2008-01-01

The Use of Satellite-Based Ocean Color Measurements for Detecting the Florida Red Tide (Karenia brevis)

Gustavo de Araujo Carvalho

University of Miami, gcarvalho@rsmas.miami.edu

Follow this and additional works at: http://scholarlyrepository.miami.edu/oa_theses

Recommended Citation


http://scholarlyrepository.miami.edu/oa_theses/116

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Theses by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
UNIVERSITY OF MIAMI

THE USE OF SATELLITE-BASED OCEAN COLOR MEASUREMENTS FOR DETECTING THE FLORIDA RED TIDE (*Karenia brevis*)

By

Gustavo de Araújo Carvalho

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

May 2008
A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

THE USE OF SATELLITE-BASED OCEAN COLOR MEASUREMENTS FOR DETECTING THE FLORIDA RED TIDE (*Karenia brevis*)

Gustavo de Araújo Carvalho

Approved:

Dr. Peter J. Minnett  
Professor of Meteorology and Physical Oceanography  
UM/RSMAS/MPO

Dr. Terri A. Scandura  
Dean of the Graduate School

Dr. Robert H. Evans  
Research Professor of Meteorology and Physical Oceanography  
UM/RSMAS/MPO

Dr. Edwards J. Kearns  
Physical Oceanographer  
National Park Service  
US Dept. of the Interior

Dr. Viva Banzon  
Assistant Scientist  
UM/RSMAS/MPO

Dr. Kendall L. Carder  
Professor of Physical Oceanography - USF  
College of Marine Science
As human populations increase along coastal watersheds, the understanding and monitoring of Harmful Algal Blooms (or red tides) is an increasingly important issue. A consistent method for accurately detecting red tides using satellite measurements would bring tremendous societal benefits to resource managers, the scientific community and to the public as well. In the West Florida Shelf, blooms of the toxic dinoflagellate *Karenia brevis* are responsible for massive red tides causing fish kills, massive die-offs of marine mammals, shellfish poisoning, and acute respiratory irritation in humans. In this work, for the first time a long-term dataset (2002~2006) the MODerate Resolution Imaging Spectroradiometer (MODIS) is compared (*i.e.*, matched-up) to an extensive data set of *in situ* cell counts of *K. brevis*; provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute. The pairing of remote sensing data with near-coincident field measurements of cell abundance was successfully used to derive the basis for the development of an alternative ocean color based algorithm for detecting the optical signatures associated with blooms of *K. brevis* in waters of the West coast of Florida. Conclusions are geographically limited to the Central West Florida Shelf during the boreal Summer-Fall (*i.e.*, the *K. brevis* blooming season). The new simpler
Empirical approach is compared with other two more complicated published techniques. Their potential is verified and uncertainties involved in the identification of blooms of *K. brevis* are presented. The results shown here indicate that the operational NOAA method for detecting red tides in the Gulf of Mexico (Stumpf et al., 2003; Tomlinson et al., 2004) performs less accurately than the other two algorithms at identifying *K. brevis* blooms. The sensitivity and specificity of the Bio-optical (Cannizzaro, 2004; Cannizzaro et al., 2008) and Empirical algorithms are simultaneously maximized with an optimization procedure. The combined use of these two optimized algorithms in sequence provides another new monitoring tool with improved accuracy at detecting *K. brevis* blooms. The ability of this Hybrid scheme ranges about 80% for both sensitivity and specificity; and the capability at predicting a correct red tides is 70%, and ~85% for non-blooms conditions. The spatial and temporal knowledge of *K. brevis* blooms can improve the direction of field monitoring to areas that should receive special attention, allowing better understanding of the red tide phenomenon by the scientific community. The relevant agencies can also develop more appropriate mitigation action plans, and public health guidance can be improved with the enhancement of sustainable coastal management strategies.
ACKNOWLEDGMENT

I am in great debt with Dr. Peter J. Minnett for extraordinarily guiding me towards a better understanding of all related matters involved with my professional life. This work only accomplished its goals because of his supportive personality. Special thanks go to Viva Banzon and Warner Baringer for generously donating time assisting me. Ed Kearns receives my gratitude not only for the scientific participation but also for his supportive words. I am grateful to Bob Evans for all of his satellite remote sensing assistance. Dr. Ken Carder, Jennifer Cannizzaro and Chuanmin Hu deserve my sincere appreciation for opening their doors for me at the very beginning.

I would like to show appreciation to Cindy Heil for generously allowing us to use the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute *in situ* database which made this work possible. The Center for Ocean and Human Health of the University of Miami sponsored this investigation and I want to thank Dr. Lora Fleming, Dr. Sharon Smith and Julie Hollenbeck. Two grants provided funding for this investigation: NSF OCE0432368 and NIEHS P50ES12736-01.

Adrienne Romanski and Lauren Zamora are particularly thanked for their fast response and constructive text corrections. With Guilherme Castelão and Rafael Schiller I have a huge dept for their patience in sharing with me their oceanographic knowledge. Without the help from these two friends I would not have made it to the end. Kay Kilpatrick, Arthur Mariano, Maria Villanueva, Ed Ryan, Angel Li, Sue Walsh, Tania Casal, and Larry Brand are among those who provided important insights.

I want to thank three extraordinary friends: Thiago Correa, Mike Lara and Gan Changlin. Thanks to the distant support received from Erika Merlim, and two other
positive energy and good soul people: Francisco de Souza and Raphaela BOMtempo. All my love is expressed to my entire family who besides cheering and praying day after day for my success on this journey, have been taking good care of my best friend: Ed Imundo Muleke Pulguento Babão, an amazing creature that used to illuminate my life with his strong autonomous personality and actions, behaving plainly as a silent human being in the shape of a dog... MUITO BEM ED. God and all other holy entities who brought to me these people and all other existing things in this world gain the very last words: Thank God!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Phytoplankton blooms</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Historical context</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Florida Red Tide</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>OBJECTIVES</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>REMOTE SENSING OF THE COLOR OF THE OCEAN</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>SATELLITE-BASED RED TIDE ALGORITHMS</td>
<td>14</td>
</tr>
<tr>
<td>4.1</td>
<td>Operational method</td>
<td>14</td>
</tr>
<tr>
<td>4.2</td>
<td>Bio-Optical technique</td>
<td>17</td>
</tr>
<tr>
<td>4.3</td>
<td>Empirical approach</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>DATASET</td>
<td>21</td>
</tr>
<tr>
<td>5.1</td>
<td>Satellite imagery</td>
<td>21</td>
</tr>
<tr>
<td>5.2</td>
<td>In situ dataset</td>
<td>23</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Chapter 1

Figure 1.1. Gulf of Mexico .................................................................  5

Figure 1.2. Historical database of the Florida Red Tide from 1957 to 2002. a) Non-bloom (green X) and red tide (red □) sampling location; b) The toxic dinoflagellate Karenia brevis (FWRI, 2002) .................................................................  8

Chapter 4

Figure 4.1. The relationship of the Operational method (Stumpf et al., 2003b; Tomlinson et al., 2004). Horizontal solid line represents a chlorophyll anomaly of 1 mg m$^{-3}$ and discriminates among flagged red tides (above the line; red shaded area) and non-blooms (below) .................................................................  16

Figure 4.2. The relationship of the Bio-optical technique (Cannizzaro, 2004; Cannizzaro et al., 2008). Vertical dashed line is 1.5 mg m$^{-3}$ of chlorophyll and the diagonal dashed line represents the calculated (Morel, 1988) values of particulate backscatter. Flagged red tides are in the lower right quadrant (red shaded area) .....................................................  18

Figure 4.3. The relationship of the Empirical approach. Dashed line represents the calculated (Morel, 1988) values of total scatter. Flagged red tides are under the line (red shaded area) .................................................................  20

Chapter 6

Figure 6.1. Satellite imaging process steps ........................................  28
Figure 6.2. Flowchart depicting the steps assessed during the *in situ* dataset analysis

Figure 6.3. Left (Light penetration mask): Samples were eliminated if below the depth range recoverable by satellite ocean color sensors, calculated with the effective penetration depth (EPD; see Appendix 5) minus 2 meters (in this example, only sample 1 was considered); Right (Bottom reflection threshold): Used to avoid bottom reflection it added 2 meters to the apparent optical depth (AOD; see Appendix 5), and only considered ETOPO2 bathymetry deeper than this value (in this example, only bottom A was considered).

Figure 6.4. Match-up analysis steps

Figure 6.5. Threshold lines used to optimize the *Bio-optical* technique (left) and *Empirical* approach (right). Dashed lines are the original threshold slope lines (Morel, 1988). Dotted lines are the thresholds slope lines (Morel, 1988) multiplied by factors of 0.33 (lower) and 3.0 (upper). On the left plot, the vertical dashed line represents 1.5 mg m\(^{-3}\) of chlorophyll. Flagged red tides are in the lower right quadrant (*Bio-optical* technique), and under the line (*Empirical* approach).

Chapter 7

Figure 7.1. Chlorophyll concentrations (mg m\(^{-3}\)) from SeaWiFS (left) and MODIS-Terra (right). Outlined area show the regions where the algorithms were applied (Figure 7.2)
Figure 7.2. Red tide envelopes (white) from the area outlined in Figure 7.1. Bio-optical technique (top); Empirical approach (bottom). SeaWiFS (August 30, 2001; left); MODIS-Terra (August 29, 2001; right). In situ data from August 25 to September 2, 2001 is overlaid. Non-bloom: green X; red tide: red □; respectively $\leq$ and $> 10^4$ cells l$^{-1}$ of *K. brevis* ................................................................. 54

Figure 7.3. The Bio-optical technique classifies everything to the right of the vertical dotted line (··· which is 1.5 mg m$^{-3}$ of Chl) and below the dashed line (--- which is Morel (1988) $b_{bp(550)}$) as red tides. In situ data from August 25 to September 2, 2001 are overlaid (non-bloom: X; red tide: □; respectively $\leq$ and $> 10^4$ cells l$^{-1}$ of *K. brevis*). SeaWiFS (August 30, 2001; left) and MODIS-Terra (August 29, 2001; right) ....................... 55

Figure 7.4. The Empirical approach classifies everything below the dotted line (··· which is Morel (1988) $b_{(550)}$) as red tides. In situ data from August 25 to September 2, 2001 are overlaid (non-bloom: X; red tide: □; respectively $\leq$ and $> 10^4$ cells l$^{-1}$ of *K. brevis*). SeaWiFS (August 30, 2001; left) and MODIS-Terra (August 29, 2001; right) ............... 55

Figure 7.5. Distribution of the original historical in situ sampling effort collected along the entire Florida coast showing increasing number of samples ........................................... 58

Figure 7.6. Temporal frequency distribution of the workable in situ dataset provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute that were within satellite scenes. Winter-Spring (blue; left); Summer-Fall (orange; right) ............................................................................................................. 62

Figure 7.7. Daily frequency distribution of the workable in situ dataset provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute that were within satellite scenes. Winter-Spring (blue; left); Summer-Fall (orange; right) ............................................................................................................. 63
Figure 7.8. Daily frequency distribution of the RED TIDE *in situ* match-ups in the Central west Florida Shelf. Winter-Spring (blue; left); Summer-Fall (orange; right) ............................................................................................................. 65

Figure 7.9. Daily frequency distribution of the NON-BLOOM *in situ* match-ups in the Central west Florida Shelf. Winter-Spring (blue; left); Summer-Fall (orange; right) ............................................................................................................. 66

Figure 7.10. Valid match-ups sampled year-round along the entire West Florida coast. *Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom). Non-bloom (left) and red tide (right) ........................................................................................................................................... 68

Figure 7.11. Valid match-ups from the entire West Florida coast in the Winter-Spring. *Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom). Non-bloom (left) and red tide (right) ........................................................................................................................................... 69

Figure 7.12. Valid match-ups from the entire West Florida coast in the Summer-Fall. *Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom). Non-bloom (left) and red tide (right) ........................................................................................................................................... 70

Figure 7.13. Valid match-ups from the Winter-Spring in the North of Florida. *Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom). Non-bloom (left) and red tide (right) ........................................................................................................................................... 72
Figure 7.14. Valid match-ups from the Winter-Spring in the South of Florida. Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom). Non-bloom (left) and red tide (right) .............................................................................................................. 73

Figure 7.15. Valid match-ups from the Summer-Fall in the North of Florida. Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom). Non-bloom (left) and red tide (right) .............................................................................................................. 74

Figure 7.16. Valid match-ups from the Summer-Fall in the South of Florida. Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom). Non-bloom (left) and red tide (right) ..... 75

Figure 7.17. Valid match-ups sampled during the Winter-Spring in the Central West Florida Shelf (CWFS). Non-bloom (left) and red tide (right). Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom) ................................................................. 77

Figure 7.18. Valid match-ups sampled during the Summer-Fall in the Central West Florida Shelf (CWFS). Non-bloom (left) and red tide (right). Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom) ................................................................. 78

