Fractal Patterns in Coastal Detections on Approaches to Canada

Min Jing Liu, Peter Dobias, and Cheryl Eisler
Centre for Operational Research and Analysis,
Defence Research and Development Canada,
101 Colonel By Dr., Ottawa, ON, K1A0K2, Canada

Abstract

Understanding spatial and temporal patterns in shipping traffic and ship movements on the approaches to Canada will enhance the ability of analysts to identify deviations from expected behaviour. The present paper expands on an earlier analysis of ship detections in Canadian coastal regions using RADARSAT-2 (RS2) and Satellite Automated Identification System (S-AIS) data. It addresses several limitations of the previous approach, and expands the analysis by looking at the properties of sub-regions, in addition to the overall coastal regions. The S-AIS is dependent on ships' self-reporting (which is mandated for cargo vessels over 300 gross tonnes and for all passenger-carrying vessels), while RS2 collects images over time using Synthetic Aperture Radar (SAR) thus limiting the need for the cooperation from tracked vessels. Employment of spatial entropy and the fractal dimension, calculated using the differential box-counting method, as measures of randomness in the spatio-temporal distributions of ship detections are presented in the paper; both quantities suggest a presence of non-random patterns of behaviour with annual changes possibly attributable to known causes (such as seasonal changes in shipping routes and fishing). Further work will include exploring additional measures, and the use of the existing measures to determine likelihood maps for expected behaviour for different vessel types.

INTRODUCTION

In order to protect their sovereign interests, coastal countries strive to monitor shipping traffic in their territorial waters and exclusive economic zones. This is required to ensure that the activities of the vessels adhere to international and national laws and regulations. Analysis of the patterns in the maritime detection data could help analysts to better understand the behaviour of vessels on the approaches to Canada in order to estimate what may constitute aberrations from expected behavioural patterns (and thus be identified as suspicious). There are two questions that the analysts attempt to answer, to which the proposed approach hopes to provide some response. The first question seeks to understand the detection reliability; in particular, what is the likelihood of false positives (detections not corresponding to actual vessels)? The second question is whether there are any vessels exhibiting some kind of anomalous behaviour? Previous research (Dobias et al., 2015) suggested that the maritime detection data taken off the East and West coasts of Canada may indeed exhibit an intrinsic non-linear geo-spatial structure reflecting non-random patterns. Understanding these patterns at the global scale, as well as for particular sub-regions, could then enable future development of probability distributions for detections in a given sub-region at a particular time of year. However, potential methodological limitations associated with the saturation of the data patterns

due to the conversion of latitude-longitude data to an integer grid were identified; these limitations reduced the confidence in the findings reported in (Dobias et al., 2015).

There are several possible approaches to understanding the intrinsic structure of the geographic distributions of vessel detections. Spatial entropy and fractal dimension were selected as possible indicators of fractal properties for RADARSAT-2 (RS2) and Satellite-based Automated Identification System (S-AIS) detections. The data can then be filtered into time periods of interest to determine if seasonal effects are a factor in the pattern development. This paper discusses the practical implementation of the proposed methodology, including the impacts of approximations inherent in mapping continuous variables on integer grids. It looks at the data along each coast overall, as well as for a selected set of sub-regions. Trends in spatial entropy and the fractal dimension are discussed, and some possible explanations in terms of known shipping and fishing trends are provided. However, this paper is not intended to provide in-depth analysis of the causes of the observed trends; its focus is on the practical implementation of the methodology for maritime data.

The paper is organized as follows. At first, the concept of spatial entropy and fractal dimension is introduced. Then, the data sources are presented, followed by a discussion of the practical implementation of the two measures. In particular, dependence of the results on data density and the convergence of these measures are explored. The observed trends for spatial entropy and the fractal dimension of ship detections along the East and West coasts of Canada are shown and discussed. This is then followed by extended analysis of spatial entropy and the fractal dimension for several coastal sub-regions. The global measures, such as the fractal dimension and spatial entropy require a relatively significant shift in the geospatial distribution of the data to produce visible variations. However, it is shown in this paper that the analysis over the time can determine what constitutes expected (normal) behaviour; unresolved subsets could then be compared to this normal to determine if they might be aberrations.

SPATIAL ENTROPY AND FRACTAL DIMENSION

In 2004, Ilachinski (2004) proposed a specific form of Shannon entropy for the spatial distribution of soldiers on a battlefield:

$$S = -\sum p_i \ln p_i \qquad (1)$$

In Eq. 1 p_i is the probability of finding a detection in the i-th cell of a grid covering the battlefield. This metric for determining the randomness of the patterns of the spatial distribution of the individual entities can be applied to ship position reports in the area of interest.

In general, a non-dispersed pattern typically yields low values of spatial entropy, while disorganized, almost uniform patterns would display high spatial entropy values. Theoretically speaking, the maximum value for spatial entropy defined by Equation (1) is $S_{max} = \ln N$, where N is the total number of soldiers on the battlefield (i.e., points in the two dimensional space). In this paper, "soldiers" are substituted by RS2 and S-AIS detections. In addition, to account for different numbers of detections at

different times (and for each coast), the value of spatial entropy was normalized so that the maximum value of $S_{max} = 1$.

In order to compute spatial entropy, the area of interest of linear size B is split into a number of sub-blocks of linear size b (their area would be b^2). Then, assuming that N_i out of N detection points are in the i^{th} sub-block, the probability of a detection being in the i^{th} block will be:

$$P_i(b) = N_i(b) / N \tag{2}$$

and Equation (1) becomes:

$$S(b) = \frac{1}{2\ln(B/b)} \sum_{i=1}^{(B/b)^2} P_i(b) \ln(1/P_i(b)) \tag{3}$$

The value of spatial entropy is dependent on the sub-block size. If the division is too coarse, most of the blocks will contain some detection points, leading to an appearance of uniformity and a consequent increase in spatial entropy. A similar effect would be achieved if there are too many blocks, and the division is too refined. In this case, most of the blocks would be empty; thus creating an appearance of greater organization of the set. However, the impact of the latter would be likely less than the impact of too coarse a subdivision. Another limitation of the practical implementation of too fine a grid is the computational cost required to evaluate each grid point. Later in this paper, the issue of the selection of an appropriate granularity of the division of the area to blocks will be discussed in greater detail.

