

Topological Data Analysis for Scoring the Periodicity of Northward and Southward Currents

Analyzing the Periodicity of Northward Surface Ocean
Velocity Through Sliding Windows and 1-Dimensional
Persistence Scoring

Elijah Guptill

June 2025

APPROVAL PAGE

TITLE: Topological Data Analysis for Scoring the periodicity
of Northward and Southward Currents
AUTHOR: Elijah Guptill
DATE SUBMITTED: June 2025

Senior Project Advisor

Signature

Mathematics Department
Chair

Signature

Contents

1 Introduction

Given time series data, we often want to determine whether the time series is periodic. [10] gives a short description of various methods which include Fourier-based methods and autocorrelation, both of which are popular. Fourier based methods often rely on Fourier coefficients relative to $\{1, \cos(nx), \sin(nx)\}_{n \in \mathbb{N}}$, or some other complete orthogonal list of $\mathcal{L}^2([0, 2\pi])$. The result is that these methods tend to favor specific kinds of curves. In the case of the complete orthogonal list $\{1, \cos(nx), \sin(nx)\}_{n \in \mathbb{N}}$, trigonometric curves are favored as more periodic. Sliding Window and 1-Persistence Scoring (SW1PerS) aims to avoid favoring a specific type of curve through topological data analysis [7]. This project aims to explore the motivation behind SW1PerS scoring and use it to identify where northward currents are most periodic within a patch of ocean near Morro Bay, California, using Near-Real Time Surface Ocean Velocity, U.S. West Coast, 6km Resolution Hourly data from the Integrated Ocean Observing System (IOOS). We will end by considering an alternative motivation for SW1PerS.

2 Topological Data Analysis

Topological data analysis aims to leverage the mathematical field of topology to quantify the shape of data. One popular technique is *persistent homology*, which utilizes algebraic topology. In essence, data points in a point cloud are connected if they all lie within an ϵ -ball with respect to a distance function. Often this is the Euclidean metric on \mathbb{R}^n where n is the number of data points. Given a specific $\epsilon > 0 \in \mathbb{R}$, homology can be used to count the number of “holes” in the data. The “persistent” part of persistent homology comes from the idea that we cannot know ahead of time what ϵ results in the best representation of the data’s “shape.” This issue is avoided by letting ϵ grow from near zero 0 to some large value so that every point in the point cloud is connected in every dimension. Persistent homology tracks when holes are “born” and “die” to quantify the shape of the data.

Sliding Window 1-D Persistent Homology Scoring (SW1PerS) endeavors to take time series, chop them into windows, and embed them in a high-dimensional Euclidean space. This embedding turns periodic functions into circles, which are 1-dimensional holes from the perspective of simplicial homology. This allows persistent homology to track the size of a hole by

recording when a hole is ‘born’ and when a hole ‘dies’. This lifetime is used to assign a periodicity score to the time series. By using homology, SW1PerS hopes to capture periodic functions that could be scored lower by other methodologies.

The background material in Sections 2.1 to Sections 2.2 are derived from [4] and Dr. Elena Dimitrova’s notes on topological data analysis.

2.1 Simplicial Complexes

Before we can see what a “hole” in data looks like, we need to learn how to represent shapes with algebra. We will start by representing shapes with building blocks called n -simplices.

Definition 2.1 (Convex Set). A subset of a Euclidean space (or more generally, an affine space over the reals) is convex if, for all pairs of points in the subset, the subset contains the whole line segment that joins them.

Example 2.2. Consider the shapes in Figure 1.

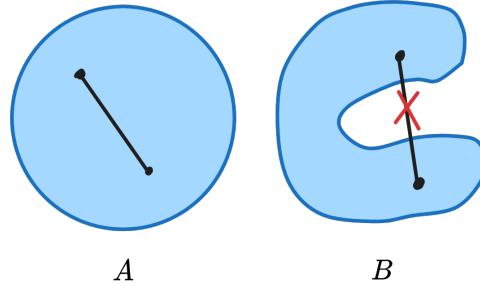


Figure 1: An example of a convex set A (a Euclidean disc in \mathbb{R}^2), and a set $B \subseteq \mathbb{R}^2$ that is not convex.

Definition 2.3 (Convex Hull). The convex hull of a shape is the smallest convex set that contains it. In other words, it is the intersection of all convex sets that contain a given subset of a Euclidean space.

Example 2.4. Consider the subsets $A = \{(0,0), (1,0)\} \subseteq \mathbb{R}^2$ and $B = \{(0,0), (1,0), (1,1)\} \subseteq \mathbb{R}^2$. Then the convex hull of A is the set $[0, 1] \times \{1\} = \{(x, y) : 0 \leq x \leq 1, y = 0\}$. Additionally, the convex hull of B is the set $\{(x, y) : 0 \leq x \leq 1, x \leq y\}$. See A and B , as well as their convex hulls in Figure 2.

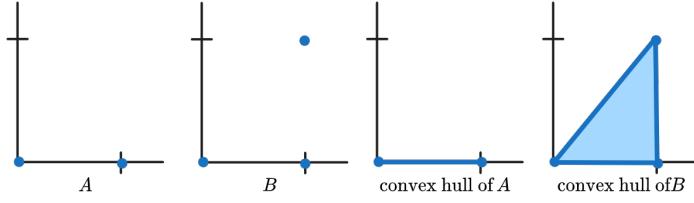


Figure 2: An example of $A = \{(0, 0), (1, 0)\} \subseteq \mathbb{R}^2$ and $B = \{(0, 0), (1, 0), (1, 1)\} \subseteq \mathbb{R}^2$, as well as their convex hulls.

Definition 2.5 (n -simplex). An n -simplex is the smallest convex subset in \mathbb{R}^n containing $n+1$ points v_0, \dots, v_n such that the vectors $v_1 - v_0, \dots, v_n - v_0$ are linearly independent. The n -simplex of v_0, \dots, v_n is denoted $[v_0, \dots, v_n]$.

Example 2.6. Figure 3 shows an n -simplex for $n = 0, 1, 2$ in \mathbb{R}^3 .

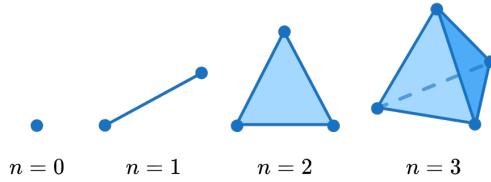


Figure 3: An example of an n -simplex for $n = 0, 1, 2$.

We can see that a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a filled-in triangle, and a 3-simplex is a filled-in tetrahedron. Given an n -simplex, $v_1 - v_0, \dots, v_n - v_0$ being independent ensures that the points of an n -simplex are linearly independent when one of the points is translated to the origin. This prevents us from picking 3 colinear points for a 2-simplex, which would result in a line (or point) instead of a triangle. Of course, this precaution extends to each dimension.

n -simplices act as building blocks with which we express shapes. If we do this algebraically, we will be able to represent holes in shapes by generators of a group.

Definition 2.7 (Simplicial Complex). A collection \mathcal{A} of subsets of a given set A is a simplicial complex if every element $\sigma \in \mathcal{A}$ has all of its subsets $\sigma' \subseteq \sigma$ in \mathcal{A} .

When the idea of a simplicial complex is applied to collections of simplices, we see that Definition 2.7 ensures that if a simplicial complex S contains a simplex, it must also contain the smaller simplices making it up.

Example 2.8. Consider the set $\{a, b, c, d, e, f\}$ and the simplicial complex

$$\mathcal{A} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{b, c\}, \{c, d\}, \{c, e\}, \{d, e\}, \{c, d, e\}, \{d, f\}, \{e, f\}\}.$$

Interpreting these subsets as simplices, see that \mathcal{A} contains the 2-simplex $[c, e, d]$ which is a filled-in triangle. Then \mathcal{A} contains its subsets. These are the triangles: $[c, e, d]$, lines: $[c, e]$, $[e, d]$, $[d, c]$, and points: $[e]$, $[d]$, $[c]$ making it up. See Figure 4

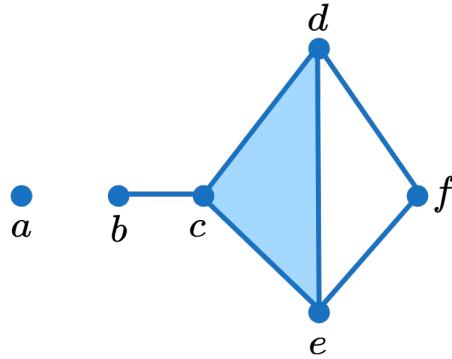


Figure 4: A representation of the simplicial complex from Example 2.8 using n -simplices.

Now that we have n -simplices, we can look at how to combine them to make more complex shapes. We will now impose a group structure upon simplices. To do so, we will need to define the negative of a simplex

Definition 2.9 (Orientation of a Simplex). We put an equivalence class on n -simplices. That is, we consider n -simplices to be the same (and have the same orientation) if there is an even permutation transforming one to the other (it takes an even number of swaps to convert one to the other). If σ_1 is an even permutation, then

$$[v_1, \dots, v_n] = [\sigma_1(v_1), \dots, \sigma_1(v_n)].$$

If σ is an odd permutation, we say that

$$-[v_1, \dots, v_n] = [\sigma_2(v_1), \dots, \sigma_2(v_n)].$$

Example 2.10. Given a 1-simplex $[v_1, v_2]$, we have $[v_2, v_1] = -[v_1, v_2]$ because it takes an odd number of swaps (1) to transform one to the other. On the other hand, given a 2-simplex $[v_1, v_2, v_3]$, we have $[v_3, v_1, v_2] = [v_1, v_2, v_3]$, because it takes an even number of swaps to convert one to the other (2: swap v_3 and v_2 then v_1 and v_3). Additionally, $[v_1, v_0, v_2] = -[v_0, v_1, v_2]$. The negative of a simplex can be interpreted as flipping its orientation. See Figure 5.

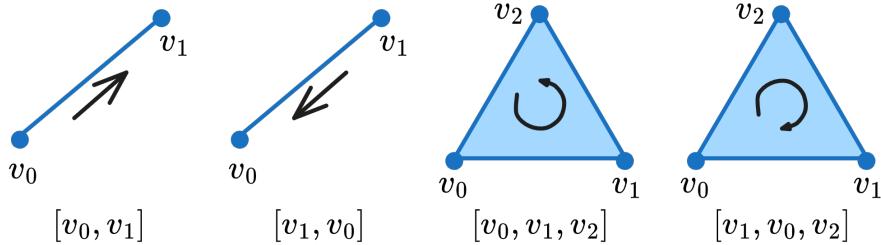


Figure 5: A figure showing the simplices $[v_0, v_1]$, $-[v_0, v_1]$, $[v_0, v_1, v_2]$, $-[v_0, v_1, v_2]$

Definition 2.11 (Space of n -chains). Given a finite simplicial complex S , the space of n -chains on S , $C_n(S)$, is the set of finite formal sums of n -simplices of S . More precisely, if $S = \{\sigma_1, \dots, \sigma_m\}$, then the elements of $C_n(S)$ can be written as formal sums of the form

$$\sum_{i=1}^m c_i \sigma_i \quad \text{where } c_i \in R$$

for some ring R . Elements of S are called n -chains.

Example 2.12. Note that Definition 2.11 allows any ring for coefficients. We pick \mathbb{Z} for simplicity. Note that coefficients of $-1, 0, 1$ have the most direct interpretation, so we focus on those. A coefficient of 0 signals the lack of that simplex, a coefficient of 1 indicates the presence of that simplex,

and a coefficient of -1 indicates the presence of that simplex with a flipped orientation. See Figure 6 to see how $[v_0, v_1, v_2] + [v_0, v_2, v_3]$ represents a clockwise oriented rectangle with vertices v_0, v_1, v_2 , and v_3 .

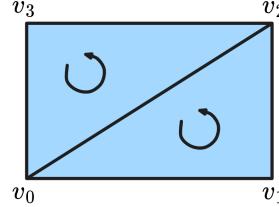


Figure 6: A representation of a clockwise oriented rectangle via the 2-chain $[v_0, v_1, v_2] + [v_0, v_2, v_3]$.

Now that we can represent shapes, we need to find a way to represent their holes. In homological algebraic topology, this is done with a boundary map. Observe that the boundary of a 2-simplex is three 1-simplices. In general, the boundary of an $n+1$ -simplex is $n+1$ n -simplices. We formally define the boundary map and give an example using 2-chains.

Definition 2.13. Given a finite simplicial complex S and $n \in \mathbb{N}$ we define the boundary map $\partial_n : C_n(S) \rightarrow C_{n-1}(S)$ as

$$\partial_n([v_0, v_1, \dots, v_n]) = \sum_{i=1}^n (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_n]$$

where \hat{v}_i indicates that v_i is excluded. Then extend ∂_n naturally onto finite formal sums of n -simplices (n -chains) to be a linear map.

Example 2.14. We compute $\partial_2([v_0, v_1, v_2])$. See that

$$\partial_2([v_0, v_1, v_2]) = [v_1, v_2] - [v_0, v_2] + [v_0, v_1] = [v_0, v_1] + [v_1, v_2] + [v_2, v_0].$$

Referring to Figure 7, we see how the boundary of the clockwise-oriented simplex $[v_0, v_1, v_2]$ is made up of the line segments running around its edges clockwise.

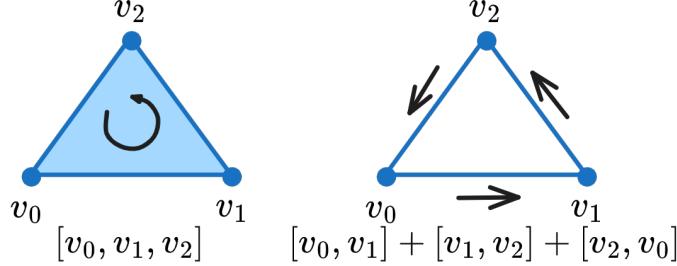


Figure 7: A depiction of $[v_0, v_1, v_2]$ and its border $\partial_2([v_0, v_1, v_2])$.

Example 2.15. We can compute the boundary of more complicated shapes, such as a rectangle, using linearity. Consider the clockwise oriented rectangle with vertices v_0, v_1, v_2, v_3 : $[v_0, v_1, v_2] + [v_0, v_2, v_3]$. Then

$$\begin{aligned}\partial_2([v_0, v_1, v_2] + [v_0, v_2, v_3]) &= \partial_2([v_0, v_1, v_2]) + \partial_2([v_0, v_2, v_3]) \\ &= ([v_1, v_2] - [v_0, v_2] + [v_0, v_1]) \\ &\quad + ([v_2, v_3] - [v_0, v_3] + [v_0, v_2]) \\ &= [v_0, v_1] + [v_1, v_2] + [v_2, v_3] + [v_2, v_0].\end{aligned}$$

Once again, we see that the boundary of the filled rectangle $[v_0, v_1, v_2] + [v_0, v_2, v_3]$ is the sum of the 1-simplices running clockwise around its edge. See Figure 8.

