

Decision support for threat detection in maritime surveillance



Jacques du Toit

Dissertation presented for the degree of
Doctor of Philosophy
in the Faculty of Science at Stellenbosch University

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 1, 2014

Abstract

The policing and monitoring of South Africa's coastline and *economic exclusion zone* is made difficult not only by the size of the area of interest, but also by the limited resources available for maritime detection and policing. As a consequence, illegal activities, such as smuggling, poaching and illegal border crossings, are often conducted with impunity. Conventional approaches to monitoring coastal areas, such as the use of patrol boats, port inspections and aircraft surveillance, may be augmented by advances in technology that are steadily contributing vast amounts of data related to maritime activity. For example, various South African agencies collect *automatic identification system* and *vessel monitoring system* transmissions, and gather additional kinematic data of maritime vessels through a number of strategically placed coastal radars. A command and control centre for actively monitoring these data (outside of the intelligence community) was established by the South African Navy in 2014.

Such centres provide surveillance operators with a real-time picture of a maritime region of interest from which they can identify relevant facts of interest through a reliance on experience and domain knowledge. The effectiveness of this process may, however, be undermined by the vast quantities of data typically under consideration, by the difficulty of identifying long-term trends in vessel kinematic behaviour and by the possibility of operator fatigue brought on by the relatively low incidence levels of activities of interest.

Effective decision support tools may play a valuable role in this context by the automatic processing of these vast collections of data, by the identification of concepts of interest and by the prediction of future occurrences of interest. It is, however, essential that such tools should be flexible enough to adapt to changes in typical vessel behaviour over time and that they should be capable of integrating new trends and new types of behaviours.

Various approaches to maritime surveillance are investigated in this dissertation from the perspectives of threat detection and anomaly identification, with particular emphasis on a systems approach to decision support. A decision support system framework that utilises rule-based and data-driven mechanisms is proposed as a means to separate the interesting from the uninteresting and to provide early warnings of potentially threatening maritime vessel behaviour to operators. This system framework is primarily concerned with kinematic data and is restricted to the identification of certain types of activities. Successful classification and, ultimately, timely prediction of potentially threatening behaviour would allow for effective policing by providing early warning to relevant entities, thus potentially leading to more effective use of available policing resources.

Uittreksel

Die patrolling en monitering van die Suid-Afrikaanse kusgebied en gepaardgaande ekonomiese eksklusiewe zone word bemoelijk deur die grootte van die tersprake area en die beperkte hulpbronne wat vir patrollie-doeleindes aangewend kan word. Gevolglik gaan onwettige aktiwiteite, soos smokkelary, stroping en onwettige immigrasie dikwels ongestraf. Konvensionele benaderings tot die monitering van kusgebiede, soos die aanwending van patrolliebote, die uitvoer van hawe-inspeksies en gereelde lugpatrollies, kan aangevul word deur tegnologiese vooruitgang wat voortdurend tot groot hoeveelhede data oor maritieme aktiwiteit bydra. Verskeie Suid-Afrikaanse agentskappe ontvang byvoorbeeld *outomatiese identifikasiestelsel* en *vaartuigmoniteringstelsel* uitsendings, en samel ook addisionele kinematiese data oor maritieme vaartuie deur middel van strategies-geplaaste kusradars in. 'n Bevel-en-beheersentrum wat hierdie inligting (buite die intelligensiegemeenskap) aktief ontleed, is in 2014 deur die Suid-Afrikaanse Vloot tot stand gebring.

Sulke sentra verskaf 'n intydse blik oor die maritieme gebied onder beskouing aan operateurs wat dan, gebaseer op hulle ervaring en omgewingskennis, relevante inligting oor vaartuie kan aflei. Die doeltreffende uitvoering van hierdie proses kan egter ondermyn word deur die tipiese groot hoeveelhede data, die moeilikheidsgraad van die identifikasie van langtermyn tendense in die kinematiese gedrag van vaartuie om die kus en die moontlikheid van operateur-uitputting as gevolg van lang periodes van relatiewe oninteressante vaartuiggedrag.

Doeltreffende besluitsteunhulpmiddels kan 'n waardevolle bydrae in hierdie konteks maak deur die ge-outomatiseerde prosessering van hierdie groot hoeveelhede data, die identifikasie van interessante vaartuiggedrag en die voorspelling van toekomstige relevante insidente. Dit is egter noodsaaklik dat sulke hulpmiddels buigsaam genoeg moet wees om te kan aanpas by veranderings in tipiese maritieme aktiwiteit oor tyd en dat nuwe tendense en tipes aktiwiteite geakkommodeer kan word.

Verskeie benaderings tot maritieme oorsig word in hierdie proefskrif vanuit die perspektiewe van die bespeuring van bedreigings en die opsporing van vreemde verskynsels ondersoek, met 'n spesifieke fokus op 'n stelselbenadering tot besluitsteun. 'n Besluitsteun stelselraamwerk wat berus op reël-gebaseerde en data-aangedrewe meganismes word as 'n hulpmiddel voorgestel waarmee interessante maritieme gedrag van oninteressante gedrag onderskei kan word om sodoende 'n vroeë waarskuwing aan operateurs met betrekking tot moontlike bedreigende maritieme aktiwiteite te kan rig. Die werking van hierdie stelselraamwerk berus hoofsaaklik op die gebruik van kinematiese vaartuigdata en is beperk tot die naspeuring van sekere soorte bedreigende gedrag. Die suksesvolle klassifikasie en tydige voorspelling van potensiële bedreigende maritieme gedrag behoort doeltreffende kusmonitering en verbeterde aanwending van die beperkte, gepaardgaande hulpbronne deur relevante kusagentskappe moontlik te maak.

Acknowledgements

The author wishes to acknowledge a number of people and institutions for their various contributions towards the completion of this work.

- Prof JH van Vuuren for the extraordinary support and for the incredible personal investment he makes in all students.
- Mr Jacques van Wyk for securing vessel data and to himself and Mr Leon Downes for being available for questions and for providing invaluable information and feedback.
- The ARMSCOR LEDGER programme, the Harry Crossley Foundation and Postgraduate Student Funding Office of Stellenbosch University for the financial assistance they provided.
- The ARMSCOR LEDGER programme and the CSIR who provided a forum for presenting and discussing this work.
- The Operations Research Society of South Africa which also provided a forum and critical feedback for this study.
- The Maties Exchange Programme for the opportunity afforded to me to study necessary courses at a foreign institution.
- The Operations Research Division of the Department of Industrial Engineering at Stellenbosch University for the wonderful research facilities that they provided.
- My colleagues in the Operations Research Division who all made contributions to the completion of this work in some way.

Table of Contents

List of Acronyms	xiii
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Background	1
1.1.1 The law of the seas	1
1.1.2 Law enforcement principles	3
1.1.3 Crimes committed at sea	4
1.2 South African maritime responsibilities	5
1.2.1 Fisheries Management	6
1.2.2 Search and Rescue	7
1.2.3 Monitoring and Enforcement	7
1.3 Problem identification	10
1.4 The thesis of this dissertation	11
1.5 Dissertation scope	12
1.6 Dissertation objectives	12
1.7 Dissertation organisation	12
2 The analysis and learning of motion patterns	15
2.1 Introduction	15
2.2 Prevailing paradigms	16
2.3 Motion patterns	18
2.3.1 Computer vision	20
2.3.2 Ecological modelling	24
2.3.3 Maritime motion patterns and threat assessment	26

2.4	Summary	33
3	A surveillance system framework in the maritime domain	35
3.1	Decision support systems and data fusion	35
3.2	Maritime surveillance	37
3.2.1	The data	39
3.2.2	Extracting meaning from the surveillance scene	40
3.2.3	The operator	40
3.2.4	Vessel designations	41
3.2.4.1	The notion of normality	41
3.2.4.2	Vessels of interest	42
3.2.4.3	Threats	42
3.3	Existing maritime surveillance DSSs	43
3.4	A newly proposed maritime surveillance DSS	44
3.4.1	The fusion component	47
3.4.2	The rule-based component	48
3.4.3	The discovery component	51
3.4.4	The activity classifier	53
3.4.5	The human machine interface	53
3.5	Summary	53
4	Rule-based system component	55
4.1	Constraints to movement and activity	55
4.2	Restrictions pertaining to small vessels	56
4.3	Restrictions pertaining to seafaring vessels	59
4.4	Rule classes	60
4.4.1	Zone infractions	63
4.4.2	Proximity alarms	67
4.4.3	Anomalous actions alarms	74
4.5	Summary	75
5	Knowledge discovery model component	77
5.1	An approach to clustering time-series	77
5.1.1	Dynamic time warping	78
5.1.2	Partitioning around medoids	84
5.2	Discovery component: A filtering approach	90

5.2.1	Density-based clustering for region discovery	91
5.2.2	Outlier removal	97
5.3	Summary	97
6	Activity classifier model component	101
6.1	Introduction	101
6.2	Hidden Markov models	101
6.3	An approach to trajectory classification	110
6.4	Summary	117
7	The Port of Cape Town: A case study	119
7.1	The data	119
7.2	Application of the origin-destination miner	125
7.3	Application of an activity classifier	129
7.4	Summary	133
8	Conclusion	135
8.1	Dissertation summary	135
8.2	An appraisal of the dissertation contributions	136
8.3	Suggestions for future work	137
A	Appendix	141
	References	143

List of Acronyms

ADSS: Adaptive Decision Support System

AIC: Akaike Information Criterion

AIS: Automatic Identification System

ASL: Australian Sign Language data set

BIC: Bayesian Information Criterion

COLREG: International Regulations for Preventing Collisions at Sea

CPA: Closest Point of Approach

CSWR: Centre for Sea Watch and Response

DAFF: Department of Agriculture, Fisheries and Forestry

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

DTW: Dynamic Time Warping

DSS: Decision Support System

DW: Data Warehouse

EEZ: Exclusive Economic Zone

FN: False Negative

FP: False Positive

GMM: Gaussian Mixture Model

GPS: Global Positioning System

HMI: Human Machine Interface

HMM: Hidden Markov Model

IMO: International Maritime Organisation

ISPS: International Ship and Port Facility Security Code

KB: Knowledge Base

KDD: Knowledge Discovery in Databases

KDE: Kernel Density Estimation

LRIT: Long-range Identification and Tracking of Ships

MDA: Maritime Domain Awareness

MMSI: Maritime Mobile Service Identity

MPA: Marine Protection Area

MRCC: Marine Rescue Co-ordination Centre

NSRI: National Sea Rescue Institute

PAM: Partitioning Around Medoids

PCA: Principal Component Analysis

SADC: Southern African Development Community

SAMSA: South African Maritime Safety Authority

SAR: Search and Rescue

SOLAS: Safety of Life at Sea

SVM: Support Vector Machine

TN: True Negative

TP: True Positive

TAC: Total Allowable Catch

TAE: Total Allowable Effort

UN: United Nations

UNCLOS: United Nations Convention on the Law of the Sea

VMS: Vessel Monitoring System

VOI: Vessel of Interest

List of Figures

1.1	South Africa’s search and rescue region of responsibility	8
1.2	South African research vessels	9
1.3	South African patrol vessels	10
1.4	The data fusion inference hierarchy	11
2.1	A simplified overview of an expert system	16
2.2	The basic pattern classification process	17
2.3	Feature selection and feature extraction	18
2.4	Generic movements patterns	20
2.5	A selection of events identified by operators which should elicit early warnings . .	32
3.1	The JDL fusion model	37
3.2	The functional processes of an integrated data mining and fusion approach	38
3.3	A simple generalized flow diagram of the surveillance process	38
3.4	Vessel designations within a maritime surveillance environment	42
3.5	Activities that may contribute to a vessel being designated as a threat	43
3.6	Methodologies employed in the newly proposed decision support system	45
3.7	An overview of the proposed decision support system	46
3.8	The track fusion component	47
3.9	Two examples of rules that may be used to generate alarms	48
3.10	A rule which may identify an occurrence of illegal fishing	49
3.11	The basic elements of a multi-agent system	50
3.12	The rule-based system component	51
3.13	The discovery component	51
3.14	The activity classifier component	53
3.15	A detailed overview of the proposed decision support system	54
4.1	Marine protection areas along the Western Cape coast	57

4.2	Marine protection areas along the Eastern Cape coast	57
4.3	Marine protection areas around Cape Point in the Western Cape	58
4.4	A decision tree describing restrictions that apply to small vessels	59
4.5	Reduced order binary decision diagram of the implication operator	62
4.6	Binary decision diagrams of zone violation and proximity rules	64
4.7	Polygonal region containment tests by way of an example	65
4.8	Regions corresponding to a navigational warning and territorial waters	66
4.9	Regions corresponding to an MPA and prohibited anchoring areas	67
4.10	Convergence patterns that may indicate a collision or a rendezvous	69
4.11	Avoidance actions for head-on and crossing collisions	70
4.12	The closest point of approach and evasive actions	72
4.13	Rendezvous and pursuit activities by way of an example	73
4.14	Examples of raid and deploy activities	74
5.1	A comparison of Euclidean and DTW time-series matching	79
5.2	DTW computed for series of different temporal characteristics	80
5.3	DTW computed for two series differing only in their temporal samples	81
5.4	A degenerate DTW alignment	82
5.5	Three global window constraints for DTW	82
5.6	A DTW alignment using a symmetric step pattern	83
5.7	A step pattern allowing points to be unmatched	83
5.8	An alignment in \mathbb{R}^2	84
5.9	A simple application of the k -means and k -medoids methods	85
5.10	A dissimilarity matrix computed using the Manhattan distance	87
5.11	Clusters determined using PAM and their accompanying silhouette plots	89
5.12	An example of a DTW dissimilarity matrix	89
5.13	A clustering of regular and irregular time-series	90
5.14	A DBSCAN clustering performed on a simple example	92
5.15	The operation of DBSCAN illustrated on two example data sets	94
5.16	The DP and radial distance method explained by way of an example	95
5.17	DBSCAN applied to thinned trajectories	96
5.18	Filtering according to origin-destination pairs and sinuosity	96
5.19	An overview of the origin-destination miner	98
6.1	The graphical mode and finite-state automatum representation of a Markov chain	103
6.2	The graphical model representation of a hidden Markov model	103

6.3	A two-state HMM model describing the use of a fair or biased coin	105
6.4	A three-state HMM model featuring bivariate Gaussian emissions	106
6.5	The generating states of the fair and biased coin HMM	107
6.6	The lattice structures that result when computing forward probabilities	109
6.7	Sequence classification using two HMMs	110
6.8	Long sequence classification using two HMMs	111
6.9	Estimating the emission distribution parameters for $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in \mathbb{R}^2	112
6.10	The AIC and BIC computed for $tr(\mathcal{U})$, $tr(\mathcal{S})$ and $tr(\mathcal{D})$	115
7.1	The time intervals at which updates occur within the case study data set	121
7.2	The frequency of speed reports collected during a five-month period	122
7.3	Preprocessing of case study data set	123
7.4	A comparison of the reported and derived headings of two vessels	124
7.5	A comparison of raw and cleaned data for October and November, 2010	126
7.6	A cropped snapshot of AIS position reports received during October 2010	127
7.7	A magnification of a region near the Port of Cape Town	127
7.8	A reduced set of tracks within a radial distance of 9 km of the Port of Cape Town	128
7.9	DP and DBSCAN applied to resampled AIS data	128
7.10	Filtering tracks with respect to sinuosity outliers	129
7.11	The medoids derived from PAM	130
7.12	The contour plots of the state-dependant distributions centred on their means . .	131
7.13	The misclassifications on the entry lane	133
7.14	A vessel deviating from a lane	133
8.1	A conceptual view of threat responsibility in South Africa	139

List of Tables

2.1	Pattern recognition applications	17
4.1	The seaward operating ranges of small vessels, as determined by their category .	56
4.2	Propositional operators	61
4.3	Truth tables for the implication operator	61
4.4	General zone violation rules	63
4.5	Specific zone violation rules	66
4.6	The general descriptions of the proximity rules	68
4.7	A summary of proximity alarms	69
4.8	Examples of Anomalous actions	74
6.1	AIC and BIC computed for a GMM	113
6.2	AIC and BIC computed for two HMMs	114
7.1	An example of AIS reports	120
7.2	AIS reports converted to SI units	120
7.3	Characteristics of AIS data observed over a five-month period	121
7.4	The reduction in tracks or updates as a result of data processing	123
7.5	An example of clustering visitation sequences	128
7.6	A frequency count of visitation sequences	129
7.7	The number of tracks in the clustered and unclustered sets	132
7.8	HMM classification results on a reduced data set	132
8.1	A comparison of extracted vessel types with descriptions available online	138

CHAPTER 1

Introduction

Contents

1.1	Background	1
1.1.1	<i>The law of the seas</i>	1
1.1.2	<i>Law enforcement principles</i>	3
1.1.3	<i>Crimes committed at sea</i>	4
1.2	South African maritime responsibilities	5
1.2.1	<i>Fisheries Management</i>	6
1.2.2	<i>Search and Rescue</i>	7
1.2.3	<i>Monitoring and Enforcement</i>	7
1.3	Problem identification	10
1.4	The thesis of this dissertation	11
1.5	Dissertation scope	12
1.6	Dissertation objectives	12
1.7	Dissertation organisation	12

1.1 Background

The rights and responsibilities of sovereign nations are codified in international law and many of the rules and regulations that are followed at sea are derived from these laws and conventions. The foundation of oceanic law provides coastal states with clear guidelines as to which areas are their responsibility, which laws apply to these areas, as well as to the areas beyond their control. Such laws clearly define crimes committed at sea and provide nations with a means to enforce them.

It is instructive to review some of these laws in order to establish a context for the activities of sea-farers and in order to gain an understanding of what is broadly permissible. This context may be further enriched by considering the historical factors that contributed to these laws.

1.1.1 The law of the seas

During the seventeenth century, the oceans of the world were subject to the *freedom of the seas* doctrine which limited the rights of sovereign coastal nations to a narrow band of sea along their

coastlines. This doctrine articulated a notion which still prevails on the *high seas* today, namely that the ocean beyond that claimed by nations is “free to all and belonging to none” [181]. Nations have sought to claim parts of the seas adjacent to their shores with a view to create a buffer zone protecting their coastlines from warships and various other intruders [181]. As nations grew in stature, the *cannon-shot* rule (roughly three *nautical miles*¹) became accepted as a measure of the size of territorial seas. Towards the end of the 1960s, some nations claimed sovereignty over the ocean up to twelve miles out to sea. A few nations resisted this trend as many straits used for international shipping would then fall under territorial claims.

Resolving the issues related to straits was one of the incentives² behind the *Third United Nations Conference on the Law of the Sea* [181]. Conflicting territorial claims, pollution at sea and increasing exploitation of the ocean and seabed, were considered threats to stability. Arvid Pardo, Malta’s ambassador to the UN, had earlier called for “an effective international regime over the seabed and the ocean floor beyond a clearly defined national jurisdiction” [181]. Furthermore, he warned that “it is the only alternative by which we can hope to avoid the escalating tension that will be inevitable if the present situation is allowed to continue.” These events led to the creation of the UN Seabed Committee and ultimately culminated in the Third United Nations Conference on the Law of the Sea [181]. The Conference on the Law of the Sea began in 1973 and resulted in the adoption of the *United Nations Convention on Law of the Sea* (UNCLOS) in 1982. South Africa is a signatory of the convention (it signed the convention on 5 December, 1984) which it later ratified³ on 23 December 1997 [26]. This convention provides a legal framework which defines the rights and responsibilities of nations with respect to, *inter alia*, jurisdiction over waters, navigation and access to seas, exploitation of resources, protection and preservation of the marine environment, and sea-bed mining [26, 181].

UNCLOS addressed a number of questions relating to sovereign territorial waters, economic activities, navigation and overflight⁴ of these waters. Coastal states may exercise sovereign rights over twelve nautical miles of ocean extending from their shores (the shoreline is measured in the general case by the low-tide mean sea level). Within this territorial sea the coastal state enjoys all rights over the water, seabed and airspace. However, the coastal state is obliged to allow innocent passage through these waters⁵. The *contiguous zone* is defined as extending from the territorial seas for another twelve nautical miles (or twenty-four nautical miles from the shore). A state has jurisdiction over this zone insofar as preventing and punishing infringements of local laws which were committed within the nation’s territory or within its territorial waters.

¹A nautical mile is a unit of length approximately corresponding to one minute of arc of latitude on the surface of the earth (as measured along a meridian on a spheroid approximating the earth’s surface). The nautical mile is equivalent to 1 852m [172].

²A gradual departure from the freedom of the seas doctrine was precipitated by the unilateral extension of territorial waters by the United States of America (1945) to the boundary of their continental shelf. Argentina followed in 1946 and claimed jurisdiction over their continental shelf and the waters above it. Chile and Peru (1947) and Ecuador claimed sovereign rights over a 200 mile zone extending from their coastlines [181]. These developments necessitated an international consensus.

³A nation is bound by the terms of a treaty once a treaty is ratified. Signing of a treaty merely requires a state to refrain from acts that would be in contravention of a treaty. The distinction is made as signatories typically have to engage their national legislative structures in order to make the treaty domestically effective. Once the appropriate national legislation is in place the treaty may be ratified [182].

⁴Aircraft enjoy the freedom to navigate the airspace above the high seas (which are typically designated as international waters).

⁵The right of innocent passage is afforded to the vessels of all states. This right allows vessels to traverse the territorial seas continuously (only stopping if in distress or if the vessel assists others in distress). In the event that a vessel calls at a port of the coastal nation or travels to and from internal waters of that state, then it will also enjoy the right of passage through territorial waters. Passage is deemed innocent if “it is not prejudicial to the peace, good order or security of the coastal state” [180, Section 3, Subsection A].

The *exclusive economic zone* (EEZ) is the stretch of ocean adjacent to the territorial sea which extends for at most 200 nautical miles from the shore. The coastal state has sovereign rights of exploration, exploitation and management of natural resources (living or not) of the seabed, the waters superjacent to (overlying) the seabed as well as rights to economic exploitation such as the production of energy from water, currents and wind. Furthermore, the convention gives the coastal nation a right to create artificial installations and structures and to conduct marine research within the EEZ. The protection and preservation of the marine environment is the duty of the coastal state⁶ [180, Article 56]. If enacting the provisions made for the EEZ causes two coastal states to have an overlapping exclusive economic zone (as is the case with many Mediterranean nations), then alternative agreements are reached between the states.

The waters beyond the EEZ which are not considered archipelagic waters are defined as the *high seas*. All nations (whether coastal or land-locked) enjoy the rights (with certain exclusions) of freedom of navigation, overflight, construction of artificial islands, fishing and scientific research [180]. They may not impose sovereign claims on any part of the high seas and these seas are to be used for peaceful purposes. This does not preclude all military activities on the high seas and states seemingly presume a moderate position in this regard in the sense that activities which states consider reasonable and which are conducted without weapons are often tolerated. Limitations are placed on military activities in the waters of coastal or foreign states by treaties such as the *Treaty of Pelindaba* which creates an African Nuclear Weapon Free Zone⁷ [80].

1.1.2 Law enforcement principles

UNCLOS provides a legislative framework which addresses and prevents threats to maritime security through the provisions of the convention and is considered a *maritime security regime*. The *International Maritime Organisation* (IMO)⁸ has formulated numerous conventions that fall into this classification. For example, the *Safety of Life at Sea Convention* (SOLAS) which requires all vessels of 300 tonnes or more to carry an *automatic identification system*⁹ (AIS) transponder onboard, and the *International Ship and Port Facility Security (ISPS) Code*.

A vessel is required to fly the flag of the state in which it is registered and may therefore be considered to be of the flag state's nationality, thus falling under the jurisdiction of that state. This jurisdiction extends to the high seas [80]. A vessel which is stateless or flagless forfeits its right of navigation of the high seas and may be boarded or interdicted by any other vessel. In the event that an authorized vessel (such as a warship) suspects a vessel of flying a false flag, this vessel may interdict or board the suspected vessel. On the high seas the right of passage is of paramount importance and vessels intercepting suspicious vessels must have a reasonable

⁶For instance, it is the duty of the coastal state to determine the total allowable catch within fisheries.

⁷The terms of this treaty state that participating nations may not, *inter alia*, station or test nuclear explosive devices (their own or those of another country) nor dump radioactive waste within their territories [183]. Whereas this treaty applies to territorial and archipelagic waters within the maritime domain, the Southeast Asian Nuclear Weapon Free Zone treaty includes the continental shelves and EEZs of coastal states [80].

⁸The IMO is a UN agency responsible for developing and maintaining international regulations which are to be followed by member states. These include regulations pertaining to safety at sea, standards for ship design and construction, pollution and compensation for those affected by it, conventions relating to the marine environment and security regimes for international shipping. Incidentally, the first incarnation of the International Convention for the Safety of Life at Sea (1914), which was agreed upon before the IMO came into existence and was created in response to the sinking of the Titanic, has been revised repeatedly by the IMO. The IMO has 170 member states of which South Africa is one [71].

⁹An AIS system is a radio frequency technology featuring coastal and satellite receivers which are capable of receiving a signal transmitted from onboard vessel transponders. These transponders also receive signals from similarly equipped nearby vessels, thus forming a better picture of their surroundings (this function may also be performed by an onboard radar).

suspicion as the burden of proof lies with the accusing state. These rights are defined in [180, Article 110] as the *rights of visit* and they are afforded to warships (and other authorized ships or aircraft) that have reasonable grounds for suspecting a vessel of being engaged in piracy, slave trade, drug trafficking, unauthorised broadcasting or if a vessel is without a flag or refuses to show its flag [80].

The greatest means of law enforcement at the disposal of a coastal state is the right of *hot pursuit*. This affords the state the right to pursue a vessel that is deemed to have infringed on laws that hold within the internal and territorial waters of the pursuing state [180, Article 111]. Such a pursuit may continue beyond the territorial sea or contiguous zone if the pursuit occurs uninterrupted. This allows coastal states to exercise the right to protect their interests against vessels that violate their laws [80]. The right of hot pursuit is also applicable to violations of laws that are in force in the contiguous zone, EEZ and on the continental shelf (whereas internal waters and territorial waters are subject to a coastal nation's domestic laws, the EEZ and continental shelf are subject to the regulations of UNCLOS). The pursuing state is required to signal the target vessel, by auditory or visual means, to stop. Should the pursued vessel enter the territorial waters of its own state or that of a third state, then the right of hot pursuit ceases¹⁰.

The boundaries of the territorial sea, contiguous zone, EEZ and continental shelf may be difficult to calculate in practice, but if the pursuing vessel determines, with the means at their disposal, that the infringement took place within a particular zone, then a pursuit is lawful. Considering that the costs involved in such a pursuit are typically high, it is only serious infringements that are pursued. Furthermore, a vessel that is wrongfully arrested or stopped outside of territorial waters has a right to seek compensation [80].

1.1.3 Crimes committed at sea

UNCLOS affords vessels innocent passage through territorial seas, but any actions considered prejudicial to peace or security of the coastal state may be considered crimes by the affected coastal state. Such activities include military-related activities, fishing activities, research or surveying activities, any acts of willful and serious pollution or any activities not congruent with direct passage [180, Article 19]. Coastal states furthermore have their own customs and immigration laws and any violations of these laws are subject to their jurisdiction whilst sea-faring states are given universal jurisdiction in the repression of piracy. UNCLOS defines piracy as illegal acts of violence or detention committed for private ends by the crew or passengers of a private craft [180].

The convention for the *Suppression of Unlawful Acts Against the Safety of Maritime Navigation* provides a clear set of definitions of acts constituting crimes at sea. These include the seizure of vessels, acts of violence against people onboard vessels, the destruction of a vessel or placing devices onboard a vessel which destroy or damage it or any of the cargo it carries (which endangers navigation)¹¹. Damaging navigational facilities and communicating false information

¹⁰Neighbouring states often seek memoranda of understanding with one another to resolve this matter so that pursuing vessels may enter the territorial waters of a neighbouring countries when engaged in hot pursuit. Alternatively, the pursuing vessel must contact the third state and seek permission on a case-by-case basis. This approach will typically render law enforcement ineffective [80].

¹¹Interestingly, sixty five percent of all reported attacks on merchant vessels are committed whilst vessels are anchored in ports (such as theft), whilst attacks against vessels on the high seas accounted for twenty seven percent of attacks in 2003. The reason for these proportions may be that boarding vessels whilst they are underway is more difficult than boarding them whilst anchored, because such an endeavour requires better organisation and equipment [29].

which endangers the safe navigation of a vessel is also considered an offense [19].

1.2 South African maritime responsibilities

South Africa is obligated, under UNCLOS and various other treaties and conventions (such as SOLAS), to observe international guidelines governing the protection of marine resources, perform search and rescue, and execute hydrographic duties¹² within designated regions. The *South African Maritime Zones Act* of 1994 adopted the relevant articles in UNCLOS which defined the territorial waters, contiguous zone, EEZ and continental shelf. This Act also applies to the surrounding waters and continental shelf of the Prince Edward islands, which are situated 540 nautical miles (approximately 1 000 km) south east of Port Elizabeth and are a part of South Africa [161]. The act defines the responsibilities of the Republic succinctly. Any law in force in South Africa is applicable to the territorial waters and airspace above it (including *common law*¹³). The right of innocent passage is necessarily guaranteed. The Republic may also exercise all powers to prevent and punish contraventions of fiscal law, customs, emigration, immigration and sanitary laws. The Republic may exercise the same rights and powers afforded it in territorial waters, within the EEZ with respect to all natural resources [103].

South Africa also plays an active role in regional waters as a member of the *Southern African Development Community*¹⁴ (SADC). In February 2012, a *Memorandum of Understanding* on maritime security cooperation was signed by three member states, namely South Africa, Mozambique and Tanzania [109]. This agreement allows the forces of these countries to patrol, search, arrest, seize and engage in hot-pursuit operations within their combined territorial waters. These joint maritime operations focus on interdicting pirates and preventing illegal activities in these waters so as to ensure security along the east coast of Africa, as well as the free flow of goods along this coastline. Due to the spectre of piracy along the east coast of Africa, the South African Navy has engaged on previous patrols along the Eastern Coast. The South African frigate, SAS Mendi, was patrolling the Mozambique channel at the inception of this agreement [109].

The coast guard function in South Africa is carried out by a number of governmental departments, the South African Navy and various non-governmental organisations. An example of a governmental organisation assuming some of these functions (although its mandate extends beyond a coast guard role) is the *South African Maritime Safety Authority* (SAMSA). It was created in 1998 and is capable of coordinating various maritime resources (such as the *National Sea Rescue Institute* (NSRI)) and charged with performing services on behalf of the government [162, 111]. These include port state control, representing South Africa at international forums, such as the IMO and SADC, certification of sea-farers and safety certification of vessels, safety

¹²South Africa is a member of the *International Hydrographic Organization* (IHO) and is responsible for the provision of navigational charts for an ocean area stretching west from the northern border of Angola, east from the northern border of Mozambique (which includes Madagascar) and south to the coast of Antarctica [161]. South Africa is also responsible for issuing weather and navigation warnings in this area.

¹³Laws which are derived by a judicial system through established legal precedents (in the South African context, common law is synonymous with *case law*). Conversely, laws enacted by a legislature are not considered common law.

¹⁴The SADC is an inter-governmental organization which promotes the interests of the region and its members. Its mission includes the promotion of equitable economic growth and socio-economic development, good governance and durable peace and security in the region. The organisation currently comprises fifteen member states [72]. These include the coastal states of Angola, Democratic Republic of Congo (which is connected to the Atlantic ocean by a 45 km stretch of coast), Namibia and South Africa in the west, as well as Mozambique and Tanzania on the east coast of Africa (the island nations of Seychelles and Mauritius are also members, but Madagascar was recently suspended due to the illegal ousting of its democratically elected government [40]).

equipment approval, registration of vessels and the provision of maritime safety information to vessels [162, 111]. SAMSA is accountable to the Minister of Transport and has the principal mandate to respond to maritime emergencies and maritime pollution, detect and assist those in distress, establish safety and environmental protection standards, and to monitor and enforce these standards.

1.2.1 Fisheries Management

The *Marine Living Resources Act* of 1998 placed fisheries management under the auspices of the Department of Environmental Affairs. Presently, fisheries management is the responsibility of the *Department of Agriculture, Forestry and Fisheries* (DAFF). The department¹⁵ is tasked with preserving and protecting the marine environment. Central to this mandate is the notion of restoring fisheries to sustainable levels and maintaining their productive capacity. It is also tasked with conservation and sustainable use of these resources [96].

South African fisheries are utilized by four broadly defined user groups. On a national level these users may be divided into subsistence, recreational and commercial fishing groups, whilst the fourth group encompasses foreign fishing enterprises. Industrial fishing enterprises are reported to occur predominantly on the west coast of South Africa, whilst recreational and subsistence fishing is more prevalent along the south and east coasts¹⁶ [14]. The *Cape hake* trawl fishery accounts for the largest portion of local commercial fishing [178]. This offshore deep-water hake fishery (the two commonly occurring species in South Africa are deep-water and shallow-water hake) operates on the west coast of South Africa in a region stretching south from the southern border of Namibia. The fishing vessels target the fish during the day when they aggregate near the ocean floor (they feed during the night) [116].

In order to ensure sustainability of South African Fisheries, control over their utilisation is required. These control measures include setting the *total allowable catch*¹⁷ (TAC) and *total allowable effort*¹⁸ (TAE), closed seasons, closed areas, gear restrictions and minimum species sizes [14, 96]. The Act stipulates that the Minister may create regulations which prescribe the procedures to be observed by foreign fishing vessels whilst in South African waters, as well as how fishing gear is stowed aboard vessels navigating those waters [96]. The DAFF is responsible for setting and enforcing the regulations pertaining to all control measures and is additionally responsible for administering fishing rights and apportioning the TAC on an annual basis between the various fishery user groups.

Another measure at the disposal of the ministry is the zoning of *marine protected areas* (MPAs). Their establishment serves to protect particular fauna and flora, and to preserve entire ecosystems. Fishing, dredging, construction of structures, waste discharge (polluting) and destruction of the natural environment is prohibited in these regions [96]. However, permission for fishing or construction may be granted by the minister if it is deemed to be appropriate or necessary. Indeed, MPAs may be divided broadly into *take* and *no-take* categories with respect to fishing in general (the aforementioned closed areas are no-take MPAs). Furthermore, distinctions are

¹⁵The fisheries branch of the department is responsible for the management of marine resources. The branch was formerly known as *Marine and Coastal Management* (MCM) and is presently based in Cape Town [110]. The abbreviation MCM is still widely used to refer to the fisheries branch.

¹⁶The waters of the Benguela current are rich in nutrients, resulting a greater availability of natural resources along the west coast, whilst the east and south coasts have a greater diversity of species [14].

¹⁷DAFF is responsible for conducting research to determine the TAC which informs the annual quotas set within the various fisheries.

¹⁸The TAE limits the resources that a fisherman/company may employ in catching fish (for instance, the number of boats or traps used).

made in take MPAs that specify who may fish in a particular MPA. For instance, in the Betty's Bay MPA, shore-angling for line fish by recreational fisherman is permitted, whilst in the Castle Rock Marine MPA, properly licensed local commercial fishing vessels are allowed to catch snoek [97]. Various notices issued by the government specify the rules governing newly defined MPAs and provide their GPS coordinates [41].

Monitoring also plays an essential role in regulating fisheries and various means are used to monitor the different fishing sectors. For example, observers may accompany fishing vessels to sea or monitor vessel landing sites. These data, and those collected through other monitoring activities, are used in inferring a stock status and recommendations of TAC¹⁹ [66]. In addition to monitoring approaches, fishing vessels are increasingly required to carry an operational *vessel monitoring system* (VMS). Vessels in the Cape Hake trawler fishery have been utilising VMS since 2002 whilst other fisheries, such as the traditional line fish and hake hand-line fisheries, followed suit in 2007 [191]. VMS technology provides spatio-temporal data and its current use in South Africa is to monitor a vessel's location with respect to MPAs. The technology is capable of reporting catch statistics too, but these features are not yet in use.

The notions of *monitoring, control and surveillance* (MCS) are central to fisheries management and provide a mechanism for combating illegal, unreported and unregulated fishing. Besides the obvious effect that illegal fishing has on fishing stocks, there is also a higher impact on the environment in terms of by-catch and incidental mortality of marine animals (such as sharks and birds) as these activities need not follow the standards applicable to legal fishing enterprises.

1.2.2 Search and Rescue

South Africa is responsible for *search and rescue* (SAR) within an IMO designated region of responsibility (otherwise referred to as a *search and rescue region* — see Figure 1.1). A *marine rescue coordination centre* (MRCC), under the management of SAMSA, facilitates rescue operations within these areas. Events that would typically fall under the province of the MRCC are aviation accidents at sea, forced landings, crews or passengers of seafaring vessels that are in distress, and maritime accidents. A multilateral agreement was signed in 2007 between Namibia South Africa, Madagascar, Mozambique, and the Comoros. This agreement established a sub-regional MRCC whose base of operations is Cape Town, South Africa (the existing South African MRCC was expanded to host the regional centre) [140].

Numerous organisations are involved in SAR operations and their efforts in the maritime domain are coordinated by the MRCC. An invaluable resource within this domain is the NSRI, a volunteer sourced non-profit organisation that boasts 92 rescue craft as well as 32 coastal and three inland dam bases [124]. The resources available through the NSRI, the South African Navy and the Air force all play a crucial role in SAR operations [102].

1.2.3 Monitoring and Enforcement

Estimates of the length of the South African coastline²⁰, including the Prince Edward and Marion Islands coastlines of 32 km and 134 km respectively, range from 2 881 to 3 924 km [21, 161, 173, 176]. The South African EEZ is estimated to cover an area of roughly a million square

¹⁹The annual TAC for demersal Cape Hake was 150 000 tonnes in 2009 [66].

²⁰The calculated length of a coastline is inversely correlated to the discrete measured length that is used in approximating it (the scale of the measurement). The smaller the length of measurement, the greater the length of the coastline. This phenomenon is known as the *coastline paradox* [199].

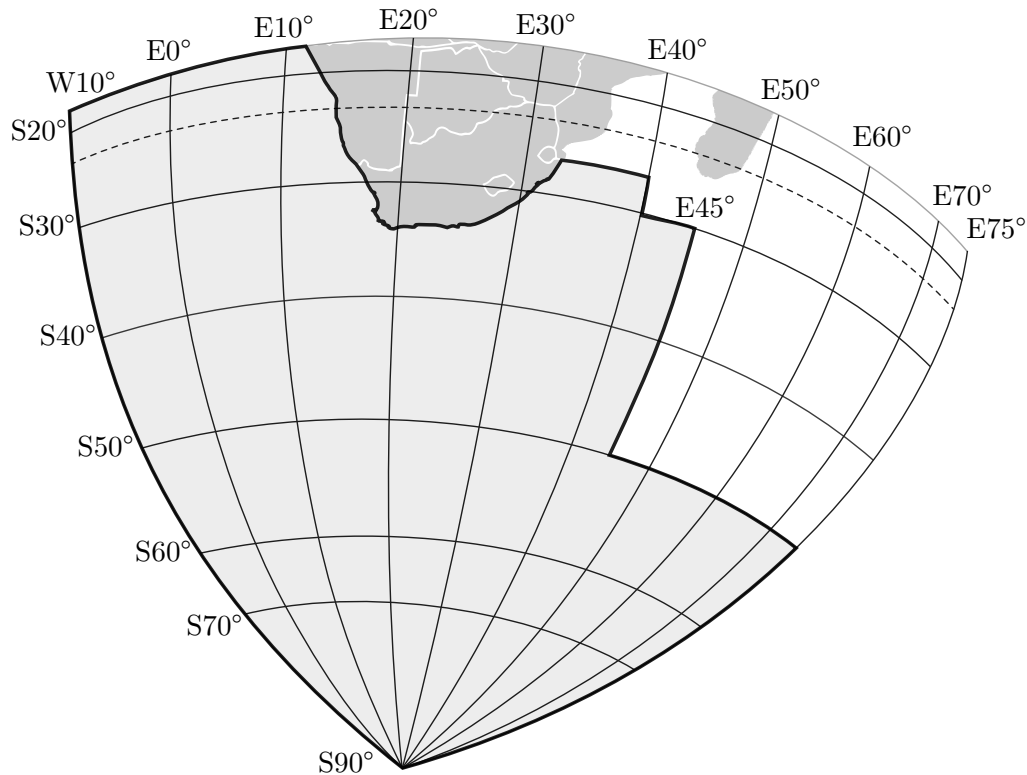


FIGURE 1.1: *South Africa's search and rescue region.*

kilometres²¹. Effective law enforcement operations across such vast distances requires effective surveillance as well as vessels capable of enforcing the laws.

The DAFF has a number of vessels under its auspices²² to perform this task. Included in this fleet are three research vessels (see Figure 1.2), four environmental protection vessels and a high speed anti-poaching vessel.

The research vessels are responsible for the monitoring of the marine environment and their responsibilities include research on marine living resources and monitoring of oceanographic conditions. The flagship vessel, *Ellen Khuzwayo*, features on-board laboratories for oceanographic studies and fish sampling, as well as acoustic equipment for fish surveys. She has a range of 200 nautical miles and is capable of remaining at sea for two weeks at a time [108].

The environmental protection vessels (or patrol vessels) act to protect South African marine resources through the prevention of poaching and protection of fisheries. The vessels are also capable of performing search and rescue operations, firefighting, towing and oil spill counter-measure functions [166]. There are three *inshore* vessels (operating up to 200 nautical miles from the coast) in the *Lilian Nyogi* class, namely the *Lilian Nyogi* herself (entered into service in 2004), the *Ruth First* (see Figure 1.3(b)) and the *Victoria Mxenge* (both entering into service in 2005). This fleet also features an *offshore* vessel (capable of patrolling beyond 200 nautical miles), namely the *Sarah Baartman*, which is capable of extended stays at sea and has patrolled

²¹If South Africa's claim to its continental shelf is successful, that area will expand by another 900 000 square kilometres [173].

²²The DAFF was created in 2009 during a restructuring of the South African cabinet. It assumed responsibility for fisheries from the *Department of Environmental Affairs and Tourism* (DEAT). The DEAT was split into the *Department of Environmental Affairs* and the *Department of Tourism* at this time [165]. Research and patrol vessels fell under the management of the DEAT and all articles relating to them reflect that.

(a) *Africana* (launched in 1982)(b) *Algoa* (launched in 1991)(c) *Ellen Khuzwayo* (launched in 2007)

FIGURE 1.2: South African research vessels [100].

as far afield as Tanzania²³ [153]. Of these patrol vessels the Sarah Baartman is the only one that currently carries an AIS transponder.

A high speed and highly manoeuvrable *interceptor* vessel, the Florence Mkhize, entered into service in 2006 to assist the larger patrol vessels in the fight against poaching²⁴. The vessel is fourteen metres long and is capable of reaching speeds of 60 knots (120 km/h — the four large patrol vessels are capable of speeds of roughly 25 knots in comparison) [99].

In addition to patrolling the waters, effective surveillance capability is necessary. There are a number of surveillance technologies in place to perform this task, such as the aforementioned VMS and AIS systems. The AIS capability recently underwent an upgrade so as to include *aids to navigation* capabilities. AIS data are transmitted using high frequency radio waves (in the VHF range of 30Mhz to 300Mhz) which are ideally suited for terrestrial communication and have a range a little further than line-of-sight. The *Centre for Sea Watch and Response* (CSWR), a department of SAMSA, was launched in 2009 and this centre has access to the

²³The Sarah Baartman participated in a joint SADC Fisheries Patrol of 5 000 nautical miles during 2005. Inspectors from Mozambique, Tanzania and South Africa were aboard the vessel which observed fifty vessels and boarded forty seven of them (three fines were issued and a single vessel ordered back to port). This operation was supported by aerial surveillance [107].

²⁴This vessel is instrumental in combating the poaching of abalone as poachers often make use of *super ducks*, themselves fast and manoeuvrable vessels. In 2006, inspectors aboard the Victoria Mxenge noticed divers in the Bird Island MPA and had to call upon the services of the Florence Mkhize as they were unable to apprehend the divers once they had boarded their super duck. The poachers were ultimately apprehended on land after they fled into a river with the Florence Mkhize in pursuit (they were apprehended by a ground-team which was in the area). The divers were arrested and fined approximately R50 000 000. [171].



(a) *Sarah Baartman* (launched in 2005) [100]. (b) *Ruth First* (launched in 2005) [150].

FIGURE 1.3: South African patrol vessels.

VMS and AIS data collected along the coasts, as well as to the *long range identification and tracking system* (LRIT). The LRIT tracking system is a satellite technology which enables the CSWR to identify and track vessels up to 1 000 nautical miles from the South African coastline. This service allows a flag state to track vessels that bear its flag globally²⁵ [27]. Besides spatio-temporal surveillance, the CSWR also monitors maritime radio communication. An integrated shore-based radar system (COASTRAD) is also operational and may be used to monitor vessel movements [178].

The South African Navy was granted operational control of the patrol and research vessels in April of 2012 and are also in the process of establishing two *Maritime Domain Awareness* (MDA) Centres. The centre based at Silvermine in Cape Town will monitor the west coast, whilst the centre based in Durban is responsible for the east coast. These centres have been created in response to the ever increasing threat of piracy which is extending southwards from the horn of Africa²⁶ [25].

1.3 Problem identification

MDA is a domain-specific situation awareness. *Situation awareness* (SAW) as a general concept is described by Endsley as “the perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” [36]. Modern surveillance systems are capable of collecting vast quantities of data. As a result, the identification of concepts of import becomes increasingly difficult for surveillance operators. More data does not necessarily translate into more information — indeed, it is a human operator’s ability to perceive relevant information from often overwhelming quantities of surveillance data that makes it useful [36]. Experienced operators rely on an array of cognitive processes in achieving better SAW. One such process is *pattern matching*, where a decision maker may recognise perceived information as an example of a class of situations [36].

In many settings, multiple sensor observations of the same phenomenon are typically aggregated into a single observation. Achieving an accurate assessment of the environment through perception therefore relies heavily on *multisensor data fusion*. Data fusion is also concerned with inferences from the observations, such as the identity of the observed intensity, activity inter-

²⁵Vessels fitted with LRIT transponders report their positions every six hours by default. The transponder can be turned off by the vessel, but it may only do so at the request of the flag state.

²⁶The first piracy attack in SADC waters was recorded in 2010 and SAMSA reports that the MRCC is receiving an increasing number of distress signals from vessels at sea and vessels calling to port in South Africa.

pretation and intent prediction (see Figure 1.4) [54]. Achieving MDA relies on data fusion to maintain a model of the observed environment external to the operator [85]. This information is then presented to human operators responsible for monitoring and issuing calls to action.

Both of these concepts rely on a *human in the loop* and although it may appear desirable to fully automate the surveillance task, most applications of this approach highlight the superior performance of humans in comparison to automated systems [145].

Advances in technology are steadily contributing to improved MDA. South Africa actively monitors AIS and VMS transmissions and monitors vessels from a number of strategically placed coastal radars. Such surveillance systems provide maritime domain operators and analysts with vast quantities of data from which the operators must distill the relevant situational facts of interest. Support tools play an invaluable role in this regard by assisting operators in processing these vast collections of data, identifying concepts of interest and predicting future occurrences. Data driven approaches from the scientific domain of pattern recognition are ideally suited to these problems and have been applied successfully to particular instances of the problem.

Considering the vast ocean area that South Africa patrols and the limited law-enforcement and rescue fleet at its disposal, a contribution may be made to maritime safety and security through improved MDA so that resources may be directed effectively.

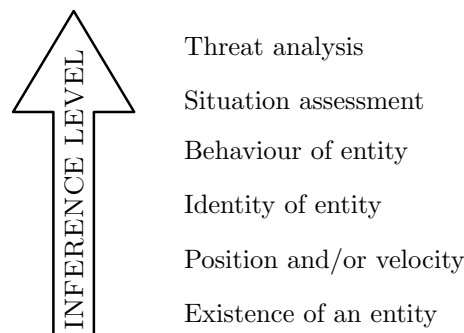


FIGURE 1.4: The data fusion inference hierarchy [54].

1.4 The thesis of this dissertation

Using data collected from sensors, a contribution may be made to coastal safety and security in the form of automated decision support based on the analysis of the spatio-temporal patterns of observed sea vessel activities so as to predict future behaviour and detect illegal activities at sea.

In particular, the inherent structure in vessel motion which manifests itself in a vessel's trajectory data may provide sufficient information to differentiate between different activities. Additionally, exploitation of such low-level information may allow for the discovery of more complex activities within the maritime environment. Using a feedback approach where an operator verifies or discards discovered activities as requiring further investigation may contribute to an effective *decision support system* (DSS) capable of adapting to and integrating new information. The thesis of this dissertation is that such a DSS may contribute directly to improved MDA.

1.5 Dissertation scope

An adaptive early-warning maritime surveillance decision support system is designed in this dissertation which is capable of supporting a maritime operator in his decision making process with regard to various threats and scenarios. Automatic decision making is not pursued as the aim of this study is to provide a support tool to an operator (following the *human in the loop* paradigm). This study is strictly limited to the maritime domain and the primary source of data is spatio-temporal AIS data. Furthermore, this study is primarily focused on *detecting* threatening behaviour and not necessarily with *quantifying the degree* to which this behaviour is threatening.

1.6 Dissertation objectives

The aim of this study is to lay the foundation for the design of a decision support system in the maritime domain. In order to achieve this goal the following objectives are pursued in this dissertation.

- I To *perform* a literature survey of approaches to trajectory classification and clustering in various scientific domains.
- II To *augment* the survey in I above with a review of approaches particularly focused on the maritime domain.
- III To *propose* a general design framework for an adaptive maritime surveillance DSS with a primary focus on spatio-temporal data. Within the realm of this objective the following aims are pursued:
 - (a) To *discuss* and *identify* threatening behaviours and how they may be addressed within the context of such a system.
 - (b) To *investigate* data driven approaches to discovering commonalities within a region of interest.
 - (c) To *achieve* trajectory classification within this framework.
- IV To *propose* methodologies by which the elements of such a decision support system may be realised.
- V To *demonstrate* the practical applicability of the DSS proposed in III above in the context of a real case study involving vessel trajectory data.

1.7 Dissertation organisation

In order to provide a methodological foundation for a surveillance DSS within the maritime domain, prevailing modelling approaches are considered in Chapter 2 by way of considering motion patterns in general. Two modelling paradigms, namely expert and data driven approaches, are considered specifically with more emphasis placed on the latter methodology. The approaches to describing and analysing trajectories in various fields are discussed in detail in Chapter 2. Applications in computer vision and ecological modelling are discussed and approaches to motion pattern analysis in the maritime domain are finally considered in fulfilment of Objective I and in partial fulfilment of Objective II of §1.6.

Chapter 3 continues with the review of literature specifically focused on the maritime domain, with particular emphasis placed on threats and what constitutes threatening behaviour. Existing maritime surveillance systems are considered and pertinent notions presented in the literature are discussed in fulfilment of Objective II of §1.6. This discussion provides a theoretical basis for the novel DSS which is presented in Chapter 3. This DSS framework combines a rule-based and data-driven approach in an effort to identify known threats and identify anomalous or unknown activities. Moreover, the proposed DSS aims to catalogue and store data for later processing, and furthermore, attempts to engage the operator at various stages of the process in fulfilment of Objective III of §1.6.

Suitable rules were identified (from the literature and from personal communications with maritime surveillance experts). These rules are discussed in some detail in Chapter 4, are used to populate the rule-based component of the DSS, and are either applied directly, modified where necessary, or newly proposed in fulfilment of Objective III(a) of §1.6. This rule set is not exhaustive and was consciously restricted to rules that can easily be applied to kinematic vessel data without relying on meta-data such as a vessel's cargo or her crew.

In Chapter 5, an approach to trajectory data mining is proposed in fulfilment of Objective III(b) of §1.6. This approach seeks to identify similar motion patterns through the application of two simple clustering techniques. The efficacy of this approach is demonstrated with respect to a synthetic data set and the system component in which it resides is discussed in detail.

Trajectory classification is pursued using these mined data in Chapter 6 by way of Hidden Markov Models. This modelling approach is discussed in detail and is demonstrated on these data in fulfilment of Objective III(c) of §1.6. The combination of the rule-based system, the data mining approach and the activity classifier provides an initial approach to populating the framework of the proposed DSS in fulfilment of Objective IV of §1.6.

Lastly, a data-driven activity classifier is derived in a special case study in Chapter 7 based on AIS data collected around the Port of Cape Town. The properties of the data are discussed and a single activity, namely that of travelling along a well-established route, is identified and modelled using the proposed methodology in fulfilment of Objective V of §1.6. Thereafter, conclusions are drawn and proposals for future work are made in Chapter 8.

CHAPTER 2

The analysis and learning of motion patterns

Contents

2.1	Introduction	15
2.2	Prevailing paradigms	16
2.3	Motion patterns	18
2.3.1	<i>Computer vision</i>	20
2.3.2	<i>Ecological modelling</i>	24
2.3.3	<i>Maritime motion patterns and threat assessment</i>	26
2.4	Summary	33

2.1 Introduction

Research into automated behaviour analysis is very active in the fields of geographic information science, computer vision and ecological modelling. These approaches are often concerned with determining behavioural patterns in vast collections of data without building explicit models of the entities under observation. Such data-driven approaches are typically trajectory-based in the sense that the results of motion (kinematic information) form the basis for the analysis. The advantages of this approach are that dynamics considerations, and therefore entity-specific details, often need not be considered and that many different entities may be observed and analysed within the same framework. However, solutions to this problem are not limited to kinematic information and it has been found that including additional information may lead to the discovery of a richer set of behaviours.

The methodologies of *expert systems* and *machine learning* are discussed in §2.2. These approaches have been applied successfully in a number of domains. Motion patterns are considered in §2.3 in greater detail. Particular attention is paid to the solutions found in the fields of computer vision (§2.3.1) and ecological modelling (§2.3.2) as they deal directly with kinematic data derived from observations. The classification of vessels and their behaviour in the maritime domain are addressed in §2.3.3.

2.2 Prevailing paradigms

The interaction of an intelligent agent with its environment has been described in detail by the artificial intelligence community. The notion of knowledge and the ability to reason successfully are essential to inducing successful behaviours in agents. These two notions are indispensable when operating within partially observed environments where problem solving agents are ultimately more successful¹ than *reflex* agents (who at best may achieve success accidentally) [149]. Agents capable of inferring unknown facts about their environment by utilising knowledge in combination with prevailing rules are known as *knowledge-based* or *logical agents*.

The domain knowledge required by the agent is encoded in a *knowledge base*. Informally, this knowledge base is composed of *sentences* which make particular assertions about the domain in question [149]. Coupled with an inference mechanism, this system allows an agent to derive new sentences from existing knowledge which enables it to take an appropriate action. The construction of a knowledge base is typically performed by knowledge engineers who work with domain experts (in the event that they are not experts themselves) to create a formal representation of the objects and relations within the domain [149]. The resulting vocabulary (formal languages are typically used) is known as the ontology of the domain. These systems are also referred to as *expert systems*. The knowledge base contains facts known about the domain which may be derived or known *a priori* (for example, the location of the agent in relation to the world in which it finds itself). As may be seen in Figure 2.1, an expert system contains encoded rules which may be used to infer new facts when presented with existing knowledge.

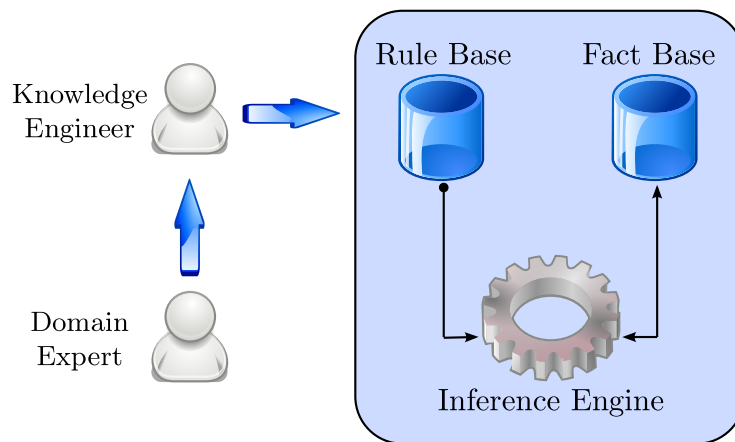
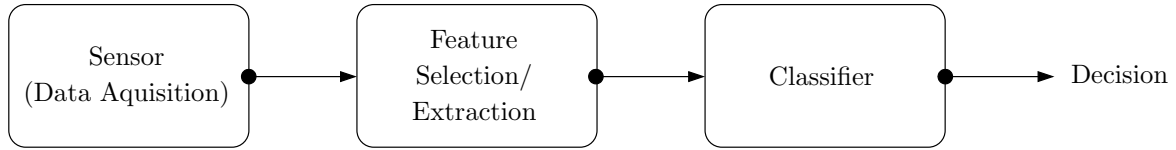


FIGURE 2.1: A simplified overview of an expert system.

These notions have been used successfully in the maritime domain and are well suited to address problem elements which are deterministic in nature. This approach has been employed in maritime anomaly detection [145] and various ontologies have been defined for situation awareness [82, 105]. A severe shortcoming of traditional expert systems is that learning new rules (without a knowledge engineer defining them explicitly) and modelling temporal data may only be achieved through numerous extensions to the system.

An alternative, and commonly used paradigm is that of statistical pattern recognition or machine learning. Pattern recognition has found application in such areas as automatic character recognition, speech recognition, medical diagnosis (detection of tumours in magnetic resonance images), biometric recognition, remote sensing and data mining [74, 198] (examples of various classification tasks are presented in Table 2.1).

¹Success in the robotics domain may be achieved if a robot navigates a scene successfully.

FIGURE 2.2: *The basic pattern classification process.*

Problem Domain	Application	Input Pattern	Pattern Classes
Bioinformatics	Sequence analysis	DNA/Protein sequence	Known types of genes/patterns
Data mining	Searching for meaningful patterns	Points in multi-dimensional space	Compact and well-separated clusters
Document Classification	Internet search	Text document	Semantic categories (e.g. business, sports)
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters, words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective/non-defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres (e.g. action, dialogue)
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorised users for access control
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory enquiry without operator assistance	Speech waveform	Spoken words

Table 2.1: *Pattern recognition applications [74].*

A *pattern* may be considered to constitute measured ‘features’ of an object where these measurements aim to capture different aspects of the object that are deemed to be relevant to the investigation (for example, measurements of the acoustic waveform in a speech recognition problem would constitute such a pattern) [198]. The intention is to discriminate between patterns in a meaningful way so as to facilitate decision making. This is achieved by determining a mapping from the feature space to the decision space. This mapping is usually described by a carefully selected model. The computation of the model parameters is referred to as *learning* as these parameters are calculated from training sets of exemplar data. Pattern recognition methods may be considered to be either *supervised* or *unsupervised* learning methods [198]. The chosen method depends on the nature of the available training data. In the former case, the data have associated class information which is used in constructing a classifier, whilst in the latter, an absence of class information means that features within the data are used to partition data into groups.

A simplification of the the classification task is presented in Figure 2.2. Once a problem has been identified and formulated, data are collected and measurements are made of the appropriate variables. An initial examination of the data may then take place in order to gain some insight into the structure of the data (this may include testing the quality of the data and gaining measures of the location and spread of the data via sample means and standard deviation).

Thereafter, appropriate variables from the measured set may be chosen (*feature selection*) or new variables may be obtained by some transformation of the original measurements (*feature extraction*) — see Figure 2.3. A training set of exemplar patterns may then be used in the design of a classifier. Lastly, the efficacy of the trained classifier may be measured by its performance with respect to an independent *test set* (parameter estimation is usually performed on a *hold-out* set or validation set, after which the ‘learned’ model is applied to a test set).

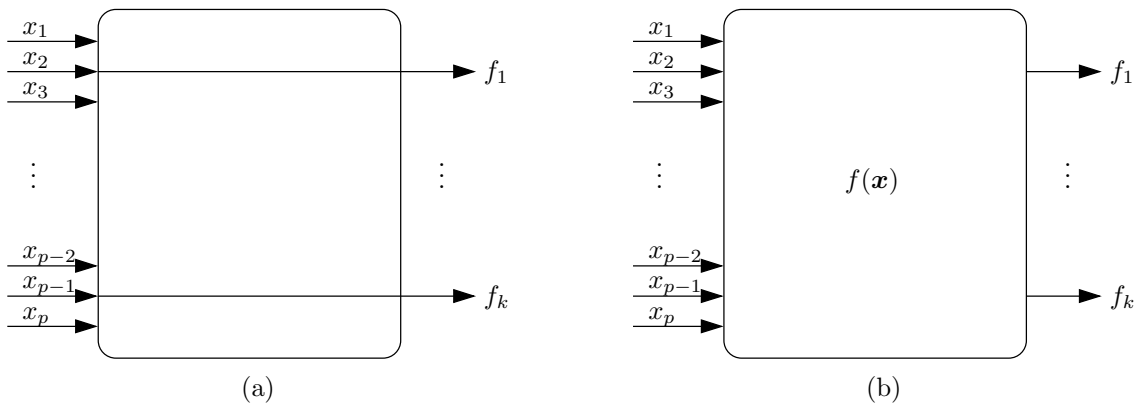


FIGURE 2.3: (a) Feature selection and (b) feature extraction.

Video and maritime surveillance are good examples of this approach. Trackers usually extract the trajectories of entities obtained via sensors. These data are typically filtered (to remove noise and outliers), resampled and smoothed. Pertinent features are selected or devised and these features are used to construct a model which is capable of identifying a future observation as representing a specific kind of motion or behaviour. The objective of such a model may be to find atypical behaviour in some context or to follow a more high-level approach by classifying activities themselves.

