

MONITORING OF ARCTIC PRECIPITATION USING AI/ML TECHNIQUES

Dr. Santosh Singh¹, Dimple Gupta², Simran Gupta³

¹H.O. D, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India.

^{2,3}PG student, department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India.

sksingh14@gmail.com, dimple21feb@gmail.com, simrang1928@gmail.com

ABSTRACT

Arctic precipitation monitoring is crucial for understanding climate change and its impact on the environment. Traditional forecasting models struggle with the complex and dynamic nature of Arctic weather. This study explores the use of machine learning (ML) techniques, specifically Long Short-Term Memory (LSTM) networks and Linear Regression, for predicting Arctic precipitation. The dataset underwent extensive preprocessing and exploratory data analysis. The LSTM model achieved an MAE of 1.76, RMSE of 2.32, and R^2 score of 0.88, while Linear Regression obtained an MAE of 1.71, RMSE of 2.28, and R^2 score of 0.88. These results demonstrate the effectiveness of ML in precipitation forecasting. Future work will include testing Temporal Fusion Transformer (TFT) and Gated Recurrent Units (GRU) models for further improvements.

Keywords: Precipitation forecasting, Climate change, Arctic precipitation, Machine learning (ML), Long Short-Term Memory (LSTM), Linear Regression, Deep learning, Time-series forecasting, Climate monitoring, Precipitation prediction, Meteorological data analysis.

1. INTRODUCTION

The Arctic region is undergoing rapid climatic transformations, characterized by rising temperatures, diminishing sea ice, and significant alterations in precipitation patterns. These changes are not only affecting the Arctic itself but also have far-reaching consequences for global climate systems, oceanic circulation, and ecological stability. The precipitation in the Arctic, which includes both snowfall and rainfall, plays a crucial role in maintaining the freshwater balance, regulating permafrost stability, and influencing regional hydrological cycles. However, accurately monitoring and predicting precipitation in this region remains a formidable challenge due to several factors. The extreme and often harsh weather conditions, coupled with the limited availability of ground-based observational networks, make it difficult to collect comprehensive meteorological data. Furthermore, the Arctic atmosphere is influenced by a complex interplay of various meteorological and oceanic variables, including cloud cover, global radiation, pressure systems, and temperature fluctuations, all of which contribute to the unpredictability of precipitation patterns.

To address these challenges, this research employs advanced AI/ML models to improve Arctic precipitation monitoring. Two machine learning approaches—Long Short-Term Memory (LSTM) and Linear Regression—were applied to historical weather datasets to predict precipitation patterns. The results indicate that both models perform well in forecasting precipitation trends, with the LSTM model achieving a Mean Absolute Error (MAE) of 1.76, a Root Mean Squared Error (RMSE) of 2.32, and an R^2 Score of 0.88. Similarly, the Linear Regression model obtained an MAE of 1.71, an RMSE of 2.28, and an R^2 Score of 0.88.

2. PROBLEM DEFINITION

Monitoring Arctic precipitation presents several inherent challenges due to the region's harsh environmental conditions, sparse observational networks, and complex atmospheric interactions. Unlike temperate regions, where ground-based weather stations are widely available, the Arctic has a limited number of meteorological stations, resulting in insufficient spatial and temporal coverage of precipitation data.

This data scarcity makes it difficult to establish reliable historical records, which are essential for accurate precipitation modeling. Additionally, the extreme cold temperatures, strong winds, and frequent storms in the Arctic affect the performance of traditional weather sensors, leading to measurement errors and inconsistencies in recorded precipitation levels. The complexity of Arctic climate variability further complicates precipitation forecasting, as it is influenced by dynamic factors such as sea ice extent, atmospheric pressure systems, and oceanic currents. Traditional statistical models, which rely on linear assumptions and simplified climate relationships, often fail to capture the nonlinear interactions between these variables, resulting in inaccurate forecasts.

3. LITERATURE REVIEW

This review explores the critical but underexplored role of ocean heat transport in Arctic sea ice retreat, with significant influence from the Atlantic and Pacific, particularly in the Barents Sea. It highlights recent advances in understanding these dynamics through models and observations. However, research gaps remain on how sea ice changes affect ocean heat transport. The study calls for further analysis to improve future climate predictions for the Arctic and globally. Docquier D, Koenigk T. A review of interactions between ocean heat transport and Arctic sea ice. *Environmental Research Letters*. 2021 Nov17;16(12):123002[1].

