审稿意见 2

First of all, thank you for dedicating your valuable time and effort to conduct a thorough review.

This paper employed multiple machine learning methods to reconstruct the salinity in the West Arctic Ocean based on easily obtained atmosphere reanalysis data and satellite-based sea ice concentration and motion. This topic is interesting and crucial, and this method can expand the spatial-temporal coverage and improve the accuracy of estimation than existing productions as the author claimed.

However, due to so much confusion about the method and verification, I have to treat this article with caution, and I believe that it now is far away from an adequate article, especially for publishing on ESSD. Here are some major comments.

1. For the method, there are 2 major questions.
    1. I cannot understand why the author using EN4 reanalysis data as a target to train his/her model. The EN4 including some uncertainties and errors cannot provide the true word's salinity. Moreover, EN4 can cover the whole area and time span this work focuses on. I think the author can use this dataset directly or simply interpolate it.

Firstly, EN4 data has been demonstrated that it can effectively analyze the salinity of the Arctic Ocean (*Hall et al.,2022*). The machine learning method is employed to train the EN4 data, with extrapolation serving three purposes on the salinity product grid: firstly, this article utilizes a machine learning approach to generate CTD data as well as EN4 data extrapolation results to generate salinity products by merging; secondly, it aims to demonstrate the selection process of input variables that can be utilized in machine learning and training; thirdly, the availability of CTD data in the east Siberian sea area is limited, therefore, the extrapolation results of salinity from EN4 data play a crucial role in generating the product.

    2. I think the author wants to highlight the machine learnings, but I don't know which role they played in the improvement of the accuracy of salinity reconstruction. Cooperation with traditional methods (e.g., optimal interpolation) is lacking in this work. Moreover, the author compared their production with EN4, which is used as a target set to train the method, and the author said their results are better, which is contrary to general knowledge. Maybe the decrease in errors comes from the application of machine learning, but the more possible reason is just the merging of CTD data.

The salinity field of the Western Arctic Ocean is taken as an example to construct a novel data mining method for polar sea areas in our paper, utilizing multiple machine learning methods that integrate multiple data sources and incorporate physical processes. We mainly used machine learning methods to train. We compared our production with EN4, and proved our results are better. Our training objective encompasses not only EN4 but also CTD, and the extrapolation results are merged to generate the salinity product. Therefore, salinity products in some aspects outperform EN4 without causing conflicts. As you said, the decrease in errors comes from the application of machine learning, and possible the merging of CTD data. The main focus of our research is to address the new method of data reconstruction in the polar regions.

2. For the verification, I feel very surprised about the 0 of RMSE for the results from the KNN method. It means that this method can perfectly reproduce your target salinity. The only possibility I can think of is that the author uses a train set the calculate RMSE instead of a verification or test set. This makes it completely impossible to evaluate the salinity reconstructed by the machine learning in this work.

First of all, we thank the reviewer for your professional questions. The datasets used for prediction from each year were randomized, as depicted in Figure 4 of the text. Subsequently, 90% of the data was selected for training purposes, constituting the training pool, while the remaining 10% was allocated for testing purposes, forming the testing pool. Indeed, we used a train set the calculate RMSE. So we added the verification results of testing pool (the table below), to get the same results. RMSE for the results from the KNN method is still about 0. Therefore, the skills in extrapolation was guaranteed.

	RF		KNN		LightGBM		Catboost		Multilinear Regression		Neural Network	
	CTD	EN4	CTD	EN4	CTD	EN4	CTD	EN4	CTD		CTD	EN4
2003	0.42	0.05	0.34	0.00	0.34	0.00	0.32	0.00	1.07	1.00	1.00	2.01
2004	0.21	0.05	0.14	0.00	0.16	0.01	0.16	0.00	0.92	1.03	0.73	2.59
2005	0.14	0.11	0.12	0.00	0.13	0.01	0.13	0.00	0.65	1.07	0.40	2.03
2006	0.10	0.07	0.11	0.00	0.12	0.00	0.12	0.00	0.69	1.00	0.45	3.11
2007	0.21	0.05	0.14	0.00	0.20	0.01	0.20	0.00	1.11	1.21	0.91	3.54
2008	0.21	0.05	0.23	0.00	0.23	0.00	0.23	0.00	1.20	1.12	0.81	1.66
2009	0.11	0.05	0.14	0.00	0.11	0.00	0.12	0.00	0.81	1.01	0.51	2.03
2010	0.21	0.11	0.23	0.00	0.22	0.01	0.22	0.00	0.94	1.20	0.59	2.94
2011	0.18	0.07	0.18	0.00	0.19	0.00	0.17	0.00	0.99	1.03	0.66	1.74
2012	0.25	0.11	0.26	0.00	0.24	0.00	0.25	0.00	0.70	1.02	0.61	2.92
2013	0.18	0.08	0.17	0.00	0.18	0.00	0.18	0.00	0.72	0.99	0.44	2.49
2014	0.16	0.06	0.16	0.00	0.15	0.01	0.15	0.00	0.43	1.08	0.36	2.51
2015	0.23	0.07	0.17	0.00	0.27	0.00	0.21	0.00	0.71	0.97	0.66	1.96
2016	0.08	0.05	0.02	0.00	0.03	0.00	0.02	0.00	0.39	1.09	0.46	2.12
2017	0.11	0.05	0.03	0.00	0.04	0.00	0.02	0.00	0.53	1.00	0.46	2.87
2018	0.12	0.06	0.13	0.00	0.15	0.00	0.13	0.00	0.58	0.99	0.33	1.76
2019	0.33	0.05	0.25	0.00	0.50	0.00	0.22	0.00	0.82	0.96	0.77	1.92
2020	1.19	0.05	0.83	0.00	1.19	0.00	1.19	0.00	1.67	1.07	1.66	3.81
2021	0.69	0.07	0.69	0.00	0.69	0.00	0.69	0.00	1.54	0.88	1.80	2.57
2022	0.20	0.06	0.21	0.00	0.20	0.01	0.21	0.00	0.78	1.06	0.60	2.90
Average	0.27	0.07	0.23	0.00	0.27	0.00	0.25	0.00	0.86	1.04	0.71	2.47
mean RMSE	**0.17**		**0.11**		**0.14**		**0.12**		**0.95**		**1.59**	