2.3 Motion patterns

Dodge *et al.* [31] provided an initial taxonomy of movement behaviour so as to lay the foundation for the development of data mining algorithms capable of addressing broader classes of motion. They argue that in order to develop generic algorithms capable of extracting motion patterns from large collections of trajectories it is necessary to identify similarities and differences that are inherent to the *motion patterns*² derived from different moving objects.

The authors proposed a conceptual framework comprising *motion parameters*, *path type*, *motion constraints* and *granularity* of measurements of motion. Motion parameters encompass *primitive parameters* (position and time), and *primary* and *secondary* derived parameters (the authors chose to name these primary and secondary derivatives). Primary derived parameters include distance, direction and spatial extent in the spatial dimension, and duration (time interval over which a particular motion is observed) and travel time in the temporal dimension. Examples of secondary derived parameters are spatial distribution, change of direction (angular velocity), sinuosity³, acceleration and approaching rate. The path travelled by an entity is classified as being either continuous or discontinuous. The authors identified four motion constraints experienced by moving entities. These include the physical capabilities of the moving entities (such

²Dodge *et al.* [31] defined motion patterns as any recognisable spatio-temporal regularity in a set of movement data.

³*Sinuosity* is a scalar index described by the ratio of distance to displacement.

as acceleration and turning circle), the spatial constraints experienced during motion (harbours and land masses), environmental constraints (such as weather conditions) and interactions from other moving entities that result in changes of the motion (such as collision avoidance). They also considered temporal granularity of measurements and the importance of domain-specific knowledge was highlighted in this regard (over and under sampling need to be avoided when resampling trajectories).

Building on this conceptual framework, the authors then presented classes of movement patterns (see Figure 2.4). *Generic patterns* are patterns that might be identified in various behaviours gleaned from movement. These patterns serve as the basic building blocks of more complex behaviours represented by *behavioural patterns* in Figure 2.4. *Co-location* in space, for instance, is a primitive pattern exhibited in the trajectories of entities which occupy common positions in space. The spatio-temporal equivalent of this notion is defined as *co-incidence in space and time*. The parameters that describe these primitive patterns may be identified in the conceptual framework as position, and position and time, respectively. The *concurrency* primitive pattern is a pattern in which entities exhibit motion which feature common motion parameter values over some period of time. This pattern may be exhibited by a tugboat with a vessel in tow or by a pilot vessel and the vessel it is guiding.

Dodge *et al.* [31] considered behavioural patterns dependent on the context of movement as well as the entity executing the movement, and so they provided only a tentative list of possible behaviours. As an example, the behavioural pattern of *fighting* features different generic patterns, amongst them concurrency and co-location in space, as well as more compound patterns in situations where groups are fighting.

Formalising patterns is conceptually useful and Dodge *et al.* [31] provided an example that utilises their classification. This example addresses the possibility of using trajectories generated by a particular class of entities, as proxy data for designing and testing a system on trajectories created by different entities (although no quantitative measures were proposed).

Essential to any pattern recognition method is a productive similarity measure. Dodge *et al.* [30] went on to propose a method for extracting movement parameters from trajectories generated by different entities. They validated their approach by attempting to classify these trajectories as originating from a particular class of entities. In keeping with their desire for generic approaches [31], their proposed method extracts generic movement parameters from the trajectories under consideration. Raw data are prepared by first removing outliers⁴. Thereafter, the data are resampled to produce equi-temporal positions by means of linear interpolation. Lastly, the data are smoothed by means of smoothing methods such as spline approximation, kernel-based smoothing or Kalman filtering (the authors do not disclose which method they employed).

Movement parameters, such as speed, acceleration, displacement and sinuosity, were computed for a trajectory. The authors proposed that these parameters first be computed on a global scale, after which trajectories should be decomposed recursively in order to find segments in which the computed parameters are approximately homogeneous. Each successive refinement was performed on movement parameter profiles (for instance, a speed-time plot) and statistics such as mean, median, standard deviation and variance were computed. The authors suggested the application of *Spearman rank correlation* to determine whether the chosen movement parameters are all relevant.

A supervised classifier was constructed using features extracted from the profile decomposition algorithm described above (the pseudo code is provided in [30]). *Principal component analysis*

⁴Measurements that reported a position at a distance greater than three standard deviations of the distances between consecutive measurements, were considered to be noise and were removed.

was used to reduce the dimensionality of the feature space before training a *support vector machine*. GPS data for motorcycles, cars, bicycles and pedestrians were obtained, as well as planar coordinates derived from eye trackers. The test data set consisted of 165 trajectories per transportation class and 115 trajectories of eye movement. The authors obtained an accuracy of 82% of cases correctly classified in the test set.

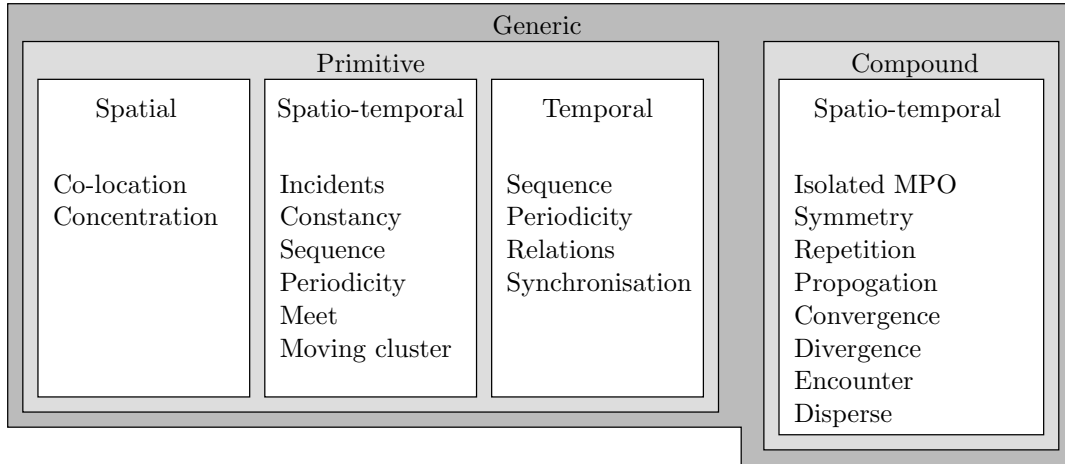


FIGURE 2.4: *Generic movements patterns, as defined in [30].*

2.3.1 Computer vision

Pusiol *et al.* [135] devised an activity classification method which employs minimal supervision during the learning phase. The activities of a single person were learned and this knowledge was used to identify the behaviour of a second person. This trajectory-based approach learns the topology of the observed scene⁵ through identification of so-called *slow regions* (designated by a section of the trajectory for which the average speed drops below a pre-defined threshold within a fixed window of measurements). The trajectory segments between slow regions may be utilised in the construction of features or so-called *primitive events*. Once slow regions have been established for trajectory segments, these segments are collapsed into single-point representations by taking the average over a segment that is deemed to be a slow region. In order to establish representative areas of activity (an activity is modelled by the primitive events present in the time window during which the activity occurs), the authors clustered these slow-points using the method of *K*-means (incorporating a Euclidean distance measure). The cluster centres were then chosen to represent the slow regions.

The scene under observation was that of a combined kitchen and living room area. This area featured a couch, table with chairs and sink/kitchen area. Typical activities in this scene included *'in the kitchen'*, *'setting the table'*, *'sitting in armchair'* and *'in corridor'*. The authors stated that they attempted to use Bayes' rule to evaluate the probability of a primitive event belonging to a particular activity, but that the volume of data was insufficient for this purpose. As a result histograms of the primitive events associated with a particular activity were constructed. Their method thus learned activities in an unsupervised manner by identifying regions of interest. These activities were then labelled to establish a *ground truth*. Activities in the test data were identified by first matching the slow regions of the test data with those of the model (Pusiol *et al.* [135] describe a method to do this). Thereafter, these activities were matched to those

⁵In computer vision applications, an area under surveillance by optical sensors is typically called a *scene*.

of the model by extracting histograms of primitive events within a sliding temporal window (the sliding window relaxed the duration of the activity as learned by the model). The authors reported favourable results in matching the activities of a second person to those of the first (in this case the ground truth).

Naftel and Kalid [119] agreed with Dodge *et al.* [30] that generic approaches to identifying motion patterns were necessary. To this end they presented an unsupervised learning scheme which utilised a *self organising map* (SOM)⁶ for performing trajectory clustering. Although concerned with video trajectories, this approach finds application in geo-referenced scenarios⁷. Instead of using the discrete measurements (usually obtained by means of a tracker) that make up the trajectory (referred to as *point-based flow vectors*) as features during the learning phase, the authors chose parametric representations of the trajectories. This *feature extraction* approach reduced the dimensionality of the feature space and made computations more efficient. Trajectories were approximated using three different methods, namely Chebyshev polynomials, Fourier series approximation and least squares approximation. For each of the three approximation techniques, an equal number of basis functions were chosen and the coefficients of these basis functions served as the features. A tracker typically returns noisy (x_i, y_i) measurements which were considered as two separate time series in [119], namely (t_i, x_i) and (t_i, y_i) . Each of these component time series were approximated and the resulting coefficients were concatenated into a feature vector. The similarity between two feature vectors was measured in terms of their proximity in feature space (determined by Euclidean distance).

The components of these feature vectors served as the inputs to the SOM which mapped similar trajectories to the same output node. When training the SOM, the authors set the number of output nodes to a value greater than the expected number of distinct patterns in the data. Similar nodes were thereafter merged until a desired number of clusters were obtained (Naftel and Kalid determined this number empirically).

The clusters, as determined by the SOM, served as categories for a k -NN classifier⁸ (Naftel and Kalid determined the value of k by means of a *leave-one-out* analysis). The SOM may be seen as performing a labeling function on the trajectories presented in the test set. The classifier was trained on these ‘labelled’ data. The distance measure utilised in classification by the authors mentioned above was that of the *Mahalanobis distance*. If a trajectory was deemed to be an outlier when considering its nearest class centre, it was classified as anomalous.

Naftel and Kalid [119] analysed the performance of their model using three data sets: labelled trajectories from the CAVIAR data set [43], the *Australian sign language* (ASL) data set [78] and data recorded by the authors. The efficacy of the approximation techniques were tested along with determining the ideal number of coefficients to use. Noise was added to the trajectories in the CAVIAR data set and the *retrieval accuracy*⁹ was tested. The authors found that Fourier series approximations performed the best in this context. The experiment was repeated using partial trajectories (sequences of measurements within trajectories were randomly discarded) and in this case least squares approximation fared the best.

⁶An SOM is a neural network in which each node in the input layer corresponds to a feature whilst each node of the output layer corresponds to a cluster in feature space.

⁷Naftel and Kalid [119] performed clustering of motion trajectories that were projected onto a camera’s image plane. Although this approach is view-dependent and is not necessarily ideal for video surveillance scenarios (the authors pointed out that a multiple viewpoint system would require training for each camera), it is well suited to centrally geo-referenced trajectories.

⁸The class membership of an observation is determined to be the same as that of the majority of its k nearest neighbours.

⁹Retrieval accuracy in this context was taken as a measure of how often an approximation of a corrupted trajectory was matched to its uncorrupted approximation using Euclidean distance measure [119].

K -means and SOM clustering methods were compared with respect to the CAVIAR data set (the data set is composed of hand-labelled trajectories). Two different feature sets were used, one comprising flow-vectors and the other Fourier series coefficients. In both cases the SOM performed better in determining an example trajectory to be a member of the correct cluster. The results were similar for partial trajectories and it was established that the parametric representation performed better than that of flow-vectors in this context.

The ASL data set has been used in a number of motion pattern applications and the authors chose to construct a classifier using these data so that their approach may be compared to existing methods. Naftal and Khalid [119] reported similar results to those obtained by Bashir *et al.* [5] and they claimed that their method was “conceptually simpler and computationally less expensive.” Lastly, anomaly detection was assessed on their recorded data set. This data set featured specific types of motion with a number of deliberately abnormal trajectories. The latter trajectories were excluded from the training set but presented to the classifier as members of the test set. These trajectories were identified as anomalous by the classifier.

Vlachos *et al.* [193] presented a method for discovering similar trajectories through hierarchical clustering and nearest neighbour classification. Their similarity measures were based on the *longest common subsequence* (LCS) problem¹⁰, which they compared to Euclidean distance and dynamic time warping¹¹ approaches. The disadvantage of the latter two methods, namely that all elements in the sequence must be matched¹², was highlighted by the authors. This property meant that outliers were considered (and in the case of Euclidean distance, they greatly skew the results) and that temporal series had to be of the same length (be composed of the same number of measurements). Consequently, the authors devised an alternative similarity function based on the LCS problem. This function considered points in space to be similar if they were deemed sufficiently close in space and time. This approach was, however, not translation-invariant, and so an attempt was made to extend the measure to address trajectories that are similar in space but have different origins. An additional similarity function was specified that considered a family of translations when computing similarity. The authors showed that a finite set of translations¹³ could be enumerated efficiently and they provided the algorithms to do so.

Once the similarity measures had been defined (the reader may consult [193] for the definitions themselves), the authors presented the hierarchical clustering approach that was used. Query trajectories were matched to trajectories or clusters in the resulting tree using a nearest neighbour approach (where the distance measure was based on the LCS problem). The authors conducted experiments testing the performance of the approximation algorithm that computes one of the similarity functions (on a data set generated from tagged marine animals). Experimental results were also presented for the classification method (the ASL data set featured as one of the test data sets). The authors reported that their method obtained timely and accurate results when computing the distance between two trajectories (using the distance measures defined in [193]). However, significant disadvantages of this approach are that trajectories featuring different origins needed to be parallel in space and that rotations of entire trajectories

¹⁰The LCS problem is the problem of finding the longest common subsequence in a set of sequences. For example, suppose $X = (3, 2, 5, 7, 6, 11, 8)$ and $Y = (2, 5, 9, 11, 10, 9, 8)$ are ordered sequences, then their *LCS* is $(2, 5, 11, 8)$.

¹¹DTW is a technique in which an optimal alignment between two time-series is sought by allowing flexibility in the time domain, subject to certain constraints [151]. This approach allows for time-series that are qualitatively similar, but quantitatively dissimilar, to be aligned [125]. An example of this would be a two trajectories that visit the same points in space, but at different points in time.

¹²This is not necessarily the case for DTW as it is possible to define a step pattern which allows points to remain unmatched in the alignment [49].

¹³Although there are infinitely many translations, Vlachos *et al.* [193] provided a heuristic technique for determining a finite set of translations.

were not considered.

Bashir *et al.* [4, 5] proposed an approach to trajectory classification in which motion trajectories are modelled using multi-modal *Gaussian mixture models* (GMMs) and *hidden Markov models* (HMMs). Trajectories were segmented at points of significant change in curvature and these subtrajectories were used as basic elements in the modelling process. *Principal component analysis* (PCA) was performed on the subtrajectories of each class (the trajectories are presumed to be ‘labelled’) and the resulting coefficients were used to train the model. Each subtrajectory was modelled using a GMM whose parameters were estimated using the *expectation maximisation* (EM) method. A new subtrajectory was categorised as similar to a particular learned GMM by means of the maximum likelihood principle.

GMMs are suitable for static scenarios, but trajectories are sequential in nature. Bashir *et al.* [4] remedied this by using HMMs to model the transitions between the subtrajectories. A class of trajectories was represented by a single learned HMM with GMM states (representing the subtrajectories) and a trajectory was deemed a member of the class, represented by the HMM that resulted in the largest log-likelihood, when evaluating the new trajectory.

The authors tested their model on two data sets, namely the ASL data set and a data set of sport videos. They trained their model on half of the data and tested it on the remaining half, reporting a classification accuracy of approximately 90%.

Johnson and Hogg [76] used a competitive, unsupervised learning scheme to describe typical behaviour within a video surveillance setting. This low-level approach was applied to the problem of pedestrian movement. Pedestrians were tracked in a scene and their (x, y) -positions, and the rates of change thereof, were used as features. A moving window was used to smooth the feature vectors so as to reduce the noise inherent in the tracking method. Euclidean distance was used as the similarity measure between flow vectors (the authors scaled the velocity and positional components so as to ensure proportional contributions to the norm).

Johnson and Hogg decided against discretising the feature space and estimating an overall density function (using a frequentest approach) on a cell-by-cell basis. They argued that such a modelling approach is not concise and that for semantic information to be attached to the distribution, one would be required to attach meaning to each cell. *Vector quantisation*¹⁴ was used instead and was implemented in a competitive learning neural network framework. The network input layer consisted of four nodes, one for each component of a feature vector, whilst the output layer consisted of k output nodes. These output nodes are the prototype vectors which serve a similar function to centroids in a k -means clustering method. The number of prototype vectors were chosen and initialised beforehand. The prototype vector most similar to the input vector is the only neuron activated (and thus updated) in a competitive learning framework and the accuracy of the model increases when more prototype vectors are used. The point density of these prototype vectors was then used to approximate the probability density function of the feature vectors (this density function is a point density function which describes the density/dispersal of vectors in the feature space).

Feature vectors were obtained at a fixed time-interval by tracking movement through successive video frames. In order to avoid situations where a slow moving object is over-represented in the feature vector probability density function, resampling was required (the same is true for fast moving objects — at a fixed sample rate the flow vectors would be sparsely distributed in space, resulting in an under-representation in the resulting probability density function). Trajectories were resampled at a constant arc length and input to the network as a sequence of vectors.

¹⁴Vector quantisation is a quantization technique typically used in competitive learning settings and lossy compression schemes.

This approach models the instantaneous aspects of the motion learned from the sequences of flow vectors. The probability density function of the vectors in the feature space was estimated by considering the volume in feature space for which a particular node is the representative (the ‘winning’ node). This made it possible to statistically measure normality of individual feature vectors. By attributing meaning to these nodes it was possible to build semantic information into the model.

Johnson and Hogg [76] modelled the sequential nature of the motion by introducing an additional layer of nodes, referred to as *leaky neurons*, to their trained network. These leaky neurons add a memory mechanism to the network by retaining a short history of their activations. The next activation of such a node depended on the previous activation, subject to a rate of decay. A sequence of flow vectors was presented to a trained network and the outputs of this network served as inputs to the leaky neuron layer. The resulting trace of activations in this layer served as a representation of the sequence of flow vectors that was of a fixed size. In this manner partial trajectories were encoded and presented to a second competitive neural network which clustered these partial trajectories.

Experimental results were limited to the clustering trajectories and no explicit validation was provided for detecting anomalous behaviours. The network configuration for clustering instantaneous data comprised four input nodes and one thousand output nodes, whilst that of the partial trajectories featured one thousand leaky neurons. The second competitive learning network consisted of a thousand input nodes and five hundred output nodes.

2.3.2 Ecological modelling

The interaction of an individual with its environment is often investigated through direct observation or by radio tracking (such as GPS tags or harmonic radar tracking). The collected telemetry data produces a wealth of movement data that may be used in understanding animal behaviour as well as the dynamics of the greater ecological system. An *individual-based* approach to modelling such systems has been gaining momentum in ecological modelling [50].

An example of this may be seen in the analysis of caribou behaviour by Franke *et al.* [44]. The authors used a *hidden Markov model* to model the spatio-temporal behaviour of the caribou. They favoured HMMs over traditional time-series approaches as the former derives the optimal state sequence from observed data (as opposed to simply predicting observations [44]).

The features chosen were *distance-between-location* and *turning-angle*. These features served as the observations and input to the HMM. Three hidden states (unobserved variables) were chosen and were deemed to correspond to three behaviours, namely bedding, feeding and relocating¹⁵. Data were obtained from the GPS collars of four individuals over a ten-day period and at a temporal sample rate of fifteen minutes. When a *GPS-fix*¹⁶ was not available (an individual might enter a cave, for instance), the *distance-between-location* (DBL) feature was standardised by dividing by the number of elapsed intervals before the next position update. The DBL features were clustered into four discrete observations: stationary ($DBL < 20m$), short distance moves ($20m < DBL < 100m$), medium distance moves ($100m < DBL < 250m$) and long distance moves ($250m < DBL$). Similarly, the turning angle feature space was partitioned into four clusters representing ahead, right, left and back.

An HMM was then constructed for each of the four caribou where DBLs and turning angles served as the observations (the parameters of each HMM were determined during the learning

¹⁵Franke *et al.* [44] state that large herbivores typically spend 90% of their time sleeping, eating or relocating.

¹⁶A GPS-fix is a GPS coordinate obtained at a specific time.

phase by application of the Baum-Welch algorithm [7]). The four caribou were all cows resident in the same area. Their motion was tracked over a period of ten days and these data were used in training the HMM. This modelling approach was compared to auto-regressive time-series models constructed for each of the caribou, but it was found that the performance of the HMM was superior. The models were evaluated using a *percentage correct* and *absolute average distance* measure. The HMMs were used to predict the next observation (using Monte Carlo sampling) and the discrepancies between these values and the original features were used in determining the values of the performance measures (the same principle was applied in evaluating the auto-regressive models).

The Markov structure of the movement models allowed analysis of the behavioural states corresponding to the hidden states. The authors were interested in the durations typically spent in these states, as well as the transitions between states. The transition probabilities allowed for the computation of such probabilities as that of relocating, given that the caribou is currently relocating in a particular time step.

Although the authors investigated spatio-temporal behaviours, the learned HMMs were restricted to the region implicit in the observed movement data and so any analysis was limited to that region. Furthermore, there are many factors that influence the movement of the caribou (such as the availability of food, a mother with a calf, and human disturbances [44]) and so a greater observation period and larger amount of data would be instructive in evaluating the efficacy of this approach.

In a similar study conducted by Guo *et al.* [52], HMMs were used in developing a cattle movement and behaviour model. Collars were fitted to six cattle and their GPS positions were recorded every 10 seconds over a period of four days. The cattle were confined to a seven-hectare paddock for the duration of the data collection. The authors used the data collected during the first two days to train their model, while the remaining data were used for testing. Utilising the GPS data and video and human observation, the authors identified *stay* regions (such as water holes and shade [52]) in which cattle typically remained for longer periods of time, and *travel* regions in between the stay regions. Hidden Markov models (for each individual) were used to model the movement of the individual in the stay regions, whilst a transit model based on an agglomerative clustering was used to model the movement through the travel regions.

The authors also chose foraging, bedding and relocating as their likely hidden states. Linear and angular speed were selected as the features which served as the observations for the HMMs. Intuitively, the suitability of these features were explained by considering that an individual in transit would typically feature little variation in the angular velocity, whilst a foraging animal would exhibit a greater variation therein [52]. The feature space was partitioned into three observations, namely:

- slow linear and angular speeds,
- fast linear and slow angular speeds, and
- slow linear and fast angular speeds.

Once the HMMs had been trained, Guo *et al.* [52] calculated the probability of obtaining a particular observation, given the current state (foraging, bedding or relocating) in order to identify which hidden state corresponds to their chosen behaviours. The transitions between the HMMs and their transit models were managed by two-dimensional Gaussian distributions which were inferred from the spatial locations occupied in a stay region by a particular animal. If an individual's next position within a stay region lay beyond three standard deviations from

its corresponding spatial Gaussian distribution, then the animal would be considered in transit and would be allocated a random velocity [52].

Using this model, the authors simulated the cattle movement and compared the spatial location of the stay and transit areas, as determined by the model, to that of the original data. They deemed the results to be similar by visual inspection. The transition probabilities internal to the HMMs were used in the analysis of the animal behaviour in a similar fashion to that in [44]. The differences between the approach adopted by Guo *et al.* [52] and those in [5, 44] are worth noting. The latter authors favoured HMMs in describing transitions between homogeneous regions or primitive elements of trajectories. In this case, most behavioural patterns occurred in the stay regions which necessitated a different modelling approach.

Patterson *et al.* [130] reviewed a more generalised stochastic modelling approach, namely that of *state-space* models¹⁷ (SSMs). HMMs are a special case of this approach (the hidden states are discrete random variables), along with linear dynamical systems such as the Kalman filter and particle filters. As highlighted by the authors, a significant advantage of using this modelling technique is that the error in the observations are modelled explicitly — thus avoiding the need to clean the data. This property is particularly useful as error-correcting methods are often *ad-hoc* and may result in a loss of structure [130].

2.3.3 Maritime motion patterns and threat assessment

An analysis of the movement patterns of recreational boat types was conducted by Pelot and Wu [131], where boats were classified as being either canoes, kayaks, motorboats or sailboats. Membership of these classes was determined through the analysis of vessel motion characteristics as embodied in their GPS recorded trajectories. Initial features derived from the GPS data included:

- mean speed (average speed as calculated over all segments of the trajectory),
- $\max_{1/20}$ (5% of the fastest segment speeds, or the three fastest in cases of too few segments),
- mean turning angle,
- total distance travelled,
- aspect ratio,
- coverage index, and
- furthest distance from the shore.

After investigating correlation between the features (discriminant analysis requires that features are not highly correlated), the total distance travelled, mean speed, mean turning angle and distance from the shore were retained. The authors had very little training data at their disposal (47 trajectories for sailboats, 17 for canoes, 21 for kayaks and 10 for motorboats) and so chose to derive the parameters of their discriminant model using all of the training data. This approach is questionable as it is expected to produce favourable results when applying the classifier to this same set of data in order to evaluate its performance. Their very successful results reflect

¹⁷A *state-space* model is a time series modelling technique which predicts future states of the system from previous states by using a process model. The current system state is then updated and refined using observations [130].

this fact. However, they did find that differentiating between canoes and kayaks is most fraught with error, as one would intuitively expect.

Maritime safety and security monitoring is typically focused on large vessels and their behaviours. Surveillance of vessel activities is of importance in identifying threats such as pollution, piracy, smuggling and terrorism [84], and studies are also being conducted on a ‘worldwide maritime surveillance’ capability [188]. Although many different types of sensors have been used to observe vessel activity, the recent *automatic identification system* (AIS) has begun to contribute large volumes of vessel telemetry data (vessels with a gross tonnage of at least 300 tonnes are required to use AIS) to the domain.

AIS systems are self-reporting and their update rates are a function of the vessel’s speed. These systems are relied upon for collision avoidance and safe navigation [142]. However, the data are unreliable as broadcasts may be fabricated by a vessel (*i.e.* false positions may be transmitted intentionally) or self-reporting may simply be deactivated during illegal operations [142].

Lane *et al.* [84] undertook preliminary research into identifying five specific anomalous behaviours discernible from AIS transmissions. This fine-grained approach to modelling specific behaviours allows for component models to be built which address particular behaviours directly whilst a fusion approach is used to generate aggregate threat values. The five anomalous behaviours considered were *deviation from a standard route*, *unexpected AIS activity*, *unexpected port arrival*, *close approach* and *zone entry*.

The assumption was made that commercial vessels typically follow the most economical route between two points (barring movement constraints such as land, shallows, traffic separation schemes and EEZs). These shortest routes were deemed to be standard routes and any deviations from these routes were considered anomalous.

Unexpected AIS activity includes a vessel reporting a position from which receivers are typically unable to receive signals, or a vessel not transmitting from an area where receivers are known to have good coverage [84]. This behaviour was modelled by placing a grid over the area of interest and calculating the probability of receiving a signal from a particular cell. The authors deemed the probability of receiving such a signal to be inversely proportional to the distance from the receiver. The authors chose a beta distribution to describe this probability. The probability of receiving a signal was estimated as a ratio of received detections to ideal detections, where ideal detections include signals that should have been detected but were not (which is estimated using interpolation). The parameters of the beta distribution were chosen to best describe these estimates (all estimates were performed on the AIS data available to the authors). Bayes’ theorem was used to calculate the probability of detecting a signal, given the number of detections within a cell, for a particular vessel. The theorem proved especially useful in this context as the prior dealt with situations where historical data contained cells with very little traffic.

Unexpected port arrivals were identified using a mathematical modelling approach as well as a database look-up process. In the latter case, if a ship arrived at a port that did not have the facilities to accommodate that vessel, then the vessel was considered anomalous [84]. A Markov model was used to model the successive ports visited by a vessel and the transition probabilities were estimated from training data.

Close approaches are an example of coupled behaviour (as defined in [31]) which may arise from an ‘illegal transfer of goods’ between two vessels [84]. Comparisons were made whilst vessels were being tracked, in order to determine whether close proximity between two vessels warranted alarm. After excluding harbour areas, the area of interest was discretized into cells once more (in order to reduce computational complexity). The kinematic quantities of vessels occupying

the same cell or adjacent cells, were compared. The authors converted the kinematic values at the closest point of approach to a probability in order to facilitate data fusion [84].

Lastly, zone entry was modelled by specifying zones (such as environmental protection areas) using polygons. Determining whether a position lies within the polygon is trivial but predicting whether a vessel is likely to enter a zone at some future time-step requires a predictive model. The authors chose to use a Gaussian distribution (over heading and speed) with variances estimated from historical data. A distribution was placed at each vessel's recently observed position. The proportion of projected tracks that intersect the zone within a particular time window represented the probability of that particular vessel entering the zone (the authors state that particle filters are better suited in describing more complex manoeuvres) [84]. Finally, an overall threat value was derived from the five models through the application of a Bayesian network.

An unsupervised learning method for anomaly detection at sea was proposed by Laxhammar [85]. The author utilised a probabilistic approach towards modelling surveillance data of sea traffic. The argument was made that monitoring a great many sea-faring vessels is intractable when relying on operators alone (given that operators are a limited and costly resource). It was the author's intention to provide a mechanism that would assist operators through the automatic identification of anomalous behaviour (Laxhammar identified speeding, anchoring, grounding and sea-drunkenness as examples of anomalous behaviour [85]). The author stated that operators wish to identify time-varying behaviours such as smuggling and poaching, as well as scenarios which appear anomalous simply because they are novel.

The modelling approach pursued by Laxhammar [85] treated trajectories as discrete points without being concerned about temporal-causality. The chosen features were modelled using *Gaussian mixture models* and in order to ease computation the area of interest was discretized into cells, computing a GMM for each one. Laxhammar based his model on previous work in which two-dimensional Gaussian distributions had been constructed on a per-cell basis employing a single feature, namely vessel velocity. Having highlighted the shortcomings of this approach (correlations between position and speed could not be analysed), he proposed a second approach that incorporated position (a four-dimensional feature space as opposed to a two-dimensional feature space). Using this extended feature space, vessels travelling in the wrong direction in a two-way sea-lane could be detected.

Using the EM method, the mean and variance of each Gaussian component may be estimated. This method finds a local maximum, is sensitive to initialisation, and does not address the problem of the ideal number of components to use in GMM. In order to deal with these shortcomings, Laxhammar implemented a greedy EM method proposed by Verbeek *et al.* [190]. Data obtained from the Swedish naval intelligence batallion was used in training, validating and testing this approach. Once parameters of the GMMs had been determined through training and validation, a position and velocity along a vessel's trajectory was presented to the model. It was deemed anomalous if it produced a likelihood lower than a particular threshold, given the GMM for that particular cell (anomalous points were deemed unlikely to have been generated from the cell density function). Laxhammar stated that instead of allowing this threshold to be set by the operator, it would be desirable to have the model learn in an on-line fashion with the assistance of an operator who could flag particular points as anomalous.

Using many of the same underlying modelling assumptions (such as a cell-based approach and that obtained data are representative of normal behaviour) as in [85], Laxhammar *et al.* [86] compared the GMM modelling (a parametric approach) of position and velocity within cells, to that of *kernel density estimation* (KDE) (a non-parametric method). Historical AIS data obtained from vessels travelling off the west coast of Sweden was used for evaluation. The

authors further emphasised a philosophical impasse in anomaly detection, namely that one is not necessarily able to define anomalous behaviour and that if one could, it would cease to be anomalous. Although this observation may appear obvious or a matter of definition, it provides a foundation for the authors' assumption that the obtained AIS data were assumed normal during training (although these data likely contain trajectories that may be considered anomalous, Laxhammar *et al.* [86]) stated that these trajectories would yield a lower likelihood in this modelling framework, and subject to a threshold, may be determined as such. The authors state, furthermore, that seeking particular anomalous situations will bias the model towards particular behaviours in favour of other equally anomalous behaviours.

The AIS data spanned a period of three weeks and were pre-processed in order to render the data set more manageable. An AIS report was included in the post-processed data set if it was received at least 200m from the last accepted report (an example of *data thinning*). Furthermore, if a vessel remained stationary for more than five minutes, it was assumed to be moored. In the latter case, or when no further reports were received, the trajectory was considered completed. Using this resampling approach, the authors extracted 36 370 trajectories which were partitioned into training and evaluation data sets on a per-cell basis (by randomly selecting four fifths of all trajectories passing through a cell for training). A threshold of 100 observations determined whether a distribution should be estimated for a particular cell (using a prior as the authors of [84] had done, may allow all cells to be considered but this would not necessarily enhance the modelling approach). The evaluation set was augmented with simulated trajectories that were constructed by randomly selecting a point in an existing trajectory and determining future speed and course subject to uniform distributions. The probability of thereafter changing the course and speed was set to $\frac{1}{10}$ (to encourage periods of coherent movement). If these sampled values resulted in a position in a cell for which a model was not constructed (due to data scarcity), then those values were discarded and new samples determined. Time was the independent variable and the interval lengths were determined by the speed (so that displacement between consecutive positions was 200m).

Laxhammar *et al.* [86] proposed two performance evaluation metrics. First, the model that attributed the largest likelihood to previously unseen normal data was deemed to be superior, and secondly, the model with the lowest response time (in terms of number of observations processed) in identifying anomalous examples, was considered superior. The authors found that the KDE approach yielded a higher median likelihood for previously unseen normal data than the GMM. However, in the anomaly detection experiments, a significant difference between the models was not present.

In closing, the authors highlighted potential limitations of their modelling approach. They stated that partitioning the data relative to origin and destination (with particular mention made of the work by Ristic *et al.* [142]), instead of considering only geographical proximity in the cells, would address particular failings of the model (for instance, minor sea lanes may be considered anomalous by the model when they cross major sea lanes). Furthermore, a general inability to exploit the structure inherent in analysing the behaviour over time was also highlighted.

Ristic *et al.* [142] used an adaptive KDE approach in constructing a model of the likelihood of observations (given a null hypothesis which assumes the training data to be normal). Using data mining methods (not discussed in [142]), motion patterns were extracted from AIS data sets. The authors chose position, velocity and origin as the features of the model. For a mined cluster j of motion patterns, the probability $p_j(\mathbf{x}, H_0)$ is approximated through adaptive KDE (where \mathbf{x} is a vector containing the velocity and speed measured at successive observations, p_j is the estimated probability density function and H_0 denotes the null hypothesis). This approach is adaptive in that window widths in regions of lower density are able to stretch, thus allowing

for a better description of the tails of probability densities.

The authors determined a ‘detector’ parameter which was used to derive a threshold from the learned distribution. A previously unseen trajectory was deemed anomalous if it violated this threshold. Tests were performed on the anomaly detection model using one hundred simulated trajectories, as well as real AIS data. In both cases, isolated test outcomes were presented in [142].

Rustic *et al.* [142] also presented a motion prediction model for predicting the state of a vessel at some time in the future, provided that the vessel is exhibiting normal behaviour, and that it will continue to do so. This was achieved by implementing a particle filter (a non-linear filtering method).

De Vries *et al.* [23] augmented vessel positional data with geographical domain knowledge and presented similarity measures to distinguish between vessel trajectories. The similarity measures were based on alignment methods. The authors observed that the regions in which vessels move possess their own semantics and that the introduction of this semantic information to the feature space may lead to the discovery of more complex behavioural patterns [23]. Examples of such geographical domain knowledge includes anchoring areas and harbours. The authors used this framework to discover behaviours in an unsupervised scheme and illustrated the proficiency of their approach through a classification experiment.

Geographical knowledge was represented in the *resource description framework*¹⁸ (RDF) which makes use of two *ontologies*¹⁹ in describing the domain. The first contains definitions of areas of interest at sea, such as anchorages and clearways (referred to as the *anchorages and clearways* ontology by the authors), whilst the second described harbours. Each concept in these ontologies was paired with a polygonal region defining its spatial extent, a unique identifier and its type.

An augmented vessel trajectory comprised a low-level trajectory component (the (x, y) -position of the vessel under a map projection), a sequence of sets of geographical labels and trajectory end-point information. The sequence of labels was constructed by testing membership of each position along the trajectory to the defined polygonal regions. As a particular position may be contained in more than one region, the corresponding sequence element contained a set of pairs containing the geographical identifiers and their types. Trajectory end-point information for the start and end of a trajectory encoded whether or not a vessel had stopped, and if so whether it was located in a harbour or near some other defined region.

The authors defined alignment-based similarity measures for each component of the augmented vessel trajectories. By computing kernels based on each of the similarity measures, namely that of positional data, geographical label data, and start and end data, it was possible to form a convex combination of the kernels that resulted in a kernel comprising all the similarity measures.

AIS data gathered from vessels within 50 km from the port of Rotterdam served as the data set. Fifty regions of six types were identified in the anchorages and clearways ontology and ninety harbours of seven types in the harbours ontology. De Vries and van Someren [24] found in an earlier study that using compressed trajectories produced better results during clustering, and this approach was followed once more.

Spectral clustering was chosen as the preferred clustering technique and a normalised graph cut was computed using weighted kernel k -means. This computation was repeated a number

¹⁸The RDF provides a framework for representing information about web resources. In particular, it provides a vocabulary and a data model that allows reasoning over RDF expressions [95].

¹⁹An ontology, in this context, provides a formal representation of knowledge within a particular domain as a set of concepts and the relationships between them [185].

of times with random initialisations as convergence was not guaranteed. The kernel weights (as formed in the convex combination described above) allowed for consideration to be directed and particular information to be included or excluded. For instance, a unity coefficient for the trajectory kernel resulted in the positional data being considered during clustering (referred to as K_{traj} by the authors). Similarly, choosing a zero coefficient for the trajectory kernel and equal fractions (one third) for the geographical kernels and start and end kernels, resulted in ontological information being considered (K_{onto}). A third combination was considered during clustering experiments, namely K_{comb} , for which the trajectory kernel received a weighting of a half and the three remaining kernels were equally weighted at one sixth. The ideal number of clusters was not known *a priori* and had to be determined empirically.

Explicit validation of the clustering method was not performed (this is expected since a ground truth not available). However, the efficacy of the similarity measures was evaluated through the classification experiment. The authors discussed three behaviours which were discovered during clustering (interested readers may consult [23] for a thorough discussion of these behaviours). The first of these behaviours, namely *anchoring in a specific anchoring area*, resulted from the K_{comb} kernel. Considering trajectory information or ontological information only resulted in vessel trajectories not ending in the same anchoring area. The remaining behaviours were *smaller ships approaching from the sea and continuing directly inland* and *vessels approaching from open sea docking in a certain part of the harbour* [23]. The inspection and interpretation of the resulting clusters allowed for semantics to be attributed to them.

The classification task set out to identify the type of vessel that created a particular trajectory (this information is recorded in AIS data). A *support vector machine* (SVM) served as the classifier and various combinations of the convex combinations of the kernels were used. The authors found that making sole use of one of the three ontology-based kernels (giving each a weight of one in different experiments) resulted in similar average classification accuracy in the lower fifty percent range. Producing a kernel by weighing the three ontological kernels equally produced an average accuracy of 66.1%. Using only the K_{traj} kernel produced results of 72.2%, whilst K_{comb} produced the best results with an accuracy of 75.4%. It is clear from this result that the combination of low-level trajectory information and domain knowledge yield the best results in this model framework [24].

Van Laere *et al.* [188] conducted workshops aimed at capturing domain knowledge from maritime surveillance experts. Although the conference proceedings in which [188] appeared were chiefly concerned with the processes involved in effectively capturing this knowledge, they also presented some of requirements and anomalies identified (the latter is of particular concern in this context). The authors conducted two workshops. The first was a ‘field study’ during which operators were presented with vessel observations (the results of this study were presented and published at different conference, but were still discussed in [188]). The operators would then provide motivations for taking particular actions. The second study was a collaborative anomaly detection workshop assisted by computer software that allowed brainstorming, voting and categorising of ideas [188]. A third workshop (conducted in Canada) was also discussed in [188] for comparative purposes.

The authors briefly highlighted the differences between *data-driven* and *knowledge-driven* approaches with respect to anomaly detection. The stated advantage of the former approach is that the models are able to recognise previously unobserved or unimagined behaviours whilst the extensive training period required is a disadvantage thereof. Through the utilisation of expert knowledge, the latter allows for a clear identification of behaviours in a robust manner. The disadvantage of this approach is that the model requires frequent updating in order to incorporate behaviours that transition from anomalous to normal (or *vice versa*) [188].

The authors referred to a few anomalies, identified during the field study, which may serve as rules in an anomaly detection system (an operator is notified in such an event). These included vessels that exceed some maximum speed, vessels that abandon a particular speed, or vessels that are stationary for a period of time. If a particular vessel enters a specific area, or if a vessel deviates from a planned route, then it was agreed that it should be flagged for operator attention. Lastly, if two vessels meet at sea or if they approach one another only to break off thereafter, then operators participating in the study wished to be notified [188].

The second study framed anomalous behaviour in operator parlance as *early warnings* [188]. Of the seventy five early warnings identified by participants, a subset of thirty one were chosen that were deemed to be of a greater significance (or simply more interesting). This subset was further partitioned into the representative categories *tampering*, *owner/crew*, *historical data*, *rendezvous*, *movement* and *cargo* (see Figure 2.5). The *tampering* category contains behaviours in which a vessel intentionally hides its present activities. The *rendezvous* category describes vessels which interact with other vessels or other objects, and the *movement* category encompasses movement-related behaviour.

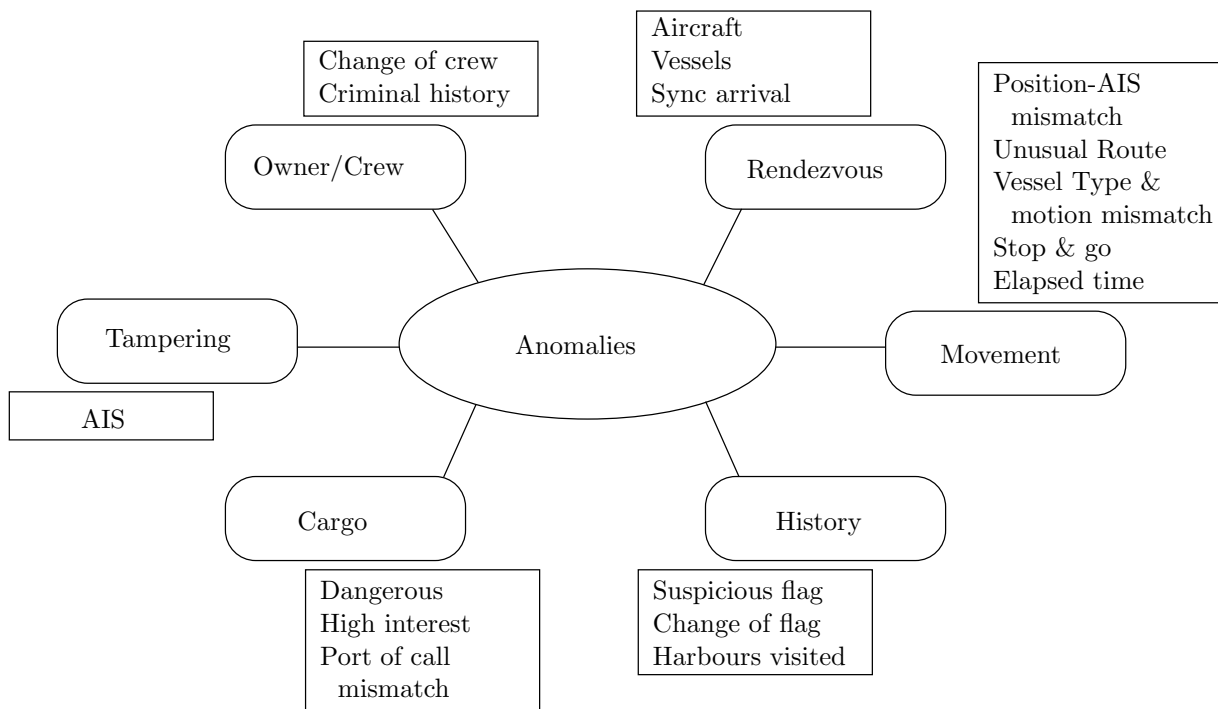


FIGURE 2.5: The events identified by operators as events of import for which they would like to receive early warnings.

Lastly, the workshop conducted in Canada (as discussed in [188]) pursued a taxonomy of maritime anomalies which were subdivided into dynamic and static anomalies, which in turn comprised kinematic or non-kinematic categories. Examples of variables used to identify kinematic anomalies included course, speed, manoeuvring, reporting and location. Whilst non-kinematic indicators included next and last port of call, cargo list, ship signature, crew lists and passengers.

2.4 Summary

The literature concerning the matching, identification or classification of motion patterns was discussed in this chapter. The differences between expert-driven and data-driven approaches were discussed in §2.2 as these methodologies are often employed in creating classification systems. Motion patterns are increasingly utilised in various domains and their classification and analysis is of importance in computer vision (§2.3.1), ecological modelling (§3.2.2) and maritime threat detection and assessment (§2.3.3). Publications in these various disciplines were discussed where particular attention was paid to applications that were primarily concerned with kinematic data. Support vector machines, neural networks, kernel density estimation, Gaussian mixture models and hidden Markov models were the most often applied techniques and most authors used labelled data sets in the construction of their classifiers. The vast majority of approaches made use of feature selection approaches in which relatively simple features were often chosen (*e.g.* speed, flow vectors, turning angle), whilst principal component analysis was successfully used in feature extraction. Instead of direct application of Euclidean distance in measuring similarity, alignment methods were found to yield promising results.

Although Dodge *et al.* [31] sought to pursue a general framework in which to identify motion patterns, a persistent problem in classification is that a classification task is often highly dependent on the data under consideration, as well as on the chosen features, similarity measure and modelling approach. In the general case, there is no model that would outperform all other models for every problem. There is invariably a trade-off in model selection where a perfect classifier is inevitably unattainable.

CHAPTER 3

A surveillance system framework in the maritime domain

Contents

3.1	Decision support systems and data fusion	35
3.2	Maritime surveillance	37
3.2.1	<i>The data</i>	39
3.2.2	<i>Extracting meaning from the surveillance scene</i>	40
3.2.3	<i>The operator</i>	40
3.2.4	<i>Vessel designations</i>	41
3.3	Existing maritime surveillance DSSs	43
3.4	A newly proposed maritime surveillance DSS	44
3.4.1	<i>The fusion component</i>	47
3.4.2	<i>The rule-based component</i>	48
3.4.3	<i>The discovery component</i>	51
3.4.4	<i>The activity classifier</i>	53
3.4.5	<i>The human machine interface</i>	53
3.5	Summary	53

Decision support systems and data fusion play an integral role in maritime surveillance systems (this is also true of other applications, such as air-defence systems). These concepts are considered in this chapter along with the constituent elements of a maritime surveillance system. Existing maritime surveillance systems are reviewed and a novel adaptive decision support system is proposed. Each of the constituent elements of this system is discussed in some detail, after which the full system architecture is finally presented in its entirety.

3.1 Decision support systems and data fusion

Decision support systems (DSSs) are information systems responsible for assisting users or operators in complex decision making tasks [157]. This goal is typically achieved by identifying possible courses of action and providing recommendations [42]. The central components of a DSS were identified in an early paper by Sprague [164] as the data subsystem, model subsystem and the user interface. Although improved solution methods have led to faster and more

responsive model subsystems within existing DSS implementations [157], the introduction of newer tools and technologies have also driven innovation in DSSs and have resulted in a large variety of support systems. In knowledge-based DSSs, for instance, artificial intelligence and expert systems are used to provide reasoning and intelligence to the decision maker [157]. Another class of DSSs, namely intelligent DSSs, are characterized by their ability to identify solution approaches, apply the appropriate decision model and interpret the solutions in order to learn from them [63].

The concept of an *adaptive DSS* (ADSS) was introduced in [63]. The formulation of an ADSS differs from that of its predecessors in that it employs unsupervised inductive learning [42, 63]. This allows an ADSS to adapt its knowledge state and generate the necessary capabilities without requiring their *a priori* definition. This approach lessens the reliance on domain experts and knowledge engineers who are instrumental in the construction of a knowledge-based DSS [63]. An ADSS, as well as many other kinds of DSSs, may also be designed to support the user more directly by monitoring user performance and usage history [42].

Implicit to the definition of a DSS as a support tool is that decisions or actions are not automatically executed by the system and that human intervention is required for actions to be taken¹. This is particularly important within military applications where a *human-in-the-loop* doctrine is typically followed in decision making processes. Furthermore, a DSS is not simply an information system which reports information, but is one that necessarily makes recommendations.

Effective decision making often depends on the availability of timely and accurate information. Access to information may empower a decision maker to take the ‘best’ course of action from a set of possible actions (this choice is contingent on beliefs, desires and goals). A decision maker invariably relies on multiple information sources which may not report particularly accurate information. The concept of *data fusion* plays a vital role in this regard. Data fusion is concerned with the combination of data from various sources, as well as the prediction or estimation of entity states [167]. In fact, Steinberg [167] states that all biological cognitive activities and automated approaches to information usage rely on data fusion. A conceptualised view of the processes involved are embodied in a framework presented by the *Joint Directors of Laboratories* (JDL) in [169]. This framework has since undergone a number of revisions and expansions which are widely used in categorizing data fusion-related functions. One of these model frameworks, called the revised JDL data fusion model, is presented in Figure 3.1 (the interested reader may consult [169] for a thorough discussion of the functions embodied within each component). Although dependent on the application, DSSs usually subsume the framework of Figure 3.1 or implement only the necessary functions of Levels 2, 3 and 4. A recent revision by Llinas *et al.* [90] is of special interest as it explores ontologies and adaptive system implementations of data fusion.

As mentioned in the discussion on ADSSs, incorporating an inductive learning process in the system design enables the system to identify new concepts without the intervention of a domain expert at various stages throughout the model improvement or adaption process. This general approach has been discussed within the data fusion community where conventional data fusion was considered to entail a deductive process of target recognition [90, 196]. The possibility of including a discovery process capable of learning unknown signatures of known targets or unknown hidden targets is discussed in [90] and a generalized architecture of such a system, receiving data from three sensors, is presented in Figure 3.2.

It may be seen that the operational data stores that are populated by the sensors in Figure 3.2, still form the basis for the conventional data fusion process which derives information about the

¹Systems that do not require human intervention are known as decision automation systems [133].

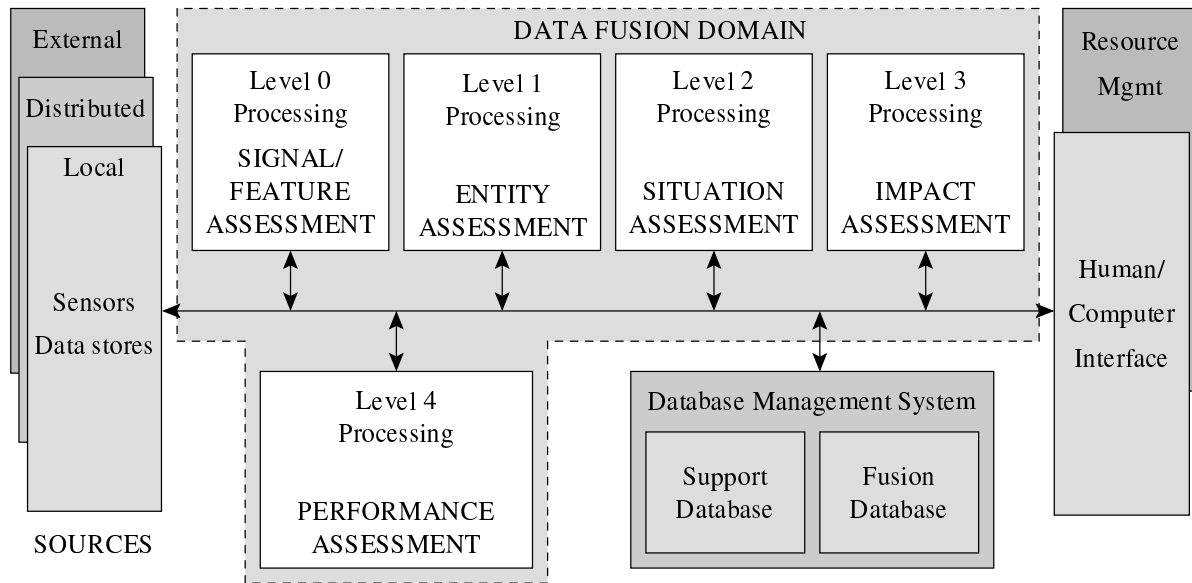


FIGURE 3.1: The JDL fusion model [168].

objects themselves (entities and their identities, as well as their tracks). Pertinent information is extracted from the operational data stores, extracted and loaded into a *data warehouse*² (DW). These data are further cleaned and transformed in order to populate an entity relationship database. This database serves as a data source for unsupervised learning approaches towards model testing. Clustering methods are used to discover structure within the data which, if yielding favourable results during validation and testing, are incorporated into the fusion process.

Llinas *et al.* [90] identified a number of challenges related to this automatic discovery component. The first of these is concerned with the capabilities of existing methods to identify patterns in the real-time data and whether these patterns extend reliably beyond local or transient phenomena which may not warrant attention. Most importantly, what qualifies as a *valid* pattern is notoriously difficult to ascertain in the open-ended clustering that takes place.

3.2 Maritime surveillance

The support systems within surveillance environments are usually tasked with finding interesting activities or entities within a mass of uninteresting data³ [15]. The work load of surveillance operators may be reduced by directing their attention to that which is important, thereby diminishing the likelihood of *information overload* [15]. Such assistance contributes to enhanced situation awareness and improved performance which would otherwise be threatened by exhausted cognitive capacity, boredom and fatigue [15, 115].

Upon review of surveillance systems in the video and maritime domain, the similarities across systems is easily identified and abstracted. In attempting to construct a generalized framework for surveillance systems in the literature, one may arrive at Figure 3.3. A distinction is made in

²A data warehouse is a database formalisation in which data from operational databases are integrated in such a way so as to provide a repository of historical and current data [68].

³In the video surveillance domain, video streams were initially relayed directly to operators. However, advances in technology resulted in a steady increase in collected video data which, in turn, led to multiple streams being displayed on a single screen [106]. Nevertheless, operator performance was still reported to be below satisfactory levels. Automation (such as motion detection) has been pursued as a result [201].

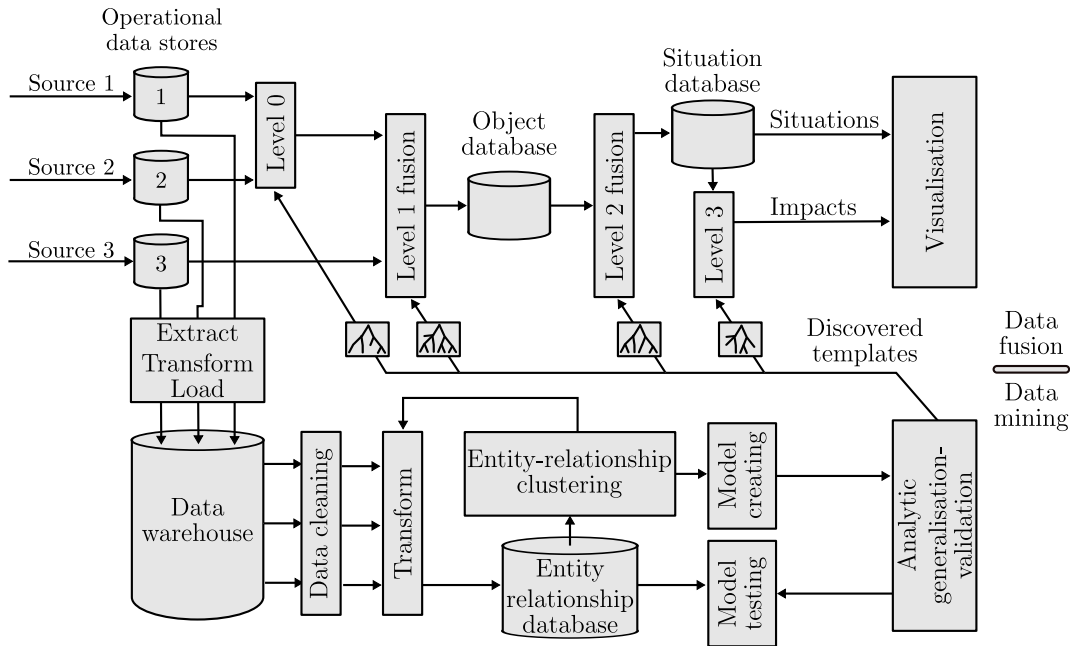


FIGURE 3.2: The functional processes of an integrated data mining and fusion approach [90].

the figure between generic system components and the cognitive processes that correspond to them. Each of the generic concepts may subsume various system or cognitive components. For instance, data acquisition and observation may involve multiple sensors whereas data management may encompass data pre-processing as well as data storage. The visualization component, or *human machine interface*, may involve analysis or interaction from the operator and may present multiple ‘views’ on numerous terminals (screens). Although rudimentary, the exposition in Figure 3.3 summarizes the broader concepts of surveillance systems.

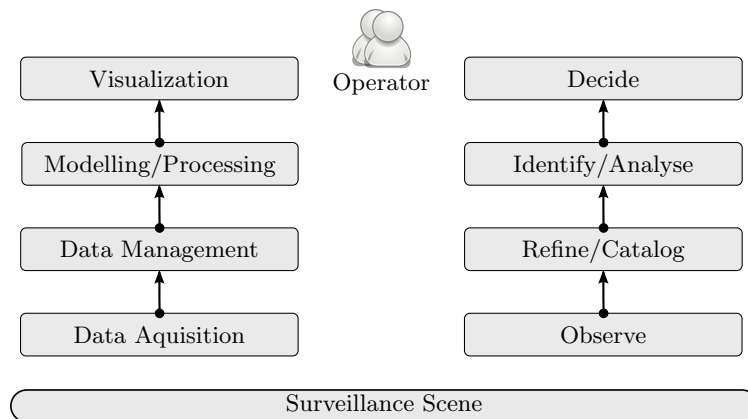


FIGURE 3.3: A simple generalized flow diagram of the surveillance process.

Within the maritime domain, the goal of a surveillance DSS is often to assist an operator in identifying vessels that are behaving strangely with respect to previously observed behaviour or with respect to expectations [15]. In order for a system to provide such support it is necessary to imbue it with a means to understand or interpret the surveillance scene (or region of interest) [115]. This requires of the system to attach certain semantics to particular events. However, the events are context-specific as behaviour that is normal for a particular type of vessel may not be normal for another. Not only do vessel activities vary between types of vessels, but they may

also vary between vessels of similar type. For instance, deep-sea hake trawlers ply their trade in a different manner from long-line tuna vessels. It is thus required that a surveillance system is capable of interpreting context-specific activities.

Guerriero *et al.* [51] identified three components they considered essential to a maritime surveillance system. These are multi-sensor signal and information processing abilities, multi-sensor fusion, tracking and anomaly detection algorithms to aid in the reduction of tracks and the identification of tracks of interest. Furthermore, the ability to learn continually from operator feedback was identified by Rhodes *et al.* [141] as a means to avoid defining specific rules to deal with all possible activities and their respective contexts. The situation awareness model of Endsley [36] is of particular interest in this domain as a maritime surveillance system ultimately contributes to improved situation awareness. This model has much in common with the cognitive component of Figure 3.3. Furthermore, the cognitive process model embodied in the celebrated OODA (*observe, orient, decide and act*) loop [8, 132] is also relevant. In particular, the assistance offered by support systems typically shortens the time it takes to act, thus shortening the decision cycle.

At the very minimum, a maritime surveillance system should be capable of processing a number of inputs that provide some notion of the surveillance scene, and produce outputs (such as alarms or vessel designations). Before a novel maritime surveillance DSS is suggested, the notions presented in Figure 3.3 are elaborated upon in order to better understand the requirements of such a system. Particular emphasis is placed on the inputs and outputs of such systems.

3.2.1 The data

Multiple data sources are typically at the disposal of a maritime surveillance system. These include data from sensors such as AIS data, video data, infrared data and coastal radar data [51]. Besides these more prevalent technologies, *synthetic aperture radar* are playing an increasingly significant role in operational surveillance systems [75]. These radar are capable of detecting oil spills and vessels in inclement weather conditions through cloud cover [75, 83].

Additional data sources include vessel watch lists, wanted vessel lists and blacklisted ship lists. These lists are typically managed by coastal states but there are efforts underway to coordinate these data. For example, the European Maritime Safety Agency, which enforces EU legislation towards reducing the risk of maritime accidents, marine pollution from vessels and the loss of human life at sea [38], publishes a list of vessels banned from EU ports [39]. These vessels are denied access to ports and anchorages for a specified period of time if they are found to be in contravention of international or EU regulations⁴.

Transmissions emanating from the Global Maritime Distress Safety System⁵ may also serve as a data source within a maritime surveillance system. Cargo manifests, crew lists and last-ports-of-call may also contain information that would be useful in achieving an accurate designation

⁴Port State Control Officers inspect vessels arriving at the ports of any of the member states of the Paris MoU (there are 27 member states, which include Canada and Russia) [174]. Vessels that are found to have deficiencies on board may, at the very least, be instructed to repair them within a certain time, or in serious cases, be detained at their port of call. Vessels which skip detention, or are guilty of multiple detentions within a specified period, will typically be banned ([174] may be consulted for the guidelines in these matters).

⁵All passenger and cargo vessels on international voyages, of a gross tonnage of 300 tonnes or greater, are required under the regulations of the International Convention for the Safety of Life at Sea (SOLAS) [70], to carry certain terrestrial and satellite radiocommunications equipment for the sending or receiving of maritime safety information and distress alerts. The Global Maritime Distress and Safety System provides an integrated communication system which utilises these radiocommunication technologies so as to provide assistance to ships in distress [138].

in an maritime surveillance system [104, 144].

