

Using Persistent Homology to Quantify a Diurnal Cycle in Hurricane Felix

Sarah Tymochko, Elizabeth Munch, Jason Dunion, Kristen Corbosiero, and Ryan Torn

Abstract—The diurnal cycle of tropical cyclones (TCs) is a daily cycle in clouds that appears in satellite images and may have implications for TC structure and intensity. The diurnal pattern can be seen in infrared (IR) satellite imagery as cyclical pulses in the cloud field that propagate radially outward from the center of nearly all Atlantic-basin TCs. These diurnal pulses, a distinguishing characteristic of the TC diurnal cycle, begin forming in the storm’s inner core near sunset each day and appear as a region of cooling cloud-top temperatures. The area of cooling takes on a ring-like appearance as cloud-top warming occurs on its inside edge and the cooling moves away from the storm overnight, reaching several hundred kilometers from the circulation center by the following afternoon. The state-of-the-art TC diurnal cycle measurement has a limited ability to analyze the behavior beyond qualitative observations. We present a method for quantifying the TC diurnal cycle using one-dimensional persistent homology, a tool from Topological Data Analysis, by tracking maximum persistence and quantifying the cycle using the discrete Fourier transform. Using Geostationary Operational Environmental Satellite IR imagery data from Hurricane Felix (2007), our method is able to detect an approximate daily cycle.

Index Terms—Topological Data Analysis, Atmospheric Science, Hurricane, Diurnal Cycle

I. INTRODUCTION

THE field of atmospheric science has numerous observation platforms that provide high space and time resolution data, but has yet to find methods which can quantify the intuitive patterns explicitly. Meanwhile, the young field of Topological Data Analysis (TDA) encompasses methods for quantifying exactly these sorts of structural intuitions seen by atmospheric scientists. This paper merges these two fields by using persistent homology, a now well-established tool in TDA, to quantify a diurnal cycle observed in a hurricane using Geostationary Operational Environmental Satellite (GOES) infrared (IR) satellite data.

Persistent homology, and more generally TDA methods, has found significant success in rather disparate applications by finding structure in data and using this insight to answer questions from the domain of interest. For instance, Giusti et al. used the homology of random simplicial complexes to investigate the geometric organization of neurons in rat brains [1]. Nicolau et al. used mapper, another tool in TDA, to discover a new subtype of breast cancer [2]. More closely

ST and EM are with the Dept. of Computational Mathematics, Science and Engineering at Michigan State University. EM is also with the Dept. of Mathematics at MSU. KC and RT are with the Dept. of Atmospheric and Environmental Sciences at University at Albany - SUNY Albany. JD is affiliated with Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and Hurricane Research Division, NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida.

Corresponding author: ST, e-mail: tymochko@egr.msu.edu.

related to this work is the use of TDA for time series analysis and image processing. This includes using persistent homology to understand periodicity in time series arising from biological [3], [4] and engineering applications [5]. There has also been a great deal of interest in using persistence for image analysis, e.g. [6], [7], [8], [9].

This paper presents an application of the use of both time series and image analysis using TDA. The diurnal cycle of tropical cyclones (TCs) has been described in previous studies [10], [11], [12], [13], [14], [15], [16], [17], [18] that provide evidence of the regularity of this cycle as well as its potential impacts. This diurnal pattern can be seen in GOES IR imagery as cyclical pulses in the cloud field that propagate radially outward from TCs at speeds of 5-10 m s⁻¹ [10], [11], [15]. These diurnal pulses, a distinguishing characteristic of the TC diurnal cycle, begin forming in the TC’s core near the time of sunset each day and appear as a region of cooling cloud-top temperatures. The area of cooling then takes on a ring-like appearance as marked cloud-top warming occurs on its inside edge and it moves away from the storm overnight, reaching several hundred kilometers from the TC center by the following afternoon. Observations and numerical model simulations indicate that TC diurnal pulses propagate through a deep layer of the TC environment, suggesting that they may have implications for TC structure and intensity [10], [11], [15], [17].

The current state of the art TC diurnal cycle measurement has a limited ability to analyze the behavior beyond qualitative observations. This paper presents a more advanced mathematical method for quantifying the TC diurnal cycle using tools from TDA, namely one-dimensional persistent homology to analyze the holes in a space. This research aims to detect the presence of the diurnal cycle in GOES IR satellite imagery and to track the changes through a time series.

The first attempt, using the naive combination of persistent homology with the GOES IR imagery, did not show the recurring pattern. Due to the drastically variable values in the IR brightness temperature data, persistent homology was not able to detect any significant structure. Looking at the data, however, there is a clear circular feature visible, so we developed more sophisticated methods to extract this structure.

In this paper, we present a method applying the distance transform and one-dimensional persistent homology, allowing us to quantify the cycle using maximum persistence. We show that using tools from TDA, we can detect cyclic behavior in the hurricane that repeats approximately every 24 hours.

II. TROPICAL CYCLONE BACKGROUND

Previous research has documented a clear diurnal cycle of cloudiness and rainfall in TCs: enhanced convection (i.e., thunderstorms) occurs overnight, precipitation peaks near sunrise, and upper-level cloudiness (i.e., the cirrus canopy) expands radially outward throughout the day, reaching its maximum areal coverage in the early evening hours [10], [11], [12], [13], [14], [15], [16], [17], [18]. To quantify the expansion and contraction of the cirrus canopy, Dunion et al. used GOES satellite IR imagery to examine the six-hour cloud-top temperature differences of major hurricanes in the Atlantic basin from 2001 to 2010 [10]. They found that an area of colder cloud tops propagated outward around $5\text{-}10 \text{ m s}^{-1}$ over the course of the day, with warming temperatures on its inner edge. More recently, in [15], Ditcheck et al. expanded Dunion et al.'s work to include all tropical cyclones in the Atlantic basin from 1982 to 2017 and found that the diurnal pulse is nearly ubiquitous, with 88% of TC days featuring an outwardly propagating pulse.