There are also many minor comments. What needs the author to pay more attention to is that many places in the MS are against the conventions of academic writing, which makes it hard to read.

Line 17: write the full name of "CTD" due to its first use in the paper. Similar issues exist in many other places, please check.

The expression of gratitude is extended to you for the reminder. the full name of "CTD" is conductivity, temperature, and depth.
The full text was carefully reviewed once again, and the following revise were taken. ITP (line 60, Ice-Tethered Profiler); CCGS (line139, Canadian Coast Guard Shipboard); NSIDC (line164, National Snow and Ice Data Center); MAE (line184, mean absolute error); FWC (line 293, Freshwater Content); WOA18 (World Ocean Atlas 2018); SLP (Sea Level Pressure).

Line 63: double full stops.

The redundant full stop has been removed, and a comprehensive inspection has been conducted.

Line 86: what's your dataset's temporal resolution?

Considering the spatial distribution of CTD data, the temporal resolution of salinity product is annual, disregarding seasonal fluctuations and focusing on interannual as well as lower frequency variations.

Line 96: I think it is better to put the section about why you focus on the Western Arctic Ocean in the Introduction than here.

We have put the reason we focus on the Western Arctic Ocean (Section 2.1 Study area original manuscript) in the introduction

Line 191: Fig. 2 is confused. It should be replotted. a), I cannot find the “Classify” and “statistic analysis” (case matters) in the text. b), Similarly, the “Physical process” and “Nearest Neighbors” (case matters) also cannot be found and they look like input variables very much. c), only 4 variables are used in the data-selecting step, which is far less than the data introduced in 2.2. And in the text, you don't seem to be doing anything with these 4 variables, while the CTD data was cleaned. This figure gives readers the opposite impression. You should make it clear to readers which are the variables used to train or build the dataset, and where are the algorithm. d), where is the WOA18 used when you create the dataset? Mark the figure or delete it.

The expression of our idea in Figure 2 may not be ideal. Thank you for reminding us. We have redrawn Figure 2. We mainly enriched the data selecting part and the machine learning part, and made some modifications to the data merging and post-calibrating.

