Simplification and Event Identification for AIS Trajectories: the Equivalent Passage Plan Method

Luis Felipe Sánchez-Heres

(*Department of Mechanics and Maritime Sciences, Chalmers University of Technology, Gothenburg, Sweden*)
(E-mail: sanchezheresluis@hotmail.com)

Two pre-processes for Automatic Identification System (AIS) trajectories commonly reported in the maritime knowledge discovery literature are trajectory simplification and event identification. Both pre-processes reduce storage and computational expenses by reducing the number of data points to be used in an analysis. This paper presents an event identification and trajectory simplification method based on behaviour identification and translation. Trajectory segments deemed to correspond to coastal or ocean navigation are translated into equivalent passage plan segments; a succinct description of the movements and behaviour of the ship. As a trajectory simplification method, it provides two main advantages over commonly used trajectory simplification methods: more meaningful simplified trajectories with better encoding of basic behaviours and the possibility to retain interesting behaviours in full resolution. As an event identification method, it is capable of differentiating between normal ocean or coastal navigating behaviour and complex or interesting behaviour, such as pilotage, reaction to a traffic conflict, or an involuntary deviation from the passage plan.

1. INTRODUCTION.   The Automatic Identification System (AIS) was devised to assist in the prevention of collisions at sea. Through radio messages, vessels broadcast their state locally and in real-time, improving their visibility to other vessels and vessel traffic services. Nevertheless, because of the introduction of AIS receiver networks and AIS message databases, AIS data is today being used in a wide range of applications.

AIS data applications can be roughly categorised into three broad groups: surveillance, accident investigation, and knowledge discovery. In surveillance applications, AIS data is used to detect abnormal or illegal vessel behaviour in real-time or periodically through deferred processes. For example, de Sauza et al. (2016) presented a methodology to identify illegal fishing through AIS data, Patroumpas et al. (2015) presented a system capable of real-time surveillance and Huang et al. (2017) presented a collision warning system. In accident investigation, AIS data encompassing a small number of vessels for a typically

short period is analysed to assist in determining the cause of an incident or accident. Wang et al. (2013) presented an example of this type of analysis. Finally, in knowledge discovery, patterns are inferred from large volumes of AIS data. What a pattern is, depends on the analysis. For example, Pallota et al. (2013) presented a method for inferring shipping routes, Rong and Mau (2013) presented a method for inferring ship manoeuvring coefficients and Kujala et al. (2009) estimated collision probabilities in the Gulf of Finland. Of the three application groups for AIS data, knowledge discovery is the most diverse, and arguably, fruitful. Patterns found through knowledge discovery have been used to improve the application of AIS data on the other groups. For example, fishing areas determined from a large number of AIS trajectories can be used for surveillance of illegal fishing (Mazzarela et al., 2014).

All AIS data applications have particular and common challenges. For example, the limitations and deficiencies of AIS data, such as sampling rate, sensor errors, unreliable values and malicious modifications pose challenges to surveillance, accident investigation, and knowledge discovery (Harati-Mokhtari et al., 2007; Hu et al., 2014; Iphar et al., 2015). For knowledge discovery, one particularly important challenge is handling the volume of AIS data.

AIS data needs to be pre-processed to reduce computational and storage expense, and to facilitate the discovery of patterns. Two pre-processes commonly reported in the maritime knowledge discovery literature are trajectory simplification and event identification. Both pre-processes reduce storage and computational expense by reducing the number of data points.

Trajectory simplification reduces the number of entries used to describe a trajectory. Down sampling and simple trajectory simplification methods based on data point significance (Berminghan and Lee, 2017) are commonly used in maritime literature (for example, Willems et al., 2009; Muthu, 2015; Li et al., 2016; Dhar, 2016; Zhang et al., 2018). Methods based on data point significance use scoring heuristics to determine whether or not a data point should be removed. While these methods are simple to implement, one must carefully consider their use in modern analyses and applications due to their significant drawbacks and the possible existence of more suitable methods (Feng and Zhu, 2016; Sun et al., 2016).

Event identification removes portions of the trajectory deemed of no interest. In other words, only the interesting trajectory segments are analysed. Event identification may also be considered a knowledge discovery process, and its results may also be the input for another knowledge discovery process. In addition to arrival and departure from ports, commonly identified events in the maritime literature are basic vessel behaviours such as slow down, stop, and turn (Pallota et al., 2013; Cazzanti and Pallota, 2015; Patroumpas et al., 2015; Qi and Zheng, 2016), and complex behaviours such as fishing and traffic conflicts (Mazzarela et al., 2014; Olindersson and Janson, 2015; Wu et al., 2016).