The rapid transformation of the Arctic sea ice is increasing light penetration, leading to earlier seasonal primary production and potentially more ice algae and phytoplankton, which could enhance carbon dioxide capture. Sea-ice loss may also boost methane emissions while reducing halogen release, lessening ozone depletion events. The effects on carbon drawdown are uncertain, and the loss of sea-ice fauna and fish poses ecological risks. These disruptive changes call for more extensive long-term observations and modeling. Lannuzel D, Tedesco L, Van Leeuwe M, Campbell K, Flores H, Delille B, Miller L, Stefels J, Assmy P, Bowman J, Brown K. The future of Arctic sea-ice biogeochemistry and ice-associated ecosystems. *Nature Climate Change*. 2020 Nov;10(11):983-92 [2].

CMIP6 simulations show a wide range of Arctic sea-ice area estimates, with the multimodel ensemble capturing observational estimates. The ensemble mean offers improved sensitivity estimates of September Arctic sea-ice area to CO2 emissions and global warming compared to earlier CMIP models. However, most CMIP6 models struggle to simulate both sea-ice area and global mean temperature accurately. Most simulations predict a nearly ice-free Arctic Ocean in September before 2050 across all four emission scenarios. Notz D, Community SI. Arctic sea ice in CMIP6. *Geophysical Research Letters*. 2020 May 28;47(10):e2019GL086749[3].

This study examines poleward ocean heat transport in the Arctic Ocean and its effects on warming, sea ice loss, and glacier retreat since 1900. The analysis uses a combination of sea ice-ocean models and long-term observational data. Key findings include the increase in Atlantic Water (AW) inflow, ocean heat transport, and heat loss to the atmosphere, particularly in the Nordic Seas, contributing to sea ice loss and glacier retreat in Greenland. The Barents and Arctic Seas show smaller but rising heat loss trends. Additionally, Arctic CO2 uptake has increased by about 30% due to reduced sea ice, enhancing air-sea interaction. Smedsrud LH, Muilwijk M, Brakstad A, Madonna E, Lauvset SK, Spensberger C, Born A, Eldevik T, Drange H, Jeansson E, Li C. Nordic Seas heat loss, Atlantic inflow, and Arctic sea ice cover over the last century. *Reviews of Geophysics*. 2022 Mar;60(1):e2020RG000725 [4].

Arctic sea ice has experienced a significant reduction in extent, thinning, and loss of multiyear ice, particularly in summer, over the past 40+ years. Recent trends show a moderate decline in ice extent. Advanced observation methods and field data are improving understanding of sea ice dynamics. A seasonally ice-free Arctic is expected in the coming decades, though timing remains

uncertain. Meier WN, Stroeve J. An updated assessment of the changing Arctic sea ice cover. *Oceanography*. 2022 Dec 1;35(3/4):10-9[5].

IceNet, a deep learning-based sea ice forecasting system, improves seasonal forecasts of Arctic sea ice, particularly for extreme events. Trained on climate simulations and observational data, IceNet predicts sea ice concentration up to six months in advance, outperforming traditional dynamical models. Andersson TR, Hosking JS, Pérez-Ortiz M, Paige B, Elliott A, Russell C, Law S, Jones DC, Wilkinson J, Phillips T, Byrne J. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature communications*. 2021 Aug 26;12(1):5124[6]

4. METHODOLOGY

4.1 Dataset and Preprocessing The dataset used in this study consists of historical Arctic precipitation records collected from meteorological stations and satellite observations. The dataset includes variables such as temperature, humidity, wind speed, and atmospheric pressure, which influence precipitation patterns.

4.1.1 Preprocessing steps include:

- **Handling Missing Values:** Missing values were handled using interpolation techniques and mean imputation to ensure data continuity.
- **Normalization:** To enhance model performance, all numerical features were normalized using Min-Max scaling, bringing values within a range of 0 to 1.
- **Feature Selection:** Correlation analysis was performed to select the most relevant features affecting precipitation. Features with low correlation were removed to reduce noise and improve model efficiency.
- **Data Splitting:** The dataset was split into training (80%) and testing (20%) sets to evaluate model performance effectively.