Data fusion techniques may then be used to provide a consolidated surveillance picture [51]. In the case of kinematic data sources, such as radar tracks and AIS tracks, a single fused track is thus produced for processing (this activity is performed within the *refine* cognitive component in Figure 3.3 and in *Level 1* of Figure 3.1).

3.2.2 Extracting meaning from the surveillance scene

Once sensor data have been acquired, fused and cleaned, it is presumed that these data provide a good approximation of what is actually occurring within the surveillance scene. Vessel motion is inadvertently affected by elements within the scene. Not only do the observed trajectories contain information that may be useful in activity analysis, but the geographical constraints to motion include sea-lanes, traffic separation schemes⁶ and areas to be avoided due to dangers posed to safe navigation [159]. Furthermore, certain navigational constraints are relevant to specific vessel classes, such as laden tankers which are required to maintain a distance of 20 miles from various points, such as Cape Agulhas and Cape Columbine [121]. Understanding the geographical factors which influence motion may therefore serve to enrich the raw vessel trajectories and provide context for the observed motion. The act of discovering regions of interest within a scene from observed data is referred to as *scene understanding* in the computer vision community [93, 115]. However, certain activities, such as man-overboard manoeuvres, will occur at sea independently of position. Investigating trajectories in a position invariant manner is therefore also deemed valuable.

Geographical data may also provide a natural segmentation mechanism for trajectories, which is beneficial as matching a number of smaller subtrajectories rather than single large trajectories is desirable⁷. This is due to the fact that as the length of a trajectory becomes significantly large, the minor variations in motion may be rendered insignificant in comparison (larger trajectory segment comparisons will share more commonality). The segmentation of trajectories into stop and move segments is an approach often used to achieve trajectory segmentation [93, 115]. Furthermore, trajectory segmentation into subtrajectories of constant motion has also been applied [187].

Data analysis may then be carried out on geographically enriched trajectories or on position invariant trajectory attributes so as to identify an activity (recall the discussion in Chapter 2 on the various methods that have been employed in this regard).

3.2.3 The operator

The role of the operator within the maritime surveillance environment is to maintain MDA through monitoring and analyzing activities within an area of interest [15]. This may be achieved by fusing and summarizing information from various data sources on an overview display [122]. False alarms should be avoided in systems in which inferences are made from this information as they can distract or misdirect the operator, causing scepticism with respect to the system [115]. This is true in numerous surveillance settings, computer intruder technologies included [54].

⁶Opposing streams of traffic are often separated into one-way lanes [159].

⁷A ferry carrying tourists to and from Robin Island typically departs and arrives at the same location. Segmenting the trajectory into subtrajectories at these origin-destination pairs is thought to enrich the trajectories (this was found to be the case in [142]).

The presentation of the information to the operator is also of paramount importance as good design strategies may improve human performance. Operators can typically select what they wish to view and systems usually take an interactive approach in which operators may perform further analysis on entities that generated certain warnings. For instance, vessel designations (as members of particular classes of interest) may be changed in [15] by operators whilst the system proposed in [144] prompts operators who may then perform further analysis on particular warnings. *Vessel traffic services* (VTSs) make extensive use of operators who monitor AIS and radar data which are overlaid on electronic charts [192]. VTSs are usually provided in port approaches, in areas of environmental sensitivity or areas which present navigational difficulties. Standard symbols for AIS target information have been defined by the IMO [69] and have been used in many of these systems [16]. VTSs operators communicate directly with vessels to coordinate their entry or exit from a harbour; they coordinate vessel movements in emerging conflict situations and they provide information to mariners upon request [200].

3.2.4 Vessel designations

A central focus in the maritime domain is *anomaly detection* [84, 85, 86, 142]. However, far fewer research articles address the concept of a threat at sea or make an attempt to quantify them ([84, 147] consider this problem). Vessel designations are discussed in the following subsections with particular focus on the designations suggested by Roy *et al.* [144]. Although these designations are categorical classes they nevertheless present an operator with richer information as opposed to a simple binary choice between anomalous or not. The surveillance outcome set (or vessel designation set) was divided into the four categories of normal, anomaly, threat and *vessel of interest* (VOI) by Roy [144]. As may be seen in Figure 3.4, intersections between the different designations are possible. Additional refinements to the designations may be made by including *hostile* or *suspect* designations [15].

3.2.4.1 The notion of normality

The concept of what is considered normal is informed by context. This is true in the definition of normality itself, whether it be statistical or normative. For example, societal norms may require adherence to a concept that is not statistically normal in the sense that the norms are concepts which need not constitute average behaviour. Similarly, what is statistically normal is population-dependent and even this may change with time. However, this dichotomy need not be conclusive as the argument has been made that normality as a biological concept is distinct from both these categories [195]. Nevertheless, this only serves to reaffirm that context is central to discussions of normality.

Within the maritime domain numerous factors external to the vessels themselves, such as geographical constraints to movement (see §3.2.2), weather conditions and tidal status, influence behaviour at sea [141]. What is considered normal behaviour for a particular class of vessel may be considered anomalous for another. For example, prevailing commercial vessel activity is expected to differ greatly from the activity of recreational vessels which may navigate closer to shore. Within the class of commercial vessels the patterns that vessels follow also differ. For example, the routes followed by passenger liners often differ from those of laden tankers as the former may engage in sightseeing activities that cause it to deviate from standard shipping routes [84].

The context of a behaviour at sea is thus informed by factors external to vessels, by the vessels themselves (different vessel types exhibit different movement capabilities) and by the activities

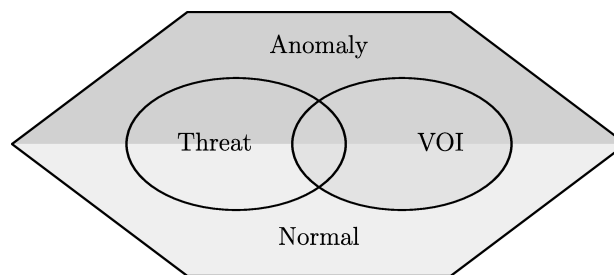


FIGURE 3.4: The division of vessel designations within a maritime surveillance environment (reproduced from [144]).

that a vessel may be engaged in (*e.g.* fishing, passenger-liner, recreational). On a finer scale it would be true that vessel behaviour depends on captain and crew as well, but this is deemed to be negligible in comparison with the aforementioned factors when considering clustering or classifications of vessels or activities.

In intrusion detection systems (in the context of computer networks [6]) and video and maritime surveillance systems, detection of abnormal behaviour is important in filtering the interesting from the uninteresting. Identifying behaviour as abnormal is sufficient in some cases to determine that a threat is present. This is particularly true in intruder detection systems as abnormal network activity or abnormal user activities would imply that an intruder has infiltrated the system. Within video surveillance environments, a direct association is often made between anomalous and dangerous behaviour (such as luggage being left unattended in airports or a pedestrian visiting a number of stationary vehicles instead of walking directly to one vehicle [127]). The same is true to a large extent within the maritime domain where it is assumed that anomalous behaviour is a strong indicator of potentially illicit behaviour [84].

3.2.4.2 Vessels of interest

Identifying particular vessels as *vessels of interest* (VOIs) or high interest vessels was identified as a natural and often used distinction at a workshop in consultation with domain experts [144]. A vessel carrying hazardous cargo would qualify as a VOI which would indicate to an operator that this vessel warrants closer attention. The US coastguard considers any vessels that may pose a significant security risk to ports or locations as VOIs. In a report released in 2009 it is stated that vessels are targeted for boarding if a relative ranking which focuses on such factors as a vessel's size, its cargo, operations and security performance, exceeds a predetermined threshold [175].

3.2.4.3 Threats

Establishing the effect that an entity may have within an environment informs its status as adversarial or threatening. This knowledge is necessary in the decision making process in order to ensure that the correct course of action is taken to mitigate this threat. Threat assessment is pursued in domains as varying as human behavioural studies, law enforcement, counter terrorism, intruder detection systems and military applications [6, 12, 35, 143]. It is a central component of *threat evaluation and weapon assignment* (TEWA) systems in the air defence domain where an enemy aircraft is deemed to be a threat (with varying degrees of certainty), or not [143]. Within the air defence environment the notion of threat is clearly defined. An aircraft that is a danger to a particular asset is deemed to be a threat to that asset.

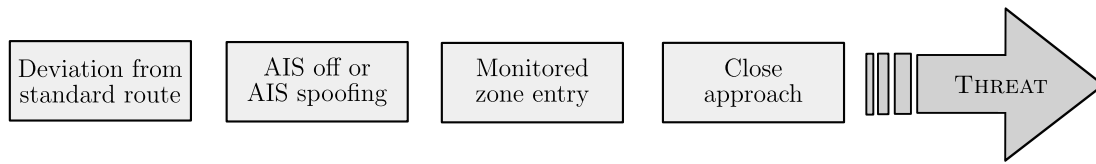


FIGURE 3.5: A number of activities contributing to a vessel being classified as a threat [84].

Activities that lend themselves to this relational explanation of a threat are activities that are threatening to installations (such as harbour infrastructure) or sovereignty (foreign military vessels entering territorial waters of the coastal nation without permission). Pirate vessels are also deemed a threat in relation to commercial shipping or with respect to the targeted vessel. Piracy may thus be seen as an economic threat to the coastal state, poaching a threat to the coastal state's marine resources and smuggling a threat to its citizenry⁸.

Within the vessel designation framework of Figure 3.4, a liquid natural gas tanker travelling past a harbour in a fashion which is considered normal, is deemed to be a vessel of interest and a threat simply due to its hazardous cargo. Should this tanker suddenly change course toward the harbour, it would constitute anomalous behaviour which is threatening (the vessel is still a VOI) [144]. Moreover, a series of actions taken by a vessel may result in it being considered a threat. An example of such a scenario is presented in Figure 3.5 where a threat results from individual activities which contribute to the certainty of threatening behaviour [84].

Threat assessment resides within level 3 of the JDL model presented in Figure 3.1 [13].

3.3 Existing maritime surveillance DSSs

The prevailing approaches to solving the modelling component of Figure 3.3 (as touched upon in §2.2) are expert systems, data-driven approaches and combinations thereof. A rule-based expert system is discussed in [104, 144, 145, 147]. A maritime domain ontology is developed by Roy [147] and a few examples of anomaly detection criteria as rules are provided. The elements of this system include the ontology, a rule set, a reasoning component capable of making inferences from the rule set and a contact history database in which maritime data are stored after having carried out data fusion. A prototype of this system was presented in [145] and further results and refinements were presented in [147]. The specification of this system and the additional publications surrounding its design are the most comprehensive in the literature. There are, however, further publications within the maritime safety and security domain that contribute to this approach in the form of ontology refinement and semi-automatic ontology extensions [22].

The ability to discover new rules is important in the context of expert systems as these systems are generally not capable of generating new rules of inference [63]. A failure to address this problem means that an expert system is limited to previously defined situations [15]. This is compounded by the fact that the possible combinations of vessels, their behaviours and their contexts render activity enumeration intractable [15].

A comprehensive or general purpose data-driven system has not been demonstrated in the literature. However, various approaches have been proposed that solve a part of the larger problem or focus on particular outcomes within maritime surveillance systems. For example,

⁸In a US congressional report [45], threats to port security include cargo containers used for smuggling, use of commercial vessels as collision weapons, the sinking of large vessels in shipping channels, attacking liquefied natural gas vessels with the intention of detonating the fuel, and the destruction or damaging of oil tankers so as to cause large-scale environmental damage or to disrupt the oil trade.

the system proposed in [84] identifies five particular activities and constructs probabilistic models to address them. These models are then fused, using Bayesian inference, into a threat value. Nevertheless, this research area is very active at present with various solutions being sought to problems such as AIS coverage and sea lane anomaly detection [86] (in addition to those discussed in §2.3.3).

Systems combining these principal methodologies have recently emerged as potentially viable solutions and their ability to detect known and unknown activities is beneficial. A framework for such a system is presented in [15] which makes use of a situation management methodology and relies on an agent-based approach. These agents execute and codify particular activities, such as pursuit, raid and smuggling, using a rule-based formalism. A data-driven approach is incorporated through the addition of an anomaly detection agent⁹. A framework that combines unsupervised and supervised learning in a traffic analysis setting was also presented in [115].

There are a few systems that are currently in production or have recently been deployed in some capacity. Three of these systems are:

- The SeeCoast port surveillance system (a BAE Systems project) addresses automated scene understanding through learned normality models of vessel activities [155]. An architecture is presented utilising rule-based and learning-based pattern recognition components.
- The SCANMARIS project implements an adaptive multi-agent system which alerts an operator of abnormal events within the surveillance picture. Alerts comprise three categorizations, namely licit, illicit or unknown [114].
- Finally the PANDA system (a DARPA initiative) aims to implement four system components, namely a motion-based pattern learning component, a prediction and activity monitoring component, an adaptive context modelling component and an anomaly processing and presentation component [134]. The stated goal of this system is the automatic evaluation of all large surface maritime vessels with the intention of determining which anomalous activities are indicative of an emerging threat.

Although some publications in the literature discuss elements of these systems, the active research conducted by governments and the companies involved is generally withheld (the interested reader may consult [104] for a brief exposition of additional systems). In the case of the PANDA project there is very limited information in the public domain.

3.4 A newly proposed maritime surveillance DSS

The system proposed in this dissertation for a maritime surveillance system follows an ADSS approach in which an unsupervised learning mechanism contributes to the adaptive nature of the system. Such an unsupervised mechanism will perform a knowledge discovery role and with the aid of the operator this knowledge may be confirmed or rejected. The system must follow the *operator in the loop* paradigm and be sensitive to changes in the surveillance scene. For example, if new regions of interest or behaviours emerge due to repeated activity, the system should be capable of discovering them and introducing them to the system (the same principle as put forth in Figure 3.2 whereby discovered templates are integrated into the fusion process). The

⁹An order of precedence is established over the agent reports so that if multiple agents report simultaneously, then the vessel is classified as engaging in whichever activity has precedence [15].

system should combine the top-down approach of expert systems with the bottom-up approach of data-driven knowledge discovery and classification. Although machine learning techniques are ultimately discovering *rules* in the general sense (approximating a mapping from an input space to an output space [113]), there are numerous known activities that are easily identified and easily formulated that do not warrant the application of advanced data-driven methods¹⁰.

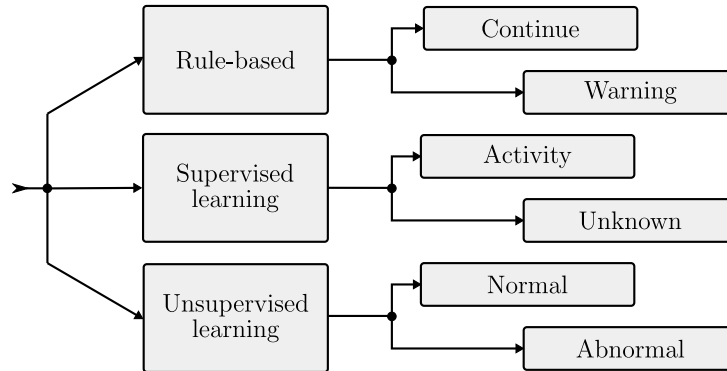


FIGURE 3.6: The different methodologies employed in the newly proposed system's decision making process along with their outputs.

The information flow over a proposed combination of the three aforementioned methodologies is presented in Figure 3.6, which forms the modelling/processing component of Figure 3.3. For the purposes of this discussion, no assumptions are made regarding the data source. Observations are presented to three components, a rule-based system component, a classifier suite and a clustering suite. These components can produce negative or positive outcomes. A negative outcome would be the determination that an observation is abnormal or that an activity is undesirable or unknown, or that a strict rule is not being adhered to. A positive outcome would solicit no warning or designation from the rule-based component, may be classified as a part of a particular activity, or would be participating within the boundaries of what is considered to be prevailing or common behaviour.

A novel framework for a maritime surveillance system is presented in this section. This framework provides the mechanism by which the outputs in Figure 3.6 may be realised. The system is divided into four processing components, namely a fusion component, a rule-based subsystem component, an activity classifier and a data mining component (referred to as a *discoverer*). The display component is embodied in the *human machine interface* (HMI). This interface provides the operator with relevant information and provides a mechanism by which the operator may provide feedback.

The fusion component processes the information received from sensors and the track updates are stored in an operational database. This operational database is directly utilised by the rule-based system component which processes the data as they become available. Data from the operational database are migrated to a DW after being cleaned and transformed into the necessary format. As tracks become available in the DW, they are classified by the activity classifier component (which also provides feedback to the HMI). The activity classifier is enriched by the data mining component which attempts to identify structure in the data stored in the DW.

¹⁰ An example of an easily formulated rule is that of whether or not a vessel bound for a South African port has submitted a *pre-arrival* report to the Maritime Rescue Coordination Centre 96 hours before its arrival at the first port (this rule applies in many shipping nations) [121]. This report is required of all foreign passenger vessels, cargo vessels of tonnage larger than 500 tons and all mobile offshore drilling units.

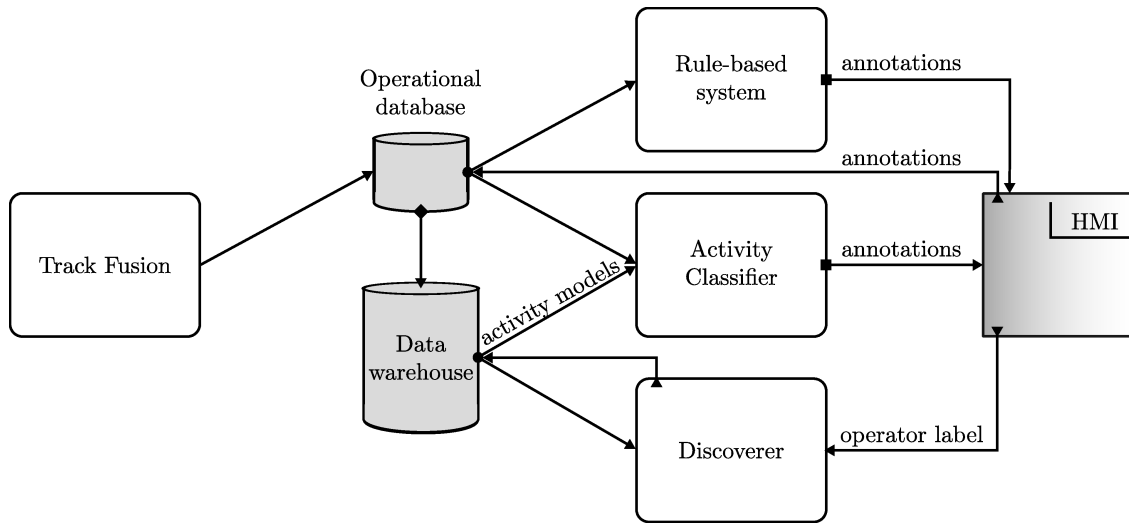


FIGURE 3.7: An overview of the proposed DSS.

The notion of a spatial horizon is required in order to truncate or partition the data into manageable subsets. An example of such a spatial horizon is a radial distance from a particular point at a port. Classifiers may be trained on data that fall within the resulting region. The behaviour of vessels close to port will not necessarily coincide with behaviours out at sea, and for this reason a spatial band centered along a curve may be employed. A good example of where this notion may prove useful is in isolating tracks of vessels travelling in the region of the EEZ. The spatial bands and horizons would need to be specified *a priori*. An additional system parameter that is required is that of a knowledge update interval which would indicate to the system when it is necessary to review existing models. This is necessary so that the system is capable of adapting to changes in vessel behaviour over time.

A central feature of this architecture is the active and passive annotation of vessel tracks. Firstly, the fusion component attributes a class to a track which it deems to be originating from a particular type of vessel. Additionally, the rule-based system annotates data points along the track which have fired particular rules. The relevant designations are then attached to each of those data points and the augmented track is stored in the DW. The activity classifier also labels a track or subtrajectory thereof as engaging in a particular type of behaviour. The aforementioned labelling of data points and tracks constitutes the passive annotation elements of the system. Active annotation is performed by the operator who may disagree with a particular designation and deem it necessary to change it¹¹. The annotations serve to enrich the tracks within the DW, may facilitate later analysis and provide a means to search for particular collections within the DW. In this manner they may serve as a potential information source that may be exploited by data mining.

An important consideration in system design is the modularity of system components. A separation of responsibilities contributes to the longevity of the system and the ease with which extensions may be made should elements need to be modified or completely replaced. Furthermore, the system should be capable of taking advantage of emerging technologies in the form of web services and ever increasing distributed processing opportunities.

¹¹The action of an operator adding a designation of his own is certainly worthy of investigation. In this approach, the operator plays a greater role in knowledge creation. However, it does run the risk of resulting in a proliferation of designations or categories. Nevertheless, this is not considered as an operational function of the system.

Each of the aforementioned four components of the newly proposed system are discussed in greater detail in the following section.

3.4.1 The fusion component

This component comprises two subsystems, namely a *track fusion* component which deals with entity location and fuses various entity locations into new or existing tracks, and an *identity fusion* component which attributes an entity type to each track, as illustrated schematically in Figure 3.8. The fusion system is concerned with producing kinematic tracks from various sensors such as radar and AIS. Additional information that may arise from static AIS data, such as whether or not a vessel is underway, may also be dealt with by adding additional capabilities. The track fusion component may use a Kalman filter [160] or particle filter [112] to fuse kinematic data from available sensors.

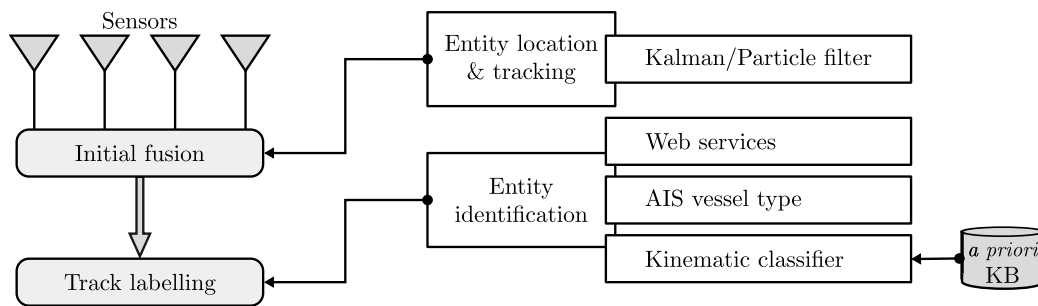


FIGURE 3.8: The track fusion component.

The entity identification component plays an important role in this system design as the activity classifier depends on its robust operation and it is tasked with determining the vessel type. Three sources of information are proposed for the determination of the vessel type. The simplest of these is the vessel type, as specified in the AIS data reports. However, these reports are subject to corruption, tampering or erroneous setup by mariners¹². With the rise of semantic web and service oriented architectures, it is expected that some vessel queries may ultimately be performed via web services and that, by harnessing such sources, further assurances of vessel types may be attained.

Lastly, a kinematic classifier aims to provide some indication of vessel type in the absence of AIS data or identifying characteristics by which to determine vessel types online. This classifier may be trained using kinematic data collected from AIS and radar¹³. The sequential nature of the vessel trajectories may be modelled by means of HMMs or conditional random fields [28]. The kinematic classifier will need to be trained and tested on data from a particular region before the system is deployed there.

Although the purpose of this component is to produce a fused track, the original data will nevertheless be retained for historical evaluation when necessary. For example, if AIS and radar

¹²An MMSI number of 10110010 is often reported by vessels which indicates a faulty initialisation of the AIS equipment. However, websites such as [203] do provide functionality for querying vessel information from their MMSI number. Provided that the MMSI number is correct, the vessel class may be confirmed in this manner.

¹³Although the AIS training/testing data will provide the vessel type, radar data will need to be annotated by another mechanism in order to enrich this training set. Spotters or visual surveillance data will play an important role in constructing a set suitable for training the kinematic classifiers. The thesis that vessel types may be differentiated by their kinematic behaviour over time would be strengthened if kinematic data is obtained not only from AIS sources.

data streams are fused, then the original data streams will be stored in a database along with the fused track.

3.4.2 The rule-based component

Many of the behaviours encountered in the maritime domain may be precisely articulated, making them suitable for direct rule-based solution methods. Two examples of rules currently in use within the South African maritime surveillance context are a minimum speed infraction rule, as presented in Figure 3.9(a), and a zone infraction rule, as presented in Figure 3.9(b).

```

if not in harbour zone then
  if not excluded vessel then
    if speed of vessel < min speed for vessel type then
      SPEED ALARM
  
```

(a) A rule enforcing a minimum speed, depending on the type of vessel.

```

if vessel in no-go zone then
  if vessel not excluded then
    NO-GO ALARM
  
```

(b) An alarm is generated when a vessel enters a closed area.

FIGURE 3.9: Two examples of rules that may be used to generate alarms.

In the interests of maintaining scalability of such a system within an environment in which data processing will only increase (through growing stores of historical data and increasing maritime activity), it is important to pursue concurrency of tasks as far as possible. The shift from single processors to multi-core processor machines has heralded the introduction of multi-threaded capabilities in various software solutions¹⁴ [170]. Individual rules within the rule-based system are ideally suited to a multi-threaded implementation as they are inherently modular (most operations may be executed independently of one another) and their workflow may easily be contained within distinct threads.

An agent-based paradigm neatly encapsulates these notions. This paradigm rose to prominence during the late 1990s and although some elements of this approach overlap with concurrent and object-oriented programming, it differs in a number of respects. Firstly, agents are considered to be autonomous and capable of coordinating and synchronising their activities¹⁵ and secondly, encounters between agents are governed by self-interest as an agent is primarily concerned with attaining its own goals [202]. The latter characteristic is interesting as agents within a large system would need to negotiate and interact in order to achieve a desired outcome (thus exhibiting social interaction).

There are numerous definitions of what constitutes an agent in the literature (see [48, 148, 149] for examples of these). These definitions may depend on the discipline or on the abilities of the agents. For example, generic agents are referred to as *software agents* in computer

¹⁴Careful consideration is required when deciding what to parallelize. For instance, parallelizing small operations introduces communication and synchronization overheads, be it memory or feedback to the calling thread, which are usually less pronounced when operations are less finely grained [170].

¹⁵In concurrent systems this coordination is generally hard-coded *a priori* [202].

```

if vessel class is fishing then
  if vessel type is trawling then
    if vessel location in EEZ then
      ILLEGAL FISHING
    else
      if vessel near EEZ then
        if vessel course toward EEZ then
          if expected departure date today then
            if vessel has history of illegal trawling then
              ILLEGAL FISHING

```

FIGURE 3.10: A rule which may identify an occurrence of illegal fishing (defined in [144]).

science. Another good example is that of *reactive agents* which respond to stimuli in their environment and are equipped only with *if-then-else* rules, whereas intelligent agents¹⁶ may be thought of as being capable of learning and modifying their behaviour [149]. Agents have been used successfully in a number of contexts, one important example area being in air traffic management systems [89]. There are, however, few examples of systems that fully realise the notion of intelligent agents [60]. Nevertheless, the agent paradigm is productive in dealing with scalability and future-proof issues as they perform asynchronously and autonomously, are naturally suited to performing their tasks on heterogeneous computing platforms and networks (provided that the necessary execution environment is available), and they allow for naturally extendable system development¹⁷ [179].

Multi-agent systems are intrinsically multi-threaded as each agent has at least one thread at its disposal [148] and communication between agents is achieved in an explicitly defined and available language (standards include the *agent communication language* and KQML). A number of the maritime anomaly detection systems that use an expert system approach, codify their rules in formal logic [65, 146]. A natural analogy within the agent paradigm is that of deductive formal logic agents. As noted in [147], a limiting factor in their logical reasoning system is the prohibitive time it takes to reach a decision (the author dealt with this problem by limiting the size of the ontology). Relaxing the strict logical restrictions is expected to result in faster rule resolution, but according to Wooldridge [202] the greatest advantage of this approach is thus lost, namely the simple, elegant and logical semantics. Not only is it time consuming to reason over large ontologies with many rules, but environmental information must also be mapped to precepts which are usually symbolic (in order to facilitate reasoning with their representations) and spatio-temporal activities are typically difficult to represent¹⁸. Nevertheless, formal logic is expressive and provides a consistent means to reason about information and to organise knowledge.

The division of responsibilities between deductive formal logic agents may avoid the situation where the reasoning process becomes protracted and the operational picture changes before a decision is reached¹⁹. An example of such an agent is provided in Figure 3.11(a). The agent

¹⁶An intelligent agent is an agent which exhibits all the characteristics of an agent, namely that it is proactive, autonomous, reactive and cooperative, but it is also capable of learning through knowledge acquisition and behaviour modification based on this knowledge.

¹⁷As noted in [197], decision makers may need to rely on information sources emanating from web sources which require an amalgamation of new and old knowledge. It is possible to add data sources to the system in Figure 3.8 by incorporating an agent with the desired functionality to deal with this new source.

¹⁸De Vries [186] developed the Space package in SWI-prolog which provides a mechanism to reason declaratively over spatial objects.

¹⁹An alternative approach is to use software agents which lack formal logic capabilities but retain most of the

observes its environment and reacts to stimuli resulting from this observation or interaction. Various information sources may provide these inputs, such as sensors or other agents. The agent is provided with a framework for understanding its world through an ontology which forms a part of its knowledge base. The knowledge base may be constructed from maritime ontologies which may be shared in part between agents, but may also be specific to some of them. The rule expressed in Figure 3.10 may, for instance, form part of a *fishing agent* who will continually check for undesirable fishing activities. In order to keep communication manageable, agents may be arranged into a federated system capable of acting as aggregators or mediators between agents and the system, as shown in Figure 3.11(a) [47]. An obvious dichotomy of *alarm* and *warning* federations may be identified. The relay of information from these federations may be managed by a controller agent which reports results to an HMI through a *user agent*. The user agent is responsible for relaying these reports and for handling queries from the user with respect to the rule-based system. Furthermore, users may request the reason for a particular designation of a vessel (for example, as an illegal fishing vessel, via the user agent who may then query the agent who generated the alarm or warning²⁰).

A multi-agent system design presents an attractive solution for the architecture of this component. Indeed, an architecture employing this design is well suited to serve as the core processor of such a system where the agents are allocated to deal with rules, user interaction, classifier results and data mining suggestions. However, in an initial design agents should be restricted to the rule-based system where specific agents are tasked with the verification of particular rules. A first-order solution may be a conventional software solution, whereas a later approach may be to follow the formal logic approach suggested in [147].

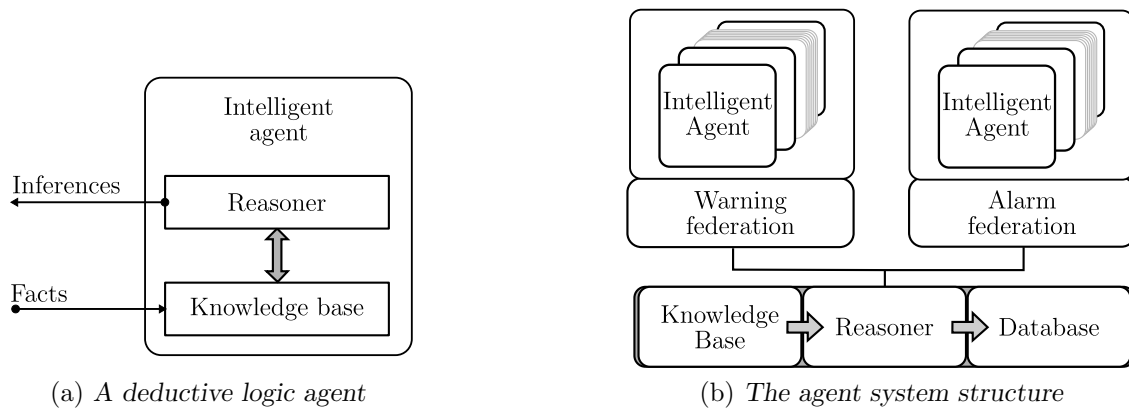


FIGURE 3.11: The basic elements of a multi-agent system tasked with issuing warnings and alarms.

The output of the rule-based system is vessel designations determined by the rules that were fired in relation to that vessel. These designations are categorical and may fall into broader classes such as navigational or spatial infringements by a vessel. The urgency attached to a particular infringement should be set *a priori* in consultation with domain experts and in this manner a hierarchy of severity may be established. As mentioned in §3.4, the points along a vessel track are annotated with designations from the rule-based system. In the event that a user provides feedback to the system which disagrees with any of the designations, these additional annotations should also be added to the track and should take precedence over the system's recommendation²¹. The rule-based system component is presented in Figure 3.12. Its primary

desirable properties, such as autonomy, pro-activity, communication and reactivity to their environment.

²⁰Mechanisms exist for agent reporting whereby an agent is capable of explaining itself to a user [57].

²¹Certain operational privileges should be granted to operators so as to ensure that system recommendations are not overridden arbitrarily or without recourse.

data source is the fused tracks from the operational database and it relies on the codified rules in the knowledge base. The outputs of this component are the designations, which result from the evaluation of real-time tracks subject to the rules, and the means by which these designations were obtained (this reporting mechanism is only brought to bear at the request of a user).

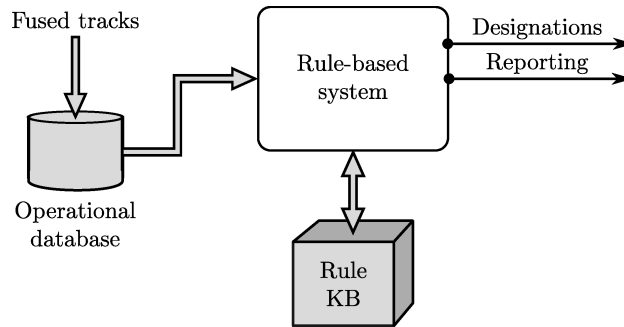


FIGURE 3.12: *The rule-based system component.*

3.4.3 The discovery component

Once the system has been deployed and a significant amount of data have been collected, the data may be partitioned using predefined spatial horizons and bands. These data are assumed to have vessel classes attributed to them by the fusion component. The data mining component should then perform automatic analysis on these data before the activity classifiers may be trained or activated.

The data mining component is instrumental in gaining an understanding of the surveillance scene. Such an understanding may be achieved by searching for exploitable structure within the data stored in the DW. This pursuit is not arbitrary; it is expected to employ clustering methods that focus on particular aspects of the data. For example, data mining may be used to cluster tracks within a chosen spatial horizon that originate within the same regions and end at the same destination (referred to as an *origin-destination miner*), as well as to search for viable partitions of tracks within the kinematic feature space. Each of these cases is illustrated schematically in Figure 3.13 and is discussed below in order to demonstrate the operation of this component.

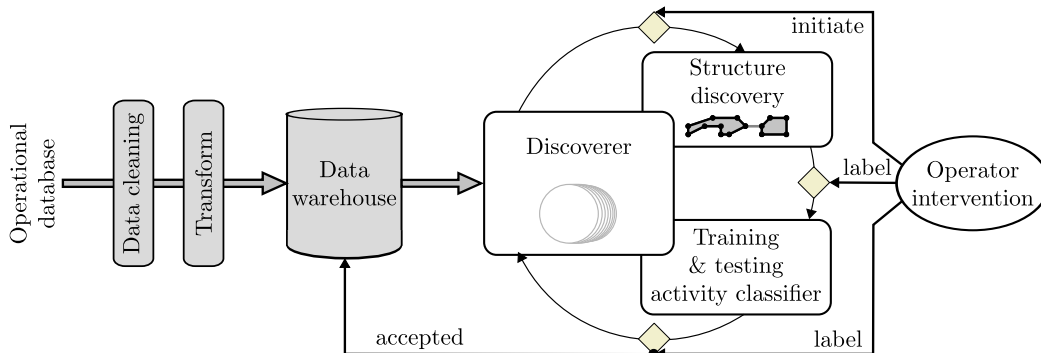


FIGURE 3.13: *The discovery component.*

Firstly, an origin-destination miner may be tasked with extracting tracks that share the same (ordered) origin-destination pairing²². These tracks should be labelled appropriately by the

²²These data mining operations are to be conducted at specified knowledge update intervals. This behaviour

system, using the designated labels as provided by the operators or known *a priori*, and may then be used to train activity classifiers. As illustrated in Figure 3.13, once a valid activity classifier is identified, the operator is once more prompted for an appropriate label and this classifier is integrated into the data warehouse. The designations that a rule-based system applies to vessels entering a marine protected area may also serve to identify a subset of data to which the origin-destination miner may be applied. Consider the scenario of ferries travelling to and from ferry terminals or docks. As these tracks share a common source and destination, they should be labelled by an origin-destination miner as such. These data may then be partitioned into a training set and test set (and in the event that there is an abundance of data, a validation set too [11]), the necessary features may be extracted or selected and the training of the classifier may commence. This optimization process should allow the classifier to adjust its parameters in order to minimize its predictive error. If the trained classifier achieves a satisfactory score with respect to the test set (this parameter must be specified *a priori*), then it should be referred to an operator for labelling. The operator may attribute the label of ‘ferry from A to B’ in the event that the origin and destination are clearly defined, or ‘ferry approaching B’ in the event that only the destination is well defined. This mechanism of utilising knowledge obtained through data mining during the training of classifiers should be repeated for various other groups of data.

Furthermore, by focussing only on the destination of a track²³, the discovery component may prove useful in identifying anchorage regions that arise beyond the regions which are dictated by nautical charts or designated by harbour authorities. If a region is discovered that does not correspond to a previously determined region, then a label for the region should be requested from an operator²⁴. This operator intervention is illustrated in Figure 3.13. If the discovered region corresponds to a region identified during the last knowledge update cycle, then the label of that corresponding region should be assumed. These regions and their labels may then be integrated into the DW. It should be noted that *a priori* information, such as the coordinates of marine protected areas, may be available. The coordinates of these regions should also be stored in the DW and may be used in the same manner as the discovered regions.

Lastly, it may be beneficial to search for differences in inter-class kinematic profiles. As mentioned in §3.2, it may be possible to differentiate between different classes of fishing vessel types. Data mining may thus be applied to the data labelled as fishing vessels with the goal of discovering a reasonable partitioning within the broader class (these labels should be attributed by the fusion component and are expected to be coarse in the sense that they may not necessarily indicate sub-class types). If a partitioning is found, then it should be suggested to an operator who may then perform further analysis in order to label these sub-types. In this case a realistic labelling should ideally be pursued.

The knowledge update interval allows the system to adapt to a changing scene and changing behaviours, albeit discretely. It is, however, still necessary to successfully negotiate knowledge expiration. It is not immediately clear when vessel tracks should be excluded from data sets that are processed or when a region that is seldomly traversed should be discarded. This problem may be circumvented in a first-order approach by discarding all discovered structures which do not reoccur, after each knowledge update interval. This should not include information that is specified *a priori*, such as marine protected area boundaries or harbours. Furthermore, it is noted that not all knowledge will expire at the same rate and it may be true that certain tracks should remain available to the learning framework for longer than others.

may be subject to change. The use of an update trigger may initially simplify the more complicated dynamics of integrating information whenever sufficient amounts have become available.

²³It is assumed that the tracks stored within the DW are divided into stop and move segments.

²⁴A working system should ideally attribute various operational privileges to different operators.

3.4.4 The activity classifier

Adopting the traditional supervised learning paradigm, data extracted by an origin-destination miner and other data mining models are used to train various classifiers. Each classifier should correspond to a particular activity and these activities may be learned over time as a result of the suggestions by the clustering component and with the assistance of operators. An example of such an activity classifier is a *hidden Markov model* (HMM) [11] trained on an identified subset of data in the DW.

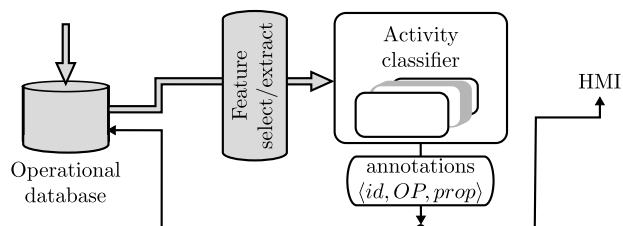


FIGURE 3.14: *The activity classifier component.*

If a particular vessel track is found to be consistent with one of the classifiers, then it should be labelled as such. This annotation should be displayed to the operator and maintained in the operational database (this process is illustrated in Figure 3.14). If the track fails to match any activity, then it should be labelled as engaging in an unknown activity²⁵. These labels may then further assist knowledge discovery through data mining.

3.4.5 The human machine interface

All operator system interactions should occur through the human machine interface. These interactions may be divided into two categories, based on information flow. Firstly, immutable data requests should be executable by an operator in a *pull* fashion. The operator should be able to request information from the rule-based system (*e.g.* to understand the reason for a vessel designation) or historical tracks from the data warehouse (*e.g.* so as to perform further analysis on a particular vessel). Secondly, the operator should be able to interact with the system in a *push* fashion whereby RBS designations or activity classifier annotations are modified. Operators should also be able to add labels of their own to particular segments of a track or provide labels to the structures found by the Discoverer.

3.5 Summary

DSSs and data fusion were briefly discussed in §3.1 as these concepts form an integral part of many surveillance systems. The constituent elements of a hypothetical maritime surveillance systems were discussed in §3.2. These included the data that might serve as input to such a system, how meaning may be extracted from a surveillance scene, the role of the operator and lastly the outputs of such a system in the form of vessel designations. Existing maritime DSSs were described in §3.3 and finally a novel maritime surveillance DSS was presented in §3.4. The rule-based system, data mining system, activity classifier and human machine interface of this DSS were described in §3.4.2–§3.4.5, respectively.

²⁵It should be appreciated that a portion of a track may not be congruent with any of the known activities, but that another portion may be. The same approach as was mentioned in the rule-based system should therefore be taken whereby individual points along the track are labelled.

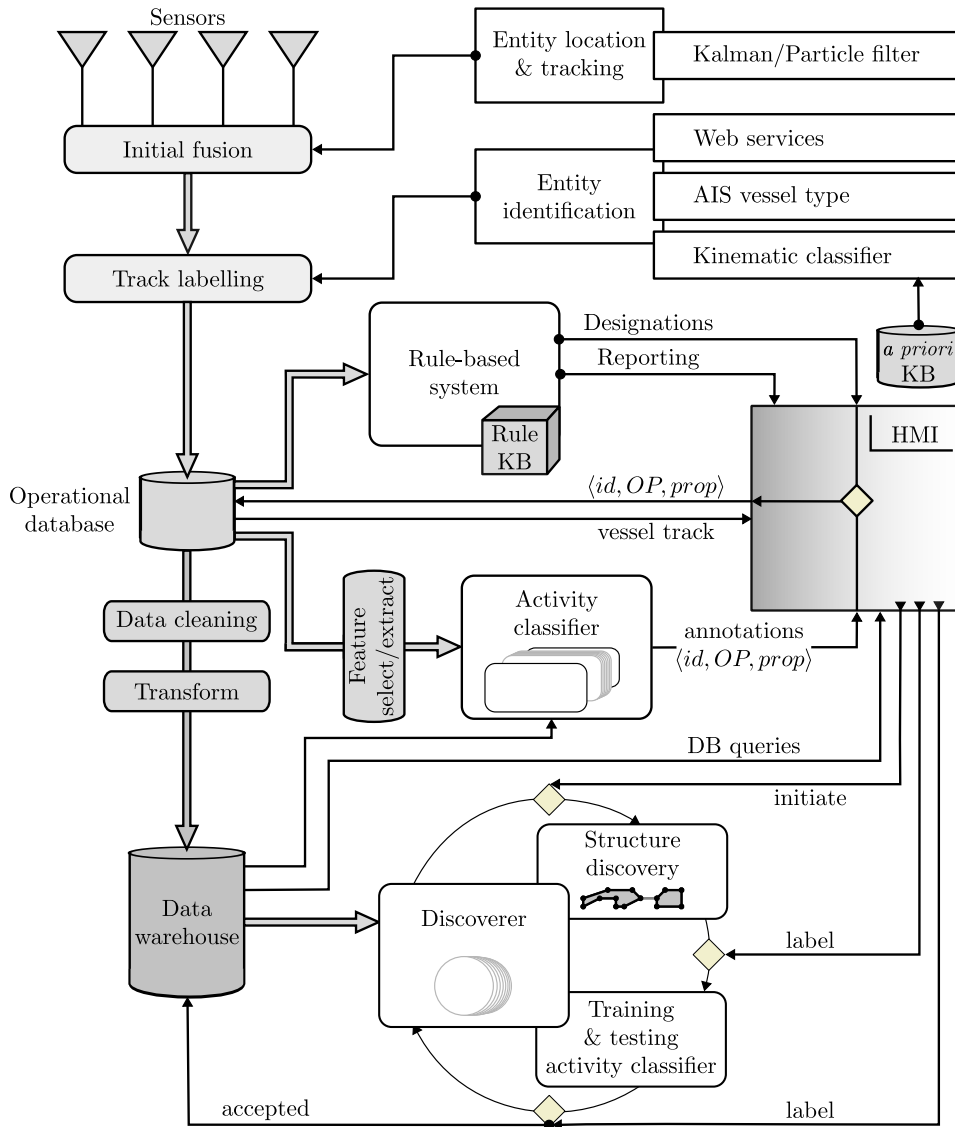


FIGURE 3.15: A detailed overview of proposed DSS.

Each of the aforementioned components may be integrated into an overall system architecture, as illustrated in Figure 3.15. In the figure, the interactions between the various components are indicated with thin black lines whilst the data flow from the initial fusion to the various databases and system components are depicted by thick filled arrows. Operator interventions are indicated using filled diamonds. The vessel track is displayed directly to the HMI from the operational database. Vessel and activity designations that are attributed to the track by the rule-based system, activity classifier, or by the operator, are also displayed in the HMI. The track is augmented with these data in the operational database. These augmented tracks are ultimately archived in the DW and they provide the historical data which the operator is able to access via database queries. These data are exploited by the data mining component which provides the training data for activity models.

CHAPTER 4

Rule-based system component

Contents

4.1	Constraints to movement and activity	55
4.2	Restrictions pertaining to small vessels	56
4.3	Restrictions pertaining to seafaring vessels	59
4.4	Rule classes	60
4.4.1	<i>Zone infractions</i>	63
4.4.2	<i>Proximity alarms</i>	67
4.4.3	<i>Anomalous actions alarms</i>	74
4.5	Summary	75

Rules which populate the rule-based system component are considered in this chapter. Various constraints to motion or activities provide the justification for some of these rules, while others are borne out of the need to identify suspicious behaviour from vessel meta-data (that is, information regarding the vessel's owners, flag state, or visited ports). An emphasis is placed on rules that rely predominantly on kinematic data as these may be collected from non-cooperative sources, such as radar. Activities pertaining to threatening behaviour are considered in greater detail. Classes of rules are identified and discussed, and the computational methods required to resolve them within their broader classes are considered. Propositional logic is used in conveying the rules when the need arises.

4.1 Constraints to movement and activity

Various maritime rules, dictated by international bodies or governments, are concerned with regulations or requirements for sea-going vessels. Such regulatory requirements include specifications for ship design or construction, vessel crewing, vessel operation and carriage of goods or people [101]. For example, South African safety regulations for small vessels categorize vessels into different classes which permit them to operate within certain ranges from the shore [98]. Vessels that are registered in one of these categories are required to carry the relevant onboard equipment and the skipper is required to be in possession of the necessary certification for operating a vessel within the relevant ranges (see Table 4.1).

The radar cross section of small vessels coupled with the scatter from the moving ocean, makes small vessels particularly difficult to track with radar. Unlike vessels of 300 tons or more, they are not required by law to carry AIS transponders and hence their positions are not necessarily

Vessel Category	Range (<i>nautical miles</i>)
A	unconstrained
B	at most 40
C	at most 15
D	at most 5
E	at most 1

Table 4.1: *Small vessel categories and the seaward ranges from the shore in which they are permitted to operate [98]. The category of vessel is determined by these seaward ranges.*

known at all times. Nevertheless, the rules that govern these vessels are considered as their activities are also of importance in the context of maritime law enforcement and there are many similarities between their activities and those that larger seafaring vessels engage in.

4.2 Restrictions pertaining to small vessels

In addition to the seaward distance constraints of Table 4.1, category E vessels may not operate further than 15 nautical miles from an approved launch site or proceed to other ports [98]. Seagoing vessels greater than three metres in length¹ are required to be registered with SAMSA into one of these categories and each registered vessel should display the registration code that is awarded to it (the category letter appears as a prefix). Vessels operating outside of these operational limits may indicate illicit activities² or vessels in distress. Although skippers are often encouraged by their resident ports to inform others of their travel plans, there are not necessarily systems in place that ensure their timely return.

Regions that smaller vessels may enter are also limited by *marine protection areas* (MPAs). As mentioned in §1.2.1, these spatially delimited regions were established in order to protect marine ecosystems and threatened species populations³. Their effective management is expected to allow for better regulation of activities within these regions whilst minimising the risk of pollution and habitat degradation [96]. South Africa has declared twenty MPAs along its continental coastline (the MPAs of the Western and Eastern Cape are shown in Figures 4.1 and 4.2, respectively). These regions cover a total of 21.5% of the coastline [163] and encompass an area of approximately 426 000 hectares⁴. The designation of these areas as *controlled*, *restricted* or *sanctuaries*, is an indicator of the types of fishing activities allowed within their boundaries. Fishing is prohibited in sanctuaries and any fishing vessel traversing a sanctuary is required to stow all fishing equipment (this includes the activity of spear fishing and spear fishing equipment). Limited fishing is allowed in restricted areas, whereas fishing is permitted in a controlled zone contingent on the fishermen being in possession of the necessary permits.

The marine protected areas around Cape Point are considered in greater detail in Figure 4.3 for the sake of discussion. The Table Mountain National Park is a controlled zone which contains six smaller sanctuary zones. Activities that may be detected via direct visual observation include

¹Vessels less than three metres in length may not go out to sea unless they do so in a designated areas, in which case they may not venture more than 1000 metres from shore.

²Rubber ducks involved in Abalone poaching in the False Bay region have been known to launch under the cover of darkness and travel across the bay to regions, often MPAs, where divers disembark [1].

³The creation of these regions, their effect on subsistence fishermen and local communities, and their continued success are the subject of a number of articles [17, 18, 163].

⁴The twenty-first MPA around Prince Edward and Marion Islands was declared in April 2013. This offshore MPA features a 12 nautical mile no-take zone extending seaward from the high water mark of the islands [139].



FIGURE 4.1: Marine protection areas along the Western Cape coast [152].

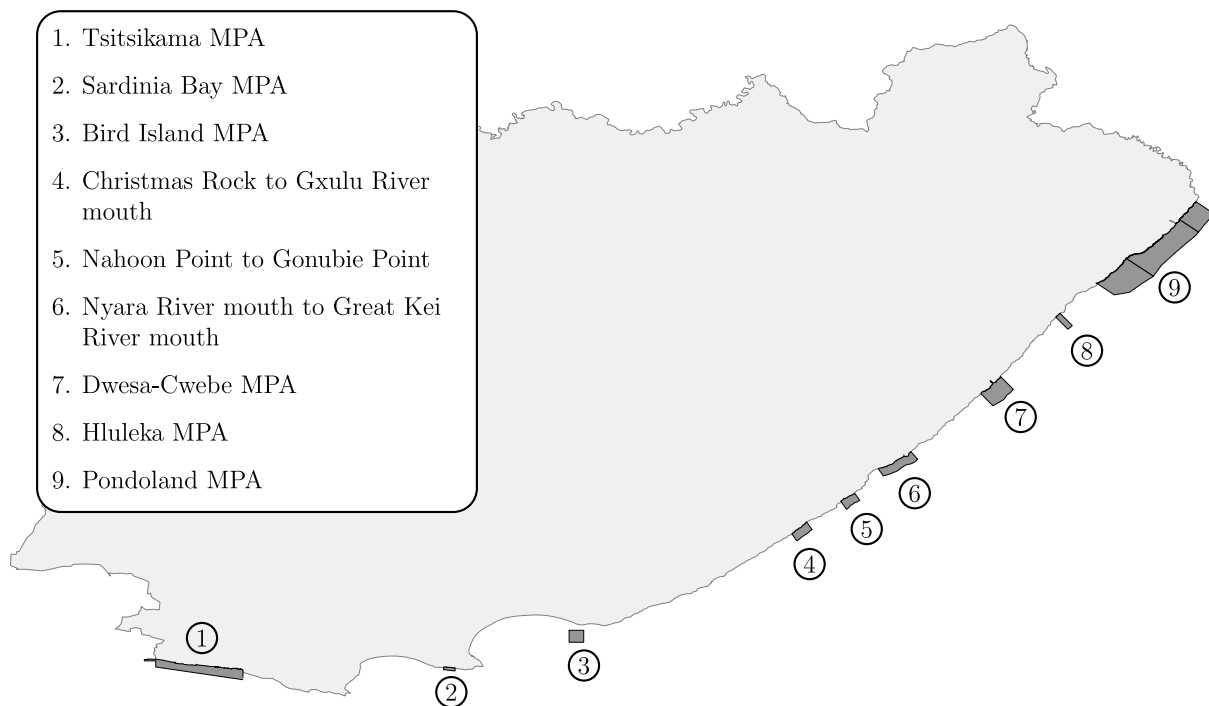


FIGURE 4.2: Marine protection areas along the Eastern Cape coast [152].

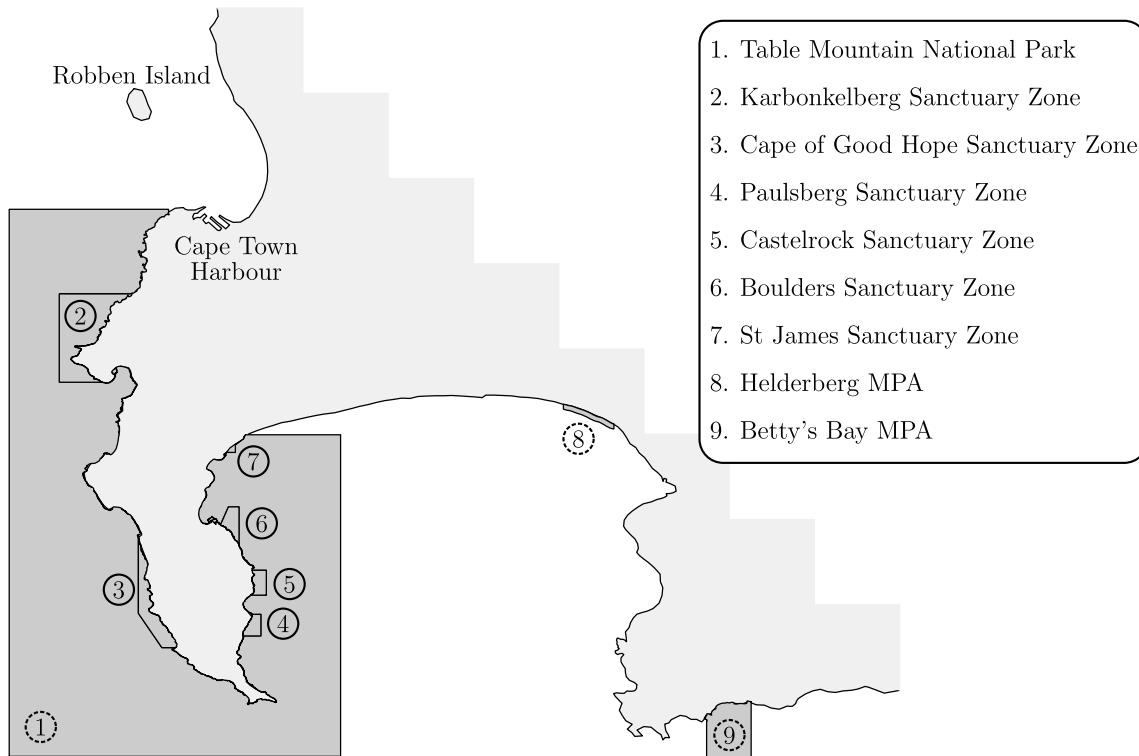


FIGURE 4.3: Marine protection areas around Cape Point in the Western Cape [152].

prohibition of personal watercraft (also known as water scooters or commonly referred to by the brand name *jetskis*), vessels from which deployed divers are expected to fly an alpha or *diver down* flag and vessels anchoring within a sanctuary for longer than 24 hours. Commercial, subsistence and recreational fishing activities may also be exhibited by vessels in the controlled zone, as vessels with the necessary permits are entitled to engage in fishing activities within this zone.

This discussion provides an indication of the basic concepts which govern acceptable actions under the aforementioned constraints. These concepts are:

- geometry of the region,
- restrictions thereof,
- behaviour therein and
- vessel exclusions.

The concepts above are conveniently expressed using atomic predicates that concisely describe the rules arising from the previous discussion. These rules are presented as binary decision trees using this formalism (see Figure 4.4). Predicates are presented, for the sake of discussion, under the assumption that methods exist for evaluating them. The root of such a tree returns true if zone z contains a vessel v . The geometries of zones may be described as closed polygonal regions. First, consider the leftmost branch of Figure 4.4. If a vessel is not contained in the zone in question, where the zone is the seaward distance from the shore, depending on the vessel category (see Table 4.1), then an alarm is issued. However, as stated above, this vessel may

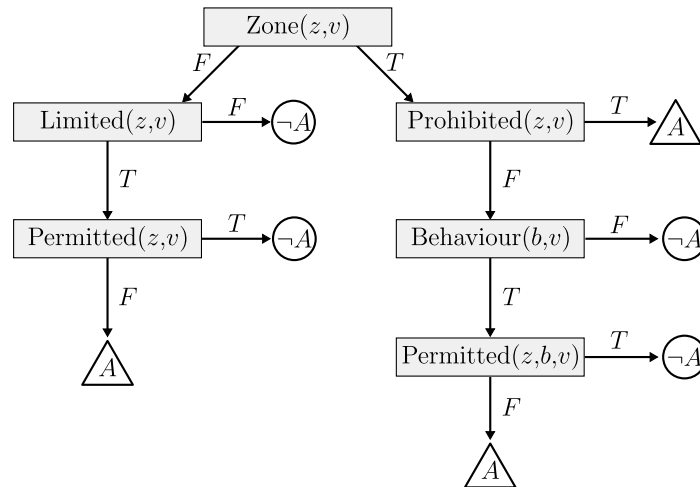


FIGURE 4.4: A binary decision tree describing restrictions that apply to small vessel classes. True, false and alarm are denoted by T , F and A , respectively.

have been granted permission to travel beyond its usual seaward limits or range. If this is not the case, then a generic alarm may be issued as the vessel is in transgression of regulations. Following the rightmost branch of the tree, if a vessel is found to be within a particular zone and is not prohibited within that zone (personal water craft would elicit the issuing of an alarm if they are found to be within one of the sanctuaries of Figure 4.3), then its behaviour is matched to known behaviours. Examples of these behaviours include *fishing*, *diving* and *moored*. If a vessel is found to be engaging in a known activity, then the possibility of exclusions is pursued. Otherwise, no alarm is issued. If a vessel is fishing or has deployed divers without the necessary permits or is found to be anchored in a region that does not permit it (or is anchored for more than 24 hours in a sanctuary), then it is not permitted to be engaging in that activity within that zone and an alarm is issued. Although the necessary systems for matching permit holders to visually identified vessels may not be in place, these rules simply codify regulations which are already in place⁵

4.3 Restrictions pertaining to seafaring vessels

Larger vessels are easier to track by radar and many of them are equipped with AIS transponders. The majority of the constraints of interest to their motion are often relevant within close proximity to the shore or close to ports. Ports present their own set of challenges to vessels, depending on their geographies, regulations and facilities. Consider the Port of Cape Town as an example. Pilotage is compulsory for large vessels wishing to dock in the harbour. These vessels are boarded by a pilot at a point roughly two miles north west of the port breakwaters⁶ [129]. As is the case with smaller vessels, mooring is forbidden in certain regions (such as VTS lanes or the port channel), but there are no designated anchorage areas outside the harbour.

Testing for vessel containment within a specified zone would establish whether or not the aforementioned restrictions are met. In fact, the geographical region defined by the coastline and

⁵There is an initiative underway in South Africa to equip crew of subsistence fishing vessels with personal GPS devices to assist in SAR. These devices may serve the dual purpose of providing update information for these small vessels [189].

⁶A vessel docking in the port is required to report to Port Control four miles from the port limits in addition to the aforementioned 96 hour call-in [129].

the EEZ boundary allows the activity of skirting and fishing to be described by the rightmost branch of Figure 4.4 (referred to as *grab and dash* in [146]). Appropriately defined geographical regions allow for vessel containment tests to be used to detect activities such as foreign navy ship entering territorial waters (which, without permission, would elicit a warning) or vessels disregarding navigational warnings issued over a region of the ocean in response to, amongst other things, inclement weather, faulty lighthouses or mining vessels.

Large vessels are also required to maintain a reasonable distance from offshore installations or from the shore. Rules testing for transgressions in this regard may also be seen as containment tests. However, proximity between seafaring vessels or between seafaring vessels and smaller vessels are typically dynamic in nature and may be considered as their own class of rules because they involve relations between vessels and not only between fixed spatial delimitations and vessels. Indeed, such events are usually of particular importance as they are indicators of smuggling or of a collision or near-collision of vessels.

Although geographical constraints also play a role in shaping activities for large vessels, there are various behaviours that are not confined to geographical regions. These activities are independent of the four concepts identified in §4.2. For example, vessels may tamper with their self-reporting mechanisms and provide false information via AIS (such as false positions) or may simply disable their devices. Such behaviour is commonly considered as anomalous [146] and rules capable of detecting this kind of behaviour have to be specified. Lastly, behaviours may not arise as a result of constraints on vessel actions or location, but may arise from other necessities. For example, a *man-overboard* manoeuvre [94] falls into neither the vessel containment nor the proximity test categories and may rather be seen as a manoeuvring anomaly.

4.4 Rule classes

The rule-based system component is concerned with identifying activities that are in contravention of restrictions and so its results are usually an affirmative of a negative activity. Three rule classes have been identified for implementation and are referred to as *zone infractions*, *proximity alarms* and *anomalous action alarms*.

