2015

# Characterizing local biological hotspots in the Gulf of Maine using remote sensing data

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

# CHARACTERIZING LOCAL BIOLOGICAL HOTSPOTS IN THE GULF OF MAINE USING REMOTE SENSING DATA

by

**MARTA M RIBERA**

B.A., Universitat Autònoma de Barcelona, 2002

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2015

Approved by

First Reader _____

Sucharita Gopal, Ph.D.
Professor of Earth and Environment

Second Reader _____

Les S. Kaufman, Ph.D.
Professor of Biology

Third Reader _____

Crystal B. Schaaf, Ph.D.
Adjunct Research Professor of Earth and Environment

## DEDICATION

To my family and friends, for encouraging me to keep swimming forward.

## ACKNOWLEDGMENTS

This thesis would not have been possible without the financial support from the EPA STAR Fellowship and Boston University's Earth and Environment Department. I want to thank my advisor, Dr. Sucharita Gopal, for her guidance and support during all these years. I also want to thank my committee members, Dr. Les Kaufman, Dr. Dana Bauer, Dr. Crystal Schaaf, and Dr. Scott Gallager, for providing the direction and feedback to finish this thesis. Also, a big thank you to Ben Cowie-Haskell and Mike Thompson at the Stellwagen Bank National Marine Sanctuary, Amber York at WHOI, Dr. Zhongping Lee and Dr. Nima Pahlevan at University of Massachusetts, Wendy Leo at MWRA, Kerry Lagueux and Brooke Wikgren at the New England Aquarium, and Dr. Zhe Zhu and Dr. Eli Melaas at Boston University for their help acquiring and processing the datasets used in this thesis. MERIS dataset in this thesis was provided by NASA and ESA. AATSR data was provided by the GHRSST.

On a personal note, I want to thank Valerie Pasquarella, Rachel Nalepa, Conor Gately, Chris Holden, and Tyler Nesbitt for taking the time to listen to my ideas and help me through this PhD process. To Andrew David and Dr. Enric Cortés from NOAA Fisheries, thank you for your mentorship during the last ten years. Big thanks to the Hofstetter and Morgan families for providing me a home away from home. Finally, I want to thank my parents, Jaume and Dolors, my sister Anna, Albert, Julia, Quim, and my partner in crime Jake, for unconditionally supporting me throughout these years.

**CHARACTERIZING LOCAL BIOLOGICAL HOTSPOTS**

**IN THE GULF OF MAINE USING REMOTE SENSING DATA**

**MARTA M RIBERA**

Boston University Graduate School of Arts and Sciences, 2015

Major Professor: Dr. Sucharita Gopal, Professor of Earth and Environment

ABSTRACT

Researchers increasingly advocate the use of ecosystem-based management (EBM) for managing complex marine ecosystems. This approach requires managers to focus on processes and cross-scale interactions, rather than individual components. However, they often lack appropriate tools and data sources to pursue this change in management approach. One method that has been proposed to understand the ecological complexity inherent in marine ecosystems is the study of biological hotspots. Biological hotspots are locations where organisms from different trophic levels aggregate to feed on abundant supplies, and they are considered a first step toward understanding the processes driving spatial and temporal heterogeneity in marine systems. Biological hotspots are supported by phytoplankton aggregations, which are characterized by high spatial and temporal variability. As a result, methods developed to locate biological hotspots in relatively stable terrestrial systems are not well suited for more dynamic marine ecosystems. The main objective of this thesis is thus to identify and characterize local-scale biological hotspots in the western side of the Gulf of Maine. The first chapter describes a new methodological framework with the steps needed to locate these types of hotspots in marine ecosystems using remote sensing datasets. Then, in the second chapter

these hotspots are characterized using a novel metric that uses time series information and spatial statistics to account for both the temporal variability and spatial structure of these marine aggregations. This metric redefines biological hotspots as areas with a high probability of exhibiting positive anomalies of productivity compared to the expected regional seasonal pattern. Finally, the third chapter compares the resulting biological hotspots to fishery-dependent abundance indices of surface and benthic predators to determine the effect of the location and magnitude of phytoplankton aggregations on the rest of the ecosystem. Analyses indicate that the spatial scale and magnitude of biological hotspots in the Gulf of Maine depend on the location and time of the year. Results also show that these hotspots change over time in response to both short-term oceanographic processes and long-term climatic cycles. Finally, the new metric presented here facilitates the spatial comparison between different trophic levels, thus allowing interdisciplinary ecosystem-wide studies.

## PREFACE

Persistent overfishing in ocean environments is severely impacting the health of these ecosystems and, in turn, the livelihoods of people depending on their resources (Chandler et al. 1996, FAO 2009). This has had many consequences including the reduction of biomass, elimination of keystone species (i.e. those species playing a central role in an ecosystem) (Botsford et al. 1997), entangling and killing of marine mammals, birds, and non-target fish species (i.e. "bycatch") (Dayton et al. 2002), destruction of benthic habitat (Auster 1998), and alteration of predator-prey relationships (Pauly et al. 1998, Christensen et al. 2003). With many trophic levels affected at the same time, marine ecosystems are losing part of their biodiversity and functionality and, in turn, the ability to return to a desirable state after a perturbation (i.e. resilience) (Worm et al. 2006). More vulnerable marine ecosystems mean that economies dependent on them are highly vulnerable as well.

Historically, managing marine resources in US waters has been contentious, partly because of the complexity inherent in all levels of the system (Steneck and Wilson 2010). From an ecological point of view, this complexity is due in part to the spatially and temporally dynamic distributions of its species (Crowder and Norse 2008). These heterogeneous distributions likely depend on a combination of factors, including topographic, oceanographic, and biological characteristics of the system (Hyrenbach et al. 2000). Moreover, the patchiness of biological aggregations is not only apparent at the surface (i.e. epipelagic zone) of the ocean, but also throughout the water column (i.e. mesopelagic zone) and at the seafloor (i.e. benthos) (Norse et al. 2005). Today, ocean

governance is further complicated by the mismatch often present between the spatial scale at which biological systems operate and the structure of human institutions (Costanza et al. 1998, POC 2003, Crowder et al. 2006). For example, the Gulf of Maine (northwest Atlantic ocean) is divided between Canadian and US waters, and the latter are further subdivided into State (coastline to 3 nautical miles offshore), Federal (3 to 200 nautical miles offshore) and International waters (beyond 200 nautical miles offshore). Even though these subdivisions made sense historically and politically at one point, they are not adequate at managing ocean resources and human activities benefiting from them.

One of the solutions proposed to deal with the complexity on both ecological and human systems is an ecosystem-based approach to management (Christensen et al. 1996, POC 2003, Pikitch 2004, Norse et al. 2005). This approach, named Ecosystem-Based Management (EBM), considers all components of an ecosystem, including the complex interactions between its different parts in the context of their environment. As part of ecosystems, humans are included in this approach, taking into account both the services they need from the system and the cumulative impacts of their actions on this system. As McLeod and Leslie (2009) pointed out, EBM principles are "grounded in the idea that ultimately we are managing people's influences on ecosystems, not ecosystems themselves". Embracing EBM principles implies a change in the way natural resources are managed, focusing on processes and cross-scale interactions instead of individual parts of the system, and planning based on unambiguous long-term objectives instead of immediate crises (Cochrane 2000, McLeod and Leslie 2009). This represents a dramatic change in approach that has yet to gain traction in policy circles, in part because the

research community has not yet developed the kinds of data and methods of dissemination that would enable managers to understand the interconnectedness between ecological, social and economic parts of the system (Murawski 2007) and act accordingly. This does not mean that current data collection efforts should be abandoned, but that data sources need to be integrated across disciplines, and analyzed at appropriate (i.e. biologically relevant) geographic scales.

From an ecological point of view, most of the complexity present in marine ecosystems is due to the temporal and spatial heterogeneity of their resources (Crowder and Norse 2008). This complicates the analysis of ecosystem processes, as spatial data have a set of unique characteristics that prevent them from being analyzed using traditional statistical tools (Anselin 1989, LeSage 1999). These can be generally explained by Tobler's First Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Thus, values from points that are closer together are likely to be more related than values from points that are farther apart. This is known in spatial statistics as "spatial dependency" or "spatial autocorrelation", and it violates the key assumption of traditional statistics that all observations are independent and follow identical distributions. Moreover, spatial data also exhibits "spatial heterogeneity" or "spatial non-stationarity", in which relationships and behaviors between variables vary across space. Consequently, errors and uncertainty vary across space as well, usually resulting in a spatially clustered distribution that violates the assumption in traditional statistics of constant error variances (i.e. homoskedasticity). When these two characteristics of spatial data are ignored and false

assumptions of independence and homogeneity are forced into the analysis of data, specification errors can occur and the results of the analysis may not be accurate (Shekhar et al. 2011).

The presence of spatial heterogeneity in marine ecosystems implies that resources are distributed unevenly across space. A consequence of this heterogeneity is the existence of biological hotspots. Biological hotspots are locations where different trophic levels temporally and spatially overlap (Hyrenbach et al. 2000, Worm et al. 2003, Davoren 2007). Numerous species from different trophic levels aggregate in these locations often to feed on abundant food supplies (Hooker and Gerber 2004), making these areas potentially critical for the long-term resilience of adjacent marine organisms. Understanding the dynamics of hotspots is a non-trivial issue for fisheries managers. Increasingly popular marine activities happening at the surface might affect processes in deeper areas (Grober-Dunsmore et al. 2008). Thus, after more data become available, different management tools, such as "vertical zoning", may need to be considered. Hotspots may also spur increased human activity, particularly if they contain commercially valuable species, and consequently be at greater risk of negative impacts. Any resulting losses of biomass in these areas could then impair adjacent populations of marine organisms that depend on them, affecting in turn other levels of the ecosystem. Therefore, prioritizing available resources for the protection of these hotspots may be warranted (Norse et al. 2005).

Ocean primary producers (mainly phytoplankton organisms) sustain these hotspots by fixing carbon dioxide dissolved in water into organic compounds and

incorporating nutrients into living tissues. Phytoplankton is a broad group of microorganisms that includes species from different divisions, such as Cyanobacteria, *Pyrrhophyta* (including dinoflagellates) or *Bacillariophyta* (including diatoms). Some species are non-motile but fast-growing (e.g. Diatoms) and others are slow-growing and motile, usually migrating vertically through the water column in response to light (e.g. Dinoflagellates). Growth of phytoplankton species is limited by several environmental factors, including light, temperature, and nutrient availability (Cibik et al. 1998). Phytoplankton live in the photic zone, which is the layer of the water column that is exposed to sunlight, as they need sun energy for the photosynthetic process. Also, similarly to other living organisms, they require various inorganic components (such as phosphorus, nitrogen, and iron) and organic nutrients (such as vitamins) to grow. Temperature can also limit the growth of primary producers, with growth rates generally increasing exponentially as temperatures increase (Eppley 1972). However, different species of phytoplankton react differently to these factors (Cibik et al. 1998). Consequently, the composition of phytoplankton assemblages is often characterized by strong spatial variability.

Overall phytoplankton biomass follows a seasonal cycle, with the lowest concentration in winter, followed by a spring bloom, a decrease during the summer, and finally a fall bloom (Thomas et al. 2003). Compared to land ecosystems, where plants are adapted to an annual cycle, phytoplankton organisms tend to have a faster cycle (usually around 100 times per year) as a consequence of their own growth rate and the grazing pressure of planktivorous species (Calbet and Landry 2004). The spring phytoplankton

bloom occurs when the rise in temperatures and higher sunlight create a thermocline trapping nutrients near the surface of the ocean, providing the right conditions for phytoplankton to grow quickly (Townsend and Cammen 1988). The input of freshwater from rivers and nutrients from coastal upwelling also help the quick increase of phytoplankton biomass. The spring bloom can last from just a few weeks to months, depending on nutrient availability and grazing pressure (Winder and Cloern 2010). Once surface nutrients are depleted and stratification prevents the influx of more nutrients to the surface, the bloom collapses and the overall phytoplankton biomass declines. The fall bloom, by contrast, is usually less intense than the spring bloom, and tends to occur during the early fall months due to an excess of nutrients near the surface of the ocean.

In general, the abundance and distribution of phytoplankton across space depend on processes that either stimulate surface productivity or concentrate productivity produced elsewhere (Mackas et al. 1985). Most of the nutrients in the Gulf of Maine come from an inflow of deep slope water through the Northeast Channel (Lynch et al. 1996). When deep currents of cold and nutrient rich water meet prominent seafloor irregularities, the water is forced to mix, promoting upwelling (Mackas et al. 1985). The rise of nutrient-rich waters to the surface stimulates primary productivity, simultaneously increasing turbulence and mixing of the water column. Topographic features can also enhance vertical mixing when they interact with tidal waves in highly stratified water (especially during the summer), producing what are called "internal waves" (Scotti and Pineda 2004). Internal waves can also concentrate phytoplankton created elsewhere by forming pockets of warm water rich in primary producers. Fronts can also increase

phytoplankton patchiness (Wolanski and Hamner 1988). These oceanographic features are characterized by waters of different temperatures and salinities that meet, producing a steep gradient that can potentially function as a barrier for plankton species. Similarly, Langmuir cells, which are produced by surface winds blowing over calm seas, can also concentrate phytoplankton in a specific location, although on a smaller scale (Stavn 1971).

The strength, timing, and location of phytoplankton blooms have the potential to determine the characteristics of the rest of the system. Higher primary production often leads to increased biomass of other species higher up the food chain (Norse et al. 2005, Worm et al. 2003, Nur et al. 2010, Incze et al. 2010). Growth and retention of phytoplankton leads to zooplankton blooms, which attract planktivorous species and ultimately larger predators. Changes in the timing of these phytoplankton blooms can also affect the success of other levels of the trophic chain. For example, certain species of zooplankton, such as the copepod in the genus *Calanus*, have adapted their life cycle, specifically the timing of their early stages, to coincide with the spring phytoplankton bloom and the resulting abundance of food (Daly and Smith 1993). If the spring bloom occurs earlier than it usually does, phytoplankton can grow without any grazing pressure during the early stages of the bloom. By the time zooplankton organisms arrive, nutrients are already mostly depleted so the phytoplankton population is not at its maximum abundance. Moreover, the rate at which zooplankton feeds on phytoplankton determines the amount of energy transferred to higher trophic levels and the quantity left either to be

consumed by microorganisms or to settle to the bottom to be used by benthic communities (Townsend and Cammen, 1988).

In recent years there has been much discussion about the potential effects of climate change on the timing and abundance of phytoplankton concentrations, and how these changes affect the rest of the system (Sarmiento et al. 2004, Navarro et al. 2011). Previous research suggests that higher temperatures may decrease global abundance of phytoplankton by increasing ocean stratification and thereby inhibiting the flow of nutrients from deeper layers of the water column (Boyce et al. 2010), although other studies suggest that such a decline in productivity might be due to changes in wind patterns (Ueyama and Monger 2005, Ji et al. 2010) or increased cloudiness (Nixon et al. 2009). Moreover, changes in the input of freshwater may change the timing of the phytoplankton blooms (Ji et al. 2007). Researchers have also pointed out that higher temperatures tend to favor cyanobacteria over other groups of phytoplankton, increasing the probability of harmful algal blooms (Paerl and Huisman 2008). These changes in timing and abundance of phytoplankton will probably trickle down the trophic chain, having consequences on the whole ocean ecosystem.

Using a combination of traditional and spatial statistics principles, the overall objective of this thesis is to locate primary productivity hotspots and study their spatial and temporal characteristics. The next sections explain the three chapters that form this thesis, which are structured as independent research projects. The main research objectives of this thesis are the following:

1. Design a methodological framework for identifying productivity hotspots in ocean systems using remote sensing data.

2. Examine the spatial and temporal variability in the magnitude of local productivity hotspots in the western Gulf of Maine

3. Explore the spatial and temporal overlap between primary producers and both surface and benthic fish organisms

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER ONE: LOCATING BIOLOGICAL HOTSPOTS IN MARINE ECOSYSTEMS USING REMOTE SENSING DATA: METHODOLOGICAL CHALLENGES AND BEST PRACTICES

## ABSTRACT[1]

The spatial heterogeneity of most ecosystems adds complexity to the management of their resources. This heterogeneity is often seen in the form of hotspots, which are generally described as aggregations of organisms in space. Hotspots are signs of underlying processes occurring in an ecosystem. Locating and characterizing these hotspots are great steps towards a better understanding of the complexity inherent in these systems. However, while hotspot analyses are common in land-based studies, they are rare over the ocean. This may be due to the highly dynamic spatial and temporal distribution of phytoplankton organisms, which directly support hotspots. Also, the challenging nature of data collection in ocean ecosystems further complicates this type of analysis. Remote sensing data can greatly benefit the analysis of hotspots in marine ecosystems. However, methods need to be adapted for the use of this type of dataset in ocean systems. To encourage researchers to study hotspots in marine ecosystems, I present a simplified methodological framework that explains the steps needed for this analysis, describes the potential challenges in ocean systems compared to land-based ones, and explores possible solutions. This study is not intended to be an extensive

[1] To be submitted to "Frontiers in Ecology and the Environment". Structure matches the publication's guidelines.

review of all aspects of hotspot analyses or remote sensing, but a way to facilitate the introduction to researchers of this type of methodology.

## INTRODUCTION

Most ecosystems are spatially heterogeneous (Levin 1992), with species and habitats distributed unevenly across space. Management plans often over-simplify this heterogeneity, creating a mismatch between the scale at which systems operate and the scale at which they are managed (Crowder et al. 2006). One of the best approaches to visualizing the uneven distribution of resources in ecosystems is the detection of hotspots. Hotspots are generally defined as locations where multiple organisms aggregate (Malakoff 2004). When these organisms represent different species, the resulting hotspots are known as "biodiversity hotspots" (e.g. Myers et al. 2000). In contrast, more general aggregations of organisms, independent of species, to feed on abundant resources result in "biological hotspots" (Davoren 2007). Both biodiversity and biological hotspots result from a combination of environmental, ecological and human processes that stimulate or aggregate resources in one particular location. Finding and characterizing hotspots therefore provides critical information on the processes driving spatial heterogeneity in ecosystems (Norse et al. 2005), and in turn may be key to successfully managing these systems.

Hotspot analyses are common in terrestrial ecosystems. For example, hotspots have been used to map land cover classifications (Gould 2000) and to find areas with

high concentrations of bird species (Seto et al. 2004). However, in ocean systems, hotspot methods have been largely under-utilized. One reason for this may be the different biological nature of marine and terrestrial resources. In the case of ocean ecosystems, biological hotspots are supported by high concentrations of phytoplankton organisms (Stevick et al. 2008). These primary producers have a life cycle estimated to be in the order of 100 times faster than their land-based counterparts (e.g. grasses and trees) (Winder and Cloern 2010). As a result, marine hotspots are highly variable in time, with phytoplankton aggregations lasting from a few days to months. Furthermore, the distribution of marine primary producers is highly dependent on dynamic oceanographic processes, such as fronts and internal waves (Mackas et al. 1985). Given that these processes aggregate productivity at various spatial scales, phytoplankton aggregations may span from a few meters to several kilometers. Thus, the high temporal and spatial variability of marine systems complicates the process of deriving local hotspots using techniques developed for more persistent and sessile land-based systems.

Another barrier to the use of hotspot analyses in ocean ecosystems is the availability of suitable data. Because of its large surface area (about 70% of the Earth's surface area) and difficult access, in-situ data collection in ocean systems is often resource-intensive and time-consuming. As a result, very few field-data sets have both the spatial coverage (i.e. size of area of study) and spatial resolution (i.e. number of data points per unit of distance) required for hotspot analysis. Remote sensing data, on the other hand, may be able to meet these requirements (Wulder and Boots 1998). Ocean satellite sensors capture the spectral reflectance of surface ocean waters, which are then

converted to chlorophyll-a concentrations (O'Reilly et al. 1998). These chlorophyll-a values are then used as a proxy for abundance of surface primary producers. The most popular satellite sensors used to study the ocean are the SeaWiFS (Sea-Viewing Wide Field-of-View Sensor), MERIS (Medium Resolution Imaging Spectrometer) and MODIS (Moderate-Resolution Imaging Spectroradiometer) sensors. These datasets have a relatively high resolution (300m – 1000m), a high repeat frequency (2-3 days), and provide global coverage. Also, since these data sources are currently available for free, they may become a very suitable option for locating ocean productivity hotspots. However, because of the different spatial and temporal characteristics of the processes occurring over the ocean compared to land, deriving ocean hotspots using this type of remote sensing dataset will require updated methods and approaches. With proper guidance, more ocean researchers may be encouraged to pursue hotspot analyses as a first step towards unraveling the spatial and temporal drivers of heterogeneity in marine ecosystems.

Here I describe a methodological framework for identifying productivity hotspots in ocean systems using remote sensing data. This framework presents a series of questions that guide the user through the steps, methodological challenges, and potential solutions of a hotspot analysis. This study does not intend to provide a thorough review of hotspot analysis or remote sensing applications. However, I hope it provides enough insight on the process of deriving ocean local hotspots so researchers are encouraged to use this type of analysis in the future for their areas of interest.

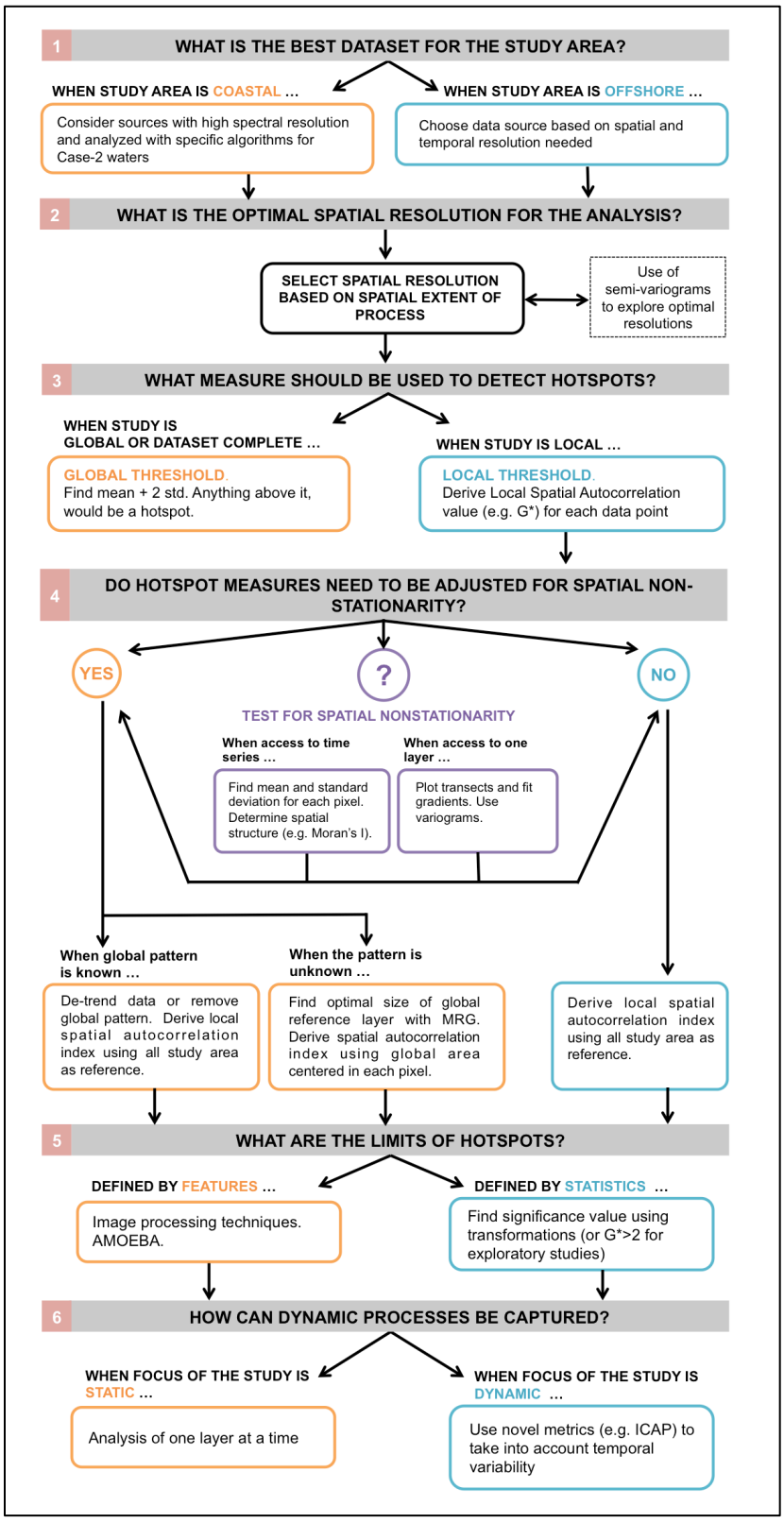


**Figure 1 – Methodological framework to derive hotspots in ocean ecosystems.**

# FINDING MARINE HOTSPOTS

I created a basic methodological framework (Figure 1) that guides users through the process and decisions needed to locate marine hotspots using remote sensing data. This framework is comprised of six structured questions that allow users to explore critical choices needed to pursue this type of analysis. While I focused this framework on identifying local productivity aggregations, these processes could be easily generalized to other types of datasets. The following sections explain each of the questions and their possible answers.