Figure 7.19. Percentage of correct identification as a function of Morel (1988) threshold slope lines multiplied by factors from 0.33 to 3.0. Red tide (top); non-bloom (center); Optimized (bottom). Bio-optical: (left); Empirical: (right). Vertical dashed line is the factor 1.0 leading to the original Morel (1988) threshold slope line. Vertical dashed lines are the best-optimized factors Long horizontal dotted line is the expected 70% optimized percentage of correct identification ............................................................................ 86
Figure 7.20. Optimization results for the *Bio-optical* technique ................................ 87

Figure 7.21. Optimization results for the *Empirical* approach ..................................... 88
LIST OF TABLES

Chapter 6

Table 6.1. Geophysical parameter library ................................................................. 26
Table 6.2. Cell counts threshold ................................................................................. 31
Table 6.3. 2-by-2 statistical table ............................................................................. 41
Table 6.4. Specificity, sensitivity, false negatives and false positives ................. 43
Table 6.5. Predictive values ..................................................................................... 45
Table 6.6. Probability ............................................................................................... 46
Table 6.7. 1st Hybrid scheme (i.e., high confidence) ............................................. 50
Table 6.8. 2nd Hybrid scheme (i.e., high confidence) ........................................... 51

Chapter 7

Table 7.1. Detection of Karenia brevis blooms in the CWFS using different algorithms (i.e., the Bio-optical technique and Empirical approach) on one SeaWiFS image (August 30, 2001) and on one MODIS-Terra image (August 29, 2001). Denominators are valid match-ups; Numerator is the total population of samples; Percentages of occurrence are in between parenthesis. Red tide (in red), false negative (in gray), non-bloom (in green), false positive (in blue). ............................................................................................................. 56
Table 7.2. This table is divided in two blocks, but the column sequence of the blocks is a continuous. Yearly frequency distribution of the *in situ* sampling effort provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute and the correlated number of valid match-ups. Here, the limit of 15,000 cell l⁻¹ was used to divide into non-bloom and red tide samples ........................................... 61

Table 7.3. Number of days (images) with valid match-ups in the Central West Florida Shelf .............................................................................................................. 64

Table 7.4. Sensitivity, specificity, false negatives, false positives (Original algorithms) .............................................................................................................. 81

Table 7.5. Predictive values (Original algorithms) ................................................. 82

Table 7.6. Probabilities (Original algorithms) ............................................................ 84

Table 7.7. Sensitivity, specificity, false negatives, false positives (Optimized algorithms) .............................................................................................................. 89

Table 7.8. Predictive values (Optimized algorithms) ................................................. 90

Table 7.9. Probabilities (Optimized algorithms) .......................................................... 91

Table 7.10. The *Hybrid* scheme .............................................................................. 92

Table 7.11. Sensitivity, specificity, false negatives, false positives (*Hybrid* algorithm) .............................................................................................................. 93

Table 7.12. Predictive values (*Hybrid* algorithm) ..................................................... 94

Table 7.13. Probabilities (*Hybrid* algorithm) ........................................................... 95
CHAPTER 1

INTRODUCTION

This Chapter describes algal blooms, providing a historical overview, in particular of the Florida Red Tides. This introduction also addresses some aspects of the toxicity of toxic phytoplankton species. An outline of the use of satellite data for detecting red tides ends this chapter.

1.1. Phytoplankton blooms

Most plankton organisms exhibit a superb splendor when observed through a microscope. But besides their exquisite beauty, they are essential to the Earth’s ecosystem. Phytoplankton are tiny near-surface-dwelling ocean plants that represent the first trophic level of the aquatic food web. Despite their diminutive size, these microalgae play a significant role in converting solar energy to chemical energy via photosynthesis, thereby forming the basis of the aquatic food web. Moreover, it has been hypothesized that they also help regulate the temperature of the Earth's surface (i.e., the CLAW hypothesis; Charlson et al., 1987).

A considerable increase in the concentration of these minute, free-floating aquatic plants occurs when the growth rates exceed mortality, causing an “algal bloom.” When algal blooms grow too large above background levels, these marine phytoplankton blooms can produce a host of negative environmental impacts (Paerl, 1988; Smayda,
Overabundances of phytoplankton are able to cause noxious effects by shading other forms of life. When a bloom collapses, lack of oxygen (i.e., anoxia) may also result from heterotrophic degradation of algal biomass causing further deleterious consequences. In addition, a small portion of phytoplankton species is capable of producing toxic substances (numbering less than 300 of nearly four thousand existing species; (Smayda 1997a). These toxic phytoplankton have been documented as a nuisance in freshwater (Bittencourt-Oliveira, 2003), brackish water (Lehman et al., 2007), and seawater (Botes et al., 2003). Where present, fish can be endangered, shellfish can be poisoned, and negative health effects can impact humans and other warm-blooded animals (Van Dolah, 2000).

Blooms of toxin-producing phytoplankton are most commonly referred to as “Harmful Algal Blooms” (hereafter HABs), but are sometimes synonymously associated with terms such as “Red Tide,” “Algae Poisoning”, “Poisoned Water” or “Water Bloom”. In this analysis, the term “Red Tide” will also be used as a synonym of HABs. However, caution must be taken with regards to any red tide-related information since popular media and some among the scientific community (e.g., Dierssen et al., 2006) frequently use this term simply to describe any large dense algal bloom, including assemblages of non-toxic species. One should also note that red tides are not typically tide-driven, and that not all red tides are dense enough to cause the water to appear murky, reddish or ruddy. Indeed, water discoloration observed during red tides can range from yellow or gold, through mahogany to blue-greenish, to even a brown or black visual aspect (Carder and Steward, 1985; Stumpf et al., 2003a; Hu et al., 2005a). In general, the water hue mainly depends on the types of blooming phytoplankton and their concentration, on light-
absorbing pigments (i.e., pigmentation) and on which the presence of other water constituents are dominant (Steidinger and Haddad, 1981; Millie et al., 1995).

Since dense human populations tend to concentrate along coastal watersheds (Small and Nicholls, 2003), any factors relating the oceans to human health can affect a considerable number of people. The occurrence of HABs along a coastal area is without doubt an issue to be viewed with concern (Kusek, et al., 1999). Consequently, monitoring the water quality for HABs in coastal regions around the world’s oceans is a pressing need and a fundamental requirement for a sustainable coastal management (Ogden, 1997; Roberts, 1997; Jameson et al., 2002). Further information on the health effects of HABs, their spatial development and techniques for monitoring them are discussed in Appendices 1-4.

A consistent method for remotely detecting red tides would bring tremendous societal benefits. Not only would it provide reliable information about the location of toxic phytoplankton blooms, greatly improving public health guidance and enhancing environmental and state resource managers sustainable coastal management strategies, it would also allow a better understanding of their distribution by the scientific community, which could also make better use of resources to study red tides.

1.2. Historical context

The social and human impacts associated with red tides have been known throughout history, with several instances described by anecdotal narrative before modern records were taken (Moore, 1977; Mitchell, 2003). Some believe that one of the earliest potential documentation of red tides is the Old Testament of the Bible. The
account depicted on the First of the Ten Plagues of Egypt in the Second Book of Moses is considered a persuasive red tide argument: “The Nile River turned into blood, the fish all died causing a terrible stench, so that Egyptians could not drink water” (Exodus, 7:14-25). To the best of our knowledge, this metaphor comparing water to blood was most likely due to some kind of pollution, rather than actual blood. Some believe that red tides were a possible interpretation of such discoloration, although as of today there has not been well-documented information on red tides in this region (Landsberg, 2002).

Further evidences of HABs exist in different locations of the globe. Fossils found in Scandinavian waters, where extinct toxic dinoflagellate cysts that existed between 500-2000 years ago suggest possible past red tide events. In the Pacific Northwest coast, Native American Indian tribes have long connected mussel poisoning with hot weather or with algae bioluminescence, probably produced by red tides (Fogg, 2002). In addition, historic seasonal mollusk harvesting controls employed in European waters during the late 16th Century were probably devised in response to red tides (Schofield et al., 1999). These and many other accounts indicate that red tides have been observed for many years before the present, affecting and impacting ancient societies and economies.

The Gulf of Mexico (Figure 1.1) has long experienced persistent red tide problems. The chronological red tide record in the Gulf of Mexico may have started in 1542. Descriptions by the Spanish explorer Álvar Núñez Cabeza de Vaca about the local Indian tribes’ shellfish harvest closure (Walsh et al., 2002) and chronicles of their season’s timetable related to fish die-offs (Landsberg, 2002) could indeed be due to red tide events. However, caution should be applied to the interpretation of Vaca’s statement because red tides are not observed annually in Texas in the present day (Wynne et al.,
2005). The next red tide-related fish kill was reported in 1648 around the Yucatan Peninsula (Mexico; ~21°N, 90°W), although it was not until 1875, in Veracruz (Mexico; ~19°N, 98°W) that an early studies linking fish kills to human respiratory irritation was made (Magaña et al., 2003).

Figure 1.1. Gulf of Mexico.

1.3. Florida Red Tide

The first officially documented Gulf of Mexico red tide within US territorial limits was registered off the West Florida Shelf in 1844, 300 years after Vaca’s testimonials (Magaña et al., 2003). Around a century later in the mid-1940s, the major dominant causative organism associated with the so-called Florida Red Tide phenomena was
identified as *Karenia brevis* (abbreviated as *K. brevis* herein, and formerly described in the literature as *Gymnodinium breve* and *Ptychodiscus brevis*; Figure 1.2.b). This dinoflagellate is the most notorious among the nearly 40 toxic species found throughout the Gulf of Mexico (Kusek et al., 1999), producing a suite of at least nine different neurotoxic brevetoxins, known collectively as “PbTxs” (from *Ptychodiscus brevis* Toxin). More details about the Florida Red tide can be found in the broad review of the ecology of *K. brevis* provided by Tester and Steidinger (1997), while Kusek et al. (1999) and Landsberg (2002) present an extensive and interesting evaluation of the phenomenon, and Magaña et al. (2003) depicts a historical sequence of red tides in the Gulf of Mexico.

In the Gulf of Mexico, *K. brevis* co-occurs with other species of the genus *Karenia* such as: *K. papilionacea*, *K. mikimotoi*, *K. selliformis*, and *K. “mexican hat”* (FWRI, 2002). These species are not easy distinguished using the morphological and genetic tools that have been applied to date (Yang et al., 2001; Culverhouse et al., 2003; Haywood et al., 2004). While *Karenia* can be found in different oceans around the globe (Chang and Ryan, 2004), the spatial distribution of *K. brevis* is mainly restricted to Gulf of Mexico waters. Sporadically, *K. brevis* events are reported in Mexico and Texas (Wynne et al., 2005), and the first *K. brevis* bloom in Louisiana, Mississippi, and Alabama was reported in 1996 (Stumpf et al., 2003b). There have not been many *K. brevis* reports outside the Gulf’s limits, and the only Caribbean outbreak was unconfirmed. However, occasionally, ocean currents have transported *K. brevis* through the Straits of Florida up the eastern US coast as far as North Carolina (Tester and Steidinger, 1997), and it may have even been observed as far as the coast of Spain (Steidinger et al., 1998).
The region where *K. brevis* most often blooms is along the West Florida Shelf, extending for almost 700 km from Key West in the south to the Panhandle coast in the north. This region receives the discharge of several rivers and coastal upwelling processes can also occur (Li and Weisberg, 1999). The near shore currents are influenced by wind forcing, tidal process, and buoyancy fluxes; off the Continental Shelf, the Loop Current and a drifter-free region called the “Forbidden Zone” (Yang et al., 1999) contribute to direct circulation. This shallow biological ecosystem exhibits considerable natural variations, including beaches, wetlands, mangroves, seagrass banks, bays, inlets, and coral barrier reefs. It is a diverse multi-use economic zone, and is important for commercial, tourism, educational, historical, and aesthetics. In addition, the Florida Keys National Marine Sanctuary, the Florida Bay and the outskirts of Everglades National Park (Porter et al., 1999) are also in this region. Thus, the water quality of the West Florida Shelf needs to be maintained in a relatively pristine state (Fourquarean and Robblee, 1999).

Historically, the Central West Florida Shelf (henceforth CWFS), between Tampa Bay (~28°N, 82.5°W) and Charlotte Harbor (~26.5°N, 82°W; Figure 1.2.a) has the highest occurrence of observed monospecific *K. brevis* blooms (FWRI, 2002). The blooming season typically begins around the Summer (~August) and persists until early Spring (~March). Although more than 80% of *K. brevis* blooms have taken place in late Summer-Fall (Walsh and Steidinger, 2001), a period concomitant with the season of hurricanes and local heavy rainfall, blooms of *K. brevis* have also been observed all year-round in this area (FWRI, 2002).
Figure 1.2: Historical database of the Florida Red Tide from 1957 to 2002.

a) Non-bloom (green X) and red tide (red □) sampling location; b) The toxic dinoflagellate *Karenia brevis* (FWRI, 2002).
CHAPTER 2

OBJECTIVES

This study is focused on the West Florida coast because of the high frequency of blooms of the toxic dinoflagellate *K. brevis*, which are responsible every year for massive red tides, causing severe fish kills, marine mammal die-offs, shellfish poisoning, and acute respiratory irritation in humans. Due to the nature of impacts associated with HABs, an effective method of detecting this type of algal bloom would directly bring benefits to many communities, e.g., coastal inhabitants, tourists, and the aquaculture, shellfish, and fisheries industries. Therefore, if the spatio-temporal knowledge of the *K. brevis* blooms can be attained from satellite-borne measurements, appropriate agencies could improve mitigation action plans, and the scientific community could make better use of field resources when studying the conditions that give rise to red tides.