An alternative metric used by Ilachinski (2004) is the fractal dimension D_F, which measures the spatial distribution of units by providing the minimum number of variables needed to specify a given pattern (Ilachinski, 2004). To calculate the fractal dimension, the overall area of the size L is again divided into equal boxes. In general, the relationship between the size of a box ε and the minimum number $N(\varepsilon)$ needed to cover all the points is a power law:

$$\mathrm{N}(\varepsilon) = (\mathrm{L}/\varepsilon)^{D_F} \tag{4}$$

where D_F is the computed via the box-counting method. For uniformly distributed points, the entire area has to be covered for all box sizes, yielding the simple $N(\varepsilon) = (L/\varepsilon)^2$ relationship characteristic for a non-fractal, two dimensional space. When Equation (4) is rewritten by taking the logarithm of both sides, the formula for D_F is obtained:

$$D_F = \lim_{\varepsilon \to 0} \frac{\ln(\mathrm{N}(\varepsilon))}{\ln(\mathrm{L}/\varepsilon)} \tag{5}$$

In practice, the value of ε will be finite, but it should be reasonably small. To calculate the fractal dimension of the ship detections, a method called differential box-counting (Sarkar and Chaudhuri, 1994) is employed in this paper. In this approach, multiple values of ε are used; the slope yielding the D_F is then found using a linear least squares fit. This approach is more robust than using a single, though small, ε value.

The fractal dimension of the system, when computed via the box-counting technique, is related to spatial entropy (Sprague and Dobias, 2008). The main difference is that spatial entropy considers the probability of finding a data point within a sub-block; hence, it is dependent on the actual point

density in sub-blocks. On the other hand, the fractal dimension considers simply the presence or absence of a point within the sub-block and thus is independent of the actual point density distribution within sub-blocks.

RS2 AND S-AIS DATA SETS

The data sets used in the analysis herein were obtained from two sources. The first source was provided by the Canadian Polar Epsilon project, with ship detections from RADARSAT-2. RS2 is a commercial, sun-synchronous polar-orbiting synthetic aperture radar satellite providing imagery-based surveillance (Vachon et al., 2014). The Canadian RS2 satellite is owned and operated by MacDonald Dettweiler and Associates. There are two primary SAR imaging beam modes used for ship detection, with 450 km and 530 km wide swaths; these were developed for the Polar Epsilon project, and are used in addition the standard imaging beam modes. The area of surveillance is constrained by the orbit and beam modes, and the persistence is limited. RS2 uses active sensing, so the detection of ships does not require regulatory participation of the observed vessels.

The second was the commercial satellite-based AIS (S-AIS) dataset. AIS is a radio-based transponder system for self-reporting, designed for enhancing safety of navigation at sea. It provides the vessel position and static identifying information for collision avoidance. AIS transponders are mandated by the United Nations Safety of Life at Sea (SOLAS) Convention (SOLAS) for all ships over 300 gross tonnes (GT), and all passenger-carrying vessels; other vessels may use it as well. The Government of Canada purchases commercial S-AIS from exactEarth® for the purpose of enhancing maritime domain awareness (exactEarth, 2014). Both datasets cover the area off the East and West coasts of Canada for the year 2014 (Figure 1). Note that there were significantly more AIS detections, since these are updated far more frequently than RS2 detections, and always cover the entire region while RS2 detections are obtained for only a subset of the monitored region for each pass.

The data consisted of latitude and longitude locations of vessels detected by RS2 or S-AIS. Assuming that the original data was a fractional data set, the Complex Adaptive Systems Analysis (CASA) software developed by Defence Research and Development Canada had previously been used to compute the spatial entropy and fractal dimensions of the detections (Dobias et al., 2015). Because the fractal dimension and spatial entropy metrics were initially implemented in CASA using an integer grid of a limited size (up to a maximum of 200x200 cells), the large datasets appeared to be more uniform than they really were. In that implementation, the data had to be mapped to an integer set using an affine transformation of the latitude and longitude values to cover as much of the grid as possible. This led to a potential saturation for large data sets (such as AIS) since multiple nearby points would be mapped to the same grid cell. In contrast, this paper reflects the analysis using more robust implementation, thus avoiding the potential saturation issues. As will be discussed below, this led to much clearer identification of non-uniform structures in the data, and enabled also the additional analysis focusing on sub-regions of Canadian coastal approaches.

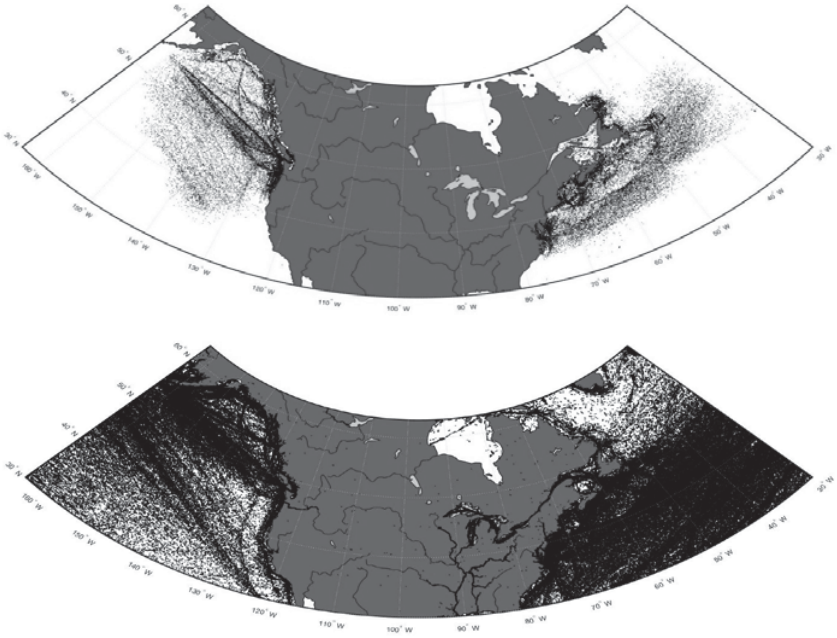


Figure 1. RS2 (top) and S-AIS (bottom) ship detections (black dots) off of the Canadian coasts for 2014.

For large datasets, such as the RS2 and AIS data (Table 1), in addition to the saturation problem, CASA had problems with insufficient memory, since it stored the entire dataset during computations. In order to bypass these limitations, the two metrics were re-implemented in MATLAB®. The main algorithms for computing spatial entropy and fractal dimension remained the same as in CASA. However, in the MATLAB® implementation, it was not necessary to use integer input, and the entire dataset did not need to be stored during the computations.

