Satellite-based Permafrost Mapping and Climate Trend Analysis in Northern North America

Daniel Monteagudo

Advisor: Ronald B. Smith, Geology & Geophysics

Second Reader: Xuhui Lee, Geology & Geophysics

April 29, 2020

A Senior Thesis presented to the faculty of the Department of Geology and Geophysics, Yale University, in partial fulfillment of the Bachelor's Degree.

Satellite-based Permafrost Mapping and Climate Trend Analysis in northern North America

Daniel Monteagudo

Abstract

The role of permafrost in the carbon cycle remains a major unknown in the study of climate change. Permafrost stores carbon as frozen organic matter within soil, but as rising global temperatures cause it to thaw, the stored carbon is able to decay and escape into the atmosphere. These thaw-based emissions may contribute to further warming and thus further thawing, in a positive feedback loop. In order to understand how severely this may impact future warming scenarios, scientists need to know where permafrost is, how much of it is melting, and how fast. This study contributes to this effort by focusing on large-scale mapping of permafrost in North America over the past 20 years. However, measuring permafrost with traditional methods (e.g. bore-holing) is difficult and expensive, meaning most permafrost datasets suffer from coarse spatial and/or temporal resolution. In this study we design a new approach to work around these measurement limitations by utilizing satellite imagery to provide nearly continuous data. We use a classification algorithm, trained using a 1997 base-map of known permafrost extent, to correlate the presence and extent of permafrost with three key MODIS-derived climatic variables —land surface temperature (LST), vegetation (NDVI), and snow (NDSI)— and two temporally constant variables— elevation (GTOPO30), and latitude. The algorithm succeeded in reproducing the base-map with an overall accuracy of 85% and kappa value of 0.75. With a three-year moving buffer, we apply this model to the last 20 years of MODIS data to produce yearly maps of permafrost extent over North America. We then analyze the spatial and temporal trends of the raw data, as well as permafrost, by using unsupervised clustering to identify regions with similar behavior over the 20-year period. We find that zones with similar climates do not necessarily have similar trends, highlighting the highly localized nature of permafrost's response to changing climates. Notably, we also observe a consistent decline in inferred regions of "continuous" permafrost, partially mirrored by an increase in "discontinuous" permafrost over the same time period, and we identify other "flashpoints" of permafrost freezing and thawing.

Contents

1. Introduction

Permafrost is defined as ground that has been frozen for at least two consecutive years, for natural climatic reasons (Van Everdingen, 1998). Permafrost regions underly 22% of exposed land in the Northern Hemisphere, and are classified by extent into the categories: “continuous” (100-90%), “discontinuous” (90-50%), “sporadic” (50-10%), and “isolated” (<10%). In total, this means permafrost actually underlies between 17 and 12 million square kilometers. Trapped within it is around 1500 gigatons of carbon (Zimov, 2006), making permafrost research a key area of study as current warming trends threaten to thaw permafrost and release this carbon into the atmosphere (Biskaborn et al, 2019). With the threat of new thaw-based carbon emissions (Schuur et al., 2015), the motivation for this study comes from the need to know how much permafrost will thaw, where, and how fast, in order to help determine how much greenhouse gas emissions we may be expected to deal with as a result.

The problem is that permafrost is difficult to study, since it is underground. This makes direct data collection, including methods such as bore holing or ice coring, expensive and difficult. As a result, permafrost datasets are often very limited in their temporal and spatial resolution. This makes it hard to perform any long term or scalable analyses on the potentially changing state of permafrost, a hindrance if we are to develop the insight we need to understand how permafrost will respond to ongoing and future climate change. However, other variables related to permafrost —climatological ones, for instance— are much easier to measure, and can be used to study permafrost indirectly (Obu et al., 2019). From satellites, one can obtain frequent and spatially-continuous measurements that effectively eliminate the need for extrapolation (Ran et al., 2015).

The first part of this project focuses on developing a methodology for predicting the presence of permafrost using such satellite measurements. In our study we focus on three key temporally changing variables —land surface temperature, vegetation, and snow cover— and two unchaining variables— elevation and latitude.

The second part of this project applies this methodology to the full 20 years of available satellite data, creating maps of inferred permafrost extent for each year. We then analyze the trends in the raw data, as well as in the permafrost maps we produced, by region. These regions were defined using clustering, to identify areas with similar characteristics and trends.

2. Data

2.1. Data Handling and Sources

This study uses Google Earth Engine (GEE) as its platform for all data manipulation, analysis, and collection processes. GEE was chosen both for its ability to process large amounts of data (achieved by outsourcing processing tasks to numerous Google server computers in parallel), and for the extensive datasets that Google has made readily available for import. This allowed us to work with the most spatially and temporally expansive dataset utilized in the field to date, to our knowledge. Our study makes use of five satellite derived datasets, and one non-satellite derived permafrost dataset. We aim to use the descriptive local information given by satellite data, which refreshes frequently and is available almost everywhere, to act as indicators of the presence of permafrost, which we verify using the much less-easily obtained permafrost dataset.

2.1.1. MODIS-derived Datasets

For our study we used data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS). Although we could have chosen newer, higher resolution imagery, or made good use of data from satellites like GRACE[1], we decided upon MODIS because of its relatively long history in the remote sensing world— with data available from as early as the turn of the century. Furthermore, the coarse spatial resolution of MODIS allowed for faster processing, which was useful in designing and adjusting this study's proof-of-concept methodology.

Land Surface Temperature (LST) was the single most important piece of information used for our permafrost modeling and predictions, due to the direct physical link between ambient temperature and the freeze-thaw state of soil (Li et al., 2019). For LST, the "MOD11A2.006 Terra Land Surface Temperature and Emissivity 8-Day Global 1km" collection was used (Fig. 1).

---

[1] Researchers have successfully used groundwater content to detect changes in the freeze-thaw state of permafrost. Furthermore, soil oxygenation (controlled primarily by wetness) plays a key role in determining whether microbes release CO2 or methane as they digest the organic matter in thawed permafrost.

Fig. 1) Twenty-year mean LST, with hillshade overlay

Fig. 2) Twenty-year mean NDSI, with hillshade overlay

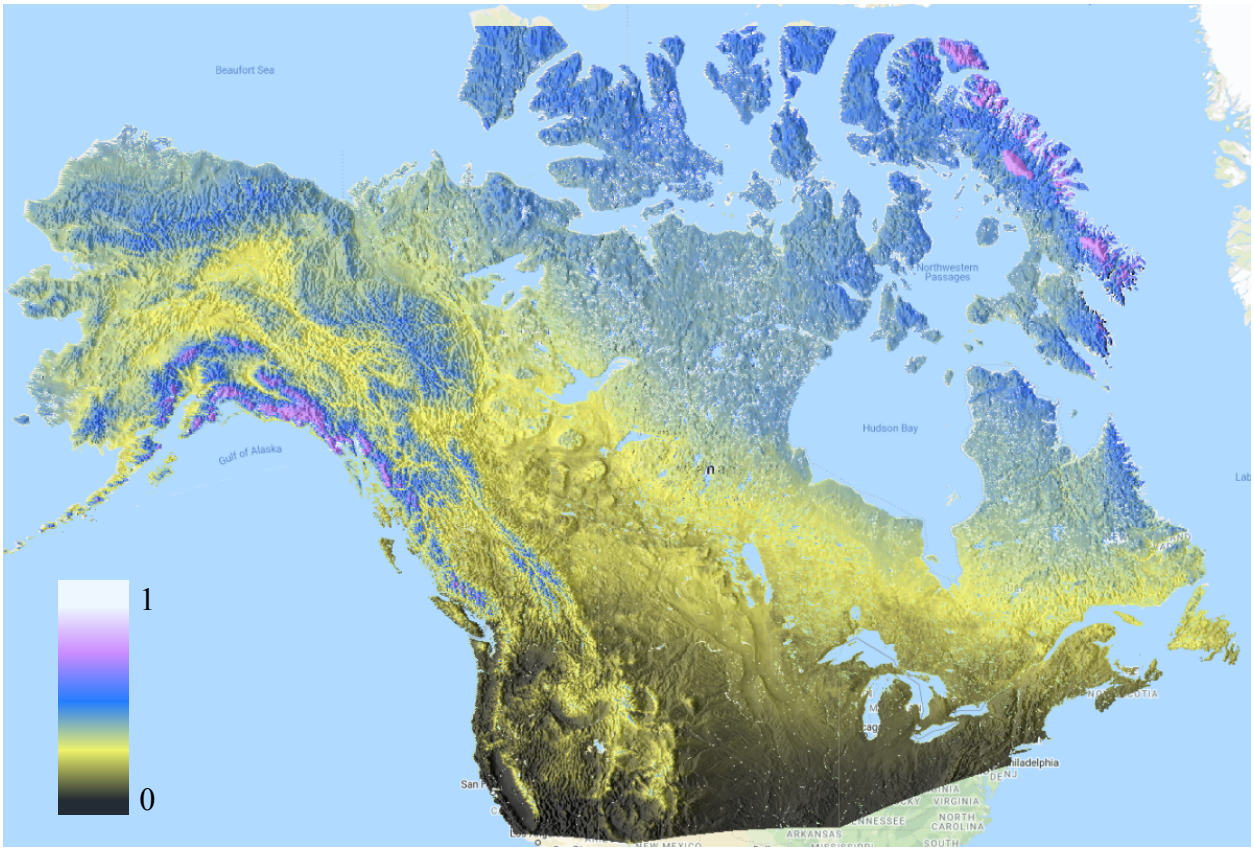
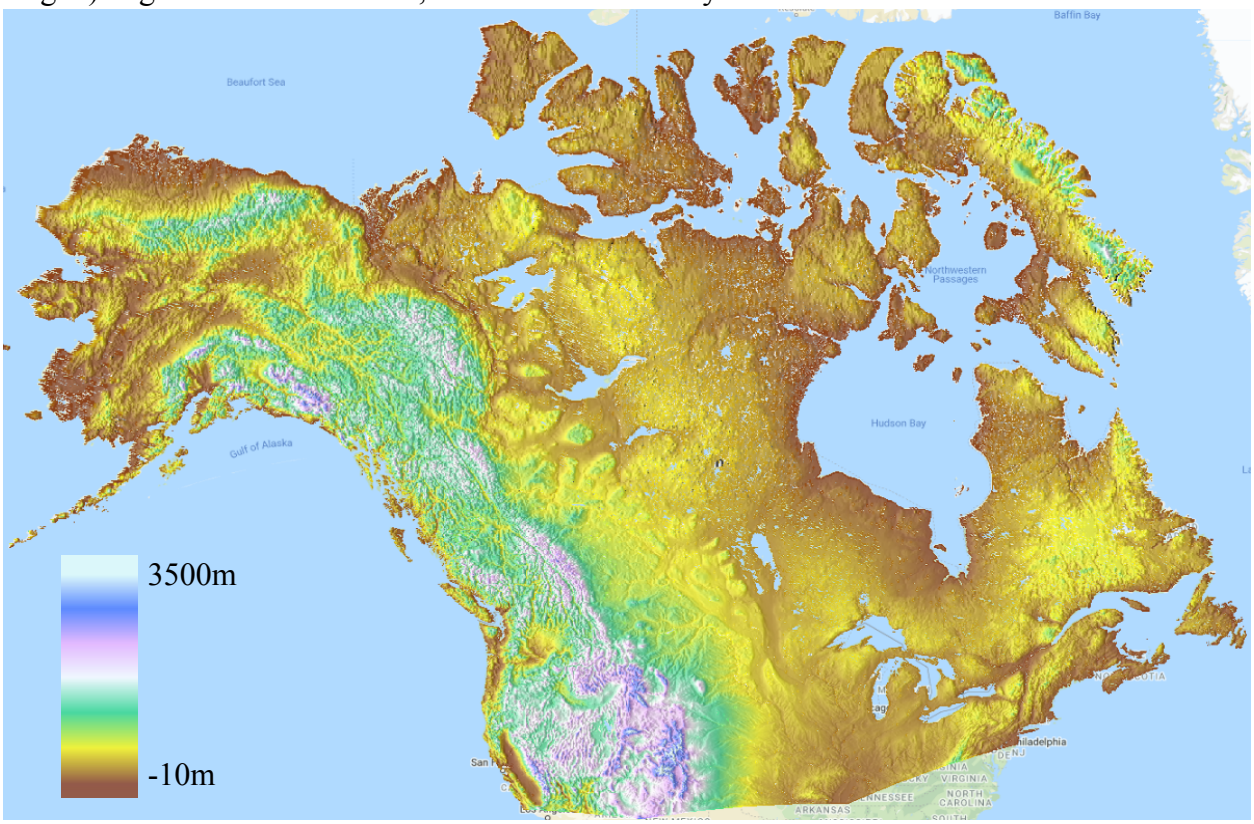


