LPHYS2268 Project
Seasonal prediction of summer Arctic sea ice extent

F. Massonnet

April 6, 2023

1 Scope of the project and practical aspects

Over the recent years, the Arctic region has attracted considerable mediatic and scientific attention. In particular, the rapid decline of sea ice has profoundly reshaped the Arctic landscape and has opened multiple opportunities, from resource extraction to shipping. As a result, the topic of seasonal Arctic sea ice prediction has become a major scientific research question since about 10 years.

1.1 Scientific goals

In this project, you will develop your own end-to-end forecasting system for September sea ice extent, based on the statistical model of your choice that is trained on sea ice extent data from observations. You will generate retrospective predictions of September sea ice extent until 2022, evaluate these predictions, and produce a new prediction for September 2023. You will also perform a short literature review to explain what are the physical mechanisms at the source of seasonal Arctic sea ice predictability.

1.2 Practical aspects

- The project is done by groups of 3 or 4 students. The members of the group can work on separate questions, but everyone should be able to understand the whole project, and answer questions on any part of the project.
- The main output of the project must be a visual medium of information that can be shared outside the course: a video, an infographic, a digital poster, a webpage. The medium should be directed toward a student who has the same level as yours but who has not followed the course. Check out existing resources[1] for inspiration, and be creative!
- An intermediate meeting with the teacher will take place one month before the deadline to check that everything is on track. Other meetings can be arranged earlier, on demand from the students.

[1] https://twitter.com/FMassonnet/status/1405871140900376584,https://twitter.com/micheltsamados/status/1334608223639244803, https://twitter.com/micheltsamados/status/1332029213055070210

- The project will count for 15 pts, as announced during the first course. **This will be your main achievement during the course, so spend your time and energy accordingly!**
- You are encouraged to have finished the technical work of the project (addressing the questions below) by the 1st of May. This will leave you enough time to start working on the visual medium before final submission.
- You are encouraged (but not obliged to) publish your scripts on public repositories like Github.
- The text below provides a set of guiding questions to run you through the scientific goal of the project – predicting September 2023 Arctic sea ice extent. These questions are there to help you structuring your visual medium, hence it's an advice to answer them all first before moving to the production of the visual.
- The visual must be delivered by Tuesday, 6th of June 2023. The language must be English.

2 Why seasonal Arctic sea ice prediction?

As explained in the first section, the topic of sea ice prediction has garnered special interest since a couple of years. As a preamble to your work, you should explain why there is a need to predict the evolution of sea ice several months in advance. This information can generally be found in the Introduction sections of scientific articles, such as Blockley and Peterson (2018) and references therein. Google Scholar is also a key resource to find articles on a given topic.

> **Question 1.** As an introduction to your work, provide a general motivation for the interest of seasonal Arctic sea ice prediction, highlighting why the need for reliable predictions is key for a range of stakeholders.

3 Mechanisms of sea ice predictability

Sea ice is a very thin layer of several meters at most, and yet is thought to be predictable well beyond weather time scales. It is therefore important to understand what are the key physical mechanisms that underpin its seasonal predictability. There are several good reviews of sea ice predictability. The ones of Guemas et al. (2014) and Chevallier et al. (2019) are excellent starting points.

> **Question 2.** Explain why scientists think there is potential to predict the September sea ice cover, and in particular its extent, already a few months earlier (in May, for instance). Describe the associated physical mechanisms of predictability.

4 Loading and visualizing sea ice extent data

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facilities (OSI SAF) publishes updated records of the monthly-mean sea ice extent in the Arctic. The data can be retrieved from the following URL: `https://osisaf-hl.met.no/v2p1-sea-ice-index`, see section "Data".

For information, sea ice extent is defined as the sum of all ocean grid cell areas in which sea ice concentration exceeds a pre-defined threshold of 15%.

> **Question 3.** Download the data and produce time series of September sea ice extent. Verify that your graph matches published time series, for example from Serreze and Meier (2019).

5 Estimating the trend line

As seen from the time series, the decline of September Arctic sea ice extent is sharp. One way to describe this decline is to compute the equation for the trend line. The equation for this line is computed by performing ordinary least square regression through the cloud of points. The slope of the trend line a is estimated as:

$$a = \frac{\sum_i \tilde{x}_i \tilde{y}_i}{\sum_i \tilde{x}_i \tilde{x}_i}$$

Here, $\tilde{x}_i$ is the year from 1979 to 2022 ($x_1 = 1979$, $\ldots$, $x_{44} = 2022$) and y_i is the September sea ice extent for the corresponding year. The tilde~denotes the anomalies, that is, the value for that year minus the 1979-2022 mean.

