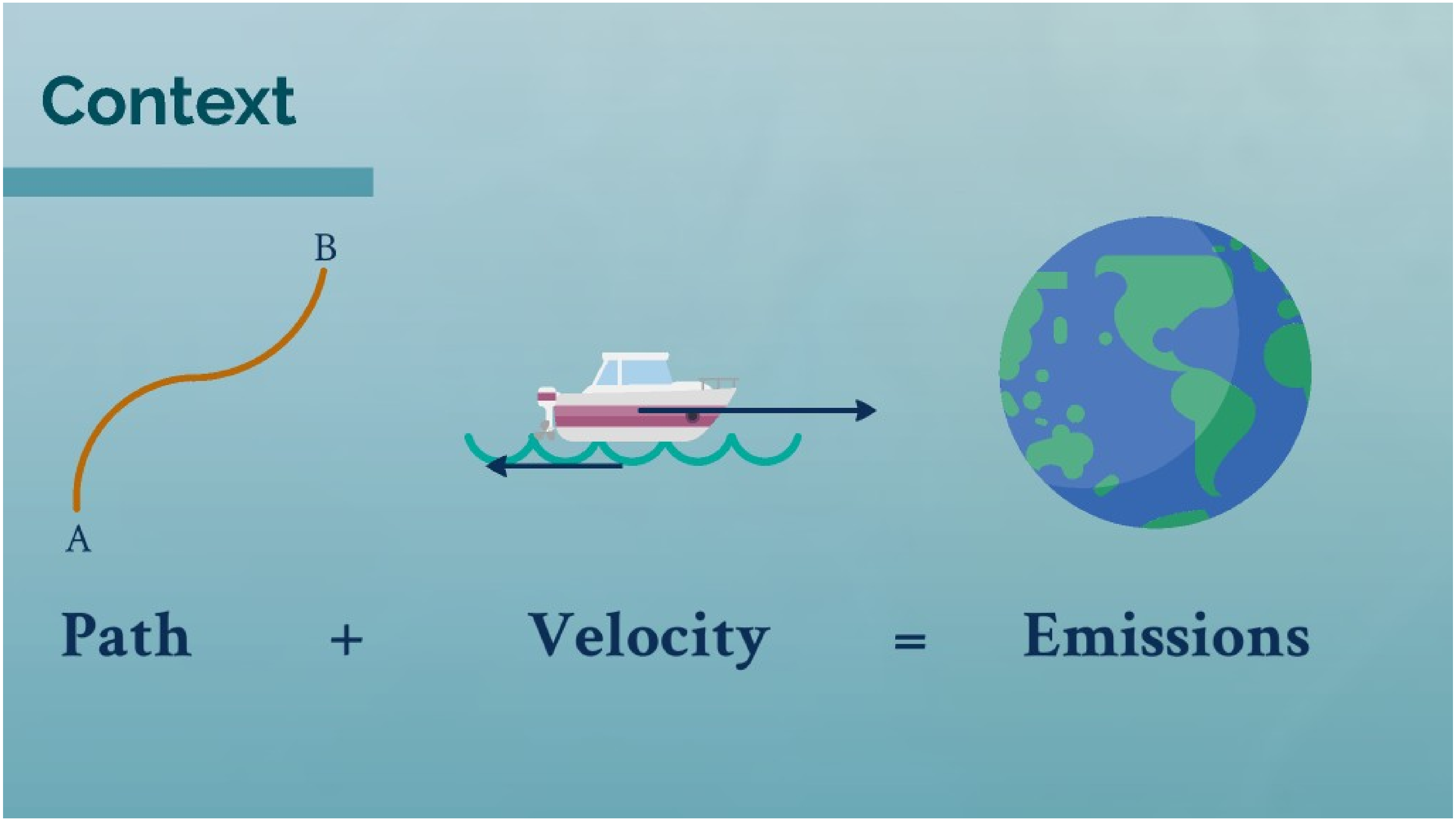

SNC·LAVALIN
Interpolating ship paths in the Port of Metro Vancouver
Eylul Aygun
Xianchan Long
Kateryna Melnykova
Cierra Moffat
Danil Platonov
Mohsen Seifi
Tianxin Tao
Shanwen Yan
PIMS BC Data Science Workshop 2018
FUTURE STEPS
SOLUTION
PROJECT GOALS AND DATASET
OUTLINE
June 8, 2018


Presentation Outline
Project Goals and Data Set
Solution
Future Steps

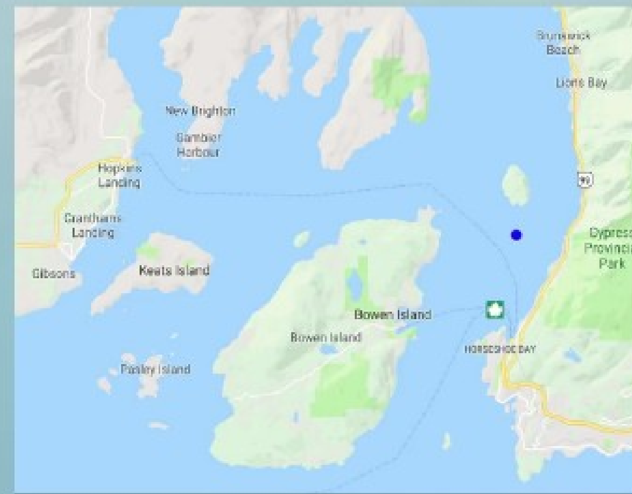

SNC·LAVALIN
Interpolating ship paths in the Port of Metro Vancouver
Eylul Aygun
Xianchan Long
Kateryna Melnykova
Cierra Moffat
Danil Platonov
Mohsen Seifi
Tianxin Tao
Shanwen Yan
PIMS BC Data Science Workshop 2018
FUTURE STEPS
SOLUTION
PROJECT GOALS AND DATASET
OUTLINE
June 8, 2018


