

# Vessel Track Recovery With Incomplete AIS Data Using Tensor CANDECOM/PARAFAC Decomposition

Changqing Liu and Xiaoqian Chen

*(College of Aerospace Science and Engineering, National University of Defense Technology, China)*  
(E-mail: chenxiaoqian@nudt.edu.cn)

Global analysis of vessel motion patterns has become possible using satellite-based Automatic Identification System (AIS). The concept of space-based AIS needs several satellites to provide complete coverage and high detection probability. However, in early development stages, often only one satellite is launched and due to its limitation of orbit and footprint, received AIS messages are discontinuous. In this paper, we have analysed real AIS data obtained by satellite to form a global maritime surveillance picture. Furthermore, we propose to take advantage of the tensor CANDECOMP/PARAFAC (CP) decomposition to analyse three mode characteristics of the data, which are location, vessel and time. For incomplete data, we exploit the link prediction technique based on tensor factorisation to recover vessel tracks in a specified area. A variant of temporal link prediction based on CP is presented. We illustrate the usefulness of exploiting the three-mode structure of AIS data by simulation, and demonstrate that the track recovery result has acceptable precision.

## KEY WORDS

1. Automatic Identification System.
2. Tensor Decomposition.
3. Link Prediction.

Submitted: 25 March 2013. Accepted: 31 May 2013. First published online: 3 July 2013.

1. INTRODUCTION. The Automatic Identification System (AIS) is traditionally a ship and shore-based maritime safety and traffic system imposed by the International Maritime Organisation (IMO). Since Wahl and Høyе (2003) from the Norwegian Defence Research Establishment, Forsvarets Forskningsinstitutt (FFI), presented the possibility of using space-based AIS receivers to help collect AIS messages, significant progress has been made in developing satellite AIS. Simply put, vessels equipped with AIS transponders transmit short omni-directional, line-of-sight messages that contain information about their identity, position, course and speed over ground, gyro course, Universal Time Coordinated (UTC) time, etc. Conventionally,

such information is received by shore stations and vessels in the vicinity to assist in vessel traffic safety and monitoring applications. However, AIS satellites can collect these data over a much broader area, hence providing a more comprehensive surveillance and management capability.

Satellites orbiting the earth have a designed orbital period and sensor footprint, so data can only be collected from vessels within the satellite footprint in one period. To avoid losing signals due to weak transmission power on a higher orbit, Low Earth Orbit (LEO) satellites are a preferred choice to collect as many messages as possible, while the shorter periods of LEO satellites can bring more vessels within the footprint. However, one satellite can only collect portions of the messages transmitted from global marine traffic. Even within the footprint, if vessel traffic is dense, some signals will be lost (Høyve et al., 2008). A possible solution is to enhance the data collecting capability by utilising a group of satellites, i.e. the concept of a “constellation”. It is estimated that full deployment of an AIS satellite constellation is due in the period 2012–2015 (Carson-Jackson, 2012). In this paper, we do not consider the concept of a constellation, but instead exploit real AIS data collected by one satellite, which, as mentioned above, are incomplete.

The incompleteness of data manifests in two forms. On the one hand, messages collected in a period are from only portions of the total vessels on the globe, mostly within the satellite footprint, yet not all. On the other hand, for some days the receiver on the satellite is powered down, thus the data obtained are not temporally continuous. The thrust of this work is that, if we can recover the temporally discontinuous pictures of global vessel traffic by extracting the inherent associative motion patterns of vessels from the spatial and temporal incomplete data, then we can schedule receivers on the satellite in a planned program to coordinate with memory and communication capability for multiple missions and improve the robustness of the satellite AIS capability. Noting that the number of vessels detected is in the thousands in a specific region and the data are temporal, we exploited the tensor decomposition technique to extract patterns of vessel motion. CANDECOMP/PARAFAC (CP) is adopted in this paper, which retains the natural data structure and yields a highly interpretable factorisation. Also, the CP model is unique under mild conditions (Kruskal, 1989), which will be convenient to our study. A gridding step is carried out to locate the vessels detected by the receiver, and therefore help establish the three way location-vessel-time array. An attempt to recover the data on missing days is made by a modified link prediction method based on tensor factorisation.

Our contributions can be summarised as follows:

- We utilise the received AIS data on a global range over a month to explicitly reveal vessel distributions on the ocean, and by this method, some busy routes are vividly detected.
- We focus on a specified region, which is a  $20^\circ \times 20^\circ$  (longitude  $\times$  latitude) area to the west of the Cape of Good Hope, establishing a three way tensor with a gridding step to obtain the location information of vessels, and extract motion patterns by performing a CP decomposition.
- We present a modified version of link prediction based on tensor decomposition to make it suitable for temporal data recovery. By utilising it on real data, we demonstrate that this method is valid and has good precision.

The paper is organised as follows: In Section 2, we briefly discuss relevant literature. In Section 3, we explain the process of decoding the AIS messages to the required form, and the location information of vessels detected over one month are demonstrated to present an appearance of the global traffic state. In Section 4, we perform a CP decomposition of the temporal AIS data tensor in a specific region. Firstly, we constructed the three-mode tensor in a location-vessel-time method, where vessels detected are denoted by their Maritime Mobile Service Identity (MMSI). The time mode has a span of more than one month with one day step size, and the location mode is vectorised by grid partition in the selected region. Secondly, we computed its CP model and obtained both overall and component level patterns. And thirdly, based on results obtained, we explain the corresponding vessel motion patterns. In Section 5, we present a modified version of link prediction based on CP decomposition, aiming at making it suitable for recovery problems, which can be considered as a combination of “forward” and “backward” predictions. Results are demonstrated to show the usefulness of this method. And in Section 6, we summarise major findings and discuss further possible work and extensions.

**2. RELATED LITERATURE.** Becky et al. (2012) have discussed the possibility of combining an AIS satellite constellation, a Synthetic Aperture Radar (SAR) satellite and an Electro-Optical (EO) satellite for routine and area surveillance, as well as in support of Maritime Domain Awareness (MDA). AIS data processing and fusion with other data sources have been studied for several years since the emergence of the concept of space-based AIS. Ristic et al. (2008) presented a statistical analysis of AIS data to find motion patterns and utilised the result to perform anomaly detection and motion prediction. Their work is based on the kernel density estimation technique, which provides an alternative way to study vessel motion patterns. However, the performance measurement provided in their work has to evaluate a probability numerically, which may not be very efficient. Høye (2004) found an empirical expression to approximate vessel detection probability, which is useful to determine the missing rate of AIS data and can be used further to improve analysis results. Bomberger et al. (2006) and Rhodes et al. (2007) apply artificial neural network methods to vessel motion pattern analysis, exploiting AIS data to train the model. Tun et al. (2007) classify vessel motion paths into different regions, based on their presented algorithm. This idea is somewhat similar to the location vectorisation step for constructing the data tensor in our paper, yet the regions obtained here are manually partitioned, making precision an issue. Several papers have exploited different techniques to perform anomaly detection for motion patterns of vessels or other mobile objects, including Laxhammar et al. (2009); Jemin et al., (2011); Laxhammar and Falkman (2010); Brax (2011); Lane et al. (2010) and Riveiro and Falkman (2010). Much of the other work concerning the concept of space-based AIS is focused on issues of vessel detection probability improvement (Eriksen et al., 2006), satellite constellation study (Becky et al., 2012; Cervera et al., 2011), and space-based AIS receiver design (Dahl, 2006). As Ristic et al. (2008) pointed out, the subject of motion pattern analysis is not unique to AIS data analysis used for maritime surveillance, actually originating in the field of computer vision. In this paper, we propose to use the tensor decomposition method to extract vessel motion patterns and

recover track lines, so we will next give a brief overview of some CP decomposition and link prediction related literature.