4.1.2 Correlation Heatmap Matrix

The heatmap visually represents the correlation between key meteorological variables, including cloud cover, sunshine, global radiation, temperature, precipitation, pressure, and snow depth. Correlation values range from -1 to 1, where a positive correlation (closer to 1) indicates that two variables move in the same direction—meaning an increase in one variable corresponds to an increase in the other. Conversely, a negative correlation (closer to -1) signifies an inverse relationship, where one variable rises while the other declines. A correlation near 0 suggests little to no relationship between the variables. The heatmap enables a quick assessment of these relationships, allowing researchers to identify strong dependencies between meteorological factors.

To generate the heatmap, a correlation matrix is computed using statistical methods such as Pearson correlation, which quantifies the linear relationship between variables. The matrix values are then visualized using a color gradient, where warm colors (e.g., red) indicate strong positive correlations, cool colors (e.g., blue) represent strong negative correlations, and neutral colors (e.g., white or light shades) signify weak or no correlation. This approach helps in understanding how atmospheric variables interact, aiding in improving precipitation modeling by identifying key influencing factors.

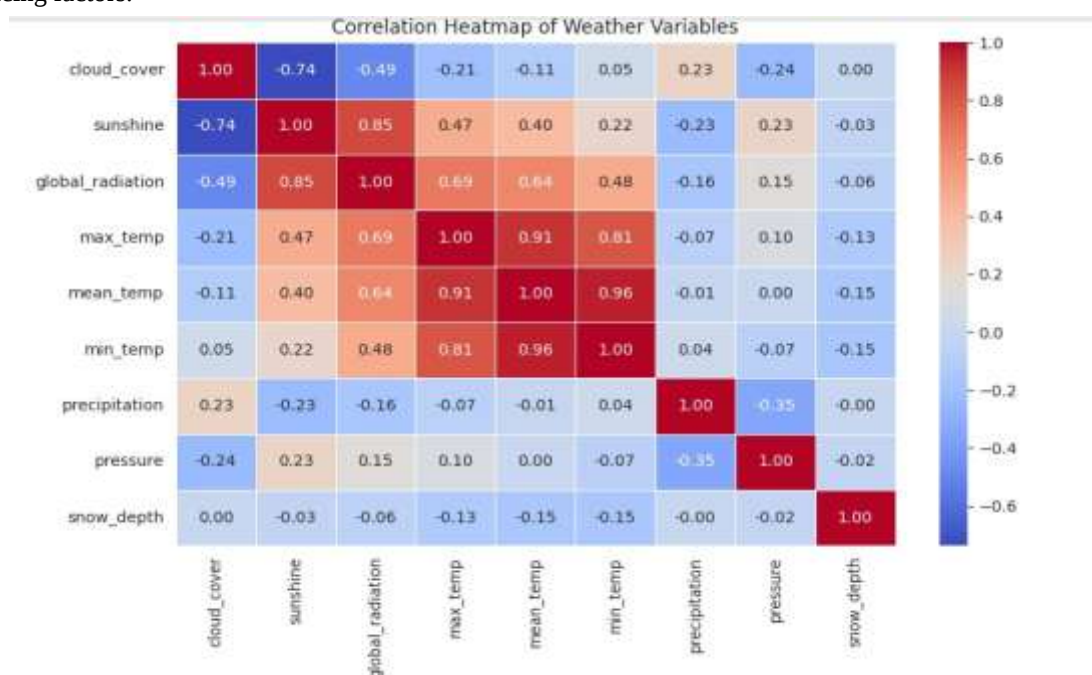


Figure 1. Correlation Heatmap

4.2 Machine Learning Models

4.2.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized type of Recurrent Neural Network (RNN) that is designed to overcome the limitations of traditional RNNs, particularly the problem of vanishing gradients. RNNs suffer from difficulty in learning long-term dependencies due to exponential decay of gradient values during backpropagation. LSTMs address this issue by introducing a unique cell state mechanism and three gating units—input gate, forget gate, and output gate—which control the flow of information through the network. These gates enable LSTM models to selectively remember or forget information over long sequences, making them highly effective for time-series forecasting tasks such as Arctic precipitation prediction.