Fig. 3) Twenty-year mean NDVI, with hillshade overlay

Fig. 4) Digital Elevation Model, with hillshade overlay



Despite the near causal relationship between LST and permafrost, there is not perfect correspondence between the two. Using other variables in addition to LST helped our data paint a more complete picture of the local physics. Snow cover, for example, changes the albedo of soil drastically. Acting as an insulator, snow can either keep soil warmer or colder than it would otherwise get, depending on the ratio of incoming solar radiation to outgoing thermal radiation (Zhang, 2005). For snow cover data, we used the Normalized Difference Snow Index (NDSI) obtained from the "MOD10A1.006 Terra Snow Cover Daily Global 500m" collection (Fig. 2).

The other key variable in our analysis was vegetation. Plants can alter soil thermodynamics in many ways, such as by providing shade or regulating water content. Based on the fact that different plants require different depths of available topsoil for their roots, incorporating vegetation data was thought to provide additional information about below-surface properties of the soil— which are still very relevant to the existence of permafrost, but which are invisible in the LST and NDSI signatures. For vegetation data, we used the Normalized Difference Vegetation Index (NDVI) taken from the "MID13Q1.006 Terra Vegetation Indices 16-Day Global 250m" collection (Fig. 3).

We also utilized an elevation dataset retrieved from the "GTOPO30: Global 30 Arc-Second Elevation" dataset (Fig. 4). Elevation was thought to play a role in terms of its effects on the other variables (for example, an inverse relationship with temperature), but also potentially in its own right as a determinant in ground water flow, and insolation (Kenner et al., 2019).

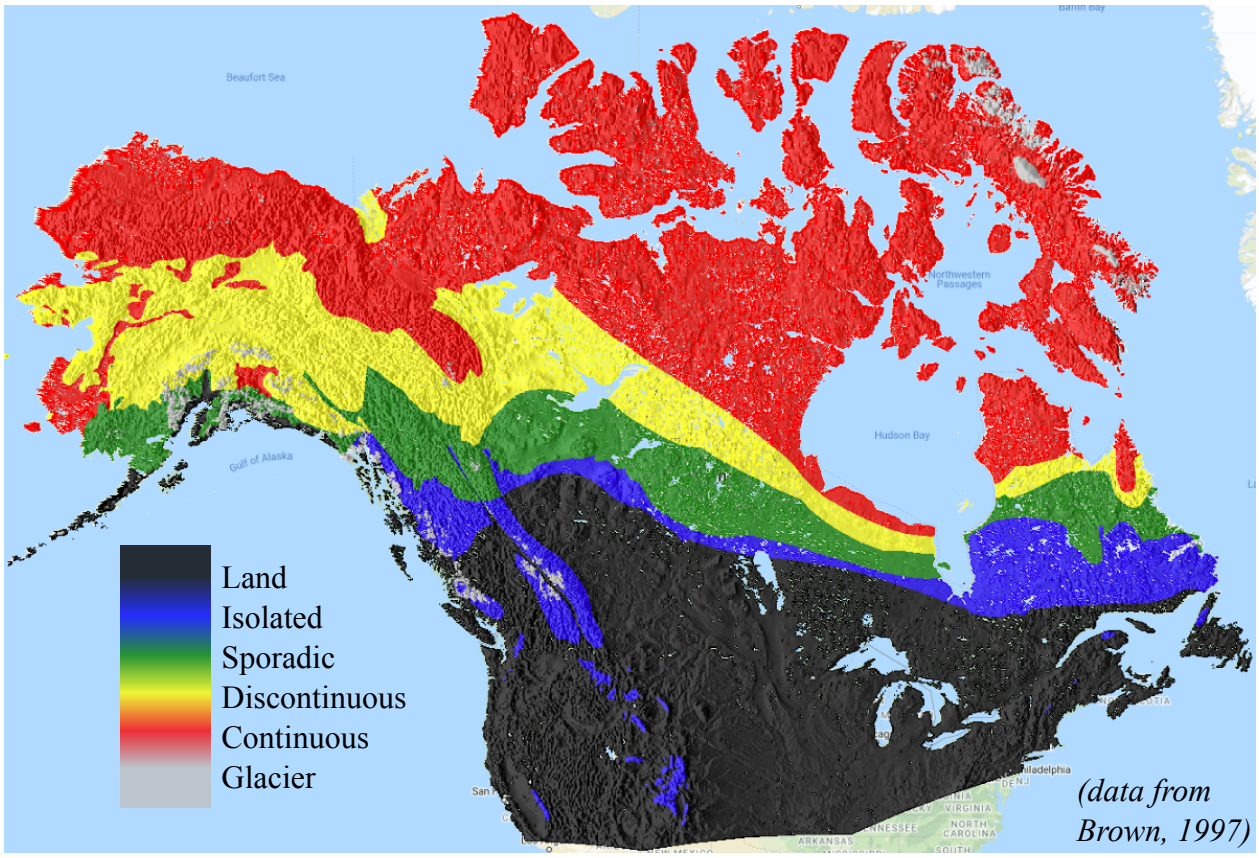
For our study we also applied a water mask (not pictured) to every image and collection before analysis. We used the the "MOD44W.006 Terra Land Water Mask Derived from MODIS and SRTM Yearly Global 250m" product for this purpose.

2.1.2. Permafrost Dataset

Much of the analysis is also based upon a permafrost map downloaded from the International Permafrost Association's website, created by Brown et al. in 1997. This map is the most recently available of its kind, and was created manually based off of the synthesis of many sources of in-situ measurements and interpolated climatic data. Originally subdivided further, this study focuses on just the broader category of "permafrost extent," which ascribes regions

with a percentage of permafrost cover, rather than precise mapping (Fig. 5). The thresholds are as follows: “continuous” (100-90% permafrost), “discontinuous” (90-50%), “sporadic” (50-10%), “isolated” (<10%), “land” (0%), and “glacier” (which we omit from our analyses).

Fig. 5) Permafrost Extent Zones, with hillshade overlay

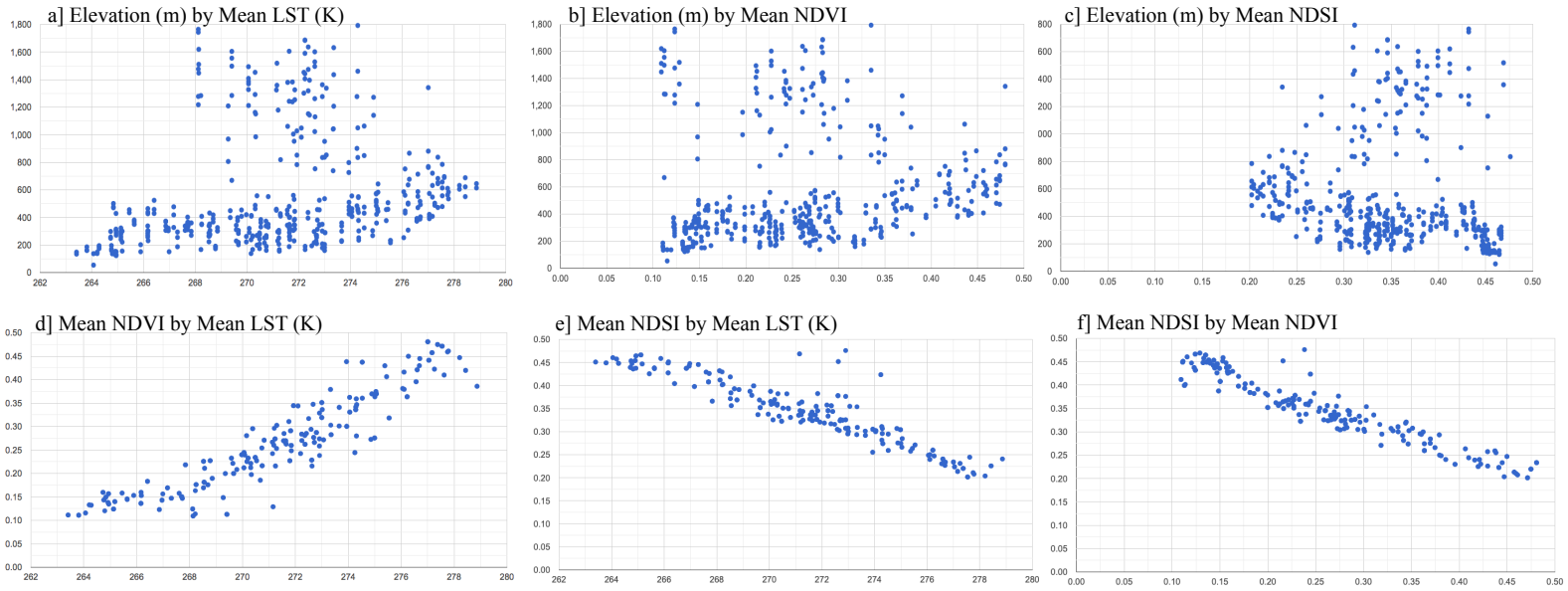


2.2. Study Area and Climatology

The study area for this project encompasses the majority of the northern region of mainland North America. A region of interest was defined, bounded to the south by the line passing roughly between Los Angeles and Philadelphia, around 37ºN, and extending northwards beyond the Northwestern Passages until reaching the Labrador Sea, Beaufort Sea, or Baffin Bay.