Tensor decompositions have applications in various disciplines. CP decomposition is a particular form that can be seen as a higher order version of matrix singular value decomposition. The CP model became popular due to the contributions of Carroll and Chang (1970) and Harshman (1970). For an introduction to many aspects of CP, refer to Kolda and Bader (2008). What led us to select CP to extract motion patterns is its uniqueness (under weak conditions) and interpretability. In recent years, software tools have been developed to facilitate the computation of tensor decomposition. The MATLAB Tensor Toolbox (Bader and Kolda, 2007) is adopted in this paper to implement the CP decomposition. Link prediction problems mainly focus on network analysis. However, in our work, by the location vectorisation operation, we transform the track recovery problem to a similar problem to link prediction. There are many strategies for link prediction problems. Node neighbourhood-based methods assume two nodes form a link if their respective neighbours have a large overlap; related work can be found in Adamic and Adar (2003). Ensemble-of-all-paths-based methods are alternative strategies; see Katz (1953). Other strategies exist, yet our approach is most related to that proposed by Dunlavy et al. (2011), where link data are temporal and the method is tensor-based. It is interesting to note that although the AIS data used in this paper are incomplete on some specific days, any vessel can only occupy one quadrilateral grid element or neighbouring elements at most at a given time range, so the problem to investigate in our study is a temporal link prediction problem, not a missing link prediction problem. One of the contributions by Dunlavy et al. (2011) is that they extend the temporal problem to a periodic temporal one, which means the time dimension has periodic patterns. For one step prediction, they give both heuristic and forecasting-based methods after CP decomposition. All the work indicates that as long as we can predict “forward”, why not “backward”? And further, by a combination of “forward” and “backward” predictions, we can recover missing temporal data by understanding motion patterns extracted with CP. The fact that there are no hindering factors for performing a backward analysis in the time dimension guarantees such a combination. In fact, we will demonstrate in subsequent sections that this idea is valid.

**3. AIS DECODING AND GLOBAL TRAFFIC REPRESENTATION.** In this section, we briefly introduce some technical characteristics of AIS codes that are relevant in this paper and explain which types of information are used.

**3.1. Issues With AIS Codes.** AIS data, typically including vessel identification (or MMSI), vessel type, speed, navigational status, draught, etc., are transmitted in the form of messages. The messages are identified by different leading codes. Some pre- and post- parts of the message codes are used for communication and auxiliary purposes. Used here is the datum part, which usually begins with the Message ID. Each piece of message is transformed to a string of 168-bit 0–1 codes. Only messages with ID 1, 2, 3 are decoded here for simplicity, and from the codes we extract the MMSI, position longitude and latitude, speed, course, and UTC time, which span from 13 May to 18 June 2012; some days inbetween are absent. The speed and course information are not directly used in data tensor construction, but for check and auxiliary purposes.

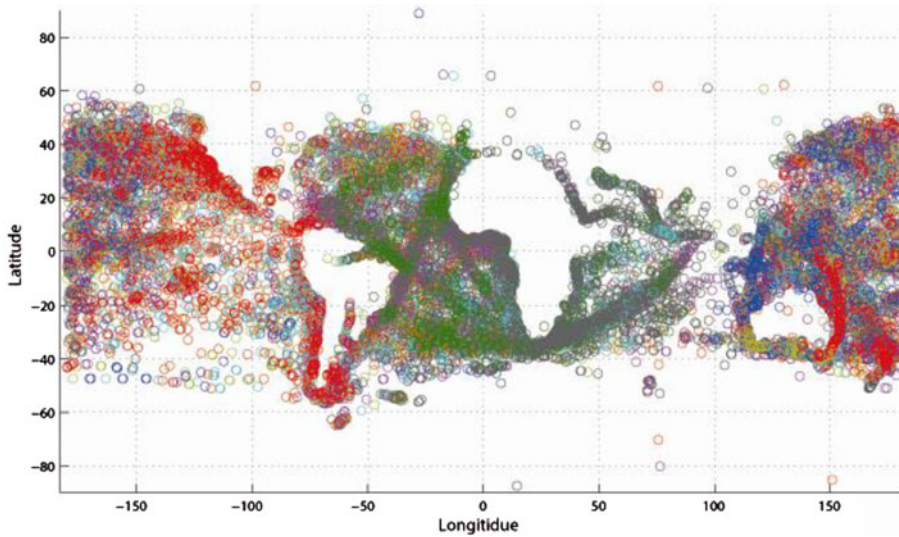


Figure 1. The global range vessel traffic state over a period of one month.

For the track recovery problem discussed in this paper, we limited the concerned region to a designated area, which is a  $20^\circ \times 20^\circ$  (longitude  $\times$  latitude) area to the west of the Cape of Good Hope. This allows us to monitor an area of interest closely, and by displacement and scalar operations, we can focus on any area on the ocean. Correspondingly, vessels outside of the concerned region are filtered. In the process of signal transmission, a small fraction of messages are truncated for abnormal reasons, and they are filtered too. Finally, we gridded the focused region, dividing it into many quadrilateral elements, and according to each vessel's longitude and latitude information, they are all distributed in the region. The magnitude of each element determines precision. This issue is discussed in detail in Section 4.

**3.2. Global Traffic Demonstration.** With accumulation of AIS data over many days, global voyage route patterns can be revealed. In Figure 1, we utilised a set of real AIS data received by satellite to depict a global vessel traffic picture. From Figure 1 the contours of Africa, South America and Australia can be clearly seen, and some busy voyage routes are also detected, for example the voyages across the Indian and Atlantic Oceans. Over a longer time period, routes that may be obstructed, especially the pole regions, may become apparent.

For a more local picture, however, it is necessary to investigate traffic patterns in detail. Fishing boats and cargo ships, for example, have very different motion patterns. Study of these patterns in a specified region may be helpful for management and safety purposes. In this paper, we mainly discuss the evolution of data characteristics in a region.

**4. REGIONAL MOTION PATTERN EXTRACTION.** In this section, to exploit the CP decomposition to perform motion pattern analysis, the procedure of constructing the tensor is illuminated first. Subsequently, we use the MATLAB Tensor Toolbox by Bader and Kolda (2007) to compute the CP model; we then explain the results. The work of this section is the basis of Section 5.

