SOPhyE Satellite Data Processing Technical Report Series: 2. Refining the Maritimes AZMP box boundaries using Self-Organizing Maps (SOMs)

Melanie Hardy, Emmanuel Devred

Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada, B2Y 4A2

2025

Canadian Technical Report of Hydrography and Ocean Sciences 387

Canadian Technical Report of Hydrography and Ocean Sciences

Technical reports contain scientific and technical information of a type that represents a contribution to existing knowledge but which is not normally found in the primary literature. The subject matter is generally related to programs and interests of the Oceans and Science sectors of Fisheries and Oceans Canada.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts.*

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications and the last number issued under each title are published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 38: Index to Publications 1981. The current series began with Report Number 1 in January 1982.

Rapport technique canadien sur l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles mais que l'on ne trouve pas normalement dans les revues scientifiques. Le sujet est généralement rattaché aux programmes et intérêts des secteurs des Océans et des Sciences de Pêches et Océans Canada.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page de titre.

Les établissements de l’ancien secteur des Sciences et Levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Vous trouverez dans l'index des publications du volume 38 du *Journal canadien des sciences halieutiques et aquatiques*, la liste de ces publications ainsi que le dernier numéro paru dans chaque catégorie. La nouvelle série a commencé avec la publication du rapport numéro 1 en janvier 1982.

Canadian Technical Report of
Hydrography and Ocean Sciences 387

2025

SOPHYE SATELLITE DATA PROCESSING TECHNICAL REPORT SERIES: 2. REFINING THE MARITIMES AZMP BOX BOUNDARIES USING SELF-ORGANIZING MAPS (SOMS)

by

Melanie Hardy, Emmanuel Devred

Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada, B2Y 4A2

Cat. No. Fs 97-18/387E-PDF ISBN 978-0-660-74655-5 ISSN 1488-5417

Correct Citation for this publication:

Hardy, M., Devred, E. 2025. SOPhyE Satellite Data Processing Technical Report Series: 2. Refining the Maritimes AZMP Box Boundaries using Self-Organizing Maps (SOMs). Can. Tech. Rep. Hydrogr. Ocean Sci. 387: vii + 16 p.

Contents

List of Tables

List of Figures

ABSTRACT

Hardy, M., Devred, E. 2025. SOPhyE Satellite Data Processing Technical Report Series: 2. Refining the Maritimes AZMP Box Boundaries using Self-Organizing Maps (SOMs). Can. Tech. Rep. Hydrogr. Ocean Sci. 387: vii + 16 p.

Chlorophyll-a concentration (Chl-a) derived from satellite ocean colour has been used to report on phytoplankton biomass in the Maritimes Atlantic Zone Monitoring Program (AZMP) at five key areas on the Scotian Shelf, namely, Georges Bank (GB), Central Scotian Shelf (CSS), Eastern Scotian Shelf (ESS), Western Scotian Shelf (WSS), and Lurcher Shoals (LS). For practical reasons, the areas were defined as rectangles that did not necessarily account for the effect of local hydrodynamic processes on the distribution of phytoplankton biomass. As a result, variation in Chl-a concentration may not reflect phytoplankton local dynamics but rather changes in water masses that may be amplified by cloud cover. A Self-Organizing Map (SOM) approach was applied to satellite-derived seasonal sea-surface temperature (SST) and Chl-a to identify the main patterns in phytoplankton dynamics on the Scotian Shelf in relation to physical forcing, and to redefine the boundaries of the AZMP Maritime boxes. The updated polygons had similar mean seasonal SST and Chl-a than the original boxes but with reduced variability as supported by smaller interquartile range. The updated polygons did not encroach on coastal areas that exhibit systematically high Chl-a as a result of contamination of the marine satellite signal by terrigenous inputs. Comparisons of the bloom metrics between the original boxes and the new SOM-derived polygons showed less extreme values in bloom characteristics for the SOM-derived polygons.

RÉSUMÉ

Hardy, M., Devred, E. 2025. SOPhyE Satellite Data Processing Technical Report Series: 2. Refining the Maritimes AZMP Box Boundaries using Self-Organizing Maps (SOMs). Can. Tech. Rep. Hydrogr. Ocean Sci. 387: vii + 16 p.

La concentration de chlorophylle-a (Chl-a) dérivée de la couleur de l'océan par satellite a été utilisée pour informer sur la biomasse phytoplanctonique dans le cadre du Programme de surveillance de la zone atlantique des Maritimes (PMZA) dans cinq zones clés du plateau néo-écossais, à savoir le banc Georges (GB), le centre du plateau néo-écossais (CSS), l'est du plateau néo-écossais (ESS), l'ouest du plateau néo-écossais (WSS) et les hauts-fonds de Lurcher (LS). Pour des raisons pratiques, ces zones ont été définies comme des rectangles qui ne tiennent pas nécessairement compte de l'effet des processus hydrodynamiques locaux sur la distribution de la biomasse phytoplanctonique. Par conséquent, la variation de la Chl-a peut ne pas refléter la dynamique locale du phytoplancton, mais plutôt les changements dans les masses d'eau qui peuvent être amplifiés par la couverture nuageuse. Une approche de carte auto-organisatrice (SOM) a été appliquée à la température saisonnière de surface de la mer (SST) et à la Chl-a dérivées des données satellites afin d'identifier les patrons principaux de la dynamique du phytoplancton sur le plateau néo-écossais en relation avec le forçage physique, et de redéfinir les limites des zones maritimes de l'AZMP. Les zones mises à jour sous forme de polygones présentent des moyennes saisonnières de SST et de Chl-a similaires à celles des boîtes originales, mais avec une variabilité réduite, comme le montre l'écart interquartile plus petit. Les polygones mis à jour n'ont pas empiété sur les zones côtières qui présentent systématiquement un niveau élevé de Chl-a en raison de la contamination du signal satellitaire marin par des apports terrigènes. Les comparaisons des métriques de floraison entre les boîtes originales et les nouveaux polygones dérivés de la SOM ont montré des valeurs moins extrêmes dans les caractéristiques de floraison pour les polygones dérivés de la SOM.

1 Introduction

Satellite remote sensing is a tool that can be used in a wide range of applications in oceanographic research and monitoring. Satellite remote sensing provides the advantages of efficiently collecting oceanographic data at a high frequency and over synoptic scales. In particular, oceanographic data such as ocean color provides valuable information on the biological activity of primary producers by retrieving the concentration of chlorophyll-a (Chl-a in $\mathrm{mg\,m^3}$) in the upper layer of the ocean. In addition to ocean color, some satellites collect the thermal signature of the surface layer of the ocean, referred to as sea-surface temperature (SST in ^{o}C). The department of Fisheries and Oceans (DFO) Maritimes region has used satellite-based SST and Chl-a to monitor changes and trends in specific regions of the Northwest Atlantic Ocean to infer the overall status of these marine ecosystems over time through the Atlantic Zone Monitoring Program (AZMP) (DFO, 2021). Areas monitored by DFO on the Scotian Shelf are referred to as the Maritimes AZMP boxes (Casault et al., 2024).

The current boxes used in the Maritimes AZMP are represented by arbitrarily-defined rectangles without accounting for local hydrodynamic effects that impact phytoplankton spatial distribution, as well as increased variance due to the systematic differences in Chl-a among water masses associated with coastal, shelf, and slope currents (Figure 1). The objective of the current study is to refine the Maritimes AZMP box boundaries to improve the ability to capture variation in phytoplankton biomass and associated metrics (i.e., bloom initiation, duration, amplitude, and magnitude) in response to local environmental forcing and to reduce variations resulting from systematic spatial heterogeneity associated with changes in water masses. The approach used to update the box boundaries involves the utilization of the SOM approach from the R package 'kohonen' (Wehrens and Kruisselbrink, 2018) to identify water masses and delineate oceanographic areas with contiguous properties based on satellite-derived seasonal SST and Chl-a, and bathymetry. In addition, we anticipate the SOM to improve the ability to quantify variation in dynamics of Chl-a concentration that are representative of the targeted subregions on the Scotian Shelf and their associated ecosystem by reducing variability due to changes in water masses.

