RESEARCH ARTICLE

Topological Analysis and the Impact of Ecological Anomalies on Sea Turtle Hatching Success

Madukpe Vine Nwabuisi[a], Nur Fariha Syaqina Mohd Zulkepli[a], Ummu Atiqah Mohd Roslan[b],*, Mohd Uzair Rusli[c]

[a]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia; [b]Special Interest Group on Mathematical Modelling and Data Analytics, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia; [c]Sea Turtle Research Unit (SEATRU), Institute of Oceanography and Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

Abstract Sea turtle hatching success is influenced by a complex combination of environmental, biological, and anthropogenic factors, making it essential to understand its dynamics for effective conservation planning. This study utilizes Ball Mapper, a topological data analysis (TDA) tool, alongside the Isolation Forest anomaly detection algorithm to investigate 10 years (2013–2023) of high-dimensional ecological data from a coastal sea turtle nesting site. The dataset includes monthly records of hatch rates, predator activity, fungal presence, and flooding events. The Ball Mapper topological graphs revealed consistent seasonal trends, with April, May, and June exhibiting the highest hatching success across years. Meanwhile, months like January and February showed consistently lower outcomes and shared structural similarities in the data topology. Isolation Forest identified months with extreme ecological stressors as anomalies; however, months such as May and June 2016 still achieved high hatching success, suggesting that the presence of some predators may have also played a role in natural biocontrol. The TDA and anomaly-based approaches provided a better understanding of the complex relationships driving hatching success, uncovering patterns not easily detected by conventional methods. By visualizing temporal and ecological variation in hatching outcomes, the research supports data-driven strategies to enhance sea turtle conservation in the face of increasing environmental variability and ecological pressure.

Keywords: Topological data analysis, sea turtle, hatching success.

Introduction

Multiple ecological, biological, and human-driven factors affect sea turtle hatching success. This early life stage is a critical determinant of population dynamics and species survival, especially given that most sea turtle species are currently classified as endangered or critically endangered globally [1,2]. Numerous biotic and abiotic factors, including temperature, predation, moisture levels, and human activities, affect hatching rates. Environmental stressors like coastal erosion, flooding, and pollutant exposure are becoming more prevalent, intensified by climate change and increased coastal development [3]. Understanding these dynamics is essential for conservation strategies, as hatchling survival directly influences juvenile recruitment and the long-term viability of sea turtle populations [4].

Field studies, such as Martins *et al.* [5], report that hatchery practices, egg viability, and environmental conditions like shading and moisture levels contribute to significant variability in hatching success. Salleh *et al.* [6] and Jani *et al.* [7] emphasized that egg exploitation, habitat degradation, and poor conservation governance have drastically reduced turtle numbers in Malaysia. These patterns are not limited to one

region; globally, sea turtle nesting is being altered by regional-scale coastal characteristics like sea surface temperatures, tidal surges, and proximity to seagrass beds, which define both nesting opportunities and hazards [2]. Quantitative approaches, especially mathematical modeling, have been pivotal in assessing sea turtle dynamics. Traditional models such as stage-structured prey-predator systems use ordinary differential equations (ODEs) to simulate interactions between eggs, hatchlings, and predators under variable ecological pressures [8].

However, these models often assume linearity and stationarity, limiting their effectiveness in representing ecological variability. Recently, advanced numerical methods enhanced by neural networks have improved predictive accuracy in simulating hatching success and interactions with anthropogenic stressors [9,10]. Field-based studies also continue to provide valuable insights. Long *et al*. [11] investigate predation pressures and evaluate management interventions like physical nest barriers and artificial incubation. Standard monitoring methods, including nest excavations and emergence counts, remain widely used but face constraints due to high spatiotemporal variability, climatic unpredictability, and species-specific nesting behaviors [3,12]. Consequently, traditional models may overlook nonlinear patterns and hidden relationships, especially in high-dimensional ecological data. To address this, the present study seeks to employ a qualitative approach, the Ball Mapper (BM) algorithm, a topological data analysis tool, to visualize complex associations in sea turtle hatching data.

The BM algorithm, introduced by Dłotko [13], enables researchers to visualize relationships within high-dimensional data through a topological graph. The graph is constructed using a single parameter, the ball radius, to cluster similar data points into nodes and connect nodes with shared points, generating a visual representation. This structure preserves the underlying shape of the dataset, making it easier to detect patterns. It has been successfully applied in environmental studies [14,15] and biological imaging [16]. In this study, BM is employed to uncover hidden ecological dynamics to identify structural patterns in sea turtle hatching success. To complement the BM, the Isolation Forest algorithm is utilized to detect anomalies in predator activity, ecological conditions, and their impact on the hatching success. This unsupervised machine learning model isolates outliers by randomly selecting features and split values, effectively distinguishing abnormal data points in high-dimensional, noisy environments. Its strengths lie in its robustness, efficiency, and minimal assumptions about data distribution [17]. Recently, Binetti *et al*. [18] applied Isolation Forest in remote sensing contexts with ecological spectral indices to detect environmental anomalies, including wildfire scars and quarry expansion.

Previously, BM algorithm has been applied in the context of air pollution [14]. As far as we concerned, BM algorithm is not yet applied to ecological data. Therefore, this study aims to apply the BM algorithm to uncover hidden, dynamic relationships between sea turtle hatching success and the seasonal or ecological variables that influence it. By preserving the topological structure of high-dimensional data, BM provides a powerful means to visualize and interpret complex ecological interactions that are often overlooked by traditional analytical methods. Additionally, the Isolation Forest anomaly detection algorithm is used to identify and quantify abnormal predator patterns that may significantly affect hatching outcomes. These data-driven tools uncover subtle patterns and rare ecological events that remain hidden using traditional methods. Insights gained from this approach are expected to enhance conservation planning, improve the effectiveness of habitat management strategies, and support adaptive responses to emerging environmental challenges, ultimately contributing to the protection and resilience of sea turtle populations in a rapidly changing world.

Materials and Methods

Data Collection

This study was conducted at the Chagar Hutang Turtle Sanctuary (CHTS), Redang Island, Malaysia, a key nesting site managed by the Sea Turtle Research Unit (SEATRU) of Universiti Malaysia Terengganu since 1993. The sanctuary operates as a long-term conservation and research station, supported by a dedicated team of rangers, biologists, and volunteers who maintain continuous biological monitoring throughout the nesting season. Each year, CHTS receives approximately 1,600–2,000 green turtle (*Chelonia mydas*) and hawksbill turtle (*Eretmochelys imbricata*) nests, with nesting commencing as early as March and concluding in late October.

Data collection at CHTS is structured into two main categories: (i) nesting data, and (ii) hatching data. The nesting dataset includes morphological measurements of nesting females, location or sector of nests, and associated environmental parameters such as shading, nest depth, and proximity to vegetation. The hatching dataset focuses specifically on hatching success, incorporating parameters

such as total number of eggs, hatched and unhatched eggs, and observed signs of disturbance or mortality factors.

All nests were monitored *in situ* and excavated following completion of incubation, which typically lasts 55–70 days depending on microhabitat conditions. Nests under shaded canopies generally incubate longer than those in open, sunlit areas. Beginning at 45 days of incubation, each nest was inspected by trained staff to assess hatching progress. If no emergence or hatchlings were detected, the nest was rechecked after three days. In cases where live hatchlings were observed partially emerged within the sand column, the nest was left undisturbed and reburied to allow natural emergence.