This paper presents the Equivalent Passage Plan (EPP) method for trajectory simplification and event identification. The EPP method is based on the following idea. During coastal and ocean navigation, the trajectory of a ship is essentially a sequence or fixed-course legs and course change manoeuvres because the ship is closely following its passage plan. Thus, trajectory segments deemed to correspond to coastal or ocean navigation can be translated into EPP segments. Doing so greatly simplifies the trajectory without degrading the description of the behaviour. Trajectory segments that are not deemed to correspond to coastal or ocean navigation, can be retained in full resolution or simplified with another
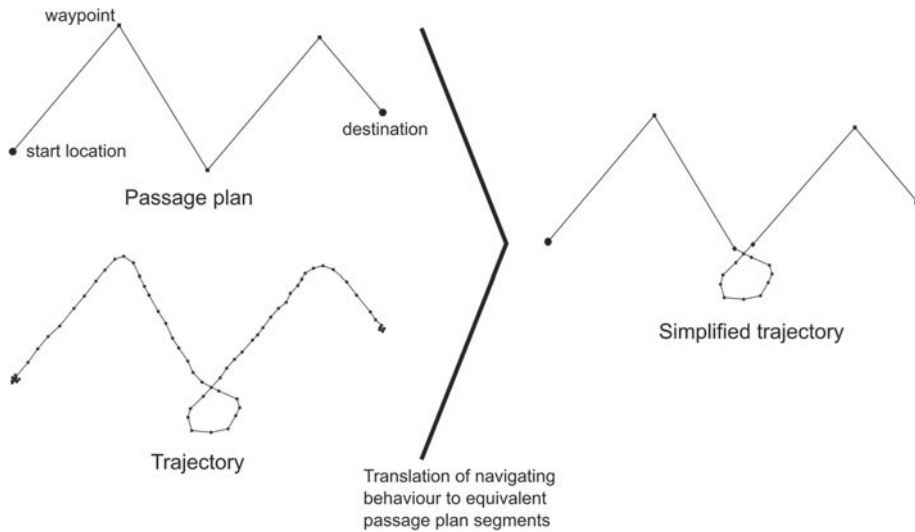
Figure 1.   Simple example of the idea behind the EPP method.

method. Figure 1 presents a simple example of this idea. The loop turn at the bottom of the trajectory cannot be classified as ocean or coastal navigation, so it is not translated. Its identification is valuable because it shows a change in behaviour, and therefore may be of special interest for the analyst. Furthermore, with the EPP method, this trajectory segment can be kept in full resolution for detailed analysis while the rest of the trajectory is simplified. The two main contributions of this method to marine traffic knowledge discovery literature are:

- The EPP method is a novel trajectory simplification method based on behaviour identification and translation. Compared to down sampling and other methods based on data point significance, the EPP method has two main advantages: first, it renders more meaningful simplified trajectories that correctly encode basic behaviour, and second, it allows for the retention, in full resolution, of trajectory segments corresponding to complex or interesting behaviour.
- The EPP method is also a novel event identification method capable of differentiating between normal ocean or coastal navigating behaviour and other types of behaviour, such as pilotage, reaction to a traffic conflict, or an involuntary deviation from the passage plan.

The paper is organised as follows: Section 2 presents the general formulation of the EPP method, and Section 3 presents a comparison of it against a trajectory simplification algorithm based on data point significance. Section 4 summarises the benefits and limitations of the EPP method.

2. METHOD DESCRIPTION.   The EPP method takes as its input a ship trajectory described by a list of points. Each point is a {longitude, latitude, timestamp} tuple and corresponds to an individual AIS message in a set. A set of AIS messages describing a
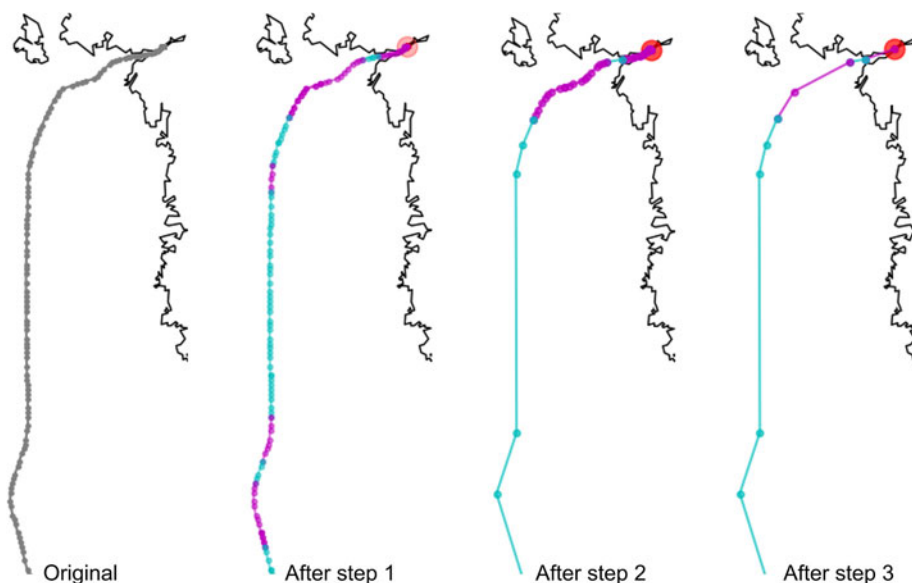
Figure 2.   Step-by-step simplification of an AIS ship trajectory with the EPP method.

trajectory consist of consecutive AIS messages with identical Maritime Mobile Service Identity (MMSI) number and less than ten minutes of delay between them.