Project Details
Dataset
Goals


Context
B
A
Path + Velocity = Emissions

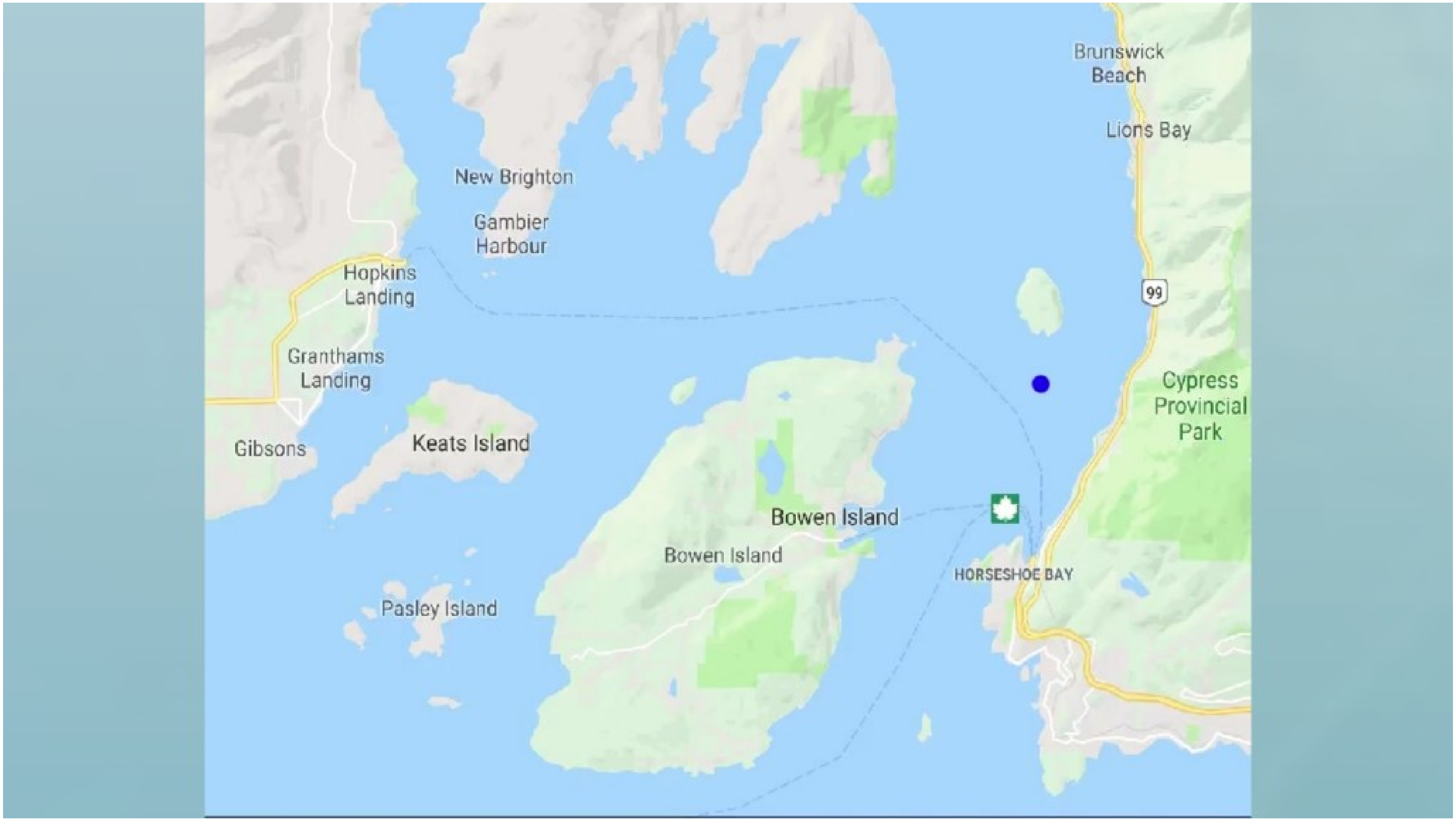
Main Goal

Predict the path of any given boat



Recognize and **fix data with large amounts of noise**

Fix ship paths so they **do not go over land**


Brunswick Beach
Lions Bay
New Brighton
Gambier Harbour
Hopkins Landing
Granthams Landing
99
Cypress Provincial Park
Gibsons
Keats Island
Bowen Island
Bowen Island
HORSESHOE BAY
Pasley Island

