Todo list

Contents

It's the author's job to make the reader's job easy - Joshua Schimel.

1 Introduction

- -

Begin the thesis by giving a general overview of the field you are conducting research in. Focus on work that has lead up to your project. Remember to cite. This could look something like: What a route planner is, maybe its definition. Then, move further on to what types of methods has been used in the field. In this case data-based and non-data based models. The middle part goes into a more specifics. Write about what you are investigating and why. Get help with referencing if needed. At the end of the introduction, simply state what the thesis aims to achieve.

- -

Before computers and computation, data was logged manually. While traveling across the worlds seas, logging wind and current information enabled ships to decrease their travel time (Lewis 1927).

Transportation on sea has become, and is still to this day one of the biggest businesses within the global trade. Maritime shipping amounts to 90 % of the world's trade transportation method (UN-Business Action Hub, United Nations 2020). Route planning is a comprehensive guide developed and used to determine the most favorable route, raise awareness of potential problems and ensure the vessel's safe passage. Constructing a route depends in most cases on knowledge of past voyages or surroundings of a route e.g. map data.

AIS-data has been used for prediction and estimation in the maritime industry. Predicting traffic flow of ships in a given region and time period (Wang, Li, and Zhang 2019). Estimating the time it takes for a vessel to arrive at a given location could be done by the information AIS-data yields. The approaches and methods used in research base their predictions on previous voyages (see chapter 2: literature review). This is one of the research areas in the maritime industry (Meijer 2017). AIS-data used in various applications such as streamline cargo transfer at harbors, avoiding collisions at sea with real-time data of speed and position of vessels within a region and helping companies to take better decisions.

Before the invention of AIS-data, conducting voyages was done based on maps.

Algorithms search through a space given start and end points. The goal is to maneuver through this space and create a route without colliding with obstacles i.e land (Hvamb 2015). The algorithms vary from applications in sciences such as informatics, mathematics and physics (Hvamb 2015), (Besse et al. 2015), (Pedersen and Fossen 2012). The latter has motivated the basis for the research done in this thesis.
In this thesis, a framework for creating and predicting routes for marine vessels will be attempted. A data-based model is provided, more specifically, a AIS-data driven one. In addition, an approach based on principles and theory in the field of fluid mechanics, more specifically potential theory. Setting up an environment given a map in which the route shall be produced by calculating streamlines will provide routes from a given starting and end point. The area of interest will be Geirangerfjorden located in Sunnmøre, Norway.

Remember to cite! and get right location of area

1.1 A brief history of navigation at sea

Before, navigating by sea was done by observing celestial bodies and determining one's position. This was before the invention of the compass. One of the first tools used in navigation was a magnetic compass. The compass feels a force by the earth's magnetic field and points to what is known as the "magnetic north". This created some confusion and inconsistent measures at sea due to unawareness of magnetic variance (difference between magnetic north and geographic north).
Jumping centuries forward in time, the twentieth century established navigational tools such as radio beacons, radar, the gyroscopic compass, and the global positioning system (GPS). Some of the tools' characteristics are:
Gyroscopic compass always points at true north
Radar(short for "radio detection and ranging") system was produced in 1935. It was used to locate objects beyond range of vision by projecting radio waves against them. This was, and still is, very useful on ships to locate other ships and land when visibility is reduced.
The U.S. navigation system known as LORAN (LOng RAnge Navigation) uses low frequency radio transmissions from so-called "master" and "slave" stations to determine a ship's position. The accuracy of LORAN is measured in hundreds of meters.
In the late twentieth century, the global positioning system (GPS) largely replaced the Loran. GPS uses the same principle of time difference from separate signals as Loran, but the signals come from satellites. As of 2002, the system consisted of

24 satellites, and gave the mariner a position with accuracy of 9 meters (30 feet) or less.

ref! over er direkte sitater! http://www.waterencyclopedia.com/Mi-Oc/Navigation-at-Sea-History-of.html

1.2 Problem statement

Objectives:
- Cleaning data
- Create a map-based algorithm
- AIS-data based algorithm

The objective of this thesis is easy to comprehend. To put it in layman's terms, how to create a route from a to b for marine vessels to traverse. A straight line between two points will not be satisfactory. Some information is required to use as input when making a model. This could of course vary and combined in order to make a model more robust or improve accuracy. Historical AIS-data (Automatic Identification System) has been and is still used to make predictions based on vessels location. This approach is realistic one as the data is based on real locations, so the safety and accuracy aspect of the route leads to well defined and trustworthy predictions. On the other hand, how is it possible to make a prediction without any data. This will be the second part of the thesis. Fluid mechanics, more specifically potential flow theory will be the main theory to which we will approach a none-data based model on. Map information will also be a requirement for the model. More detailed objectives can be found below:

1.2.1 AIS-data driven

Given a dataset containing vessel information i.e., how is it possible to construct a route. Analyzing the data will be the first step. Looking into how to structure the data in order to find patterns which the model can be based on.

1.2.2 Fluid mechanical approach

Make a vector field given a map and construct a route with flow theory from fluid mechanics. Each potential represents different objects in the map. The necessary representations of a map will be objects which needs to be avoided together with start and end points of route.