Rules may be stated concisely and simply stated using propositional or predicate calculus. For example, the sentence

If the vessel is in zone A and is not permitted to be there, then an alarm is issued,

may be expressed in propositional logic as

$$\text{InZoneA} \wedge \neg \text{Permitted} \Rightarrow \text{Alarm}. \quad (4.1)$$

Complex propositions may thus be constructed from propositional variables and operators (see Table 4.2). For example, InZoneA represents the proposition that *the vessel is in zone A*, whilst application of the unary negation operator to the proposition, namely $\neg \text{InZoneA}$, represents the fact that *the vessel is not in zone A*. The truth of the statement made in (4.1) depends on the truth of the precedent, antecedent and the implication operator. The truth table for the implication operator is presented in Table 4.3. The equivalence of $p \Rightarrow q$ and $\neg p \vee q$ is demonstrated in the truth table⁷.

⁷Any binary function featuring propositional operators may be expressed as a function composed of conjunctions, disjunctions or negations.

Operator	Symbol
negation	\neg
conjunction	\wedge
disjunction	\vee
exclusive disjunction	\oplus
implication	\Rightarrow
equivalence	\Leftrightarrow

Table 4.2: Propositional operators [67].

p	q	$p \Rightarrow q$	$\neg p \vee q$
0	0	1	1
0	1	1	1
1	0	0	0
1	1	1	1

Table 4.3: Truth tables for the implication operator.

The limited expressive power of propositional calculus is revealed when attempting to describe many objects in complex environments as it is necessary to create a proposition for each object or state thereof, as well as the various entities to which it may be related. Using expressions as propositions which evaluate to true or false overcomes some of these limitations. The predicates in Figure 4.4 are examples of such expressions. The predicate $\text{InZone}(z, v)$ states that vessel v is in zone z whilst $\neg\text{InZone}(z, v)$ states the opposite. As is the case where a proposition represents facts, these statements allow one to reason about zone containment without providing the actual mechanism by which this containment may be determined. Furthermore, the addition of the existential and universal quantifiers, \exists and \forall respectively, allow for reasoning about groups of objects. For example,

$$\text{ZoneAlarm}(\text{active}) \iff \exists v : \{\text{vessels}\}; \exists z : \{\text{zones}\} \bullet \text{InZone}(z, v) \wedge \neg\text{Permitted}(z, v)$$

states that the issuing of a zone alarm occurs if and only if a vessel has entered a zone in which it is not permitted. Indeed, any expression which may be evaluated to true or false may be considered a predicate [67]. An example of such an expression is

$$\text{CurrentSpeed} - \text{PreviousSpeed} < 0,$$

where these variables are not propositions as before, but variables representing numbers. Set inclusion is another example of such a predicate.

For the sake of completeness, the minor differences between the conceptual binary decision tree depicted in Table 4.4 and the conventional data structure representation of binary trees are considered. The binary tree representation for the implication operator is depicted in Figure 4.5(a). The explicit labelling of the arcs as either *true* (1) or *false* (0) is replaced by the convention that arcs are drawn as solid or dashed lines, respectively. The leaves always represent a true or false result whilst the internal nodes represent decisions or tests. Expressing a binary function (in this case implication) in this manner completely specifies the function for all possible assignments to the boolean variables (it essentially encodes the truth table in Table 4.3). Determining whether a boolean function is satisfiable for an assignment of values may be determined directly from the tree. This approach unfortunately requires the construction of a tree that grows exponentially in nodes.

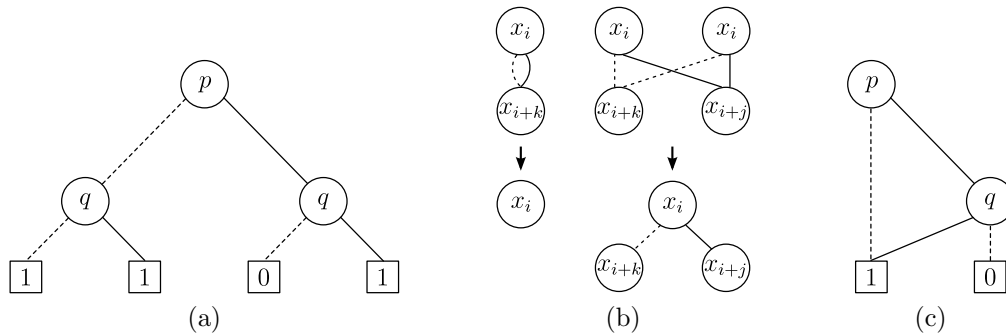


FIGURE 4.5: (a) The binary tree representation of the implication $p \Rightarrow q$ (where a solid line denotes a true assignment to the node from which it emanates and a dotted lines represents a false assignment). (b) Two properties used in the reduction of these trees and (c) a reduced order binary decision diagram representation of the implication $p \Rightarrow q$.

Binary decision diagrams offer an elegant alternative to this representation (also referred to as *reduced order binary decision trees* [81]). Variable ordering is imposed and the conventional binary decision tree is reduced by ensuring that node successors are always distinct and that all isomorphic sub-trees are merged (see Figure 4.5(b)). This means that redundant tests of boolean variables do not occur and that the resulting tree is canonical — it is a unique binary decision tree.

Although predicate calculus provides a more expressive framework and is used in various applications, such as model checking and system specification [67], a propositional approach to describing the rules is followed here for the clarity of exposition that it provides⁸. A number of sets are nevertheless defined so as to provide context for the objects that are of importance to the rule system in practice.

Objects within the system are considered to be described by a set of *properties*. A property is an ordered *attribute-value* pair which is attached to an object. For example, $(callsign, ZARH02)$ would be an attribute associated with a particular vessel. Objects within the surveillance scene may then be described by the ordered tuple

$$\text{obj} := (\text{id}, \text{properties}),$$

where *id* represents a unique identifier awarded to a particular object within the system and the *properties* descriptor captures the attribute-value pairs associated with that object. Geographical objects within a scene may then be defined as a set of objects

$$\mathbb{G} := \{(\text{id}, \text{properties})\}. \quad (4.2)$$

Principal amongst the properties attached to this object is a *geometry* attribute that defines the location of a geographical object (in the case of the geometry being described by a single point) or the geometry thereof. For instance, a lighthouse or buoy may feature a longitude and latitude pair as values associated with the geometry attribute, whereas a geographical region would feature an ordered list of such pairs. The set of vessels, denoted by \mathbb{V} , within the surveillance environment may also be expressed as a set of objects, but in this case the attribute-value pairs

⁸There are existing approaches in the maritime and video surveillance domain that make extensive use of ontologies and *first-order logic*, particularly in reasoning about non-kinematic data. As the focus of this study is primarily on kinematic data and since these approaches are well documented, propositional logic is used when necessary to convey the essence of a rule, instead of using the more expressive predicate or formal logic. Indeed, these rules will initially be implemented in conventional programming languages as opposed to logic programming languages and so the need to use predicate or first-order logic is not yet of critical importance.

describe, amongst other properties, current and historical position, speed, and heading, as well as attributes awarded by the system, such as vessel designations. Vessel types known to the system constitute the set

$$\mathbb{S} := \{\text{cargo, fishing, military, ferry, passenger, pleasure, SAR, tug, \dots}\} \quad (4.3)$$

and known behaviours make up the set

$$\mathbb{B} := \{\text{pursuit, fishing, loitering, drifting, anchored, \dots}\}. \quad (4.4)$$

The explicit definition of vessel and behaviour types provides a taxonomy of these concepts within the system. No distinction is made at present between derived behaviours (such as computationally determining whether a vessel is fishing) or reported behaviours (such as AIS status reports on large vessels where the vessel may report being anchored or underway).

Adopting this basic framework and reasoning conceptually using propositional calculus, the three rule classes mentioned above are discussed below. Methods for resolving some of the activities are explored, but the persistence of an alarm and severity thereof are not discussed in this chapter. Instead, all alarms are treated simply as notifications of the fact that a rule has resulted in an alarm being activated.

4.4.1 Zone infractions

The rules presented in Table 4.4 have a similar underlying structure and are considered to fall within the class of *zone infraction* rules. A region $z \in \mathbb{G}$ may be described by a collection of properties. In addition to a polygonal region specified by the geometry attribute, there are five additional properties that determine how vessels and vessel types interact with z . These attributes include prohibited sets of vessels, vessel types and behaviours. Allowance is made for the possibility that a particular vessel (not contained in the prohibited set of vessels) may be exempt from restrictions imposed on the latter two concepts. For example, fishing may be prohibited within an MPA for all but a few vessels which have been granted permission to fish in that MPA.

Zone infractions	Description
Zone entry	Vessel within a restricted zone
Behaviour violation	Vessel engages in a restricted or prohibited behaviour

Table 4.4: General zone violation rules.

The general zone violations in Table 4.4 may be described succinctly using propositional statements where the proposition InZ represents the fact that a vessel is in the zone under consideration. Suppose that the fact that a vessel, vessel type or behaviour may be prohibited is represented by VPro , VPro and BPro , respectively. Lastly, suppose that the exceptions to vessel type and behaviour restrictions are represented by VPer and VPer , respectively. Then

$$\neg \text{InZ} \Rightarrow \neg A, \quad (4.5)$$

$$\text{InZ} \wedge \text{VPro} \Rightarrow A, \quad (4.6)$$

$$\text{InZ} \wedge \neg \text{VPro} \wedge \text{VPro} \wedge \neg \text{VPer} \Rightarrow A, \quad (4.7)$$

$$\text{InZ} \wedge \neg \text{VPro} \wedge \text{VPro} \wedge \text{VPer} \Rightarrow \neg A, \quad (4.8)$$

$$\text{InZ} \wedge \neg \text{VPro} \wedge \neg \text{VPro} \wedge \neg \text{BPro} \Rightarrow \neg A, \quad (4.9)$$

$$\text{InZ} \wedge \neg \text{VPro} \wedge \neg \text{VPro} \wedge \text{BPro} \wedge \neg \text{VPer} \Rightarrow \neg A, \quad (4.10)$$

$$\text{InZ} \wedge \neg \text{VPro} \wedge \neg \text{VPro} \wedge \text{BPro} \wedge \text{VPer} \Rightarrow A, \quad (4.11)$$

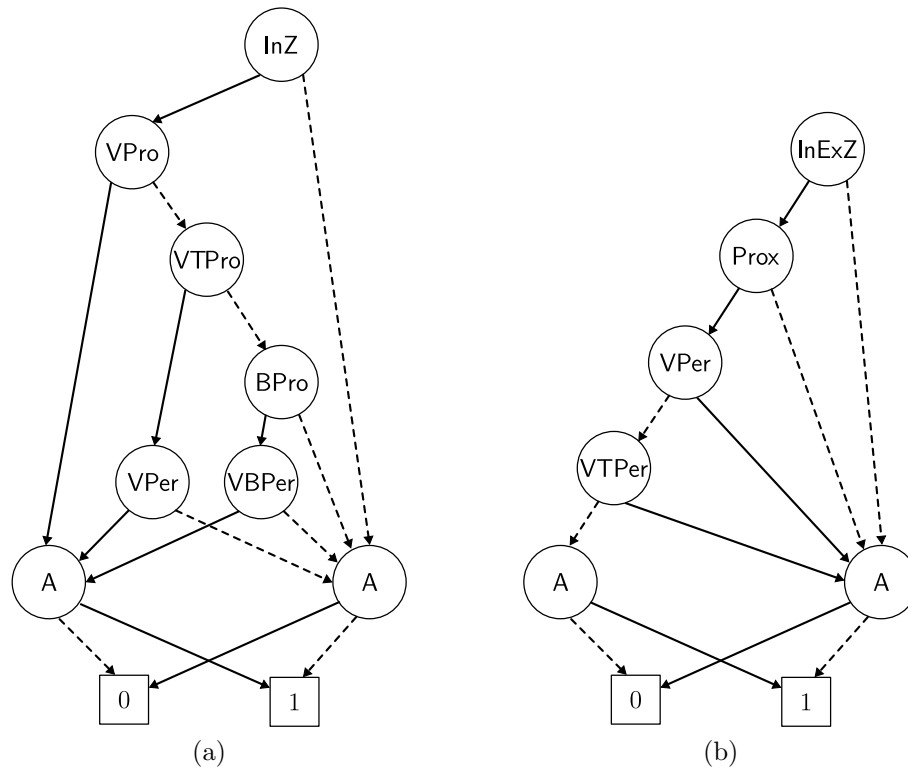


FIGURE 4.6: The binary decision diagram describing the conjugation of (a) the zone violation rules in (4.5)–(4.11) and the proximity rules in (4.13)–(4.16).

where A denotes the event of issuing an alarm. The binary decision diagram of this rule set is presented in Figure 4.6(a). The binary decision diagram demonstrates the truth of the rules explicitly. It may be seen that a vessel within the zone in question which is prohibited from being within the zone must result in an alarm being active (an inactive alarm does not satisfy the boolean equations).

The implementation of rules that are specific cases of the two general forms listed in Table 4.4 requires a means to determine if a vessel lies within a particular zone. The containment of a vessel within a geographical zone may be determined using methods from computational geometry. The problem is analogous to determining whether or not a specified or query point lies within a closed polygon described by an ordered collection of points $P_0, P_1, \dots, P_{n-1}, P_n = P_0$. This technique is often applied in computer graphics applications and geographical information systems, and various solutions exist for this decision problem.

A common solution to this problem is known as the *odd-even* or *parity test*, which is motivated by the *Jordan Curve Theorem*. This theorem states that a simple closed curve divides the points in the plane not on the curve into two distinct domains of which the curve is the common boundary [46, 73]. A special case of this theorem is obtained when the simple closed curve is a simple polygon, *i.e.* a non-intersecting polygon. In this case the set of points in the plane not on $\mathcal{P} = \{P_0, \dots, P_n\}$ is the union of two disjoint sets such that points in different sets must meet at \mathcal{P} . These disjoint sets are then considered to contain points of odd or even parity [73]. The parity test is performed by determining whether a line extending from a query point Q to a reference point R outside the polygon crosses \mathcal{P} an odd or even number of times. If it crosses \mathcal{P} an even number of times, it is considered to be outside the polygon, or inside otherwise (see Figure 4.7(a)). Calculations are often simplified by choosing the line to be collinear with the x -axis and degenerate cases, namely when a polygonal vertex lies along the line, are dealt with by

treating vertices on the line as though they are slightly above it (this method was first published in [158] and later corrected in [53]).

A second commonly used approach employs the notion of a *winding number* $\omega \in \mathbb{Z}$ of a point with respect to a closed curve. A non-zero winding number indicates that a query point lies within a closed curve. In the continuous case, this quantity is calculated via integration of the differential of the angle resulting from traversal of the curve in a counterclockwise direction (since the curve is closed, integration necessarily yields a value of $2\pi\omega$). In the discrete case, the winding number $\omega = f(Q, \mathcal{P})$ of a query point Q with respect to the polygon \mathcal{P} is the number of revolutions made around Q while visiting each vertex in \mathcal{P} (where $P_n = P_0$). This quantity may be calculated as the sum of the signed angles between QP_i and QP_{i+1} . The query point Q may be considered, without loss of generality, to be at the origin, in which case the winding number may be determined by

$$f(Q, \mathcal{P}) = \frac{1}{2\pi} \sum_{i=0}^{n-1} \arccos \frac{\langle P_i, P_{i+1} \rangle}{\|P_i\| \|P_{i+1}\|} \times \text{sign} \begin{vmatrix} P_i^x & P_{i+1}^x \\ P_i^y & P_{i+1}^y \end{vmatrix}. \quad (4.12)$$

Computing the inverse trigonometric function and the norm in (4.12) are costly operations which make a direct application of this approach prohibitive for polygonal areas defined by many vertices [59, 64].

The interior of a self-intersecting polygon is defined relative to these two methods. In Figure 4.7(a), the interior pentagon is considered to be exterior to the polygon using the parity test. In contrast, the winding number about Q in Figure 4.7(b) is non-zero and so Q is contained within the pentagon.

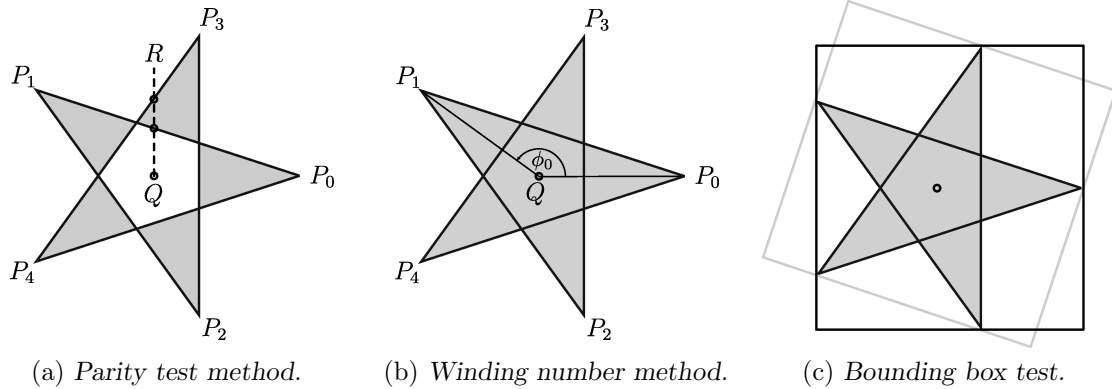


FIGURE 4.7: The interior of the self-intersecting pentagram as determined using (a) the parity test and (b) the winding number method (here $\phi = \frac{4}{5}\pi$ and $\omega = \frac{5}{2\pi}\phi = 2$). (c) Containment may first be tested within the bounding box without the need for all the vertices of the enclosed polygon.

The use of point-in-polygon methods to determine whether a vessel is within the territorial zone is particularly wasteful considering the number of points that need to be evaluated. Alternative approaches, such as oriented areas or closest point tests, however, require some knowledge of which points to test in order to establish whether the vessel has indeed crossed the boundary. An elegant solution is that of space partitioning and the application of strip trees is particularly suitable in this case [3]. Nevertheless, point-in-polygon tests are used in this study and unnecessary evaluations are avoided by first testing the containment of a point within a minimum area oriented bounding box before further calculations are performed⁹. In the case of Figure 4.7(c), the axis-aligned bounding box is a minimal area bounding box.

⁹The C function `lnPoly()` is used to perform the parity test [126].

Rules that are special cases of the two general forms listed in Table 4.4 are presented in Table 4.5. In each case the various decisions within the binary decision diagram still need to be made. The first two rules are easily implemented using the point-in-polygon approach and determining whether the vessel is permitted to be there by virtue of the attributes of the region. The last two rules require the behaviours of fishing and anchoring to be resolved.

Rule	Description
Navigation warning zone	Warnings issued to vessels to avoid dangerous regions
Military vessel alarm	Foreign military vessels within territorial waters without permission
Fishing alarm	Vessel engaging in fishing activity within restricted region
Anchoring alarm	Vessel comes to a stop in a region that forbids anchoring

Table 4.5: Specific zone violation rules.

The four rules of Table 4.5 are discussed in greater detail in the following example.

Example 4.1 (Zone infractions) *The dark grey regions in Figures 4.8 and 4.9 define the exclusion zones in this example.*

Navigation warning zone. *The dark shaded region in Figure 4.8(a) demarcates a region subject to a navigation warning (coastal navigation warning 422 of 2013, issued by the South African Hydrographic Office [123]). All vessels are required to keep clear of this region. The coordinates of the trapezium provide the value of the geometry attribute. As all vessels are requested to keep clear, the prohibited set may be considered to contain all vessels currently within the surveillance scene. Suppose no exceptions are made (resulting in empty sets for the permitted attribute), then any vessel found within the zone will result in an alarm being issued (this is clear from Figure 4.6(a) and (4.6)).*

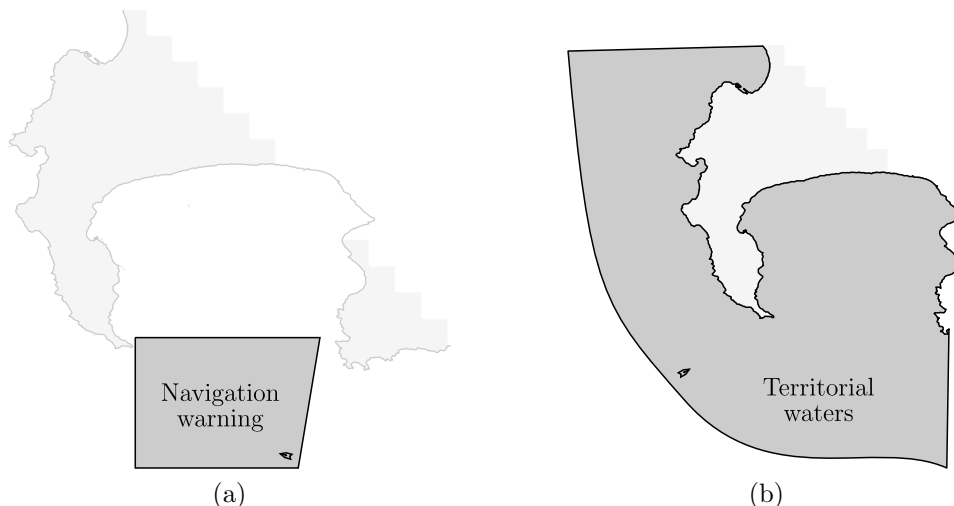


FIGURE 4.8: Geographical areas corresponding to (a) a navigation warning and (b) the territorial waters around Cape Point.

Military vessel alarm. *Suppose the vessel in Figure 4.8(b) is a foreign military vessel. Furthermore, suppose that the relevant authorities have not been notified and that the vessel has thus not been permitted to enter the territorial waters. Using the previously*

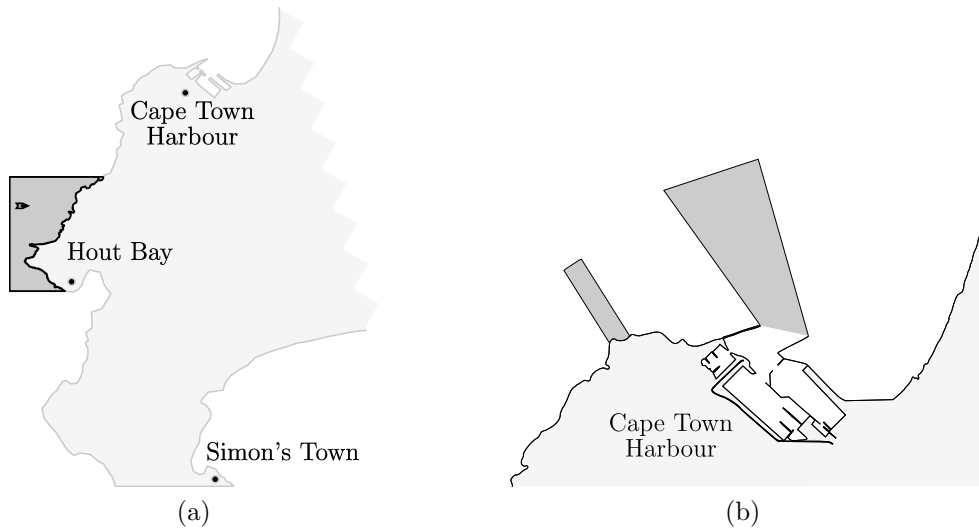


FIGURE 4.9: Geographical regions corresponding to (a) the Karbonkelberg marine protection area and (b) regions near Cape Town harbour in which anchoring is prohibited.

discussed framework, the attributes attached to the territorial waters zone may prohibit military vessels but allow exceptions for specific military vessels permitted to enter South African waters¹⁰.

Fishing alarm. The Karbonkelberg sanctuary is considered in Figure 4.9(a). Suppose that only two commercial fishing vessels v_1 and v_2 are permitted to fish in the region, then the zone features attribute-value pairs such as $(BehaviourProhibited, \{fishing \in \mathbb{B}\})$ and $(BehaviourExceptions, \{(fishing, \{v_1, v_2\})\})$. All other vessels found to be engaging in the activity of fishing would result in an alarm being issued. A fishing activity is deemed to occur when the speed of a fishing vessel drops below a particular threshold for a specified time.

Anchoring alarm. Anchoring is prohibited in the dark shaded regions in Figure 4.9(b). If a vessel is deemed to have come to a stop in these regions (i.e. the speed drops below a specified threshold), and no exceptions are made, then an alarm is issued. In the small vessel case, mooring to a buoy is often prohibited. In this framework a zone may be attributed to the buoy and any vessel found to come to a stop within that zone may be deemed to have moored to the buoy. ■

This formalism of zone infractions includes behaviours since their violation depends on the zone and its attributes. A distinction is made in [146] between alarms generated by behaviours and those generated by entering areas which are off-limits.

4.4.2 Proximity alarms

Proximity rules describe scenarios at sea that involve more than one vessel (proximity to a geographical object is described by zone violation constructions). Using the nomenclature of Dodge *et al.* [30], who provided a taxonomy of motion patterns, the nature of these rules is made clear. The primitive pattern of *co-location* in space (and time) describes the fundamental

¹⁰The South African military vessels would be included in such an exception list.

nature of the proximity rules whilst the compound spatio-temporal patterns of *divergence* and *convergence* describe potential smuggling scenarios and collision events.

Proximity Alarms	Description
Convergence	Vessels converge to same position in space and time
Divergence	Vessels disperse from within close proximity

Table 4.6: *The general descriptions of the proximity rules.*

As was the case with the zone rules, one may wish to exclude certain vessels or vessel types. Furthermore, there are areas in which these rules should preferably not be utilised. For instance, within a harbour area, vessels may often travel in close proximity to one another and the use of these rules in such an area would result in numerous false alarms¹¹. Considering the general structure of proximity alarms, suppose a subset of \mathbb{G} has been selected for querying by a proximity rule, and let InExZ be a proposition expressing the fact that the vessels under consideration lie outside these geographical regions. If it is assumed that Prox is a proposition representing the fact that a proximity threshold has been breached, then

$$\neg \text{InExZ} \Rightarrow \neg A \quad (4.13)$$

$$\text{InExZ} \wedge \neg \text{Prox} \Rightarrow \neg A \quad (4.14)$$

$$\text{InExZ} \wedge \text{Prox} \wedge (\text{VPer} \vee \text{VTPer}) \Rightarrow \neg A \quad (4.15)$$

$$\text{InExZ} \wedge \text{Prox} \wedge \neg(\text{VPer} \vee \text{VTPer}) \Rightarrow A, \quad (4.16)$$

where the meanings of VTPer , VPer and A are as defined before¹². The binary decision diagram representing this class of rules is shown in Figure 4.6(b). The formulation of this rule set does not strictly differentiate between the exclusion of a vessel or a vessel type (as may be seen in (4.15) and (4.16)). This approach is taken as it may be desirable simply to exclude all pilot vessels, for example. However, one need only exclude particular vessels and ignore the vessel type specification if a more finely grained approach is required¹³.

The divergence pattern describes the *vessel deployed* rule of Table 4.7, whilst the *convergence* pattern describes the *rendezvous*, *pursuit*, *raid* and *collision* rules.

A distinction is made between the convergence patterns due to the difference in the kinematic indicators that result in their identification. The rendezvous activity, referred to as *meet-at-sea* by Lane *et al.* [84], is considered first. Three possible scenarios arise when considering a rendezvous of two vessels at sea: a vessel may be approaching an anchored or near-stationary vessel (see Figure 4.10(a)), two moving vessels may be heading directly towards one another (Figure 4.10(b)) or two vessels may be approaching the same future position (Figure 4.10(c)). Importantly, these vessels will be slowing in order to meet the other vessel or the anchored vessel. It is this difference which separates it from the collision rule. Detecting a rendezvous requires a sensitivity to the decrease in relative velocity between two entities. An impending collision

¹¹In the South African context, the harbour limits, as specified by Transnet, may be used to determine the boundaries of a harbour exclusion zone.

¹²It should be noted that the occurrence of a proximity event depends on the interaction between two vessels and that if either vessel lies within the exclusion sets for vessels or vessel types, then an alarm will not be issued. For example, a more accurate description of (4.15), involving vessels A and B , would thus be

$$\text{InExZ} \wedge \text{Prox} \wedge (\text{VPerA} \vee \text{VTPerA} \vee \text{VPerB} \vee \text{VTPerB}) \Rightarrow \neg A.$$

¹³A similar approach may be followed in the implementation of the zone infraction rules. However, it may be desirable to review whether or not a vessel is permitted in the zone before performing further evaluations on the vessel type.

Proximity Alarms	Dependencies or indicators	Exclusions
Rendezvous	Vessel approaching anchored vessel or both slowing	Law enforcement
Collision	Vessels set to closely pass one another	SAR, Pilots, Tugs
Pursuit	Similar headings and pursuer travelling at speed	Hot pursuit
Raid	Vessel approaches large vessel from shore	Law enforcement
Vessel deployed	Vessel deploys smaller craft	Permitted vessel

Table 4.7: A summary of proximity alarms, their indicators and activities which may falsely trigger them.

remains a possible collision regardless of any changes in speed (this is not the case when the heading changes). For this reason, a collision alarm may be seen as a precursor to a rendezvous alarm.

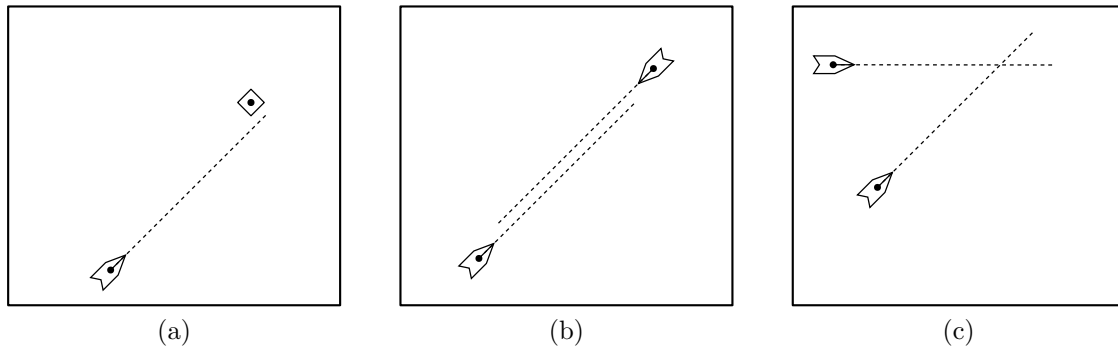


FIGURE 4.10: Convergence patterns that may indicate a collision or a rendezvous. A vessel approaching an anchored vessel is depicted in (a), whilst two vessels on collision courses are shown in (b) and (c).

A possible collision may be determined by performing a linear prediction of the vessels' positions under the assumption of constant continued velocity (the collision rule is applied to sea-faring vessels as smaller vessels have a higher manoeuvrability). This simple method may be used to calculate the *closest point of approach* (CPA) which is required to be greater than a predefined threshold if it is to be determined that the vessel is not in danger of a collision. The closest point of approach may be calculated by using the relative velocity of the two vessels under consideration and fixing it in the local coordinate system of one of the vessels. Suppose two vessels $a, b \in \mathbb{V}$ are at respective positions (p_x, p_y) and (q_x, q_y) , with velocity vectors \mathbf{v}_a and \mathbf{v}_b . Then the relative velocity between the two vessels with respect to vessel a is $\mathbf{v} = \mathbf{v}_b - \mathbf{v}_a$. The position of vessel a is considered to be the origin of the relative coordinate system and the position of vessel b in this system is $\mathbf{s} = (q_x - p_x, q_y - p_y)$. The position over time of vessel b may then be described by

$$\mathbf{r}(t) = \mathbf{s} + \mathbf{v}t$$

within the coordinate system centered at the position of vessel a . The closest point of approach lies at the position for which the magnitude of $\mathbf{r}(t)$ is a minimum. This minimum is achieved at time

$$t_c = \frac{s_x v_x + s_y v_y}{v_x^2 + v_y^2}. \quad (4.17)$$

A better understanding of the measures mariners take to avoid a collision may be gained from the Convention on the International Regulations for Preventing Collisions at Sea (COLREG) [20]. These regulations specify which vessels are expected to give-way, as well as the manner in which these actions should be taken. The vessels on a direct collision course in Figure 4.10(b) would

both take evasive action by steering to starboard (see Figure 4.11(a)). Vessels are encouraged to take decisive action in all collision avoidance manoeuvres so that their actions are clearly observed by the other vessel (be that observation by sight or by radar). If two vessels are expected to pass too closely to one another then the *give-way vessel* is required to perform evasive manoeuvres. The vessel situated to the port side of a second vessel is deemed to be the *give-way* vessel and is required to turn to starboard to avoid collision (see Figure 4.11(b)). The other vessel (called the *stand-on* vessel) may continue in its current direction (unless it observes that the give-way vessel is not acting appropriately). If there is uncertainty about the second vessel's position (*e.g.* heavy fog) then vessels are expected to slow and utilise their fog horns in establishing safe passage. Importantly, these are the only conditions under which vessels are encouraged to reduce their speed in order to avoid a collision.

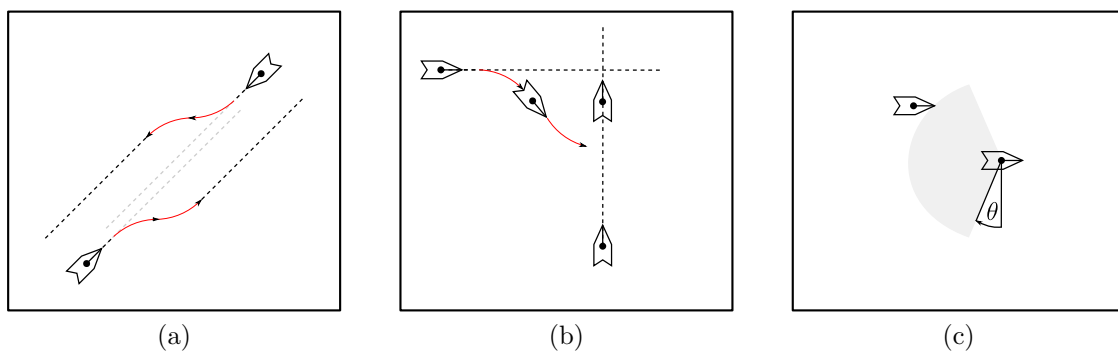


FIGURE 4.11: Actions to be taken in order to avoid (a) a head-on collision and (b) a collision when crossing. A vessel approaching another within the grey region shown in (c), that is, the region 22.5° behind midship, is considered to be passing that vessel.

It is useful to consider the collision rule as composed of strong and weak variants. The strong collision rule determines that a collision will take place if the CPA and the time at which it occurs are both within a stricter predefined interval than that of the weak collision rule. The weak collision rule is thus concerned with activities that are likely to occur over longer periods of time and at greater distances. The reason for this difference is made clearer when considering the remaining convergence patterns in greater detail.

The rendezvous rule relies on the stronger variant of the collision rule and may be expressed as

$$SCol \wedge BSlow \Rightarrow Prox, \quad (4.18)$$

where $SCol$ and $BSlow$ represent respectively the facts that two vessels are on course for a strong collision and are both slowing. It is necessary to consider the velocity updates of at least the previous time in determining whether the vessels are slowing (*e.g.* a moving average, where the window size depends on the update rate, may be used to determine whether this is the case).

The pursuit rule relies on the weaker variant of the collision rule. It is assumed that the pursuing vessel is a small vessel which is non-cooperative (*i.e.* it is not equipped with an AIS transponder). Furthermore, it is further assumed that the pursuing vessel is travelling faster than its target and in the same direction. If $SmVess$, $SmFast$, $RDDec$, $TMove$, and $WCol$ respectively represent the facts that the pursuing vessel is a small vessel, is travelling faster than a specified threshold, that the relative distance between the two vessels is decreasing, that the target is moving, and that a weak collision is detected, then

$$SmVess \wedge SmFast \wedge RDDec \wedge TMove \wedge WCol \Rightarrow Prox. \quad (4.19)$$

The heading of the pursuing vessel is ignored in this formulation (contrary to the pursuit rule proposed in [15]) as its similarity with the target's heading is captured in the collision test¹⁴. In order to avoid false detections it may be desirable to exclude small vessels that are beyond the horizon line of sight of the large vessel (this is not done here).

The primary purpose of the raid rule is to identify vessels departing from the shore on an intercept course with a sea-faring vessel. If the sea-faring vessel is large, it will be unable to change its course or speed significantly in the time that it will take a small vessel that is travelling at speed to reach it. The addition of a concept governing the first appearance of a vessel within the surveillance scene is beneficial in formulating this rule. A vessel may be identified as trustworthy should it emerge from an trusted launch site (for example, many small fishing vessels may launch from a particular site and travel directly out to sea, thus generating a number of false alarms if the site is not accounted for and a collision is predicted with vessels out at sea). Suppose AccO represents the fact a small vessel launches from a trusted location (referred to here as *accepted origin*). Then

$$\text{SmVess} \wedge \neg \text{AccO} \wedge \text{SmFast} \wedge \text{Wcol} \Rightarrow \text{Prox.} \quad (4.20)$$

The divergence rule of vessel deployment makes use of this first appearance concept which essentially provides some indication of trustworthiness if a vessel is deployed from an unconventional location. In this case a vessel is deployed within the vicinity of a larger vessel and moves away from it thereafter. The heading of this vessel does not matter initially, but it may be prudent to take this into account. For instance, a vessel heading to an unconventional location on shore would surely be a better indicator of illicit behaviour than one heading to a trusted location. However, in both instances an inspection may be warranted and particularly so in the latter case where a vessel may attempt to use a trusted location to mask its activities. The introduction of a safety zone centred on the vessel further establishes this activity in the proximity of the vessel. If the fact that a vessel is found to be within such a safety zone is represented by InVZone , then the rule may be expressed as

$$\text{SmVess} \wedge \text{InVZone} \wedge \neg \text{AccO} \Rightarrow \text{Prox.} \quad (4.21)$$

This rule does not preclude the possibility that a vessel may be underway when a smaller vessel is deployed.

Example 4.2 (Proximity alarms) *Each of the rules in Table 4.7 are illustrated here with respect to hypothetical kinematic data.*

Collision. *Suppose a threshold for the CPA is chosen as $C_\tau = 0.6 \text{ km}$, and that the collision rule is active when two vessels are within a 6 km range of one another¹⁵. A vessel positioned at (1, 7), travelling eastwards at a speed of 11 knots, is predicted to have a CPA of roughly 0.36 km with a vessel positioned at (7, 1) and travelling north at a speed of 12 knots (see Figure 4.12(a)). If it is assumed that the identities of these vessels are known and that they are not excluded by virtue thereof, then a collision alarm will be issued. In this case, $t_c \approx 17$ minutes and $\|\mathbf{r}(t_c)\| \approx 0.36 \text{ km}$.*

Suppose the give-way vessel follows the COLREG conventions and takes action to avoid the danger of a close approach (this evasive manoeuvre is illustrated in Figure 4.12(b)).

¹⁴Although the target vessel is explicitly represented by TMove in (4.19), it is also implicitly captured in WCol . This rule may be thought of as being evaluated from the vantage point of the target vessel.

¹⁵These thresholds are based on the work of Lee [10] who posed questions to naval officers about actions taken to avoid collisions. It was found that in good visibility subjects often took evasive action before vessels were within a range of approximately 8 km of one another.

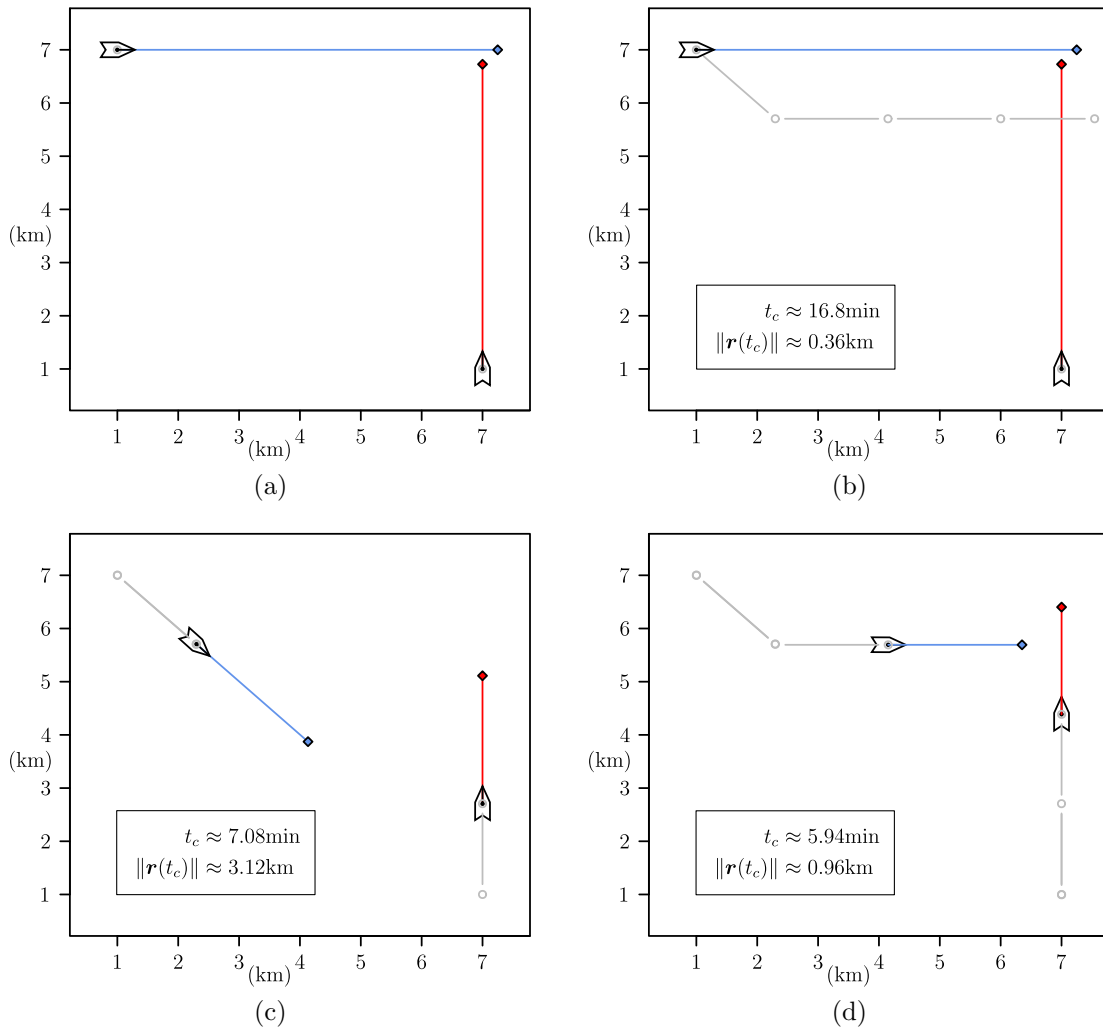


FIGURE 4.12: The closest point of approach of (a) two vessels is indicated at the termination of their predicted linear paths. Should the give-way vessel deem the closest point of approach to be undesirable, it executes an evasive manoeuvre (b) and (c), which, upon return to its original course, (d) has changed the proximity prediction favourably.

Suppose it changes its bearing with respect to the north-south meridian to 45° (see Figure 4.12(c)) whilst the stand-on vessel continues in its current direction. The time to the CPA decreases to $t_{c,1} \approx 7.08$ minutes whilst $\|\mathbf{r}(t_{c,1})\| \approx 3.12$ km. After returning to its original course in Figure 4.12(d), the closest point of approach is $\|\mathbf{r}(t_{c,2})\| \approx 0.96$ km. The collision rule would issue an alarm only for the configuration of Figure 4.12(a) since the CPA is below the specified threshold.

Rendezvous. Consider the scenario depicted in Figure 4.13(a) in which two vessels are on a collision course. A stricter threshold of 0.5 km is chosen for the CPA between the two vessels. Each successive update of the two vessels' tracks reveals that they are both slowing and that the possibility of a collision is not diminishing. This scenario satisfies (4.18) and the vessels are deemed to be engaging in a rendezvous.

Pursuit. Suppose a bulk carrier is travelling at a speed of 13 knots along a course of 45° at time t_0 . Suppose a smaller vessel is detected at the same time at position (0,0) and is deemed to be travelling northwards initially. The small vessel is travelling at a speed of 24 knots and changes its headings over the following two time steps to be travelling in roughly

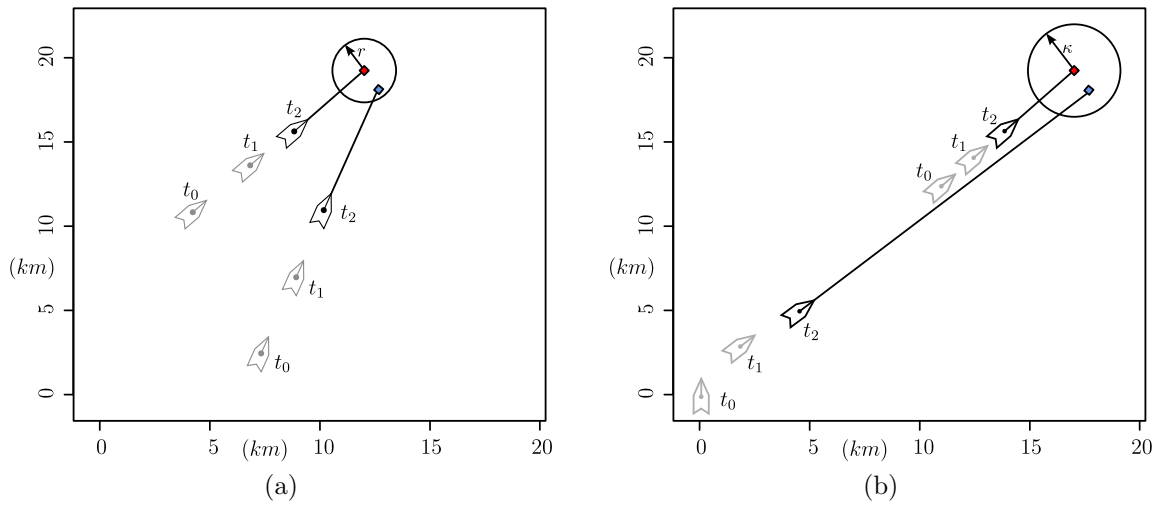


FIGURE 4.13: (a) Two vessels which are both slowing and are determined to have a CPA of r are considered to be engaging in a rendezvous (updates at t_1 and t_2 are used in the calculation). (b) A small vessel travelling at speed is deemed to be pursuing a bulk carrier when the calculated CPA is within a distance of κ kilometres and the distance between the vessels has been decreasing.

the same direction as the bulk carrier. Furthermore, suppose that a weak collision CPA for a bulk carrier is set as $\kappa = 2$ km and that a small vessel is considered to be travelling at speed if it exceeds the threshold of 20 knots. Using the last two kinematic updates, the fact that the distance between the vessels is decreasing is easily determined (this necessitates at least two updates before this calculation can be performed). This will result in (4.19) being true and (4.15) being activated (provided that none of the exclusions of Figure 4.6(b) are satisfied).

Raid. Suppose that a bulk carrier is travelling north at a speed of 13 knots roughly 15 km from shore (see Figure 4.14(a)) and suppose, furthermore, that a small vessel launches from an untrusted location along the coastline and travels at speed directly out to sea. The rectangular area $z \in \mathbb{G}$ in Figure 4.14(a) may be a zone around a harbour which is characterised by an attribute designating it as trusted. A vessel launching from z would result in (4.20) being false (it should be recognised that evaluating AccO could be rendered unnecessary if it is assumed that all accepted origins would be included in a set of zones that are tested during a first evaluation of InExZ in the binary decision diagram of Figure 4.6(b)). In the scenario in Figure 4.14(a), a weak collision is also detected which means that all the conditions of (4.20) are satisfied and the small vessel is deemed to be engaging in a raid on the bulk carrier.

Vessel deployed. Suppose an anchored vessel deploys a small vessel within a few kilometres of the shore (this scenario is depicted in Figure 4.14(b)). The small vessel's first appearance, at time t_0 , is determined to lie within a disk of radius r centered at the anchored vessel. If the anchored vessel is not considered to be a trusted location for the origin of another vessel, then (4.21) is satisfied. ■

Whereas zone infraction rules capture restrictions to activities within static geographical zones and proximity alarms identify coupled vessel behaviour, it is necessary to identify activities by individual vessels that are considered anomalous within the context of the surveillance scene.

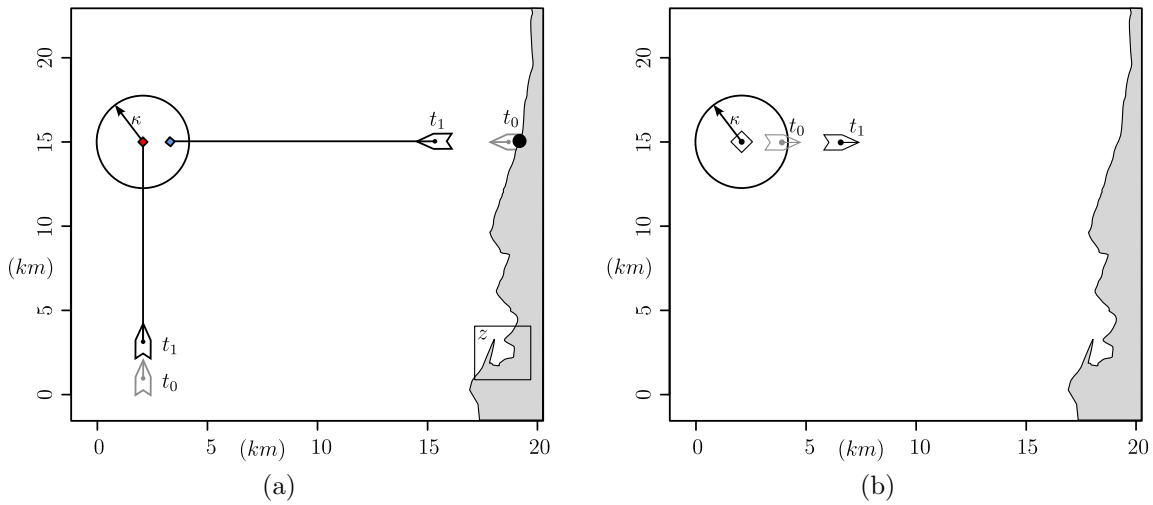


FIGURE 4.14: (a) A small vessel departing from an untrusted location on a collision course with a large vessel is determined to be engaging in a raid activity. (b) A small vessel is first detected within close enough proximity of an anchored vessel that it is deemed to have been deployed by the anchored vessel.

4.4.3 Anomalous actions alarms

A number of activities or actions that are considered anomalous are presented in Table 4.8. These rules typically rely on a single property being in evidence and so their expression using predicate logic is not always necessary. All but the final action presented in Table 4.8 rely on kinematic data and all of these rules apply to large vessels. Even though the emphasis is on kinematic data, there is also a reliance on the additional information provided by AIS transmitters.

Anomalous actions	Description
Call-in failure	Pre-arrival information not submitted to port
Vessel is spoofing	Self-reporting systems presenting false information
Considerable heading change	A vessel changes its heading significantly
Excessive vessel speed	A vessel is travelling faster than is typical
Vessel changed destination	A vessel changes AIS destination port whilst underway

Table 4.8: *Anomalous actions.*

A call-in failure (see §3.4) may be detected by consulting the destination field of a vessel and determining whether, assuming linear motion under constant velocity, the vessel will arrive at its intended destination within a time less than the time required by port authorities to submit a pre-arrival report. The intended destination may be garnered from the voyage information that is transmitted via AIS. If $\text{DestPort}X$ denotes the fact that the intended port of call of a vessel is port X , $\text{In}\tau\text{RangeOf}X$ represents the fact that the expected time of arrival of a vessel at the port is less than some time τ and that $\text{PreArr}X$ describes the fact that a vessel has submitted a pre-arrival report to port X , then

$$\neg\text{PreArr}X \wedge \text{DestPort}X \wedge \text{In}\tau\text{RangeOf}X \Rightarrow A. \quad (4.22)$$

Voyage information transmitted via AIS also provides an estimated time of arrival that may be used in conjunction with the estimate or instead of it.

A spoofing vessel may be detected by comparing the kinematic updates provided by the self-reporting AIS system and measurements acquired from non-cooperative sources such as radar.

If the updates from cooperative data sources differ significantly from those obtained by a radar for a certain duration, then spoofing may be expected. As only a single property needs to be determined to hold true (in a propositional sense), the expression of the rule is omitted.

It is uncommon for a large vessel that is underway to change course significantly. The third rule of Table 4.8 addresses this situation. A predetermined threshold is set for what is considered to be an acceptable change of course whilst a vessel is underway. Suppose $\text{In}\omega\text{Diff}$ represents the fact that a vessel's heading remains within an angular difference of ω with respect to a previous update (or series of updates) and that Move represents the fact that a vessel is underway. Then the anomalous action of a vessel executing a manoeuvre featuring a considerable change of heading may be expressed as

$$\neg\text{In}\omega\text{Diff} \wedge \text{Move} \Rightarrow A. \quad (4.23)$$

The speed at which a vessel travels depends on the type of vessel and on the fuel costs. Vessels on long voyages typically restrict their speeds to economical values and so vessels of the same type operating contrary to this expectation may be considered anomalous. Suppose IsTypeX and ExceedSpX represent respectively the facts that the vessel currently under consideration is of type X and that it is travelling at a speed greater than the commonly observed speed (with the addition of a tolerance) for that vessel type, then

$$\text{IsTypeX} \wedge \text{ExceedSpX} \Rightarrow A. \quad (4.24)$$

Voyage information transmitted via AIS contains a field specifying the destination of a vessel and alteration of this field whilst a vessel is underway is considered to be an anomalous action. An innocuous reason for changing the destination may be that the skipper of a vessel failed to change the field upon leaving the port at which the vessel was previously moored. If DestPortX and Move are as described before (see (4.22) and (4.23)), and DestChanged represents the fact that the destination has just been changed, then

$$\text{Move} \wedge \text{DestPortX} \wedge \text{DestChanged} \Rightarrow A. \quad (4.25)$$

4.5 Summary

A number of rules were proposed in this chapter for inclusion in the rule-based component of the decision support system proposed in §3. These rules were motivated generally as constraints to vessel movement and activities (§4.3). The constraints pertaining to small vessels (§4.2) and sea-faring vessels were discussed specifically (§4.3). Three rule classes were introduced in §4.4, namely zone infractions (§4.4.1), proximity alarms (§4.4.2) and anomalous action alarms (§4.4.3). The zone infraction rules were primarily concerned with vessel behaviour relative to static geographical objects whilst the coupled behaviour of stationary or moving vessels was captured in the proximity alarms. These rules were primarily focussed on kinematic data, but in the case of anomalous alarms, the additional data provided by AIS transmitters in the form of voyage information were used. Although many of these rules are mentioned in the literature (see §2.3.3), the rules described in this chapter are deemed to be of significance in the South African context. Furthermore, methods for determining whether a vessel lies within a polygonal region were discussed and a means to determine the CPA of two vessels was provided. The remaining computational approaches necessary for determining behaviours, such as fishing or anchoring, were not dealt with in this chapter.

CHAPTER 5

Knowledge discovery model component

Contents

5.1	An approach to clustering time-series	77
5.1.1	<i>Dynamic time warping</i>	78
5.1.2	<i>Partitioning around medoids</i>	84
5.2	Discovery component: A filtering approach	90
5.2.1	<i>Density-based clustering for region discovery</i>	91
5.2.2	<i>Outlier removal</i>	97
5.3	Summary	97

Knowledge discovery in databases (KDD) is an active field of research in which the objective is to extract useful knowledge from data. Data mining forms an integral part of the process which is described in [92] as iterative and interactive. The process comprises data acquisition, data preprocessing, data transformation, data mining and evaluation, and finally interpretation. Maimon *et al.* [92] present data mining paradigms as a dichotomy of verification methods and discovery methods; the former is concerned with hypothesis testing whilst the later attempts to discover an hypothesis from within a large set of hypotheses. The process of data mining may include prediction in the form of classification or regression as part of the broader goal.

A clustering or unsupervised approach is pursued in this chapter as a sub-problem of the broader KDD and data mining process. The objective is to provide a mechanism for realising an *origin-destination* miner as proposed in the framework presented in §3.4. To this end, a simple partitioning method that employs a shape matching or sequence alignment measure is investigated as an approach to clustering time-series. A simple origin-destination miner is proposed which utilises a density-based spatial clustering technique to determine origin-destination pairs. Features are derived by which to perform outlier removal so as to aid in the extraction of prevalent tracks within a cluttered surveillance environment.

5.1 An approach to clustering time-series

A situation may often arise in which data elements are not associated with a particular class or target, or that this information is not known *a priori*. Clustering techniques may be used to discover structure within the data in such cases. Clustering, or unsupervised learning, aims to assign similar elements to nearly homogeneous groups. This problem may be stated generally as the problem of partitioning a set of items into non-empty disjoint subsets (fuzzy clustering

is characterised by relaxing the disjointness criterion) [177]. Instrumental to this approach is the definition of a suitable similarity measure. The outcome of a clustering task depends on the chosen clustering technique, as well as the similarity measure.

Clustering methods are overwhelmingly concerned with static data, that is, data for which the feature values do not change over time [88]. These clustering methods may be categorised into partitioning methods, hierarchical methods, density-based methods, grid-based methods and model-based methods [55]. Time-series clustering methods invariably attempt to modify these methods to accommodate time-series data or otherwise the data are converted to static representations so that existing static methods may be employed [88]. The latter approach is followed here. A *dynamic time warping* (DTW) distance measure capable of comparing time-series is employed. This distance measure is utilised by a partitioning method.

5.1.1 Dynamic time warping

DTW is an alignment technique which determines a measure of dissimilarity between two time-series by allowing contractions or dilations of the temporal axis. This distance measure is also often utilised in shape matching¹. The efficacy of the Euclidean distance as a measure of similarity between two time-series is often diminished by distortions in the temporal domain. A calculation of the Euclidean distance between pairwise successive points may result in poor results when the time-series are similar in shape, but differ in their sample rate. Consider, for example, the application of Euclidean distance and DTW in Figures 5.1(a) and 5.1(b), respectively. The illustrated discrete time-series $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$, for $n = 27$, are sinusoidal in shape and differ primarily in that the sections highlighted in grey occur later in X than in Y . Allowing for a distortion of the time axis results in a successful matching of these sections and an improved overall matching of the curves. DTW is more robust than Euclidean distance, which does not perform as well with noise and scaling [49].

The time-series $X = (x_1, \dots, x_q)$ and $Y = (y_1, \dots, y_r)$, where possibly $q \neq r$, are compared by evaluating a local dissimilarity function² and determining the association between the points of each series, subject to global constraints. In the general case where the entries of X and Y are elements of some feature space \mathcal{F} , the dissimilarity function may be defined as

$$d: \mathcal{F} \times \mathcal{F} \mapsto \mathbb{R}_+.$$

Euclidean or Manhattan distance functions are among the distance functions that may be employed in the specific case. Pairwise evaluation of the dissimilarity function yields a local distance or dissimilarity matrix $D \in \mathbb{R}_+^{q \times r}$. A *warping path* $\phi(k)$ of length $\ell \leq \min\{r, q\}$ is a sequence of tuples $\phi(k) = (\phi_x(k), \phi_y(k))$ which index the elements of the matrix and thus associate the elements of X and Y by remapping their time indices [49]. In this way the temporal dimension is expanded or contracted. A warping path is considered valid if it is monotonically increasing (*i.e.* $\phi_x(k+1) \geq \phi_x(k)$ and $\phi_y(k+1) \geq \phi_y(k)$) and if it satisfies a continuity constraint. This constraint takes the form of step rules that dictate which values $\phi(k+1)$ may assume, given the values of $\phi(k)$ and $\phi(k-1)$. An additional boundary constraint, namely that $\phi(1) = (1, 1)$ and

¹Shape matching measures are very common in computer vision settings and are employed in object classification within a scene. These measures typically aim to identify similar objects in which shape is often considered an equivalence class under rotations and translations [154]. DTW is used in this context by transforming a shape to a single time-series. This series may be derived by measuring the length of a rotating ray, positioned at the centroid of a shape.

²A *dissimilarity function* is synonymous with a *distance function* in the sense that the function yields smaller values if objects are closer to one another.

