Abstract

Given ship trajectory data for a region, this paper proposes a physics-guided approach to detect anomalous trajectories. This problem is important for detection of illegal fishing or cargo transfer, which cause environmental and societal damage. This problem is challenging due to the presence of gaps in trajectories. Current state-of-the-art approaches either ignore the gaps or fill them using simple linear interpolation, which underestimates the ship’s possible locations during the gap. This paper proposes a novel physics-guided gap-aware anomaly detection test that incorporates physical constraints using a space-time prism. The proposed approach is evaluated with a case study using Marine Cadastre data of ships traversing in the Aleutian Islands region of Alaska in October 2017. A trajectory that could have traversed a marine protected area is correctly flagged by the proposed approach for investigation.

1 Introduction

Ship trajectory data refers to sequences of time-stamped records containing geographic locations and other ship status measures. This data frequently contain gaps (as shown in Figure 1) caused by a ship’s failure to report its locations at a predefined frequency. This paper aims to develop an anomaly detection test that utilises a physics-guided space-time prism model to interpolate trajectory gaps.

Anomaly detection on ship trajectories is of societal importance as it can be used to assist sea traffic surveillance operations to identify anomalous behaviour in a timely fashion. However, a large number of ship trajectories contain gaps, which could be intentional or accidental. For instance, Figure 1 shows a large gap in the trajectory of a Panamanian commercial fishing vessel entering the Galapagos Marine Reserve. The ship was suspected of illegal fishing activities (Oceana Mar 2018). Efficient methods are needed to detect unusual ship behavior that warrants further investigation by authorities.

The current literature is either not gap aware [(Liu, Souza, and Sydow 2014), (Laxhammar 2008), (Venskus et al. 2017), (Venskus et al. 2019), (Liu et al. 2015), (Ristic 2014)] or uses linear interpolation to fill these gaps [(Pallotta, Vespe, and Bryan 2013), (Fernandez Arguedas, Pallotta, and Vespe 2018)]. However, linear interpolation has limitations in modeling the position of moving objects because it assumes that motion follows a straight path and it does not incorporate the object’s kinematic properties (Miller 1991). To the best of our knowledge, our work is the first framework that uses a physics-guided principle, namely a space-time prism (Miller 1991), to interpolate gaps in anomalous trajectory detection.

Interpolating gaps is a challenging problem as the ship can travel at various velocities and stop in multiple locations leading to a large number of possible paths. This paper leverages physical constraints of motion by applying spatial and temporal constraints to bound a ship’s possible locations when data is missing.

This approach yields a more realistic trajectory gap model than previous work, thus likely to improve the anomaly detection.

Contributions: The main contribution of the paper is a novel physics-guided gap-aware anomaly test for trajectories. We evaluate the proposed test using a case study on...
the Marine Cadastre dataset. The case study shows that the proposed test can detect an anomalous trajectory which is missed by existing methods.

**Workshop Relevance:** This paper is relevant for at least two PGAI-AAAI-20 workshop topics:

- Use of physical constraints (e.g., maximum velocity in space-time prisms) or priors in supervised and unsupervised AI methods [e.g., hierarchical density-based spatial clustering of applications with noise (HDBSCAN) (McInnes, Healy, and Astels 2017) and anomaly detection].
- Hybrid constructions of physics-based (e.g., space-time prism) and machine learning models (e.g., HDBSCAN and anomaly detection).

**Scope:** This paper addresses anomaly detection of ship trajectories containing gaps. It focuses on maximum speed-based space time prisms to model the gaps. Consideration of more detailed physics (e.g., acceleration) and properties, such as ship-size, measurement errors, land and shallow waters, falls outside the scope. The paper uses a case study to evaluate the proposed approach. Ground-truth or synthetic data constructed with known ground truth is not considered.

**Relation to Artificial Intelligence:** The 2019 update of the National Artificial Intelligence Research and Development Strategic Plan (Kratsios, Córdova, and Walker 2019) describes Data Analytics as one of the main long-term investments that are needed to advance AI. “Further investigation of multimodality machine learning is needed to enable knowledge discovery from a wide variety of different types of data (e.g., discrete, continuous, text, spatial, temporal, spatio-temporal, graphs).” Hence, work presented in this paper is of direct relevance to the AI community.

**Outline:** The paper is structured as follows: In Section 2, we briefly explain the space-time prism concept. In Section 3, we outline our proposed approach. Section 4 describes the evaluation of results. We review the broad literature of anomalous trajectory detection in Section 5. Section 6 concludes the paper with discussion of the future work.