The central hypothesis of the present research is to determine if blooms of *K. brevis* are associated with explicit changes in the optical properties of the water (McLeroy-Etheridge and Roesler, 1999; Schofield et al., 2006) that can be identified using satellite measurements. Thus, for the first time a long-term dataset of MODerate Resolution Imaging Spectroradiometer (MODIS; Salomonson et al., 1989; Esaias et al., 1998) measurements is used to study blooms of *K. brevis* in West Florida. The overall aim is to perform a match-up evaluation (i.e., pairing the remote sensing dataset with near-coincident *in situ* measurements of cell abundance) to determine the ability of satellite-
based algorithms to detect specific optical signatures associated with *K. brevis* blooms. The published algorithms tested are: the **Operational method** (Stumpf et al., 2003; Tomlinson et al., 2004), and the **Bio-optical technique** (Cannizzaro, 2004; Cannizzaro et al., 2008). A third algorithm, introduced in this study and which we refer to as the **Empirical approach**, was developed to determine if a simpler approach can be as effective as the more complicated published algorithms.

The specific objective is to use an extensive *in situ* database to determine the algorithms accuracy at detecting the optical signatures associated with *K. brevis* blooms. A spatial and temporal investigation aims to assess the uncertainties associated with detecting *K. brevis* from space-borne measurements in terms of regional properties along the full extent of the West Florida coast and seasonal effectiveness. Through further development or functional modifications, an additional goal is to accomplish an empirical optimization to increase the sensitivity and specificity of the algorithms, so as to improve the confidence and predictability of the detection of *K. brevis* bloom using satellite data.
CHAPTER 3

REMOTE SENSING OF THE COLOR OF THE OCEAN

The total radiance acquired by space-borne visible sensors (L_{sat}(\lambda)) is profoundly influenced by light scattered by air molecules and aerosols (Gordon and Clark, 1981). The value of L_{sat}(\lambda) is also affected by other factors such as specular reflection of the Sun by the ocean’s surface (Gordon and Wang, 1994). In fact, most of the signal exiting the top-of-atmosphere is comprised of contributions from the atmosphere (e.g., ~90% in the blue (\lambda of ~440 nm) and ~95% in the green (\lambda of ~550 nm); Gordon, 1998). As a result, corrections on a pixel-by-pixel basis are required to retrieve the other small portion (5~10%) as accurately as possible in order to recover the radiance backscattered out of the water just beneath the sea surface, the L_{w}(\lambda) (Gordon, 1978; 1997; Gordon and Wang, 1994; Siegel et al., 2000). In the western Gulf of Mexico, when re-suspension occurs, a method to split benthic algae Chl from the water column Chl must be applied (Wynne et al., 2005). In the eastern Gulf of Mexico, global Chl band-ratio algorithms (e.g., O’Reilly et al., 2000) tend to overestimate Chl up to a factor of two, and if atmospheric correction is not well applied this error can be increased by another factor of two (Ruddick et al., 2000; Stumpf, 2001). Hence, with certain assumptions (Deschamps et al., 1983; Franz et al., 2007a), L_{sat}(\lambda) can be described as the sum of the desired information coming out of the water (i.e., the water-leaving radiance; L_{w}(\lambda); Gordon and Wang, 1992a; 1992b), plus any additional radiance that may contribute to the signal (i.e., L_{sky}(\lambda)):
\[ L_{\text{sat}(\lambda)} = \tau(\lambda) \times L_{\text{w}(\lambda)} + L_{\text{sky}(\lambda)} \quad (1) \]

where \( \tau(\lambda) \) is the atmospheric diffuse transmittance.

In optical remote sensing of seawater, the intrinsic color of the ocean (sometimes referred to as the visible sea spectral reflectance) is usually described in terms of spectral remote sensing reflectance \( (R_{\text{rs}(\lambda)}; \text{Zaneveld}, 1982; 1995; \text{Olzews} \text{i} \text{k and Dareck}, 1999) \). In fact, this quantity is generally used to characterize optical properties of the water, and is defined by Gordon et al. (1988) as the ratio of the upwelling radiance \( (i.e., L_{\text{w}(\lambda)}) \) to the incident light (or downwelling solar irradiance; \( E_{\text{d}(\lambda)} \); Gregg and Carder, 1990; Pilewskie et al., 1998; 2000). Also, as light at any given wavelength interacts with any water body, it is absorbed or scattered depending upon the composition medium (Kirk, 1994). Another way to express \( R_{\text{rs}(\lambda)} \) is by means of inherent optical properties of the water \( (i.e., \text{properties that only depend on the content of the water despite light conditions}; \text{Hoge et al.}, 2001) \) which includes the absorption \( (a(\lambda); \text{Lee et al.}, 1996) \) and scattering \( (b(\lambda); \text{Maffione and Dana}, 1997) \) coefficients. The \( b(\lambda) \) can be further separated into two scattering component: forward \( (b_f(\lambda)) \) and backward \( (b_b(\lambda)) \). Thus, the \( R_{\text{rs}(\lambda)} \) is also controlled by the ratio of light backscattered by the ocean surface \( (b_b(\lambda)) \) to the absorbed light \( (a(\lambda)) \). In a simplified manner (Sathyendranath and Platt, 1997; Carder et al., 1999), the \( R_{\text{rs}(\lambda)} \) can be expressed by:

\[ R_{\text{rs}(\lambda)} = \frac{L_{\text{w}(\lambda)}}{E_{\text{d}(\lambda)}} \approx \alpha \times \left( \frac{b(\lambda)}{a(\lambda)} \right) \quad (2) \]
where $\alpha$ is a dimensionless constant (Carder et al., 1999). However a wide range of values of $\alpha$ are commonly found in the literature, from 0.33 (Morel and Prieur, 1977) to 0.083 (Kirk, 1994). Inherent optical properties of the water are not produced directly from the satellite measurements, and details on the derivation using $R_{rs(\lambda)}$ can be found in Lee et al. (1994) and Carder et al. (1999). In general, the more derived the product, the more error it has (Bailey and Werdell, 2006).

Because $L_w(\lambda)$ varies at a specific location with the sea surface light field, conditions depend upon the time of day as well as for the time of the year, and so a normalized water-leaving radiance ($nL_w(\lambda)$; Gordon, 2005) is used in lieu of $L_w(\lambda)$. Having very little dependence on the solar zenith angle and viewing geometry (Gordon and Clark, 1981), the $nL_w(\lambda)$ it is calculated by multiplying the $R_{rs(\lambda)}$ by the mean extraterrestrial solar irradiance ($F_0(\lambda)$), which is well known at any given wavelength at the top-of-atmosphere (Souto et al., 2005), such that:

$$nL_w(\lambda) = F_0(\lambda) \cdot R_{rs(\lambda)}$$  \hspace{1cm} (3)

To summarize, the primary optical variables ($i.e.$, $R_{rs(\lambda)}$, $L_w(\lambda)$, $nL_w(\lambda)$) along with several existing spectral bio-optical algorithms ($e.g.$, Gordon et al., 1983; Carder et al., 1999; 2004; O’Reilly et al., 2000) can be used to estimate inherent optical properties of the water and geophysical parameters such as Chl. Blondeau-Patissier et al. (2004) provide a review of some bio-optical algorithms used to derive Chl, and Appendix 1 provides details about inherent optical properties of the water, optically active constituents and other remote sensing-related matters.
CHAPTER 4

SATELLITE-BASED RED TIDE ALGORITHMS

Chapter 4 describes the pre-existing satellite-based algorithms for detecting harmful algal blooms, and introduces the proposed alternative methodology.

4.1. The official NOAA monitoring system: OPERATIONAL method

Hu et al. (2005a; 2008) present an interesting association regarding the Chl content of K. brevis cells, based on the work of Evens et al. (2001). Assuming ~10 Chl pigments per cell, the background population of K. brevis in the Gulf of Mexico (~100 cells l\(^{-1}\); Geesey and Tester, 1993) would comprise about 0.001 mg m\(^{-3}\) of Chl. However, this amount of Chl is less than local Gulf of Mexico offshore levels (~0.1 mg m\(^{-3}\); Muller-Karger et al., 1991; Gilbes et al., 1996), and even below global Chl average values (~0.24 mg m\(^{-3}\); Gregg and Conkright, 2002).

Using a single CZCS image, Tester et al. (1998) demonstrated that the minimum threshold concentration to detect blooms of K. brevis from space is on the order of 10\(^5\) cells l\(^{-1}\). Taking this threshold together with the aforementioned pigmentation’s assumption, blooms of 10\(^5\) cells l\(^{-1}\) would correspond to approximately 1 mg m\(^{-3}\) of Chl, edging out mesotrophic and eutrophic oceanic province limits (Antoine et al., 1996). In addition, laboratory studies have shown that 10\(^5\) cells l\(^{-1}\) of K. brevis contain ~1 mg m\(^{-3}\) of Chl (Stumpf et al., 2003b).
The use of the Chl content of *K. brevis* cells provide insights about the use of Chl anomalies to detect these type of blooms in region where they monspecifically occur, like the West Florida coast. As climatologic satellite-based fields of geophysical parameters are widely used in oceanographic studies, e.g., SST (Palacios, 2004), Thomas (2000) and Stumpf (2001) recommend the use of Chl anomalies to investigate possible occurrences of blooms of *K. brevis* in the Gulf of Mexico. NOAA’s CoastWatch program has implemented this suggestion, issuing operational Chl anomaly bulletins (HabFS, 2007) to detect possible blooms of *K. brevis* in the Gulf of Mexico from SeaWiFS data (Stumpf et al., 2003b; Tomlinson et al., 2004). In the present study, NOAA’s official Chl anomaly method will be referred to as the **OPERATIONAL method**.

The *Operational* method, takes into account the difference (i.e., anomaly) of the current satellite Chl image and a Chl running mean of the “previous” 60-days. In order to avoid Chl mean biases in the presence of slowly changing blooms, a two week-long window is left between the analyzed image and the mean (e.g., if the analyzed image was acquired in March 14 of 2004, the mean would be calculated from images from February and January). While determining the Chl mean, in the time binning process, at least 10 pixels had to be considered per bin. In this way, using the Chl pigment-relation, and laboratory analyses, satellite Chl anomalies of 1 mg m\(^{-3}\) are assumed to represent an increase of 10\(^5\) cells of *K. brevis* per liter from the two months mean field. Consequently, the *Operational* method flags possible red tides events whenever Chl anomalies are > 1 mg m\(^{-3}\) (Figure 4.1).

A feature regarding the implementation of this method is the use of ancillary meteorological information (e.g., wind) that is aggregated to the knowledge of previous
blooms to helps the interpretation of the analyst (Fisher et al., 2005). However, uncertainty stems from some aspects of this monitoring algorithm. The foremost is that Chl anomaly does not correspond to the presence of any specific algae, as many plants besides toxic algae species contain Chl pigments (Garver et al., 1994). Hence, caution should be taken regarding the results of the Operational method, given that any major bloom-forming photosynthetic group can give raise to such Chl anomalies $> 1 \text{ mg m}^{-3}$. Furthermore, atmospheric effects (Ruddick et al., 2000), turbidity (Jupp et al., 1994; Hu et al., 2000; 2006a), and even bottom reflection (Ohde and Mueller, 2001; Gould and Arone, 1998; Lee et al., 2001; D’Sa et al., 2002) can also contaminate satellite measurements of Chl.

Figure 4.1: The relationship of the Operational method (Stumpf et al., 2003b; Tomlinson et al., 2004). Horizontal solid line represents a chlorophyll anomaly of 1 mg m$^{-3}$ and discriminates among flagged red tides (above the line; red shaded area) and non-blooms (below).
In summary, the Operational method is not sufficient to detect the presence of a *K. brevis* bloom with certainty, just speculating the presence of these types of blooms when a Chl anomaly of 1mg m\(^{-3}\) or greater is detected. Also, the detectable cell threshold of this method (i.e., 10\(^5\) cells l\(^{-1}\)) is therefore two orders of magnitude higher than state agency shellfish bed closure limits of 5x10\(^3\) cells l\(^{-1}\) (FDA, 2005) based on NSP threats, and an order of magnitude higher than those defined by other selected algorithms.

### 4.2. The pre-existing satellite-base algorithm: *BIO-OPTICAL* technique

Since Chl is present in all plants, one method for optically detecting and distinguishing HAB species from non-toxic species could capitalize on unique spectral features associated with HAB (Garver et al., 1994; Schofield et al., 1999). Based on the assumption that HABs species experience low grazing rates as they accumulate in a localized area, due to the toxins they contain (Shuman and Lorenzen, 1975; Chattopadhyay et al., 2002; Prince et al., 2006a; 2006b), the proliferation of HAB species could cause unique changes in the optical characteristics of the water column (*e.g.*, lack of detritus). Thus, Cannizzaro (2004) and Cannizzaro et al. (2008) developed a classification technique for detecting *K. brevis* based on these optical changes. This algorithm will be referred to as the *BIO-OPTICAL technique*.