Despite the consistent signature and documentation of this diurnal cloud signature, open questions remain as to how the diurnal cycle is linked to inner-core convective processes and whether it is a column-deep phenomenon or mainly tied to upper-level TC cloud dynamics related to incoming solar radiation [12], [13], [14], [15]. Investigating these questions is relevant to TC forecasting as the diurnal cycle of clouds and rainfall has implications for forecasting storm structure and intensity, as evidenced by the diurnal cycle in objective measures of TC intensity and the extent of the 50-kt wind radius documented by Dunion et al. Additionally, and especially relevant to the current work, most of the papers above have identified the pulse using subjective measures of cloud-top temperature change and timing [10], [15]. The current work seeks to quantify the pulse to determine its true periodicity using persistent homology, a topological tool that is particularly effective at capturing the type of patterns visible in the pulse.

III. MATH BACKGROUND

Persistent homology is a tool from the field of TDA which measures structure in data. This data can start in many forms, including as point clouds or, as in the case of this work, as a function on a domain. In this section, we will briefly review the necessary background to understand cubical homology and persistent homology, and refer the interested reader to [19], [20], [21], [22], [23] for a more complete introduction. Additionally, we will introduce tools used in our method, including the distance transform, mathematical morphology, and the Fourier transform.

A. Cubical complexes

In this section, we largely follow Chapter 2 of [19] with the caveat for the informed reader that because we use homology with \mathbb{Z}_2 coefficients, we can be lazy about orientations of cubes. In addition, our data consists of 2D images, so we need only define cubes up to dimension 2.

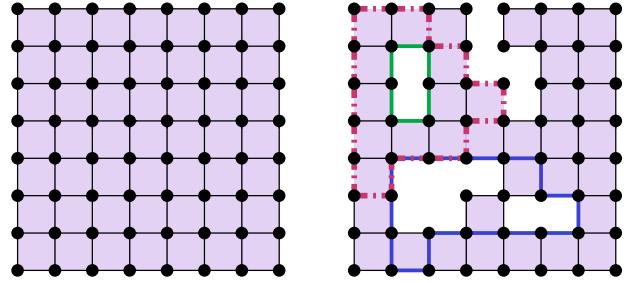


Fig. 1. An example of K for representing a 7×7 matrix is shown at left. A subset of this complex is shown at right. The red dashed loop and the green bold loop represent equivalent classes in $H_1(M_r)$. The bold blue loop represents a different equivalence class in $H_1(M_r)$. For this example, $H_1(M_r)$ has rank 3.

An elementary interval is a closed interval $I \subset \mathbb{R}$ of the form $[\ell, \ell+1]$ or $[\ell]$ for $\ell \in \mathbb{Z}$, which are called nondegenerate and degenerate respectively. An elementary cube $Q \in \mathbb{R}^2$ is a product of elementary intervals $Q = I_1 \times I_2$. The dimension of Q , $\dim(Q)$, is the number of nondegenerate components of Q . Note that 0-dimensional cubes are just vertices at the points on the lattice $\mathbb{Z} \times \mathbb{Z}$ in \mathbb{R}^2 , 1-dimensional cubes are edges connecting these vertices, and 2-dimensional cubes are squares. Let \mathcal{K} denote the set of all elementary cubes in \mathbb{R}^2 and $\mathcal{K}_d \subset \mathcal{K}$ the set of d -dimensional cubes. A set $X \subset \mathbb{R}^2$ is cubical if it can be written as a finite union of elementary cubes. Then we denote the associated cubical complex as $\mathcal{K}(X) = \{Q \in \mathcal{K} \mid Q \subset X\}$, with the d -dimensional subset denoted $\mathcal{K}_d(X) = \{Q \in \mathcal{K}(X) \mid \dim(Q) = d\}$. If $Q \subseteq P$, then we say Q is a face of P , denoted $Q \leq P$. If $Q \subsetneq P$, then Q is a proper face of P , denoted $Q < P$, and is additionally a primary face of P if $\dim(Q) = \dim(P) - 1$.

A greyscale image, or more generally an $m \times n$ matrix, can be viewed as a function $M : D \rightarrow \mathbb{R}$ where

$$D = \{(i, j) \mid 0 \leq i < m, 0 \leq j < n\}.$$

We will model this as a function defined on a particularly simple cubical set $K = \mathcal{K}([0, m] \times [0, n])$; see the left of Fig. 1 for an example. For simplicity, we denote by $s_{i,j}$ the square $[i, i+1] \times [j, j+1]$.

So, given a matrix M , we equivalently think of this data as a function $M : K \rightarrow \mathbb{R}$ where we set $M(s_{i,j})$ equal to the matrix entry $M_{i,j}$ and set $M(P) = \min_{s_{i,j} > P} M(s_{i,j})$ for all lower dimensional cubes P . Note that we will abuse notation and use M to denote both the original matrix and the view of this matrix as a function with domain K .

B. Distance transform

The distance transform is a tool used in image processing and is computed on binary images [24], [25]. It is used in various applications in many disciplines, such as guiding robots to navigate obstacles [26], computing geometric representations such as Voronoi diagrams [27] as well as being a useful method in many other image processing tools [28].

This operation assigns each pixel in the foreground (pixels with value 1) a value based on its distance to a pixel in the background (pixels with value 0). This can be computed using

various distance metrics, most frequently the L_1, L_2 or L_∞ distance.

Using the notation from the previous section, $s_{i,j}$ represents the pixel, (i, j) in the image represented as a matrix of pixels, M . Given any $s_{i,j} \in M$

$$\min d(s_{i,j}, x)$$

where x is a 0-valued pixel and d is any distance metric. Given two pixels, s_{i_1,j_1}, s_{i_2,j_2} we calculate the L_∞ distance, also called the chessboard distance, between them as

$$d(s_{i_1,j_1}, s_{i_2,j_2}) = \max\{|i_2 - i_1|, |j_2 - j_1|\}.$$

This defines a distance on the pixels, which are the 2-cells in the cubical complex. The distance can be extended to the lower dimensional cells in the same manor as described in Sec. III-A.