The SOM, also commonly known as the Kohonen network or map, is a method that uses an unsupervised artificial neural network designed to identify patterns using competitive training (Kohonen, 2001). This method has demonstrated its universal capability at recognizing patterns to a wide range of applications such as speech recognition, cloud classification, and data mining, as well as ocean applications including improving Chl-a estimates and ocean color classification (Oja et al., 2003). SOMs are similar to k-means clustering with the additional advantage of preserving data topology (i.e., map consistency for satellite data) by keeping the same neighborhoods (Wehrens et al., 2007). The SOM requires the user to manually test the number of nodes (i.e., number of clusters) and decide what fits best for the data being trained. It's important to note that the nodes are randomly initialized before training a data set, such that no matter the number of times the initial data set is trained, the results will always be similar but not identical (Lobo, 2009). Unsupervised SOM approaches are particularly adapted to ocean colour satellite data, as they solve non-linear problems, they are robust in handling noisy or missing data (e.g., due to cloud cover), and they do not require previous assumptions about the data distribution (Richardson et al., 2003).

2 Data and Methods

2.1 Region of Interest

The present study focuses on the five AZMP boxes of the DFO Maritimes region that are located on the Scotian Shelf within an area bounded by 55 to 71 °W and 40 to 50 °N (Figure 1). These boxes include Georges Bank (GB), Central Scotian Shelf (CSS), Eastern Scotian Shelf (ESS), Western Scotian Shelf (WSS), and Lurcher Shoals (LS).

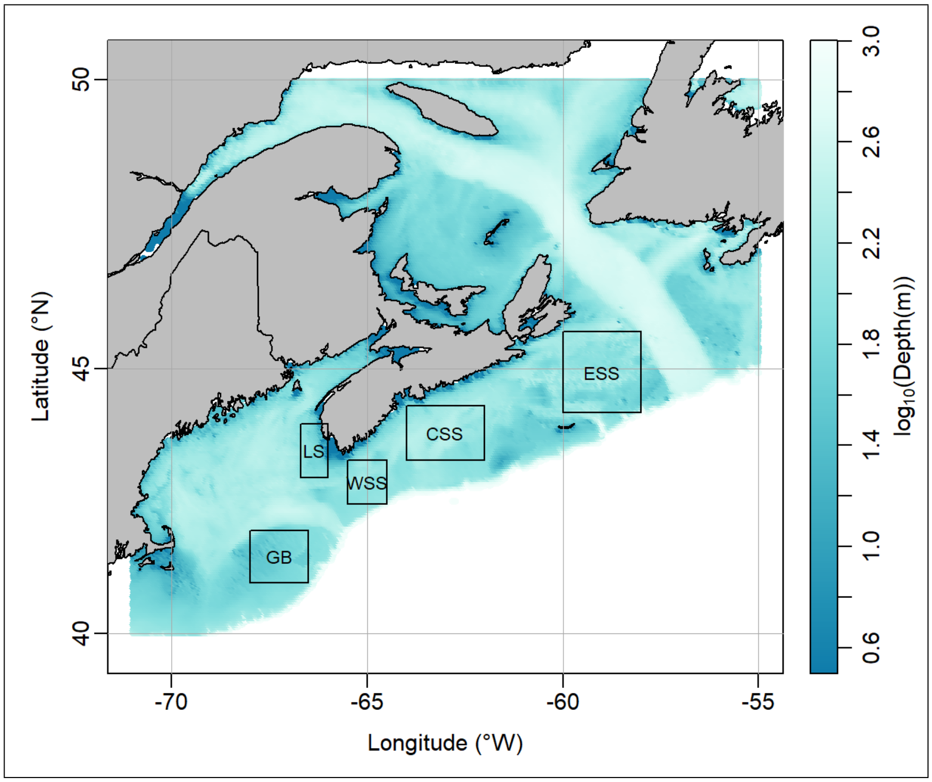

Figure 1: Map of study area including bathymetry shown on a $\log_{10}$ scale and the five AZMP Maritime boxes that are used for reporting of phytoplankton bloom conditions.

2.2 Satellite Data

2.2.1 Sea-surface Temperature (SST) and Chlorophyll-a (Chl-a)

Global seasonal SST and Chl-a Level-3 binned images at 4-km resolution from the MODerate resolution Imaging Spectroradiometer (MODIS) on the Aqua platform were downloaded directly from the National Aeronautic and Space Administration (NASA) website (`https://oceancolor.gsfc.nasa.gov`). The global images were subset to a pan-Canadian grid that includes the Scotian Shelf. The Chl-a concentration was derived by NASA from the standard OC3M algorithm ((O'Reilly et al., 1998)), a detailed description of the algorithm can be found here (`https://modis.gsfc.nasa.gov/data/dataprod/chlor_a.php`, accessed on May 23rd 2024). Table 1 displays the time span of the seasonal MODIS data for both SST and Chl-a.

2.2.2 Bathymetry

Bathymetry data was retrieved from the 'pancan_bath_4km' data set in the 'oceancoloR' package (Clay, 2022). This bathymetry data set was selected given that its binning scheme is consistent with the seasonal

Table 1: Temporal coverage of MODIS data in each season, including the year and day of year (doy).

Season	Year	Day of Year
Spring (SP)	2003 to 2020	080 to 171
Summer (SU)	2002 to 2020	172 to 263
Fall (FA)	2002 to 2020	264 to 354
Winter (WI)	2002 to 2020	355 to 079

SST and Chl-a data binning. More information on the Level-3 product binning scheme can be found the NASA website (https://oceancolor.gsfc.nasa.gov/resources/docs/technical/ocean_level-3_binned_data_products.pdf).

2.3 Methods

All data handling and computational aspects of this project were coded in R language (http://cran.r-project.org/) and are available on a Github repository. Access to the repository is available on demand (https://github.com/BIO-RSG).

2.3.1 Self-Organizing Maps (SOMs)

The R package 'kohonen' was used to produce the SOM for the study area. The input data layers were prepared for the SOM by filtering the data as the input requires that all the data layers have the same number of bins. Bins that contain seasonal Chl-a values greater than 50 $\mathrm{mg\,m^{-3}}$ or bathymetry values greater than 1000 m were removed. Bathymetry data greater than 1000 m was filtered out to focus our SOM analysis on the Scotian Shelf where the Maritime boxes are located. Each bin was checked to ensure that they contained finite values in all input data layers (i.e., seasonal SST, seasonal Chl-a, and bathymetry). Based on the above criteria, a total of 17023 bins were flagged out of 45630 possible bins, giving a remaining 28580 bins for the study area SOM. Seasonal Chl-a and bathymetry values were $\log_{10}$-transformed to decrease the range of variation and decrease the concentrated coastal effects. Chl-a and SST input layers were arranged into two four-column matrices where each column corresponded to a season. In each bin (containing seasonal SST and Chl-a and bathemetry values), each variable was scaled using the following equation:

$$x_{i,scaled} = \frac{(x_i - \bar{x})}{\sigma(x_i)}$$

where $\bar{x}$ corresponds to the mean of $x_1, x_2, ...x_N$ and $\sigma(x_i)$ corresponds to the standard deviation of $x_1, x_2, ...x_N$, with N the total number of bins.

The *somgrid()* function was used to set up a hexagonal topology for the grid of nodes (i.e., center of a cluster) and assign codebook vectors (i.e., bathymetry vector and matrices of seasonal SST and Chl-a) to each node to be used in training. A hexagonal grid was chosen (Figure 2) as it provided smoother maps than a rectangular grid due to an increase in neighbors for a given hexagonal grid cell, six neighbors are considered during the training process while only four neighbors are considered in a rectangular grid. The learning rate was set to decline linearly from 0.05 to 0.01 over a maximum number of iterations, which was dependent on grid size. The maximum number of iterations was determined by manually testing iteration values until the relative distance to the closest unit for each data layer plateaus during training (see 'Training' plot in Figure 2). Weights were added to each property to ensure that the results would be mainly driven by Chl-a and SST (and not bathymetry), such that the input data layers were given weights of 0.4, 0.4, and

0.2 for Chl-a, SST, and bathymetry, respectively. The input data layers, along with the above parameters mentioned (grid, iterations, learning rate, and user weights) were applied to the *supersom()* function where the data training takes place. The *supersom()* function was applied over the *som()* function as it allows for multiple data layers and the use of several variables.