Once hatchling emergence was confirmed, or the incubation period had concluded without emergence, the nest was completely excavated. Eggshells were counted to determine the number of successfully hatched eggs, and unhatched eggs were categorised by developmental stage where possible. Any evidence of nest disturbance or mortality factors was recorded, including signs of predation by ants, crabs, maggots, or monitor lizards; fungal colonisation; root intrusion; flooding; maternal disturbance; and physical damage to eggs.

Hatching success was calculated as the percentage of eggs producing live hatchlings that successfully emerged, following the formula:

$$\text{Hatching Success (\%)} = \frac{\text{Number of Hatched Eggs}}{\text{Total Number of Eggs}} \times 100$$

Ball Mapper Algorithm

The BM algorithm, introduced by Dlotko [13], is a tool in Topological Data Analysis (TDA) and a variant of the conventional Mapper algorithm. The BM algorithm groups data points into subsets using a collection of closed balls, determined by a single parameter $\varepsilon > 0$, the ball radius. The set of balls iteratively covers the entire point cloud through the Max-min algorithm method, formed by an $\varepsilon - net$. This ball acts as a covering tool, clustering data points based on proximity. In the BM algorithm, nodes represent the points clustered within each ball, and edges are drawn between nodes when their corresponding balls overlap, creating a topological graph. By focusing only on the ball radius, this approach reduces the complexity of managing multiple parameters, making the algorithm easier to implement while still preserving the topological structure of the dataset. Figure 1 presents a flowchart illustrating the implementation process of the BM algorithm on a point cloud.
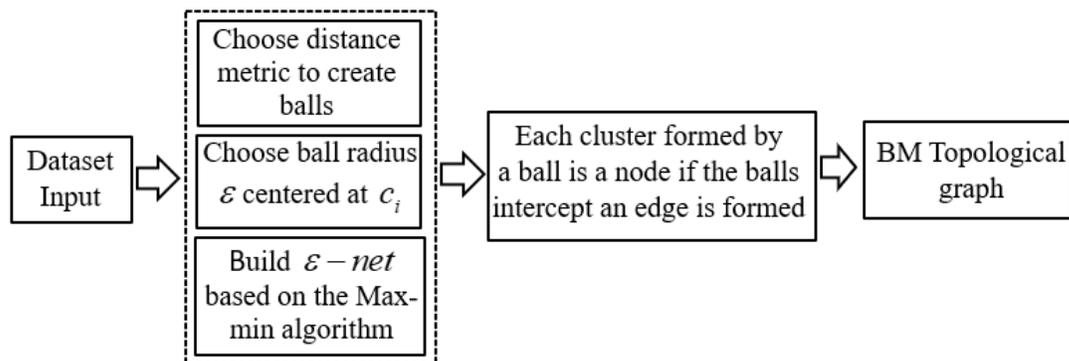

Figure 1. Ball Mapper flow chart

Given a dataset or point cloud X, the mathematical expression for the BM algorithm is given as follows:

i.   Cover: a point cloud X is covered by a collection of closed balls $B_\varepsilon(c_i) = \{x_i \in X : d(x_i, c_i) \le \varepsilon\}$, this implies that an net is built using a Max-min algorithm to form balls of radius $\varepsilon > 0$ centered at an arbitrary point c_i , where $\{x_i, c_i\} \in X$ and $i, j = \{1, \dots, n\}$. For every point $x_j \in X$ and $x_i \in X$ there exists a ball $B_\varepsilon(c_j) = \{x_j \in X : d(x_j, c_j) \le \varepsilon\}$ and $B_\varepsilon(c_i) = \{x_i \in X : d(x_i, c_i) \le \varepsilon\}$. Each ball is formed by covering points $x_i \in X$ within the distance:

$$d(x, c) = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2} \ , \tag{1.1}$$

where $x = (x_1, \dots, x_n)$ and $c = (c_1, \dots, c_n)$. Note that the distance between the points in a ball is within the threshold:

$$d(x_i, c_i) \leq \varepsilon. \tag{1.2}$$

ii. Ball mapper graph: an abstract graph is constructed where the collection of balls, $\{B_\varepsilon(c_i), B_\varepsilon(c_j), \ldots, B_\varepsilon(c_n)\}$ covered the entire X. Each ball forms a node N and if $B_\varepsilon(c_i) \cap B_\varepsilon(c_j) \neq \emptyset$ then an edge E is formed connecting the nodes, hence the Ball mapper graph $G = (N, E)$.

Application of the BM to Hatching

Before the implementation of the BM algorithm, the variables in the dataset $X \subset \mathbb{R}^n$ are converted to point clouds or vectors according to their corresponding months as given in the example:

Monthly$(A_i) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$
$A_i \subset X$ where $x_i \in X$.
The variables are:
x_1 : Total hatched eggs
x_2 : Total unhatched eggs
x_3 : Number of nests
x_4 : Number of Crabs
x_5 : Number of Maggots
x_6 : Number of Ants
$x_1 + x_2$: Total number of eggs laid
x_7 : Hatching success rate
$$x_7 = \frac{x_1}{x_1 + x_2} \tag{1.3}$$

The numerical values of the vectors are normalized using a min-max scaler, and the standard Euclidean distance is applied to measure the proximity between points across different months. The goal is to illustrate how the predictive variables for hatching success, including hatched eggs, unhatched eggs, maggots, crabs, ants, and other predators, vary over the months. The size of the nodes is determined based on the specific variable of interest, allowing for a clearer visualization of how these predictive variables change over time. This is done using the following procedure: each node of the BM graph corresponds to a ball $B_\varepsilon(c_i)$, $c_i \in X$ and a landmark point in the point cloud for building $\epsilon -$net balls using the Max-Min algorithm. Therefore, $x_i = B_\varepsilon(c_i) \cap X$, where$X$ is the dataset. Balls covering the data points within a threshold $d(x_i, c_i) \leq \varepsilon$ (radius of the ball) are formed to cover data points corresponding to each month, if the distance between the months is within the same threshold, the balls overlap, resulting in the connectedness of such nodes. The value of each ball, or the size of the nodes in the BM graph, is determined by one of the predictive variables $(x_1, x_2, x_3, x_4, x_5, x_6)$ in this case, the predictive variable of interest is the number of hatched eggs. This study applies the BM algorithm using the pyBallMapper Python library [19] to visualize topological structures. This approach provides an effective way to analyze and represent the dataset in a simplified manner.

Isolation Forest Algorithm

The Isolation Forest (IF) algorithm is an unsupervised method for detecting anomalies based on the idea that anomalies are easier to isolate than normal points. It begins with random sampling to reduce computational load. Then, it constructs multiple isolation trees by randomly selecting features and splitting values between their maximum and minimum. This partitioning continues until points are isolated or a maximum tree depth is reached [20]. The path length $m(l(x))$ (the number of splits needed to isolate a point x) is key. An anomaly score $s(x, n)$ is calculated as:

$$s(x, n) = 2^{-\left(\frac{m(l(x))}{E(n)}\right)},$$

where $m(l(x))$ is the average path length over a collection of isolation trees, and $E(n)$ is the expected path length in a binary search tree of size n, approximated by:

$$E(n) = 2H(n-1) - \frac{2(n-1)}{n}, \tag{2.1}$$

with H_i being the harmonic number. Finally, a decision function is used to classify a point as an anomaly if:

$$D(x, n) = \begin{cases} 1, & if \ s(x, n) \geq \omega. \ (anomaly) \\ 0, & if \ otherwise. \quad (normal) \end{cases}' \tag{2.2}$$

where $\omega = 0.5$ is a threshold (the value is conventionally acceptable). Points with shorter path lengths and higher anomaly scores are classified as anomalies. Isolation Forest looks for records that are "easier to separate" from the crowd because they have strange values in features.