Table 1. Monthly detection numbers from RS2 and AIS sensors.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
RS2	2204	6455	7926	9549	13287	15837	13593	13079	13081	11329	10984	5439
AIS	265879	237846	283510	374585	474499	484496	404727	597132	546326	524952	296507	n/a

SPATIAL ENTROPY OF COASTAL DETECTIONS (BROKEN DOWN BY COAST)

As is mentioned in the introduction, one of the issues for a practical implementation of the fractal dimension and spatial entropy was the dependence of their values on the size of the grid. This is especially true in the case of spatial entropy because, unlike the fractal dimension, it only

used one length of the grid blocks, b. Therefore the dependence of spatial entropy values S on the grid size was investigated. This dependence can alternatively be expressed as a dependence of S on the number of grid cells, or a dependence on average detection point density.

The number of sub-blocks on the pattern edge, x, can be expressed in terms of the grid block linear size b as:

$$x = b / B \tag{6}$$

where B is the linear size of the entire pattern. In order to understand the dependence of spatial entropy on the number of cells $S = S(x)$, a test dataset corresponding to randomly (near-uniformly) distributed points was first analyzed. A caveat needs to be included here: while randomly-generated points will be distributed nearly uniformly, there will always be some clustering. Thus the expected value of spatial entropy would be slightly below the maximum value $S_{max} = 1$ that would be obtained for uniformly distributed data.

The analyzed pattern covered a two dimensional battlefield with the length of pattern edge of 100. The entire data set contains 10,000 data points; as is mentioned above, randomly generated data will actually feature some degree of clustering; therefore the expected value of S should be less than one. Figure 2 shows the shape of the spatial entropy (blue diamonds) for different value of x corresponding to different average point densities. The secondary axis shows the percentage of grid cells with no point in them (red dashes).

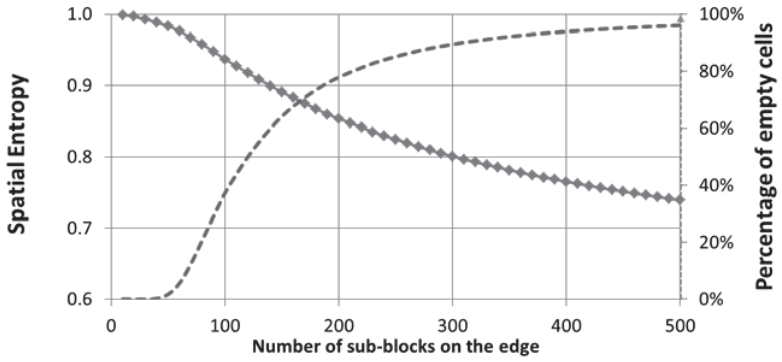


Figure 2. Spatial entropy for a random, near-uniformly distributed pattern.

If there were too few boxes (the average point density too high), the spatial entropy values remained close to one, creating an appearance of a uniformly distributed set. On the other hand, if there were too many boxes (with the corresponding average point density very low), spatial entropy would drop much more than expected for a random set. It was determined that the most reasonable values were obtained around the inflexion point corresponding to densities between one and two points per grid cell. Thus a rule-of-thumb using an average point density of two points per grid cell was employed in the analysis. For the analysis purposes a common grid resolution was required across all data but the number of detection varied from month to month. Thus, in order to account for the varying monthly detection numbers, and to enable using the same value across all months, the number of grid cells was defined by an expression:

$$x = \mathrm{int}\left(\min\left(\sqrt{N_i / 2}\right)\right) \qquad (7)$$

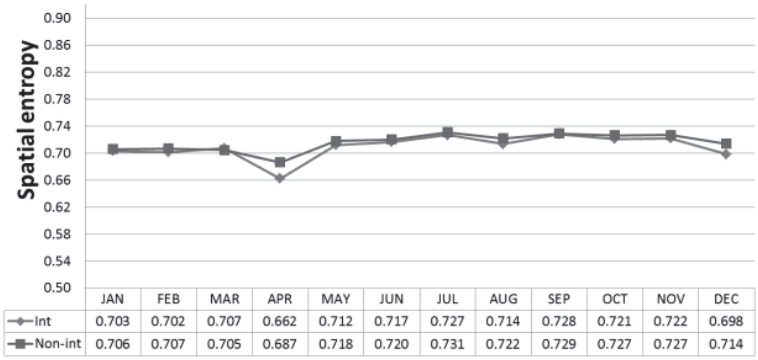
where N_i is the number of detections for the i^{th} month (i =1, 2, 3, … , 12); the *min*() function is evaluated across all months.

There are two possible approaches for implementing spatial entropy calculations in MATLAB®. One possibility, called Option 1, is to convert the data into an integer set *a priori*, and then map the resulting integer values onto a grid. This may lead to artificial clustering, if several points' coordinates are rounded to the same integer. This would be the likely scenario, since the latitudes and longitudes of the detections fall into a very limited range. For example, on the West coast values fall between 30–65 degrees North and 110–180 degrees West, resulting in only approximately 2400 possible integer combinations. For thousands of data points in the S-AIS set this could represent a serious problem. To mitigate this problem, the data set was stretched over a greater range of values using an affine transformation:

$$x' = ax + b \quad \text{and} \quad y' = cy + d \qquad (8)$$

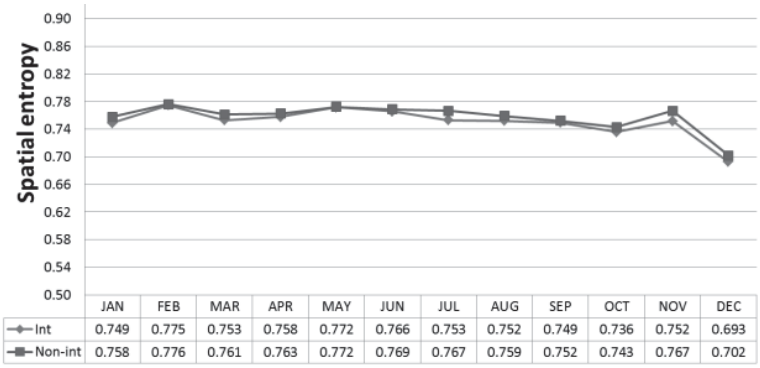
where a, b, c, and d are some pre-determined parameters. This approach is generally less expensive computationally.