C. Homology

Homology [20], [29] is a standard tool in algebraic topology which provides a vector space¹ $H_k(X)$ for each dimension $k = 0, 1, 2, \dots$ for a given topological space X . The different dimensions measure different properties of the space. In particular, for this work we are interested in 1-dimensional homology; i.e. when $k = 1$. The 1-dimensional homology group measures the number of loops in the space; equivalently, we can think of this as the number of holes in the space. In particular, if we look at the black region in each of the examples in Fig. 2, the rank of the first homology for each is (1,1,1,2,2,3).

The exact definition of homology is as follows. For any cubical set L (which for the purposes of this discussion will always be a subset of K), we have sets giving the cubes of different dimensions: $\mathcal{K}_i(L)$ for $i = 0, 1, 2$. An i -chain is a formal linear combination of i -simplices in L ,

$$c = \sum_{Q_j \in \mathcal{K}_i(L)} a_j Q_j,$$

with coefficients $a_j \in \mathbb{Z}_2$. We can of course add these objects by setting $(\sum a_j Q_j) + (\sum b_j Q_j) = \sum(a_j + b_j)Q_j$ and multiply by a constant. Thus, the collection of all i -chains forms a vector space $C_i(L)$.

We define a linear transformation

$$\delta_i : C_i(L) \rightarrow C_{i-1}(L)$$

called the boundary map, by setting $\delta_i(Q) = \sum P$ where the sum² is over the primary faces $P < Q$. The kernel of δ_1 , $\text{Ker}(\delta_1)$, (that is, the set of elements of $C_1(L)$ which map to 0) is generated by closed loops in L . The image of δ_2 , $\text{Im}(\delta_2)$, is generated by boundaries of 2-cells. Then the 1-dimensional homology group is defined to be $H_1(K) = \text{Ker}(\delta_1)/\text{Im}(\delta_2)$. An element of this group $\gamma \in H_1(L)$, represents an equivalence class of loops which can differ by collections of 2-cells; see the right side of Fig. 1 for an example of two loops which are equivalent in $H_1(L)$.

¹Normally a group, however, we are working with field coefficients.

²Again, notice that because we are working with \mathbb{Z}_2 coefficients, the bookkeeping normally needed for orientation is unnecessary.

D. Persistent Homology

For a *static* space L , $H_1(L)$ measures information about the number of loops. Persistent homology takes as input a *changing* topological space, and summarizes the information about how the homology changes.

Let an $m \times n$ \mathbb{R} -valued matrix M be given. Fix a function value $r \in \mathbb{R}$ and let $M_r = f^{-1}(-\infty, r]$. That is, M_r is the subset of squares in $K_{m \times n}$ which have value at most r in the matrix, along with all edges and vertices which are faces of any included square. M_r is often called a sublevel set of M . See the right side of Fig. 1 for an example of the structure of K and see Fig. 2 for M_r regions corresponding to the matrix on the left of Fig. 2. This shows M_r drawn in black for $r = 0.6, 0.9, 1.03, 1.1, 1.2$, and 1.23.

These spaces have the property that $M_r \subseteq M_s$ for $r \leq s$, thus we can consider the sequence

$$M_{r_1} \subseteq M_{r_2} \subseteq \cdots \subseteq M_{r_k} \quad (1)$$

for any set of numbers $r_1 < r_2 < \cdots < r_k$. This sequence of spaces is called a filtration. For each of the spaces, we can compute the homology group $H_p(M_{r_i})$. The inclusion maps of Eqn. (1) give rise to linear maps

$$H_p(M_{r_1}) \rightarrow H_p(M_{r_2}) \rightarrow \cdots \rightarrow H_p(M_{r_k}).$$

It is these maps that we study to understand how the space changes. In particular, when we are focused on 1-dimensional homology ($k = 1$) as in this study, a loop is represented by an element $\gamma \in H_1(M_{r_i})$. We say that this loop is born at r_i if it is not in the image from the previous space; that is, $\gamma \notin \text{Im}(H_1(M_{i-1}) \rightarrow H_1(M_i))$. This same loop dies at r_j if it merges with this image in M_{r_j} ; that is, $\gamma \in \text{Im}(H_1(M_{i-1}) \rightarrow H_1(M_j))$ where we abuse notation by using γ to both refer to the class in $H_1(M_{r_i})$ and the image of this class under the sequence of maps in $H_1(M_{r_j})$. We refer to $r_j - r_i$ as the lifetime of the class.

A persistence diagram, as seen in the right of Fig. 2, is a collection of points where for each class which is born at r_i and dies at r_j is represented by a point at (r_i, r_j) . The intuition is that a class which has a long lifetime is far from the diagonal while a class with a short lifetime is close. In many cases, a long lifetime loop implies that there is some sort of inherent topological feature being found, and thus that this point far from the diagonal is important, while short lifetime loops are likely caused by topological noise due to sampling or other errors in the system. In the example of Fig. 2, there is a prominent off-diagonal point which shows that the function defined by the matrix has a circular feature. Thus, a common measure for looking at the persistence diagram when investigating a single, circular structure is the maximum persistence, defined as

$$\text{MaxPers}(D) = \max_{(r_i, r_j) \in D} r_j - r_i$$

for a given persistence diagram, D .

