Simulation Settings

We created the initial conditions and settings of the simulation with the goal of creating simulated surfaces which are an accurate representation of the types of surfaces that we see in 1980 - 2010, but with further research, we believe that it may be possible that other suitable initial conditions exist and may be even better for creating simulated surfaces. As an example, perhaps some researchers may want to design a simulation where the north and south ends of waves 'curl' instead of forming a largely vertical pattern. Another example could be that some wish to impose additional or different types of noise beyond that which we instituted.

Some differences which some may think are important or necessary may include turning the simulated surface into a half-spheroid or another similar shape. While this is not wrong, the result will be the same, as the homology of the surface is the same whether it is created as a four-sided surface or a hemisphere. Different conditions or settings which create a more 'accurate' depiction of the surface are totally acceptable, but the results of any analysis will be generally identical. The only potential issue we need to resolve when performing TDA on a real isobaric surface is to make sure to "cut" the hemisphere at the latitude with the highest average height. This will have the effect of making sure that the surface is highest at the edges, just like the simulated surfaces.

One note that we make here is that some may wish to create a simulation which simulates the behavior of the northern hemisphere or globe over the entire time course, or at least for multiple days at once. We do not criticize this notion, but we do offer one perspective; no matter whether a simulation is created for one day or multiple days, we have yet to encounter examples where estimation of wavenumber is based on information outside of a single isobaric plot. From this, we believe that creating simulated surfaces in isolation is appropriate: surfaces are created in roughly the same way that they are assessed, which should lead to few or no issues with using a simulation created in the manner we used. Again, others may find positive results with a different simulation; we encourage the use of any tool that can improve estimation of wavenumbers through TDA and the sequential filtration.

4.5.2 Alternative Metrics for Estimating Wavenumbers

In Section 4.3, we discussed three different metrics for estimating the number of Rossby waves in a surface. We used these metrics along with a training and test dataset to evaluate the accuracy of each metric. While the best of these metrics,

χ^2, has a very reasonable measure of accuracy at over 86% accuracy, we readily acknowledge that seahorse plots and their summaries as in Table 4.1 may be summarized in numerous ways. We describe one avenue for exploration here and acknowledge the possibility for other metrics as well.

In contrast to the measures used in Section 4.3, we'd like to mention one metric focused on the heart of the plot. This cluster typically has a known structure; namely, we observe concentric regions, with the innermost regions indicating higher feature numbers. If we take ω to be the maximum wave number in the central region and then count all pixels which meet two criteria:

a) the pixel shares a side with the central cluster (which starts at just single point)

b) the pixel's wave number satisfies $w_i \geq \omega$.

By using the above criteria on all pixels in the plot (or by iterating around the points connected to the initial heart), the central cluster of pixels will grow to include all points that are connected to the center where $w_i \in \{\omega, \omega-1, \omega-2\}$. The resulting cluster typically will have the appearance of three concentric rings. Since this approach only considers pixels which are connected to the main central cluster, it will be completely unaffected by any tail behavior. We can compute this metric using

$$p_{center,i} = \frac{C_i}{n_1 + n_2 + ... + n_8 + n_{9+}},$$

Where C_i is the pixel count of the set of connected pixels using the above criteria. One could formulate this metric for a different range of feature counts, although we found that two consistently yielded the best results.

We believe that such a metric could serve as part of an approach for estimating wavenumbers, but the current description does not lead to a clear set of

values to compare for differing candidate wavenumbers. With the three metrics we used, it was clear how to go from the values produced by that metric to an estimated wavenumber. With this general approach, we do not have the same path which makes the estimation process unclear.

4.5.3 Number of Hearts

In examining seahorse plots over many days from 1980-2010 as well as other years, we occasionally witnessed a somewhat strange phenomenon. Instead of a plot which featured a single central region (which we've sometimes called a 'heart'), we have occasionally witnessed two distinct central regions. We will say that a seahorse plot has a second central region if it meets the following criteria:

1. the region has a highest feature number w' and two concentric regions with feature numbers $w' - 1$ and $w' - 2$ which are unconnected from the first central region.