Results and Discussion

Each node in the BM graphs represents a month with a cluster of similar characteristics, such as predation factors, total unhatched eggs, and total hatched eggs. The size of each node reflects the total quantity of hatched eggs, while the node's color represents the hatching success rate based on the color bar; the color gradient from dark purple to yellow visually captures the variation in hatching success, with darker colors indicating lower success and lighter colors indicating higher success rate. Edges between nodes suggest that the months share similar hatching patterns, such as comparable predation factors, though not necessarily hatching success rates. Lastly, the numbers in the nodes correspond to the months (e.g., January, February, etc.), allowing for easy identification of months with various degrees of similarity, total hatched eggs, and hatching success rate. The following sections present and analyze the annual BM graphs and the ecological anomalies impacting the hatching success rate.

Hatching success BM graph for 2013 and 2014

Figure 2 presents the hatching success BM graphs for 2013 and 2014, showcasing variations in success patterns.
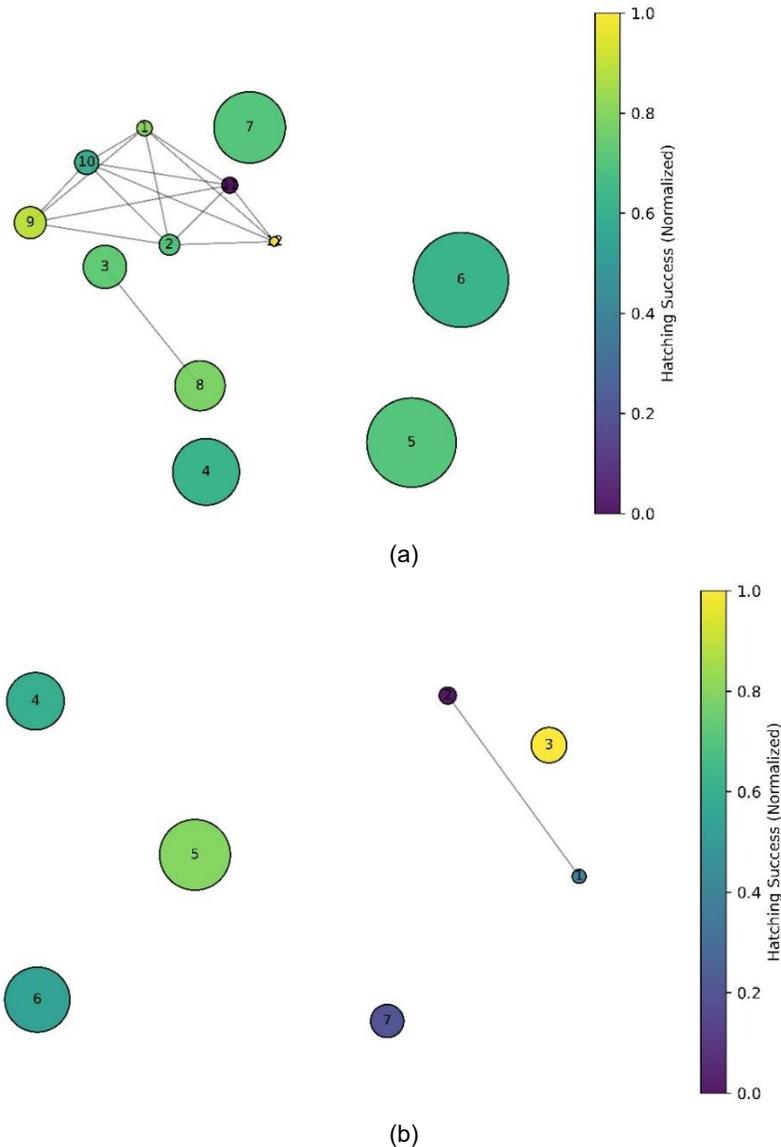

(a)

(b)

Figure 2. Hatching success BM graph for (a) 2013 and (b) 2014

From Figure 2a, the 2013 BM graph reveals distinct variations possibly influenced by environmental factors and predator presence. Nodes with high hatching success, such as Node 12, representing

December, have the highest hatching success rate, followed by Node 9, representing September. In contrast, Node 11 achieved the lowest hatching success, suggesting a high possible level of predator activity relative to the total number of eggs laid in November, while other months exhibit varied success rates. Connections between Nodes 1, 2, 9, 10, 11, and 12 representing January, February, September, October, November, and December 2013, respectively, as well as between Nodes 3 and 8, representing March and August 2013 reveal strong similarities in the number of predator and number of eggs hatched.

However, singleton nodes like Nodes 4, 5, 6, and 7 represent unique clusters with distinct predator activity and total hatched eggs. The graph also shows that the highest number of hatched eggs occurred between May and June 2013, with May achieving better hatching success. From Figure 2b, the 2014 BM graph highlights that the highest hatching success occurred in March, followed by May, while the lowest was recorded in February and July. The highest number of eggs hatched was observed in May, followed by June 2014. Apart from January and February, which displayed similarities, the other months analyzed exhibited distinct behaviors attributed to the presence of predators and other factors. Further analysis of the BM graphs for subsequent years is provided in the following sections.

Hatching Success BM Graph for 2015 and 2016

Figure 3 illustrates the hatching success BM graphs for 2015 and 2016, highlighting variations in success patterns.
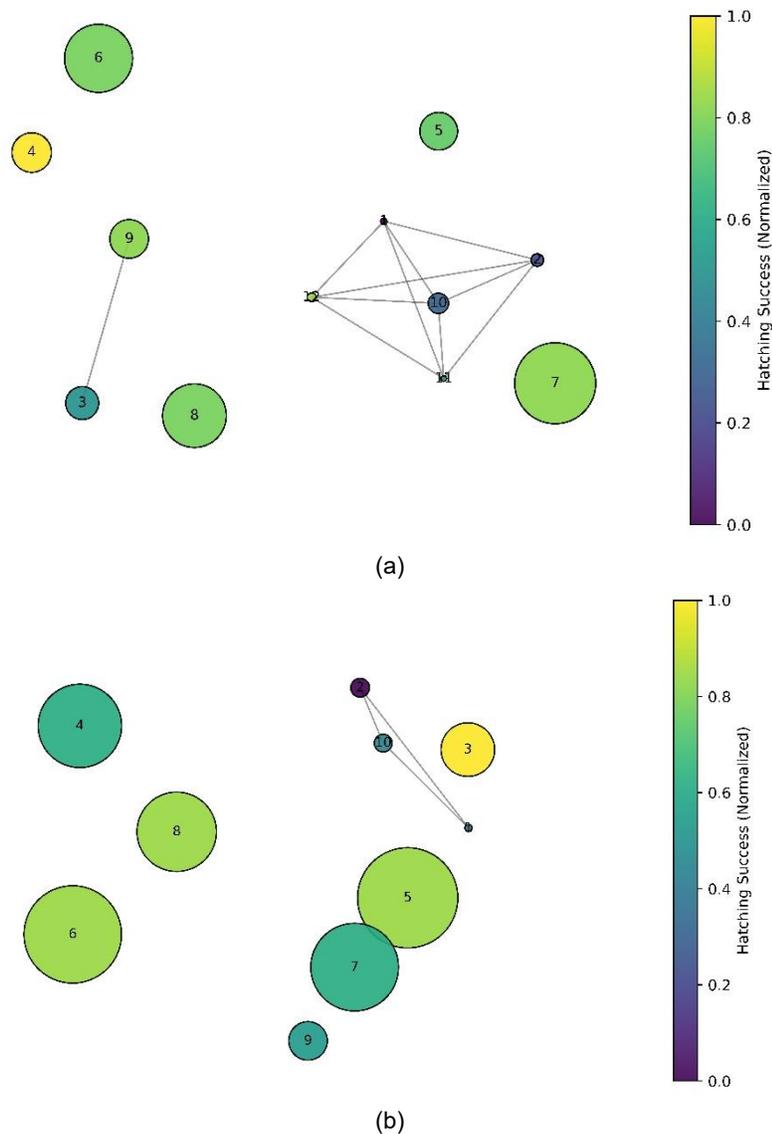