## *1. WHAT IS THE BEST DATASET FOR THE STUDY AREA?*

The choice of source of remote sensing data for hotspot analyses depends in part on the location of the study area. Different sensors have very different characteristics, from the placement of their bands to their resolution, making them suited for certain locations but not recommended for others (see Turner et al 2003 for a review). The choice of sensor is especially critical in coastal study areas. In shallow areas along the coasts, light travels through clear ocean waters, reaches the seafloor, and it is often reflected back towards the surface (Lee et. al. 1998). As a result, the radiation that reaches the satellite sensor comes from the surface of the ocean confounded by the properties of the rest of the water column. This contaminated surface reflectance signal typically results in an overestimation of chlorophyll-a values. Sensors with higher spectral resolution (i.e. narrower bands) are more capable of separating the specific

frequencies reflected by phytoplankton pigments (Bricaud et al. 1999). Consequently, these types of sensors may be able to mitigate the problem of light reflected by seafloor, so the results in coastal waters may have less uncertainty. However, even when using these sensors, there is still a possibility some of the results may be affected.

Coastal waters are also known for being turbid (i.e. high number of dissolved particulates) (Maritorena et al. 1994). Turbid waters (i.e. "Case-2 waters") tend to reflect light at a higher rate than clear waters (i.e. "Case-1 waters") (Morel and Prieur 1977). Because many remote sensing algorithms are calibrated using clear water properties, the resulting chlorophyll-a values in turbid areas may not be fully accurate. New algorithms have been tested to be able to overcome some of the consequences of both turbid (e.g. Schroeder et al. 2007) and shallow waters (e.g. Lee et al 1999). However, there is still much discussion on the accuracy of some of these methods, so more research is needed.

Due to its high spectral resolution and the placement of its bands along the electromagnetic spectrum, MERIS is often recommended in coastal areas (Doerffer et al. 1999). However, the European Space Agency lost contact with MERIS in April 2012, so newer studies will likely be limited to the use of MODIS data. While MODIS may be able to accurately reflect phytoplankton abundance in open ocean waters (McClain 2009), in coastal locations it is important to either consider newer algorithms to analyze its datasets (e.g. Dall'Olmo et al. 2005) or take into account the potential error on the results.

## *2. WHAT IS THE OPTIMAL SPATIAL RESOLUTION FOR THE ANALYSIS?*

Hotspot analyses capture processes that either stimulate or aggregate resources in a particular location (Mackas et al. 1985). Some of these processes have a global influence, with resulting aggregations spanning a few kilometers, while others have a local effect, with the size of the resulting hotspots being just a few meters. The scale of the process under consideration will determine the scale of the study and the spatial resolution of the dataset that needs to be used (Woodcock and Strahler 1987). The choice of spatial resolution is key for correctly locating and characterizing hotspots. For example, low-resolution datasets (e.g. 10 km) may be able to detect large regional scale seasonal blooms (Song et al. 2010), but may miss local scale processes such as internal waves (Scotti and Pineda 2004).

Satellite sensors provide data at native spatial resolutions (e.g. 300m for MERIS), but researchers have the choice of resampling datasets to a broader resolution before running hotspots analyses to better match the scale of the process under study. While the selection of the appropriate spatial resolution of an object-based analysis like hotspot identification is often left to the experienced eye of the researcher in most ecological analyses, there is a procedure to help make this choice. This procedure, which was first described by Woodcock and Strahler (1987), uses semi-variograms to explore the change in variance of an image as its spatial resolution decreases. The optimal spatial resolution is the one that maximizes the variance between cells. This methodology has been effectively used in many studies on land ecosystems (e.g. Curran 1988), where the features are often well defined. However, for ocean ecosystems, the results may be more

difficult to interpret, as illustrated by Figure 2. Phytoplankton aggregations have ill-defined borders and are controlled by processes of different spatial scales occurring simultaneously, so the resulting semi-variogram does not show just one peak where variance is maximized but many different peaks at different resolutions. However, these results still provide valuable information as an initial exploration of a dataset, so Woodcock and Strahler's procedure should still be recommended for any marine hotspot analysis.

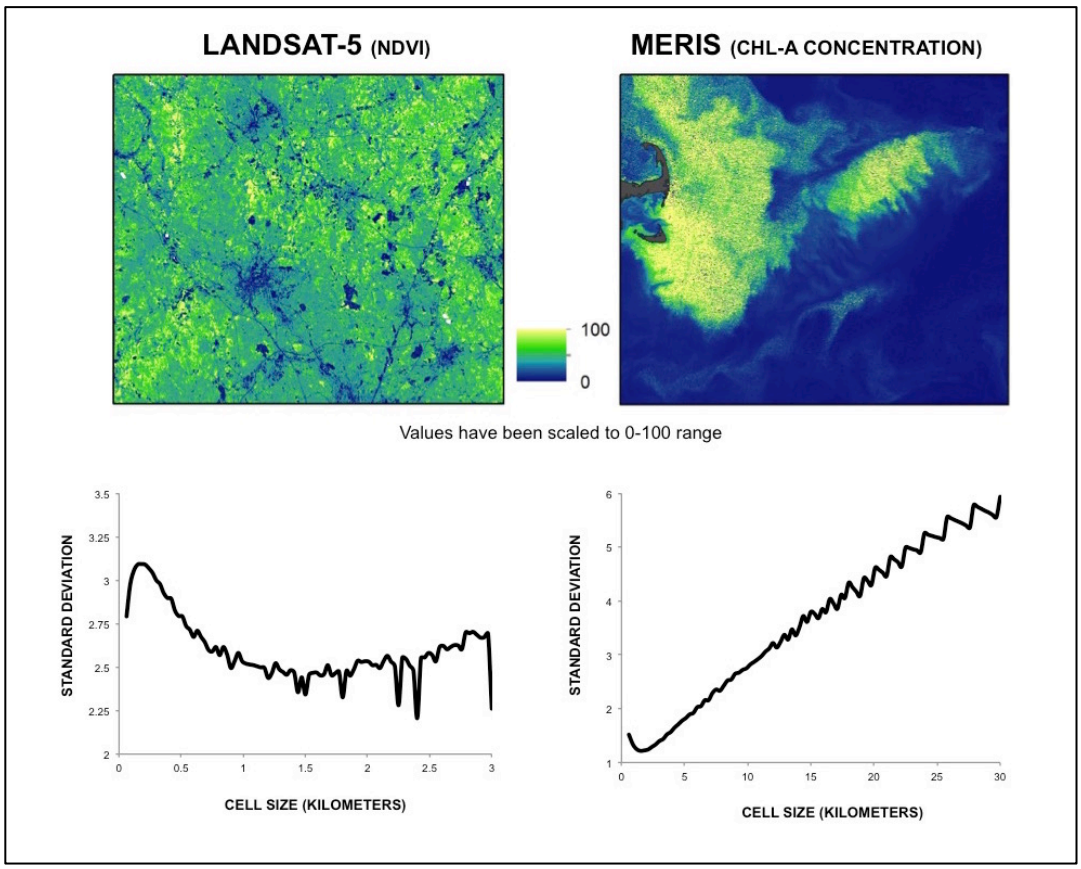


**Figure 2 – Comparison of the size of features found on remote sensing images over land and ocean. Both images were taken around the state of Massachusetts (USA): (left) NDVI values from Landsat-5 (30m resolution), (right) chlorophyll-a concentration from MERIS (300m resolution). For comparison purposes, both variables have been scaled to a 0-100 range. Plots represent semi-variograms resulting from the process described in Woodcock and Strahler (1987).**

## *3. WHAT MEASURE SHOULD BE USED TO DETECT HOTSPOTS?*

The process of locating hotspots generally involves the comparison between the value of a data point and the value of a threshold. This threshold may be set "globally", using all available data (e.g. Link et al. 2013), or "locally", using only the data of the locations in the vicinity of the data point (e.g. LeDrew et al. 2004). Figure 3 shows a visual representation of each type of threshold. In the case of a global threshold, any data point that has a value above the threshold (often mean plus two standard deviations) is considered a hotspot. Global-threshold methods are useful when researchers have access to a complete dataset (i.e. full range of possible values) or they are certain that the processes driving hotspots are constant across the region. However, for many local scale studies, this is rarely the case.

Local threshold methods consider a hotspot any data point that has similar values to its neighbors. Unlike in the global threshold approach where any value above a given threshold is considered a hotpot, using the local definition an isolated high-value point would not be considered a hotspot. For ecological studies, one of the most effective ways to locate hotspots using local thresholds is to use indices of "spatial autocorrelation" (Dormann et al. 2007). These indices take advantage of the fact that many ecological datasets are "spatially dependent" or "spatially auto-correlated", meaning that data points that are closer together are likely to be more related than points that are farther apart (Tobler 1970). A point whose value is similar to the values of its neighbors shows significant spatial autocorrelation. If this point also has a relatively high value, then it may be considered a hotspot. Several local autocorrelation indices exist (see Sokal et al.

1998 for a review), although the most frequently used for ecological studies is the G* index by Getis and Ord (1992). This index may be standardized and reported as a z-score (Ord and Getis 1995), which allows its comparison across space and time. The G* index requires both the mean and standard deviation of the study area (i.e. "global reference parameters"), which are assumed constant across space.

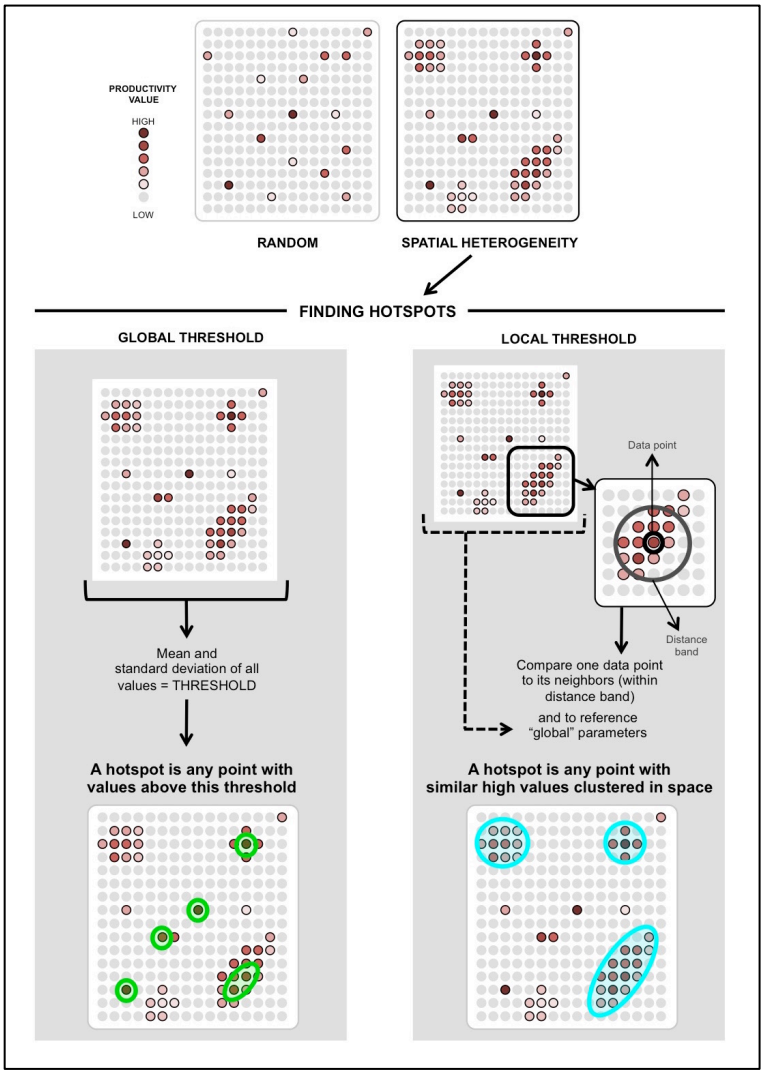


**Figure 3 – Visual comparison of hotspots depending on the threshold used: (Left) global threshold, (right) local threshold**

## *4. DO HOTSPOT MEASURES NEED TO BE ADJUSTED FOR SPATIAL NON-STATIONARITY?*

In ocean ecosystems, different areas are often governed by different ecological processes (Windle et al. 2010). For example, productivity values in near-shore areas are often consistently higher than those on the open ocean due to oceanographic processes such as upwelling. This overarching global gradient may mask local aggregations of productivity, for example the ones created by local fronts. This creates productivity aggregations that change in magnitude, timing and spatial scale across space. Statistically, this manifests in ocean datasets in the form of "spatial non-stationarity". The effect of non-stationarity is especially problematic when using local spatial autocorrelation indices, because they require global reference parameters that are assumed constant across space. If the data are non-stationary, hotspots may only be located in areas with higher means and variances, usually near the coast, and aggregations offshore could be ignored.

The effect of spatial non-stationarity is difficult to detect in remote sensing datasets. Usually, the best way to know there is a problem is to calculate the mean and standard deviation across several years for each pixel and then map all the values. If these resulting global parameters show any spatial pattern or trend (i.e. they are a function of the location), then the data may indeed be spatially non-homogeneous. However, this process requires a long time-series of data layers, which some researchers may not be able to obtain. When using just one data layer, the simplest way to determine if datasets follow a global gradient may be to plot the values along different transects (e.g. inshore

to offshore) and inspect the results for trends (Fortin and Dale 2005). Figure 4 shows a hypothetical example. Semi-variograms may also be used for this purpose, with unbounded variograms (i.e. not plateauing as distance increases) signaling spatial non-stationarity in the dataset.

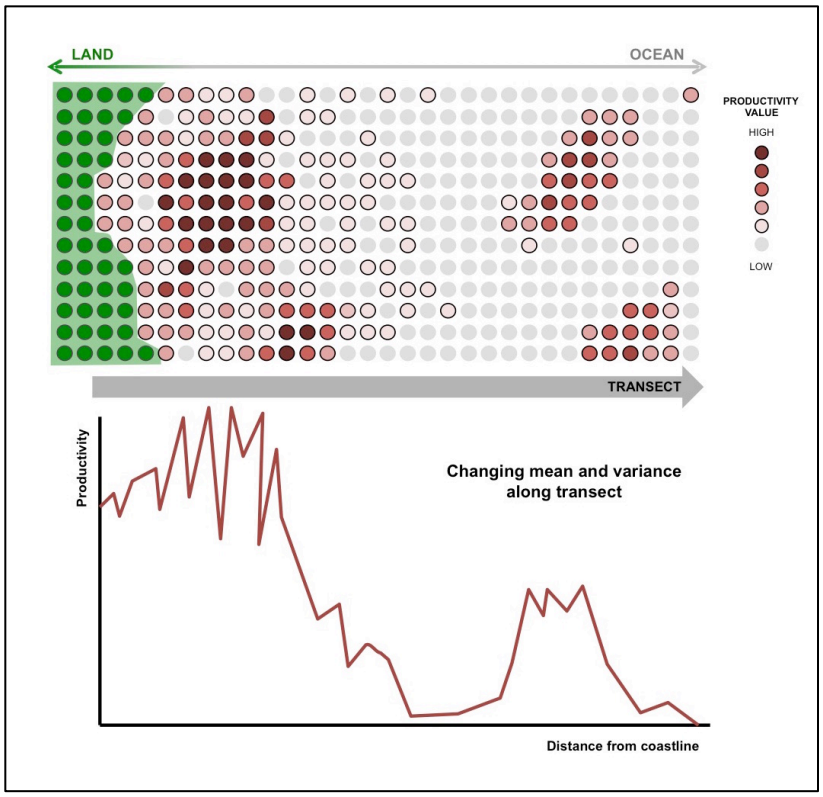


**Figure 4 – Visual representation of spatial non-stationarity (adapted from Fortin and Dale 2005)**

If spatial non-stationarity is suspected in the dataset due to a known global trend, the easiest solution is to de-trend the data and derive local spatial autocorrelation values from the resulting values. However, when the pattern is not known, there are two main types of methodological approaches to take to calculate local spatial autocorrelation indices (see box 1 for a case study). The first type comprises all "spatial partitioning"

methods (Fortin and Dale 2005), which include methods such as spatial clustering or boundary delineation. They separate the overall area of study into smaller homogeneous (i.e. stationary) areas and G* indices are derived using only the values within each area. The second type of methods, which Mackenzie (2007) called the "Multi-region G*", are designed exclusively for the G* index (Ord and Getis 2001). This methodology derives a different circular "global" reference area centered on each data point. The size of each global area is selected based on its homogeneity, which is then used to derive G* for each pixel. Spatial partitioning methods work best when the borders of the hotspots are well defined. When the borders appear more like a gradient than a line, this type of methodology may show some problematic values at the edges between partitions (see box 1). Multi-region G* methods may be the best solution for hotspots with ill-defined borders. However, most of the methods of this second group are not fully tested, and need to be compared to field validation data to ensure its accuracy at detecting hotspots.

---

*BOX 1 – CASE STUDY: COMPARISON OF METHODS TO LOCATE HOTSPOTS IN THE PRESENCE OF SPATIAL NON-STATIONARITY*

I used a full-resolution MERIS data layer from the Gulf of Maine to compare the results when using the following methods to locate hotspots in the presence of spatial non-stationarity (see Figure 5):

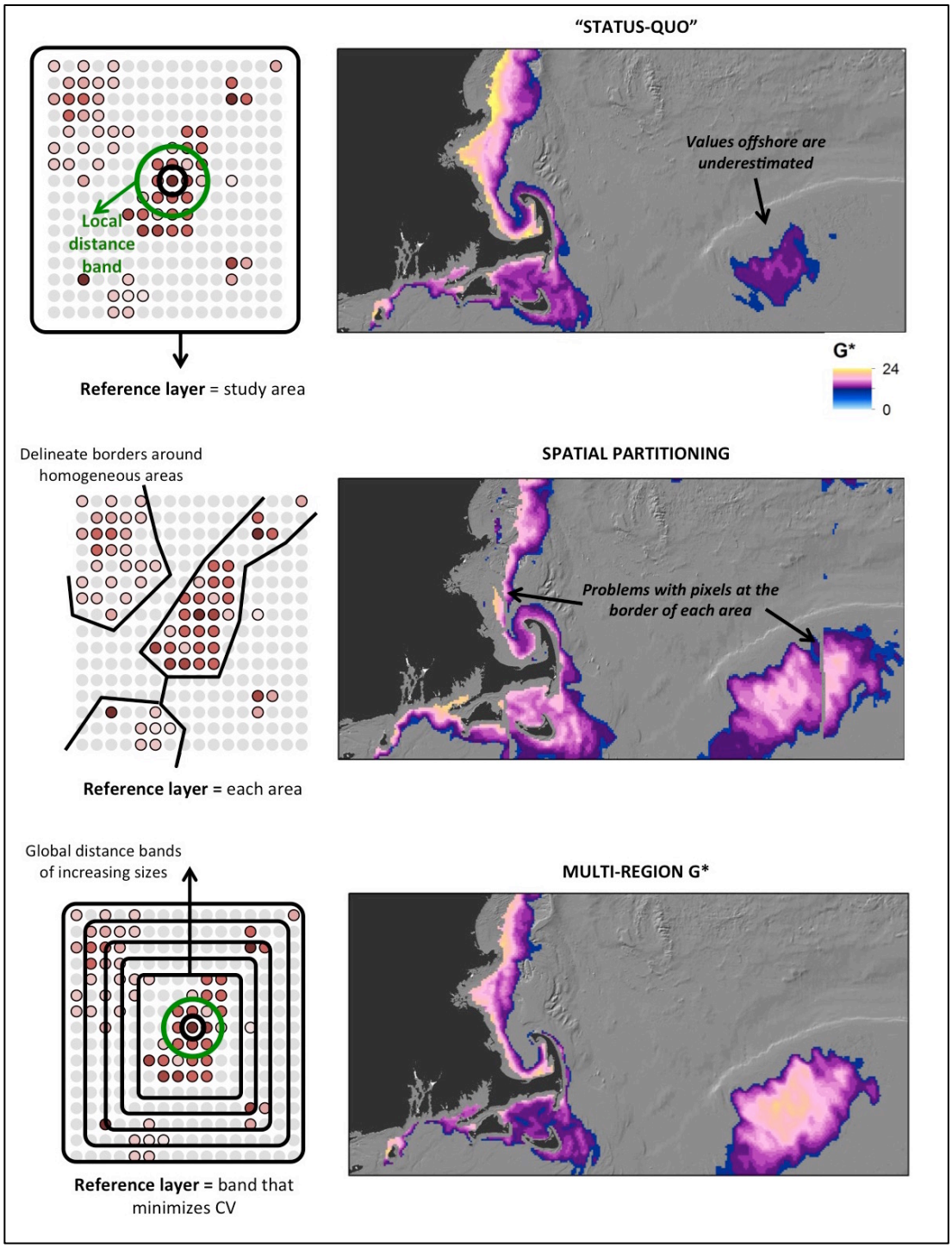


**Figure 5 – (Left) Visual representation of the different methodologies. (Right) Comparison of the resulting hotspots for each methodology. Only pixels with G* value above 2 are shown.**

- *Spatial partitioning using delineating boundary methods*: Use fuzzy c-means clustering to separate the area in different homogeneous zones. Calculate the global parameters exclusively for each zone and use them to derive G* values. (see Fortin and Dale 2005 for more information)
- *Multi-region G* method*: For each pixel, create different global distance bands of increasing sizes. For each distance band, calculate the coefficient of variation (CV=standard deviation/mean). The distance band at which the CV value is minimized is the one used as a reference layer to derive the value of the G* index. This is a modification of the process defined in Mackenzie (2007).
- *"Status-quo":* Assume spatial homogeneity.

---

## *5. WHAT ARE THE LIMITS OF HOTSPOTS?*

Once spatial autocorrelation values are derived, it is often necessary to determine the borders around hotspots. Finding the limits of local marine hotspots may be done by either using image-processing techniques to define features or by determining a threshold of statistical significance. Due to the ill-defined nature of borders around hotspots, image-processing techniques may be challenging. To facilitate border detection, Aldstadt and Getis (2006) created a new process called AMOEBA that iteratively detects the cells that belong to each hotspot based on the absolute value of the G* index. This method, however, is very computationally intensive so it may not be advisable for high-resolution remote sensing layers, especially when the area of study is wide.

A threshold of statistical significance is often complicated to determine with local spatial autocorrelation indices due to the problem of multiple testing with points that are not independent. A common proposed solution for this effect is the Bonferroni adjustment, which divides the significance level (e.g. 0.05) by the number of tests. However, in data layers where the points are spatially correlated, this adjustment tends to be very restrictive, creating a very high critical threshold value. Rogerson (2002) proposed a different adjustment, called M* adjustment, that uses Gaussian kernels to calculate a new critical significance value. However, more research is needed to determine whether this adjustment is accurate for ocean studies. Meanwhile, if the purpose of the study is just exploratory, Nelson and Boots (2008) recommended to treat any value above 2 as a hotspot due to the fact that G* values are considered z-scores.

## *6. HOW CAN DYNAMIC PROCESSES BE CAPTURED?*

The concept of a hotspot, especially on land ecosystems, is often assumed to be static, meaning that the processes that create these hotspots are thought to be constant throughout time. However, most marine ecosystems are very dynamic (Mann and Lazier 1996). The abundance of phytoplankton organisms depends on a combination of environmental conditions that change over time, such as temperature and nutrient availability. In temperate locations, a major component of the temporal variability in phytoplankton abundance is seasonal, with two major phytoplankton blooms during spring and fall (Thomas et al. 2003). The variability that goes beyond this expected

seasonality is often called “anomalies” and, even though it is less predictable than seasonal variability, it is as important at determining the distribution and abundance of phytoplankton populations (Kahru et al. 2012).