2. this second feature occupies at least 20% as many pixels as the first central region (this number may need revision).

This type of structure, which we'll call a 'two-heart' plot, does not follow the expected structure of these plots. Part of the issue here is that with two central clusters, one may naturally wonder whether the Rossby waves are showing up in the first heart, the second heart, both, or neither. The truth is that there is not a single correct answer to this question. When we look at these two-heart plots, it does not seem that there is a clear pattern in how to estimate Rossby waves on these days. We believe that estimating wave numbers on these days is difficult or even impossible, and we leave it as an open question. Currently, we would perhaps suggest that parties especially interested in estimation of wavenumbers

may want to adapt their solutions to return a result of "Inconclusive" for any two-heart days.

4.5.4 Other Rossby Waves

While the additional topics one may investigate with ideas from this work may expand beyond topics already listed, we do want to list one additional applications for this work or something similar to it. The first is analysis of Rossby or Rossby-like waves outside of earth. In Covey and Schubert [4], the authors describe the presence of Rossby waves on Venus, while Solar Rossby waves are modeled and described in Dikpati et al. [7]. While the specific dynamics of Rossby waves on each of these objects may be different than the Rossby waves analyzed here, we do note that our work may have a natural extension to Rossby waves in other places.

4.5.5 Manifold Deformation

At its core, this work has provided a tool for describing the "texture" of a manifold. We used these tools for describing deformations in isobaric surfaces. None of this work is restricted to studying isobaric surfaces. The use of sequential filtrations could be a candidate method for describing the number and type of deformations on a manifold. Staying within the realm of studying various manifolds on Earth, we point oceanography as a natural arena for use of sequential Morse filtrations. Instead of going "up" from the surface, we could envision response values within a region (such as an individual ocean) as the depth below sea level where a certain pressure value or temperature is reached. Alternatively, the response value could be based on surface temperature, windspeed, current speed, or other features of the ocean. Some of these data may be available via the Argo satellite system Program [21] or from other sources. These data may help

those who study oceans to gain insights about deformations in manifolds of the above response values, which is a very natural extension of the work already included in this dissertation.

4.5.6 Access to Turnkey results

Part of the benefit we provide in this work is the development of tools which provide estimated Rossby wave frequencies for each day of NCEP/NCAR data. One tool which we do not present in this work, but which we plan on developing, is a web portal where researchers can provide various inputs such as: start date, end date, desired frequencies, length of the given wavenumber, choice of metric ($p_{prev}, p_{Tprev}, \chi^2$ or others), and others. The portal would then return to the user the range or ranges of dates where this method yields the given number of waves. To our knowledge, such a turnkey dataset does not exist. We are very excited to share this result publicly. Developing and maintaining this portal would also enable us to update estimates with additional metrics and different simulation settings.

Chapter 5

Conclusion

In this dissertation, we've identified an ongoing issue in atmospheric science research. Namely, many researchers are interested in identifying the number of Rossby waves in the polar jet stream given an isobar such as the 500 hPa surface. Those in the field have gone to far as to state that meaningful conclusions about climate, weather, and other events are dependent on the method used to summarize various attributes of these waves. We've shown how this question can be considered as a question about the number of persistent deformations in the surface using Topological Data Analysis(TDA). Our work here has focused on refining current TDA tools to analyze this type of data, developing a simulation to create an accurate representation of 500 hPa surfaces, and the development of three separate metrics whose accuracy points to encouraging potential for such a tool.

The Morse filtration in TDA is used to measure the number of persistent features of a given dimension and has been used in a variety of applications. When the response surface is dominated by a single feature, the "shape" of those data is not very compelling from a homological standpoint. While TDA has often been described as a way to analyze the "shape" of data, the nature of

Rossby waves requires us to know about the "texture" of the data instead. We developed a TDA tool which we call sequential Morse filtration which helps to measure persistence across two different scales. We can measure the persistence of features in the isobar across latitudes (one scale) as well as filtration heights (a second scale). This multiscale approach to TDA enables us to look at the "texture" of surfaces like the 500 hPa surface of geopotential heights. Part of this development involves the development of a new visualization tool for sequential Morse filtration which we've named the seahorse plot which serves as a way to visualize multiscale persistence.