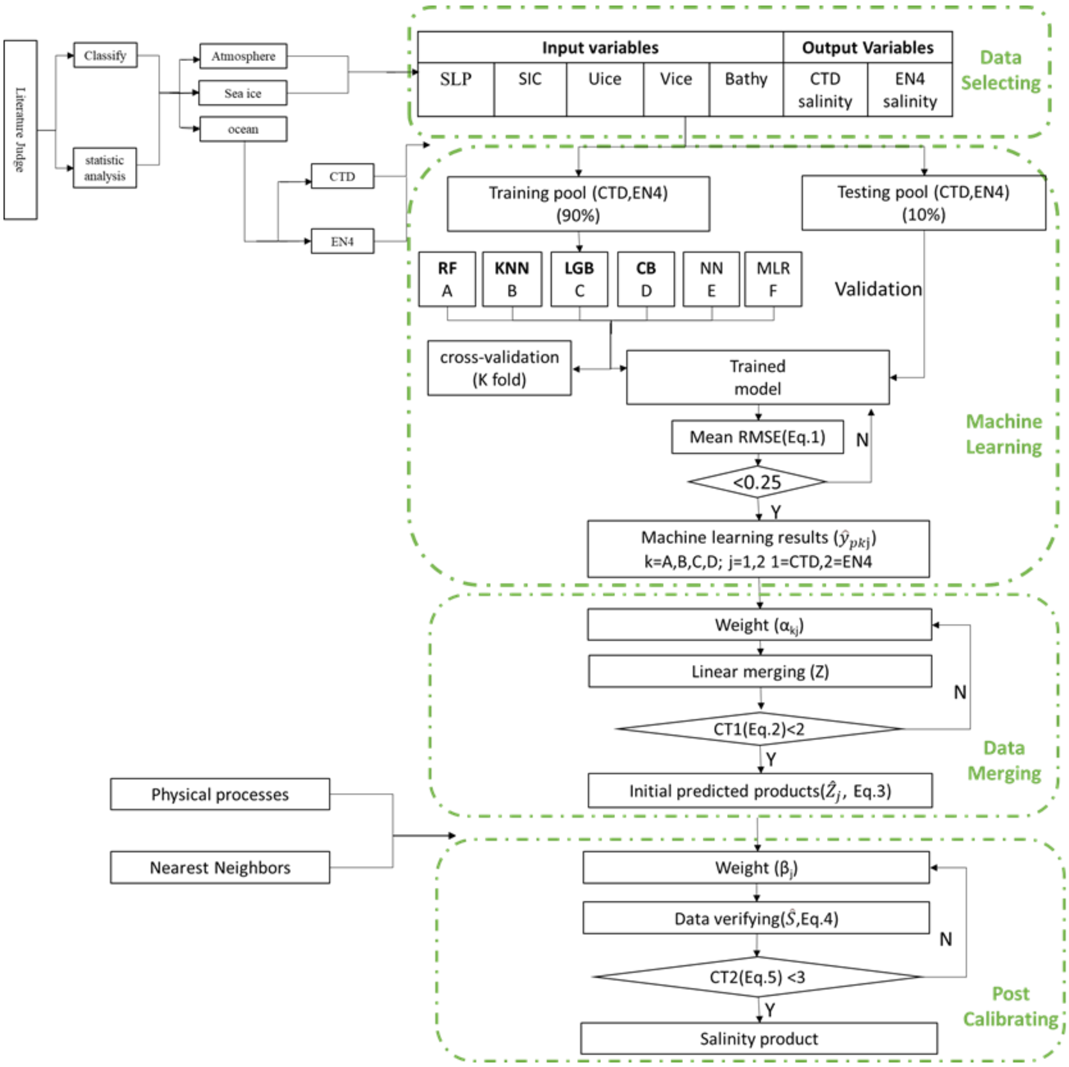


Figure 2 Procedure for improving the salinity field in the Western Arctic Ocean through a data mining-based machine learning method.

(a) “classify” and “statistic analysis” are used to choose the final input variables. The process of “classify” and “statistic analysis” involves identifying the influential factors that impact the salinity of the arctic ocean, as discussed in previous literature. In terms of thermodynamics, the melting and freezing processes of sea ice have a significant effect on salinity. Additionally, from a dynamic perspective, both ice-ocean stress and air-ocean stress contribute to salinity redistribution within the ocean. Therefore, we propose incorporating variables such as sea level pressure(SLP), sea ice concentration(SIC), and sea ice drift speed(Uice,Vice).

(b) The “Physical process” and “Nearest Neighbors” are that “The salinity product is generated through the post-calibrating, when there are CTD measured data around the grid point, the salinity value of the point is formed by merging the EN4 prediction results and the CTD prediction results according to weights; otherwise, the salinity value of the point is taken as the EN4 prediction result (line 258-261,original manuscript)”. The post-calibrating included the concepts of "Physical process" and "Nearest Neighbors".

(c) The data introduced in 2.2 include CTD salinity data (from ITP, LSSL, WOD18, UDASH), EN4 salinity data, SLP (from ERA5), SIC (from NSIDC), Uice (from NSIDC), Vice (from NSIDC). The input variables include SLP, SIC, Uice, Vice. The output variables include Salinity (EN4 and CTD). So the data introduced in the 2.2 used to train(90%) and test(10%). Indeed, we don't do anything with these 4 variables, while the CTD data was cleaned. Because these 4 input variables provided by different institution are after Quality control. We think there are reliable. The time scale of the variables needs to be adjusted to align with that of CTD factors in machine learning. There is also ORAS5 data for the validation of salinity product in section2. WOA18 data is used to validate the generated salinity products as well as ORAS5 data.

Line 122 In this section, I think two things were done: 1, introduce the data used in this work; 2, data clean (selecting). It would be better to divide them into two paragraphs. A lot of data are introduced here, but it is confusing which one is used to train, which one is used to create the dataset, and which one is used to evaluate. Reorganize them according to their purpose.