1.3 Constraints

In the thesis when it analyzing the data set, we will assume land is the only obstacle for each vessel. In reality there could be several vessels traversing the same area at the same time which could lead to a change in course. Could be useful to mention shipping rules on the sea!!

2 Literature review

- *Try to present the route prediction method as a problem of many solutions. Each method with its own limitation. By combining AIS and none-AIS approaches the predictions made will hopefully be stronger and more robust. Weight the contribution to route planning, not patterns nor collision avoidance, motion patterns, anomaly detection. important factors, but not of interest in review section*
- -

In this section an overview of the research done in the path-planning field is presented. An attempt to present the methods systematically beginning with methods with applied mathematics, physics and finishing with AIS-based models. Lastly, an overview presented in a pros and cons table to summarize the methods will be done.

The research area constituting route predicting has been applied in multiple disciplines such as, mathematics, physics and informatics.

This literature review will focus on algorithm/methods produced for predicting trajectories for marine vessels. The scope of this review will be on research emphasizing what each model uses as input, how the trajectory is generated and what the output is. An attempt to categorize each model into a list of pros and cons will be done.

How many years have you focused on?!?!?! within a time frame of 5 years as the models in present time are using realistic data. This leads to accurate predictions.

Depending on method some sort of input to model is required. The input often leads to an output of certain character. Here, mainly inputs will have the type of start and end points or AIS-data. The models dependent on start and end-point

(Hvamb 2015) presents three approaches for route planners. These methods requires a map. The first one using a Voronoi diagram. Making connecting lines between vertices of polygons or points describing obstacles e.g islands. Then, perpendicular lines at their half distance are drawn and connected creating Voronoi lines. Each line not intersecting the obstacles are suggested routes.

The next method is based on Rapidly-exploring Random Trees (RRT). Searching an area by sampling points a graph is produced. A bias toward to the goal makes the model possible for convergence.

The final method presented in (Hvamb 2015) is applies a Probabilistic roadmap and uses this for planning routes for marine vessels.

(Pedersen and Fossen 2012) makes the first route planner by applying potential flow theory on marine vessels. Setting up potentials with different flow characteristics, a flow pattern emerges. From this, following certain streamlines lead to the desired goal. In addition, a vessel guidance scheme is derived making sure the vessel gets to its destination.

(Besse et al. 2015) analyzes taxi-drivers' patterns in San Francisco with GPS-data. Predicting the destination of the taxi's by their choice of route with a method derived from (Besse et al. 2015) a Symmetrized Segment-Path Distance (SSPD).

(Pallotta, Vespe, and Bryan 2013) implements a unsupervised learning scheme (Traffic Route Extraction and Anomaly Detection) to predict routes in different locations. By clustering AIS-data into waypoints for certain areas in a route. Vessels are then classified with state when entering a waypoint.

Method	Author	Pros	Cons
Voronoi diagram	Hvamb 2015	• Based on map i.e need polygons to describe obstacles • No need for AIS data (may be useful if wanting to straighten out route)	• Initial route may not be realistic • Must apply additional algorithm for getting e.g shortest route • Computational heavy to generate
RRT	Hvamb 2015	• Able to provide several routes from same or different starting points • Several trees can be initiated at once • No need for AIS data • possible to optimize w.r.t other metrics • A bias parameter can decrease number of calculations	• Struggles in finding paths in narrow passages • Exploration method is symmetric which causes unecessary computations
PRM	Hvamb 2015	• Does not requires AIS data • Based on map • Can compute multiple routes • Can implement a straightening out technique for shorter routes	• Finding the shortest route, which is often the case, requires an additional algorithm • Computing multiple routes comes at a higher computational cost

Potential flow	Pedersen and Fossen 2012	• No need for AIS-data • Not computational heavy • Produces multiple routes by visualizing streamlines	• Must adjust strength-parameters for optimal map representation • Streamlines could be very close to objects and must therefore be selected carefully • If flow is non-uniform around a vessel leads to deviation from the streamline
SSPD	Besse et al. 2015	• Does not requires labeled data • Implementation able with cython for optimal computation speed	• Not easy to verify how much data is required for a good model
TREAD	Pallotta, Vespe, and Bryan 2013	• No prior information needed • Amount of data is huge • Model able to predict anomalies	• Not a good model for traffic with a low density of vessels • Certain amount of preprocessing is necessary for good and precise motion patterns to be made

3 Theory

- -

Worth mentioning what the criterias are for setting up a potential. From Navier stokes equation and conservation of mass to streamfunction. No curl $->$ Potentialflow

- -

In this section the mathematics used in the thesis is presented. The theory will mostly be from the field of fluid mechanics. A brief introduction to the equation which governs the properties and motion of fluids will be given, namely Navier-Stokes equations. From the general Navier-Stokes equations, some characteristics of fluid will be assumed (the fluids viscosity) and yield new equation which will be the main equations used in this thesis. These equations are part of fluid mechanics named potential theory.
Visualization is a key component the research on which we will conduct. Therefore, tools for using vector fields and plotting are a necessity. A more detailed description can be found in (Gjevik 2009) and (Acheson 2003). The data used will also be described in this section.