A threshold of 10\(^4\) cells l\(^{-1}\) is used by the *Bio-optical* technique to define a detectable *K. brevis* bloom. This is an order of magnitude more sensitive than that of the *Operational* method. With a multi-year/multi-season shipboard radiometric bio-optical dataset collected mostly as part of the ECOHAB program, a considerable remote sensing reflectance (R\(_{rs(\lambda)}\)) drop was noted in going from waters containing low *K. brevis* cells, as
compared with $R_{rs(\lambda)}$ in areas where higher *K. brevis* concentrations were present. Variations in the particulate backscatter ($b_{bp(\lambda)}$) were observed to cause this $R_{rs(\lambda)}$ change.

The *Bio-optical* technique proposes a classification scheme to detect *K. brevis* based on the simultaneous presence of two criteria: high Chl and low $b_{bp(\lambda)}$. The estimates of $b_{bp(\lambda)}$ and Chl are calculated with the Carder algorithm (Carder et al., 1999). Thus, red tides are flagged by *Bio-optical* if satellite-measured Chl values are > 1.5 mg m\(^{-3}\) and values of satellite-measured particulate backscatter ($b_{bp\text{SAT}(550)}$) are concurrently lower than calculated (Morel, 1988) values of particulate backscatter ($b_{bp\text{M}(550)}$). Meanwhile, cases not meeting these criteria are identified as non-blooms. Figure 4.2 depicts this relationship.

![Figure 4.2: The relationship of the Bio-optical technique (Cannizzaro, 2004; Cannizzaro et al., 2008). Vertical dashed line is 1.5 mg m\(^{-3}\) of chlorophyll and the diagonal dashed line represents the calculated (Morel, 1988) values of particulate backscatter. Flagged red tides are in the lower right quadrant (red shaded area).](image-url)
The only results of this technique that have been published using satellite imagery are derived from data from one SeaWiFS image, but it successfully separated among the absence and presence of \textit{K. brevis} blooms. In this single satellite application, a relaxed temporal window within few days of the analyzed overpass was employed on its \textit{in situ}-satellite match-up, justified because blooms can last more than a day or two in the same region.

### 4.3. The new, simple, alternative: \textit{EMPIRICAL} approach

The present investigation introduces an experimental approach to identify blooms of \textit{K. brevis} in the Gulf of Mexico. Since inherent optical properties of the water (such as $b_{bp(\lambda)}$) are a derived satellite product and not directly measured by ocean color sensors, the performance of an alternative method was established taking into account Chl and a single water-leaving radiance band ($L_{w(\lambda)}$). This method, henceforth referred to as the \textit{EMPIRICAL approach}, is derived from MODIS-Aqua data and validated with \textit{in situ} cell counts.

Gordon and McCluney (1975) demonstrated that maximum EPDs for Case-2 waters occur approximately 1.5 m for wavelengths around 600 nm. In order to identify \textit{K. brevis} blooms in the West Florida Shelf, the \textit{Empirical} approach utilizes the 550 nm wavelength, which penetrates coastal waters most deeply out of the available current satellite ocean color sensors (Sackmann and Perry, 2006). At 550 nm the absorption is somewhat low and attenuation is primarily due to water molecules backscatter (Cannizzaro, 2004).
The *Empirical* approach was designed as a simple algorithm to differentiate waters with *K. brevis* blooms from non-blooming conditions through comparisons between the values of two different properties: the $L_{w(550)}$ and the total scatter ($b_{(550)}$). Through the development of this algorithm, an long-term investigation of *in situ* cell counts validated the accuracy of the relation between Chl and $L_{w(550)}$. For an event to be classified as a red tide by the *Empirical* approach, it must have the satellite-measured $L_{w(550)}$ below that related to the total scatter ($b_{M(550)}$) calculated by Morel (1988). The other cases are determined to be non-bloom. The cell count threshold used to determine the presence or absence of a bloom is $15\times10^3$ cells l$^{-1}$. Figure 4.3 depicts this relationship.

![Empirical Approach Graph](image)

**Figure 4.3:** The relationship of the *Empirical* approach. Dashed line represents the calculated (Morel, 1988) values of total scatter. Flagged red tides are under the line (red shaded area).
CHAPTER 5

DATASET

The satellite dataset and in situ database used in the present investigation are presented in this Chapter.

5.1. Satellite imagery

The National Aeronautics and Space Administration’s (NASA) proof-of-concept Coastal Zone Color Scanner (CZCS; Hovis et al., 1980; Evans and Gordon, 1994) was put in orbit in 1978 onboard the NIMBUS-7 satellite. Intended to be an experimental mission, CZCS was expected to operate for ~1 year, but it actually acquired information for almost ten years before it stopped operation in 1986.

The success achieved by CZCS prompted NASA to launch other ocean color satellites. After a hiatus of more than ten years following the demise of CZCS, in August of 1997, the Sea-viewing Wide Field-of-view Sensor (SeaWiFS; Hooker et al., 1992; Hooker and McClain, 2000; McClain et al., 2004) lead the way for the next generation of visible satellite sensors (IOCCG, 2007). SeaWiFS was designed specifically for ocean color applications, having a noteworthy mechanism to avoid sun-glint: an along-track scanner tilt of ±20°.

CZCS and SeaWiFS were the heritage instruments of the widely used ocean color remote sensor, the NASA Earth Observing System’s (EOS) MODerate Resolution
Imaging Spectroradiometer (MODIS; Salomonson et al., 1989; Esaias et al., 1998). Two nearly identical MODIS instruments are currently operating on the Terra and Aqua satellites, launched in December 1999 and May 2002. MODIS is well suited for global monitoring with frequent temporal coverage, with each instrument having one to two days site revisit, above and below 30° of latitude respectively (Anderson et al., 2003). MODIS was designed for multi-purpose and multi-disciplinary Earth imaging capability (e.g., atmosphere, land and ocean applications), offering high-quality retrievals such as solar-stimulated phytoplankton Chl fluorescence relative to a baseline (i.e., fluorescence line height; Letelier and Abbott, 1996; Gower and Borstad, 2004; Huot et al., 2005), and simultaneous physical (e.g., SST) and biological (e.g., Chl) measurements (Denman and Abbott, 1994).

Both SeaWiFS and MODIS have swaths centered on the spacecraft’s ground track and have very similar ocean color bands, however those of MODIS are narrower than those of SeaWiFS (Esaias et al., 1998). With regards to the signal-to-noise ratio, MODIS is approximately a factor of three more sensitive than SeaWiFS (Gower and Borstad, 2004).

The analysis conducted in this investigation is focused on measurements acquired by MODIS, mainly for four reasons: 1) No one had performed a long-term analysis using MODIS imagery in the study of HABs in the Gulf of Mexico; 2) SeaWiFS was intended for a five-year mission, and was already well beyond its design lifetime when this investigation started in 2006, and at the beginning of 2008, its geolocation mode stopped working properly; 3) The MODIS mission serves as a prototype, representing a potential
linkage between the ongoing and the next generation of satellite instruments (IOCCG, 2007); and 4) MODIS data are freely available near-real-time.

The MODIS on NASA’s Aqua satellite (Collection 4) was preferred over the instrument on Terra. MODIS-Terra has been shown to have significant uncertainties and instabilities, such as an instrumental artifact consisting of data striping perpendicular to the satellite track caused by differences between mirror sides and detectors (Hoge et al., 2003). It is important to note that when using MODIS data for ocean color applications, a 2-fold consideration should be emphasized: 1) Because MODIS-Aqua corrections are more complete (Franz et al., 2000); and 2) MODIS-Terra has a degraded temporal trend in calibrating the shortest ocean color band wavelength with profound impact on its signal (Barnes et al., 2003). However, MODIS-Terra data were used to replicate published results derived using SeaWiFS data (Cannizzaro 2004; Cannizzaro et al., 2008) in a case study that predated the launch of Aqua (Section 6.1.1).

5.2. **In situ dataset**

To confirm the detection of HABs in the eastern Gulf of Mexico on satellite measurements, an *in situ* database (1956–2006) of the presence and abundance of the genus *Karenia* was used. This *in situ* compilation, gathered and distributed by the Florida Fish and Wildlife Conservation Commission’s Fish and Wildlife Research Institute, covers the entire West Florida Shelf. This sampling effort was collected over time by a variety of institutions and programs, lacking consistent standardization processes, *e.g.*, sampling collection methods or counting procedures (FWRI, 2002; Hu et al., 2008).
All entries of this database contain a handful of information in common: sampling date, site location (i.e., latitude and longitude), Karenia concentration abundances in cells l\(^{-1}\), and the depth that the sample was collected. The precision of the cell counts concentration in this database, indicating algal presence, had the minimum value of 333 cells l\(^{-1}\) (Hu et al., 2008). Some extra ancillary data were sometimes available, and could include Chl levels, water temperature, salinity, nutrients, and the presence of other non-toxic phytoplanktonic species, but this information was not available for the majority of the samples. As a result, the assessment of these additional water quality variables for further applications was limited. Furthermore, an important piece of information needed for validating satellite data was not accessible for most in situ stations: the sampling time. Neither was the value of site bathymetry, another vital piece of information need in ocean color validation studies in near shore regions. Other relevant aspects for matching satellite ocean color measurements with in situ data were not documented, e.g., water stratification and bloom patchiness. And even though Global Position System equipment is now easily available, the validity of the field station coordinates cannot be assured.

In addition, whereas some monitoring programs that provided information to this database had routine site-revisits, many contributions were done as response monitoring with no regular site distribution. Also, the collection depth of the samples varied as well. These and other characteristics of this in situ database introduce a risk of spatial and temporal sources of uncertainties.
CHAPTER 6

METHODS

Chapter 6 introduces methods used to process satellite images and the procedures to treat in situ data. The in situ-satellite collocation match-up analysis is described as well. This Chapter also focuses on analyses of the dataset using the analyzed pre-existing satellite-based algorithms. The optimization procedure is also elucidated, so is the combined hybridized scheme.

6.1. Satellite imagery

Individual daily granules (i.e., files containing five minutes worth of data covering 2000 by 2330 km) of MODIS-Aqua data (Level-1A files) spanning more than four years (beginning in June 2002 through the end of October 2006) were downloaded from the Ocean Color website (OCW, 2007). To reduce the volume of the data, the area of interest (22°N~32°N of latitude and 77°W~92°W of longitude; Figure 1.2) was extracted from 1665 scenes at a spatial resolution of 1 km² at the nadir point.

All satellite images used in this investigation were processed at the University of Miami's Rosenstiel School of Marine and Atmospheric Science (RSMAS) Remote Sensing Laboratory (RRSL) using the most up-to-date algorithms and software: the NASA-provided SeaWiFS Data Analysis System (SeaDAS package version 5.0.3 - SeaDAS, 2007; Baith et al., 2001). In-house codes were written in Interactive Data
Language (IDL) and UNIX shell to automatically process and analyze data. A Dell 1950, dual core 2.33 GHz Xeon processor with 2GB of memory running Fedora core 5 was used for data processing. Image analysis was done on a Macintosh Dual 500 MHz Power PC G4 with 1.38 GB SDRAM of memory.

Individual Level-1A files were radiometrically processed and geometrically corrected to produce Level-2 images (i.e., L2 file) using the Multi-Sensor Level-1 to Level-2 SeaDAS processing code (MSL124 version 5.2.3; Baith et al., 2001; Feldman and McClain, 2005). Each L2 file included a large geophysical parameter library (Table 6.1).

### Table 6.1: Geophysical parameter library.

<table>
<thead>
<tr>
<th>Bands</th>
<th>Parameter</th>
<th>Inherent optical properties of the water (Carder et al., 1999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>412</td>
<td>( b_{\lambda} )</td>
<td>Total scatter</td>
</tr>
<tr>
<td>443</td>
<td>( b_{bp(\lambda)} )</td>
<td>Particulate backscatter</td>
</tr>
<tr>
<td>488</td>
<td>( a_{\lambda} )</td>
<td>Total absorption</td>
</tr>
<tr>
<td>531</td>
<td>( a_{ph(\lambda)} )</td>
<td>Absorption due to phytoplankton</td>
</tr>
<tr>
<td>551</td>
<td>( a_{dg(\lambda)} )</td>
<td>Absorption due to detritus and CDOM</td>
</tr>
<tr>
<td>667</td>
<td>( nL_{\lambda} )</td>
<td>Normalized water-leaving radiance</td>
</tr>
<tr>
<td>678</td>
<td>( L_{\lambda} )</td>
<td>Water-leaving radiance</td>
</tr>
<tr>
<td>748</td>
<td>( R_{\lambda} )</td>
<td>Remote sensing reflectance</td>
</tr>
<tr>
<td>869</td>
<td>Chl</td>
<td>Chlorophyll (O’Reilly et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>( K_{490} )</td>
<td>Downwelling diffuse attenuation coefficient</td>
</tr>
<tr>
<td></td>
<td>FLH</td>
<td>Fluorescence line height</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>Sea surface temperature</td>
</tr>
<tr>
<td></td>
<td>Solar and satellite zenith angles</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Solar and satellite azimuth angle</td>
<td></td>
</tr>
</tbody>
</table>
Three procedures were employed to minimize storage requirements and to optimize processing time while generating the L2 files: 1) Only the best quality flag data (flagged “0” for good) were used; 2) Three default MSL12,4 masks were applied: land (pixels over land); cloud (pixels having albedo at 865 nm >1.1%); and hi-lt (pixels with radiances > knee value in one or more bands causing reduced precision); and 3) Because inconsistencies in the atmospheric correction are experienced in the ocean color dataset for high zenith angles (both solar and satellite), no pixel was considered having satellite zenith angles > 60° (Bailey and Werdell, 2006). Other masking actions employed are described in section 6.3.