Yes, the problem of overfitting has been considered in the process of machine learning in this paper. The datasets used for prediction from each year were randomized Subsequently, 90% of the data was selected for training purposes, constituting the training pool, while the remaining 10% was allocated for testing purposes, forming the testing pool. Your question is very meaningful, and we have added a summary at the end of 2.2

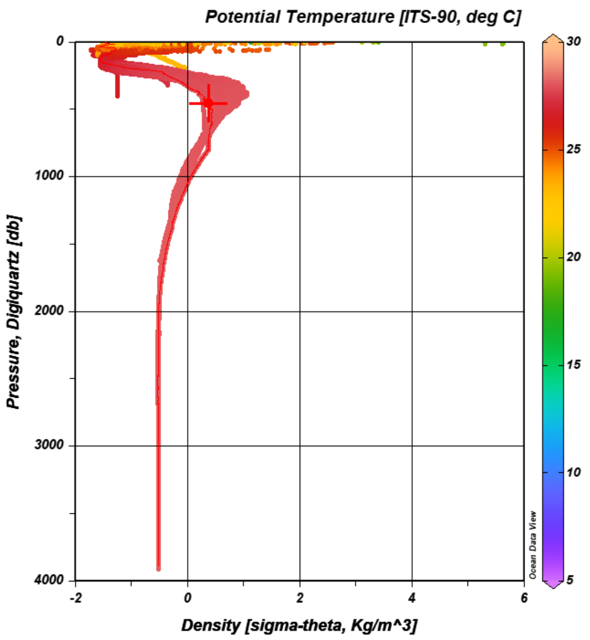
Line 127 “the data with flags 0 and 1 based on the quality control provided by the data itself”: what does it mean?

WOD18 provides quality-controlled data, all data in the WOD are associated with as much metadata as possible, and every ocean data value has a quality control flag associated with it. Flag 0 means accepted value, Flag 1 means range outlier (outside of broad range check).

0: accepted value
1: range outlier (outside of broad range check)
2: failed inversion check
3: failed gradient check
4: observed level bullseye flag and zero gradient check
5: combined gradient and inversion checks
6: failed range and inversion checks
7: failed range and gradient checks
8: failed range and questionable data checks
9: failed range and combined gradient and inversion checks

Line 141: Why the data in 2004 are ignored?

The CTD data of LSSL collected during the 2004 expedition was not utilized. The potential temperature (shadow) and density values of the CTD data in 2004 are evidently anomalous (refer to the figure below), suggesting a potential issue with the data storage process, thus rendering them unsuitable for use.



Line 143: Enlarge the size of years. Also, in Fig 9 and 10.

Fig9 and Fig10 have been revised (see the mannuscript_re2).

Line 151: Does the quality of the reconstructed data vary across seasons?

The question you raised is indeed insightful. Considering the aforementioned aspects, it is important to note that our salinity data only provides annual results due to the requirements of data volume. Consequently, seasonal variations in the reconstructed data cannot be accounted for at present. This limitation has prompted us to contemplate its development as a future endeavor.

Line 154: Fig 3 instead of “Fig 2”.

The manuscript's writing errors have been brought to my attention, and I sincerely appreciate your contribution. The manuscript has been revised, and we check the whole paper.

Line 157: Why did you choose the EN4? You should tell readers more about it, for example, if the WOD data is assimilated.

The absence of CTD data in the East Siberian Sea region, as depicted in Figure 1. There is a significant disparity in the fundamental physical environment compared to that of the Canadian Basin. Consequently, it is inappropriate to extrapolate the results of CTD training to encompass the entire West Arctic Ocean. Specifically, substantial errors are expected when applying these results to the East Siberian Sea region. To mitigate this issue, we introduce EN4 data for salinity estimation where measurements are lacking. In such cases, careful consideration must be given to account for variations in physical processes across different sea areas.

Line 174: which Neural Network did you use?

In this paper, Neural Network we used is Multi-layer Perceptron (MLP) from scikit-learn

Line 195 “the prediction using EN4 data is better than using CTD data”: just means that the results based on EN4 are closer to its target data. There may be some errors in EN4. Don’t say that it is better.

“The prediction using EN4 data is better than using CTD data” means that in the verification results of the test pool (10%), it is shown that the RMSE of EN4 is smaller than that of CTD (Table 1). Thanks for your reminding, I have changed the words in the article.

Line 199: target values instead of “True values”.

Thanks for your reminding again, I have changed the words in line 199 as well as figure 4a and 4b.

Line 204: It is so surprising that the RMSE is 0. I guess you may just show the results of the training set. You should explain what causes these 0.