An alternative (designated here as Option 2) would be to use a non-integer grid, using the actual values for points' coordinates and testing whether the actual values of the points' coordinates fell inside a grid cell. This approach, while more accurate and robust, is computationally more intensive since it requires testing each of the cells for the presence of points. Figure 3 and Figure 4 show a comparison of spatial entropy for RS2 data calculated via both methods for the East and West coasts (respectively). From the similarity between the two plotted lines for both coasts, it is apparent that for the selected average point density of two per cell, the implication of using the integer grid leads to very little error when compared to the use of the non-integer grid. However, this error would increase if the average density varied significantly from two points per cell.



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Int	0.703	0.702	0.707	0.662	0.712	0.717	0.727	0.714	0.728	0.721	0.722	0.698
Non-int	0.706	0.707	0.705	0.687	0.718	0.720	0.731	0.722	0.729	0.727	0.727	0.714

Figure 3. Spatial entropy for RS2 detections for the East coast (integer vs. non-integer representation).



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Int	0.749	0.775	0.753	0.758	0.772	0.766	0.753	0.752	0.749	0.736	0.752	0.693
Non-int	0.758	0.776	0.761	0.763	0.772	0.769	0.767	0.759	0.752	0.743	0.767	0.702

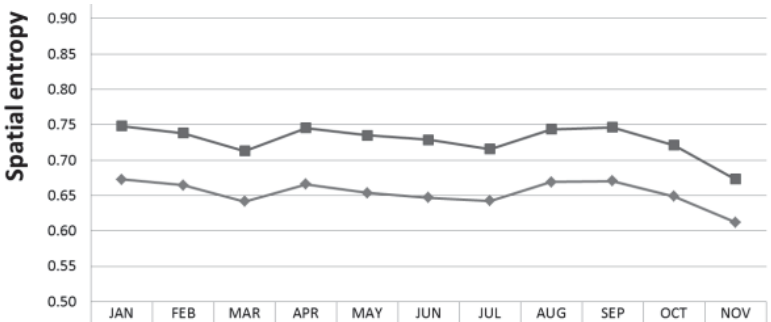
Figure 4. Spatial entropy for RS2 detections for the West coast (integer vs. non-integer representation).

Spatial entropy for the East coast (Figure 3) shows little variability throughout most of the year, with the value holding at approximately $S \cong 0.7$. This suggests a possible presence of some non-random behaviour (patterns) in the detections. There is a noticeable drop in April (indicating an emergence of a pattern), then entropy increases again; the summer values are slightly higher than December-February values. It was earlier proposed that the changes might be due to the start of a fishing season (Dobias et al., 2015). However, more detailed analysis of area subsets is needed to ascertain the nature of the change.

For the West coast (Figure 4), the spatial entropy values remain comparable to those for the East coast ($S \cong 0.7$). However, the variability is smaller, with the exception of a small drop in November (again, possibly due to salmon fishing season). This may be a reflection of slightly less significant weather impacts on the West coast. More detailed analysis looking at subsets of the overall area will be discussed in a following section.

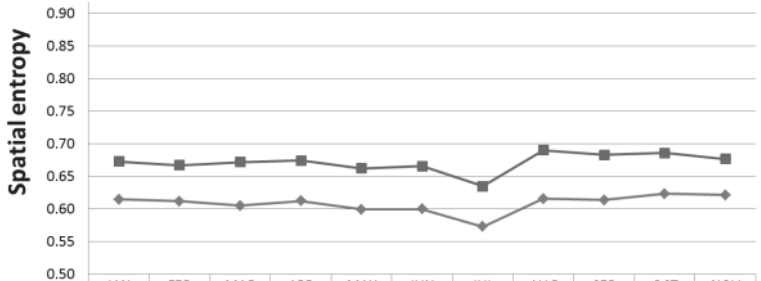
In the work by Dobias et al. (2015), due to the use of a suboptimal (too coarse) grid in CASA, the spatial entropy values were slightly higher (implying more uniform distribution on the used grid). The difference was relatively small for RS2 data, but rather significant for S-AIS data, as shall be discussed below.

Spatial entropy of S-AIS data (Figure 5 and Figure 6, for the East and West coasts, respectively) exhibits a much greater difference between the two computational approaches. While the trend is practically identical between the two options (the use of integer approximation, blue line, and the use actual latitude and longitude values, red line), the use of the integer approximation yielded noticeably lower values of entropy suggesting a greater level of organization. Since there are two orders of magnitude more data points for S-AIS than for RS2, it is possible that, in this case, the approximation forced too many data to particular grid cells despite the fact that the overall point density remained the same. Since the overall trend is the same between the two lines, and in both cases the values are within the expected range (i.e., there is no apparent effect of a grid saturation), due to the ~20% lower computational cost, the integer approximation is still preferred.



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
Int	0.673	0.665	0.641	0.666	0.654	0.647	0.642	0.669	0.670	0.649	0.612
Non-int	0.748	0.738	0.713	0.745	0.735	0.729	0.716	0.744	0.747	0.721	0.673

Figure 5. Spatial entropy for S-AIS detections for the East coast (integer vs. non-integer representation).



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
Int	0.615	0.612	0.605	0.612	0.599	0.600	0.573	0.616	0.614	0.624	0.622
Non-int	0.673	0.667	0.672	0.675	0.663	0.666	0.635	0.691	0.683	0.686	0.677

Figure 6. Spatial entropy for S-AIS detections for the West coast (integer vs. non-integer representation).

The shape of the curves again suggests that there are some changes in the degree of organization in the detections in the course of a year. For the East coast, there is a spring decrease in entropy about a month before a similar change in RS2 data. The difference between the two sets may be due to the difference in target vessels. While fishing vessels may have been driving the RS2 trend, they are not required to use AIS. Thus the S-AIS trend might be driven by a different class of vessels (commercial shipping). Likewise, the West coast AIS pattern differs somewhat from the RS2 one. There is a slight drop in July followed by an increase toward the end of the year. Again, a possible explanation of the discrepancy may be the difference between the targets collected by the two sensors. The analysis of the sub-regions (further below) provided additional insights into this difference in trends.

FRACTAL DIMENSION OF COASTAL DETECTIONS (BROKEN DOWN BY COAST)

As discussed in an earlier section, the fractal dimension is a slightly more robust metric with respect to the subdivision of the analyzed domain into blocks than spatial entropy. This stems from the fact that the calculation of the fractal dimension employs a range of the sizes of a sub-block ε. However, there still are some dependencies of the fractal dimension on the minimum and maximum values of ε. This range can be expressed as a range

of admissible numbers of sub-blocks ($x = L/\varepsilon$). Several options for the coarsest division (i.e., for the x_{min} values) were tested; the differences were negligible, therefore it was decided that, for simplicity, the smallest admissible division of the area would be 2x2 sub-blocks (i.e., $x_{min} = 2$). To facilitate the calculation, the number k of different block sizes (i.e., the number of different values of ε the $N(\varepsilon)$ fit) was kept between 20 and 40; these points were evenly distributed.

