Bayesian Unsupervised Machine Learning Approach to Segment Arctic Sea Ice Using SMOS Data

Christoph Herbert1,2, Adriano Camps1,2, Florian Wellmann3, and Mercedes Vall-llossera1,2

1CommSensLab, Universitat Politècnica de Catalunya (UPC) and Institut d’Estudis Espacials de Catalunya (IEEC/CTE-UPC), Barcelona, Spain, 2Barcelona Expert Center (BEC), Barcelona, Spain, 3Institute for Computational Geoscience and Reservoir Engineering, RWTH Aachen University, Aachen, Germany

Abstract Microwave radiometry at L-band is sensitive to sea ice thickness (SIT) up to ~ 60 cm. Current methods to infer SIT depend on ice-physical properties and data provided by the ESA's Soil Moisture and Ocean Salinity (SMOS) mission. However, retrieval accuracy is limited due to seasonally and regionally variable surface conditions during the formation and melting of sea ice. In this work, Arctic sea ice is segmented using a Bayesian unsupervised learning algorithm aiming to recognize spatial patterns by harnessing multi-incidence angle brightness temperature observations. The approach considers both statistical characteristics and spatial correlations of the observations. The temporal stability and separability of classes are analyzed to distinguish ambiguous from well-determined regions. Model uncertainty is quantified from class membership probabilities using information entropy. The presented approach opens up a new scope to improve current SIT retrieval algorithms, and can be particularly beneficial to investigate merged satellite products.

Plain Language Summary Remote sensing techniques are commonly used to provide maps of sea ice thickness (SIT). Methods to obtain these maps are based on the sea ice composition and on the signal measured by satellite. Sea ice Composition is spatially complex and changes during its formation and melting. Currently used data from observations of ESA's Soil Moisture and Ocean Salinity (SMOS) mission depend on several sea ice parameters, which hinders good estimation of almost any specific sea ice parameter. In this work, a new method to combine the information contained in SMOS brightness temperature data is investigated, with the aim to divide the Arctic region into a number of smaller areas - so called classes. Useful information about sea ice is contained in the spatial and statistical distribution of SMOS data, which are collected at different incidence angles. The relationship between the observations and the statistical properties of the obtained classes allow an assessment of its degree of separability and uncertainty. How classes change in time is used to estimate their temporal stability. The presented approach can be used to investigate the link between a variety of spatial datasets to improve current SIT products, and it can be applied in other scientific fields.

1. Introduction

Global warming effects are amplified in the Arctic, and sea ice has been declining with the annual minimum for September 2020 ending up being the second lowest in the 42-year satellite record (NSIDC, 2020). Sea ice governs heat transfer and influences atmospheric circulation, which is particularly important because low- and mid-latitude's climates are closely related to polar climate (Francis & Vavrus, 2012; Overland & Wang, 2010). Monitoring of both sea ice concentration (SIC), as the fraction of sea-ice cover within an observed cell, and sea ice thickness (SIT), are necessary for a consistent determination of sea ice dynamics. Microwave radiometry is independent of daylight and at lower microwave frequency it is mostly unaffected by atmospheric conditions. The emissivity in the microwave spectrum depends on the dielectric properties of sea ice, which are a function of its physical composition including salinity, density, surface temperature, and surface roughness. The collected signal is emitted from a radiating layer which depends on the penetration depth at sensor frequency. A snow layer on top of sea ice influences the radiative properties of sea ice and the energy received by the sensor. The contribution of snow to the emitted signal is less in the lower microwave spectrum, and the separability of surface properties, such as open water and sea ice including SIT, is—in theory—feasible.
Several algorithms have been developed, based on spaceborne brightness temperature ($T_b$) observations and their corresponding sensitivity to sea ice, to retrieve SIC and SIT at Arctic scale. Sensors operating at higher microwave frequencies have smaller penetration depth into sea ice and enable to discriminate sea ice and water at the surface, which allows the monitoring SIC and sea ice extent (SIE). Satellite-based maps of Arctic SIC have been generated since 1979 starting with the launch of the Scanning Multichannel Microwave Radiometer (SMMR) and follow-on missions carrying sensors such as the Special Sensor Microwave Imager/Sounder (SSMIS), and the Advanced Microwave Scanning Radiometer 2 (AMSR2). Algorithms are based on the contrast of $T_b$ using observations at frequencies $\sim 19$ GHz and $\sim 37$ GHz (Comiso & Nishio, 2008; Lavergne et al., 2019; Parkinson et al., 1999). Low-frequency microwave radiometry observations at L-band (1.4 GHz) are sensitive to thin sea ice up to 0.6 m depth. ESA’s Soil Moisture Ocean Salinity (SMOS) mission (Font et al., 2009; Kerr et al., 2010) provides multi-incidence angle full-polarization $T_b$ maps at L-band, and various products have been developed based on SMOS data retrieving thin SIT at Arctic scale (Gupta et al., 2019; Huntemann et al., 2014; Kaleschke et al., 2016; Ricker et al., 2017; Tian-Kunze et al., 2014).