The *intercept* b of the trend line is then found by taking advantage of the fact that the trend line passes through the center of mass of the data:

$$b = \overline{y} - a\overline{x}$$

> **Question 4.** Compute the equation for the trend line in the form
>
> $$Y = b + aX$$
>
> What is the September 2023 ($X = 2023$) forecast based on simple extrapolation of the trend?

6 Event definition

The September sea ice extent in itself is an abstract number, so it is convenient to place it in comparison to pre-defined values. To this end, we need to define an "event". Here are three proposed events:

1. September Arctic sea ice extent will be below the trend line

2. September Arctic sea ice extent will be less than 4.5 million km^2
3. September Arctic sea ice extent will be less than previous year

> **Question 5.** For the remainder of the project, choose yourself one of the three above events. From the observed record (1979-2022), identify the year(s) for which the even happened, and report the corresponding observed event frequencies.
> *Two important remarks:*
>
> 1. The estimated observed frequencies should rely only on past data, not on future data. For example, when determining from observations if the September 2000 sea ice extent is below the trend line (event 1), the trend line to be considered is that based on the 1979-1999 data, not 1979-2022.
> 2. For events 1 and 3, determining if the event happens in observations requires two and one year of past data, respectively, to estimate the trend and the value of previous year. The event occurrence analysis should therefore only start in 1981.

7 Statistical forecasting system

You are now going to produce retrospective forecasts of September sea ice extent, based on data up to May each year (included). You are free to choose whatever method looks best to you.

One popular, recommended method is the "anomaly persistence forecast": the September forecast of year x_i ($x_i \geq 1981$) is equal to the September 1979-$(x_i - 1)$ average plus the May anomaly for year x_i relative to the May 1979-$(x_i - 1)$ average:

$$y^{\text{forecast}}_{\text{Sept},x_i} = \overline{y^{\text{observed}}_{\text{Sept},1979\to x_{i-1}}} + \left(y^{\text{observed}}_{\text{May},x_i} - \overline{y^{\text{observed}}_{\text{May},1979\to x_{i-1}}}\right) \tag{1}$$

(here, y denotes sea ice extent).

Eq. 1 makes it possible to produce single-value forecasts, that is, one number for each year. However, as we have seen in the course, a forecast should always be expressed as a probability density function (pdf), to reflect uncertainty in the model, in the initial conditions or in the forcing. Thus, your predictions from each year computed in the previous section cannot come as single numbers as suggested from Eq. 1.

In the case of an anomaly persistence forecast, you could assume that the forecast pdf for year x_i will be a gaussian random variable $\mathcal{N}\left(\mu^{\text{forecast}}_{x_i}, \sigma^{2\ \text{forecast}}_{x_i}\right)$. The mean will be the value calculated above:

$$\mu^{\text{forecast}}_{x_i} = \overline{y^{\text{observed}}_{\text{Sept},1979\to x_{i-1}}} + \left(y^{\text{observed}}_{\text{May},x_i} - \overline{y^{\text{observed}}_{\text{May},1979\to x_{i-1}}}\right) \tag{2}$$

To evaluate the variance, one must identify where uncertainty could come from in our model (Eq. 1). If we assume that the observations are free of measurement errors, then uncertainty will come from the fact that the two time averages (the first and third terms on the right hand side in Eq. 2) are subject to sampling variability. We recall that the variance of the sample mean is equal to the variance of the sample divided by the number of samples. If we further assume that the errors for the May mean and the September means are uncorrelated, we could estimate the variance of our forecast for year x_i as:

$$\sigma_{x_i}^{2\ \text{forecast}} = \frac{1}{n_i}\left(\mathsf{Var}\left[y^{\text{observed}}_{\text{Sept},1979\to x_{i-1}}\right] + \mathsf{Var}\left[y^{\text{observed}}_{\text{May},1979\to x_{i-1}}\right]\right) \tag{3}$$

where n_i is the number of years considered ($n_i = x_i - 1979$).