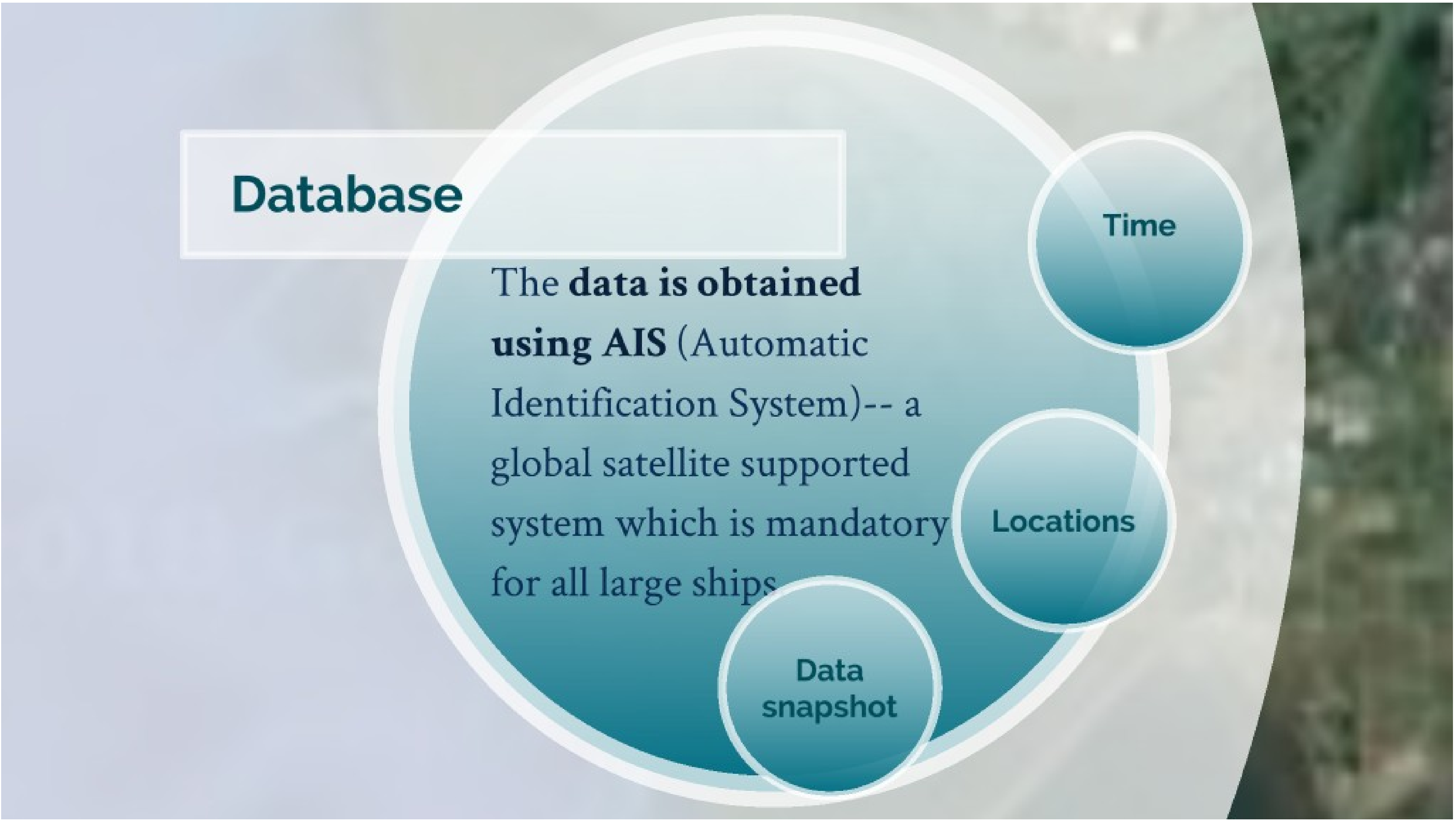
Data

Current Data

Past 10 days	Every 10s -1 day	~460 Boats

Historial Data

Past 87 days	Every 5min -1 day	~2100 Boats

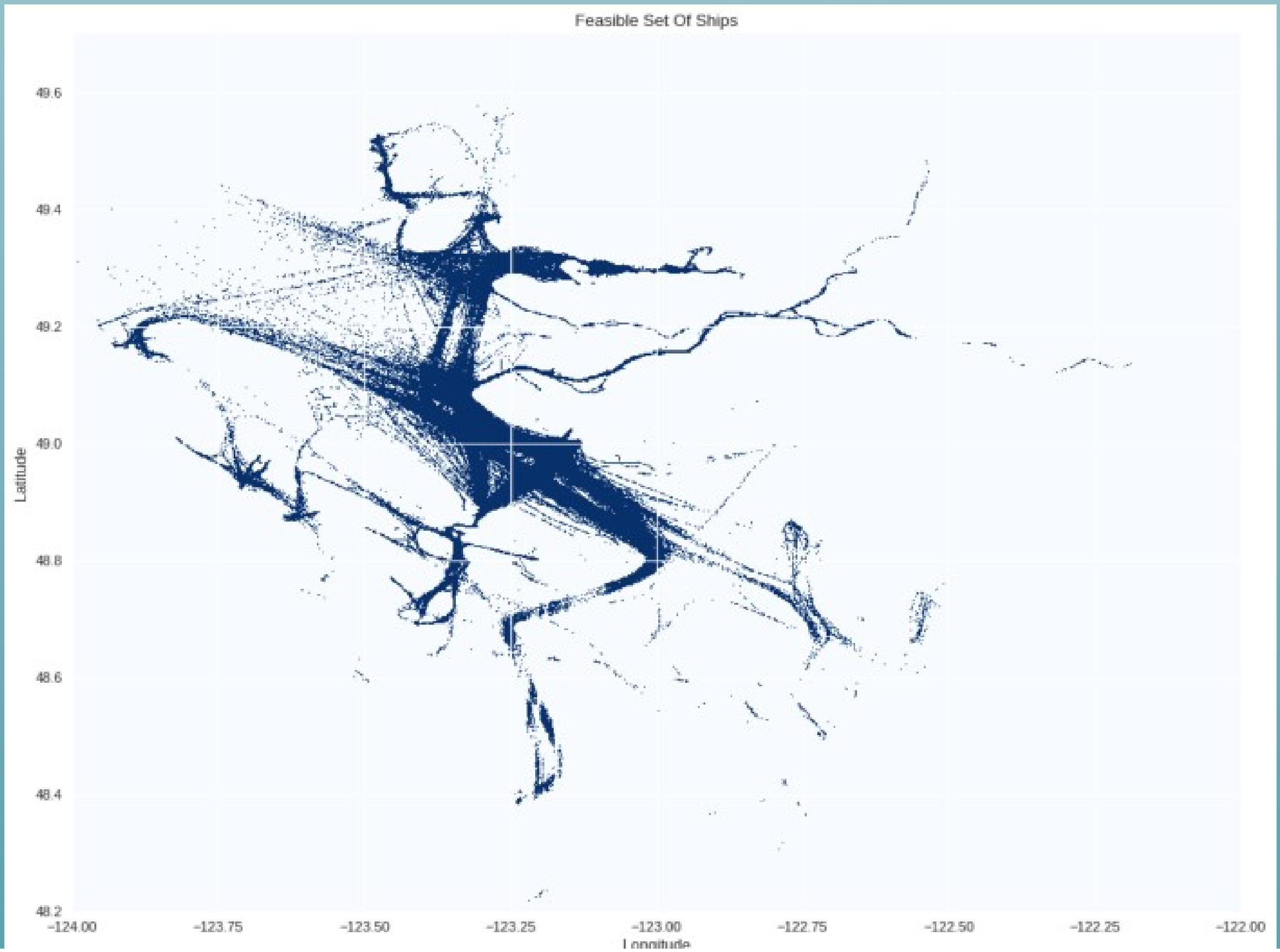

Database
The data is obtained using AIS (Automatic Identification System)-- a global satellite supported system which is mandatory for all large ships.
Time
Locations
Data snapshot

Data snapshot

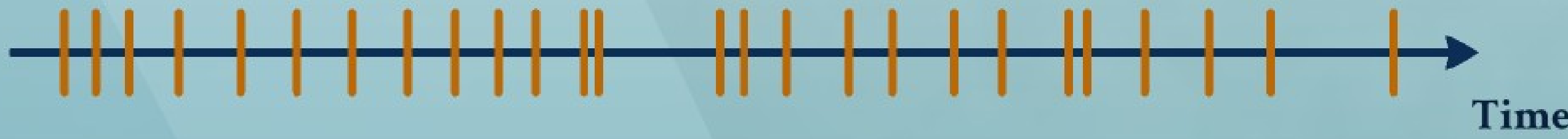
	UserID	NavigationalStatus	ROT	SOG	PositionAccuracy	Longitude	Latitude	COG	TrueHeading	TimeStamp	ReceivedTime	rowid
0	316005621	0	-127.0	0.1	0	-122.77156	49.23065	32.6	511.0	24	2013-10-22 01:05:24.510	311057489
1	316018851	0	0.0	1.7	0	-123.05445	49.29853	67.0	110.0	24	2013-10-22 01:05:25.400	311057490
2	316003679	2	-127.0	0.1	0	-123.10751	49.31308	150.8	511.0	24	2013-10-22 01:05:25.853	311057491
3	316014621	0	127.0	12.3	0	-123.09534	49.29940	210.7	210.0	26	2013-10-22 01:05:26.027	311057492
4	316005721	0	-127.0	0.1	0	-123.10684	49.31094	166.9	511.0	25	2013-10-22 01:05:26.620	311057493

	Observation	Days	Ships
Delta_Current	2560654	10	460
Delta_History	2296035	83	2192
NewWest_Current	3810941	10	525
NewWest_History	2750853	83	2217

Locations


Feasible Set Of Ships
Latitude
Longitude
49.6
49.4
49.2
49.0
48.8
48.6
48.4
48.2
−124.00
−123.75
−123.50
−123.25
−123.00
−122.75
−122.50
−122.25
−122.00