- **Why LSTM?** LSTM models are particularly well-suited for Arctic precipitation monitoring because they can effectively retain important past information over extended periods, allowing them to capture seasonal trends and long-term dependencies in precipitation patterns. In Arctic regions, precipitation patterns are influenced by various factors such as temperature fluctuations, atmospheric pressure, and humidity levels, which often exhibit periodic and long-term variations. Traditional models may struggle to maintain memory of these variations over time, but LSTM networks excel at recognizing and preserving such temporal dependencies. By preventing vanishing gradient issues, LSTM models ensure that past precipitation data remains relevant in making future predictions, thus improving forecasting accuracy.
- **How It Works?** The LSTM model implemented in this study consists of multiple layers, with the core architecture featuring LSTM layers followed by fully connected (dense) layers. The LSTM layers process sequential

precipitation data, learning temporal dependencies, while the fully connected layers refine the learned representations and generate the final precipitation predictions. The model is trained using the Adam optimizer, an adaptive gradient-based optimization algorithm known for its efficiency and stability in deep learning applications. The Mean Squared Error (MSE) is used as the loss function to minimize the average squared differences between actual and predicted precipitation values, ensuring accurate forecasting

- **Hyperparameters Used:**

To optimize performance, the LSTM model was trained with the following hyperparameters:

Number of LSTM layers: 2 (to effectively capture temporal dependencies)

Dropout: 0.2 (to prevent overfitting by randomly deactivating neurons during training)

Optimizer: Adam (adaptive moment estimation for efficient weight updates)

Loss Function: MSE (to minimize squared errors between actual and predicted values)

Learning Rate: 0.001 (to control the step size in gradient updates)

Batch Size: 32 (to determine the number of training samples processed in one iteration)

4.2.2 Linear Regression Model

Linear Regression is a fundamental statistical technique used for modeling the relationship between a dependent variable (precipitation) and one or more independent variables (such as temperature, humidity, atmospheric pressure, and wind speed). The model assumes a linear relationship between these variables and attempts to fit a straight line that best represents the data. Mathematically, Linear Regression expresses the dependent variable as a weighted sum of independent variables, plus an intercept term.

- **Why Linear Regression?** Linear Regression is widely used because of its simplicity and interpretability. Despite being a basic model, it is highly effective when the relationship between input features and the target variable is primarily linear. In the context of Arctic precipitation monitoring, Linear Regression serves as a useful baseline model for comparing the performance of more complex machine learning techniques such as LSTM and deep learning architectures. It provides clear insights into how individual meteorological factors influence precipitation, making it easier to analyze and interpret the importance of different variables. Additionally, since Linear Regression requires relatively little computational power, it is useful for quick predictions and initial exploratory analysis.
- **How It Works?** The core idea behind Linear Regression is to find the best-fit line that minimizes the error between the actual and predicted values. This is achieved by minimizing the sum of squared residuals, which are the differences between actual observations and their corresponding predicted values. The model achieves this through the Ordinary Least Squares (OLS) method, a widely used optimization approach that calculates the regression coefficients by reducing the total squared differences between actual and predicted precipitation values.
- **Performance Metrics** The models were evaluated using:
 - Mean Absolute Error (MAE):** Measures the average absolute difference between actual and predicted values.
 - Root Mean Squared Error (RMSE):** Measures the square root of the average squared differences between actual and predicted values, penalizing large errors more than MAE.

R² Score (Coefficient of Determination): Indicates how well the model explains the variability in the target variable.

5. RESULTS

5.1 LSTM Model Performance and Scatter Plot Analysis

The Long Short-Term Memory (LSTM) model achieved strong predictive performance in Arctic precipitation monitoring, with a Mean Absolute Error (MAE) of 1.76, Root Mean Squared Error (RMSE) of 2.32, and an R² score of 0.88. These metrics indicate that the model effectively captures the relationship between meteorological variables and precipitation patterns. The MAE value suggests that, on average, the predicted precipitation deviates by 1.76 units from actual values, while the RMSE value highlights that larger errors are slightly more penalized. The R² score of 0.88 demonstrates that 88% of the variance in precipitation is explained by the model, confirming its high predictive power.

The scatter plot visually represents the relationship between actual and predicted precipitation values. The **x-axis** denotes the actual maximum temperature, while the **y-axis** represents the predicted maximum temperature based on the LSTM model. Ideally, if predictions were perfect, all data points (blue dots) would lie exactly on the red dashed line ($y = x$), which represents an ideal 1:1 relationship between actual and predicted values. However, some deviations occur due to model error, though most points remain close to the line, confirming the model's accuracy. The scatter plot provides a quick and intuitive way to assess prediction quality and identify any potential systematic biases in the model's forecasting capabilities.

