CanCoast 2.0 Data Analyses: National Scale Model Sensitivity

Prepared by
Tim Webster PhD, Thomas Allen
Applied Geomatics Research Group
NSCC, Middleton
Tel. 902 825 5475
email: tim.webster@nscc.ca

Submitted to:
Public Services and Procurement Canada,
Geological Survey of Canada (Atlantic), GSCA

How to cite this work and report:

Webster, T., Allen, T., 2022. CanCoast 2.0 Data Analyses: National Scale Model Sensitivity. Technical report, Applied Geomatics Research Group, NSCC Middleton, NS. Public Services and Procurement Canada, Geological Survey of Canada (Atlantic), GSCA.

Executive Summary

CanCoast 2.0 is a collection of data created by the Geological Survey of Canada summarizing characteristics of Canada's shoreline that are relevant to the determination of coastal sensitivity to change over time into six variables. These data have been worked with in two papers of interest to the project, Manson et al. (2019) and Hatcher and Manson (2021). The Applied Geomatics Research Group (AGRG), Nova Scotia Community College were commissioned to reproduce and build upon the methods and results of the Hatcher and Manson (2021) paper to create a new Coastal Sensitivity Index (CSI) dataset for all of Canada using a μ-statistics approach. This CSI dataset ranks the sensitivity of a given section of the coastline with a unitless score, higher being more sensitive and lower being less sensitive compared to the other sections in the dataset. Using information from the original papers, the publicly available CanCoast 2.0 data, updated data and R and Python scripts written by Scott Hatcher and provided by Dr. Gavin Manson, and ArcGIS Pro, two new CSI products were generated using all six coastal sensitivity indicator variables. One of these new products represents the early century, or present, and the other represents late century, or future coastal sensitivity from modelled data. After these initial data were generated, an additional six datasets for each of the early and late century data were generated based on the CSI variables, excluding a different variable each time. Maps and score distribution histograms were generated using ArcGIS Pro for each of these products at a small scale and large scale, one for the entire country and one focused on the Atlantic region to obtain a more detailed look at changes across CSI products. Furthermore, fields were added to these early and late century datasets summarizing the difference between the variable excluded CSI_{μ} scores and the original six-variable scores, the variance among scores in the variable-excluded runs were calculated, and the variance field was mapped. Results were similar across both century datasets in terms of high sensitivity areas, with more change in scores for variable-excluded runs in the late century data resulting from the larger time span the data represents, allowing for more potential change. Areas of particularly high sensitivity across all runs include Banks Island, Victoria Island, and the Queen Elizabeth Islands north of Nunavut and the Northwest Territories, as well as large parts of Cape Breton around the Bras d'Or Lake in Nova Scotia.

Table of Contents

Table of Figures

1. Introduction

The purpose of this project is to replicate the methods and expand upon the results of Hatcher and Manson (2021), which itself expands upon the work done in Manson et al. (2019). These projects took spatial information which characterizes Canada's coastline and used them to calculate scores that serve as indices of coastal sensitivity. *Figure 1* and *Figure 2* show the reader the extent of this data, both at a coarse nation-wide scale and a more detailed Atlantic Canada-focused scale, respectively.

The spatial information for the variables used in the coastal sensitivity scores was obtained from the Government of Canada's CanCoast 2.0 site, where it is distributed for use as six shapefiles (one per variable), and was supplemented by updated versions of the data made available to AGRG by Dr. Gavin Manson. These variables, along with the names of their corresponding shapefiles, can be found in *Table 1*. Each of these variables were produced by assigning attributes to the national-scale shoreline vector, CANCOAST_SHORELINE_V3. These attributes were then ranked from 1 to 5 for each segment to serve as indicators of sensitivity, with higher numbers indicating a greater contribution to overall coastal sensitivity.

Variable	**Shapefile Name(s)**	**CSI_{μ} Version**
Change in Sea Level	CANCOAST_SEALEVEL2006_2020_V3	Early Century
	CANCOAST_SEALEVEL2006_2100_V3	Late Century
Significant Wave Height	CANCOAST__WAVEHEIGHTSEAICE1996_2005_V1	Early Century
	CANCOAST_WAVEHEIGHTSEAICE2090_2099V_1	Late Century
Ground Ice	CANCOAST_GROUNDICE_V1	Shared
Coastal Materials	CANCOAST_MATERIAL_V2	Shared
Backshore Slope	CANCOAST_SLOPE_V5	Shared
Tidal Range	CANCOAST_TIDALRANGE_V6	Shared

Table 1 List of variables used in CSI calculations and the corresponding shapefiles containing their data.

Both Manson et al. (2019) and Hatcher and Manson (2021) use what is known as a µ (or mu) statistics approach to calculate their coastal sensitivity scores. Higher numbers indicate more sensitive sections of coastline, while lower scores indicate that a given section of coastline is less sensitive compared to the others. As opposed to using more traditional methods such as the geometric mean to calculate sensitivity scores for this data, the µ-statistics method determines scores by ranking coastal segments relative to one another. When compared to two other commonly used methods in Hatcher and Manson (2021), the µ-statistics method both offers more sensitivity in final segment scores and less sensitivity to the propagation of error. To offer a brief outline of the process, coastal segments are first grouped by similarity in variable ranks, then groups are scored in terms of superiority – 0 for identical/unranked, 1 for superior, and -1 for inferior – and then finally these scores are added up, giving each row a final CSI_{μ} score.

To add to the work done in these two previous CanCoast papers, this project will replicate the methods for applying μ statistics scoring to the CanCoast data, then re-run the scoring function on the data a total of six times, each time excluding one of the indicator variables from the calculations. As seen in *Table 1*, there are two shapefiles for each of the "Change in Sea Level" and "Significant Wave Height" variables, corresponding to what we will refer to as "Early Century" and "Late Century" timeframes. As advised by Dr. Manson, these two versions of the variables will not overlap and are not compatible with one another, so early century and late century datasets were constructed to be scored separately, meaning the scoring methods needed to be repeated once per dataset. For the sake of clarity, *Table 1* also indicates to which CSI_{μ} Version each shapefile contributes. These two datasets and six different indicator-excluded scores will be analysed and visualized using ArcGIS Pro to investigate the effect of their exclusion on coastal sensitivities.

Figure 1 A national-scale overview showing the extents of the CanCoast Marine Shoreline 3.0 data.

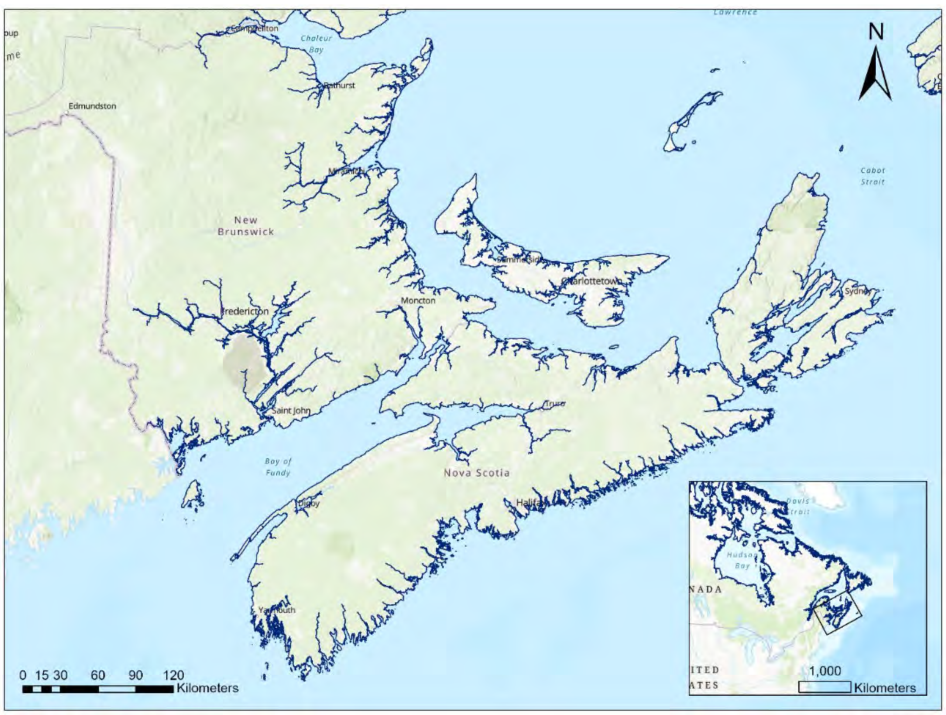

Figure 2 An Atlantic Canada-scale overview showing the extents of the CanCoast Marine Shoreline 3.0 data in greater detail.

2. Methods

2.1 Shapefile and Spatial Data Preparation

After obtaining the CanCoast shapefile data from Dr. Manson, the first step of processing was to import the data into an appropriate working environment. For the sake of ease of use and time efficiency, Esri's ArcGIS Pro version 2.9.1 was used to build a file geodatabase (GDB) and import the CanCoast shapefiles as feature classes. Working with this data within a GDB allowed for much faster processing vs working with shapefiles for data of this size and scale due to the way the data is stored and indexed spatially.

The next step after all the data has been imported was to join the six variable shapefiles based on their spatial data. CanCoast Shoreline V3 was used as a base to which all of the other variable scores were joined, as according to Manson et al. (2019), each of these variable shapefiles were originally made by assigning scores to segments of the CanCoast Shoreline V3 dataset. This means that all of the segments of the variable shapefiles line up with the segments found in CanCoast Shoreline V3. The Spatial Join tool in ArcGIS Pro was used a total of twelve times, six times on the early century dataset and another six times on the late century dataset, once for each variable.

After the joins were completed the data was inspected for unexpected values to ensure no errors were made during joining that would interfere with correct CSI_{μ} score calculation. This was achieved by opening the attribute tables for the early and late century datasets in ArcGIS Pro, then sorting the indicator variable score columns from highest-to-lowest and lowest-to-highest. Initially, score values of 0 were found in the Ground Ice indicator variable columns, but these were found to be true zero-value segments upon inspection of the source shapefile for those areas. Furthermore, there were some sections in the CanCoast Shoreline V3 vector data that did not exist in every indicator variable dataset, so they were reported as "Null" after the join. Rather than removing these areas entirely, the null values were converted to "0" for compatibility with the project's script using ArcGIS Pro's Calculate Field tool seen in *Figure 3*.

Figure 3 An example of the Calculate Field window used to convert Nulls to 0 after joining all the shapefiles together.

Finally, the last step of preparing the data for CSI_{μ} calculation was to run the Pairwise Dissolve tool in ArcGIS Pro, using the six variable score fields as the dissolve fields parameter. The Pairwise Dissolve tool was chosen over the standard dissolve tool for its improved performance in ArcGIS Pro according to Esri's documentation. The purpose of using this tool is to combine adjacent coastal segments with identical indicator variable scores into longer segments. This greatly reduced the number of shoreline segments in the joined feature class from more than 15 million down to 723,072 features in the early century dataset, and 723,727 in the late century dataset. Due to the volume of data being worked with, this reduction was very important in reducing processing time in the scripting portion of the project.

2.2 Script Work

Due to the size and nature of the data being worked with in this project, several language options were investigated in order to complete the required script, including R, C++, and Python. Of these three languages, Python version 2.7 was settled on for its ease of use and compatibility with Esri's ArcPy library, allowing for direct interaction with feature classes in geodatabases.

Initial scripting materials, provided by Dr. Manson, had been written in Python and R by Scott Hatcher. In these scripts, the portion responsible for interaction with the CanCoast data feature dataset and grouping of rows was written in Python, and the portion for the superiority and µ-scoring logic was written in R. To convert this latter portion to Python, the R script was first analysed to determine the methods and order of operations for score generation, testing on small matrices generated within the script that resembled small-scale versions of the final data to be processed. After gaining a sufficient understanding of the script, it was re-written in Python and again tested on small matrices generated within the script to ensure results were as expected based on the methods laid out in Hatcher and Manson (2021). Next, this adapted scoring script was incorporated into Scott Hatcher's original Python script, and it was tested again to ensure expected behaviour, this time on a subset of the actual data analysed in the completion of this project.

The Python script for this project and the original script it was based on makes use of what are known as "libraries" to expand upon the capabilities of the Python programming language. The three main libraries used in this project's script include NumPy, Pandas, and Esri's ArcPy. NumPy serves to add array and matrix functionality on a scale much larger than what is capable of standard Python data types. Pandas offers the DataFrame object, which is used in this script for its efficient grouping and data merging functions. Lastly, ArcPy is used for its compatibility with spatial data types and spatial data stored in spatial databases, such as the file geodatabase used in this project.

To perform the µ-scoring method, the Python script first loaded the aforementioned Python libraries, then set the system environment to point to the geodatabase containing the CanCoast data, the feature dataset, and the variables of interest to the scoring algorithm. Next, for the sake of speed, the contents of feature class were read into a NumPy array and converted into a Pandas DataFrame. This is the point at which the first step of the µ-stats scoring methods from Hatcher & Manson (2021) was applied. The rows in the DataFrame were reduced, or grouped, into rows of identical variable scores. This newly reduced data was then passed along to the function responsible for performing the superiority testing, adapted from R, where all rows were tested against one another and their scores added up to reach the final coastal sensitivity index score. This new list of scores was then referenced back to the original rows of the feature dataset, a new column to store the data was inserted, and each of the original sections were updated with their own scores. A loop was built into the script to enable it to run an additional six times on the data, to create the modified CSI_{μ} scores for each variable type after creating the first "control" dataset with all six variables included. The final script was run a total of two times, once for the early century and once for the late century datasets.