E. Mathematical morphology

Mathematical morphology is a broad field based on analyzing shapes of objects using mathematical tools from areas

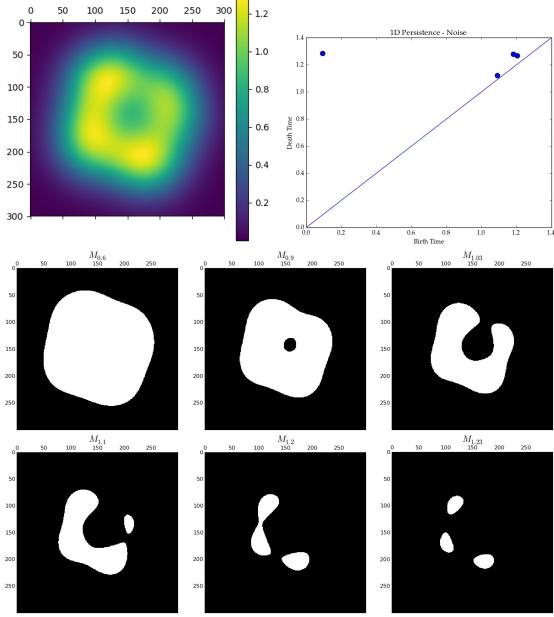


Fig. 2. An example matrix, M (top left) and corresponding persistence diagram (top right). Second and third row: The black portions are sublevel sets, M_r , where $r = 0.6, 0.9, 1.03, 1.1, 1.2$ and 1.23 . The existence of a point far from the diagonal in the persistence diagram shows that there is a prominent circular structure; while the other points are caused by the noise in the circle.

including set theory and geometry. This field is particularly useful for analyzing geometric structure in images and image processing. Specific applications include classification of digital images of cancerous tissue [30], restoration of old films [31], and ridge detection in finger prints [32].

Two major tools in mathematical morphology are erosion and dilation. Both tools involve a kernel moving through a binary image. In erosion, a pixel in the original image will remain a 1 only if all pixels under the kernel are 1's, otherwise it becomes a 0. This process removes small clusters of pixels, often considered noise, and pixels near the boundary.

Dilation is the opposite of erosion. A kernel moves through the binary image and a pixel is assigned a 1 if at least one pixel under the kernel is a 1, otherwise it is assigned a 0. Therefore, erosion followed by dilation will remove noise and rebuild the area around the boundary. This process of erosion followed by dilation is called opening. We use opening in Sec. V-C to test the influence of noise on our method. The choice of kernel for these methods can vary in size and shape depending on the application. For a more intensive explanation of these methods as well as the mathematical properties, we direct the reader to [33].

Opening is included in the python module `cv2`. Opening is specifically implemented using the function `cv2.morphologyEx` using `cv2.MORPH_OPEN` as the second input.

F. Fourier transform

The Fourier transform is a common method for investigating the periodicity of time series. It does so by decomposing a wave into a sum of sinusoids with different frequencies. We

will provide a short description of the Fourier transform; for a more detailed explanation, see [34]. Given a real valued function $f: \mathbb{R} \rightarrow \mathbb{R}$, the Fourier transform is

$$f(v) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i vt} dt.$$

This converts a function from the time domain to the frequency domain. In particular, when working with discrete data, we use the discrete Fourier transform. Let T be the time between discrete samples, then let $t_k = k * T$ where $k = 1, \dots, N - 1$. Then the discrete Fourier transform is

$$F_n = \sum_{k=1}^{N-1} f(t_k) e^{-2\pi i nk/N}.$$

The inverse discrete Fourier transform can then be calculated as

$$f(t_k) = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{2\pi i kn/N}.$$

The discrete Fourier transform reveals periodic components of the input data as well as the strength of each periodic component.

The power spectrum can be estimated using the discrete Fourier transform by calculating the square of the absolute value of the Fourier transform, $|F_n|^2$. Plotting this gives a visualization of the strength of each frequency of the periodic components in the input data. See Fig. 5 for an example. If the strongest frequency is f_k , then the period for this component is $1/f_k$.

IV. METHOD

The data was given in the form of two sets of storm-centered GOES IR ($10.7 \mu\text{m}$) satellite imagery. These two data sets have the same native spatial resolution, but differ in temporal resolution. The first set (hereafter the GOES-12 dataset), utilizes brightness temperatures derived directly from GOES-12 4-km IR satellite imagery and consists of data in hourly increments, spanning 2 to 4 September 2007 with the exception of 0415 UTC and 0515 UTC each day (due to the GOES-12 satellite eclipse period). Imagery was remapped such that each pixel has a spatial resolution of 2 km and each image covers a total area of approximately $1500 \text{ km} \times 1500 \text{ km}$. This is represented as a 752×752 matrix. The second data set is the GridSat-GOES and consists of data in 3-hour increments, spanning 31 August to 6 September 2007 with the exception of 0600 UTC each day [35]. Each pixel has a resolution of approximately 8 km and each image covers a total area of approximately $2400 \text{ km} \times 2400 \text{ km}$. This is represented by a 301×301 matrix. This data is cropped to a 191×191 matrix to approximately match the area covered by the first set of data. The cropped version covers a total area of approximately $1530 \text{ km} \times 1530 \text{ km}$.

The GridSat-GOES data set requires some additional processing. A different normalization is used with this data; thus, in order to convert it, the following equation is applied to the GridSat-GOES brightness temperatures

$$\frac{(\text{Original} \cdot 0.01 + 200.0) - 22.858}{0.919565}.$$

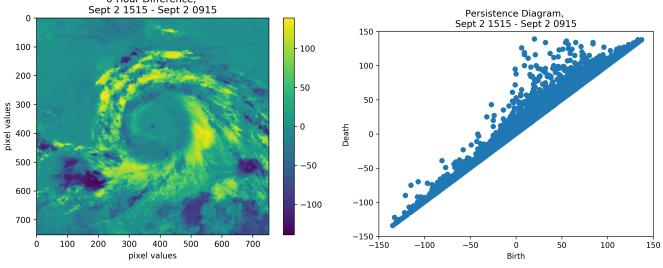


Fig. 3. Hurricane 6-hour time difference, $M(t)$, (left), and corresponding persistence diagram (right). The persistence diagram does not show any significant off diagonal point, thus it is not picking up the circular structure visible in the image.

Some images in the GridSat-GOES data set also contain missing values where the brightness temperature for certain pixels was not recorded and is instead assigned a fill value. In order to prevent these values from impacting our results, we interpolate values for these pixels. For a given pixel with a missing value, we compute the average value of a 5×5 grid centered at the pixel, not including the pixels in this range that also have missing values.

