General comments:

In summary, this research aims to develop a transformer-based U-net model to improve prediction skill of Arctic sea ice concentration at seasonal lead times. The primary novel contribution of this work is the use of a state-of-the-art deep learning architecture with a custom domain-specific loss function as well as including spring sea ice thickness as an input predictor to overcome the common spring predictability barrier (SPB) faced by previous numerical and DL-based models that decrease prediction skill of predictions made before May.

The authors choose to use a transformer based architecture to capture spatiotemporal patterns in the input data. The model inputs include multiple sea-ice related variables over 3-6 months and outputs the sea ice concentration of the next six months, specifically looking at predictions for June-Sept at 6 month lead times. The authors develop a novel loss function that combines both a standard DL loss function (mean squared error) and a domain-specific NIEE loss which accounts for spatial similarity in model predictions and ground truth. The model follows a standard DL training procedure and includes three years of unseen data to test their model performance. The authors evaluate their model using standard DL performance metrics along with Binary Accuracy (BACC) which accounts for accurate spatial distributions.

The authors succeed in improving prediction skill when compared to the benchmark Persistence model and dynamical ECMWF model. For Sept ice prediction specifically, the ECMWF model performs better at smaller lead times but the author's model shows marked performance improvement at a lead time of 3-6 months, thus overcoming the spring predictability barrier. The authors perform necessary ablation tests to investigate the role of sea ice thickness in their model. They show a model that includes sea ice thickness as input outperforms a model without sea ice thickness at seasonal lead times of 3-6 months. When testing their model on unseen data, the authors use an ensemble model approach by averaging the results of 20 separately trained models.The authors also test their model against a previous state of the art deep learning CNN based approach, IceNet. They show their model produces better ACC scores at longer lead times. Their model produced lower BACC scores but can capture more individualized/local or extreme characteristics in comparison to IceNet which tends to produce smoothed out results.

Overall, this manuscript provides a significant improvement to the modeling of sea ice concentration by providing a novel approach to increase prediction skill at seasonal lead times to overcome the SPB. The authors clearly state their motivations for the project and outline their novel contributions. The authors discuss previous modeling approaches in the field and clearly showcase their significant scientific results where their model outperforms previous approaches.The contributions of this paper is twofold, one in utilizing a novel transformer based approach that the authors state has not been previously used in this field and second in highlighting the impact of spring sea ice thickness on improving prediction accuracies, revealing potential scientific insights that need to be studied further. Below are a few suggestions to improve reproducibility and presentation quality.

Specific comments:

In line 52, the authors state that experiments show the reason for decrease in prediction skill before spring is due to ice motion and growth in the winter. It is unclear which experiments in the literature the authors are referring to. Including a citation here would help justify this statement.

In line 79, a citation for the IceNet model (Andersson et al 2021) is missing. Since working with the IceNet model is a significant portion of the manuscript, perhaps including more information about the IceNet model and how it differentiates from the author's transformer based model would provide more context to the reader in this section.

In lines 108-109, the authors state that multiple experiments were conducted to determine the length of the model input features. It is unclear what experiments or methods the authors tried. Was this determined by manual tuning using their cross-validation strategy or did the authors employ an automated grid-search type strategy. The authors can also include if domain knowledge or previous literature informed the choice of input lengths of the data.

In section 4.7, when comparing the transformer based model to the CNN-based IceNet model, the authors state they used identical training and testing settings to perform fair comparisons. It is unclear whether the authors used the same 20 trained ensemble model approach they had used for their transformer model for the IceNet model. If so, specifying whether they used an ensemble approach or singular-model approach for IceNet would clarify this for the reader. If the authors did not use a similar ensemble approach, the authors should justify this choice.

In line 269, it is unclear how the authors transformed the IceNet output to match the continuous scale (for e.g restructuring only the final output layer). Having this additional context would help with reproducibility of this experiment.

For Figures 2, 4, 6 and 8, it is slightly confusing to the reader at first how to interpret these results, specifically the Difference column. Perhaps including the caption that red signifies improvement in accuracy and blue signifies a decrease would aid in understanding especially because in Figures 3 and 5 the opposite color scheme is used (red = high error, blue - lower error).

Technical comments:

In line 20 of the abstract, 'SIE' is referenced without expanding the full word.

In line 24, there is a space missing before the in-text citation.

In Figure 2, the score percentages are hard to read and their text size could be increased.