(a)

(b)

Figure 3 Hatching success BM graph for (a) 2015 and (b) 2016

In 2015, Node 4 showed the highest hatching success, reflecting minimal predator interference in April, followed by Nodes 7 and 12, while Nodes 1, 2, and 10 exhibited significantly low success rates, with January recording the lowest. Strong connections between Nodes 1, 2, 10, 11, and 12, as well as between Nodes 3 and 9, indicate similarities in predator activity and total number of eggs, whereas singleton nodes like Node 4 represent unique patterns. In 2016, Figure 3b, Node 3 demonstrated the highest hatching success in March, followed by May and June, which recorded the highest number of hatched eggs, while Node 2, representing February, showed the lowest hatching success. February shares similarities with January and October, as evidenced by connected nodes, while Nodes 4 and 3 appear more isolated, reflecting distinct patterns. The subsequent years' BM graphs are presented I the next section.

Hatching Success BM Graph for 2017 and 2018

Figure 4 presents the hatching success BM graphs for 2017 and 2018, highlighting variations in success patterns across the two years. In the 2017 graph (Figure 4a), Node 5, representing May 2017, demonstrates the highest hatching success and the largest number of eggs hatched, followed by Node 4, representing April 2017. Conversely, Node 9, corresponding to September 2017, shows the lowest hatching success. The nodes are largely isolated, with no connections observed, indicating distinct differences in the number of eggs hatched and the presence of predators across the months of 2017. In the 2018 graph (Figure 4b), Node 9, representing September 2018, recorded the highest hatching success, followed by Nodes 5 and 7. In contrast, Node 4, representing April 2018, shows the lowest hatching success, followed by January. A linear connection between Nodes 2 and 10 suggests similarities in predator activity or the number of hatched eggs for those months. In contrast, the remaining nodes are largely independent, reflecting unique patterns of predator influence and egg-hatching outcomes. Section 3.5 presents the BM hatching success graphs for the years 2019 and 2020.
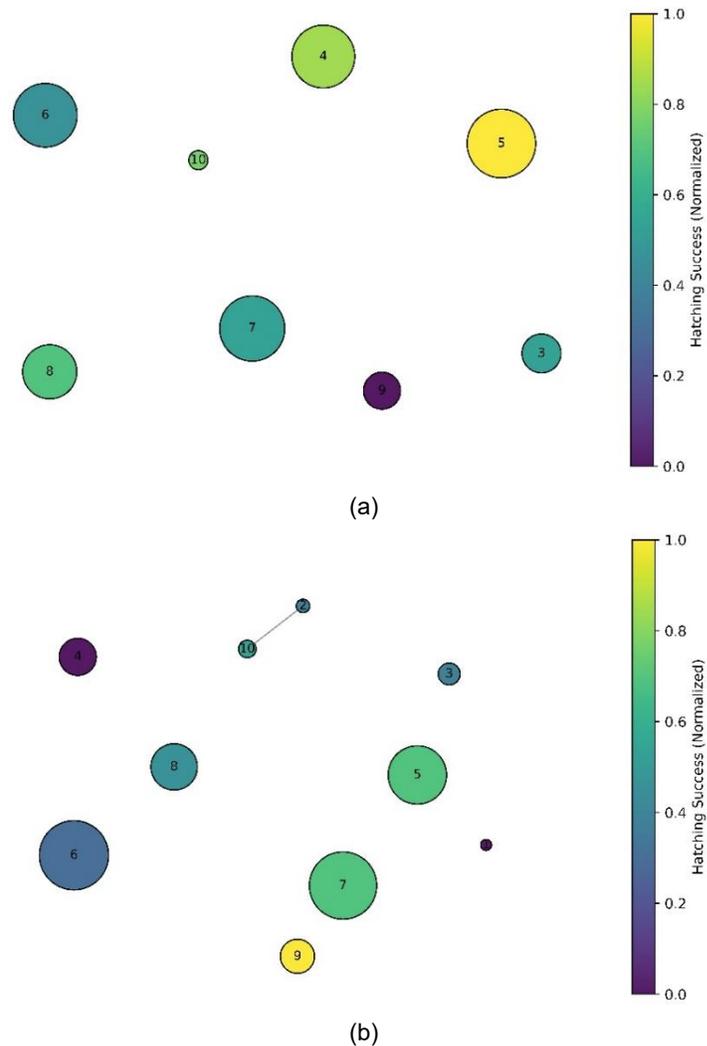

(a)

(b)

Figure 4 Hatching success BM graph for (a) 2017 and (b) 2018

Hatching Success BM Graph for 2019 and 2020

Figure 5 shows the hatching success BM graphs for 2019 and 2020. In the 2019 graph (Figure 5a), Node 10, representing October 2019, demonstrates the highest hatching success, followed by Nodes 4 and 7. In contrast, Nodes 8 and 1, representing August and January 2019, show the lowest hatching success. The connectedness between Nodes 1, 2, and 10 indicates similarities in predator activity or hatched eggs for these months, with October being more similar to February than January. However, other nodes, such as Node 6 with the highest number of hatched eggs, followed by May 2019, are largely isolated, reflecting distinct hatching patterns and predator influences. In the 2020 graph (Figure 5b), Node 10, representing October 2020, exhibits the highest hatching success. followed closely by Node 2, corresponding to February 2020. Similarly, Node 1, representing January 2020, shows the lowest hatching success. Notably, strong connections are observed between Nodes 1, 2, 10, and 11, suggesting shared patterns in predator activity or environmental conditions for these months. Other nodes remain more isolated, indicating unique ecological dynamics affecting hatching success during those months. The following section presents the BM graphs for the years 2021, 2022, and 2023.
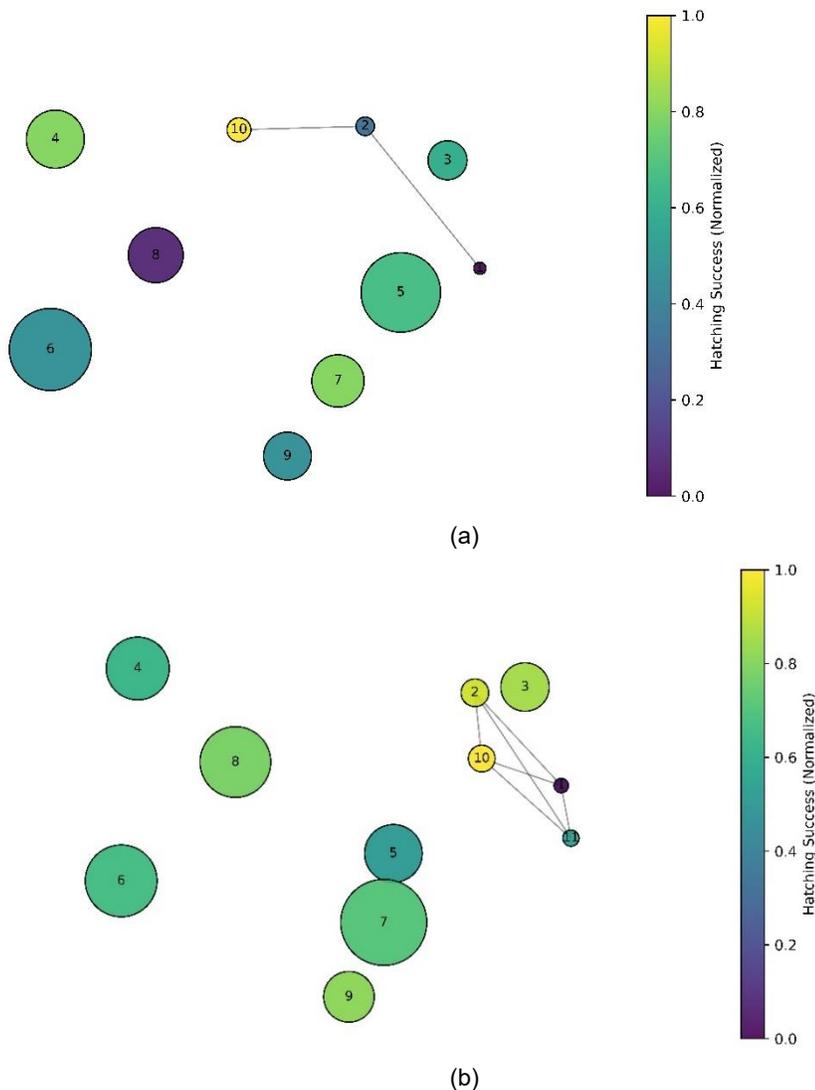