The semi-analytical ocean color algorithm (Carder et al., 1999; 2004) was used to derive the inherent optical properties (Table 6.1), and Chl was estimated using the global band-ratio OC3M (O’Reilly et al., 2000). To replicate the Operational method, Chl anomalies were derived by subtracting a 60-day Chl running mean from the analyzed Chl image, both mapped to the same projection. A 2-week long window was left between the analyzed image and the mean. The Chl mean was generated with space bin resolution of 1km, and had to have at least 10 pixels in each bin.

Although the Chl anomaly match-up of the Operational method was done on mapped images, further match-ups from the other selected algorithms were performed using satellite coordinates (i.e., non-mapped L2 files). Ultimately, for visualization of each L2 file, quick-looks (e.g., Chl, SST) were then created in the Portable Network Graphics format (png) and mapped covering the region of interest (i.e., the West Florida coast; Figure 1.2) on a common grid system (i.e., cylindrical equidistant projection). Figure 6.1 depicts the satellite imaging process steps.
Figure 6.1: Satellite imaging process steps.
6.1.1. Preliminary application of the algorithms to detect *K. brevis* blooms

To verify consistency with the published *Bio-optical* technique of Cannizzaro (2004) and Cannizzaro et al. (2008), the same SeaWiFS image (from August 30, 2001) was used to replicate the results. Since ocean color measurements are dependent on clear skies, by applying this technique to data from additional ocean color satellite sensors (*e.g.*, MODIS) this could increase temporal sampling by increasing the chance of acquiring cloud-free images. Use of a multi-sensor approach requires uniformity in the methodology outcomes, and a demonstration of the equivalence of the results from different sensors. At that date, MODIS-Aqua had not yet been launched, so the replication using MODIS data was performed on a MODIS-Terra image, and the closest clear-sky image was used, from August 29, 2001.

The detectable *K. brevis* bloom threshold was the same as used by the *Bio-optical* technique (*i.e.*, $10^4$ cells l$^{-1}$), but some other adjustments were implemented. The same *in situ* dataset (FWRI, 2002) was used to match with the satellite imagery, however with a more relaxed temporal window (nine days instead four; from August 25 to September 2, 2001). In lieu of the Carder algorithm (Carder et al., 1999), the Garver-Siegel-Maritorena (GSM01; Maritorena et al., 2002) semi-analytical ocean color algorithm was used to derive $b_{bp(\lambda)}$; and Chl estimates were calculated with the global band-ratio algorithms: OC4 for SeaWiFS and OC3 for MODIS (O’Reilly et al., 2000).

In addition to comparing the *Bio-optical* technique in SeaWiFS and MODIS-Terra imagery, the *Empirical* approach, introduced in the present investigation, was also evaluated in the same manner, and applied to data from both sensors.
6.2. In situ dataset

The historical in situ database of the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute details the sampling efforts targeting the genus Karenia over the last 50 years (1956~2006). Even though this database extends back to mid 1950’s, the work reported here was limited to the comparisons within the available observations of the MODIS-Aqua mission. Congenerics of Karenia brevis (e.g., K. papilionacea, K. mikimotoi, K. mexicanhat, and K. selliformis) are known to occur in the Gulf of Mexico. However, because K. brevis is the most abundant species (FWRI, 2002), and the most notorious in terms of its toxic effects (Kusek et al., 1999), the match-up analysis conducted here took into account only K. brevis samples. Therefore, a number of in situ records of other Karenia species were excluded. Thus, hereafter, any reference to the “in situ database” refers specifically to the subset limited to data on K. brevis taken between June of 2002 to October of 2006.

Based on cell count per liter, the in situ database was used as a binary classification scheme recording samples representing conditions with the absence or presence of K. brevis blooms. Therefore, blooms of K. brevis were defined as having a cell count exceeding 15,000 cells l⁻¹ (Table 6.2). A value of “111” was used whenever Karenia cells were not present (i.e., meaning absence of cells or “0” cell count) to allow data to be plotted in logarithmic-scaled plots.

Site bathymetry data are of fundamental importance in using ocean color satellite data, because of possible bottom reflection influence on the satellite signal in shallow areas (Gould and Arone, 1998; Lee et al., 2001; Ohde and Mueller, 2001; Ohde and Siegel, 2001; D’Sa et al., 2002). However, since this information was not included in the
database for most *in situ* stations, the ETOPO2 (Earth TOPOgraphy database at 2-minute resolution; Geodas, 2007) was used to determine local water column depth for all stations. But because the distance between each ETOPO2 point is almost 4 km, erroneous values for some *in situ* stations sited near shore occurred (*i.e.*, altitudes in lieu of depths). Consequently, such stations were excluded from the analysis.

<table>
<thead>
<tr>
<th>Non-bloom</th>
<th><strong>≤ 15,000 cells l⁻¹</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Red tide</td>
<td><strong>&gt; 15,000 cells l⁻¹</strong></td>
</tr>
</tbody>
</table>

### Table 6.2: Cell counts threshold.

**Binary classification scheme of cells of *Karenia* spp. per liter**

#### 6.2.1. *In situ* quality assurance masks

- **LAND PROXIMITY MASK:** To avoid taking into account pixels matched-up with stations prone to errors (*e.g.*, bottom reflection and atmospheric correction) derived from shallow regions with terrigenous sediment load, a buffer zone was created around the Florida shoreline. Since satellite imagery is subject to geolocation inaccuracies (Patt, 2002), but because MODIS images have geolocation accuracy of 150 m at nadir (Wolfe et al., 2002), all *in situ* stations within distances of half pixel (*i.e.*, 500 m) from the coast were eliminated. This mask also helped to reduce the chance of direct contamination by land effects in the satellite pixel.

- **ETOPO2 MASK:** This mask removed any samples from the *in situ* database collected within 2 m of the bottom as given by ETOPO2 bathymetry to prevent possible bottom reflectance in the satellite signal (Lee et al., 1999a; 199b; Ohde and Siegel, 2001). The value of 2 m was chosen to be more cautious regarding the use of the ETOPO2 bathymetry.
6.2.2. Spatio-temporal uncertainties

The uncertainties associated with the space-borne detection of *K. brevis* blooms were determined for distinct geographic locations and different periods of the year. Based on the visual analysis of the historical red tide database, the West Florida coastline was divided into three sub-regions: Central West Florida Shelf (CWFS), North (N-FL), and South (S-FL), with the latitudes of 28.25°S (north of Tampa Bay) and 25.75°S (south of Cape Romano) establishing the limits between the CWFS and N-FL, and CWFS and S-FL. The database was divided into two periods: a less favorable period (*i.e.*, Winter-Spring) and the blooming season (*i.e.*, Summer-Fall). At first, the data analysis was performed for the entire region and throughout the whole year. Then, to verify the spatio-temporal performance of the detecting *K. brevis* blooms using satellite measurements, the relationship between regions was explored within each of the two seasons.

Figure 6.2 depicts the process steps assessed during the *in situ* dataset analysis.

6.3. Match-up analysis

A match-up analysis consists primarily of comparisons between field data and related satellite measurements. In another words, match-up pairs of *in situ* attributes and data acquired from space are used to search for specific signatures within the satellite measurement that are in accordance to what is really present in the water at a certain location and time. Thus, this investigation matched-up *in situ* abundances of *K. brevis* with collocated satellite information to determine the accuracy of the remote-sensed measurements.
Figure 6.2: Flowchart depicting the steps assessed during the *in situ* dataset analysis.

Because the exact sampling time was not listed for many stations within the *in situ* database, the match-ups were temporally relaxed to a 24h window. Hence, the satellite pixel and collocated field station were matched if acquired within the same day.
The present research basically used SeaDAS to conduct the *in situ*-satellite pairing, but the version utilized did not allow calculating the statistics (e.g., mean and standard deviation) of the pixels within a multi-pixel box (Hu et al., 2001; Bailey and Werdell, 2006). Therefore, the pairs of match-ups were only derived from a full resolution pixel covering the location of the *in situ* data point, and a pixel box investigation could not be performed in order to compare results. This pixel-by-pixel evaluation was carried and used to train the selected algorithms at detecting the presence or absence of *K. brevis* blooms.

In order to increase the quality of the match-ups, stringent flagging procedures were developed. A mask was applied to maintain the confidence of valid *in situ*-satellite pairs and its acceptable quality:

- **LIGHT PENETRATION MASK:** This mask assured that all *in situ* samples were within the depth range recoverable by satellite ocean color sensors (Bailey and Werdell, 2006). Samples were only considered if the computed satellite pixel value of K_{490} was positive and < 1, and the calculated effective penetration depth (EPD; see Appendix 5) minus two meters was deeper than the actual collection depth (Figure 6.3). The value of two meters was chosen to ensure the *in situ* samples were taken at a depth shallow enough for them to contribute to the satellite signal, given that there are possible uncertainties in the calculation of K_{490}. 
Figure 6.3: **Left (Light penetration mask):** Samples were eliminated if below the depth range recoverable by satellite ocean color sensors, calculated with the effective penetration depth (EPD; see Appendix 5) minus 2 meters (in this example, only sample 1 was considered); **Right (Bottom reflection threshold):** Used to avoid bottom reflection it added 2 meters to the apparent optical depth (AOD; see Appendix 5), and only considered ETOPO2 bathymetry deeper than this value (in this example, only bottom A was considered).

Two alternative tests were designed to determine whether the application of a bottom reflection threshold could retain good data obtained from shallow regions. This is important due to the fact that from a human point of view the worst problems caused by red tides are along shorelines (Roesler and McLeroy-Etheridge, 1998), and also because the satellite ocean color datasets face limitations whenever applied near shore (Lee et al., 1999; Ohde and Siegel, 2001; Hyde et al 2007).
- **BOTTOM REFLECTION THRESHOLD:** 1) As in other studies of algal blooms using satellite data (e.g., Subramaniam et al., 2002; Hu et al., 2008), a “shallow water mask” was applied constraining the use of valid match-up pairs to deep waters (*i.e.*, >10 m); and 2) A more relaxed approach allowing the account of near shore stations, eliminated match-ups if the calculated apparent optical depth (AOD; see Appendix 5) plus two meters was shallower than the stipulated local ETOPO2 bathymetry (Figure 6.3). The value of two meters was selected to provide a margin in the use of the ETOPO2 bathymetry.

To determine the effects on spatial coverage, the space-borne results were tested with two alternative satellite zenith angle thresholds:

- **SATELLITE ZENITH ANGLE THRESHOLD:** 1) In order to avoid atmospheric correction-related errors that increase for viewing angles > 45° (Ding and Gordon, 1994), valid match-up pairs were restricted to smaller angles; and 2) A less strictive approach that accepts a wider satellite zenith angle range (*i.e.*, up to 60°) was used to explore the possibility of increased special coverage.

Even though the selected algorithms only utilize few geophysical variables (*i.e.*, Chl, Chl anomalies, $b_{bp(\lambda)}$ and $L_{w(\lambda)}$) the entire geophysical parameter library thus produced (Table 6.1) was filtered for questionable values at the pixel, to avoid failing to identify errors in the atmospheric correction. These additional filters screened the match-ups pairs to ensure that it did not contain any inappropriate data. If any of the geophysical parameters within the pixel did not fit the quality assurance of the additional filters, this pixel was not used. Figure 6.4 depicts the match-ups process steps. The following filters were applied to the Level-2 geophysical parameter library:
- **CHLOROPHYLL-FLUORESCENCE FILTER:** Some published satellite-based studies of algal blooms have optimal Chl windows (*e.g.*, from 0.5 to 3.0 mg m\(^{-3}\) in Subramaniam et al., 2002) or Chl thresholds (*e.g.*, > 1.5 mg m\(^{-3}\) in Cannizzaro, 2004). In this investigation, pixels were restricted to satellite-retrieved Chl > 0 and < 100 mg m\(^{-3}\). In addition, only reasonable positive physical values for fluorescence line height were retained (in W m\(^{-2}\) μm\(^{-1}\) sr\(^{-1}\)).