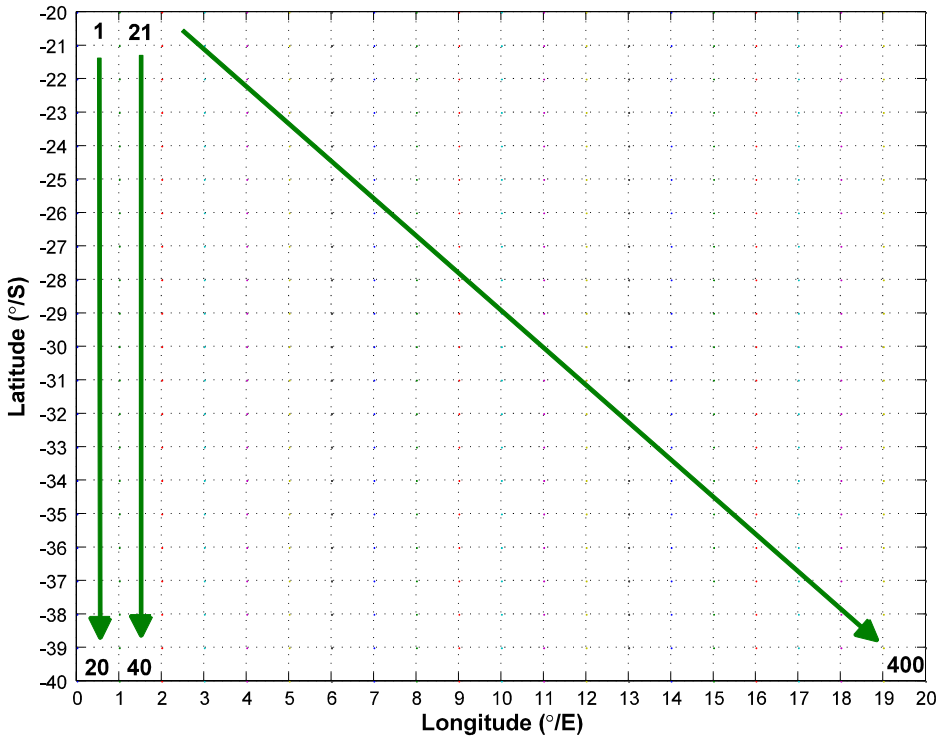


Figure 2. Grid partition in a  $20^\circ \times 20^\circ$  (longitude  $\times$  latitude) region to the west of the Cape of Good Hope.

4.1. *Tensor Construction.* A vessel's position data over time are organised as a tensor  $T$  of size  $I \times J \times T$ , where the three modes are location, vessel and time respectively. The region selected for surveillance is a  $20^\circ \times 20^\circ$  (longitude  $\times$  latitude) area to the west of the Cape of Good Hope. Instead of using longitude and latitude simultaneously to denote a vessel's position, we divide this region by use of gridlines, see Figure 2, referred to in this paper as location vectorisation.

The elements of tensor  $T$  are determined by

$$T(i, j, t) = \begin{cases} 1 & \text{if in element } i, \text{ there is vessel } j \text{ on day } t, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Noting the size of the elements can affect precision of analysis, we set it as  $1^\circ \times 1^\circ$  initially, taking speed and course of vessels into account, which can ensure that most vessels sail through several adjacent elements in subsequent days. The 400 elements are numbered as in Figure 2 and then arranged along the first mode of tensor  $T$ . A total of 726 vessels obtained from decoding the AIS messages with ID 1, 2, 3 are arranged along the second dimension according the number order of their MMSI. The time span is from 13 May to 18 June 2012, with the data for 23–26 May and 5 June missing.

4.2. *CP Decomposition Results.* We computed the CP model of tensor  $T$ , which has 50 components, utilising an Alternating Least Squares (ALS) algorithm from the



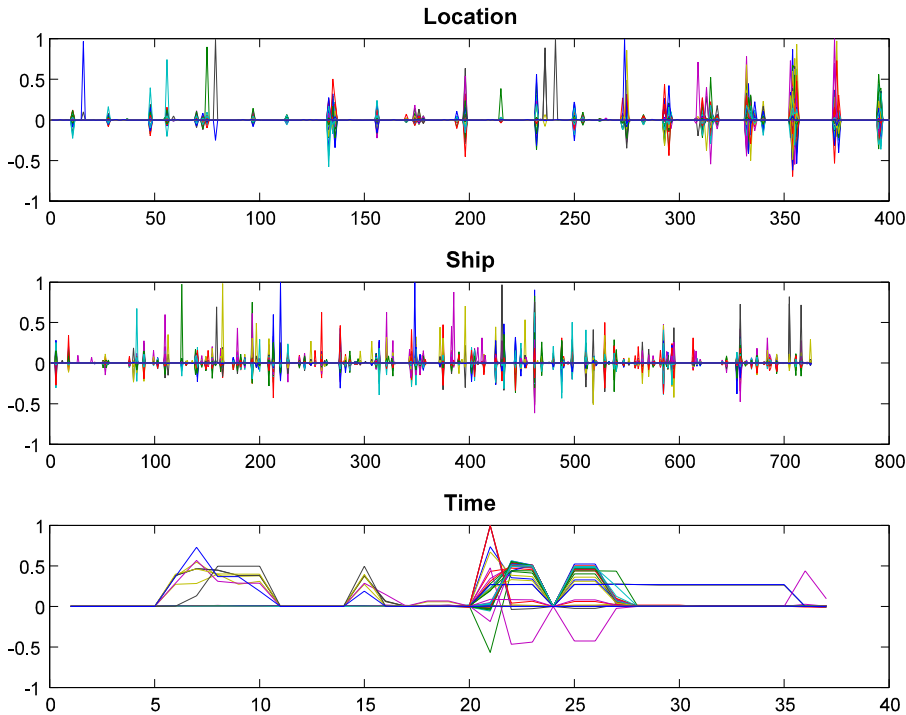


Figure 3. CP decomposition results of tensor  $T$ , with all 50 factors for each mode plotted overlying.

Tensor Toolbox by Bader and Kolda (2007), i.e.

$$T \approx \sum_k^{50} \lambda_k \mathbf{a}_k \circ \mathbf{b}_k \circ \mathbf{c}_k \tag{2}$$

where  $\circ$  denotes outer product computation,  $\lambda$  scales each component, and  $\mathbf{a}_k, \mathbf{b}_k, \mathbf{c}_k$  are factors. In Figure 3, the results are separately plotted, overlying all the obtained 50 factors together for the three modes, location, vessel and time. On the top panel for the location mode, we see that the result is in a discrete type as a whole, with some adjacent elements clustered together. A possible interpretation will be given in the next subsection. The middle mode for vessel is more complex than that for location. In order to extract the inherent patterns from it, we take two examples out of the 50 components, see Figures 4 and 5. In the bottom mode in Figure 3, we can determine the evolution of bustle pattern caused by all 726 detected vessels in this region of interest, noting the data for 23–26 May (day 11–14) and 5 June (day 24) are lost, which leads to an interval and a point of zero on the pattern. Moreover, the pattern of the first 5 days is also null which is actually due to the region of interest being outside of the satellite footprint during this period. How to recover this lost part of the pattern is our main focus in this paper, and we will deal with this in Section 5. Another point that needs attention is that in the CP model implemented here, the non-negative constraint is not imposed.

