Climate Crisis Ai Hackathon | Jan 22-24

#challenge-sea-ice

Estimate the velocity of drifting sea ice to help us better understand the dynamical response of the Arctic Ocean to a changing climate.

A kaggle-style challenge proposed by Charles Brunette & Bruno Tremblay, McGill University

Background

Arctic sea ice drifts under the influence of winds and ocean currents.

Ice motion is important - up to several kilometers a day. Tracking the movement of sea ice can be applied to different research questions, including the transition from older sea ice to a younger seasonal ice pack, the transport of ice-rafted sediments, or pollutants in the context of an oil spill, and the risk assessment associated with navigation and marine operations in the Arctic. In the McGill sea-ice group, we use ice motion tracking to understand how the dynamics of sea ice can be used as a seasonal predictor for the summer ice conditions. As sea ice experiences convergent motion, it ridges and forms thick piles of sea ice, or to the contrary, divergent ice motion leads to the formation of open water areas or leads. These changes in the ice thickness distribution of the pack ice consequently influence the melt of sea ice in the summer.

Information about the velocity of sea ice is required for performing sea ice tracking; and in a perfect world, we would like this information to be available everywhere in the Arctic in space and time. The most reliable information about ice motion comes from passively drifting buoys that were deployed in the Arctic over the last few decades. These GPS-equipped instruments provide a very precise measure of ice motion, but unfortunately offer limited spatial coverage. Remote sensing offers a better spatial coverage; image processing on sequences of satellite imagery can be used to reconstruct ice motion. Still, these satellite-derived ice motion vectors may not be available all year long due to the sensitivity of the different instruments. When neither of these two observational sources of ice motion is available, we need to make a best estimate of the sea ice drift to fill the gaps. As the atmospheric forcing is one of the main drivers of the circulation of sea ice, we can use information about the winds for estimating the sea ice drift.

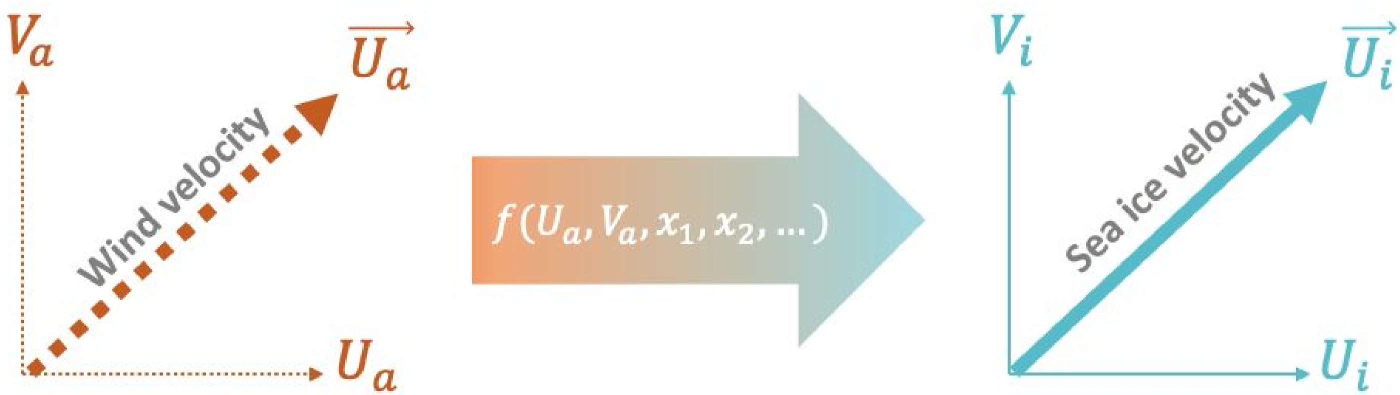
Objectives

In this challenge, your task is to use an AI approach for building the model that best reproduces the buoy drift, based on the wind fields. For doing so, you will be provided with a forty years data record of drifting Arctic buoys, and co-located winds, and additional environmental parameters (such as sea ice conditions: sea ice concentration, sea ice thickness). If successful, your model could be used to produce sea ice motion estimates that will be combined to other observations of sea ice motion and help the polar research community. **We are interested in collaborating on future projects!**

Deliverables & evaluation

- Using the DRIFT_DATA_TRAINING.csv data, build a model that estimates daily u-components and v-components of sea ice velocity (U_i and V_i), based on wind velocity (U_a and V_a) and the other parameters that are provided.
- Return a list of ice motion estimates (U_i and V_i) from the test data provided in DRIFT_DATA_TEST.csv; this output will be compared to buoy drift observations to evaluate the performance of your model.
- The evaluation will be based on the root-mean-square error for:
    - Error on sea ice speed: $\Delta|U| = |U(estimate)| - |U(buoy)|$
        - where speed is defined as: $|U| = \sqrt{[U^2 + V^2]}$
    - Directional error: $\Delta\theta = \theta(estimate) - \theta(buoy)$