- **SEA SURFACE TEMPERATURE FILTER:** As the water observed during upwelling events in the eastern Gulf of Mexico can reach temperature as low as 15°C (COOPS, 2007), this value was used as a threshold to eliminate pixels. This is justified because temperatures lower than the established threshold could be the result of unresolved cloud contamination, which could compromise a valid match-up.

- **INHERENT OPTICAL PROPERTY FILTER 1:** Pixels were retained as valid match-ups only if they had reasonable positive physical values for all inherent optical properties (Table 6.1) in bands 412, 443, 488, 531, 551 and 667; while for bands 768, 748 and 869, only pixels with value equal to zero were considered.

- **PRIMARY OPTICAL VARIABLE FILTER 2:** Pixels having unphysical negative values or > 100 for nL\(_{\lambda}\), R\(_{rs}\)(\(\lambda\)) or L\(_{\alpha}\)(\(\lambda\)), for any band, were eliminated from the match-up analysis.
Figure 6.4: Match-up analysis steps.
6.4. Spatio-temporal investigation

Following the definition of the valid in situ-satellite match-up pairs, an investigation evaluating the spatial and temporal accuracy of the satellite-based algorithms was conducted. The overall results of this investigation indicated whether sufficient data (i.e. valid match-ups) were available for spatial and temporal conclusions to be made, as well as it provided information of how well the algorithms performed in each region and period.

For this investigation, the three algorithms were analyzed with the in situ dataset divided into two subsets base on 15,000 cells $l^{-1}$ for red tide and non-bloom separation (Table 6.2). However, to facilitate comparisons with the Operational method, an evaluation of this method was also conducted with its original threshold to define a detectable bloom: $10^5$ cells $l^{-1}$.

At first, scatter-plots were constructed to test comprehensively the coherence of the entire dataset (i.e., data sampled all along the West Florida coast throughout the years of the analyzed period), followed by a temporal analysis containing data separated into two seasons: the blooming season and the off-season (see 6.2.1.). After that, a more detailed spatio-temporal coherence approach looked separately into data collected during each of the two seasons within different locations (i.e., North-FL, South-FL, and Central West Florida Shelf - CWFS; see 6.2.1.). The statistical performance of each subset, expressed as percentages of correct identification, was then evaluated by region and by season.
6.5. Detection accuracy assessment

Subsequent to the spatio-temporal investigation, which defines where and when enough information was available to draw conclusion about the algorithms’ performances, their individual success rates were analyzed. The accuracy of the outcomes was determined through statistical inference. The findings presented in this section were based on the approaches of Congalton (1991), Sokal and Rohlf (1995), Mangnusson and Mourão (2004), and the references therein.

The basic statistical assessments can provide resource managers with tools for making appropriate decisions while using the outcomes of the algorithms. To fully understand and to better extract useful information from the statistical assessment, they should not be reviewed separately, thus avoiding giving the wrong impression about the correct accuracy. Instead, the statistical assessments should be viewed together. The relative frequencies in the training classifier dataset give rise to the results of the statistical assessments. The validation of the accuracy of the algorithms would be expected to hold true with an independent or withheld dataset.

The statistical assessments were also performed with the in situ dataset divided into two subsets based on 15,000 cells l\(^{-1}\) for red tide and non-bloom separation (Table 6.2). However, the original cell count threshold to define a detectable bloom (i.e., \(10^5\) cells l\(^{-1}\)) of the Operational method was also used in the evaluation of this algorithm, facilitating comparisons with other published results.

Table 6.3 (i.e., 2-by-2 table or error matrix) explains each of the relationships between the frequencies of occurrence of each event.
Table 6.3: 2-by-2 statistical table.

<table>
<thead>
<tr>
<th></th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Red tide</td>
<td>Non-bloom</td>
</tr>
<tr>
<td>Satellite-based</td>
<td>Red tide</td>
<td>A</td>
</tr>
<tr>
<td>algorithm outcome at</td>
<td>Non-bloom</td>
<td>C</td>
</tr>
<tr>
<td>detecting red tide</td>
<td>SUM</td>
<td>A+C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalence of red tides</th>
<th>( \frac{(A+C)}{(A+B+C+D)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds of red tides</td>
<td>( \frac{(A+C)}{(B+D)} )</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>( \frac{(A+D)}{(A+B+C+D)} )</td>
</tr>
</tbody>
</table>

6.5.1. First statistical assessment

Verify the prevalence and odds of a red tide occurring in the analyzed dataset (i.e., valid match-ups). The prevalence of red tide (i.e., population of known red tide samples (A+C; Table 6.3) divided by the total population of samples (A+B+C+D; Table 6.3)) is a measure of the observed red tide frequency; while the odds of red tide (i.e., population of known red tide samples (A+C; Table 6.3) divided by the population of known non-bloom samples (B+D; Table 6.3)) represent the probability of red tide cases related to the non-bloom population.
6.5.2. Second statistical assessment

Measure of the overall accuracy in order to know the algorithm’s success rate at correctly identifying cases. This descriptive value can be derived by taking the summation of the major diagonal of Table 6.3 (i.e., A+D) divided by the total number of samples (i.e., A+B+C+D). These findings provide to resource managers information about the success rate at identifying correct cases (i.e., total algorithm accuracy at predicting correct field conditions).

6.5.3. Third statistical assessment

Determine how well the algorithm correctly identifies an observed condition (i.e., the absence or presence of K. brevis blooms), and how many times the algorithm failed. Table 6.4 explains these relationships taking into account the columns based on the in situ observation. A measure of how good is the algorithm outcome was determined by the number of times it correctly flagged red tide samples among the red tide population (i.e., sensitivity; Table 6.4), and by the number of times it correctly identified non-bloom samples among the non-bloom population (i.e., specificity; Table 6.4). The algorithm failure was measured when known red tide samples are not flagged (i.e., false negatives; Table 6.4), and if known non-bloom samples were classified as red tide (i.e., false positives; Table 6.4). It is good to bear in mind that sensitivity and specificity are independent variables, so are false negatives and positives. But sensitivity is coupled with false negatives, while specificity with false positives. This assessment sought to answer four questions:
What is the algorithm’s:

3.a: Success rate of correctly flagging known red tide samples? (Sensitivity)

3.b: Failure rate in flagging known red tide samples? (False negatives)

3.c: Success rate at correctly identifying known non-bloom samples? (Specificity)

3.d: Failure rate in identifying known non-bloom samples? (False positives)

<table>
<thead>
<tr>
<th>$3^{rd}$ statistical assessment: DETECTED CASES</th>
<th>In situ observation of the presence or absence of red tide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite-based algorithm outcome at detecting red tide</td>
<td>Red tide</td>
</tr>
<tr>
<td>Red tide</td>
<td>Sensitivity: $A/(A+C)$</td>
</tr>
<tr>
<td>Non-bloom</td>
<td>False negatives: $C/(A+C)$</td>
</tr>
<tr>
<td>SUM</td>
<td>$A+C$ (100%)</td>
</tr>
</tbody>
</table>

Information on success flagging among *a priori* known samples (*i.e.*, Questions 3.a and 3.c) could offer useful information to resource managers while developing mitigation action plans. On the other hand, the cases where the algorithms are inaccurate are not welcome for different reasons. Errors in flagging a red tide (*i.e.*, false negatives; Question 3.b) are not desirable from the resource managers point of view due to possible public health implications this outcome can cause; whereas the failure to identify non-blooms
(i.e., false positives; Question 3.d) are not financially desired, since no red tide will be present and an erroneous announcement may cause economic losses (e.g., due to loss of tourism or unnecessary closure of shellfish beds). However in a false positive case, undesirable public health risks will not follow, and this case is more acceptable in relation to a false negative case.

6.5.4. Fourth statistical assessment

Relied on the actual outcome of each algorithm (i.e., predictive values). Table 6.5 explains these relationships taking into account the rows based on the algorithms outcome at detecting red tide. It focused to answer four questions:

Of the samples:
4.a: Flagged by the algorithm as red tide, how many are actually red tide?
4.b: Flagged by the algorithm as red tide, how many are in fact non-bloom?
4.c: Not classified by the algorithm as a red tide, how many are in fact non-bloom?
4.d: Not classified by the algorithm as a red tide, how many are actually red tide?

The positive predictive value (i.e., Question 4.a) reflects the proportion of known red tide samples that are correctly flagged, and the higher this value, the more accurate is the algorithm at detecting a red tide sample. In the same manner, the higher the negative predictive value (i.e., Question 4.c), more non-bloom samples are correctly identified. In the opposite way, if taking into account the inverse of the predictive values, the larger accuracies are found with lower values. The inverse of the negative predictive value (i.e.,
Question 4.b) means that the red tide samples are not detected, and the inverse of the positive predictive value (i.e., Question 4.d) gives rise to non-bloom samples that were flagged as red tides.

Table 6.5: Predictive values.

<table>
<thead>
<tr>
<th>4th statistical assessment: PREDICTIVE VALUES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite-based algorithm outcome at detecting red tide</td>
<td>Red tide</td>
<td>Non-bloom</td>
</tr>
<tr>
<td>Red tide</td>
<td>Positive predictive value: $A/(A+B)$</td>
<td>Inverse of positive predictive value: $B/(A+B)$</td>
</tr>
<tr>
<td>Non-bloom</td>
<td>Inverse of negative predictive value: $C/(C+D)$</td>
<td>Negative predictive value: $D/(C+D)$</td>
</tr>
<tr>
<td></td>
<td>Question 4.a</td>
<td>Question 4.c</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Measures of predictive values are statistically important variables as it represents the probability of accurate detection (i.e., correct identification of red tides and non-blooms cases) among the “gold standard” conditions (i.e., observed presence or absence of \( K. \) brevis blooms). However, the predictive values strongly depend on the red tide prevalence. With this assessment, resource managers would be able to measure the
confidence level of the algorithm predictions. In other words, it would give them an idea of the algorithms’ detection accuracy.

6.5.5. Fifth statistical assessment

Measure the probability (Table 6.6) of each case: correct identification of non-blooms or red tides; and the chances of getting false positives or false negatives. The events are normalized in relation to the total population, and the sum of the 4 possible options (5.A: red tide correctly flagged; 5.B: read tide not correctly detected; 5.C: non-blooms correctly identified; and 5.D: non-blooms identified as red tides; Table 6.6) has to add up to 1. The outcomes of this assessment are not equally likely to occur if the in situ dataset has unequal numbers of samples for each subset (i.e., prevalence is not 50:50).

<table>
<thead>
<tr>
<th>5th statistical assessment: PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite-based algorithm outcome at detecting red tide</td>
<td>Red tide</td>
<td>Non-bloom</td>
</tr>
<tr>
<td>Red tide</td>
<td>5.A= A/(A+B+C+D)</td>
<td>5.B= B/(A+B+C+D)</td>
</tr>
<tr>
<td>Non-bloom</td>
<td>5.C= C/(A+B+C+D)</td>
<td>5.D= D/(A+B+C+D)</td>
</tr>
<tr>
<td>SUM</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Since the red tide location is in fact unknown and because the algorithms outcomes may not be clear-cut, the best approach to enhance the likelihood of delineating red tides using space-borne measurements is a converse interpretation of these statistical assessment. An algorithm to achieve a detection excellence must have two characteristics simultaneously. Firstly, it should address the problem, meaning correctly detect red tides and non-bloom (i.e., owing reasonable sensitivities and specificities, including acceptable false negatives and false positives). Secondly, to achieve user (e.g., resource manger, scientist, general public) confidence, it should have good predictive values. If one of these characteristics is attained but not the other, the assessment of the accuracy of the algorithm is handicapped to some extent, and its outcomes should be used with caution due to the likelihood of additional uncertainties.

6.6. Optimization of the algorithms

Following the statistical assessment, an empirical optimization procedure was performed to increase the confidence of the detection of *K. brevis* blooms using space-borne measurements. This optimization was only performed on the Bio-optical and Empirical algorithms and analyzed with the in situ dataset divided into two subsets based on 15,000 cells l⁻¹ for red tide and non-bloom separation (Table 6.2), and aimed to reach satisfactory results of correct red tide detection (i.e., sensitivity) simultaneously with correct non-bloom identification (i.e., specificity).

Given that each algorithm has a particular set of remote-sensed parameters (e.g., Bio-optical: Chl and b_{bp(λ)}; Empirical: Chl and L_{w(λ)}) used along with specific calculated threshold slope lines (Morel, 1988) to flag the presence (i.e., red tide) or absence (i.e.,
non-bloom) of *K. brevis* blooms, the identification success rate changes by adjusting the thresholds. Therefore, the optimization involved multiplying the slope lines by a factor varying from 0.33 to 3.0 (as depicted on Figure 6.5). The slope lines were initially set low (when multiplied by 0.33), and subsequently underwent a 0.01 step-by-step incremental increase up to and including the higher factor of 3.0. This increase included its original stage (i.e., with the factor equal to 1.0).