4.3. *Pattern Interpretation.* In the selected region of interest, vessels from or to Europe and South America around the Cape of Good Hope form three main routes,

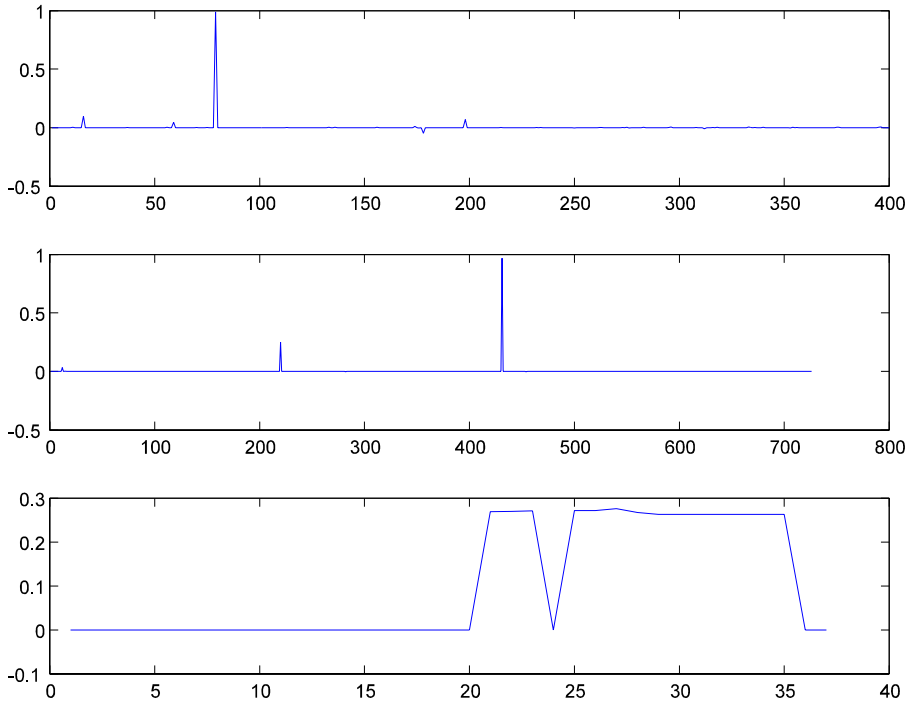


Figure 4. The #42 component of the CP decomposition.

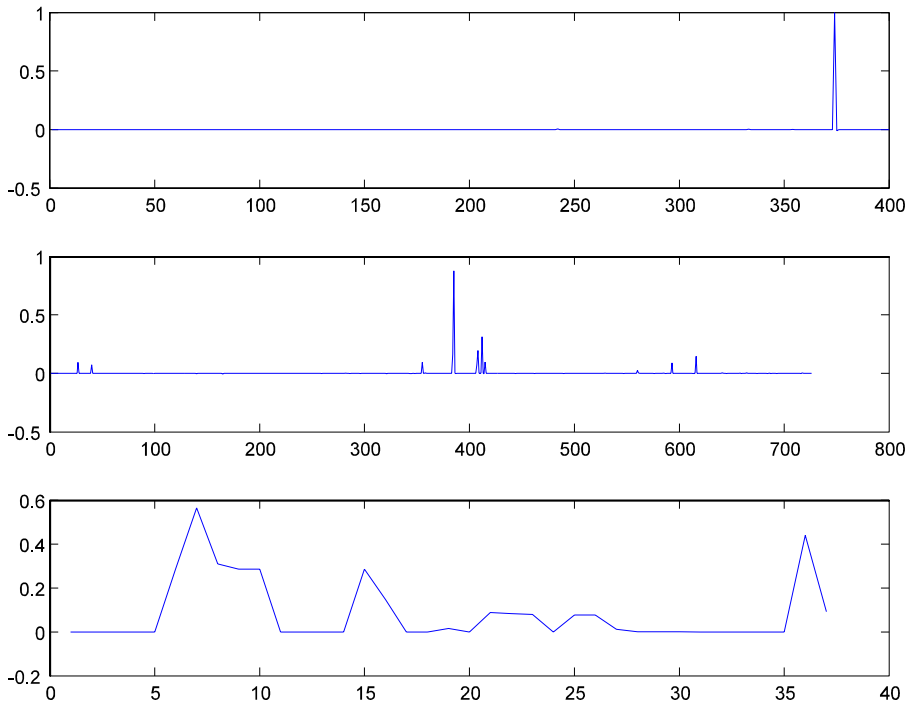


Figure 5. The #47 component of the CP decomposition.



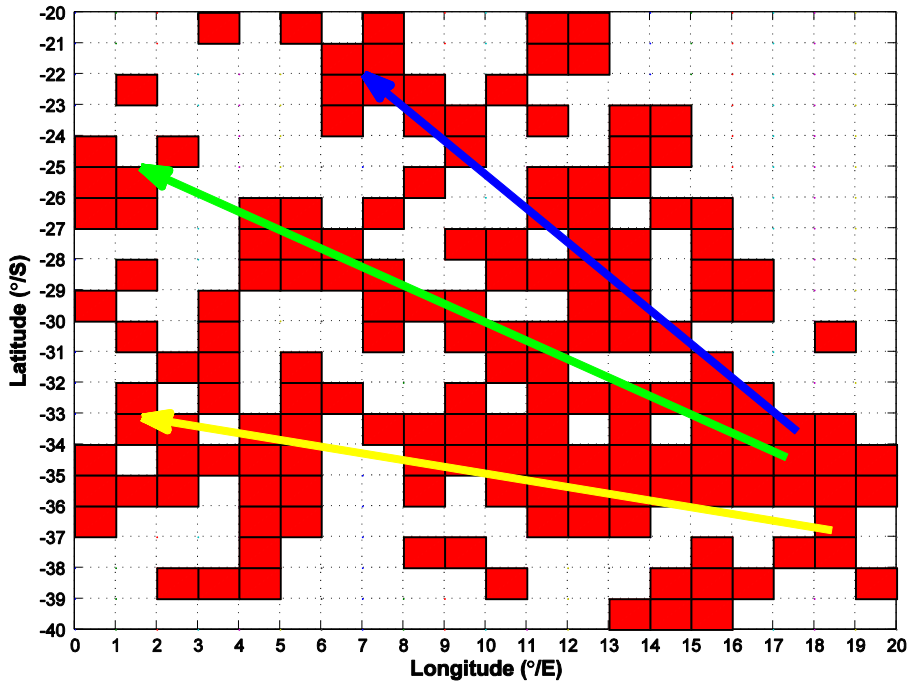


Figure 6. Approximation of the original routes with extracted patterns by CP. Quadrilateral element size is  $1^{\circ} \times 1^{\circ}$ .

i.e. European direction, Amazon direction, and Rio de Janeiro direction from the natural Cape, which can be clearly seen in Figure 1. Noting the elements are vectorised along the column of the region, which has been partitioned into a set of  $1^{\circ} \times 1^{\circ}$  quadrilateral elements, adjacent elements composing this route may be discretized, hence generating the pattern in the top panel of Figure 3. However, by the extracted pattern, we can to some extent approximate the original route with a precision determined by the size of element first designated. Figure 6 illustrates a coarse grained representation of the motion pattern in this region, while in Figure 7 the real pattern is shown for comparison. Note the dense dots to the right of the uppermost route denote vessels along the coastline, which do not form a specific route, see Figure 1 for reference. As an example, Figures 4 and 5 illustrate the 42nd and 47th components of the CP model. In Figure 4, the vessel with MMSI “477013300” which is numbered 431 in the vessel list locates in the No. 79 element over the period from 2 to 16 June. Although the missing 5 June data is obvious, the pattern extracted is only trivially affected. The same interpretation is true for Figure 5, but attention should be paid to the time mode, which spans the range of missing data, i.e. 23–26 May (day 11–14) and 5 June (day 24).