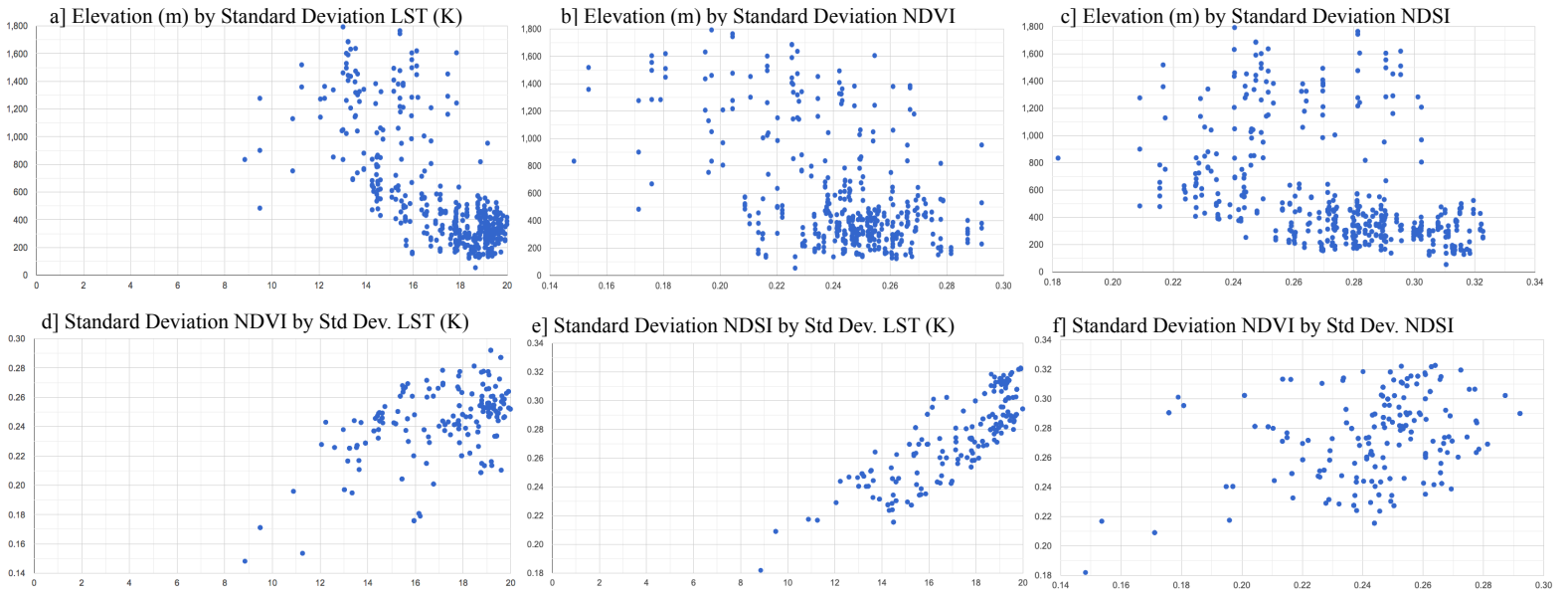
In order to better understand the climatology of the area, preliminary investigations were performed, which included the mapping and visual examination of the four main variables of interest (LST, NDVI, NDSI, elevation). Twenty-year means were calculated and plotted against

Fig. 6) Correlation of Means — *Scatterplots show first variable in each pairing on the y-axis, the second on the x-axis.*



each other in order to understand the correlations between variables (Fig 6). In general, it was found that LST and NDVI correlated positively with elevation, while NDSI correlated with it negatively (a-c). LST, NDVI, and NDSI were all tightly correlated with each other as well, with NDSI correlating negatively again with the other two (d-f). This confirmed physical expectations that vegetation should grow better in warm temperatures, and that snow cover should increase with lower temperatures. Twenty-year standard deviations were also mapped and plotted (Fig. 7) in an attempt to understand the seasonal cycles of the variables. All three variables saw higher

Fig. 7) Correlation of Standard Deviations — *See Fig. 6 for description. Variable pairings plotted per pixel.*



variance at lower altitudes (a-c). The standard deviation of LST correlated positively with that of both NDVI and NDSI, suggesting that changes in these variables happen in parallel (d,f).

Other correlations were plotted as well (Fig 8), with some possible explanations. Notably, it was found that mean LST & NDVI correlated negatively with the standard deviation of NDSI (e,f). It may be said then that at high temperatures, there is little variance in snow cover because there is rarely snow, and as temperatures decrease, snow becomes more and more frequent, increasing the standard deviation. However, if this were the case, we would expect to see the standard deviation fall again at very low temperatures, when snow cover becomes almost constant. The signal is repeated with NDVI because it is directly correlated with LST. A similar explanation may be attributable to [a] and [b], where mean NDVI falls and mean NDSI rises with increasing standard deviation of LST. This may be due to the fact that the LST seasonal cycle increases with latitude. As latitude increases, it gets colder, explaining the corresponding changes in NDVI & NDSI. However, these two plots (a,b) appear to be parabolic, which complicates this theory. This climatological analysis helped inform later decisions in the study.

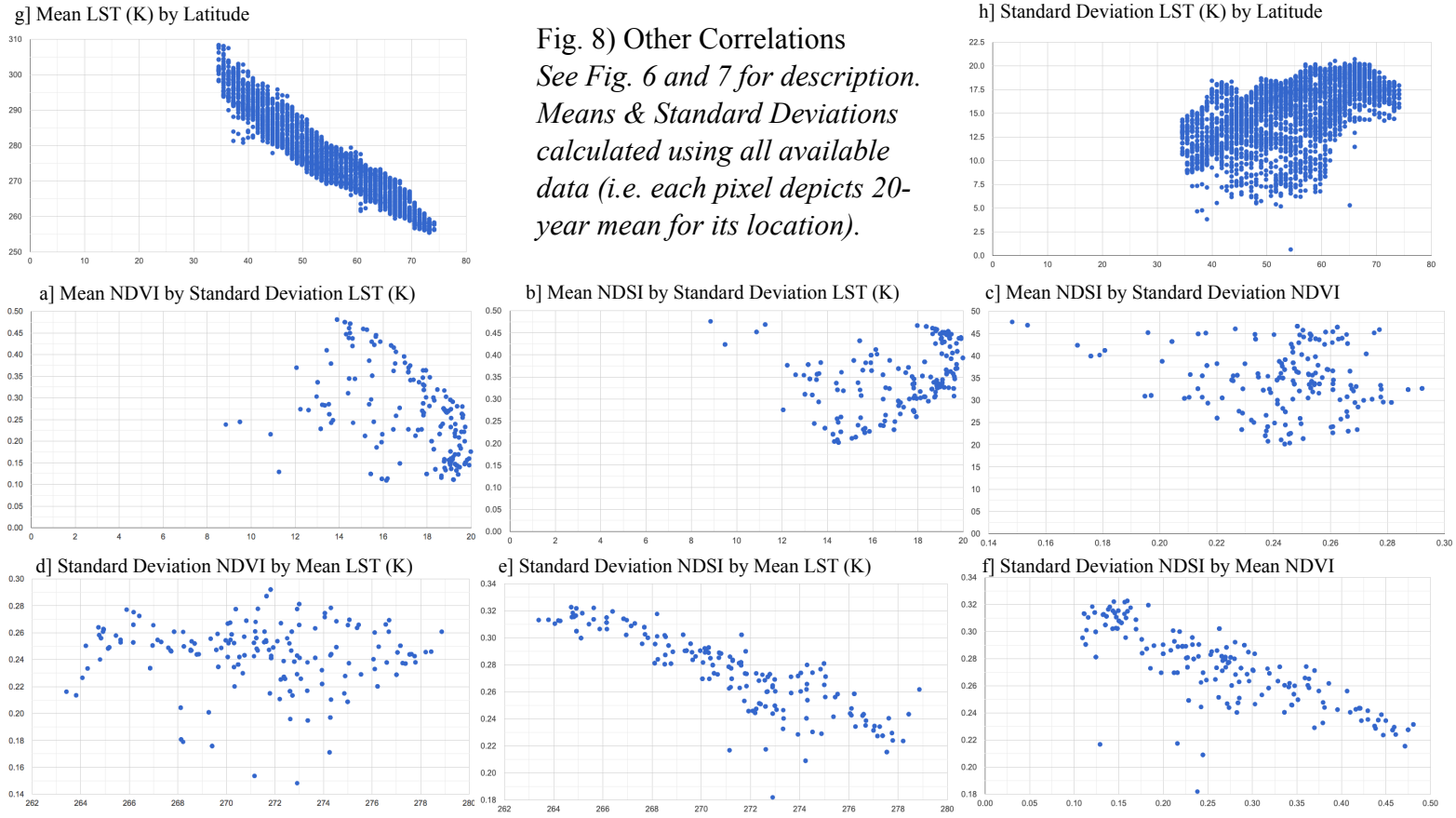


Fig. 8) Other Correlations
See Fig. 6 and 7 for description. Means & Standard Deviations calculated using all available data (i.e. each pixel depicts 20-year mean for its location).