3.1 Vector

A vector is used to describe a quantity with direction and magnitude. From a point of reference this can be represented in a coordinate system. Mathematically this can be written as:

$$\vec{r} = [x, y, z]$$

The arrow indicates that r is of the type vector. The first component moves along the x-axis, second along the y-axis and the third along the z-axis. The three components indicate that the vector $\vec{r}$ exists in a three-dimensional space. This can visualized as seen in the figure below:

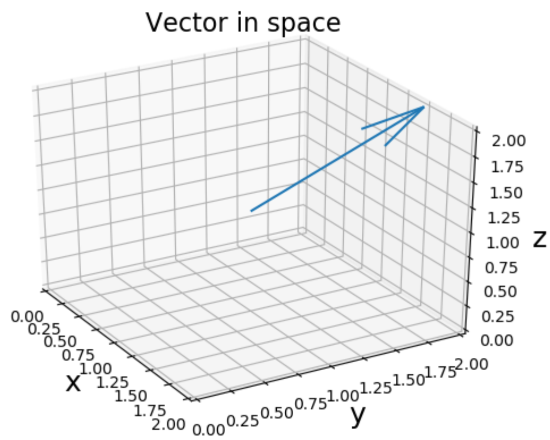

Figure 1: shows a typical vector representation. The arrow shows a vectors direction, it's length represents it's magnitude.

3.2 Vector field

Vector fields are useful when wanting to describe velocities, forces or accelerations within a domain. The field are represented with vectors which are functions of spacial coordinates and time. Mathematically this is written as:

$$\vec{F} = \vec{F}(x, y, z, t)$$

This can further be written as:

$$\vec{F}(x, y, z, t) = (F_x(x, y, z, t), F_y(x, y, z, t), F_z(x, y, z, t))$$

F_x, F_y, F_z are scalar fields. The field is called stationary if it is independent of time. See figure below for a stationary vector field.

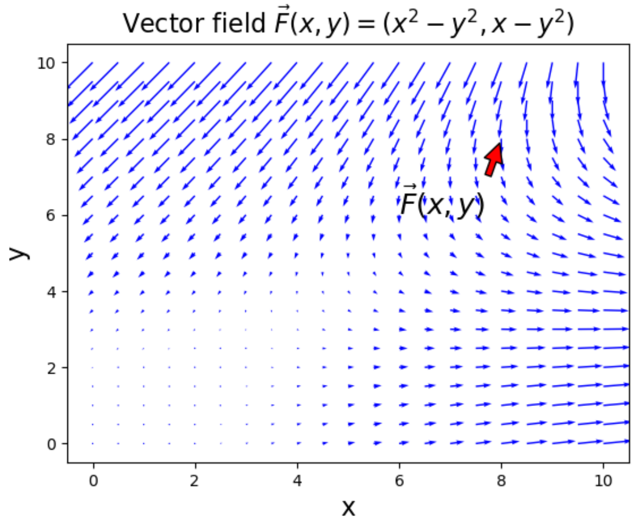

Figure 2: Here a vector field is visualized F(x, y) taking in two parameters x and y and calculates the field from [0-10] in both x and y-direction. The arrow size is the magnitude at that specific point, and also indicating the fields direction..

3.3 Navier-Stokes equation

The Navier-Stokes equations are derived using the laws of several quantities which are conserved. The quantities include mass, momentum and energy.
The continuity equation arise from the law of conservation of mass which says that mass can not be created or disappear. The general equation of continuity is stated below:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{u}) = 0$$

where ρ is the fluid density, $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}]$ the gradient and $\vec{u}$ the velocity. We will assume that the properties of fluids looked at are incompressible. This means the fluid's density is constant. The continuity equation for incompressible fluid states:

$$\nabla \cdot \vec{u} = 0$$

Where both density terms are excluded first by statement above and dividing by the density ρ.
This leads up to the next equation, the momentum equation. It states that a fluid's density times acceleration is proportional to the forces acting upon the fluid. Since

we are only worried about incompressible fluids the momentum equation for such a fluid is:

$$\frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \nabla)\vec{u} = -\frac{\nabla p}{\rho} + \mu \nabla^2 \vec{u} + f_v$$

This is the Navier-Stokes equation for an incompressible fluid. The left side hand states the inertia term. The terms $\frac{\partial \vec{u}}{\partial t}$ and $(\vec{u} \cdot \nabla)\vec{u}$ are the local and convective acceleration respectively. The right hand side is pressure gradient $\frac{\nabla p}{\rho}$, viscous forces $\mu \nabla^2 \vec{u}$ and external forces f_v. As for the continuity equation we will make some assumption which in turn will produce equation which will be used here. The fluid which will be focused on are stationary and inviscous $\mu = 0$. These assumptions leads further to potential flow theory which will be covered next. The next section will be the main focus of the physics model of the thesis.