Combining with patterns extracted in other components, we conclude that it is viable to recover the lost route information by applying some extension technique to the incomplete data. Generally, the route information is contained in the first factors of CP decomposition, while the second factors group vessels in neighbouring elements, and the time factors reflect evolving characteristics of vessels. From the examples above, it can be seen that the CP decomposition extracts AIS data patterns well.

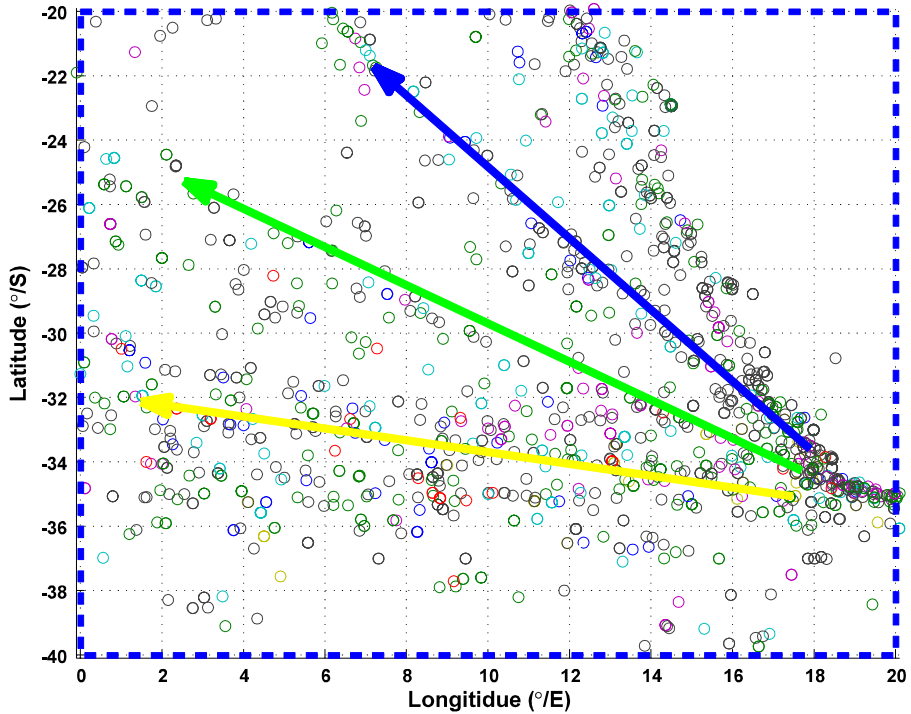


Figure 7. Real motion pattern of the 726 vessels in the region of interest over the time span.

5. TRACK RECOVERY FROM INCOMPLETE DATA. Based on the CP decomposition of the AIS data tensor and potential viability of recovering lost information from extracted patterns, we present a revised version of the link prediction technique based on CP decomposition in this section to accomplish the task of track recovery for vessels.

5.1 *Revised Link Prediction Based on CP.* Dunlavy et al. (2011) present heuristic and forecasting-based prediction methods that utilise temporal information extracted by CP; their tensor-based methods are applied to data with periodic patterns. Enlightened by their work, we propose a revised link prediction method based on CP decomposition to make it suitable for the track recovery application here. However, data in our focus have no periodic patterns, but a multiple-step based “recovery”, i.e. forward and backward predictions, is needed to account for missing data over several days. Therefore, we present a merging sequential one step prediction approach to handle this problem.

For multiple time steps missing data, motion patterns of vessels are affected more by recent data in light of motion coherence; therefore we use the latest 3 days’ data to predict a subsequent day. We construct an  $I \times J \times 3$  tensor  $T^D$  first which contains the latest 3 days’ data, then compute the CP decomposition using Tensor Toolbox, resulting

$$T^D = \sum_k^{50} \lambda_k^D \mathbf{a}_k^D \circ \mathbf{b}_k^D \circ \mathbf{c}_k^D \quad (3)$$

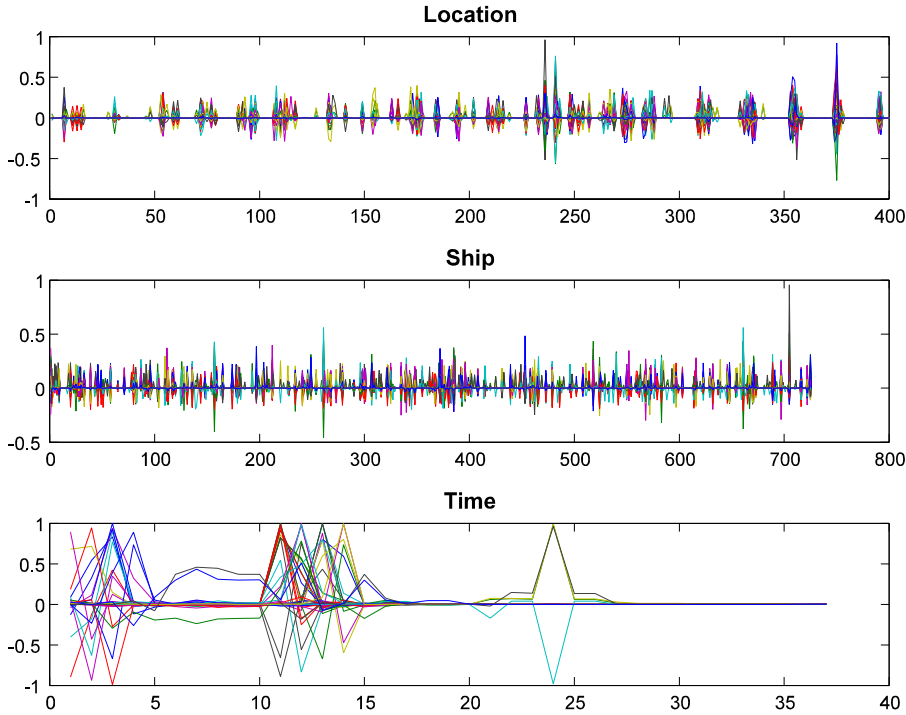


Figure 8. CP decomposition results of the full tensor  $F$ , with all 50 factors for each mode plotted overlying. Location and vessel patterns are dense, compared with those in Figure 3, and missing data in time mode are recovered by a combination of “forward” and “backward” predictions.

Here, the superscript denotes the last index of the three days in the original  $I \times J \times T$  tensor. And further, we give a one-step prediction score for day  $D+1$  as follows,

$$S^{D+1} = \sum_k^{50} \gamma_k^D \lambda_k^D \mathbf{a}_k^D \mathbf{b}_k^{DT} \tag{4}$$

where

$$\gamma_k^D = \frac{1}{3} \sum_t^3 \mathbf{c}_k^D(t) \tag{5}$$

Subsequently, as for day  $D+2$ , we construct a new  $I \times J \times 3$  tensor  $T^{D+1}$ , using the latest 2 days’ real data and obtained score on day  $D+1$ . The remaining mathematics carry through as before, until we obtain predictions for all days. Note that as this proceeds, data used to construct new tensors may be all from obtained scores, so long as the missing period is temporally equal or greater than 4 days, such as 23–26 May given in Section 4.