3. Methodology

3.1. Designing the Permafrost Prediction Model

The first part of the study was aimed at developing a model to predict the presence and extent of permafrost based on the variables detailed above. This work was done based off of the 1997 reference map of permafrost by Brown et al., which we simplified from its original format by aggregating regions based solely upon permafrost extent, as defined and depicted above (Fig. 5).

3.1.1. Threshold Classification

First, the potential of the study variables to predict permafrost was tested using simple threshold classifications. Characteristic thresholds —delineating between classes of permafrost extent based on the value of a single variable— were determined and chosen based on scatter plots (Fig. 9) obtained by overlaying 2001-2004 satellite data over the reference map of permafrost extent. Predictive maps of permafrost extent were made using each variable by itself, with four thresholds defined for each one, resulting in five classes (Fig. 10).

Alone, both elevation and latitude thresholds were highly ineffective as permafrost predictors. LST, NDVI, and NDSI, on the other hand, could replicate some of the correct spatial distribution of permafrost. Of these, LST performed the best (Fig. 11).

3.1.2. Supervised Classification

The next method attempted to combine the input of multiple variables at a time, with the hope that this would provide greater nuance and result in better classification overall. In order to this, we used supervised classification techniques to train an algorithm to recognize the signature of permafrost within the non-permafrost data that it was given. We trained this classifier against the 1997 reference map using the euclidean minimum-distance method, again using climate data from 2001-2004. When LST, NDVI, and LST were combined in this way, the accuracy of the model improved. With latitude included, the accuracy increased further. At this stage, including

Fig. 9) Per-Variable Pixel Distributions by Permafrost Region

Error bars show one standard deviation from mean in each category.

elevation did not improve the results. As such, the map and confusion matrix of the LST, NDVI, NDSI, and Latitude trial are included, along with those of the one-variable threshold trials from the previous section (Figs. 10, 11).

These preliminary results indicated that LST was the main indicator of permafrost extent, but that the other variables could still provide extra information that helped the model. It was initially unclear how valuable the contributions of NDVI and NDSI would be in addition to LST data. Given the high correlations explored above in the data section, they were expected to be fairly redundant. However, NDVI and NDSI actually proved to be complementary to LST, despite each performing worse individually. This suggested that the two variables might capture additional aspects of the physics involved in permafrost formation. The physical reason for latitude's effectiveness in the model was less clear, since its 'signature' should have been picked up by the other variables (i.e. higher latitudes = lower insolation = lower LST & NDVI, and higher NDSI). Elevation should have operated similarly, acting as more of a forcing agent on the other variables, and the reason for its detrimental impact on the model's accuracy was unclear.

Fig. 10) Predictive Permafrost Map — Initial Test Trails

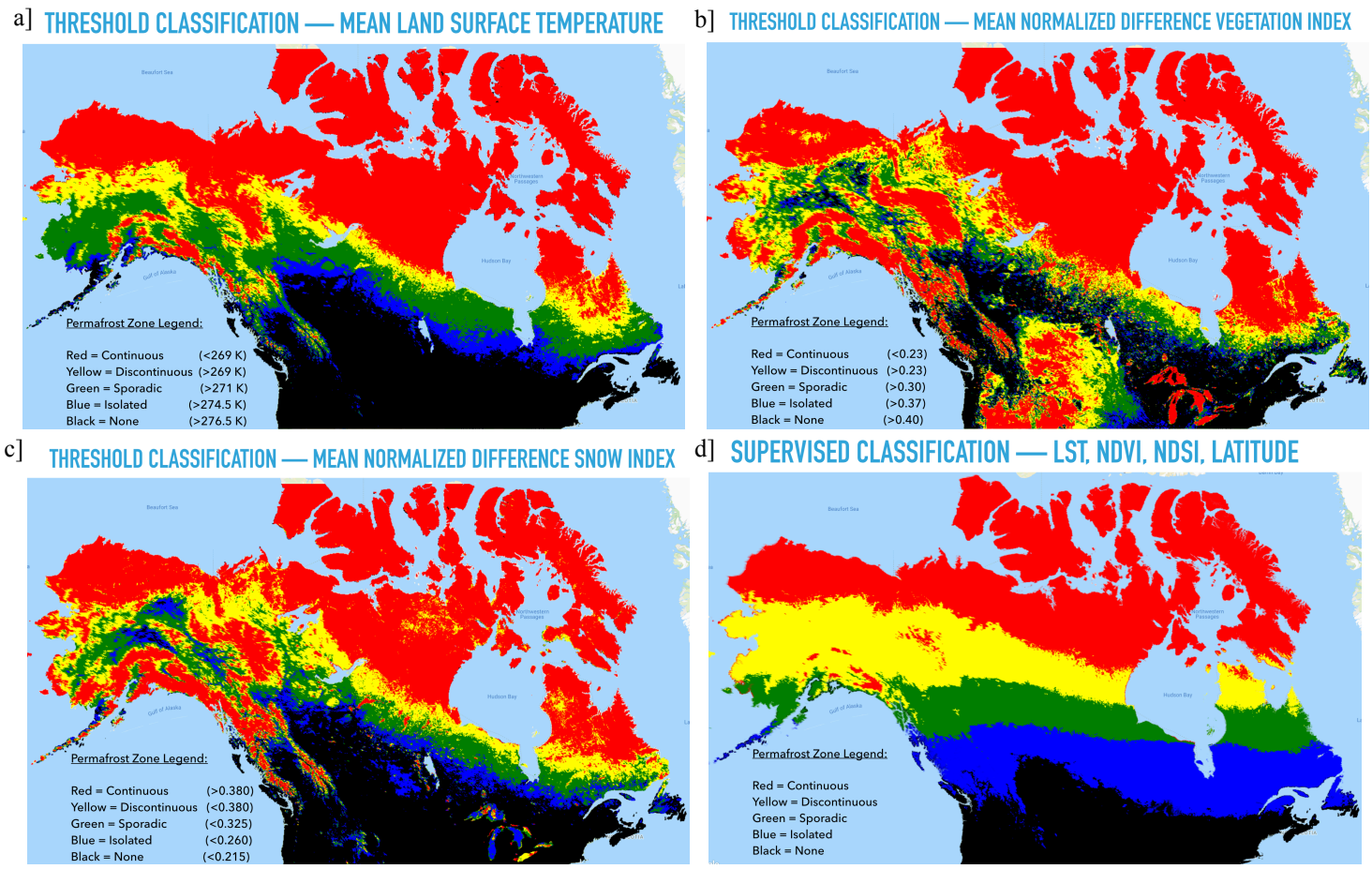


Fig. 11) Confusion Matrices
L = Land, I = Isolated, S = Sporadic, D = Discontinuous, C = Continuous.
Sampled pixels are placed in columns based on actual class, and rows based on predicted class. Therefore any pixels outside of the main diagonal have been misclassified.

LST-based Classifier

	L	I	S	D	C
L	904	337	4	3	0
I	10	117	69	42	0
S	0	111	94	70	2
D	0	5	96	250	12
C	0	6	20	131	661

Accuracy: 0.688
Kappa: 0.589
Consumers:
L: .989
I: .203
S: .332
D: .504
C: .979
Producers:
L: .724
I: .491
S: .339
D: .688
C: .808

NDVI-based Classifier

	L	I	S	D	C
L	754	112	93	189	100
I	87	43	39	53	16
S	63	65	46	64	39
D	26	71	66	131	69
C	0	15	34	125	644

Accuracy: 0.549
Kappa: 0.394
Consumers:
L: .810
I: .140
S: .165
D: .233
C: .741
Producers:
L: .604
I: .180
S: .166
D: .360
C: .787

NDSI-based Classifier

	L	I	S	D	C
L	1062	165	3	4	14
I	39	83	39	54	23
S	1	143	33	35	65
D	0	99	68	120	76
C	0	9	27	93	689

Accuracy: 0.674
Kappa: 0.553
Consumers:
L: .963
I: .166
S: .194
D: .392
C: .794
Producers:
L: .850
I: .348
S: .119
D: .330
C: .842

Supervised Classifier

	L	I	S	D	C
L	905	312	26	5	0
I	11	180	46	1	0
S	0	45	195	36	1
D	0	11	49	291	12
C	0	0	24	150	644

Accuracy: 0.752
Kappa: 0.674
Consumers:
L: .988
I: .328
S: .573
D: .602
C: .980
Producers:
L: .725
I: .756
S: .703
D: .801
C: .787

3.1.3. Validation and Model Improvement

3.1.3.1. Internal Model Improvements

After testing the viability of both methods of classification, the supervised classification model's prediction capabilities appeared to be more desirable than those of the threshold-based model. The next stage of the project was therefore to 1) improve the internal workings of the supervised classification model and 2) attempt to understand the physical nature of the key variables' influences on it. The first step involved applying a more extensive water mask than the one used in the original reference map. Section 2.1.1. of this report describes this mask, which was applied consistently for the remainder of the study. This was particularly important for reducing error caused by the Great Lakes, which were left unmasked by the original reference map (included in the "land" class since they lay outside the permafrost zone), and were therefore consistently misclassified by the model. The original classification trials described above were redone with the new mask, and accuracy increased across the board.

The next thing done was to chose the best classification algorithm for the supervised model. The initial trials had been non-normalized, euclidian minimum distance classifications[2]. As such, we normalized the datasets in order to un-skew the calculations of euclidian distance. This was done by multiplying LST values by a factor of 0.02, latitude by 0.0028, and elevation by 0.00033, in order to reduce their ranges to that of the relevant values of NDVI and NDSI. In other words, all the data was now scaled roughly to values between 0 and 1, in order to prevent high-value or high-range inputs (such as LST, typically ranging from 230-320ºK) from dominating relatively less-dynamic inputs (such as NDVI, ranging between just -1 and 1) in the euclidian space. This dramatically improved the results. In particular, normalizing the data turned elevation into a useful variable, such that most of the classification performed from then on was improved by the inclusion of both latitude and elevation (which we will call "spatial" variables, as opposed to the "measured" variables). Next, various trials were performed using the mahalanobis maximum likelihood classification[3]. This method improved accuracy and kappa values significantly compared to euclidian minimum distance, but resulted in very visible misclassification of various permafrost zones as (perennially unfrozen) land. Despite this flaw, mahalanobis was chosen as the preferred classification scheme for the remainder of the study, due to its much more robust method of normalization. Spectral angle classification was also tested, but discarded as an option because its results were not useful. With these internal improvements —a new water mask, normalized datasets, and switching from euclidean to mahalanobis classification (which also allowed for effective incorporation of elevation data)— the model's accuracy was brought up to 0.848, with a kappa value of 0.749.