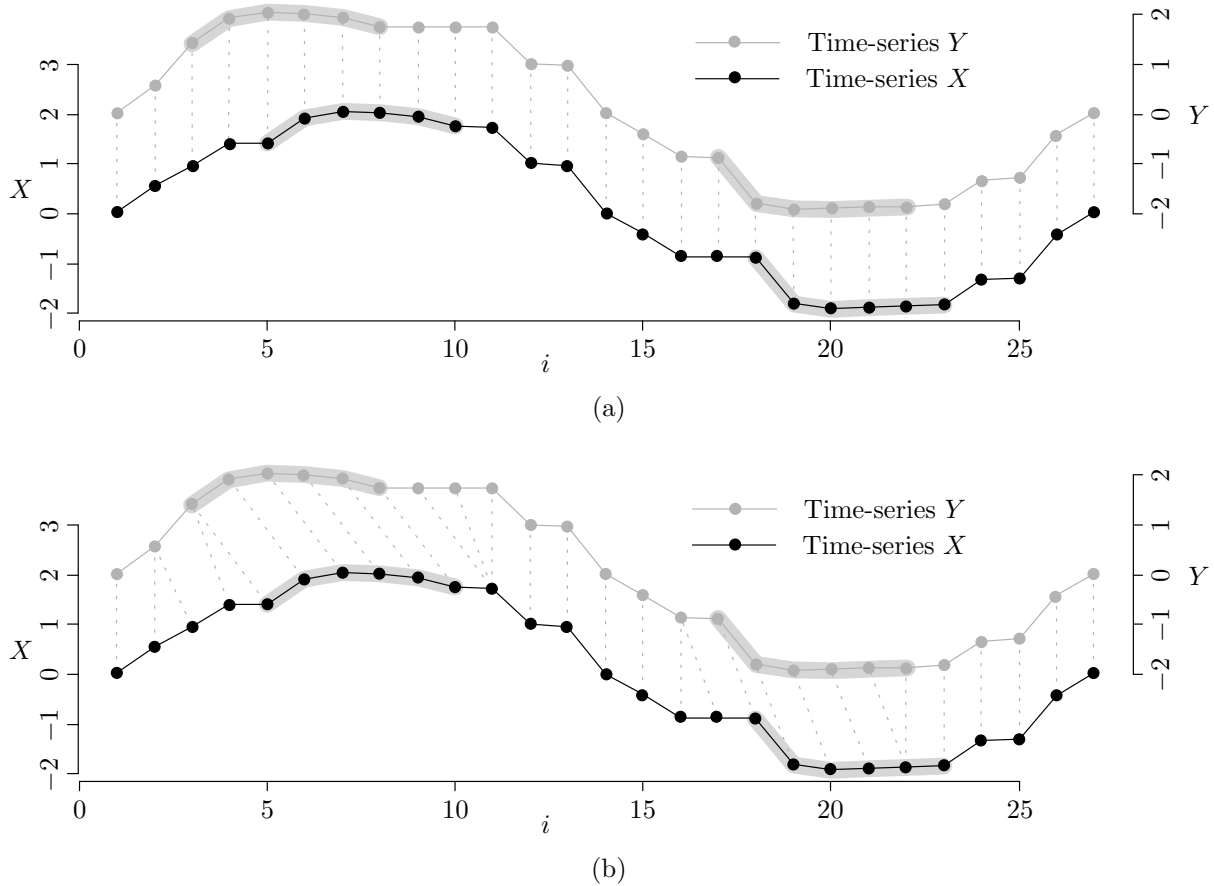


FIGURE 5.1: A comparison of (a) a Euclidean distance matching and (b) a dynamic time warping between two similar sinusoidal time-series X and Y (the latter being offset by 2 along the y -axis). The time-series determined by X and Y are nearly identical but for the sections of the curve indicated in grey and the few points that are altered by the shift along the temporal axis (indicated here by the index i) of these sections. It may be seen that dynamic time warping successfully matches these sections in the two curves.

$\phi(\ell) = (r, q)$, or that the series must be matched from beginning to end, is also often employed (not in the case of partial matching) [49].

The accumulated distance between X and Y may be calculated from ϕ as

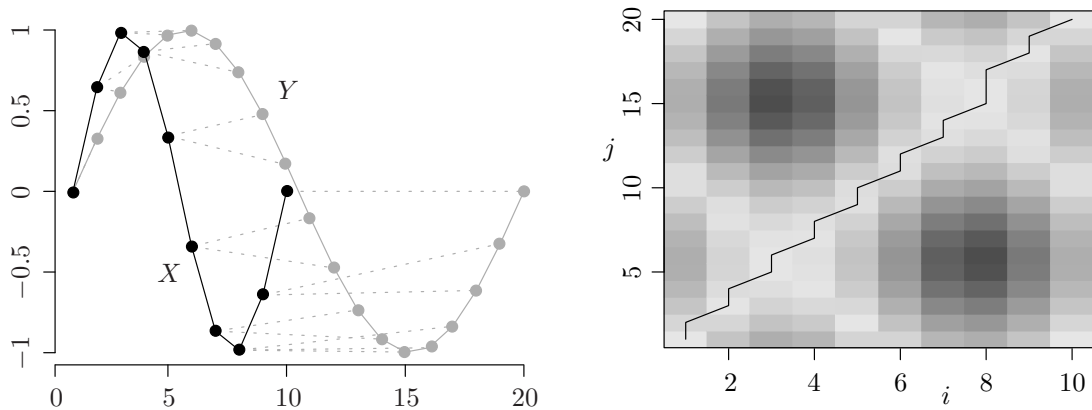
$$d_{\phi}(X, Y) = \sum_{k=1}^{\ell} d(\phi_x(k), \phi_y(k)), \quad (5.1)$$

and an optimal warping path $\phi^*(k)$ is a path of smallest cumulative distance

$$d_{\phi^*}(X, Y) = \min_{\phi} d_{\phi}(X, Y). \quad (5.2)$$

This value of $d_{\phi^*}(X, Y)$, called the *DTW distance*, is typically normalised by the length of the path. An optimal path (that is, a path of minimal cumulative distance) may be computed in $\mathcal{O}(rq)$ time by employing dynamic programming. A recurrence relation is utilised which determines the accumulated distance at a particular step as the sum of the distances to the element currently under consideration, and the minimum of the accumulated distances of adjacent elements [88]. The local distance matrix computed between the curves X and Y in Figure 5.2(a), is presented as an example in Figure 5.2(b). The global dissimilarity matrix which results from the aforementioned process is presented in Figure 5.2(c) (in this case the optimal path is unique).

- Consider two sinusoidal time-series that are identical in shape but differ in their period. Let $X = \sin\left(\frac{2\pi}{9}(i-1)\right)$ and $Y = \sin\left(\frac{2\pi}{19}(j-1)\right)$ where $i = 1, \dots, 10$ and $j = 1, \dots, 20$. The contraction of the time axis results in a warping path that associates the majority of consecutive pairs of points in Y with a single point in X , as shown in Figure 5.3(a). The local cost matrix depicted in Figure 5.3(b) agrees with the expectation that the greatest local distances occur at the points of highest (peak) and lowest (trough) amplitude. The warping path is shown on the dissimilarity matrix where it may be seen that only two values of X fail to map to a pair of points in Y , namely X_8 which is associated with Y_{15}, Y_{16} and Y_{17} , and X_{10} which is associated with Y_{20} .



(a) An alignment of $X = \sin\left(\frac{2\pi}{9}(i-1)\right)$ (black line) and $Y = \sin\left(\frac{2\pi}{19}(j-1)\right)$ (grey line) where $i = 1, \dots, 10$ and $j = 1, \dots, 20$.

(b) The global dissimilarity matrix. The darker the shade of a cell, the larger the distance between the corresponding samples.

FIGURE 5.3: DTW for a pair of sinusoidal curves where one of the curves features twice as many temporal samples as the other.

In both of the aforementioned cases it is assumed that the independent parameter is time and that the curves are sampled equi-temporally. In instances where it is necessary to compare two-dimensional spatial curves, the pairwise distance measure reduces to a Euclidean norm. ■

It is clear that the warping path of two identical time-series will lie along the diagonal of the global cost matrix. As time-series become less similar, this path will move further from the diagonal. A problem arises when computing an alignment that excessively duplicates elements in one time-series in an effort to find an optimal matching. This may commonly occur when the series are of significantly different lengths, if the shape of one of the series favours such a mapping (e.g. the two time-series may be of two significantly different shape classes) or if noisy data are present. Consider the comparison performed in Figure 5.4. This alignment may be considered degenerate or pathological in the sense that approximately 85% of the points are aligned with two points. The warping path associated with this alignment, presented in Figure 5.4(b), lies significantly off the diagonal of the dissimilarity matrix. The implementation of local and global constraints to the warping path make it possible to avoid such degenerate alignments.

A global constraint to the warping path may be instituted by constraining the path to a permissible region in the dissimilarity matrix. These regions are referred to as *warping windows* (see Figure 5.5) and many of them have been proposed in the literature [49]. Three commonly used windows, the *Sakoe-Chiba band*, the *slanted band* and *Itakura parallelogram*, are presented in Figure 5.5(a), 5.5(b) and 5.5(c) respectively. The use of these windows also places restrictions

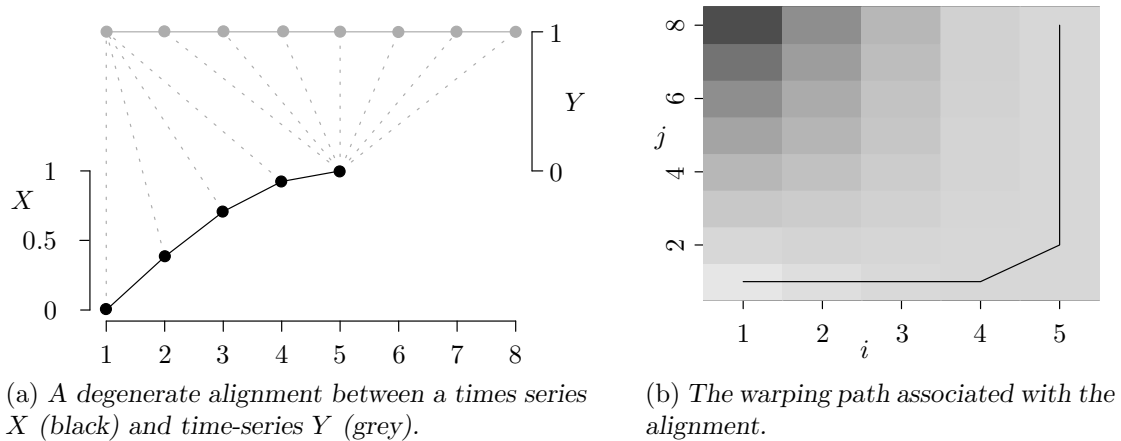


FIGURE 5.4: (a) A poor correspondence is found between the time-series $X_i = \sin(\frac{\pi}{8}(i-1))$ ($i = 1, \dots, 5$) and $Y_j = 1$ ($j = 1, \dots, 8$). (b) The repeated associations with the first point of X and last point of Y are shown in the global dissimilarity matrix.

on the possible disparity in the number of data points per series. For example, the Sakoe-Chiba band runs along the main diagonal and has a fixed width of $w \in \mathbb{N}$. An alignment that adheres to matching the beginning and end points of the two time-series is infeasible in this instance when the lengths of X and Y differ by more than w .

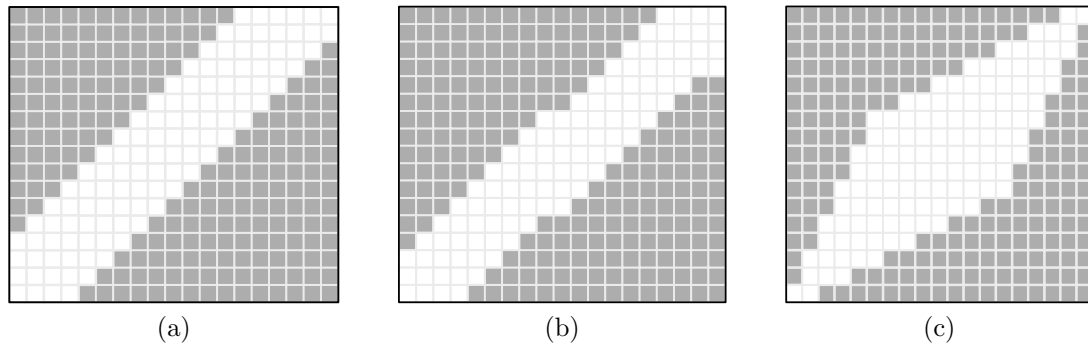


FIGURE 5.5: Three global window constraints, namely (a) the Sakoe Chiba window, (b) the slanted band window and (c) the Itakura window.

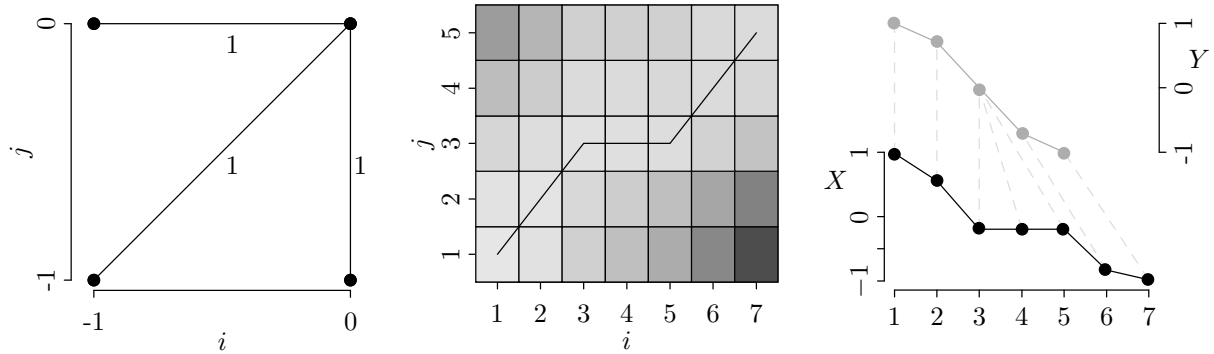
Local constraints may be instituted through the selection of a particular step pattern. These patterns govern the mechanism by which the dynamic programming approach populates the global dissimilarity matrix and their use in conjunction with window widths may result in scenarios where two time-series are incomparable⁴. The time-series $Y_i = \cos(\frac{\pi}{4}(i-1))$ for $i = 1, \dots, 5$ and

$$X = \left(1, \frac{\sqrt{2}}{2}, 0, 0, 0, -\frac{\sqrt{2}}{2}, -1\right)$$

are compared using a step pattern in Figure 5.6(a). This comparison differs from that of (5.3) in Example 5.1 only in the cost associated with matching successive points along the diagonal of the dissimilarity matrix. Nevertheless, the optimal path, shown in Figure 5.6(c), subject to the step patterns in (5.3) and Figure 5.6(a), remains the same. The requirement of matching all points in time-series may be relaxed by using the step pattern of Figure 5.7(a). The resulting

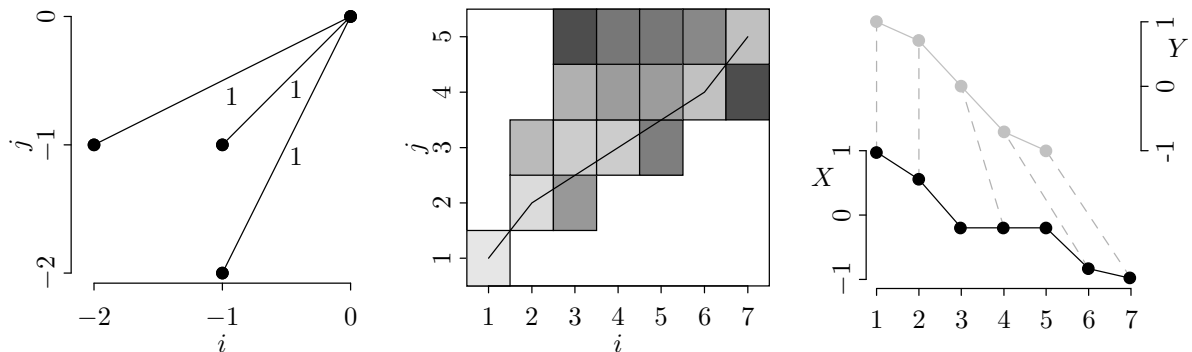
⁴A benefit of these constraints is that the comparison between two time-series may be abandoned without performing unnecessary computations once it is clear that the time-series cannot be compared.

warping path aligning X to Y is presented in Figure 5.7(c) where it may be seen that points along one of the curves remain unaligned.



(a) The step pattern used in computing the values in the dissimilarity matrix with associated costs. (b) The dissimilarity matrix computed via the symmetric step pattern in (a). (c) The alignment resulting from the symmetric step pattern in (a).

FIGURE 5.6: An alignment (c) between $X = (1, \frac{\sqrt{2}}{2}, 0, 0, 0, -\frac{\sqrt{2}}{2}, -1)$ and $Y_i = \cos(\frac{\pi}{4}(i-1))$ for $i = 1, \dots, 5$ which uses (a) a simple symmetric step pattern. The resulting warping path and global dissimilarity matrix is presented in (b).



(a) A step pattern which allows points to remain unaligned. (b) X_3 and X_5 are unmatched whilst all Y_j values are aligned. (c) The resulting alignment with two unmatched points.

FIGURE 5.7: (a) A step pattern which (c) permits points to remain unmatched in the pursuit of (b) an optimal warping path.

Although DTW has grown in prominence in matching univariate time-series, it is just as easily applied in higher dimensions. For example, two objects travelling along precisely the same curve in a Cartesian plane, but at different speeds, would be considered similar in this form of shape matching. Consider the alignment of the two parametric logistic curves⁵ $\mathbf{r}(t)$ and $\mathbf{q}(s)$ in Figure 5.8 where $t = -6, -5, \dots, 5, 6$ and $s = -6, -5.5, \dots, 5.5, 6$. The alignment of these curves is presented in the figure, where the curves lie in the x - y plane and time extends along the z -axis.

⁵The logistic function is defined as

$$f(x) = \frac{1}{1 + e^{-t}}$$

and the parametric curves $\mathbf{r}(t)$ and $\mathbf{q}(t)$ are defined as $(t, f(t))$, for some t .

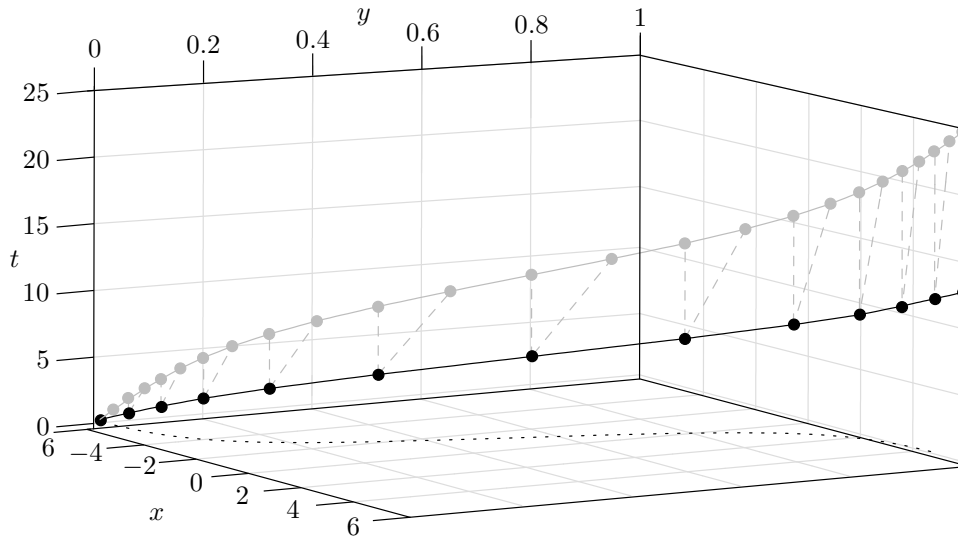


FIGURE 5.8: The alignment computed between two parametric logistic functions $\mathbf{r}(t)$ and $\mathbf{q}(s)$ in \mathbb{R}^2 . The former traverses the parameter interval twice as quickly as the latter. The temporal axis lies along the z -axis.

5.1.2 Partitioning around medoids

The partitioning methods of k -medoids and k -means are closely related. In the method of k -means, the membership of a query object to a cluster is determined by the proximity of that object to the cluster centroid. However, in the method of k -medoids a cluster representative is selected, which is a member of the cluster, and the membership of a query object to the cluster is determined by proximity to this representative. This cluster representative is referred to as a *medoid*⁶ and it may be considered to be located within a cluster, minimizing the average distance from the representative to all other objects within the cluster [79]. The difference between the approach of these methods is illustrated in Figure 5.9. Although the method of computing medoids is sensitive to noise and outliers, these effects are less pronounced than those experienced when computing cluster centroids [177]. Nevertheless, both methods attempt to minimise the average distance between cluster members and their cluster centroids or representatives, thus minimising the average dissimilarity of elements of the same cluster (in the case of k -means this is often achieved by using the squared distance as a distance measure) [79, 177].

These clustering techniques are referred to as *hard* partitioning techniques since an object may only be a member of a single cluster (whereas membership to multiple clusters is possible in fuzzy clustering techniques). The number of clusters is seldom known *a priori* and the methods of k -means and k -medoids are not guaranteed to find a globally optimal clustering. To this end, various values of k (the number of clusters) are typically considered in an attempt to determine an acceptable value of the parameter (this is an example of *model selection* for which cross validation may be used [11]).

The problem of *partitioning around medoids* (PAM) may be described elegantly as a facility location problem in which the locations of a number facilities are to be chosen from among a set of possible locations so as to minimize the total distance from all demand points to their nearest facilities [79]. Determining the number of facilities as well as their location, with respect to a given set of demand points, is an NP-hard bi-objective optimization problem. In the specific

⁶In the one-dimensional case, the medoid of a cluster is equivalent to its median. This is, however, not true in higher dimensions.

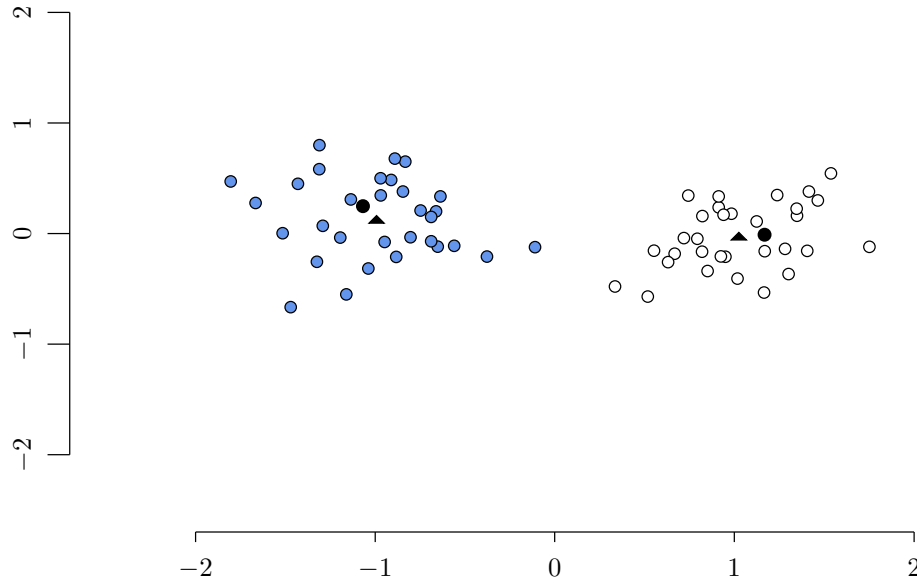


FIGURE 5.9: The k -means and k -medoids clustering techniques applied to clusters generated from two bivariate Gaussian distributions. The centroids are represented by triangles whilst the medoids are distinguished from the remaining points by black circles. Points that belong to the same cluster are similarly coloured (grey or white).

case of k -medoids clustering, the facility locations and demand points are drawn from the same set of elements.

Suppose $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is a set of objects and that the dissimilarity between \mathbf{x}_i and \mathbf{x}_j is denoted by $d(i, j) \in \mathbb{R}_+$. The decision to assign object \mathbf{x}_j to the cluster represented by \mathbf{x}_i is denoted by $\alpha_{ij} \in \mathbb{Z}_2$ whilst the selection of object i as a representative is denoted by $\beta_i \in \mathbb{Z}_2$, where a binary value of 1 for α_{ij} or β_i denotes the affirmative, while the binary value of 0 for these variables denotes the negative. In this optimization problem the objective is to

$$\text{minimise } \sum_{i=1}^n \sum_{j=1}^n d(i, j) \alpha_{ij} \quad (5.4)$$

subject to the constraints

$$\sum_{i=1}^n \alpha_{ij} = 1, \quad j = 1, \dots, n, \quad (5.5)$$

$$\alpha_{ij} \leq \beta_i, \quad \begin{array}{l} i = 1, \dots, n, \\ j = 1, \dots, n, \end{array} \quad (5.6)$$

$$\sum_{i=1}^n \beta_i = k, \quad (5.7)$$

$$\alpha_{ij}, \beta_i \in \{0, 1\}, \quad \begin{array}{l} i = 1, \dots, n, \\ j = 1, \dots, n. \end{array} \quad (5.8)$$

Constraint set (5.5) ensures that \mathbf{x}_j is assigned to a single representative object \mathbf{x}_i , while constraint set (5.6) ensures that this assignment only takes place if the object \mathbf{x}_i is a representative.

Constraint (5.7) fixes the number of representative objects as k , while constraint set (5.8) restricts the values of the decision variables α_{ij} and β_i to binary values. The value of k may either be considered fixed (*i.e.* a parameter whose value is specified) or variable.

It is possible to solve this problem exactly for relatively small data sets and the problem is computationally much cheaper to solve when the value of k is fixed. Various heuristic methods have been proposed in the literature for solving (5.4)–(5.8) with a fixed value of k [120, 156, 177]. In one such approach cluster representatives are selected randomly and these initial estimates are iteratively improved by substituting representatives with non-representative objects in the data set, and determining whether the cluster quality has improved as a result of the substitution. Let the subset $\mathcal{C} \subset \mathcal{X}$ denote the set of cluster representatives (such that $|\mathcal{C}| = k$). Furthermore, let \mathcal{I} and $\mathcal{I}_{\mathcal{C}}$, be index sets such that for $\mathbf{x}_i \in \mathcal{X}$, $i \in \mathcal{I}$ and for $\mathbf{x}_j \in \mathcal{C}$, $j \in \mathcal{I}_{\mathcal{C}}$. Then $i \in \mathcal{I} - \mathcal{I}_{\mathcal{C}}$ for all \mathbf{x}_i in $\mathcal{X} - \mathcal{C}$. The overall cluster quality is determined by evaluating the function

$$J(\mathcal{C}, \mathcal{U}) = \sum_{j \in \mathcal{I}_{\mathcal{C}}} \sum_{i \in \mathcal{I}_{\mathcal{X}-\mathcal{C}}} \gamma_{ij} d(\mathbf{x}_i, \mathbf{x}_j), \quad (5.9)$$

where

$$\gamma_{ij} = \begin{cases} 1 & \text{if } d(\mathbf{x}_i, \mathbf{x}_j) = \min_{q \in \mathcal{I}_{\mathcal{C}}} d(\mathbf{x}_i, \mathbf{x}_q), \text{ for } i = 1, \dots, n \\ 0 & \text{otherwise.} \end{cases}$$

The γ_{ij} values ensure that only the distances from the elements closest to their representatives contribute to the sum. The heuristic approach described above proceeds by swapping an element from the medoid set with an element from the set of remaining objects and computing the change in the function value in (5.9). These $k(n - k)$ swap sets are denoted by \mathcal{C}_{ij} ($i = 1, \dots, k$ and $j = 1, \dots, n - k$), where the element $\mathbf{x}_i \in \mathcal{C}$ is swapped with $\mathbf{x}_j \in \mathcal{X} - \mathcal{C}$, and used in computing the change in the objective function

$$\Delta J_{ij} = J(\mathcal{C}_{ij}, \mathcal{U}_{ij}) - J(\mathcal{C}, \mathcal{U}). \quad (5.10)$$

If $\Delta J_{ij} < 0$, then \mathbf{x}_j replaces \mathbf{x}_i in the global medoid set. The process reaches a local minimum when no further reductions can be made to (5.9) (*i.e.* when $\Delta J_{ij} \geq 0$).

Example 5.2 (k -Medoids clustering)

Suppose the points of $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_5\}$ are

$$(\mathbf{x}_1^T, \dots, \mathbf{x}_5^T) = \begin{pmatrix} 1 & 1 & 1 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}. \quad (5.11)$$

Using the Manhattan metric in two dimensions, namely $d((a_x, a_y), (b_x, b_y)) = |a_x - b_x| + |a_y - b_y|$, the dissimilarity matrix in Figure 5.10(a) is obtained. Furthermore, suppose $\mathcal{C} = \{\mathbf{x}_4, \mathbf{x}_5\}$ are arbitrarily chosen as initial medoids, then

$$J(\mathcal{C}, \mathcal{X} - \mathcal{C}) = d(\mathbf{x}_1, \mathbf{x}_4) + d(\mathbf{x}_2, \mathbf{x}_4) + d(\mathbf{x}_4, \mathbf{x}_3) = 15.$$

The datum \mathbf{x}_5 forms a singleton cluster and does not contribute to the sum.

In order to determine whether this initial clustering may be improved, the six possible swaps are performed between the elements in the medoid set and its exclusion. These are $\mathcal{C}_{41}, \mathcal{C}_{42}, \mathcal{C}_{43}, \mathcal{C}_{51}, \mathcal{C}_{52}$ and \mathcal{C}_{53} . In this case,

$$\begin{aligned} \Delta J_{41} &= J(\mathcal{C}_{41}, \mathcal{U}_{41}) - J(\mathcal{C}, \mathcal{U}) \\ &= (d(\mathbf{x}_1, \mathbf{x}_2) - d(\mathbf{x}_2, \mathbf{x}_4)) + (d(\mathbf{x}_1, \mathbf{x}_3) - d(\mathbf{x}_4, \mathbf{x}_3)) + (d(\mathbf{x}_4, \mathbf{x}_5) - d(\mathbf{x}_1, \mathbf{x}_4)) \\ &= -11. \end{aligned} \quad (5.12)$$

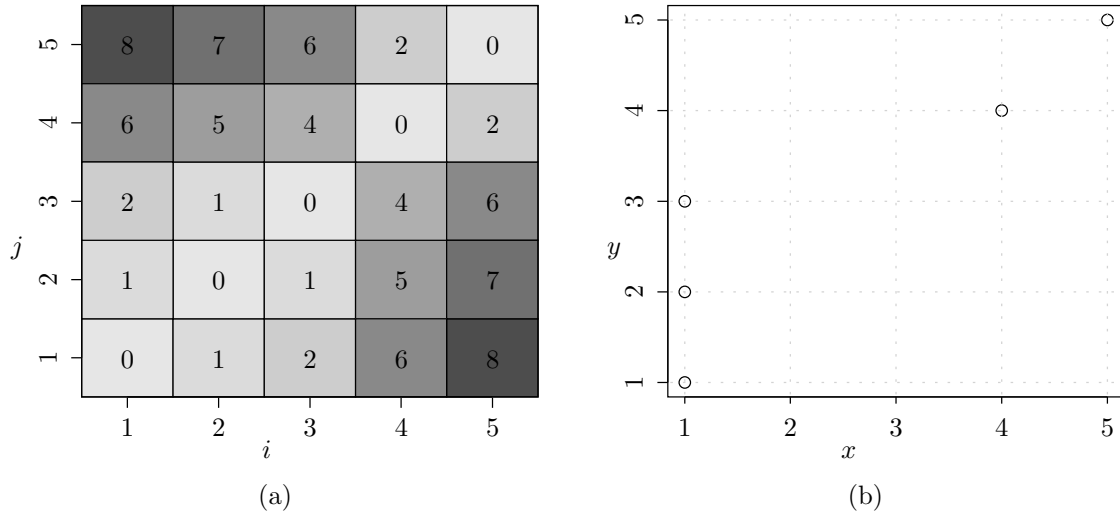


FIGURE 5.10: (a) The dissimilarity matrix, with respect to the Manhattan distance, of the points in (5.11) and (b) their representation in the Cartesian plane.

results in the largest decrease of the function value in (5.9), and so the swap \mathcal{C}_{41} is performed. It may be noted that $\Delta J_{51} = \Delta J_{41}$, $\Delta J_{52} = \Delta J_{42}$ and $\Delta J_{53} = \Delta J_{43}$. This is due to the fact that choosing \mathbf{x}_4 or \mathbf{x}_5 are equivalent as the remaining three points in Figure 5.13(b) are sufficiently distant from either of them.

The medoids that result from this swap operation are $\mathcal{C} = \{\mathbf{x}_2, \mathbf{x}_4\}$. Attempting a second swap operation does not yield a decrease in (5.9) and so the process terminates at a local minimum value of J . ■

The computation of a single ΔJ_{ij} value requires $(k - n)$ evaluations of the distance function whilst each iteration requires $k(k - n)$ computations over each of the possible selections of k elements from the set of n . Therefore, the time complexity of this approach is $O(k(k - n)^2)$, which is quadratic in the number of points to be clustered. This approach is acceptable when n is sufficiently small, but it becomes untenable as n grows. As a consequence, algorithms which make use of randomized sampling are often used in favour of PAM for large data sets⁷ [79].

In a well-known variation of the approach in Example 5.2, the initial k -medoids are selected more systematically [61]. The first medoid, \mathbf{x}_i , is selected as one being closest to all points in the data set. That is, \mathbf{x}_i is selected such that

$$\operatorname{argmin}_i d(\mathbf{x}_i, \mathbf{x}_j)$$

for all $j \neq i$. All subsequent medoids are chosen in such a manner as to be as distant as possible from the initial medoid.

A convenient method of interpretation or validation of the results of PAM is provided by the notion of *silhouette widths* [79]. Silhouette widths are obtained for individual objects by determining the difference between the average dissimilarity between the objects and the objects within the nearest cluster, and the average dissimilarity of an object to all other objects within the cluster to which it is assigned. Suppose \mathbf{x}_i is assigned to cluster \mathcal{C}_k . Then the average

⁷The *Clustering large applications* (CLARA) algorithm achieves a time complexity which is quadratic only in the number of samples. However, on modern computers PAM can comfortably be executed on data sets containing thousands of points.

intra-cluster dissimilarity between \mathbf{x}_i and the remaining $\mathbf{x}_j \in \mathcal{C}_k$ is

$$a(i) = \frac{1}{|\mathcal{C}_k|} d(i, j), \quad \text{for all } \mathbf{x}_j \in \mathcal{C}_k, \quad (5.13)$$

where $d(i, j)$ is the distance measure in (5.9). The cluster \mathcal{C}_ℓ that is considered closest to the element \mathbf{x}_i is determined as the one for which the neighbour cluster dissimilarity

$$b(i) = \min_{\ell \neq k} \frac{1}{|\mathcal{C}_\ell|} d(i, j), \quad \text{for all } \mathbf{x}_j \in \mathcal{C}_\ell. \quad (5.14)$$

Using the average intra-cluster dissimilarity $a(i)$ and the neighbour cluster dissimilarity $b(i)$, the silhouette width for an element \mathbf{x}_i is given by

$$s(i) = \frac{b(i) - a(i)}{\max\{b(i) - a(i)\}}. \quad (5.15)$$

It is clear that the assignment of \mathbf{x}_i to \mathcal{C}_k is arbitrary if $a(i) = b(i)$, and that the membership of \mathbf{x}_i to cluster \mathcal{C}_k is well founded if $b(i) \gg a(i)$.

Example 5.3 (Silhouette plots)

Consider the bivariate data of Figure 5.9 which forms two distinguishable clusters of 32 elements each. The elements of these two clusters are normally distributed. That is, $X \sim \mathcal{N}(\boldsymbol{\mu}_X, \Sigma)$ and $Y \sim \mathcal{N}(\boldsymbol{\mu}_Y, \Sigma)$, where

$$\boldsymbol{\mu}_X = (-1, 0), \quad \boldsymbol{\mu}_Y = (1, 0), \quad \text{and } \Sigma = 1/3\mathbf{I}$$

and $\mathbf{I} \in \mathbb{R}^{2 \times 2}$ is the identity matrix. The silhouette plot for these data is presented in Figure 5.11(a). The average intra-cluster silhouette widths, denoted by $\hat{s}(j) = \sum_{i \in \mathcal{C}_j} \frac{1}{|\mathcal{C}_j|} s(i)$, are $\hat{s}(1) \approx 0.67$ and $\hat{s}(2) \approx 0.73$. The point $X = (-0.1097, -0.1213)$ in Figure 5.9 has the smallest silhouette width of approximately 0.14. It is clear that the points are easily separable in this case and the large $\hat{s}(i)$ values confirm this (the average of all silhouette widths is approximately 0.7)⁸.

Suppose the variance of the two distributions is changed to $\sigma^2 = 2/3$ (see Figure 5.11(b)). The data are then clustered into two clusters \mathcal{C}_1 and \mathcal{C}_2 with $|\mathcal{C}_1| = 28$ and $|\mathcal{C}_2| = 36$. In this case the average silhouette width over all the data points diminishes to 0.48, while $\hat{s}(1) \approx 0.46$ and $\hat{s}(2) \approx 0.49$. ■

The k -medoids approach is well-suited for use in conjunction with DTW since it does not require computation of the average time-series⁹. The pairwise DTW distances between time-series X_1, X_2, \dots, X_n (see (5.2)) may be precomputed to determine a DTW distance matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that

$$\Lambda_{ij} = d_{\phi^*}(X_i, X_j).$$

This distance matrix may then be utilised as a look-up table for the PAM algorithm.

⁸An average of the intra-cluster silhouette widths, $\hat{s}(i)$ for all i , which is greater than 0.5 is considered to be a reasonable clustering [79]. As this value decreases it becomes less sensible to speak of the data as being clustered in such a way that its elements may be said to fall definitively within a particular cluster.

⁹An approach to computing the average time-series may be found in [56].

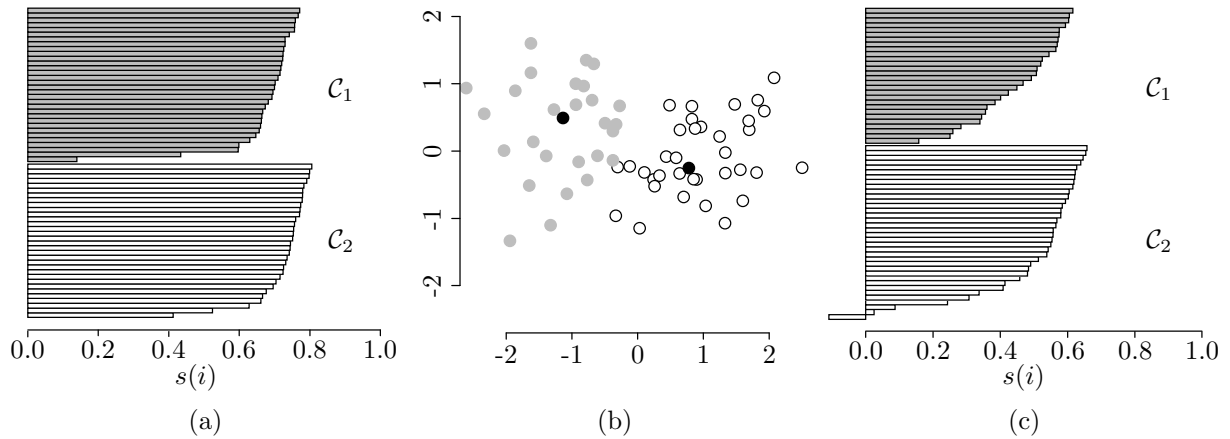


FIGURE 5.11: (a) The silhouette plot resulting from clustering the data of Figure 5.9 using PAM for $k = 2$. (b) Data with a higher variance is clustered into two clusters by PAM and (c) the accompanying silhouette plot illustrates the poorer discrimination between the clusters.

Example 5.4 (DTW and PAM)

Consider the data of Figure 5.12(a), comprising three different directions along which objects travel at a particular speed, or twice as slowly. Objects may continue travelling in the direction in which they are travelling or they may diverge to travel in a north-easterly or south-easterly direction. Suppose the trajectories in these directions are represented by \mathcal{S} , \mathcal{U} and \mathcal{D} , respectively (as indicated in Figure 5.12(a)), that each set has a cardinality of four and that in each set two of the objects are travelling twice as fast as the remaining two.

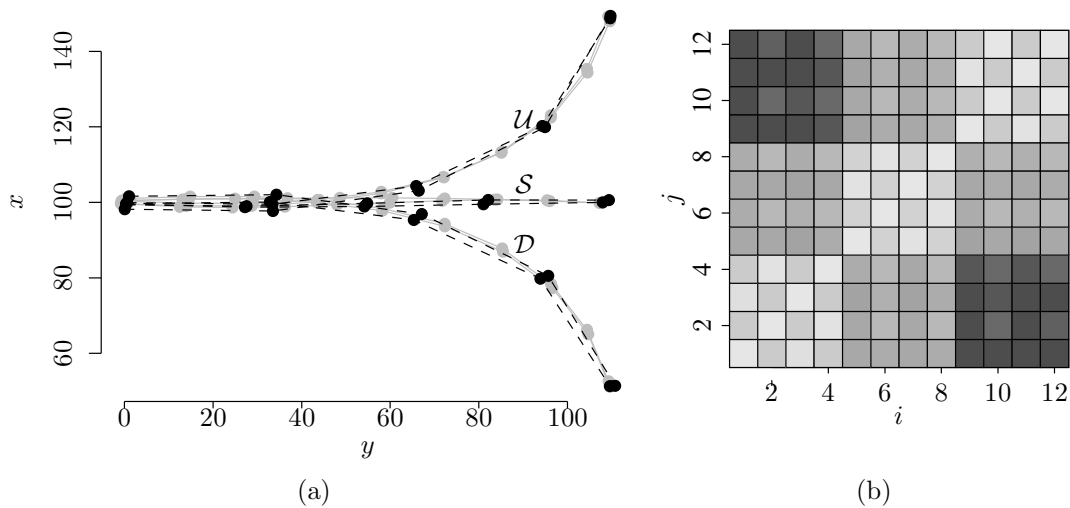


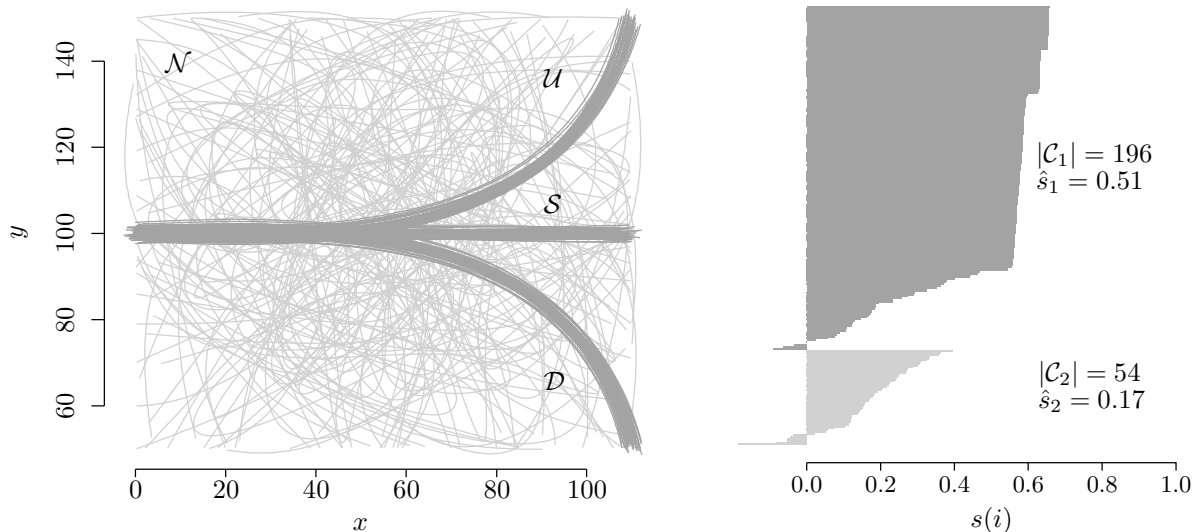
FIGURE 5.12: (a) Twelve two-dimensional time-series and (b) their corresponding DTW dissimilarity matrix Λ .

The dissimilarity matrix, $\Lambda \in \mathbb{R}^{12 \times 12}$, that results from pairwise application of DTW is presented in Figure 5.12(b). The time-series are ordered according to their membership to one of the three sets \mathcal{S} , \mathcal{U} or \mathcal{D} and alternate in length. The resulting shading of the matrix reveals that there is a clear distinction under the DTW measure and PAM successfully recovers the three clusters with a resulting average silhouette width of 0.65. ■

Using PAM in conjunction with DTW to cluster time-series of relatively similar shape and number of samples has been shown to work satisfactorily in the simple example above. However, the situation of structured data occurring within less structured data is of more interest as the clustering of such data is expected to require additional data processing.

5.2 Discovery component: A filtering approach

If a time-series is considered to represent the trajectory of a particle moving through space, then the method of alignments by DTW may be considered to measure whether particles moving through the same space at different rates (within some reasonable tolerance) are travelling according to similar time-series. These trajectories may typically occur within a cluttered data set which requires outlier identification and removal. It is clear from the first simple PAM clustering of Figure 5.9, that the data are clearly separable. However, in the event that spurious points clutter the observation space, PAM is expected to yield a poor clustering (for the simple reason that it will attempt to cluster all data¹⁰). To this end, outlier removal is most often a precursor to performing clustering on a data set. In the data mining methodology proposed in [92], this step falls within the data-preprocessing stage. It is noted that outlier removal need not only be applied to the raw data (for example, on the positional data used in the previous examples), but may also be performed on features that are derived from the data.



(a) A data set containing regular and irregular time-series.

(b) The silhouette plot resulting from the application of PAM with $k = 2$.

FIGURE 5.13: (a) A data set containing regular (\mathcal{U} , \mathcal{S} and \mathcal{D}) and irregular time-series (\mathcal{N}) and (b) the resulting silhouette plot when applying PAM with $k = 2$.

Consider the data set in Figure 5.13(a). For the sake of discussion it is assumed that the data comprise four types of trajectories, the highly regular sets \mathcal{U} , \mathcal{S} and \mathcal{D} , which contain time-series progressing in north-easterly, easterly, and south-easterly directions respectively, and the irregular set \mathcal{N} . The irregular set features a hundred time-series whilst the remaining three data sets each has a cardinality of fifty. An attempt to uncover this structure by direct application of PAM with $k = 4$ and using DTW as the dissimilarity measure, results in an approximate average silhouette width of 0.32. Determining the number of clusters k that maximises the

¹⁰This is observed in [34] when using PAM to cluster vessel trajectories.

average silhouette width (approximately 0.43) results in two clusters being identified¹¹. This result is intuitively appealing as it appears to be discriminating between the regular and irregular data. The discovered clusters, say \mathcal{C}_1 and \mathcal{C}_2 , have cardinality of 196 and 54 respectively (Figure 5.13(b)). The latter cluster is particularly poorly supported and although it contains only time-series from the irregular data set, it fails to contain them all. If an outlier removal approach is followed where all time-series below a particular threshold are discarded, then it is possible to retain only the structured data (the lower bound for discarding all the irregular data is 0.55). Although this approach may appear attractive, its use is not easily justified as the silhouette widths of the individual time-series depend on a particular clustering and it may require a few iterations of this process to identify poorly performing data with respect to some value of k . There is still no guarantee that a considerable portion of the structured data will not be discarded during this process¹².

The above approach is naive but illustrates the difficulty of clustering noisy data without performing outlier removal. The definition or purpose of the clustering or data mining task is integral in determining whether the results are valid. The task in the application considered in this dissertation is to discover predominant structures within trajectory data that are spatially relatively compact. All but the set \mathcal{N} are considered the target of the above data mining exercise and it is shown in this section that a simple and robust approach to mining these time-series is achievable through extraction of the origin-destination pairs of each time-series, or more generally, regions of interest.

5.2.1 Density-based clustering for region discovery

Determining a clustering of objects by virtue of a measure of collective proximity is precisely the notion that was evident in the discussion of k -means and k -medoids in §5.1.2. However, a disadvantage of these approaches is that all points need to be clustered. An approach to circumventing this problem may be to perform an initial clustering and then to discard points that are distant enough according to some measure, or to discard distant points during the clustering process. An example of the former approach might be clustering via non-parametric density estimation whereby a function which fits the data is computed as the sum of kernel functions which are centered on each datum. A threshold value may then be chosen which allows for the data to be divided into clustered and non-clustered data [2]. An example of the latter method is in use in the database community and relies on an iterative approach to constructing clusters in which points are discarded during the clustering process if they do not have sufficient numbers of observations within some predefined proximity. This method was proposed by Ester *et al.* [37] and is known as the *density-based spatial clustering of applications with noise* (DBSCAN) method.

DBSCAN is well suited to large data sets and is capable of discovering clusters of arbitrary shape¹³ [37]. In order to quantify density it is necessary to provide a threshold value for differentiating between dense and non-dense regions. DBSCAN achieves this by employing the notion of an ε -neighbourhood of a point, and the number of elements that this neighbourhood contains, denoted by ρ . A point is deemed to belong to a cluster if its ε -neighbourhood contains ρ points.

¹¹Incidentally, if this approach is followed in Example 5.4, then six clusters are favoured since then the average silhouette width is approximately 0.84.

¹²In this example, selecting an average silhouette width threshold of 0.57 discards the majority of the time-series in \mathcal{D} whilst retaining \mathcal{U} and \mathcal{S} . Indeed, performing the same calculations on four similarly constructed data sets of equal cardinality, namely 100, results in $k = 10$ clusters if the objective is to maximise silhouette width.

¹³The methods of k -means and k -medoids are capable of discovering convex clusters whilst Gaussian mixture models [11] and spectral clustering [194] are examples of methods that are able to find non-convex clusters.

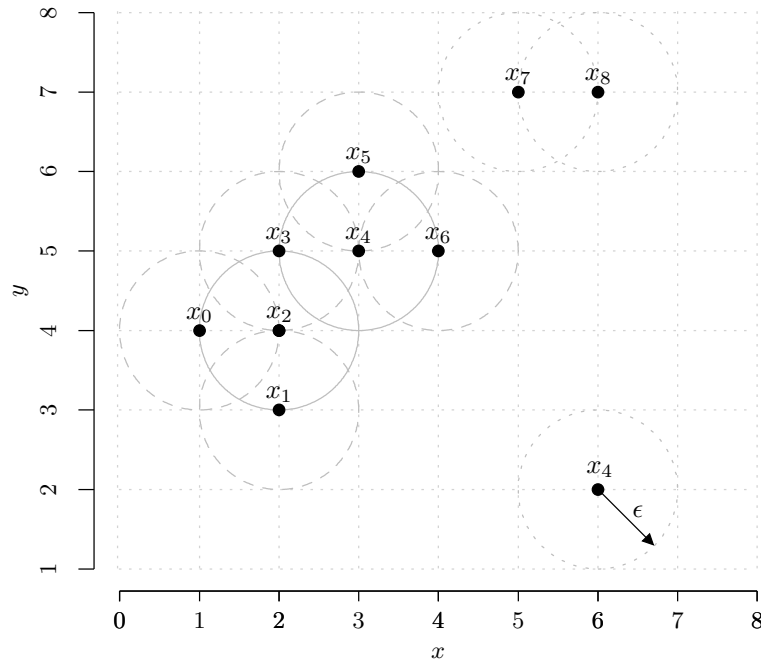


FIGURE 5.14: A single cluster is found amongst a set of points in the Cartesian plane using the Euclidean distance metric with $\varepsilon = 1$ and $\rho = 4$. The core point neighbourhood boundaries are depicted using solid lines, boundary points using dashed lines, and noise points using dotted lines.

Using these parameters, Ester *et al.* [37] distinguish between *noisy*, *boundary* and *core points*.

For some distance measure $d(x, y)$ over a data set \mathcal{X} , the neighbourhood of a point $p \in \mathcal{X}$ is

$$N_\varepsilon(p) = \{q \in \mathcal{X} \mid d(p, q) \leq \varepsilon\}. \quad (5.16)$$

If $|N_\varepsilon(p)| \geq \rho$, then p is considered a *core point*. A point $q \in \mathcal{X}$ is considered to be *directly density-reachable* from p , if $q \in N_\varepsilon(p)$ and p is a core point. Consider, for example, the points presented in Figure 5.14 and assume a Euclidean distance metric with $\varepsilon = 1$ and a support of $\rho = 4$. Then x_0, x_1, x_3 and x_5, x_6 are boundary points which are directly density-reachable from the core points x_2 and x_4 , respectively. This relationship is not symmetric since the core points are not directly density-reachable from any of the boundary points. However, this is not the case for core points and sequences of core points that are pairwise directly density-reachable. In this manner, an arbitrary sequence of points in which each point lies in the neighbourhood of the preceding point establishes a relationship between the initial and final point of the sequence. The final point is said to be *density-reachable* from the initial point if such a sequence of points exists (*e.g.* x_6 in Figure 5.14 is density-reachable from x_2). Furthermore, Ester *et al.* [37] introduced the symmetric and transitive relation of density-connectedness in which boundary points may be related to one another. Two points $p, q \in \mathcal{X}$ are *density-connected* if there is a datum $c \in \mathcal{X}$ such that p and q are density-reachable from c (*e.g.* x_6 is density-connected to x_1).

A cluster, in the context of the DBSCAN algorithm, is defined in terms of the notions of density-reachability and density-connectedness. A cluster $\mathcal{C} \subseteq \mathcal{X}$ has the properties that an arbitrary point within the cluster is density-reachable from any other point in the cluster, and that all points in the cluster are density-connected (with respect to $\varepsilon, \rho \in \mathbb{R}$)¹⁴. Points that do not

¹⁴Although this is the definition provided by Ester *et al.* [37], it may be noted that the requirement that all points should be density-reachable from one another is already captured in the property of being density-connected. Consequently, it may be said that a collection of points form a cluster if and only if all elements in it

fall within a cluster are considered to be noise. The points $\mathbf{x}_7, \mathbf{x}_8$ and \mathbf{x}_4 in Figure 5.14 are considered to be noise.

The computation of clusters by the DBSCAN algorithm is performed iteratively. A core point x that has not yet been visited acts as the seed to a new cluster. If the neighbourhood of x contains more than ρ points, then all points that are density-connected to x are determined and assigned to the cluster. Otherwise, x is considered to be noise. It may be that a boundary point was considered first and that, since no points are density-reachable from it, it may initially be classified as noise. However, if a core point which is density-connected to it is selected during future iterations, this status will be changed and the boundary point will be included in a cluster. Visiting all points in a data set containing n elements, and performing range queries subject to ϵ , results in a worst-case time complexity of $O(n^2)$. Using space indexing techniques, such as R^* -trees¹⁵, this time complexity may, however, be lowered to $O(n \log n)$.

The DBSCAN algorithm¹⁶ proceeds by starting with an arbitrary point and determining the set of neighbours within a radius of ϵ . If the number of neighbours is greater than ρ , then this point is a core point and acts as the seed of a cluster. Each of these neighbourhood points is then evaluated in a similar fashion to determine whether they are themselves core points or boundary points. If a point does not satisfy the core point conditions, then it is labelled as noise and another point is considered as a potential seed. This process continues until all the points in the data set have been visited.

Example 5.5 (DBSCAN) *DBSCAN is applied to two data sets in this example. The iterative nature of the algorithm is first explored, after which the allocation of clusters that feature fewer than the required number of points is demonstrated.*

(a) Suppose that $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{14}\}$, where

$$(\mathbf{x}_1^T, \dots, \mathbf{x}_{14}^T) = \begin{pmatrix} 1 & 2 & 2 & 2 & 3 & 3 & 4 & 5 & 6 & 6 & 6 & 6 & 7 & 7 \\ 4 & 3 & 4 & 5 & 4 & 5 & 1 & 6 & 2 & 3 & 5 & 6 & 3 & 4 \end{pmatrix}. \quad (5.17)$$

Furthermore, suppose that no spatial indexing has been performed and that the points are ordered lexicographically, as presented in (5.17). For $\epsilon = \sqrt{2}$ and $\rho = 4$, the algorithm begins by considering \mathbf{x}_1 . The neighbourhood $N(\mathbf{x}_1) = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ is determined and the cluster \mathcal{C}_1 is initialised with \mathbf{x}_1 since it is a core point and has not yet been allocated to a cluster. Each of the points in the neighbourhood of \mathbf{x}_1 are evaluated (since they have not yet been visited) in a similar fashion to determine whether additional points may be added to the cluster. This iterative evaluation results in the cluster $\mathcal{C}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_6\}$ which contains only core points (this is illustrated in Figure 5.15a). Evaluating the point \mathbf{x}_7 reveals that its neighbourhood contains too few points for the point to seed a cluster and so it is labelled as noise. The same is true for \mathbf{x}_8 , and the point which is considered next, \mathbf{x}_9 . Evaluation of \mathbf{x}_{10} reveals the point to be a core point and all the points in its neighbourhood are evaluated. The label of \mathbf{x}_9 is changed to reflect that it is a member of the newly formed cluster \mathcal{C}_2 , by virtue of being in the neighbourhood of a core point¹⁷. The remaining points in $N(\mathbf{x}_{10})$, namely \mathbf{x}_{13} and \mathbf{x}_{14} are both core points which are added to the newly seeded cluster, \mathcal{C}_2 . There is no need to evaluate the neighbours of \mathbf{x}_{13}

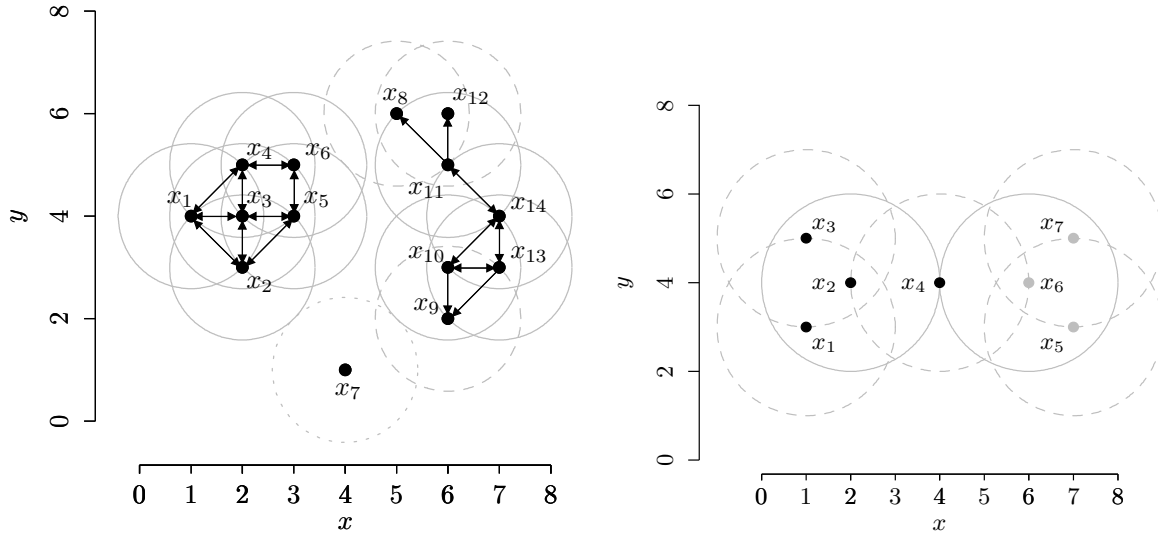
are density-connected.

¹⁵An R^* -tree is a common spatial indexing scheme that is similar to a KD -tree in that points are stored according to the subdivisions of space into which they fall.

¹⁶The DBSCAN implementation in the R “fpc” package is used in this dissertation [61].

¹⁷If a boundary point is contained within the neighbourhood of a core point, then it is added to the cluster (but there is no need to evaluate its neighbours iteratively).

iteratively as they have all been evaluated, but these evaluations continue for $\mathbf{x}_{11} \in N(\mathbf{x}_{14})$ and $\mathbf{x}_8, \mathbf{x}_{12} \in N(\mathbf{x}_{11})$. Thereafter, all the points in the data set have been visited and two clusters, $\mathcal{C}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_6\}$, $\mathcal{C}_2 = \{\mathbf{x}_8, \mathbf{x}_9, \dots, \mathbf{x}_{12}\}$ and a single noise point, \mathbf{x}_7 , have been identified subject to the chosen values of ϵ and ρ .



(a) The boundaries of the neighbourhoods for core points, boundary points and noise points are presented as solid, dashed and dotted lines, respectively.

(b) Two clusters (the points of which are distinguished by black or grey filled circles) that share a common boundary point are identified by DBSCAN. The neighbourhood boundaries of the points follow the same convention as in (a).

FIGURE 5.15: DBSCAN applied to (a) the fourteen points in (5.11) subject to the parameter values $\epsilon = \sqrt{2}, \rho = 4$ and (b) the nine points in (5.18), subject to the parameter values $\epsilon = 2$ and $\rho = 4$.

(b) Suppose next that $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_7\}$, where

$$(\mathbf{x}_1^T, \dots, \mathbf{x}_7^T) = \begin{pmatrix} 1 & 1 & 2 & 4 & 6 & 7 & 7 \\ 3 & 5 & 4 & 4 & 4 & 3 & 5 \end{pmatrix}, \quad (5.18)$$

and that $\epsilon = 2$ and $\rho = 4$. The algorithm proceeds as before and \mathbf{x}_2 is discovered as the first core point — all of its neighbours are assigned to the cluster that it seeds (as illustrated in Figure 5.15(b)). However, once \mathbf{x}_6 has been considered, it too is found to have four neighbours, but \mathbf{x}_4 is a boundary point and has already been attributed to another cluster. This core point still results in the creation of a cluster although \mathbf{x}_4 is not attributed to it. This is consistent with the formulation of a cluster in the context of DBSCAN. ■

The notion of using origin or destination as additional features in extracting similar trajectories from a scene is an approach which has been applied successfully in motion pattern analysis [142, 115] (see §2.3.3). De Vries *et al.* [23], for example, further enrich trajectory kinematic data by associating subtrajectories with the geographical regions through which they pass, but also include origin and destination knowledge. The approach followed by Morris *et al.* [115] uncovers entry and exit zones in traffic surveillance scenes. In this instance, DBSCAN is used to identify origin and destination pairs in the data, but also to provide a filtering mechanism by which to extract trajectories that share common regions in space.

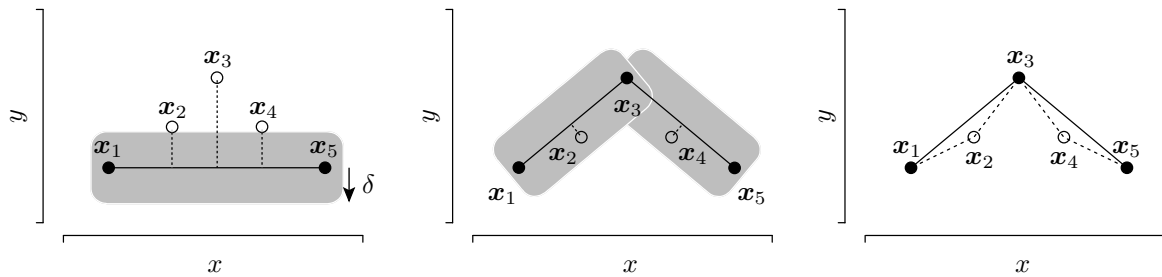
Consider again the data set of Figure 5.13(a). It is sufficient and useful to consider simplified incarnations of these trajectories. Trajectory segmentation or simplification may be used to this

end in reducing the number of points of a piecewise linear path. Such approaches are common in simplifying GIS computations, facilitating visualisation of GIS data and in trajectory resampling in motion pattern analysis. Two simple such subsampling techniques are that of *Douglas-Peucker* (DP) [58] and radial distance poly-line simplification.