Methods to invert for geophysical parameters such as SIT can be categorized into statistic and process-based modeling. The choice of a suitable technique directly depends on the modeling applications and the available data. Process-based techniques are based on thermodynamic or radiative transfer models and attempt to reproduce the processes of sea ice growth and require detailed knowledge on the sea ice composition, including layering and possible snow loading. Sea ice properties are estimated from empirically determined parameters of different ice types, and sea ice is under continuous transformation showing regional and seasonal variability. Therefore, information about the sea ice heterogeneity is not available and accurate models need collocated and concurrent observations from multiple sources to prevent ambiguous inversion. Models can be subject to over-simplification, and model uncertainty is difficult to estimate, especially at Arctic scale considering an entire year. Validation capability is also limited due to sparsely collected in situ and airborne data. SIT retrieval algorithms perform well during Arctic freeze-up (Kaleschke et al., 2016), whereas high heterogeneity of sea ice during summer melt and limited spatial resolution of satellite observations make SIT estimation highly ambiguous. Therefore, SIT maps of sufficient quality are only available from mid-October to mid-April.

In this study, a statistic-based approach is investigated to segment Arctic sea ice, assuming that independent information on its properties are captured in the SMOS multi-incidence angle $T_b$. However, since SMOS observations have low spatial resolution (35–50 km), small-scale information on the surface heterogeneity (i.e., different ice types) is lost. The aim is to yield a framework to reveal spatial patterns from differences and similarities in the sensitivity of $T_b$ observations to sea ice properties using an unsupervised learning algorithm. A Bayesian inferential model based on Gaussian Mixture Models (GMM) and Hidden Markov Random Fields (HMRF) considers both the statistical characteristics and the spatial correlations of the observations (Wang et al., 2017). The Arctic region is reduced to a relevant number of spatial classes, while keeping the probabilistic distribution for subsequent cluster analysis and uncertainty quantification. Spatial information is provided in terms of a latent field in physical space and statistical information is indicated in the covariances of the obtained classes in feature space. A direct inference of sea ice properties, particularly at the ocean-ice-boundary, is ambiguous because $T_b$ measurements are sensitive to both SIC and thin sea ice. It cannot be distinguished whether an observed surface consists of thin sea ice located on top, or of thicker ice adjacent to sea water. Therefore, observations consisting of open water, and low SIC are corrected using SIC maps of the OSI-401-b product, provided by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The polarization ratio (PR) between horizontally and vertically polarized values is selected for segmentation to increase the sensitivity to sea ice signatures by reducing the effect of physical surface temperature.

2. Data and Methods

In this study, PR maps at multi-incidence angles are obtained from SMOS $T_b$ observations and OSI-401-b SIC maps, and are used to segment the Arctic ocean into sub-regions based on different sea ice properties. The proposed unsupervised machine learning approach is based on a Bayesian inference framework (Wang et al., 2017). The aim is to indicate patterns in a latent field in physical space according to the most relevant
Tb observations. The temporal evolution of these patterns can be analyzed in terms of cluster separability and correlation of the input features to investigate the corresponding sea ice signatures.