3.1.3.2. Physical Model Improvements

At this stage, the focus turned to selecting the most appropriate inputs for the model that would best capture the physics of permafrost formation and thaw. To this end, experimentation

---

[2] Euclidean Minimum Distance classification plots data points with *n* input bands in an *n*-dimensional space, and assigns new points to the nearest cluster in that space.

[3] Mahalanobis classification uses a similar *n*-dimensional space, but includes built in normalization. In addition, it considers variances and co-variances of variables within clusters, to determine the percentage likelihood that a new point belongs to a cluster or not.

began with the datasets that aimed to utilize their seasonality more fully. The first such strategy was to convert the 8-day composite LST data into “thawing degree days” (TDD). This was done by taking the difference above zero degrees Celsius of every pixel in every LST image, and multiplying by 8 to convert to degree days. All negative values (from days where temperature was below zero) were set to 0. Then, for every pixel in the same location, these degree days were summed over three years (2001-2004, to be consistent with the date range of previous trials) to find the total number of thawing degree days— a measure of how much heat energy may have been absorbed over that span of time. Classification was then performed using this variable (TDD) in lieu of LST. When this was done alongside the other measured and spatial variables (i.e. with NDVI, NDSI, latitude, and elevation), TDD performed (marginally) better than LST, raising overall accuracy from 0.848 to 0.852, and kappa from 0.749 to 0.754.

The next strategy attempted to use standard deviation of LST rather than the mean, in order to make a classification based on seasonality instead of temperature alone. This was done over the same three year period as above, in conjunction with the same measured and spatial variables. The resulting overall accuracy and kappa values were nearly as good as the corresponding TDD trials: 0.852 and 0.753, respectively.

The third strategy was to run similar classifications as those attempted above, but using only data from either the winter months DJF, or the summer months, JJA. This was expected to reduce the noise that an entire year’s worth of data produces, and hone in on just the changes that mattered most (warming extent in summer months, and cooling extent in winter months). These time filters were applied across the board to all the measured variables, and classifications were run using each of the three LST-based variables (mean, standard deviation, and TDD), all alongside the spatial variables. Using TDD in either winter or summer produced worse results than with full-year data. For mean LST, using just summer months worsened results, but winter produced very good results with an accuracy of 0.860 and kappa of 0.764. Lastly, using standard deviation of LST in the summer was worse as well, but the winter trial produced the best results yet seen with overall accuracy 0.873 and kappa 0.785.

While some of these input modifications improved the model's performance, we decided not to use them for our eventual analysis. This is because the improvements were only marginal, and the mixing of spatial and temporal signals would have made the trends difficult to interpret.

3.2. Data Collection and Analysis

In the previous section, we trained our classification and prediction model on data from 2001-2004, the earliest three full years of MODIS data available. The model produced a map of inferred permafrost extent for those years, which matched the 1997 reference map with 85% accuracy. In this section we describe the second part of our study, in which we applied our model to the rest of the available data, in order to generate inferred maps of permafrost extent for each year from 2001 to the present. We used a three-year sliding window to reduce noise, such that the latest map generated uses data from the start of 2017 to the start of 2020. The inputs used for the model were mean LST, mean NDVI, mean NDSI, latitude, and elevation— all treated with a water mask. The model used the mahalanobis maximum likelihood classification method to assign pixels to their permafrost extent classes. We then analyzed the raw data alongside the inferred permafrost extent maps to identify and examine any long term trends.

3.2.1. Clustering to Define Regions of Analysis

We chose to organize our trend analysis by region, in order to identify zones which had similar characteristics and trends. Rather than attempting to do this by hand through visual analysis, we employed unsupervised classification (or "clustering") to more rigorously sort the data, and find phenomena within the data that we may not have noticed otherwise.

Clustering works similarly to the supervised classification described in section 3.1.2., plotting each data point in an *n*-dimensional space —where *n* is the number of input bands— and assigning it to a class (or "cluster") defined within that space. The difference is the lack of training data with which to define those classes. Instead, it defines its own automatically with a

Fig 12.) Clustering Regions, produced from flattened time-series of LST-means

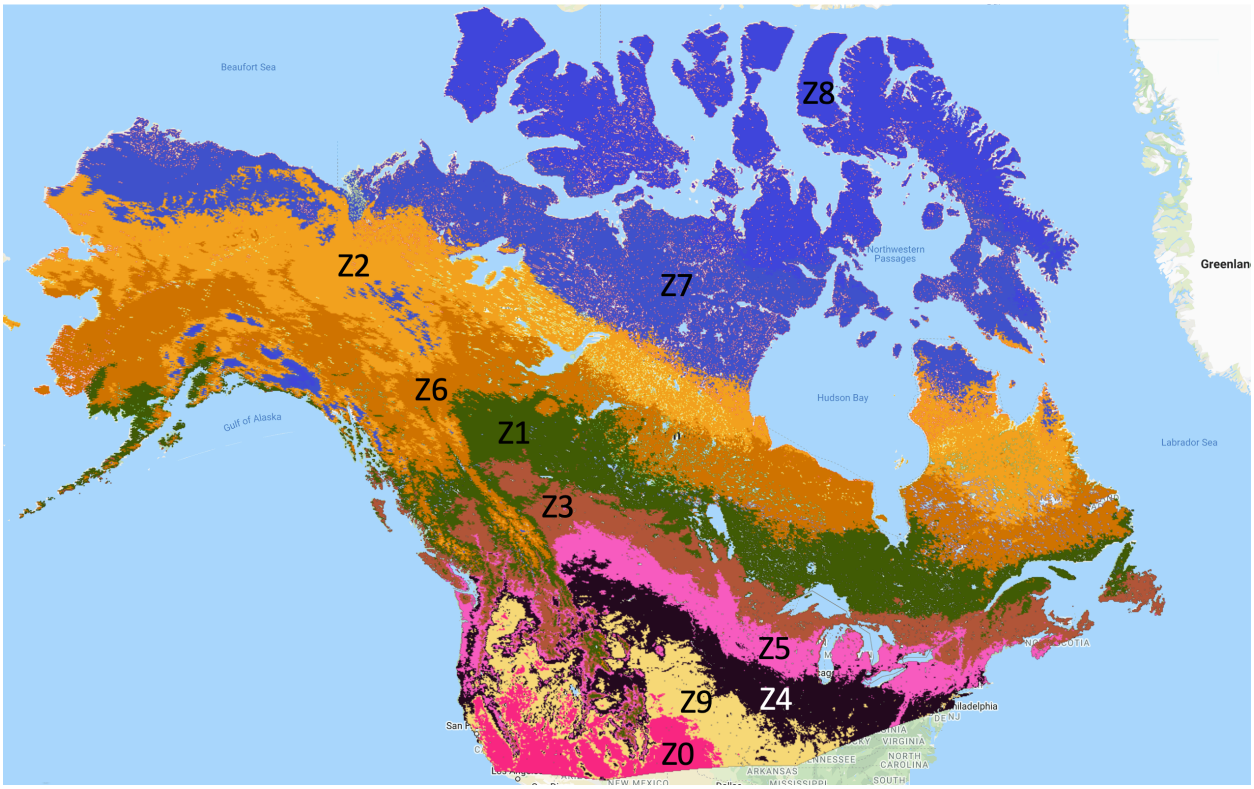
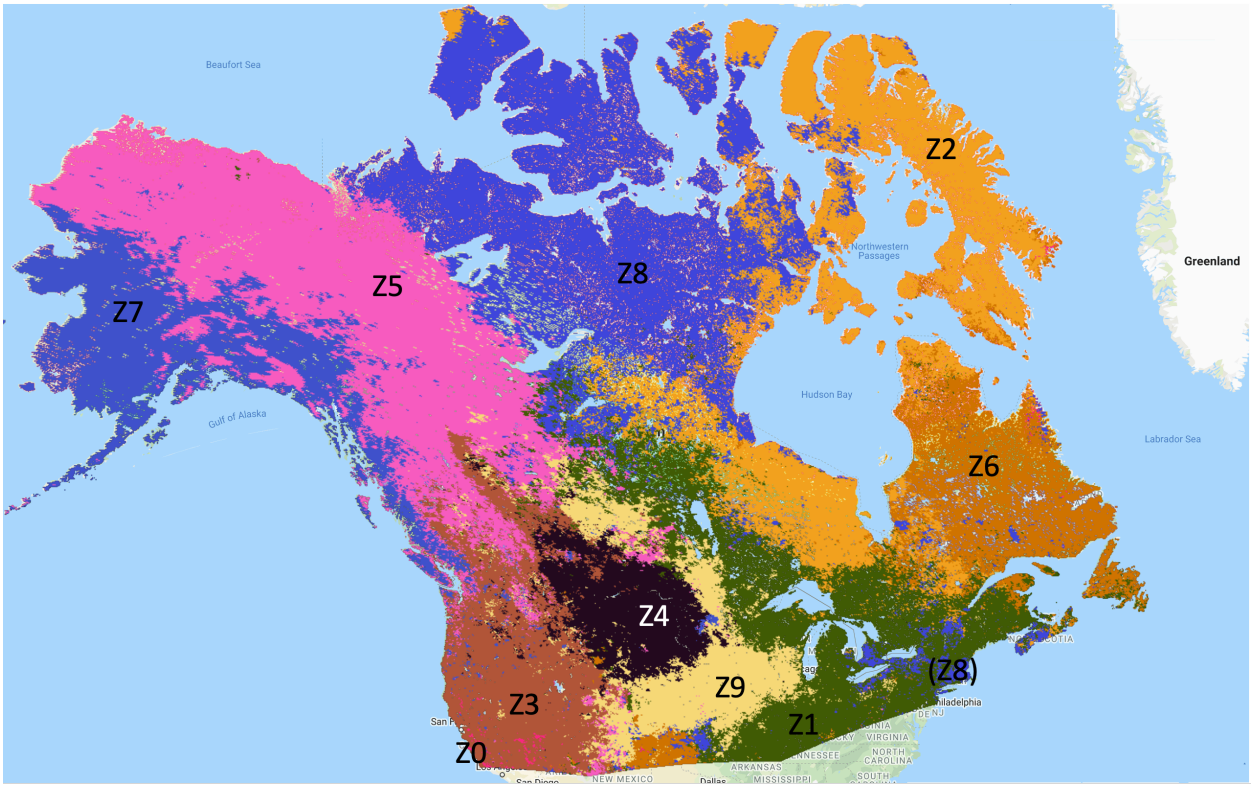


Fig 13.) Clustering Regions, produced from flattened time-series of LST-anomalies



clustering algorithm. We used the K-means algorithm[4] available in GEE[5], choosing to set $k = 10$ in order to define 10 classes— double the number of permafrost extent categories defined earlier.