3.4 Potential flow

Streamlines, also called field lines is useful to visualize the flow in a fluid. This is represented as a vector field. By assuming that the flow of the fluid is steady, that is, the velocity does not change with time at a fixed point in space, it is possible to find these streamlines. Mathematically the velocity can be written as:

$$\vec{u} = \frac{d\vec{x}}{ds} \tag{1}$$

Where $\vec{u} = [u(x,y,z), v(x,y,z), w(x,y,z)]$, $d\vec{x} = [dx, dy, dz]$ and ds is a small change in the curve s. Streamline curves are defined as having same direction as the velocity vector at each point. Then the cross product between the velocity vector and curve s is zero:

$$\vec{u} \times d\vec{x} = 0 \tag{2}$$

Writing out the components gives:

$$(vdz - wdy)\vec{i} = 0$$

$$(wdx - udz)\vec{j} = 0$$

$$(udy - vdx)\vec{k} = 0$$

This gives the equations:

$$\frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w} = ds \tag{3}$$

For a 2-dimensional description the equation for an incompressible flow is :

$$\nabla \cdot \vec{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{4}$$

$\vec{u}$ is the velocity vector is a function of u and v, $\vec{u} = [u, v]$. u and v are the velocity components in x and-y direction, respectively. $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ is the differential operator nabla. We can relate the velocity vector to the stream functions as:

$$u = \frac{\partial \psi}{\partial y} \tag{5}$$

$$v = -\frac{\partial \psi}{\partial x} \tag{6}$$

ψ is the stream function. Putting (5) and 6) into (4) yields:

$$\nabla \cdot \vec{u} = \frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial^2 \psi}{\partial x \partial y} = 0 \tag{7}$$

Also assuming the field is irrotational $\nabla \times \vec{u}$, the laplace equation is obtained:

$$\nabla \times \vec{u} = \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial x^2} = \nabla^2 \psi = 0 \tag{8}$$

Where eq (5) and eq (6) are substituted in as expressions for the velocity components. The laplacian has some important properties which will be used here. The laplacian operator is a linear operator which says that if two different stream functions are solutions to eq (8) then the sum is also a solution. This is known as the superposition principle .

ref her

Equation (1) is satisfied. Furthermore, eq (2) can give us the conditions on what the stream function must look like in order for this to be true. Putting eq (5) and eq (6) into eq (2) and assuming this is for a 2 dimensional flow the stream function yields:

is this correct?

$$udy - vdx = \frac{\partial \psi}{\partial y} dy + \frac{\partial \psi}{\partial x} dx - = 0 \tag{9}$$

$$\partial\psi = 0 \tag{10}$$

$$\psi = Constant \tag{11}$$

ψ is constant along the streamline. All stream functions satisfying eq (7) are valid choices for ψ. Solutions of ψ are: sinks, sources, doublets. The velocity components represented in polar coordinates are:

$$v_r = -\frac{1}{r}\frac{\partial\psi}{\partial\theta} \tag{12}$$

$$v_\theta = \frac{\partial\psi}{\partial r} \tag{13}$$

and the laplacian in polar coordinates:

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2} = 0 \tag{14}$$

$r = \sqrt{x^2 + y^2}$ is the radius and θ is the angle from origo.

3.4.1 Uniform

A uniform flow has constant velocity components u and v in x and y direction respectively. By the relation from eq(5) and eq(6) this satisfies eq(7) and is therefore a solution

3.4.2 Sink and source

Consider a source flowing radially outwards with a steady rate. By finding how much is flowing outwards we can find the velocity components of this type of flow. By integrating the flow over a circle with radius r this yields:

$$Q = \int_0^{2\pi} v_r r d\theta = 2\pi r v_r \tag{15}$$

Q is the flow strength. A flow which emits fluid isotropically i.e is uniform in all directions. This type of flow has the characteristics of a source and has the stream function:

$$\psi_{source} = \frac{Q}{2\pi}\theta \tag{16}$$

Needless to say, this is also a solution of the laplace equation (8). The corresponding velocity components take the form:

$$u_{source} = \frac{Qx}{2\pi(x^2 + y^2)} \tag{17}$$

$$v_{source} = \frac{Qy}{2\pi(x^2 + y^2)} \tag{18}$$

Where eq(12) is equated with eq (15) for the radial velocity. The stream function satisfies the laplace equation eq(8). A flow which absorbs fluid isotropically is called a sink and has the same stream function as a source but with a positive sign:

$$\psi_{sink} = -\frac{Q}{2\pi}\theta \tag{19}$$

The velocity components share the same form as the source but with opposite sign which is expected.

$$u_{sink} = -\frac{Qx}{2\pi(x^2 + y^2)} \tag{20}$$

$$v_{sink} = -\frac{Qy}{2\pi(x^2 + y^2)} \tag{21}$$

3.4.3 Doublets

Here we derive the stream function equation for a doublet. This is where a sink and source coincide at the same point. By placing a sink in a point, a is small distance, (a, 0) with angle θ_2 and a source in (-a,0) with angle θ_1. This gives the stream function using the superposition principle:

$$\psi = -\frac{Q}{2\pi}(\theta_1 - \theta_2) \tag{22}$$

Now, take the tangent on each side:

$$\tan\Big(-\frac{2\pi\psi}{Q}\Big) = \tan(\theta_1 - \theta_2) = \frac{\tan(\theta_1) - \tan(\theta_2)}{1 + \tan(\theta_1)\tan(\theta_2)} \tag{23}$$

Further we rewrite $\tan(\theta_1)$ as $\frac{y}{x-a}$ together with $\tan(\theta_1)$ as $\frac{y}{x+a}$ and insert this into eq (19).