Table 2: Parameters used in the different SOM models.

Grid Size	# of Nodes	# of Iterations
3x3	9	200
4x4	16	400
5x5	25	600
6x6	36	800

The SOM output consists of the training information and classification results, such as node classes and euclidean distances (i.e., sum of squares). Each bin in the study area is ultimately assigned a 'winning' node and the arrangement of the nodes in the study domain dictates spatial clustering of the input data (Figure 2).This process was repeated for four different grid sizes (Table 2) and the optimal grid size was chosen based on its performance. The percent coverage of the overlapping nodes for each Maritime box, as well as comparison of the original box data with the assigned node data, was used as a performance matrix. Winning nodes for the optimal grid size are thereafter referred to as 'Max Nodes'.

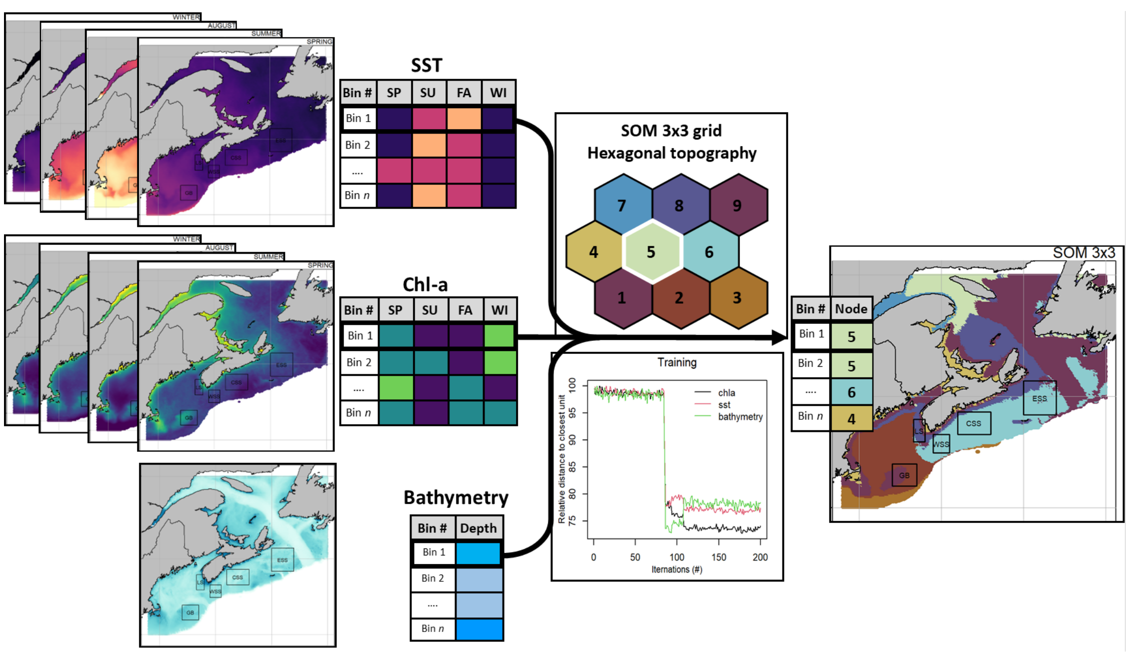

Figure 2: Illustration of the SOM mapping method. The bin node classes are then mapped with the Maritime boxes. For this example, bin 1 is assigned Node 5 for a 3x3 grid sized SOM.

2.3.2 Linear Regression of Seasonal Chl-a and SST in Maritime boxes and Max Nodes

The mean seasonal SST and Chl-a were computed for the Maritime boxes by taking the arithmetic mean of the bin data within the respective box boundaries. The mean seasonal SST and Chl-a were also computed for the Max Nodes by taking the arithmetic mean of all bins assigned to that respective node in the study area. The mean bathymetry was also computed for each Maritime box and respective Max Node and visually compared to the 1:1 line. This was repeated for all SOM grid sizes (3x3, 4x4, 5x5, and 6x6) to detect potential differences between the Maritime boxes and Max Node properties. The SOM grid size that provided the Maritime boxes with Max Nodes that had sufficient overlap with the current Maritime boxes (> 50% coverage) as well as similar seasonal properties was selected to update the boundaries of the Maritime boxes, which we will be referred to as the 'SOM-derived polygon'.

2.3.3 Comparison of SST and Chl-a in Maritime boxes and SOM-derived polygons

The original Maritime boxes were plotted on top of the selected SOM, then using the Max Node boundaries assigned to that respective Maritime box, the SOM-derived polygons were then manually created on the selected SOM image with the chosen grid size using the *drawPoly()* function (R 'raster' package). This allowed the boundaries of the original box to be redefined but also to preserve the broad location of the original box. When possible, SOM-derived polygons were drawn to contain a similar amount of bins (i.e., similar surface area) as the original Maritime box (i.e., similar surface area). The seasonal distribution and variance of SST and CHLA in the Maritime boxes and SOM-derived polygons were compared using a box plot approach, including the median, upper and lower quartiles, whiskers, and potential outliers.

Further, spring bloom characteristics were compared using scorecards, a simple mean to visualize departure from mean conditions. The Maritimes AZMP report on the spring bloom status by using four indices to describe the spring bloom timing and intensity, these indices are the initiation, duration, magnitude, and amplitude. For this report, we focused on the initiation, duration, and amplitude. These scorecards provide a compact format for comparison and are used in Maritimes AZMP reporting (Casault et al., 2024).

3 Results and Discussion

3.1 SOM results

The SOM analysis of various grid sizes provided four emerging regions that remained independent of the total number of nodes, namely, the Gulf of Saint Lawrence, coastal areas, the Scotian Shelf, and the Gulf of Maine (Figure 3 a)). As the number of nodes increased from 9 (3x3 grid) to 36 (6x6 grid), each of these four main regions were subdivided into smaller regions.

The results from each SOM analysis are shown in Figure 3. The 6x6 grid (Figure 3 d) showed the lowest percent cover for all Maritime boxes except for the WSS box with the second highest percent cover (88.6%). The GB box exhibited the highest coverage of a node (64.7% for node 3) with a SOM trained with a 5x5 grid (Figure 3 c), while the LS box reached the highest coverage of a node (86.6% for node 14) with a SOM trained with a 4x4 grid (Figure 3 b). For the Scotian Shelf boxes (i.e., CSS, ESS, and WSS), the percent coverage was highest for the SOM trained with a 3x3 grid with values of 97.6, 81.4 and 98.6%, respectively (Figure 3 a). In general, the percent coverage of a node by a box generally decreased as the number of nodes increased (increased grid size). Table 3 displays the Max Node and percent coverage for each Maritime box from each SOM output.