To fully describe the on-surface movements of a ship, a trajectory must capture the temporal changes of the ship's velocity, orientation, and location. The input trajectory to the EPP method describes only temporal changes in location. The reason for this is that despite AIS having the capacity to describe temporal speed and orientation changes (heading, rate of turn, speed and course over ground), this information can seldom be used as it is often missing or unreliable (Harati-Mokhtari et al., 2007). Nevertheless, the ship's course, orientation and velocity can be approximated by assuming that the ship moves forward with a constant speed along a rhumb line joining any two adjacent points. Thus, for each trajectory segment defined by a pair of adjacent points, the ship's velocity, orientation and course can be calculated, and for each pair of these two-point segments, a course change can be determined.

Figure 2 shows how a ship trajectory is simplified with the EPP method. The steps shown in the figure are:

(1) Segment the trajectory into three basic behaviours: stop, fixed-course sailing, and turn.
(2) Identify basic behaviour segments corresponding to coastal and ocean navigation and translate them into EPP segments.
(3) If possible or desired, simplify the rest of the trajectory with another method.

In step 1, the trajectory is segmented into basic behaviours with an algorithm based on the work of Buchin et al. (2011). The algorithm uses three main functions TEST, SPLIT and SEGMENT. Figures 3 and 4 present pseudocodes for these functions. The function TEST accepts a trajectory segment (a subset of points) as input and returns *True* if the entire segment and any subsegment of it is identified as one of the basic behaviours (stop,

```
function TEST(subtrajectory)
    courses ← courses calculated from subtrajectory
    course_changes ← course changes calculated from courses
    if length(subtrajectory) = 1 then
        return True
    end if
    if all vertices in subtrajectory are enclosable in a circle of radius r then
        return True                                    ▷ Stop behaviour
    end if
    if the range of values in courses ≤ d degrees then
        return True                                    ▷ Fixed course behaviour
    end if
    if course_changes does not contain a set of k adjacent course changes
with values ≤ d degrees then
        return True                                    ▷ Turn manoeuvre behaviour
    end if
    return False
end function
```

Figure 3.   Pseudocode of the TEST function.

fixed-course sailing, or turn manoeuvre) and *False* otherwise. A trajectory segment can be identified as a stop if all its points can be enclosed by a circle small enough to guarantee a lack or near lack of movement. A trajectory segment can be identified as fixed-course sailing when the courses calculated from each pair of adjacent points segments are within a narrow range of values and cannot be identified as a stop. A possible description of a turn manoeuvre is a trajectory segment that does not contain fixed-course segments and is not a stop. For this algorithm, this specification is too restrictive. If used, a trajectory segment corresponding to the transition between turning directions (for example, a ship turning to port starts turning to starboard) is identified as a fixed-course segment. This issue is resolved by allowing sets of maximum k adjacent segments, that would be identified as fixed-course segments. Stop, fixed-course and turn are not the only basic behaviours that can be identified, but they are the most convenient due to their simplicity.

The SEGMENT function takes a full trajectory as input and returns a list of trajectory segments corresponding to the basic behaviours. The segmentation consists of calling the SPLIT function until the trajectory cannot be split into more basic behaviours.

The SPLIT function is a modified binary search algorithm. It takes a trajectory segment as input and returns i, the index of the first point in the given trajectory segment where there is a transition between basic behaviours (that is, the function TEST returns *True* for $[0 : i]$ and *False* for $[0 : i + 1]$). In Figure 3, the first if-statement warrants an explanation. The if-statement returns *True* when the trajectory consists of only one point. This condition is necessary so that the SPLIT function works correctly. Because the SPLIT function is a binary search algorithm, the TEST function must be monotonic, so it cannot return *False* for the trajectory segment $[0 : 1]$ and *True* for $[0 : 2]$. The if-statement resolves this problem.

In step 2, the identification and translation is done with a TRANSLATE function. This function encodes what is considered as coastal and ocean navigation and the desired level of detail for the EPP. The level of detail can be defined through two questions: How accurate should the EPP describe the ship's path (spatial accuracy) and how accurate should the EPP capture the ship's location in time (temporal accuracy)? In this implementation of the EPP method, the ship trajectory segments deemed to correspond to coastal or ocean

```
function SEGMENT(trajectory)
    segments ← new List
    current ← trajectory
    loop
        i ← SPLIT(current)
        segments.insert(current[0 : i])
        if i + 1 = length(current) then
            break
        else
            current ← current[i : end]
        end if
    end loop
    return segments
end function

function SPLIT(trajectory)
    split ← False
    first ← 1
    last ←length(trajectory)
    while first ≤ last AND NOT split  do
        i ← (first + last) DIV 2
        pass_A ← TEST(trajectory[0 : i])
        pass_B ← TEST(trajectory[0 : i + 1])
        if pass_A AND NOT pass_B then
            split ← True
        else
            if pass_A AND pass_B then
                first ← i + 1
            else
                last ← i − 1
            end if
        end if
    end while
    return i
end function
```

Figure 4.   Pseudocode of the SEGMENT and SPLIT functions.