The DP method approximates a path by a straight-line segment joining the first and last point of the path (Figure 5.16(a)). If all remaining points lie within a predefined distance from this line segment, then this approximation is accepted. Otherwise, a point which lies furthest from the line is selected and the method is applied recursively to the subsequences resulting from it (Figure 5.16(b) and Figure 5.16(c)).

The radial distance method is a brute-force approach which, starting with the first point of a path, removes all subsequent points that lie within a specified radial distance of the point that is currently under consideration. Thereafter, the next point in the sequence which has not been subjected to such removals, becomes the focal point and the process continues until the final point is reached.

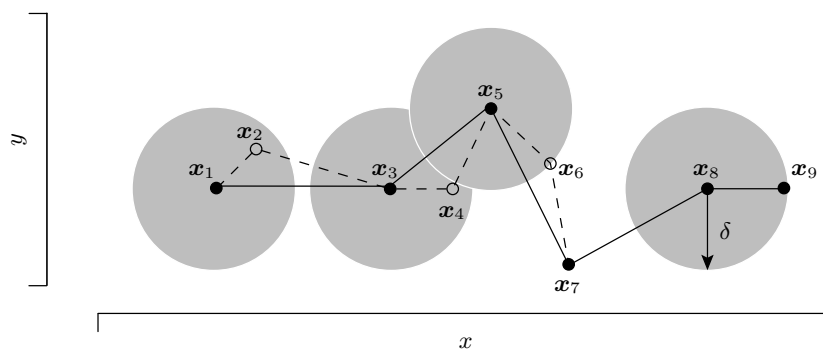
Both these approaches include the start and end point in the final trajectory and may be susceptible to including spurious points if they have not yet been removed from the data set.



(a) The point x_3 lies the furthest from the line segment joining x_1 and x_5 .

(b) The polyline is recursively simplified, and x_2 and x_4 are discarded as they lie within the specified tolerance.

(c) The simplified polyline contains x_1, x_3 and x_5 .



(d) The points x_2, x_4 and x_6 all lie within the closed neighbourhood defined by their predecessors and the tolerance of δ , and are discarded.

FIGURE 5.16: (a)-(c) The DP method applied to five data points with a tolerance of δ . (d) Radial distance thinning applied to nine data points with a tolerance of δ .

Example 5.6 (Trajectory mining using DP and DBSCAN)

Consider the data set of Figure 5.13(a) once more and suppose that the trajectories are reduced using the DP method with a threshold of $\delta = 10$. The resulting points are overlaid on the

original trajectories in Figure 5.17(a) and are processed by DBSCAN as independent points. DBSCAN identifies 5 spatially dense clusters, shown in Figure 5.17(b), subject to the parameter values $\epsilon = \delta$ and $\rho = 32$. Incidentally, the same number of clusters are identified for $\epsilon \in [5, 12]$ and $\rho \in [10, 32]$. The tolerance of the DP method may also be reduced to $\delta = 5$ without a change in the number of clusters. Performing data extraction of trajectories originating and terminating in

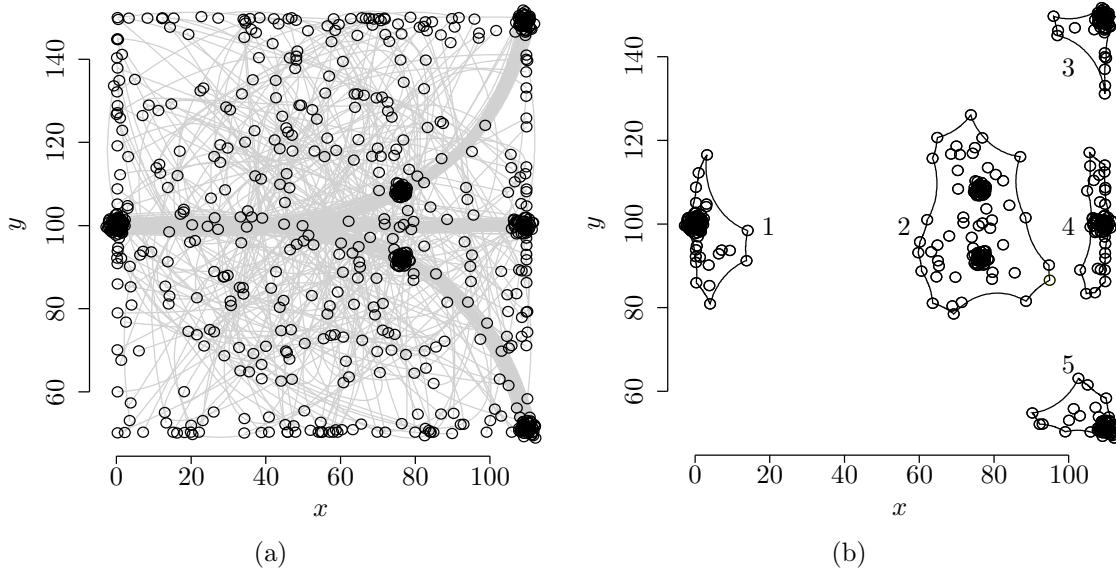
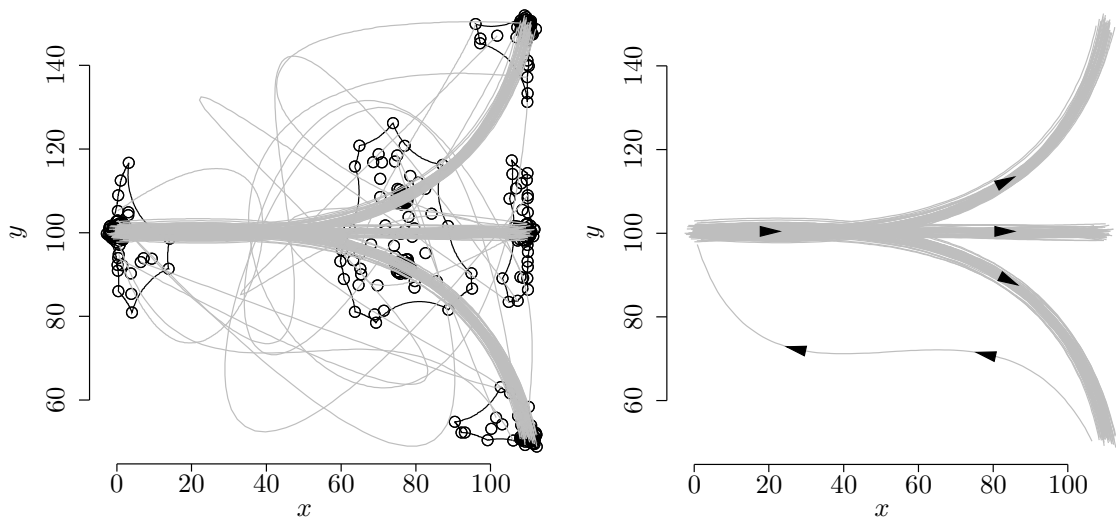


FIGURE 5.17: (a) The trajectories in the $\{\mathcal{U}, \mathcal{S}, \mathcal{D}, \mathcal{N}\}$ data set are thinned using the DP method with a threshold of $\delta = 10$ and (b) the resulting points are clustered using DBSCAN subject to $\epsilon = \delta = 6, \rho = 12$.

one of these clusters produces the reduced data set of Figure 5.18(a). This set has a cardinality of 162, reduced from 250. The number of trajectories from the unstructured subset account for twelve of this total whilst the remaining trajectories are in the sets \mathcal{U}, \mathcal{S} and \mathcal{D} of Figure 5.13(a), each of cardinality 50. ■



(a) Trajectories originating and terminating within a DBSCAN cluster are retained. (b) Trajectories which have sinuosity values outside the 1.5 inter quartile range are discarded.

FIGURE 5.18: (a) Filtering according to origin-destination pairs and (b) sinuosity.

The choice of parameters for the DP method and DBSCAN is instrumental to the success of the approach illustrated in the example above. Methods may be devised for computing suitable values for these parameters, subject to a cost function, and for determining an ideal value for the tolerance of the DP method (by minimizing the error of the approximations to the trajectories themselves), but this is not pursued here. The data extracted using the origin-destination filter approach of Example 5.6 is expected to contain trajectories which differ from the majority of trajectories in this reduced data set and it is necessary to consider additional features of the trajectories in order to arrive at a reasonable approximation of predominant trajectories¹⁸. Taking advantage of derived attributes of these prevailing trajectories and removing all trajectories which may be considered outliers with respect to these attributes allows for further filtering of the data set.

5.2.2 Outlier removal

The notion of *sinuosity* captures two attributes of a trajectory, namely distance (curvilinear length) and displacement, as the ratio

$$\frac{\sum_{i=1}^{n-1} \|\mathbf{x}_{i+1} - \mathbf{x}_i\|}{\|\mathbf{x}_n - \mathbf{x}_1\|}, \quad (5.19)$$

where $\mathbf{x}_1, \dots, \mathbf{x}_n$ are the discrete points of a trajectory of length n and $\mathbf{x}_n \neq \mathbf{x}_1$. The closer the motion of an object is to a straight line, the closer its sinuosity is to unity. Trajectories that are similar in sinuosity may be very dissimilar in their shape — two trajectories sharing end points would have the same displacement but may travel between their end points in a fashion which makes them of equal curvilinear length but of different shape. For example, there are infinitely many trajectories which never intersect the line connecting the two end points which would be of equal sinuosity to a curve that crosses this line a certain number of times. Once the data set has been thinned using DBSCAN and trajectories which feature similar sinuosity values are retained, it is possible to apply PAM and discard all data points which do not comply with a silhouette width of 0.5 as the final computation in this process. The fact that PAM clusters all data is less of a concern on the reduced data set as the remaining data are more similar and the clusters are well supported.

Example 5.7 (Outlier removal and trajectory extraction)

Calculating the sinuosity of each trajectory in the filtered data set of Figure 5.18(a), and discarding all trajectories which have a sinuosity value outside the 1.5 quartile range, produces the reduced data set of Figure 5.18(b). The trajectory which is distant from all others has a low silhouette width with respect to any of the three clusters when clustering with PAM and DTW. It is assigned to the medoid of the trajectory set \mathcal{D} and has a silhouette width of -0.01 . The average silhouette widths per cluster are $\hat{s}(\mathcal{U}) = 0.74$, $\hat{s}(\mathcal{S}) = 0.79$ and $\hat{s}(\mathcal{D}) = 0.61$. ■

5.3 Summary

A filtering approach for extracting a subset of trajectories from a larger data set was presented in this chapter, utilising DTW as a dissimilarity measure and PAM and DBSCAN as clustering

¹⁸If the unstructured data set \mathcal{N} in Figure 5.13(a) contains many more curves than those present in \mathcal{U} , \mathcal{S} and \mathcal{D} , then DBSCAN should find one large cluster and this step would be redundant. This is to be expected as the trajectories in \mathcal{N} would then become the predominant trajectories in the data set in the sense that their down-sampled incarnations result in a set of points which sufficiently cover the scene.

techniques. DTW was reviewed in §5.1.1 as a means to align two time-series which may be similar in shape but differ in sample rates. PAM was described in §5.1.2 as a classical optimization problem and a heuristic approach for solving it was described. The notion of silhouette widths was introduced as a means to perform cluster validation and the difficulties of clustering time-series in the presence of noise was illustrated using an artificially generated data set. DBSCAN was reviewed in §5.2.1 and applied to these data so as to extract trajectories by way of discovering regions of interest. Trajectories were thinned using the DP-method which means that DBSCAN will be certain to discover regions where trajectories begin and end, if these trajectories remain sufficiently close to one another. The derived attribute of sinuosity was introduced in §5.2.2 and used to distinguish between the common and less common trajectories via outlier removal. Finally, the remaining data were clustered using PAM and a poorly clustered trajectory was identified by its unsatisfactory silhouette width.

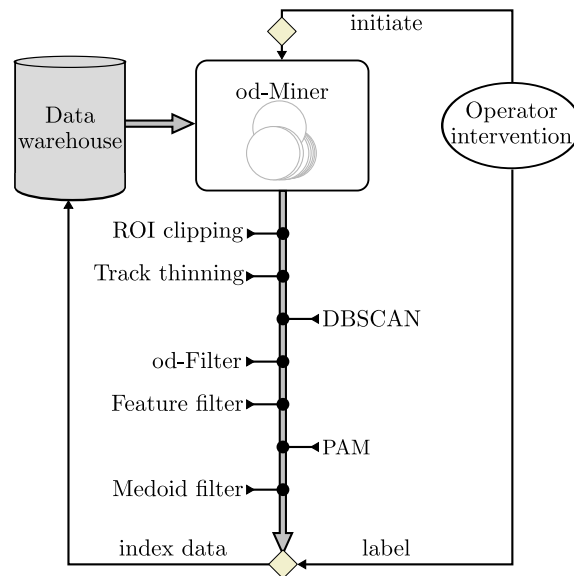


FIGURE 5.19: An overview of the origin-destination miner.

Based on the discussion in this chapter the methodology in Figure 5.19 is proposed for population of the discovery component of the system presented in Figure 3.15. This approach describes a particular clustering task that uses origin-destination pairs as the primary mechanism by which to reduce the data set. Using outlier removal as a means to further reduce the data is expected to work satisfactorily in circumstances where the task may be characterised by the extraction of prevailing activities¹⁹. The origin-destination miner, referred to as *od-miner* in Figure 5.19, is initiated by an operator or analyst. This task necessarily requires a region of interest to be specified *a priori*. In this chapter, the data set featuring \mathcal{U} , \mathcal{S} , \mathcal{D} and \mathcal{N} in Figure 5.13(a) may be considered to be clipped by a rectangular viewport specified by $[0, 120] \times [50, 150]$. These data may then be thinned using a down-sampling technique such as the DP-method. DBSCAN may be applied to discover origin-destination pairs within the region of interest and these pairs may be used to further reduce the number of trajectories for consideration (the *od-filter* step in Figure 5.19). Outliers with respect to derived features may then be removed, after which PAM may be used to identify clusters within the data (the number of clusters may be derived from the unique origin-destination pairs that occur in the data). Lastly, trajectories that are considered

¹⁹For example, a data set containing the fishing activities where fishing vessels are restricted to operating during daylight hours would likely feature a few trajectories for which the vessels return well after dark. This data mining task would constitute a different component that would be added to the discovery component of the overall system.

outliers with respect to these medoids may be discarded and the operator may intervene to approve and label the clusters with names such as “journey from A to B”. These trajectories may then be indexed in the data warehouse as having these additional properties and may thus be extracted by the activity classifiers at a later stage.

CHAPTER 6

Activity classifier model component

Contents

6.1	Introduction	101
6.2	Hidden Markov models	101
6.3	An approach to trajectory classification	110
6.4	Summary	117

6.1 Introduction

The participation of an entity in an activity defined by its spatio-temporal characteristics may be determined from a single datum or from a sequence of observed trajectory data. An example of the former is the event that a vessel enters a no-go area, as was discussed in the rule-based system of §4.4.1. An example of the latter is a vessel travelling along a well-established route towards a particular destination. A significant amount of information is captured in the temporal nature of sequential data and attributing a particular activity to a sequence of updates may require knowledge of the historical updates of a trajectory in relation to the historical updates of other trajectories.

The statistical model-based technique of *hidden Markov models* (HMMs) [87] is explored as a means to describe classes of trajectories in this chapter. HMMs are ideally suited as classifiers of temporal sequences where the parameters of an HMM are estimated from training data [93]. This approach has been applied successfully in various contexts where HMMs act as representatives of the class on which they were trained [5, 44]. These classes are expected to be informed by data gleaned from the surveillance scene. The activity of travelling along a well-established route is considered in detail and an activity classifier is constructed to populate the activity classifier component in the decision support system of §3.4. This classifier obtains its training data from the origin-destination miner of §5.2 and seeks to identify vessels as participating in this derived activity.

6.2 Hidden Markov models

A simple approach to time-series analysis may be undertaken by considering observations in a sequence to be wholly independent of one another. This approach is unsatisfactory since it fails

to exploit correlations that may exist between observations within the sequence, particularly those that occur within close proximity [11]. For example, change in vessel speed is often self-correlated; a vessel that is slowing down is likely to continue doing so and less likely to alternate between acceleration and deceleration. These dependencies may be described in a probabilistic framework¹ as a joint distribution over the observations and if certain assumptions are possible about the casual dependencies between observations, the computation of this distribution may be reduced in complexity. Suppose a series of T observations is represented by the random variables X_1, X_2, \dots, X_T . Then the joint distribution over these variables may be written as

$$p(X_1, \dots, X_T) = \prod_{i=1}^T p(X_i | X_1, \dots, X_{i-1})$$

by application of the product rule for probabilities². If it is assumed that each observation is independent of all but the one preceding it, then this expression reduces to

$$p(X_1, \dots, X_T) = p(X_1) \prod_{i=2}^T p(X_i | X_{i-1}). \quad (6.1)$$

An ordered sequence of such random variables $\{X_t\}_{t=1}^T$ constitutes a *stochastic process*. In the event that the index t is discrete, as in the expressions above, this process is referred to as a discrete-time stochastic process. A sequence of observations may thus be modelled as realisations of such a stochastic process in which dependencies between variables may in general be arbitrarily far reaching. If the random variables are discrete, then the process is referred to as a *Markov chain* in which serial dependencies characterise the process. More precisely, the collection $\{X_t : t \in \mathbb{N}\}$, where $X_t \in \mathcal{S} = \{s_1, s_2, \dots, s_m\} \subset \mathbb{N}$, forms an n -th order Markov chain if

$$p(X_t = s_t | X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, \dots, X_1 = s_1) = p(X_t = s_t | X_{t-1} = s_{t-1}, \dots, X_{t-n} = s_n). \quad (6.2)$$

The expression in (6.1) describes a first-order Markov chain and the accompanying graphical model³ is presented in Figure 6.1(a). The discrete values to which the random variables map are commonly referred to as *states*. Successive states of a first-order Markov chain are not independent⁴ and the distribution from which an observation is drawn depends on the previous observation (this property is referred to as the *Markov property*). For example, consider a ferry travelling repeatedly between two locations and suppose this activity is reduced to the actions of underway or passenger pick-up/drop-off. These events may be described by a first-order Markov chain with two states where $a_{ij}(t) = p(X_{t+1} = j | X_t = i)$ determines the probability of transitioning from state i to state j (see Figure 6.1(b)).

A transition distribution may in general be a function of the states, as well as, of the parameter t . A simplifying assumption is often made that these conditional distributions are independent of t .

¹Random variables over some domain are represented here by upper-case letters (*e.g.* X) whilst specific values from their respective domains are represented as lower-case letters (*e.g.* $X = x$). The distinction between discrete and continuous random variables is made only where necessary and the shorthand notation $p(X = x) = p(x)$ denotes the probability mass function in the former case, and the probability density function in the latter case. For continuous random variables, this shorthand notation represents the fact that the probability of x occurring is determined over an infinitesimal neighbourhood of x .

²Let X and Y be random variables. Then their joint distribution may expressed as $p(X, Y) = p(Y | X)p(X)$ [11].

³A joint probability distribution $p(X_1, \dots, X_T)$ may be represented by a graph whose vertices correspond to random variables X_1, \dots, X_T and whose edges represent direct dependencies between these variables. In the case of Bayesian networks, these graphs are acyclic and directed.

⁴Two random variables X and Y are said to be independent if $p(X, Y) = p(X)p(Y)$. That is, the occurrence of X does not depend on that of Y and *vice versa*.

That is, $p(X_{t+\tau} = j | X_t = i)$ for all $\tau \in \mathbb{N}$ and the Markov chain is said to be *time-homogeneous*⁵. In this case, the transition probabilities are neatly encapsulated in a transition matrix $A \in \mathbb{R}^{m \times m}$ where the transition from state i to j is described by element a_{ij} . These probabilities may be inferred from observations by maximising the log-likelihood of the observation sequence. Solving the constrained optimization problem by application of Lagrange multipliers (see [11] for a description of the process) reveals that the maximum likelihood of transitioning from state i to j is simply the number of times this transition has been observed divided by the number of times the process was in state i .

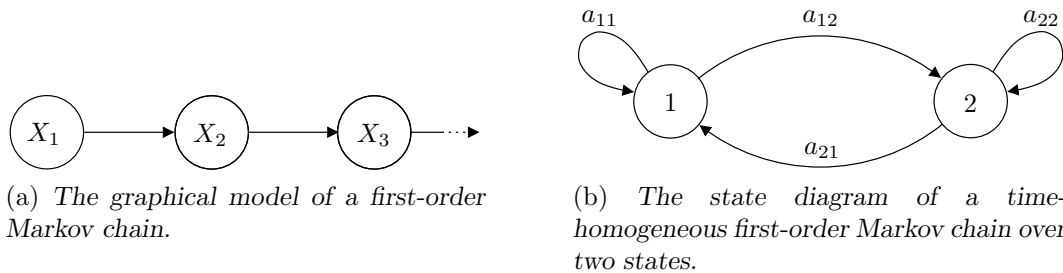


FIGURE 6.1: (a) The graphical model of a Markov chain (b) which may be viewed as a finite-state automatum with probabilistic state transitions [9].

A first-order Markov chain remains limited in that it is unable to capture potential dependencies over several successive observations. The inclusion of these dependencies leads to higher order Markov chains (as specified in (6.2)) which become increasingly difficult to estimate as the dependencies between states increase. If the observations are discrete random variables, then the number of parameters required to specify the model grows exponentially in the number of values the random variables may assume. An n -th order Markov chain with m states requires the estimation of $m^n(m-1)$ parameters [11].

A more general model framework may be obtained by the introduction of latent or hidden variables in which parameter estimation remains tractable. This framework accommodates a class of models which is referred to as *state space models*. These models are characterised by the fact that an unobservable variable Z_t is introduced for each observation X_t and that the hidden variables form a Markov chain. The graphical model is presented in Figure 6.2 and the distributions that it describes have the property that the state Z_{t+1} is conditionally independent⁶ of Z_{t-1} , given the state Z_t . The sequence of variables Z_1, Z_2, \dots, Z_t thus form a discrete-time, first-order Markov chain [9].

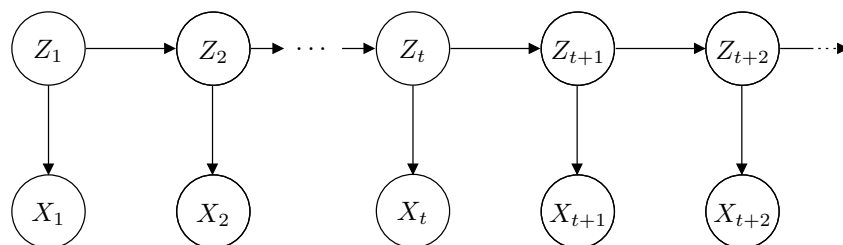


FIGURE 6.2: The graphical model of an HMM.

Importantly, there are no conditional independence relations between the observations them-

⁵A chain that exhibits time-homogeneity is referred to as a *stationary chain* in [9] whilst MacDonald *et al.* [91] refer to a *stationary chain* as a process which remains constant over an initial distribution over the states.

⁶Suppose X, Y and Z are random variables such that $p(X | Y, Z) = p(X | Z)$. Then X is said to be *conditionally independent* of Y , given Z ; this notion is denoted by $X \perp\!\!\!\perp Y | Z$ [11].

selves and the prediction of X_{t+1} depends on all preceding observations, *i.e.* $p(X_{t+1} | X_1, X_2, \dots, X_t)$ [11]. However, an assignment to Z_t results in X_t being independent of every other variable in the HMM (*i.e.* $X_t \perp\!\!\!\perp \{\mathbf{X}^{t-1}, \mathbf{Z}^{t-1}\} | Z_t$, where the bold face variable denotes a vector of variables, $\mathbf{X}^t = (X_1, X_2, \dots, X_t)$ and the negation of an index $\mathbf{X}^{(-t)}$ denotes $\{X_i : i \in \mathbb{N}_T\} - \{X_t\}$) which greatly simplifies inference and parameter estimation. The simplest incarnation of an HMM may be summarised by the conditional independence properties

$$p(Z_t | \mathbf{Z}^{(t-1)}) = p(Z_t | Z_{t-1}) \quad (6.3)$$

and

$$p(X_t | \mathbf{X}^{(t-1)}, \mathbf{Z}^{(t)}) = p(X_t | Z_t) \quad (6.4)$$

for all $t \in \mathbb{N}$ [91]. Enforcing these properties enables the joint distribution over the sequence to be expressed as

$$p(\mathbf{X}^{(T)}, \mathbf{Z}^{(T)}) = p(Z_1) \left(\prod_{t=2}^T p(Z_t | Z_{t-1}) \right) \left(\prod_{t=1}^T p(X_t | Z_t) \right).$$

This expression may be derived directly from the graphical model in Figure 6.2.

The Markov chain $\{Z_t\}$ is referred to as the *parameter process* and the observation process $\{X_t\}$ is commonly referred to as the *state-dependent* process. The state-dependent distribution from which X_t is drawn depends⁷ on the state Z_t . An HMM is parameterised by the transition probabilities of the Markov chain, the state-dependent distributions associated with each of these hidden states, and an initial distribution over the Markov states at $t = 1$. That is, an HMM is specified by a set $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ of hidden states and an associated transition probability matrix $A = [a_{ij}]$ that describes the probability of transitioning from state s_i to s_j , *i.e.*

$$a_{ij} = p(Z_{t+1} = s_j | Z_t = s_i) \text{ for all } 1 \leq i, j \leq N, \text{ where } a_{ij} \geq 0.$$

Furthermore, a probability distribution describing the probability that X_k may be observed when the system is in state s_j is captured in an *emission* matrix $B = [b_j(k)]$, where

$$b_j(t) = p(X_t | Z_t = s_j) \text{ for all } 1 \leq j \leq N \text{ and all } 1 \leq t \leq T.$$

Finally, the initial state distribution is represented by $\boldsymbol{\pi} = [\pi_i]$, where $\pi_i = p(Z_1 = s_i)$ for all $1 \leq i \leq N$ [136]. The shorthand notation $\Theta = (A, B, \boldsymbol{\pi})$ is used when referring to a particular HMM.

Example 6.1 (HMMs as generators)

Two HMMs are considered in this example. The first features discrete state-dependent distributions, while the second features continuous state-dependent distributions.

- (a) *Suppose a series of coin tosses are produced by a fair or biased coin and that $X_t \in \{H, T\}$ are the outcomes of the coin tosses (the biased coin favours tails with a probability of $p(T) = 0.9$). Furthermore, suppose that the coin used to generate a particular observation is unknown. Let the states of the Markov chain be denoted by $\mathcal{S} = \{F, B\}$, which represent the fair and biased coins, and suppose the transition and emission matrices are defined as*

$$A = \begin{matrix} & F & B \\ \begin{matrix} F \\ B \end{matrix} & \begin{bmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{bmatrix} \end{matrix} \quad \text{and} \quad B = \begin{matrix} & H & T \\ \begin{matrix} H \\ T \end{matrix} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix} \end{matrix}, \quad (6.5)$$

respectively. Furthermore, suppose the initial distribution is $\boldsymbol{\pi} = \frac{1}{2}(1, 1)$. A sequence of observations $\mathbf{X}^{(500)}$ is sampled⁸ from the model and is represented as crosses in Figure 6.3

⁷If independence between the states of the parameter process is assumed, then each time-slice of the HMM is a Gaussian mixture model [91].

⁸This method of sampling is commonly referred to as ancestral sampling [11].

that indicate whether or not a sample is heads or tails. The states of the underlying Markov process are presented in a two-tone grey band in which light grey denotes the fact that the state corresponding to the fair coin generated the observation, whilst the darker grey regions denote the fact that the state corresponding to the biased coin is responsible for the observation. The probability of remaining in the state corresponding to the fair coin is 0.99, whilst the probability of transitioning from that state to the state corresponding to the biased coin is 0.01.

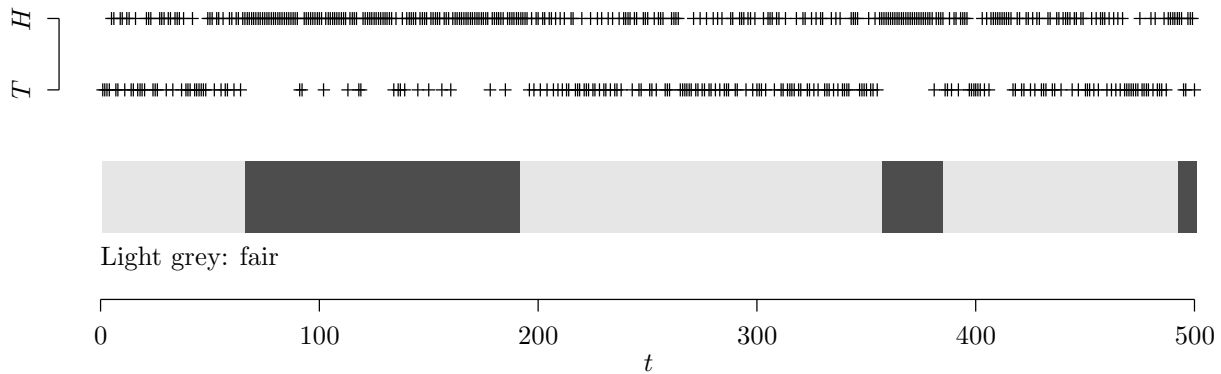


FIGURE 6.3: A two-state HMM model describing the use of a fair or biased coin [62].

- (b) Suppose a particle moves in the Cartesian plane in such a manner that it visits three locations in a counter-clockwise manner and that its position is distributed according to a Gaussian distribution centered on each of these locations.

The three locations are considered to map to the states of a Markov chain and the state-dependent distributions are the bivariate Gaussian distributions

$$\mathcal{N}_1((0, 0), 0.1 \times I_2), \mathcal{N}_2((3, 0), 0.05 \times I_2) \text{ and } \mathcal{N}_3((1.5, 3), 0.1 \times I_2),$$

where $I_2 \in \mathbb{R}^{2 \times 2}$ is the identity matrix. Furthermore, suppose the transition matrix is specified as

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.6 & 0.4 \\ 0.99 & 0 & 0.01 \end{bmatrix}, \tag{6.6}$$

and that the particle always originates in the region corresponding to the first state of the Markov chain (the initial state distribution is therefore $\pi = (1, 0, 0)$).

A sequence of $\mathbf{X}^{(100)}$ is sampled from this model and the Markov states responsible for the observation are presented in Figure 6.4(a), whilst the observed sequence is presented in Figure 6.4(b). It is clear from the the transition matrix in (6.6) that the particle never lingers in the region assigned to first Markov state and that the propensity for lingering in the third state is evident in the observations.

Using a three-component mixture model to describe the data in Figure 6.4(b) will not be sufficient as the sequential ordering will be lost due to the assumption that the observations are independent. ■

A preferable approach to the explicit definition of the parameters of the HMMs in Example 6.1 is the estimation of these parameters from observations. These parameters may be determined

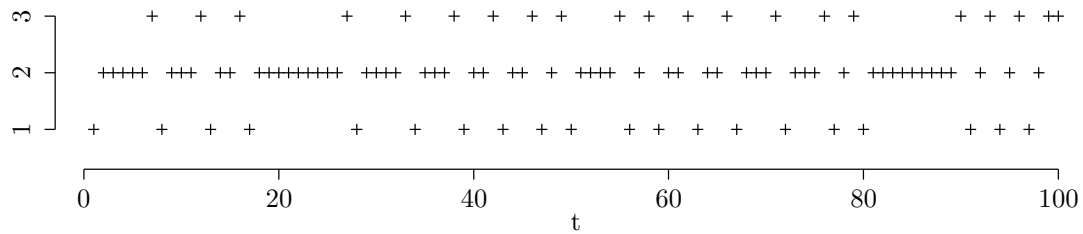
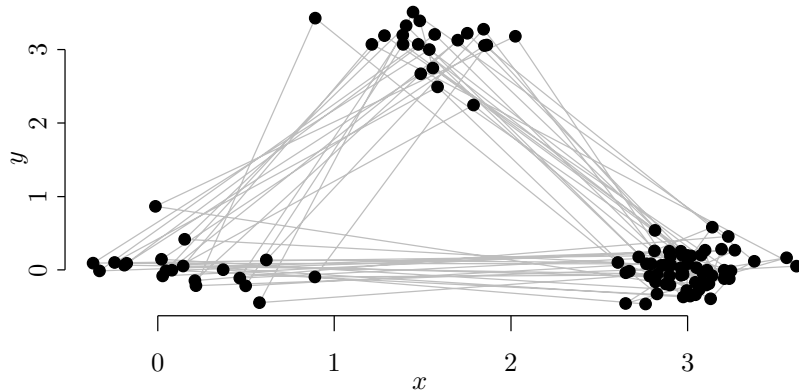

 (a) The states generating emissions at each $t \in (1, 100)$.

 (b) The emissions which yield (x, y) -positions in the Cartesian plane.

FIGURE 6.4: A three-state HMM model featuring bivariate Gaussian emissions.

from a data set of observations \mathcal{X} by maximum likelihood estimation. That is, the parameters Θ that maximize the marginal distribution

$$p(\mathcal{X} | \Theta) = \sum_{\mathcal{Z}} p(\mathcal{X}, \mathcal{Z} | \Theta) \quad (6.7)$$

may be estimated. This produces the likelihood of the data, given the parameters Θ . A direct maximisation of (6.7), however, leads to an expression with no closed-form solutions. Instead, a general iterative framework for maximum likelihood estimation in models featuring latent variables is employed. This approach is known as the *expectation-maximisation method* [11]. Given some initial estimate Θ_0 of the model parameters, the expectation of the posterior distribution over the latent variables and the current estimate of the parameters is computed in the *E-step* (i.e. $E[p(\mathcal{Z} | \mathcal{X}, \Theta_0)]$). This expectation is maximised in the *M-step* and yields updated parameters Θ_1 which are used to recompute the posterior estimates. This process is repeated until suitable convergence in the log-likelihood or in the parameter values is observed [11]. This method does not guarantee convergence to a global optimum and finding suitable initial choices for the transition and emission matrices contributes to convergence. The general framework described above has come to encapsulate many approaches to maximum-likelihood that were developed independently in various applications. In the case of HMMs, the *Baum-Welch* algorithm was developed [136].

The task of parameter estimation, often referred to as *learning*, is cited in [136] as one of three commonly performed inference tasks on HMMs. The remaining two tasks are *evaluation* and *decoding*. The evaluation task answers the critical question as to whether or not a given sequence is likely to have been generated by a particular HMM. If the model parameters for a given HMM are contained in Θ , as before, then these parameters completely specify the model. The evaluation task aims to determine the likelihood of a series of observations $\mathbf{X}^{(t)}$, given the model (i.e. $p(\mathbf{X}^{(t)} | \Theta)$). Lastly, the decoding task determines the state sequence which best

explains an observed sequence. Two commonly used optimality criteria for quantifying this notion are to determine the state sequence that maximises the posterior distribution $p(\mathbf{X}^{(t)}|\mathbf{Z})$ or to determine the most likely complete sequence which maximises the posterior distribution. An implementation of the latter approach is known as the *Viterbi* algorithm [136]. The difference may be better understood when considering the fair-biased HMM in Example 6.1, where the transmission and emission matrices are

$$A = \begin{matrix} & F & B \\ F & \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \\ B & \end{matrix} \quad \text{and} \quad B = \begin{matrix} & H & T \\ H & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{3}{5} \end{bmatrix} \\ T & \end{matrix}, \quad (6.8)$$

respectively. If the sequence

$$\mathbf{X}^7 = (H, H, T, T, T)$$

is observed, then the most likely state sequence according to the Viterbi method is B, B, B, B, B , whilst the sequence featuring the most likely states at each value $t = 1, 2, \dots, 5$ is F, F, B, B, B . The state sequences determined by these methods are compared in Figure 6.5 over the sequence generated in Example 6.1.

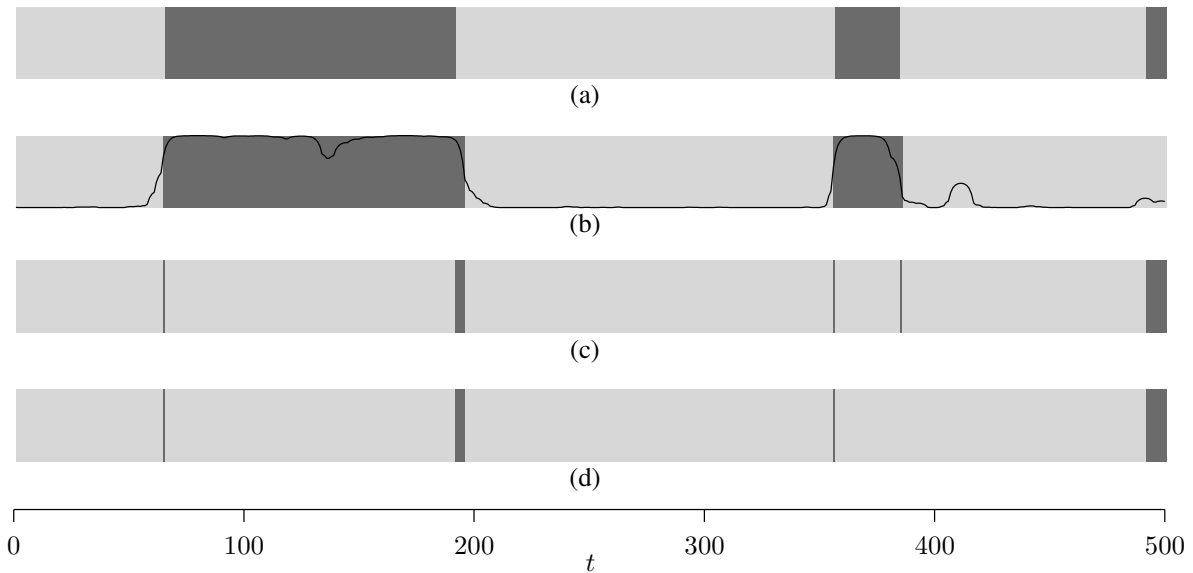


FIGURE 6.5: (a) The states corresponding to either the fair coin (light grey) or the biased coin (dark grey) that generated the sequence $\mathbf{X}^{(500)}$ presented in Example 6.1. (b) The most probable state path (identical colouring to (a)), as computed using the Viterbi algorithm, and the posterior probability curve determined by the forward-backward algorithm. (c) The difference between the generating hidden states and the most probable path. (d) The difference between the generating hidden states and the states as determined using the posterior probability (a probability threshold of 0.5 is used to determine whether a state generated the observation).

The evaluation task is revisited as it presents some insight into the methods that are used in resolving the three aforementioned inference problems. Let $\mathbf{X}^{(T)} = \mathbf{x}^{(T)}$ be an observed sequence. Then the marginal distribution

$$p(\mathbf{x}^{(T)} | \Theta) = \sum_{\mathbf{Z}} p(\mathbf{x}^{(T)} | \mathbf{Z}, \Theta) p(\mathbf{Z} | \Theta) \quad (6.9)$$

determines the probability of the sequence $\mathbf{x}^{(T)}$. The fact that observations are independent when the states are known allows the likelihood term in (6.9) to be expressed as the product of

emission probabilities, *i.e.*

$$\begin{aligned} p(\mathbf{x}^{(T)} | \mathbf{Z}^{(T)}, \Theta) &= \prod_{t=1}^T p(x_t | z_t, \Theta) \\ &= b_{z_1}(x_1) b_{z_2}(x_2) \cdots b_{z_T}(x_T). \end{aligned} \quad (6.10)$$

The prior distribution is easily computed by application of the Markov property and reduces to

$$p(\mathbf{Z}^{(T)} | \Theta) = \pi_{z_1} \prod_{i=1}^{T-1} a_{z_i z_{i+1}} \quad (6.11)$$

$$= \pi_{z_1} a_{z_1 z_2} a_{z_2 z_3} \cdots a_{z_{T-1} z_T}. \quad (6.12)$$

The combination of these terms in (6.9) yields the expression

$$p(\mathbf{x}^{(T)} | \Theta) = \sum_{z_1, z_2, \dots, z_T} \pi_{z_1} b_{z_1}(x_1) a_{z_1 z_2} b_{z_2}(x_2) \cdots a_{z_{T-1} z_T} b_{z_T}(x_T) \quad (6.13)$$

which requires N^T summations over all possible states and roughly $2T$ evaluations for each transition and emission probability. Direct summation over all possible states therefore results in an algorithm with time complexity of $O(TN^T)$, which is computationally infeasible for long sequences (*i.e.* large values of T) [136]. Utilising a dynamic programming approach leads to an improved approach for computing the probability of (6.13). This approach proceeds by computing ever-increasing sub-sequences and reusing the results of these calculations in subsequent calculations. These sub-sequences are represented by

$$\alpha_t(i) = p(x_1, x_2, \dots, x_t, z_t = s_i | \Theta),$$

which are referred to as *forward probabilities*. Representing the latent variables of an HMM at each time slice as points on a lattice illustrates how dynamic programming may be used. Consider a sequence of three observations $\mathbf{X}^{(t)} = (H, H, T)$, which are assumed to be generated by the HMM of Example 6.1(a). Considering these states across the $t = 1, 2, 3$ time slices produces the lattice in Figure 6.6(a). The forward probabilities at $t = 1$ are $\alpha_1(1) = \pi_1 b_1(X_1)$ and $\alpha_1(2) = \pi_2 b_2(X_1)$ since π is the initial state distribution. At the next time instant, $t = 2$, the forward probabilities are

$$\alpha_2(1) = (a_{21}\alpha_1(1) + a_{11}\alpha_1(2)) b_1(X_2) \quad \text{and} \quad \alpha_2(2) = (a_{12}\alpha_1(1) + a_{22}\alpha_1(2)) b_2(X_2).$$

If the probability for $p(\mathbf{X}^2 | \Theta)$ is sought, then it is evident from (6.13) that this quantity is the sum of $\alpha_2(1)$ and $\alpha_2(2)$. Returning to the sequence of three observations, the paths that the dynamic programming approach considers during calculation of $\alpha_3(1)$ is shown in Figure 6.6(b). This approach is commonly referred to as the *forward algorithm* and reduces the time complexity to $O(N^2T)$.

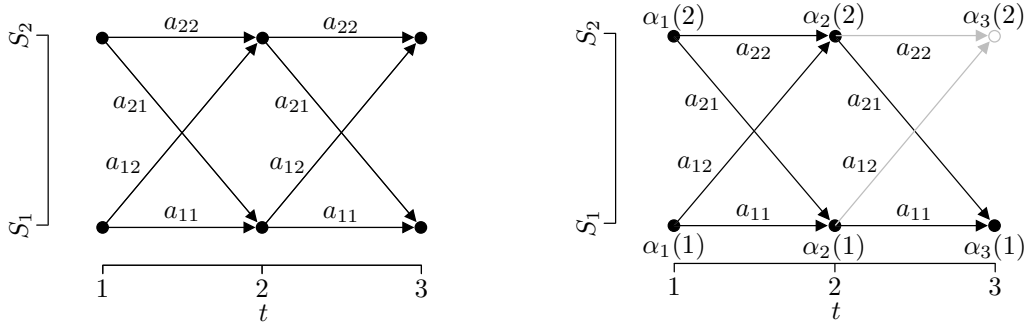
A backward probability may be defined in a similar fashion as

$$\beta_t(i) = p(x_1, x_2, \dots, x_t | z_t = s_i)$$

and its calculation also follows a dynamic programming approach. Forward probabilities are used in resolving the inference problem of evaluation, whilst backward probabilities are used in the Viterbi method, and model learning requires their combined usage [136].

When performing evaluation of long observation sequences, the probability in (6.9) diminishes due to the repeated multiplication operations. This calculation is susceptible to *underflow*⁹,

⁹Underflow is the phenomenon that occurs when the result of a floating point operation is smaller in magnitude than the smallest value that can be represented on a computer. The IEEE standard for double precision numbers, which is in use on most computing platforms, specifies the smallest representable number to be 2.22×10^{-308} .



(a) The state paths and their accompanying transitions are embodied in a lattice structure.

(b) The four state paths on which the computation of $\alpha_3(1)$ depends are presented as black edges.

FIGURE 6.6: The lattice structures resulting from a dynamic programming approach to computing the probability in (6.9).

thus requiring the application of a logarithmic transformation. This transformation preserves the monotonicity of the function it acts upon, thus preserving extrema whilst ensuring that small numbers may still be represented in computer memory.

Example 6.2 (Sequence classification)

Consider two HMMs Θ_1 and Θ_2 , where Θ_1 represents the HMM in Example 6.1 and Θ_2 comprises the transition and emission matrices

$$A_2 = \begin{matrix} & F & B \\ F & \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{9}{10} & \frac{1}{10} \end{bmatrix} \\ B & \end{matrix} \quad \text{and} \quad B_2 = \begin{matrix} & H & T \\ H & \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ 0.02 & 0.98 \end{bmatrix} \\ T & \end{matrix}, \quad (6.14)$$

respectively, and a prior distribution of $\pi_2 = (\frac{1}{2}, \frac{1}{2})$. In a competing model setting, the model which best explains a sequence is considered to be the generating model. That is, if $p(\mathbf{X}^{(t)} | \Theta_i) > p(\mathbf{X}^{(t)} | \Theta_j)$, then the model corresponding to Θ_i is selected as the generating model. Consider the sequences

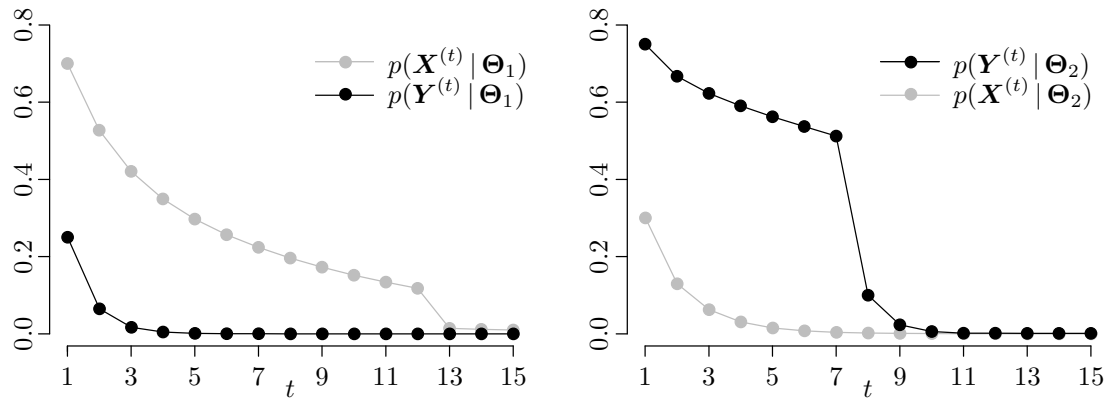
$$\mathbf{X}^{(15)} = (T, T, \dots, T, H, T, T) \quad (6.15)$$

and

$$\mathbf{Y}^{(15)} = (H, H, H, H, H, H, H, T, T, T, T, H, H, H, H), \quad (6.16)$$

each generated by the HMMs defined by Θ_1 and Θ_2 , respectively. If the forward algorithm is used to compute the probability of observing these sequences with respect to each of the models for $t = 1, 2, \dots, 15$, the results in Figures 6.7(a) and 6.7(b) are obtained. The probability of the sequence, given a model, is expected to be higher when evaluated with respect to its generating model, than with respect to any other competing model. As is illustrated in Figure 6.7(a), $\mathbf{X}^{(t)}$ is classified as more likely to have been generated by Θ_1 than Θ_2 , and in Figure 6.7(b) the converse is true when considering the model Θ_2 . However, this is not always the case. In this example, any sequences that have identical symbols $\mathbf{X}^{(t)} = \mathbf{Y}^{(t)}$, for $t = 1, \dots, T$, will be assigned to the HMM which favours that sequence (Θ_1 is more likely to emit the symbol T whilst Θ_2 is more likely to emit the symbol H).

As the length of an observed sequence grows, the probability of observing it decreases. This agrees with intuition as the probability of observing a particular sequence in the space of all sequences diminishes as this space grows. Despite this fact, it is still possible to discern between two long



(a) Evaluation of sequences $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ with respect to the HMM defined by the parameters Θ_1 .

(b) Evaluation of sequences $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ with respect to the HMM defined by the parameters Θ_2 .

FIGURE 6.7: A comparison of the probabilities of the sequences (6.15) and (6.16) evaluated at $t \in 1, 2, \dots, 15$ with respect to two different HMMs.

sequences as differences remain in these probabilities. This is illustrated in Figure 6.8(a) and 6.8(b), where two sequences of 500 observations are compared with respect to the models¹⁰ Θ_1 and Θ_2 . ■

The previous example illustrates how a competitive model setting may be used to identify observation sequences, but in each case the HMM topologies and parameters are set *a priori*. An alternative approach to inducing these topologies from training data is subsequently considered where HMMs are trained on multiple sequences of two-dimensional trajectories and these derived models are used to classify trajectories that were not used in training.

6.3 An approach to trajectory classification

The classification of spatial trajectories using hidden Markov models is widespread in the literature (see §2). Choosing a suitable model structure is often determined in an *ad-hoc* fashion or in some cases, by experimental evaluation. The features on which the time-series are modelled inevitably depend on the application area. A simple approach to modelling time series using HMMs is pursued in this section where the features are selected as Cartesian coordinates of the trajectory. This simplified approach is investigated so that properties of the modelling approach may be addressed without the additional complication of selecting different features. Indeed, this direct approach is also applied in traffic and pedestrian trajectory modelling [93, 115].

The data that are mined by the origin-destination miner of the clustering component in §5 may be used to train a collection of HMMs. Considering that the time-series feature continuous observations, HMMs with continuous observations densities should be used¹¹. Gaussian densities may be chosen as the emission densities. A consequence of assuming continuous emission distributions is that parameter estimation proves more challenging and for this reason it is necessary to provide good initial estimates of these distributions in order to facilitate convergence of the EM algorithm [136].

¹⁰Incidentally, the exponentiation of the log-likelihoods for the sequence $\mathbf{X}^{(500)}$ are $p(\mathbf{X}^{(500)}|\Theta_1) = 3 \times 10^{-139}$ and $p(\mathbf{X}^{(500)}|\Theta_2) = 2 \times 10^{-195}$.

¹¹It would be possible to quantize observations and use a discrete emission HMM.

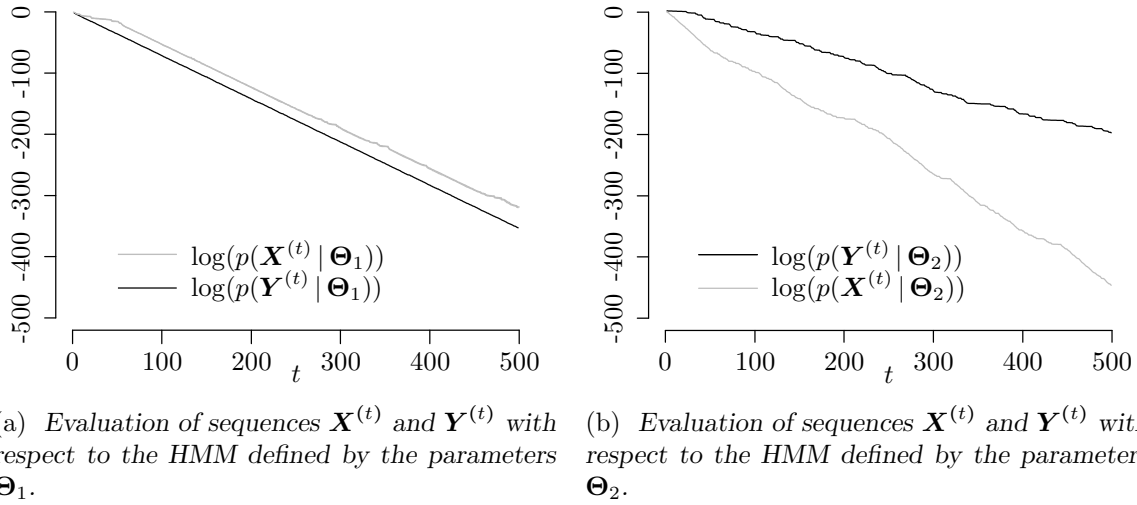


FIGURE 6.8: A comparison of the log-likelihoods of the sequences (6.15) and (6.16) evaluated at $t \in 1, 2, \dots, 15$ with respect to two different HMMs.

The medoid computed by PAM in §5.2.2 may serve as the reference series for a particular HMM. The initial estimates for the state-dependant distributions may be estimated relative to this trajectory. Linear interpolation of the medoid may be performed so that series can be segmented into portions on which the sample mean and covariances may be computed¹². The sample mean $\bar{\mathbf{x}}$ of a collection of N observations, where $\mathbf{x} \in \mathbb{R}^d$, is

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^d \mathbf{x}_i \quad (6.17)$$

and the sample covariance is

$$\mathbf{Q} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (6.18)$$

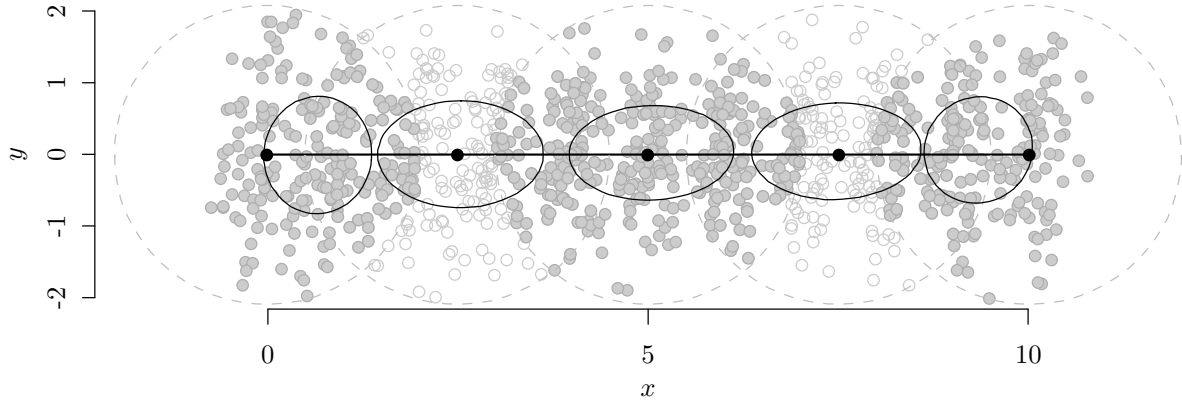
Suppose that m states are desired for the underlying Markov chain and that the medoid is denoted by $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$. Then these collections of observations for each of the desired states are determined to be the points which lie within a disk of radius

$$\delta_h = \frac{1}{M} \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{x}_{i+1}\|,$$

centered on the points determined by interpolation. Consider a collection of one hundred straight lines originating in the region of $(0, 0)$ and terminating in the region of $(10, 0)$, presented in Figure 6.9(a). Suppose that five initial estimates are sought for a six-state HMM with Gaussian emission that will model these data. The reference curve, presented as the black line in Figure 6.9(a), is interpolated to provide the points $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_5$. The point clouds comprise all points within the disk $D_{\delta_h}(\mathbf{y}_i) = \{\mathbf{x} \mid d(\mathbf{y}_i, \mathbf{x}) < \delta_h\}$. These calculations serve to provide initial estimates to the EM method (as these values are refined during training, their heuristic method of derivation is not necessarily a disadvantage). This process is also followed in Figure 6.9(b) in which eight initial estimates for emission distributions are sought on a variation of the data set

¹²Use of the radial distance method discussed §5.2.1 for this task is not suitable as this approach is not guaranteed to produce the desired number of states.

\mathcal{U} from §5. The EM algorithm converges¹³ and the means and covariances it determines through training are shown in Figure 6.9(c).



(a) Six mean and covariance tuples $(\bar{x}_i, \mathbf{Q}_i)$ estimated from one hundred lines which originate in the region of $(0,0)$ and terminate in the region of $(10,0)$. Only the points within a disk of radius δ_h centered at the i -th interpolated point contribute to the estimate of $(\bar{x}_i, \mathbf{Q}_i)$.

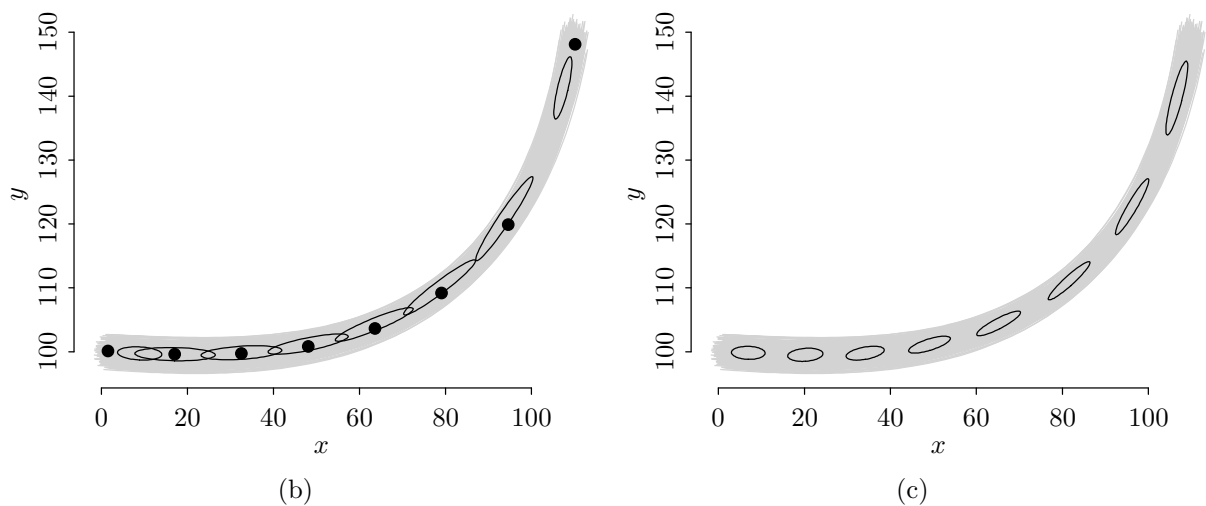


FIGURE 6.9: (a) Six overlapping point clouds are used to calculate initial estimates of the emission distribution parameters of a six-state HMM. (b) This process is repeated for a variation of the data set \mathcal{U} . Points along the medoid which result from linear interpolation are depicted as filled circles whilst the initial estimates resulting from these data are presented as black ellipses centered at the sample means. (b) Parameter estimation, using the series depicted in grey, yields the refined mean and covariances of each of the emission distributions.

The question naturally arises from Figure 6.9(c) as to whether or not the number of hidden states is appropriate for modelling the series in \mathcal{U} (assuming that the bivariate Gaussian distributions are used). *Model selection* may be used to resolve this question and suitable criteria are required for comparison of various models [91]. Minimising the log-likelihood is the criterion employed by *cross-validation* in selecting the most suitable model, whereas the *Bayesian information criterion* (BIC) and the *Akaike information criterion* (AIC) additionally penalise model complexity¹⁴. If

¹³One hundred random initialisations were chosen for the sake of comparison and none of these executions converged.

¹⁴The notion of favouring parsimony over more complex solutions or models is a prevailing notion throughout the literature. This is captured in *Occam's Razor* and often applied as a method of model selection (see [32] for a critical review of the advantages and disadvantages of its use in modelling).

the log-likelihood of a fitted model is denoted by $\log L$, p denotes the number of free parameters and T is the number of observations, then

$$AIC = -2\log L + 2p \quad (6.19)$$

and

$$BIC = -2\log L + p \log T. \quad (6.20)$$

A model which yields the smallest value for AIC or BIC may be considered better than other choices of models fitted to the data. In each case the fit of the model to the data is captured in the log-likelihood term whilst the complexity is captured in either $2p$ or $p \log T$. The number of parameters p for an m -state HMM with d -dimensional observations may be determined by

$$md \left(1 + \frac{d+1}{2}\right) + m(m-1) + (m-1). \quad (6.21)$$

The first term describes the number of parameters necessary to specify the emission distributions for each of the m states (since each emission distribution has a mean of dimension d and covariance matrix of $d(d+1)/2$). The second term describes the number of parameters that need to be estimated in the transition matrix ($m(m-1)$ as opposed to m^2 since each row sums to unity) and the third term specifies the number of parameters in the initial distribution $\boldsymbol{\pi}$. Selecting the number of observations T in (6.20) is not always obvious. This value is often simply taken to be the number of elements in the sample under consideration.

Besides determining the number of states in an HMM, the above information criteria may also be used to determine suitable state-dependant distributions if a number of choices are available [91]. For example, suppose an appropriate distribution is sought to model the data of Figure 5.9, where these data are observations from a particular state-dependant distribution. Modelling these data using a GMM requires an estimation of the number of Gaussian components in the mixture. Although it is clear from the figure that these data are bimodal, this will often not be known *a priori*. The results of computing the AIC and BIC for mixtures of one to five components are presented in Table 6.1. It may be seen that the correct number of components are recovered as both criteria are minimal when considering two mixture components. Thereafter, the penalty term overcomes the decreases in the log-likelihood values¹⁵.