2.1. SMOS Multi-Incidence Angle Tb Data

ESA’s SMOS mission was originally designed to provide global and frequent maps of soil moisture and ocean salinity, but measurements also showed sensitivity to different sea ice properties (thin SIT and SIC). The SMOS satellite is equipped with the Microwave Imaging Radiometer with Aperture Synthesis (MIRAS), an interferometric radiometer operating at L-band (∼1.4 GHz) that acquires multi-incidence angle (0–60°) full polarimetric Tb in ascending (6 a.m.) and descending (6 p.m.) sun-synchronous orbit (Corbella et al., 2005). Tb maps are retrieved with a radiometric resolution between 0.8 - 2.2 K, a spatial resolution of ∼35 km at center of field of view, and a revisit time of ∼1–3 days (Kerr et al., 2001). The retrograde polar orbit (98.42° inclination and 758 km altitude) limits the observations to a maximum latitude of ∼84°, resulting in missing values around the poles (“polar hole”). The input data set for this study is given by the SMOS Level 1B data product consisting of the Fourier components of Tb in the antenna polarization reference frame. The high jump discontinuities in Tb between land and sea observations lead to oscillations after image reconstruction at coastal areas (Gibbs phenomenon). These contaminated zones, as well as continental land mass, were removed in the data product. Ascending and descending SMOS observations show only small differences in Tb. Therefore, Tb of both orbits are averaged. A daily multi-angular data set with 2° sampling is created similar to Gabarró et al. (2016) with Tb provided in horizontal and vertical polarization.

2.2. Input Features Selection

The study period includes the late summer melt and the first half of the freeze up period from September 1 to December 31, 2016. A 5-day composite of Tb is considered to guarantee full coverage of the Arctic Ocean. Pixels of Tb images either consist of sea ice with ($0 < $SIC$ \leq 1$), or purely consist of open water (SIT = 0). At frequencies <117 GHz, Planck’s equation can be simplified using the Rayleigh-Jeans approximation, and Tb is the product of the ice emissivity ($\epsilon$) and the physical temperature ($T_{phys}$) with an error <1% (Ulaby et al., 1986). $T_{phys}$ varies depending on the atmospheric conditions among the Arctic and is non-negligible in the lower microwave spectrum. Therefore, input data for segmentation are selected with the objective to correct for SIC and to reduce the effect of spatial and temporal variability of $T_{phys}$ on Tb. In addition, direct inference of specific sea ice properties, particularly at the ocean-ice-boundary, is ambiguous by the fact that Tb can be sensitive to both SIC and thin SIT, and different ice types can be mixed in low-resolution SMOS images.

In a first step, $T_{tb\ SI}$ was determined from the observed Tb, SIC, and the freezing point of seawater ($T_{h\ OW}$) (Equation 1).

\[
T_b = \alpha T_{h\ SI} + (1 - \alpha) T_{h\ OW} \quad \text{with} \quad \alpha \in [0,1] \quad \text{and} \quad T_b = \epsilon T_{phys}
\]  

Hereby, OSI-401-b SIC maps are provided in a polar stereographic projection grid at 10 km resolution and are regridded and upsampled to SMOS resolution using k-d-tree resampling - a fast nearest-neighbor-interpolation method (Bentley, 1975). Images are resampled by recursively splitting the grid into subsets to evaluate the nearest neighbors within a defined radius of influence according to the resolution of the measurements. $T_{h\ OW}$ are determined at different incidence angles and polarizations by evaluating the coldest values obtained for observations with low SIC located at latitudes above 75°N. SIC is often underestimated with respect to SIT, resulting in an overestimation of Tb, which particularly influences the segmentation of areas covered by thin ice along sea ice edges. Therefore, a SIC threshold of $\alpha = 0.5$ was chosen to provide an open water mask and to exclude observations classified with low SIC, which limits the overestimation error.