The procedure above can be seen as a “forward” prediction. However, the data collected after the missing time span can also be used to accomplish a “backward” prediction, taking full advantage of information on both sides. There is no essential difference in the mathematics to perform such a process, only to inverse the direction

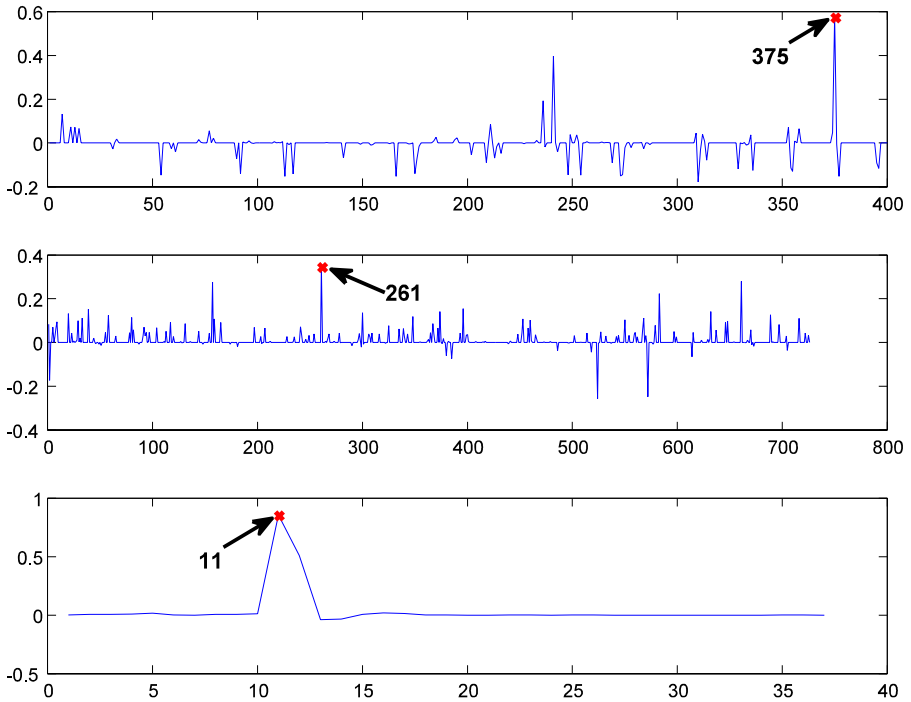


Figure 9. The #6 component of the CP decomposition of  $F$ , which specially illustrates a recovered pattern in the missing period from day 11 to day 14.

of constructing  $I \times J \times 3$  tensors, i.e. for a missing period over day  $D+1$  to  $D+L$ , we compute scores  $S^{D+L}$  for day  $D+L$  by constructing the tensor using real data on day  $D+L+3$ ,  $D+L+2$ ,  $D+L+1$  first.

For the task of recovery, we merged the forward and backward predictions to recover the data. Neither of the two forms has an advantage over the other, so we took the mean of “forward” and “backward” score arrays to construct recovery tensor  $M$  of size  $I \times J \times L$ , where

$$M(:, :, j) = \frac{1}{2}(S^j + S^j) \quad j = D+1, D+2, \dots, D+L \quad (6)$$

in the MATLAB notation. Then the original data tensor with vacancies due to data missing was filled with the above recovery tensor fragment to form a new “full” one,  $F$  which is used to regenerate vessel tracks.

5.2. *Simulation.* Based on the revised link prediction method proposed above, we illustrate the results in Figures 8 to 11. Figure 8 illustrates a full view of patterns produced with the full tensor decomposition. Compared with that in Figure 3, it is obvious that both the location and vessel patterns are denser than before, which means that more elements are occupied and some vessels are recovered to show up in elements during the missing days. The bottom subplot more clearly demonstrates this, where a pattern of days that are originally absent in Figure 3 are present already. The intense rise and fall in the time mode pattern is due to the fact that many more new links have appeared in the result.

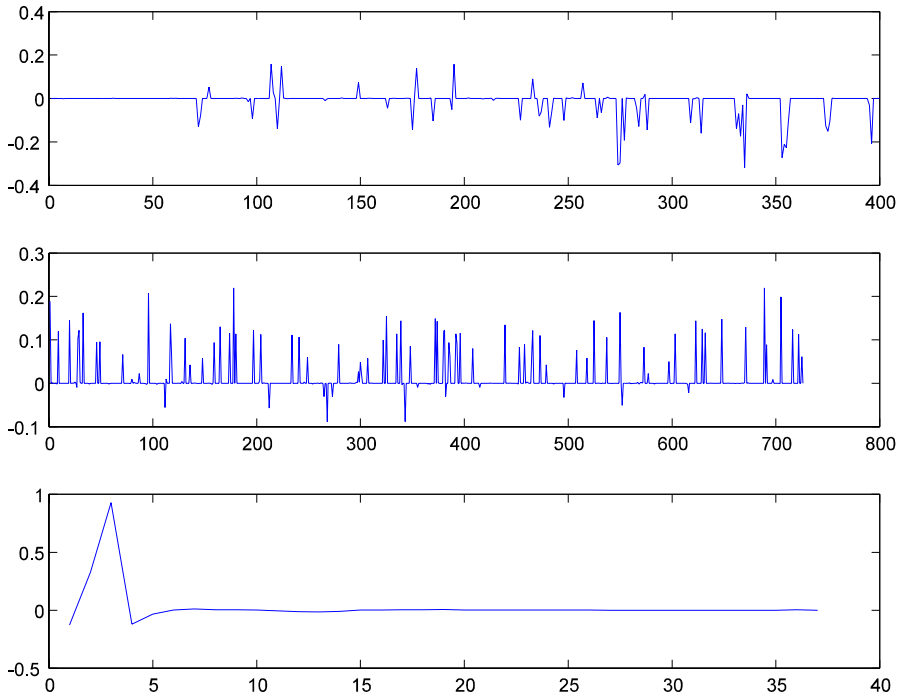


Figure 10. The #15 component of the CP decomposition of  $F$ , which specially illustrates a recovered pattern in the missing period from day 1 to day 5.

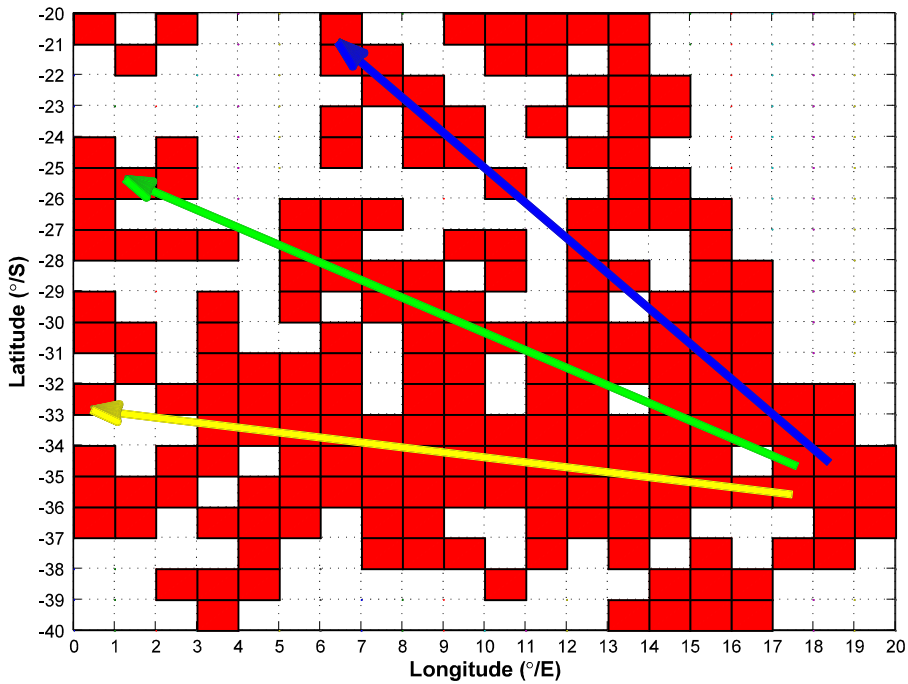


Figure 11. Recovery of the original three routes with patterns extracted by CP of  $F$ .