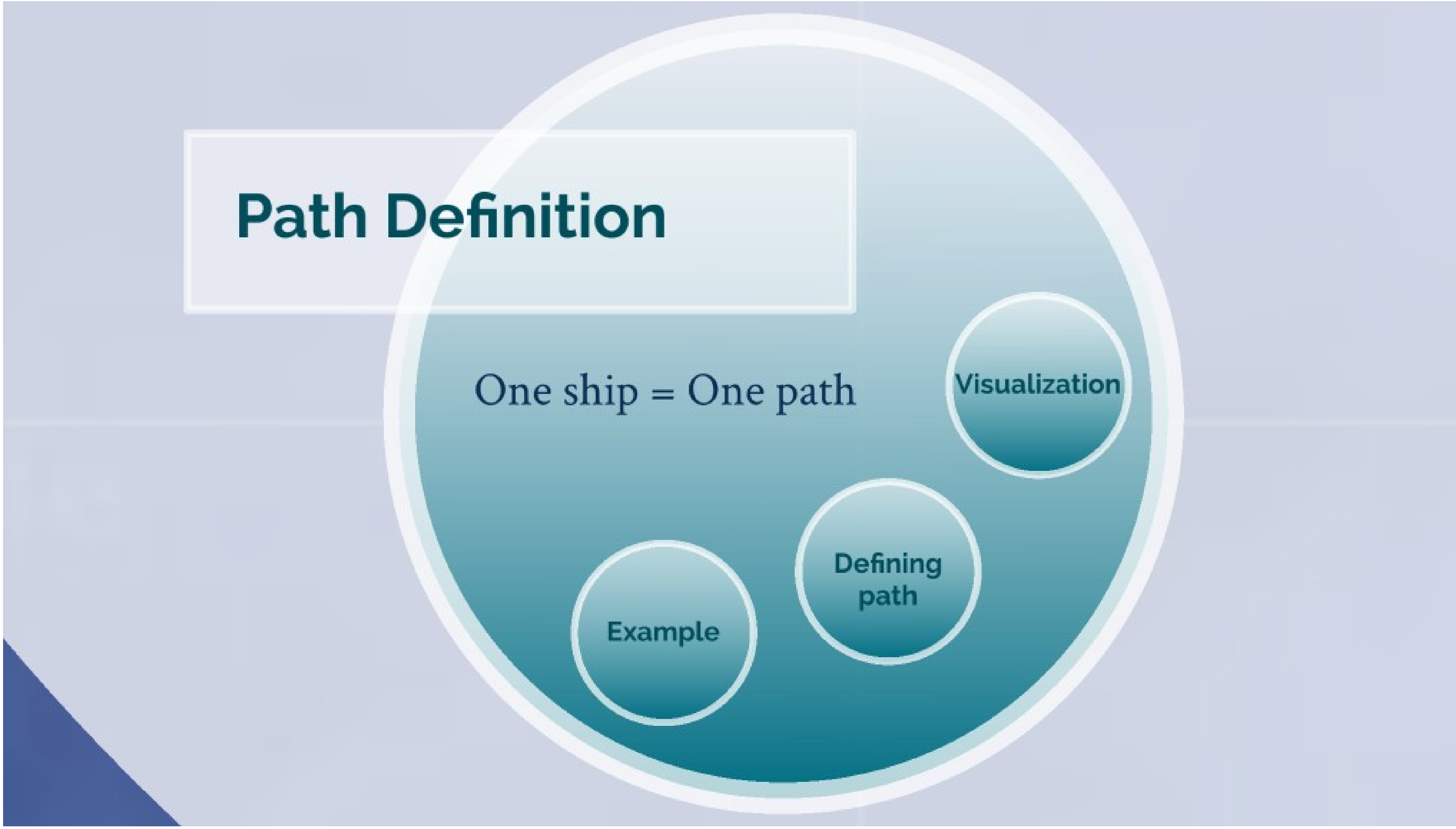
Time - Sparcity of Data



Data is not linearly spaced with time


SNC · LAVALIN
Interpolating ship paths in the Port of Metro Vancouver
Eylul Aygun
Xianchan Long
Kateryna Melnykova
Cierra Moffat
Danil Platonov
Mohsen Seifi
Tianxin Tao
Shanwen Yan
PIMS BC Data Science Workshop 2018
FUTURE STEPS
SOLUTION
PROJECT GOALS AND DATASET
OUTLINE
June 8, 2018

Solution
Path definition
Regularize time
Locate missing samples
Interpolate path
Historical data reconst


Path Definition
One ship = One path
Visualization
Defining path
Example

Example

Ship ID = 218042000

Definition of path

(1) If there is **no data for the ship for at least 1 hour** and the last known location of the ship is **close to boundary of Port of Metro Vancouver**, then we **assume that the ship leaves Metro Vancouver area**. New path starts from theat point.

(2) We **split the data in two chunks** if there is no data for more than **2 hours**.

Path visualization

Ship ID = 218042000

Regularize time

Recall: samples are taken irregularly.

Solution: granularity of time. We set time accuracy to 6 minutes.

Details: For multiple samples at the same time, we store the average

Missing samples: where

Since **data is not taken regularly**, what samples do we consider to be missing?

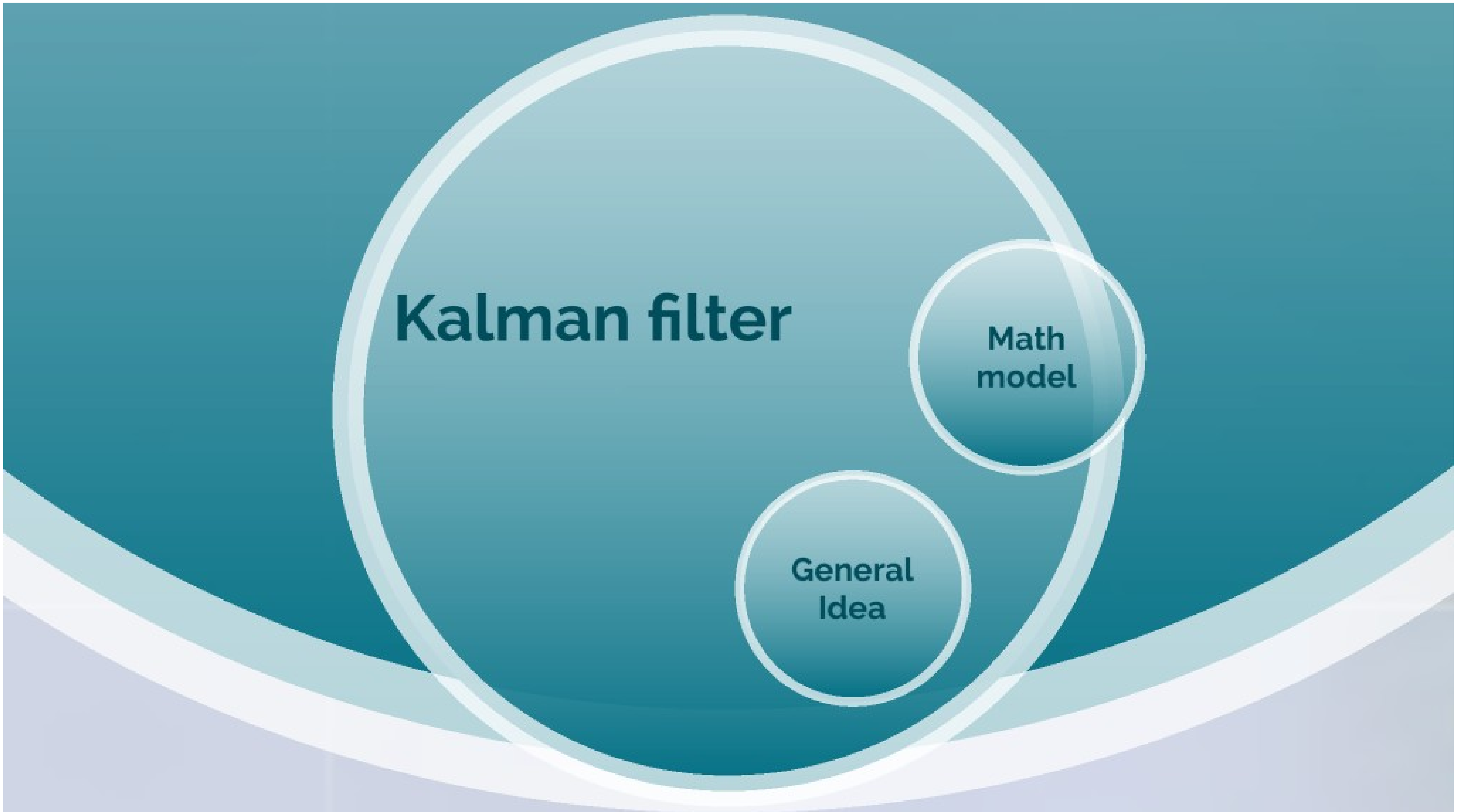
Now, with granular and regularized time, **the sample is missing if no data was taken for 6 minutes**.