navigation are meant to be used in route planning analyses where detailed information of the ship's turns is not necessary, and neither is the exact location of the ship along the route at a given time; therefore, the EPP segments do not need to capture these details. Figure 5 presents the conditions and manner in which this implementation of the TRANSLATE function identifies stops and behaviour corresponding to coastal or ocean navigation, and then translates them into stop points and EPP segments. The identification is done through the basic behaviours described by the TEST function. EPP segments are sets of fixed-course
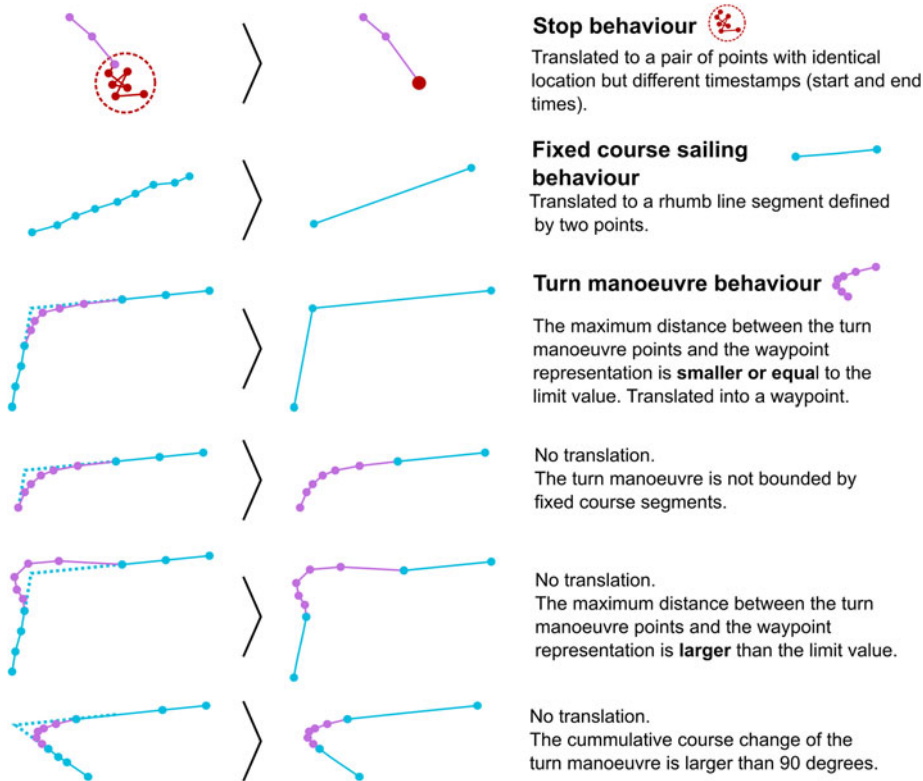
Figure 5.   Visual explanation of the mapping performed by the TRANSLATE function.

behaviours and turn manoeuvres that can be considered to be normal or uneventful course change manoeuvres. Therefore, not all turn manoeuvres are translated. This is why, in Figure 5, only one turn manoeuvre is translated. Unnecessarily wide or very large course change manoeuvres are seldom normal or uneventful. Furthermore, if a trajectory starts or ends with a turn manoeuvre, it is impossible to know whether it was a normal course change manoeuvre. To address this issue, such turn manoeuvres at the beginning or end of a trajectory are not translated. Untranslated turn manoeuvres may correspond to pilotage, collision avoidance manoeuvres or course corrections. A pseudocode for the TRANSLATE function is not presented here for brevity. Interested readers are referred to the actual code hosted in the following repository: https://github.com/LuisFe/EPP.

At a glance, the EPP method may bear some resemblance to trajectory simplification methods based on data point significance. One such method was presented by Dhar (2016). In his algorithm, data points describing course changes or speed above a threshold value are deemed as significant and retained. All other data points are discarded. The result of using changes in heading as a significance scoring heuristic is a simplified trajectory that ignores small changes and represents turn manoeuvres above the threshold value with one or several data points. Similarly, the TEST function also contains a threshold value for course changes, but its use and the end result is significantly different. The TEST function describes how to segment the trajectory, but not how to simplify it. The plot labelled "After

step 1" in Figure 2 shows the identified behaviours in different colours. At this point in the EPP method, the trajectory has not been simplified to any extent. Step 2 is where simplification may take place, as the TRANSLATE function only simplifies segments deemed to correspond to coastal or ocean navigation or stops. The result of this method is a trajectory that may be simplified to different extents. Turn manoeuvres, for example, are described as waypoints if they are part of coastal or ocean navigating behaviour and retained in full resolution otherwise.

Without any further simplification of the non-translated trajectory segments, the magnitude of the simplification achieved through the EPP method is primarily limited by the portion of the trajectory that corresponds to coastal and ocean navigation, as well as stops. Secondary limitations are the complexity of the trajectory and the desired details to be captured by the EPP segments. For example, a course change manoeuvre can be represented in an EPP segment as a waypoint or as an arc. In this implementation, course changes are represented as waypoints, but other implementations of this method could easily have higher levels of detail.

3. RESULTS AND DISCUSSION. This section presents comparisons between ship trajectories processed with the EPP method and two trajectory simplification methods based on data point significance and a discussion of the advantages and drawbacks of the EPP method.

The first comparison is with the Douglas-Peucker (DP) line simplification algorithm (Douglas and Peucker, 1973). The DP algorithm was used for two reasons. First, it is commonly used to simplify AIS trajectories (Willems et al., 2009; Muthu, 2015; Li et al., 2016; Dhar, 2016; Zhang et al., 2018). Second, the resulting simplified trajectories exhibit some of the expected drawbacks from trajectory simplification methods based on data point significance.