Figure 7 shows the dependence of the fractal dimension of a set of 10,000 randomly distributed points on the maximum value of x for various values of x_{min}. The solid lines represent the fractal dimension for given interval $[x_{min}, x_{max}]$. The secondary axis shows the percentage of the grid cells with no detection point in them (dashed line).

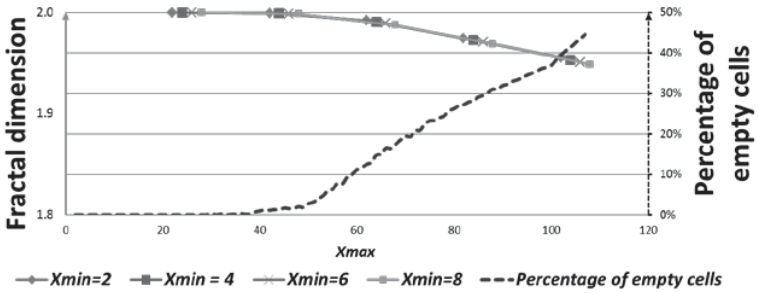


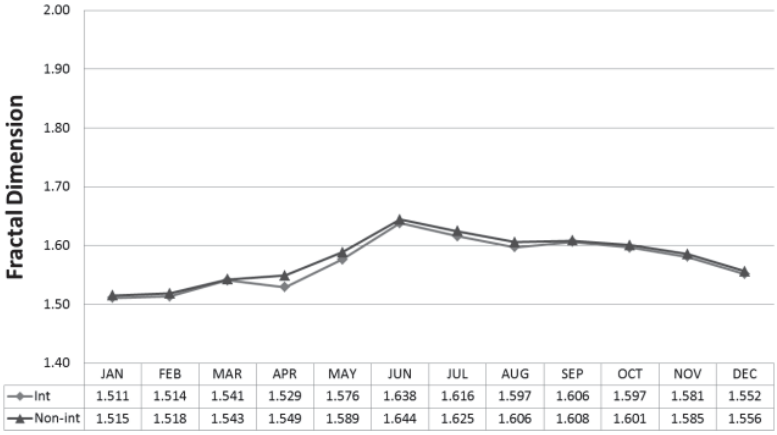
Figure 7. Dependence of D_F of randomly distributed data on the number of sub-blocks.

Both too coarse and too fine division would lead to an appearance of uniformity; in the former case, (almost) each cell would contain some points, thus $D_F \approx 2$, in the latter case the space would be (almost uniformly) empty and thus the $D_F \approx 0$. It was decided to use x_{max} corresponding to the average data density of one data point per sub-block which meant that approximately 50% of the cells would not be empty. This is somewhat different from the criterion employed for spatial entropy; the difference can be attributed to the different information content of each quantity, as discussed earlier. Thus, for the analyzed S-AIS and RS2 datasets the value of x_{max} was calculated as:

$$x_{\max} = \min(\sqrt{N_i}) \qquad (9)$$

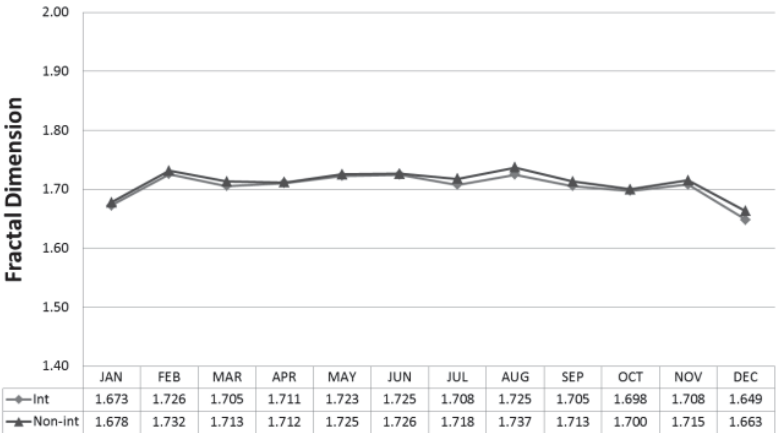
where N_i is the number of detections in the i^{th} month; $i = 1, 2, 3, \ldots, 12$.

Again, there were the same two possible computational approaches: one employing the integer representation of the detection data using affine transformation to expand the real domain to a wider range of integers (Equation (8)), and the second option using the actual (non-integer) values for latitude and longitude. Figure 8 and Figure 9 show the comparison of the two representations for the East and West coasts, respectively. It is apparent that, like for spatial entropy, for RS2 data the difference is negligible.



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Int	1.511	1.514	1.541	1.529	1.576	1.638	1.616	1.597	1.606	1.597	1.581	1.552
Non-int	1.515	1.518	1.543	1.549	1.589	1.644	1.625	1.606	1.608	1.601	1.585	1.556

Figure 8. Fractal dimension for RS2 detections for the East coast (integer vs. non-integer approach).



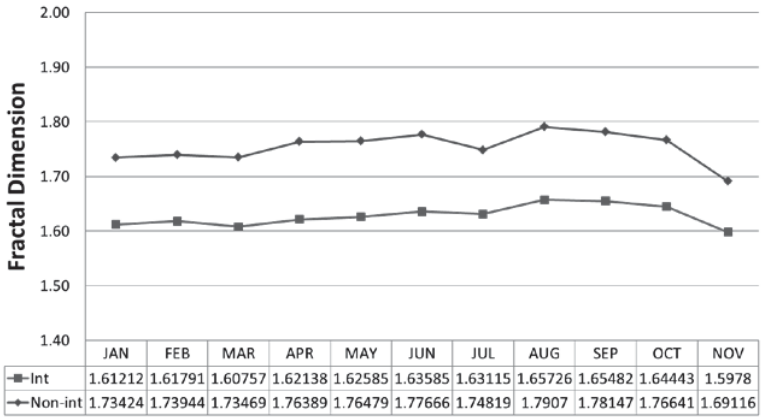
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Int	1.673	1.726	1.705	1.711	1.723	1.725	1.708	1.725	1.705	1.698	1.708	1.649
Non-int	1.678	1.732	1.713	1.712	1.725	1.726	1.718	1.737	1.713	1.700	1.715	1.663

Figure 9. Fractal dimension for RS2 detections for the West coast (integer vs. non-integer approach).