Interpolating Path

Accuracy

pykalman

Kalman filter


Kalman filter
Math model
General Idea

General Idea of the Kalman Filter

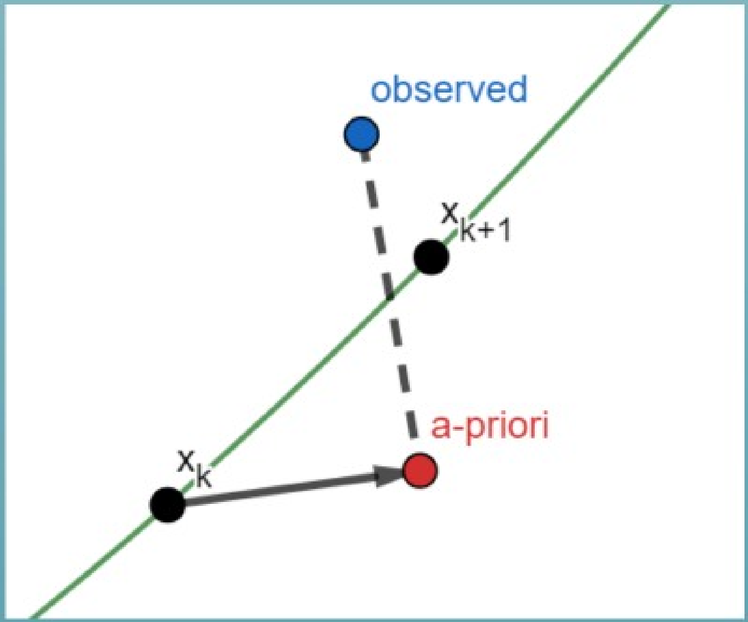
Suppose that we know ship location, velocity and acceleration. Can we predict new location?

$$\begin{bmatrix} x_{k+1}^{\text{a-priori}} \\ v_{k+1}^{\text{a-priori}} \\ a_{k+1}^{\text{a-priori}} \end{bmatrix} = \begin{bmatrix} x_k & + & v_k \Delta t & + & a_k \frac{\Delta t^2}{2} \\ & & v_k & + & a_k \Delta t \\ & & & & a_k \end{bmatrix} + \begin{matrix} \text{process} \\ \text{noise} \end{matrix}$$

Now suppose that we also observe the ship location. Can we refine the reconstruction?

$$\begin{bmatrix} \widetilde{x}_{k+1} \\ \widetilde{v}_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ v_k \end{bmatrix} + \begin{matrix} \text{sampling} \\ \text{noise} \end{matrix}$$

Can we refine the interpolation?

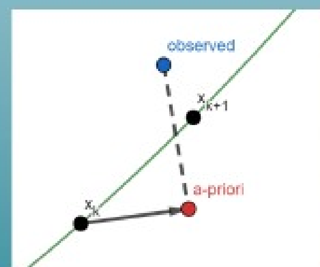


Consider a linear combination of a-priori prediction and observed point

Kalman Filter: Math formulation

$$x_{k+1}^{\text{a-priori}} = T\, x_k^{\text{true}} + \text{process noise}$$

$$x_{k+1}^{observed} = M\, x_{k+1}^{true} + \text{sampling noise}$$



$$x_{k+1}^{\text{a-post}} = x_{k+1}^{\text{a-priori}} + K(H x_{k+1}^{\text{observed}} - x_{k+1}^{\text{a-priori}})$$

Using the Kalman Filter with pykalman

pykalman is a Python library that implements Kalman filter.

Pros:

1. pyklaman handles missed entries

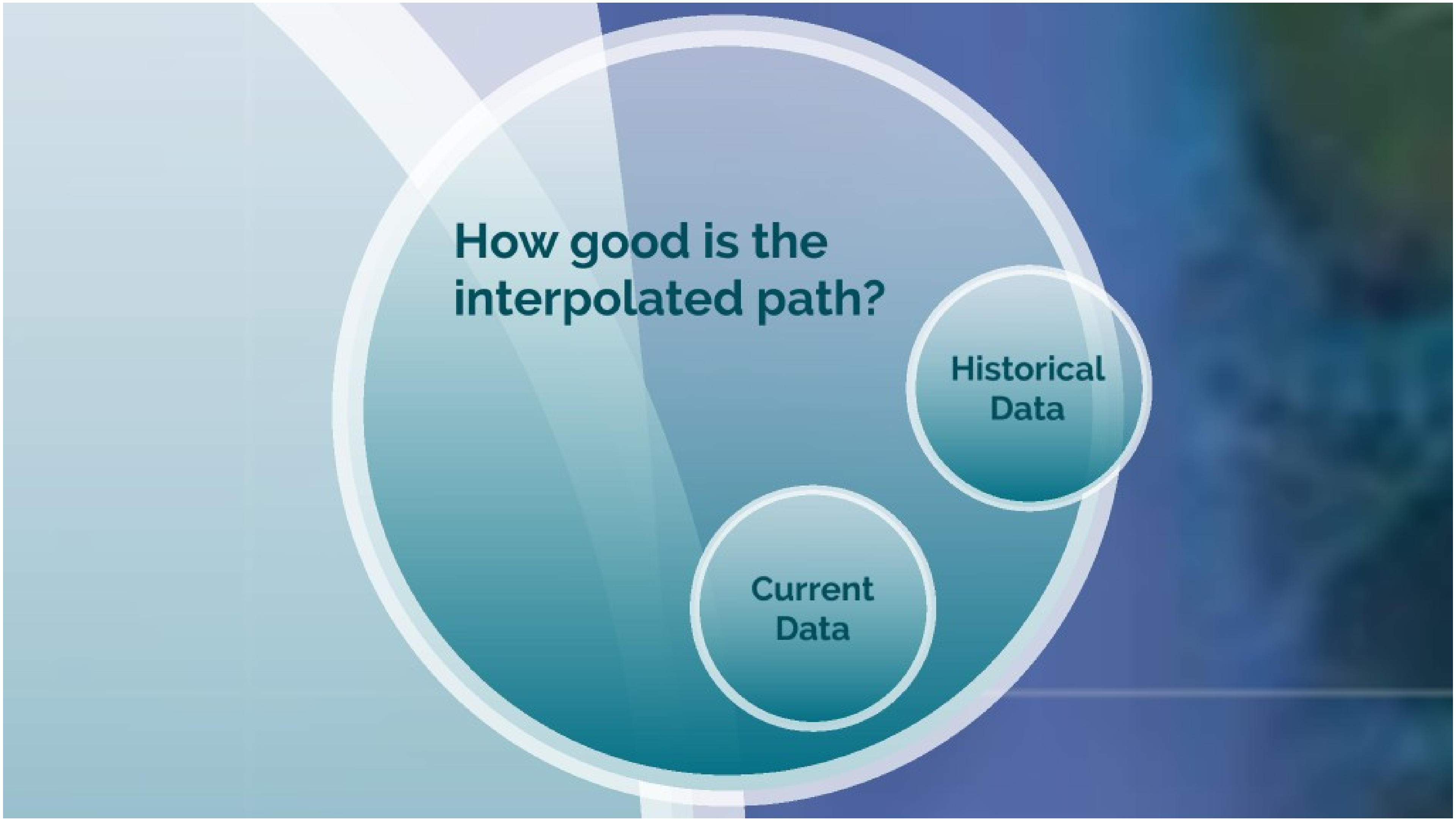
2. pykalman approximates covariance matrices of process noise and measurement