2.3 Comparing Generated CSI Values

A number of approaches were used to compare the number of different CSI_{μ} scoring results across the two century datasets. Fourteen maps were created for each century dataset using similar methodology to Hatcher & Manson (2021), one map for the CSI_{μ} score including all variables and one for each of the CSI_{μ} scores excluding a variable. Using ArcGIS Pro, the CSI_{μ} scores were classified with the Natural Breaks (Jenks) method into 5 classes and assigned labels ranging from very low to very high sensitivity. As well, histograms plotting the distribution of the scores for each CSI_{μ} score run were made with ArcGIS Pro's statistics tool to serve as another visual aid to qualitatively assess changes in sensitivity across runs.

After the initial mapping and plotting of the script generated CSI data, ArcGIS Pro's Calculate Field tool was used to generate more statistics to help characterize the spatial distribution of the data. Firstly, for each of the variable excluded CSI_{μ} scores, the difference between these scores and the original six-variable control was calculated and stored in their own fields. This allowed for the mapping of the change in CSI_{μ} scores resulting from variable exclusion. Secondly, the Calculate Field tool was used along with a Python code block seen in *Figure 4* to calculate and store the variance for each shoreline segment among variable excluded CSI_{μ} scores. The standard deviations for these segments were then calculated by simply taking the square root of variance field that had just been calculated. These new fields allowed the mapping of the coastal segments that saw the most change in CSI_{μ} scores across all six of the variable exclusion runs for the early and late-century datasets.

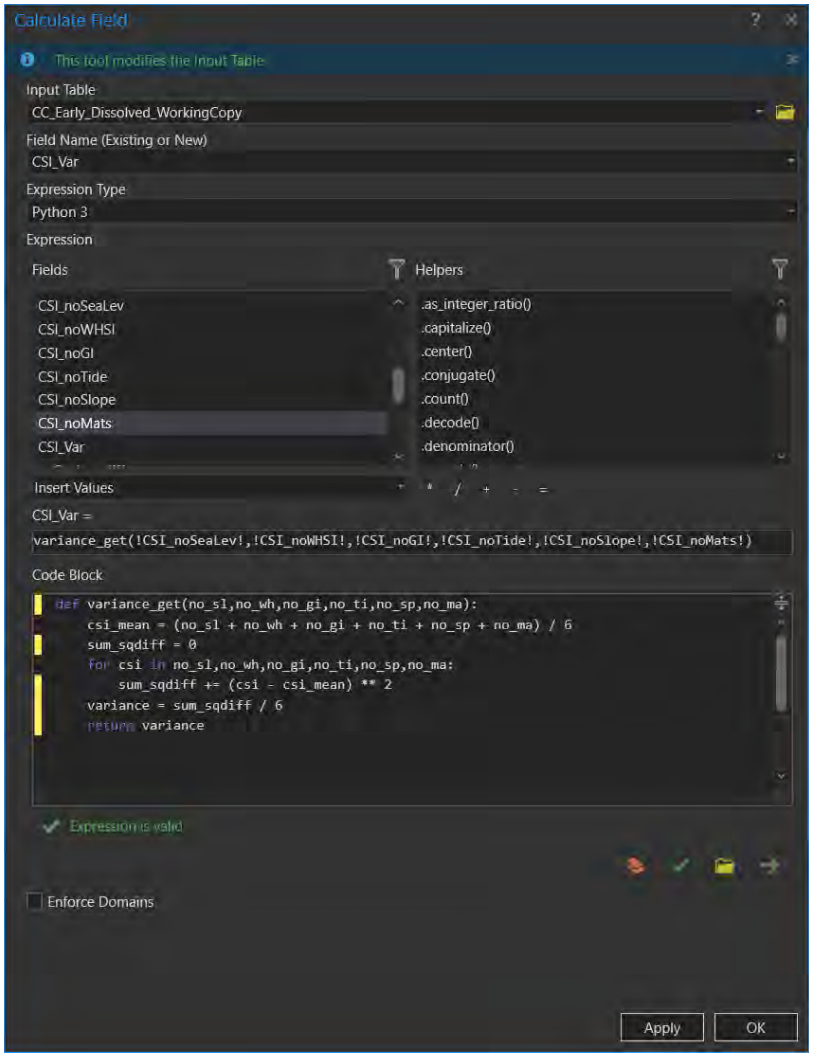

Figure 4 The ArcGIS Pro Calculate Field window with the Python code block written for calculating CSI_{μ} score variance.

3. Results

Looking through the products and data created in processing yielded some interesting results that will be investigated in greater depths in the sections to follow. As may have been expected, the exclusion of certain variables resulted in large-scale changes in the modeled sensitivity across Canada. These can be seen in both the score distribution histograms, such as the ones in *Figure 5*, and the maps, such as the one in *Figure 8*. *Figure 5* shows there are also differences in sensitivity across the two century datasets, with the standard deviation being much higher in the late-century compared to the early-century, meaning the CSI_{μ} scores are more spread out across the data rather than being more closely clustered around the mean. The results of this distribution can be seen as well in *Figure 6*, where less of the map in the late-century is dominated by one sensitivity class compared to the early-century.

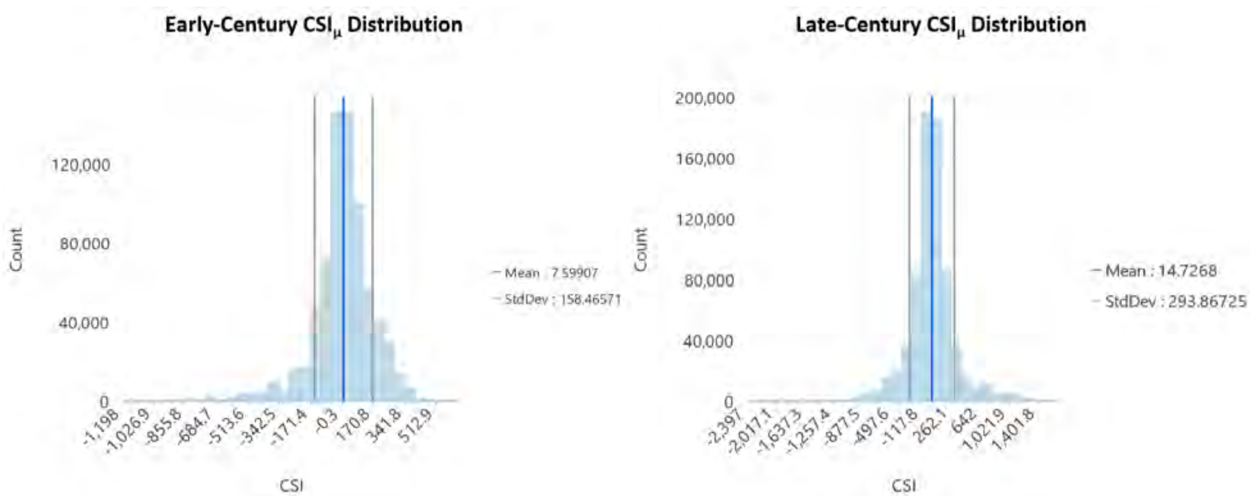

Figure 5 Comparison of the distribution of the 6-variable CSI_{μ} scores between the Early and late-century datasets.

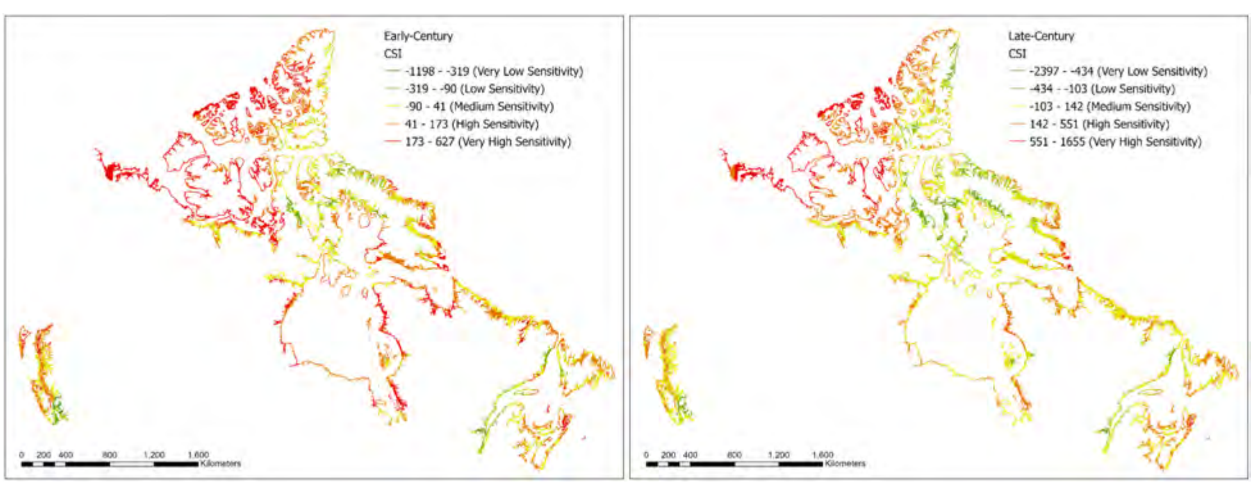

Figure 6 A national-scale visual comparison of CSI_{μ} scores across early-century and late-century six-variable datasets.

3.1 Early Century

Starting with the early-century data, one may note that the distribution for the change in sea level excluded CSI_{μ} scores are identical to the six-variable CSI_{μ} scores. This is not due to error, as in the original "CANCOAST_SEALEVEL2006_2020_V3" shapefile input data (see *Table 1*), all the indictor values were identical and of rank 3 (see *Table 2*). As a result, there was no change in CSI_{μ} scores between these two runs as that variable never had any influence on row/segment superiority. However, there was plenty of change in the rest of the variables, seen both in their histograms and spatially via their maps. Additionally, during review some areas started to stand out for their relative lack of change across runs, as can be observed in the following figures. In particular, the northwestern-most islands in the Northwest Territories and Nunavut remain largely within areas of high to very-high sensitivity across all runs, though interestingly they are also in one of the areas of highest score variance in the country (see *Figure 22*). The coasts around Cape Breton in Nova Scotia provide another example of such an area (see *Figure 23*).

FID	Shape	SLChange	SLScore	Year
0	Polyline	-13.171958	3	2020
1	Polyline	-13.171958	3	2020
2	Polyline	-13.171958	3	2020
3	Polyline	-13.171958	3	2020
4	Polyline	-13.171958	3	2020
5	Polyline	-13.171958	3	2020

Table 2 Sample data from the CANCOAST_SEALEVEL2006_2020_V3 shapefile attribute table.

It is possible to infer whether a variable had a detrimental or protective effect on the scores for an area by observing whether that area had changes from run-to-run and by comparing it to the six-variable control run. As an example, comparing *Figure 15* vs *Figure 9* shows how much more of Nova Scotia is now in the high to very-high sensitivity range after excluding ground ice ranks from the scoring method. With ground ice ranks included, the area will have a great deal more low-ranking indicators in comparison with areas with ground ice such as those farther north, keeping it lower in relative sensitivity. We can see that the opposite applies as well from the slope-excluded run, with sensitivities being lower in *Figure 19* than the control run. This would suggest that slope is a significant factor in contributing to sensitivity for this area as its exclusion improves CSI_{μ} scores.

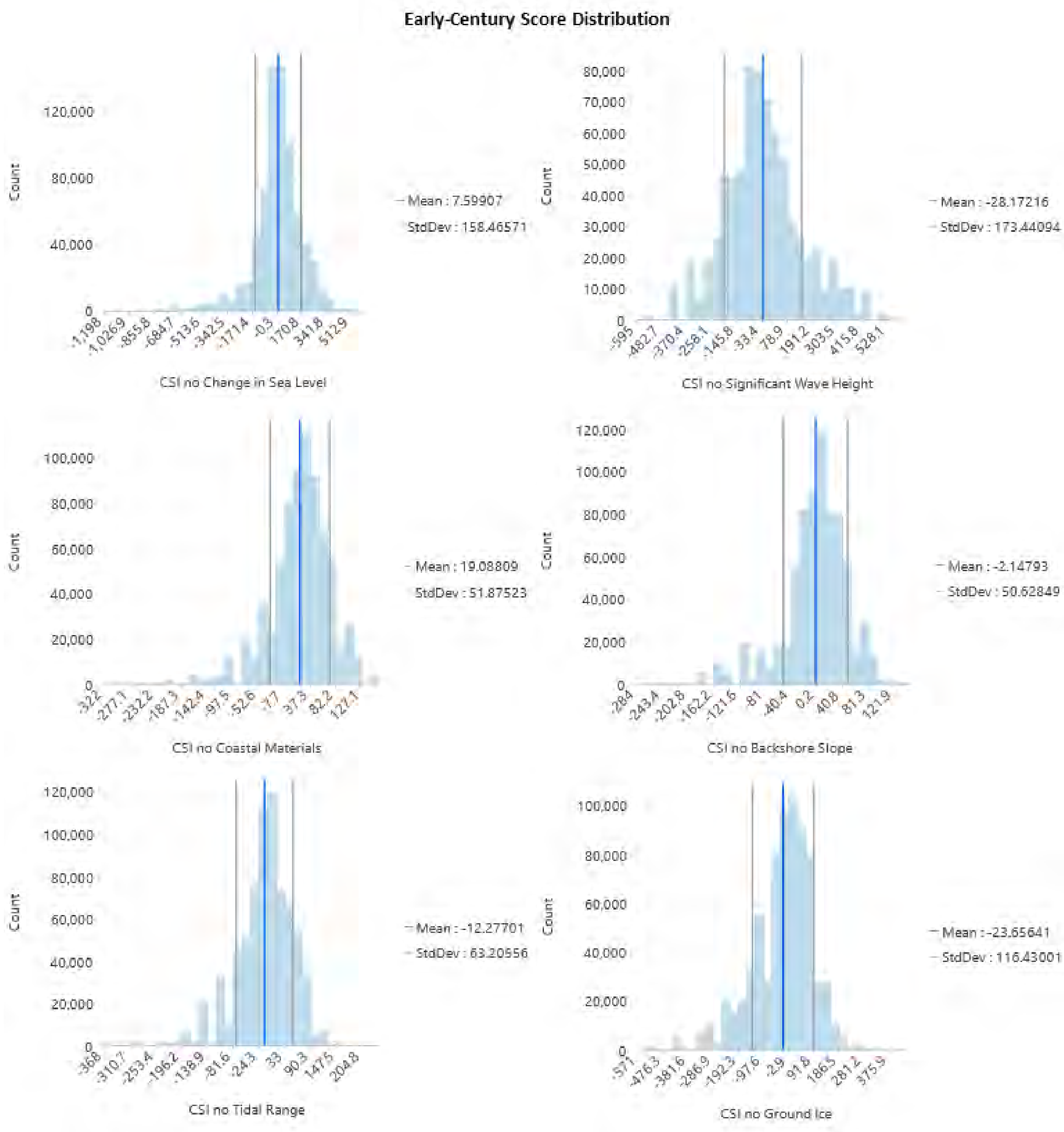

Figure 7 Plots of the distributions of CSI_{μ} scores among the variable-excluded runs in the early-century dataset.