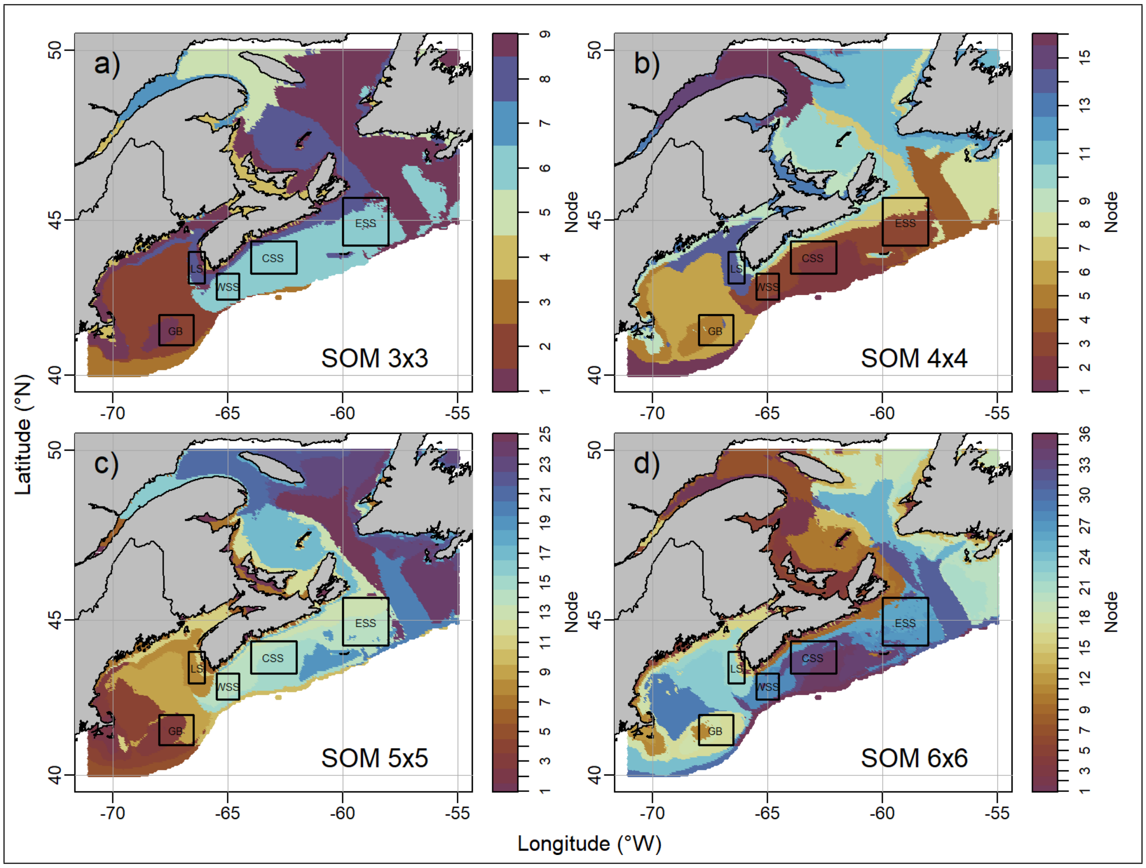

Figure 3: Mapped nodes from each SOM output. The grid size used in the SOM is indicated on each map by the black line rectangles The Maritime boxes are also plotted and labelled.

Table 3: Max Node and percent coverage of each Maritime box for each SOM grid size.

Box	3x3		4x4		5x5		6x6	
	Max node	Coverage (%)	Max node	Coverage (%)	Max node	Coverage (%)	Max node	Coverage (%)
GB	2	56.7	5	58.8	3	64.7	17	46.4
CSS	6	97.6	2	58.8	15	65.1	33	54.5
ESS	6	81.4	3	69.7	14	64.4	26	52.3
WSS	6	98.6	3	87.5	14	86.3	28	88.6
LS	8	65.3	14	86.6	8	80.6	22	79.1

For the SOM with a 3x3 grid, three unique Max Nodes occured in each of the Maritime boxes, with the Scotian Shelf boxes sharing the same Max Node. For both the SOM 4x4 and 5x5 grids, four unique Max Nodes occured in each of the Maritime boxes, with ESS and WSS sharing the same Max Node. It is expected that the ESS and WSS boxes share the same Max Node as the annual and seasonal mean of SST and Chl-a, as well as the average depth, are very similar (Figure 4). Finally, with the SOM 6x6 grid, none of the Maritime boxes shared the same Max Node; however the percent cover of the Max Node was smallest, with the exception of the WSS box, which exhibited the second highest coverage (88.6%, Figure 3). Figure 4 shows the seasonal mean of SST and Chl-a along with the mean bathymetry for all original Maritime boxes. The mean depth of the AZMP boxes varies between 60 (LS) and 160 (CSS) meters. All five boxes show a strong SST seasonal cycle with a variation of 10°C or more between winter and summer, except the LS box, which shows the coolest mean temperature and the smallest seasonal variation, likely due to its proximity to the shore and shallow bathymetry, which temper the seasonal cycle. Chl-a in LS remains relatively high and consistent in all seasons, likely an artifact of terrigenous runoffs that contaminate the satellite signal due to being near shore. All other boxes showed a strong seasonal cycle with highest Chl-a occurring in spring and lowest Chl-a occurring in summer.

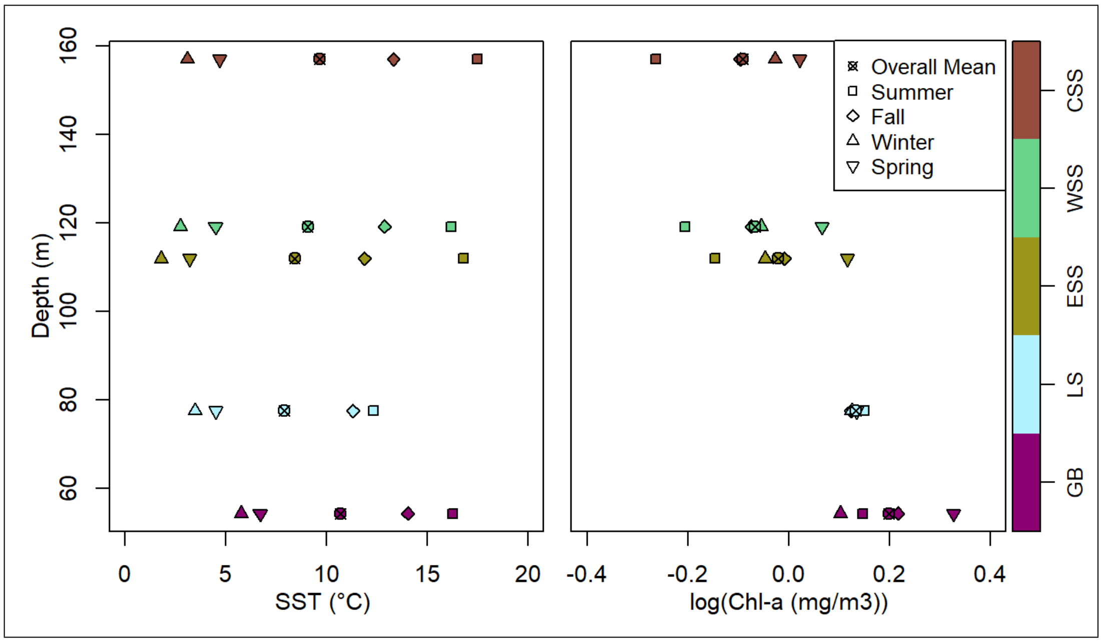

Figure 4: Seasonal and overall mean of SST, Chl-a, and depth for the Maritime boxes. Bathmetry was reverted from log scale for this plot.

The mean distance or quality (as expressed by kohonen) of each Max Node is presented in Figure 5 and represents the average distance between the Max Node and all its members (i.e., bin data values that belong to that node). In general, the trend in the mean distance decreases as the grid size increases. The lower the mean distance, the better the quality or representation of bins in the node. This decreasing trend is expected as the higher the number of grid cells, the closer the members will be to their codebook vectors. The mean distance is reduced by more than half when the number of nodes increases from 9 to 36.

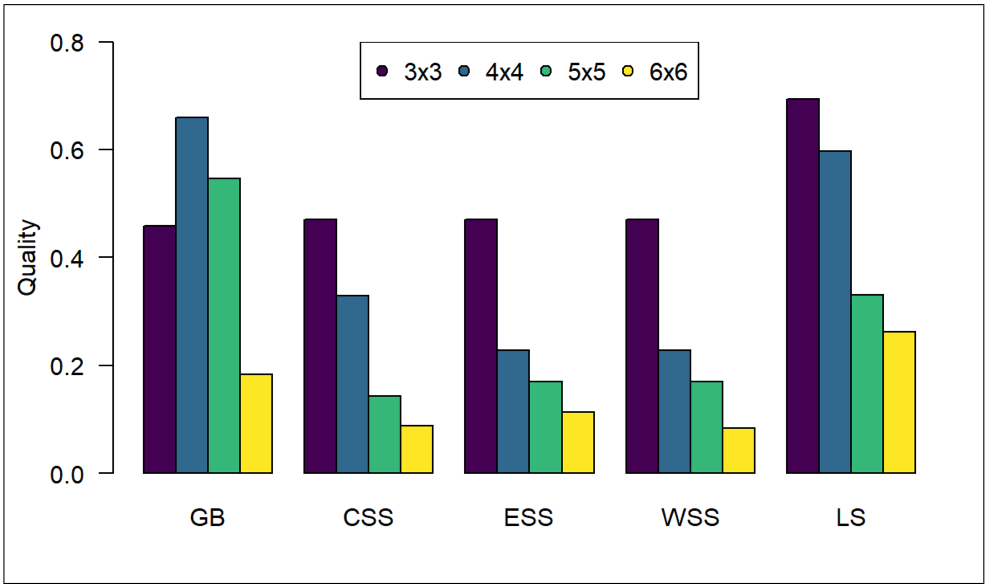

Figure 5: The mean distance or quality of bins assigned to the respective Max Node of each box from each SOM output.