The DP algorithm simplifies lines described by polylines according to a threshold parameter τ. If the distance between a point and a line passing through two other points is smaller than τ, the point is deemed insignificant and thus removed. This means that the DP algorithm can simplify a complex line to different levels. The higher the value of τ, the simpler the line becomes. For AIS trajectory simplification with the DP algorithm, there is no standard value for the τ parameter. The value depends on the analysis and the trajectories. Some research, however, has tried to create guidelines (Zhang et al., 2016).

Figures 6–8 present comparisons between the EPP method and the DP algorithm. For each of the example trajectories, the DP algorithm could have easily rendered simplified trajectories with higher or lower resolution than the ones shown since the level of simplification depends on the value of the τ parameter. The τ value used for each trajectory was chosen as follows. First, the Hausdorff distance, a numerical measure of line similarity, was used to determine the similarity between the simplified trajectories obtained from both methods and the original trajectories. Then, for each example trajectory, the τ parameter was adjusted so that the simplified trajectory obtained from the DP algorithm had the same similarity value as the one obtained from the EPP method. This procedure guarantees that the simplified trajectories obtained from both methods describe at an equivalent level of detail the on-surface movements of the ship.

The second comparison is with the Open Window Spatiotemporal Algorithm (OPW-SP) presented by Meratnia and Rolf (2004). This algorithm was used because it is a trajectory
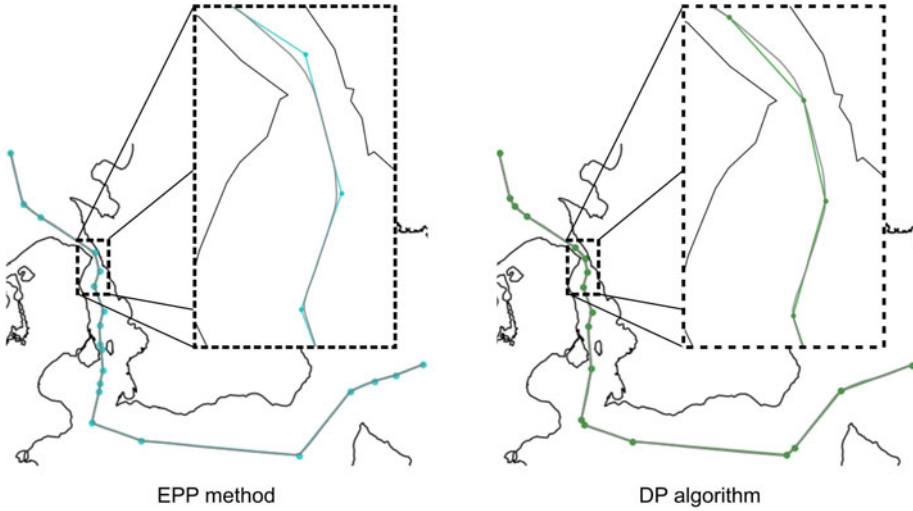
Figure 6.   Comparison between simplified trajectories obtained with the EPP method and the DP algorithm for an AIS ship trajectory that can be fully translated into an EPP. (Original trajectory shown in light grey.)
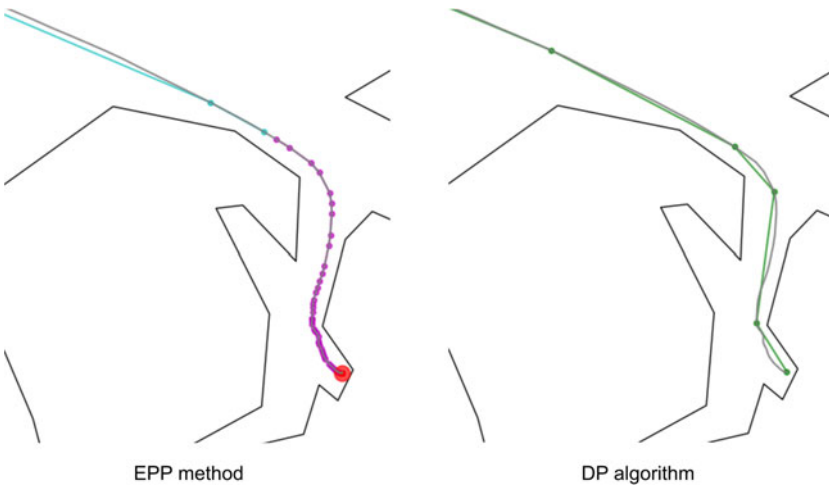

Figure 7.   Comparison between simplified trajectories obtained with the EPP method and the DP algorithm for an AIS ship trajectory that contains an approach to a port and a stop. (Original trajectory shown in light grey.)