The TC diurnal pulse is propagating outward through the day; thus, in order to see the movement and changes in the GOES satellite brightness temperature, we consider the difference in matrices six hours apart [10], [15]. For all times t , given the original brightness temperature image $S(t)$, we compute the six-hour differences, $M(t) = S(t+6) - S(t)$; see Fig. 3 for an example. While circular features are visually prominent in the data, simply using persistence on the difference data, as discussed in Sec. III, did not show any relevant features due to the extreme differences in the function values between the circular sections. The right side of Fig. 3 shows the corresponding persistence diagram where no significant circular structure is detected.

Thus, we define a new function on the difference matrix using the following method. Fix a threshold μ and let $M(t)_\mu$ be the subset of $M(t)$ which has function value less than μ . This method results in a binary matrix defined entry-wise as,

$$M(t)_\mu[i, j] = \begin{cases} 1 & \text{if } M(t)[i, j] < \mu \\ 0 & \text{if } M(t)[i, j] \geq \mu. \end{cases}$$

We will address this choice of threshold in Sec. V-B; however, we will focus on the case $\mu = 70$ degrees for most of our analysis. Note that because $M(t)$ is a difference of two images, the threshold is not isolating all pixels above a certain temperature, but rather those pixels that increase in value by at least 70 degrees over the six hours.

We apply the distance transform to this binary matrix, as described in Section III-B, giving a new matrix $D(t)$.

To calculate the distance transform, we use the python submodule `scipy.ndimage`, specifically the function `distance_transform_cdt` with the chessboard metric. Therefore, each entry in $D(t)$ corresponds to the minimal distance to an entry where $M(t)[i, j] \geq \mu$. The distance transform $D(t)$ is then scaled by the resolution for each dataset in order to convert the units to kilometers instead of pixels. For

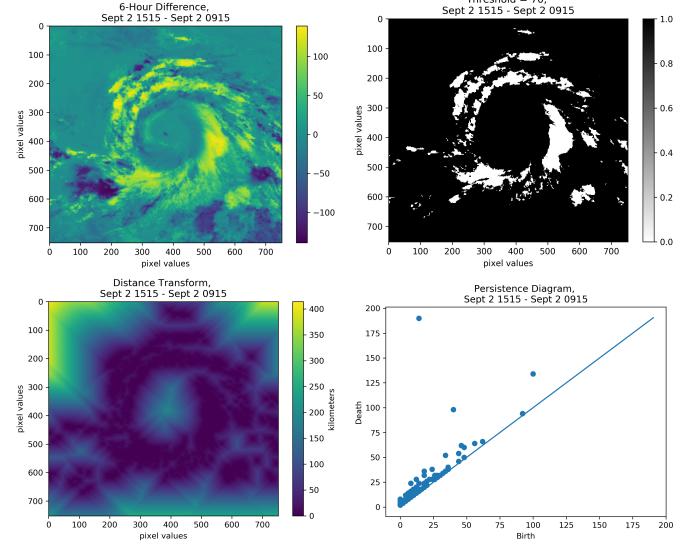


Fig. 4. Example of 6 hour difference, $M(t)$ (top left), thresholded subset, $M(t)_\mu$ where $\mu = 70$ (top right), distance transform function (bottom left) and corresponding persistence diagram (bottom right). This example comes from the GOES-12 data set.

the GridSat-GOES data, we scale by a factor of 8 km/pixel and for the GOES-12 data, we scale by a factor of 2 km/pixel. We then compute sub-level set persistence on the function $D(t)$ using the `cubtop` method in Perseus [21], [36]. Fig. 4 shows an example of each step in the method.

For each six-hour difference in each data set, we apply the steps described above, then calculate maximum persistence as defined in Sec. III-D. By plotting maximum persistence over time, we can see how the most prominent circular feature changes through the progression of the day and life of the TC. The red and blue solid lines in Fig. 6 represent this plot for both data sets. The plot shows an oscillatory pattern for both data sets that appears to repeat approximately daily.

V. RESULTS

A. Maximum persistence

In order to quantify the oscillatory pattern seen in Fig. 6, we use the discrete Fourier transform (DFT). The DFT is used to calculate the most prominent frequency in the data, in order to determine how often the cyclic behavior repeats. Note, to use the discrete Fourier transform the time steps must be equal; however, because of the missing times in our data, this is not the case. Therefore, we approximate the maximum persistence at these values by adding a point along the line between the times immediately before and after the missing time. Additionally, we truncate the maximum persistence to only include the days where we have the data for the entire day. This means truncating the GridSat-GOES data set to include only 1-4 September, and the GOES-12 data set to include only 1-2 September. The discrete Fourier transform was calculated using the python submodule `numpy.fft`.

We first calculate the Fourier transform using the function `fft`, then calculate the frequency bins using `fftfreq`. Using this information we plot the approximate power spectrum (the

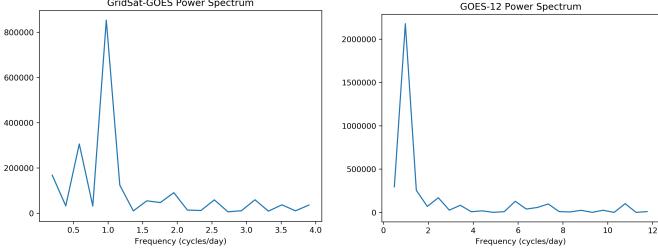


Fig. 5. Power spectrum for each data set. The highest peak on the left plot occurs at approximately 0.976 cycles/day, while the peak on the right occurs at approximately 0.979 cycles/day.