3.2 Seasonal Linear Regression

The slope and R^2 values for the linear regression of the seasonal SST of the SOM polygons against the Maritime boxes showed little changes as a function of the SOM grid sizes for all Maritime boxes with the exception of LS. The LS box in the SOM 3x3 grid size showed the largest slope of 1.34±0.27 and an R^2 of 0.89. The larger grid sizes for the LS box showed little changes in the slope and an R^2 of 0.99 (Table 4). This is likely due to the homogeneity of SST on the Scotian Shelf.

Table 4: Summary statistics for linear regressions of the seasonal SST of the Max Node versus the Maritime box for each SOM grid size.

	3x3					4x4				
Box	Slope	SE	Intercept	SE	adj. R^2	Slope	SE	Intercept	SE	adj. R^2
GB	1.04	0.11	-0.59	1.28	0.97	1.05	0.08	-0.43	0.92	0.98
CSS	0.99	0.01	-0.54	0.07	1.00	0.98	0.01	0.64	0.09	1.00
ESS	0.95	0.02	0.97	0.21	1.00	0.97	0.02	0.74	0.26	1.00
WSS	1.05	0.04	-0.51	0.41	1.00	1.07	0.03	-0.77	0.37	1.00
LS	1.34	0.27	-2.62	2.41	0.89	0.95	0.06	0.79	0.52	0.99
	5x5					6x6				
Box	Slope	SE	Intercept	SE	adj. R^2	Slope	SE	Intercept	SE	adj. R^2
GB	1.04	0.07	-0.37	0.85	0.99	1.03	0.07	-0.29	0.85	0.98
CSS	0.98	0.01	0.50	0.08	1.00	1.01	0.00	-0.29	0.03	1.00
ESS	0.97	0.02	0.67	0.25	1.00	0.99	0.01	0.11	0.13	1.00
WSS	1.07	0.04	-0.84	0.37	1.00	1.04	0.02	-0.35	0.17	1.00
LS	0.97	0.06	0.68	0.51	0.99	0.97	0.06	0.70	0.50	0.99

For seasonal mean Chl-a, the Scotian Shelf boxes showed minimal changes in slope and showed an R^2 of 1.00 for all grid sizes. The GB box also displayed minimal changes in the slope as a function of a grid size except for the SOM 6x6 grid where the slope decreased to 0.88±0.12. The LS box in the SOM 3x3 showed a negative slope of -1.45±3.36, while it ranged from 4.14 to 4.56 for the large grid sizes. The

R^2 remained relatively high (i.e. > 0.75) for all boxes, with the exception of LS, which showed low R^2 for all SOM grids (i.e., R^2 <= 0.4) (Table 5). Lastly, the scatter plot of mean bathymetry values for all boxes except LS in the 4x4 and 5x5 SOM fell closest to the 1:1 line.

Table 5: Summary statistics for linear regression of the seasonal Chl-a of the max node region versus the Maritime box for each SOM grid size. SE is the standard error and adj. R^2 corresponds to the adjusted R^2.

	3x3					4x4				
Box	Slope	SE	Intercept	SE	adj. R^2	Slope	SE	Intercept	SE	adj. R^2
GB	1.10	0.12	-0.18	0.03	0.96	1.09	0.34	0.06	0.07	0.75
CSS	0.94	0.06	-0.00	0.01	0.99	1.10	0.20	-0.04	0.03	0.90
ESS	1.00	0.31	-0.07	0.03	0.75	1.05	0.20	-0.03	0.02	0.90
WSS	1.04	0.17	-0.02	0.02	0.93	1.05	0.09	0.02	0.01	0.98
LS	-1.45	3.36	0.39	0.45	-0.37	4.14	2.38	-0.38	0.32	0.40
	5x5					6x6				
Box	Slope	SE	Intercept	SE	adj. R^2	Slope	SE	Intercept	SE	adj. R^2
GB	1.10	0.29	0.06	0.06	0.82	0.88	0.12	-0.13	0.03	0.95
CSS	1.04	0.13	-0.04	0.02	0.95	1.02	0.07	0.02	0.01	0.99
ESS	1.07	0.14	-0.03	0.01	0.95	1.03	0.04	0.05	0.00	1.00
WSS	1.05	0.08	0.02	0.01	0.98	1.03	0.07	0.00	0.01	0.99
LS	4.46	2.67	-0.48	0.36	0.37	4.56	2.65	-0.49	0.35	0.39

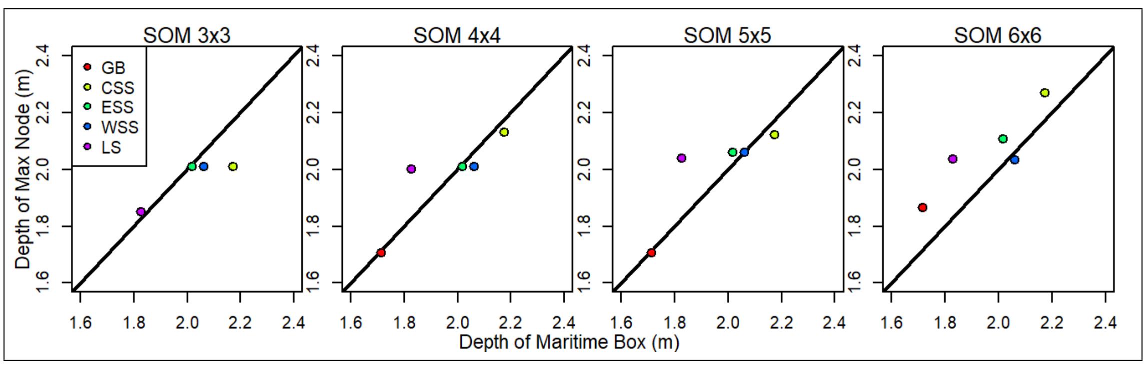

Figure 6: Scatter plots for the mean depth of the respective max node versus Maritime box for each SOM output. Solid black line indicates the 1:1 line. Note that the depth is in logrithmic scale.

3.3 SOM-derived Maritime polygons

The linear regression results of the seasonal mean SST displayed similar slope and R squared results for the SOM 4x4, 5x5, and 6x6 outputs while the SOM 3x3 displayed a larger slope (1.34±0.27) for the LS box. The linear regression results of the seasonal mean Chl-a displayed similar slopes for the SOM 4x4 and 5x5 outputs while the SOM 3x3 showed a negative slope (-1.45±3.36) for the LS box and the SOM 6x6 results displayed the largest slope for the LS box as well as the lowest slope (0.88±0.12) for the GB box. The mean bathymetry data in the SOM 4x4 and 5x5 shows similar results that lies closest to the 1:1 line (Figure 6). Given these results and the agreement with the AZMP boxes as indicated by the percentage cover (Table 3), the SOM 3x3 and 6x6 were discarded from further analysis and only the SOM 4x4 and 5x5 grids were tested to delineate the new maritime boxes. Both the SOM 4x4 and SOM 5x5 outputs provided a total of 4 unique Max Nodes, with ESS and WSS boxes overlapping within the same node (Table 3).

The boundaries of the boxes were refined manually to roughly follow the contour of the Max Node the original box was assigned to and moving away from a rectangular shape to a polygon shape (Figures 7 and 8). The surface area between the original boxes and updated polygons were attempted to be kept consistent. The GB and ESS boxes were the most reduced in surface area to ensure their boundaries would remain within their Max Node boundaries (Table 6).

Table 6: Number of bins or pixels within the Maritime boxes and SOM-derived polygons.