Table 1. Links Predicted With a Forward Form of the Proposed Method.

Day #	Location	Vessel	Possible links	Real links	Density (%)
10	35	52	1300	54	4.15
<b>11</b>	30	36	1080	107	<b>9.91</b>
12	36	33	1188	133	11.2
13	30	60	1800	452	25.11
<b>14</b>	33	60	1980	683	<b>34.49</b>
15	29	41	1189	42	3.53

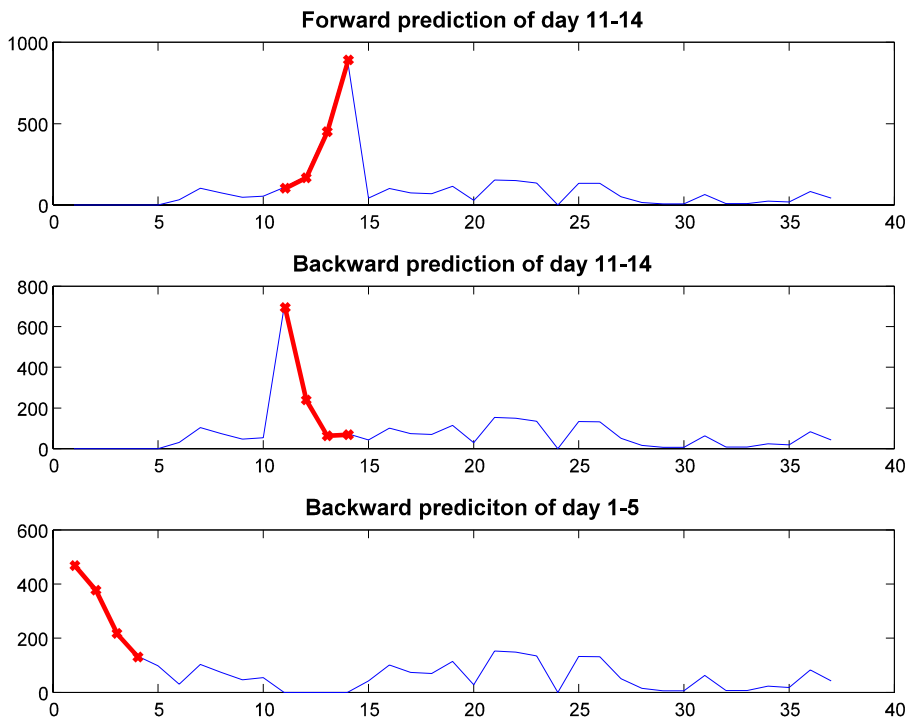


Figure 12. Trend of the number of links in forward and backward forms separately during the missing day periods.

Furthermore, in order to discover the underlying mechanism of how the pattern is recovered, we demonstrate two CP decomposition components of the full recovered tensor  $F$  in Figures 9 and 10, where one reveals which vessels are located in which quadrilateral elements from day 11 to day 14, i.e. 23–26 May, and the other reveals which vessels are located in which elements during the first 5 days. With many other components containing such analogous information composited together, a more legible route pattern is given in Figure 11.

5.3. *Discussion.* When predicting links in a further time step using the method proposed in this paper, new links will arise, especially as the sequential iteration proceeds. This causes the fluctuation shown in the bottom of Figure 8. Table 1, for example, illustrates the number of links in the successive days from 22–27 May, predicted using the forward form of the proposed method, noting the four missing

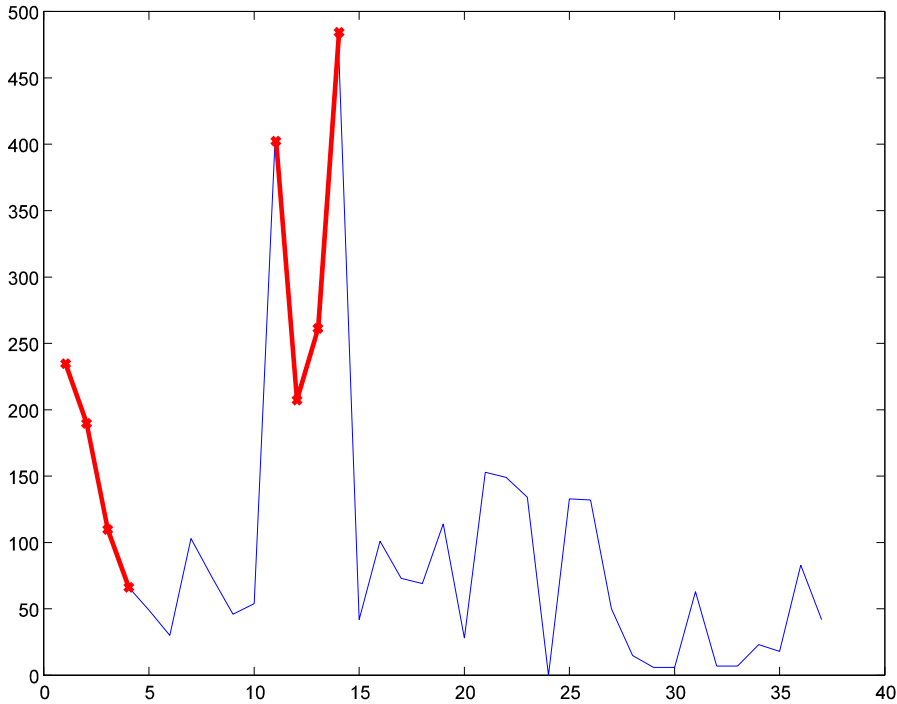


Figure 13. Number of links during missing days with forward and backward forms merged, while the first five days only filtered.

days therein. It can be seen that the proportion of predicted links increases in further prediction steps, in the forward form. The backward form of prediction has an inverse trend, see Figure 12. Thus, we take the mean of two forms to alleviate the effect of monotony. For the other form of missing data, i.e. the first five days due to the region being out of the footprint of satellite, we filter out those with scores less than two. The merged result is shown in Figure 13.

In the bottom subplot of Figure 8, the extracted pattern reflects well that inherent in the recovered tensor  $F$ , which are in accordance with Figure 13. If the missing data recovered are added to real data, we see more vessels are detected to be active as shown in the middle subplot of Figure 8, which will make a more explicit pattern of vessel track routes. It can be seen that detected elements are denser in Figure 11 than in Figure 6.