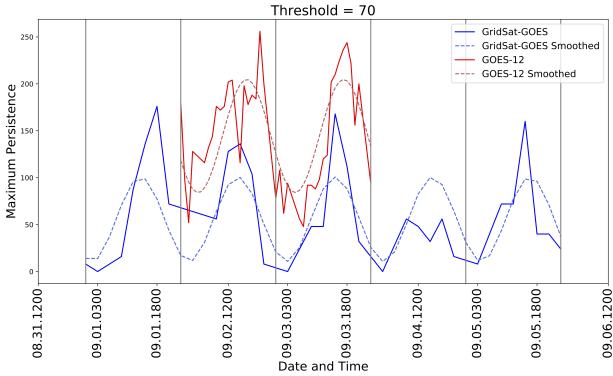


Fig. 6. Maximum persistence plotted over time for both data sets using threshold $\mu = 70$ in addition to the smoothed versions, created using inverse Fourier transform. Gray vertical lines separate days according to UTC.

square of the absolute value of the amplitude) for each data set, as can be seen in Fig. 5. Note, our data is all real, so the power spectrum will be symmetric for positive and negative frequencies; therefore, we only need to look at the positive frequencies. Picking the frequencies corresponding to the highest peaks in each of the power spectrums gives a frequency of 0.976 cycles/day for the GridSat-GOES data set, and 0.979 cycles/day for the GOES-12 data set. This means that the cycle is repeating every 24.6 hours for the GridSat-GOES data set, and every 24.5 hours for the GOES-12 data set.

Using the most prominent frequency for each data set, we calculate the inverse Fourier transform using `ifft` and plot these smoothed versions over the original data in Fig. 6. These smoothed curves closely resemble the patterns exhibited by the original maximum persistence versus time plots; therefore, these approximately 24 hour patterns seen in the plots are actually the most prominent signals in the data. This verifies the claim that our method is detecting a daily cycle.

B. Choice of threshold

We claim that the choice of threshold is not difficult and that the approximately daily cycle will be detected for a variety of thresholds. In Fig. 7 are surfaces that represent maximum persistence versus time for thresholds, $\mu \in \{25, 30, \dots, 75\}$. There is a clear periodic pattern for both data sets regardless of the threshold chosen in this range. In fact, for all thresholds tested between 35 and 75, the period is consistent at 24.6 hours

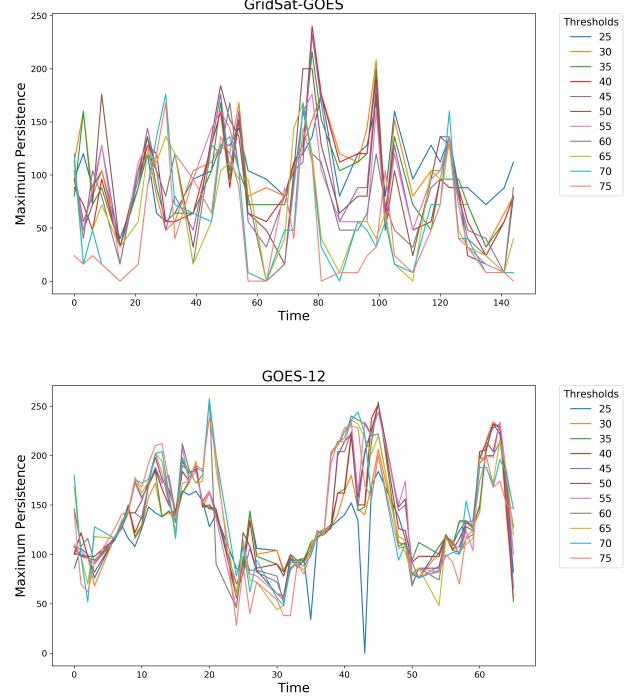


Fig. 7. Maximum persistence vs time plot for all thresholds $\mu \in [25, 75]$ such that $\mu = 5n$ where $n \in \mathbb{N}$.

for the GridSat-GOES data set and 24.5 hours for the GOES-12 data set. For $\mu = 25, 30$, the Fourier transform is unable to pick up the daily pattern in the GridSat-GOES data set, and instead just sees one period for the entire five days. Because a wide range of threshold values consistently detect a daily pattern, we conclude the choice of threshold does not need to be precise.

C. Removal of Noise

While the above method detects a daily cycle, there are some instances where the six-hour differencing introduces noise because of varying behavior in the center of the hurricane. The top images in Fig. 8 shows an example of this. There is a small area of pixels in the center of the image that are above the threshold, causing the distance transform to fill in the center of the circular region. This dot in the center will change the value of the maximum persistence; therefore, we used the method of opening described in Sec. III-E using a 8×8 pixel kernel for the GOES-12 data set and a 2×2 pixel kernel for the GridSat-GOES data set to remove noise such as these center pixels. Note, the difference in size of the kernel is due to the differences in spacial resolution between the two data sets. We use the python module `cv2` for these computations. Opening is specifically implemented using the function `cv2.morphologyEx` using `cv2.MORPH_OPEN` as the second input. The bottom images in Fig. 8 show the results when opening is used on the thresholded matrix and then the distance transform is applied. The center pixels are removed so the distance transform is no longer filled in.

Using this extra step in the method, we recalculate maximum persistence for all times and the estimated period using

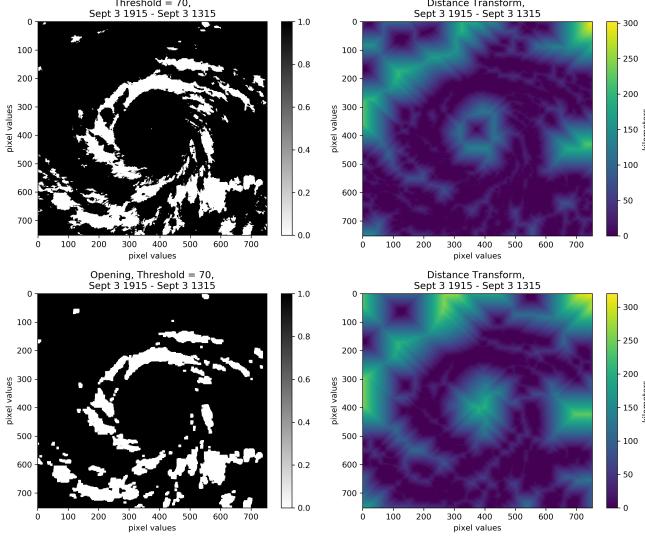


Fig. 8. The top two images are an example of the threshold and the corresponding distance transform where noise in the center of the hurricane causes the distance transform to fill in. The bottom two images are an example of the threshold after opening and the corresponding distance transform. This example comes from the GOES-12 data set. Note, the noise in the center of the hurricane is so small it may not be visible in the image in the top left; however, it is clear in the distance transform on the top right that there are pixels above the threshold in that region.