In a second step, to account for variations in $T_{phys}$, the polarization ratio (PR) is computed as the normalized difference between vertically and horizontally polarized values ($T_{h\ SI\ V}$ and $T_{h\ SI\ H}$) as follows
which reduces to the emissivities of sea ice with the advantage of enhancing the sensitivity to the actual sea ice properties. $T_{\text{SI}H}$ is higher than $T_{\text{SI}H}$ with larger differences for increasing incidence angles. Also, the emissivity depends on the optical path length through sea ice, and PR increases for observations at higher incidence angles. Figure 1a indicates the relation between PR and SIT as a function of incidence angle and sea ice type using a radiative transfer model (Menashi et al., 1993). PR's obtained for high incidence angles show sufficient sensitivity range over ice-covered area with values reaching from 0 (thick ice, saturation) to ~0.4 (thin ice) and its distribution depends on the observed period. Selecting PR values for high angles increases the content of independent information about sea ice, whereas values at lower angles are more...
likely to contain redundant information, which may lead to segmentation biases. Therefore, PR maps at 40°, 48° and 56° are used as input features for segmentation.

2.3. Bayesian Unsupervised Machine Learning Algorithm

A Bayesian unsupervised machine learning approach Wang et al. (2017) is employed, previously applied to extract patterns of subsurface heterogeneity from geophysical multi-source data (Herbert et al., 2019; Wang et al., 2019). A Gaussian Mixture Model (GMM) is used to fit $N$ data points (image pixels) in an $M$-dimensional space ($M$ number of features) to find an optimal set of multivariate Gaussian distributions ($L$ classes). The distributions are parametrized by their means $\mu_{l\theta}$ and covariances $\Sigma_{l\theta}$ for each cluster $l$ and incidence angle $\theta$. Since features originate from satellite observations, a Hidden Markov Random Field (HMRF) is used to consider the statistical characteristics of data points in feature space as well as their spatial dependencies. A directional smoothing coefficient $\beta$ accounts for anisotropy conditions with the assumption that neighboring pixels are more likely to belong to the same class. The segmentation results in a latent field $x$ of hidden variables, which indicates the most probable class membership as well as the probability $p(x_i \mid l)$ of each pixel $i$ to belong to class $l \in L$. The segmentation procedure is described in detail in Wang et al. (2017). The model parameters ($\mu$, $\Sigma$, $\beta$) as well as the latent field $x$ are obtained through Bayesian optimization in an iterative sampling process using a Markov Chain Monte Carlo (MCMC) approach after an initial Expectation-Maximization step. The convergence of the segmentation and the model accuracy can be monitored using the ratio between the number of misclassified data points to the total number of data points. Prior to segmentation, the number of classes was predefined regarding the distribution of PR values. During late summer melt, only two significant classes are expected, comprising the remaining thick multi-year ice and regions of thinner ice. With the start of the freeze-up period after mid-September, newly formed sea ice and hence high PR values become more abundant to be captured in the segmentation. The choice of a new significant class is further approved by a posteriori evaluation of cluster separability.

2.4. Cluster Analysis

Results of the Bayesian segmentation are analyzed regarding the obtained patterns in physical space, and the location and orientation of clusters in feature space. The information-theoretic measure of entropy ($H$) is used to provide model uncertainty. It was initially defined by (Shannon, 1948) in the context of communication and since then has been adapted to geosciences (Goodchild et al., 1994; Wellmann & Regenauer-Lieb, 2012). It is used to distinguish well-classified from uncertain regions and is defined by

$$H(x) = -\sum_{l=1}^{L} p(x_i \mid l) \log(p(x_i \mid l)).$$

where $p(x_i \mid l)$ denotes the probability in the physical space of pixel $i$ to belong to class $l$. $H$ can reach values close to zero (pixel clearly assigned to one class) and $H_{max} = L[1 - \log(L)]$ (uniform distribution for $L$ classes).