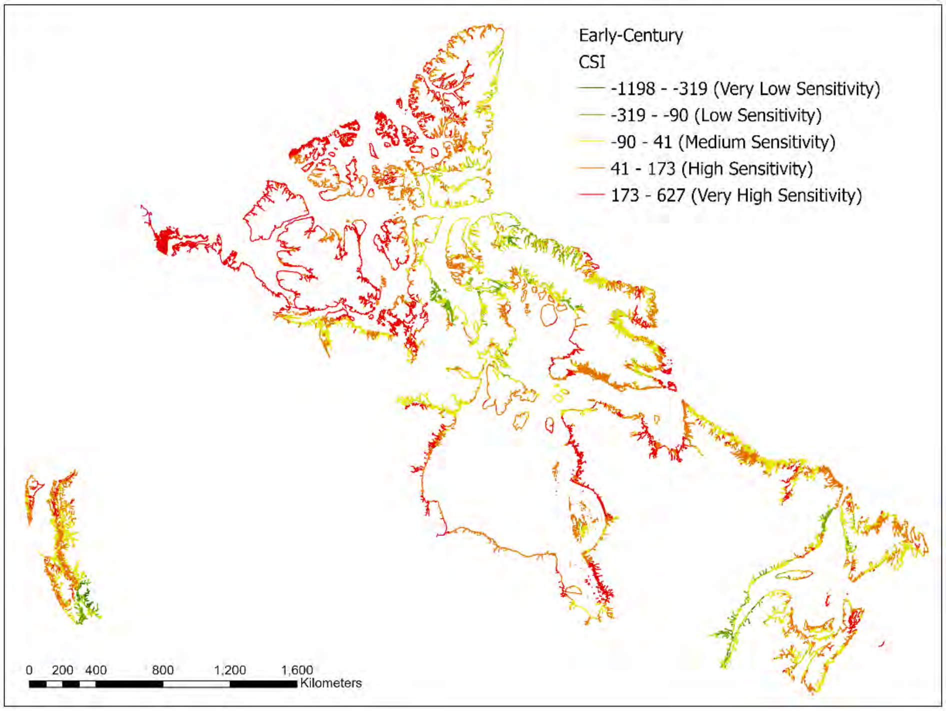

Figure 8 A National overview of the early-century CSI_{μ} scores.

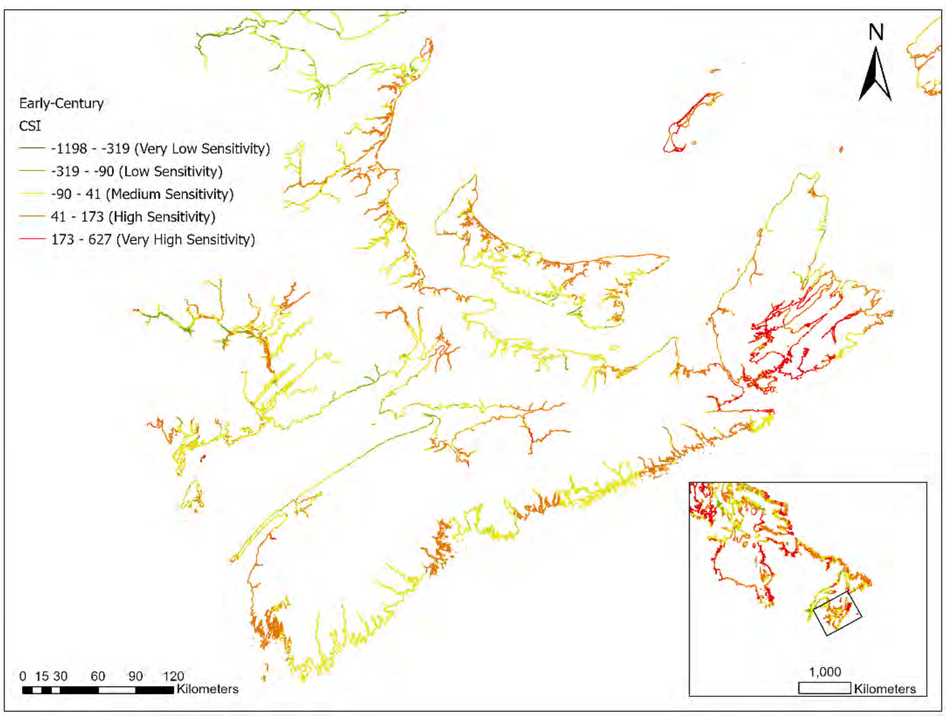

Figure 9 An Atlantic overview of the early-century CSI_{μ} scores.

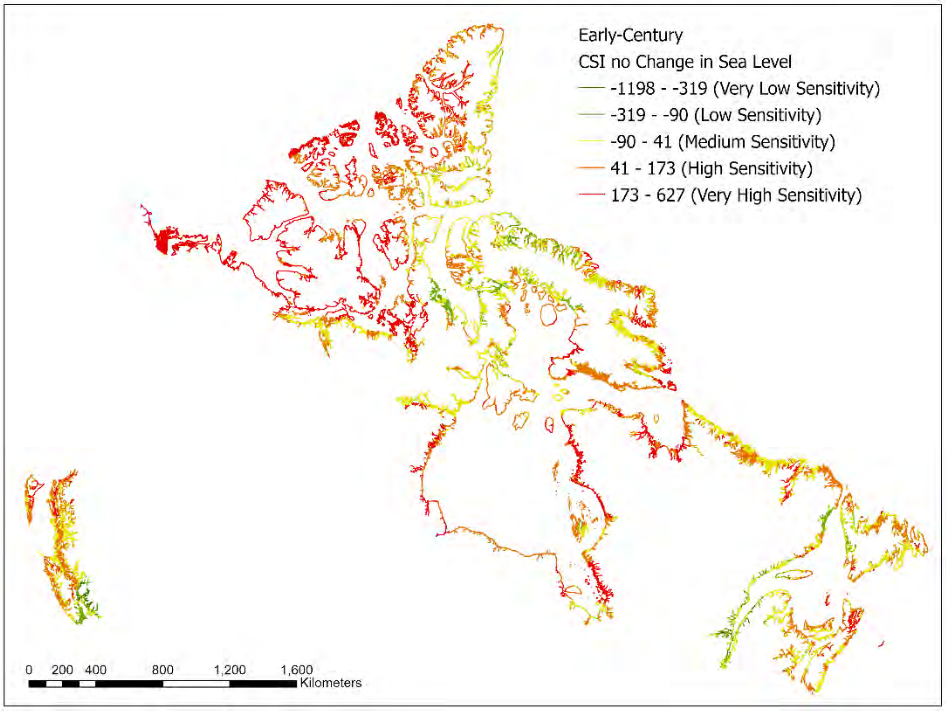

Figure 10 A National overview of the early-century CSI_{μ} scores with the Change in Sea Level 2006-2020 variable excluded.

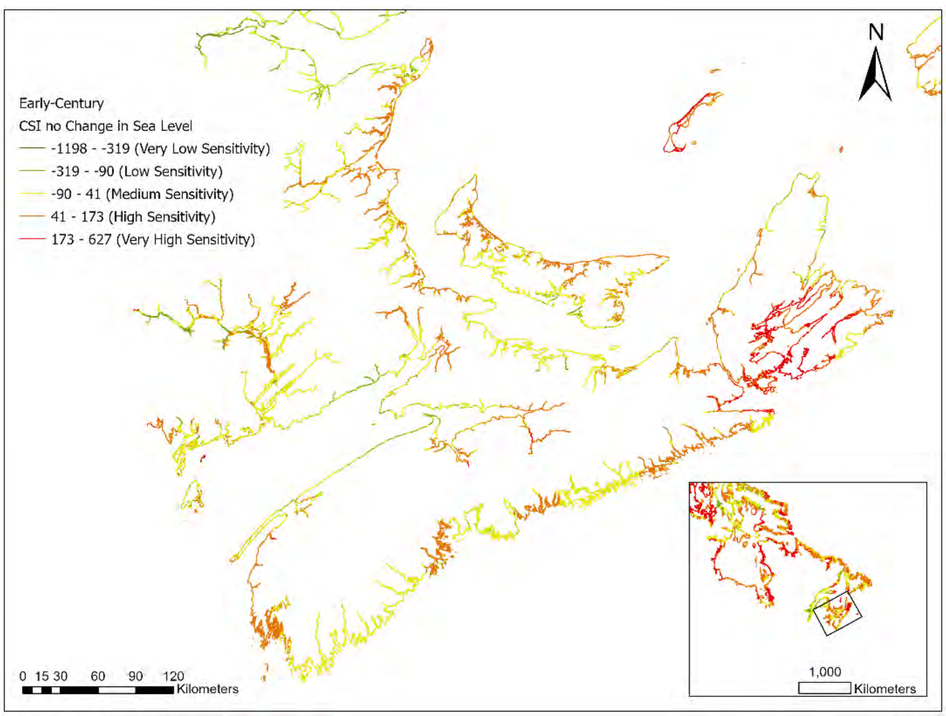

Figure 11 An Atlantic overview of the early-century CSI_{μ} scores with the Change in Sea Level 2006-2020 variable excluded.

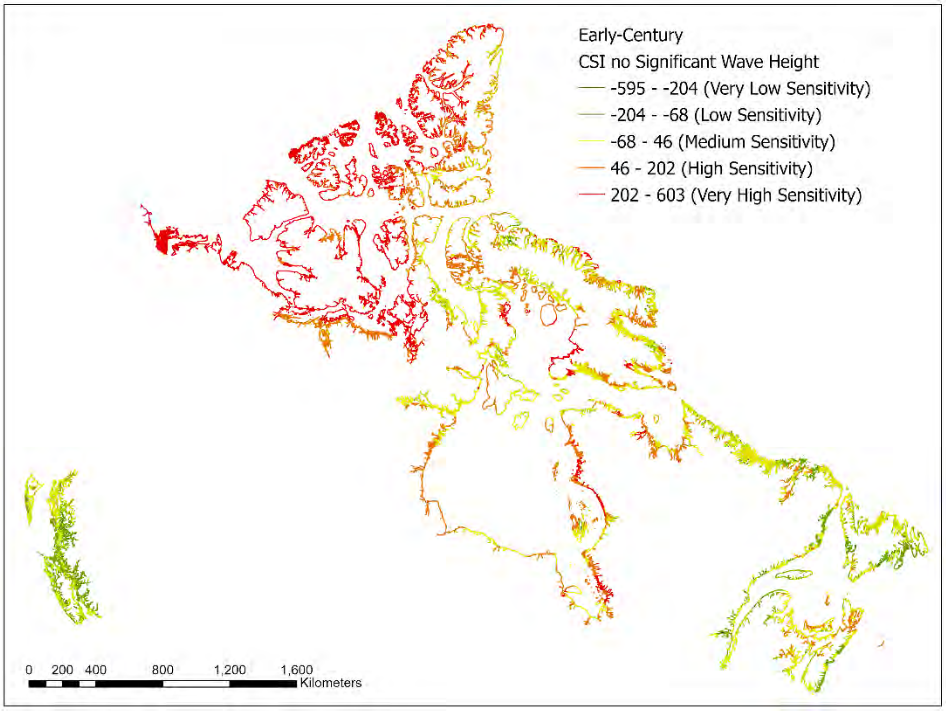

Figure 12 A National overview of the early-century CSI$_\mu$ scores with the Significant Wave Height Including Sea Ice Effects 1996-2005 variable excluded.

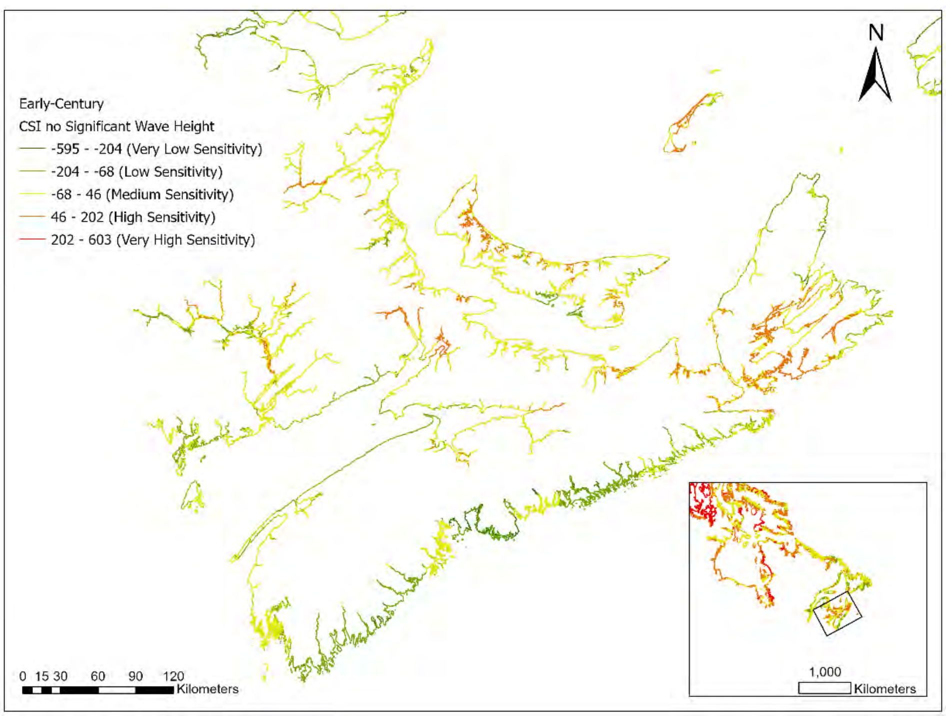

Figure 13 An Atlantic overview of the early-century CSI$_\mu$ scores with the Significant Wave Height Including Sea Ice Effects 1996-2005 variable excluded.