(a)

(b)

Figure 5 Hatching success BM graph for (a) 2019 and (b) 2020

Hatching Success BM Graph for 2021, 2022 and 2023

Figure 6 presents the hatching success BM graphs for 2021, 2022, and 2023, highlighting disparities in hatching success patterns. In the 2021 graph (Figure 6a), in Node 4, the highest hatching success is observed, while Nodes 1 and 2 show the lowest success rates. Similarities are observed between Nodes 1 and 10 and between Nodes 2 and 3. June records the highest number of hatched eggs, reflecting favorable conditions during that month. In the 2022 graph (Figure 6b), Node 1, representing January

2022, achieves the highest hatching success, while Node 2, representing February 2022, records the lowest success rate. Node 6 is notable for having the highest number of hatched eggs.
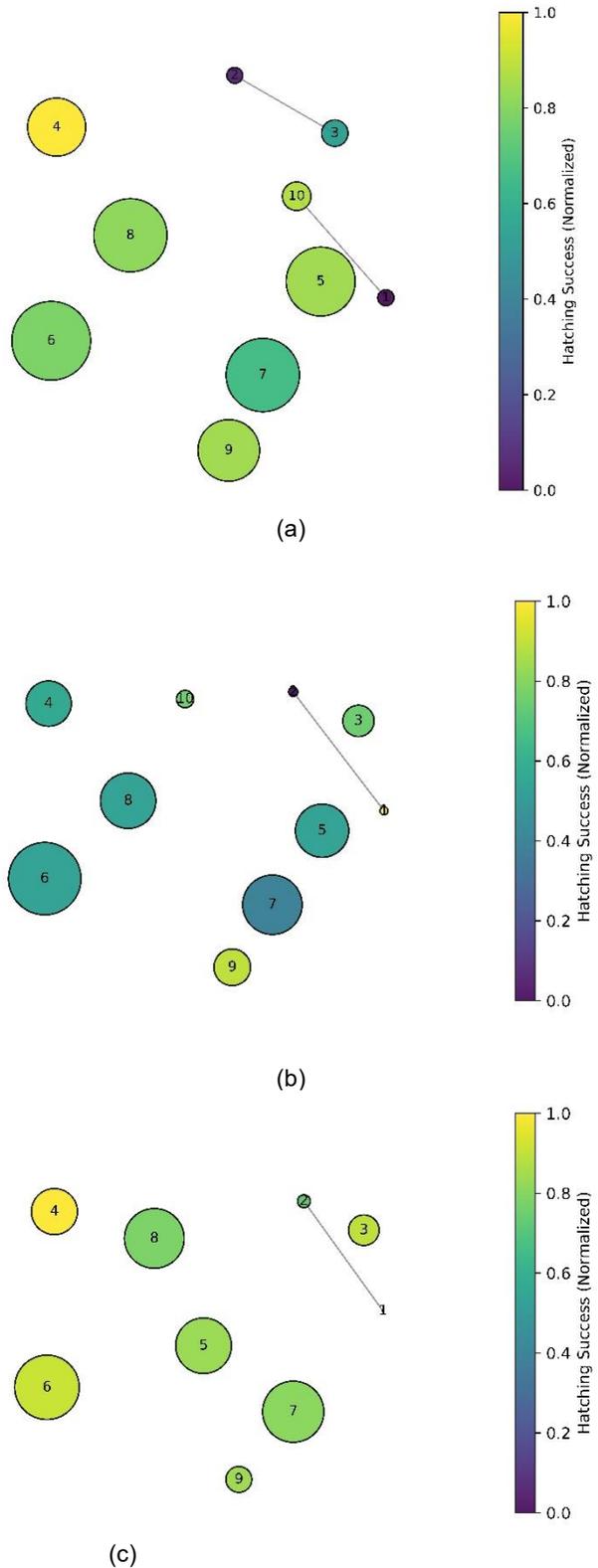

(a)

(b)

(c)

Figure 6 Hatching success BM graph for (a) 2021 (b) 2022 and (c) 2023

Finally, Figure 6c shows the 2023 graph, Node 6, corresponding to June 2023, records the highest number of hatched eggs, whereas Node 4, representing April 2023, achieves the highest hatching success. Similarities are noted between Nodes 1 and 2 in terms of patterns or conditions. Collectively, these graphs illustrate month-to-month variations in hatching success influenced by predator presence and environmental conditions, with certain months consistently demonstrating better outcomes than others. Section 3.6 presents a BM hatching success graph based on a combined dataset spanning 10 years, from 2013 to 2023.