Sea ice properties, to which SMOS multi-incidence $T_b$ are sensitive to, are dissimilar between classes and show similarities within the same class. Clusters in feature space are investigated regarding their location and orientation by analyzing the model parameters $\mu$ and $\Sigma$. The correlation coefficient $\rho$ quantifies the intra-cluster cohesion and can be used to distinguish between informative and redundant observations (Benesty et al., 2009). It is derived for each cluster from $\Sigma$ in two-dimensional marginal space between features $j$ and $k \in \{40^\circ, 48^\circ, 56^\circ\}$

$$\Sigma = \begin{bmatrix} \Sigma_{\theta} & \Sigma_{\rho} \\ \Sigma_{\rho} & \Sigma_{\theta} \end{bmatrix} = \begin{bmatrix} \sigma_i^2 & \sigma_i \sigma_j \rho_{jk} \\ \sigma_j \sigma_i \rho_{jk} & \sigma_j^2 \end{bmatrix},$$

where $\sigma_{jk}$ correspond to the standard deviations with respect to feature $j, k$ and $\rho_{jk} = \rho_{kj}$ denote the correlation coefficients between two features, given by
The Geometric Separability Index (GSI) (Thornton, 1998) is a distance-based measure to analyze inter-cluster separability and is widely used for cluster interpretation (Greene, 2001; Mthembu & Marwala, 2008). GSI compares all \( N \) data points with their nearest neighbor regarding their class membership and is defined by

\[
GSI(f) = \frac{\sum_{i=1}^{N} (f(x_i) + f(x_i') + 1) \mod 2}{N}, \quad \text{with} \quad f(x_i) = \begin{cases} 1, & \text{if } x_i' = x_i \\ 0, & \text{if } x_i \neq x_i' \end{cases}
\]

where \( f \) is a binary target function, and \( x_i' \) is the nearest neighbor of \( x_i \) in the feature space of pixel \( i \). GSI \( \in [0.5, 1] \) and for values reaching its lower or upper limit, clusters are completely entangled or ideally separable, respectively. In this study, both global and cluster-specific separability are estimated. Global separability is computed based on Euclidean distance for all data points, and cluster-specific separability is obtained based on Mahalanobis distances \( (x_i - \mu)\Sigma^{-1}(x_i - \mu)^T \), considering the data points and covariances of the specific cluster (Mahalanobis, 1936). GSI is investigated during the study period to evaluate the dynamics of the underlying sea ice properties and the stability of the segmentation.

3. Results

Arctic sea ice is segmented independently using 5-day composites into classes during the periods of late summer melt and early freeze up from September 1 to December 31, 2016. The latent field in physical space and the corresponding multivariate Gaussian distributions of data points in feature space are presented as an example for the segmentation step interval October 24 – 28, 2016 (Sections 3.2 and 3.1). The temporal evolution of the clusters is evaluated in Section 3.3. Class membership and separability are assessed in Section 3.4 to indicate cluster stability and performance of the algorithm.

3.1. Latent Field of Classes in Physical Space

Figures 1c and 1d show the resulting latent field and the model uncertainty quantified by information entropy, respectively. The latent field indicates spatial patterns, which are acquired from the final iteration of the segmentation by assigning the class with highest probability to every pixel. Pixels with the probability to belong to two or more clusters have larger entropy and reflect therefore uncertain pixels. These pixels comprise regions at the boundary between classes and pixels, which are generally difficult to assign to any cluster. In the latter case, these pixels may point out sub-regions with different sea ice properties (anomalies), which are characterized with high model uncertainty.

The segmented spatial patterns are compared to those of the SMOS L3 Sea Ice Thickness product, provided by the Alfred Wegener Institute (AWI) for Polar and Marine Research (Tian-Kunze et al., 2014). SIT mean values were computed according to the indicated spatial classes in each segmentation step, and averaged values are determined during freeze up from October 15 to December 31, 2016. The three classes can be associated to different ice thickness (in meters) of 1.24 ± 0.10, 0.54 ± 0.24, and 0.13 ± 0.07, respectively. The classes are labeled as (0 thick ice up to sensor saturation) (1 transition zone with higher thickness variability, containing various ice types), and (2 newly formed thin ice).

3.2. Clusters in Feature Space

The multivariate Gaussian distributions in marginal features space between 48° and 56° are illustrated in Figure 1b. The correlation between the input features is generally higher for thick ice, resulting in a well-determined cluster with high intra-cluster cohesion. In contrast, newly formed thinner ice shows less correlation between input features. This enables discrimination of classes with similar surface characteristics, to which multi-incidence angle observations show a different signature. However, sea ice is a complex...
medium and sea ice growth can occur under rougher or calmer ocean conditions, causing newly formed ice to be heterogeneous. These differences in the origin of sea ice formation might be captured in the input features, indicated by a broader distribution in marginal features space. On the contrary, the structures of multi-year thick ice appear more homogeneous.