## 2 Space-Time Prism

The paper’s approach utilizes a physics guided space-time prism to interpolate gaps, which is described by Equation 1 in three parts.

\[
\begin{align*}
t_p & \leq t_i \leq t_q \quad (1a) \\
(x_i - x_p)^2 + (y_i - y_p)^2 & \leq (t_i - t_p)^2 \times V_{\text{max}}^2 \quad (1b) \\
(x_i - x_q)^2 + (y_i - y_q)^2 & \leq (t_q - t_i)^2 \times V_{\text{max}}^2 \quad (1c)
\end{align*}
\]

Figure 2 is an illustration of a 3-dimensional space-time prism with \(x\) and \(y\) representing its position and \(t\) representing time. Consider an object moving from an initial point \(P\) represented by \((x_p, y_p, t_p)\) to the next point \(Q\), \((x_q, y_q, t_q)\). The possible positions of the object at any time between \(t_p\) and \(t_q\) are in a region of bounded space-time, determined by:

- Spatial constraints, such as the need to be in location \((x_p, y_p)\) at time \(t_p\) and location \((x_q, y_q)\) at time \(t_q\).
- Maximum velocity the ship can travel, \(V_{\text{max}}\).

**Time spent (\(t_q - t_p\)).**

At any particular time \(t_i\) bounded by equation 1a, the object’s possible locations are determined by:

1. The positions the object could have reached from the origin \(P\), bounded by \(V_{\text{max}}\) and the elapsed time \((t_i - t_p)\). This is represented by Equation 1b and illustrated by the top facing cone (blue) in Figure 2.
2. The positions the object could be in so that it can reach the destination \(Q\) in the remaining time \((t_q - t_i)\), bounded by \(V_{\text{max}}\). This constraint is represented by Equation 1c and illustrated by the bottom facing cone (red) in Figure 2.

![Figure 2: Spatial projection of a space-time prism (Miller 1991)](image)

The intersection of these cones gives a space-time prism which represents all the possible locations of the object at time \(t_i\). This intersection is depicted by the black ellipse shown in Figure 2. The spatial projection of this ellipse in the \(x\)-\(y\) plane gives the location bound for the object at time \(t_i\),

\[
\begin{align*}
(t_i - t_p) & \leq (t_q - t_p) \quad (2a) \\
(t_q - t_i) & \leq (t_q - t_p) \quad (2b) \\
[(x_i - x_p)^2 + (y_i - y_p)^2] + [(x_i - x_q)^2 + (y_i - y_q)^2] & \leq 2(t_q - t_p)^2 \times V_{\text{max}}^2 \quad (2c)
\end{align*}
\]

In order to get the location bound for the time range \((t_p - t_q)\), ellipse represented by Equation 2c is used. This equation is in the form of the two focus definition of ellipse. Ellipse can also be described as a fixed sum of distances from 2 foci. Equation 2c is obtained by modifying the sum of equations 1b and 1c using equations 2a and 2b.

## 3 Proposed Approach

We introduce a physics-guided gap-aware anomaly detection test that consists of four steps. (1) Anomalous Region Extraction; (2) Trajectory interpolation; (3) Calculation of a Physics-Guided Gap Aware (PGA) Anomaly Score; (4) Deciding anomaly based on PGA. A trajectory can be represented by a multidimensional time series (Aggarwal 2013) or a sequence of multi-dimensional points (Lee, Han, and Whang 2007). In this paper, we work with the latter.
3.1 Rasterization

The trajectory points are rasterized into 2-dimensional raster cells with latitude (vertical) and longitude (horizontal) as their axes. A raster is a grid where each cell contains information. The longitude and latitude resolution of each cell are predefined hyper-parameters. Larger values of these hyper-parameters will lead to trajectory points from a wider area to be grouped into the same cell, leading to a less precise modeling of the area, and vice versa.

3.2 Anomalous Region Extraction

Given the historical trajectory points within an area, we first cluster them based on positional attributes (latitude and longitude) using HDBSCAN (McInnes, Healy, and Astels 2017). This step results in the latitude and longitude points frequently traversed, representing common vessel traffic patterns.