Ten Years Combined Hatching Success BM Graph

Figure 7 presents the hatching success BM graph of the combined data set for 2013 to 2023, illustrating the hatching success patterns and relationships across the years. Nodes with consistently high hatching success, such as April, May, and June across multiple years, are prominent. Conversely, nodes like January, February, October, and November exhibit consistently lower hatching success. The connections between nodes reveal significant similarities in hatching patterns and predator presence across specific months and years. For instance, tightly clustered connections among nodes like January, February, and October across the years 2014, February 2018, and February 2020 suggest shared environmental or ecological conditions that influenced similar hatching outcomes. Additionally, the yearly topological graphs reveal that the highest hatching success volume was recorded during the first six months, with March, April, May, and June standing out as the most prominent months.
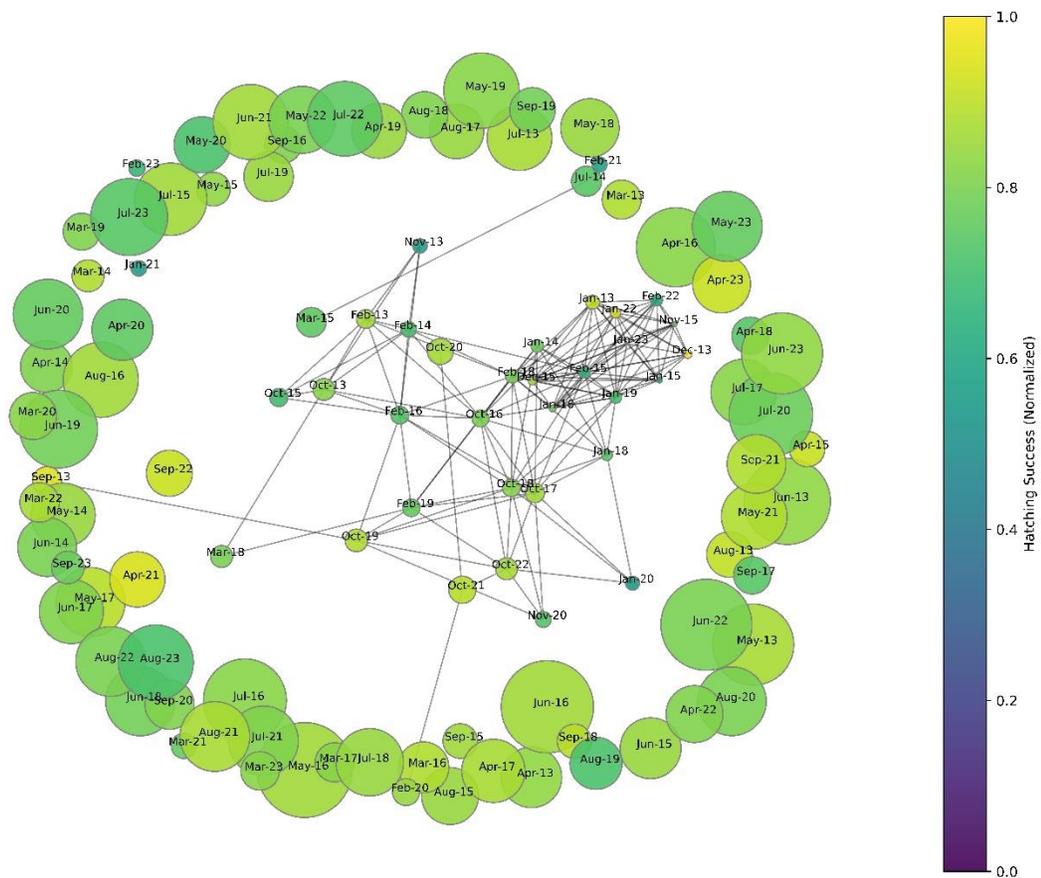

Figure 7 Hatching success BM graph for 2013 to 2023 combined

Effect of Ecological Anomalies on Hatching Success Rates

As shown in Table 1, the anomalies detected span several months from different years, each marked by distinct ecological stress patterns. In April 2016, there were elevated levels of maggots (986), fungi (714), and 673 crabs, contributing to a hatching success of 0.79. May 2016 and June 2016 continued this trend with high predator loads and increased disturbance, particularly June, which experienced 504 flood events yet still achieved a hatching success of 0.82. May 2020 and June 2020 also recorded high levels

of monitor lizard interference (110 and 59, respectively) and fungi (2,120 and 4,820), which likely contributed to lower success in May (0.70) compared to June (0.74). In July 2020, a dense concentration of crabs and fungi appeared again, maintaining a similar success rate of 0.75. May 2021 saw a combination of high lizard activity and fungi, yet it managed the highest hatching success among anomalies (0.84). The anomalies in May 2022, June 2022, July 2023, and August 2023 were primarily driven by extreme ant counts and fungi, with July and August notably exceeding 4,000 fungal counts and recording slightly lower success rates of 0.72 and 0.70, respectively. These findings highlight the complexity of ecological stressors, showing that even under severe biological and environmental pressures, some months still sustained moderate to high hatching success.

Table 1 Ecological and Biological Characteristics of Anomalous Months (2016–2023) and Corresponding Hatching Success

Month-Year	Total number of eggs	Ants	Crabs	Maggots	Monitor Lizard	Roots	Fungi	Disturbed by the mother	Flood	Hatching Success
Apr-16	20548	22	673	986	6	13	714	2	4	0.79
May-16	28754	8	871	612	10	15	1305	72	10	0.82
Jun-16	26970	2	625	675	1	33	1622	26	504	0.82
May-20	11308	121	255	269	110	0	2120	105	0	0.7
Jun-20	18620	167	174	546	59	40	4820	64	0	0.74
Jul-20	23672	186	396	271	40	83	4010	91	1	0.75
May-21	13438	48	350	173	123	2	1346	44	10	0.84
May-22	15232	304	10	66	0	155	953	4	1	0.77
Jun-22	28032	102	2	276	0	31	3242	3	108	0.77
Jul-23	21000	6315	2	637	14	5	4241	0	0	0.72
Aug-23	20269	5871	34	584	25	31	4592	0	0	0.7

The comparison between anomalous and normal months in Figure 8 reveals striking ecological differences that offer insight into disturbances affecting sea turtle hatching success. Anomalous months, identified by Isolation Forest, consistently show elevated levels of biological stressors such as ants, crabs, maggots, monitor lizards, roots, fungi, and flooding. Notably, maggot and fungi count in these months are significantly higher, with fungi often exceeding 4000 units. This suggests that microbial infestation and organic decay may be critical indicators of vulnerability. Additionally, anomalies frequently exhibit extreme predator presence, especially ants and crabs, indicating potential nest invasions that may interfere with successful embryonic development. Flooding also appears as a key disruptive factor, with some anomalies recording over 500 flood events compared to minimal disturbance in normal months.

Interestingly, hatching success in anomalous months does not always show a corresponding decline. While one might expect lower outcomes due to intense ecological stress, the median hatching success in anomalies remains clustered around 75% to 82%, sometimes even exceeding the variability observed in normal months. This result suggests that certain years may possess mitigating factors, such as favorable microclimates, maternal behaviors, or protective nesting strategies, that buffer against otherwise harsh ecological conditions. Overall, these findings emphasize the complexity of sea turtle hatching dynamics, where not all biological intensity results in failure, but rather signals an ecological state worthy of closer investigation and targeted conservation.
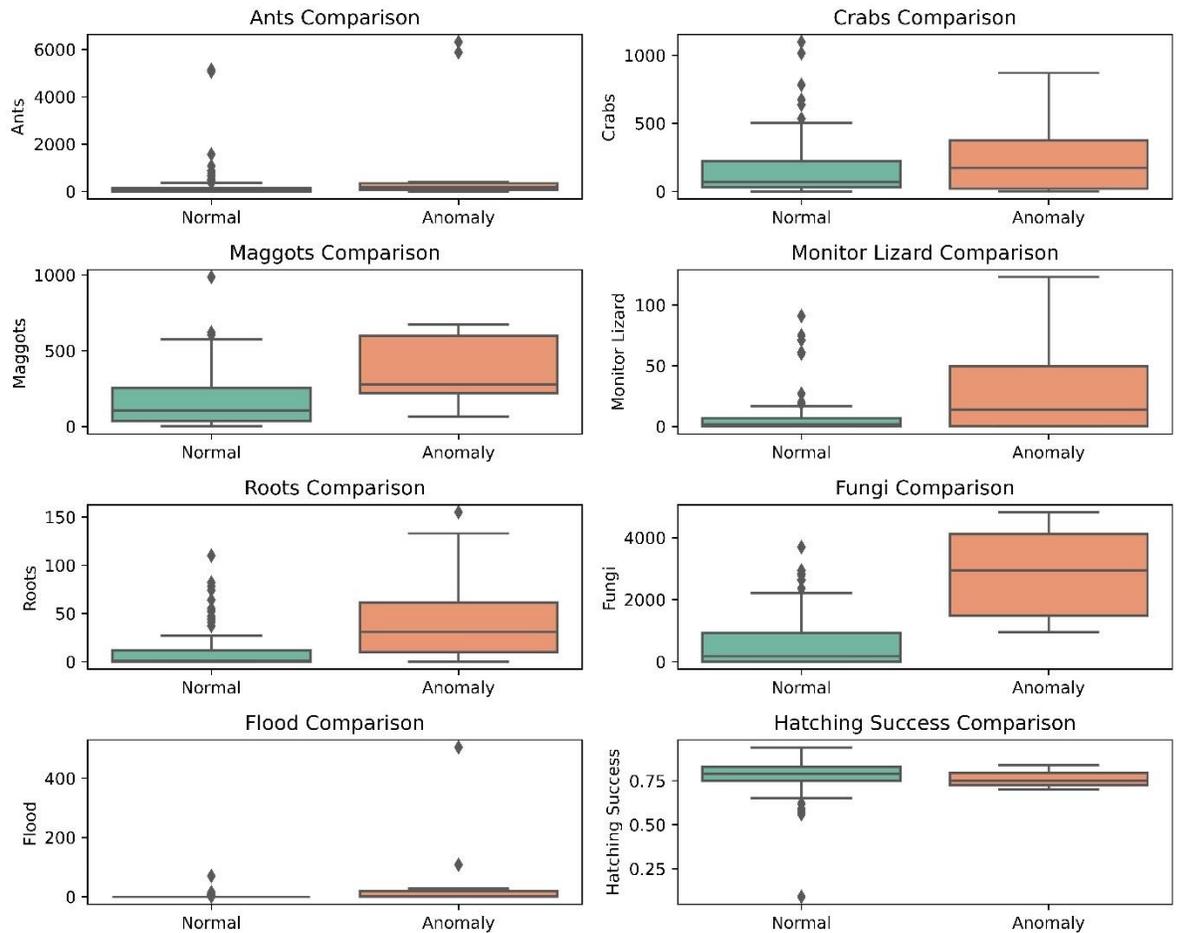
Figure 8 Comparative Boxplots of Ecological Variables and Hatching Success Between Anomalous and Normal Months