simplification algorithm based on data point significance that considers the temporal dimension, in contrast with the DP algorithm that does not. In the OPW-SP algorithm, a trajectory point is discarded if it does not inform of a speed change over a certain threshold or if its removal renders a simplified trajectory that describes a similar location at the same time within a distance error. Figure 9 presents a comparison between the EPP method and the OPW-SP algorithm using different speed and distance error thresholds. As one would expect, the higher thresholds render a more simplified trajectory.
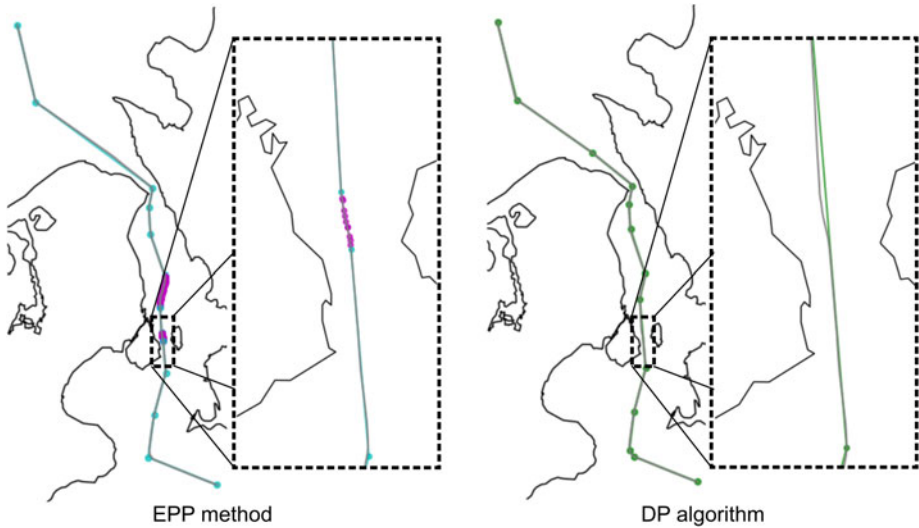
Figure 8.   Comparison between simplified trajectories obtained with the EPP method and the DP algorithm for an AIS ship trajectory that contains a course correction manoeuvre. (Original trajectory shown in light grey.)
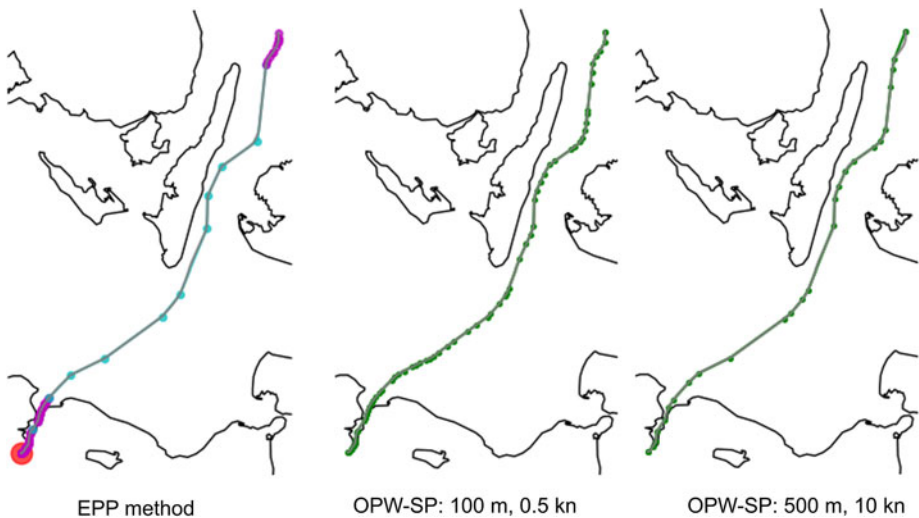

Figure 9.   Comparison between simplified trajectories obtained with the EPP method and the OPW-SP algorithm with different speed and distance error thresholds. (Original trajectory shown in light grey.)

3.1.   *EPP renders more meaningful simplified trajectories.*   Figures 6–9 present trajectories simplified with the EPP method and the DP and OPW-SP algorithms. With the EPP method, line segments and points (that is, rhumb lines and waypoints in the EPP) of translated segments correspond to fixed-course sailing and course change manoeuvres, respectively. With the DP and OPW-SP algorithms this correspondence is not guaranteed. Figure 6 presents a trajectory that could be translated entirely into an EPP. At a glance, the

simplified trajectories obtained with the EPP method and the DP algorithm look very similar, but closer examination shows that the EPP method provides superior correspondence between the simplified trajectory and the underlying behaviours. The same issue can be observed in Figure 9. The line segments and points of the simplified trajectories rendered by the OPW-SP algorithm do not correspond to fixed-course sailing and turn manoeuvres regardless of the level of simplification. The simplified trajectories obtained with the DP and OPW-SP algorithms approximate the originals, but the underlying behaviour is lost.

Figures 7 and 9 illustrate another important difference. The EPP method can recognise stops; the DP and OPW-SP algorithms cannot. In the figures, the ends of the simplified trajectory can be interpreted as stops only because they end at a shore. A stop without such context would be essentially erased by the DP and OPW-SP algorithms.

Simplifying AIS trajectory segments to EPP segments consisting of waypoints and stops is similar, but not equal, to reducing it into sequences of critical points as proposed by Potroumpas et al., (2015). The EPP segments provide a different encoding of the behaviour that is more suitable for some types of analyses (for example, route planning analyses).

Overall, the DP and OPW-SP algorithms simplify trajectories without regard to the underlying behaviour, rendering simplified trajectories where each point only describes the position of the ship in time and space. The EPP method, on the other hand, simplifies the trajectory according to the identified behaviours, rendering a simplified trajectory that encodes the underlying behaviours: each point describes more than just the position of the ship in time and space. This characteristic is particularly valuable for the analysis of individual ship trajectories since the behaviour is highlighted.