### 3.3. Temporal Evolution of Clusters

Figure 2 shows the temporal evolution of cluster mean values and standard deviations (StDev) in marginal feature space for $\theta = 56^\circ$, and the distribution of PR and the corresponding class membership at three particular dates (September 4–8, October 24–28, and December 24–28, 2016).

Figure 2. Temporal evolution of clusters. (a) Cluster mean values and standard deviations at 56° and mean Arctic temperature for latitudes > 75°N. (b) PR distribution with respect to class membership at three particular dates (September 4–8, October 24–28, and December 24–28, 2016).
Figure 3a shows the evolution of the number of pixels per class membership for filtered sea ice (SIC > 0.5) in comparison to the total. SIE comprises sea ice cover for SIC > 0.15 and daily data was downloaded from the data archive of the National Snow and Ice Data Center (NSIDC, 2020). Deviations and offsets between SIE and the total pixel counts are due to missing values within the “polar hole” and contaminated zones at the sea-land boundary. The increase of the total number of pixels is equivalent to a monthly growth rate in SIE of about $2.5 \times 10^6$ km$^2$. The number of pixels consisting of newly formed ice is broadly stable, whereas the number pixels classified as transition zone are slightly increasing during freeze up. As sea ice grows, thick sea ice becomes more abundant, leading to a log-normal-shaped PR distribution with increasing expected value (Figures 2b and 3). Although thin ice becomes less representative in the data during freeze up, the algorithm is still capable of separating three classes as long as sea ice formation continues.

### 3.4. Separability of Clusters

Global and cluster-specific separability are shown in Figure 3b. The solid lines show the GSI for a choice of two classes in late summer melt and three classes during freeze up. High global separability is achieved during the entire study period with values around 0.9. The cluster-specific GSI indicates separable classes with mean values of 0.95, 0.83, and 0.83 for thick ice, transition zone and new sea ice, respectively. During the freeze up, thin ice starts to overlap with the transition zone and a threshold of minimum GSI needs to be defined to specify the appropriate number of classes for each segmentation step. For comparison, GSI is shown for the end of the summer melt for a segmentation with three classes (dashed lines). In this case, classes highly overlap and the choice of two initial clusters from the beginning of the study period leads to higher separability.
4. Discussion

A novel approach is evaluated to obtain sea ice maps from SMOS observations using Bayesian segmentation. The estimation of a constant number of stable and separable classes revealed periods, when $T_b$ observations show similar sea ice signatures. The information content obtained by linking $T_b$ data at multiple incidence angles and polarizations is reduced to a number of most significant classes, with good inter-cluster separability. The corresponding spatial patterns, which are indicated in the latent field result, can be used to extract the heterogeneity of the underlying sea ice properties.

Information entropy points out both uncertain zones between segmented classes and anomalies which can form sub-classes. As an example, ponded sea ice during summer melt has different surface characteristics, which may result in a further discriminable class only during that particular period. Since cluster mean values represent the most significant observations at every segmentation step, their temporal evolution can be used to define dynamic tie points. These tie points can be analyzed to investigate how sensitive input features respond to changes in sea ice signatures. So far, the algorithm was applied to observations comprising the late summer melt and the beginning of the freeze up and classes were mainly analyzed during the formation of sea ice. The computational cost scales linearly with the segmented period, and current findings suggest an extension of the observation period to assess the potential and limitations of the method during an entire annual cycle, and to evaluate its consistency over multiple years.

The implemented method serves as a framework to integrate multi-source datasets and is capable of recognizing patterns by considering the statistical characteristics and spatial correlations. The relationship of satellite observations at multiple frequencies can be used to select an appropriate set of input features and to enhance the sensitivity to ice-physical parameters, such as SIT. A combination of the presented approach with a physics-based inference model build upon the estimated distribution of classes may increase the retrieval accuracy of existing large-scale sea ice products.