Note : Vector components of sea ice drift and wind are provided in the data. Schematically, your task is:



ADDITIONAL RULES :

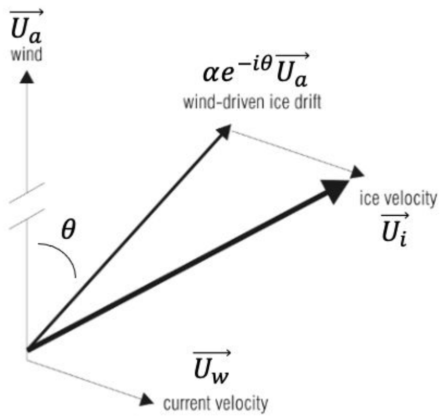
- **<u>You cannot use future information (*t+1*) to predict *U_i*(t), the velocity at time *t*.</u>**
- **<u>The buoy ID cannot be used as an input.</u>**

A (very) short sea ice drift primer

In global climate models and coupled ice-ocean models, sea ice is generally treated as a viscous-plastic fluid, and sea dynamics are represented by a Navier-Stokes equation, with a specific rheology that defines the mechanical properties of sea ice. While offering a full picture of sea ice processes, the drawback of these representations of sea ice is that they are challenging to implement and computationally very intensive.

For studying certain features of the sea ice system, such as wind-driven sea ice circulation, we can make a set of assumptions that greatly simplify the problem while still capturing most of the dynamical features. For instance, if we consider ice motion at equilibrium and the first order balance between the wind stresses and the ocean drag, based on the physics we can derive a **linear relationship between sea ice velocity (Ui), wind velocity (Ua) and the ocean currents (Uw)**. This classical representation is what we call the **free drift of sea ice**:

$$\overrightarrow{U_i} = \alpha e^{-i\theta}\overrightarrow{U_a} + \overrightarrow{U_w}$$



(adapted from Lepparanta, 2011)

Where α **is a wind-ice transfer coefficient that takes a value between 1-2%** (converting wind speed in m/s to sea ice speed in cm/s); and θ **is a turning angle between the wind direction and the sea ice drift direction**, due to the Coriolis effect because of the Earth's rotation, the wind turning angle usually ranges between 20-40°. This simple free drift can capture somewhere up to 70% of the variability of sea ice motion (Thorndike & Colony, 1982). While this can serve as an inspiration, we are eager to see what new solutions will come up from your AI/ML approaches to this problem!

TIPS & COMMENTS:

- **The u and v components of velocity are coupled (via angle θ), keep this in mind!**
- **Observations of ocean currents under the ice are not available.**

References: All about sea ice (NSIDC); Lepparanta, The drift of sea ice (2011); Thorndike & Colony, Sea ice motion in response to geostrophic winds (1982).

Data

The data provided to you for this challenge is a table containing a **list of data points** obtained from drifting buoys deployed in the Arctic Ocean over the last 40 years (International Arctic Buoy Programme). The list is sorted timewise. For every buoy data point, at the same timestamp and buoy location, we retrieve additional parameters from other datasets that can help build a model for the ice drift. Note that all the physical parameters represent **daily averages**.

Quicklook : Here are a few entries from the data table:

	1 year	2 month	3 day	4 doy	5 x_EASE	6 y_EASE	7 u_buoy	8 v_buoy	9 id_buoy	10 u_ERA5	11 v_ERA5	12 sic_CDR	13 h_cs2smos	14 h_piomas	15 d2c
424311	2019	12	30	364	180.9766	187.9224	3.5643	-9.3813	9990	-1.4221	-7.3527	1.0000	1.7522	1.7069	549.3100
424312	2019	12	30	364	235.0014	157.5116	-0.0049	0.0051	74645	-2.9429	2.7542	0.5438	NaN	0.0431	44.8075
424313	2019	12	30	364	193.9023	173.3384	0.5298	-11.7238	81140	-2.4314	-5.5106	1.0000	1.4140	1.5967	683.6004
424314	2019	12	30	364	209.7856	171.4870	-0.8406	3.1051	12200	0.0188	4.9470	1.0000	1.2176	1.2989	312.4070
424315	2019	12	30	364	140.6385	133.5584	-1.4234	1.3588	118	-4.4376	4.6231	1.0000	1.5611	1.7128	489.5808
424316	2019	12	30	364	152.5892	163.5544	-4.3706	4.8471	32110	-0.1750	4.5827	1.0000	2.5245	2.1838	460.4839
424317	2019	12	30	364	193.2827	173.8235	-0.0046	-12.5942	66760	-2.5688	-5.8641	1.0000	1.4104	1.6289	697.4674
424318	2019	12	30	364	180.2735	160.3775	-5.4600	-11.7731	45951	-4.5424	-6.1481	1.0000	1.6443	1.8141	988.1774
424319	2019	12	30	364	141.8630	105.4669	0	0	54740	-0.4751	-4.2679	0.5842	NaN	0.6209	24.5304
424320	2019	12	30	364	108.2035	152.6802	0.0051	-0.0099	32461	1.2913	6.3075	0.2453	NaN	0.6524	9.4769
424321	2019	12	30	364	147.4731	137.3698	-1.4340	1.3228	97270	-3.9410	2.2662	1.0000	1.4543	1.8017	538.9984
424322	2019	12	30	364	213.4726	213.3748	-0.8195	31.2830	74637	0.1414	12.2723	0.9323	0.5430	0.7570	110.0075
424323	2019	12	30	364	191.6005	180.4456	-0.2918	-10.6455	85220	-2.6176	-5.4127	1.0000	1.4505	1.6984	737.5270

You will find herein a description of the different variables included in the data table [units in bracket]. The data table is self-sufficient; however, if you are interested in looking into more details into the source data for the different variables, we are including links to the original datasets. Also note that the coordinate and the velocity components refer to the EASE-grid, which is defined after the data description.

year : **year**

month: **month**

day: **day in the month**

doy : **day-of-year**

- day of year ranging from 1 to 364

x_EASE : **x cartesian coordinate**

- x-position of the daily drift estimate. This is a cartesian coordinate on the EASE-grid.

y_EASE : **y cartesian coordinate**

- y-position of the daily drift estimate. This is a cartesian coordinate on the EASE-grid.

u_buoy : u-component of the sea ice velocity [cm/s]

- u-component of the ice velocity from the sea ice drifting buoy. It is oriented in the x-direction of the EASE-grid.
- The buoy data is maintained by the International Arctic Buoy Programme

v_buoy : v-component of the sea ice velocity [cm/s]

- v-component of the ice velocity from the sea ice drifting buoy. It is oriented in the y-direction of the EASE-grid (downward positive).

id_buoy : numerical buoy identifier

- Identify individual buoys. This is of interest for tracking the displacement of a single buoy, or sorting data. However, **the buoy identifier should not be used as an input for the model**.

u_ERA5 : u-component of the wind [m/s]

- u-component of the 10 m wind speed, obtained from the ERA5 reanalysis dataset, oriented in the x-direction of the EASE-grid.
- In short, a reanalysis is a model that reproduces the historical state of the atmosphere, land, and ocean, by being run with a focus on the assimilation of observations. The ECMWF-ERA5 is a state of the art reanalysis.

v_ERA5 : v-component of the wind [m/s]

- v-component of the 10 m wind speed, obtained from the ERA5 reanalysis dataset, oriented in the y-direction of the EASE-grid (downward positive)

sic_CDR : sea ice concentration [%]

- Sea ice concentration describes the fraction of a grid-cell covered with sea ice. A value of 0 indicates open water and a value of 1 is 100% ice cover.
- The sea ice concentration data comes from the National Snow and Ice Data Center Climate Data Record.

h_CS2SMOS : sea ice thickness [m]

- Observational sea ice thickness. The Cryosat-2 / SMOS data comes from a merged sea ice thickness dataset that combines observation from satellite altimetry (Cryosat-2) and passive microwave remote-sensing. Available since 2010, October through April.
- **This field is not available continuously.** Therefore it might be of limited use as an input, but one could conceive using it as an element of comparison for the ice thickness from the reanalysis (or even training data for the thickness from the reanalysis!)

h_piomas : sea ice thickness [m]

- Reanalysis sea ice thickness. From the Pan-Arctic Ice and Ocean Assimilation System (PIOMAS), a state of the art ice volume reanalysis that provides ice thickness information.
- The advantage of the PIOMAS ice thickness is that it's continuously available in space and time, making it a very relevant input for the model.

d2c: distance to coast [km]

- Distance from the buoy coordinate to the nearest coastline. The accuracy is +/- 25 km given the resolution of the grid.
- Coastal effects can significantly affect sea ice drift. The interaction of sea ice with coastlines is challenging to represent in models, and it is unclear whether it can be well captured in a ML setup, but it is well worth the try!