> **Question 6.** Produce a set of retrospective forecasts of September sea ice extent for each year between 1981 and 2022, based on the method of your choice. Make sure that the forecasts are produced in an "operational" context, that is, without knowledge of data after May of the year for which a forecast is produced (May can be included in the data). For example, when re-forecasting September 2000, your prediction can only use data until May 2000 included. To summarize your findings, produce a graph with the forecasts and their uncertainty; for example, the mean value and a 2 standard deviation confidence interval.

8 Retrospective probabilistic forecast of the event

Now that you have a full characterization of forecast uncertainty for each year, you can report the probability that the event chosen in Section 6 happens for that year. As a reminder, if the variable Y is distributed as $\mathcal{N}(\mu;\sigma^2)$, then the probability that Y is below a certain threshold y is

$$\begin{aligned} \text{Prob}(Y \leq y) &= \text{Prob}\left(\frac{Y-\mu}{\sigma} \leq \frac{y-\mu}{\sigma}\right) & (4)\\ &= F_Y\left(\frac{y-\mu}{\sigma}\right) & (5)\end{aligned}$$

where F_Y is the cumulative distribution function (cdf) of the $\mathcal{N}(0,1)$ distribution. The function F_Y does not exist in analytical form but all softwares can estimate it. For example, if the argument of F_Y is 1.96, then

- in Python: `from scipy.stats import norm; norm.cdf(1.96)`
- in R: `pnorm(1.96)`
- in Matlab: `normcdf(1.96)`

> **Question 7.** Evaluate the probability (a number between 0 and 1) that the even chosen happens according to your forecast, for each year between 1981 and 2022. Plot your results, and add the observed frequencies for each year calculated in Section 6

9 Verification of retrospective forecasts

For each year, you can verify whether or not (1 or 0) the chosen event did indeed occur. This is how you computed the observed frequency of the event in Section 6. So, for each year between 1981 and 2022, you have now a probability that the event occurs p_i (between 0 and 1), and a verifying observations o_i for the event (0 or 1).

Question 8. Compute the Brier Score of your forecasts over 1981-2022, namely,

$$BS = \frac{1}{n}\sum_{i}^{n}(p_i - o_i)^2$$

Also, compute the reference Brier Score BS_{ref} for the "cheap" forecast wherein p_i is replaced by the observed frequencies of occurrence of the event calculated in Section 6.
Finally, derive the Brier Skill Score,

$$BSS = \frac{BS - BS_{\text{ref}}}{0 - BS_{\text{ref}}}$$

Discuss your results, and the relevance of using a statistical model instead of using observed frequencies.

10 Forecast for 2023

The May 2023 value will be published on the data website on the first or second of June.

Question 9. Provide a forecast pdf for September 2023 sea ice extent, and report the probability that the chosen even will occur.

Congratulations! You have just made your first forecast. You are encouraged to submit it to the Sea Ice Prediction Network, a community effort to collect all forecasts of September sea ice extent. More on information on `https://www.arcus.org/sipn`.

11 References

All References are available on Moodle (Project / Project references)

E. W. Blockley and K. A. Peterson. Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness. The Cryosphere, 12(11): 3419-3438, Oct. 2018. ISSN 1994-0424. doi: 10.5194/tc-12-3419-2018.

M. Chevallier, F. Massonnet, H. Goessling, V. Guemas, and T. Jung. The Role of Sea Ice in Sub-seasonal Predictability. In Sub-Seasonal to Seasonal Prediction, pages 201-221. Elsevier, 2019. ISBN 978-0-12-811714-9. doi: 10.1016/B978-0-12-811714-9.00010-3.

V. Guemas, E. Blanchard-Wrigglesworth, M. Chevallier, J. J. Day, M. Déqué, F. J. Doblas-Reyes, N. S. Fuckar, A. Germe, E. Hawkins, S. Keeley, T. Koenigk, D. Salas y Melia, and S. Tietsche. A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales: Arctic Sea-Ice Predictability and Prediction. Quarterly Journal of the Royal Meteorological Society, 142(695): 546-561, 2014. ISSN 00359009. doi: 10.1002/qj.2401.

Serreze, M. C., Meier, W. N. (2019). The Arctic's sea ice cover: Trends, variability, predictability, and comparisons to the Antarctic. Annals of the New York Academy of Sciences, 1436(1), 36–53. https://doi.org/10.1111/nyas.13856