The application of the BM algorithm alongside the Isolation Forest anomaly detection method offered complementary insights into patterns and anomalies in sea turtle hatching success, with each method uncovering distinct aspects of the underlying ecological dynamics. The BM graphs revealed that months like April, May, and June consistently displayed high hatching success and egg hatch volume across multiple years. This aligns with the study by Ceolotto *et al.* [21], which identified mid-nesting seasons as peak periods for reproductive output due to favorable temperature and moisture conditions. The clustering of months with similar ecological profiles enabled the identification of periods characterized by optimal predator control or environmental stability. Conversely, months like January, February, and October frequently exhibited low hatching success and shared connections that suggest common stressors such as cooler temperatures and increased disturbance, an observation consistent with Dimitriadis *et al.* [22], who reported that early-season nesting is often more vulnerable to climatic variability.

The BM graphs' visual topological structure also reinforced the importance of predator activity, with months that had unique predator compositions often forming isolated nodes. For instance, months dominated by ant or crab activity displayed distinct topological separation, corroborating the studies by Long *et al.* [11] and Stokes *et al.* [23], who found that ant and crab predation significantly disrupts embryonic development. Moreover, months with strong connections in BM graphs, such as January and February, often had moderate to low hatching success despite differences in predator types, suggesting that other latent environmental variables may be influencing the outcomes, such as flooding or microbial growth. This complexity reinforces Farmer *et al.* [12]'s critique that empirical methods struggle to account for the compounded effects of multiple interacting stressors in field data.

The anomaly detection results from Isolation Forest further revealed months with extreme biological or ecological conditions, including elevated predator counts and fungal infestations. Yet, interestingly, many

of these "anomalous" months did not correspond with the lowest hatching success. For instance, despite significant predator and fungal pressures, hatching success rates in May and June 2016 remained high, exceeding 80%. This suggests that hatching resilience may be influenced by unobserved protective factors such as natural biocontrol, nest depth, maternal behavior, or favorable microclimates. These findings challenge the deterministic assumptions often embedded in classical mathematical models, which typically predict linear reductions in hatching success with increased predator load [9]. Instead, the results support the view that sea turtle hatching success is shaped by complex, nonlinear interactions as revealed through topological data analysis approaches.

From an ecological perspective, the consistent identification of April, May, and June as both topological and performance outliers supports the strategic importance of these months for conservation interventions. Protection efforts such as predator barriers, shading structures, or selective nest relocation can be prioritized during these months to maximize conservation returns. Conversely, the similarity clusters of low-performing months like January and October transition periods between dry and wet seasons in many ecosystems (especially tropical and subtropical ones), seasonally-timed interventions to address common stressors such as flooding or microbial infestation. Using data to divide the nesting season into important periods helps conservation programs, especially those with fewer resources, to work more efficiently and protect more nests during the long nesting season. As Lewison *et al.* [24] noted, integrating ecological and data science insights is increasingly critical for adaptive conservation management in the 21st century.

In summary, this research demonstrates the application of the BM algorithm in visualizing non-obvious ecological relationships and the strength of anomaly detection in detecting complex nesting conditions. These methods offer a robust, complementary approach for advancing predictive analytics in sea turtle conservation. The findings provide a better understanding of hatching patterns than traditional statistical models, supporting earlier claims that sea turtle reproduction is influenced by many complex, interacting factors. The observed resilience in some anomalous months underscores the adaptability of sea turtle reproductive strategies and highlights the need for more granular, localized studies of nest-level factors in future research.

Conclusions

This study demonstrates the effectiveness of combining Ball Mapper, a topological data analysis tool, with the Isolation Forest anomaly detection algorithm to explore the multifactorial nature of sea turtle hatching success over 10 years. The analysis revealed consistent seasonal patterns, with April, May, and June exhibiting high hatching success, while January and February were consistently associated with lower performance. The Ball Mapper graphs successfully highlighted clusters of similar ecological conditions, while the anomaly detection method uncovered months with extreme predator or environmental stress. Interestingly, some anomalous months displayed unexpectedly high hatching success, suggesting that the presence of certain predators may have contributed to natural biocontrol by regulating other, more harmful nest intruders. These findings support the notion that sea turtle reproductive outcomes are shaped by non-linear and context-dependent interactions, emphasizing the importance of advanced analytical tools in conservation ecology.

While these methods are used to reveal complex ecological patterns, some natural limitations remain. Data gaps in certain months and years may have slightly limited the ability to fully capture long-term trends. Although Ball Mapper and Isolation Forest effectively captured hidden structures and anomalies, future improvements could involve systematically evaluating the sensitivity of ball radius parameters and contamination thresholds to different ecological contexts, ensuring optimal model calibration. Moreover, future studies should aim to integrate high-frequency, nest-level environmental monitoring, including continuous measurements of sand temperature, moisture dynamics, and microbial profiles, to capture microhabitat variability with greater precision. Continued refinement and expansion of these approaches will be crucial for advancing predictive conservation strategies and ensuring the long-term resilience of sea turtle populations in a dynamic environmental landscape.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

References

[1] Monsinjon, J. R., Wyneken, J., Rusenko, K., López-Mendilaharsu, M., Lara, P., Santos, A., Marcovaldi, M. A. D., Fuentes, M. M., Kaska, Y., Tucek, J., Nel, R., Williams, K. L., LeBlanc, A., Rostal, D., Guillon, J., & Girondot, M. (2019). The climatic debt of loggerhead sea turtle populations in a warming world. *Ecological Indicators, 107*, 105657. https://doi.org/10.1016/j.ecolind.2019.105657.