We used clustering to define regions based on variations in LST data, as the most salient variable with regards to permafrost. We tried the clustering two different ways, differentiated by the input data we gave to the clusterer in each method. For the first method, we calculated a series of 17 mean-LST images, one for every 3-year window (starting with 2001-2004 and ending with 2017-2020, as with the permafrost maps). We then stacked them all on top of each other into one 17-band image, so that each pixel of this new image would contain data from the entire time-series across its bands. This final image is what was given to the clusterer, and resulted in classes defined mostly by differences in their mean LST values (Fig. 12, 14a).

For the second method, we followed the same procedure, but after calculating yearly LST means, we subtracted these from the 20-year mean to get yearly LST anomalies. Thus, the bands of each pixel in the 17-layer image contained *only* information about change over time at that point. The resulting classes were thus defined as areas with similarly shaped LST trends (Fig. 13, 15a), rather similar mean LST values. With this method, one of the classes that was generated was extremely small, and not distinctive enough to warrant any conclusions as to why this was the case. As a result, we neglected the region 'Z0' from the rest of our analysis.

4. Results

Our study was focused on analyzing trends in twenty years of LST, NDVI, and NDSI raw data, as well as trends in the maps of inferred permafrost extent that we produced. The raw data is easily accessible online in the GEE catalogue. We have included the image time-series of our

---

[4] The k-means algorithm picks k data points to become the initial centroids of k classes, assigning all other data points to the class with the nearest centroid. Once the classes have been established, it takes the mean of the data points within each class to determine new centroids, and then assigns all the data points to new classes based on these centroids. This repeats iteratively until the centroids no longer move, at which point the classes have been optimized to have the minimum amount of intra-class variance.

[5] The k-means clusterer available in GEE actually uses "k-means++" (Arthur & Vassilvitskii, 2006), which is a variation on the standard k-means algorithm that optimizes the initial choice of centroids. This ensures that every time the clusterer is run on the same inputs, the same outputs are achieved.

generated maps of permafrost extent in appendix A. The results of our regional analysis of this data are plotted in the rest of this section.

4.1. Raw Data Trends By Region

Fig 14.) Raw Data Trends — *Regions defined by clustering on LST means*

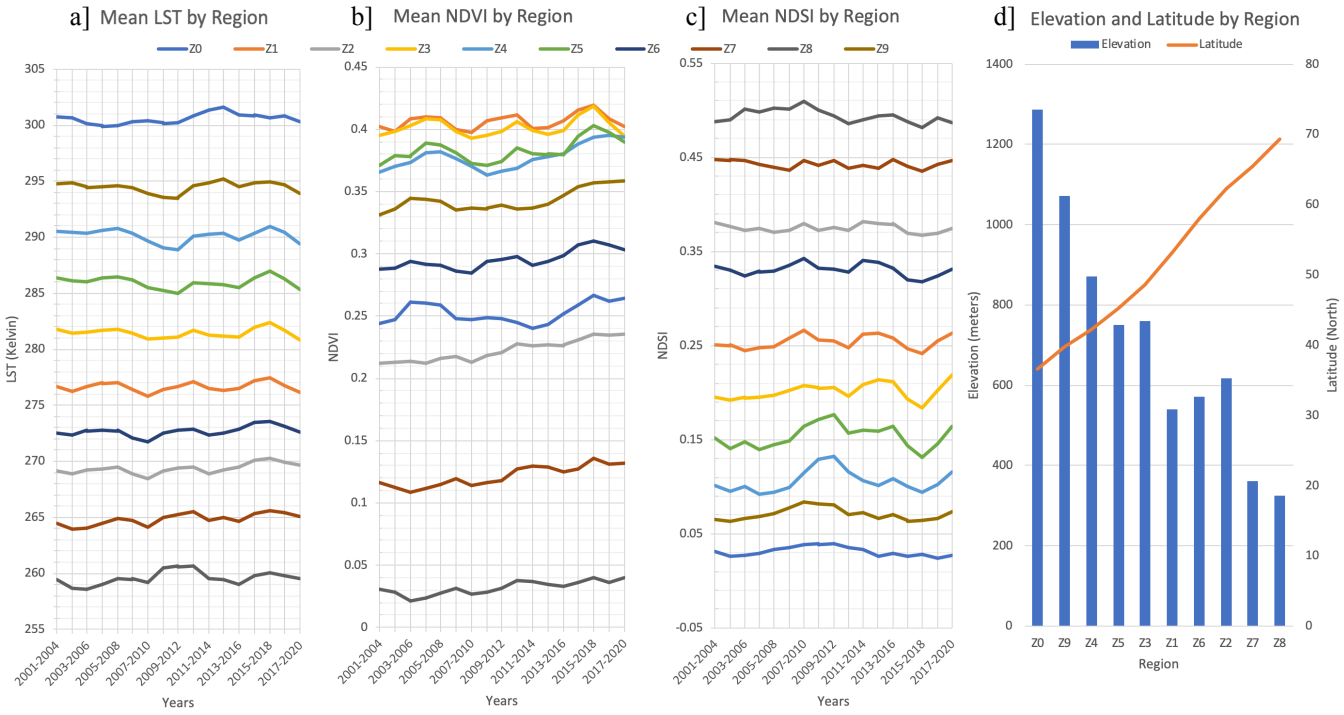
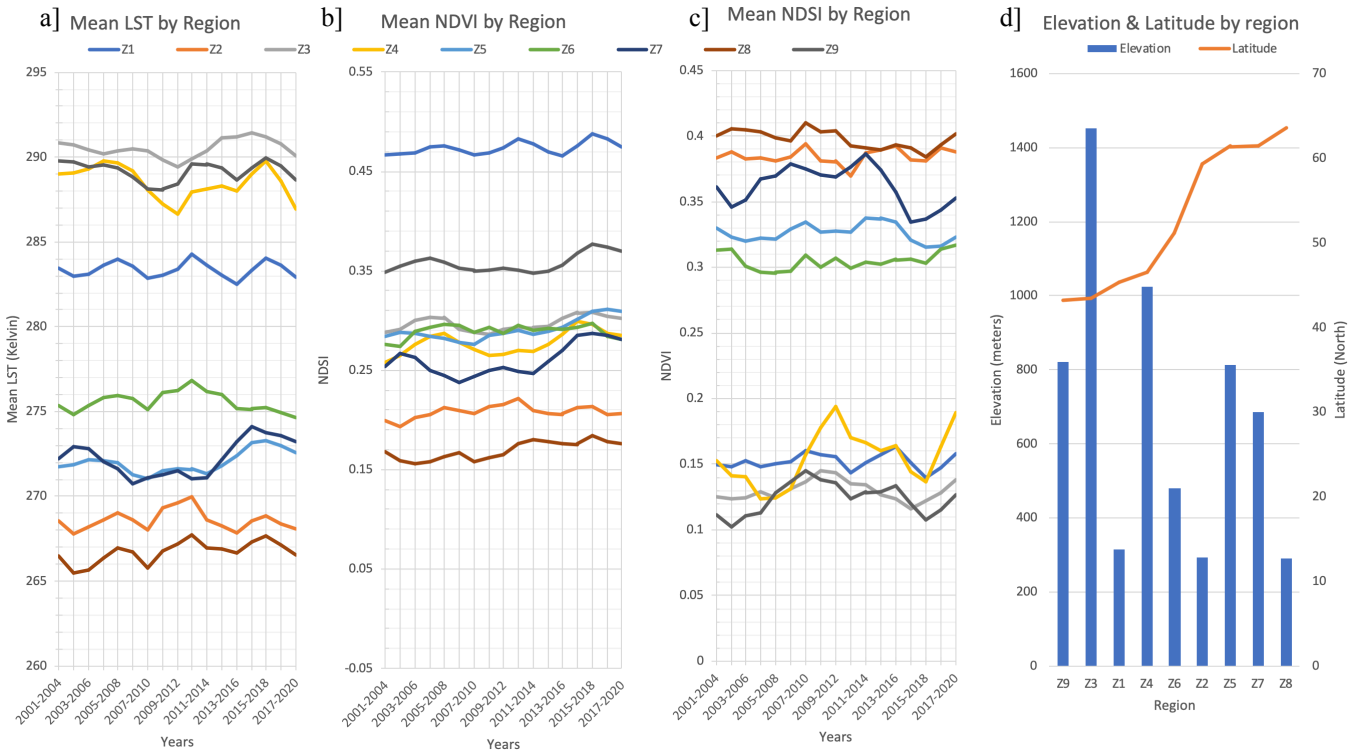