In the past few years, there has been a move towards redefining hotspots depending on their predictability over time. Hotspots are defined as areas that have a high probability of presenting favorable conditions for the presence of high productivity at one point during the year (Etnoyer et al. 2004). New metrics have taken this concept and applied it to dynamic ocean systems, defining hotspots based on either the high incidence of positive anomalies (Suryan et al. 2012) or their seasonal magnitude (Chapter 2, this thesis). These metrics facilitate the comparison of information from different trophic levels (Winiarski et al. 2013). Also, because these are based on time series, it is possible to study the year-to-year changes on a location or long-term climatic processes.

## CONCLUSIONS

Oceans are complex systems, in part because of the spatially and temporally dynamic distributions of their species (Crowder and Norse 2008). For many years, researchers have been proposing an ecosystem-based approach to management as a way to take into account the complex interactions between marine species within the context of their environment (Christensen et al. 1996). However, translating these principles into practical plans has proven challenging (Arkema et al. 2006). Because local hotspots provide information on the processes driving spatial heterogeneity in marine ecosystems,

developing systematic approaches to mapping marine hotspots may be a critical first step towards characterizing and unraveling the complexity inherent in marine systems.

In essence, hotspots highlight ecosystem processes and cross-scale interactions instead on individual parts of the system. Therefore, the results of marine hotspot analysis can be used to start conversations on effective changes in management approach. Stakeholders easily understand the concept of a hotspot, as they see it in the field every day. In fact, fishermen often target these locations (Hooker and Gerber 2004). Therefore, understanding where and how hotspots form will help ensure the long-term protection of biodiversity and ecological progresses. The study of hotspots will also provide the means to study potential long-term changes in these aggregations. For example, it is possible to understand whether climate change is affecting the number and magnitude of feeding aggregations in an ecosystem. I believe hotspot analyses, especially when combined with remote sensing data, will allow researchers to test hypothesis that they could not consider before. I am hoping that the simplified methodological framework presented here will open the door to new possible studies than in the long-term benefit the management of extremely valuable ocean resources.

# CHAPTER TWO: LOCATION, MAGNITUDE, AND TEMPORAL VARIABILITY OF LOCAL PRIMARY PRODUCTIVITY HOTSPOTS IN THE GULF OF MAINE

## ABSTRACT[2]

Identification of primary productivity hotspots may be a necessary step toward ecosystem-based management goals, as these often signal underlying processes that aggregate or stimulate resources in a particular location. However, previously used metrics to locate these hotspots are not easily adapted to local marine datasets, in part due to the high spatial and temporal variability of phytoplankton populations. The objective of this study was to identify local-scale primary productivity hotspots in two separate regions in the western side of the Gulf of Maine using remote sensing chlorophyll-a data (from MERIS sensor), and to study their variability in space and time. For this reason, I first defined a new hotspot metric, the *Index of Cumulative Anomalous Productivity* (ICAP), which identified as a hotspot any area that consistently exhibited high-magnitude anomalies (i.e. residuals from seasonal pattern) through time, a sign of highly dynamic communities. I used this metric to identify hotspots for every "biological season", which were defined based on phenology. I then determined the seasonal, inter-annual and long-term variability of the hotspots in these areas by fitting Fourier curves to each pixel's time series. Finally, I tested the uncertainty of our results and validated them using field data. Our results revealed that both the location and the magnitude of hotspots are highly

[2] To be submitted to "Ecography". Structure matches the publication's guidelines.

dependent on the season and the area of study. Our data also suggest that the variability of the magnitude of these aggregation areas in time depends on where these hotspots are located and during what time of the year they occur. I argue that this new hotspot index compliments existing global measures as it helps managers understand the dynamic characteristics of a complex marine system. It also provides a unique metric that is easily compared across space and between different trophic levels, which may facilitate future ecosystem-wide studies.

## INTRODUCTION

The identification of hotspots is becoming an important step in marine ecosystem based management efforts (Hooker and Gerber 2004). Hotspots are generally defined as locations where organisms aggregate (Nelson and Boots 2008), and they often signal underlying processes driving spatial patterns that would not occur by chance. Understanding the nature of these processes may provide information on the complexity present within the system (Getis and Boots 1979), which may in turn help managers with their efforts towards applying ecosystem-based principles to their management plans. Hotspots have already been used in studies to identify feeding grounds (Nur et al. 2011), study the effect of human activity on overall ecosystem health (Davoren 2007), and help prioritize marine areas to protect (Worm et al 2003). However, while the benefits of this type of analysis are clear, there are multiple challenges in the process of locating hotspots, especially in marine ecosystems.

The methodology used for locating hotspots often varies depending on the scale of the study. In global scale studies, hotspots are usually located by comparing the value of each data point with a global threshold, which is often the mean plus one standard deviation of all available values (e.g. Myers et al 2000). This processes is frequently used to locate global areas with high number of different species (i.e. "biodiversity hotspots") and in turn in need of increased conservation efforts. The process of locating hotspots using a global threshold is successful when the whole Earth's surface can be used to establish an appropriate threshold. However, the results may be difficult to interpret in regional or local scale studies, where the threshold must be set using the limited data available or subjectively determined values (e.g. Link et al. 2013). For these local scale studies, Wulder and Boots (1998) recommended the use of spatial statistics to locate hotspots. In this case, hotspots are defined as concentrations of similarly high values in space (Ord and Getis 1995) and they are located using local spatial autocorrelation indices, which compare the value of each location to its neighbors. Locations with both positive spatial autocorrelation and a significantly high value of a variable are considered hotspots. When these autocorrelation indices are applied to abundance or productivity data, the resulting hotspots signal locations where organisms aggregate in space to feed on abundant supplies (Worm et al. 2003), which in ecological studies are often referred to as "biological hotspots".

The characterization of biological hotspots in marine ecosystems is challenging, in part because of the high temporal and spatial variability of the resources supporting them. Marine biological hotspots are usually supported by high abundance of primary

producers, which attract other species higher up the food chain to feed (Norse et al. 2005). As a result, characteristics of phytoplankton populations directly determine the characteristics of the resulting hotspots. Aggregations of primary producers are dependent on a combination of environmental conditions, such as temperature and nutrient availability, that either stimulate or aggregate primary productivity at varying spatial and temporal scales (Mackas et al. 1985). These environmental conditions are both spatially and temporally dynamic (Mann and Lazier 1996), making the associated biological hotspots highly variable as well. In temperate locations, part of this variability is due to the seasonal fluctuations of phytoplankton abundance, which are often characterized by a low concentration in winter, followed by a spring bloom, a decrease during the summer and finally a fall bloom (Thomas et al. 2003). The strength and timing of the blooms may vary from year to year, but the general seasonality is largely consistent over time. Variability beyond this seasonal pattern, however, is more difficult to predict (McGowan et al. 1996). This variability, often called “anomalies”, can be in the form of either long-term global trends (e.g. climate change), year-to-year variations caused by large-scale atmospheric cycles (e.g. North Atlantic Oscillation), or short-term variations caused by various oceanographic processes (Wolanski and Hamner 1988).

Previous studies of ocean biological hotspots have been static, only considering aggregations during a particular moment in time (e.g. Barrell and Grant 2013). However, to fully understand the processes driving the complexity in marine systems, there is the need to incorporate the temporal variability of marine resources into the analysis of biological hotspots. For this reason, Suryan et al. (2102) proposed to redefine biological

hotspots as areas that have a higher probability of presenting positive productivity anomalies over time. These should signal areas of increased and predictable transfer of organic matter between trophic levels, which may in turn predict the locations where predators aggregate. The authors created an index, the Frequency of Chlorophyll Peaks Index (FCPI), which measured the proportion of days a location had a positive anomalous value of chlorophyll-a (i.e. positive residual) compared to the regional seasonal pattern. The FCPI uses remote sensing information, so it is simple to adapt to different areas across the globe and can be easily compared to datasets from other trophic levels. However, this index does not take into account the magnitude of the anomalies, so it has shown limited applicability when comparing areas with very different overall mean chlorophyll values, such as coastal and offshore locations (Suryan et al. 2012, Winiarski et al. 2013). Also, because it does not use spatial statistics measures to account for the variability across space, it is affected by the presence of outliers in remote sensing datasets.

In this paper, I examine the spatial and temporal variability in magnitude of local productivity hotspots in the western Gulf of Maine. I define a novel hotspot index that uses local spatial statistic indices to detect spatial concentrations of anomalous chlorophyll-a values. This new metric characterizes different locations based on their ability to either stimulate phytoplankton growth or concentrate it at a higher rate compared to the regional pattern. The remainder of this paper is separated in three sections: first, I define this new hotspot metric and apply it to locate primary productivity hotspots; second, I determine the seasonal, inter-annual and long-term variability of the

identified hotspots; and third, I calculate the uncertainties of our model and validate the results using field data.

## METHODOLOGY

### *STUDY AREA*

The area of study for this project spans the western third of the Gulf of Maine (see Figure 6). The Gulf of Maine is one of the most productive and complex temperate areas in the world, making it especially sensitive to human disturbance (Auster et al. 1996). The habitats present in this semi-enclosed sea support an estimated 3,317 species of flora and fauna (Thompson 2010), which in turn support a fishing industry of about 20,000 commercial fishermen, 10 million yearly visitors (with activities ranging from whale watching to birding, recreational fishing, and boating), and multiple transportation and energy industries. The study area is centered over Stellwagen Bank and includes some of the most popular fishing spots in the area (e.g. Jeffrey's Ledge). The Stellwagen Bank is a complex temperate system of mid-water and benthic habitats which supports over 575 marine species (SBNMS 2010). The popularity of this area as a fishing and tourist spot is due not only to its high biological productivity and diversity but also to its close proximity to Boston (25 miles) and its neighboring towns. It is also one of the most popular spots in the world for whale watching. Since 1992, this bank has been part of the Stellwagen Bank National Marine Sanctuary (SBNMS). I divided the study area into two separate areas, north and south, with the limit between the two being the latitude of Chatham, MA. This step was necessary because areas north and south of Cape Cod are

characterized by different regional oceanographic processes, which I assumed would result in different seasonal patterns of productivity and confound hotspot identification within each region.

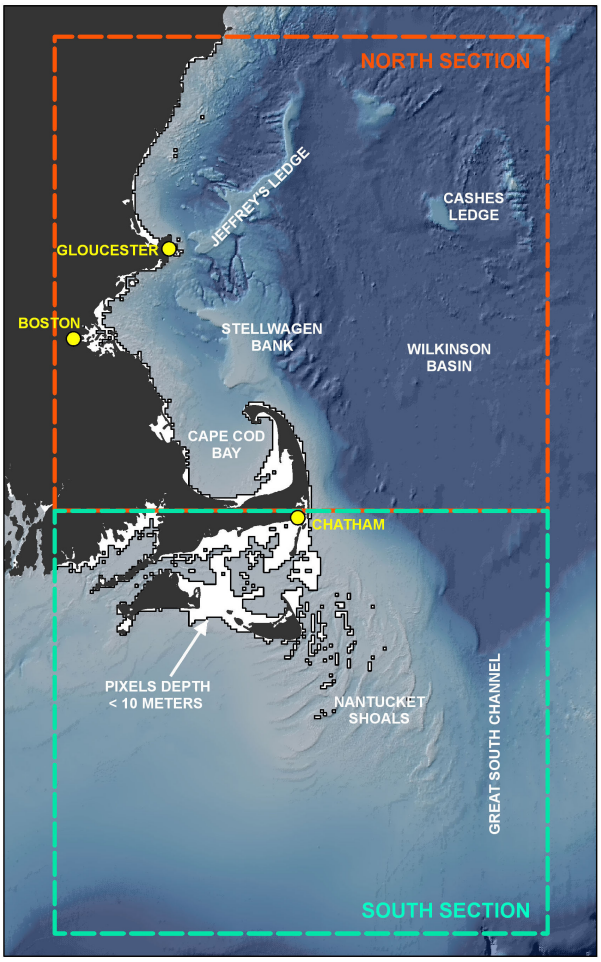


**Figure 6 – Map showing study area and some of the principal topographic features. The area of study was separated in two sections: north (orange dashed line) and south sections (turquoise dashed line). White-background areas represent locations with depths shallower than 10m.**

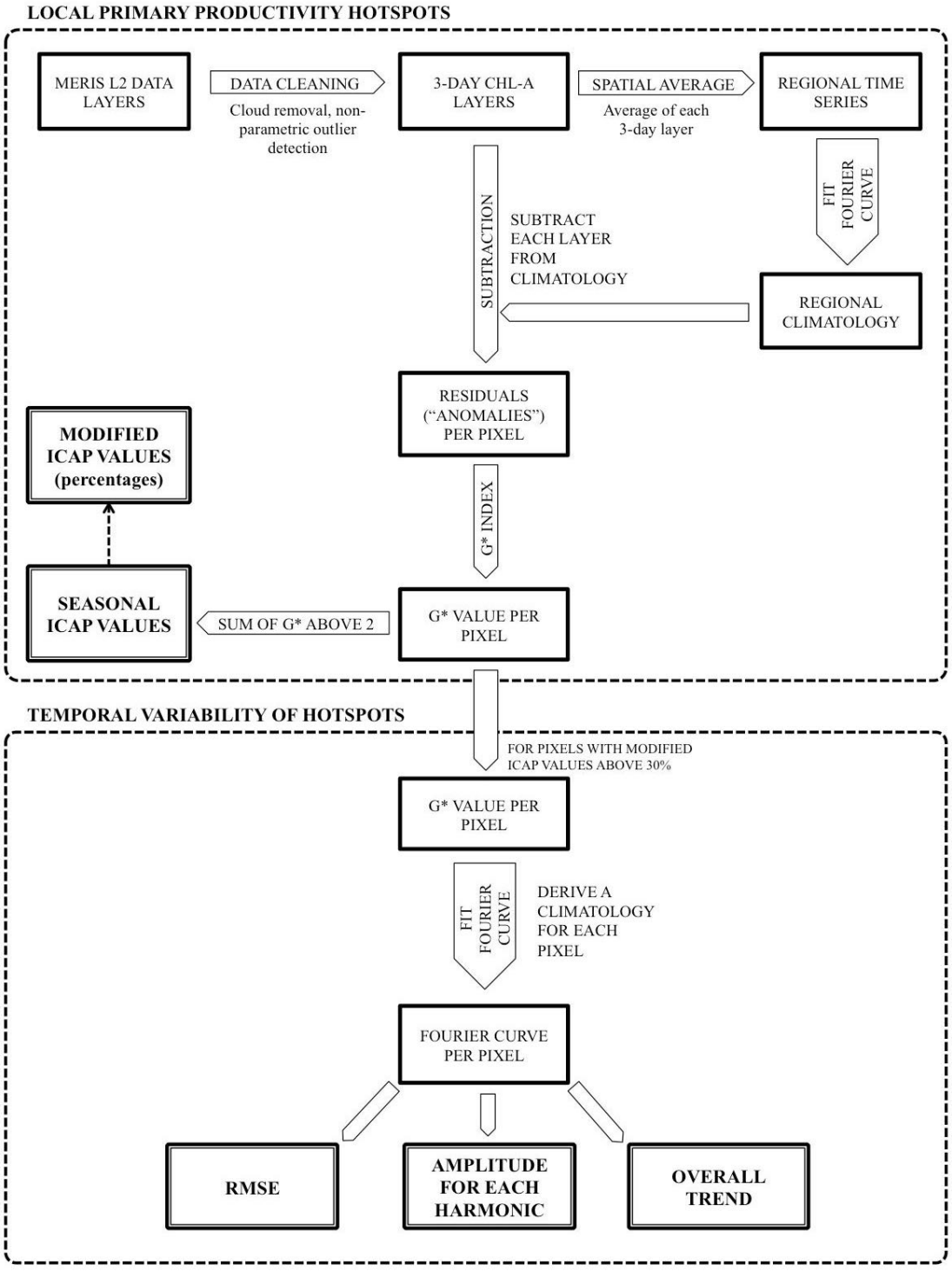


**Figure 7. - Graphic representation of the methodology in this study**

*LOCAL PRIMARY PRODUCTIVITY HOTSPOTS*

I used chlorophyll remote sensing data from the Medium Resolution Imaging Spectrometer (MERIS) as a proxy for phytoplankton abundance. I acquired all available reduced resolution MERIS Level-2 imagery (a total of 1661 images) collected between April 2002 and April 2012. These reduced resolution datasets are 3-day composites with a spatial resolution of 1200m. They were acquired by the European Space Agency (ESA) and processed by the National Aeronautics and Space Administration (NASA). I screened each 3-day level-2 data layer for incorrect values and missed clouds. The screening process included the removal of chlorophyll-a values above 45 $mg/m^3$, a threshold representing the maximum-recorded concentration in the area of study (Hyde et al. 2007). I also removed values within a two-pixel distance of any cloud and in locations with water depths shallower than 10m, to prevent incorrect chlorophyll values due to either shallow waters or proximity to land (Maritorena et al. 1994). To eliminate the possibility of missed clouds and cloud-shadows, I also fitted a curve to each pixel's time series using a LOWESS non-parametric algorithm (Cleveland 1981). Its residuals were then converted to z-scores, and values above 5 and below -5 were considered outliers and thus removed from the dataset. Finally, all layers were both collocated to the same reference grid so the coordinates for the center of all pixels were the same in every layer, and mean-centered (i.e. subtract overall all-time mean from each layer) to remove any possible global spatial gradient.

I located seasonal hotspots using a new metric, the *Index of Cumulative Anomalous Productivity* (ICAP), which it is a measure of the magnitude of productivity

anomalies within each pixel and across time. The derivation of this index (see Figure 7) is a process adapted from Suryan et al. (2012). It includes four main steps: (1) deriving a regional climatology (i.e. seasonal pattern); (2) calculating the anomalies (i.e. residuals) for each pixel and time period; (3) deriving local spatial autocorrelation values for each pixel and time period; and (4) adding all the significant spatial autocorrelation values within a season. For this study, instead of defining seasons based on astronomical dates, I defined seasons based on the phenology of the regional seasonal blooms. For the remainder of the paper, I will refer to these seasons as "biological seasons". The main advantages of using biological seasons over astronomical ones are two-fold: first, they better match the life cycle of phytoplankton species; and second, they may help locate areas where the timing of the blooms is consistently different than the regional norm. The start and end dates of both spring and fall blooms were determined based on the regional climatology, using the methodology described in Siegel et al. (2002). This results in four different "biological seasons": Spring bloom, summer break, fall bloom, and winter break.

*Step 1 – Regional climatology*. I first derived the spatial average (i.e. average value of all pixels within a layer) from each 3-day MERIS layer. The result was a regional time series, with one data value for each time period ($\bar{x}_t$). I then obtained a regional climatology ($\bar{y}_t$) by fitting a Fourier curve to this time series. I chose the Fourier curve to fit this time series because it represents well a repetitive seasonal pattern with two main peaks throughout the year. Also, the parameters of the Fourier curve can be used to derive the characteristics of the regional seasonality. For this study, the fourier

curve included an intercept value, an overall trend, and up to 2 pairs of sine and cosine values that represent the different harmonics (12-month and 6-month harmonics). To ensure the best-fit possible, I allowed the model to choose the number of harmonics needed based on the resulting AIC (Akaike Information Criterion). As an example, this is the equation with two harmonics:

$$\bar{y}_t = a_0 + a_1\bar{x}_t + a_2\, cos(w\bar{x}_t) + a_3\, sin(w\bar{x}_t) + a_4\, cos(2w\bar{x}_t) + a_5\, sin(2w\bar{x}_t)$$

$$w = 2\pi/365.25 \qquad (1)$$

This curve was fitted using the "lasso-glm" function in MATLAB (Mathworks, Inc.). Lasso is a shrinkage regression method (Tibshirani 1996), which predicts coefficients that are biased to be small, thus reducing potential errors due to the presence of outliers.

*Step 2 – Anomalies.* This step involves subtracting the regional climatology from each pixel's time series to derive 3-day anomaly layers. Positive anomalies signal concentrations of chlorophyll-a that are above what is expected for the region.

*Step 3 – Local spatial autocorrelation.* From the anomalies values, I derived local spatial autocorrelation values for each pixel using Getis and Ord (1992) G* index. This index measures the proportion between the values within a certain distance from a point and the variance for the rest of the study area. It was modified by Wulder and Boots (1998) for its use with remote sensing data and standardized so the resulting values are z-

scores, thus comparable across space and time. The equation to derive this form of the G* index is:

$$G_i^* = \frac{\sum_j w_{ij}(d)p_j - W_i^* \bar{p}}{s\left[W_i^* \frac{n - W_i^*}{n-1}\right]^{1/2}}$$

$$W_i^* = \sum_j w_{ij}(d) \quad , \quad \bar{p} = \sum_j p_j \quad , \quad s^2 = \frac{\sum_j p_j^2}{n - \bar{p}^2} \qquad (2)$$

where $p_i$ is the positive anomaly value at pixel *i*. $w_{ij}$ is the weigh matrix (in this case, a 3 by 3 matrix of all ones). The values of $\bar{p}$ and *s* are the mean and standard deviation of all anomalies within a layer. To ensure that the distribution of the resulting $G_i$* is normal, it is recommended that this index is only derived when a pixel has at least 8 surrounding data points (Griffin et al. 1996). This step resulted in G* values for each pixel and date.

*Step 4 – ICAP values*. I aggregated all significant $G_i$* values for each pixel and biological season. Following the recommendations by Wulder and Boots (1998), I considered any G* value above 2 as being a significant hotspot. Due to data gaps created by the presence of clouds (specially during winter), I calculated 15-day averages of the $G_i$* indices prior to calculating the ICAP value per season. Also, to take into account the variability between years, I calculated the ICAP for each year and season, and then averaged all the yearly results. The resulting ICAP values represent the extra productivity available within each pixel, compared to the regional pattern. Researchers may use these indices to compare values across pixels and across seasons. However, it is possible that in some regions the biological seasons have different number of days, which makes the comparison between seasons challenging. In this case, I converted the ICAP index to

percentage values (named here "Modified ICAP"). These percentages represent the ratio between ICAP values and the maximum potential ICAP value. This maximum potential value is the value that would occur if every 15-day group had a "high anomaly" (median plus one standard deviation of all $G_i$* values within a region).

*TEMPORAL VARIABILITY OF HOTSPOTS*

To determine the spatial changes in the temporal variability of biological hotspots, I derived a seasonal curve from the G* values for each pixel and day. Following the same methodology above, I fitted a Fourier curve including an intercept value, an overall trend, and up to three harmonics (12, 6 and 3-month harmonics) depending on the resulting AIC. From each pixel's fitted curve, I derived the amplitude (i.e. seasonal variability), the trend (i.e. long-term variability), and root-mean square error (i.e. short-term variability). The amplitude (A) is represented by the square root of the sum of squares of the coefficients representing the 12 month, 6 month, and 3 month harmonics. For each harmonic, the amplitude is calculated as follows:

$$A_{12} = \sqrt{{a_2}^2 + {a_3}^2} \quad , \quad A_6 = \sqrt{{a_4}^2 + {a_5}^2} \quad , \quad A_3 = \sqrt{{a_6}^2 + {a_7}^2} \qquad (3)$$

where $a_2$ to $a_7$ are the coefficients from the Fourier curve equation corresponding to each of the harmonics. The 12-month amplitude represents the differences between the first half of the year and the last. The 6-month amplitude represents the differences between seasons. The 3-month amplitude represents the persistent within-season changes (i.e.

fluctuation within seasons that occurs in most years). The trend (i.e. linear change over time) is represented by the coefficient $a_1$. Finally, the root-mean square error (RMSE) is the square root of the squared residuals divided by the total number of observations. RMSE symbolizes short-term changes in productivity that are not repetitive every year. It can also be considered as the year-to-year variability.