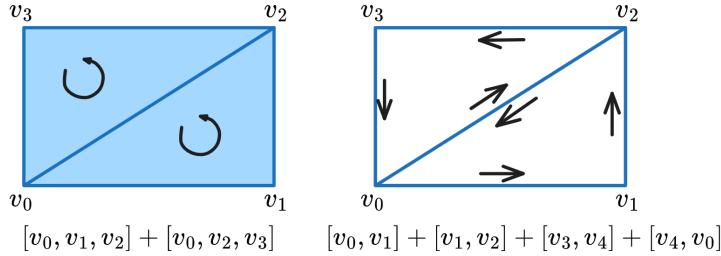


Figure 8: A depiction of $[v_0, v_1, v_2] + [v_0, v_2, v_3]$ and its boundary $\partial_2([v_0, v_1, v_2] + [v_0, v_2, v_3])$.

Now that we can represent a shape with an n -chain and compute its boundary with the boundary map ∂_n , we will look at how to identify holes.

Example 2.16. We shall look and see that line segments represented by 1-chains have a boundary consisting of the final vertex minus the starting vertex. Consider the line segment in Figure 9. See that

$$\begin{aligned}\partial([v_0, v_1] + [v_1, v_2] + [v_2, v_3] + [v_3, v_4]) \\ = (v_1 - v_0) + (v_2 - v_1) + (v_3 - v_2) + (v_4 - v_3) = v_4 - v_0.\end{aligned}$$

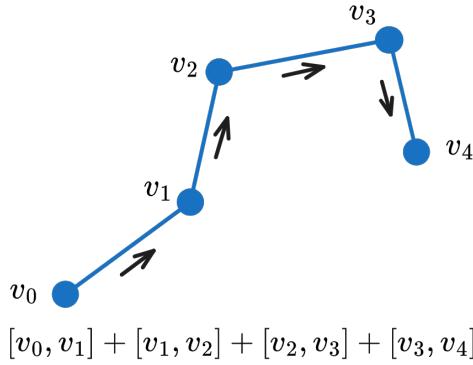


Figure 9: A visual depiction of the 1-chain $[v_0, v_1] + [v_1, v_2] + [v_2, v_3] + [v_3, v_4]$ representing a line segment.

Example 2.16 indicates that the boundary of a line segment with the same start and end point will be zero.

Example 2.17. Consider the simplicial complexes

$$\begin{aligned}\mathcal{A} &= \{\{v_0, v_1, v_2\}, \{v_0, v_1\}, \{v_1, v_2\}, \{v_2, v_3\}, \{v_0, v_3\}, \{v_0\}, \{v_1\}, \{v_2\}, \{v_3\}\} \\ \mathcal{B} &= \{\{v_0, v_1\}, \{v_1, v_2\}, \{v_2, v_3\}, \{v_0, v_3\}, \{v_0\}, \{v_1\}, \{v_2\}, \{v_3\}\}\end{aligned}$$

with spaces of n -chains C_0, C_1, C_2, \dots and C'_0, C'_1, C'_2, \dots , respectively. Additionally, let $\partial_n : C_n \rightarrow C_{n-1}$ and $\partial'_n : C'_n \rightarrow C'_{n-1}$ be the boundary maps for the corresponding chain groups for each $n \in \mathbb{N}$.

Recall once again Figure 7. We see that $\partial_2([v_0, v_1, v_2]) = [v_0, v_1] + [v_1, v_2] + [v_2, v_0]$. Observe that $[v_0, v_1, v_2] \in \mathcal{A}$ and so $\partial_2([v_0, v_1, v_2]) = [v_0, v_1] + [v_1, v_2] + [v_2, v_0] \in \text{Im } \partial_2 = \{\partial_2(x) : x \in C_2\}$. Then we see that the 1-D loop $[v_0, v_1] + [v_1, v_2] + [v_2, v_0]$ is the boundary of a higher-dimensional object within the chain C_2 coming from the simplicial complex \mathcal{A} . Visually, this corresponds to the inside of the loop being filled in.

On the other hand, observe that $\partial_1([v_0, v_1] + [v_1, v_2] + [v_2, v_0]) = 0$. This is because the boundary of a line segment is the final vertex minus the first, recall Example 2.16. Because this segment is a loop, we get zero. This may be interpreted as n -chains with zero boundary being closed objects. Then 1-chains with zero bound are closed loops. We will think of these closed loops as potential 1-D holes. Observe that $[v_0, v_1] + [v_1, v_2] + [v_2, v_0] \in \ker \partial_1 = \{x \in C_1 : \partial_1(x) = 0\}$. Note that $[v_0, v_1] + [v_1, v_2] + [v_2, v_0] \in C_1$ and $[v_0, v_1] + [v_1, v_2] + [v_2, v_0] \in C'_1$.

Now observe that $[v_0, v_1, v_2] \notin B$. This corresponds to the the loop $[v_0, v_1] + [v_1, v_2] + [v_2, v_0]$ not being filled in. Then $[v_0, v_1] + [v_1, v_2] + [v_2, v_0]$ may be interpreted as a loop or a hole relative to B because it is not the boundary of a higher-dimensional object, and it closes on itself. On the other hand notice that $[v_0, v_1, v_2] \in \mathcal{A}$ and so $[v_0, v_1] + [v_1, v_2] + [v_2, v_0]$ is the boundary of a higher dimensional object with respect to \mathcal{A} . From this example,e we build the intuition that “1-holes” correspond to 1-chains that are loops (or sums of loops; $\in \ker \partial_1$) that are not the boundary of 2-chains ($\notin \text{Im } \partial_2$). Then “ n -holes” will be considered elements of $\ker \partial_n$ that are not in $\text{Im } \partial_{n+1}$.

Lemma 2.18. *Given a finite simplicial complex S with chain groups $C_0(S)$, $C_1(S)$, $C_2(S)$, ... it is the case that $\partial_n \circ \partial_{n-1} = 0$ for all $n \in \mathbb{N}$.*

We could show that $\ker \partial_n$ is a subgroup of C_n . Using Lemma 2.18 we can show that $\text{Im } \partial_{n+1}$ is a normal subgroup of $\ker \partial_n$. This allows us to define a quotient group. This effectively takes all the “closed shapes” in our simplicial complex and identifies the “closed” shapes coming from the boundary of a higher-dimensional shape with zero. In the case of 1-chains, we consider the group of all finite formal sums of loops and identify loops that are the boundary of a 2-dimensional object with zero.

Definition 2.19. Given a simplicial complex S with chain groups $C_0(S)$, $C_1(S)$, ..., the n -th simplicial homology group is the quotient group

$$H_n(S) = \ker \partial_n / \text{Im } \partial_{n+1},$$

where $\ker \partial_n : C_n \rightarrow C_{n-1}$ and $\text{Im } \partial_{n+1} : C_{n+1} \rightarrow C_n$.

The n th homology group encodes topological information in a group structure, and in the case of H_1 , the generators of this group represent the independent 1-dimensional “holes” or loops in the topological space X .

Definition 2.20 (Betti Number). Given a finite simplicial complex S , the space of n -chains is finitely generated. Then the n -th simplicial homology group can be written as a direct sum of copies of \mathbb{Z} and a torsion part T :

$$H_n(S) = \mathbb{Z}^k \oplus T.$$

Then k is the n -th Betti number.

The Betti number can be used as a rough numerical summary of $H_n(S)$, counting the number of “ n -dimensional holes” in our space.

2.2 Persistent Homology

Now that we have given intuition as to how homology counts holes, we will give a brief overview of persistent homology.

Definition 2.21 (Metric Space). A metric space is a pair (M, d) where M is a set and d is a distance function $d : M \times M \rightarrow \mathbb{R}$ satisfying the following properties:

1. $d(x, y) \geq 0$ for all $(x, y) \in M \times M$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$ for all $(x, y) \in M \times M$
4. $d(x, y) \leq d(x, z) + d(z, y)$ for all $(x, y, z) \in M \times M \times M$

Recall that we want to use homology on real-world data, but every metric space with a finite number of points has trivial homology groups. This is a problem because every point cloud of data from the physical world is finite.

We must find a way to inflate our data so that we may calculate its homology groups. One solution to this problem is quite intuitive.

Definition 2.22 (Rips Complex). Let (P, d) be a metric space where P is a point set. Given a real number $r > 0$, the Vietoris-Rips or Rips complex on P is the simplicial complex $\mathcal{R}^r(P)$ such that a simplex σ is in $\mathcal{R}^r(P)$ if and only if $d(p, q) \leq r$ for every pair of vertices of σ .

In essence, we pick a distance r and connect points that are all within distance r . These correspond to the creation of n -simplices of different dimensions depending on how close the points are.

Example 2.23. In Figure 10 we see the Vietoris Rips Complex for various values of r . When r is small, no points are close enough to be connected. As r increases, pairs of points that are within distance r are connected to form 1-simplices, triples of points form 2-simplices, and so on.

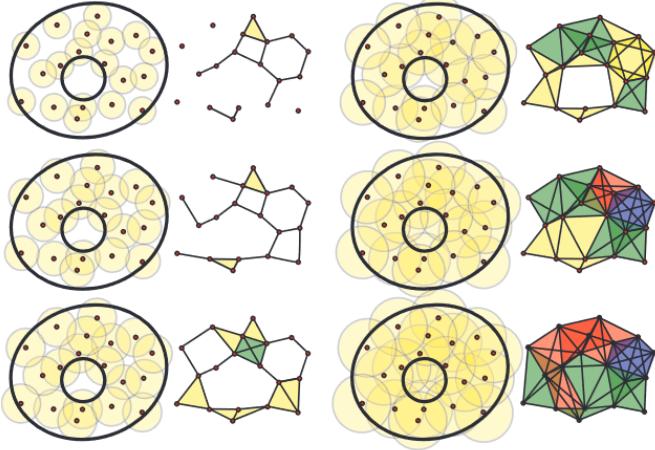


Figure 10: A Rips complex shown for different values of r .

Now we have seen that the Rips complex changes for different values of r . The question naturally arises: what value of r should we use? Note that when r is smaller than the minimum distance between any two points (in our finite collection of data points), the Rips complex is a collection of disjoint points. For large enough r all points will be within distance r of each other, and so the Rips complex is just one high-dimensional simplex. Instead of finding a single value of r to use, we can let r grow from near zero until the entire point cloud is merged into a single simplex. We can track the Betti numbers as r grows to get an idea of how different dimensional holes appear and disappear in the Rips complex over r .

From a theoretical perspective, we should note that there is another simplicial complex that can be constructed from data called the Čech complex. It is easier to show that the Čech complex is a good topological representation of a point cloud versus the Rips Complex. The fact the Čech complex can approximate a topological space well is called the Nerve Theorem and gives credence to studying data through homology groups. An additional proposition shows that the Rips Complex can approximate the Čech Complex well. Because it is easier to work with the Rips complex, it is the complex we shall consider.

We now introduce the idea of a filtration.

Definition 2.24 (Filtration). A filtration of a simplicial complex, K , is a nested sequence of sub-complexes starting at the empty set and ending with the full simplicial complex, i.e.,

$$\emptyset \subseteq K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K$$

Example 2.25. The following example is taken from [8]. See Figure 11 which shows a filtration of the point cloud $S = \{(0, 0), (1, 3), (2, -1), (3, 2)\} \subseteq \mathbb{R}^2$.

1. When $r < \sqrt{5}$ none of the balls of radius $r/2$ intersect and so $\mathcal{R}^r(S)$ is four points.
2. When $r = \sqrt{5}$, the balls of radius $r/2$ centered at $(0, 0)$ and $(2, -1)$ intersect which means we add a 1-simplex between $(0, 0)$ and $(2, -1)$. Similarly, we add a 1-simplex between $(1, 3)$ and $(3, 2)$.
3. When $r \in (\sqrt{5}, \sqrt{10})$, no additional balls of radius $r/2$ intersect so $\mathcal{R}^r(S) = \mathcal{R}^{\sqrt{5}}(S)$.
4. When $r = \sqrt{10}$, we add a 2-simple between $(0, 0)$, $(1, 3)$ and $(2, -1)$, $(3, 2)$.
5. When $r = \sqrt{13}$, we add two 2-simplices between $(0, 0)$, $(1, 3)$, $(2, -1)$ and $(1, 3)$, $(2, -1)$, $(3, 2)$.
6. When $r \in (\sqrt{13}, \sqrt{17})$ we have that $\mathcal{R}^r(S) = \mathcal{R}^{\sqrt{13}}(S)$.
7. When $r = \sqrt{17}$ we add a 3-simplex.
8. When $r > \sqrt{17}$, $\mathcal{R}^r(S) = \mathcal{R}^{\sqrt{17}}(S)$.

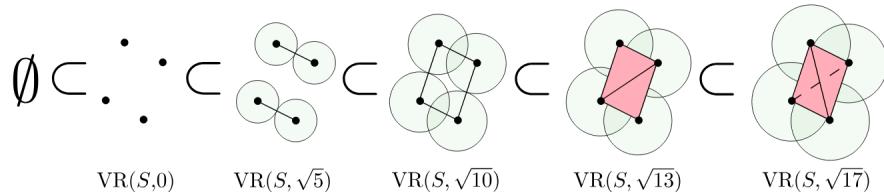


Figure 11: A visual representation of the filtration $S = \{(0, 0), (1, 3), (2, -1), (3, 2)\} \subseteq \mathbb{R}^2$ [8].

Given a filtration, we can see that simplices appear at different values of r . This inspires the following definitions.

Definition 2.26 (Birth). For a filtered complex K and subcomplexes K_{i-1} , K_i , a topological feature $x \in H_p(K_i)$ is born at i if $x \notin H_p(K_{i-1})$.

Definition 2.27 (Death). For a filtered complex K and subcomplexes K_{i-1} , K_i , a topological feature $x \in H_p(K_{i-1})$ dies at i if $x \notin H_p(K_i)$. A feature will also die if the feature merges with a feature born earlier in the filtration.

Definition 2.28 (Persistence Interval). For a given topological feature x with birth point i and death point j , the persistence interval for the feature is given by $[i, j]$. If $j = \infty$, the component does not die during the filtration.

Definition 2.29 (Persistence Barcode). A persistence barcode is a graphical representation of the persistence intervals as a collection of horizontal line segments where the horizontal axis corresponds to the parameter r and whose vertical axis represents an ordering of topological features (homology generators).