$$\tan\Big(-\frac{2\pi\psi}{Q}\Big) = \frac{\frac{y}{x-a} - \frac{y}{x+a}}{1 + \frac{y^2}{(x^2-a^2)}} \tag{24}$$

After cleaning up the equation reads:

$$\tan\left(-\frac{2\pi\psi}{Q}\right) = \frac{2ya}{x^2+y^2-a^2} \tag{25}$$

Taking the inverse tangent on both sides of and remembering that a is small distance gives small values in the argument of arctan. For small values of a gives:

$$-\frac{2\pi\psi}{Q} = \arctan\left(\frac{2ya}{x^2+y^2-a^2}\right) \approx \frac{2ya}{x^2+y^2-a^2} \tag{26}$$

Solving for the stream function yields:

$$\psi = -\frac{Qay}{\pi(x^2+y^2-a^2)} \tag{27}$$

Now, as a tends toward 0, Q will tend to infinity. In this case, let a $->$ 0 and Q $->\infty$ so that there product is constant. This can be written as:

$$\lim_{\substack{a\to 0\\ Q\to\infty}} -\frac{Qay}{\pi(x^2+y^2-a^2)} = -\frac{ky}{\pi(x^2+y^2)} \tag{28}$$

k is the strength of the doublet.
The stream function equation for a doublet reads finally:

$$\psi_{doublet} = -\frac{ky}{\pi(x^2+y^2)} \tag{29}$$

The velocity components for a doublet is found from eq (5) and eq (6):

$$u = \frac{k(y^2-x^2)}{(x^2+y^2)^2} \tag{30}$$

$$v = -\frac{kyx}{(x^2+y^2)^2} \tag{31}$$

Figures (2-4) visualize different streamline behaviors from different solution of laplace equation ()8).

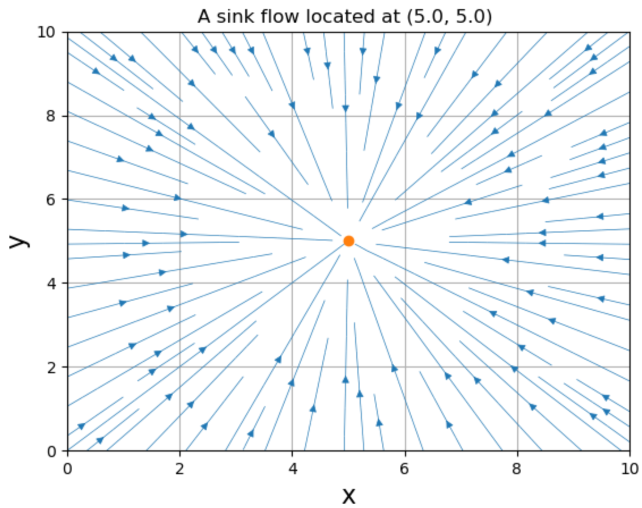

Figure 3: A solution of the laplace equation produces streamlines directed radially inward to it's origin (orange point). The velocity decreases inversely proportional from it's distance from the origin. This flow characteristic is called a sink flow. This flow is useful when representing end point of routes.

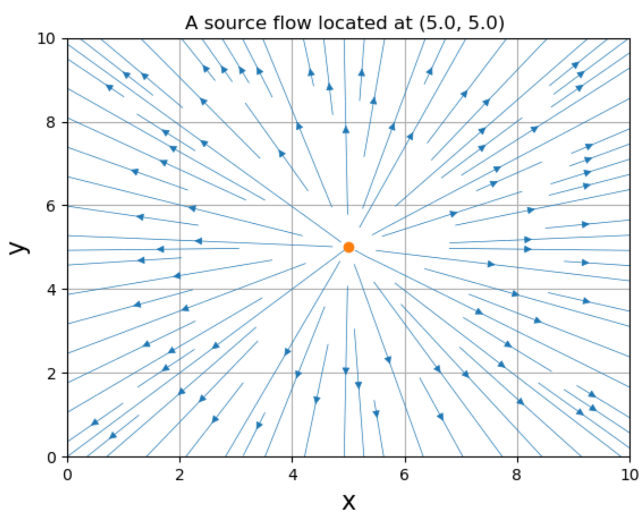

Figure 4: This flow characteristic is called a source flow. The velocity is points outward from it's origin (orange point). This flow is useful in starting points

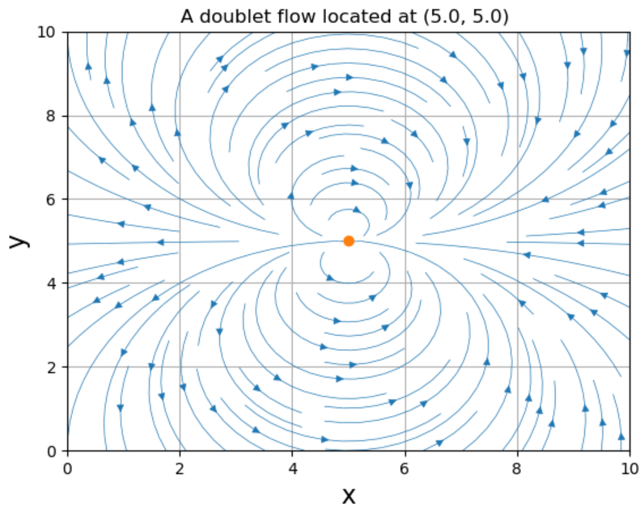

Figure 5: A doublet emerges from a sink and source coinciding in the point. A doublet represents objects i.e islands and land.