These points are rasterized with the predefined latitude and longitude resolution. Each raster cell is assigned a weight that quantifies how unlikely it is that the region was traversed. Higher weights indicate less-traversed regions. The formulation of each cell’s weight is detailed in Equation 3.

\[ R_{x,y} = \frac{\sum_{i=1}^{N} 1 - \max\{p_{ik} : k = 1, ..., K\}}{N} \]  

where \( R_{x,y} \) is the weight at cell \((x,y)\) and \( N \) is the count of points that belongs to cell \((x,y)\). For each trajectory point, HDBSCAN assigns \( K \) number of probabilities \( (p_{ik}) \), where \( K \) is the number of clusters. The maximum probability assigned to a point quantifies the strength of its membership to the assigned cluster. To calculate the anomaly of a trajectory point’s location, we subtract 1 from the strength of the membership \( p_{ik} \). Finally, the weight of each raster cell is the mean of the anomaly of all trajectory points in it.

The output of this step is a 2-dimensional anomalous region raster \( R \) with each cell assigned a weight indicating how anomalous the region it represents. An example is shown in Figure 5. This example is discussed in detail in Section 4. The number in each raster cell is its weight. The cells coloured yellow have a weight of 1, representing the most anomalous regions in the area and the cells coloured purple have a weight of 0, representing the least anomalous regions.

3.3 Trajectory Interpolation

Given a particular ship trajectory, its points are rasterized into 2-dimensional grid cells with the same resolution parameters used in Section 3.2. When a time gap larger than a predefined threshold \((T_{hr})\) occurs between two consecutive points, space-time prism interpolation is performed to approximate the ship’s probable locations (refer to Section 2).

Each trajectory raster cell will be assigned a weight value that represents the probability of the ship being present in that cell’s latitude and longitude range. Weight \( T_{x,y} \) of cell \((x,y)\) is decided by the following rules:

1. For cells without interpolation (i.e., that represent a region with a reported trajectory point), \( T_{x,y} = 1 \).
2. For cells which are flagged by space-time prism interpolation as possible locations of ship, \( T_{x,y} \) is directly proportional to the time gap and inversely proportional to the number of interpolated cells.
3. For cells with no real trajectory nor interpolated trajectory in it, \( T_{x,y} = 0 \).

Equation 4 details the computation of weight \( T_{x,y} \) of a cell \((x,y)\): With \( r_{xy} \) as the number of real reported trajectory points and \( i_{xy} \) as the number of interpolated trajectory points in cell \((x, y)\) within the time gap units \( \Delta t \). \( T_{x,y} \) is defined as follows.

\[
T_{x,y} = \begin{cases} 
1 & r_{xy} \geq 1 \\
\frac{\Delta t}{r_{xy} + \tau_{thr}} & r_{xy} = 0 \land i_{xy} \geq 1 \\
0 & r_{xy} = 0 \land i_{xy} = 0 
\end{cases} \]  

(4)

The output of this step is a 2-dimensional trajectory raster \( T \), where each cell represents how likely it was for the ship to be in that region.

An example of raster \( T \) is shown in Figure 4c. There is a gap in the ship’s trajectory data starting from the cell marked with the orange box to the cell marked with the red box. Using a space-time prism, it was found that the ship could have been present in all locations represented by the cells inside the pink box with a probability of 0.4.

3.4 Physics-Guided Gap-Aware Anomaly (PGA) Score

The output of subsections 3.2 and 3.3 are 2-dimensional raster cells with weights that quantify the degree of anomaly of parts of the region \( R \) and probabilities that a ship has been there \( (T) \), respectively. We multiply \( R \) and \( T \) to get a 2-dimensional score raster \( S \) as shown in Equation 5. Higher scores in the cells of \( S \) indicate that the trajectory is likely to visit an anomalous region (e.g., marine sanctuary) that is not a normal shipping lane. When the probability that a ship has been in a region represented by cell \((x,y)\) is high (indicated by high \( T_{x,y} \)) and that region is anomalous (high \( R_{x,y} \)), its anomaly score \( S_{x,y} \) is high, and vice-versa. Figure 6c contains an example of score raster \( S \).

\[ S = R \times T \]  

(5)

An overall anomaly score is calculated by taking the sum of \( S \), and dividing it by the root of its area, as represented by Equation 6. We call this a Physics-Guided Gap-Aware Anomaly (PGA) score.

\[ PGA = \frac{\sum_{x=1}^{r} \sum_{y=1}^{c} S_{x,y}}{\sqrt{r \times c}} \]  