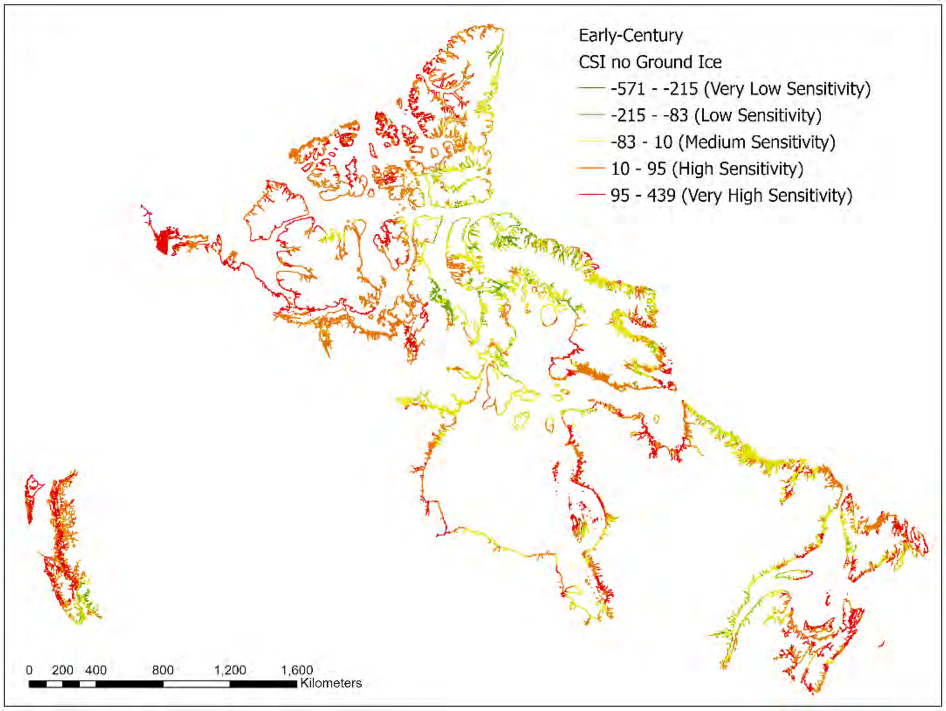

Figure 14 A National overview of the early-century CSI_{μ} scores with the Ground Ice variable excluded.

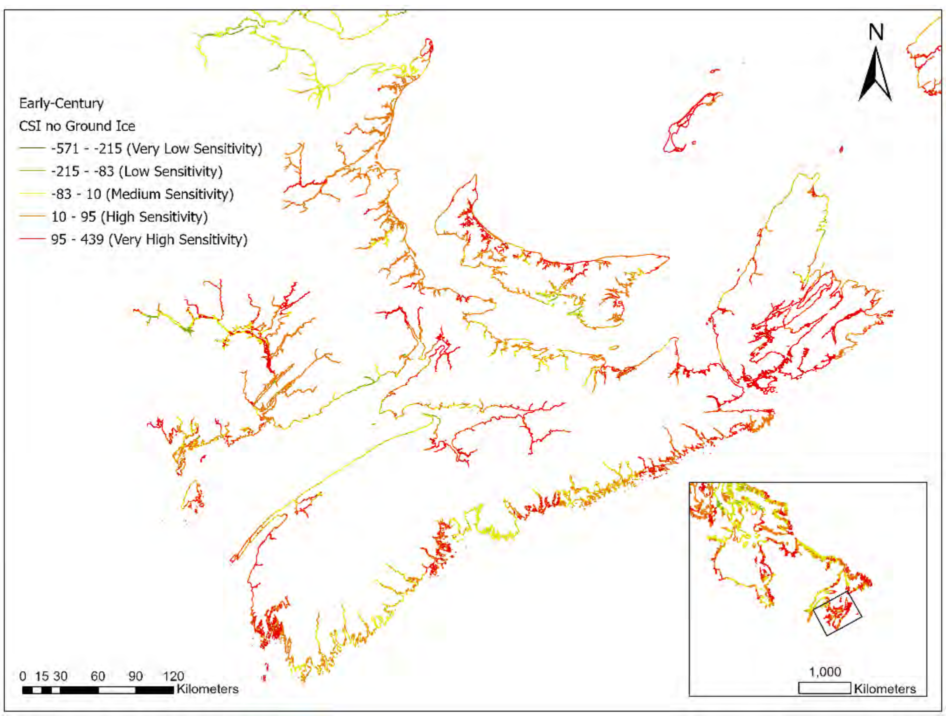

Figure 15 An Atlantic overview of the early-century CSI_{μ} scores with the Ground Ice variable excluded.

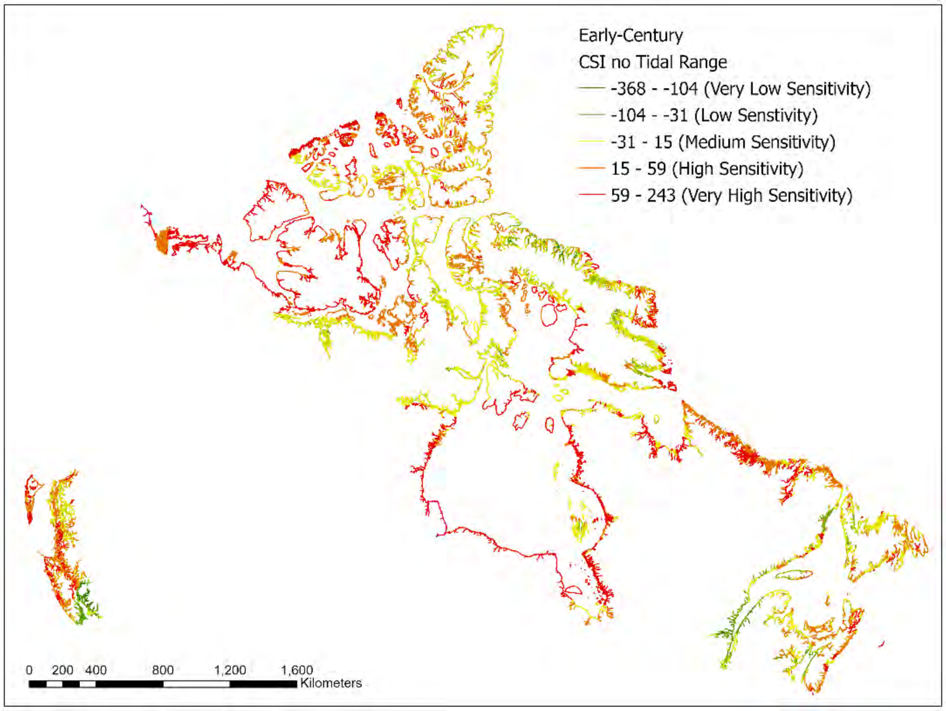

Figure 16 A National overview of the early-century CSI_{μ} scores with the Tidal Range variable excluded.

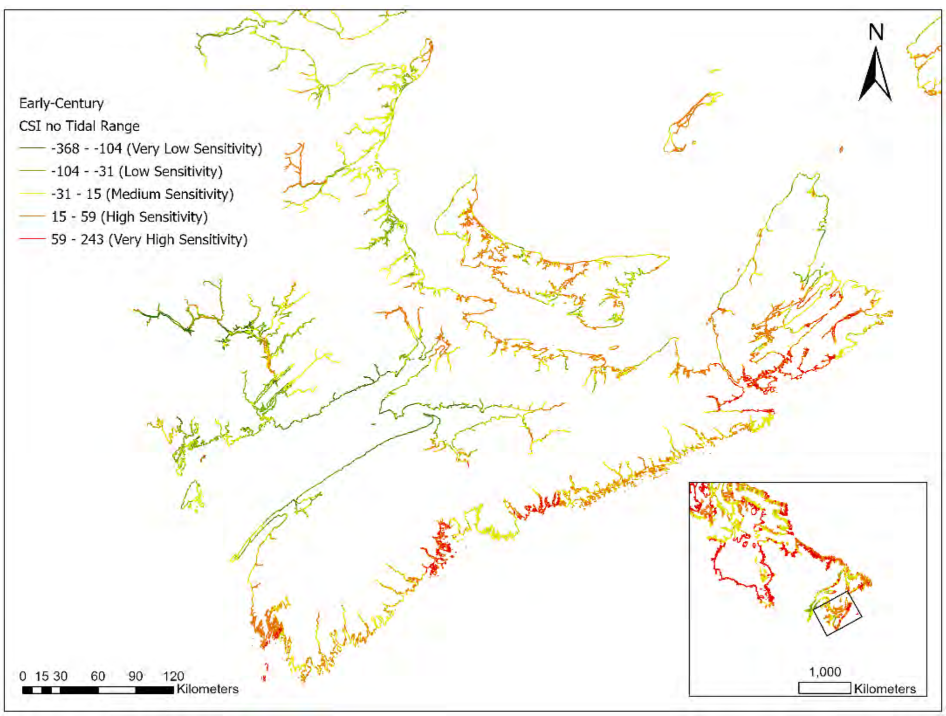

Figure 17 An Atlantic overview of the early-century CSI_{μ} scores with the Tidal Range variable excluded.

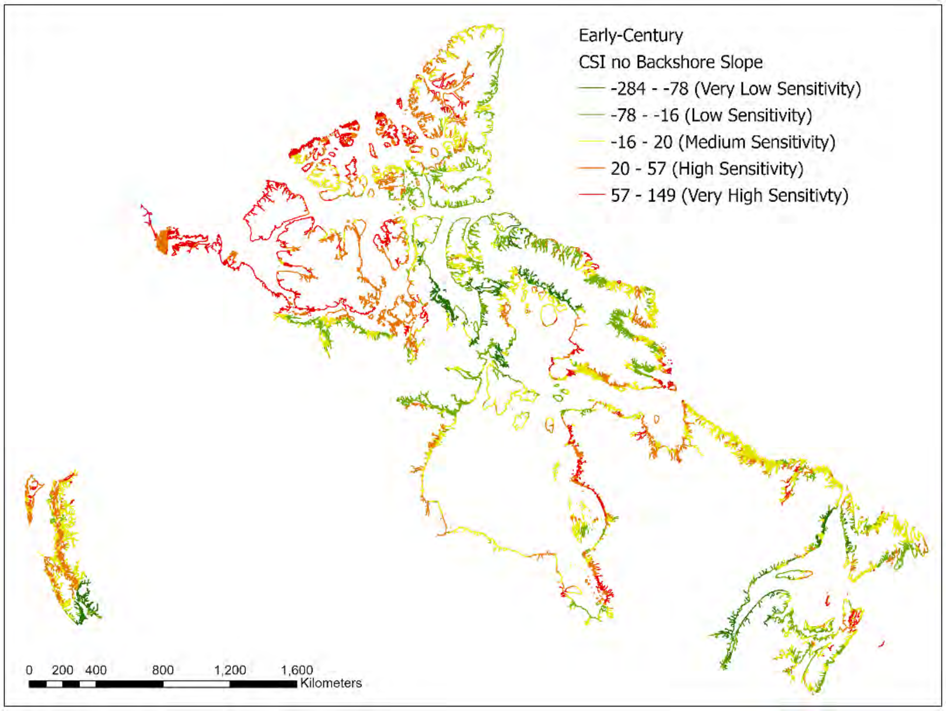

Figure 18 A National overview of the early-century CSI$_\mu$ scores with the Backshore Slope variable excluded.

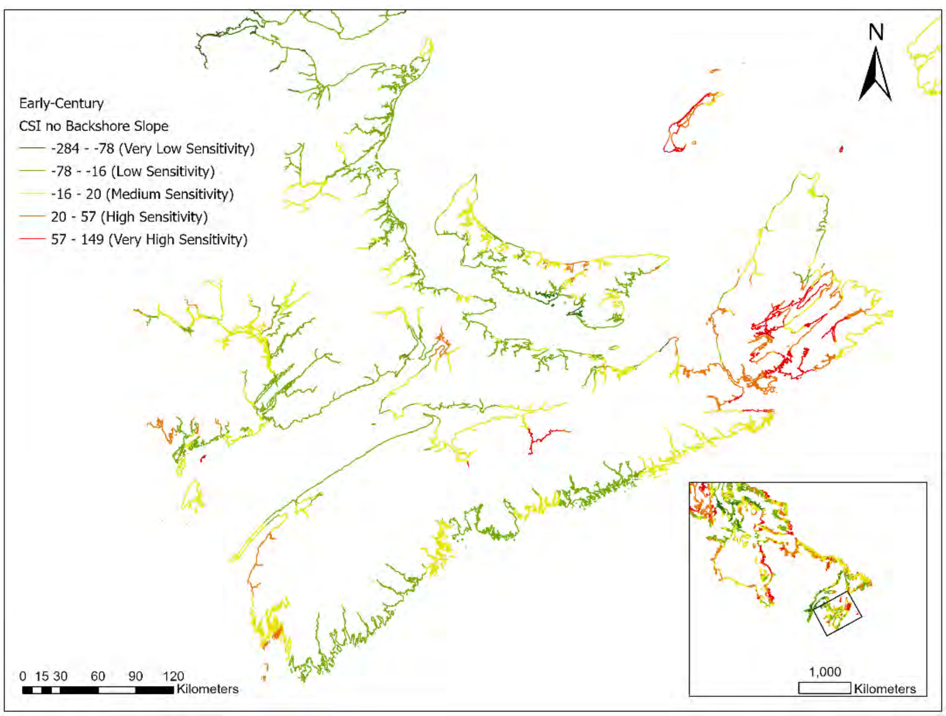

Figure 19 An Atlantic overview of the early-century CSI$_\mu$ scores with the Backshore Slope variable excluded.