3.2. *EPP can identify interesting behaviour and can keep it at full resolution.* Figures 7–9 show a fundamental difference between the methods: the EPP method can treat segments of the trajectory differently according to their underlying behaviour, while the DP and OPW-SP algorithms cannot. In some cases, this difference may be a significant advantage. Segments that cannot be deemed to correspond to coastal and ocean navigation may be of great interest for knowledge discovery. Questions such as: does abnormal behaviour happen? where does it happen? why does it happen? and what kind of behaviour is it? may be unanswerable if the simplification algorithm oversimplifies the trajectory.

In Figure 7, the ship is entering a port and changing its course constantly, a behaviour typical of pilotage. With the EPP method, port entry is deemed not to correspond to ocean or coastal navigation, so it is stored in full resolution, enabling detailed analysis in the future. The DP algorithm, on the other hand, greatly simplifies the port approach with the current value. If the value of was changed in order to reduce the simplification, the whole trajectory would be stored at a higher resolution, and not only the port entry.

In Figure 8, the EPP algorithm does not translate a manoeuvre by the ship to adjust its course, so it is retained at full resolution. The DP method, on the other hand, disregards it with the used value. If a smaller value had been used, the manoeuvre would not have been erased, but it would not have been identified either.

In Figure 9, one can appreciate that these issues apply also to the OPW-SP algorithm. With 500 metres and 10 knots as thresholds, the OPW-SP algorithm disregards the manoeuvre at the northernmost section of the trajectory. This manoeuvre is not disregarded when 100 metres and 0.5 knots are used as thresholds, but it is not identified either and it leads to a higher resolution in the entire trajectory.

An important detail is that the current implementation of the EPP method does disregard speed changes in the translated segments. The OPW-SP algorithm does not. Nevertheless,

as mentioned in Section 2, the EPP method disregards the temporal dimension in the translated segments by design. Future implementations could identify speed changes by modifying the TEST function.

If AIS ship trajectories are to be used in detailed trajectory analyses (for example, Wang et al., 2013), being able to identify and retain complex trajectories in full resolution is of great advantage. The semantic method proposed by Potroumpas et al. (2015) does not have this capacity. Furthermore, the EPP method is more than just a trajectory simplification method, it is an event identification method.

3.3. *Limitations and performance of the EPP method.* The main limitation of the EPP method is that it is only suitable for AIS trajectories of ships that engage in coastal or ocean navigation. If a ship's trajectories cannot be deemed to correspond to coastal or ocean navigation, there will be no simplification or event identification beyond stops. Pilot boats are a good example. This type of vessel does not sail with a fixed-course over relatively long distances, so no segments of their trajectory can be translated to EPP segments.

Depending on the application, the data reduction capacity of the EPP method may be also a considerable limitation. Without the use of complementary trajectory simplification methods (see step 3 in Section 2), the only case when the EPP method can achieve data reduction comparable to trajectory simplification methods based on data point significance is when a ship trajectory consists only of coastal and ocean navigation behaviour. Figure 6 presents such a case. In it, the DP algorithm and the EPP render simplified trajectories consisting of 18 and 20 points respectively. If the ship trajectory includes other types of behaviour, the data reduction achieved through the EPP will be smaller because of the retention of segments at full resolution (see Figures 7, 8 and 9). Even if a complementary simplification method is used, the data reduction of the EPP method will still be smaller than other methods. The EPP method is also slower than the DP algorithm, but faster than the author's implementation of the OPW-SP algorithm. As a simple comparison, in a computer with a 1.7 GHz processor, the EPP method takes around 5 seconds to identify behaviours and simplify 2,000 AIS messages, the DP algorithm takes between 0.17 and 1.1 seconds depending on the threshold value, and the OPW-SP algorithm takes around 10 seconds.

Overall, if more meaningful simplified trajectories or the identification of potentially abnormal behaviour is of no interest, the EPP method is not a suitable trajectory simplification algorithm.

4. CONCLUSIONS. The EPP method is a novel trajectory simplification and event identification method based on behaviour identification and translation. It has two main advantages over trajectory simplification methods based on data point significance. First, the simplified trajectories are more meaningful. Trajectory segments deemed to correspond to coastal and ocean navigation are translated to EPP segments. The EPP segments themselves express that the displayed trajectories correspond to normal ocean or coastal navigation, while the encoding guarantees correspondence between waypoints in the EPP segments and course change manoeuvres, as well as rhumb lines and fixed-course sailing. Non-behaviour-based methods do not provide this guarantee, as they do not account for underlying behaviour in the simplification. Second, the EPP method can treat segments of the trajectory differently according to their underlying behaviour. This characteristic can be used to identify complex and interesting behaviours from being unidentified or erased due

to simplification. When this characteristic is exploited, the EPP method works as an event identification method, and as such, it is the first one capable of differentiating between normal ocean or coastal navigating behaviour and other types of behaviour, such as pilotage, reaction to a traffic conflict, or an involuntary deviation from the passage plan. The main limitation of the EPP method is that a ship trajectory will not be simplified unless the ship engages in coastal or ocean navigation. Despite its advantages, if more meaningful simplified trajectories or the identification of complex and abnormal behaviour is of interest, the EPP method is not a suitable trajectory simplification algorithm because of lower data reduction capabilities, longer running time, and higher implementation effort.