### *UNCERTAINTY ANALYSIS AND VALIDATION*

Most of the uncertainty in the results stems from the presence of clouds in the dataset. Consequently, in this study I wanted to characterize how much the results change due to missing data points. I first chose the pixel with the highest number of cloud-free observations. Then, I randomly selected samples from this time series to generate vectors with different lengths (400 different vectors with lengths ranging from 100 to 500 cloud-free observations). I fitted Fourier curves to each increasingly sized sample vector, and from these fitted curves I derived amplitude values for each of the harmonics and trend.

Validation of the model results is challenging since there is no field database that has both a wide spatial coverage to calculate hotspots and a long time series that matches the MERIS remote sensing dataset. In fact, the only database available in the area that has a persistent record of surface chlorophyll-a concentration across time is the Environmental Monitoring and Mapping System (EM&MS) database managed by the Massachusetts Water Resource Authority (MWRA). The MWRA has been collecting monitoring data from different sites across Massachusetts Bay to study the effects of the placement of the new wastewater outfall (which started in September 2000) on the water

quality and organisms around the Bay (Werme et al. 2012). This dataset includes 10-years of monthly surface chlorophyll-a concentration for 23 stations across Massachusetts Bay. The MWRA database has several limitations for this study. First, samples were collected once every month, thus the temporal resolution is lower compared to the remote sensing dataset. Also, most stations have gaps in data collection, especially during winter months. Most stations are also close to the coastline, so their yearly patterns are very similar. In fact, only 10 stations are both within our study area and have enough data-points to be used as validation data. Finally, these field stations are not close enough to be able to calculate G* indices, so the results cannot be compared one-to-one with ICAP values. However, even with all these limitations, it is possible to use the MWRA dataset to test the ability of the MERIS data to capture the seasonal variation within each field station.

I adopted the same methodology used to derive ICAP indices, with a few exceptions. I used the same climatology derived in step 3 above, although I reduced its temporal resolution by averaging the values for each month. This climatology was subtracted from each station's chlorophyll-a values, which were previously mean-centered. To be able to compare these results to the G* values, I converted the resulting anomalies to z-scores. I fitted a Fourier curve with 2 harmonics to these anomalies. Finally, to measure the uncertainty on the fitted amplitude of each station, I derived yearly amplitudes by calculating the difference between the maximum and minimum anomaly value for each year and calculated the standard error of the resulting values.

# RESULTS

## *LOCAL PRIMARY PRODUCTIVITY HOTSPOTS*

The regional seasonal curves (i.e. climatologies) resulting from fitting Fourier curves to each study area's spatially averaged information are shown in Figure 8. Based on the resulting AICs from each fit, the model chose the 12-month and the 6-month harmonic in both study areas (north and south). Visually, the northern climatology has its lowest value during the winter break, followed by a spring bloom peak, a dip during the summer months, and finally a peak during the fall bloom, which is slightly smaller than the spring peak. On the other hand, the southern climatology's lowest value occurs during the summer break, with a fall peak that is less noticeable compared to the northern climatology. The overall fluctuation of the northern climatology is also higher compared to the southern climatology. The trends for both curves are positive (i.e. values increasing across time), although the one for the northern area is of slightly higher magnitude than the southern trend.

I derived the start and end dates of the phytoplankton blooms from the climatology curves. These dates were used to define "biological seasons". For the northern areas, spring bloom season started on day number 70 (March 11$^{th}$), summer break started on day number 170 (May 19$^{th}$), fall bloom started on day number 239 (August 27$^{th}$), and winter break started on day number 316 (November 12$^{th}$). On the other hand, for the south area spring bloom season started on day 32 (February 1$^{st}$), summer break on day 153 (May 2$^{nd}$), fall bloom on day 257 (September 14$^{th}$), and winter break on day 312 (November 8$^{th}$).

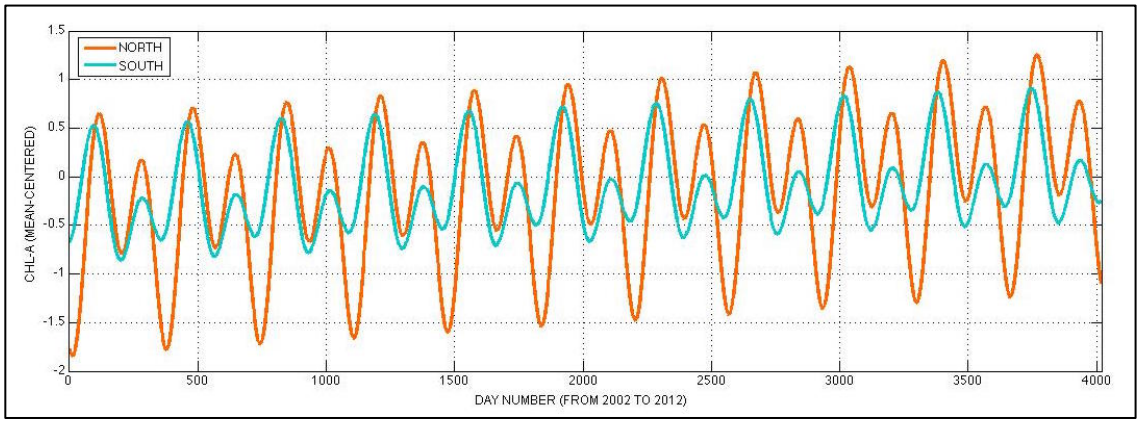


**Figure 8. - Plot of regional climatologies for both study areas (north in orange, south in turquoise).**

The ICAP indices for the northern and southern study areas are shown in Figures 9 and 10, respectively. During the spring bloom, the northern area has a concentration of anomalous values north of Stellwagen Bank (an area known as Tillies Bank) and around Cape Ann (where Gloucester is located). High ICAP values are also located along the western edge of Stellwagen bank and over its southwest corner. Once summer break arrives, very few locations have high ICAP values, except the area just offshore from Boston and just north of Cape Ann. The situation during the fall bloom shows most of the productivity concentrated along the coastal regions of New Hampshire, Maine and Massachusetts. Finally, during the winter break, the greatest spatial concentration of anomalous values is located in Cape Cod Bay. Spring bloom ICAP indices for the southern study area show a high concentration of anomalous chlorophyll values on the Nantucket Shoals. During the summer break, only locations within Buzzards Bay show high concentrations of anomalies. During the fall bloom and winter break, productivity

anomalies are concentrated on the southwestern end of Cape Cod, and near the coasts of Rhode Island and Connecticut.

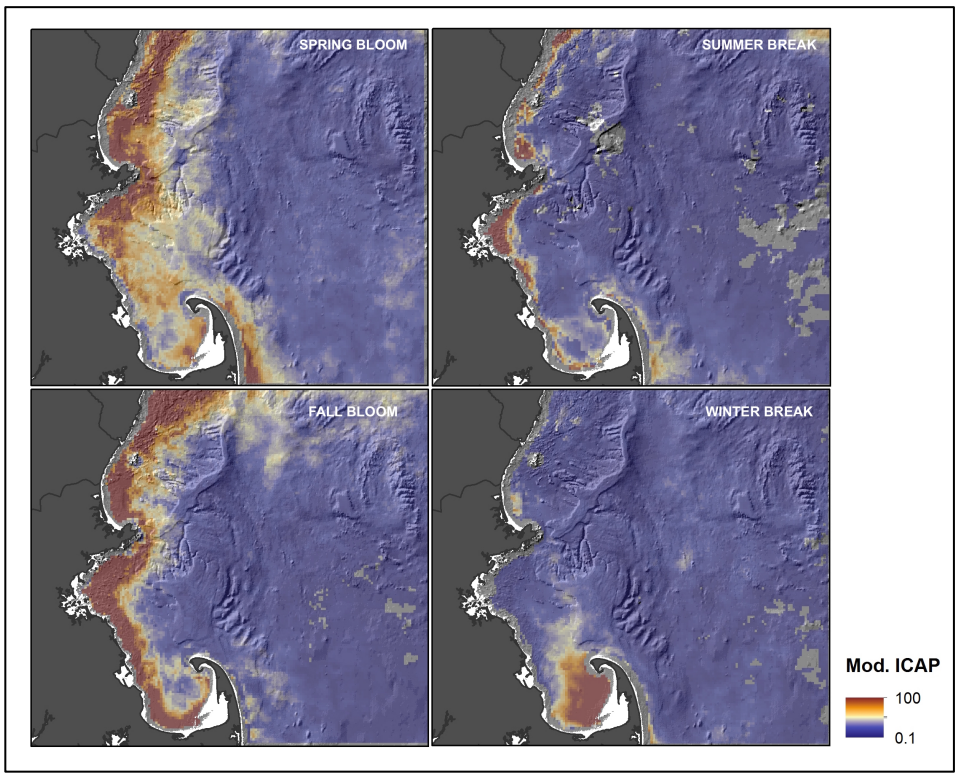


**Figure 9. - Modified ICAP values for the northern study area. (Top-left) Spring bloom, (top-right) summer break, (bottom-left) fall bloom, and (bottom-right) winter break. Values range from a persistent high spatial concentration of anomalous values (dark brown) to a low concentration (dark purple). Gray colored background signal no-data values**

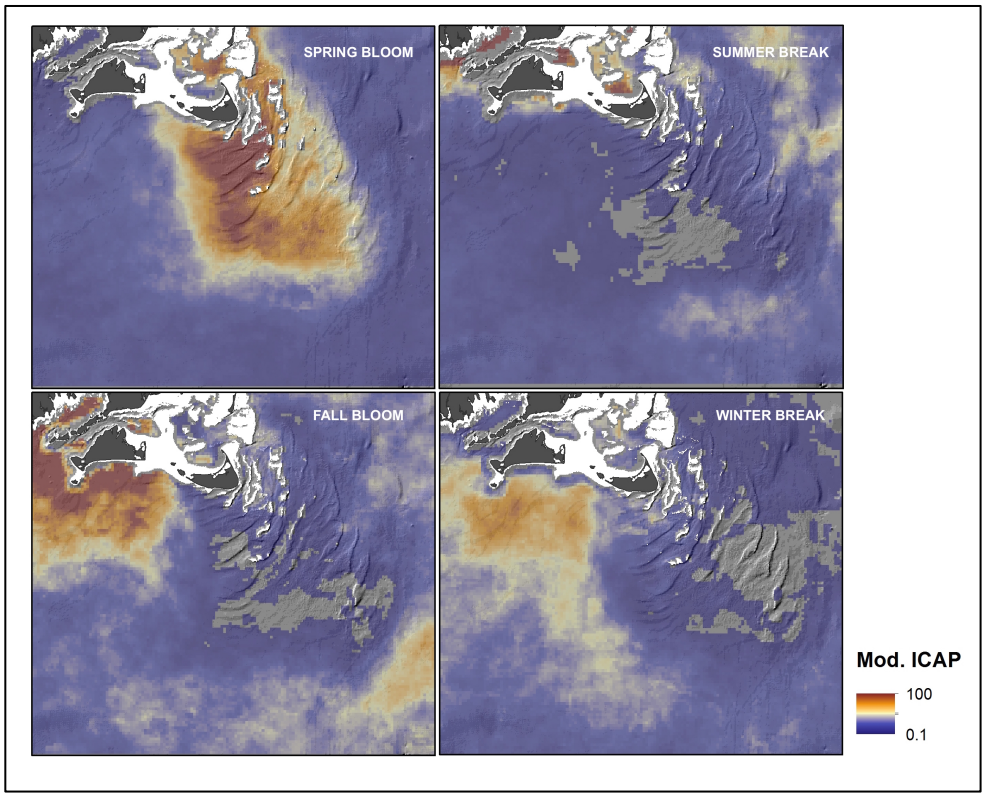


**Figure 10. – Modified ICAP values for the southern study area. (Top-left) Spring bloom, (top-right) summer break, (bottom-left) fall bloom, and (bottom-right) winter break.**

## *TEMPORAL VARIABILITY OF HOTSPOTS*

I fitted a Fourier curve to each pixel's time series of $G_i^*$ values. During this process, the model chose the optimal number of harmonics to ensure a better fit. The resulting number of harmonics chosen signal the overall seasonal patterns for each pixel. For example, in the northern study area, most of the pixels with only the 12-month harmonic (Figure 11, blue color) are located along a band east of Stellwagen Bank. This band mainly coincides with the location of the Western Maine Coastal Current, which brings cold nutrient-rich waters from the northern areas of the Gulf of Maine. Models with only this harmonic represent seasonal patterns with only one bloom during the year.

This band continues to the south for a few kilometers until it disappears off the eastern side of the study area. Most of the remainder of the area has a combination of pixels with the 12, 6, and 3 month harmonics, and pixels with only 12 and 6-month harmonics. These two cases represent areas with two major blooms, although the pixels with the 3-month harmonic have blooms with significantly different magnitudes.

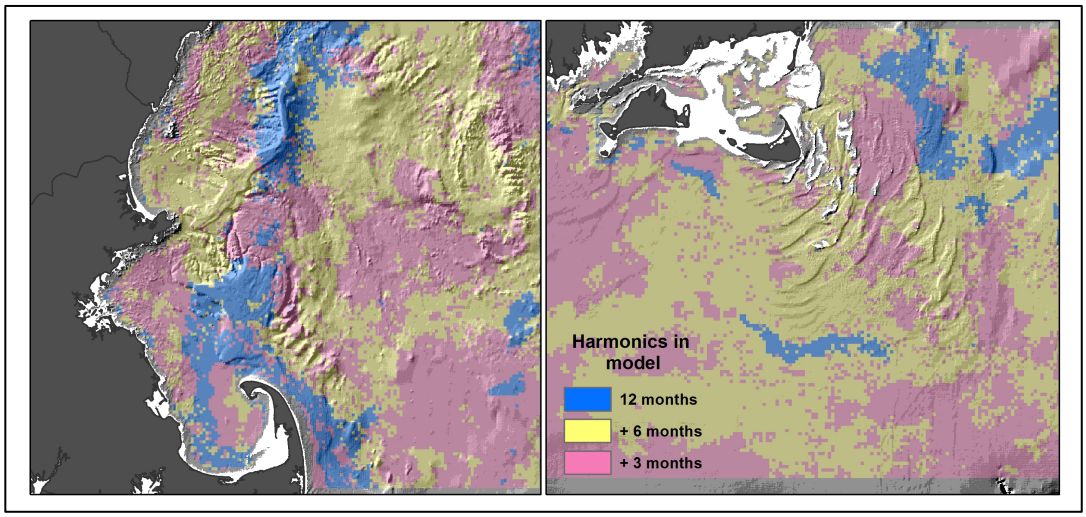


**Figure 11. – Maps showing the type of model fitted for each pixel in the north (left) and south (right) study areas. Blue colored pixels have models with only a 12-month harmonic. Yellow pixels have models with 12 and 6-month harmonics. Finally, pink pixels have models with 12, 6 and 3-month harmonics.**

The maps in Figure 12 show the amplitude values for each of the harmonics for the northern study area. Locations with a high amplitude for the 12-month harmonic have higher differences in productivity between the first 6-months of the year and the last 6-months. For the northern area, these pixels are concentrated along the coast and over Cape Cod Bay. Conversely, the locations with the lowest 12-month amplitude are along locations known for high mixing of the water column, which allows productivity to maintain a high level throughout the seasons. A high value of the 6-month harmonic indicates a high difference between seasons. In this case, pixels with high amplitude are

mainly concentrated north and south of Cape Ann and within Cape Cod Bay. The fact that Cape Cod bay has high amplitude for both the 12 and 6 month harmonic signals that only one biological season (in this case winter) has high productivity compared to the rest of the year. Finally, the amplitude of the 3-month harmonic represents variability within a season. However, unlike RMSE values (see below), this variability is persistent across the years. Pixels with high 3-month amplitude are concentrated along the areas offshore of Boston and Provincetown in Massachuetts, and Portland in Maine.

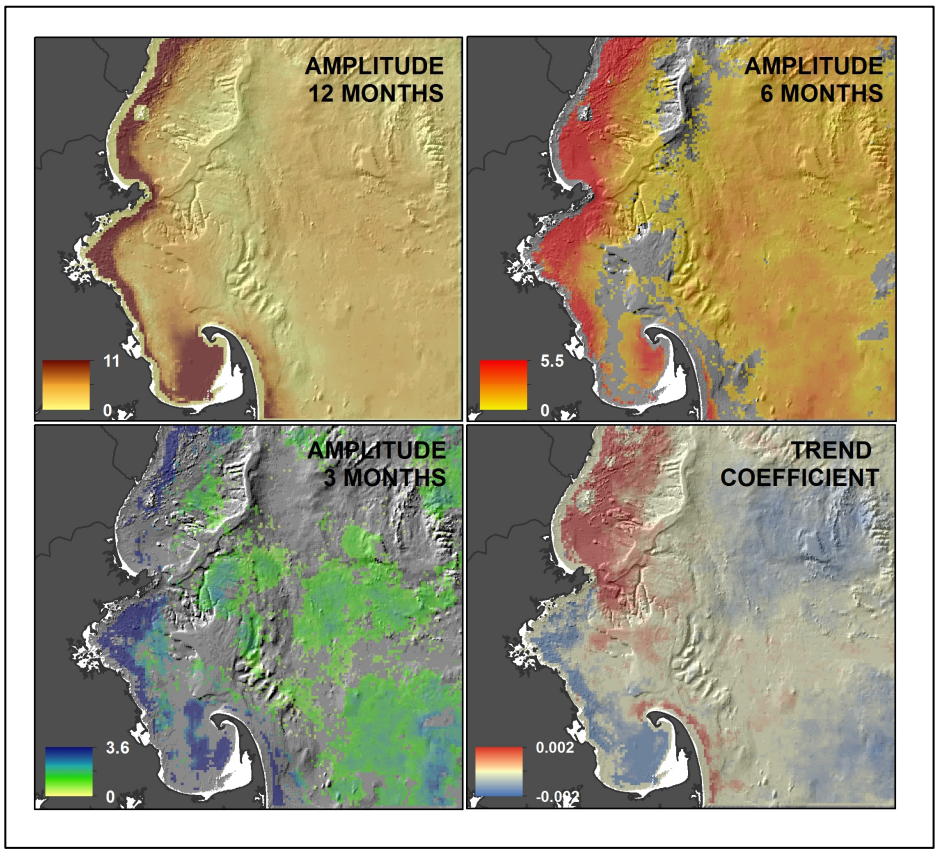


**Figure 12 –Magnitude of the amplitudes for each of the Fourier harmonics and trend values for each pixel: (from left to right, top to bottom) 12 month harmonic, 6 month harmonic, 3 month harmonic and trend coefficient. Pixels with white background signal models where the value is zero**

Figure 12 also shows the trend ($a_1$) coefficient values (representing long-term variability) for each pixel in the northern study area. Overall, trend values for all seasons are centered at zero and range from $-2 \times 10^{-3}$ to $2 \times 10^{-3}$. Generally, negative trend values appear within Massachusetts Bay and over Cape Cod Bay. Positive trend values are located north of Cape Ann and along the coast of New Hampshire and Maine. Most of the offshore areas east of Stellwagen bank have trend values close to zero. The boxplots in Figure 13 show RMSE values for pixels with modified ICAP values over 25% separated by biological season and study area. RMSE values represent the variability that is not repetitive. I considered these values as the year-to-year variability. RMSE values in both regions are lower during the fall bloom and the winter break than throughout the rest of the year. The highest RMSE values are found during the northern area's summer break, thus signaling a highly dynamic system during this time of the year.

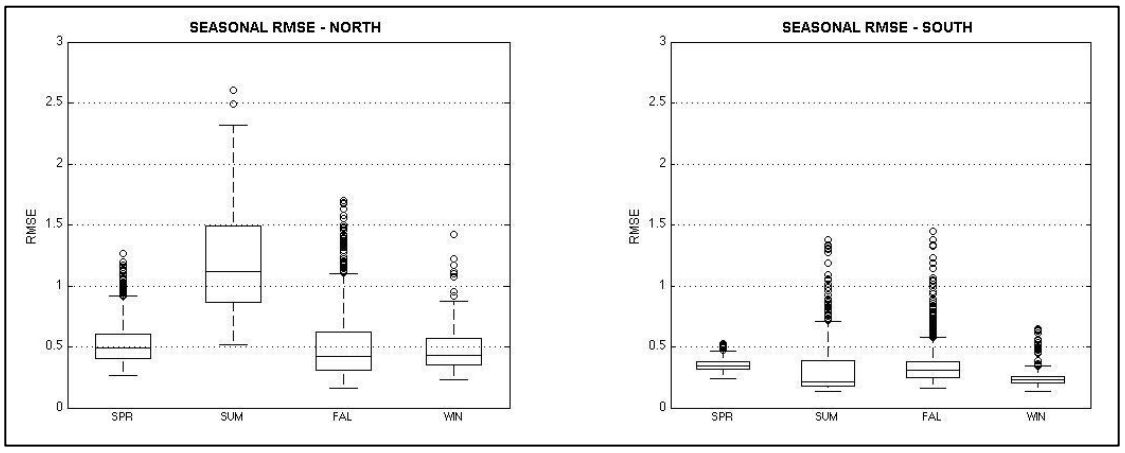


**Figure 13.- Boxplots showing the range of RMSE values within each biological season and study area. Only values for pixels with modified ICAP values above 25% within each season are shown. (Left) Northern area, (right) southern area.**

*UNCERTAINTY ANALYSIS AND VALIDATION*

Figure 14 shows the resulting amplitudes (12 months and 6 months) and trend coefficients when I vary the number of samples in a time series. The pixel I chose (with 437 cloud-free observations, the highest of all pixels) had a fitted curve that had originally 3 harmonics and a negative trend. Because the Fourier curve is fitted using a regularization method such as lasso, a coefficient is set to zero when it is either not needed to fit the curve or there is not enough data to find its value. Therefore, as the number of cloud-free observations decreases, I would expect the number of zero values to increase (represented in blue cross markers on the scatter plots). Also, as the number of observations decreases, I expect the variance for the results to increase, thus signaling an increased uncertainty.

The regularization effect can be seen in all scatter plots in Figure 14, but especially in the plot for the trend coefficient (bottom-left). In this case, as the number of cloud-free observations decreases, the trend value slightly fluctuates in the positive range. However, when the number of observations falls bellow 300, the lasso-glm is unable to correctly fit a trend value in most cases and sets the coefficient to zero. For the few cases the model fits a trend, these values show a high variance, even reaching negative values. The scatter plots for the amplitude values for the 12 month and 6 month harmonics show two very different situations. The results for the 12-month harmonic show a smaller variance, thus smaller uncertainty, as the number of cloud-free observations decrease. When the number of observations falls below 300, the model either chooses to set the coefficients to zero or fits a value within the 95% confidence interval. Conversely, the

values for the amplitude of the 6-month harmonic show a high variance as the number of observations decrease. In very few cases, the model sets the coefficients to zero.