5. Conclusion

In this work, Arctic sea ice is classified using a Bayesian unsupervised learning approach by making full use of the information about sea ice properties contained in the PR of SMOS multi-incidence angle $T_b$ data. Sea ice properties are considered anisotropic as well as regionally and seasonally variable among the Arctic and $T_b$ cannot be assumed to be sensitive to similar properties over an entire year. Therefore, both statistical characteristics of observations are evaluated and the segmentation is carried out by means of a discretized number of spatially regularized classes. The number of classes was determined a priori from the PR distribution, and it was verified a posteriori using GSI. Model uncertainty was determined using information entropy and enabled to distinguish well-determined from uncertain regions. High global separability was achieved considering two classes during late summer melt and three classes during freeze up, respectively. A comparison with existing SMOS-SIT maps indicated that classes can be attributed to SIT ranges. During late summer melt, two classes could be attributed to remaining thick ice and a transition zone, showing differences in the correlations of the input features. With the beginning of the formation of new thin ice during freeze up, an additional class could be discriminated based on the occurrence of higher PR values. However, the decrease in relative abundance of newly formed ice to the total sea ice during freeze up resulted thin sea ice to be less significant and led to higher overlap between classes. The underlying sea ice properties and the corresponding variation in PR have to be better understood to draw conclusions of the obtained classes, considering an entire annual cycle of Arctic sea ice formation and melting.

Data Availability Statement

Data from October 15 to December 31, 2016 were obtained from AWI (https://smos-diss.eo.esa.int/socat/L3_SIT_Open).

References

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1–4). Springer.
Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM, 18*(9), 509–517.

Csiszár, I. (2017). ERAS: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/home. Copernicus Climate Change Service Climate Data Store (CDS). (Accessed: 2020-01-10)

Comiso, J. C., & Nishio, F. (2008). Trends in the sea ice cover using enhanced and compatible ams-r, ssmi, and ssnir data. *Journal of Geophysical Research, 113*(C2), 1–7. https://doi.org/10.1029/2007JC004257

Corbella, I., Torres, F., Camps, A., Collander, A., Martin-Neira, M., Ribó, S., et al. (2005). Miras end-to-end calibration: Application to ssnir l1 processor. *IEEE Transactions on Geoscience and Remote Sensing, 43*(5), 1126–1134

Font, J., Camps, A., Borges, A., Martin-Neira, M., Boutin, J., Reul, N., et al. (2009). Smos: The challenging sea surface salinity measurement from space. *Proceedings of the IEEE, 98*(5), 649–665.

Francis, J. A., & Vavrus, S. J. (2012). Evidence linking arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters, 39*, L08601. https://doi.org/10.1029/2012GL051000

Gabarró, C., Pla Resina, J., Turiel, A., Portabella, M., Martínez, J., Olmedo, E., & González, V. (2016). Arctic sea ice concentration estimation with smos data. In *European Space Agency (ESA) Living Planet Symposium, Paper 912 - Session title: Cryosphere Posters*, 9–13 May 2016, Prague, Czech Republic.

Goodchild, M., Chih-Chang, L., & Leung, Y. (1994). Visualizing fuzzy maps. *Visualization in Geographical Information Systems, 158–167.

Greene, J. (2001). Feature subset selection using thornton’s separability index and its applicability to a number of sparse proximity-based classifiers. In *Proceedings of annual symposium of the pattern recognition association of South Africa*. Retrieved from http://www.prasa.org/proceedings/2001/prasa01-01.pdf

Gupta, M., Gabarró, C., Turiel, A., Portabella, M., & Martinez, J. (2019). On the retrieval of sea-ice thickness using smos polarization differences. *Journal of Glaciology, 65*(251), 481–493.

Herbert, C., Wellmann, F., Wang, H., & Hebel, C. v. (2019). Extracting heterogeneity of subsoil from geophysical measurements using unsupervised learning algorithms. In *Geophysical Research Abstracts (Vol. 21, EGU2019-17054). EGU General Assembly 2019.