(6)

where \( S_{x,y} \) is the anomaly score of each cell in raster \( S \). \( r \) and \( c \) represent the number of rows and columns in \( S \) respectively. For the given example in Figure 6c, PGA calculation is detailed as follows:

\[ PGA = \frac{\sum_{x=1}^{15} \sum_{y=1}^{15} S_{x,y}}{\sqrt{15 \times 15}} = \frac{9.1}{15} \approx 0.607 \]

A PGA score signifies how likely the ship was to be in an anomalous region. It has the following lemmas:
Lemma 1: The range of the PGA score is bounded by 0 and the root of the area of score raster $S$.

$$0 \leq PGA \leq \sqrt{r \times c}$$

(7)

For the given example in Figure 6c, Lemma 1 is as follows.

$$0 \leq PGA \leq 15$$

Lemma 2: The range of the PGA score is more tightly bounded by the length of the trajectory.

$$0 \leq PGA \leq \frac{n_t}{\sqrt{r \times c}}$$

(8)

$$0 \leq n_t \leq r \times c$$

(9)

where $n_t$ represents the number of raster cells the trajectory traverses through (after prism interpolation). For the given example in Figure 6c, $n_t$ is the count of all non-zero cells in the score raster $S$ and Lemma 2 is as follows.

$$0 \leq PGA \leq \frac{51}{15} \Rightarrow 0 \leq PGA \leq 3.4$$

3.5 Deciding Anomaly Based on PGA

After calculating the PGA score, anomalous trajectories can be detected by defining a threshold. The threshold can be decided according to the PGA range found using Lemma 2. Trajectories with a score above the threshold are anomalous. For the given example in Figure 6c, using a threshold of 0.5, the trajectory is deemed to be anomalous.

4 Evaluation

We compared the performance of the proposed approach with existing methods. The key question addressed in the evaluation was “Can the proposed interpolation based on a space-time prism detect an anomalous trajectory that is missed by methods based on simple linear interpolation or no interpolation?”

Data: Automatic Identification System (AIS) data for the Aleutian Islands region of Alaska in October 2017 was used from Zone01 of the Marine Cadastre’s dataset. This data consists of 126,931 records, each representing an AIS signal received. A case study was performed on the trajectory of a ship for the date of October 23, 2017. The considered trajectory has many gaps, including one for 12.8 hours which is shown in Figure 3a. As the data for the ship is missing near the Aleutian Islands Coral Habitat Protection Areas (dark blue region), it should be flagged for investigation. Figure 3a is the original trajectory without interpolation, red line in Figure 3b shows a linear interpolation of the gap, and the red ellipse in Figure 3c shows the space-time prism interpolation. Point P is the last trajectory point reported before the time gap and Q is the first trajectory point reported after the time gap. No interpolation and linear interpolation do not consider the protected area (dark blue regions) as possible locations of the ship, hence they do not flag its trajectory as anomalous. However, space-time interpolation shows that the ship could have traversed the protected area during the time gap and should be flagged as anomalous.

Candidate methods: We evaluated anomaly detection with PGA on trajectories containing gaps by comparing the proposed space-time prism approach, with that of existing methods, namely linear interpolation and no interpolation. We used the same methodology as the proposed approach, by replacing the interpolation stage (Section 3.3) by Linear Interpolation and No Interpolation.

The trajectory was rasterized as shown in Figure 4a. The X-axis corresponds to longitude with a resolution of 23.94 minutes and the Y-axis corresponds to latitude with a resolution of 6.36 minutes. The value in each cell represents the probability of the ship traversing in that cell’s latitude and longitude range. For example, the cells with ‘1’ represent points with continuous trajectory (no gap in data). The cells highlighted with an orange and red box represent the start and stop locations respectively. The time gap between the two locations was 12.8 hours.

Anomaly scores for the region were calculated using the trajectory data corresponding to all ships for October 2017, as shown in Figure 5.

Possible paths that the ship could have traversed in the time gaps were estimated using the space-time prism (Section 3.3). The possible trajectories are represented by green rectangles in Figure 4c. The anomaly score for the interpolated trajectory was calculated, as shown in Figure 6c. The PGA for the trajectory was found to be 0.648 by calculation method detailed in Section 3.3. Using a threshold of
0.5, this trajectory was declared to be anomalous by the proposed method and was flagged for investigation.

4.1 Comparison with Linear Interpolation
In Figure 4b, we notice that linear interpolation does not consider all possible scenarios. For the time gap corresponding to the missing data between the orange and red boxes, linear interpolation does not consider the scenario where the ship could be present in cells such as (row 9, column 4) with possible causes such as halting. Figure 6b shows the anomaly score for the trajectory with linear interpolation. Its PGA after linear interpolation is 0.368, which is less than what was computed using the proposed approach. Furthermore, with the threshold of 0.5, this trajectory will not be detected as anomalous and not be flagged for investigation.