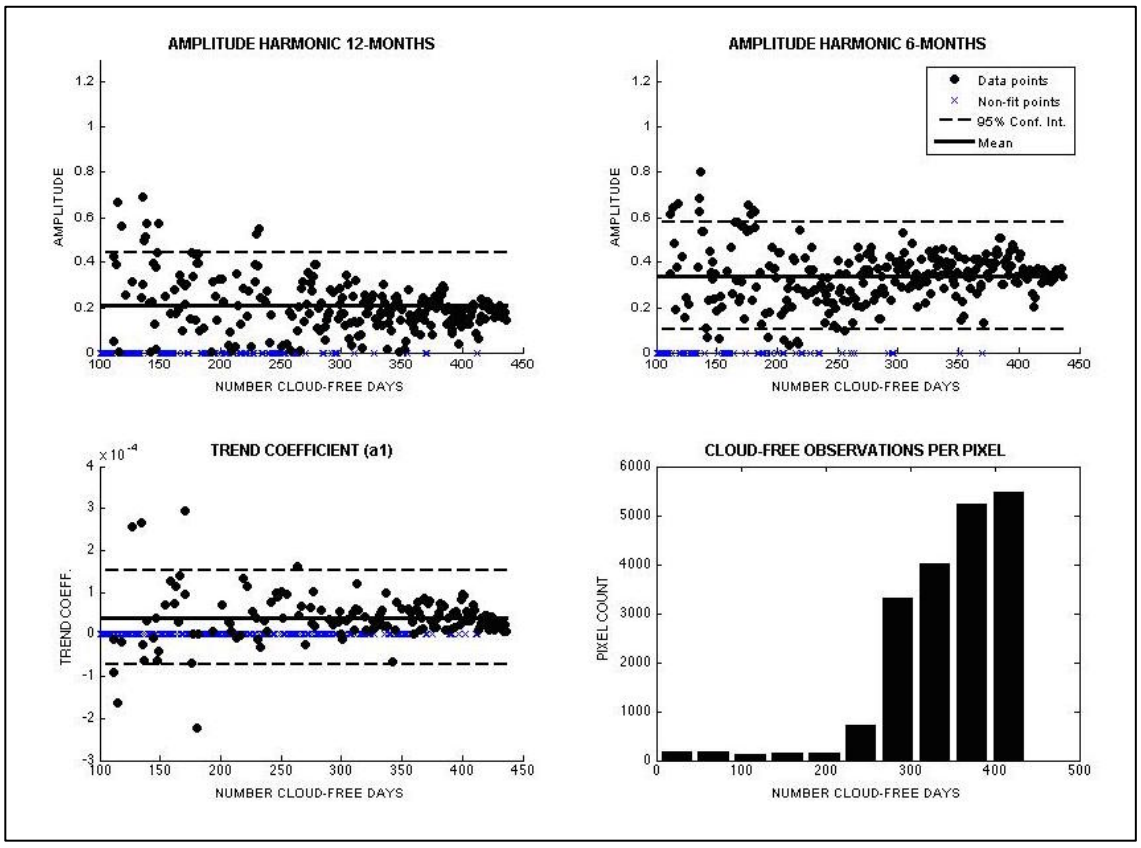


**Figure 14.- Scatter plots showing the amplitude for the 12-month harmonic (top-left), the 6-month harmonic (top-right) and trend (bottom-left) when different number of cloud-free days are used to fit the Fourier curve. Blue crosses signal runs where the lasso-glm model chose to set the variable to zero. Dotted lines represent the bounds of the 95% confidence interval and solid lines signal the mean. (Bottom-right) Histogram showing the number of cloud-free days per pixel.**

Finally, I compared the amplitudes derived from the MERIS remote sensing data to the amplitudes derived using the MWRA field data (Figure 15). The scatter-plot shows all stations but one clustered within a small range of amplitude values. As previously explained, this lack of a wide range of values is one of the main limitations of using the

MWRA dataset for validation. Also, because I was unable to calculate the G* index due to the high distance between stations, the values from the MWRA dataset cannot be compared 1:1 to the values derived from the MERIS dataset. However, I was able to fit a linear regression to the resulting amplitude values. The regression line has an equation with a positive slope (2.62) and an intercept near zero (0.36). The R-square for this regression is 0.829, signaling a tight fit.

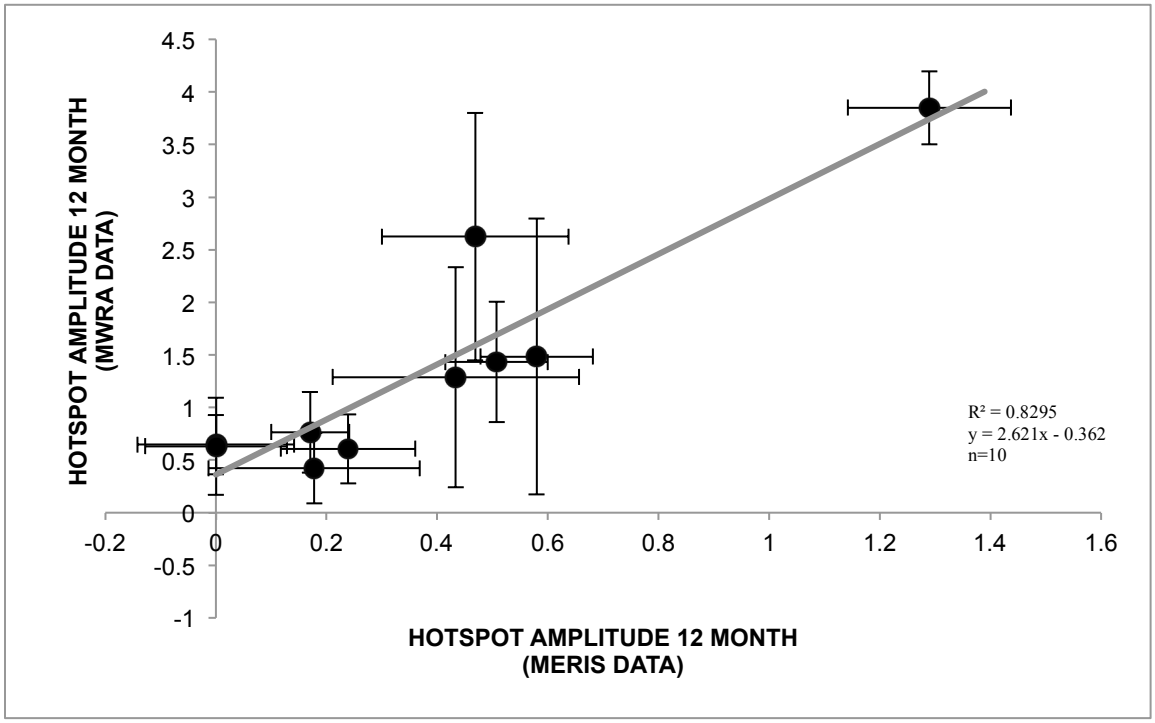


**Figure 15.- MERIS-derived yearly amplitude of hotspot values versus the MWRA-derived amplitude for each of the MWRA field stations. The error bars represent ±1 standard error of the mean of the yearly amplitudes (maximum – minimum of all values per year)**

## CONCLUSIONS

In this study I presented a novel index to locate and study local productivity hotspots in a marine ecosystem. Because ICAP values are based on anomalies and not

absolute productivity values, they show a clear picture of the areas where processes concentrate productivity at a higher rate than what is expected for the region. When used on studies related to marine ecosystems, ICAP values provide various advantages over other previously used metrics. For example, compared to averages of chlorophyll-a concentrations, ICAP values provide information on the temporal persistence of hotspots in each area, which help researchers studying very dynamic systems. Also, ICAP values are compared to indices based on percentages of positive anomalies (e.g. Suryan et al. 2012), the former are able to separate distinct areas based on the magnitude of their anomalies, addressing one of the main limitations from percentage-based measurements (Winiarski et al 2013).

The ICAP index also facilitates the testing of research hypothesis related to spatial and temporal heterogeneity in marine ecosystems. In this study, the maps of the seasonal ICAP values in both study areas of the Gulf of Maine support the hypothesis that the magnitude and spatial distribution of local primary productivity hotspots vary by season. They also support the idea that different processes of different scales aggregate phytoplankton at each area. For example, above-average spring blooms in the northern area occur in areas with high topographic changes, while fall hotspots concentrate near coastal areas where there is a release of nutrients that stimulate phytoplankton abundance. ICAP values may also be used to locate areas where the timing of the phytoplankton blooms is persistently different than the rest of the region. One clear example of this is the high magnitude hotspot during the winter break in Cape Cod Bay. This persistent hotspot signals an early spring bloom in the area, which has been recorded in the past in

various field data studies (e.g. Bigelow et al. 1940). The derivation of the regional climatologies prior to calculating ICAP values also provides information on the distinct processes occurring at a regional scale. It is important to ensure that the climatologies are properly fitted and that areas with different regional processes are studied separately. For example, if our two study areas were to be studied together, hotspot values between blooms in the south would have been overestimated.

The ability to track productivity over time for each pixel allows us to not only understand what the normal pattern is for the area but also to determine when a location persistently departs from this normal. The results from our area of study show that the inter-annual and long-term variability of hotspots are both location and season dependent. The simplicity of the Fourier algorithm provides the means of comparison between different sources of temporal variability. Especially interesting is the comparison between the 3-month amplitude values and RMSE values, as they can describe how predictable the within-season variability may be. For example, in certain areas in the middle of Stellwagen Bank, the 3-month amplitude is relatively low while the RMSE is relatively high. This signals an area that does not show persistent high concentrations of productivity, although occasional ephemeral local processes may concentrate productivity. Trend values are useful at determining the effects of long-term climate change on patterns of marine productivity, although with only ten years of data, the results in this study should be interpreted cautiously. However, even within the available data, certain patterns are starting to emerge, which correlate to changes users have started to see in the field. For example, for the past few years there has been a decline in the

abundance of seabirds in the southwestern corner of Stellwagen Bank, while seabird abundance has increased further south near Chatham, MA (personal communication with researchers in the Stellwagen area). This effect can be seen in the results from this study, especially when spatially plotting the fitted-curves trend values.

The main objective of this research was to explore the spatial and temporal complexity in a marine ecosystem. I wanted to provide the basis to formulate new hypotheses and a metric to be able to test them in the future. Our novel cumulative hotspot index, ICAP, was able to detect biologically relevant productivity aggregations using remote sensing data. This information combined with the coefficients from the per-pixel fitted Fourier curves provided a clearer picture on the complex dynamic processes occurring in the western Gulf of Maine than existed previously. The results from this study also strengthen our belief that the study of primary productivity hotspots may be an important tool at understanding and monitoring the consequences from long-term processes such as climate change. Besides, given that hotspots are often targeted by fishermen and other resource extractors, they are of critical importance to fisheries managers. I hope this research offers the first step for ecosystem wide analyses, which may help modify how marine ecosystems are managed in the future.

.

# CHAPTER THREE: LOCAL BIOLOGICAL HOTSPOTS IN THE GULF OF MAINE: SPATIAL MATCH/MISMATCH BETWEEN PRIMARY PRODUCTIVITY AND FISH ABUNDANCE

## ABSTRACT[3]

Primary productivity values from remote sensing sources have been widely used in the past to predict aggregations of top predators in marine ecosystems, even though these organisms feed on species several levels higher up the trophic chain. High productivity has also been used as a predictor of overall benthic biomass, even though the connection between the two is not direct. With this inclusion of productivity information to prediction models, researchers assume that some locations facilitate the spatial and temporal overlap between the availability of food resources and the species that depend on them. These locations, known as "biological hotspots", often originate with aggregations of phytoplankton, which attract organisms higher up the trophic chain. However, because there is no direct trophic transfer between them, it still remains unclear whether there is in fact an overlap between primary producers and both pelagic and benthic predators. Previous attempts to test this hypothesis, often known as the 'Spatial Match/Mismatch Hypothesis', have had inconclusive results, due in part to the metrics used to account for aggregations of primary producers. In this study, I explored the spatial match between surface productivity and fish abundance in the western Gulf of Maine. I tested the hypothesis that aggregations of primary producers spatially correlate

[3] To be submitted to "Marine Ecology – Progress Series". Structure matches the publication's guidelines.

to both pelagic and benthic fish abundance. I also explored whether primary productivity is a significant predictor of fish abundance, together with abiotic variables such as sea surface temperature and topography. I quantified surface productivity using the Index of Cumulative Anomalous Productivity (ICAP) applied to MERIS remote sensing data. Benthic and pelagic fish abundances were derived from the Vessel Trip Report (VTR), a fisheries-dependent data source. Results show a significant spatial correlation between pelagic fish abundance and aggregations of primary productivity for both spring and fall seasons. Spatial correlations were also significant between benthic fish abundance and primary productivity hotspots during spring months, but not during fall. Abiotic variables alone were strong predictors of both overall surface productivity and benthic abundance, signaling the possibility that environmental conditions are the cause of the spatial overlap between surface and benthic resources.

## INTRODUCTION

Biological hotspots are generally defined as locations where high number of organisms from different trophic levels aggregate (Worm et al. 2003, Malakoff 2004). These are usually areas that provide the right conditions for the lower levels of the food chain to grow and where predators aggregate to feed on abundant resources. Locating and characterizing these biological hotspots provides information on the drivers of spatial heterogeneity in marine ecosystems, and it is considered a good first step towards understanding the complexity inherent in these systems (Wulder and Boots 1998). Biological hotspots are supported by high concentrations of phytoplankton organisms,

which provide the resources that attract organisms higher up the trophic chain. These primary producers may also determine the spatial and temporal characteristics of the resulting hotspots (Norse et al. 2005, Nur et al. 2010, Incze et al. 2010). In general, the abundance and distribution of phytoplankton across space depend on either environmental conditions that stimulate surface productivity, such as temperature or nutrient availability, or oceanographic processes that concentrate productivity produced elsewhere (Mackas et al. 1985). Topographic features can enhance aggregations of productivity by facilitating processes such as upwelling or internal waves (Scotti and Pineda 2004). Due to the high temporal and spatial variability of these processes, biological hotspots tend to be highly variable as well.

The concept of biological hotspots implies there is a temporal and spatial overlap between organisms and the resources they depend on. For example, aggregations of seabirds often temporally and spatially overlap with surface aggregations of forage fish, thus creating a biological hotspot (Davoren 2007). Some researchers have referred to this as the "Spatial match-mismatch hypothesis" (Durant et al. 2007). This hypothesis states that the survival of a predator is dependent on the spatial and temporal overlap between its distribution and the one of its prey. While this hypothesis has been difficult to test directly, the spatial overlap between predators and resources is often assumed in many studies, especially in ecosystem-wide models (e.g. Steele et al. 2007). In fact, this overlap is often assumed even when the groups of organisms do not share a direct trophic link. For example, phytoplankton abundance has been repeatedly used for predicting the location and abundance of both pelagic (e.g. Worm et al. 2003) and benthic predators

(e.g. Wei et al 2010), even though these predators do not feed directly on primary producers. The use of surface primary productivity to predict abundance of other organisms either assumes that the conditions that facilitate aggregations of phytoplankton organisms may also encourage aggregations of other species (Reese and Brodeur 2006), or that the productivity not consumed may deposit and enhance secondary productivity in the long term (Townsend and Cammen 1988).

Understanding the spatial overlap between primary producers and other tropic levels is not a trivial matter. Recent studies have described changes in timing, abundance and location of phytoplankton blooms due to climate change (Behrenfeld et al. 2006), which may result in shifting species' distributions towards northern latitudes (Jones and Cheung 2014). This in turn may produce a spatial mismatch between the locations of the resources and the organisms depending on them, which has already been reported in land ecosystems (Schweiger et al. 2008). However, the study of this spatial overlap in marine ecosystems has been especially problematic in part because the high temporal and spatial variability of phytoplankton organisms. For example, Grémillet et al. (2008) concluded there was a mismatch between surface and benthic productivity, although indices used to describe primary productivity did not facilitate the comparison in highly dynamic ecosystems. In the past few years, a few researchers have proposed a shift towards redefining biological hotspots from the point of view of primary producers, taking into account their temporal and spatial heterogeneity (Suryan et al. 2012). From this point of view, hotspots are defined as areas that exhibit productivity aggregations at a higher frequency compared to the rest of the region. These are areas that facilitate the growth or

concentration of phytoplankton organisms, and in turn encourage the transfer of energy from primary producers to the rest of the trophic chain. Suryan et al. (2012) and later Ribera et al. (Chapter 2, in this thesis) adapted this new definition of biological hotspots into novel metrics that locate hotspots based on the frequency and magnitude of productivity anomalies (i.e. residuals) above the expected regional seasonal pattern.

The main objective of this study was to test whether there is a spatial overlap between primary producers and both surface and benthic fish organisms. For this reason, we used anomaly-based metrics on remote sensing datasets to define biological hotspots, and explored the advantages this type of metric has over previously used approaches. This study was structured in two main parts: first, we used remote sensing and fisheries dependent data to test whether abiotic variables alone drive aggregations of both pelagic and benthic organisms; and second, we tested whether adding surface productivity to the model better accounts for variability in abundance of surface and benthic fish species.

## METHODS

For this study, we followed the methodological steps represented in Figure 16. This study includes four main steps: (1) location of biological hotspots using remote sensing datasets, (2) exploration of the spatial correlation between primary productivity hotspots and fish abundance data from both the surface and bottom of the water column using semi-variograms, (3) fitting a model predicting fish abundance using only abiotic variables (temperature and topographic variables), and (4) adding surface productivity hotspot information to this model to determine whether it adds any predictive power.

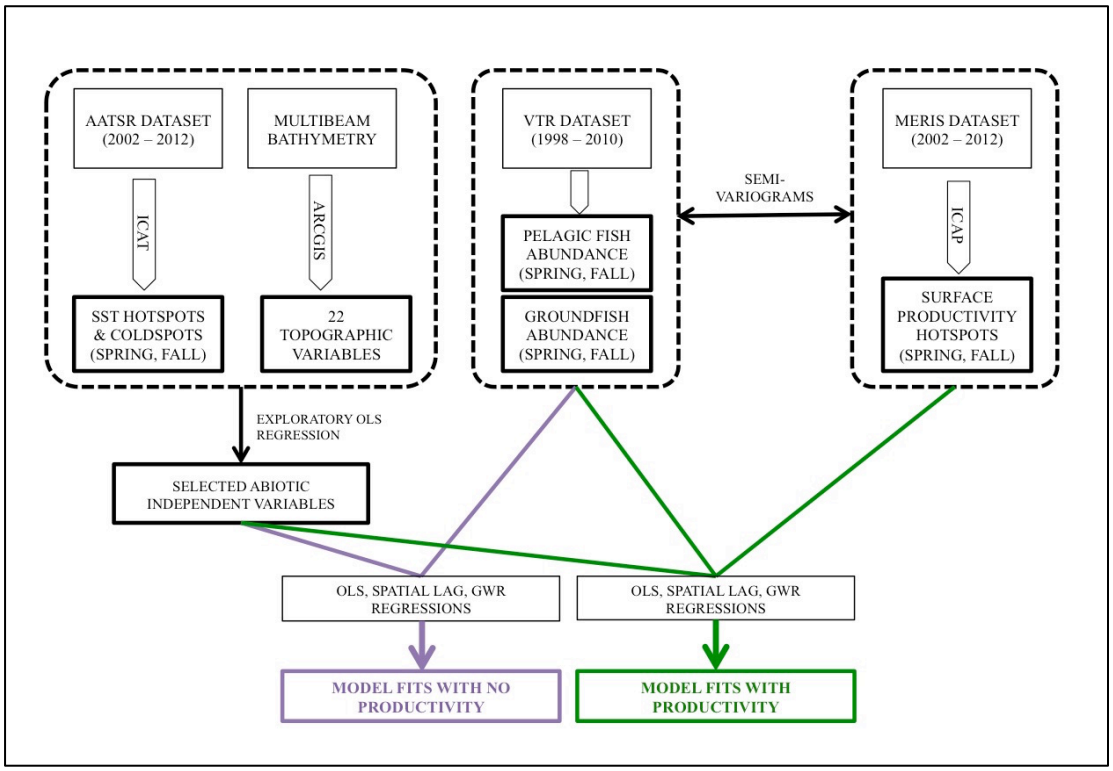


**Figure 16. – Visual representation of the methodological steps in this thesis**

I chose to run this study in a section of the Western Gulf of Maine (Figure 17). This area is centered over the Stellwagen National Marine Sanctuary, which is located about 30 miles offshore from the city of Boston (MA). This is a highly productive temperate system, well known as a fishing area for both pelagic and benthic species. Figure 17 shows both the limits of the study area and the limits of the regional area. The regional area is the reference extent for the calculation of hotspot values. The limits of the study area correspond to a statistical fishing area known as area 514 (see below).

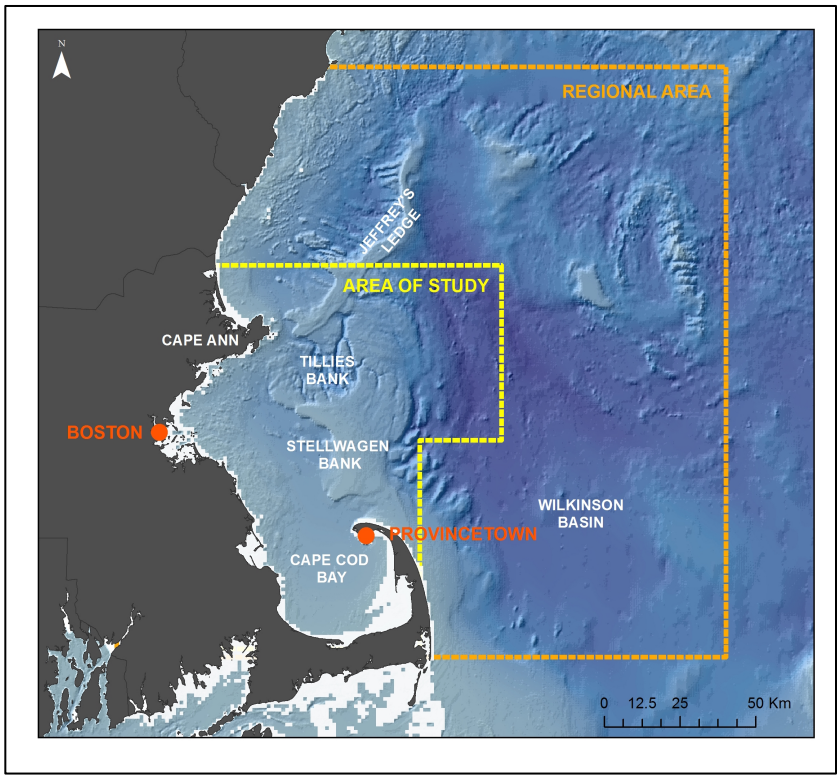


**Figure 17. - Map showing the area of study (yellow), which corresponds to VTR statistical area 514, and the regional area (orange) used as a reference to calculate hotspots and coldspots. White-colored cells are locations with depths shallower than 10 meters and thus not included in the study area.**

## *LOCATING PRIMARY PRODUCTIVITY HOTSPOTS*

I used chlorophyll-a concentration derived from the Medium Resolution Imaging Spectrometer (MERIS) sensor as a proxy for phytoplankton abundance. MERIS is a sensor on the satellite ENVISAT managed by the European Space Agency (ESA). It collects ocean-color data at a 1200m resolution every 3 days. I used all available level-2 data from April 2002 to April 2012 (1661 layers). I processed all layers using BEAM (software created by Brockmann Consult for ESA), removing pixels flagged for clouds, high-glint, and ice. All layers were also collocated to the same reference grid so the

coordinates for the center of all pixels were the same in every layer. I removed any value above 45mg/l, as I considered this threshold to represent the maximum concentration of chlorophyll-a recorded in the area of study (Hyde et al. 2007). For each layer, the pixels contiguous to clouds (i.e. 2 pixel buffer around each cloud) and the ones with a depth below 10 meters were also removed. Finally, to eliminate the possibility of missed clouds and cloud-shadows, I fitted a *lowess* non-parametric curve to each pixel's time series. Its residuals were then converted to z-scores, and values above 5 and below -5 were removed from the dataset.

I used the *Index of Cumulative Anomalous Productivity* (ICAP) created in chapter 2 to determine whether each pixel may be considered a hotspot. This index is based on anomalies (i.e. residuals) above the regional seasonal pattern. It is a measure of both the persistence (i.e. the probability a location becomes a hotspot) and the strength (i.e. the magnitude of the anomalies over time) of productivity in a location. It is based on an index created by Suryan et al. (2012) named *Frequency of Chlorophyll Peaks Index* (FCPI). However, unlike the FCPI index, which defines a hotspot as an area with a frequent presence of high anomalies (above the 80% confidence interval), the ICAP considers both the spatial structure of each data layer to determine whether a location is considered a hotspot and the magnitude of the anomalies.

I followed the methodology described in chapter 2 to derive the ICAP values. The process involves five main steps: (1) mean-center all values (subtract the all-time mean from each pixel), (2) calculate a spatial average from each layer and derive a regional climatology (i.e. seasonal pattern); (3) derive the anomalies (i.e. residuals) for each pixel

and time period by subtracting each pixel's time series to the regional climatology; (4) calculate the local spatial autocorrelation value (Getis and Ord's local G* index) for each pixel; and (5) add all the significant positive spatial autocorrelation values within a season. I calculated the ICAP indices for both spring and fall seasons. However, following the same process in chapter 2, I defined the seasons based on the start of the spring and fall blooms in the region. For each of the seasons, I also derived the FCPI and an overall mean value, and compared the results.