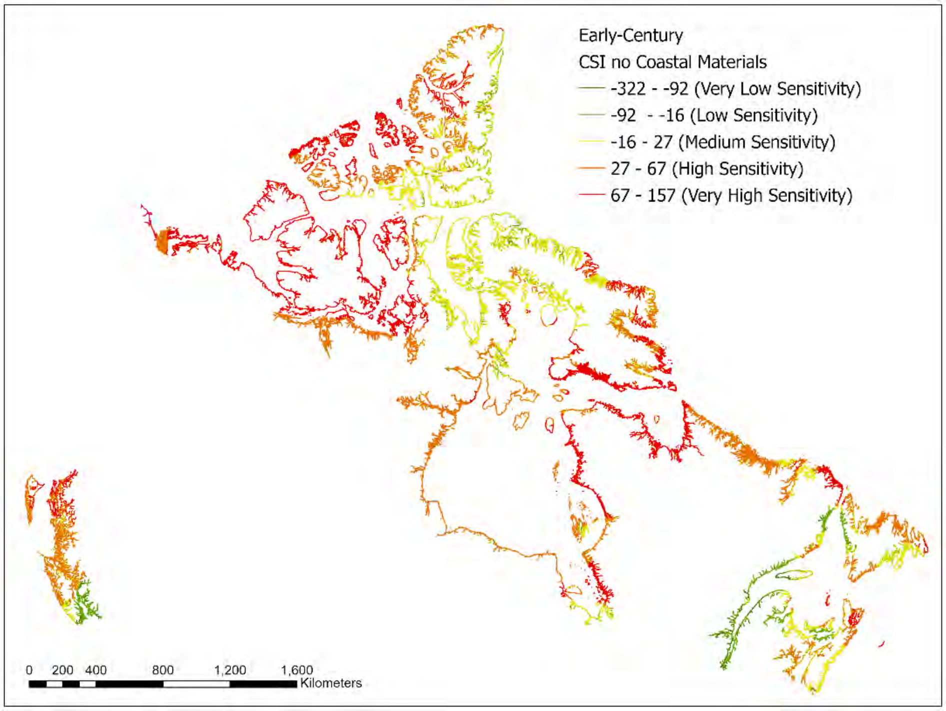

Figure 20 A National overview of the Early-Century CSI_{μ} scores with the Coastal Materials variable excluded.

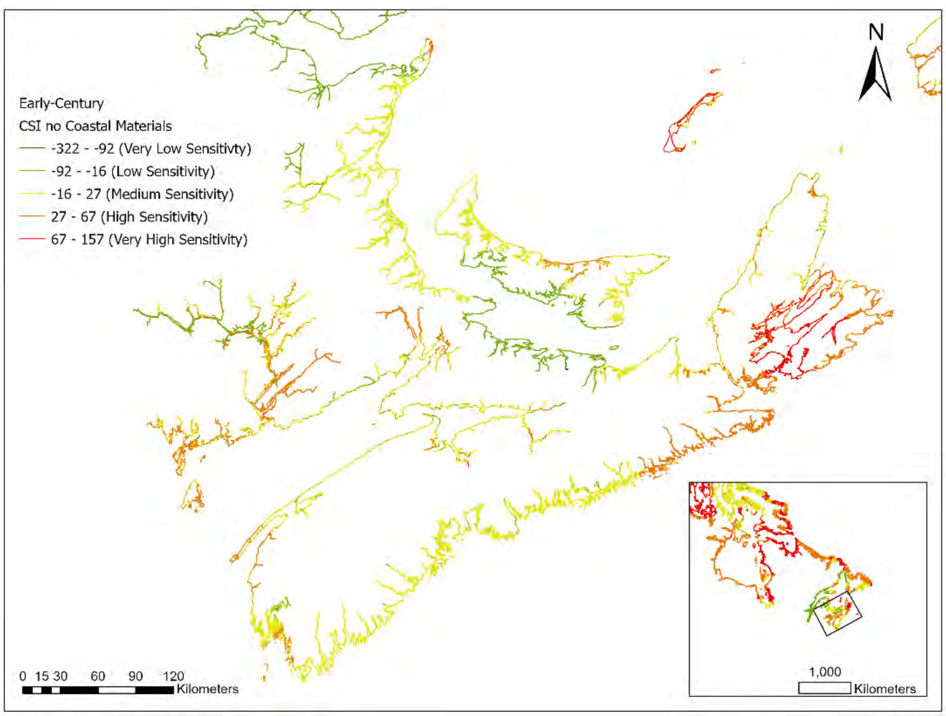

Figure 21 An Atlantic overview of the early-century CSI_{μ} scores with the Coastal Materials variable excluded.

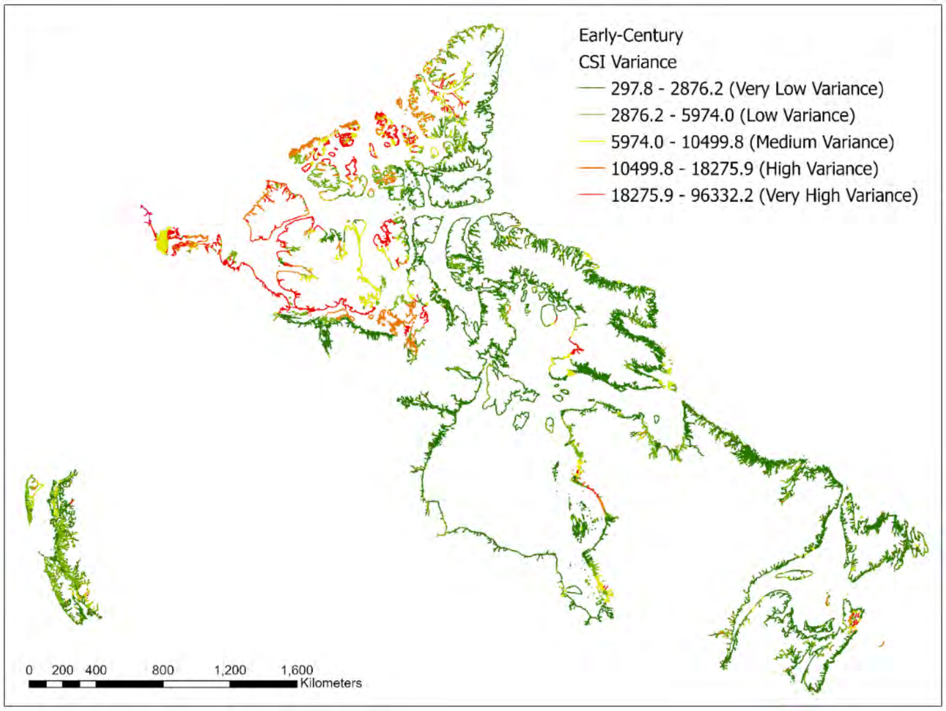

Figure 22 A national overview of the variance in early-century variable-excluded CSI_{μ} scores.

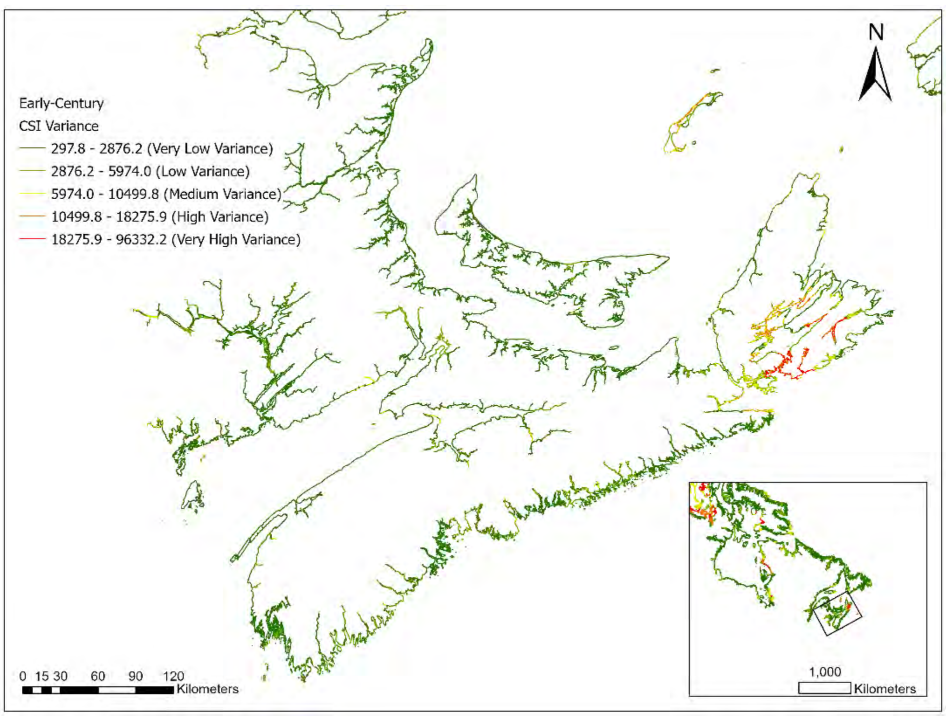

Figure 23 An Atlantic overview of the variance in early-century variable-excluded CSI_{μ} scores.

3.2 Late Century

The late-century data displayed the same scale of changes and thus the same logic as used in the early century section could be applied to interpret the results. Note that for this group of score calculations, the change in sea level run does differ from the control run as the indicator values from "CANCOAST_SEALEVEL2006_2100_V3" do vary spatially, likely owing to greater modelled change in sea level over nearly 100 years vs the more limited 14 years in the "CANCOAST_SEALEVEL2006_2020_V3" data used for the early-century.

Again, as with the early century dataset, the islands north of Nunavut and the Northwest Territories and areas of Cape Breton saw both some of the highest CSI_{μ} scores and the highest variation in scores across runs, seen in the maps and histograms. Following the trend set by the early century data, the ground ice variable-excluded score run once again pushed many more segments into the very high and high sensitivity classes of scores (see *Figure 32*).

Interestingly, when compared to the early-century data, variance of the variable-excluded CSI_{μ} scores across the country appears to have increased overall. Looking at the distribution histogram plots for the variance in *Figure 24* confirms that this is the case. The mean variance across the entire country is roughly doubled going into the late century dataset. As the only differences in the two datasets comes from the wave height and sea level, it can be inferred that these variables are responsible for this jump in variance. As these variables now represent change over a longer period of time, especially in the case of the sea level variable, their contribution to the overall variance has increased as well.

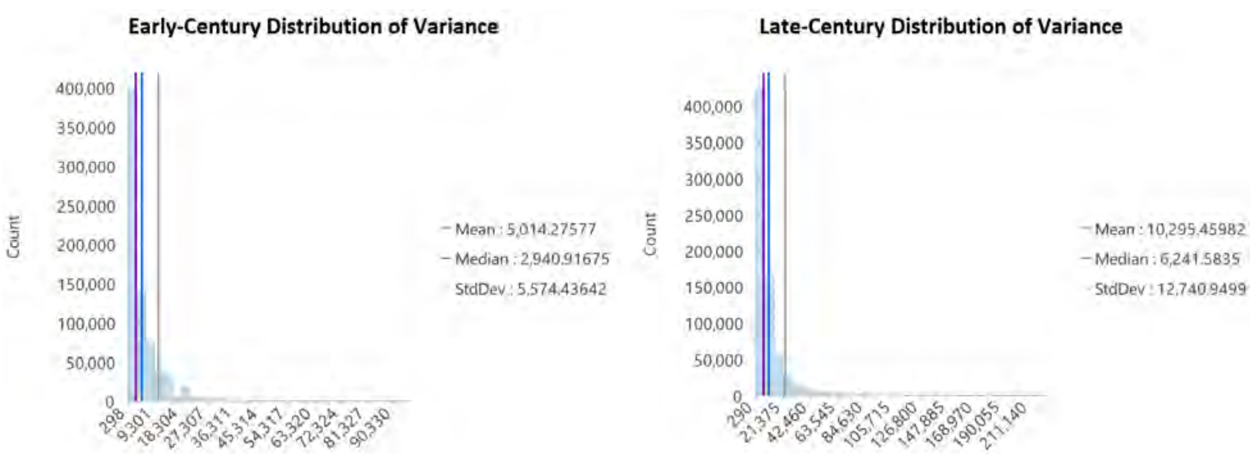

Figure 24 Plots of the distributions of the variance in variable excluded runs in the early and late century datasets.