For the task of single vessel tracking, it is possible to determine which element they are located in during the period when real data are absent. When analysing each component obtained from the CP decomposition of tensor  $F$ , which vessels are in which elements is clear, as demonstrated in Figures 9, 10. For example, in the 6th component of the CP model illustrated in Figure 9, we know that on day 11, the vessel numbered 261 (whose MMSI is 354623000, actually) is most likely in element 375. Note an element of  $1^\circ \times 1^\circ$  in the region of interest is an area of approximately  $100 \text{ km} \times 100 \text{ km}$ , which is the same magnitude of the least range a vessel will manoeuvre in a day, after checking the speed data of all the vessels, so the results obtained in this paper have an acceptable precision.



6. **CONCLUSION.** In this paper, we use the tensor decomposition tool to analyse a set of AIS data. Traditionally, shore-based AIS data are used for short-range maritime traffic service and safety applications. However, the new concept of space-based AIS has extended this capability to a global range. We utilised the real data received by satellites over a period of a month or so to depict a global vessel traffic pattern to demonstrate its advantage over traditional techniques. Nevertheless, satellite-based AIS has limitations. In the early stage of satellite AIS, with only one satellite in orbit and the restriction due to its orbit and footprint may lead to incomplete data. The analysis of these data is the main focus of this paper, and in particular, we discuss the task of track recovery of missing data.

We constructed three-way tensors to represent location, vessel identity and time information. Before that, we first proposed to vectorise the location information to a set of gridded elements, making the first mode of the tensor one-dimensional. By performing the CP decomposition, we primarily extracted motion patterns inherent in the data. In the case of incomplete data, we proposed to take advantage of link prediction techniques to first recover the full data tensor, and then solve the problem of which vessels are in which elements during the missing days. A revised version of link prediction based on CP is given, where we merge the forward and backward prediction links to handle the problem of track recovery. Simulation results illustrate the applicability of this method.

The concept and techniques of satellite-based AIS are promising, yet how to effectively exploit the data received in the global range is challenging. In this paper, we have innovatively applied the CP decomposition that is efficient in the field of data mining to AIS data analysis.

#### ACKNOWLEDGMENT

The authors would like to thank Mr. Lihu Chen, for his help in AIS message decoding.

#### REFERENCES

- Adamic, L.A. and Adar, E. (2003). Friends and Neighbors on the Web. *Social Networks*, **25**(3): 211–230.
- Bader, B. W. and Kolda, T. G. (2007). MATLAB Tensor Toolbox, Version 2.2. <http://csmr.ca.sandia.gov/tgkolda/TensorToolbox/>.
- Becky, C., Foley, K.C. and Chandler, S. (2012). The Ability of a Small Satellite Constellation to Tip and Cue Other Commercial Assets. *Proceedings of 26<sup>th</sup> Annual AIAA/USC Conference on Small Satellites*, Logan, UT.
- Bomberger, N.A., Rhodes, B.J., Seibert, M. and Waxman, A.M. (2006). Associative Learning of Vessel Motion Patterns for Maritime Situation Awareness. *Proc. Int. Conf. Information Fusion*, Florence, Italy.
- Brax, C. (2011). *Anomaly Detection in the Surveillance Domain*. University of Skovde.
- Carroll, J.D. and Chang, J.J. (1970). Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalisation of ‘Eckart-Young’ Decomposition. *Psychometrika*, **35**, 283–319.
- Carson-Jackson, J. (2012). Satellite AIS - Developing Technology or Existing Capability? *Journal of Navigation*, **65**(2): 303–321.
- Cervera, M.A., Ginesi, A. and Eckstein, K. (2011). Satellite-based Vessel Automatic Identification System: a Feasibility and Performance Analysis. *Int. J. Satell. Commun. Network.*, **29**, 117–142.
- Dahl, O.F.H. (2006). Space-Based AIS Receiver for Maritime Traffic Monitoring Using Interference Cancellation, Norwegian University of Science and Technology.
- Dunlavy, D.M., Kolda, T.G. and Acar, E. (2011). Temporal Link Prediction Using Matrix and Tensor Factorisations. *ACM Trans. Knowl. Discov. Data*, **5**(2).

- Eriksen, T., Høye, G., Narheim, B. and Meland, B.J. (2006). Maritime Traffic Monitoring Using a Space-Based AIS Receiver. *Acta Astronautica*, **58**, 537–549.
- Harshman, R.A. (1970). Foundations of the PARAFAC Procedure: Models and Conditions for an “Explanatory” Multi-modal Factor Analysis. *UCLA working papers in phonetics*, **16**, 1–84.
- Høye, G. (2004). Observation Modelling and Detection Probability for Space Based AIS Reception-Extended Observation Area. FFI/RAPPORT-2004/04390, Forsvarets Forskningsinstitutt.
- Høye, G.K., Eriksen, T., Meland, B.J. and Narheim, B.T. (2008). Space - Based AIS for Global Maritime Traffic Monitoring. *Acta Astronautica*, **62**, 240–245.
- Jemin, G., John, C.L., Tarunraj, S. and Adam, F.M. (2011). Anomaly Detection Using Context-Aided Target Tracking. *Journal of Advances in Information Fusion*, **6**(1): 39–56.
- Katz, L. (1953). A New Status Index Derived From Sociometric Analysis. *Psychometrika*, **18**(1): 39–43.
- Kolda, T.G. and Bader, B.W. (2008). Tensor Decompositions and Applications. *SIAM Review*.
- Kruskal, J.B. (1989). Rank, Decomposition, and Uniqueness for 3-way and N-way Arrays. *Multway Data Analysis*, Amsterdam, 7–18.
- Lane, R.O., Nevell, D.A., Hayward, S.D. and Beaney, T.W. (2010). Maritime Anomaly Detection and Threat Assessment, *2010 13th Conference on Information Fusion*, 1–8.
- Laxhammar, R., Falkman, G. and Sviestins, E. (2009). Anomaly Detection in Sea Traffic - a Comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *Proceedings of 12th International Conference on Information Fusion*, Seattle, WA, 756–763.
- Laxhammar, R. and Falkman, G. (2010). Conformal Prediction for Distribution - Independent Anomaly Detection in Streaming Vessel Data. *Proceedings of Stream KDD'10, ACM*, Washington, DC, 47–55.
- Rhodes, B.J., Bomberger, N.A. and Zandipour, M. (2007). Probabilistic Associative Learning of Vessel Motion Patterns at Multiple Spatial Scales for Maritime Situation Awareness. *Proc. Int. Conf. Information Fusion*, Quebec, Canada.
- Ristic, B., La Scala, B., Morelande, M. and Gordon, N. (2008). Statistical Analysis of Motion Patterns in AIS Data: Anomaly Detection and Motion Prediction. *Proceedings of the 11<sup>th</sup> International Conference on Information Fusion*, 1–7.
- Riveiro, M. and Falkman, G. (2010). Supporting the Analytical Reasoning Process in Maritime Anomaly Detection : Evaluation and Experimental Design. *Proceedings of the 14<sup>th</sup> IEEE International Conference on Information Visualisation (IV'10)*, 170–178.
- Tun, M.H., Chambers, G., Tan, T. and Ly, T. (2007). Maritime Port Intelligence Using AIS Data. *Proc. RNSA Security Technology Conference*, Melbourne, Australia.
- Wahl, T. and Høye, G.K. (2003). New Possible Roles of Small Satellites in Maritime Surveillance. *Proceedings of Fourth IAA Symposium on Small Satellites for Earth Observation*, Berlin, Germany.