We added the verification results of testing pool (the table below), to get the same results. RMSE of the EN4 testing pool results from the KNN method is still about 0. The reason behind this has been identified. The KNN method utilizes the target values of the K nearest training samples to make predictions on the regression values of the predicted samples. In other words, the regression value is estimated based on sample similarity. The regression value of the sample to be tested can be measured in different ways: (1) by employing the ordinary arithmetic average algorithm for K nearest neighbor target values (weights='uniform'), and (2) by utilizing a weighted average approach for K nearest neighbor target values that takes into account the difference in distance(weights='distance'). We choose the second way in this paper. The EN4 testing pool consists of 10% EN4 gridded data. When the second method is employed for training and verification, the predicted value obtained through training corresponds to the data point (distance=0). Consequently, the RMSE is 0.

Line 207: “prediction” or reconstruction? Prediction is a good usage for this method, but it looks like you didn’t do it in this work.

The prediction is made based on the regression relationship obtained from the training pool (90%). The predicted value is then compared with the target value in the testing pool(10%), and the RMSE (Table 1) is calculated. Four machine learning methods are selected based on their respective root mean square errors. It has been used in the article. Reconstruction refers to the extrapolation of the relationship obtained from the training pool that has been validated by the testing pool.

Line 216 “In the same year, some machine learning predictions are good while others are poor.”: it is better to give me a statistical, quantitative result rather than such a description.

In the same year, some machine learning predictions are good while others are poor (Table 1). The quantitative result support this is lines 212-216 in the original manuscript. We add the quantitative result in lines 217-219 in original maniscript. For example, in 2020, the mean RMSE of RF (0.32) and KNN (CTD (0.45)) were poor, but the predictions of LGB (0.14) and CB (0.14) were good.

Line 226: In this work, you always show me examples, but a better way is to give readers a statistical result. I think it is not so hard.

Thanks for your reminding. Figure 5 shows the spatial distribution of the error between the predicted value and the target value obtained by different machine learning methods. Here we supplement the results of the statistical analysis (MAE) of the performance of four selected machine learning methods in the testing pool (10%)。

Table. MAE between the predicted salinity and target salinity of four selected machine learning methods in the testing pool

	RF		KNN		LightGBM		Catboost	
	CTD	EN4	CTD	EN4	CTD	EN4	CTD	EN4
2003	0.22	0.03	0.12	0.00	0.14	0.00	0.13	0.00
2004	0.12	0.03	0.07	0.00	0.07	0.00	0.07	0.00
2005	0.05	0.04	0.03	0.00	0.04	0.00	0.04	0.00
2006	0.05	0.03	0.05	0.00	0.06	0.00	0.06	0.00
2007	0.07	0.03	0.06	0.00	0.06	0.00	0.06	0.00
2008	0.09	0.03	0.07	0.00	0.09	0.00	0.10	0.00
2009	0.05	0.03	0.05	0.00	0.06	0.00	0.05	0.00
2010	0.09	0.05	0.08	0.00	0.09	0.00	0.09	0.00
2011	0.09	0.04	0.07	0.00	0.09	0.00	0.08	0.00
2012	0.10	0.05	0.09	0.00	0.10	0.00	0.10	0.00
2013	0.07	0.03	0.06	0.00	0.06	0.00	0.06	0.00
2014	0.05	0.03	0.04	0.00	0.04	0.00	0.04	0.00
2015	0.09	0.03	0.04	0.00	0.08	0.00	0.06	0.00
2016	0.04	0.03	0.00	0.00	0.01	0.00	0.01	0.00
2017	0.06	0.03	0.01	0.00	0.02	0.00	0.01	0.00
2018	0.04	0.02	0.03	0.00	0.04	0.00	0.03	0.00
2019	0.11	0.03	0.07	0.00	0.11	0.00	0.07	0.00
2020	0.20	0.03	0.14	0.00	0.18	0.00	0.17	0.00
2021	0.11	0.04	0.06	0.00	0.08	0.00	0.08	0.00
2022	0.07	0.02	0.06	0.00	0.07	0.00	0.07	0.00
Average	0.09	0.03	0.06	0.00	0.08	0.00	0.07	0.00

Line 245: Maybe here you can tell readers how to merge and post-calibrate data first, and then discuss how to calculate uncertainty. Give me more details.

Firstly, we selected four of the six machine learning methods according to the verification results (Table 1). The EN4 and CTD data were trained and validated using four machine learning methods, respectively, and extrapolated to obtain the extrapolation results of the product grid (EN4_RF, CTD_RF, EN4_KNN, CTD_KNN …). The reconstruction results of EN4 and CTD were obtained by performing a weighted average based on the Mean Absolute Error (MAE) of the four machine learning methods. This is our data merging step.