Definition 2.30 (persistence diagram). A persistence diagram is a graphical representation of the persistence intervals as a collection of points in the first quadrant of the Cartesian plane. Birth time is plotted along the x -axis while death time is plotted along the y -axis. Each topological feature is plotted at its (birth, death) point. A line is plotted along the diagonal as all points must lie above it. A far distance from the diagonal indicates that a topological feature persisted for a long time.

Example 2.31. Let $S = \{(0, 0), (1, 3), (2, -1), (3, 2)\} \subseteq \mathbb{R}^2$ from Example 2.25. In Figure 12 we plot its corresponding persistence barcode. In Figure 13, we plot the corresponding persistence diagram.

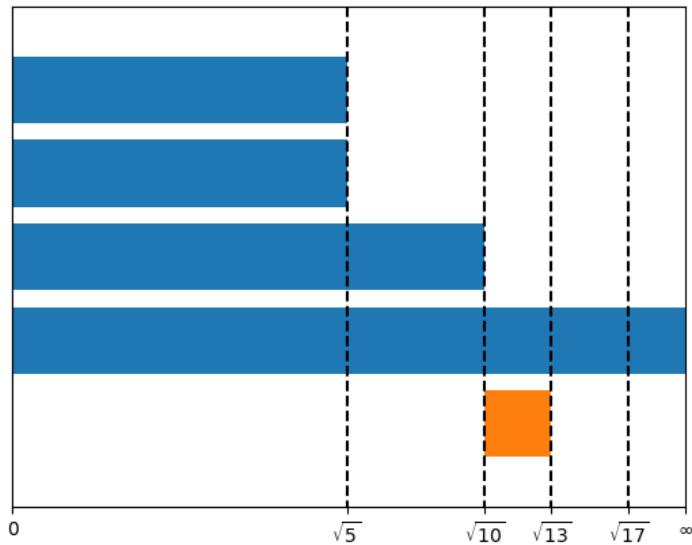


Figure 12: A persistence barcode corresponding to $S = \{(0, 0), (1, 3)(2, -1), (3, 2)\} \subseteq \mathbb{R}^2$. The blue bars correspond to topological features in H_0 , while the orange bars correspond to topological features in H_1 .

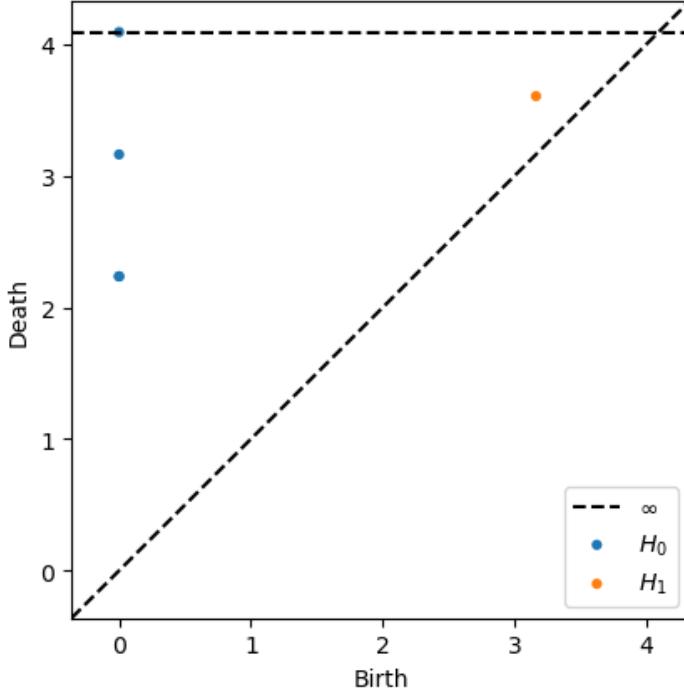


Figure 13: A persistence diagram corresponding to $S = \{(0, 0), (1, 3)(2, -1), (3, 2)\} \subseteq \mathbb{R}^2$.

The barcode can be thought of as tracking Betti numbers as r changes. This gives a rough numerical summary of when topological features (holes) appear and disappear in each homology group.

2.3 Application of 1-Dimensional Persistent Homology to the Periodicity of Time Series

Sliding windows and 1-persistence scoring (SW1PerS) aims to use homology to assign a periodicity score to a time series [7]. The general idea follows: given a time series, use splining to construct a curve, choose a window size and a number of points $m + 1$ to sample in each window. This number $m + 1$ is called the embedding dimension. This is because we sample each window $m + 1$ times and embed the sample in \mathbb{R}^{m+1} . We then choose some number of windows spread across the domain of f . This constructs a point cloud, on which we may perform 1-persistent homology. It will turn out that loops within the point cloud correspond to periodic functions. By finding

the largest persistence of a 1-D topological feature in H_1 we can assign a periodicity score.

Definition 2.32. Let M be a positive integer and τ a positive real number with $f : [a, b] \rightarrow \mathbb{R}$. The sliding-window embedding of f based at $t \in \mathbb{R}^{m+1}$ is the point

$$SW_{m,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t + \tau) \\ \vdots \\ f(t + M\tau) \end{bmatrix} \in \mathbb{R}^{M+1}.$$

By allowing t to take on different values, we can create a sliding-window point cloud for f . $M\tau$ is the window size of this sliding-window embedding.

In order to create a sliding-window point cloud we may sample points from $[a, b - M\tau]$ for values of t . One inspiration for the sliding-window point cloud is Taken's Theorem. It implies that a smooth attractor can be topologically reconstructed by the embedding of sliding windows in \mathbb{R}^{M+1} .

Example 2.33. We consider the example $f : [0, 2\pi] \rightarrow \mathbb{R}$ where $f(x) = \cos(Lx)$. We choose $M = 1$ and let τ vary to see its effect in Figure 14.

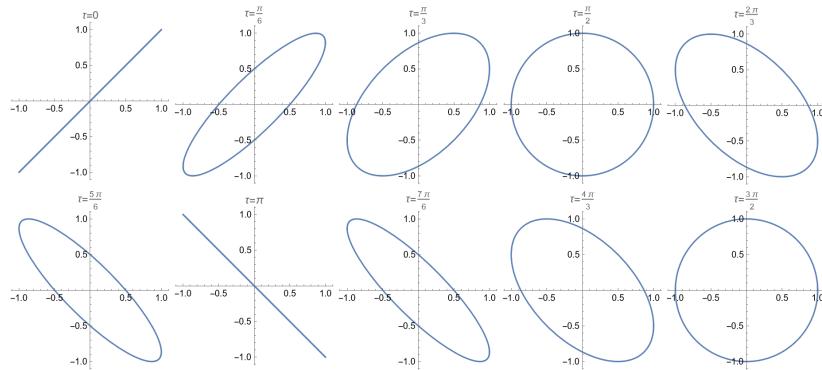


Figure 14: We plot $(\cos(t), \cos(t + \tau))$ for $t, \tau \in [0, 2\pi]$.

From Figure 14 we can see that $SW_{M,\tau}f(t)$ generally traces out a loop. For a fixed τ , we hope to sample t from $[0, 2\pi - M\tau]$ and plug these values into $SW_{M,\tau}f(t)$, creating a point cloud of the corresponding circle. Then we can use 1-persistent homology to detect the presence of and size of these circles. However, observe that the choice of τ greatly affects the shape of $SW_{M,\tau}f(t)$. We see that a particular value of τ is the roundest, and so

it will be the easiest to identify with persistent homology. More precisely, the most round $SW_{M,\tau}f(t)$ in Figure 14 will have the longest persistence interval.

In Example 2.33, we arbitrarily chose our embedding dimension and varied τ . We need to know how to pick these values to best measure periodicity. The following example will suggest how to do so and provide motivation for the SW1PerS method.

Example 2.34. We add some details to the motivation for SW1PerS as outlined in [6]. Let $M + 1 \in \mathbb{N}$ be an arbitrary embedding dimension and $\tau \in \mathbb{R}$ an arbitrary step size. Using the sum of cosines formula,a we have that

$$\begin{aligned} SW_{M,\tau}f(t) &= \begin{bmatrix} \cos(Lt) \\ \cos(Lt + L\tau) \\ \vdots \\ \cos(Lt + LM\tau) \end{bmatrix} \\ SW_{M,\tau}f(t) &= \cos(Lt) \begin{bmatrix} 1 \\ \cos(L\tau) \\ \vdots \\ \cos(LM\tau) \end{bmatrix} - \sin(Lt) \begin{bmatrix} 1 \\ \sin(L\tau) \\ \vdots \\ \cos(LM\sin) \end{bmatrix} \\ &= \cos(Lt)u - \sin(Lt)v \end{aligned}$$

where $u_n = \cos(Ln\tau)$ and $v_n = \sin(Ln\tau)$ for $n = 0, 1, \dots, M$. Observe that we can write

$$SW_{M,\tau}f(t) = \begin{bmatrix} u_0 & -v_0 \\ u_1 & -v_1 \\ \vdots & \vdots \\ u_M & -v_M \end{bmatrix} \begin{bmatrix} \cos(Lt) \\ \sin(Lt) \end{bmatrix}.$$

We can determine how round the curve $SW_{M,\tau}f(t)$ is with an isometry. Let $\langle \cdot, \cdot \rangle$ be the standard Euclidean inner product. Let $x(t) = [\cos(Lt), \sin(Lt)]^T$. Observe that

$$\begin{aligned} \|SW_{M,\tau}f(t)\|^2 &= \langle SW_{M,\tau}f(t), SW_{M,\tau}f(t) \rangle \\ &= \langle \cos(Lt)u - \sin(Lt)v, \cos(Lt)u - \sin(Lt)v \rangle \\ &= \langle \cos(Lt)u, \cos(Lt)u \rangle - \langle \cos(Lt)u, \sin(Lt)v \rangle \\ &\quad - \langle \sin(Lt)v, \cos(Lt)u \rangle + \langle \sin(Lt)v, \sin(Lt)v \rangle \\ &= \cos^2(Lt)\|u\|^2 - 2\cos(Lt)\sin(Lt)\langle u, v \rangle + \sin^2(Lt)\|v\|^2 \end{aligned}$$

Additionally, let

$$A = \begin{bmatrix} \|u\|^2 & -\langle u, v \rangle \\ -\langle u, v \rangle & \|v\|^2 \end{bmatrix}$$

Note that this is actually the Gram matrix for the vectors u, v with respect to $\langle \cdot, \cdot \rangle$. It follows that

$$\begin{aligned} \langle x(t), Ax(t) \rangle &= \left\langle \begin{bmatrix} \cos(Lt) \\ \sin(Lt) \end{bmatrix}, \begin{bmatrix} \|u\|^2 & -\langle u, v \rangle \\ -\langle u, v \rangle & \|v\|^2 \end{bmatrix} \begin{bmatrix} \cos(Lt) \\ \sin(Lt) \end{bmatrix} \right\rangle \\ &= \left\langle \begin{bmatrix} \cos(Lt) \\ \sin(Lt) \end{bmatrix}, \begin{bmatrix} \cos(Lt)\|u\|^2 - \sin(Lt)\langle u, v \rangle \\ -\cos(Lt)\langle u, v \rangle + \sin(Lt)\|v\|^2 \end{bmatrix} \right\rangle \\ &= \cos^2(Lt)\|u\|^2 - 2\cos(Lt)\sin(Lt)\langle u, v \rangle + \sin^2(Lt)\|v\|^2. \end{aligned}$$

It is well known that Gram matrices are hermitian and positive semi-definite, then A has real eigenvalues. From the spectral theorem, there exists an orthogonal matrix $B \in M_2(\mathbb{R})$ and diagonal matrix $\Lambda = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \in M_2(\mathbb{R})$ so that $A = B\Lambda^2B^T$. It follows that

$$\langle x(t), Ax(t) \rangle = \langle x(t), B^T\Lambda^2Bx(t) \rangle = \langle \Lambda Bx(t), \Lambda Bx(t) \rangle = \|\Lambda Bx(t)\|^2.$$

Combining the past two observations, we get that $SW_{M,\tau}f(t) \mapsto \Lambda Bx(t)$ is an isometry. We can now use the roundness of $\Lambda Bx(t)$ to determine the roundness of $SW_{M,\tau}f(t)$. It is not difficult to show that a 2×2 orthogonal matrix is a scalar times rotation. Without loss of generality, we may assume B to be a rotation. This is because B is orthogonal, and we can factor out any scalar in the expression $A = B\Lambda^2B^T$ to get a new matrix B' that is just a rotation. Then suppose that B is a rotation by $\alpha \in \mathbb{R}$. We now have a nice expression for $\Lambda Bx(t)$.

$$\begin{aligned} \Lambda Bx(t) &= \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \cos(Lt) \\ \sin(Lt) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos(Lt)\cos(\alpha) - \sin(Lt)\sin(\alpha) \\ \sin(Lt)\cos(\alpha) + \cos(Lt)\sin(\alpha) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos(Lt + \alpha) \\ \sin(Lt + \alpha) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\lambda_1} \cos(Lt + \alpha) \\ \sqrt{\lambda_2} \sin(Lt + \alpha) \end{bmatrix}. \end{aligned}$$

Then $\Lambda Bx(t)$ traces out an ellipse with minor axes of length $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$. We will now find the eigenvalues λ_1 and λ_2 . For a 2×2 matrix, its eigenvalues can be expressed in terms of the trace and determinant. That is

$$\lambda = \frac{1}{2} \left(\text{Tr}(A) \pm \sqrt{\text{Tr}(A)^2 - \det(A)} \right).$$

Note that the $\text{Tr}(A) = \|u\|^2 + \|v\|^2$ and $\det A = \|u\|^2\|v\|^2 - \langle u, v \rangle^2$. We can use Lagrange's trigonometric formula to compute:

$$\langle u, v \rangle = \sum_{n=0}^M \cos(Ln\tau) \sin(Ln\tau) = \frac{1}{2} \sum_{n=0}^M \sin(2Ln\tau) = \frac{\sin(L(M+1)\tau) \sin(LM\tau)}{2 \sin(L\tau)}.$$

We end up with eigenvalues:

$$\begin{aligned} \lambda_1 &= \frac{(M+1) + \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right|}{2} \\ \lambda_2 &= \frac{(M+1) - \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right|}{2}. \end{aligned}$$

Our curve $\Lambda Bx(t)$ will be the round when the minor axis are equal. This occurs when

$$\begin{aligned} \frac{(M+1) + \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right|}{2} &= \frac{(M+1) - \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right|}{2} \\ \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right| &= - \left| \frac{\sin(L(M+1)\tau)}{\sin(L\tau)} \right| \\ \sin(L(M+1)\tau) &= 0 \\ L(M+1)\tau &= \pi k, \quad \text{for } k \in \mathbb{Z}. \end{aligned}$$

Now suppose that

$$M\tau = \left(\frac{M}{M+1} \right) \frac{2\pi}{L}.$$

It follows that

$$L(M+1)\tau = 2\pi.$$

Observe that when M is large, $M\tau$ is almost the period of $f(t) = \cos(Lt)$, which is $2\pi/L$. This shows that when the window size is near to the period of $f(t)$, $\Lambda Bx(t)$ is roundest and so $SW_{M,\tau}f(t)$ is roundest due to the isometry between them.