4 AIS-data

A concise introduction of the data used in the thesis is presented. Also, a motivation as to why AIS data was introduced in the first place is mentioned.

Autonomous Identification System data, for short AIS-data, is a self reporting system installed on maritime vessels. It sends out information about a vessels characteristics in real time. Some characteristics are: a vessels position, speed over ground (SOG), course over ground (COG) and vessel identification. The receiver on the other end, typically other ships within a vicinity or vessel traffic services (VTS), use the information for safety measures such as avoiding collisions.

The development of AIS-data come after the environmental catastrophe of the Exxon Valdez oil spillage in 1989 (ref: marineinsight, 2019). In 2004 the International Maritime Organization (IMO) made it mandatory for all vessels with a weight over 300 gross tons to install a AIS transceiver (ref: IMO, 2020). The huge amounts of data being logged has made analyzing AIS-data a useful tool within the maritime industry. Being able to predict arrival and waiting times at ports, turnaround times and make accurate decisions in advance regarding risk assess-

ment is a benefit for the parties involved.
Managing data at hand are important aspects as to how it is going to be used. It is also crucial to keep in mind that in most cases when operating with data-dependent is being aware of its limitations ref!!!.
In this thesis, an extraction within a specified area will be made of the AIS-data. The data gathered is from DNV (Det Norske Veritas). The AIS-data will help gain insight and perspective on the route done by vessels as this is actual route done by vessels. As AIS-data is installed by many different vessel types, an interesting view is to view what vessel types take which route depending on the environment. This could benefit the none-data model more accurate by a more specific choice of vessel types. A drawback is the loss of data in this approach.
References in chapter:
https://www.marineinsight.com/maritime-history/the-complete-story-of-the-exxon-valdez-oil-spill/
http://www.imo.org/en/OurWork/safety/navigation/pages/ais.aspx
https://www.portvision.com/news-events/press-releases-news/how-to-use-ais-data-to-your-advantage

4.1 Preprocessing AIS-data

AIS-data provides several variables at a given time. These variables are:

- International Maritime Organization (IMO) number: is a identification number for vessels and is unique for every vessel.

- Longitude and latitude: make up vessel positions in a geographic coordinate system. Their units are angles and make up unique points on a sphere. In a cartesian coordinate system the longitude and latitude would represent x and y coordinates respectively.

- Heading: is based on directions on a compass. Used for finding vessels' course measured in angles where 0°and 360 °is true north, 90 °true east, 180 °true south and 270 °true west.

- Speed Over Ground (SOG): is a vessel speed in magnitude relative to the earth's surface. SOG is measure in knots or(1.852 km/h).

- Ship type: specifies the ships characteristic e.g. cargo ship, passenger ship, tug boat, tank ships, coast guard ship, pleasure craft, high-speed craft, towing vessel, fishing boats, sailing vessels, ekranoplans. If the vessel is not of this ”other” is registered.

- Length: the vessels maximum length.

- Time stamp position: is the time when data was registered. Measured with an accuracy up to seconds.

4.1.1 Data structure

The key elements in a tidy data set is being consistent, easy to access and work with and exploit. Another element to take into consideration is how the computer will be able to extract values. This will help when it comes to analyzing the data (Wickham 2014).

Each row represent a observation and each column represent a variable. This is the preferred method how to organize data according to(Wickham 2014). Below is an image of how the AIS-data is organized from file:

	IMONumber	TimestampPosition	Longitude	Latitude	Heading	SpeedOverGround	ShipType	Length	unix
1	9536521	2018-06-04 00:00:04	10.270005	63.47954	271	8.3	cargo_ships	89	1528070404
2	9536521	2018-06-04 00:10:05	10.21974	63.4794083333333	268	8.1	cargo_ships	89	1528071005
3	9536521	2018-06-04 00:20:06	10.1665833333333	63.47936	268	9	cargo_ships	89	1528071606
4	9536521	2018-06-04 00:30:14	10.108645	63.4789383333333	269	9.3	cargo_ships	89	1528072214
5	9536521	2018-06-04 00:40:24	10.05038	63.4776783333333	274	9.1	cargo_ships	89	1528072824
6	9536521	2018-06-04 00:50:24	9.99481333333333	63.4795516666667	289	8.9	cargo_ships	89	1528073424
7	9536521	2018-06-04 01:00:04	9.94693	63.489025	307	8.2	cargo_ships	89	1528074004
8	9536521	2018-06-04 01:10:06	9.91273833333333	63.5058183333333	325	8.4	cargo_ships	89	1528074606

Figure 6: The first 8 rows of the data set is shown. Each observation is a row number with columns representing the vessel’ characteristic/information at a given time.

As seen from figure (6) the structure of the data is up to standard with the guide lines from (Wickham 2014). Next, we calculate the velocity components of each observation by using 2 columns from the data set.