Hunttemann, M., Heygster, G., Kaleschke, L., Krumpen, T., Mäkynen, M., & Drusch, M. (2014). Empirical sea ice thickness retrieval during the freeze up period from smos high incident angle observations. *The Cryosphere, 8*(2), 439–451. https://doi.org/10.5194/tc-8-439-2014

Kaleschke, L., Tian-Kunze, X., Maasß, N., Beitsch, A., Wernecke, A., Miernecki, M., et al. (2016). Smos sea ice product: Operational application and validation in the barents sea marginal ice zone. *Remote Sensing of Environment, 180*, 264–273. https://doi.org/10.1016/j.rse.2016.03.009

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., et al. (2010). The smos land mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE, 98*(5), 668–687.

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. A. M. J., & Font, T. J. (2001). Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing, 39*(8), 1729–1735.

Lavergne, T., Sorensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., et al. (2019). Version 2 of the eumetsat osi sof and esa cci sea-ice concentration climate data records. *The Cryosphere, 13*(1), 49–78. https://doi.org/10.5194/tc-13-49-2019

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (India), 21*, 49–55.

Menashi, J. D., St. Germain, K. M., Swift, C. T., Comiso, J. C., & Lohanick, A. W. (1993). Low-frequency passive-microwave observations of sea ice in the weddell sea. *Journal of Geophysical Research, 98*(C12), 22569–22577. https://doi.org/10.1029/93JC02058.

Mthembu, L., & Marwala, T. (2008). A note on the separability index. *Cornell University, arXiv preprint arXiv:0812.1107v1*

NSIDC. (2020). Arctic sea ice at minimum extent for 2020 (September 2020). USA. NASA National Snow and Ice Data Center. (Accessed: 2020-10-08)

Overland, J. E., & Wang, M. (2010). Large-scale atmospheric circulation changes are associated with the recent loss of arctic sea ice. *Tellus A, 62*(1), 1–9. https://doi.org/10.1111/j.1600-0870.2009.00461.x

Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. J., & Comiso, J. C. (1999). Arctic sea ice extents, areas, and trends, 1978-1996. *Journal of Geophysical Research, 104*(C9), 20837–20856. https://doi.org/10.1029/1999JC900082.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., & Haas, C. (2017). A weekly arctic sea-ice thickness data record from merged cryosat-2 and smos satellite data. *The Cryosphere, 11*, 1607–1623. https://doi.org/10.5194/tc-11-1607-2017.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.

Thornton, C. (1998). Separability is a learners best friend. In *Proceedings of annual symposium of the pattern recognition association of South Africa*. Retrieved from http://www.prasa.org/proceedings/2001/prasa01-01.pdf

Tian-Kunze, X., Kaleschke, L., Maasß, N., Mäkynen, M., Serra, N., Drusch, M., & Krumpen, T. (2014). Smos-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification. *The Cryosphere, 8*, 997–1018. https://doi.org/10.5194/tc-8-973-2014.

Ulaby, F. T., Moore, R., & Pung, A. (1986). *Microwave remote sensing: Active and passive. Volume 3 From theory to applications (Book).* Research supported by NASA, NSF, and DOD, Dedham, MA: Artech House, Inc. (Remote Sensing, 4 (1986)).

Wang, H., Wellmann, F., Zhang, T., Schaff, A., Kanig, R. M., Verweij, E., et al. (2019). Pattern extraction of topsoil and subsoil heterogeneity and soil-crop interaction using unsupervised bayesian machine learning: An application to satellite-derived ndvi time series and electromagnetic induction measurements. *Journal of Geophysical Research: Biogeosciences, 124*(6), 1524–1544. https://doi.org/10.1029/2019JG005046.

Wang, H., Wellmann, J. F., Li, Z., Wang, X., & Liang, R. Y. (2017). A segmentation approach for stochastic geological modeling using hidden Markov random fields. *Mathematical Geosciences, 49*(2), 145–177.

Wellmann, J. F., & Regenauer-Lieb, K. (2012). Uncertainties have a meaning: Information entropy as a quality measure for 3-d geological models. *Tectonophysics, 526*, 207–216. https://doi.org/10.1016/j.tecto.2011.05.001