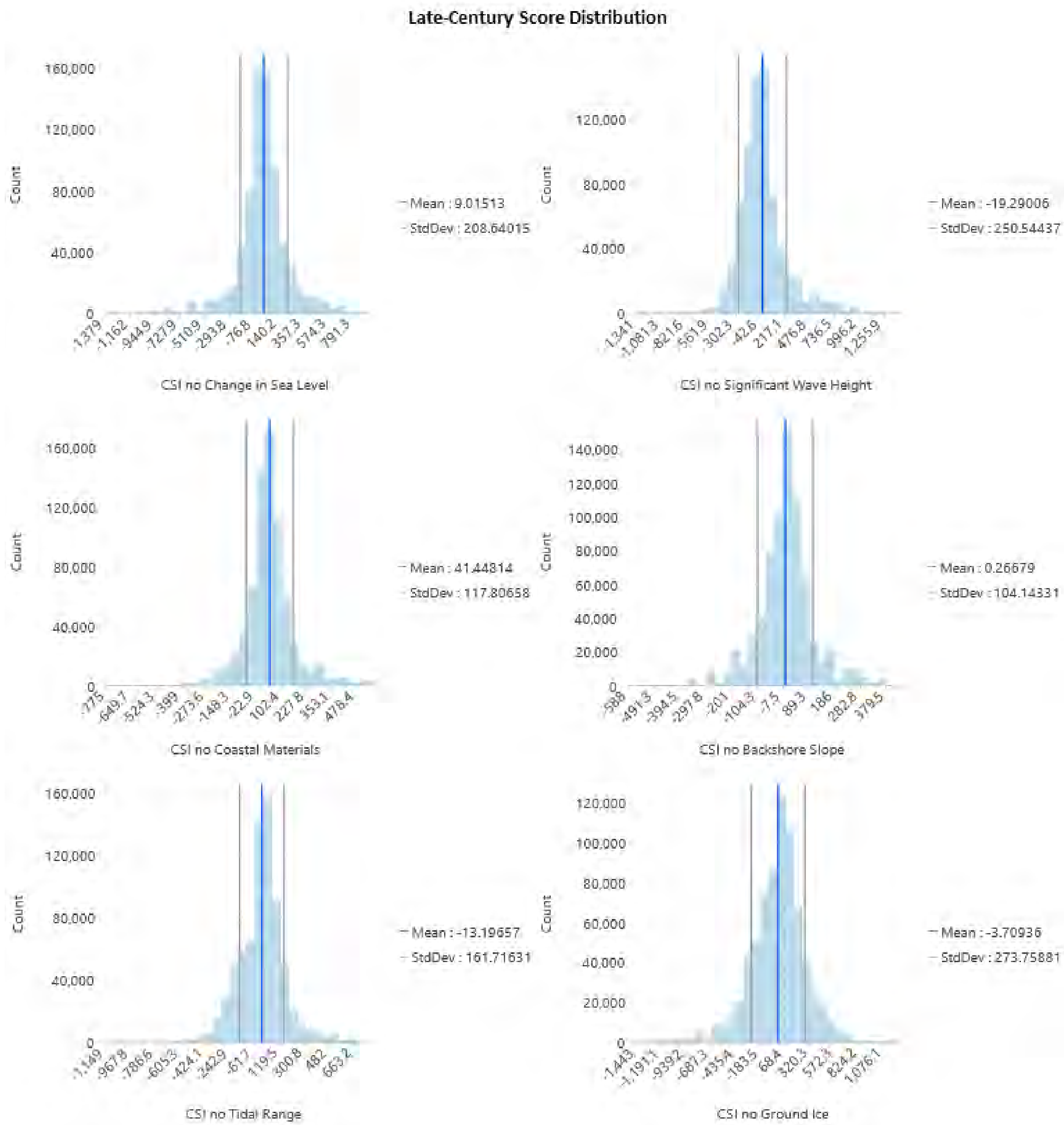

Figure 25 Plots of the distributions of CSI$_{\mu}$ scores among the variable-excluded runs in the late-century dataset.

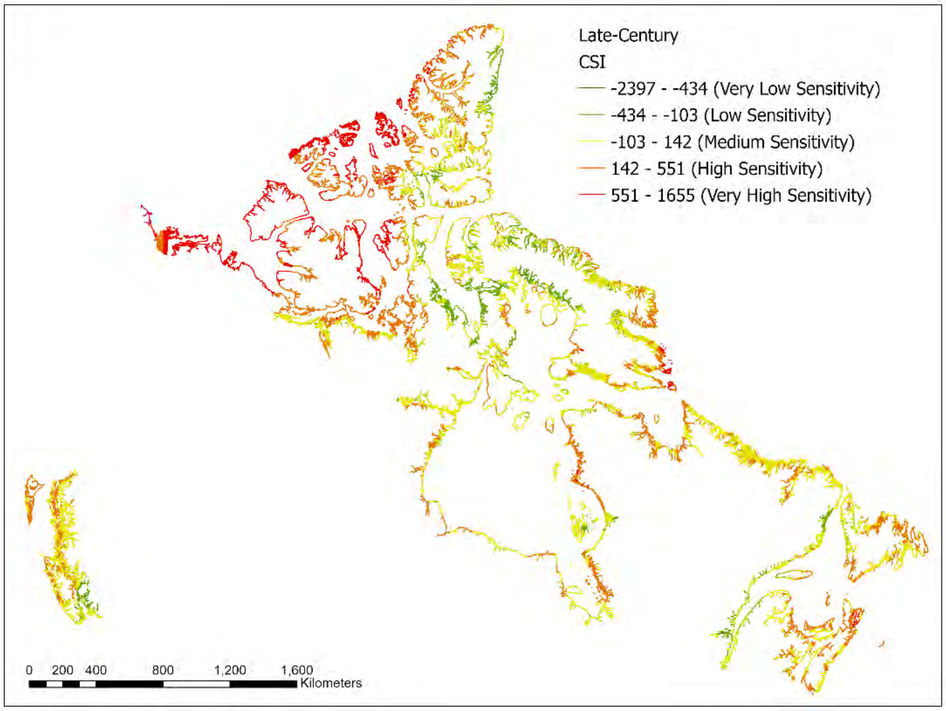

Figure 26 A National overview of the late-century CSI_{μ} scores.

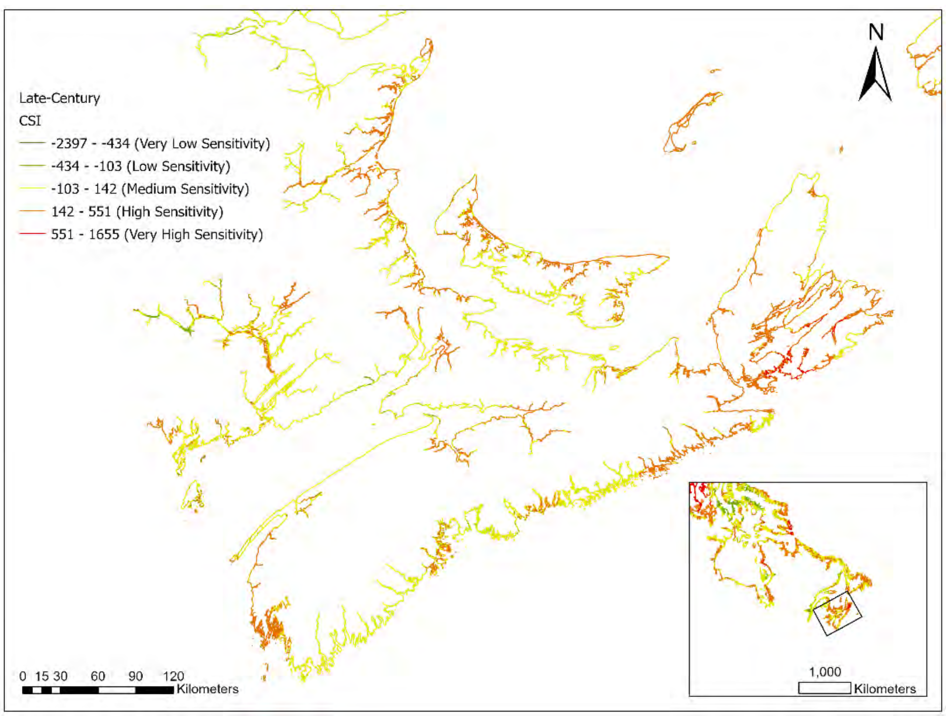

Figure 27 An Atlantic overview of the late-century CSI_{μ} scores.

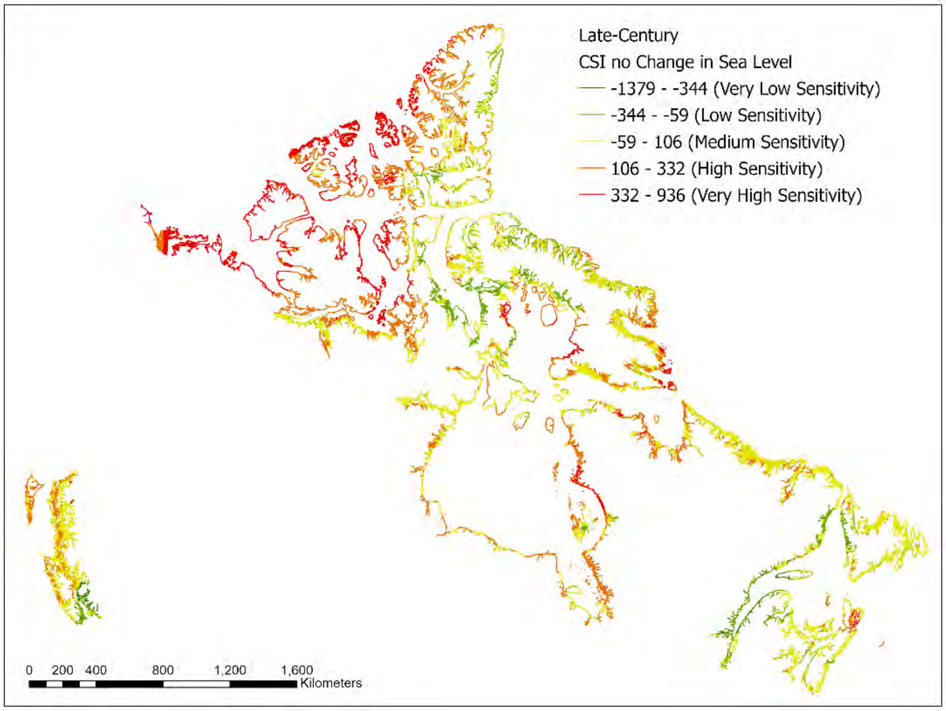

Figure 28 A National overview of the late-century CSI_{μ} scores with the Change in Sea Level 2006-2099 variable excluded.

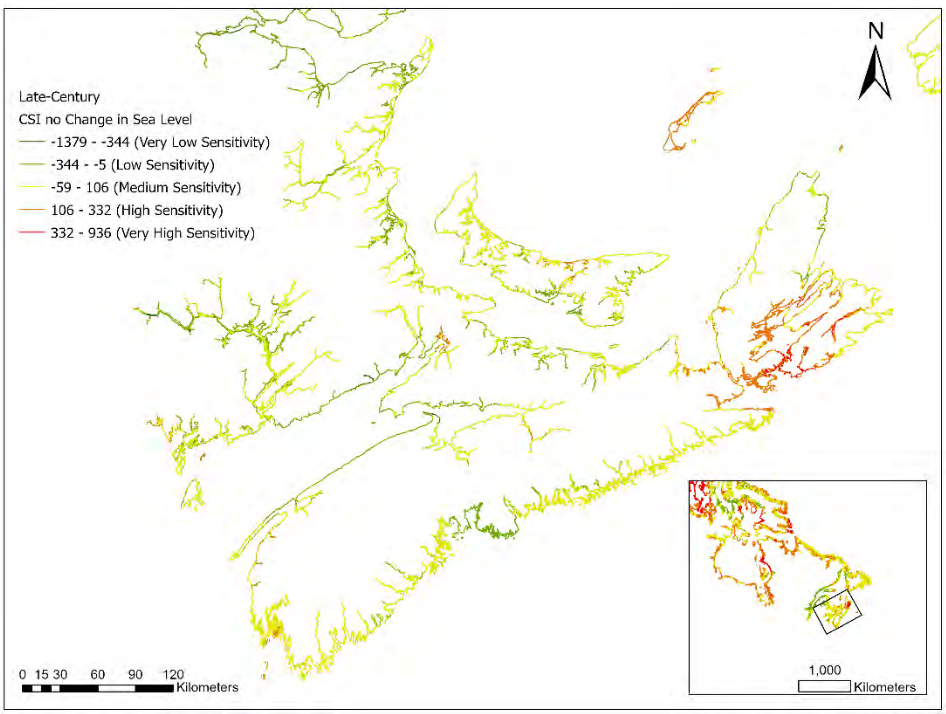

Figure 29 An Atlantic overview of the late-century CSI_{μ} scores with the Change in Sea Level 2006-2099 variable excluded.

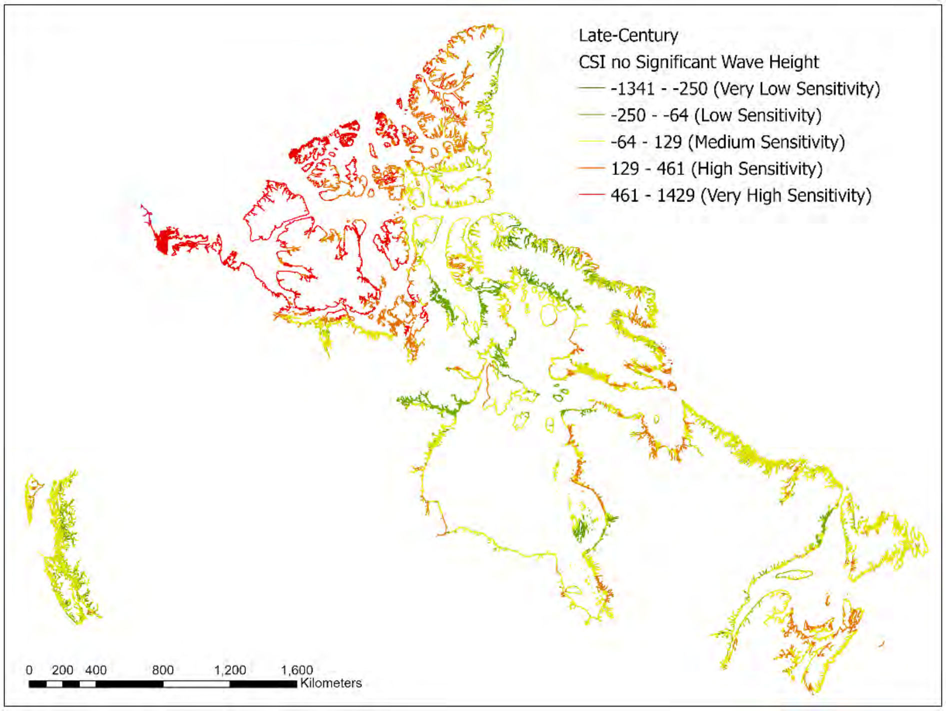

Figure 30 A National overview of the late-century CSI$_{\mu}$ scores with the Significant Wave Height Including Sea Ice Effects 2090-2099 variable excluded.