4.1.2 Calculating velocities

Velocities are useful units to calculate. This grants the possibility to visualize a vessels direction. Here, we will see which vessels travel to and from Trondheim. Vessels may take detour, traveling through Trondheim to then converge to Trondheim after a while. From figure (6) the velocities can be found by using columns: heading and speedoverground. The velocities can be found with these two equations:

$$v_{lon} = SOG \cdot \sin(heading) \tag{32}$$

$$v_{lat} = SOG \cdot \cos(heading) \tag{33}$$

v_{lon} and v_{lat} are the velocity components of the corresponding longitude and latitude direction, respectively. SOG is the speed over ground and heading is the vessels angle in radians.

4.1.3 Route extraction

When visualizing the data can be done in many ways. Often as the data set is big, plotting every data point can be messy and misleading. Here, some information about how one route emerges and what classifies a route is of interest. We want to see where each vessel travels and which path it takes. By looking at the timestamp column in the data set, there is often a trend of logged data from a certain time interval. This could indicate the start and end of a route. Keeping in mind that it could also mean some else e.g. turning of the tracking system, malfunction in the power etc.

A check can be done with the velocities v_{lon}, v_{lat}. If the direction for each time series has the same direction they follow the same route.

5 Maps and coordiante systems

The knowledge of the earth's shape has been known by cartographers for the past 2000 years (Snyder 1987). When traveling far it was necessary to know where and distance of travel. How to measure distances on the globe has been needed for many years when people needed to navigate from one place to the other.

This is mentioned here as some of the data we will be working with is logged as longitude and latitude coordinates. A short motivation why these quantities are used on maps will be mentioned in the coming sections. Additionally, how convert a spherical surface onto a rectangular surface we be focus in the following section.

5.1 Geographic coordinate system

A coordinate system is useful tool when working with positional data. This gives insight distances in multiple dimensions and also from the use of mathematics grants the possibility of calculating positions from a given reference point or relative to given points.

When determining positions on a spherical shaped object e.g. the earth, one must take into account the curvature and how distances are measured. The measures of longitude and latitude have been implemented and defined in such a way to make a structured way to split up the globe. These are angular measurements. There are several ways of defining latitude and longitude: geocentric , astronomical and geographic. The latter is most widely used.

Latitude, also known parallels are measured from north to south direction with an interval of 180 °. Equator is located at 0 °. The earth is not a perfect sphere. This affects the relative lengths of latitudes. 1 °of latitude amounts to 110.567km at Equator and 111.699 km at the poles (Encyclopedia britannica 2020).

Longitude, also known as meridians are measured from west to east with an interval of 180 °. The prime meridian, which is where 0 °is defined, runs through Greenwich, London The relative distances are equal for 1 °of longitude and equals 111.32 km (Encyclopedia britannica 2020). Longitude and latitude coordinates make up a geographic coordinate system. Positional information can be determined in terms of the prime meridian and the Equator. See figure below:

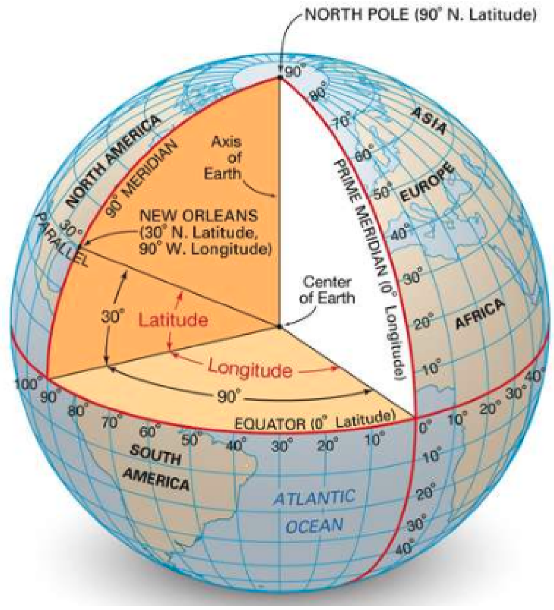

Figure 7: Here we see how each angle og longitude and latitude is defined from the earth's center. Figure gathered from (Encyclopedia britannica 2020).

5.2 Map selection

The data used in the thesis are logged from the surface of a sphere i.e. the earth. When it comes time to visualize the data, this is done on flat surface. Some background is needed in order to find the appropriate representation. In this section, an introduction to maps and the different ways to represent maps in two dimensions are presented here. As we know the earth has a spherical shape, but in geography class the earth was usually shown to us on a two dimensional map. This is known as a map projection. If you tried to flatten the earth onto a rectangular plane and also maintaining properties such as geometry, area size and distance you would not succeed (see figure **??**). An example is trying to peel an orange and make it's shape rectangular. Converting from three to two dimensions leads to distortion (Snyder 1987). What properties to preserve is chosen by the way am projection is done.

Figure 8: Here, the earth flattened out without any distortion. Linking together the pieces comes at the price of distortion. Figure gathered from (Šavrič and Kennedy, Melita 2020).

5.3 Projections

When deciding which projection to use, it is useful to consider what it is most important quantity to preserve. Below we state some projection properties which are commonly considered:

Equal-area preserves the area-size of objects. This is useful when wanting to see the relative size of multiple objects.

Conformal preserves the angle between points. This is important when navigating at sea and when using large- scaled maps.