Now that we have covered the different variables, at the risk of being repetitive:

Your objective is to reproduce the u_buoy and v_buoy components of sea ice drift velocity, based on the u_ERA5 and v_ERA5 components of the wind as input parameters, plus other useful input parameters.

For the purpose of this challenge, we have split the full data record into a training dataset (80%) and a testing dataset (20%). You will be provided with two data tables. **DRIFT_DATA_TRAIN.csv** is your training dataset, with 339,478 entries; and **DRIFT_DATA_TEST.csv**, with 84,870 entries, from which we have hidden the u_buoy, v_buoy, and id_buoy parameters, is the data you will use to produce your final estimates of the u and v components of ice motion that will be evaluated.

Note on licencing:

- ***We ask you kindly not to distribute the data outside of this challenge. The data is presently being submitted to a journal as part of a scientific publication!***

Reference grid

We use a polar Equal-Area Scalable Earth grid (EASE-grid) as a referential for x-y cartesian coordinates and the u- and v- components of velocity. This Arctic grid has dimensions of 361 x 361, and a nominal resolution of 25km (note that the buoy data is provided with a sub-grid cell precision). This type of grid is characterized by grid cells of equal area and roughly the same width and height, which is very convenient for working with vector quantities in polar regions. The figure below shows the domain and the orientation of the u and v vector components in the EASE-grid referential.

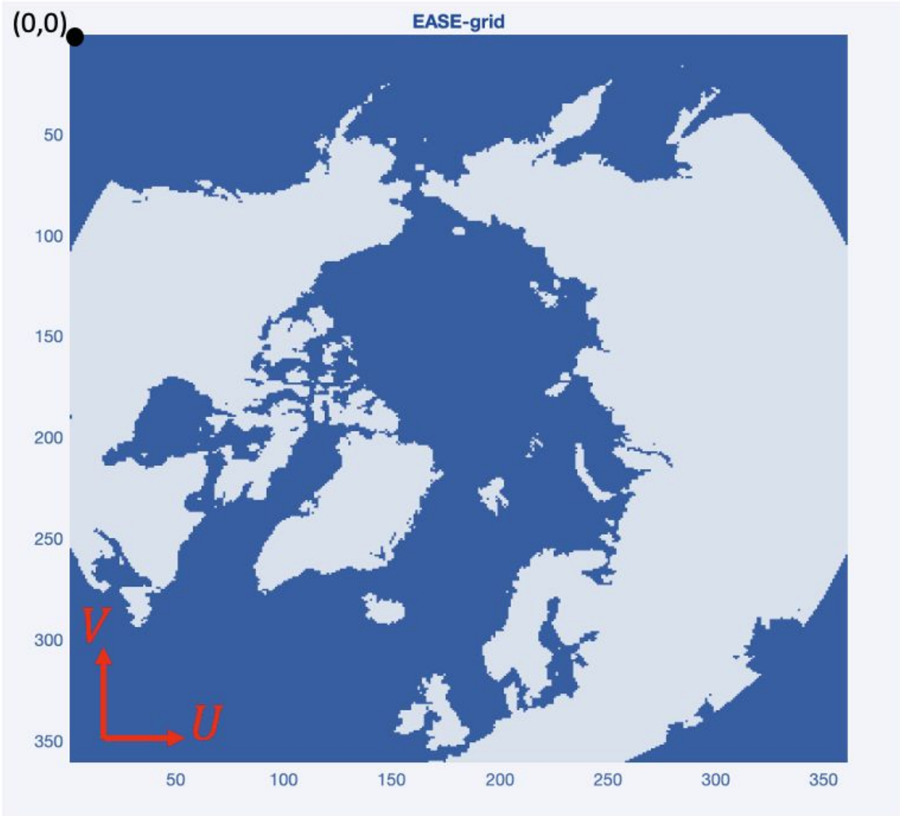
Note that you are not required to provide any type of gridded output for realizing this challenge. Anyhow, if someone is interested in working with the geographical coordinates (e.g.: for mapping), we also provide the latitudes and longitudes corresponding to the cartesian coordinates of the grid:

latitude_EASE.csv

- A 361 x 361 grid containing the polar EASE-grid latitudes values.

longitude_EASE.csv

- A 361 x 361 grid containing the polar EASE-grid longitudes values



Contact

For more information on this challenge, please do not hesitate to contact me:

- **Email**: charles.brunette@mail.mcgill.ca
- **Slack**: #challenge-sea-ice