From Example 2.34 we build the idea that if $f(t)$ is a periodic function so that $f(t+w) = f(t)$ for all t , then $SW_{M,\tau}f(t)$ is roundest when $M\tau$, the window size, is close to the period of $f(t)$, which is w . We want to pick τ so that this is the case. Then we can measure this roundness with the largest persistence interval obtained from a rips complex filtration over a sampling of the curve $SW_{M,\tau}f(t)$: $\{SW_{M,\tau}f(t_0), SW_{M,\tau}f(t_1), \dots, SW_{M,\tau}f(t_m)\}$ where $m \in \mathbb{N}$.

2.4 The SW1PerS Algorithm

We will now describe SW1PerS as developed in [7].

Algorithm 1 SW1PERS $((f_1, \dots, f_S), (t_1, \dots, t_S), M, n, L, p)$

Input: A time series with observations f_1, \dots, f_S corresponding to time values $0 < t_1 < t_2 < \dots < t_S$. $M + 1$ is the embedding dimension chosen so that $M + 1 \geq \frac{S}{L}$. n is the number of evenly spaced points sampled from $t_S - M\tau$. $\tau = \frac{t_S - t_1}{(M+1)L}$ is chosen to maximize the roundness $SW_{M,\tau}f(t)$ when f is L -periodic. L is the expected periodicity. A function is L -periodic if $f(t + \frac{t_S - t_1}{L}) = f(t)$ for all $t \in [t_1, t_S - M\tau]$. p is the prime p for which the coefficients \mathbb{Z}_p are used to compute 1-persistent homology of $SW_{M,\tau}f(t)$. p should be chosen so that $p > \frac{M}{L}$.

Output: The L -periodicity score, which is in $[0, 1]$. Note that 0 is the most L -periodic while 1 is the least L -periodic.

Definitions: Let $\tau = \frac{t_S - t_1}{(M+1)L}$.

- 1: $\tau \leftarrow \frac{t_S - t_1}{(M+1)L}$
 - 2: $f(t) \leftarrow$ A cubic splining of (f_1, \dots, f_S) and (t_1, \dots, t_S) .
 - 3: $X \leftarrow SW_{M,\tau}f\left((0, \frac{t_S - m\tau}{n}, 2\frac{t_S - m\tau}{n}, \dots, t_S - M\tau)\right)$
 - 4: $\bar{X} \leftarrow \frac{X - \text{mean}(X)}{\|X - \text{mean}(X)\|}$.
 - 5: $\text{mp}(\bar{X}) \leftarrow$ the maximum persistence length of a 1-homological feature computed from the 1-persistent homology of \bar{X} using coefficients in \mathbb{Z}_p .
 - 6: **return** $1 - \frac{\text{mp}(\bar{X})}{\sqrt{3}}$
-

In Algorithm 1 we ensure that $M + 1 \geq S/L$ so that topological information is preserved [7]. Additionally, we choose $p > M/2$ so that the coefficients in \mathbb{Z}_p do cause the homology groups to miss desired periods [6]. Note that in the calculation of \bar{X} , the re-centering by the mean and normalization are applied point-wise to each window within the point cloud.

We used standard packages such as numpy [3] for numerical computations and Ripser [2] for the computation of 1-persistent homology to implement SW1PerS in Python. In Examples 2.35, 2.36, 2.37, and 2.38 we sample 50 evenly space points from the domain of each function. Then we compute the SW1PerS algorithm and plot periodicity versus window size. The parameters $M = 50$, $n = 200$, $L = \frac{t_S - t_1}{(M+1)\tau}$, and $p = 53$ are used. Recall that the window size is $w = M\tau$.

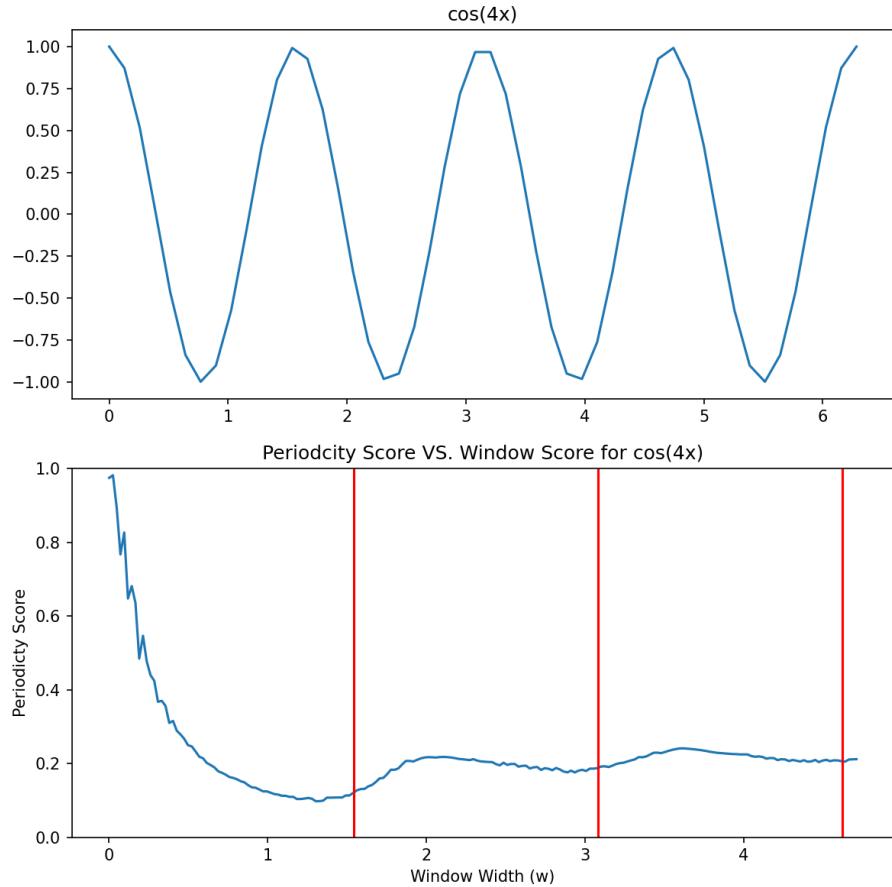


Figure 15: (Top) $\cos(4x)$ is graph over $[0, 2\pi]$. (Bottom) A graph of SW1PerS periodicity versus window size for $\cos(4x)$. Three red vertical lines are plotted on the same graph at $w = (M/(M+1))(2\pi/4)$, $(M/(M+1))2(2\pi/4)$, $(M/(M+1))3(2\pi/4)$.

Example 2.35. From Example 2.33, we expect the periodicity score to be

minimized when $M\tau$ is about $(2\pi/4)$, $2(2\pi/4)$, $3(2\pi/4)$. Referring to Figure 15, we see that the periodicity does reach local minimum when $M\tau$ is near $(2\pi/4)$, $2(2\pi/4)$, $3(2\pi/4)$. As noted in [9], see that the periodicity score drops low once the window size is at least half the period of the function. We should note that the periodicity score slowly rises as the window size increases to larger multiples of the expected period

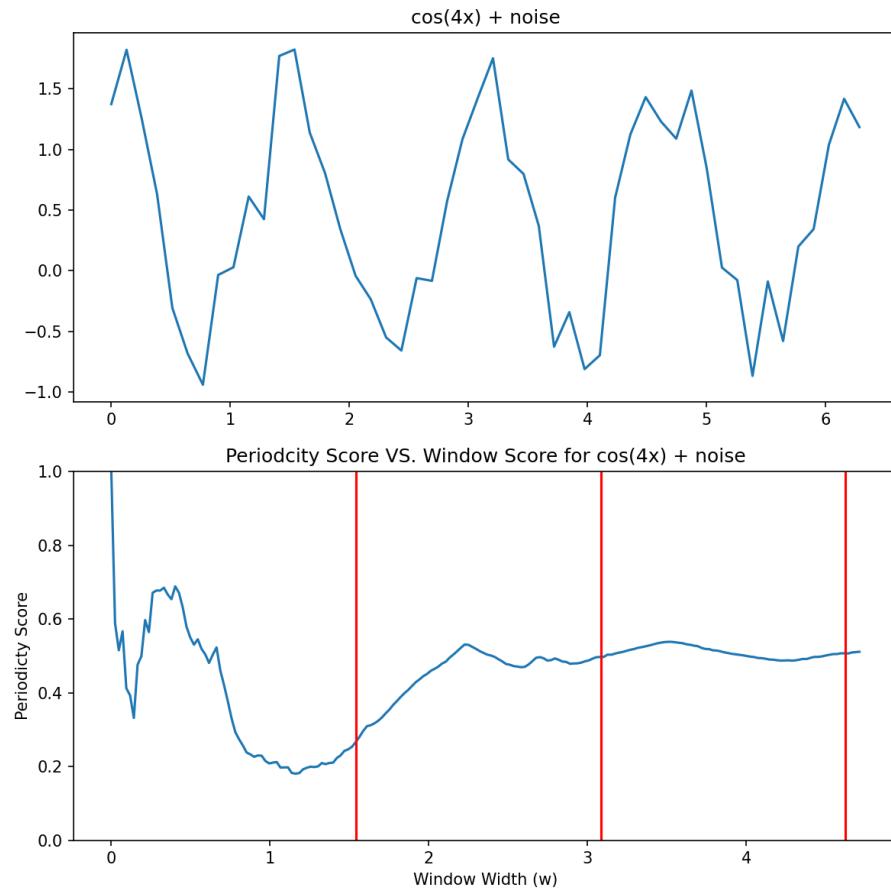


Figure 16: (Top) $\cos(4x) + \text{noise}$ is graph over $[0, 2\pi]$. (Bottom) A graph of SW1PerS periodicity versus window size for $\cos(4x)$ with random noise from $[0, 1]$. Three red vertical lines are plotted on the same graph at $w = (M/(M+1))(2\pi/4)$, $(M/(M+1))2(2\pi/4)$, $(M/(M+1))3(2\pi/4)$.

Example 2.36. Due to the noise added to $\cos(x)$, we should not be surprised that the local minima have seemed to stray from $w = (2\pi/4)$,

$2(2\pi/4)$, $3(2\pi/4)$ when comparing to Example 2.35. Despite the noise, SW1PerS is able to recognize that the curve is periodic when the window size is near the period. However, when the window size grows to multiples of the expected period, the periodicity score rises drastically.

One of the hopes of the SW1PerS algorithm is to identify periodicity stemming from non-trigonometric curves. We will now look at some examples of periodic curves that are not continuous.

Example 2.37. In Figure 17, we see that the periodicity score is minimized near window sizes $w = 1, 2, 3$. Then SW1PerS is able to detect periodicity in discontinuous periodic functions.

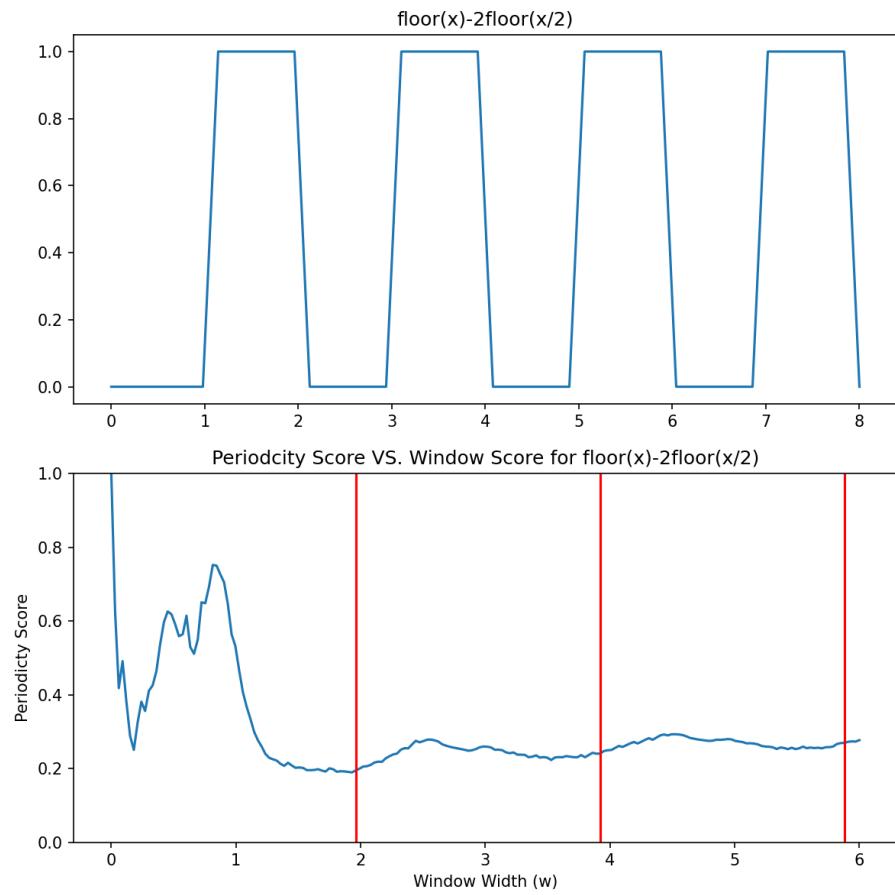


Figure 17: (Top) $\text{floor}(x) - 2\text{floor}(x/2)$ is plotted over $[0, 2\pi]$. (Bottom) A plot of SW1PerS periodicity versus window size for $\text{floor}(x) - 2\text{floor}(x/2)$. Three red vertical lines are plotted on the same graph at $w = (M/(M+1))2, (M/(M+1))4, (M/(M+1))6$.