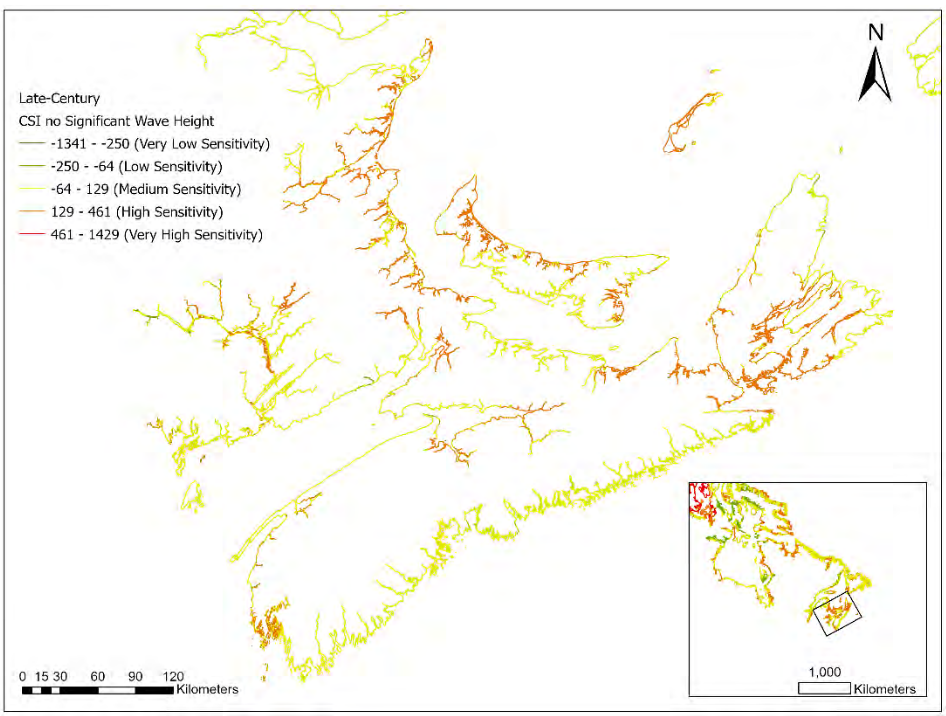

Figure 31 An Atlantic overview of the late-century CSI$_{\mu}$ scores with the Significant Wave Height Including Sea Ice Effects 2090-2099 variable excluded.

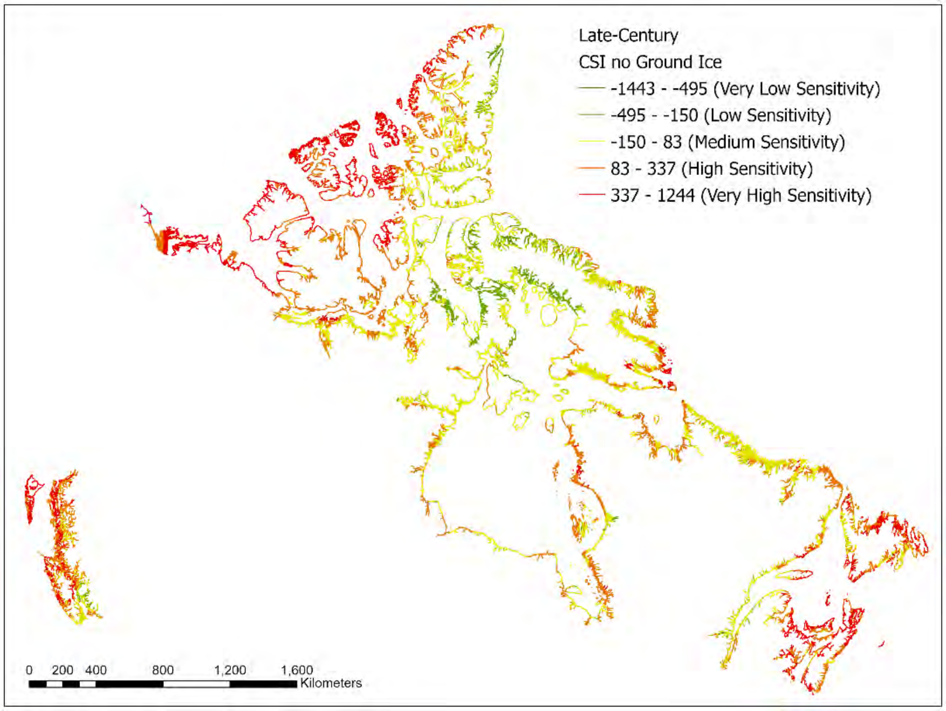

Figure 32 A National overview of the late-century CSIμ scores with the Ground Ice variable excluded.

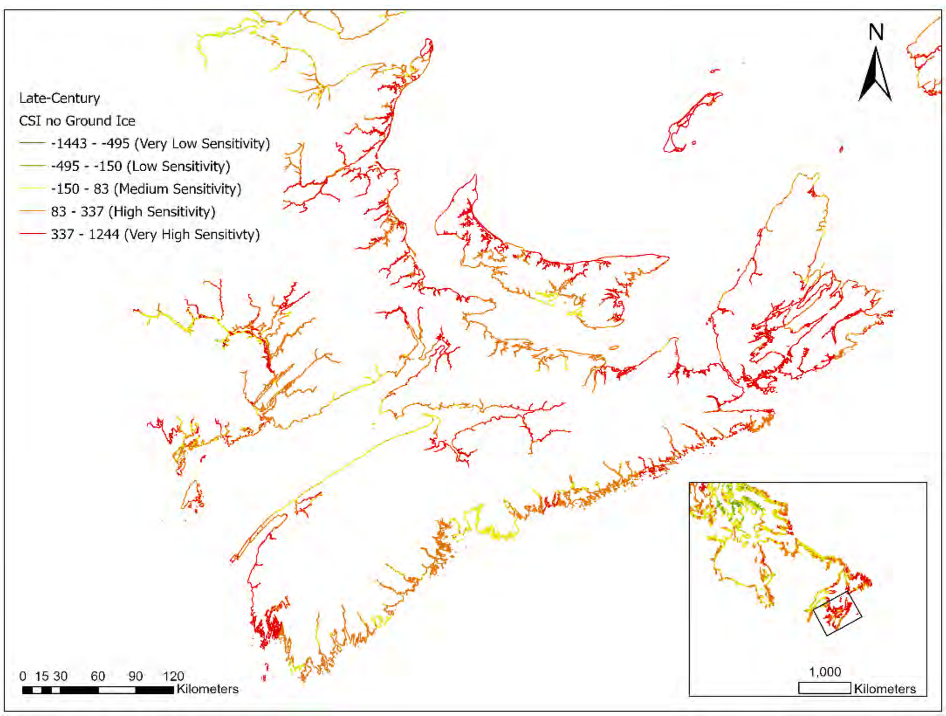

Figure 33 An Atlantic overview of the late-century CSIμ scores with the Ground Ice variable excluded.

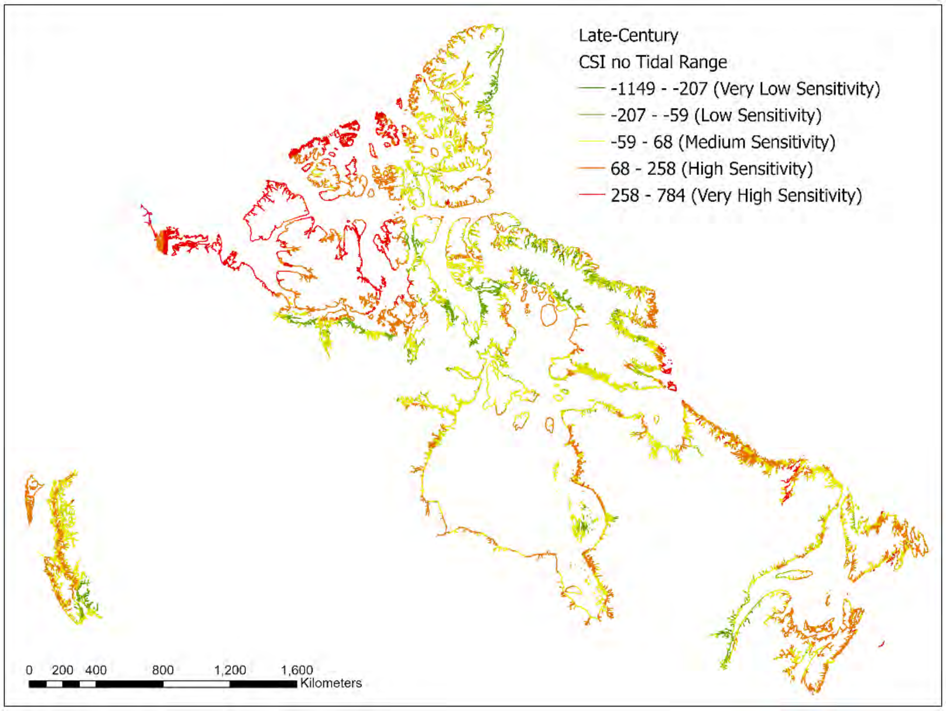

Figure 34 A National overview of the late-century CSI$_{\mu}$ scores with the Tidal Range variable excluded.

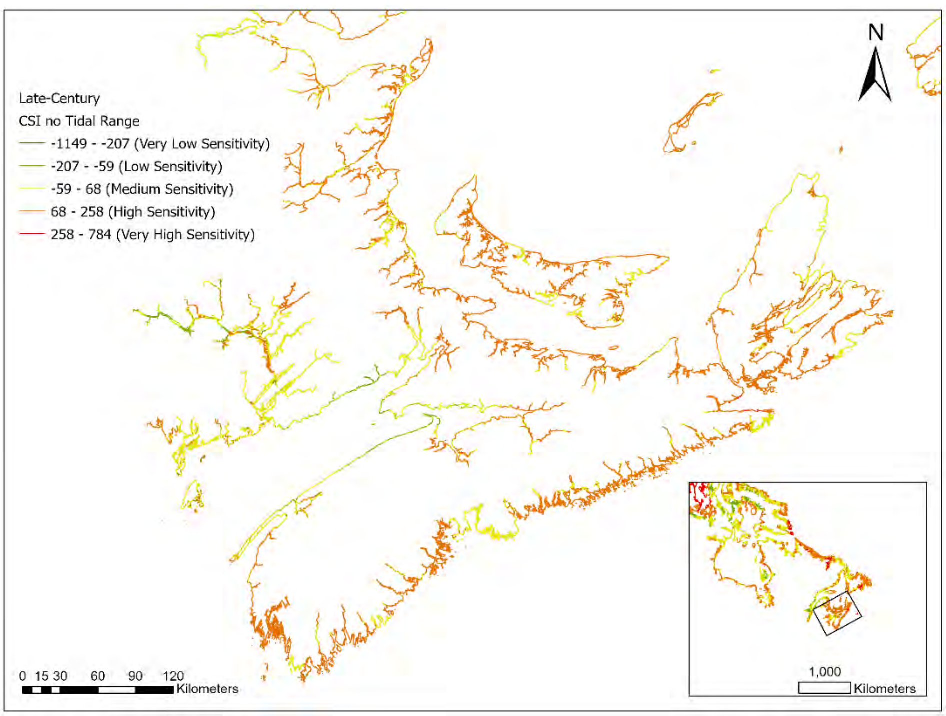

Figure 35 An Atlantic overview of the late-century CSI$_{\mu}$ scores with the Tidal Range variable excluded.

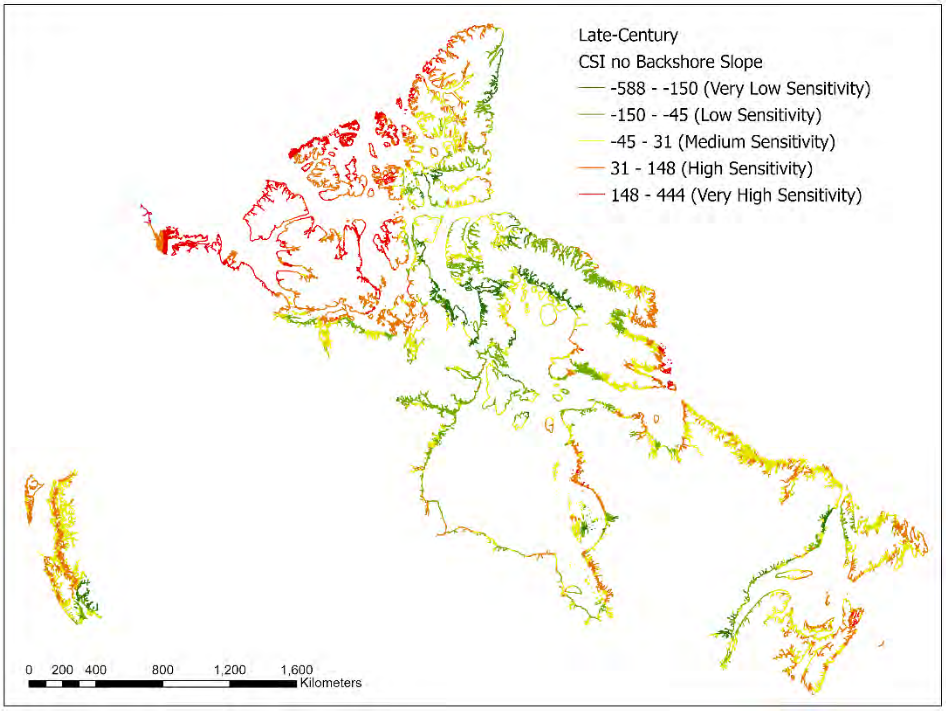

Figure 36 A National overview of the late-century CSI_{μ} scores with the Backshore Slope variable excluded.