3. pykalman also smooths the path

Cons:

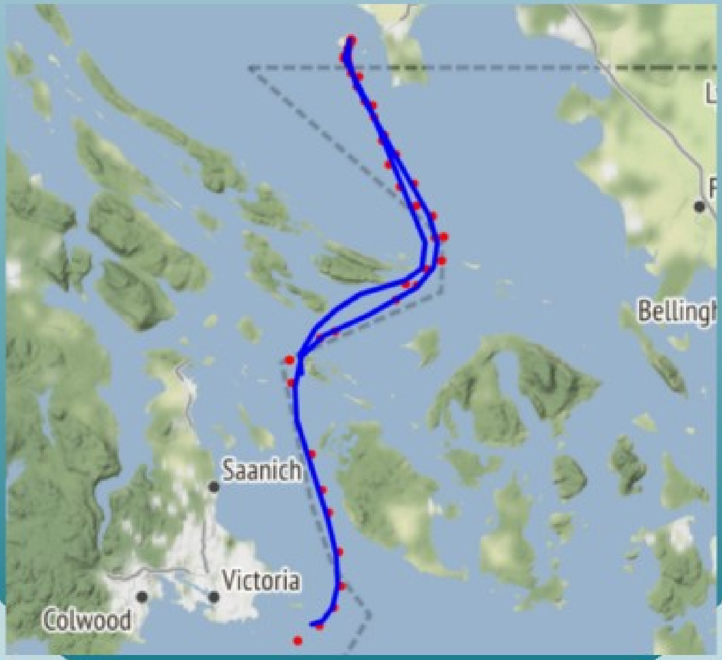
1. pykalman does not allow to inject different distribution of the measurement noise over path

2. Noise is assumed to be Gaussian

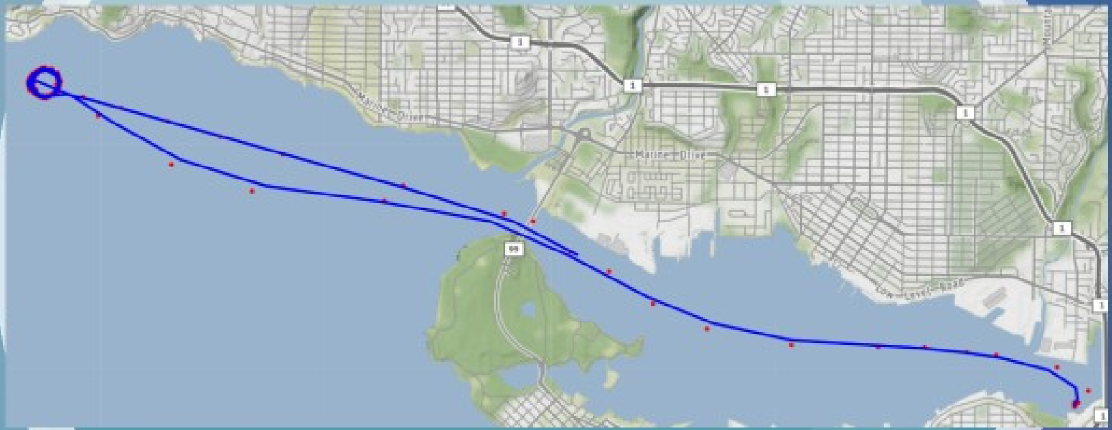


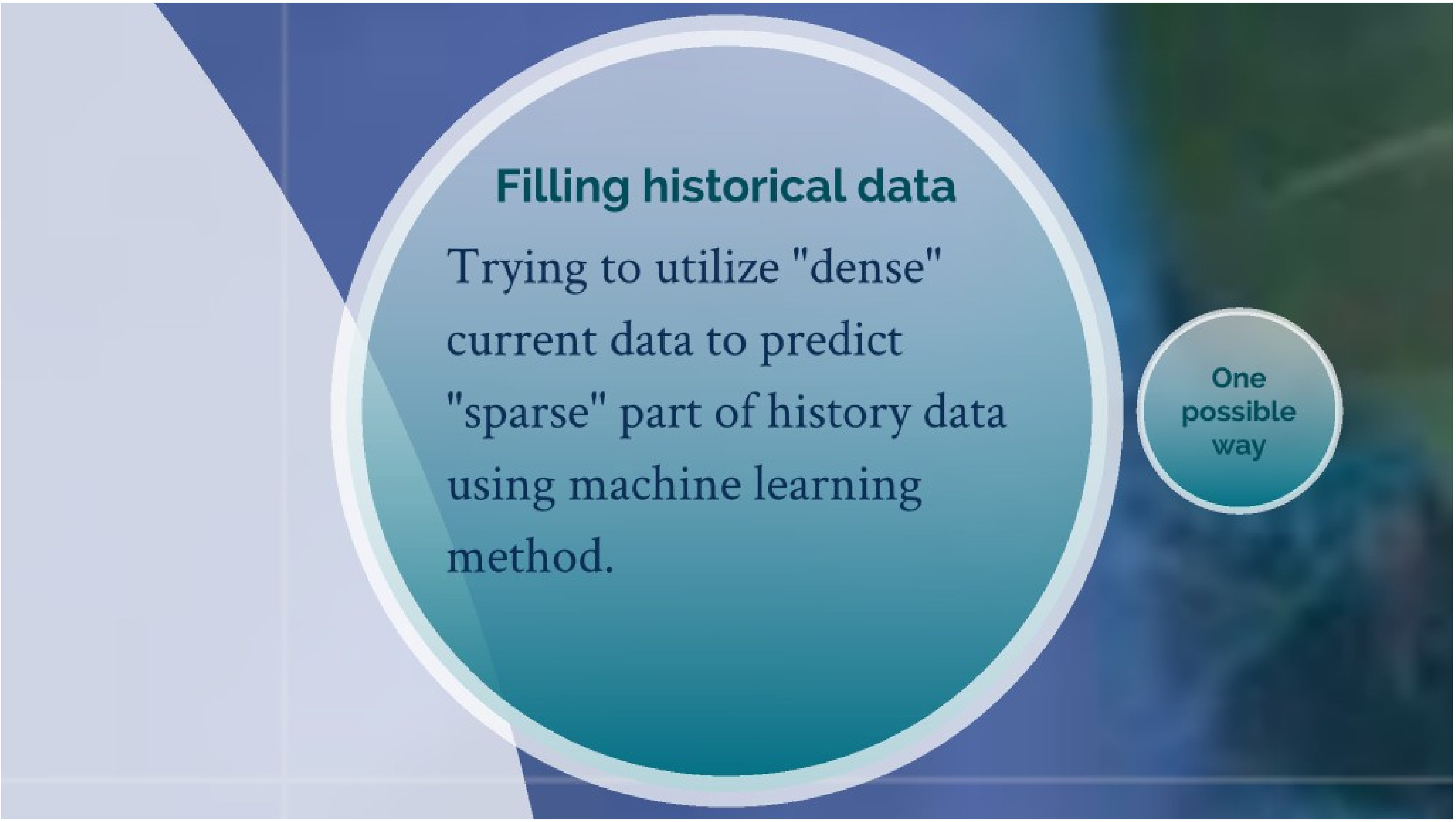
How good is the interpolated path?
Historical Data
Current Data