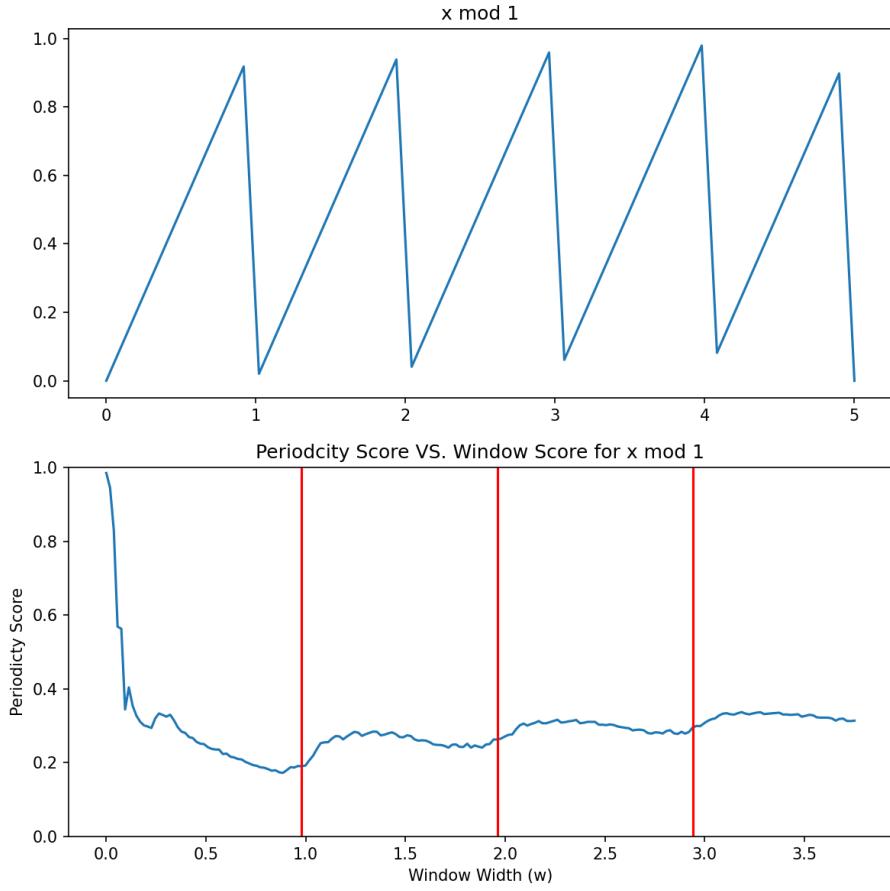


Figure 18: (Top) $x \bmod 1$ is plotted over $[0, 5]$. (Bottom) A plot of SW1PerS periodicity versus window size for $x \bmod 1$. Three red vertical lines are plotted on the same graph at $w = (M/(M+1))1, (M/(M+1))2, (M/(M+1))3$.

Example 2.38. In Figure 18, we see that the periodicity score is minimized near the window sizes $w = 1, 2, 3$, and SW1PerS is able to detect the periodicity within square waves.

Overall, Examples 2.35, 2.36, 2.37, and 2.38 show the importance of using the optimal window size.

3 Results and Discussion

We will now investigate where the northward surface ocean velocity is most periodic within a patch of ocean near Morro Bay, California, USA. Specifically, we will use the the Near-Real Time Surface Ocean Velocity, U.S. West Coast, 6km Resolution Hourly data collected by IOOS and lying within the region bounded by latitudes 34.5°N to 35.1°N and longitudes -122.4°W to -120.8°W [5]. From this point forward, this region will be called the study patch. The velocity is recorded at each point within the study patch every hour. We will examine the northward/southward velocity data from 2014 to 2024 on three different time scales.

3.1 Periodicity Over A Week

We begin by examining the northward velocity data from 2020-01-01 00:00:00 to 2021-01-01 00:00:00. The data is evenly split into 52 chunks, meaning that each chunk is approximately a week long. Figure 19 shows the periodicity score for each point within the study patch calculated over the time period 2020-01-01 00:00:00 to 2021-01-08 00:00:00. Additionally, the pointwise mean, standard deviation, lower quartile, median, and upper quartile are shown. SW1PerS is computed multiple times with parameters $M = 58$, $n = 200$, $L = 3, 4, 5, 15$, and $p = 61$, with the minimum periodicity score being taken for each point. From now on, such a figure will be called the periodicity score graph for the corresponding time segment. Note that the parameters are chosen to follow the optimal parameters specified in Algorithm 1. We choose $L = 3, 4, 5, 15$ to search for periodicity on the scale of days to half days.

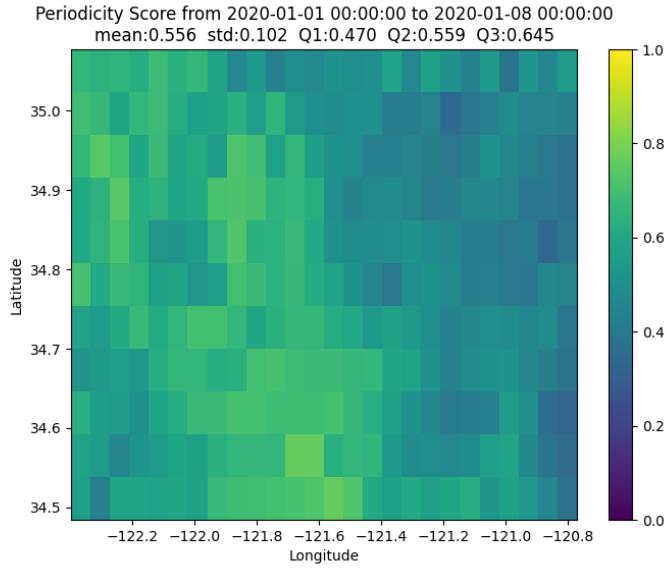


Figure 19: This is a graph of the minimum periodicity scores for each point in the study patch computed by SW1PerS with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

Referring to Figure 19, we can see that the mean periodicity score for 2020-01-01 to 2021-01-08 is 0.556. This is not very periodic. However, we can see on the right-hand side of the figure that there is lower periodicity. This is investigated further. In Figure 20 we graph the 20 most periodic velocity curves over the time period from 2020-01-01 to 2021-01-01.



Figure 20: A graph of the 20 most periodic velocity curves from 2020-01-01 to 2020-01-08. SW1PerS was computed with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

Referring to Figure 20, we can see that the most periodic curve had periodicity 0.328, which is significantly lower than the mean of 0.556. This indicates that different parts of the study patch exhibit different levels of periodicity. This can also be seen from the right half of Figure 19, which is bluer than the left. This indicates a lower periodicity score. The velocity curves in 20 appear to exhibit periodicity on the scale of days/weeks, but do not appear to oscillate rapidly each day.

In Figure 20, we graph the 20 least periodic northward velocity curves over the time period 2020-01-01 to 2021-01-08.



Figure 21: A graph of the 20 least periodic velocities from 2020-01-01 to 2020-01-08. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

The most notable feature of Figure 21 is that many of the least periodic curves are lacking data. As part of SW1PerS, this missing data was splined in and may be having a larger effect on the periodicity score than expected. Additionally, note that these least periodic velocity curves have a periodicity score near .7, which lies above the third quartile.

Having examined the periodicity graph for one time segment in the study patch, we would like to compare time segments. In Figure 22, we plot the periodicity score graph for each of the 52 time chunks. From now on, this type of figure will be called the matrix of periodicity score graphs

for a corresponding time scale.

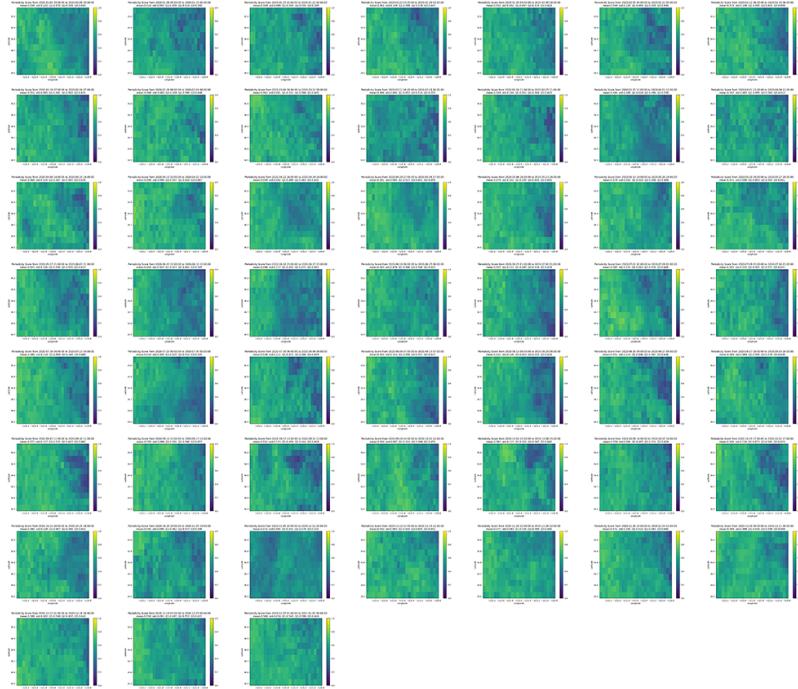


Figure 22: This is a matrix of periodicity score graphs for each of the 52 time segments from 2020-01-01 to 2021-01-08 (see Figure 19). SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

In Figure 22, we can see that the eastward side of the study patch exhibits more periodicity due to the more common appearance of blue hues. Note that the eastern side of the time patch is closer to shore. Previously, we remarked that the velocity curves with lower periodicity often had missing data. Additionally, because the western side of the time patch is farther from shore is more common for data to be missing. This is because the data is collected by radio towers on the shore. For this reason, the study patch may only exhibit more periodicity on the eastern side due to missing data.

Now we shall examine the most periodic curves found on a weekly time scale from 2020-01-01 to 2021-01-01 and compare their shapes (we chose five time periods that exhibit low periodicity scores and vary in shape, but not the 5 time segments with the five lowest periodicity scores). See Figures 23, 24, 25, 26, and 27.

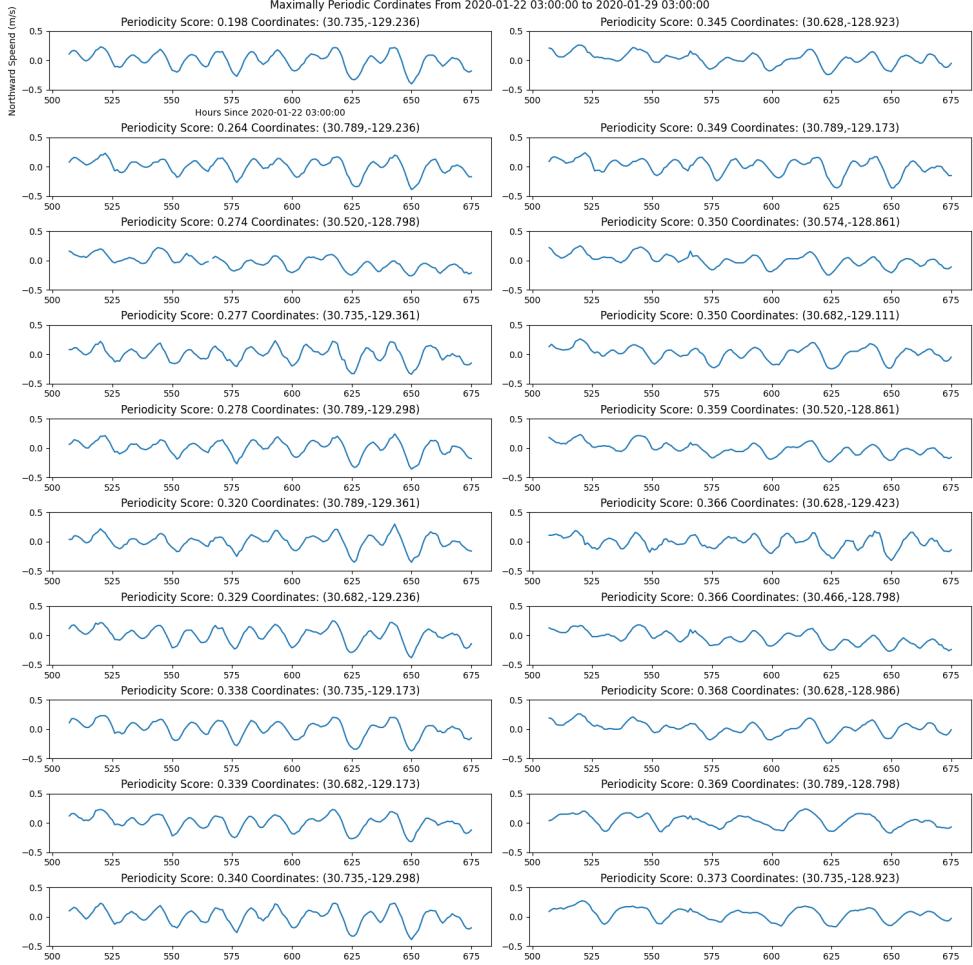


Figure 23: A graph of the 20 most periodic velocity curves from 2020-01-22 to 2020-01-29. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

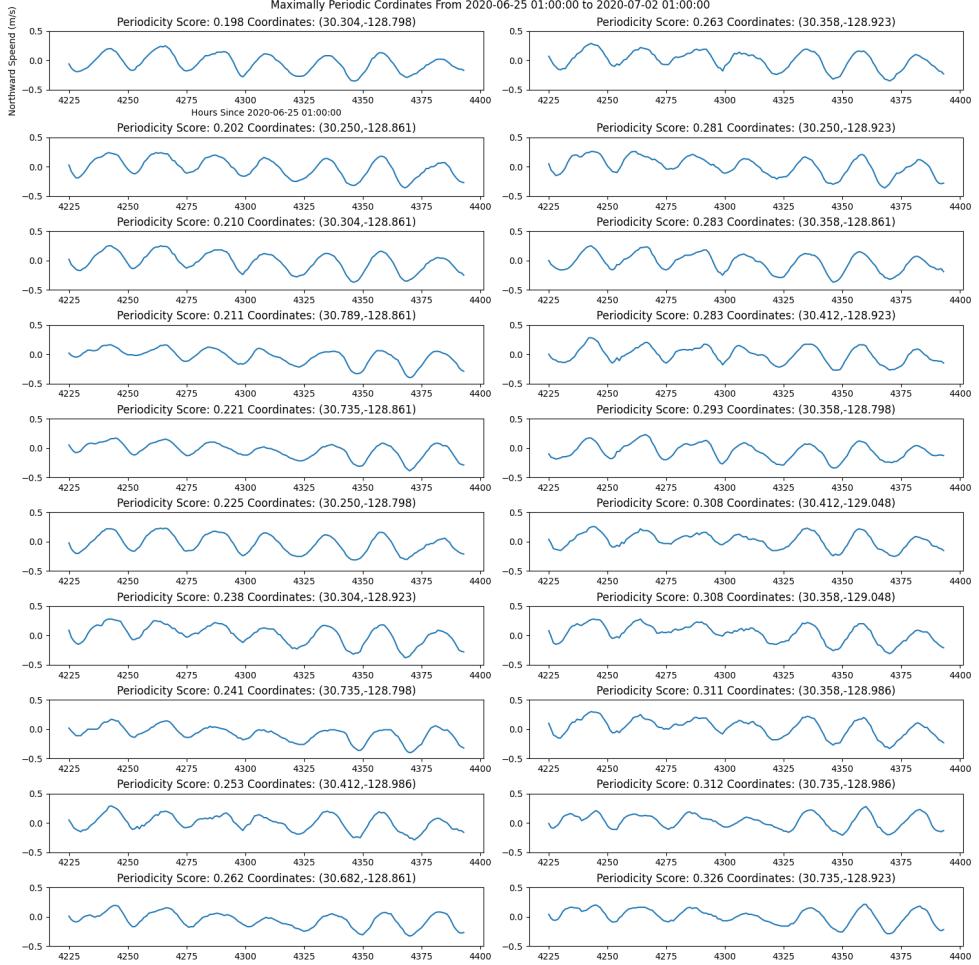


Figure 24: A graph of the 20 most periodic velocity curves from 2020-05-025 to 2020-07-02. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