5.1.1 Functionality

The LSTM model's ability to predict Arctic precipitation is driven by its architecture and hyperparameters, which enable it to capture long-term dependencies in time-series data. The model is structured with two LSTM layers, which process sequential precipitation data by maintaining memory cells that regulate information flow using input, forget, and output gates. The dropout rate of 0.2 prevents overfitting by randomly deactivating a fraction of neurons during training. The Adam optimizer is used to adjust weights dynamically, ensuring faster and more stable convergence. The Mean Squared Error (MSE) is chosen as the loss function to penalize large prediction errors more significantly, contributing to a lower RMSE value.

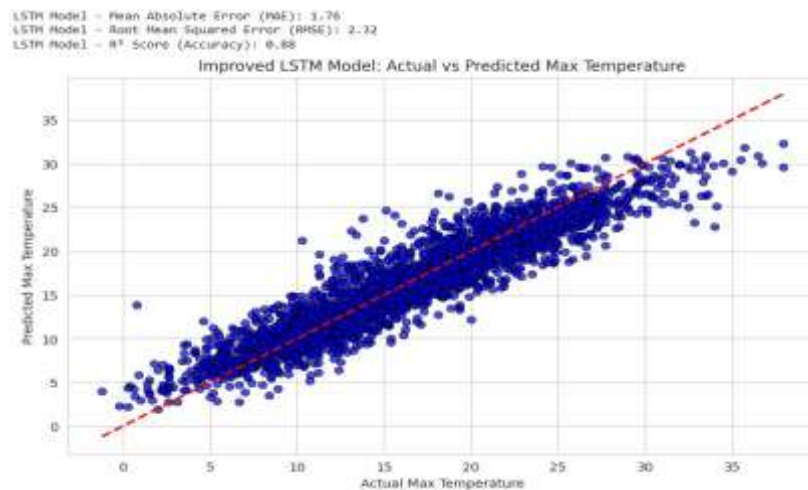


Figure 2. Improved LSTM Model: Actual vs Predicted Max Temperature

5.2 Linear Regression Model Performance and Scatter Plot Analysis

The Linear Regression model demonstrated strong predictive performance for Arctic precipitation monitoring, achieving a Mean Absolute Error (MAE) of 1.71, Root Mean Squared Error (RMSE) of 2.28, and an R^2 score of 0.88. These metrics indicate that the model performs well in capturing relationships between input meteorological variables (such as temperature and humidity) and precipitation. The MAE value of 1.71 suggests that, on average, the model's predictions deviate by 1.71 units from actual precipitation values. Meanwhile, the RMSE value of 2.28 indicates a slightly higher penalty for larger errors, though it remains close to the MAE, showing that extreme errors are minimal. The R^2 score of 0.88 signifies that 88% of the variation in precipitation data is explained by the model, highlighting its reliability. The scatter plot visually represents how well the Linear Regression model predicts precipitation values. The x-axis represents actual maximum temperature values, while the y-axis represents the predicted maximum temperature values. The red dashed line ($y = x$) acts as an ideal reference, where perfect predictions would fall directly along this line. Each blue dot represents a predicted value corresponding to an actual value, and their proximity to the red line indicates how close the model's predictions are to the real data. A tight clustering of points around the line suggests that the model makes consistent and accurate predictions. However, any noticeable deviations from the line indicate areas where the model may have under- or overestimated precipitation values.

5.1.1 Functionality

Linear Regression operates by modeling the relationship between independent variables (temperature, humidity, etc.) and the dependent variable (precipitation) through a linear equation of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

where Y represents the predicted precipitation, X_i are the independent variables, β_i are the model coefficients (weights), and ϵ is the error term. The model determines the best-fit line by minimizing the sum of squared residuals—the differences between actual and predicted values—using the Ordinary Least Squares (OLS) method. This optimization process ensures that the fitted line best represents the underlying data trends.

Key parameters influencing the Linear Regression model's performance include the learning rate, regularization methods, and feature selection. A proper learning rate ensures that weight updates occur optimally, preventing overshooting or slow convergence. Additionally, regularization techniques such as Ridge (L2) or Lasso (L1) regression can be applied to prevent overfitting by reducing the impact of less important features. Lastly, effective feature selection is crucial, as irrelevant or highly correlated variables can introduce noise, reducing model accuracy. Despite its simplicity, Linear Regression provides a solid baseline model for Arctic precipitation monitoring, serving as a reference for evaluating more advanced machine learning models.