## *ABIOTIC VARIABLES*

### *Sea surface temperature hotspots and coldspots*

Sea surface temperature (SST) information was derived from the Advanced Along Track Scanning Radiometer (AATSR) dataset. Like MERIS, this radiometer is also on the ENVISAT and managed by ESA. The AATSR features a dual-view of the ocean's surface, providing one image at nadir (perpendicular to Earth's surface) and one forward view. This ensures the accurate detection of low-level clouds, including fog. Also, the on-board calibration system ensures that the temperature is accurately recorded and that the processing of the resulting data does not need to be compared to field measurements. I used Level-2P data, which was processed by the Group for High Resolution Sea Surface Temperature (GHRSST). The AATSR L2P dataset already came pre-processed, with pixels near the coast and around clouds removed.

The AATSR dataset provides SST temperature at a 1km spatial resolution and a repeat frequency of 3 days. This data was resampled to 1200m to match the resolution of

MERIS. To derive SST hotspots, I followed the same methodology used in the case of chlorophyll-a hotspots. These hotspots represent areas that show high frequency of positive temperature anomalies throughout a season. I called these ICAT values (Index of cumulative anomalous temperature). I also derived SST "coldpots", which I defined as areas that consistently show low temperature anomalies (with G* values below -2).

*Topographic features*

I used multibeam bathymetry to assess the topographic characteristics of the seafloor. This dataset was collected by Valentine et al. (2000) and provides depth information for each cell. It has a native spatial resolution of 10 meters. Using multibeam bathymetry, I also derived information on the shape of the seafloor. Particularly, I measured the Slope, Aspect and Benthic Position Index (BPI) (Figure 18). The first two variables were calculated using ArcGIS 10.1 and its Spatial Analyst Extension. The latter measure was calculated using ArcGIS and the Benthic Terrain Modeler toolbox (Wright et al. 2005). The slope is measured as the maximum rate of change from the value of a cell to its eight surrounding cells. It is measured in degrees, from 0 (flat) to 90 (fully vertical drop). Aspect determines the direction the slope is facing (i.e. North). It is measured clockwise in degrees, from 0 to 360. Finally, I also derived the distance in meters from each data point to the coastline.

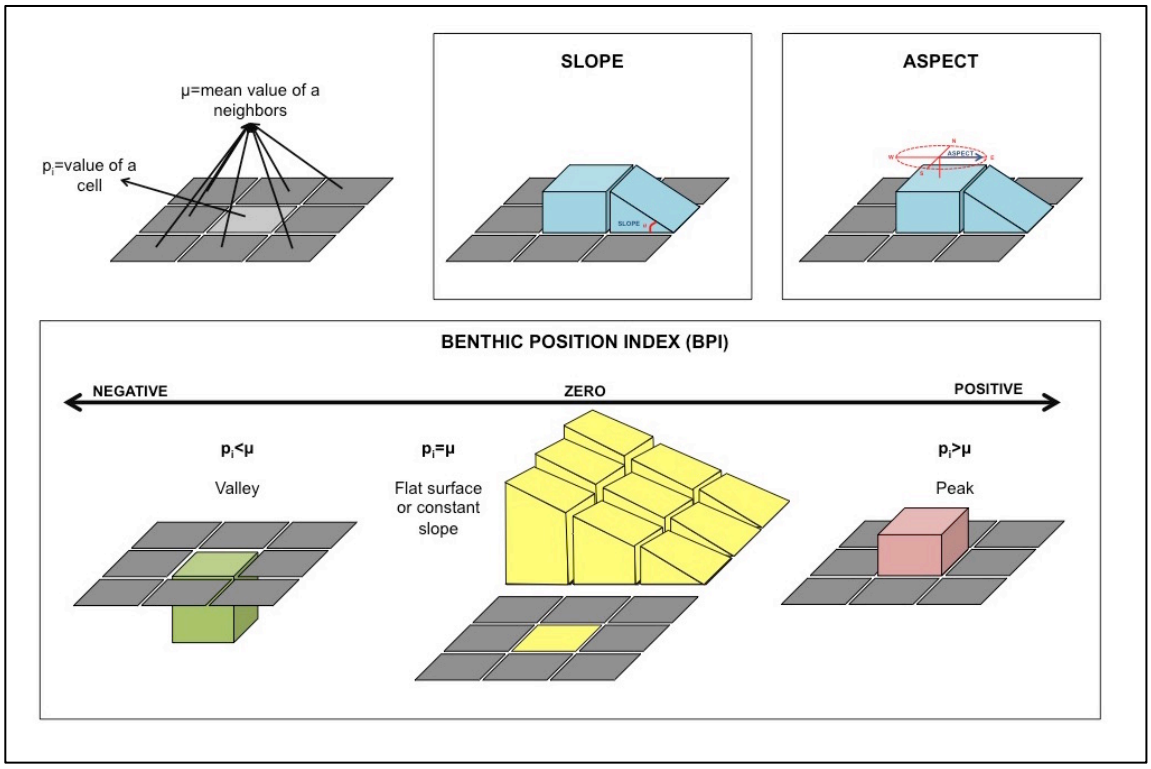


**Figure 18 – Visual representation of Slope, Aspect and Bathymetric Position Index (BPI)**

Bathymetric Position Index (BPI) is a measure of the position of a cell relative to its neighbors, and it is used to determine whether a cell is located in a ridge or a valley (Figure 18, bottom). It is analogous to the Topographic Position Index (TPI) used in land-based studies. A cell with a positive BPI value represents an area shallower than its surrounding cells, signaling a possible ridge or peak. On the other hand, locations with negative BPI values typically signal valleys, which are areas deeper than their surrounding cells. Locations with BPI values close to zero represent either flat areas (slope near zero) or areas where the slope is constant (slope of a cell significantly greater than zero). The BPI index is scale-dependent, so its values depend on the amount of cells

used to calculate it. I calculated BPI indices at two separate scales: fine scale (using a buffer of 3 cells) and broad scale (using a buffer of 25 cells).

Multibeam bathymetry and its derived products have a spatial resolution of 10m. However, to be able to compare them to chlorophyll-a or temperature data, they need to be resampled at a resolution of 1200m. Often, this resampling is done by averaging all values within a 1200m cell. However, in this case I not only calculated the mean, but also the standard deviation, range, maximum and minimum values within each 1200m for each of the derived products. For example, cells with high range values of BPI may signal areas where features change from peaks to valleys within the same 1200m cell.

### *FISH ABUNDANCE*

For this study, I used catch per unit effort (CPUE) derived from the Vessel Trip Report (VTR) dataset (Orphanides and Magnusson 2007) as a proxy for fish abundance. VTR, also known as logbook, is a report each fishing vessel is required to fill after each trip, or within a trip if there is a change in gear or fishing area (based on statistical areas). Most federally managed fisheries (except lobster fishery) are included in this dataset. Each report includes, among other information, the date and time of the trip, the location fished, the type of gear used, the species caught, and the total amount landed. VTR is a fisheries-dependent self-reported dataset, and as a result, its use as a proxy for fish abundance has a series of limitations. First, the locations of each of the sampling points are not random, as they depend on where the fishermen travelled. Therefore, the whole

area of study has not been sampled with the same effort. Also, the voluntary reporting nature of this dataset adds more uncertainty to the results.

I used VTR records from 1998 to 2010 in the reporting zone number 514, which includes Stellwagen Bank, Cape Cod Bay and the southern part of Jeffrey's ledge. Identifying information such as the vessel permit number or name were removed prior to analysis. I aggregated data to 1200m grids and cleaned for clearly incorrect locations (i.e. on land). I only considered records for gears classified as trawls, including bottom and mid-water trawls. I separated the records in two main groups based on the type of species caught: pelagics, also known as forage fish (e.g. atlantic herring), and groundfish (e.g. atlantic cod). These groups were based on the classification used in VTR records. For each group, I calculated CPUE as the total catch divided by the number of hours the trawl was in the water. I considered total catch as the sum of the number of individuals caught (both the ones kept and the ones discarded) within a cell and season. I derived CPUE for every year and season, and then calculated a seasonal average.

### *EXPLORING SPATIAL CORRELATION BETWEEN VARIABLES*

I first explored the spatial correlation between ICAP values for spring and fall, and benthic and pelagic fish abundances using semi-variograms (Curran 1988). Semi-variograms compare the value of a point with the values within a spatial lag distance. Semi-variograms are useful to visualize the spatial pattern of the variability in the dataset. In this case, I first run a linear regression between each pair of dependent-independent variables (e.g. groundfish abundance ~ ICAP spring), and then fitted the variogram to the

residuals. This way, I could explore the spatial pattern of the correlation between the variables. Spatially correlated variables should result in semi-variograms where values are smaller at shorter spatial lags, and increase as the spatial lag distance increases, until they reach a plateau. To fit the experimental variograms, I used R statistical software (version 3.1.1) with the function "variogram" from the package "gstat". I binned results every 1200m, and standardized the variables prior to fitting the semi-variograms.

### *PREDICTING FISH ABUNDANCE FROM ABIOTIC AND PRODUCTIVITY INFORMATION*

To test whether adding information on primary productivity hotspots improves the predictions of both benthic and pelagic fish abundances, I run two separate linear regressions: one with only abiotic variables (topography and SST) and one including both abiotic variables and ICAP values. However, prior to running the predictive models, I determined the best combination of abiotic independent variables for each dependent variable. I run multiple Ordinary Least Squares regressions (OLS) for each dependent variable (abundance of ground-fish during spring and fall seasons, and abundance of pelagics during spring and fall seasons) with all possible combinations of independent variables, including all topographic variables and SST hotspots and coldspots. Prior to running each OLS regression, the Belsley, Kuh, and Welsch's collinearity diagnostics were run (Belsey et al. 1980) to ensure only combinations of independent variables with weak collinearity (i.e. condition indices lower than 30) were tested. Also, all variables were standardized prior to running the regressions. This resulted in 404 overall model

runs. I compared the models using the Akaike Information Criteria (AIC), and choose the model that minimized this value.

For the prediction of fish abundances from abiotic variables, I tested three types of linear models: OLS, Spatial lag, and Geographic weighted regression (GWR). OLS fits a regression line to the data by minimizing the size of the prediction errors. This type of model assumes that the observations are independent of one another. However, often in ecological studies, the value of a data point depends on the values of its neighbors, a property known as "spatial autocorrelation" or "spatial dependence". When this occurs, the residuals of the OLS regression are not randomly distributed in space, showing often a clustered pattern. One way to include this spatial dependence in a regression model is through adding a spatial lag relationship between the points (i.e. weighted information from the immediate neighbors for each location). This results in what is known as the "Spatial lag model". While OLS and Spatial lag models fit one regression line for the study area, GWR fits a regression line for each data point and its neighbors (Brunsdon et al. 2002). GWR is often recommended to account for the spatial variation in the relationship between dependent and independent variables. This occurs when different ecological processes occur at different spatial locations. For each dependent variable, I fitted the three types of regression models using the combination of independent variables that had the best AIC values when tested through exploratory OLS regression. OLS and Spatial-lag models were fitted using the application GEODA, created by Anselin et al. (2006). GWR models were fitted using the application GWR4, created by Nakaya et al.

(2005). I then run the same models adding seasonal ICAP values. All these model runs were compared using AIC values.

# RESULTS

## *LOCATING SURFACE PRODUCTIVITY HOTSPOTS*

Following the methodology described in Chapter 2 in this thesis, I derived cumulative hotspot values (ICAP) for both spring and fall seasons. The limits of the seasons were set based on the phenology of the regional climatology. Spring started on day number 70 (March 11$^{th}$) and ended on day number 170 (May 19$^{th}$), and fall started on day number 239 (August 27$^{th}$) and ended on day number 316 (November 12$^{th}$). I plotted ICAP values to visually compare the results for each season (Figure 19). Spring values are concentrated mainly near Cape Ann, over Tillies' Bank, the southwestern corner of Stellwagen Bank, and over Cape Cod Bay. On the other hand, fall values are overall higher than spring values and they are mainly concentrated along the coast.

I compared ICAP values with two other popular productivity metrics: the FCPI by Suryan et al. (2012) and the seasonal average chlorophyll-a value. Figure 20 shows the results for the spring season. Visually, the map with ICAP values exhibits more defined hotspots areas than the one for the mean values. Because ICAP values are derived using the G* index, they also have less isolated high or low isolated values compared to the FCPI map. Also, FCPI does not take into account the value of the anomaly, just the fact that it surpasses a threshold, so the values are more homogeneous across the area.

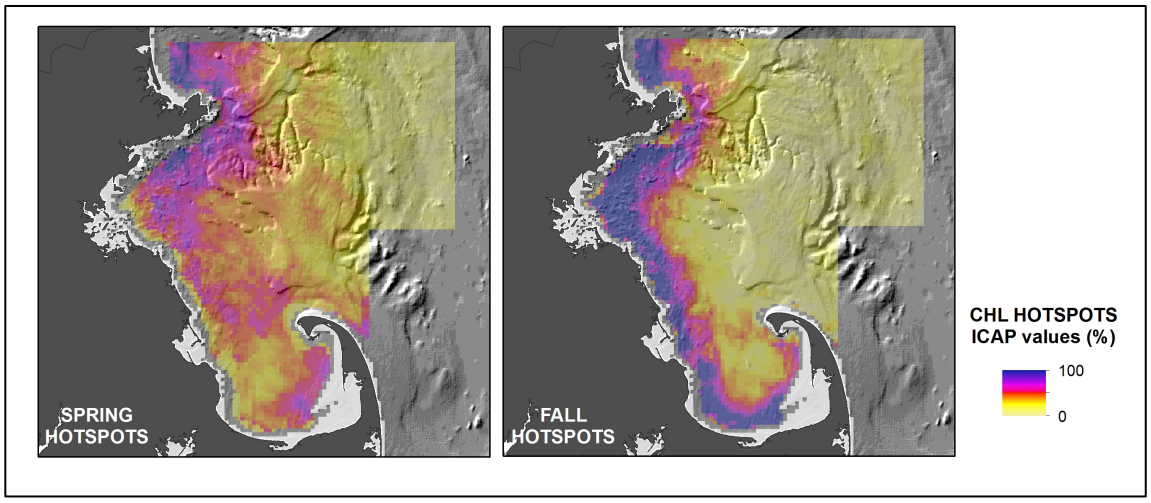


**Figure 19. – Cumulative seasonal productivity hotspot values (ICAP): (left) Spring, (right) Fall. Values range from a persistent high spatial concentration of anomalous values (dark purple) to a low concentration (yellow). White-colored cells are locations too shallow to be included in the study area.**

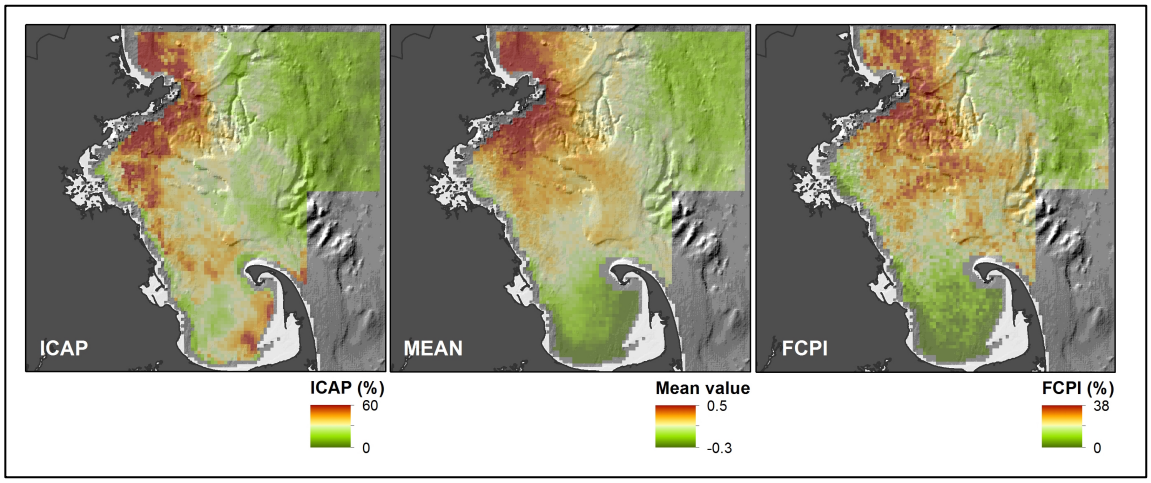


**Figure 20. – Comparison between different hotspot metrics for Spring. From left to right: ICAP value, seasonal mean and FCPI. White-colored cells are locations too shallow to be included in the study area.**

## *ABIOTIC VARIABLES*

### *Sea surface temperature hotspots and coldspots*

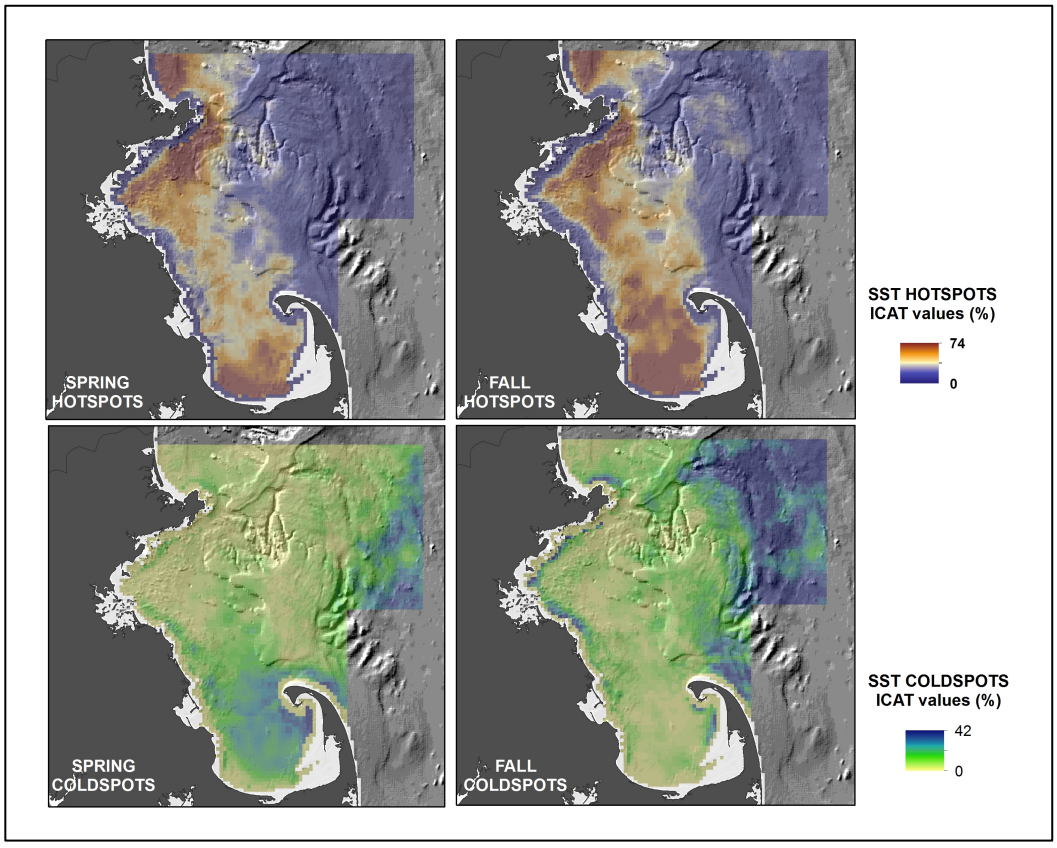


**Figure 21. – Sea surface Temperature cumulative Hotspots (top) and Coldspot (bottom) values, for both Spring (left) and Fall (right) seasons. White-colored cells are locations too shallow to be included in the study area.**

I used the same methodology described above to derive hotspots and coldspots of sea surface temperature. Figure 21 shows these ICAT values for both spring and fall seasons. During the spring bloom, positive temperature anomalies (i.e. values higher than the regional seasonal pattern) are concentrated around Cape Ann, part of Cape Cod Bay, and the Southwest corners of Stellwagen Bank. Opposite to what occurs in other

locations, the water column during this time of the year is colder in the surface than at the bottom, so warmer anomalies at the surface may signal upwelling events mixing the water column. On the other hand, the cold temperature anomalies are concentrated around Provincetown and offshore, east of Stellwagen Bank. During the fall bloom, high temperature anomalies strengthen in Cape Cod Bay and remain the same around Cape Ann and over the southwest corner of Stellwagen Bank. However, the low temperature anomalies are mainly concentrated east of Stellwagen Bank, just south of Jeffrey's ledge.

*Topographic features*

From multibeam bathymetry layers, I derived depth, slope, aspect, and BPI values at two spatial scales (broad and fine). Figure 22 shows the comparison between the BPI values at both scales. These maps show the native resolution of the multibeam bathymetry layers (10m). BPI values above zero signal areas that may be considered peaks. At a fine scale, these are located mainly along the border of Stellwagen Bank and Jeffrey's ledge, and as small features just offshore of Boston. At a broad scale, high BPI values cover most of the major topographic features in the study area, such as Tillies Bank and Jeffrey's ledge. BPI values below zero signal areas that may be considered valleys. At a fine scale, these BPI values draw a border along the main topographic features. However, at a broad scale, negative values are concentrated in between the banks at Tillies, and surrounding the main features of Stellwagen Bank and Jeffrey's ledge.

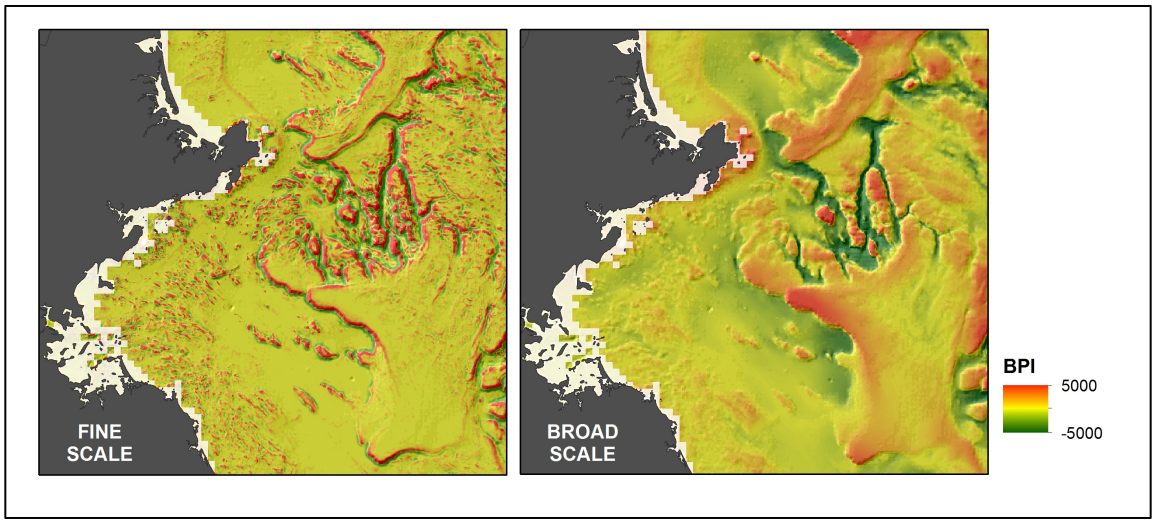


**Figure 22. – Comparison between BPI at a fine scale (left) and broad scale (right). High BPI values show peaks (red) and low values show valleys (green). Yellow colored cells signal areas with BPI close to zero. White cells are locations outside of the study area.**

*FISH ABUNDANCE*

Figure 23 shows average CPUE values for each season (spring and fall) and species group (groundfish and pelagics). The pink dotted line represents the limits of the West Gulf of Maine Closure. This closure, which has been in place since 1998, prevents any bottom fishing within its borders. As a result, for all seasons, the highest CPUE values for bottom species are along the western border of Stellwagen Bank. In spring, the highest abundance of groundfish concentrates towards the northwest corner of Stellwagen Bank. On the other hand, abundance values are more evenly distributed during fall months, mainly over Tillies bank, and Stellwagen bank's northwest and southwest corners. In the case of pelagic fish, due to seasonal closures and the fact that very few boats are responsible for most of the catch, in spring there are very few cells with CPUE above zero. These are mostly concentrated north of Cape Ann. During fall months,

pelagic CPUE values are mainly concentrated over Stellwagen's southwest corner, over Jeffrey's ledge and north of Cape Ann.