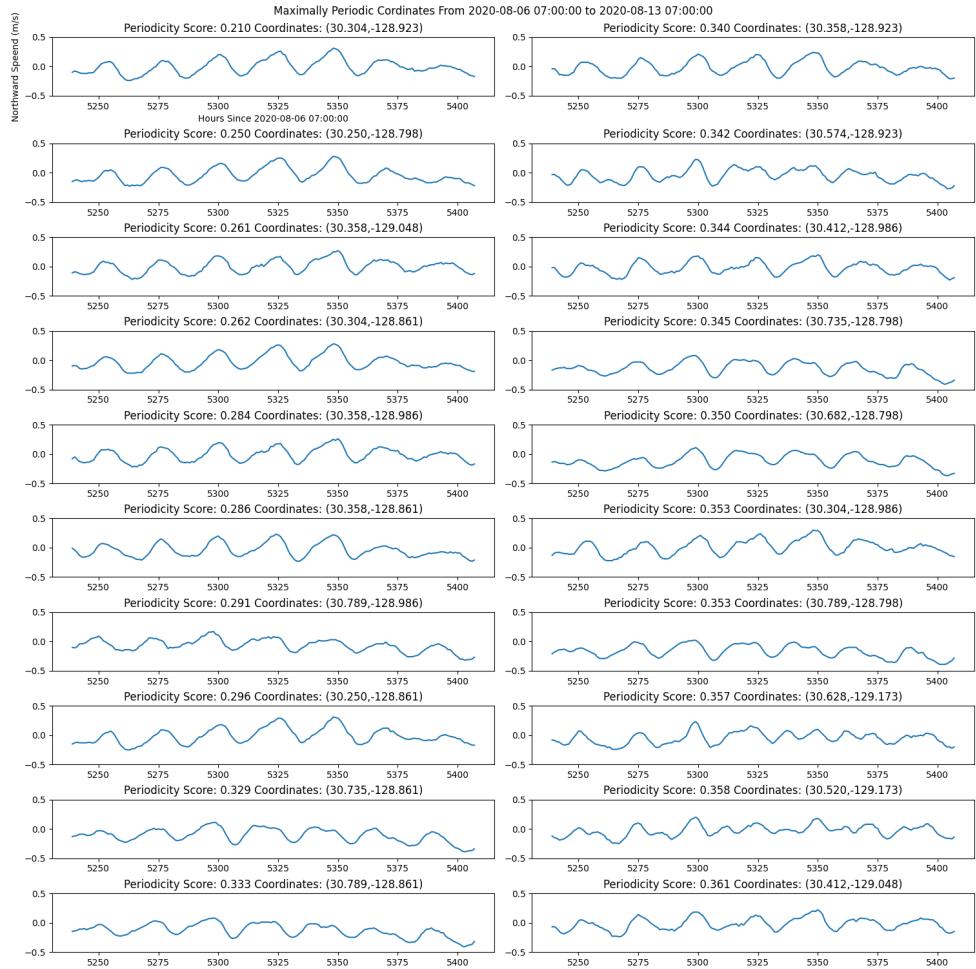


Figure 25: A graph of the 20 most periodic velocity curves from 2020-08-06 to 2020-08-13. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

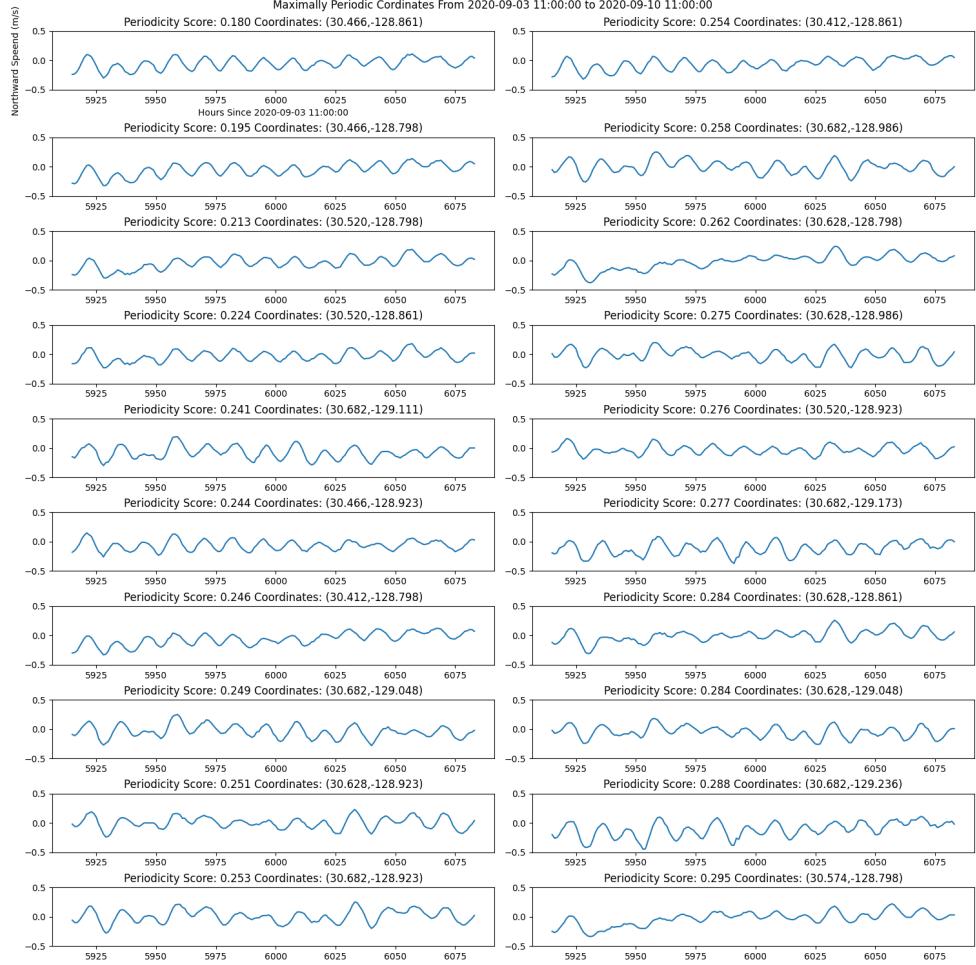


Figure 26: A graph of the 20 most periodic velocity curves from 2020-09-03 to 2020-09-10. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

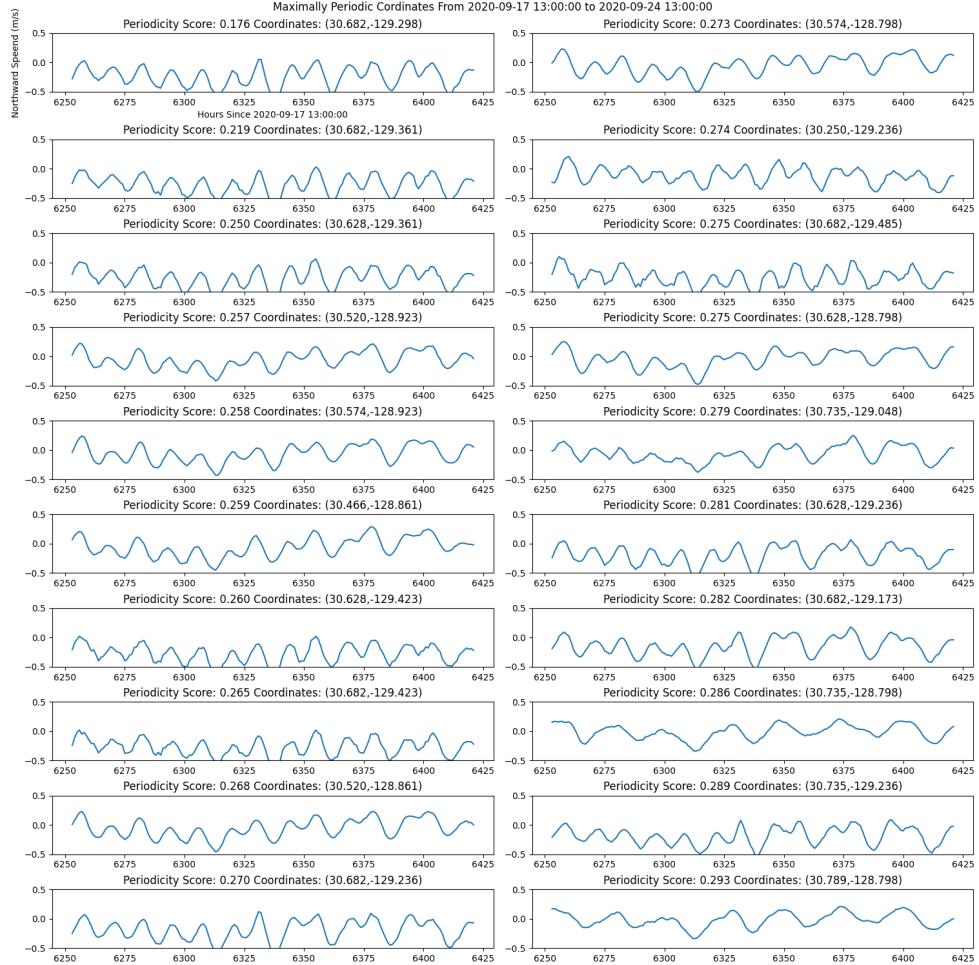


Figure 27: A graph of the 20 most periodic velocity curves from 2020-09-17 to 2020-09-24. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 3, 5, 15, 30$, and $p = 61$.

In Figures 23 and 27 we can see a periodic curves with two apparent frequencies, in Figures 25 and 26 we can see more triangular periodic curves, and finally in 24 we can see a more trigonometric curves (although they may just be more triangular curves). This demonstrates SW1PerS ability to detect different types of periodicity. This reveals that northward velocity within the study patch can exhibit a 24 hour period with two local maximums, a 12 hour period with one local maximum, or a 24 hour period with one maximum. Compare Figures 23, 26, and 25, respectively.

3.2 Periodicity Over a Month

We examine the northward surface ocean velocity data from 2020-01-01 00:00:00 to 2023-01-01 00:00:00. The data is evenly split into 36 chunks, meaning that each chunk is approximately a month long. We begin by examining the matrix of periodicity score graphs with two sets of parameters. In Figure 28, we use parameters $M = 360$, $n = 200$, $L = 2, 3, 5, 7$, and $p = 61$ to look for periodicity on the scale of weeks. In Figure 29, we use parameters $M = 360$, $n = 200$, $L = 15, 30, 35, 40$, and $p = 61$ to look for periodicity on the scale of days. Parameters are mostly chosen to be optimal, as described in Algorithm 1, but due to computational complexity, we pick a smaller prime than suggested: $p = 61$.

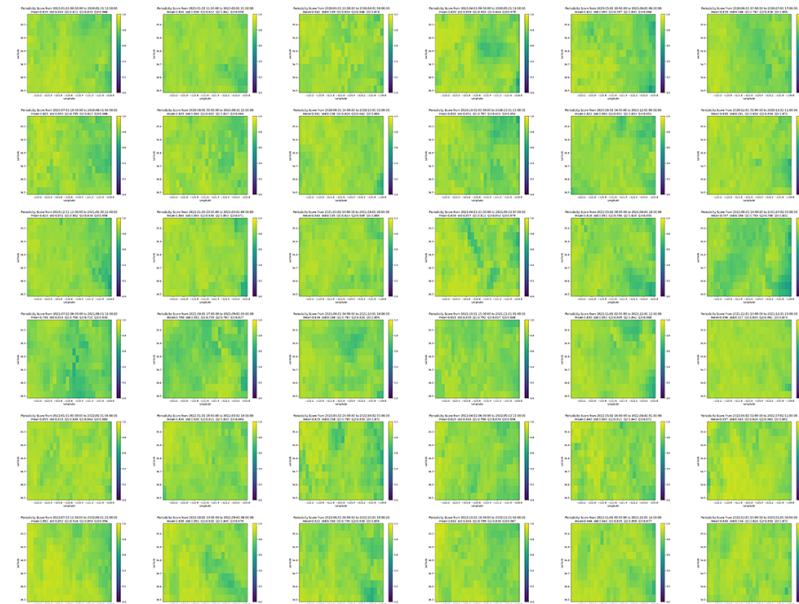


Figure 28: This is a matrix of periodicity score graphs for each of the 52 time segments from 2020-01-01 to 2021-01-08. SW1PerS was run with parameters $M = 360$, $n = 200$, $L = 2, 3, 5, 7$, and $p = 61$.

Due to the bluer hue in Figure 28 versus Figure 29, we see that the average periodicity score is lower for each time period when using $L = 15, 30, 35, 40$ versus $L = 2, 3, 5, 7$. This indicates more periodicity on the time scale of days than weeks. However, when comparing to Figure 22, we see that Figure 29 is still much more yellow. This generally indicates a low level of periodicity compared to the periodicity observed between days in

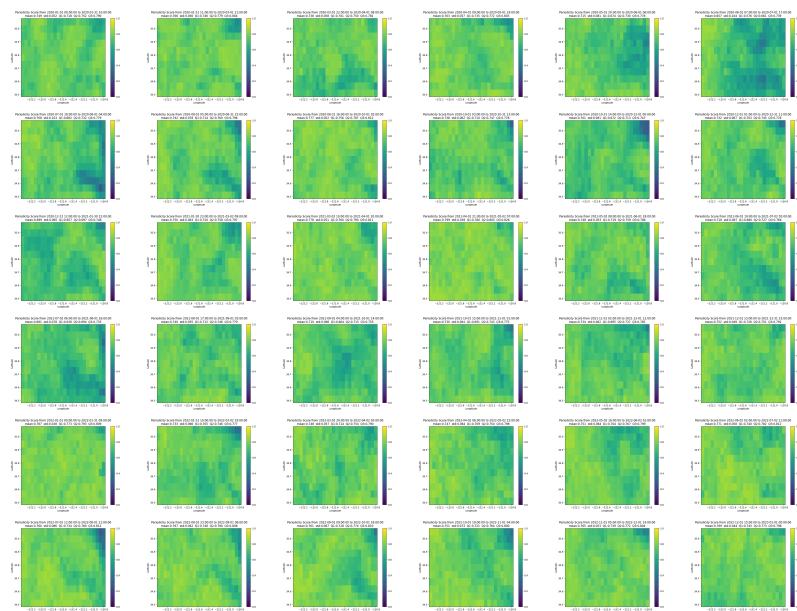


Figure 29: This is a matrix of periodicity score graphs for each of the 36 time segments from 2020-01-01 to 2023-01-01 (see Figure 19). SW1PerS was run with parameters $M = 360$, $n = 200$, $L = 15, 30, 35, 40$, and $p = 61$.