	Number of Bins (#)		
Box	Orignal	SOM 4x4	SOM 5x5
GB	645	423	470
CSS	830	833	777
ESS	1193	809	756
WSS	351	358	339
LS	268	286	287

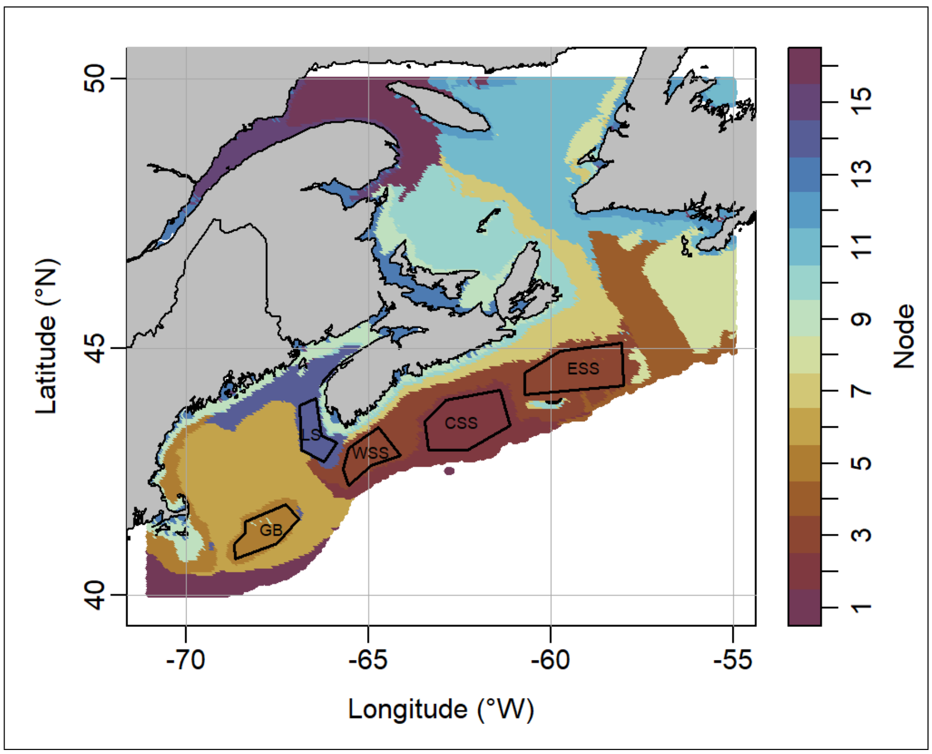

Figure 7: SOM derived Maritime boxes for the 4x4 grid size in the region of interest.

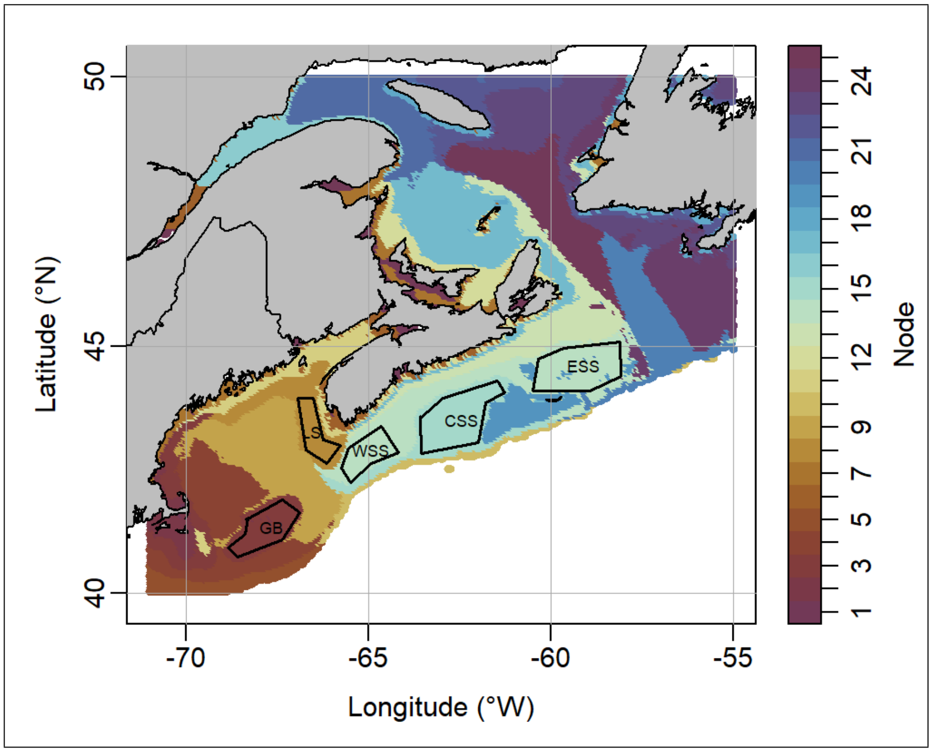

Figure 8: SOM derived Maritime boxes for the 5x5 grid size in the region of interest.

As demonstrated by the linear regression (Tables 4 and 5) and the box plots (Figure 9), the mean seasonal SST and Chl-a remained consistent between the original, the SOM 4x4-derived and SOM 5x5-derived boxes. The main advantage being that the interquartile ranges for the SOM-defined boxes are much smaller, fewer outliers for the SST in the LS box. There are also fewer outliers in the SOM-defined boxes than in the original boxes.

In the Chl-a box plots, all SOM-derived Maritime boxes show smaller interquartile when compared to original maritime boxes. For all boxes except GB, which showed no outliers, the number of outliers in the SOM-derived polygons decreased compared to the original boxes. The median Chl-a of the SOM-derived polygons are also similar in all Maritime boxes.