Figure 6.5: Threshold lines used to optimize the Bio-optical technique (left) and Empirical approach (right). Dashed lines are the original threshold slope lines (Morel, 1988). Dotted lines are the thresholds slope lines (Morel, 1988) multiplied by factors of 0.33 (lower) and 3.0 (upper). On the left plot, the vertical dashed line represents 1.5 mg m$^{-3}$ of chlorophyll. Flagged red tides are in the lower right quadrant (Bio-optical technique), and under the line (Empirical approach).

At each increment stage, the red tide and non-bloom success identification rate was recalculated and recorded. The stage having the maximum sensitivity along with the maximum specificity was set to establish the new separation level deciding if samples
were to be classified as red tides or as non-blooms. Therefore, the optimization determined the best threshold slope lines (Morel, 1988). Subsequently, the statistical assessments were computed again determining more equilibrated and reasonable results. This recalculation of the statistics gave an optimized likelihood of correctly identifying red tides and non-bloom situations.

6.7. The combined Hybrid scheme

Following the empirical optimization procedure, the detection of *K. brevis* blooms was evaluated using the optimized algorithms together analyzed with the *in situ* dataset divided into two subsets based on 15,000 cells l\(^{-1}\) for red tide and non-bloom separation (Table 6.2). This combined algorithm (hereafter referred to as Hybrid scheme) tried to increase the confidence at detecting the blooms of *K. brevis* by establishing a simultaneous detection scheme based on the abilities of the optimized algorithms. This Hybrid scheme was tested at two levels. Firstly it focused on a higher predictability scheme; and secondly it targeted a higher confidence scheme, as follows:

- The sensitivity of the 1\(^{st}\) Hybrid scheme (i.e., high predictability; Table 6.7) was determined by the cases where one or both algorithms correctly classified a known red tide sample; on the other hand, it recognized a case of false negative only if both algorithms failed to classify a known red tide sample as a red tide. While the specificity of the 1\(^{st}\) Hybrid scheme was determined by the cases where one or both algorithms classified a known non-bloom sample as non-bloom; on the other hand, it recognized a
case of false positives only if both algorithms failed to classify a known non-bloom sample as a non-bloom.

- The sensitivity of the 2nd Hybrid scheme (i.e., high confidence; Table 6.8) was determined only by the cases where both algorithms correctly classified a known red tide sample; on the other hand, it recognized a case of false negative if one or both algorithms failed to classify a known red tide sample as a red tide. While the specificity of the 2nd Hybrid scheme was determined only by the cases where both algorithms correctly classified a known non-bloom situation; on the other hand, it recognized a case of false positives if one or both algorithms failed to classify a known non-bloom sample as a non-bloom.

Table 6.7: 1st Hybrid scheme (i.e., high confidence).

<table>
<thead>
<tr>
<th>1ST HYBRID SCHEME: HIGH PREDICTIBILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Red tide</strong></td>
</tr>
<tr>
<td>Known red tide sample is DETECTED by BOTH OR ONLY by one algorithm</td>
</tr>
<tr>
<td><strong>False negative</strong></td>
</tr>
<tr>
<td>ONLY if a known red tide sample FAILS to be detected by BOTH algorithms</td>
</tr>
</tbody>
</table>

| **Non-bloom**                          |
| Known non-bloom sample is DETECTED by BOTH OR ONLY by one algorithm |
| **False positive**                     |
| ONLY if a known non-bloom sample FAILS to be detected by BOTH algorithms |

The same statistical assessments of section 6.5 were computed for both confidence levels of the Hybrid scheme providing information of its ability to detect blooms of *K. brevis*. 


Table 6.8: 2\textsuperscript{nd} Hybrid scheme (i.e., high confidence).

<table>
<thead>
<tr>
<th>2\textsuperscript{nd} HYBRID SCHEME: HIGH CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Red tide</strong></td>
</tr>
<tr>
<td>False negative</td>
</tr>
<tr>
<td><strong>Non-bloom</strong></td>
</tr>
<tr>
<td>False positive</td>
</tr>
</tbody>
</table>
CHAPTER 7

RESULTS

The results obtained by this investigation are quantified in Chapter 7. A preliminary comparison among the algorithms and the ways to reach the valid pairs of \textit{in situ}-satellite match-ups open this Chapter. The spatio-temporal distribution of the valid match-ups is shows. Statistical assessments are performed to verify the accuracy of the selected algorithms, followed by the results of the optimizations and the combined hybridized scheme.

7.1. Satellite imagery

The published results of Cannizzaro (2004) and Cannizzaro et al. (2008) were satisfactorily replicated using the two algorithms (\textit{i.e.}, Bio-optical technique and the Empirical approach) on data from both SeaWiFS and MODIS-Terra. Figure 7.1 shows a similar Chl patterns on data from both sensors. \textit{In situ} samples have confirmed elevated concentration of \textit{K. brevis} cells, and Figure 7.2 depicts the special coherence of the red tide envelopes from the study area outlined in Figure 7.1. The spatial coherence produced by both algorithms was similarly and satisfactorily comparable. The two algorithms (\textit{Bio-optical}: Figure 7.2 - top; and \textit{Empirical}: Figure 7.2 - bottom) almost equally flagged the red tide on data derived from both sensors (SeaWiFS; Figure 7.2 - left; and MODIS-Terra; Figure 7.2 - right). Most red tide match-ups fall inside the shaded envelopes.
created by the algorithms, and the majority of the non-blooms match-ups (X) are outside the envelopes (Figure 7.2).

Figure 7.1: Chlorophyll concentrations (mg m$^{-3}$) from SeaWiFS (left) and MODIS-Terra (right). Outlined area show the regions where the algorithms were applied (Figure 7.2).

Table 7.1 summarizes the statistical results of the Bio-optical technique and Empirical approach, and also shows the observed false positive and false negative cases. Figure 7.3 displays the relationship between $b_{bp(\lambda)}$ versus Chl, which is the basis of the Bio-optical technique: 70% of the red tides were flagged in the SeaWiFS data, and 100% in the MODIS-Terra data. The Bio-optical non-bloom detection in the SeaWiFS data was 94%, and 70% in the MODIS-Terra data. Figure 7.4 presents the relationship between $L_{w(\lambda)}$ versus Chl, which is the basis of the Empirical approach: both sensors had similar accuracy (~95%) for red tide identification and about 85% for non-bloom cases.
Figure 7.2: Red tide envelopes (white) from the area outlined in Figure 7.1.

*Bio-optical* technique (top); *Empirical approach* (bottom).

SeaWiFS (August 30, 2001; left); MODIS-Terra (August 29, 2001; right).

*In situ* data from August 25 to September 2, 2001 is overlaid.

Non-bloom: green X; red tide: red □; respectively ≤ and > $10^4$ cells l$^{-1}$ of *K. brevis*. 
Figure 7.3: The Bio-optical technique classifies everything to the right of the vertical dotted line (···· which is 1.5 mg m$^{-3}$ of Chl) and below the dashed line (--- which is Morel (1988) $b_{bp(550)}$) as red tides. In situ data from August 25 to September 2, 2001 are overlaid (non-bloom: $\times$; red tide: $\square$; respectively $\leq$ and $> 10^4$ cells l$^{-1}$ of $K. brevis$).

SeaWiFS (August 30, 2001; left) and MODIS-Terra (August 29, 2001; right).

Figure 7.4: The Empirical approach classifies everything below the dotted line (···· which is Morel (1988) $b_{b(550)}$) as red tides. In situ data from August 25 to September 2, 2001 are overlaid (non-bloom: $\times$; red tide: $\square$; respectively $\leq$ and $> 10^4$ cells l$^{-1}$ of $K. brevis$).

SeaWiFS (August 30, 2001; left) and MODIS-Terra (August 29, 2001; right).
Table 7.1: Detection of *Karenia brevis* blooms in the CWFS using different algorithms (i.e., the *Bio-optical* technique and *Empirical* approach) on one SeaWiFS image (August 30, 2001) and on one MODIS-Terra image (August 29, 2001).

Denominators are valid match-ups; Numerator is the total population of samples; Percentages of occurrence are in between parenthesis.

**Red tide (in red), false negative (in gray), non-bloom (in green), false positive (in blue).**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Bio-Optical technique</th>
<th>Empirical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite</td>
<td>Sensor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SeaWiFS</td>
<td>MODIS-Terra</td>
</tr>
<tr>
<td>Red tide</td>
<td>14/20</td>
<td>3/50</td>
</tr>
<tr>
<td></td>
<td>(70%)</td>
<td>(6%)</td>
</tr>
<tr>
<td>Non bloom</td>
<td>6/20</td>
<td>47/50</td>
</tr>
<tr>
<td></td>
<td>(30%)</td>
<td>(94%)</td>
</tr>
<tr>
<td>Red tide</td>
<td>19/20</td>
<td>7/50</td>
</tr>
<tr>
<td></td>
<td>(95%)</td>
<td>(14%)</td>
</tr>
<tr>
<td>Non bloom</td>
<td>1/20</td>
<td>43/50</td>
</tr>
<tr>
<td></td>
<td>(5%)</td>
<td>(86%)</td>
</tr>
</tbody>
</table>

The Carder semi-analytic algorithm (Carder et al., 1999) was originally used to calculate $b_{bp}(\lambda)$ and Chl, but not many valid retrievals were produced for the $b_{bp}(\lambda)$ in the MODIS-Terra image, probably due to the fact that MODIS-Terra is not sufficiently well calibrated for ocean color measurements (Franz et al., 2000; 2005; Barnes et al., 2003). Thus, in order to compare both algorithms and both sensors, the $b_{bp}(\lambda)$ from Garver-Siegel-Maritorena (GSM01; Maritorena et al., 2002) was used instead. But when the Carder algorithm was used to derive $b_{bp}(\lambda)$ on the SeaWiFS image, the successful red tide detection increased from 70% to 90%, but the non-bloom detection slightly decreases (from 94% to 88%). This improvement on the results of SeaWiFS application occurred because the $b_{bp}(\lambda)$ retrievals of the Carder algorithm were smaller than the $b_{bp}(\lambda)$ GSM01,
and the basis of the *Bio-optical* technique to flag red tides is in fact to have low $b_{bp(\lambda)}$.

Thus, since the *Bio-optical* technique already had 100% confidence on MODIS-Terra data while using the $b_{bp(\lambda)}$ GSM01, it is probable that if $b_{bp(\lambda)}$ derived with the Carder algorithm were retrieved, and lower, the results would have been the same or even better. On the other hand, the non-bloom detection could be reduced for the same reason.

### 7.2. *In situ* dataset

An analysis throughout the *in situ* sampling distribution of the original historical *in situ* database (1956~2006, containing ~76000 entries) provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute, showed an increase in the number of stations in the course of the last decade, especially in the last 2 years (Figure 7.5). This sampling distribution indeed influenced the number of valid match-ups.

Not all stations (~21,000 samples) spanning the analyzed MODIS-Aqua lifetime mission period (Table 7.2: Column A) could be used in this study. A sizeable number of these records (~50%) were eliminated because the present investigation only considered *Karenia brevis* cell counts. This elimination clearly showed the strong *K. brevis* dominance among the other congeneric *Karenia* found in the Gulf of Mexico. Due to the use of the ETOPO2 dataset to provide site bathymetry, another fraction of stations (~800) also was not used.

Further data exclusions occurred by applying two masks (see 6.3). All stations located within buffer zone around the Florida coast of half of the MODIS pixel size (*i.e.*, 500m) were eliminated (numbering ~700). About 500 samples were discarded because
they were close to the bottom (ETOPO2 mask). Additional samples were removed during the match-up analysis. These findings, together with the temporal distribution of the workable dataset are explained in the next section.

Figure 7.5: Distribution of the original historical *in situ* sampling effort collected along the entire Florida coast showing increasing number of samples.
7.3. Match-up analysis

Since a relaxed temporal match-up window of 24h was used to collocate field data with satellite measurements, the first step of the match-up analysis was to verify that there was a satellite scene on the same day of the in situ sample. Thus, 7929 samples had a satellite image on the same day (as shown in column B of Table 7.2). However, 1783 in situ stations did not have a corresponding satellite pixel on the same day, and the total in situ samples within satellite scenes were then reduced by 22.5%, leaving the workable dataset with 6146 samples (Table 7.2: Column C). Figure 7.6 and Figure 7.7 illustrate the temporal distribution of the workable in situ dataset. This workable dataset was comprised of 44.7% (n=744; Table 7.2: Column D) of the initial 1665 MODIS-Aqua images with appropriate information for the match-up analysis during the almost 4 years of this study.

The next step was to establish pixel quality. If the pixel was masked as cloud, it was eliminated. A light penetration depth mask and the additional filters eliminated many cloud-free pixels. The bottom reflection threshold cautiously avoided accounting for this effect and found no argument for cutting off the analysis shallower water valid match-up pairs (i.e., >10 m). The satellite zenith angle threshold tests did not produce an argument for cutting off data between 45° and 60°, because the inclusion of valid match-ups with angles greater than 45° did not cause a deterioration of the results. Thus, these two alternative thresholds tests increased special coverage.