Section 3.1. Figure ?? is the periodicity score graph for the time segment with the lowest average periodicity score from the 36 segments spanning 2020-01-01 to 2023-01-01. This turns out to be the 18th segment.

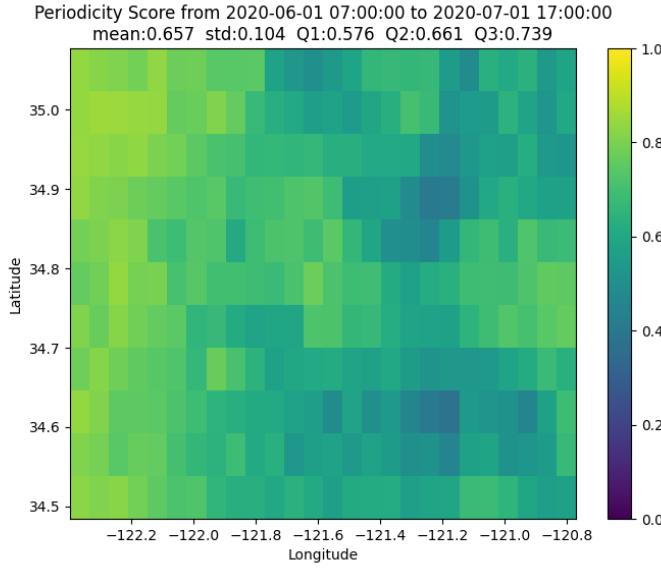


Figure 30: This is a graph of the minimum periodicity score computed by SW1PerS with parameters $M = 360$, $n = 200$, $L = 15, 30, 35, 40$, and $p = 61$.

Looking at Figure ??, we can see the mean and median periodicity scores from 2021-06-01 07:00:00 to 2021-07-01 17:00:00 are 0.657 and 0.661. These are both quite high, indicating a low periodicity score for northward surface ocean velocity on the time scale of multiple days.

In Figure ?? we show the 20 most periodic velocities from time segment 18.

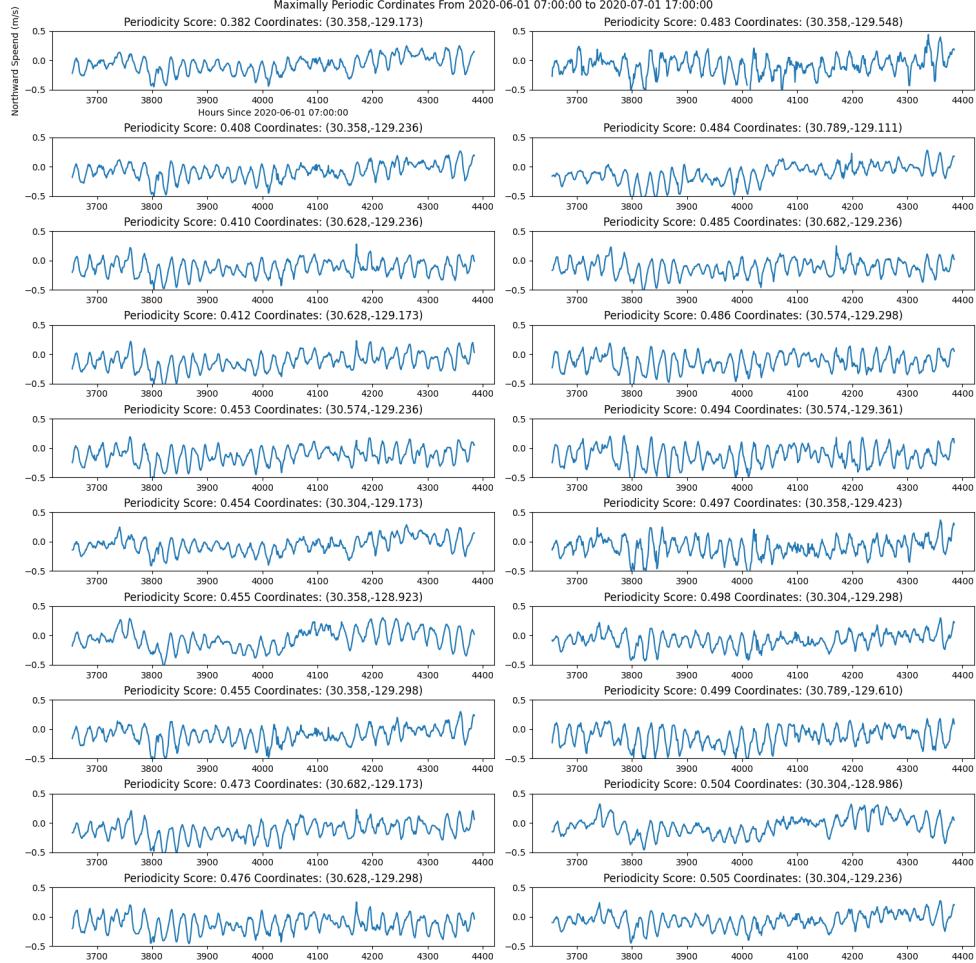


Figure 31: A graph of the 20 most periodic velocities from 2020-06-01 07:00:00 to 2020-07-01 17:00:00. SW1PerS was run with parameters $M = 58$, $n = 200$, $L = 15, 30, 35, 40$, and $p = 61$.

In Figure ?? we can see periodic oscillations on the scale of 16-20 hours. This is more rapid than the values of L that we checked. Still, we get a minimum periodicity score of 0.382, which is much lower than the corresponding mean of 0.657.

3.3 Periodicity Over a Year

We examine the northward surface ocean velocity data from 2014-01-01 00:00:00 to 2024-01-01 00:00:00. The data is evenly split into 10 chunks,

meaning that each chunk is approximately a year long. We will begin by examining the matrix of periodicity score graphs for each time period with two sets of parameters. In Figure ??, we use parameters $M = 4380$, $n = 200$, $L = 2, 3, 5, 6$, and $p = 61$ to look for periodicity on the scale of months. In Figure ??, we use parameters $M = 4380$, $n = 200$, $L = 365, 730$, and $p = 61$ to look for periodicity on the scale of 12 to 24 hours. Parameters are mostly chosen to be optimal as described in Algorithm 1, but due to computational complexity, we pick a smaller prime than suggested: $p = 61$.

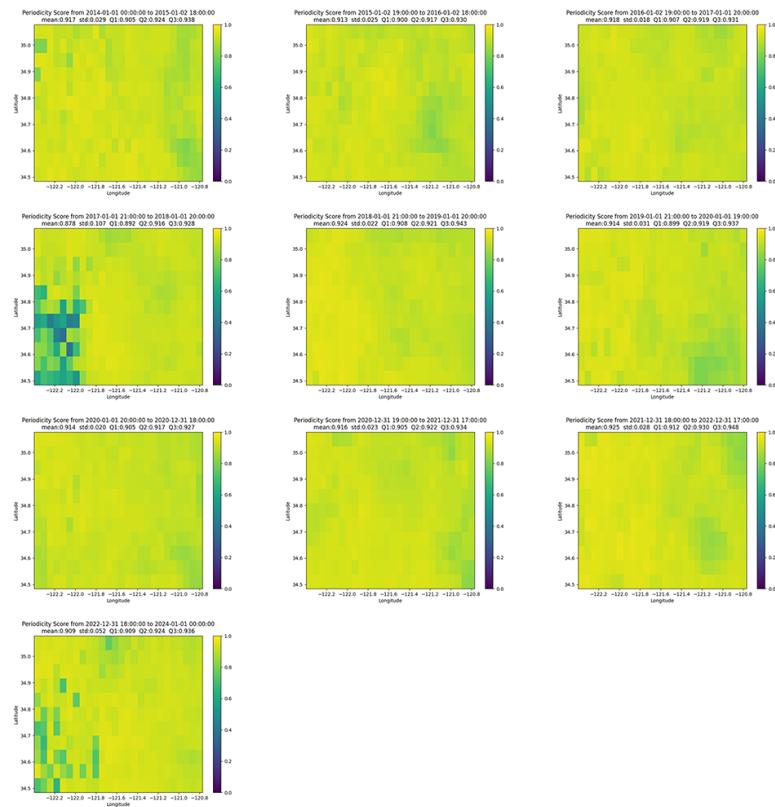


Figure 32: This is a matrix of periodicity score graphs for each of the 52 time segments from 2014-01-01 to 2024-01-01. SW1PerS was run with parameters $M = 4380$, $n = 200$, $L = 2, 3, 5, 6$, and $p = 61$.

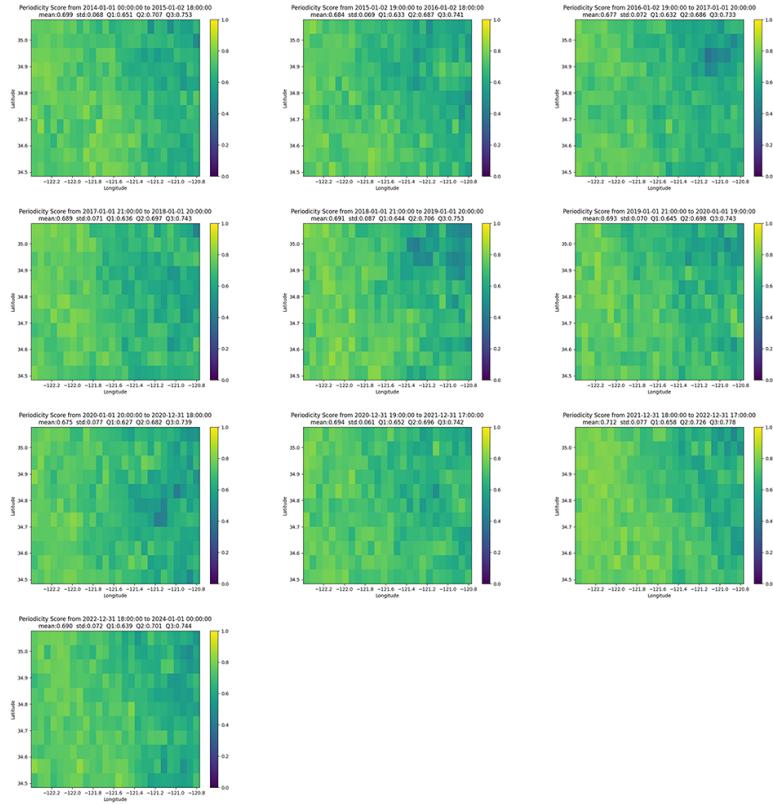


Figure 33: This is a matrix of periodicity score graphs for each of the 52 time segments from 2014-01-01 to 2024-01-01. SW1PerS was run with parameters $M = 4380$, $n = 200$, $L = 365, 730$, and $p = 61$.

In Figure ??, the periodicity score matrix for each time period is almost purely yellow, which indicates a complete lack of periodicity on the time scale of months. Note that the region of blue in the 4th time segment may be disregarded due to missing data (See Figure ??).

On the other hand, in Figure ?? we see a decent amount of blue in each of the periodicity graphs for each time segment. This indicates that SW1PerS is still able to detect periodicity with 12 to 24 hours (these are similar periods to Section 3.1) within a much longer time segment with much more data than in Section 3.1.

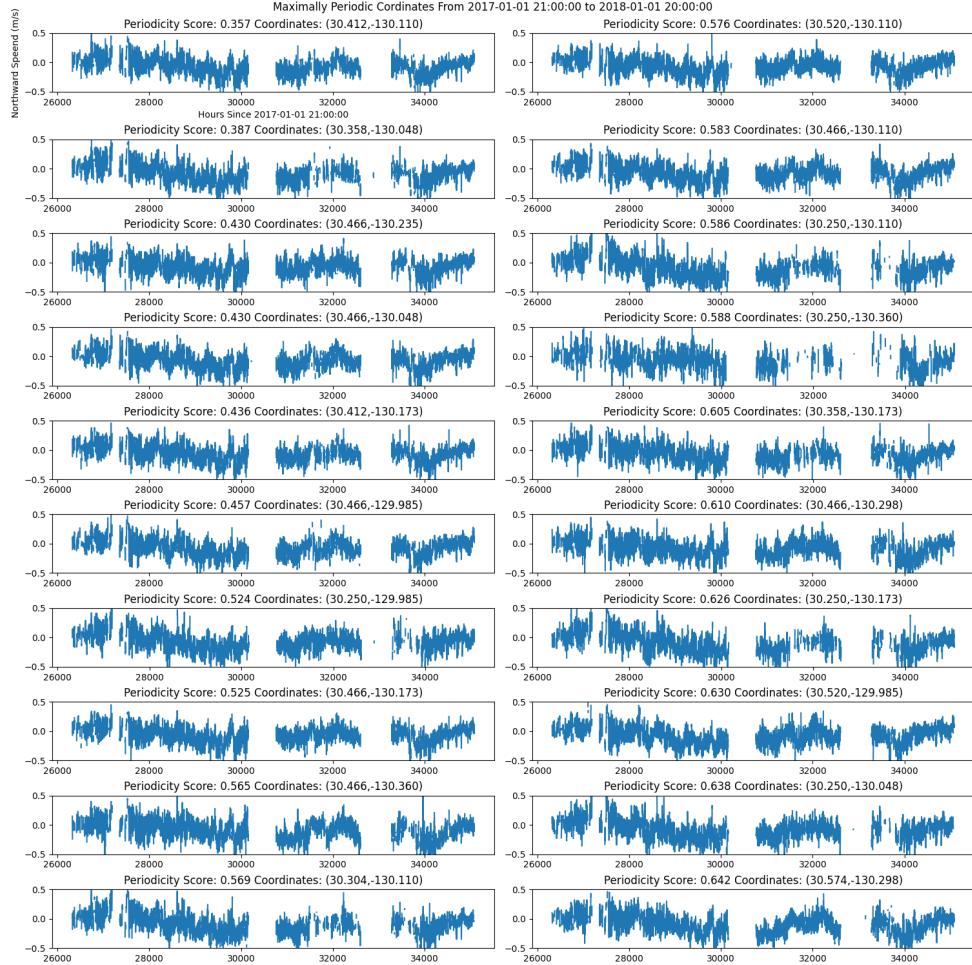


Figure 34: A graph of the 20 most periodic velocities from 2017-01-01 21:00:00 to 2018-01-01 20:00:00. SW1PerS was run with parameters $M = 4380$, $n = 200$, $L = 2, 3, 5, 6$, and $p = 61$.