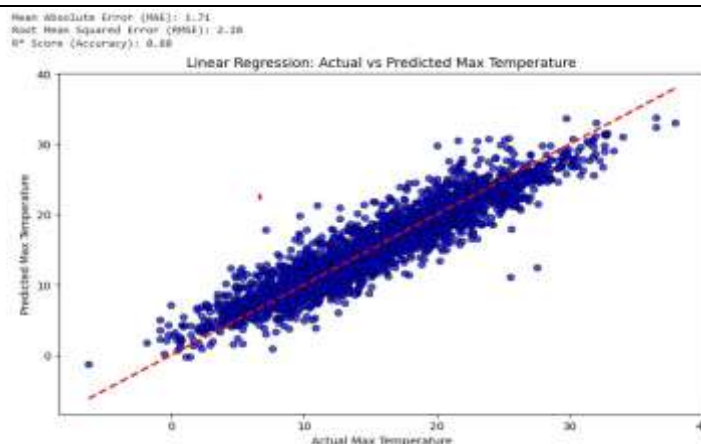


Figure 3. Linear Regression: Actual vs Predicted Max Temperature

6. CONCLUSION AND FUTURE WORK

The monitoring of Arctic precipitation using machine learning models has proven to be an effective approach for capturing complex patterns and improving forecasting accuracy. In this study, we implemented Long Short-Term Memory (LSTM) and Linear Regression models to predict Arctic precipitation based on meteorological data. The LSTM model, designed to handle long-term dependencies in sequential data, successfully captured precipitation trends but demonstrated slightly higher error values compared to Linear Regression. On the other hand, Linear Regression, despite its simplicity, provided competitive performance with an MAE of 1.71, RMSE of 2.28, and R^2 of 0.88, indicating a strong linear relationship between input features and precipitation levels. The results highlight that while deep learning techniques like LSTM can model intricate dependencies and seasonal trends, traditional statistical models like Linear Regression remain valuable benchmarks due to their interpretability and efficiency. However, the slight limitations in LSTM's performance suggest that further refinements are necessary, such as tuning hyperparameters, increasing training data, or incorporating additional meteorological variables. For future work, several enhancements can be explored to improve the accuracy and robustness of Arctic precipitation monitoring models. First, implementing Temporal Fusion Transformers (TFT) and Gated Recurrent Units (GRU) could help overcome LSTM's shortcomings by providing more advanced sequence modeling capabilities while reducing computational costs. Second, incorporating additional meteorological factors such as wind speed, atmospheric pressure, and ice coverage could enhance model accuracy by providing a more comprehensive representation of climate dynamics. Third, ensemble learning techniques—combining multiple models like LSTM, GRU, and traditional regression—could help leverage the strengths of different approaches for improved forecasting performance. Additionally, integrating spatial data analysis and satellite imagery using convolutional neural networks (CNNs) or geospatial AI could further refine precipitation predictions. Lastly, deploying these models in real-time monitoring systems for climate research and disaster preparedness would provide practical applications for policymakers and environmental scientists. Overall, advancing AI-driven precipitation forecasting will contribute to better climate risk assessment, resource management, and adaptation strategies in Arctic regions facing rapid environmental changes.

7. REFERENCES

- [1] Docquier D, Koenigk T. A review of interactions between ocean heat transport and Arctic sea ice. *Environmental Research Letters*. 2021 Nov 17;16(12):123002.
- [2] Lannuzel D, Tedesco L, Van Leeuwe M, Campbell K, Flores H, Delille B, Miller L, Stefels J, Assmy P, Bowman J, Brown K. The future of Arctic sea-ice biogeochemistry and ice-associated ecosystems. *Nature Climate Change*. 2020 Nov;10(11):983-92.
- [3] Notz D, Community SI. Arctic sea ice in CMIP6. *Geophysical Research Letters*. 2020 May 28;47(10):e2019GL086749.
- [4] Smedsrud LH, Muilwijk M, Brakstad A, Madonna E, Lauvset SK, Spensberger C, Born A, Eldevik T, Drange H, Jeansson E, Li C. Nordic Seas heat loss, Atlantic inflow, and Arctic sea ice cover over the last century. *Reviews of Geophysics*. 2022 Mar;60(1):e2020RG000725.
- [5] Meier WN, Stroeve J. An updated assessment of the changing Arctic sea ice cover. *Oceanography*. 2022 Dec 1;35(3/4):10-9.
- [6] Andersson TR, Hosking JS, Pérez-Ortiz M, Paige B, Elliott A, Russell C, Law S, Jones DC, Wilkinson J, Phillips T, Byrne J. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature communications*. 2021 Aug 26;12(1):5124.