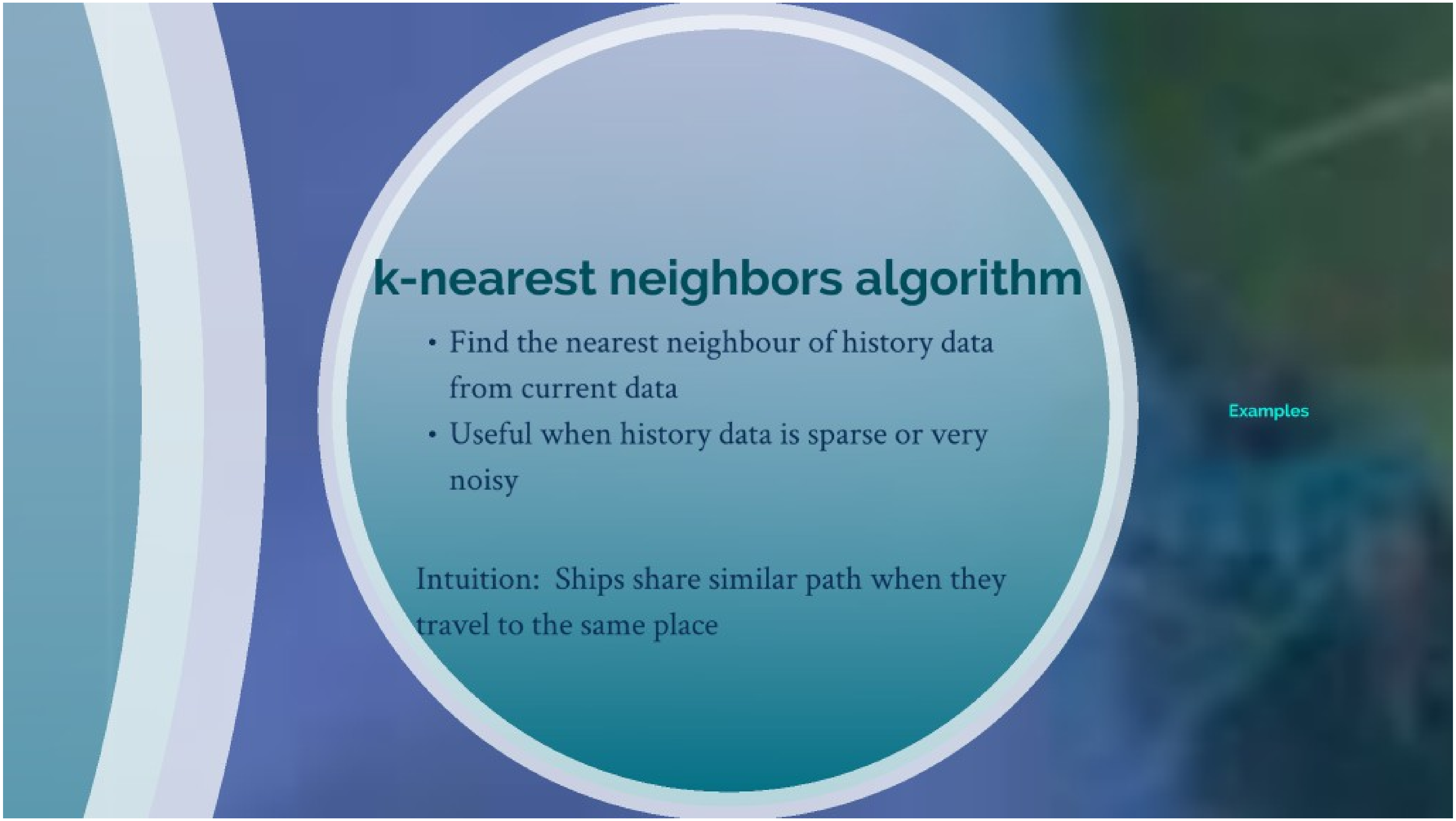
Interpolation of Current Data



Interpolation on Historical Data



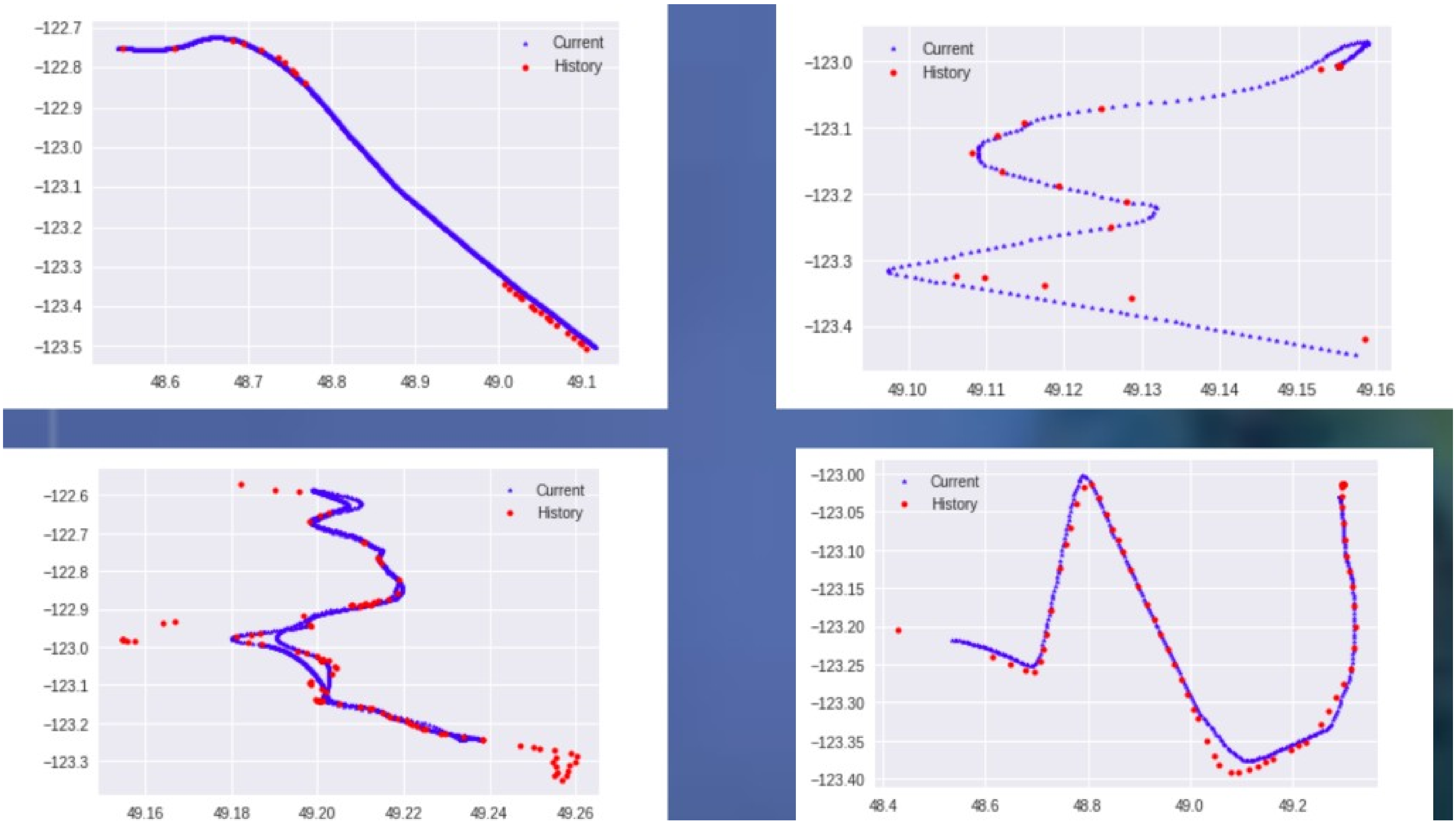

Filling historical data
Trying to utilize "dense" current data to predict "sparse" part of history data using machine learning method.
One possible way


k-nearest neighbors algorithm
• Find the nearest neighbour of history data from current data
• Useful when history data is sparse or very noisy
Intuition: Ships share similar path when they travel to the same place
Examples

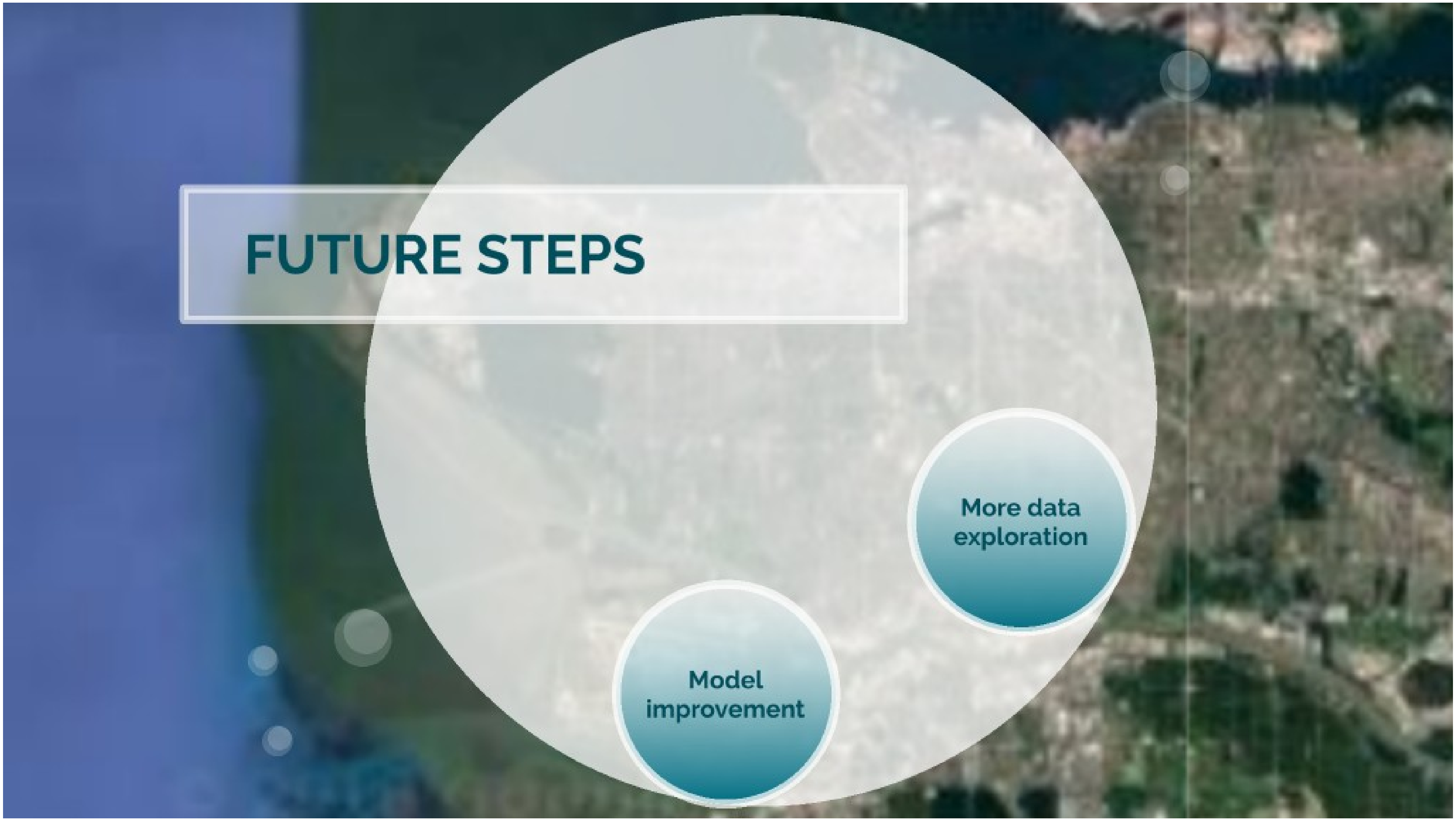
e nearest neighbour of histo

ırrent data

when history data is sparse o


Current
History
Current
History
Current
History
Current
History


SNC·LAVALIN
Interpolating ship paths in the Port of Metro Vancouver
Eylul Aygun
Xianchan Long
Kateryna Melnykova
Cierra Moffat
Danil Platonov
Mohsen Seifi
Tianxin Tao
Shanwen Yan
PIMS BC Data Science Workshop 2018
FUTURE STEPS
SOLUTION
PROJECT GOALS AND DATASET
OUTLINE
June 8, 2018


FUTURE STEPS
More data exploration
Model improvement

Model improvement

Model improvement:

1.Measurement noise should be uniformly distributed rather than Gaussian.

2. We need to relax the assumption that noise is indentically distributed.

There are packages of modified noise however they are not able to handle missing data.

More data exploration

1. What are the most common ship paths? Does ship repeats its path?
2. Any correlation of location, ship type, etc. with frequency of measurements?
3. Can we come up with a method of suggesting ship path even if no data was taken over really long period? (Previous trips, similar paths, etc.)


SNC·LAVALIN
Interpolating ship paths in the Port of Metro Vancouver
Eylul Aygun
Xianchan Long
Kateryna Melnykova
Cierra Moffat
Danil Platonov
Mohsen Seifi
Tianxin Tao
Shanwen Yan
PIMS BC Data Science Workshop 2018
FUTURE STEPS
SOLUTION
PROJECT GOALS AND DATASET
OUTLINE
June 8, 2018