In Figure ??, we show the periodicity graph for the time period with the lowest mean periodicity. This is the third time period spanning 2016-01-02 19:00:00 to 2017-01-01 20:00:00. We see an average periodicity of 0.677. This is much less than the periodicity scores seen in ???. This indicates that the northward velocity curves in the study patch exhibit high periodicity on the daily time scale and little periodicity on the monthly time scale, according to SW1PerS.

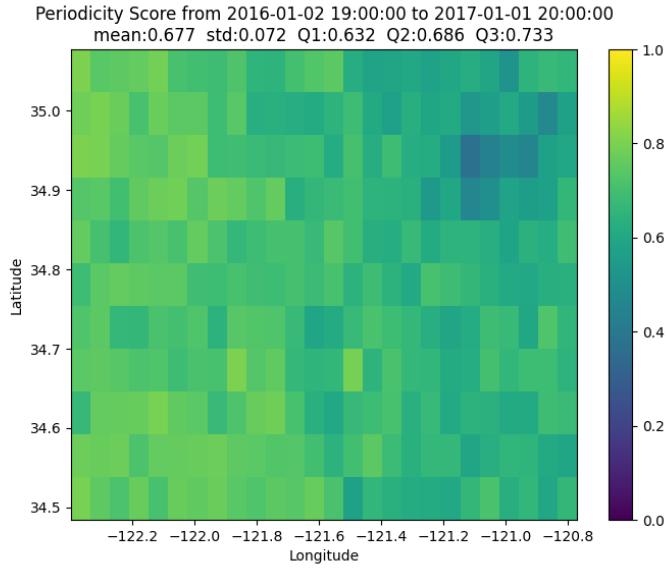


Figure 35: This is a graph of the minimum periodicity score computed by SW1PerS with parameters $M = 4380$, $n = 200$, $L = 365, 730$, and $p = 61$.

3.4 Summary of Periodicity in the Study Patch

Across each time scale Figures 22, 28, 29, ??, and ?? indicate that the study patch is more periodic on the eastern side, versus the western side. This difference could be due to noise, as the western side has more missing data because it is farther away from radio stations on the shore. Additionally, Figures 23, 24, 25, 26, and 27 reveal that SW1PerS was able to detect different types of periodic curves on the time scale of days. We found that SW1PerS did not detect significant periodicity on the time scale of months. However, SW1PerS was still able to find periodicity in the study patch on the time scale of days when considering the data for a whole year. SW1PerS's failure to find periodicity on the scale of months may be related to the window size being larger than a day when L is adjusted to look for periodic behavior over months. See Figure 16, where the periodicity score rapidly increases as the window size becomes bigger than the expected period.

4 Closing Remarks

4.1 Alternative Intuition for Sliding Window 1-Persistent Homology

We will now take a detour to explore an alternative formulation of sliding window 1 persistence homology scoring. It should be noted that [6] was not able to show that the roundness of $SW_{M,\tau}f(t)$ is maximized when the window size $M\tau$ approximates the period of $f(t)$. Instead, Perea and Harer showed that the roundness of $SW_{M,\tau}\cos(Lt)$ and $SW_{M,\tau}\sin(Lt)$ is maximized when the window size matches the period. They hoped to extend this result to all periodic functions because every periodic function can be represented as an infinite sum of sines and cosines via its Fourier series. Additionally, part of the SW1PerS algorithm relies on splining to fill in missing data. This assumes that the spline will be an accurate approximation of the true observation curve. In this case, why not work directly with the splined curve and the standard inner product induced norm on $\mathcal{L}^2([0, 2\pi])$? (Note that this interpretation would not make use of Taken’s Theorem and the reconstruction of state space.) We would like to recover the idea that the roundness of $SW_{M,\tau}f(t)$ is maximized when the window size approximates the period of f , as seen in Example 2.33. The following results are written with general invariant measures instead of just Lebesgue measure to demonstrate the generality of this perspective. We will not delve into the definition or background of the Lebesgue integral. However, the general reasoning applies to standard Riemann integrals, and so the reader will not lose much by considering the results for Riemann integrals for the purposes of motivation.

Theorem 4.1. *Let (X, S, μ) be a measure space, where X is a linearly ordered abelian Archimedean group and μ is a translation-invariant measure. If $f : X \rightarrow \mathbb{R}$ is an S -measurable function and there exists $w \in X$ so that $f(x + w) = f(x)$ for all $x \in X$, then*

$$\int_{[0,w)} f(x + t) d\mu(x) = \int_{[0,w)} f(x) d\mu(x) \quad \forall t \in X.$$

Proof. Suppose $f : X \rightarrow \mathbb{R}$ is an S -measurable function and there exists $w \in X$ so that $f(x + w) = f(x)$ for all $x \in X$. Observe that μ being

translationally invariant implies that

$$\int_{[0,w)} f(x+t) d\mu(x) = \int_{[t,w+t)} f(x) d\mu(x).$$

Due to the Archimedean property of X , there exists $n_0 \in \mathbb{Z}$ so that $n_0w < t \leq n_0w + w$. It follows that $n_0w + w < w + t$. Then

$$\int_{[t,w+t)} f(x) d\mu(x) = \int_{[t,n_0w+w)} f(x) d\mu(x) + \int_{[n_0w+w,w+t)} f(x) d\mu(x).$$

Again, because μ is translationally invariant, it follows that

$$\int_{[n_0w+w,w+t)} f(x) d\mu(x) = \int_{[n_0w,t)} f(x+w) d\mu(x). \quad (1)$$

Recall that $f(x+w) = f(x)$ for all $x \in X$. We have that

$$\int_{[n_0w,t)} f(x+w) d\mu(x) = \int_{[n_0w,t)} f(x) d\mu(x). \quad (2)$$

Combining (??) and (??) we get that

$$\begin{aligned} \int_{[t,w+t)} f(x) d\mu(x) &= \int_{[t,n_0w+w)} f(x) d\mu(x) + \int_{[n_0w+w,w+t)} f(x) d\mu(x) \\ &= \int_{[t,n_0w+w)} f(x) d\mu(x) + \int_{[n_0w,t)} f(x) d\mu(x) \\ &= \int_{[n_0w,n_0w+w)} f(x) d\mu(x) \\ &= \int_{[0,w)} f(x+n_0w) d\mu(x) \\ &= \int_{[0,w)} f(x) d\mu(x). \end{aligned}$$

□

In Theorem ?? we see that translating a periodic function integrated over the length of its period does not change the value of the integral.

Theorem 4.2. *Let (X, S, μ) be a measure space, where X is a linearly ordered abelian Archimedean group and μ a translation-invariant measure.*

Consider $\mathcal{L}^2(X) = \{f : X \rightarrow \mathbb{R} : \int_X f^2 d\mu < \infty\}$. Let $\langle f, g \rangle_b = \int_{[0,b]} fg d\mu$. Then define $rf(t) : X \rightarrow \mathcal{L}^2(X)$ by

$$rf(t) = f(x + t).$$

If there exists $w \in X$ so that $f(x + w) = f(x)$ for all $x \in X$, then

$$\|rf(t)\|_w = \sqrt{\langle rf, rf \rangle_w} = \sqrt{\int_{[0,w)} f(x)^2 d\mu(x)}.$$

This shows that $\|rf(t)\|$ is a constant function.

Given a measurable function $f : X \rightarrow \mathbb{R}$, Theorem ?? implies that $rf(t)$ lies on the ball of radius $\sqrt{\int_{[0,w)} |f(x)|^2 d\mu(x)}$ within $\mathcal{L}^2(X)$. Then $r_{f,w}$ must be perfectly round, lying within a sphere. The length we integrate over corresponds to the window size, and we have somewhat recovered the result from Example 2.33. The next example will emphasize this fact.

Theorem 4.3. Let $X = \mathbb{R} \setminus [0, 1]$ and μ be Lebesgue measure. If $f : X \rightarrow \mathbb{R}$ is a continuous function and $w \in [0, 1]$ is the least positive real number so that $f(x + w) = f(x)$ for all $x \in X$, then $rf : [0, w] \rightarrow \mathcal{L}^2(X)$ is injective and $r_{f,b}([0, w])$ is isometric (and homeomorphic) to $\{x \in \mathbb{R}^2 : \|x\| = R\}$.

Proof. Note that because singletons are measure zero under Lebesgue measure, we can replace the half-open intervals in Theorem ?? with closed intervals. We first show that $rf : [0, w] \rightarrow \mathcal{L}^2(X)$ is injective. Suppose that $t_1, t_2 \in [0, w]$ so that $rf(t_1) = rf(t_2)$. Then $f(x + t_1) = f(x + t_2)$ for all $x \in X$. Now let $x \in X$ and observe that $x - t_0 \in X$. Then

$$f((x - t_0) + t_1) = f((x - t_0) + t_0) \implies f(x + (t_1 - t_0)) = f(x)$$

for all $x \in X$. Notice that $t_1 - t_0 \in [0, w]$ because $0 \leq t_0 \leq t_1 \leq w$. Either $t_1 - t_0 = 0$ or $t_1 - t_0 \neq 0$. In the latter case, $t_1 - t_0$ violates the minimality of the period w so it must be that $t_1 - t_0 = 0$ and $t_1 = t_0$. We have now shown that $rf : [0, w] \rightarrow \mathcal{L}^2(X)$ is injective.

It is clear that $rf : [0, w] \rightarrow rf([0, w])$ is an injective surjection, so its inverse $(rf)^{-1} : r_{f,w}([0, w]) \rightarrow [0, w]$ exists. Now define $\phi : rf([0, w]) \rightarrow \{x \in \mathbb{R}^2 : \|x\| = R\}$ by $\phi(y) = R \exp(2\pi i (rf)^{-1}(y)/w)$. Note that $(rf)^{-1}$ is a bijection and $R \exp(2\pi i t/w) : [0, w] \rightarrow \{x \in \mathbb{R}^2 : \|x\| = R\}$ is a bijection. Then ϕ is a bijection as the composition of bijections is a bijection. It

follows from Theorem ?? that $\|\phi(y)\| = R = \|R \exp(2\pi i r_{f,w}^{-1}(y)/w)\|$; so ϕ is an isometry. Finally, note that isometries are continuous, so ϕ is a continuous bijection. Because ϕ is an isometry, so is its inverse, so ϕ^{-1} . Then ϕ^{-1} is also a continuous bijection. We have that ϕ is an isometric homeomorphism; we have shown the result. \square

4.2 Conclusion

We found that SW1PerS was able to detect different types of daily periodicity within the Near-Real Time Surface Ocean Velocity, U.S., 6km Resolution Hourly data collected by the IOOS. Further study could be done to determine whether these different types of periodic curves are correlated with their location within the study patch or the time of year. SW1PerS found the eastern portion of the study patch to be more periodic than the western portion, though this could be due to noise. In order to address this issue, we could implement a moving average as mentioned in [7] or we could rerun SW1PerS on the Near-Real Time Surface Ocean Velocity, U.S., 6km Resolution, 25-hour, monthly, or yearly data averages collected by the IOOS. Additionally, we could follow [9] and implement the SW1PerS algorithm to compute periodicity over the whole velocity field at once instead of per data point. Finally, we could use the result from Theorem ?? to investigate a version of SW1PerS homology using windows embedded in $\mathcal{L}^2([0, 2\pi])$ with the metric induced by the standard inner product. The main point to figure out here is how to normalize the point cloud so that periodicity scores for different window sizes are comparable.

The code and data used in this project can be found at [here](#). The following repository was a useful resource for performing topological data analysis in Python [1].

References

References

- [1] *Github - ctralie/TDALabs: Pedagogical Jupyter notebooks for topological data analysis. Topics include basic shapes, sliding window audio/video, lower star filtrations, 3D shapes, spaces of images patches — github.com/ctralie/TDALabs.* [Accessed 14-06-2025].

- [2] GitHub - scikit-tda/ripser.py: A Lean Persistent Homology Library for Python — github.com. <https://github.com/scikit-tda/ripser.py>. [Accessed 14-06-2025].
- [3] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [4] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2019.
- [5] National Oceanic and Atmospheric Administration, Department of Commerce - Near-Real Time Surface Ocean Velocity, U.S. West Coast, 6km Resolution — catalog.data.gov. <https://catalog.data.gov/dataset/near-real-time-surface-ocean-velocity-u-s-west-coast-6km-resolution5>. [Accessed 14-06-2025].
- [6] Jose A. Perea and John Harer. “Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis”. In: *Found Comput Math* 15.3 (June 1, 2015), pp. 799–838. ISSN: 1615-3383. DOI: 10.1007/s10208-014-9206-z. URL: <https://doi.org/10.1007/s10208-014-9206-z> (visited on 01/17/2025).
- [7] Jose A. Perea et al. “SW1PerS: Sliding Windows and 1-Persistence Scoring; Discovering Periodicity in Gene Expression Time Series Data”. In: *BMC Bioinformatics* 16.1 (Aug. 16, 2015), p. 257. ISSN: 1471-2105. DOI: 10.1186/s12859-015-0645-6. URL: <https://doi.org/10.1186/s12859-015-0645-6> (visited on 01/04/2025).
- [8] *Rips Filtration Tutorial for the R-TDA Package* — comptag.github.io. https://comptag.github.io/rpackage_tutorials/2019/07/tda-rips-tutorial.html. [Accessed 13-06-2025].
- [9] Christopher J. Tralie and Jose A. Perea. *(Quasi)Periodicity Quantification in Video Data, Using Topology*. Jan. 21, 2018. DOI: 10.48550/arXiv.1704.08382. arXiv: 1704.08382 [cs]. URL: <http://arxiv.org/abs/1704.08382> (visited on 06/14/2025). Pre-published.

- [10] Muneera Yousif et al. “Periodicity Analysis of Most Time Series Methods: A Review”. In: *2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. 2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). Nov. 2024, pp. 1–7. DOI: 10.1109/ISMSIT63511.2024.10757242. URL: <https://ieeexplore.ieee.org/document/10757242> (visited on 06/11/2025).