In the work by Dobias et al. (2015), the data saturation – driven by the implementation of the fractal dimension in CASA – led to the values of for S-AIS detections to converge close to the value of two, suggesting a near-uniform distribution. This contradicted visual observation of the data structure (as seen in Figure 1). The adjusted implementation described herein, led to improved results. The values of D_F for the East coast detections ranged between 1.5 and 1.6; the values for the West coast detections were slightly higher, around 1.7. These values of the fractal dimension for S-AIS correspond to the ones obtained for the RS2 data ($D_F \approx$ 1.6 for both the East and West coast). A contributing factor for lower values on the East coast may be more complex coastal topography of that coast. Along the East coast, there is a noticeable increase in the D_F values during summer months. This might be attributable to higher shipping density in parts of the region during summer months. The low variability of D_F for the West coast detections is consistent with the findings for spatial entropy.

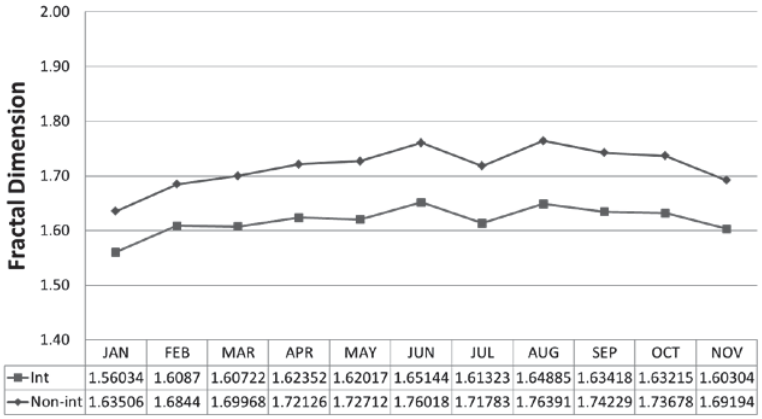
Figure 10 and Figure 11 show a comparison of results for the two implementation options (the integer and non-integer representations of the S-AIS data) for the East and West coasts, respectively. Similarly to the

results for spatial entropy, the integer representation led to a decrease in D_F values, but the overall trend remained the same. As for spatial entropy, the use of the integer representation to calculate the fractal dimension yields approximately a 15-20% savings in computation time.



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
Int	1.61212	1.61791	1.60757	1.62138	1.62585	1.63585	1.63115	1.65726	1.65482	1.64443	1.5978
Non-int	1.73424	1.73944	1.73469	1.76389	1.76479	1.77666	1.74819	1.7907	1.78147	1.76641	1.69116

Figure 10. Fractal dimension for S-AIS detections for the East coast (integer vs. non-integer approach).



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
Int	1.56034	1.6087	1.60722	1.62352	1.62017	1.65144	1.61323	1.64885	1.63418	1.63215	1.60304
Non-int	1.63506	1.6844	1.69968	1.72126	1.72712	1.76018	1.71783	1.76391	1.74229	1.73678	1.69194

Figure 11. Fractal dimension for S-AIS detections for the West coast (integer vs. non-integer approach).

Interestingly, the S-AIS data show much smaller variability on the East coast than they do on the West coast. On the East coast there is a slight increase over the summer months, likely corresponding to increased shipping activity in the northern Atlantic. Along the West coast, there is a similar increase; in both cases D_F exhibits a slight drop in July. In order to better ascertain the causes for this trend, deeper analysis of sub-regions was conducted and is discussed in following sections.

SUB-REGION ANALYSIS OF SPATIAL ENTROPY AND FRACTAL DIMENSION OF COASTAL DETECTIONS

For each coast, the map was split into several sub-regions, and the spatial entropy and the fractal dimension of those sub-regions were calculated. The

West coast was divided into two sub-regions and the east coast was divided into three sub-regions (Figure 12). The fractal dimension and spatial entropy were implemented individually for each sub-region following the same approach to determining box sizes as discussed earlier.

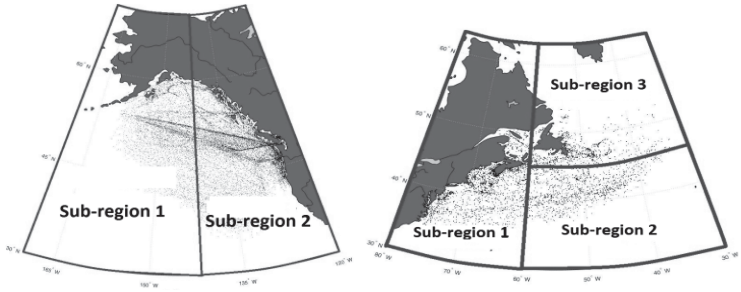


Figure 12. Sub-regions for the East coast (right) and the West coast (left).

Figures 13 and 14 show spatial entropy for the sub-regions and overall for both East and West coast detections, respectively. As is already mentioned earlier, spatial entropy for the East coast shows a noticeable drop in April (Figure 13, purple x's). The same trend appears in sub-regions 1 and 3 (near the coast-line). In contrast, the values for sub-region 2 (open ocean) is more flat. This supports the previous assertion that the changes may be due to a commencement of the fishing season leading to an occurrence of non-random clusters in the area (Dobias et al., 2015). Later in the year the fishing traffic would be supplemented by pleasure crafts leading to a decrease in the organization (an increase in entropy). For the West coast overall (Figure 14, green triangles) spatial entropy shows a slight increase in February followed by a slight drop in March. Overall, the trend is flat. Interestingly, this is mostly consistent with sub-region 2 (closer to the coast), while sub-region 1 (further from the shore line) shows an almost opposite trend. There is no readily apparent explanation for this behaviour.

The opposing trend between the sub-regions for the West coast highlights an important fact. The relationship between the spatial entropy of sub-regions and that for the entire region is not linear. For instance, for the West coast (Figure 14), spatial entropy for both sub-regions increased (indicating a more uniform distribution) between September and October, but the value for entire region decreased (thus indicating less uniform distribution). This may have been caused by a shift in relative distribution of detections between the two sub-regions. In practical terms it suggests that a single sub-dividing of the region might not be sufficient to gain comprehensive insights into the system dynamics, and it may be necessary to use multiple different sub-divisions to gain further insights.