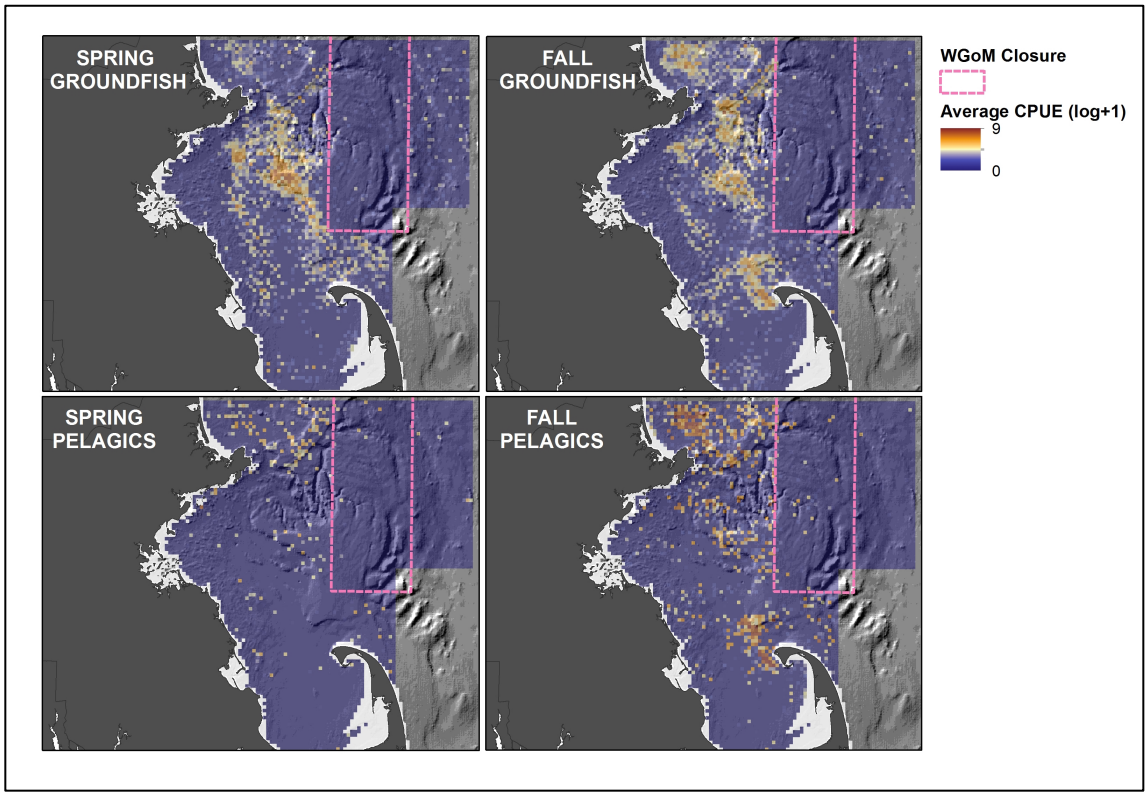


**Figure 23. – Average CPUE derived from VTR dataset: (top) Values for groundfish species, and (bottom) values for pelagic species.**

*EXPLORING SPATIAL CORRELATION BETWEEN VARIABLES*

To explore the spatial correlation between each dependent variable and ICAP values, I fitted experimental semi-variograms with the residuals from the regressions comparing each pair of variables. Figure 24 shows the resulting scatterplots. Plots that show an initial increase in variance as distance increases may be considered to show

spatial correlation between the variables. This is the case for both ground-fish abundances and fall pelagic abundance, but not for spring pelagic abundance. These results may be explained by the quality of VTR data for pelagic species. Due to fisheries regulations and seasonal closures, there are very few cells with abundance values above zero during spring months. The semi-variogram for fall pelagic abundance shows an initial increase in variance but it rapidly levels off. Therefore, in this case there may be spatial correlation between productivity and fish abundance at local scales, within a 10-kilometer radius, but not beyond this distance. I must note that the high y-intercept (i.e. nugget) in all variograms may signal a high error in the regression results.

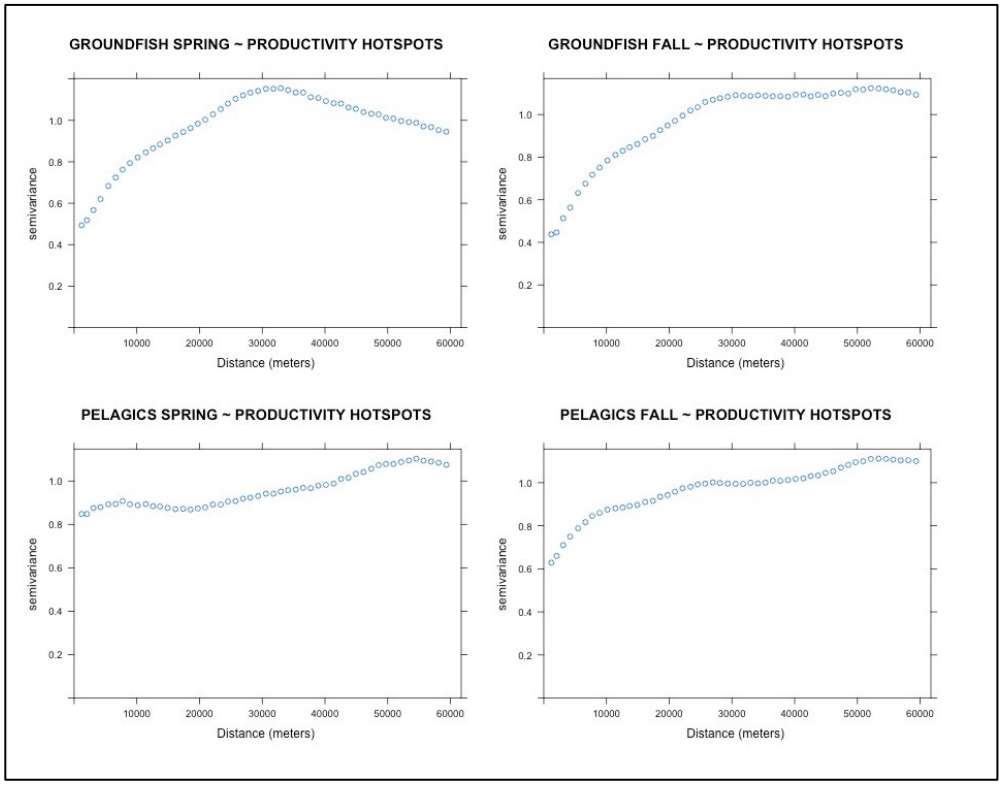


**Figure 24. – Semi-variograms comparing each dependent variable to ICAP values for each season.**

## *PREDICTING FISH ABUNDANCE FROM ABIOTIC AND PRODUCTIVITY INFORMATION*

Before fitting all predictive models, I determined the best combination and number of independent variables for each dependent variable. For this reason, I run multiple OLS regressions with all possible combinations of topography and SST values as independent variables, and fish abundance for each group as dependent variables. For each dependent variable, I chose the model run that had the smallest AIC value. Table 1 shows the resulting models. For example, in the case of spring ground-fish abundance, the best model included sea-surface temperature hotspots and coldspots, standard deviation of depth values, mean of slope values, minimum BPI value at fine scale, distance to coast and range of BPI values at a broad scale. This combination of variables signal that ground-fish abundance values may be correlated to big changes in topographic features. The variable "aspect" was not included in any of the resulting models. Also, due to the small number of spring pelagic abundance values above zero, I chose to not include this group in further tests.

**Table 1 – Best fit models from exploratory OLS regressions. The dependent variables for these regressions are: (Top to bottom) abundance of groundfish for spring and fall, and abundance of pelagics for spring and fall. For each dependent variable, this table only shows the model run that had the lowest AIC value.**

| | GROUNDFISH - SPRING | | |
|---|---|---|---|
| | AIC | 19921.9 | |
| | ADJ. $R^2$ | 0.114374 | |
| | STANDARD ERROR | 1.39575 | |
| | F- STATISTIC | 105.81 | *** |
| | MORAN'S I (Resid.) | 65.5514 | *** |
| INDEPENDENT VARIABLES | SST HOTSPOT | 0.2801 | *** |
| | DEPTH (St. Dev.) | 0.4867 | *** |
| | SLOPE (Mean) | -0.7165 | *** |
| | BPI Fine (Min.) | -0.3198 | *** |
| | SST COLDSPOT | -0.1592 | *** |
| | DISTANCE COAST | -0.1065 | *** |
| | BPI Broad (Range) | 0.1722 | *** |

| | GROUNDFISH - FALL | | |
|---|---|---|---|
| | AIC | 20288 | |
| | ADJ. $R^2$ | 0.13924 | |
| | STANDARD ERROR | 1.44144 | |
| | F- STATISTIC | 132.283 | *** |
| | MORAN'S I (Resid.) | 73.6952 | *** |
| | SST COLDSPOT | -0.0258 | *** |
| | BPI Broad (Min.) | -0.1757 | *** |
| | SST HOTSPOT | 0.2586 | *** |
| | SLOPE (Max.) | 0.3753 | *** |
| | BPI Fine (Range) | -0.3896 | *** |
| | DEPTH (St. Dev.) | 0.2171 | *** |
| | DISTANCE COAST | -0.1067 | *** |

| | PELAGICS - SPRING | | |
|---|---|---|---|
| | AIC | 13282.2 | |
| | ADJ. $R^2$ | 0.018334 | |
| | STANDARD ERROR | 0.778212 | |
| | F- STATISTIC | 17.6646 | *** |
| | MORAN'S I (Resid.) | 20.4231 | *** |
| INDEPENDENT VARIABLES | SST COLDSPOT | -0.0130 | *** |
| | DISTANCE COAST | -0.0044 | *** |
| | BPI Broad (Max.) | 0.0057 | *** |
| | BPI Fine (Max.) | -0.0086 | *** |
| | DEPTH (Max.) | 0.0030 | *** |
| | SLOPE (Max.) | 0.0040 | *** |

| | PELAGICS - FALL | | |
|---|---|---|---|
| | AIC | 21296.6 | |
| | ADJ. $R^2$ | 0.053045 | |
| | STANDARD ERROR | 1.57537 | |
| | F- STATISTIC | 52.9824 | *** |
| | MORAN'S I (Resid.) | 49.9016 | *** |
| | DISTANCE COAST | -0.0100 | *** |
| | DEPTH (St. Dev.) | 0.0226 | *** |
| | SLOPE (Mean) | -0.0397 | *** |
| | BPI Fine (St. Dev.) | 0.0209 | *** |
| | BPI Broad (Mean) | 0.0069 | *** |
| | SST COLDSPOT | -0.0128 | *** |

| SIGNIFICANCE LEVELS: | *** | < 0.01 | ** | < 0.05 | * | < 0.1 |
|---|---|---|---|---|---|---|

I run OLS, spatial lag, and GWR regression models for each of the "winning" combinations of independent variables. Table 2 shows the resulting AIC values in each case. In all cases, AIC values for both GWR and spatial lag models were smaller than the corresponding ones for OLS regressions, indicating a better model fit. Because GWR fits a regression line to each data point and its neighbors, I considered GWR results to represent the "local" relationship between variables. In fact, in this study, the optimal bandwidth selected for all GWR models was 3600m. On the other hand, because Spatial lag models only fit one regression line for the whole area of study, its results may be considered a representation of the "global" relationship between variables.

Table 2 also shows resulting AIC values when I added ICAP values as an independent variable to each model run. In the case of spring ground-fish abundance, adding ICAP values significantly improves the fit for both spatial lag and GWR models. On the other hand, for fall ground-fish abundance, the AIC values for spatial lag models do not change when adding productivity information, but they decrease for GWR model fits. Also, the t-values for the ICAP coefficients in this case are not significant. These results signal the lack of a significant "global" relationship between fall ground-fish abundance and surface productivity. For pelagic fall abundance values, productivity values significantly improve all three model runs. Finally, Figure 25 shows predicted spring ground-fish abundance values from both Spatial lag and GWR model runs. Both model runs do a good job at predicting values at the northwestern edge of the Stellwagen Bank. However, because of the presence of the West Gulf of Maine closure, the spatial lag model does not correctly predict values on the eastern side of the bank. It is possible

that adding the variable "distance to closure" as an independent variable may improve the overall spatial lag model fit.

**Table 2 – Comparison of AIC values from different model runs (OLS, Spatial Lag and GWR) before and after ICAP values are added as independent variables.**

| | | OLS | | SPATIAL LAG | | GWR | |
|---|---|---|---|---|---|---|---|
| | | NO CHL | WITH CHL | NO CHL | WITH CHL | NO CHL | WITH CHL |
| GROUNDFISH | SPRING | 19921.90 | 19770.70 | 17506.30 | 17490.00 | 12325.09 | 9275.77 |
| | FALL | 20288.00 | 20237.00 | 17396.20 | 17396.20 | 10527.95 | 9963.36 |
| PELAGICS | FALL | 21296.60 | 21287.60 | 19890.10 | 19888.70 | 11387.37 | 9495.65 |

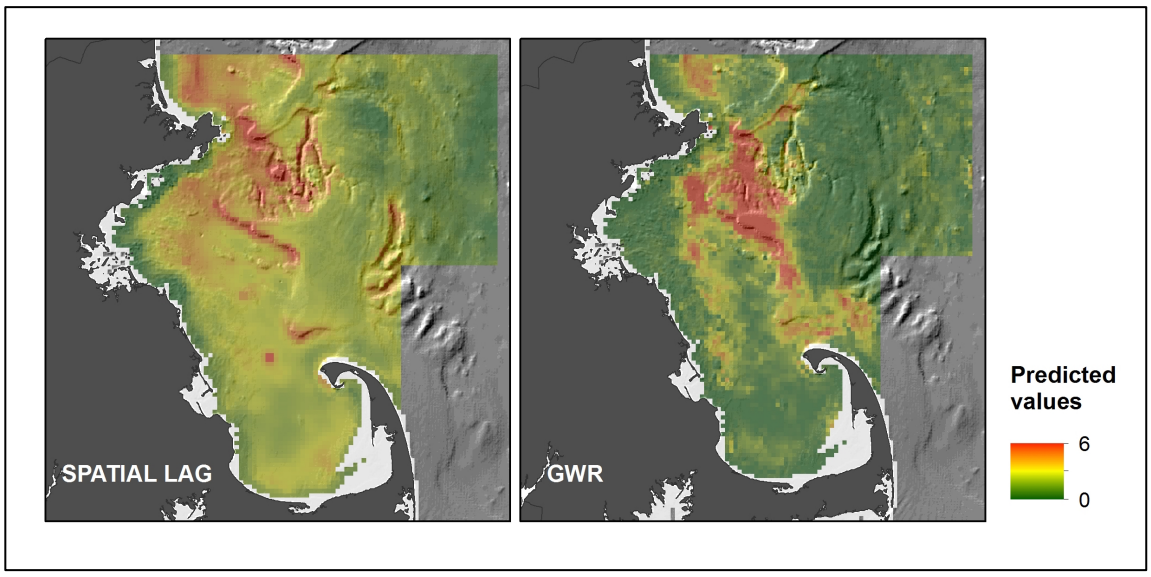


**Figure 25. – Comparison of predicted spring bottom fish abundance between Spatial lag (left) regression and GWR (right) regressions.**

# CONCLUSIONS

The main objective of this study was to determine whether adding primary productivity hotspot information to a model improves the spatial prediction of both pelagic and benthic fish abundance. Through this process, I tested novel metrics for both primary productivity and SST hotspots, which improved the comparison between the different sources of data. These metrics are based on the tendency certain areas have to show high or low valued anomalies over time. They do not provide values of productivity or temperature at a moment in time, but they identify areas that provide the right conditions to either stimulate productivity or concentrate the one produced elsewhere. This new way of defining biological hotspots allows the comparison between datasets from different trophic levels and in turn provides the basis for ecosystem wide studies. I also run three different types of regression models to allow for spatial dependence between the variables tested.

Results from this analysis show a seasonal variability in the relationship between productivity hotspots and abundance of species from other trophic levels. When focusing on benthic fish, adding productivity information does improve the fit of the model during spring months. However, during fall months, the "global" model fit does not change when adding productivity. One way to explain these results is through the influence topography has on both surface and bottom productivity. Spring phytoplankton blooms occur when increased water temperatures and sunlight create a thermocline that trap nutrients near the surface of the ocean, thus providing the optimal conditions for phytoplankton growth. As a result, surface productivity hotspots during the spring are

often found in areas with big topographic features, since upwelling events enhance the mixing of the water column, and increase the flow of nutrients and higher temperatures to the surface. At the same time, topographic features also create favorable conditions for many ground-fish species to grow, providing protection and sources of food. As a result, areas with significant topographic features do have both high surface productivity and high benthic abundance.

A second explanation for this significant positive correlation may be the process of "bentho-pelagic coupling". Surface productivity that is not consumed during the spring bloom may be deposited on the seafloor, enhancing secondary productivity (Townsend and Cammen 1988). Even though the fact that only Spring results are significant does not support this hypothesis, I need more information to fully test it. Surface productivity values need to be spatially and temporally compared to productivity values across the water column. Currently, very few datasets provide the resolution and spatial extent needed for this type of analysis. One dataset that may be able to fit these requirements is the one collected by HabCam. The HabCam (Gallager et al. 2005) is a non-motorized vehicle operated by the Woods Hole Oceanographic Institute (WHOI) that is towed from a ship. While HabCam was build to take pictures of the ocean seafloor for habitat characterization, it also includes equipment that measures sea surface temperature, salinity, dissolved oxygen and chlorophyll-a concentration. In the following months, I am planning to use this dataset to test the hypothesis that surface productivity increases overall benthic productivity.

Prediction of fall pelagic species does also improve when adding productivity. While most pelagic species do not feed on phytoplankton, they eat organisms (e.g. zooplankton) that do. Therefore, areas that tend to have higher concentrations of productivity throughout the seasons may attract forage fish species. Our results for spring pelagic abundance were limited by the availability of data. The reduced availability of data for spring months compared to fall is mainly due to fisheries regulations in the area, including seasonal closures of pelagic fisheries and limits on days at sea. In general, using fisheries-dependent datasets for this type of analysis presents a series of limitations. First, sampling is not done randomly across the area, so the same effort was not applied in every single location. Fishermen decide where to fish based not only on resources but also on other socio-economic factors, such as previous knowledge of the area, cost of fuel, presence of protected areas, and other fisheries regulations. Also, because of temporary closures of fisheries due to low stocks, the sampling was not constant throughout the twelve years of analysis. Another big limitation of VTR data comes from its self-reporting nature. Previous studies have pointed out the lack of accuracy when it comes to discarded fish abundance for all gear types. It is possible to use LPUE (Landing per Unit Effort) instead, which only includes the individuals that are kept and landed. However, I decided to include both kept and discarded individuals to account for species that are not commercially valuable (e.g. sand-lance). Also, the fact that all yearly values are averaged into a final result does reduce the uncertainty of the final abundance values.

This study explored the spatial relationship between productivity aggregations and aggregations of other organisms. The results in this analysis support the concept of

biological hotspots. Because of high abundance of organisms, these biological hotspots often spur increased human activity, particularly if they contain commercially valuable species. As a result, they are at greater risk of negative impacts due to overfishing. Any resulting losses of biomass in these areas could then impair adjacent populations of marine organisms that depend on them, affecting in turn other levels of the ecosystem. Therefore, prioritizing available resources for the protection of these biological hotspots may be warranted. Interdisciplinary studies such the one presented here may provide managers with the information and tools that would help them determine the level of human activity permitted in these areas to ensure that they are healthy and sustainable in the future.

.

# CONCLUDING REMARKS

The main objective of this thesis was to identify and characterize biological hotspots in the Gulf of Maine. In the first chapter, I described a simplified methodological framework to guide users through the process of locating these local-scale hotspots. This framework described the challenges when locating hotspots in marine ecosystems compared to terrestrial ones, and proposed new solutions researchers may want to consider for future projects. In the second chapter, I defined a new metric, the Index of Cumulative Anomalous Productivity (ICAP), which identified hotspots based on their predictability over time and the magnitude of their productivity anomalies. Using time series analysis, I also described the seasonal, inter-annual and long-term variability of these hotspots. Finally, in the third chapter of this thesis, I compared these ICAP values to both abiotic variables (e.g. sea surface temperature) and fish abundance information. This ecosystem-wide study tested the spatial correlation between aggregations of primary producers and fish from both benthic and pelagic areas of the ocean. Results showed that certain areas of the Gulf of Maine facilitate aggregations of primary producers and both benthic and pelagic fishes, creating biological hotspots, though these organisms do not have a direct trophic link.

This thesis relied extensively on satellite remote sensing data, mainly as a proxy for both primary productivity abundance and sea surface temperature. Satellite datasets provided a wide spatial coverage, fine spatial resolution, and high temporal repetition, which benefited the study of biological hotspots. However, this type of data source also

has limitations. High concentrations of clouds limit the amount of information available. Data may also be lost because of high water column turbidity, which frequently occurs in coastal locations. While the lack of data because of clouds cannot be controlled, incorrect data values due to high turbidity may be prevented in the future by using more advanced algorithms to process remote sensing data. These algorithms are specifically created for Case-2 waters. Also, because the European Space Agency lost contact with MERIS in 2012, we were only able to analyze 10 years worth of data. Over the past few years, many researchers have started to use the Moderate Resolution Imaging Spectroradiometer (MODIS) operated by the National Aeronautics and Space Administration to derive productivity estimates. Future studies in a variety of fields would benefit from an increased effort to build and deploy sensors specially made for deriving ocean water properties, which would increase the availability and accuracy of data used for studies such as the thesis presented here.

This thesis was also limited by the availability of fisheries data. The VTR dataset relies on fishermen collecting accurate data on number of fish, species caught, and locations where fishing took place. Moreover, fishing effort is not homogeneous across the study area, as it depends on fisheries regulations and the overall behavior of fishermen. This adds uncertainty to the results of this study. Federal agencies, such as the National Oceanographic and Atmospheric Administration (NOAA), have been collecting fisheries independent information for the past few years, which would directly benefit studies such as the ones presented here. However, currently these data sources are only available to federal employees. This study was also limited by the availability of

validation data. The MWRA dataset is currently the only one that offers the minimum temporal resolution necessary to compare field and remote sensing hotspot values in the Gulf of Maine. For interdisciplinary studies such the ones presented in this thesis, there is a need for coordinated field collection efforts that acquire information from different levels of a marine ecosystem and that are designed following the spatial and temporal requirements of hotspot analyses.

The results from this thesis will be added to a decision support tool to facilitate the management of marine resources in the Gulf of Maine. This tool, the Marine Integrated Decision Analysis System (MIDAS), was created by a team of Boston University researchers lead by Prof. Suchi Gopal and Prof. Les Kaufman. MIDAS was first designed to manage marine protected areas in Belize and was subsequently adapted for Gulf of Maine waters (Gopal et al. 2015). This interactive decision support tool was developed to improve the effectiveness of marine spatial planning in the Gulf of Maine by translating complicated model results into simple visualizations, allowing stakeholders to run scenarios based on their own parameters. The results from this thesis will allow MIDAS to adapt its simulation model to include spatial and temporal heterogeneity in phytoplankton populations, which in turn may ensure that the model more closely represents the real dynamics of marine systems in the area. The results of this thesis should also serve as the foundation for researchers to identify and test other hypothesis on the processes driving spatial heterogeneity in coastal systems, which in turn may help managers apply ecosystem–based principles to their managements plans.

# BIBLIOGRAPHY

----. 2003. America's Living Oceans: Charting a Course for Sea Change. Pew Oceans Commission, Arlington, Virginia.

----. 2009. The State of World Fisheries and Aquaculture 2008. Food and Agriculture Organization of the United Nations (FAO).

----. 2010. Stellwagen Bank National Marine Sanctuary Final Management Plan and Environmental Assessment. U.S. Department of Commerce. National Oceanic and Atmospheric Administration. Office of National Marine Sanctuaries. Silver Spring, MD.

Aldstadt, J., Getis, A. 2006. Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis* 38: 327-343.

Anselin, L. 1989. What is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis. Institute of Mathematical Geography, Ann Arbor, ML.

Anselin, L., Syabri, I., Kho, Y. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38:5–22.

Arkema, K.K., Abramson, S.C., and Dewsbury, B.M. 2006. Marine ecosystem-based management: from characterization to implementation. *Frontiers in Ecology and the Environment* 4(10): 525–532.