The uncertainty of the data in this article includes three parts. The first part is the uncertainty generated during the machine learning process. The four machine learning methods we selected independently calculated the uncertainty (rMAE) during the validation of the test pool. After averaging these four uncertainties, we found that the surface salinity prediction generated by CTD had an uncertainty of 0.24%, while the one generated by EN4 had an uncertainty of 0.02%. The second part is the uncertainty generated during the data merging process. In the mergence of reconstruction results extracted by four machine learning methods, we computed initial predicted values based on EN4 and CTD, respectively. The second component of uncertainty is represented by the computational average of rMAE between the initial predicted values and the extracted reconstruction results obtained from the four machine learning methods. The third part is the uncertainty generated during the post-calibrate process. Reconstruction results based on CTD and EN4 data will be merged, taking into account the different physical processes in various sea areas and the availability of at least three CTD data points in the product grid. In doing so, we calculated the mean uncertainty (rMAE) between the final product and the reconstruction results based on CTD and EN4, respectively.

Line 253: how to get the weights "a"? Also, the weights beta needs more description.

Weight "a" contains the result of the arithmetic average (both 0.25) and the result of the weighted average (Based on EN4 and CTD respectively):

1.Restructed values of four machine learning methods (A1, A2, A3, A4)

2. weights "a"

α=1-3*(A1/(A1+A2+A3+A4) ;β=1-3*(A2/(A1+A2+A3+A4)

γ =1-3*(A3/(A1+A2+A3+A4) ;λ=1-3*(A4/(A1+A2+A3+A4)

3. initial Salinity production =α*A1+β*A2+ γ *A3+ λ*A4

The initial predicted values are first calculated based on EN4 and CTD using equal weights, and then the rMAE is computed between these predictions and the extrapolated results obtained from different machine learning methods.

Weight beta is similar to weight a. Weight beta contains the result of the arithmetic average (both 0.25) and the result of the weighted average:

Line 258: where is the 1st-3rdpost-calibrating?

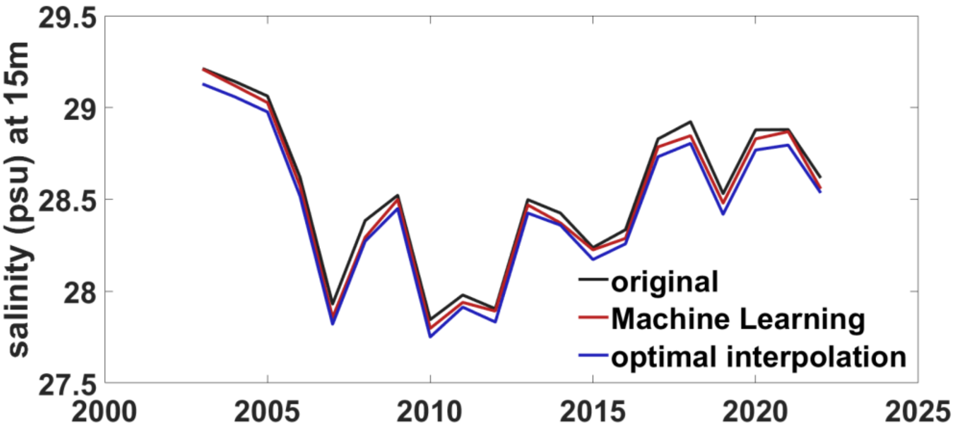
It's a typo. It means the fourth step is post-calibrating, corresponding to the procedure step. It has been revised in the manuscript.

Line 259 "when there are CTD measured data around the grid point": the "around" needs to be quantified.

The "around" has been quantified in the manuscript. "When there are at least three CTD measurements available in the vicinity of the grid point".

Line 261: I'm a bit confused about three things. a), why are you using machine learning to reconstruct salinity based on the EN4. I assume you're trying to improve its resolution, but you don't show the advance of the method. You need to compare it with traditional methods like optimal interpolation. b), you only used the reconstructed data based on CTD which is close to the in-situ observations. I think this may be due to the increasing error away from the buoy, but this also requires evidence. You need to show the reader at what distance the salinity error based on CTD reconstruction is greater than EN4. c), if you use the optimal interpolation instead of machine learning to reconstruct salinity based on the CTD data, how much error is going to increase?

(a) We used machine learning to reconstruct salinity based on the EN4. The results of the proposed comparison between optimal interpolation and machine learning methods are appended below. Compared with traditional methods, the machine learning method has the advantage of more accurate reconstruction of salinity. The MAE of EN4 salinity reconstruction by machine learning method (0.04psu) is significantly smaller than that by traditional method (0.09psu).