[2] Christiaanse, J. C., Antolínez, J. A. A., Luijendijk, A. P., Athanasiou, P., Duarte, C. M., & Aarninkhof, S. (2024). Distribution of global sea turtle nesting explained from regional-scale coastal characteristics. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-023-50239-5.

[3] Cassill, D. L., & Watkins, A. (2022). Nest-site choice by loggerhead sea turtles as a risk-management adaptation to offset hatching failure by unpredictable storms and predators. *Frontiers in Ecology and Evolution, 10*. https://doi.org/10.3389/fevo.2022.850091.

[4] Gammon, M., Fossette, S., McGrath, G., & Mitchell, N. (2020). A systematic review of metabolic heat in sea turtle nests and methods to model its impact on hatching success. *Frontiers in Ecology and Evolution, 8*. https://doi.org/10.3389/fevo.2020.556379.

[5] Martins, S., Ferreira-Veiga, N., Rodrigues, Z., Querido, A., Loureiro, N. D. S., Freire, K., Abella, E., Oujo, C., & Marco, A. (2021). Hatchery efficiency as a conservation tool in threatened sea turtle rookeries with high embryonic mortality. *Ocean & Coastal Management, 212*, 105807. https://doi.org/10.1016/j.ocecoaman.2021.105807.

[6] Salleh, M. H. M., Esa, Y., Salleh, S. M., & Sah, S. A. M. (2022). Turtles in Malaysia: A review of conservation status and a call for research. *Animals, 12*(17), 2184. https://doi.org/10.3390/ani12172184.

[7] Jani, J. M., Jamalludin, M. A., & Long, S. L. (2020). To ban or not to ban? Reviewing an ongoing dilemma on sea turtle egg trade in Terengganu, Malaysia. *Frontiers in Marine Science, 6*. https://doi.org/10.3389/fmars.2019.00762.

[8] Khairuddin, M. A. A., Roslan, U. A. M., Safuan, H. M., & Rusli, M. U. (2021). Mathematical model of the population dynamics of sea turtles at Chagar Hutang Beach, Malaysia. *Journal of Sustainability Science and Management, 16*(4), 327–336. https://doi.org/10.46754/jssm.2021.06.024.

[9] Pratama, D. A., Bakar, M. A., Roslan, U. A. M., Surono, S., & Salhi, A. (2024). Numerical solutions of sea turtle population dynamics model using restarting strategy of PINN-Adam. *Results in Applied Mathematics, 22*, 100457. https://doi.org/10.1016/j.rinam.2024.100457.

[10] Roslan, U. A. M., Jailani, M. S. O., Jailani, O., & Rusli, M. U. (2019). Stability analysis for the dynamics of a new sea turtle–human interaction model. In *Proceedings of the International Conference on Computational Science and Technology (ICCST 2019)* (pp. 1–10). https://www.researchgate.net/publication/338469369.

[11] Long, S. L., Gan, J. Y. S., Loke, W. Q., & Yek, S. H. (2023). Predatory ants invading sea turtle nests on the east coast of Peninsular Malaysia. *Insectes Sociaux, 70*, 365–372. https://doi.org/10.1007/s00040-023-00925-4.

[12] Farmer, N. A., Powell, J. R., Morris, J. A., Soldevilla, M. S., Wickliffe, L. C., Jossart, J. A., MacKay, J. K., Randall, A. L., Bath, G. E., Ruvelas, P., Gray, L., Lee, J., Piniak, W., Garrison, L., Hardy, R., Hart, K. M., Sasso, C., Stokes, L., & Riley, K. L. (2022). Modeling protected species distributions and habitats to inform siting and management of pioneering ocean industries: A case study for Gulf of Mexico aquaculture. *PLOS ONE, 17*(9), e0267333. https://doi.org/10.1371/journal.pone.0267333.

[13] Dłotko, P. (2019). Ball mapper: A shape summary for topological data analysis. *arXiv*. https://doi.org/10.48550/arxiv.1901.07410.

[14] Madukpe, V. N., Zulkepli, N. F. S., Noorani, M. S. M., & Gobithaasan, R. U. (2025). Comparative analysis of Ball Mapper and conventional Mapper in investigating air pollutants' behavior. *Environmental Monitoring and Assessment, 197*(2). https://doi.org/10.1007/s10661-024-13477-2.

[15] Akuaka, G. O., Haris, H., Madukpe, V. N., Zarkasi, K. Z., Furusawa, G., & Hamid, B. A. A. (2025). Visualization of physicochemical parameters' behavior in leachate, baseliner, and surface water during dry and rainy seasons at a sanitary landfill. *Environmental Monitoring and Assessment, 197*(5). https://doi.org/10.1007/s10661-025-13961-3.

[16] Valerio, J., Vasconcelos-Filho, J. E., Stosic, B., de Oliveira, W. R., Santan, F. M., Antonino, A. C. D., & Duarte-Neto, P. J. (2024). Topological analysis of the three-dimensional radiodensity distribution of fish otoliths: Point sampling effects on dimensionality reduction. *Micron, 188*, 103731. https://doi.org/10.1016/j.micron.2024.103731.

[17] Togbe, M. U., Chabchoub, Y., Boly, A., Barry, M., Chiky, R., & Bahri, M. (2021). Anomaly detection using isolation in concept-drifting data streams. *Computers, 10*(1), 13. https://doi.org/10.3390/computers10010013.

[18] Binetti, M. S., Uricchio, V. F., & Massarelli, C. (2025). Isolation Forest for environmental monitoring: A data-driven approach to land management. *Environments, 12*(4), 116. https://doi.org/10.3390/environments12040116.

[19] Gurnari, P. (2022). *pyBallMapper: Python implementation of the Ball Mapper algorithm* [Software]. GitHub. https://github.com/Gurnari/pyBallMapper.

[20] Bâra, A., Văduva, A. G., & Oprea, S. (2024). Anomaly detection in weather phenomena: News and numerical

data-driven insights into climate change in Romania's historical regions. *International Journal of Computational Intelligence Systems, 17*(1). https://doi.org/10.1007/s44196-024-00536-2.

[21] Ceolotto, L., Mazzariol, S., Pietroluongo, G., Centelleghe, C., Mazzoldi, C., & Barausse, A. (2025). Modeling the impacts of natural and human factors on the hatching success of the loggerhead sea turtle (*Caretta caretta*) along the coasts of Italy. *PLOS ONE, 20*(4), e0320733. https://doi.org/10.1371/journal.pone.0320733.

[22] Dimitriadis, C., Karditsa, A., Almpanidou, V., Anastasatou, M., Petrakis, S., Poulos, S., Koutsoubas, D., Sourbes, L., & Mazaris, A. D. (2022). Sea level rise threatens critical nesting sites of charismatic marine turtles in the Mediterranean. *Regional Environmental Change, 22*(2). https://doi.org/10.1007/s10113-022-01922-2.

[23] Stokes, H. J., Esteban, N., & Hays, G. C. (2023). Predation of sea turtle eggs by rats and crabs. *Marine Biology, 171*(1). https://doi.org/10.1007/s00227-023-04327-9.

[24] Lewison, R., Oro, D., Godley, B., Underhill, L., Bearhop, S., Wilson, R., Ainley, D., Arcos, J., Boersma, P., Borboroglu, P., Boulinier, T., Frederiksen, M., Genovart, M., González-Solís, J., Green, J., Grémillet, D., Hamer, K., Hilton, G., Hyrenbach, K., … Yorio, P. (2012). Research priorities for seabirds: Improving conservation and management in the 21st century. *Endangered Species Research, 17*(2), 93–121. https://doi.org/10.3354/esr00419.