REFERENCES

Berminghan, L. and Lee, I. (2017). A framework of spatio-temporal trajectory simplification methods. *International Journal of Geographical Information Science*, **31**(6), 1128–1153.

Buchin, M., Driemel, A., Van Kreveld, M. and Sacristán, V. (2011). Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science*, **3**, 33–63.

Cazzanti, L. and Pallota, G. (2015). Mining Maritime Vessel Traffic: Promises, Challenges, Techniques. *OCEANS 2015-Genova*, Genova, Italy. IEEE. 1–6.

de Souza, E. N., Boerder, K., Matwin, S. and Worm, B. (2016). Improving fishing pattern detection from satellite AIS using data mining and machine learning. *PloS one*, **11**(7), e0158248.

Dhar, S. (2016). *Addressing Challenges with Big Data for Maritime Navigation: AIS Data within the Great Lakes System.* Doctoral Dissertation, University of Toledo.

Douglas, D.H. and Peucker, T.K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, **10**(2), 112–122.

Feng, Z. and Zhu, Y. (2016). A survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access*, **4**, 2056–2067.

Harati-Mokhtari, A., Wall, A., Brooks, P. and Wang, J. (2007). Automatic Identification System (AIS): data reliability and human error implications. *The Journal of Navigation*, **60**(3), 373–389.

Hu, Q., Cai, F., Yang, C. and Shi, C. J. (2014). An algorithm for interpolating ship motion vectors. *International Journal on Marine Navigation and Safety of Sea Transportation*, **8**(1), 35–40.

Huang, C., Hu, S., Kong, F. and Xi, Y. (2017). Pre-warning system analysis on dynamic risk of ship collision with bridge at restricted waters. 4th *International Conference on Transportation Information and Safety (ICTIS)*, Alberta, Canada. IEEE.

Iphar, C., Napoli, A. and Ray, C. (2015). Detection of false AIS messages for the improvement of maritime situational awareness. In *OCEANS'15 MTS/IEEE*, Washington, United States of America, IEEE. 1–7.

Kujala, P., Hänninen, M., Arola, T. and Ylitalo, J. (2009). Analysis of the marine traffic safety in the Gulf of Finland. *Reliability Engineering & System Safety*, **94**(8), 1349–1357.

Li, Y., Liu, R. W., Liu, J., Huang, Y., Hu, B. and Wang, K. (2016). Trajectory compression-guided visualization of spatio-temporal AIS vessel density. 8th *International Conference in Wireless Communications & Signal Processing (WCSP16)*, Yangzhou, Jiangsu, China. IEEE. 1–5.

Mazzarella, F., Vespe, M., Damalas, D. and Osio, G. (2014). Discovering vessel activities at sea using AIS data: Mapping of fishing footprints. 17th *International Conference on Information Fusion (FUSION)*. Salamanca, Spain. IEEE. 1–7.

Meratnia, N., & Rolf, A. (2004). Spatiotemporal compression techniques for moving point objects. In *International Conference on Extending Database Technology*, Berlin, Germany. Springer.

Muthu, S. S. (2015). Visualization, statistical analysis, and mining of historical vessel data. Doctoral dissertation, University of New Brunswick.

Olindersson, F. and Janson, C. (2015). Development of a software to identify and analyse marine traffic situations. *International Conference on Marine Simulation and Ship Manoeuvrability (MARSIM)*, Newcastle, United Kingdom.

Pallotta, G., Vespe, M. and Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, **15**(6), 2218–2245.

Patroumpas, K., Artikis, A., Katzouris, N., Vodas, M., Theodoridis, Y. and Pelekis, N. (2015). Event Recognition for Maritime Surveillance. *18th International Conference on Extending Database Technology (EDBT)*, Brussels, Belgium. 629–640.

Qi, L. and Zheng, Z. (2016). A measure of similarity between trajectories of vessels. *Journal of Engineering Science and Technology Review*, **9**(1), 17–22.

Rong, H. and Mou J. (2013). Predict maneuvering indices using AIS data by ridge regression. *International Workshop on Next Generation Nautical Traffic Models, Delft, The Netherlands.* 102–111.

Sun, P., Xia, S., Yuan, G. and Li, D. (2016). An Overview of Moving Object Trajectory Compression Algorithms. *Mathematical Problems in Engineering*, Article ID 6587309.

Wang, Y., Zhang, J., Chen, X., Chu, X. and Yan, X. (2013). A spatial–temporal forensic analysis for inland–water ship collisions using AIS data. *Safety Science*, **57**, 187–202.

Willems, N., Van De Wetering, H. and Van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, **28**(3), 959–966.

Wu, X., Mehta, A. L., Zaloom, V. A. and Craig, B. N. (2016). Analysis of waterway transportation in Southeast Texas waterway based on AIS data. *Ocean Engineering*, **121**, 196–209.

Zhang, S. K., Liu, Z. J., Cai, Y., Wu, Z. L. and Shi, G. Y. (2016). AIS trajectories simplification and threshold determination. *The Journal of Navigation*, **69**(4), 729–744.

Zhang, S. K., Shi, G. Y., Liu, Z. J., Zhao, Z. W. and Wu, Z. L. (2018). Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Engineering*, **155**, 240–250.