The mean salinity at 15m in the BG region based on EN4 (defined as BG box, Proshutinsky et al.,2009:170 °W-130 °W,70.5 °N-80.5 °N)

(b) In the process of generating the final salinity product, we utilized not only the data reconstructed based on CTD measurements but also the data reconstructed using EN4. Additionally, we incorporated a distance criterion and employed a weighted average method to derive the ultimate salinity product when the grid point was in proximity to at least three CTD profiles. The steps for the mergence of reconstructed products based on CTD and EN4 into products are as follows:

Case 1: There are at least three CTD profiles around grid points:

Calculate the absolute MAE_{CTD} (MAE_{EN4}) of the difference between the mean salinity of the measured points around the grid point and the CTD (EN4) salinity predicted by the grid point

$MAE_{CTD} < MAE_{EN4}$

s_product=salinity_pred_CTD

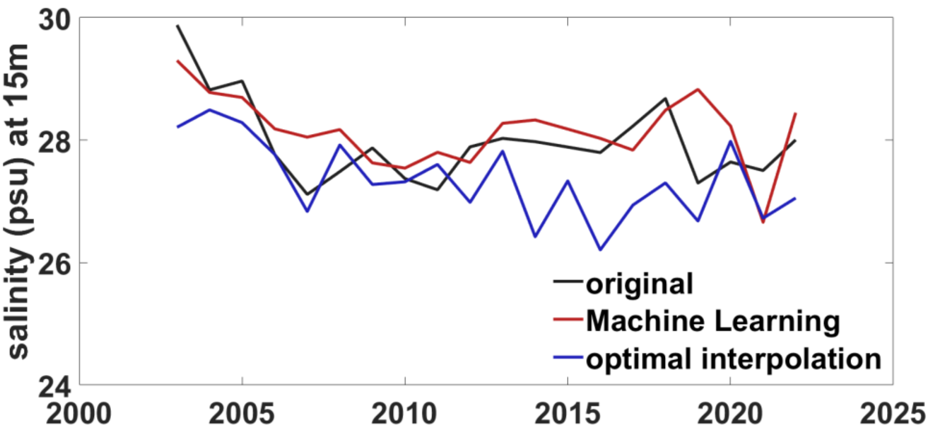
$MAE_{CTD} > MAE_{EN4}$

s_product=(1-MAE_{CTD} /MAE_{CTD} +MAE_{EN4})*salinity_pred_CTD +(1-MAE_{EN4} /MAE_{CTD} +MAE_{EN4})*salinity_pred_EN4

Case 2, there is no CTD profiles near the grid point

s_product=salinity_pred_EN4

(c) Thanks to your suggestions, we incorporated optimal interpolation for reconstructing the CTD salinity, achieving an MAE of 0.73 psu. This value is slightly larger than the absolute error of 0.52 psu obtained from the machine learning.



The mean salinity at 15m in the BG region based on CTD (defined as BG box, Proshutinsky et al.,2009:170 °W-130 °W,70.5 °N-80.5 °N)

Line 270: I don't suggest using percentiles to record uncertainty. The salinity of the ocean is always high, so the proportions of errors are always low and the percentile statements are misleading. For the FWC, you calculate the proportion of errors, but there are differences between the reconstructed salinity and 34.8 psu. These two similar variables are quite different, potentially confusing readers.

The proposition is commendable, and percentages can be misleading when conveying uncertainty. The uncertainty is transformed into the true salinity bias. Words in the manuscript (line270-278; line295) and Figure 6 has been revised by us.

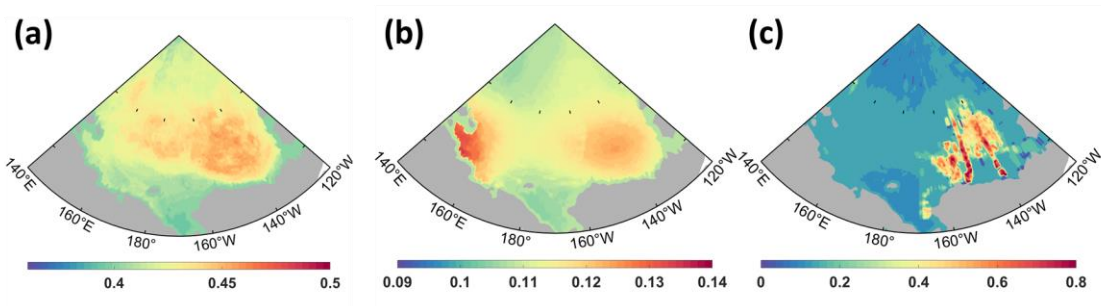
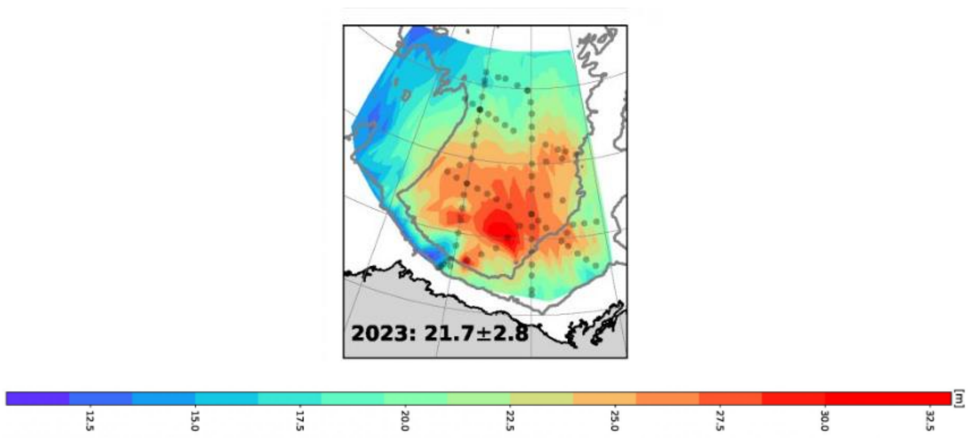