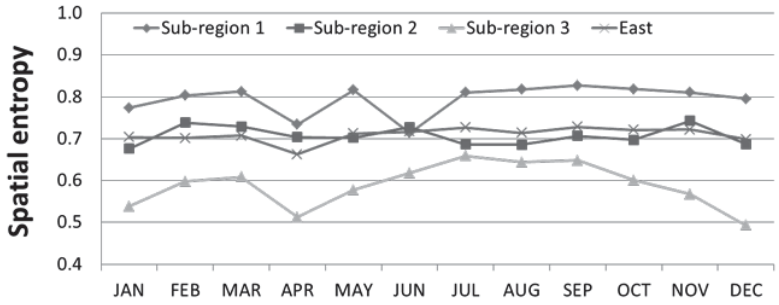


Figure 13. Spatial entropy in RS2 data for the East coast and its sub-regions.

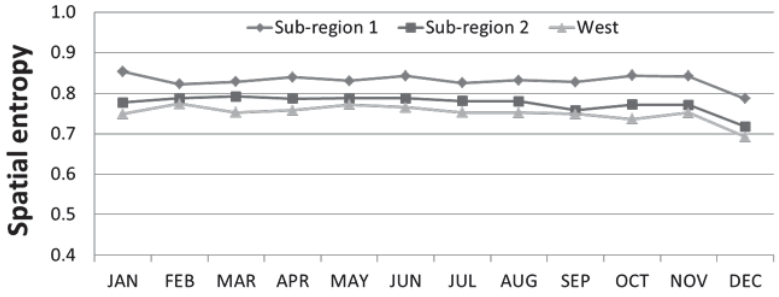


Figure 14. Spatial entropy in RS2 data for the West coast and its sub-regions.

The analysis of spatial entropy in S-AIS data (Figures 15 and 16) further confirms the non-linear nature of spatial entropy relative to sub-regions. For the East coast, sub-region 2 (outer, southern area) seems to have a largely stabilizing effect, showing very little variability during the course of the year. Sub-regions 1 and 3 have opposing trends in the spring, while they become aligned in the late summer and in the fall. This might be caused by the significant impact of weather (e.g., melting ice) that would open northern Atlantic to shipping in the spring. For the West coast, the outer area (sub-region 1) appears to be the primary driver of the overall spatial entropy trend. This is the opposite from the observations made for the RS2 data, and thus it reinforces the claim made earlier that the difference between the S-AIS and RS2 trends is possibly caused by the difference in target vessels for each of these sensors.

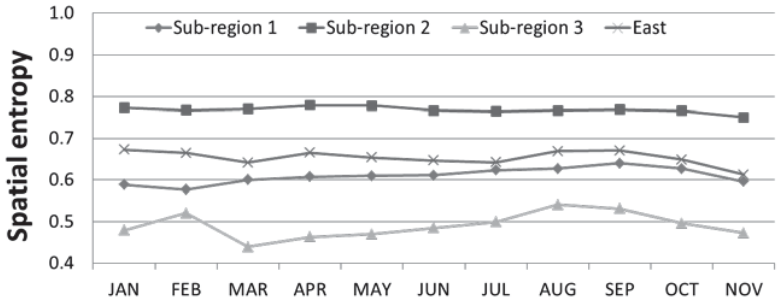


Figure 15. Spatial entropy in S-AIS data for the East coast and its sub-regions.

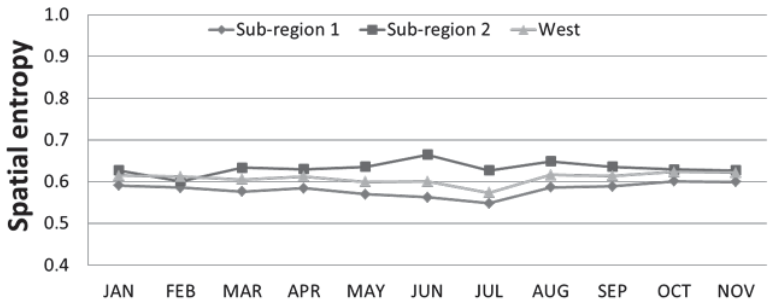


Figure 16. Spatial entropy in S-AIS data for the West coast and its sub-regions.

In the previous section, it was noted that there was a noticeable increase in the D_F for the East coast during summer months while the West coast variability was much smaller. Figures 17 and 18 show the fractal dimension of RS2 detections for sub-regions and overall for both coasts. Again, the relationship between the sub-regions and the coasts overall seems to be non-linear. The significant peaks in the sub-region 1 of the East coast (Figure 17, blue diamonds) might suggest that the main drivers for this area would be fishing and recreational shipping. In contrast, the trend in the sub-region 3 (increase in summer) is probably related to the climatic effects (ice, hurricane season).

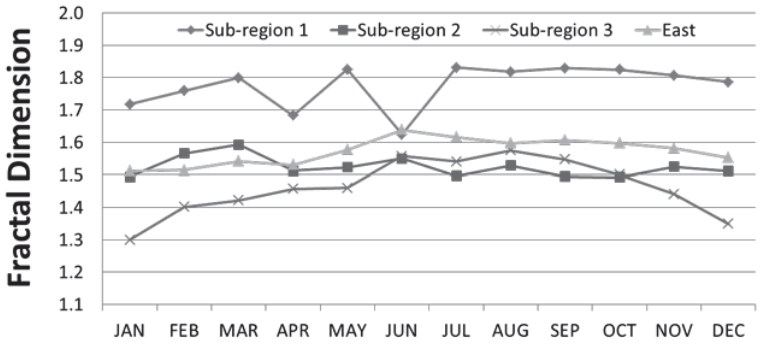


Figure 17. Fractal dimension of RS2 data for the East coast and its sub-regions.

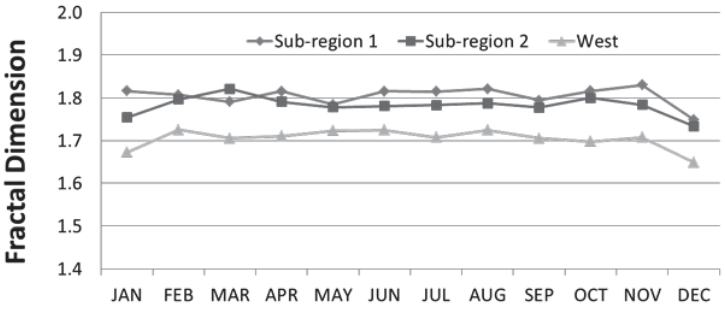


Figure 18. Fractal dimension of RS2 data for the West coast and its sub-regions.