Equidistant maintains the distance between points. Preserving this quantity comes in handy when wanting to compare several distances.

Compromise considers some distortion from all properties, but attempting to balancing it.

There are three types of methods commonly used when it comes to map projection: cylinder, cone and plane.

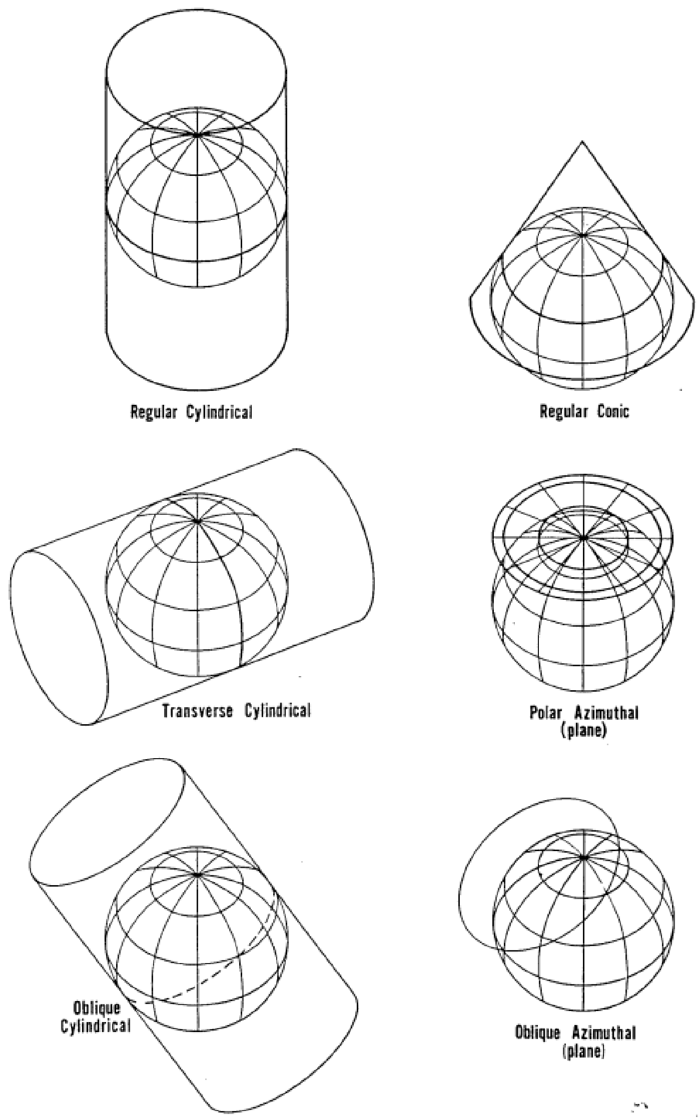

Figure 9: Here we see how the different projections are made from a spherical surface to a flat surface. Figure gathered from (Snyder 1987).

References

Acheson, D.J (2003). *Elementary Fluid Dynamics*. Oxford.

Besse, Philippe et al. (Aug. 2015). "Review and Perspective for Distance Based Trajectory Clustering". en. In: *arXiv:1508.04904 [cs, stat]*. arXiv: 1508.04904 [cs, stat].

Encyclopedia britannica (2020). *Latitude and Longitude*. https://www.britannica.com/science/latitude.

Gjevik, Bjørn (2009). *Innføring i Fluidmekanikk*.

Hvamb, Knut (2015). "Motion Planning Algorithms for Marine Vehicles". en. In: p. 153.

Lewis, Charles Lee (1927). *Matthew Fontaine Maury, the Pathfinder of the Seas*. United States naval institute.

Meijer, Ricardo (2017). "ETA Prediction Predicting the ETA of a Container Vessel Based on Route Identification Using AIS Data". PhD thesis.

Pallotta, Giuliana, Michele Vespe, and Karna Bryan (June 2013). "Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction". en. In: *Entropy* 15.12, pp. 2218–2245. ISSN: 1099-4300. DOI: 10.3390/e15062218.

Pedersen, Morten D. and Thor I. Fossen (2012). "Marine Vessel Path Planning & Guidance Using Potential Flow". en. In: *IFAC Proceedings Volumes* 45.27, pp. 188–193. ISSN: 14746670. DOI: 10.3182/20120919-3-IT-2046.00032.

Šavrič, Bojan and Kennedy, Melita (2020). *Map Projections*. https://www.google.com/url?sa=i&url=https
G4borMyEPS8um7oLQP&ust=1582711305852000&source=images&cd=vfe&ved=0CAIQjRxqFwo

Snyder, John (1987). *Map Projections- a Working Manual*.

UN-Business Action Hub, United Nations (2020). *Business.Un.Org*. https://business.un.org/en/entities/13.

Wang, Xuantong, Jing Li, and Tong Zhang (Dec. 2019). "A Machine-Learning Model for Zonal Ship Flow Prediction Using AIS Data: A Case Study in the South Atlantic States Region". en. In: *Journal of Marine Science and Engineering* 7.12, p. 463. ISSN: 2077-1312. DOI: 10.3390/jmse7120463.

Wickham (2014). "Tidy Data". In: *Journal of statistical software*, p. 24.