Components	AIC	BIC	$\log L$
1	3402.45	3426.99	-1696.22
2	3027.53	3081.51	-1502.76
3	3031.31	3114.75	-1498.65
4	3035.80	3148.68	-1494.90
5	3039.37	3181.70	-1490.68

Table 6.1: The results of the AIC and BIC for a GMM when applied to the data of Figure 5.9.

As stated in [91], the problem of estimating the number of states of an HMM is neither trivial nor settled. The authors in that text recommend using both criteria as a guide in model selection. Using sequences generated from the discrete two-state HMM for the biased coin of Example 6.1(a), and from the three-state HMM with Gaussian emissions in Example 6.1(b), HMMs with 2, 3, 4 and 5 states are estimated from these generated sequences. The resulting AIC, BIC and log-likelihood values are presented in Tables 6.2(a) and 6.2(b), respectively. In both cases, AIC and BIC effectively identify the number of states of the original HMMs. It may

¹⁵The number of parameters in an m -component mixture model may be computed using (6.21) without the term $m(m-1)$.

be seen that the log-likelihood decreases in both cases, thus illustrating the utility of the model complexity penalty terms. The fit of the model, as determined by the likelihood, of an HMM with m states always improves with increased states [91].

States	$\log L$	AIC	BIC	States	$\log L$	AIC	BIC
2	-305.80	621.60	642.68	2	-189.70	405.41	439.28
3	-304.43	630.87	677.23	3	-38.05	122.10	182.02
4	-302.54	643.08	723.15	4	-31.97	133.94	225.12
5	-300.63	659.26	781.49	5	-23.53	145.06	272.71

(a)

(b)

Table 6.2: The log-likelihood, AIC and BIC criteria for (a) the HMM with bivariate Gaussian observations in Example 6.1(b) and (b) the HMM with bivariate Gaussian observations in Example 6.1(b).

Finally, the data set featuring the north-easterly (\mathcal{U}), easterly (\mathcal{S}), south-easterly (\mathcal{D}) and the less structured data (\mathcal{N}) is considered. It is assumed that the three structured data clusters have been labelled as the activities of travelling from within the region of $[0, 120] \times [50, 150] \times$ in each of the aforementioned directions. These clusters are considered to be routes.

Example 6.3 (Trajectory model estimation)

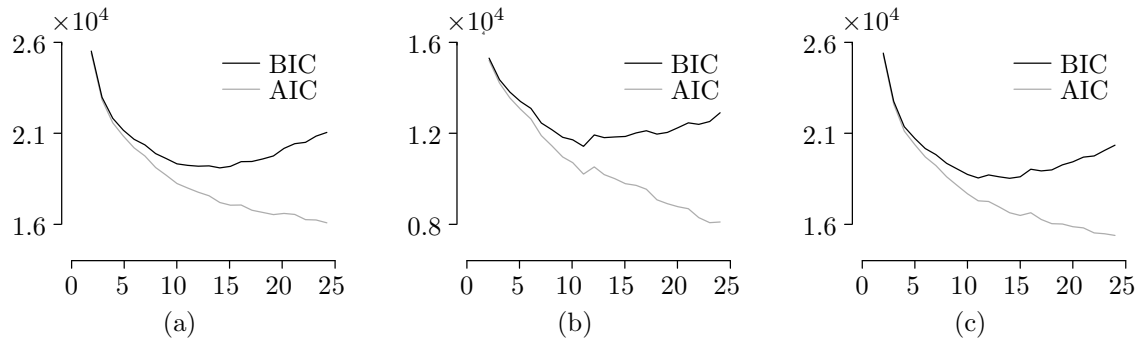
The clusters identified by the od-miner of §5.2.2 provide the training and testing data for a corresponding HMM. Each set \mathcal{U} , \mathcal{S} , \mathcal{D} and \mathcal{N} comprises 498 sequences that are subdivided into trajectories representing objects that are travelling at three different speeds, thus resulting in sequences of varying length. These classes of 168 trajectories each comprise sequences of approximately eleven, seventeen and twenty eight observations. The state-dependent distributions are chosen as bivariate Gaussian distributions¹⁶. Training data are derived from the \mathcal{U} , \mathcal{S} , \mathcal{D} and \mathcal{N} . These training sets are denoted by $tr(\mathcal{U})$, $tr(\mathcal{S})$, $tr(\mathcal{D})$ and $tr(\mathcal{N})$, where $tr(\mathcal{X})$ is a function selecting a subset of elements from a set \mathcal{X} . Each training set has a cardinality of one hundred and is selected as a random permutation of the original data sets.

The initial estimates for the means and covariances of each HMM are calculated for the \mathcal{U} , \mathcal{S} and \mathcal{D} sets using the method depicted in Figure 6.9(a). These HMMs¹⁷ are denoted by $\Theta_{\mathcal{U}}$, $\Theta_{\mathcal{S}}$ and $\Theta_{\mathcal{D}}$. A background or noise model $\Theta_{\mathcal{N}}$ is estimated for the data in \mathcal{N} . A reference curve is not used to provide initial estimates to this HMM since the clustering method of §5.1.2 was not applied. Rather, K -means is used to initialise the HMM corresponding to these data.

The AIC and BIC are used to provide an indication as to how many hidden states to use. These criteria are computed for 2, 3, ..., 23 hidden states for \mathcal{U} , \mathcal{S} and \mathcal{D} , the results of which are presented in Figure 6.10. The penalty terms in the AIC fail to overcome the log-likelihood values when computing this criterion over all three data sets, whilst the BIC deems 10 hidden states to be appropriate for the models $\Theta_{\mathcal{U}}$, $\Theta_{\mathcal{S}}$ and $\Theta_{\mathcal{D}}$. Similarly, ten states are favoured by the BIC for $\Theta_{\mathcal{N}}$. Although an agreement between these two criteria is often sought, it is possible to conduct further analysis by computing the error rate of an HMM in a classification task for varying numbers of hidden states (an example of this approach may be found in [137]).

¹⁶If Gaussians are chosen for the observation distributions, then whether or not to use a mixture distribution or a simple bivariate distribution may be determined using the aforementioned information criteria. An arbitrary selection of a point cloud along any of the routes reveals that a bivariate Gaussian is preferred over mixture densities with two or more components.

¹⁷The HMMs modelling the three structured data sets are initialised with K -means for the sake of comparison. The initial centroids are chosen at random, and compared to those derived using the aforementioned reference curve method. Forty initial estimates were calculated for each of the models and the resulting log-likelihood values were inferior to those produced using the reference method in every case.


 FIGURE 6.10: The AIC and BIC computed for the (a) $\text{tr}(\mathcal{U})$ (b) $\text{tr}(\mathcal{S})$ and (c) $\text{tr}(\mathcal{D})$ data sets.

Considering that three different velocities are present in the training data, it is expected that the transition matrix will be a sparse matrix with a right non-zero bandwidth. In this example the topology is learnt directly from the data. It may be seen from the resulting transition matrix

$$A_{\mathcal{U}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \left[\begin{array}{cccccccccc} 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.59 & 0.41 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.58 & 0.42 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.53 & 0.47 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0.75 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.17 & 0.53 & 0.30 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.48 & 0.52 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.00 & 1.00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 \end{array} \right] \end{matrix}$$

for the HMM $\Theta_{\mathcal{U}}$ with ten hidden states that a left-to-right topology is induced. The remaining two structured HMMs $\Theta_{\mathcal{S}}$ and $\Theta_{\mathcal{D}}$ exhibit the same left-to-right topology. This is not the case for the noise HMM $\Theta_{\mathcal{N}}$ which only exhibits five non-zero transitions in its transition matrix. ■

Using the HMMs of Example 6.3, two approaches to classifying trajectories may be pursued. Firstly, HMMs may be trained on the structured data and thresholds for determining class membership may be selected for each HMM. Secondly, in addition to the HMMs trained on the structured data, an HMM may be estimated from the unstructured data set and class membership may be resolved by determining which HMM produces the highest probability of having generated a provided sequence. This model may be considered as a *background* model which models the noise in the scene¹⁸ and the approach is akin the competitive classification approach employed in Example 6.2. The threshold for class membership to class \mathcal{C} in the former approach may be determined directly as

$$\min\{t^{-1} \log P(\mathbf{X}_i^{(t)} | \Theta_{\mathcal{C}}) | \mathbf{X}_i^{(t)} \in \mathcal{C}\}. \quad (6.22)$$

This approach relies on the training set being representative of the class and would require a large quantity of training data. It does not address the possibility that a trajectory from the unstructured class lies above the threshold when being evaluated by the derived HMM. In an effort to address this, an experimental approach may be taken by choosing the threshold

¹⁸Speech recognition pursues an analogous methodology where *filler* or *garbage* models describe unknown portions of speech or captured sound which is not speech.

which provides a suitable inaccuracy or misclassification rate for a given classifier. Suppose the classification of a trajectory \mathbf{X} by an HMM trained on a class with class label C is defined by the function $h_C(\mathbf{X})$, then the misclassification rate of an HMM may be defined as

$$\text{error}(\Theta_C) = \frac{1}{|\mathcal{C}|} \#\{\mathbf{X} \in \mathcal{C} \mid h_C(\mathbf{X}) \neq C\}. \quad (6.23)$$

It is not necessarily true that a classifier will achieve an error rate of zero and threshold values are often chosen with respect to the trade-off between the *sensitivity* and *specificity* of a classifier. These notions are defined in relation to the successful and unsuccessful classifications of observations in a binary classification test. That is, an observation correctly attributed to a class by a classifier is considered a *true positive* (TP), whilst an observation which is erroneously attributed to that class is referred to as a *false positive* (FP). Conversely, an observation which is erroneously determined to be a non-member of the class is referred to as a *false negative* (FN) whilst the correct determination of a non-class member as such is considered to be a *true negative* (TN). The sensitivity of a classifier is defined in relation to the *true positive rate* of

$$\frac{TP}{TP + FN},$$

whilst the specificity is the *false positive rate*

$$\frac{FP}{FP + TN}.$$

A decision threshold that results in a satisfactory trade-off between these two values may be chosen¹⁹.

Example 6.4 (Trajectory classification)

The recommendations of the BIC are followed and a topology featuring fourteen hidden states is chosen for Θ_U and Θ_D , whilst ten states are chosen for Θ_S (see Figure 6.10). Validation sets, denoted by $va(\mathcal{X} - tr(\mathcal{X}))$ for $\mathcal{X} \in \{U, S, D, N\}$, comprising one hundred randomly selected elements, are used to evaluate the derived HMMs.

Classifying the trajectories of the validation sets derived from the structured data using the threshold formulation presented in (6.22) and the learnt models Θ_U , Θ_S and Θ_D of Example 6.3, results in a single misclassification. The misclassified trajectory is a member of $va(\Theta_U)$ that fails to meet the threshold values of any of the three HMMs and is attributed to the noise class as a result. The failure to correctly classify this element is a result of the brittleness of the thresholds determined for each model²⁰.

An alternative to this approach is to make use of a noise HMM and follow a competitive classifier approach. The inclusion of this fourth model results in no misclassifications on the validation set. However, this result is a fortunate consequence of the random seed that is used in determining the partitioning of the data into training and validation sets. Repeating the experiment where the random seed is allowed to vary results in a single trajectory in Θ_U being misclassified for roughly five percent of the runs. Although a perfect classification may not be possible in many cases, the simplicity of the data in this example may warrant such an expectation. Further investigation reveals that the false positives result from the trajectories being erroneously attributed to Θ_N , the

¹⁹The notions of specificity and sensitivity are well described in medical literature. In diagnostic tests for terminal diseases, it is often important to choose a threshold that does not result in many false positives.

²⁰Incidentally, selecting training sets of 50, 100, ..., 250 elements results in 1, 1, 6, 4 and 0 misclassifications respectively.

noise HMM. The large variances associated with the state-dependant distributions of this HMM coupled with the fact that there are fewer states over which to perform inference, results in this model more easily being seen to generate these false positives. In most applications (see [85, 179]) the number of hidden states in HMMs are chosen to be equal. If this approach is followed and the average BIC value is chosen for the number of hidden states for all models, then the error rate of all the classifiers is zero. ■

6.4 Summary

Hidden Markov models were discussed by way of Markov chains in §6.2. Some of the conditional independence properties which characterise these models were discussed and the state-space graphical model structure was presented. Discrete and continuous HMMs were constructed and used to generate sequences. Model parameter estimation and forward and backward probabilities were discussed. Finally, sequence classification was considered where two HMMs act as classifiers within a competitive classifying framework.

Applying this modelling approach to trajectory data provides a mechanism for classifying the data gathered by the *od-miner* of §5.2, thus completing the cycle of the *discovery component* in the decision support system framework presented in Figure 3.15. HMMs may be trained on the mined data by utilising the calculated medoids to provide an initial estimate for the state-dependent parameters of the models. A favourable topology may be chosen for each of the models by application of the BIC. Model performance may be determined with respect to a test set and the error rates may be presented to an operator. If the results are satisfactory, these HMMs may be deployed within the system to classify vessel trajectories.

CHAPTER 7

The Port of Cape Town: A case study

Contents

7.1 The data	119
7.2 Application of the origin-destination miner	125
7.3 Application of an activity classifier	129
7.4 Summary	133

In this chapter, the discovery component and activity classifier component of the decision support system proposed in Chapter 3 are considered in some detail in the context of its application to a portion of a maritime vessel data set collected over a five-month period in the region of the Port of Cape Town in South Africa. The attributes of the data set are described and the various approaches toward resolving the model components introduced in §3.4.3 and §3.4.4, and elicited further in Chapters 5 and 6, are applied to a subset of these data as a means of illustrating the practical workability of the system framework proposed in §3.4. The two components mentioned above are implemented using the methods proposed in Chapters 5–6. Finally, the efficacy of these approaches are discussed in some detail in respect of the available data.

7.1 The data

AIS data collected from a coastal antenna situated in the Port of Cape Town are considered in this chapter. All vessels of a gross tonnage of at least three hundred tons that embark on international voyages are required by the IMO SOLAS regulations to carry transponders that automatically provide information to coastal authorities and other vessels [70]. Furthermore, all domestic vessels of a gross tonnage of at least five hundred tons and all passenger vessels, irrespective of size, are also required to carry this equipment. AIS transponders typically provide automatic updates describing the identity, type, position, course, speed and navigation status of the vessel on which it is installed [184]. The AIS aims to provide an additional tool to mariners to avoid collision, particularly in inclement weather which often renders conventional radar ineffective. The signal coverage of an AIS antenna depends on the height of the antenna and may have a range of 20 nautical miles when placed on vessel radar poles.

AIS data are transmitted via *very high frequency* (VHF) channels at update intervals that depend on the speed of a vessel and the change in its course (moored vessels broadcast their position less frequently). In addition to vessel identifiers and motion-related data, static or voyage

information is also transmitted, albeit less frequently. This voyage information includes the unique vessel IMO number, a call sign, the vessel name, the vessel type, the vessel's destination port and its estimated time of arrival at the port of destination. Whereas the kinematic portion of an AIS message¹ is automatically determined from a vessel's navigational sensors, the voyage information is completed manually. Two classes of AIS transponders are currently in use, which differ in their adherence to certain performance standards. Vessels required by the IMO to carry transponders are typically fitted with *Class A* transponders whilst less expensive *Class B* transponders are often fitted to smaller vessels.

The position report of a Class A AIS transponder includes the navigational status of the vessel (*e.g.* powered vessel is under way, vessel is anchored, vessel is moored or vessel is engaged in fishing), the rate of turn of a vessel, the speed of the vessel, the longitudinal and latitudinal position of the vessel, the vessel heading and a time stamp. This study is primarily concerned with the positional reports of AIS data which are encoded after reception at a base station into a proprietary binary file format, the specification of which is provided by the South African Navy. An example of three positional reports is presented in Table 7.1.

MMSI	Speed	Longitude	Latitude	Course	Time stamp
239477000	109	18.212 53	-34.138 95	160	2010-10-02 17:33:17
239477000	107	18.247 05	-34.225 83	162	2010-10-02 18:03:37
239477000	108	18.248 08	-34.228 70	162	2010-10-02 18:04:37

Table 7.1: Three example records featuring a nine-digit unique vessel identifier, the vessel speed (in units of 10 knots), the latitudinal and longitudinal vessel position (presented in decimal degrees notation), the vessel course (measured in degrees from true north) and a time stamp. This data format is identical to that of a popular live vessel reporting website [100].

The convention is followed in the subsequent analysis of this chapter that SI units are used to represent kinematic quantities. The time stamps are converted to seconds in *POSIX* time², the speed values are represented in metres per second, and the course values measured in radians from true north. Positional data are projected onto a Cartesian plane by way of a map projection, in this case an *equidistant conic map projection*. This projection has the property that distances within a limited locality of the focal point of the projection are preserved. The quantities in the positional reports of Table 7.1 are repeated in these SI units in Table 7.2.

MMSI	Speed	x	y	Course	Time stamp
239477000	5.607	461 527.8	-3 807 230	2.792	1 286 033 597
239477000	5.504	464 223.5	-3 817 033	2.827	1 286 035 417
239477000	5.555	464 302.4	-3 817 356	2.827	1 286 035 477

Table 7.2: The speed (*m/s*), course (radians), time stamp (seconds) and positional data projected using an *equidistant conic projection* with a meridian of approximately 13.183 degrees and latitudes of -29.383 and -37.481 degrees (the latitude values specify the secant intersection of the cone with the sphere).

The data used in this case study spans the five months of October 2010 to February 2011. The numbers of unique vessels that were observed during each month are shown in Table 7.3, along with the average number of trajectories, referred to as a *tracks*³, which were observed per day⁴.

¹An AIS message is a string of bits which encode the values related to the type of report.

²POSIX time, or Unix time, is the number of seconds that have elapsed since 1 January, 1970.

³The term *track* refers to the path followed by a vessel and it comprises a number of updates. An example of an update is a GPS fix.

⁴The average number of tracks per day is computed since a single track may span several days.

The values presented in Table 7.3 were calculated before performing any data preprocessing and without any restrictions being placed on the minimum number of updates necessary to constitute a track. The raw data were encoded in binary files on a daily basis, thus splitting

Month	Number of unique vessels per month	Average number of tracks per day
Oct 2010	903	72.19
Nov 2010	865	74.66
Dec 2010	854	73.70
Jan 2011	949	85.87
Feb 2011	896	94.89

Table 7.3: The number of unique vessels observed over a five-month period as well as the average number of vessel tracks observed per day.

tracks between data files corresponding to different days. This practice explains why there are more tracks on average than there are unique vessels. Furthermore, vessels (such as tug boats and fishing vessels) that use the port as their operating base may contribute several distinct tracks over the course of a month.

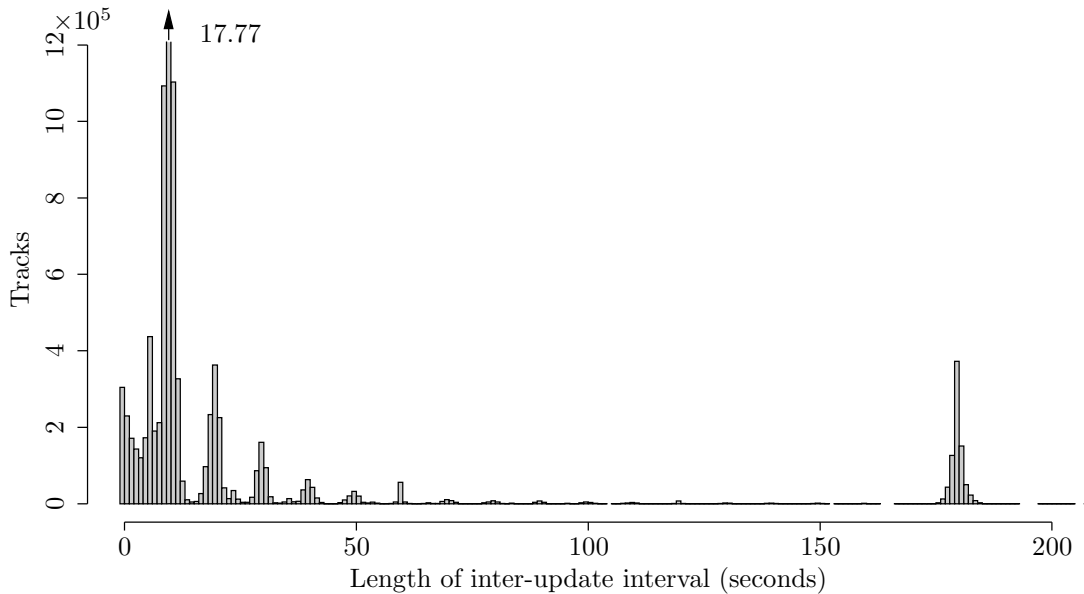


FIGURE 7.1: The time intervals at which updates occur within the data set.

The reporting interval of Class A AIS devices is three minutes for moored vessels, whilst vessels that are under way commonly transmit positional updates every two to fifteen seconds. Class B devices provide updates every thirty seconds, but if the speed of the host vessel drops below one metre per second, then a three-minute inter-update interval is adopted. The reporting intervals were computed over the entire data set with respect to each track (in this instance, tracks comprising single updates were discarded). It may be seen in Figure 7.1 that a number of updates coincide approximately with the three-minute boundary, as expected, but that the majority of updates occur at nine-second inter-update intervals. Figure 7.1 also reveals that there are a number of duplicate time stamps in the data set. More specifically, 3.2% of the updates in Figure 7.1 are duplicates. Further investigation of the time stamps revealed that an additional percentage of update intervals are negative.

In comparison, vessel speed updates exhibited far fewer spurious observations, as may be seen in

Figure 7.2. A large portion of the reported speeds are zero, but that is to be expected for vessels that are anchored or moored. The fact that this does not correspond to a greater number of update intervals of three minutes in Figure 7.1 is a consequence of the fact that an AIS transponder generates a report every three minutes when the navigational status is manually configured to the moored state.

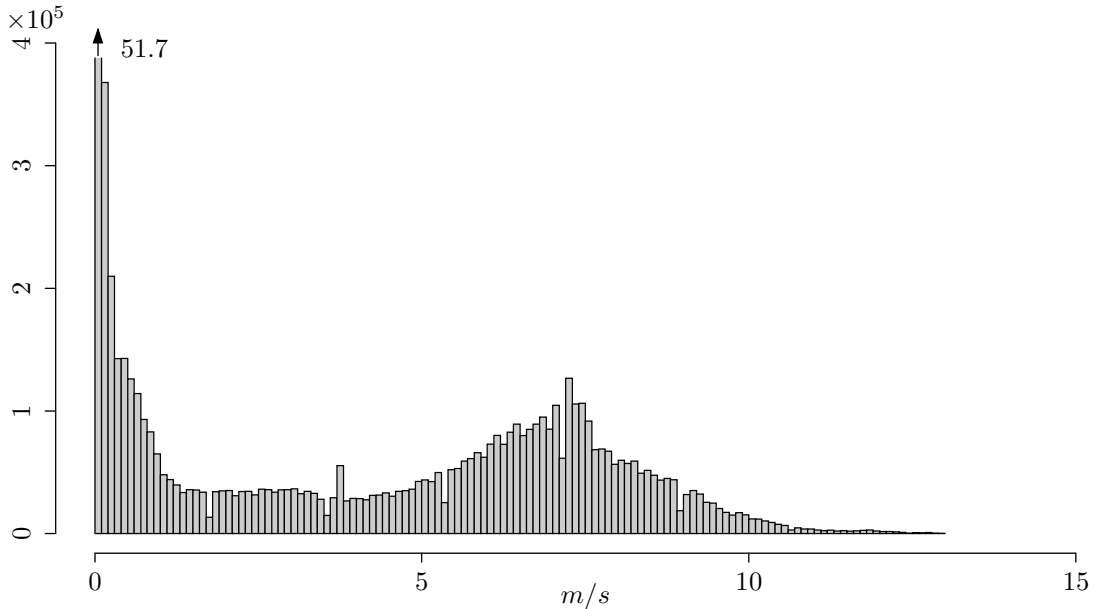


FIGURE 7.2: The frequency of speed reports over all tracks collected during the five-month period.

The data set comprises a total of 12 910 daily tracks consisting of 11 578 572 updates. It is clear from Figure 7.1 that data preprocessing was required. The following ten steps were performed during the process of data cleaning and data preprocessing, as illustrated in Figure 7.3:

Step 1. Parameter bounding was enforced during the conversion to SI units. The longitude and latitude values were constrained by the map projection library while the non-negativity of the course (referred to as heading) and speed values was enforced (for the i -th report of a track, these are denoted by h_i and s'_i respectively) in the sense that tracks containing updates violating these constraints were discarded.

Step 2. Similarly, tracks containing non-monotonically increasing time stamps were also discarded.

Step 3. In order to avoid introducing artificial segmentation of tracks generated by the same vessel as a result of the data storage format, daily segments of tracks sharing the same MMSI number were merged⁵.

Steps 4–5. Additional data cleaning operations were performed in the guise of duplicate filtering, where successive updates containing either duplicate time stamps (step 4 in Figure 7.3), or duplicate position and heading values when the vessel speed is greater than zero (step 5 in Figure 7.3), were removed.

⁵It became clear during this study that MMSI numbers are not an ideal choice for identifying vessels uniquely as these values may be attributed to more than one vessel over a period of time. Using the IMO number is a better choice.

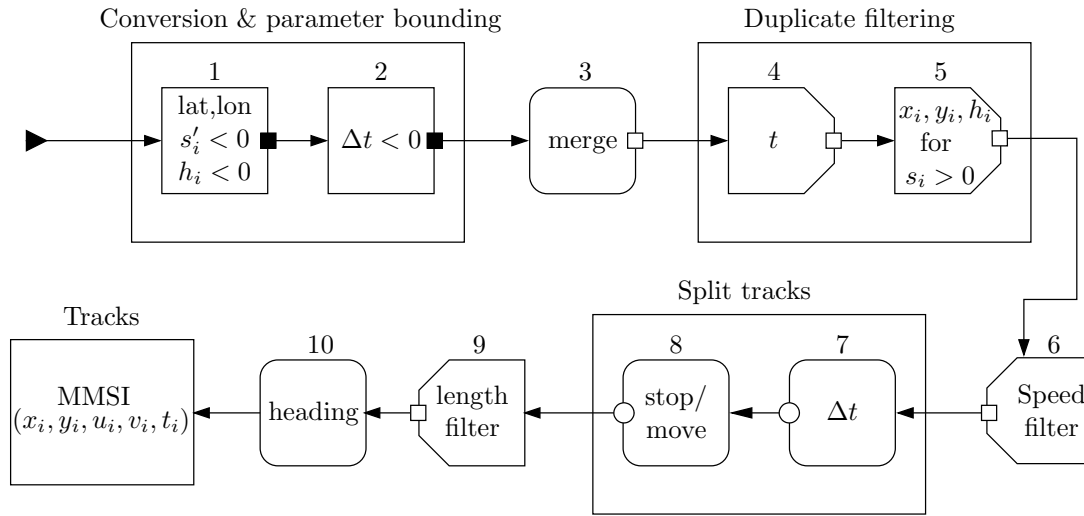


FIGURE 7.3: The data cleaning and preprocessing operations. Filled squares describe junctions that result in entire tracks being discarded, unfilled squares indicate that tracks have been reduced and circles indicate that the number of tracks have increased at that juncture.

Step 6. Two speed threshold filters were applied which discard updates subject to a common threshold of 36 m/s (70 knots). The selection of this threshold was based upon the speeds attainable by some of the fastest vessels currently in operation⁶. This approach was favoured over an outlier removal approach due to the fact that fast vessels would be discarded by the latter method as these vessels will often be under-represented in the data set⁷. The first filter discards any updates with reported speeds in excess of the threshold, whilst the second filter calculates the average linear velocity between two successive updates, and discards updates that have a derived speed in excess of the threshold. This step may be considered as the last step of the data cleaning phase (steps 1–6 in Table 7.3) and these data may be stored as tracks within a database.

Processing phase	Tracks (Υ)	Updates (ζ)	% of ζ discarded
0	12 910	11 578 572	–
1	12 073	10 355 092	10.57
2	11 583	9 580 571	6.7
3	2 994	9 580 571	–
4	2 994	9 480 388	0.87
5	2 994	9 020 664	3.97
6	2 994	9 009 232	0.1
7	75 108	9 009 232	–
8	169 414	9 009 232	–

Table 7.4: The number of tracks and updates in the data set after application of the first eight data cleaning and preprocessing steps presented in Figure 7.3.

⁶The Norwegian Skjold-class corvettes are capable of speeds in excess of 30 m/s in calm seas, and 20 m/s in rough seas. Some super yachts are capable of reaching speeds of 35 m/s .

⁷In the case study data set it was found that outlier removal produced relatively low thresholds, even if the updates of zero speed were not considered. There is simply a dearth of low-speed reports, as may be seen in Figure 7.2.

Steps 7–8. Two operations were applied partitioning tracks into parts were applied. The first of these operations (step 7 of Figure 7.3) is based on the time that elapses between updates (tracks were separated into different parts if the time interval was chosen as 1080 s, or 18 minutes in Table 7.4), and the second (step 8 of Figure 7.3 partitions tracks into stop and move segments with respect to a minimum threshold speed (selected as 0.54 m/s in Table 7.4). Vessels reporting to be travelling at these speeds were deemed to have stopped.

Step 9. An optional length filter may be applied to further reduce the data by excluding tracks that are not of a sufficient length (step 9 in Figure 7.2). For the implementation of the origin-destination miner, discarding tracks of a limited number of updates would be computationally beneficial. Tracks of fewer updates are, however, introduced by the partitioning process of steps 7 and 8, but approximately 80% of the tracks remaining after having applied steps 7 and 8 of the data preprocessing process were found to have a total number of updates in the interval of [1, 31].

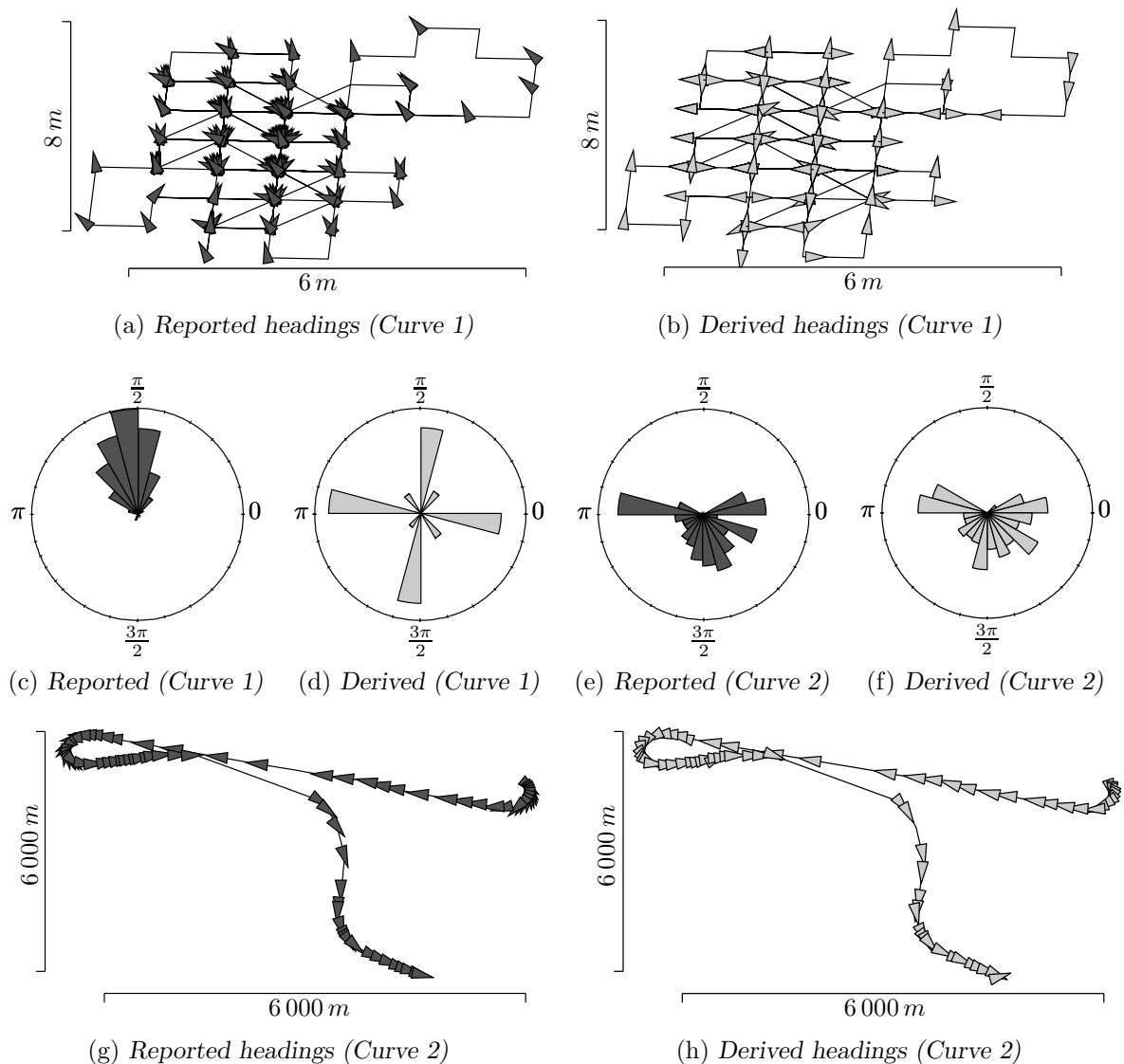


FIGURE 7.4: A comparison of the reported and derived headings of two vessels, represented by oriented triangles (light-grey and dark-grey respectively). This comparison is performed for an anchored vessel in (a)–(b), and for a vessel that is under way in (g)–(h). The associated normalised histograms of heading for the former is shown in (c)–(d), and for the latter in (e)–(f).

Step 10. Lastly, the reported heading h_i and speed s_i were replaced with a heading vector (u_i, v_i) . The reported headings were derived from successive position reports. Although there were no significant differences between the reported headings and derived headings when a vessel was travelling at speed, this was not the case for vessels that were stationary. This phenomenon is depicted in Figure 7.4 where the reported headings and derived headings of a moored vessel and a vessel underway are compared (reported headings are depicted by the darker shade of grey). The derived headings of the anchored vessel in Figure 7.4(b) are more evenly distributed in the corresponding normalised histogram of the heading plot (Figure 7.4(d)) whilst the reported heading in Figure 7.4(a) is as expected (Figure 7.4(c)). Comparing the headings of a vessel that is underway reveals that there is little difference between their headings in Figure 7.4(g) and Figure 7.4(h), and in the normalised histograms of the heading plots. Derived headings may prove useful in refining the mechanism by which vessels that have stopped are determined⁸.

Steps 7, 8, 9 and 10 of Figure 7.3 were typically applied as needed. For instance, it was often required to segment curves with respect to a different inter-update time interval in step 7 of the analysis. The culmination of the data cleaning process (steps 1–6), track segmentation (steps 7–8) and heading derivation (step 10) was a set of tracks, each consisting of an ordered list of updates, the i -th of which is denoted by $\zeta_i = (x_i, y_i, u_i, v_i)$, where x_i and y_i represent position coordinates, and u_i and v_i denote velocity coordinates.

For illustrative purposes, the tracks recorded during the months of October and November 2010 are considered in greater detail in the remainder of the chapter, so as to avoid excessive data clutter. The majority of tracks observed during these two months that were discarded on account of non-monotonically increasing time stamps, were recorded during the first two weeks of the month of October⁹. Raw (unprocessed) position data for the months of October and November 2010 are shown in Figure 7.5(a) and 7.5(c). The data preprocessing in steps of 1–7 described above were applied to produce the cleaned data sets in Figures 7.5(b) and 7.5(d). The considerable distances at which AIS signals are still received by the antenna is clear in these figures. This range is reported to depend on weather conditions [189]. It is also evident that occlusions hamper the collection of reports in the lower regions of Figures 7.5(b) and 7.5(d).

A partition of tracks into stop and move segments (step 8 of Figure 7.2) was performed on these data. Furthermore, all updates received within a radial distance of 24 km of Cape Town harbour were retained. In addition to cropping operations, it is often necessary to perform clipping or masking operations which remove all tracks within a specified region. The clipping operation is useful when discarding reports from the harbour areas as they are typically numerous. Vessel traffic lanes are immediately apparent in Figure 7.5 (where the direction of vessel motion is indicated by arrows). It is also apparent that vessels anchor offshore in addition to within the harbour itself. Magnifying the region within the black rectangle of Figure 7.5 reveals the interesting motion patterns that result from these vessels, as shown in Figure 7.6.

7.2 Application of the origin-destination miner

The origin-destination miner, as described in Chapter 4, is considered in this section as a modelling approach that may reside within the discovery component of the proposed DSS. Contrary

⁸A reduced number of reports were plotted for the sake of illustration.

⁹Personal communication with a maritime expert revealed that an error had been discovered in the timestamps attributed to some of the tracks recorded during this period [189]. Most of these tracks were simply duplicated in that a 24 hour period.

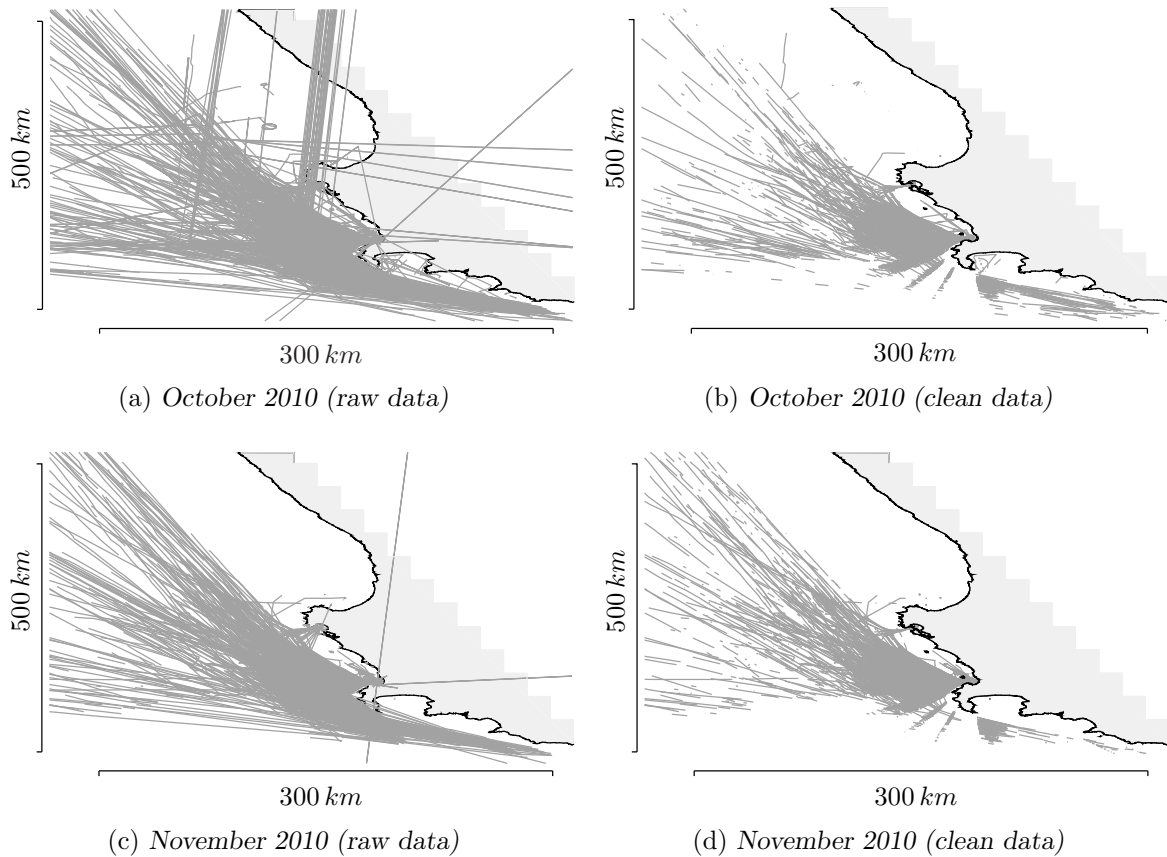


FIGURE 7.5: Comparison of raw and cleaned data for October and November, 2010.

to traffic surveillance applications, vessel trajectories are geographically less constrained. Furthermore, it is expected that vessels travelling along established routes should be well represented in origin and destination clusters¹⁰. The reduced subset of data for the months of October and November 2010 described in §7.1 is considered in this section. These data are subject to a minimum length constraint for each of the reported tracks (thirty two reports were required within a radial distance of 24 km of Cape Town in this case). Trajectories featuring fewer updates were discarded and the remaining data were linearly interpolated, assuming constant velocity between vessel reports, at intervals of sixty seconds. Interpolation makes comparison by way of DTW more effective whilst the possibility that long line segments between updates are introduced, is reduced by the application of the length filter. More importantly, it induces reasonable state durations in the HMM. Furthermore, vessel reports received within close proximity the harbour were removed by applying a region mask. Vessels within the harbour travel at very low speeds, which results in an over representation of kinematic reports from that area. The resulting tracks are depicted in Figure 7.8 as dark grey line segments, whilst the portions of the tracks that were removed on account of the harbour mask, are presented as light grey line segments.

The approach proposed in §5.2 was followed. The Douglas-Peucker poly-line simplification technique was applied to reduce trajectories to a subset of points with respect to a tolerance of 1000 m, and with respect to the radial viewport of 9 km. Thereafter, DBSCAN was applied to the resulting points, where the minimum cluster support was specified as thirty two points and

¹⁰A DSS would likely apply viewport clipping or have particular models in operation over certain areas of interest. These viewports would produce origin and destination regions at their boundaries.

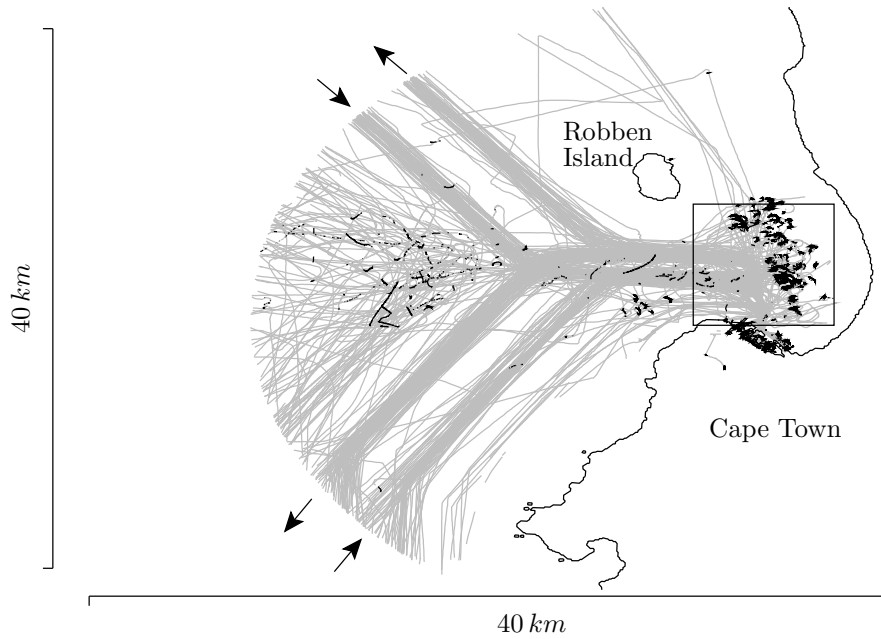


FIGURE 7.6: AIS position reports during the month of October 2010 within a 24 km radius from the Port of Cape Town, partitioned into stop (black) and move (grey) segments .

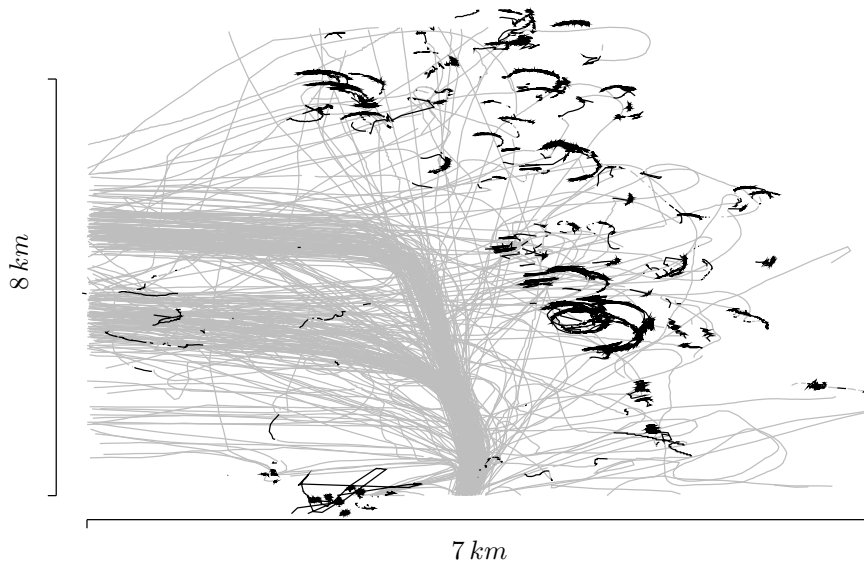


FIGURE 7.7: A magnification of the region contained within the rectangle in Figure 7.6.

the neighbourhood radii as 400 m. These operations resulted in the five clusters presented in Figure 7.9(a), with all remaining points being attributed to a noise cluster. A limited number of examples of the cluster sequences are presented in Table 7.5. Tracks which have the same origin and destination are not considered, and any tracks that feature points attributed to the noisy cluster are also omitted. It is noted that a more fine grained approach may be applied in cases where there is an abundance of data, where tracks are segmented with respect to their intersection with intermediate clusters.

Upon completion of the spatial clustering phase, the operator or analyst is expected to intervene and assign labels to the regions, as described in the system framework of Figure 3.15. These

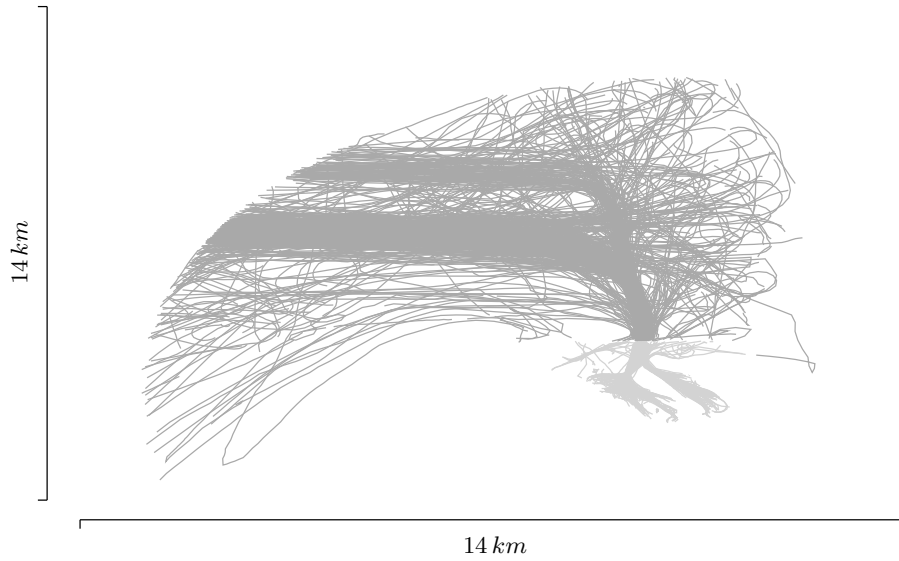


FIGURE 7.8: A reduced set of AIS position reports within a 9 km radius of the Port of Cape Town. The portions of the tracks that fall within a harbour mask are discarded (light grey line segments).

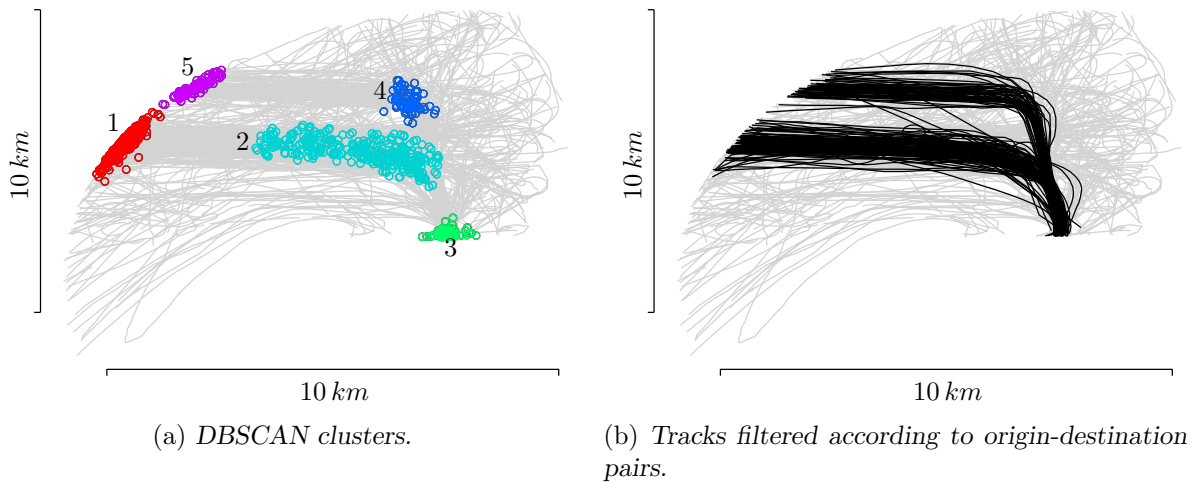


FIGURE 7.9: (a) Application of the DP technique and DBSCAN results in five clusters. (b) Filtering the tracks with respect to the cluster pairs (1,3) and (3,5) produces two track clusters.

labelled or selected regions serve to further reduce the data set. The discovery component may also make recommendations based on the frequency of observed sequences. The frequencies of the twelve most popular sequences in Table 7.5, are shown in Table 7.6, where it may be seen that, in the absence of the noise cluster, the origin-destination pairs of (1,3) and (3,5)

Cluster visitation sequences							
123	345	123	120	120	03	12	223
345	12	23	00	523	123	340	10
003	100	00	05	00	03	12	23
01	10	023	10	12	23	1003	1203

Table 7.5: A subset of cluster visitation sequences that are attributed to tracks with respect to the clustering of Figure 7.9(a), where the noise cluster is denoted by zero.

feature prominently. The former cluster pair captures vessel tracks that arrive at the edge of the viewport in Figure 7.9(a) and navigate to the harbour, whilst the latter describes vessels departing from the harbour and travelling to the edge of the viewport. The clusters of tracks that result from these pairs are referred to as *entry* and *exit* clusters, and are highlighted in Figure 7.9(b).

	Sequence frequency											
Sequence	123	00	10	345	03	12	23	120	100	023	223	003
Count	101	89	75	56	52	44	33	31	23	18	18	14

Table 7.6: The twelve most frequently occurring visitation sequences that result from the application of DBSCAN to the data in Figure 7.8.

An additional filter of sinuosity was applied and any tracks which achieved a sinuosity ratio outside the 1.5 inter quartile range were discarded. The extracted sequences for the entry and exit clusters, as well as those tracks that were deemed to be outliers with respect to sinuosity, are shown in Figure 7.10(a) and Figure 7.10(b), respectively. Lastly, these data were partitioned using PAM and DTW. No further curves were, however, discarded as the silhouette plot revealed the clustering to be satisfactory. The medoids that resulted from this clustering are depicted in Figure 7.11.

These tracks were then used to provide training data and testing data for two HMMs, as described in more detail in the next section.

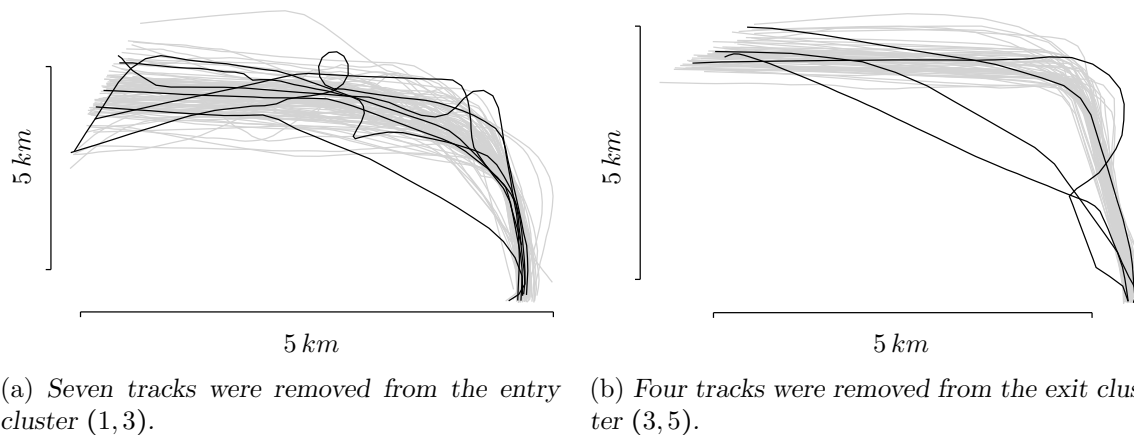
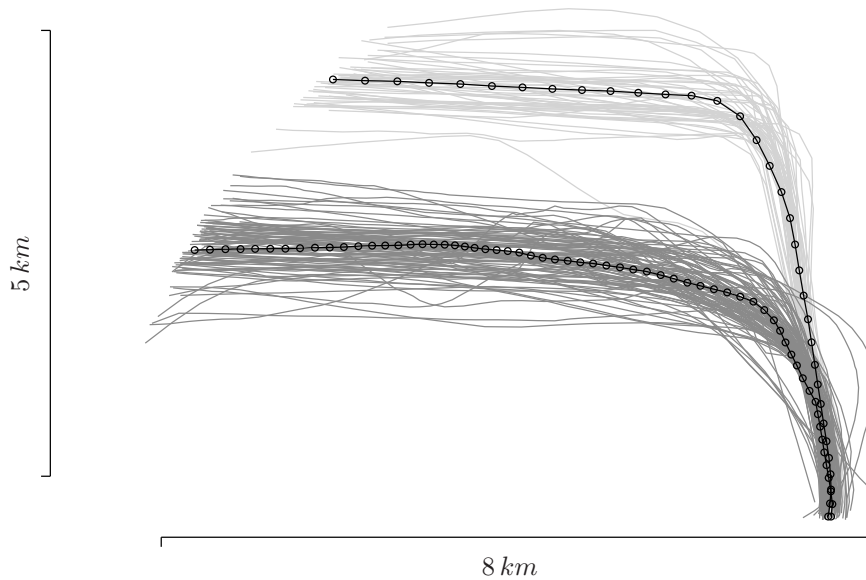


FIGURE 7.10: Outlier tracks with respect to the sinuosity measure (black) were not considered for training.

7.3 Application of an activity classifier

A *left-to-right* HMM structure was utilised to capture the sequential nature of the vessel tracks in the entry and exit clusters uncovered in §7.2 using the methodology described Chapter 6. Each cluster of tracks was modelled using an HMM featuring continuous state-dependent distributions. These state-dependent distributions were modelled as multivariate Gaussian distributions over vessel updates. The choice of distribution was informed by application of the BIC on arbitrary point clouds that were created with respect to the medoids (see §6.3) and by a similar approach followed in traffic surveillance [93]. Determining a suitable number of hidden states for the HMM using BIC was complicated by the fact that the number of observations per track is not

FIGURE 7.11: *The medoids determined by PAM and DTW.*

the same for all tracks within a particular cluster. Although heuristic approaches may be sought to resolve this problem (for example, by selecting a subset of tracks of similar sample length as a mechanism to guide the initial model selection search), it is ultimately necessary to perform iterative model selection thereafter (as mentioned in Example 6.3). As a reduced set served as the data set and the goal was to show workability of the approach, the number of states was chosen arbitrarily. Contrary to the approach taken in [93], where only position and velocity were used, additional models with bivariate Gaussian state-dependent distributions, were also considered for the sake of comparison.

Four HMMs were trained on the entry and exit clusters of Figure 7.12(a). The HMMs Θ_{en4} and Θ_{ex4} were allocated twenty multivariate Gaussian state-dependent distributions for modelling observations (x_i, y_i, u_i, v_i) , and Θ_{en2} and Θ_{ex2} featured nine bivariate Gaussian state-dependent distributions modelling only positional observations (x_i, y_i) . The entry and exit cluster data were partitioned into a training set and a test set,¹¹ with seventy five percent of instances serving as training data and twenty five percent serving as test cases (see Table 7.8). The medoids of Figure 7.11 provided the reference curves upon which the individual updates were partitioned into point clouds, and from which the sample means and covariances were estimated (see §6.3 for a description of the method). The sampled mean-centred contour plots of the bivariate Gaussian state-dependent distributions for Θ_{ex2} and Θ_{en2} are shown in Figure 7.12(a). The estimated estimation means and covariances, as determined by the convergence of the EM algorithm to a local optimum, are presented for in Figure 7.12(b)). Similarly, the sample means and covariances of the state-dependent distributions of Θ_{ex2} and Θ_{en2} are presented in Figure 7.12(c) and their refined estimates are shown in Figure 7.12(d).

Vessels approaching the harbour tend to reduce their speed whilst those that are exiting have no reason to do so. It may be seen from the estimates that the majority of states are concentrated in the narrow harbour entry channel (there are more vessel reports within that confined region). In modelling this region well, the EM algorithm is able to attain a greater log-likelihood value

¹¹In an earlier approach [33], these data were partitioned into a training set, validation set and test set, after which the threshold to determine class membership was chosen based on the performance of the HMMs with respect to their respective validation sets. However, the threshold may be selected directly from the training data since the worst performing track is used as the threshold value.

over the sequences. Fewer vessel reports emanate from the latter part of the exit cluster on account of the vessels departing the harbour at greater speed. This phenomenon is also present in the estimated parameters of the corresponding exit HMMs, where the contour plots are more distant from one another.

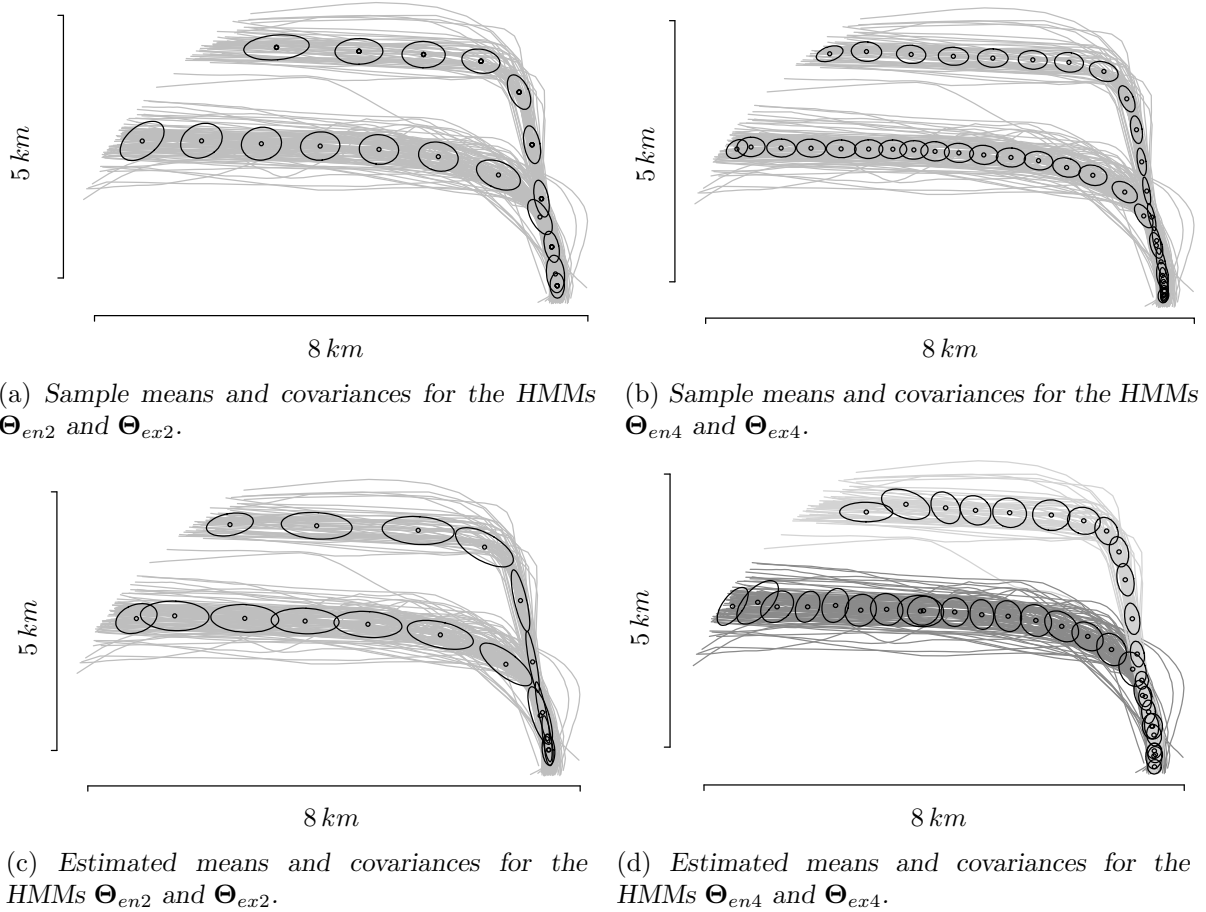


FIGURE 7.12: A two-dimensional representation of the sample means and covariances for the state-dependent distributions of the HMMs. (c)–(d) The means and covariances of the state distributions as determined by model training.

The initial state distributions of the HMMs were chosen as $\boldsymbol{\pi} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \dots, 0)$ so as to account for the possibility that the first vessel report of a vessel departing the harbour may be deemed to have more likely been generated by a state other than the first. This initial distribution is re-estimated by the EM algorithm and in the case of the entry HMMs it reverts to $(1, 0, \dots, 0)$. However, for the exit HMM, Θ_{ex4} , this is not the case¹².