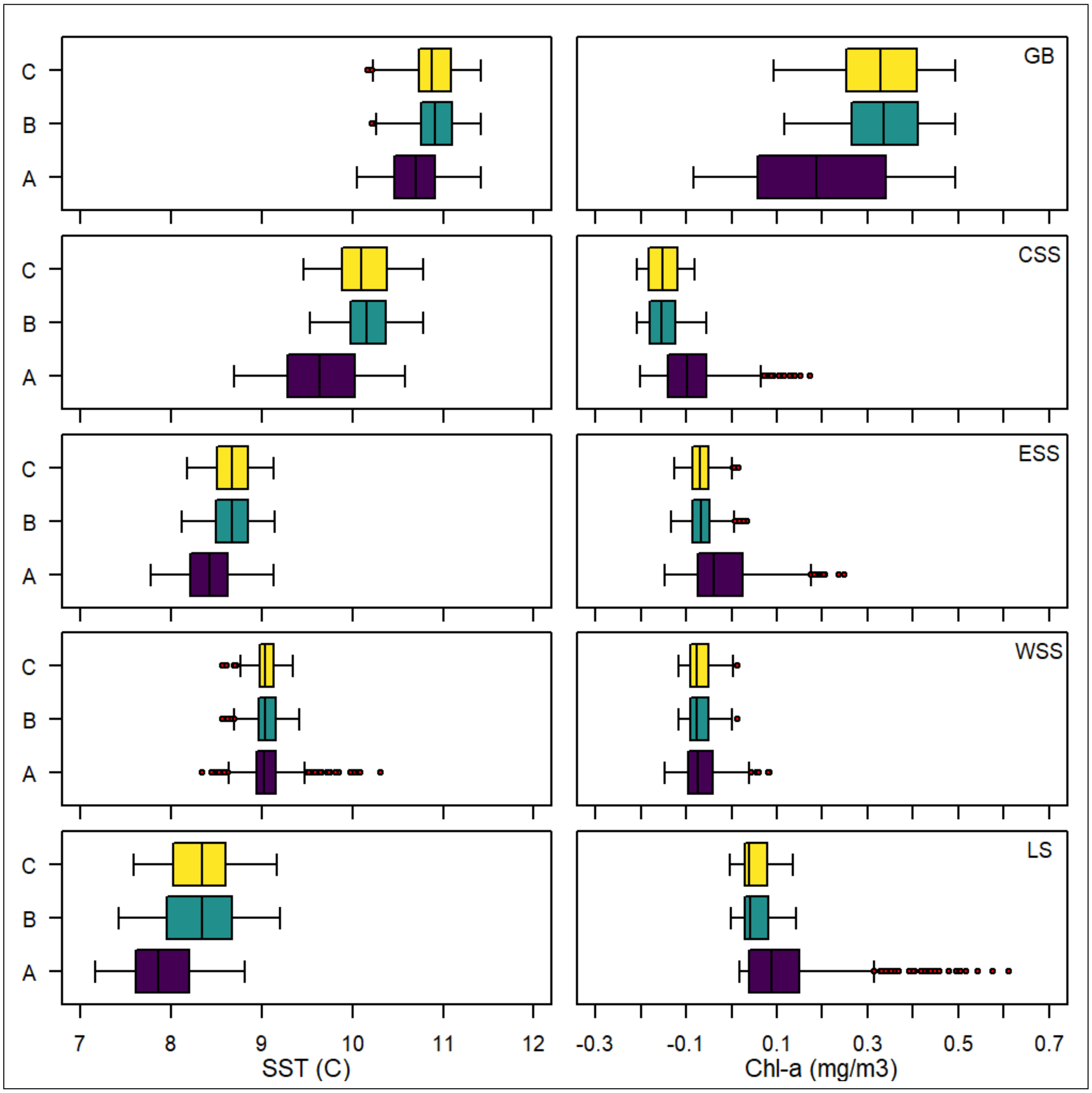

Figure 9: Box plot for mean SST and Chl-a of each bin within the Maritime boxes. The labels A, B, and C represent the original boxes (purple), SOM-derived 4x4 polygons (teal), and SOM-derived 5x5 polygons (yellow). Each row of plots are labelled in the top corner of the Chl-a plots (right column of plots). Red points indicate potential outliers.

3.4 Implication of Refined Box Boundaries for AZMP reporting

The main goal of the study was to improve the representation of the marine ecosystems composing the Scotian Shelf and provide a more accurate seasonal and inter-annual Chl-a signal free of contamination from adjacent ecosystems. This was achieved by defining areas that 1) avoided coastal regions where band-ratio derived Chl-a algorithms perform poorly and 2) did not cover fronts and permanent oceanographic features, as revealed by the SOM analysis. The SOM boxes correspond rather to an update and refinement of the current AZMP boxes than the identification of specific marine ecosystems.

Inter-annual variation in the spring bloom metrics are reported as anomaly (i.e., deviation to normal condition) that are visualized in scorecards. These scorecards were computed for three spring bloom metrics (initiation, duration, and amplitude) for each of the boxes and polygons (Figures 10, 11, 12, 13, and 14). An unusual event is defined when an index anomaly is outside the ±1.5 standard deviation of the mean. The number of unusual events was reduced for all the three parameters from a total of 38 to 27 suggesting that the SOM boundaries of the Scotian Shelf boxes improved the ability to accurately describe the spring bloom in different locations without accounting for changes in water masses and account for the effect of local hydrodynamic processes. For instance, the number of years with a longer bloom duration than normal (i.e., outside ±1.5 standard deviation range) decreased from 15 to 10 with the SOM derived boxes.

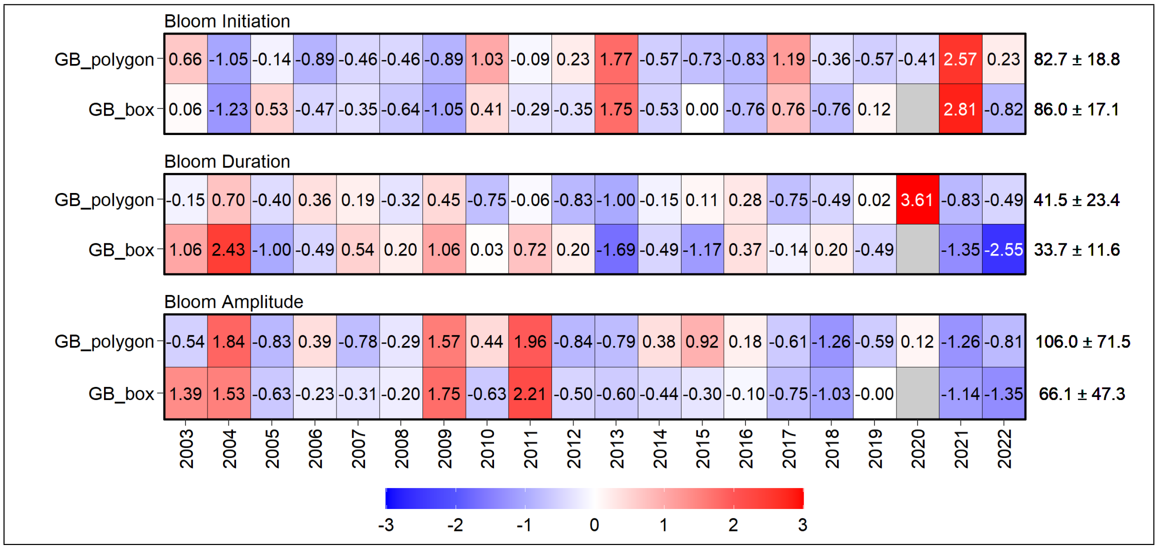

Figure 10: Scorecards for the phytoplankton spring bloom initiation, duration, amplitude and magnitude for the GB box and polygon.

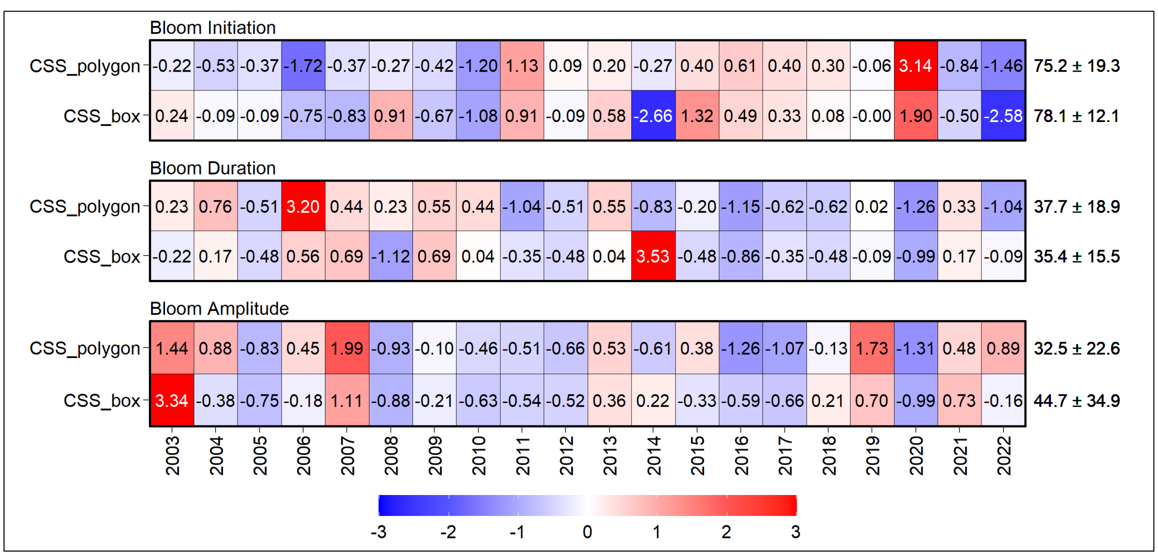

Figure 11: Scorecards for the phytoplankton spring bloom initiation, duration, amplitude and magnitude for the CSS box and polygon.