Fig 15.) Raw Data Trends — *Regions defined by clustering on LST anomalies*



4.2. Permafrost Trends By Region

Fig 16.) Permafrost Trends — *Regions defined by clustering on LST means*

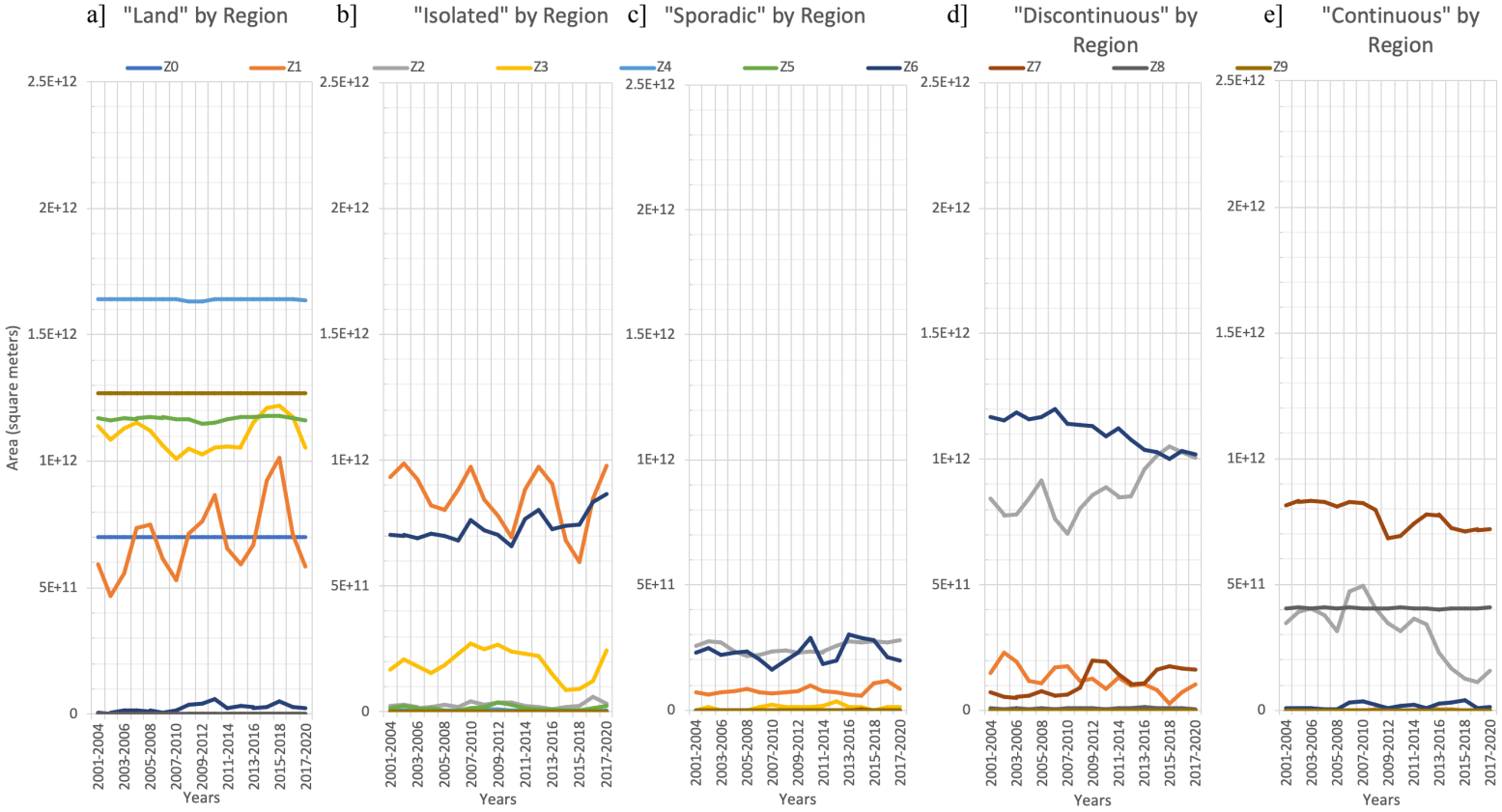
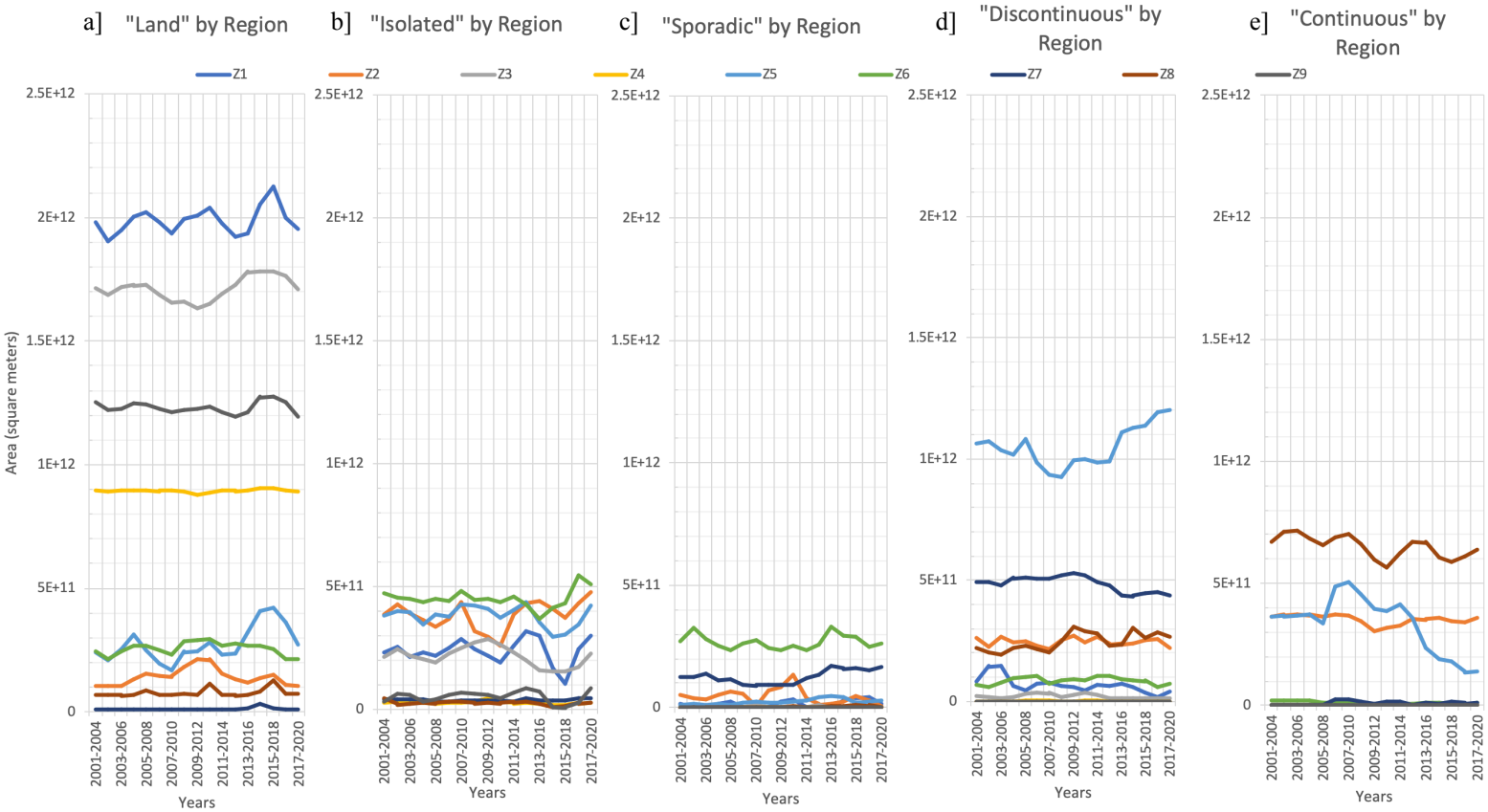


Fig 17.) Permafrost Trends — *Regions defined by clustering on LST anomalies*



5. Discussion

5.1. Efficacy of Approach

One of the main goals of this study was to determine the feasibility of predicting permafrost extent based on satellite-derived measurements. The results of the supervised classification approach we designed and utilized above were promising in this regard. Both in terms of the accuracy and kappa statistics, and in terms of the visual quality of the maps produced, the performance of the model was satisfactory.

The statistics measured the extent to which the model was able to reproduce the original reference map that it was trained off of, and showed an overall accuracy of 84.8%, with a kappa value of 0.749. This brings the model into the range of substantial (but not strong) correlation with the reference map. By adjusting the inputs to the model, better correlation was achieved (maximum accuracy achieved was 87.3%, and kappa 0.785), although these versions of the model were not utilized due to the increased difficulty they would have introduced in interpreting the results (because of mixed spatial and temporal signals). These versions did, however, demonstrate the possibility of continued future improvements to the model.

The visual results seen in the model-generated maps of inferred permafrost extent were also promising. At a glance, they showed that the model was indeed staying relatively true to the original spatial distribution of permafrost classes seen in the reference map. Furthermore, when applied to subsequent years, the maps produced by the model continued to have similarly spaced (and similarly contiguous) classes as it did when applied to the first (training) year of data. This indicated that the model was not overly trained, and that it was behaving as it was supposed to.

The model did seem to have particular trouble conflating “sporadic” permafrost with “discontinuous” (and “isolated”) permafrost, as is visible in the encroachment of the yellow (and blue) zones into what should be the green zone on the reference map.

5.2. Interpretations and Implications of Trends

The results of the unsupervised classification to find regions of interest and the analysis that followed highlighted first and foremost the highly localized, and non-generalizable nature of environmental responses to climate change. That the two sets of regions —one defined by mean values and the other by trend shapes— did not line up with each other suggests that areas with ostensibly similar mean climates can have varying sensitivities to climatic forcing. The regions defined by mean were stacked mostly horizontally (Fig. 12), which reflects latitude's impact on LST. The regions defined by trend, however, cut across each other more vertically and in pockets (Fig. 13), possibly suggesting an influence of continentality. Elevation is also very different between trend-defined regions, suggesting that geographic features like mountain ranges or simply altitude also influence how a region is affected by climate change.

Plotting the raw data revealed surprisingly few, or at least weak, long term trends, but did shed light on the relationships between some of the key variables. The inverse relationship between LST and latitude was well known and expected, but confirmed here by observing that regions with higher mean LST had lower mean latitudes (Fig. 14a,d). Similarly, regions with high LST had low NDSI, and temporal changes in LST were reflected in upside-down mirror by temporal changes in NDSI (a,c). Both of these correlations were observed again even when the regions were not themselves defined by mean value (Fig. 15a,c,d), with some exceptions for latitude that seem to be explained by differences in elevation. Overall, NDVI was distributed among regions less systematically than the other variables.

Permafrost trends were analyzed by plotting the land area covered by each class of permafrost extent within each region. It should be noted that not every region was the same size, so the vertical stacking of lines in Fig. 16 & 17 does not necessarily indicate which regions were predominantly covered by one or another class of permafrost extent. For instance, region Z3 appears above Z0 in Fig. 16a, indicating that it has more "land" pixels, even though Z3 is actually north of Z0 (Fig. 12) and contains "isolated" and "sporadic" pixels (Fig. 16b,c), while Z0 is purely "land."

Both Fig. 16 & 17 show instances of permafrost in certain regions shifting up a class (progradation) or down a class (degradation). Progradation is analogous to net freezing and an increase in percent permafrost cover, while degradation is analogous to net thawing and a decrease in percent permafrost cover. These processes can be observed whenever the contours of a particular region's trend line in one class are mirrored upside-down by the same region's trend line in an adjacent class. This mirroring is particularly visible in Fig. 16 with regions Z1 & Z3 on the border between "land" and "isolated," and with regions Z2 & Z7 on the border between "discontinuous" and "continuous."