Figure 6. Spacial pattern of sea surface salinity uncertainty (psu) during the data merging (a, CTD; b, EN4) and post-calibrating.

Line 288: Please do a brief introduction about BGEP.

The Beaufort Gyre is one of the Arctic Ocean's primary circulation features, storing and transporting freshwater, sea ice, and heat across the Canadian Basin, and is a critical part of the regional and global climate system.Since 2003, during a time of unprecedented change, the Beaufort Gyre Exploration Project has provided continuous monitoring of conditions in the region and established a strong foundation that is vital for understanding the current state and future trajectories of the Arctic Ocean environment.

BGEP freshwater content is the Estimation of liquid freshwater content (FWC) of the Beaufort Gyre region (BGR) are computed following Proshutinsky et al. (2009) using CTD, XCTD, and UCTD profiles collected each year.The FWC is calculated using optimal interpolation on a 50-km square grid between 70˚N and 80˚N, and 130˚W - 170˚W, and where water depths exceed 300 m.



FWC in the BGR based on hydrographic measurements in 2003, The black dots indicate the locations of observational sites.(Figure from https://www2.whoi.edu/site/beaufortgyre/data/freshwater-content-gridded-data/)

Line 334: why the freshwater decreasing after 2011 supports the recent major freshening event for 2012 to 2016.

The recent major freshening event from 2012 to 2016 in North Atlantic (Holliday et al.,2020). The freshwater from the Beaufort Gyre into the North Atlantic make the North Atlantic fresh (Zhang et al.,2021). So we think that “the freshwater decreasing after 2011 supports the recent major freshening event for 2012 to 2016”.

Line 384: I think it will be clearer to write as “sea level pressure from ERA5 and sea ice concentration and motion from NSIDC”.

The comments you provided in the manuscript have been revised, and we are grateful for your input.

Line 385: where did you use the ETOPO1?

The etopo1 bathymetric data was utilized in both the machine learning training process and the screening of the CTD salinity profile. In machine learning training process we use etopo1 as the input profiles, which is determined by the physical process. Western arctic ocean salinity is influenced by circulation, bathymetric play a great role in circulation (Wind-Driven Flow Along f / H Contours, see Figure8 in Timmermans and Marshall,2020). While we screened of the CTD profiles, The validity of the data is confirmed for profiles that encompass over 90% of the whole water column in shelf and slope areas.

Line 392: you should not add any new results in the Summary and please cite Fig. 11 in this paragraph.

Ok, Figure 11 and its conclusion have been deleted in the summary and inputted section machine-learning. we cited Fig. 11 in the summary. “The importance of various factors varies when predicting salinity in both EN4 and CTD datasets. Interestingly, both datasets consistently highlight sea level pressure as the primary influential factor for surface salinity prediction, while sea ice concentration emerges as the main determinant when forecasting salinity at a depth of approximately 200m (corresponding to the halocline base) (Fig. 11). The reconstruction of salinity data in the western Arctic Ocean holds significant scientific value. However, further research is needed to incorporate other variables that influence salinity, such as the Pacific Ocean inflow the and the ventilation process in the Chukchi Sea, into the salinity data reconstruction process.”

Line 405: I cannot understand the meaning of “trend” and also don’t know why you discuss the relationship between salinity and freshwater here. If you want to show me something, a figure about it is necessary.

The term "trend" refers to the process of reconstructing salinity based on CTD measurements, specifically focusing on water depth. The significance of bathymetric at 15m is approximately 0.06, while it increases to around 0.17 at a depth of 200m. However, the reconstruction data obtained from EN4 yields contrasting results, indicating that the importance of bathymetric at 15m is approximately 0.2 but decreases to about 0.08 at a depth of 200m.

Line 411: at the beginning of the paper, you said that the greatest advantage of your dataset is that the salinity in recent years is included, but here you say that the greatest advantage is your result is more accurate. I agree that both of them are important and you are able to do them, but changing your big problem in one paper is improper.

The question you pose holds significant importance. Our primary focus lying in enhancing the reliability of salinity data reconstructed through machine learning methodologies, thereby aiming to advance this concept. The salinity in recent years is included is neccesarry.