A vessel track, $\Upsilon = \zeta_0, \dots, \zeta_n$ was deemed to be a member of class \mathcal{C}_i if the probability of the sequence of observations, given the corresponding HMM Θ_i , was less than the selected threshold for the i -th class. In this case, the HMMs Θ_{en2} and Θ_{en4} model the entry class, while Θ_{ex2} and Θ_{ex4} model the exit class.

Each HMM was estimated on the corresponding training instances in Table 7.8 and membership of a track to one of the classes was granted if the track overcame the acceptance threshold for that particular class. This threshold was determined directly from the training data as the

¹²The HMM toolbox that was used in this application considered zero values as constants that did not require estimation during the optimisation process [118].

	Total tracks	Training set	Test set
Entry cluster	97	74	23
Exit cluster	53	40	13
Unclustered	280		

Table 7.7: *The number of vessel tracks in the clustered and unclustered sets.*

minimum log-likelihood attained on instances in the training set, given the model. That is, a track Υ was deemed to be a member of the class described by the HMM Θ_i when

$$\log p(\Upsilon|\Theta_i) > \min_{X \in \text{tr}(\mathcal{D})} \log(p(X|\Theta_i)).$$

This method of determining the acceptance threshold is simplistic, but serves the purpose of demonstrating the viability of this approach¹³. A *one-versus-all* classifying approach was taken to classify tracks from the test sets of Table 7.8 and tracks that were not clustered during the data mining process of §7.2. The results of this experiment are presented in Table 7.8. Consider the HMM Θ_{en2} for the sake of explanation. Of the twenty three instances in the test set of the entry class, all of them were correctly classified as members of the entry class and none of them were incorrectly classified as non-members — that is, the classifier produced a *true positive* (TP) count of 23 and a *false negative* (FN) of zero. Attempting to classify the test data for the exit class using HMM Θ_{en2} , resulted in a *true negative* (TN) count of thirteen (all the tracks in the exit class) and a *false positive* (FP) count of zero. This is to be expected since the HMMs are very restrictive with respect to which states are possible starting states. Considering that the tracks from the two classes originate at such distant locations, this should always be the case.

	Test sets				Unclustered set	
	TP	TN	FP	FN	TN	FP
Θ_{en2}	23	13	0	0	230	50
Θ_{ex2}	12	23	0	1	276	4
Θ_{en4}	23	13	0	0	236	44
Θ_{ex4}	11	23	0	2	276	4

Table 7.8: *The classification results of the entry and exit test sets for the two groups of HMMs, as well as classification results for the unclustered set.*

The performance of both formulations of the entry classifiers on the entry data was markedly poorer than that of the classifiers trained on the exit data. The reason for this was the manner in which vessels approaching the harbour behave. As they await a berth, they tend to reduce speed in the entry lane, and in some instances, perform loops whilst waiting. These low speeds meant that these tracks would often be subdivided during the stop move preprocessing step. The tracks which qualified as false positives of the classifier associated with Θ_{ex4} are presented in Figures 7.13(a) and 7.13(b). It is important to recall that the classes were implicitly defined by the clustering step and so classifier performance is measured relative to that partition.

Lastly, the possibility of identifying a track in real-time is demonstrated by way of an example. A track was selected from the unclustered data set and the threshold value for Θ_{en4} was used. The track of Figure 7.14 is evaluated at each update. The updates for which it is correctly classified by the entry HMM are indicated by open circles, whilst the updates for which it is not longer described by the HMM are indicated by solid circles. When considering partial tracks, a more

¹³Alternatively, a threshold range may be determined, as in [77].

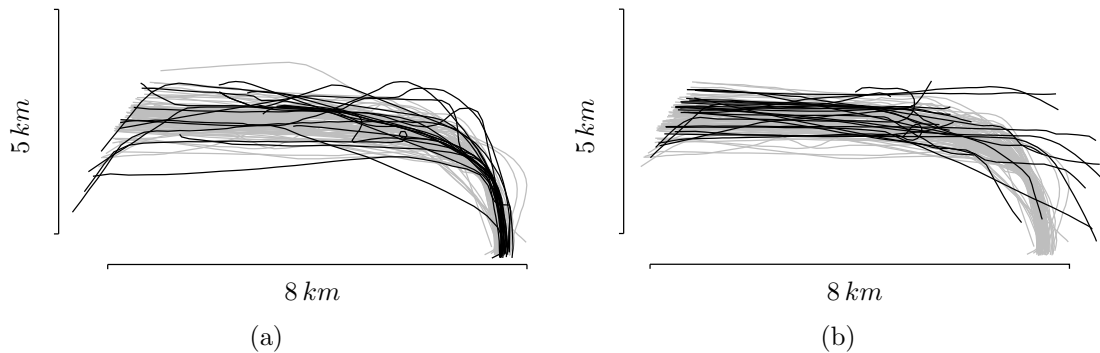


FIGURE 7.13: The misclassifications of Θ_{en4} with respect to the calculated threshold. (a) The tracks that end in the harbour area and (b) the tracks that do not terminate at the harbour.

reasonable approach to evaluating the acceptance threshold may be to scale the log-likelihoods with respect to the track length. However, the brittleness of the threshold will remain. As was discussed in §6.3, the construction of a noise model may also be pursued.

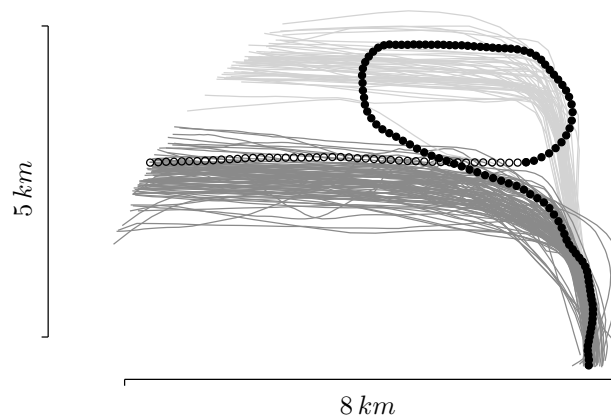


FIGURE 7.14: A vessel deviating from its lane is classified at $t = 43$ as no longer being a member of the entry class associated with the HMM Θ_{en4} .

7.4 Summary

The purpose of this chapter was to demonstrate, in the context of a case study involving real maritime vessel data, the workability of the decision support framework proposed in Chapter 3.

AIS data collected over a five-month period in the region of Cape Town were analysed. The properties of the vessel reports were discussed and some details of the reporting frequency and message contents within these data were considered. A ten-step approach to cleaning and processing the data was proposed, motivated and discussed. This approach was applied to the data set to illustrate the effect it had on the number of vessel tracks and vessel reports ultimately retained. The quality of the course information was also investigated.

Thereafter, an origin-destination miner was applied to a subset of AIS data from the months of October and November 2010. DBSCAN was used to identify entry and exit clusters that coincide with Cape Town harbour and the edges of a 9 km viewport. Following the methodology of Chapter 5, these clusters were used to select a subset of trajectories that were further reduced

by outlier removal with respect to sinuosity. PAM was used to recover the final clusters and these data served as training data for HMMs.

HMMs were subsequently trained with respect to the entry and exit clusters based on both two and four observations per datum. The difference in the performance of these models were found to be small in this case. The classification results were presented for each HMM with respect to a cluster test set. Finally, a single example of real-time classification was provided using this simple framework.

CHAPTER 8

Conclusion

Contents

8.1	Dissertation summary	135
8.2	An appraisal of the dissertation contributions	136
8.3	Suggestions for future work	137

A summary of the research conducted in this dissertation is presented in the first section of this chapter. Thereafter, a brief appraisal of the dissertation contributions is conducted in §8.2, and suggestions for future work are made in §8.3.

8.1 Dissertation summary

The rights and responsibilities of coastal nations in terms of ensuring safe, innocent passage of international vessels in their waters and protecting the marine environment were briefly discussed in Chapter 1 within the broader context of international treaties. In addition to the effects that these agreements had on the zoning of waters around coastal states, the provisions made for law enforcement (such as the principle of hot pursuit) and crimes committed at sea, were described. Within this context, South Africa's maritime responsibilities were discussed insofar as they extend to fisheries management, search and rescue, monitoring and enforcement and operations within South African and Mozambican waters. The disparity between available resources for pursuing safety and security within South Africa's vast area of responsibility has contributed to the identification of the problem studied in this dissertation. Finally, the dissertation scope and objectives were outlined.

The current state of the art in respect of the analysis and learning of motion patterns was reviewed in Chapter 2 in fulfilment of dissertation Objective I of §1.6. Two general methodologies towards classification systems were described, namely the use of expert systems and the adoption of machine learning approaches. Particular attention was paid to behaviour analysis in the fields of computer vision and ecological modelling as well as in the maritime domain. These analyses typically rely on kinematic data. Various notions that are central to the topic of the dissertation were also introduced in this chapter by way of a discussion surrounding maritime motion patterns and threat assessment (in fulfilment of dissertation Objective II of §1.6).

Maritime surveillance was considered from a systems perspective in Chapter 3. The uses of and reasons for pursuing a decision support system and data fusion were briefly discussed and constituent elements of a typical maritime surveillance system were reviewed. This discussion

included a high-level consideration of the notion of normality and threats within a maritime surveillance scene. Existing maritime surveillance systems were described and a novel maritime surveillance decision support architecture was proposed in fulfilment of dissertation Objective III of §1.6. The constituent rule-based component, discovery component and activity classifier component of this system were discussed in detail (in fulfilment of dissertation Objectives III(a)–III(c) of §1.6).

Approaches toward populating the system components presented in Chapter 3, were proposed in Chapters 4–6 in fulfilment of dissertation Objective IV of §1.6. The rule-based component was considered in Chapter 4 and three rule classes, namely zone infractions, proximity alarms and anomalous actions alarms, were identified for inclusion in the decision support system of Chapter 3 by considering geographical and other constraints to motion. Examples of rules were provided for each of the rule classes. These rules address the problem of identifying known threats.

In Chapter 5, a simple data mining approach was suggested for inclusion in the decision support system designed in Chapter 3 and its efficacy was demonstrated with respect to a synthetic data set. This approach was specifically designed to extract origin destination pairs within a particular region of interest. This was achieved by thinning the trajectories in the data set and then performing DBSCAN to identify spatial clusters. These clusters were used to reduce the data set by requiring that a trajectory should begin and end in one of these clusters. Additional data thinning was performed using the feature of sinuosity and the resulting data set was clustered using a partitioning around medoids method. These medoids were considered to be route representatives.

Sequential data modelling using the framework of hidden Markov models was investigated in Chapter 6. Properties of the models were discussed and a few examples were presented. These examples demonstrated the generative and classification properties of these models in the context of a synthetic data set. Their use in competitive classifying tasks were considered in these examples, as well as the efficacy of their use as classifiers when adopting thresholds to classify trajectories.

A case study of vessel reports collected over a five-month period in the region of Cape Town was conducted in Chapter 7 in fulfilment of dissertation Objective V of §1.6. These data were considered in some detail, revealing some of the properties of the data set and the vessel reports. Data considerations were limited to the kinematic portion of the AIS reports. The data pre-processing operations were presented and their application to the data was discussed. The data mining approach described in Chapter 4 was used to extract a subset of data and the time-series classification approach of Chapter 6 was applied to these data.

8.2 An appraisal of the dissertation contributions

Although the literature review of motion patterns in Chapter 2 does not contain novel concepts, it was intentionally broad in its review of behaviour analysis in an effort to explore similarities between various approaches in different application areas. There is a succinct introduction to the maritime domain in the literature which contains a review of selected methods [104].

The problem of maritime surveillance was investigated from a kinematic point of view with specific emphasis on the South African context in Chapters 1 and 4. South African maritime policing resources, surveillance capabilities and the reiteration of some of the rules used in the existing South African context, were brought together herein.

A novel DSS framework for maritime threat detection was proposed in Chapter 3. This framework was informed by existing approaches by human operators and by apparent requirements within the South African context. The acquisition of such a system was pursued by the South African Navy in 2014, but no such system was in operation at the time of writing this dissertation. The notion that the system proposed in this dissertation implicitly acts as a data collector is beneficial in the South African context where there is presently limited capability in terms of maritime surveillance. System components were populated with rules and the use of data mining and a feature classifier were proposed as examples of functional system components. This system was presented at various conferences [33, 34] and the meetings held with domain experts at the Institute for Maritime Technology. The resulting questions and discussions informed some of the subsequent design choices.

The novel concept of a weak or strong collision was introduced as a means to allow greater flexibility in existing collision rules. This concept was used to derive alternatives to the raid and pursuit rules, as they appear in the literature. The novel notion of trusted locations was also introduced as a mechanism to avoid false alarms for these rules. Three new rule classes were proposed and discussed from a theoretical standpoint within the context of existing taxonomies in the literature.

A simple, yet novel, trajectory mining approach was applied to AIS data collected around the port of Cape Town. While the classification approach of using HMMs for trajectory classification is well documented in the literature and its use is widespread across various disciplines, this application was the first of its kind in a South African context. This approach shares the use of DBSCAN with that of a very recent paper [128] published within the maritime domain.

The research documented in this dissertation resulted in a simple code-base within the R environment capable of performing much of the data processing necessary for analysis. Although the literature abounds with different methods to trajectory classification, it was found that few implementations of these approaches are available for experimentation.

8.3 Suggestions for future work

The anomaly detection approaches discussed in Chapter 2 very often assume that the observations within a particular data set are normal, while anomalous activities are classified as deviations from the models learnt on these data sets. The approach taken in this dissertation was to reduce the data set through data mining and to construct classifiers on these reduced data set. This may still be seen as anomaly detection. A single detector of this type was implemented in this study and additional detectors may be investigated. These detectors would require additions to the discovery component of the system framework, as well as to the classifier component. Examples of these detectors may be an anchoring detector, or fishing detector. Instances of the former appear in the processed data after the stop/move preprocessing step.

Additional functionality to integrate vessel types may be pursued so as to provide a natural segmentation of data into tracks per vessel class. The specification for the binary voyage data was acquired towards the end of the study and although the initial extraction functionality was implemented, vessel types were therefore not integrated into the analysis code-base. An initial investigation into the vessel types revealed that there were discrepancies between the reported vessel types and the vessel types recorded on a popular vessel tracking website [100]. As a consequence, an approach to deriving accurate vessel types using online sources such as [100] in conjunction with the reported vessel types, may prove to be beneficial. A comparison of the information obtained from vessel reports and that of the vessel types and descriptions available

in [100] is presented in Table 8.1.

Vessel	Reported Vessel Type	Web Vessel Type [100]	Web Description [100]
232365000	Cargo (hazardous category A)	Container	Container ship
235051174	Cargo (hazardous category C)	Dry Cargo	Container ship
236335000	Tanker	Tanker	oil/chemical tanker
257583000	Tug	None	Multi purpose offshore vessel
304010886	Cargo (hazardous category A)	Dry Cargo	General Cargo ship
353776000	Passenger	None	Crane ship
353878000	Cargo (hazardous category A)	Dry Cargo	Container ship
354093000	Cargo	None	Bulk carrier
356946000	Cargo	Dry cargo	Container ship
372848000	Tanker	Tanker	Crude oil tanker
377383000	Dredging or underwater ops	Dredger	Mining
477323000	Cargo (general)	Cargo	General cargo
538003274	Cargo (hazardous category A)	Container	Container ship
601045000	Other	Other	Fishery research vessel
601524000	Tug	Tug	Tug
636013898	Cargo	Dry Cargo	Container ship
636014280	Cargo	Dry Cargo	Decommissioned
636014671	Tanker	Tanker	Floating storage/production
636091291	Cargo (hazardous category A)	Dry Cargo	Container ship
636091312	Cargo (hazardous category A)	Dry Cargo	Container ship
636091429	Other	Unknown	anchor handling vessel
636091567	Cargo (hazardous category D)	Dry Cargo	Container ship
636092108	Tanker	Tanker	Oil/chemical tanker
645167000	Cargo	Cargo	Cable layer — decommissioned or lost
657211000	Tanker	Tanker	Oil/chemical tanker

Table 8.1: *Initial vessel type extraction was explored and the reported vessel types were compared with the descriptions of an online resource [100].*

Suitable approaches to data storage and representation should be investigated. Although the data warehouse approach advocated in this dissertation will reduce search or query times within a large data set, there may be many benefits to choosing an underlying data representation infrastructure which is well suited to temporal and geographical data. Various recent technologies are available to fulfil the role of data representation within a data warehouse, but it will be beneficial to perform a quantitative comparison between these technologies.

The decision support system for threat detection proposed in this dissertation may be considered as an underlying system within a larger maritime surveillance decision support system. Where the former identifies threats, the latter should quantify them in the sense that certain threats are of more significance than others. Meetings with subject matter experts at the Institute for Maritime Technology in Simon's Town revealed that threats exhibit varying degrees of importance to different role players within the South African context and that attributing levels of importance to them without taking the responsibilities of the various organisations into account, is not meaningful¹. For instance, the conventional threats in Figure 8.1 may be arranged on a continuum between SAMSA, the South African Navy and the DAFF. A threat evaluation DSS would need to provide decision support contingent on the responsibilities of these organisations. In addition to providing threat evaluation decision support, the larger system may provide resource allocation decision support. These three systems, namely a threat detection system, a threat evaluation system and a resource allocation system, may provide a comprehensive solution to decision support in the maritime domain. The benefit of the threat detection system's agnostic approach to threats is that data may be collected in the interim and categorised so

¹The subject matter experts elected to avoid ranking threats in a general sense.

as to inform the development of the remaining systems. For example, predictive zone entry, as performed by Ristic *et al.* [142], would be a good future avenue of investigation as a model component of a threat evaluation system (the threat detection system simply notes that the event occurs). The track annotation mechanism within the threat detection system proposed in this dissertation may be used to produce a database featuring numerous examples for further analysis.

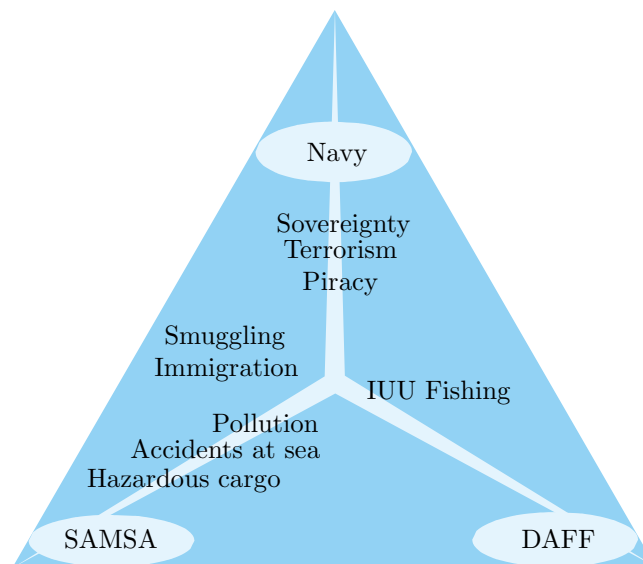


FIGURE 8.1: A conceptual view of the arrangement of a limited set of threatening behaviours with respect to the organisations under which the responsibility of action falls. The medians partition the triangle into regions of responsibility. For example, smuggling and illegal immigration may be attributed an equal level of importance by the South African Navy and SAMSA, whereas threats to sovereignty elicit a greater need for response from the South African Navy.

Research into a track manager component which fuses radar and AIS tracks is required. Although there are publications on this subject and track stitching is an active field of research, these approaches need to be integrated with the proposed framework. The effect that this will have on the quality of the data should be immediately beneficial. Furthermore, radar data are collected along the South African coast and these data may be leveraged in an attempt to devise methods for effectively determining vessel types from the observed tracks. AIS tracks, for which the vessel types are known, may serve to provide a ground truth in this regard.

The rule set proposed in Chapter 4 may be expanded and application of the rules to the data set is necessary. This will be beneficial in enhancing an understanding of the vessel behaviours in the region of interest. The scope of this study was broad in the sense that the pursuit of a system framework did not allow for comparative studies of the various modelling approaches. Further research in this regard, as well as an expansion of the HMM approach so as to consider a hierarchy of models or the use of particle filters, is necessary. Ultimately, the system interactions, such as the operator in the loop, making decisions and curating data, may only be tested once there is a sufficient modelling capability within the broader system.

Finally, refinements to the proposed data mining and classification methods are necessary in terms of a sensitivity analysis with respect to the various parameters on which these methods are contingent.

APPENDIX A

Contents of the accompanying compact disc

A brief description of the contents of the compact disc included with the dissertation is given in this appendix. The compact disc contains various scripts that were required to test the methods described in Chapter 5 and 6, as well as the code required to clean and process the AIS data used in Chapter 7.

Existing implementations of dynamic time warping, DBSCAN, PAM and HMMs were used in this dissertation. The scripting environments of MATLAB and R were utilised for the purpose of data formatting and library interactions and the code written in these environments included procedures for track cropping and masking, polyline simplifications, feature derivation and data cleaning and preprocessing. A large portion of the code was initially written in MATLAB, but much of it was later migrated to the R environment, which has a richer set of libraries available for pattern recognition.

The code responsible for the conversion of the binary data files acquired from the Institute for Maritime Technology was written in C++ and has been included despite the fact that the data files may not be shared. A track class and compilation scripts are also included¹. Lastly, during the analysis process BASH scripting was used to process the ASCII files that resulted from the binary data conversions.

The compact disc contains the following five directories:

Data sets. This directory contains synthetic data used in Chapters 5 and 6, as well as coastline data and shape files for MPAs.

Code. This directory contains four subdirectories, namely “MATLAB”, “cplusplus”, “R”, and “BASH”, which contain the various implementations in each of those languages. The code is documented and test cases are provided for many of the functions.

Libraries. This directory contains a limited number of libraries used in this dissertation. Access to libraries within the R environment is well supported by the CRAN repositories and so they are not included.

Documentation. Documentation files describing the source code are provided in this directory.

Dissertation. This directory contains an electronic copy of the dissertation.

¹Due to poor processing speeds in MATLAB, an effort was made to process the data in C++. However, R was found to be sufficiently powerful and so this approach was abandoned.

References

- [1] ANDERSON F, 2010, *Defence, peace, safety and security at the Council for Scientific and Industrial Research*, [Personal Communication], Contactable at fanderso@csir.co.za.
- [2] AZZALINI A & TORELLI N, 2007, *Clustering via non-parametric density estimation*, *Statistics and Computing*, **17(1)**, pp. 71–80.
- [3] BALLARD DH, 1981, *Strip trees: A hierarchical representation for curves*, *Communications of the Association of Computing Machinery*, **24(5)**, pp. 310–321.
- [4] BASHIR F, KHOKHAR AA & SCHONFELD D, 2007, *Object trajectory-based activity classification and recognition using hidden Markov models*, *IEEE Transactions on Image Processing*, **16(7)**, pp. 1912–1919.
- [5] BASHIR F, QU W, KHOKHAR A & SCHONFELD D, 2005, *HMM-based motion recognition system using segmented PCA*, *Proceedings of the IEEE International Conference on Image Processing (ICIP 2005)*, Genoa, pp. 1288–1291.
- [6] BASS T, 2000, *Intrusion detection systems and multisensor data fusion*, *Communications of the Association of Computing Machinery*, **43(4)**, pp. 99–105.
- [7] BAUM LE, PETRIE T, SOULES G & WEISS N, 1970, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, *The Annals of Mathematical Statistics*, **41(1)**, pp. 164–171.
- [8] BETHMANN RC & MALLOY KA, 1989, *Command and control: An introduction*, (Unpublished) Technical Report, Defense Technical Information Center, Fort Belvoir (VA).
- [9] BILMES JA, 2006, *What HMMs can do*, *IEICE Transactions on Information and Systems*, **89(3)**, pp. 869–891.
- [10] BIN L, 2006, *Behaviour of ship officers in manoeuvring to prevent a collision*, *Journal of Marine Science and Technology*, **14(4)**, pp. 225–230.
- [11] BISHOP C, 2007, *Pattern recognition and machine learning*, 1st Edition, Springer, New York (NY).
- [12] BORUM R, FEIN R, VOSSEKUIL B & BERGLUND J, 1999, *Threat assessment: Defining an approach for evaluating risk of targeted violence*, *Behavioural Sciences & the Law*, **17(3)**, pp. 323–337.
- [13] BOURY-BRISSET AC, 2004, *Ontological approach to military knowledge modelling and management*, (Unpublished) Technical Report, Defense Technical Information Center, Fort Belvoir (VA).
- [14] BRANCH GM & CLARK BM, 2006, *Fish stocks and their management: The changing face of fisheries in South Africa*, *Marine Policy*, **30(1)**, pp. 3–17.

- [15] BRAX C & NIKLASSON L, 2009, *Enhanced situational awareness in the maritime domain: An agent-based approach for situation management*, Proceedings of the SPIE: Intelligent Sensing, Situation Management, Impact Assessment, and Cyber-Sensing, Orlando (FL), pp. 3–13.
- [16] CHANG SJ, 2004, *Development and analysis of AIS applications as an efficient tool for vessel traffic service*, Proceedings of the Oceans MTTs/IEEE Techno-Ocean Conference, Kobe, pp. 2249–2253.
- [17] CHRISTIE P, 2004, *Marine protected areas as biological successes and social failures in Southeast Asia*, Proceedings of the American Fisheries Society Symposium, Bethesda (MA), pp. 155–164.
- [18] CHUENPAGDEE R, PASCUAL-FERNÁNDEZ JJ, SZELIÁNSZKY E, ALEGRET J, FRAGA J & JENTOFT S, 2013, *Marine protected areas: Re-thinking their inception*, Marine Policy, **39**, pp. 234–240.
- [19] *Convention for the Suppression of Unlawful Acts against the Safety of Maritime Navigation*, [Online], [Cited April 25th, 2012], Available from <http://treaties.un.org/doc/db/Terrorism/Conv8-english.pdf>.
- [20] *Convention on the International Regulations for Preventing Collisions at Sea*, 1972, [Online], [Cited August 23rd, 2013], Available from <http://www.imo.org/About/Conventions/ListOfConventions/Pages/COLREG.aspx>.
- [21] COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH, 2007, *AwareNet: Persistent, ubiquitous surveillance technologies for enhanced national security*, [Online], [Cited June 7th, 2010], Available from www.csir.co.za/dpss/pdf/protect%5C_waters.pdf.
- [22] DE VRIES GKD, MALAISÉ V, VAN SOMEREN M, ADRIAANS P, SCHREIBER G, NIJHOLT A, PANTIC M, POEL M & HONDORP G, 2008, *Semi-automatic ontology extension in the maritime domain*, Proceedings of the 20th Belgian-Netherlands Conference on Artificial Intelligence, Enschede, pp. 265–272.
- [23] DE VRIES GKD, VAN HAGE WR & VAN SOMEREN M, 2010, *Comparing vessel trajectories using geographical domain knowledge and alignments*, Proceedings of the 10th IEEE International Conference on Data Mining Workshops, Sydney, pp. 209–216.
- [24] DE VRIES GKD & VAN SOMEREN M, 2010, *Clustering vessel trajectories with alignment kernels under trajectory compression*, Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Barcelona, pp. 296–311.
- [25] DEAN W, 2012, *Navy to establish maritime awareness centres*, [Online], [Cited April 27th, 2012], Available from http://www.defenceweb.co.za/index.php?option=com_content&view=article&id=24437:navy-to-establish-maritime-awareness-centres&catid=51:Sea&Itemid=106.
- [26] DEPARTMENT OF INTERNATIONAL RELATIONS AND COOPERATION, 2003, *UNCLOS*, [Online], [Cited April 3rd, 2012], Available from <http://www.dfa.gov.za/foreign/Multilateral/inter/unclos.htm>.
- [27] DEPARTMENT OF TRANSPORT, 2009, *Address by Minister of Transport, Mr Sibusiso Ndebele, at the opening ceremony for the Centre for Sea Watch and Response, Cape Town*, [Online], [Cited April 19th, 2012], Available from <http://www.info.gov.za/speech/DynamicAction?pageid=461&sid=5291&tid=5436>.

- [28] DIETTERICH T, 2002, *Machine learning for sequential data: A review*, Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, London, pp. 15–30.
- [29] DILLON DR, 2005, *Maritime piracy: Defining the problem*, SAIS Review, **25(1)**, pp. 155–165.
- [30] DODGE S, WEIBEL R & FOROOTAN E, 2009, *Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects*, Computers, Environment and Urban Systems, **33(6)**, pp. 419–434.
- [31] DODGE S, WEIBEL R & LAUTENSCHÜTZ AK, 2008, *Towards a taxonomy of movement patterns*, Information Visualization, **7(3)**, pp. 240–252.
- [32] DOMINGOS P, 1999, *The role of Occam's Razor in knowledge discovery*, Data Mining and Knowledge Discovery, **3(4)**, pp. 409–425.
- [33] DU TOIT J & VAN VUUREN JH, 2014, *Semi-automated maritime vessel activity detection using hidden Markov models*, Proceedings of the 43rd Annual Conference of the Operations Research Society of South Africa, Parys, pp. 71–78.
- [34] DU TOIT J & VAN VUUREN JH, 2012, *Towards coastal threat evaluation decision support*, Proceedings of the 41st Annual Conference of the Operations Research Society of South Africa, Buffelsdrift, pp. 31–39.
- [35] ELZINGA P, POELMANS J, VIAENE S, DEDENE G & MORSING S, 2010, *Terrorist threat assessment with formal concept analysis*, Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI), Vancouver, pp. 77–82.
- [36] ENDSLEY MR, 2000, *Theoretical underpinnings of situation awareness: A critical review*, pp. 3–32 in ENDSLEY MR & GARLAND DJ (EDS), *Situation Awareness Analysis and Measurement*, CRC Press, Boca Rotan (FL).
- [37] ESTER M, KRIEGEL H, SANDER J & XU X, 1996, *A density-based algorithm for discovering clusters in large spatial databases with noise*, Proceedings of the Knowledge Discovery and Data Mining, Portland (OR), pp. 226–231.
- [38] EUROPEAN MARITIME SAFETY AGENCY, *European Maritime Safety Agency (EMSA)*, [Online], [Cited February 27th, 2013], Available from <http://www.emsa.europa.eu/>.
- [39] EUROPEAN MARITIME SAFETY AGENCY, *THETIS: Refusal of access*, [Online], [Cited February 27th, 2013], Available from <https://portal.emsa.europa.eu/web/thetis/refusal-of-access>.
- [40] *Extraordinary Summit of SADC Heads of State and Government*, 2009, [Online], [Cited April 17th, 2012], Available from <http://www.sadc.int/index/browse/page/477A>.
- [41] *FAO country profile, fishery legislation*, [Online], [Cited April 18th, 2012], Available from http://www.fao.org/fishery/countrysector/FI-CP_ZA/5/en.
- [42] FAZLOLLAHI B, PARIKH MA & VERMA S, 1997, *Adaptive decision support systems*, Decision Support Systems, **20(4)**, pp. 297–315.
- [43] FISHER R, 2002, *CAVIAR project*, [Online], [Cited June 17th, 2011], Available from <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.
- [44] FRANKE A, CAELLI T & HUDSON RJ, 2004, *Analysis of movements and behavior of caribou (Rangifer tarandus) using hidden Markov models*, Ecological Modelling, **173(2-3)**, pp. 259–270.

- [45] FRITTELLI J, 2008, *Port and maritime security: Background and issues for Congress*, pp. 11–42 in VESKY JP (ED), *Port and Maritime Security*, Nova Science Publishers, New York (NY).
- [46] FULTON W, 1995, *Algebraic topology: A first course*, Springer, Chigaco (IL).
- [47] GENESERETH MR & KETCHPEL SP, 1994, *Software agents*, Communications of the Association of Computing Machinery, **37(7)**, pp. 48–53.
- [48] GILBERT D, 1997, *Intelligent agents: The right information at the right time*, (Unpublished) Technical Report, IBM Corporation.
- [49] GIORGINO T, 2009, *Computing and visualizing dynamic time warping alignments in R: The DTW package*, Journal of Statistical Software, **31(7)**, pp. 1–24.
- [50] GRIMM V & RAILSBACK SF, 2005, *Individual-based modeling and ecology*, Princeton Univ Press, Princeton (NJ).
- [51] GUERRIERO M, WILLETT P, CORALUPPI S & CARTEL C, 2008, *Radar/AIS data fusion and SAR tasking for maritime surveillance*, Proceedings of the 11th International Conference on Information Fusion (FUSION'08), Cologne, pp. 1–5.
- [52] GUO Y, POULTON G, CORKE P, BISHOP-HURLEY G, WARK T & SWAIN D, 2009, *Using accelerometer, high sample rate GPS and magnetometer data to develop a cattle movement and behaviour model*, Ecological Modelling, **220(17)**, pp. 2068–2075.
- [53] HACKER R, 1962, *Certification of Algorithm 112: Position of point relative to polygon*, Communications of the Association of Computing Machinery, **5(12)**, pp. 606.
- [54] HALL DL & LLINAS J, 1997, *An introduction to multisensor data fusion*, Proceedings of the Institute of Electrical and Electronics Engineers, **85(1)**, pp. 6–23.
- [55] HAN J, KAMBER M & PEI J, 2006, *Data mining: Concepts and techniques*, Morgan Kaufmann, San Francisco (CA).
- [56] HAUTAMAKI V, NYKANEN P & FRANTI P, 2008, *Time-series clustering by approximate prototypes*, Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), Tampa (FL), pp. 1–4.
- [57] HAYNES SR, COHEN MA & RITTER FE, 2009, *Designs for explaining intelligent agents*, International Journal of Human-Computer Studies, **67(1)**, pp. 90–110.
- [58] HECKBERT PS & GARLAND M, 1997, *Survey of polygonal surface simplification algorithms*, (Unpublished) Technical Report, Carnegie-Mellon University, Pittsburgh (PA).
- [59] HEINES E, 1994, *Graphics Gems: IV*, Morgan Kaufmann, San Francisco (CA).
- [60] HENDLER J, 2007, *Where are all the intelligent agents?*, IEEE Intelligent Systems, **23(3)**, pp. 2–3.
- [61] HENNIG C, 2014, *Flexible procedures for clustering*, R package version 2.1-7, [Online], [Cited February 10th, 2014], Available from <http://CRAN.R-project.org/package=fpc>.
- [62] HOLMES S, 2012, *Hidden Markov models — The unfair casino*, [Online], [Cited July 20th, 2013], Available from <http://www.stanford.edu/class/stats366/hmmR2.html>.
- [63] HOLSAPPLE CW, PAKATH R, JACOB VS & ZAVERI JS, 1993, *Learning by problem processors: Adaptive decision support systems*, Decision Support Systems, **10(2)**, pp. 85–108.
- [64] HORMANN K & AGATHOS A, 2001, *The point in polygon problem for arbitrary polygons*, Computational Geometry, **20(3)**, pp. 131–144.

- [65] HUNTER A, 2009, *Belief modeling for maritime surveillance*, Proceedings of the 12th International Conference on Information Fusion (FUSION'09), Seattle (WA), pp. 1926–1932.
- [66] HUTCHINGS L, AUGUSTYN CJ, COCKCROFT A, VAN DER LINGEN C, COETZEE J, LESLIE RW, TARR R, OOSTHUIZEN H, LIPINSKI MR & ROBERTS MR, 2009, *Marine fisheries monitoring programmes in South Africa*, South African Journal of Science, **105(5-6)**, pp. 182–192.
- [67] INCE DC & INCE D, 1993, *Introduction to discrete mathematics, formal system specification, and Z*, Oxford University Press, New York (NY).
- [68] INMON WH, 2002, *Building the data warehouse*, J. Wiley, Indianapolis (IA).
- [69] *Interim guidelines for the presentation and display of AIS target information*, [Online], [Cited January 25th, 2013], Available from http://www.imo.org/blast/blastDataHelper.asp?data_id=5391&filename=217.pdf.
- [70] *International Convention for the Safety of Life at Sea (SOLAS), 1974*, 2001, [Online], [Cited February 18th, 2013], Available from [http://www.imo.org/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-\(SOLAS\)-1974.aspx](http://www.imo.org/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS)-1974.aspx).
- [71] *International Maritime Organization*, 2011, [Online], [Cited April 10th, 2012], Available from <http://www.imo.org/About/Pages/Default.aspx>.
- [72] *Introducing SADC*, 2012, [Online], [Cited April 17th, 2012], Available from <http://www.sadc.int/index/browse/page/715>.
- [73] JACKSON S, 1968, *A development of the Jordan curve theorem and the Schoenflies theorem for polygons*, The American Mathematical Monthly, **75(9)**, pp. 989–998.
- [74] JAIN AK, DUIN RPW & MAO J, 2000, *Statistical pattern recognition: A review*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **22(1)**, pp. 4–37.
- [75] JOHANNESSEN JA, 2000, *Coastal observing systems: The role of synthetic aperture radar*, Johns Hopkins APL Technical Digest, **21(1)**, pp. 41–48.
- [76] JOHNSON N & HOGG D, 1996, *Learning the distribution of object trajectories for event recognition*, Image and Vision Computing, **14(8)**, pp. 609–615.
- [77] JUSTINO EJ, BORTOLOZZI F & SABOURIN R, 2002, *The interpersonal and intrapersonal variability influences on off-line signature verification using HMM*, Proceedings of the XV Brazilian Symposium on Computer Graphics and Image Processing, Fortaleza, pp. 197–202.
- [78] KADOUS MW, 2002, *Temporal classification: Extending the classification paradigm to multivariate time series*, PhD Thesis (draft), Computer Science and Engineering, University of New South Wales, Sydney.
- [79] KAUFMAN L & ROUSSEEUW PJ, 2009, *Finding groups in data: An introduction to cluster analysis*, John Wiley & Sons, Hoboken (NJ).
- [80] KLEIN N, 2011, *Maritime security and the law of the sea*, Oxford University Press, New York (NY).
- [81] KNUTH DE, 2006, *The art of computer programming, Volume 4, Fascicle 1: Binary decision diagrams*, Addison-Wesley Professional, Upper Saddle River (NJ).
- [82] KOKAR MM, MATHEUS CJ & BACLAWSKI K, 2009, *Ontology-based situation awareness*, Information Fusion, **10(1)**, pp. 83–98.

- [83] LABORATORIES SN, *What is synthetic aperture radar?*, [Online], [Cited August 08th, 2012], Available from <http://www.sandia.gov/radar/whatis.html>.
- [84] LANE RO, NEVELL DA, HAYWARD SD & BEANEY TW, 2010, *Maritime anomaly detection and threat assessment*, Proceedings of the 13th International Conference on Information Fusion (FUSION'10), Edinburgh, pp. 1–8.
- [85] LAXHAMMAR R, 2008, *Anomaly detection for sea surveillance*, Proceedings of the 11th International Conference on Information Fusion (FUSION'08), Cologne, pp. 55–62.
- [86] LAXHAMMAR R, FALKMAN G & SVIESTINS E, 2009, *Anomaly detection in sea traffic—A comparison of the Gaussian mixture model and the kernel density estimator*, Proceedings of the 12th International Conference on Information Fusion (FUSION'09), Seattle (WA), pp. 756–763.
- [87] LI C & BISWAS G, 1999, *Clustering sequence data using hidden Markov model representation*, Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology, Orlando (FL), pp. 14–21.
- [88] LIAO WT, 2005, *Clustering of time series data — A Survey*, Pattern Recognition, **38**(11), pp. 1857–1874.
- [89] LJUNGBERG M & LUCAS A, 1992, *The OASIS air-traffic management system*, Proceedings of the 2nd Pacific Rim International Conference on Artificial Intelligence (PRICAI), Seoul, pp. 185–190.
- [90] LLINAS J, BOWMAN C, ROGOVA G, STEINBERG A, WALTZ E & WHITE F, 2004, *Revisiting the JDL data fusion model II*, (Unpublished) Technical Report, Defense Technical Information Center, Fort Belvoir (VA).
- [91] MACDONALD IL & ZUCCHINI W, 2009, *Hidden Markov and other models for discrete valued time series*, CRC Press, Boca Raton (FL).
- [92] MAIMON OZ & ROKACH L, 2005, *Data mining and knowledge discovery handbook*, Springer, New York (NY).
- [93] MAKRIS D & ELLIS T, 2005, *Learning semantic scene models from observing activity in visual surveillance*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, **35**(3), pp. 397–408.
- [94] *Man overboard procedures*, [Online], [Cited August 12th, 2013], Available from <http://coastguardsouthernregion.wikispaces.com/file/view/Man+Overboard+Proceedures.pdf>.
- [95] MANOLA F ME & MCBRIDE B, 2004, *RDF Primer*, [Online], [Cited July 20th, 2011], Available from <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [96] *Marine Living Resources Act*, 1998, [Online], [Cited April 18th, 2012], Available from <http://www.info.gov.za/view/DownloadFileAction?id=70675>.
- [97] *Marine Living Resources Act: Declaration of areas as marine protected areas*, 2000, [Online], [Cited April 18th, 2012], Available from <http://www.info.gov.za/view/DownloadFileAction?id=66014>.
- [98] *Marine Notice No. 13*, 2011, [Online], [Cited July 15th, 2013], Available from <http://www.samsa.org.za/sites/samsa.org.za/files/MN%2013%20of%202011.0.pdf>.
- [99] *Marine protection reinforced: South Africa introduces speed chase vessel as part of 14th World Oceans Day Celebrations*, 2006, [Online], [Cited April 13th, 2012], Available from <http://www.info.gov.za/speeches/2006/06060910151002.htm>.

- [100] *MarineTraffic project*, 2012, [Online], [Cited April 16th, 2012], Available from <http://www.marinetraffic.com/ais/>.
- [101] *Maritime New Zealand*, [Online], [Cited July 15th, 2013], Available from <http://www.maritimenz.govt.nz/Rules>.
- [102] *Maritime Rescue Coordination Centre (MRCC)*, 2010, [Online], [Cited April 19th, 2012], Available from <http://www.samsa.org.za/maritime-rescue-coordination-centre>.
- [103] *Maritime Zones Act 15 of 1994*, 1994, [Online], [Cited April 13th, 2012], Available from <http://www.info.gov.za/acts/1994/a15-94.pdf>.
- [104] MARTINEAU E & ROY J, 2011, *Maritime anomaly detection: Domain introduction and review of selected literature*, (Unpublished) Technical Report, Defense Technical Information Center, Fort Belvoir (VA).
- [105] MATHEUS CJ, KOKAR MM & BACLAWSKI K, 2003, *A core ontology for situation awareness*, Proceedings of the 6th International Conference on Information Fusion, Cairns, pp. 545–552.
- [106] MCCAHL M & NORRIS C, 2002, *CCTV in Britain*, (Unpublished) Technical Report, Centre for Criminology and Criminal Justice, University of Hull, Hull.
- [107] *Minister Marthinus van Schalkwyk welcomes home the Sarah Baartman: First-ever joint SADC fisheries patrol a great success*, [Online], [Cited April 27th, 2012], Available from <http://www.info.gov.za/speeches/2005/05071116151001.htm>.
- [108] *Minister of Environmental Affairs and Tourism Marthinus van Schalkwyk launches the Ellen Khuzwayo research vessel in Cape Town*, 2007, [Online], [Cited April 12th, 2012], Available from <http://www.info.gov.za/speeches/2007/07092014451001.htm>.
- [109] *Minister Sisulu signs memorandum of understanding (MOU) on maritime security cooperation with Tanzania and Mozambique*, 2012, [Online], [Cited April 16th, 2012], Available from <http://www.info.gov.za/speech/DynamicAction?pageid=461&sid=24925&tid=55706>.
- [110] MINISTRY OF AGRICULTURE, FORESTRY AND FISHERIES, 2012, *Media statement from the Ministry of Agriculture, Forestry and Fisheries on the mooted relocation Fisheries Branch to Pretoria*, [Online], [Cited April 17th, 2012], Available from <http://www.info.gov.za/speech/DynamicAction?pageid=461&sid=25499&tid=58575>.
- [111] MINISTRY OF TRANSPORT, *Speech by Minister of Transport, Mac Maharaj, at the inauguration of the South African Maritime Safety Authority (SAMSA), Hatfield Gardens, Pretoria, 7 April 1998*, [Online], [Cited April 26th, 2012], Available from <http://www.info.gov.za/speeches/1998/000725545p4001.htm>.
- [112] MITCHELL HB, 2007, *Multi-sensor data fusion: An introduction*, Springer, Berlin.
- [113] MITCHELL T, 2009, *Machine learning: An algorithmic perspective*, Chapman and Hall/CRC, Boca Raton (FL).
- [114] MOREL M, GEORGÉ JP, LITTAYE A, JANGAL F, NAPOLI A, GIRAUD MA & LEBREVELEC J, 2008, *ScanMaris: Automatic detection of abnormal vessel behaviours*, Proceedings of the International Symposium 3AF: Europe and the Transformation of its Forces, Paris, pp. 1–8.
- [115] MORRIS B & TRIVEDI M, 2011, *Trajectory learning for activity understanding: Un-supervised, multilevel, and long-term adaptive approach*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33(11)**, pp. 2287–2301.

- [116] *MSC Fisher Fact Sheet: South Africa hake trawl fishery*, 2009, [Online], [Cited April 18th, 2012], Available from <http://www.msc.org/track-a-fishery/certified/south-atlantic-indian-ocean/south-africa-hake-trawl-fishery>.
- [117] MÜLLER M, 2007, *Information retrieval for music and motion*, Springer, Berlin.
- [118] MURPHY K, 1998, *Hidden Markov Model (HMM) toolbox for Matlab*, [Online], [Cited June 11th, 2013], Available from <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>.
- [119] NAFTEL A & KHALID S, 2006, *Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space*, *Multimedia Systems*, **12(3)**, pp. 227–238.
- [120] NASCIMENTO M, TOLEDO F & CARVALHO A, 2012, *A hybrid heuristic for the k-medoids clustering problem*, *Proceedings of the 14th International Conference on Genetic and Evolutionary Computation*, Philadelphia (PA), pp. 417–424.
- [121] NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY, 2011, *Sailing directions planning guide: South Atlantic Ocean and Indian Ocean*, Springfield (VA).
- [122] NILSSON M, VAN LAERE J, ZIEMKE T & EDLUND J, 2008, *Extracting rules from expert operators to support situation awareness in maritime surveillance*, *Proceedings of the 11th International Conference on Information Fusion (FUSION'08)*, Cologne, pp. 1–8.
- [123] *Notices to Mariners: 2013 Series*, [Online], [Cited August 16th, 2013], Available from http://www.sanho.co.za/notices/2013_series.htm.
- [124] *NSRI: ACSA Sponsorship*, 2011, [Online], [Cited April 19th, 2012], Available from <http://www.nsri.org.za/2011/04/acsa-sponsorship/>.
- [125] OATES T, FIROIU L & COHEN PR, 1999, *Clustering time series with hidden Markov models and dynamic time warping*, *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning*, Stockholm, pp. 17–21.
- [126] O'ROURKE J, 1998, *Computational geometry in C*, Cambridge University Press, Cambridge (MA).
- [127] OWENS J & HUNTER A, 2000, *Application of the self-organizing map to trajectory classification*, *Proceedings of the IEEE International Workshop on Visual Surveillance*, Dublin, pp. 77–83.
- [128] PALLOTTA G, VESPE M & BRYAN K, 2013, *Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction*, *Entropy*, **15(6)**, pp. 2218–2245.
- [129] PANARGO SHIPPING LTD, *Cape Town port information*, [Online], [Cited July 15th, 2013], Available from <http://www.panargo.co.za/wp-content/uploads/2011/06/ct.pdf>.
- [130] PATTERSON TA, THOMAS L, WILCOX C, OVASKAINEN O & MATTHIOPOULOS J, 2008, *State-space models of individual animal movement*, *Trends in Ecology & Evolution*, **23(2)**, pp. 87–94.
- [131] PELOT R & WU Y, 2007, *Classification of recreational boat types based on trajectory patterns*, *Pattern Recognition Letters*, **28(15)**, pp. 1987–1994.
- [132] PLEHN MT, 2000, *Control warfare: Inside the OODA loop*, (Unpublished) Technical Report, Defense Technical Information Center, Fort Belvoir (VA).
- [133] POWER DJ, 2011, *What is decision automation?*, [Online], [Cited January 17th, 2013], Available from <http://dssresources.com/faq/index.php?action=artikel&id=6>.

- [134] *Proposal Information Package (PIP). Predictive Analysis for Naval Deployment Activities (PANDA)*, [Online], [Cited February 5th, 2013], Available from <https://www.fbo.gov/utills/view?id=4b9eec01c0278620e50704ed51e24785>.
- [135] PUSIOL G, BREMOND F & THONNAT M, 2009, *Trajectory based primitive events for learning and recognizing activity*, Proceedings of the IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), Kyoto, pp. 1081–1088.
- [136] RABINER LR, 1989, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the Institute of Electrical and Electronics Engineers, **77(2)**, pp. 257–286.
- [137] RABINER LR, LEVINSON SE & SONDHI MM, 1983, *On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition*, Bell System Technical Journal, **62(4)**, pp. 1075–1105.
- [138] *Radiocommunications*, 2013, [Online], [Cited February 18th, 2013], Available from www.iemo.org/OurWork/Safety/RadioCommunicationsAndSearchAndRescue/Radiocommunications/Pages/Default.aspx.
- [139] *Regulations for the management of the Prince Edward Islands marine protected area*, [Online], [Cited July 30th, 2012], Available from <http://www.info.gov.za/view/DownloadFileAction?id=187610>.
- [140] *Remarks at the signing ceremony of the Southern Africa sub-regional Maritime Rescue Co-ordination Centre (MRCC) by Mr Jeff Radebe (MP) Minister of Transport, Waterfront Cape Town*, 2007, [Online], [Cited April 19th, 2012], Available from <http://www.info.gov.za/speeches/2007/07011810151001.htm>.
- [141] RHODES BJ, BOMBERGER NA & ZANDIPOUR M, 2007, *Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness*, Proceedings of the 10th International Conference on Information Fusion (FUSION'07), Québec, pp. 1–8.
- [142] RISTIC B, LA SCALA B, MORELANDE M & GORDON N, 2008, *Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction*, Proceedings of the 11th International Conference on Information Fusion (FUSION'08), Cologne, pp. 40–46.
- [143] ROUX JN & VAN VUUREN JH, 2007, *Threat evaluation and weapon assignment decision support: A review of the state of the art*, ORiON, **23**, pp. 151–186.
- [144] ROY J, 2008, *Anomaly detection in the maritime domain*, Proceedings of the SPIE Conference on Optics and Photonics in Global Homeland Security IV, Orlando (FL), pp. 1–14.
- [145] ROY J, 2009, *Automated reasoning for maritime anomaly detection*, Proceedings of the NATO Workshop on Data Fusion and Anomaly Detection for Maritime Situational Awareness (MSA 2009), La Spezia, pp. 15–17.
- [146] ROY J, 2010, *Rule-based expert system for maritime anomaly detection*, Proceedings of the SPIE: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense, Orlando (FL), pp. 1–12.
- [147] ROY J & DAVENPORT M, 2010, *Exploitation of maritime domain ontologies for anomaly detection and threat analysis*, Proceedings of the International Conference on Waterside Security (WSS), Carrara, pp. 1–8.
- [148] RUDOWSKY IS, 2004, *Intelligent agents*, Communications of the Association for Information Systems, **14(14)**, pp. 275–290.

- [149] RUSSELL SJ & NORVIG P, 2003, *Artificial intelligence: A modern approach*, 2nd Edition, Prentice Hall, Upper Saddle River (NJ).
- [150] *Ruth First*, [Online], [Cited May 03rd, 2012], Available from http://www.sa-transport.co.za/ships/tugs_pilot_harbour/ruth_first.html.
- [151] SALVADOR S & CHAN P, 2007, *Toward accurate dynamic time warping in linear time and space*, *Intelligent Data Analysis*, **11(5)**, pp. 561–580.
- [152] SANBI BIODIVERSITY PLANNING UNIT, *National spatial biodiversity assessment*, [Online], [Cited October 22nd, 2012], Available from <http://bgis.sanbi.org/nsba/project.asp>.
- [153] *Sarah Baartman off-shore environmental protection vessel returns to South Africa after successful transboundary fisheries patrol*, 2009, [Online], [Cited April 13th, 2012], Available from <http://www.info.gov.za/speeches/2009/09032616151003.htm>.
- [154] SCHMIDT FR, TÖPPE E, CREMERS D & BOYKOV Y, 2007, *Efficient shape matching via graph cuts*, *Proceedings of the Energy Minimization Methods in Computer Vision and Pattern Recognition*, Ezhou, pp. 39–54.
- [155] SEIBERT M, RHODES BJ, BOMBERGER NA, BEANE PO, SROKA JJ, KOGEL W, KREAMER W, STAUFFER C, KIRSCHNER L & CHALOM E, 2006, *SeaCoast port surveillance*, *Proceedings of the SPIE Conference on Photonics for Port and Harbor Security II*, Kissimmee (FL), pp. 1–8.
- [156] SHENG W & LIU X, 2006, *A genetic k-medoids clustering algorithm*, *Journal of Heuristics*, **12(6)**, pp. 447–466.
- [157] SHIM J, WARKENTIN M, COURTNEY JF, POWER DJ, SHARDA R & CARLSSON C, 2002, *Past, present, and future of decision support technology*, *Decision Support Systems*, **33(2)**, pp. 111–126.
- [158] SHIMRAT M, 1962, *Algorithm 112: Position of point relative to polygon*, *Communications of the Association of Computing Machinery*, **5(8)**, pp. 434.
- [159] *Ships' routing*, [Online], [Cited January 28th, 2013], Available from <http://www.imo.org/ourwork/safety/navigation/pages/shipsrouting.aspx>.
- [160] SICILIANO B & KHATIB O, 2008, *Springer handbook of robotics*, Springer, Berlin.
- [161] SIKO M, 1996, *South Africa's maritime interest and responsibilities*, [Online], [Cited April 13th, 2012], Available from <http://www.issafrica.org/pubs/ASR/5No2/5No2/SAMaritime.html>.
- [162] SOUTH AFRICAN MARITIME SAFETY AUTHORITY, *SAMSA*, [Online], [Cited April 26th, 2012], Available from <http://www.samsa.org.za/about>.
- [163] SOWMAN M, HAUCK M, VAN SITTERT L & SUNDE J, 2011, *Marine protected area management in South Africa: New policies, old paradigms*, *Environmental Management*, **47(4)**, pp. 573–583.
- [164] SPRAGUE JR RH, 1980, *A framework for the development of decision support systems*, *MIS Quarterly*, **4(4)**, pp. 1–26.
- [165] *Statement by President Jacob Zuma on the appointment of the new Cabinet*, 2009, [Online], [Cited April 12th, 2012], Available from http://www.info.gov.za/events/2009/new_cabinet.htm.

- [166] *Statement by the office of Marthinus van Schalkwyk, Minister of Environmental Affairs & Tourism: Honouring SA's heroines — SA receives first of four new environmental protection vessels*, 2004, [Online], [Cited April 13th, 2012], Available from <http://www.info.gov.za/speeches/2004/04111615451001.htm>.
- [167] STEINBERG AN, 2001, *Data fusion system engineering*, Aerospace and Electronic Systems Magazine, IEEE, **16(6)**, June, pp. 7–14.
- [168] STEINBERG AN & BOWMAN CL, 2004, *Rethinking the JDL data fusion levels*, Proceedings of the National Symposium on Sensor Data Fusion (JHUAPL), Laurel (MD), pp. 1–18.
- [169] STEINBERG AN, BOWMAN CL & WHITE FE, 1998, *Revisions to the JDL data fusion model*, American Institute of Aeronautics and Astronautics, New York (NY).
- [170] SUTTER H & LARUS J, 2005, *Software and the concurrency revolution*, Queue, **3(7)**, pp. 54–62.
- [171] *Ten alleged poachers arrested, super duck confiscated after high speed chase*, [Online], [Cited April 14th, 2012], Available from <http://www.info.gov.za/speeches/2008/08012516151001.htm>.
- [172] *The International System of Units (SI)*, 2006, [Online], [Cited April 5th, 2012], Available from http://www.bipm.org/utis/common/pdf/si_brochure_8_en.pdf.
- [173] *The maritime environment*, [Online], [Cited April 19th, 2012], Available from http://www.navy.mil.za/SANGP100/SANGP100_CH02.pdf.
- [174] *The Paris Memorandum of Understanding on Port State Control*, [Online], [Cited February 27th, 2013], Available from http://www.parismou.org/Organization/2010.12.27/Memorandum_of_Understanding.htm.
- [175] *The United States Coast Guard's program for identifying high interest vessels*, [Online], [Cited January 28th, 2013], Available from http://www.oig.dhs.gov/assets/Mgmt/OIG_09-107_Sep09.pdf.
- [176] *The World Factbook*, [Online], [Cited January 22nd, 2013], Available from <https://www.cia.gov/library/publications/the-world-factbook/geos/sf.html>.
- [177] THEODORIDIS S & KOUTROUMBAS K, 2006, *Pattern recognition*, 3rd Edition, Academic Press, San Diego (CA).
- [178] TILNEY R & PURVES MG, 1999, *The status of integrated fisheries monitoring in South Africa*, Proceedings of the International Conference on Integrated Fisheries Monitoring, Sydney, p. 343.
- [179] TURBAN E, KING D & WANG J, 2003, *Introduction to e-commerce*, Prentice Hall, Upper Saddle River (NJ).
- [180] UNITED NATIONS, 2001, *United Nations Convention on the Law of the Sea of 10 December 1982*, [Online], [Cited April 3rd, 2012], Available from http://www.un.org/Depts/los/convention_agreements/texts/unclos/closindx.htm.
- [181] UNITED NATIONS, 2011, *United Nations Convention on the Law of the Sea (A historical perspective)*, [Online], [Cited April 3rd, 2012], Available from http://www.un.org/Depts/los/convention_agreements/convention_historical_perspective.htm.
- [182] UNITED NATIONS, 2012, *United Nations Treaty Collection*, [Online], [Cited April 9th, 2012], Available from <http://treaties.un.org/pages/Overview.aspx?path=overview/glossary/page1.en.xml>.

- [183] UNITED NATIONS, 2012, *United Nations Office for Disarmament Affairs*, [Online], [Cited April 10th, 2012], Available from <http://disarmament.un.org/treaties/t/pelindaba>.
- [184] US COAST GUARD NAVIGATION CENTER, *Automatic Identification System Overview*, [Online], [Cited July 27th, 2014], Available from <http://www.navcen.uscg.gov/?pageName=AISmain>.
- [185] USCHOLD M & GRUNINGER M, 1996, *Ontologies: Principles, methods and applications*, Knowledge Engineering Review, **11(2)**, pp. 93–136.
- [186] VAN HAGE WR, WIELEMAKER J & SCHREIBER G, 2010, *The space package: Tight integration between space and semantics*, Transactions in Geographical Information Systems, **14(2)**, pp. 131–146.
- [187] VAN HAGE W, DE VRIES GKD, MALAISÉ V, SCHREIBER G & VAN SOMEREN M, 2009, *Spatial and semantic reasoning to recognize ship behaviour*, Proceedings of the International Semantic Web Conference, Chantilly (VA).
- [188] VAN LAERE J & NILSSON M, 2009, *Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance*, Proceedings of the 12th International Conference on Information Fusion (FUSION'09), Seattle (WA), pp. 171–178.
- [189] VAN WYK J, 2011, *Radar lab at the Institute for Maritime Technology*, [Personal Communication], Contactable at ljvanwyk@gmail.com.
- [190] VERBEEK J, VLASSIS N & KRÖSE B, 2003, *Efficient greedy learning of Gaussian mixture models*, Neural Computation, **15(2)**, pp. 469–485.
- [191] *Vessel Monitoring System (VMS) installation deadline reminder for traditional line fish and hake handline sector*, 2007, [Online], [Cited April 19th, 2012], Available from <http://www.info.gov.za/speeches/2007/07033012451001.htm>.
- [192] *Vessel Traffic Services*, [Online], [Cited January 25th, 2013], Available from <http://www.imo.org/ourwork/safety/navigation/pages/vesseltrafficservices.aspx>.
- [193] VLACHOS M, KOLLIOS G & GUNOPULOS D, 2002, *Discovering similar multidimensional trajectories*, Proceedings of the 18th International Conference on Data Engineering, San Jose (CA), pp. 673–684.
- [194] VON LUXBURG U, 2007, *A tutorial on spectral clustering*, Statistics and Computing, **17(4)**, pp. 395–416.
- [195] WACHBROIT R, 1994, *Normality as a biological concept*, Philosophy of Science, **61(4)**, pp. 579–591.
- [196] WALTZ EL, 1998, *Information understanding: Integrating data fusion and data mining processes*, Proceedings of the IEEE International Symposium on Circuits and Systems, Los Alamitos (CA), pp. 553–556.
- [197] WANG H, 1997, *Intelligent agent-assisted decision support systems: Integration of knowledge discovery, knowledge analysis, and group decision support*, Expert Systems with Applications, **12(3)**, pp. 323–335.
- [198] WEBB AR, 2011, *Statistical pattern recognition*, 2nd Edition, Wiley, Chichester.
- [199] WEISSTEIN EW, *Coastline paradox*, [Online], [Cited April 19th, 2012], Available from <http://mathworld.wolfram.com/CoastlineParadox.html>.
- [200] WIERSMA E & MASTENBROEK N, 1998, *Measurement of vessel traffic service operator performance*, AI & Society, **12(1)**, pp. 78–86.

-
- [201] WOO H, JUNG YM, KIM J.-G & SEO JK, 2010, *Environmentally robust motion detection for video surveillance*, IEEE Transactions on Image Processing, **19(11)**, pp. 2838–2848.
- [202] WOOLDRIDGE MJ, 2002, *An introduction to multiagent systems*, Wiley, Chichester.
- [203] *World Shipping Register*, [Online], [Cited March 14th, 2013], Available from <http://www.world-ships.com/>.