4.2 Comparison with No Interpolation
Figure 6a shows the anomaly score for the given trajectory with no interpolation. This candidate method does not consider the path the ship would have traversed between the start and stop positions. It only calculates an anomaly score for the end points, leading to an underestimated PGA of 0.101. Thus, no interpolation doesn’t detect the trajectory as anomalous and doesn’t flag it for investigation.

4.3 Can the proposed approach detect an anomalous trajectory that is missed by the traditional methods?
Assume that the threshold of 0.5 is suitable to distinguish anomalous trajectories from the ones following the typical traffic pattern. The other candidate methods (linear and no interpolation) would have failed to detect it, because they give scores below 0.5. The proposed approach was able to correctly detect the anomalous trajectory in the case study because it considered all possible locations the ship could have traversed through. The proposed approach gave a score of 0.6, which is above the threshold.

5 Discussion
Broader literature review: Current approaches to ship anomaly detection vary mainly in the methods to extract the normal traffic patterns from historical data, and how these patterns are represented. (Pallotta, Vespe, and Bryan 2013) and (Liu, Souza, and Sydow 2014) used DBSCAN to cluster together similar trajectories. (Liu, Souza, and Sydow 2014) modified the DBSCAN algorithm to cluster trajectory points with additional parameters such as speed and direction variances. As such similar points are not only close location-wise, but their speed and direction are also similar at that location. (Laxhammar 2008) used Gaussian Mixture Models as the cluster model and Expectation-Maximization (EM) as clustering algorithm for normalcy extraction. (Venskus et al. 2017) and (Venskus et al. 2019) used Self-organizing Map (SOM) neural network technique to extract normal lane clusters.

The representation of normalcy also differs across the literature. (Liu et al. 2015) represents normal trajectories as vector points along the trajectory, which consist of five attributes: average course over ground, average speed over
Conjectures: We would like to explore the following conjectures:

1. The range of PGA can be bounded by the maximum speed of the ship. The number of locations the ship traverses in the time gap is found using a space-time prism (refer Section 2) which is bounded by the ship’s maximum speed. Decreasing the ship’s maximum speed tightens the constraints of the space-time prism by decreasing the RHS of Equations 2c.

2. The proposed approach will identify all the regions flagged by linear interpolation as anomalous (anomaly score in score raster S > 0). However, the anomaly scores are also dependent on the probability of a ship being present in a particular location (weight in trajectory raster T). Thus, we need to explore if all the trajectories flagged by linear interpolation is a subset of the trajectories flagged by the proposed approach.

6 Conclusions and Future Work

A physics-guided gap-aware approach is proposed to detect anomalous trajectory. The proposed approach is relevant for sea surveillance, and can potentially aid border protection or maritime administration services. The novelty of the proposed approach is using a more realistic physics-guided method to interpolate gaps in trajectories prior to anomaly detection. Existing approaches are not as realistic as they either disregard the gaps or perform a simple linear interpolation, which underestimates the set of possible locations. A comparison of the proposed approach with the current methods was done using a case study. The result shows that space-time prism interpolation can detect a potentially anomalous trajectory that was missed by the current methods.

In the future, we want to explore additional physical constraints such as adding maximum acceleration, non-uniform distribution for the interpolated region (e.g., using straight line probabilistic approach or monte-carlo simulation), ship-size and measurement errors to our interpolation method to model the gaps more precisely. Also, we want to improve the evaluation methodology by generating synthetic data with defined ground-truth labels and using metrics such as precision, recall and AUROC/AUPRC scores. Conjecture 1 defined in Section 5 needs to be further explored to tighten the range of the physics-guided gap-aware anomaly score. Further investigation of conjecture 2 in Section 5 will lead to a better understanding of the relationship between the proposed approach’s interpolation result and existing methods.

7 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 1029711, 1737633, IIS-1320580, IIS-0940818, IIS-1218168, 1916518, and 1901099, the USDOD under Grants No. HM1582-08-1-0017 and HM0210-13-1-0005, the Advanced Research Projects Agency-Energy (ARPA-E), USDOE under Award No. DE-AR0000795, the NIH under Grant No. UL1 TR002494, KL2 TR002492, and TL1 TR002493, the USDA under Grant No. 2017-51181-27222. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency.
thereof.

References