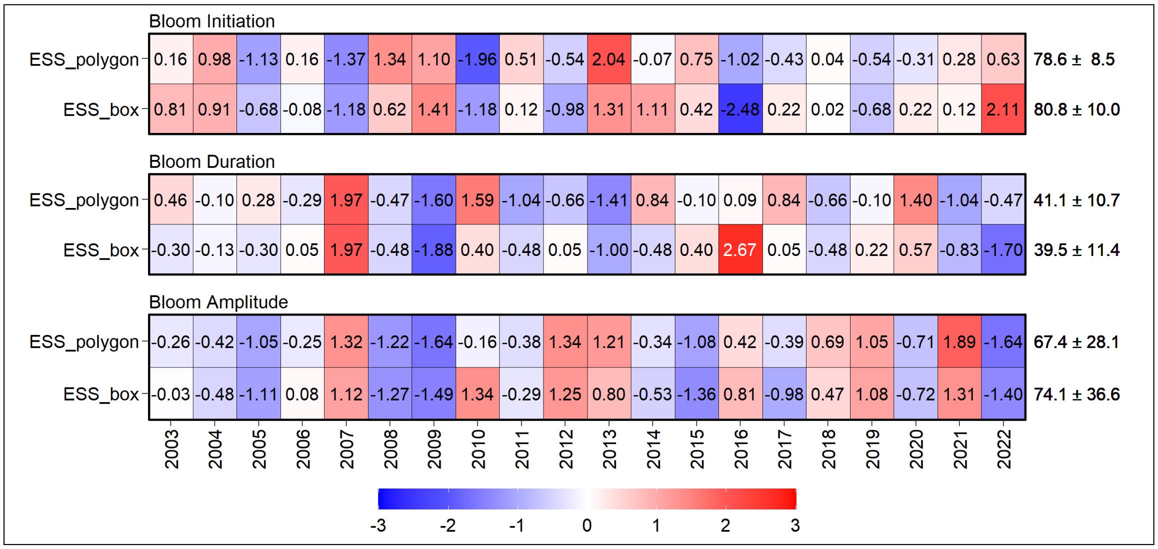

Figure 12: Scorecards for the phytoplankton spring bloom initiation, duration, amplitude and magnitude for the ESS box and polygon.

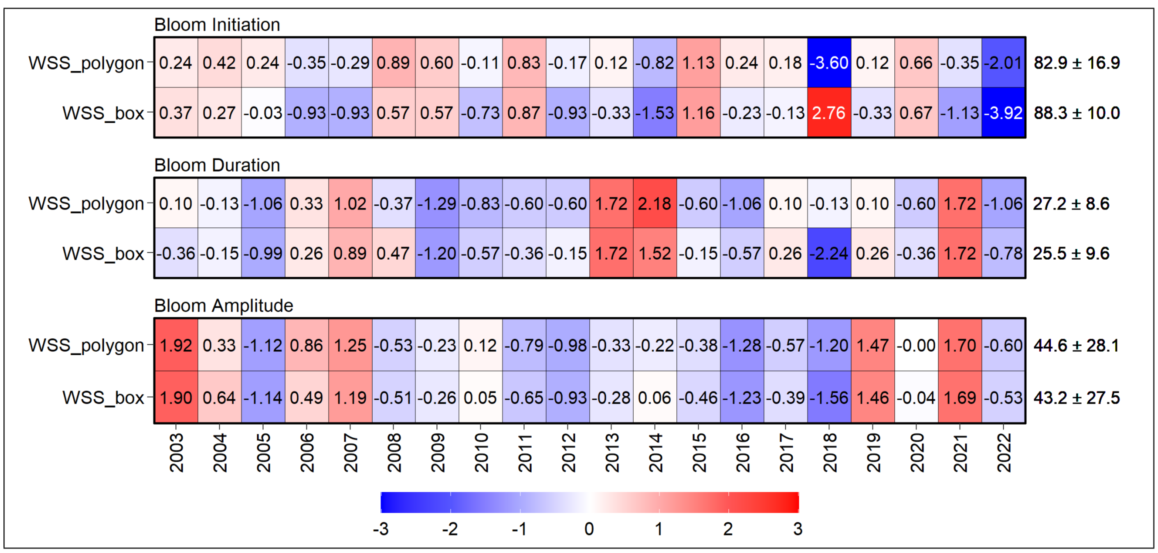

Figure 13: Scorecards for the phytoplankton spring bloom initiation, duration, amplitude and magnitude for the WSS box and polygon.

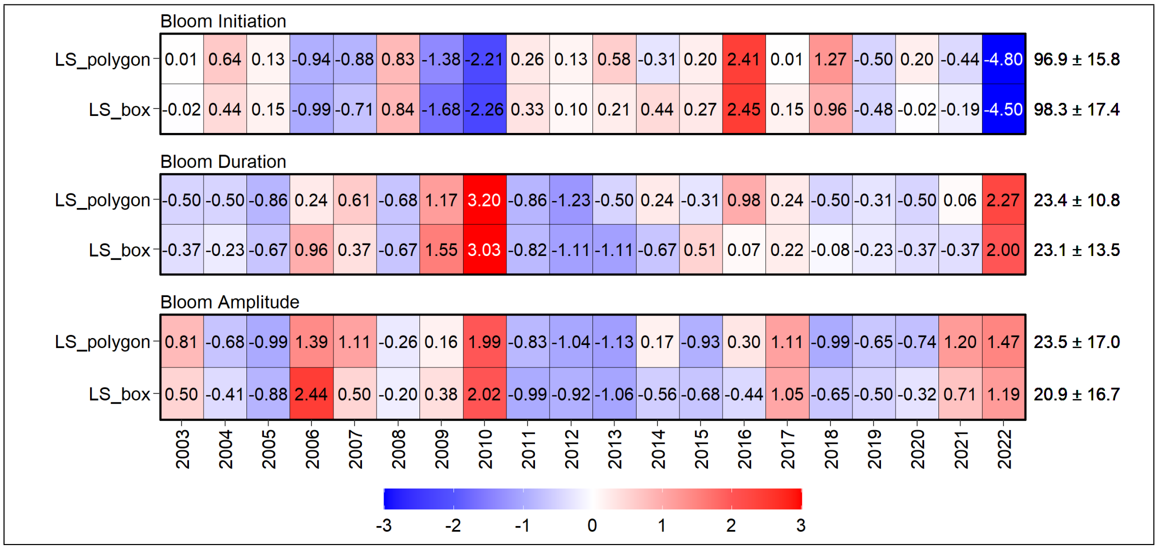

Figure 14: Scorecards for the phytoplankton spring bloom initiation, duration, amplitude and magnitude for the LS box and polygon.

4 Conclusion

The current reports aimed at refining the boundaries of the boxes for satellite Chl-a data analysis on the Scotian Shelf for the DFO Maritime regions. We found that an objective approached, Self Organizing Maps, applied to seasonal SST and Chl-a, weighted by bathymetry, provided a definition of the boxes that is more representative of mesoscale and large scale spatial patterns in environmental variability, by avoiding coastal areas and overlapping different marine ecosystems. While the new boxes remained consistent with the original definition, as attested by their mean annual properties, the mean Chl-a within the boxes showed less variation in the seasonal and interannual signal. This improvement was also reflected in the spring bloom metrics computed for the AZMP reports with less anomalous events observed when the new definition was use. We recommend using the new polygons informed from our SOM analysis in future reporting of the Maritimes AZMP.

Acknowledgement

We want to thank Andrea Hilborn for SOM coding support and Stephanie Clay for producing bloom metric data.

References

Casault, B., Beazley, L., Johnson, C., Devred, E., and Head, E. (2024). *Chemical and Biological Oceanographic Conditions on the Scotian Shelf and in the Eastern Gulf of Maine During 2022*. Fisheries and Oceans Canada, Maritimes Region.

Clay, S. (2022). *oceancolouR: Various Functions for Satellite and Data Analysis*. R package version 0.0.0.9000.

DFO (2021). Oceanographic conditions in the Atlantic zone in 2020. *DFO Can. Sci. Advis. Sec. Sci. Advis. Rep.*, (2021/026).

Kohonen, T. (2001). Self-organizing maps (third ext. edition). *New York*.

Lobo, V. J. (2009). Application of self-organizing maps to the maritime environment. In *Information Fusion and Geographic Information Systems*, pages 19–36. Springer.

Oja, M., Kaski, S., and Kohonen, T. (2003). Bibliography of self-organizing map (som) papers: 1998–2001 addendum. *Neural computing surveys*, 3(1):1–156.

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C. (1998). Ocean color chlorophyll algorithms for seawifs. *Journal of Geophysical Research: Oceans*, 103(C11):24937–24953.

Richardson, A. J., Risien, C., and Shillington, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59(2-3):223–239.

Wehrens, R., Buydens, L. M., et al. (2007). Self-and super-organizing maps in r: the kohonen package. *Journal of Statistical Software*, 21(5):1–19.

Wehrens, R. and Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software*, 87(1):1–18.