For the West coast (Figure 18), there are no such pronounced variations for either of the two sub-regions. For this coast, the overall trend seems to be

driven by the relative shifts in the shipping density between the inner and outer area.

The analysis of the S-AIS data for the sub-regions confirms earlier observations. For the East coast (Figure 19) sub-region 3 has the most varied behaviour; the other sub-regions are flat, which is consistent with the overall behaviour. Interestingly, for the West coast outer region (sub-region 1) there is a steady increase for the summer and fall likely corresponding to an overall increase in shipping traffic once the storm season in the North Pacific ends. On the other hand, the overall trend is almost entirely driven by the inner sub-region suggesting that the variability might be due to various fishing seasons along the coast.

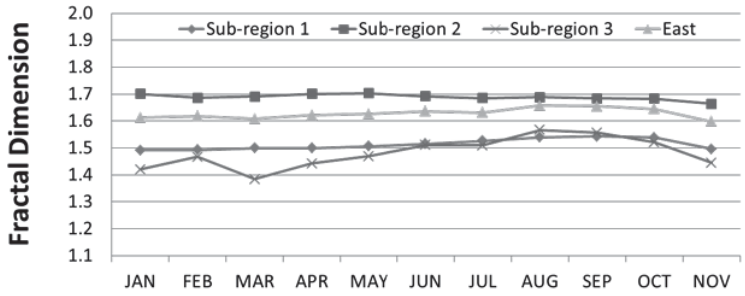


Figure 19. Fractal dimension of S-AIS data for the East coast and its sub-regions.

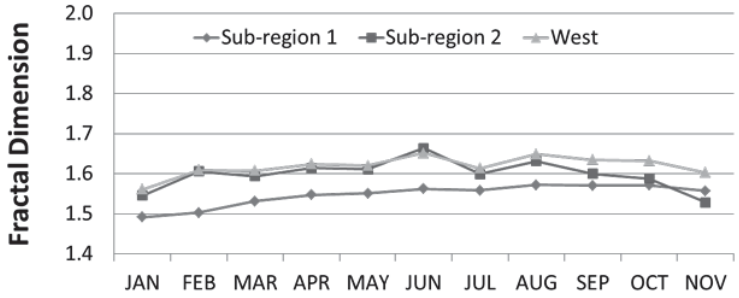


Figure 20. Fractal dimension of S-AIS data for the West coast and its sub-regions.

SUMMARY AND CONCLUSIONS

Spatial entropy and the fractal dimension were used to analyze spatial trends in maritime detection data obtained from RS2 and S-AIS. This work addressed implementation issues identified in earlier work and it also extended previous analysis by exploring trends for a set of sub-regions to better understand the driving factors behind the overall coastal trends.

The following four general observations have been made; the first two concern the implementation of the measures, while the other two are about the trends and behaviours exhibited by the detections:

- The size of the grid blocks employed to calculate both spatial entropy and the fractal dimension need to be related to the detection density;

- It is possible to employ an integer approximation of the latitude and longitude to reduce computational demand; doing so might under-estimate the resulting values, but the trends will still be reliable;
- There is a non-linear relationship between the sub-regions and the coastal areas as a whole. This implies that multiple different subdivisions might be required in order to sufficiently capture system dynamics; and,
- Due to the different nature of target vessels for RS2 and S-AIS, it may not be possible to relate the trends in the two types of detections.

The key focus of the analysis at this stage is the temporal (month-to-month) change in the patterns. These changes were quite consistent for varying grid size thus showing that they were not simply a relic of changing detection numbers. As a secondary benefit, it was demonstrated that using integer representation of the data, while giving different absolute value of spatial entropy and the fractal dimension, yielded temporal variation in the trends consistent with the more accurate method using latitude/longitude data directly. The advantage of the integer approximation was approximately 20% more efficient computation; the latter was a significant consideration for AIS data due to their volume.

Overall, the current results appear promising; the next possible step in this research is to look at multi-year data, comparing trends for corresponding months, followed by the analysis of month-to-month variations across multiple years. Additional follow-up steps might include removing zeroth-order pattern (e.g., main shipping routes) and analyzing properties of the background detection distribution. Possible future work could also include more in-depth analysis of sub-regions. For instance smaller sub-regions could be used in different combinations. In addition, the ship detections could be categorized by vessel type or size (if the information is available). Lastly, the trends can be combined with detections from other sensors as well as with the contextual information from other sources (historical weather and ice patterns, merchant shipping trends and routes, fishing regulations, law enforcement, etc.).

Additional value of the methodology, apart from the temporal analysis of global patterns as has been done in this paper, might be in the comparison of large subsets of the detections (e.g., detections corresponding to a particular vessel type) with each other and with the overall detection distribution; the methodology is likely to be of limited applicability to the assessment of individual detections. However, understanding of larger subsets from different sensors may for instance provide important indications with respect to whether particular unresolved detections from certain sensors are likely to be real or false positives (i.e., the sensors showing detections which do not correspond to real vessels). In addition, by combining these measures for multiple sensors with contextual information, it may be possible to develop maps of normal behaviour that will enable rapid estimate of potential aberrations. This will be a subject of further research.

REFERENCES

Dobias, P., Horn, S., Liu, M.J., Eisler, C., Sprague, K.B. (2015). Use of fractal-based approaches in the assessment of the Canadian recognized maritime picture, in *Proceedings from 32nd International Symposium on Military Operational Research*, Royal Holloway, Egham, Surrey, United Kingdom.

(exactEarth, 2014) exactEarth (2014). "exactEarth receives $19.2M AIS data contract from the Government of Canada" (online), retrieved from http:// www.exactearth.com/media-centre/recent-news/216-canadian-govt-pr-sept-2014 (Access Date: 26 May 2015).

Ilachinski, A. (2004). *Artificial War: Multiagent-Based Simulation of Combat*, Singapore: World Scientific.

Sarkar N., Chaudhuri, B.B. (1994). An efficient differential box-counting approach to compute fractal dimension of image, *IEEE Transanctions on Systems, Man and Cybernetics*, 24, pp 115-120.

Sprague, K.B., Dobias, P. (2008). Modeling the Complexity of Combat in the Context of C2, *The International C2 Journal*, 2(2).

(Vachon et al., 2014) Vachon, P.W., Kabatoff, C., Quinn, R. (2014). Operational ship detection in Canada using RADARSAT, *Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE International, IEEE.

(SOLAS) *Safety of Life at Sea (SOLAS) Convention,* Chapter V, Regulation 19.