We then developed a simulation to replicate the behavior of isobaric surfaces. This simulation has beneficial properties. It has an intuitive construction which allows for clear discourse between practitioners and statisticians, which is crucial in discussions for development of current and future approaches for analysis. The simulation is modular, which enables researchers to add new layers or remove existing layers as a method for envisioning these data in the presence or absence of various phenomena. Additionally, this simulation method is flexible; it is based on historical data but allows for the creation of surfaces which not represented frequently, such as very high wavenumbers. We then used this simulation to form a basis for measuring and estimating wavenumbers in cases where the true wavenumber is known; something we cannot do with existing data.

We have proposed three different metrics for this estimation process. All three have some intuitive application, but performed at different levels. The best measure, which operates similar to the familiar χ^2 value for a frequency table, obtains a prediction accuracy over 86%. This metric proves to be a promising overall metric for analyzing 500 hPa geopotential height data, which can enable the creation of a turnkey dataset for estimated wavenumbers using current

geopotential height data. We expect that the use of this estimation method will create greater cohesion among atmospheric scientists as it provides an clear and logical framework for estimating wavenumbers and helps to serve as an automated and accurate method for such estimation using sequential Morse filtrations.

This work is among the first approaches to multiscale persistence using a TDA framework and can be used to provide accessible and universal tools for Rossby wave estimation. We look forward to helping to promote adoption of sequential Morse filtrations for Rossby wave estimation. This work advances TDA by expanding descriptions of overall persistence of features using Morse filtrations towards describing more localized persistence using sequential Morse filtrations.

We also describe future directions for related work, including potential avenues for improving these estimation methods. We have noted some trends in seahorse plots which may merit further study, like the phenomenon of multiple hearts in a seahorse. We believe that the development of additional metrics for estimation remains an open question. This work also has natural extensions to studying Rossby waves in places other than Earth, and we also point to the use of sequential Morse filtrations more generally, as in the case of oceanographic data or any other manifold whose texture is of interest. We look forward to using sequential Morse filtrations in providing a turnkey dataset for Rossby wavenumbers thoughout all years with NCEP/NCAR data and maintaining development of related tools.

Bibliography

[1] A. Bjorner. *Topological methods*. Elsevier, 1995.

[2] Peter Bubenik, Gunnar Carlsson, Peter T. Kim, and Zhiming Luo. Statistical topology via morse theory, persistence and nonparametric estimation. *Journal of Machine Learning Research*, 2009. doi: 10.1090/conm/516.

[3] Gunnar Carlsson. Barcodes: The persistent topology of data. *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, 46:255–308, 04 2009.

[4] Curt Covey and Gerald Schubert. Planetary-scale waves in the venus atmosphere. *Journal of Atmospheric Sciences*, 39(11):2397 – 2413, 1982. doi: 10.1175/1520-0469(1982)039<2397:PSWITV>2.0.CO;2.

[5] Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. *Proc. Sympos. Point-Based Graphics*, 06 2004. doi: 10.2312/SPBG/SPBG04/157-166.

[6] Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. *Proc. Sympos. Point-Based Graphics*, 06 2004.

[7] Mausumi Dikpati, Peter A. Gilman, Subhamoy Chatterjee, Scott W. McIntosh, and Teimuraz V. Zaqarashvili. Physics of magnetohydrodynamic rossby waves in the sun. *The Astrophysical Journal*, 896(2):141, Jun

2020. ISSN 1538-4357. doi: 10.3847/1538-4357/ab8b63. URL https://iopscience.iop.org/article/10.3847/1538-4357/ab8b63.

[8] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.