Due to the regions being sorted by increasing LST (and latitude), Fig. 16 proves very insightful in showing us the thermal and geographic flashpoints for permafrost thaw. Z1 and Z3 are adjacent to each other (Fig. 12), and are the only two classes to exhibit change in the "land" and "isolated" categories, showing that soil reaches the threshold for permafrost formation within their geographic bounds. The mean LST associated with these regions is between 275 and 282ºK, on the cusp of freezing. The trend of the changes indicate degradation at this boundary. Similarly, the adjacent regions of Z2 and Z7 are the flashpoint for the degradation of "continuous" to "discontinuous" permafrost, with an average LST range of 264 to 270ºK. On the other hand, the boundaries between "isolated," "sporadic," and "discontinuous" are less clear for two reasons. The first reason is that permafrost has two options at these boundaries, progradation or degradation, so the mirroring is less obvious, as the signal may be split. The second reason is that these permafrost classes were the most likely to be conflated with each other, according to visual analysis of the model's training trial with the original reference map. Despite this, evidence of degradation is discernible in region Z6, where permafrost goes from "discontinuous" to "isolated," skipping "sporadic".

Fig. 17 does not provide as many interpretable insights, but does give further evidence to the support the observation that "continuous" permafrost zones, in particular, are degrading heavily into "discontinuous" zones.

6. Summary

Our study aimed to develop a methodology with which to map permafrost using satellite-derived measurements, and to examine trends in the past twenty years of this data to understand how permafrost and related environmental factors have responded to ongoing climate change. Using supervised classification, we were able to develop a model that reproduced its training data with an overall accuracy of 85% and kappa value of 0.749, with demonstrated room for improvement. We applied this model to the remaining data to generate yearly maps of inferred permafrost extent from 2001 to the present.

Defining regions of interest to organize our analysis, we found that regions with similar means do not necessarily have similar trends. Analyzing our raw data region by region highlighted the strong inverse relationships between LST and NDSI and between LST and latitude, while NDVI and elevation remained less systematically correlated to the other variables. Analyzing our inferred permafrost data by region helped identify "flashpoints" for permafrost degradation (net thawing) and progradation (net freezing) among regions defined by mean LST. Degradation from "isolated" permafrost to "land" (0% permafrost) was seen roughly between the 275 and 282ºK isotherms, while degradation from "continuous" to "discontinuous" permafrost was seen roughly between the 264 and 270ºK isotherms.

Our results offer a proof of concept for our permafrost mapping methodology, which future work can improve upon. Using more advanced satellite imagery, including higher resolution imagery (Duguay et al., 2005) or ground penetrating sensors (Whitley et al., 2018), the accuracy of the model can be improved. If our results are robust, this methodology can easily be extended to be useful to climate scientists studying permafrost-thaw based greenhouse gas emissions. With some extra calculation, our results can be used to quantify the area amount of permafrost that has thawed or frozen in the past twenty years, and from there —with further research on air-soil gas fluxes— this information can be utilized to determine how much greenhouse gas emissions may have occurred during this time period. This will help climate scientists better understand how to incorporate the effects of permafrost thaw into predictions of future climate change, which will in turn help shape policy aimed at reducing carbon emissions by the required amount.

7. Acknowledgements

First and foremost, I would like to thank my advisor, Ron Smith, for sharing with me his time, statistical and climate-related knowledge, and general wisdom— without which none of this work would have been possible. A special thanks to the YCEO lab at Yale for giving me the resources and platform to conduct my research, and in particular to TC Chakraborty for some critical assistance early on teaching me to use Google Earth Engine. I would like to thank everyone else in the Geology & Geophysics department who helped make my education possible, especially my DUS Mary-Louise Timmermans for four years of encouragement. Lastly, a warm thank you to all my friends at Yale, for keeping me sane and motivated, and to my parents, for their love and support during the time I spent working on this project and also during the entirety of the rest of my life.

8. References

Arthur, D., & Vassilvitskii, S. (2006). *k-means++: The advantages of careful seeding. Stanford.*

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., ... & Allard, M. (2019). *Permafrost is warming at a global scale.* Nature communications, 10(1), 264.

Brown J, Ferrians OJ Jr., Heginbottom JA, and Melnikov ES 1997 *Circum-Arctic map of permafrost and ground ice conditions*. US Geological Survey Reston, CP-45.

Duguay, C. R., Zhang, T., Leverington, D. W., & Romanovsky, V. E. (2005). *Satellite remote sensing of permafrost and seasonally frozen ground.* GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION, *163*, 91.

Kenner, R., Noetzli, J., Hoelzle, M., Raetzo, H., & Phillips, M. (2019). *Distinguishing ice-rich and ice-poor permafrost to map ground temperatures and ground ice occurrence in the Swiss Alps.* The Cryosphere, 13(7), 1925–1941. doi: 10.5194/tc-13-1925-2019

Li, A., Xia, C., Bao, C., & Yin, G. (2019). *Using MODIS Land Surface Temperatures for Permafrost Thermal Modeling in Beiluhe Basin on the Qinghai-Tibet Plateau.* Sensors, 19(19), 4200. doi: 10.3390/s19194200

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., ... & Khomutov, A. (2019). *Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale.* Earth-Science Reviews.

Ran, Y., Li, X., Jin, R., & Guo, J. (2015). *Remote sensing of the mean annual surface temperature and surface frost number for mapping permafrost in China*. Arctic, Antarctic, and Alpine Research, 47(2), 255-265.

Schuur, E. A., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., ... & Natali, S. M. (2015). *Climate change and the permafrost carbon feedback*. Nature, 520(7546), 171.

Van Everdingen RO 1998 *Multi-language glossary of permafrost and related ground ice terms*. International Permafrost Association, National Snow and Ice Data Center, University of Colorado, Boulder.

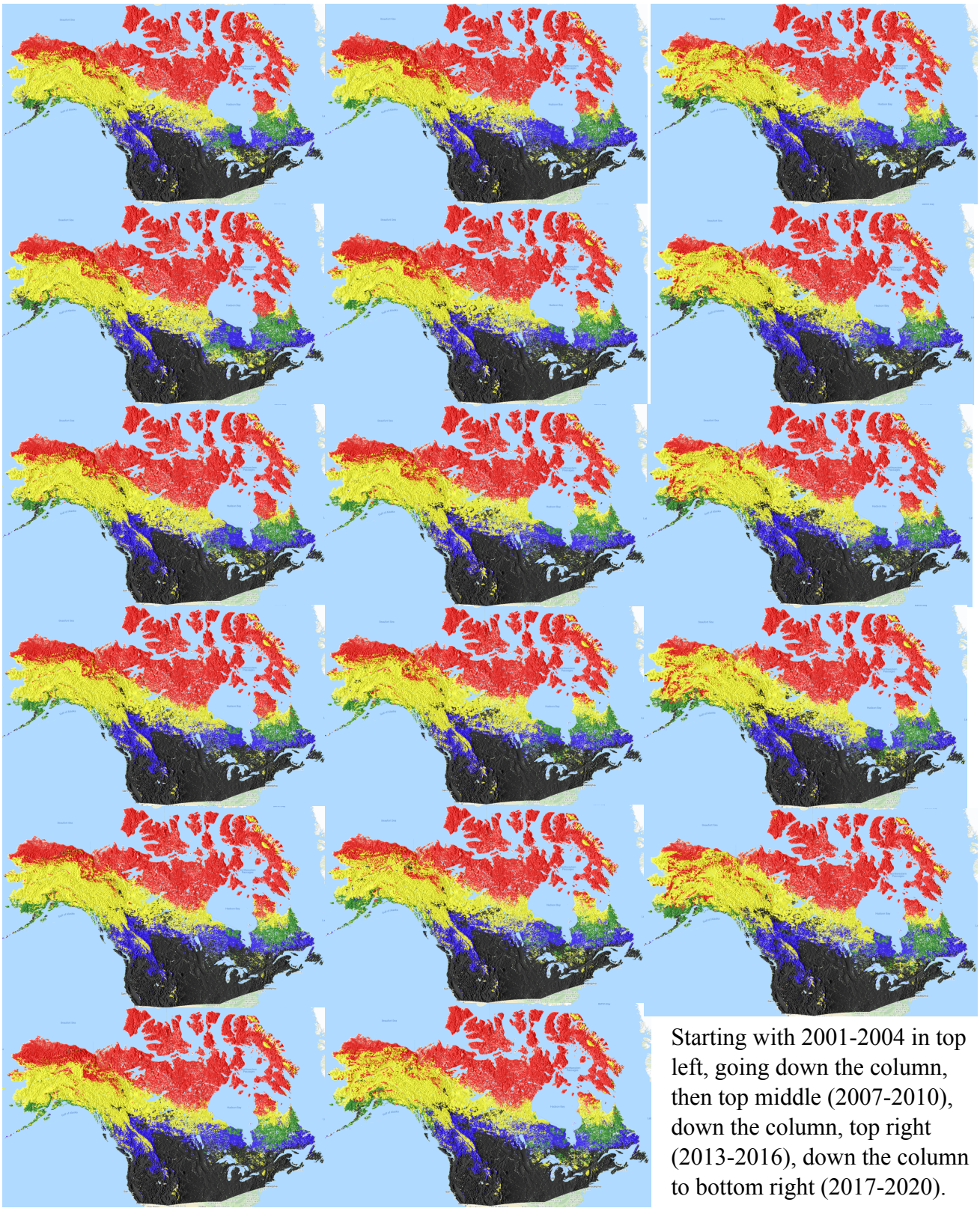
Whitley, M.A.; Frost, G.V.; Jorgenson, M.T.; Macander, M.J.; Maio, C.V.; Winder, S.G. *Assessment of LiDAR and Spectral Techniques for High-Resolution Mapping of Sporadic Permafrost on the Yukon-Kuskokwim Delta, Alaska.* Remote Sens. **2018**, 10, 258.

Zhang, T.: *Influence of the seasonal snow cover on the ground thermal regime: an overview*, Rev. Geophys., 43, 1–23, 2005.

Zimov, S. A., Schuur, E. A., & Chapin III, F. S. (2006). *Permafrost and the global carbon budget*. Science(Washington), 312(5780), 1612-1613.

9. Appendices

9.1. Appendix A: Image Time-Series: Permafrost Extent



Starting with 2001-2004 in top left, going down the column, then top middle (2007-2010), down the column, top right (2013-2016), down the column to bottom right (2017-2020).