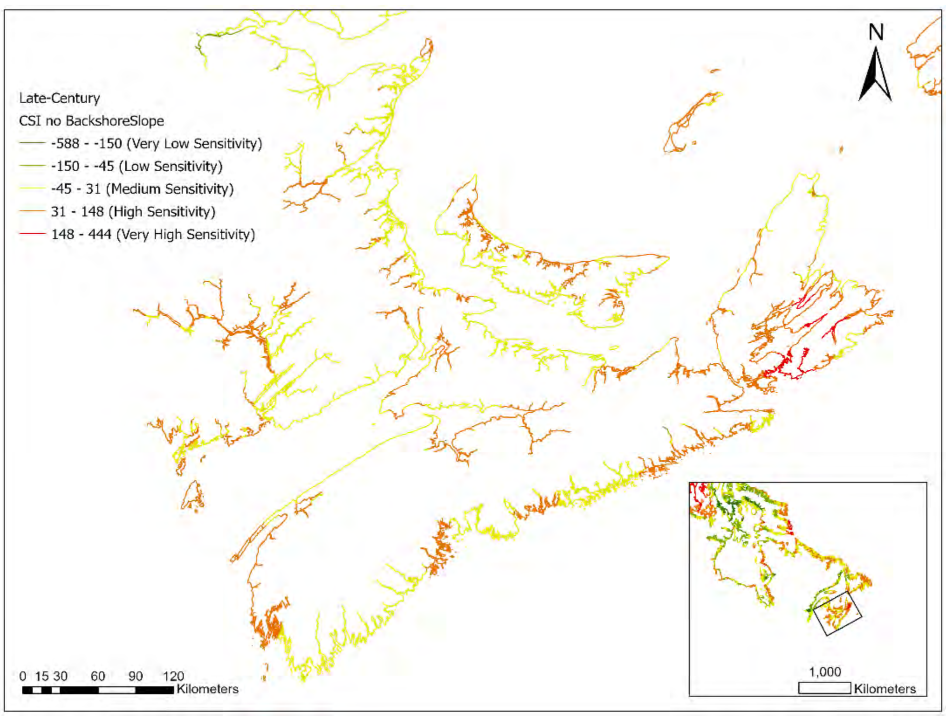

Figure 37 An Atlantic overview of the late-century CSI_{μ} scores with the Backshore Slope variable excluded.

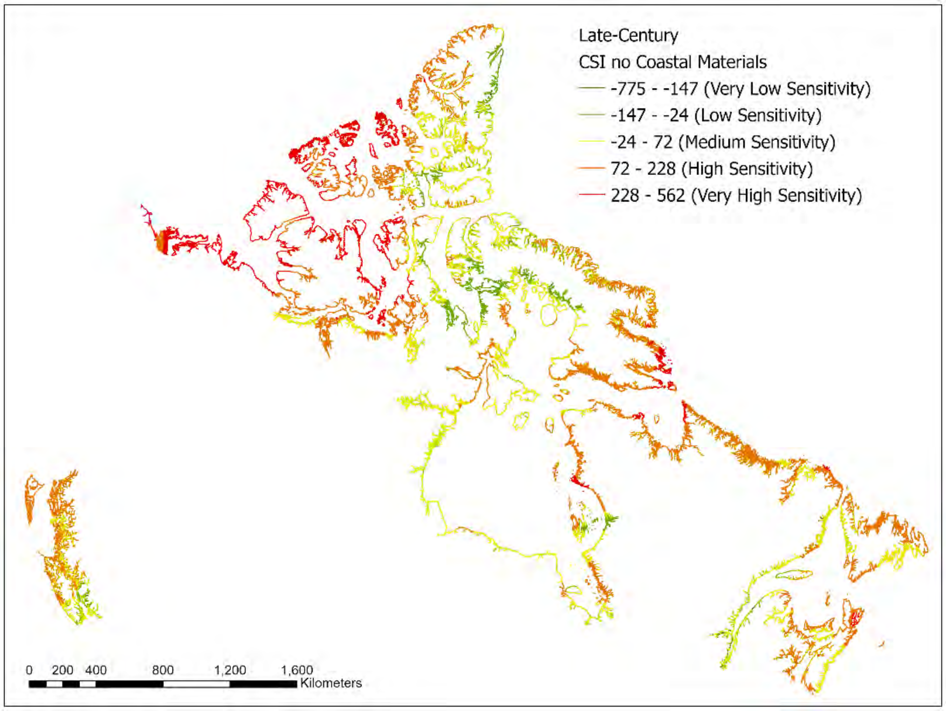

Figure 38 A National overview of the late-century CSI$_{\mu}$ scores with the Coastal Materials variable excluded.

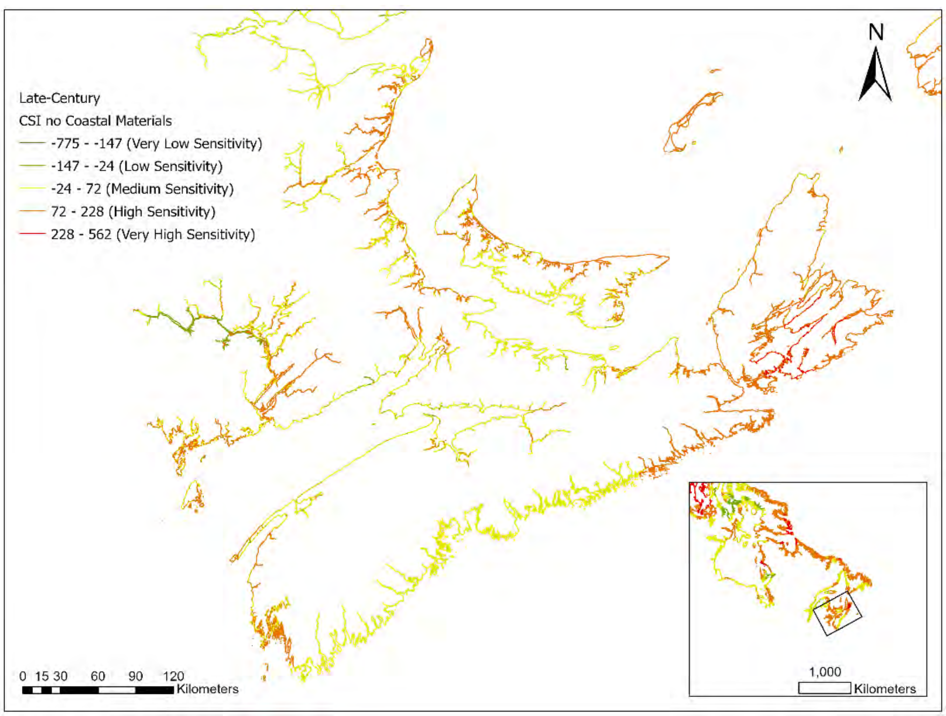

Figure 39 An Atlantic overview of the late-century CSI$_{\mu}$ scores with the Coastal Materials variable excluded.

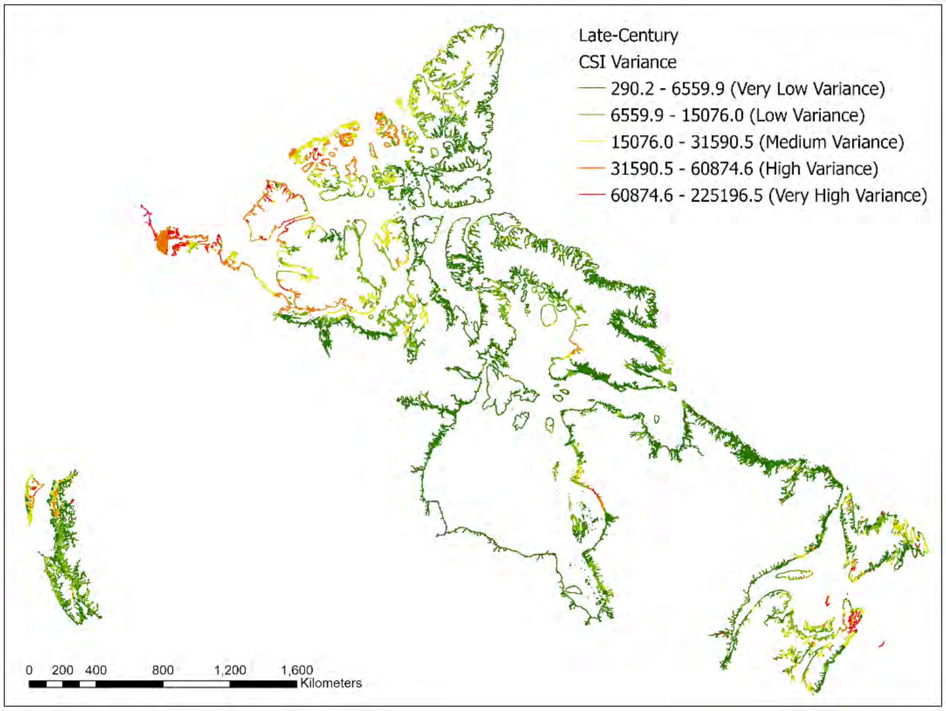

Figure 40 A national overview of the variance in early-century variable-excluded CSI_{μ} scores.

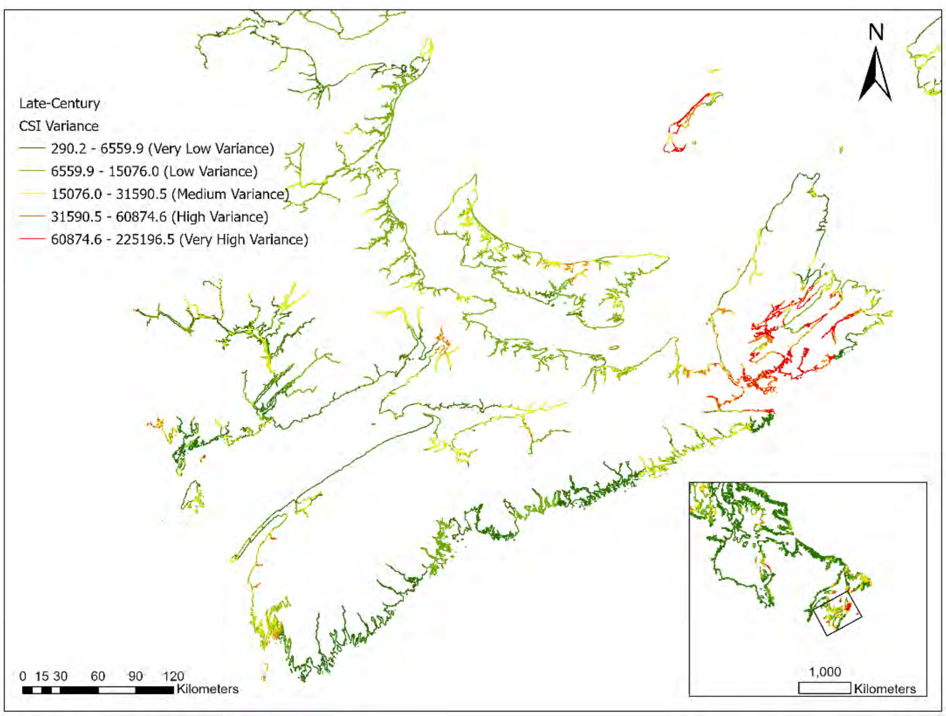

Figure 41 An Atlantic overview of the variance in early-century variable-excluded CSI_{μ} scores.

4. Conclusion

This project has successfully replicated the scripting methods to apply the µ-scoring procedure from Hatcher and Manson (2021) using the original CanCoast 2.0 shapefiles and updated shapefile data provided to AGRG. The data was prepared for scoring by combining the original indicator variable shapefiles using ArcGIS Pro, splitting it into early century and late century datasets, then inputting it into a Python script that calculates the CSI_{μ} scores. Lastly, as a form of spatial analysis of the data, plots were made, stats on CSI_{μ} scores including variance and standard deviation were calculated and score distribution histograms were produced within ArcGIS Pro. From these products, quantitative and qualitative analyses can be performed to see both the areas that are influenced the most by changes in these input variables, and which variables seem to have the biggest influence on overall scores by how much they have changed compared to the original control CSI_{μ} scores. In particular, Banks Island, Victoria Island, and the Queen Elizabeth Islands north of Nunavut and the Northwest Territories, and the coastal regions of Cape Breton around the Bras d'Or Lake in Nova Scotia stand out for being consistently high in sensitivity throughout all the runs. In terms of indicator variables, the ground-ice excluded runs stand out as having pushed many more coastal segments into the very high sensitivity range of scores.

5. Acknowledgements

We would like to thank Geological Survey of Canada (Atlantic), GSCA for the funding support. Many thanks to both Dr. Gavin Manson for providing support and guidance during the project work and Scott Hatcher for writing the original scripts added to in this project, as well as their work done in developing the original methods upon which the project was based. We would like to thank AGRG’s Kevin McGuigan and Matt Roscoe for their guidance in choosing the most time-efficient scripting and processing methods, and lastly a big thank you the team of researchers responsible for collecting and creating the data used to complete the original CanCoast 2.0 data.

6. References

Hatcher, S.V., Manson, G.K. (2021) A revised coastal sensitivity index for Canada's marine coasts calculated using nonparametric statistics.

Manson , G.K., Couture, N.J., James, T.S. (2019) CanCoast 2.0: data and indices to describe the sensitivity of Canada's marine coasts to changing climate.