[9] Jennifer A Francis and Stephen J Vavrus. Evidence linking arctic amplification to extreme weather in mid-latitudes: Arctic links to mid-latitude weather. *Geophysical Research Letters*, 39, Mar 2012. ISSN 00948276.

[10] Robert Ghrist. Barcodes: The persistent topology of data. *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, 45, 02 2008. doi: 10.1090/S0273-0979-07-01191-3.

[11] T. Hinks, X. Zhou, K. Staples, B. Dimitrov, A. Manta, T. Petrossian, P. Lum, C. Smith, J. Ward, P. Howarth, A. Walls, S.D. Gadola, , and R. Djukanović. Multidimensional endotypes of asthma: topological data analysis of cross-sectional clinical, pathological, and immunological data. *Lancet*, Suppl 1:S42, 2015.

[12] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, Roy Jenne, and Dennis Joseph. The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437 – 472, 1996. doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/bams/77/3/1520-0477_1996_077_0437_tnyrp_2_0_co_2.xml.

[13] Yohai Kaspi and Tapio Schneider. Winter cold of eastern continental boundaries induced by warm ocean waters. *Nature*, 471(7340):621–624, Mar 2011. ISSN 1476-4687.

[14] Nicole Lazar and Joseph Kadane. Movies for the visualization of mcmc output. *Journal of Computational and Graphical Statistics*, 11:863–874, 01 2012. doi: 10.1198/106186002376.

[15] Michael E. Mann, Stefan Rahmstorf, Kai Kornhuber, Byron A. Steinman, Sonya K. Miller, and Dim Coumou. Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*, 7(1), March 2017. doi: 10.1038/srep45242. URL `https://doi.org/10.1038/srep45242`.

[16] Chul Moon, Scott A. Mitchell, and Matthew Andrew. Statistical inference over persistent homology predicts fluid flow in porous media. *Water Resources Research*, 55:9592–9603, Nov 2019. doi: 10.1029/2019WR025171.

[17] Vladimir Osychny and Peter Cornillon. Properties of rossby waves in the north atlantic estimated from satellite data. *Journal of Physical Oceanography*, 34(1):61 – 76, 2004. doi: 10.1175/1520-0485(2004)034<0061:PORWIT>2.0.CO;2. URL `https://journals.ametsoc.org/view/journals/phoc/34/1/1520-0485_2004_034_0061_porwit_2.0.co_2.xml`.

[18] Lum P., Singh G., Lehman A., and et al. Extracting insights from the shape of complex data using topology. *Sci Rep*, 3, 2013. URL `https://doi.org/10.1038/srep01236`.

[19] Eric Palmen. The role of atmospheric disturbances in the general circulation. *Quarterly Journal of the Royal Meteorological Society*, 77:337–354.

[20] Mark E Piana, 2022. URL `https://groups.seas.harvard.edu/climate/eli/research/equable/hadley.html`.

[21] International Argo Program. Temperature, salinity, pressure, and biogeochemical profile data from globally distributed argo profiling floats, by month since april 2003 for the global argo data repository, containing data from 1995-09-07 to present.

[22] James A Screen and Ian Simmonds. Exploring links between arctic amplification and mid-latitude weather: Arctic warming and mid-latitude weather. *Geophysical Research Letters*, 40:959–964, Mar 2013.

[23] James A. Screen and Ian Simmonds. Amplified mid-latitude planetary waves favour particular regional weather extremes. *Nature Climate Change*, 4(8):704–709, June 2014. doi: 10.1038/nclimate2271. URL `https://doi.org/10.1038/nclimate2271`.

[24] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. ISBN 978-3-905673-51-7. doi: 10.2312/SPBG/SPBG07/091-100.

[25] Peter Effertz Thomas LeBlanc, Joe Halvorson. Rossby wave theory: An observational waves study. URL `https://meteor.geol.iastate.edu/~peffertz/portfolio/454%20Waves%20Report.pdf`.

[26] NASA Viz. NASA Viz: Aerial Superhighway, February 2012. URL https://svs.gsfc.nasa.gov/10902.