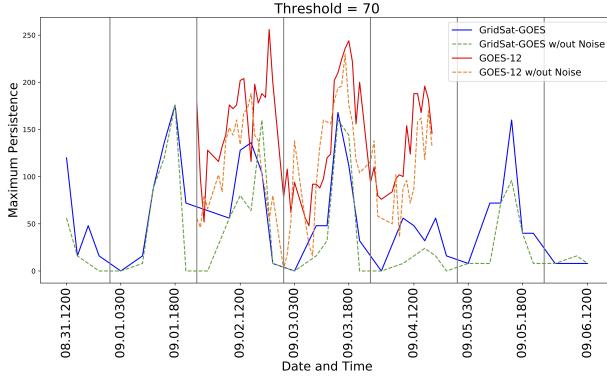


Fig. 9. Maximum persistence plotted over time for both data sets using threshold $\mu = 70$ in addition to the versions using opening to remove noise. Gray vertical lines separate days according to UTC.

Fourier transforms in the same manner described in Sec. V-A. Fig. 9 shows maximum persistence plotted vs time for the both data sets using our original method, and the method including the additional opening step. While the new maximum persistence values vary a little from the originals, the general oscillatory behavior remains unchanged. Using the Fourier transform, we get the same results as in Sec. V-A; the cycle still repeats every 24.6 hours for the GridSat-GOES data set, and every 24.5 hours for the GOES-12 data set. Thus, the noise is not affecting the periodicity we are detecting and the opening step is therefore unnecessary.

VI. DISCUSSION

This paper presents a novel method for detecting and analyzing the diurnal cycle of tropical cyclones using methods

from TDA. While current state of the art TC diurnal cycle measurement is mostly qualitative, our method provides a mathematically advanced method for automatic detection and measurement. While our method involves a choice of a parameter for the threshold, we present evidence that the choice is not important as there are many possible choices of thresholds that would yield the same results. Additionally, our method is robust to noise in the GOES satellite imagery as blurring the image does not yield different periodicity than the original method.

While this method was only tested on Hurricane Felix (2007), it was able to detect almost identical patterns in two sets of GOES satellite data for that hurricane that varied in spatial and temporal resolution. In the future, we hope to apply this analysis to more TCs and apply it to other satellite channels to further test our method.

ACKNOWLEDGMENT

The authors would like to thank William Dong, who was instrumental in creating code for the first iteration of this project. The work of ST and EM was supported in part by NSF grant DMS-1800446; EM was also supported in part by NSF CMMI-1800466. The work of JD was supported in part by the Office of Naval Research Program Element (PE) 0601153N Grant N000141410132 and by NOAA Grant NA14OAR4830172 from the Unmanned Aircraft Systems (UAS) Program. KC was supported by NSF Grant AGS-1636799.

REFERENCES

- [1] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, “Clique topology reveals intrinsic geometric structure in neural correlations,” *Proceedings of the National Academy of Sciences*, 2015. [Online]. Available: <http://www.pnas.org/content/early/2015/10/20/1506407112.abstract>
- [2] M. Nicolau, A. J. Levine, and G. Carlsson, “Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011. [Online]. Available: <http://www.pnas.org/content/108/17/7265.abstract>
- [3] J. A. Perea and J. Harer, “Sliding windows and persistence: An application of topological methods to signal analysis,” *Foundations of Computational Mathematics*, pp. 1–40, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s10208-014-9206-z>
- [4] A. Deckard, R. C. Anafi, J. B. Hogenesch, S. B. Haase, and J. Harer, “Design and analysis of large-scale biological rhythm studies: a comparison of algorithms for detecting periodic signals in biological data,” *Bioinformatics*, vol. 29, no. 24, pp. 3174–3180, 2013. [Online]. Available: <http://bioinformatics.oxfordjournals.org/content/29/24/3174.abstract>
- [5] F. A. Khasawneh and E. Munch, “Chatter detection in turning using persistent homology,” *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 527–541, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0888327015004598>
- [6] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, “On the local behavior of spaces of natural images,” *International Journal of Computer Vision*, vol. 76, pp. 1–12, 2008, 10.1007/s11263-007-0056-x. [Online]. Available: <http://dx.doi.org/10.1007/s11263-007-0056-x>
- [7] V. Robins, M. Saadatfar, O. Delgado-Friedrichs, and A. P. Sheppard, “Percolating length scales from topological persistence analysis of micro-CT images of porous materials,” *Water Resources Research*, vol. 52, no. 1, pp. 315–329, jan 2016.
- [8] M. Saadatfar, H. Takeuchi, V. Robins, N. Francois, and Y. Hiraoka, “Pore configuration landscape of granular crystallization,” *Nature Communications*, vol. 8, no. 15082, 2017.