Auster, P. J., et al. 1996. The impacts of mobile fishing gear on seafloor habitats in the gulf of Maine (Northwest Atlantic): Implications for conservation of fish populations. *Reviews in Fisheries Science* 4(2): 185–202.

Auster, P.J. 1998. A conceptual model of the impacts of fishing gear on the integrity of fish habitats. *Conservation Biology*, 12(6), pp.1198–1203.

Barrell, J., Grant, J. 2013. Detecting hot and cold spots in a seagrass landscape using local indicators of spatial association. *Landscape Ecology* 28: 2005–2018.

Belsley, D. A., Kuh, E., Welsch, R. E. 2005. Regression diagnostics: Identifying influential data and sources of collinearity (Vol. 571). John Wiley & Sons.

Behrenfeld, M. J., et al. 2006. Climate-driven trends in contemporary ocean productivity. *Nature* 444:752–755.

Bigelow, H. B., et al. 1940. Phytoplankton and planktonic protozoa of the offshore waters of the Gulf of Maine. *Transactions of the American Philosophical Society* 31(3):

149–237.

Botsford, L.W., Castilla, J.C. and Peterson, C.H. 1997. The management of fisheries and marine ecosystems. *Science*, 277, pp.509–515.

Boyce, D.G., Lewis, M.R., and Vetter, B. 2010. Global phytoplankton decline over the past century. *Nature*, 466(7306), pp.591–596.

Bricaud, A., Morel, A., and Barale, V. 1999. MERIS potential for ocean colour studies in the open ocean. *International Journal of Remote Sensing* 20 (9): 1757-1769

Brunsdon, C., Fotheringham, A. S., Charlton, M. E. 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 28: 281–298.

Calbet, A., and Landry, M.R. 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnology and Oceanography*, 49(1), pp.51–57.

Chandler, M., Kaufman, L., and Mulsow, S. 1996. Human Impact, Biodiversity and Ecosystem Processes in the Open Ocean. In *Functional Roles of Biodiversity: A Global Perpective*, H. A. Mooney et al., eds. John Wiley & Sons Ltd., pp. 431–474.

Christensen, N.L., et al. 1996. The Report of the Ecological Society of America Committee on the Scientific Basis for Ecosystem Management. *Frontiers in Ecology and the Environment* 6 (3): 665-691.

Christensen, V. et al. 2003. Hundred-year decline of North Atlantic predatory fishes. *Fish and Fisheries*, 4(1), pp.1–24.

Cibik, S.J. et al. 1998. Massachusetts Bay plankton communities: characterization and discussion of issues relative to MWRA's outfall relocation. *Massachusetts Water Resource Authority, Boston, MA. Report ENQUAD 98-08.* pp.1–145.

Cleveland, W.S. 1981. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician* 35(1): 54.

Cochrane, K.L. 2000. Reconciling sustainability, economic efficiency and equity in fisheries: the one that got away? *Fish and Fisheries*, 1(1), pp.3–21

Costanza, R. et al. 1998. Principles for Sustainable Governance of the Oceans. *Science*, 281(5374), pp.198–199.

Crowder, L.B., et al. 2006. Resolving mismatches in US ocean governance. *Science* 313: 617–618.

Crowder, L., Norse, E. 2008. Essential ecological insights for marine ecosystem-based management and marine spatial planning. *Marine Policy* 32: 772–778.

Curran, P.J. 1988. The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment* 24: 493–507.

Dall'Olmo, G., et al. 2005. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sensing of Environment* 96: 176-187

Daly, K. L., Smith Jr, W. O. 1993. Physical-biological interactions influencing marine plankton production. Annual Review of Ecology and Systematics 24:555–585.

Davoren, G. K. 2007. Effects of Gill-Net Fishing on Marine Birds in a Biological Hotspot in the Northwest Atlantic. *Conservation Biology* 21(4): 1032–1045.

Dayton, P.K., Thrush, S., and Coleman, F.C. 2002. Ecological Effects of Fishing in Marine Ecosystems of the United States. *Pew Oceans Comission*.

Doerffer, R., Sorensen, K., Aiken, J. 1999. MERIS potential for coastal zone applications. *International Journal of Remote Sensing* 20(9): 1809–1818.

Dormann, C.F., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609–628.

Durant, J. M., et al. 2007. Climate and the match or mismatch between predator requirements and resource availability. *Climate Research* 33:271.

Etnoyer, P., et al. 2004. Persistent Pelagic Habitats in the Baja California to Bering Sea (B2B) Ecoregion. *Oceanography* 17(1): 90–101.

Eppley, R.W. 1972. Temperature and phytoplankton growth in the sea. *Fishery Bulletin*, 70(4), pp.1063–1085.

Fortin, M.J., Dale, M.R.T. 2005. Spatial analysis: a guide for ecologists. Second Edition. Cambridge University Press.

Getis, A., Boots, B. 1978. Models of spatial processes: an approach to the study of point, line and area patterns. Cambridge University Press.

Getis, A., Ord, J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24(3): 189–206.

Gallager, S. M., et al. 2005. High Resolution Underwater Imaging and Image Processing for Identifying Essential Fish Habitat. NOAA Technical Memorandum NMFS-F/SPO-68.

Gopal, S., et al. 2015. Modeling Coastal and Marine Environmental Risks in Belize: the Marine Integrated Decision Analysis System (MIDAS). *Coastal Management*. In press.

Gould, W. 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications* 10(6): 1861–1870.

Grémillet, D., et al. 2008. Spatial match-mismatch in the Benguela upwelling zone: should we expect chlorophyll and sea-surface temperature to predict marine predator distributions? *Journal of Applied Ecology* 45:610–621.

Griffin, P., et al. 1996. Regional patterns of affirmative action compliance costs. *The Annals of Regional Science* 30: 321–340.

Grober-Dunsmore, R. et al. 2008. Vertical Zoning in Marine Protected Areas: Ecological Considerations for Balancing Pelagic Fishing with Conservation of Benthic Communities. *Fisheries*, 33(12), pp.598–610.

Hooker, S. K., Gerber, L. R. 2004. Marine Reserves as a Tool for Ecosystem-Based Management: The Potential Importance of Megafauna. *BioScience* 54(1): 27–39.

Hyde, K.J.W., et al. 2007. Validation of SeaWiFS chlorophyll a in Massachusetts Bay. *Continental Shelf Research* 27: 1677–1691.

Hyrenbach, K.D., Forney, K.A., Dayton, P.K. 2000. Marine protected areas and ocean basin management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 10(6), pp.437–458.

Incze, L. et al. 2010. Connectivity of lobster (*Homarus americanus*) populations in the coastal Gulf of Maine. part ii: coupled biophysical dynamics. *Fisheries Oceanography*, 19(1), pp.1–20.

Ji, R. et al. 2007. Influence of ocean freshening on shelf phytoplankton dynamics. *Geophysical Research Letters*, 34.

Ji, R. et al. 2010. Marine plankton phenology and life history in a changing climate: current research and future directions. *Journal of Plankton Research*, 32(10), pp.1355–1368.

Jones, M. C., Cheung, W.W.L. 2014. Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES Journal of Marine Science, ICES Journal of Marine Science: Journal du Conseil, fsu172.*

Kahru, M., et al. 2012. Spatial and temporal statistics of sea surface temperature and chlorophyll fronts in the California Current. *Journal of Plankton Research* 34(9):

749–760.

LeDrew, E.F., et al. 2004. A spatial statistical operator applied to multidate satellite imagery for identification of coral reef stress. *Remote Sensing of Environment* 91: 271–279.

Lee, Z., et al. 1998. Hyperspectral Remote Sensing for Shallow Waters. I. A Semianalytical Model. *Applied Optics* 37 (27): 6329-6338.

Lee, Z., et al. 1999. Hyperspectral Remote Sensing for Shallow Waters. 2. Deriving Bottom Depths and Water Properties by Optimization. *Applied Optics* 38 (18): 3831-3843.

LeSage, J.P. 1999. Spatial econometrics. Morgantown, WV: Regional Research Institute, West Virginia University.

Levin, S.A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* 73(6): 1943–1967.

Link, H., et al. 2013. Are Hotspots Always Hotspots? The Relationship between Diversity, Resource and Ecosystem Functions in the Arctic. *PLoS ONE* 8 (9): e74077.

Lynch, D.R. et al. 1996. Comprehensive coastal circulation model with application to the Gulf of Maine. *Continental Shelf Research*, 16(7), pp.875–906.

Mackas, D.L., et al. 1985. Plankton Patchiness: Biology in the Physical Vernacular. *Bulletin of Marine Science* 37(2): 652–674.

Mackenzie, I.K. 2007. Modifying a Local Measure of Spatial Association to Account for Non-Stationary Spatial Processes (MS dissertation). Victoria, BC: University of Victoria.

Malakoff, D. 2004. Marine Science: New Tools Reveal Treasures at Ocean Hot Spots. *Science* 304: 1104–1105.

Mann, K., Lazier, J. 1996. Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans. Blackwell Science.

Maritorena, S., et al. 1994. Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. *Limnology and Oceanography* 39(7): 1689–1703.

McClain, C.R. 2009. A Decade of Satellite Ocean Color Observations. *Annual Review of Marine Science* 1: 19–42.

McGowan, J.A., et al. 1996. Plankton patterns, climate, and change in the California Current. *California Cooperative Oceanic Fisheries Investigations Reports* 37: 45–68.

McLeod, K., and Leslie, H. 2009. Ecosystem-based management for the oceans. Island Press, Washington, DC, USA/London, UK

Morel, A., Prieur, L. 1977. Analysis of variations in ocean color. *Limnology and Oceanography* 22(4): 709–722.

Murawski, S.A. 2007. Ten myths concerning ecosystem approaches to marine resource management. *Marine Policy*, 31(6), pp.681–690.

Myers, N., et al. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nakaya, T., et al. 2005. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine* 24:2695–2717.

Navarro, G. et al. 2011. Seasonal-to-interannual variability of chlorophyll- a bloom timing associated with physical forcing in the Gulf of Cádiz. *Advances in Space Research*, 50, pp.1164–1172.

Nelson, T.A., Boots, B. 2008. Detecting spatial hot spots in landscape ecology. *Ecography* 31: 556–566.

Nixon, S.W. et al. 2009. The impact of changing climate on phenology, productivity, and benthic–pelagic coupling in Narragansett Bay. *Estuarine, Coastal and Shelf Science*, 82(1), pp.1–18.

Norse, E.A., et al. 2005. Place-Based Ecosystem Management in the Open Ocean. In: Norse, E. A., and Crowder, L. B. (ed.), Marine Conservation Biology: the Science of Maintaining the Sea's Biodiversity. Marine Conservation Biology Institute.

Nur, N. et al. 2010. *Wildlife Hotspots in the California Current System*, Petaluma, CA: PRBO Conservation Science.

Nur, N., et al. 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecological Applications* 21(6): 2241–2257.

Ord, J.K., Getis, A. 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27(4): 286–306.

Ord, JK, Getis, A. 2001. Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation. *Journal of Regional Science* 41(3): 411–432.

Orphanides, C. D., and G. M. Magnusson. 2007. Characterization of the northeast and mid- Atlantic bottom and mid-water trawl fisheries based on vessel trip report (VTR) data. *Northeast Fisheries Science Center, NMFS, NOAA, US Department of*

*Commerce, Woods Hole, Massachusetts.*

O'Reilly, J.E., et al. 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research* 103(C11): 24937-24953.

Paerl, H. W., Huisman, J. 2008. Blooms like it hot. Science 320:57–58.

Pauly, D. et al. 1998. Fishing down marine food webs. *Science*, 279, pp.860–863.

Pikitch, E.K. et al. 2004. Ecosystem-Based Fishery Management. *Science*, 305(5682), pp.346–347.

Reese, D. C., Brodeur, R. D. 2006. Identifying and characterizing biological hotspots in the northern California Current. *Deep Sea Research Part II: Topical Studies in Oceanography* 53:291–314.

Rogerson, P.A. 2002. Change detection thresholds for remotely sensed images. *Journal of Geographical Systems* 4: 85–97.

Sarmiento, J.L. et al. 2004. Response of ocean ecosystems to climate warming. *Global Biogeochemical Cycles*, 18(3), GB303.

Schroeder, T., et al. 2007. Atmospheric correction algorithm for MERIS above case-2 waters. *International Journal of Remote Sensing* 28(7): 1469–1486.

Schweiger, O., et al. 2008. Climate change can cause spatial mismatch of trophically interacting species. *Ecology* 89(12), 3472-3479.

Scotti, A., Pineda, J. 2004. Observation of very large and steep internal waves of elevation near the Massachusetts coast. Geophysical Research Letters 31(22)

Seto, K.C., et al. 2004. Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing* 25(20): 4309–4324.

Shekhar, S., et al. 2011. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 193-214.

Siegel, D.A., et al. 2002. The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. *Science* 296: 730–733.

Sokal, R.R., Oden, N.L., Thomson, B.A. 1998. Local spatial autocorrelation in biological variables. *Biological Journal of the Linnean Society* 65: 41–62.

Song, H., et al. 2010. Phenology of phytoplankton blooms in the Nova Scotian Shelf-Gulf

of Maine region: remote sensing and modeling analysis. *Journal of Plankton Research* 32(11): 1485–1499.

Stavn, R. H. 1971. The horizontal-vertical distribution hypothesis: Langmuir circulations and Daphnia distributions. Limnology and Oceanography 16(2): 453–466.

Steele, J. H., et al. 2007. Balancing end-to-end budgets of the Georges Bank ecosystem. *Progress in Oceanography* 74:423–448.

Steneck, R.S., and Wilson, J.A. 2010. A fisheries play in an ecosystem theater: challenges of managing ecological and social drivers of marine fisheries at multiple spatial scales. *Bulletin of Marine Science*, 86(2), pp.387–411.

Stevick, P.T., et al. 2008. Trophic relationships and oceanography on and around a small offshore bank. *Marine Ecology - Progress Series* 363: 15–28.

Suryan, R. M., et al. 2012. New approach for using remotely sensed chlorophyll a to identify seabird hotspots. *Marine Ecology - Progress Series* 451: 213–225.

Thomas, A. C., et al. 2003. Satellite-measured phytoplankton variability in the Gulf of Maine. *Continental Shelf Research* 23: 971–989.

Thompson, C. 2010. The Gulf of Maine in Context. Gulf of Maine Council on the Marine Environment.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B* (Methodological) 58(1): 267–288.

Tittensor, D. P., et al. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature* 466: 1098–1101.

Tobler, W.R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46: 234–240.

Townsend, D.W., Cammen, L.M. 1988. Potential importance of the timing of spring plankton blooms to benthic-pelagic coupling and recruitment of juvenile demersal fishes. *Biol. Oceanogr*, 5(3), pp.215–228.

Turner, W., et al. 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution* 18(6): 306–314.

Ueyama, R., Monger, B.C. 2005. Wind-induced modulation of seasonal phytoplankton blooms in the North Atlantic derived from satellite observations. *Limnology and Oceanography*, pp.1820–1829.

Valentine, P. C., et al. 2000. Sun-illuminated sea floor topographic map and perspective

view imagery of Quadrangles 1-18, Stellwagen Bank National Marine Sanctuary off Boston, Massachusetts. U.S. Geological Survey Open-File Report 99-363.

Wei, C.-L., et al. 2010. Global patterns and predictions of seafloor biomass using random forests. *PLoS ONE* 5:1–12.

Werme, C., et al. 2012. 2011 Outfall Monitoring Overview. Boston: Massachusetts Water Resources Authority. Report 2012-11.

Winder, M., Cloern, J.E. 2010. The annual cycles of phytoplankton biomass. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 3215–3226.

Windle, M.J.S, et al. 2010. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science* 67: 145–154.

Winiarski, K. J., et al. 2013. Spatially explicit model of wintering common loons: conservation implications. *Marine Ecology - Progress Series* 492: 273–283.

Wolanski, E., Hamner, W.M. 1988. Topographically Controlled Fronts in the Ocean and Their Biological Influence. *Science* 241: 177–181.

Woodcock, C.E., Strahler, A.H. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21: 311–332.

Worm, B., et al. 2003. Predator diversity hotspots in the blue ocean. *Proceedings of the National Academy of Science* 100(17): 9884–9888.

Worm, B. et al. 2006. Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science*, 314(5800), pp.787–790.

Wright, D. J., et al. 2005. ArcGIS Benthic Terrain Modeler [a collection of tools used with bathymetric data sets to examine the deepwater benthic environment]. *Oregon State University, Davey Jones' Locker Seafloor Mapping/Marine GIS Laboratory and NOAA Coastal Services Center.*

Wulder, M., Boots, B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing* 19(11): 2223–2231.

.

# CURRICULUM VITAE

**MARTA M RIBERA**

675 Commonwealth Ave (#130), Boston, MA
(352) 256-6321 | martamribera@gmail.com

---

**EDUCATION**

**PHD IN GEOGRAPHY**
Boston University. Boston, MA. Proj. graduation: May 2015.

**B.A. IN BIOLOGY (Minor in Ecology)**
Universitat Autònoma de Barcelona. Barcelona, Spain

**FELLOWSHIPS AND AWARDS**

**EPA STAR FELLOWSHIP**
Awarded from 2011 to 2014

**OUTSTANDING TEACHING AWARD**
Awarded for academic years 2009/2010 and 2010/2011

**PROFESSIONAL EXPERIENCE**

**RESEARCH ASSISTANT**
BOSTON UNIVERSITY, BOSTON, MA
September 2011 – Present

Research for doctoral thesis called "Characterizing local biological hotspots using remote sensing data". Spatial analysis of ocean remote sensing data and multibeam/backscatter data. Data sources included phytoplankton samples, fisheries dependent (VTR, VMS) and independent data (NOAA Bottom Trawl), and pictures of the ocean floor. Use of neural networks and fuzzy methods for data clustering. Design of a decision support tool.

**TEACHING FELLOW**
BOSTON UNIVERSITY, BOSTON, MA
September 2008 – August 2011

Teaching assistant for undergraduate introductory class: GE101: Natural Environment – The atmosphere. Four weekly labs of 30 students each. Design of new laboratory exercises and material. Supervision of new teaching assistants. Teaching full lecture when professor absent. Seminars on use of spatial data in marine ecosystems for Marine GIS and Introduction to GIS courses.

**GIS ANALYST / FISHERIES BIOLOGIST**
NATIONAL MARINE FISHERIES SERVICE.
PANAMA CITY, FL.
March 2004 – August 2008

GIS specialist. Multibeam mapping, spatial analysis and habitat classification of MPAs in SE coast of United States and Gulf of Mexico. Database design and maintenance. Map creation for various publications. IT helpdesk and Website design. Scientific crew in monitoring cruises on the Gulf of Mexico and South Atlantic MPAs. Juvenile Reef fish study in St. Andrews Bay.

**ADDITIONAL EXPERIENCE**

**GULF STATES MARINE FISHERIES COMMISSION. BOTTOM MAPPING GROUP.**
New Orleans, LA. 2005 – 2007

**RMA VOLUNTEER. St. ANDREWS BAY SEAGRASS MONITORING**
Panama City, FL. 2004 – 2007

**MARINE RESOURCES VOLUNTEER.**
**PARQUE NACIONAL ISLAS GALÁPAGOS.**
Santa Cruz, Ecuador. July – August 1999

**VOLUNTEER EDUCATOR AND SUPERVISOR.**
**YOUTH RECREATION CENTER CEPITU.**
Barcelona, Spain. 1996 – 2002

**OTHER LANGUAGES**

Proficient in Spanish, Catalan.
Basic French.

**TECHNICAL SKILLS**

Proficient in ArcGIS products (including Spatial Analyst, 3D analyst and ModelBuilder), MATLAB, BEAM (ESA), SEADAS (NASA), Access and other Microsoft Office Products, Dreamweaver, and Photoshop. Working knowledge of HTML, R, SQL programming. Basic knowledge of Caris, ENVI, Visual Basic, BASH and Python programming.

**PUBLICATIONS**

Gopal, S., Kaufman, L., Pasquarella, V., **Ribera**, M., Holden, C., Shank, B. 2015. Modeling Coastal and Marine Environmental Risks in Belize: the Marine Integrated

Decision Analysis System (MIDAS). *Coastal Management*. In press.

Patel, H., Gopal, S., Kaufman, L., Carleton, M., Holden, C., Pasquarella, V., **Ribera**, M., Shank, B. 2010. MIDAS – A Spatial Decision Support System for monitoring Marine Management Areas. *International Regional Science Review*, 34(2), pp.191–214.

Carlson, J., **Ribera**, M., Conrath, C., Heupel, M. 2010. Habitat use and movement patterns of bull sharks Charcharhinus leucas determined using pop-up satellite archival tags**.** *Journal of Fish Biology*, vol. 77 (3).

Harter, S., **Ribera**, M., Shephard, A., Reed, J. 2009. Assessment of fish populations and habitat on Oculina Bank: examination of a deep-sea coral MPA off eastern Florida. *Fishery Bulletin* 107(2): 195-206.

Cortés, E., Arocha, F., Beerkircher, L., Carvalho, F., Domingo, A., Heupel, M., Holtzhausen, H., Neves, M., **Ribera**, M., Simpfendorfer, C. 2009. Ecological Risk Assessment of pelagic sharks caught in Atlantic pelagic longline fisheries. *Aquatic Living Resources*, vol. 22 (1): 25-34

Harter, S., David, A., **Ribera**, M. 2008. Survey of coral and fish assemblages on Pulley Ridge, SW Florida. *NMFS Panama City Laboratory Contribution 08-10.*

Siegfried, K., **Ribera**, M., Hale, L., and Carlson, J. 2006. Expanded take estimates of snapper-grouper from the shark bottom longline fishery within proposed marine protected areas in the South Atlantic Fishery Management Council's Amendment 14 to the Snapper-Grouper Federal Management Plan. *NMFS Panama City Laboratory Contribution 06-16.*

Siegfried, K., **Ribera**, M., Hale, L., Carlson, J. 2006. Expanded take estimates of coastal shark from the shark bottom longline fishery within proposed Type II Marine Protected Areas in the South Atlantic Fishery Management Council's Amendment 14 to the Snapper-Grouper Fishery Management Plan. *NMFS Panama City Laboratory Contribution 06-20.*

## PRESENTATIONS

**Ribera**, M., Gopal, S, Kaufman, L., Cowie-Haskell, B. 2014. Local productivity hotspots in the western Gulf of Maine: strength, persistence and correlation to fishing effort. *Regional Association for Research on the Gulf of Maine (RARGOM) Annual Conference.* Boston (MA), September 30, 2104.

**Ribera**, M. 2013. Connecting the dots between surface productivity and the rest of Stellwagen Bank: One step toward

ecosystem-based management of a coastal ecosystem. *RARGOM Annual Conference*. Portsmouth (NH), October 8, 2013.

**Ribera**, M. 2012. Spatial characteristics of bentho-pelagic coupling in Stellwagen Bank (Gulf of Maine). Poster. *American Fisheries Society (AFS) Annual Conference*. Minneapolis (MN), August 19-23 2012.

Boumans, R., M. **Ribera**. 2010. The Multi-scale Integrated Model of Ecosystem Services (MIMES) and its application in Massachusetts. *Decision Support Tool Developer Workshop*. Palo Alto, CA, October 4-5 2010.

**Ribera**, M. 2010. MIDAS: A Spatial Decision Support System for Monitoring Marine Management Areas. *American Association of Geographers Annual Conference*. Washington, DC, April 14-18 2010

Cortés, E., M. Heupel, M. **Ribera** and C. Simpfendorfer. 2008. Productivity and Susceptibility Analysis (Ecological Risk Assessment) of Atlantic sharks. *88th Annual Meeting of the American Society of Ichthyologists and Herpetologists (ASIH), 24th Annual Meeting of the American Elasmobranch Society (AES)*, Montreal, Canada, July 23-28 2008.

**Ribera**, M. 2008. Combining fisheries and geological data using GIS: the West Florida Reserves. *AFS Annual Conference*, Florida Chapter.

**Ribera**, M. 2006. Efficacy testing of MPAs: the West Florida Shelf and the South Atlantic Bight. Poster. *GIS conference: GIS and Ocean mapping in support of Fisheries Research and management*. MIT, Cambridge, MA.

## ACADEMIC SERVICE

Organizer of Graduate Student Seminars (2009 – 2013)
Member of the American Fisheries Society (AFS) and the American Geophysical Union (AGU)