- [9] V. Robins, P. J. Wood, and A. P. Sheppard, "Theory and algorithms for constructing discrete Morse complexes from grayscale digital images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1646–1658, Aug. 2011.
- [10] J. P. Dunion, C. D. Thorncroft, and C. S. Velden, "The tropical cyclone diurnal cycle of mature hurricanes," *Monthly Weather Review*, vol. 142, no. 10, pp. 3900–3919, 2014.
- [11] J. P. Dunion, C. D. Thorncroft, and D. S. Nolan, "Tropical cyclone diurnal cycle signals in a hurricane nature run," *Monthly Weather Review*, vol. 147, no. 1, pp. 363–388, 2019.
- [12] E. L. Navarro and G. J. Hakim, "Idealized numerical modeling of the diurnal cycle of tropical cyclones," *Journal of the Atmospheric Sciences*, vol. 73, no. 10, pp. 4189–4201, 2016.
- [13] M. E. O'Neill, D. Perez-Betancourt, and A. A. Wing, "Accessible environments for diurnal-period waves in simulated tropical cyclones," *Journal of the Atmospheric Sciences*, vol. 74, no. 8, pp. 2489–2502, 2017.
- [14] X. Tang and F. Zhang, "Impacts of the diurnal radiation cycle on the formation, intensity, and structure of hurricane edouard (2014)," *Journal of the Atmospheric Sciences*, vol. 73, no. 7, pp. 2871–2892, 2016.
- [15] S. D. Ditcheck, J. Molinari, K. L. Corbosiero, and R. G. Fovell, "An objective climatology of tropical cyclone diurnal pulses in the atlantic basin," *Monthly Weather Review*, no. 2018, 2018.
- [16] J. P. Kossin, "Daily hurricane variability inferred from goes infrared imagery," *Monthly Weather Review*, vol. 130, no. 9, pp. 2260–2270, 2002.
- [17] K. D. Leppert and D. J. Cecil, "Tropical cyclone diurnal cycle as observed by trmm," *Monthly weather review*, vol. 144, no. 8, pp. 2793–2808, 2016.
- [18] J. Steranka, E. B. Rodgers, and R. C. Gentry, "The diurnal variation of atlantic ocean tropical cyclone aoud distributioninferred from geostationary satellite infrared measurements," *Monthly weather review*, vol. 112, no. 11, pp. 2338–2344, 1984.
- [19] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational homology*. Springer Science & Business Media, 2006, vol. 157.
- [20] A. Hatcher, *Algebraic Topology*. Cambridge University Press, 2002.
- [21] K. Mischaikow and V. Nanda, "Morse theory for filtrations and efficient computation of persistent homology," *Discrete & Computational Geometry*, vol. 50, no. 2, pp. 330–353, September 2013.
- [22] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. American Mathematical Society, 2010.
- [23] E. Munch, "A user's guide to topological data analysis," *Journal of Learning Analytics*, vol. 4, no. 2, 2017.
- [24] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *J. ACM*, vol. 13, no. 4, pp. 471–494, Oct. 1966. [Online]. Available: <http://doi.acm.org/10.1145/321356.321357>
- [25] A. Rosenfeld and J. Pfaltz, "Distance functions on digital pictures," *Pattern Recognition*, vol. 1, no. 1, pp. 33 – 61, 1968. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0031320368900137>
- [26] Y. T. Chin, H. Wang, L. P. Tay, H. Wang, and W. Y. Soh, "Vision guided agv using distance transform," in *Proceedings of the 32nd ISR (International Symposium on Robotics)*, vol. 19. Citeseer, 2001, p. 21.
- [27] C. Arcelli and G. S. di Baja, "Computing voronoi diagrams in digital pictures," *Pattern Recognition Letters*, vol. 4, no. 5, pp. 383 – 389, 1986. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0167865586900607>
- [28] G. Borgefors, "Distance transformations in digital images," *Computer Vision, Graphics, and Image Processing*, vol. 34, no. 3, pp. 344 – 371, 1986. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0734189X86800470>
- [29] J. R. Munkres, *Elements of Algebraic Topology*. Addison Wesley, 1993.
- [30] J. Thiran and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *IEEE Transactions on biomedical engineering*, vol. 43, no. EPFL-ARTICLE-86388, pp. 1011–1020, 1996.
- [31] L. Joyeux, S. Boukir, B. Besserer, and O. Buisson, "Reconstruction of degraded image sequences. application to film restoration," *Image and Vision Computing*, vol. 19, no. 8, pp. 503 – 516, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885600000913>
- [32] C. D. Lake, R. M. Lougheed, and J. H. Beyer, "Morphological algorithm for ridge extraction in fingerprint images," in *Image Algebra and Morphological Image Processing IV*, E. R. Dougherty, P. D. Gader, and J. C. Serra, Eds. SPIE, June 1993.
- [33] J. Serra, *Image Analysis and Mathematical Morphology*, ser. Image Analysis and Mathematical Morphology. Academic Press, 1984, no. v. 1. [Online]. Available: <https://books.google.com/books?id=kfY-ngEACAAJ>
- [34] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.
- [35] K. R. Knapp and S. L. Wilkins, "Gridded satellite (gridsat) goes and conus data," *Earth System Science Data*, vol. 10, no. 3, pp. 1417–1425, 2018.
- [36] V. Nanda, "Perseus," <http://www.sas.upenn.edu/~vnanda/perseus/>, July 2015.

Sarah Tymochko is a PhD student at Michigan State University in the Department of Computational Mathematics, Science, and Engineering. Her primary research interests include topological data analysis and machine learning. Previously, she received a Bachelor of Arts in Mathematics from College of the Holy Cross in Worcester, MA.

Elizabeth Munch received her PhD in Mathematics from Duke University in 2013. Her research is in topological data analysis and its applications. Prior to joining the faculty of Michigan State University, she was an Assistant Professor in the Department of Mathematics and Statistics at the University at Albany - SUNY, and a Postdoctoral Fellow at the Institute for Mathematics and its Applications at the University of Minnesota for the 2013-2014 thematic year on applications of topology. She also holds a Master of Arts in Mathematics from Duke University, a Bachelor of Science in Mathematics from the University of Rochester, and a Bachelor of Music in Harp Performance from the Eastman School of Music.

Jason Dunion received his PhD in Atmospheric Science from the University at Albany-SUNY in 2016. His research interests include development of satellite remote sensing tools for monitoring tropical cyclones and Saharan dust storms, climatology of the tropical Atlantic atmosphere and tropical cyclogenesis. He also holds a Master of Science in Atmospheric and Oceanic Science from the University of Wisconsin-Madison and a Bachelor of Arts in Geography and Geology from the University of New Hampshire.

Kristen Corbosiero received her PhD in Atmospheric Science from the University at Albany in 2005. Her research interests include tropical cyclone structure and intensity change, lightning observation systems, and the North American Monsoon System. Kristen was a faculty member at the University of California Los Angeles from 2007-2011, holds a Master of Science from the University at Albany and a Bachelor of Science with Distinction from Cornell University.

Ryan Torn is a Professor in the Department of Atmospheric & Environmental Sciences at the University at Albany. He received his PhD from the University of Washington.