In the end, only ~13% the workable dataset provided acceptable samples (n=778; Table 7.2: Columns E, F). These acceptable samples (i.e., valid match-ups pairs) provided the basis used to train the satellite-based algorithms at detecting K. brevis
blooms. However, the procedure of the Operational method required a complex chlorophyll anomaly calculation (see 4.1 and 6.1), which also eliminated additional data from the valid match-up pairs, leaving only 680 samples to use with this algorithm.

The region with the highest historical occurrence of observed monospecific K. brevis blooms (i.e., the Central West Florida Shelf; CWFS) contained ~68% of the total valid in situ-satellite match-ups pairs (n=529; Table 7.2: Column G). Taking cells count threshold of 15,000 cells l⁻¹ to separate into non-bloom from red tide cases (Table 6.2), about 2/3 of the samples found in the CWFS were non-blooms, while the other 1/3 was represented by red tides samples (Table 7.2: Columns H, I).

The great majority (~70%) of CWFS valid match-ups came from the blooming season, i.e., the Summer-Fall (n=371; Figure 7.7; Figure 7.8; Figure 7.9; Table 7.2: Columns J, K). This region and period contained ~65% of the non-bloom valid match-ups, and almost 80% of the red tide cases. This concentration of red tide valid match-ups in the Summer-Fall is in accordance with other related literature (e.g., Walsh and Steidinger, 2001).

Even though the workable in situ samples appeared to be well distributed over time (Figure 7.6 and Figure 7.7), temporal biases indeed appeared after the definition of the valid match-up pairs containing samples only from the CWFS (e.g., Figure Figure 7.8 and Figure 7.9).
Table 7.2: This table is divided in two blocks, but the column sequence of the blocks is a continuous. Yearly frequency distribution of the *in situ* sampling effort provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute and the correlated number of valid match-ups. Here, the limit of 15,000 cell l\(^{-1}\) was used to divide into *non-bloom* and *red tide* samples.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original in situ sampling effort</td>
<td>Workable in situ samples</td>
<td>Number of satellite scenes used on the match-up analysis</td>
<td>Valid match-ups on the entire West Florida coast</td>
<td>Percentages of valid match-ups on the entire West Florida coast</td>
</tr>
<tr>
<td></td>
<td>2002 (*)</td>
<td>1905 (959)</td>
<td>624 *</td>
<td>81</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>1916</td>
<td>1448</td>
<td>183</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>1153</td>
<td>953</td>
<td>113</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>1880</td>
<td>1500</td>
<td>180</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>2021</td>
<td>1621</td>
<td>182</td>
<td>181</td>
</tr>
<tr>
<td>Total (*)</td>
<td>8875 (7929)</td>
<td>6146</td>
<td>744</td>
<td>778</td>
<td>12.7%</td>
</tr>
</tbody>
</table>

*After June = C/(B) = D/1665 = E/C

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valid match-ups on the Central West Florida Shelf</td>
<td>Valid match-ups from throughout the whole year in the CWFS</td>
<td>CWFS Summer-Fall valid match-ups</td>
<td>NON BLOOM</td>
<td>RED TIDE</td>
</tr>
<tr>
<td></td>
<td>NON BLOOM</td>
<td>RED TIDE</td>
<td>NON BLOOM</td>
<td>RED TIDE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002 (*)</td>
<td>60</td>
<td>55</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>112</td>
<td>80</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>94</td>
<td>83</td>
<td>11</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>147</td>
<td>51</td>
<td>96</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>116</td>
<td>78</td>
<td>38</td>
<td>33</td>
</tr>
<tr>
<td>Total (*)</td>
<td>529</td>
<td>347</td>
<td>182</td>
<td>226</td>
<td>145</td>
</tr>
</tbody>
</table>

*After June = G/E = H/G = I/G = J/H = K/I

67.9% | 65.3% | 34.4% | 65.1% | 79.7%
Figure 7.6 presents histograms illustrating the temporal distribution of the 7929 workable *in situ* samples matched within satellite scenes spread throughout the entire West Florida coast. No data are shown before June 2002 or after October 2006, because this was the start of the MODIS-Aqua mission, and the end of data used in this investigation. It is evident that the greater concentration of samples occurred during the Summer-Fall, which is usually the blooming-season. Figure 7.7 depicts the same temporal distribution showing the distribution of the sampling effort, but on a daily basis.

**Figure 7.6:** Temporal frequency distribution of the workable *in situ* dataset provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute that were within satellite scenes. Winter-Spring (blue; left); Summer-Fall (orange; right)
Figure 7.7: Daily frequency distribution of the workable *in situ* dataset provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute that were within satellite scenes.

Winter-Spring (blue; left); Summer-Fall (orange; right)
Figure 7.8 depicts the temporal distribution of the valid \textit{in situ}-satellite red tide match-ups pairs from the CWFS. During the Summer-Fall (Figure 7.8; left column), the prominent 2005 concentration of cloud-free pixels with red tides was conspicuous. This represents more than 60% of the samples (89 of 145) coming from 17 satellite images from different days (Table 7.3). The other \sim10\% of red tides in the CWFS came from 2002, 2003 and 2004 together. The Winter-Spring (Figure 7.8; right column) had a low incidence of valid red tide match-ups: only 37 were registered over the entire period of this analysis.

For the non-bloom cases in the CWFS, 65\% of the valid match-ups were within the Summer-Fall. During this season, each of the analyzed years had an average of \sim20\% of the valid match-ups (Figure 7.9; left column), accounting for almost 70 satellite images, averaging 20\% per year (Table 7.3).

Table 7.3: Number of days (images) with valid match-ups in the Central West Florida Shelf.

<table>
<thead>
<tr>
<th></th>
<th>Season</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red tide</td>
<td>Winter-Spring</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Summer-Fall</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>17</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>23</td>
<td>13</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Season</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-bloom</td>
<td>Winter-Spring</td>
<td>1</td>
<td>16</td>
<td>12</td>
<td>6</td>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Summer-Fall</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>19</td>
<td>69</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>15</td>
<td>28</td>
<td>25</td>
<td>17</td>
<td>35</td>
<td>120</td>
</tr>
</tbody>
</table>
Figure 7.8: Daily frequency distribution of the **RED TIDE in situ** match-ups in the Central west Florida Shelf. Winter-Spring (blue; left); Summer-Fall (orange; right)
Figure 7.9: Daily frequency distribution of the **NON-BLOOM in situ** match-ups in the Central west Florida Shelf. Winter-Spring (blue; left); Summer-Fall (orange; right)
7.4. Spatio-temporal investigation

The selected algorithms (i.e., Operational; Bio-optical and Empirical) were then spatially and temporally tested using the 778 valid match-up pairs that were left after the quality control of the in situ data outlined in the previous two sections. The spatio-temporal results presented in this section separated samples into two subsets (i.e., non-blooms and red tides) taking 15,000 cells l\(^{-1}\) as a threshold (Table 6.2) for the three algorithms. Although, the Operational method results are also shown using its original threshold (100,000 cells l\(^{-1}\)).

The first step of this investigation included all valid match-up pairs (i.e., year-round samples collected all along the West Florida coast), and demonstrated that the three algorithms had success identification rate of non-bloom >70% (Figure 7.10 left). However, the Empirical and Bio-optical algorithms had moderate success identification rates of red tide: ~61% (Figure 7.10 bottom right) and ~53% (2\(^{nd}\) row from the bottom right in Figure 7.10), respectively. The Operational method had very reduced red tide detection accuracy with both thresholds: ~35% (1\(^{st}\) and second rows from top right in Figure 7.10).

The second step in this investigation looked into the seasonal success rate over the entire West Florida coast during the Winter-Spring. All three algorithms had non-bloom detection exceptionally high (>90%; Figure 7.11 left), although the very few (n~40) red tide valid match-ups were poorly identified (<20%; Figure 7.11 right). Using the original threshold of Operational method, this reduced even more the number of red tide samples (n=18), which limits conclusions to be made with such small number of samples (Figure 7.11 top right).
Figure 7.10: Valid match-ups sampled year-round along the entire West Florida coast. *Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom). Non-bloom (left) and red tide (right).
Figure 7.11: Valid match-ups from the entire West Florida coast in the Winter-Spring. 
Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), 
Bio-optical (2nd row from bottom) and Empirical (bottom). 
Non-bloom (left) and red tide (right).
Figure 7.12: Valid match-ups from the entire West Florida coast in the Summer-Fall. 
*Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), 
*Bio-optical* (2nd row from bottom) and *Empirical* (bottom). 
Non-bloom (left) and red tide (right).
The third step in this investigation looked into the seasonal success rate over the entire West Florida coast during the Summer-Fall. For non-bloom samples, a detection rate of >60% for the three algorithms was found (Figure 7.12 left). The red tide success identification rates (Figure 7.12 right) were alike those for Figure 7.10 (right), but slightly higher for all algorithms.

Summarizing these first three steps of this spatio-temporal investigation containing data from the entire West Florida coastline: none of the three algorithms had good accuracy for detecting red tides in the Winter-Spring, but the Bio-optical and the Empirical algorithm had moderate red tide identification during the Summer-Fall. Independent of the period of the year of the count threshold utilized, the Operational method could not adequately flag red tides.

After analyzing the valid match-ups for the entire West Florida coast, the next step of the spatio-temporal investigation was to seasonally take into account data from two regions: North-FL and South-FL.

Very few non-bloom valid match-ups was registered for the North-FL during the Winter-Spring, constraining any conclusion to be drawn (Figure 7.13 left), and not a single sample with red tide came out of the match-up analysis within this region-period (Figure 7.13 right).

In the South-FL during the Winter-Spring, the Bio-optical and Empirical algorithms detected 100% of the non-bloom samples, while the Operational detected ~85% for both thresholds (Figure 7.14 left). An extraordinarily small number of valid red tide match-ups (n=7) appeared in this region-period (Figure 7.14 right), and only two samples had more than 100,000 cell l⁻¹, preventing any conclusion from being made.
Figure 7.13: Valid match-ups from the Winter-Spring in the North of Florida. 
Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), 
Bio-optical (2nd row from bottom) and Empirical (bottom). 
Non-bloom (left) and red tide (right).
Figure 7.14: Valid match-ups from the Winter-Spring in the South of Florida. 
Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), 
Bio-optical (2nd row from bottom) and Empirical (bottom). 
Non-bloom (left) and red tide (right).
Figure 7.15: Valid match-ups from the Summer-Fall in the North of Florida. Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom). Non-bloom (left) and red tide (right).
Figure 7.16: Valid match-ups from the **Summer-Fall** in the **South** of Florida. 
*Operational* (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), *Bio-optical* (2nd row from bottom) and *Empirical* (bottom).
Non-bloom (left) and red tide (right).
During the Summer-Fall in the North-FL, the non-blooms were identified by the *Empirical* only 50% of the time, by *Bio-optical* about 83%, and no matter which threshold was used, the *Operational* detected almost 80% of the non-bloom samples (Figure 7.15 left). A small number of red tide points restrict any conclusion for this region and period (Figure 7.15 right).

All algorithms had a correct detection rate above 80% for non-bloom match-ups of the South-FL in the Summer-Fall, except for the *Operational* with its original threshold that detected ~64% (Figure 7.16 left). The *Bio-optical* and *Empirical* algorithms completely failed to detect red tide samples in this region and period (<15% for both), but even though there was a small number of red tide samples (n~30), the *Operational* had a red tide detection rate of about 67% (Figure 7.16 right). The same rate of identification was observed for use of the original threshold *Operational* method, but with even less samples (n=9).

After analyzing the valid match-ups for the North-FL and South-FL regions, the next step of the spatio-temporal investigation was to seasonally take into account data from the Central West Florida Shelf (CWFS).

The analysis of the valid match-up pairs of the CWFS during the Winter-Spring shows an excellent non-bloom success rate (>90%) for all algorithms (Figure 7.17 left). However, all three algorithms break down while detecting red tides in this region and period (Figure 7.17 right).
Figure 7.17: Valid match-ups sampled during the Winter-Spring in the Central West Florida Shelf (CWFS). Non-bloom (left) and red tide (right).

Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom).
Figure 7.18: Valid match-ups sampled during the Summer-Fall in the Central West Florida Shelf (CWFS). Non-bloom (left) and red tide (right).

Operational (top row: 100,000 cells/liter; and 2nd row from top: 15,000 cells/liter), Bio-optical (2nd row from bottom) and Empirical (bottom).
The scatter-plots containing pairs of non-bloom match-ups from the CWFS during the Summer-Fall illustrate that the *Operational* method had the best success identification rate of non-bloom samples (~84%) independent of the threshold used. The *Bio-optical* technique detected about 65% of the non-blooms, and the *Empirical* approach ~52% (Figure 7.18 left). For the red tide samples within this region and period the *Empirical* approach had the best success identification rate (~85%), then the *Bio-optical* had ~76%, and the *Operational* a low accuracy of only ~26% (Figure 7.18 right) independent of the threshold used.

**7.5. Detection accuracy assessment**

The spatio-temporal results of the preceding section showed that the valid match-ups could only supply sufficient information about the Central West Florida Shelf (CWFS) during the blooming-season (*i.e.*, Summer-Fall). Thus, the performances of the statistical assessments showed in this section were solely applied to this region and period. The non-bloom and red tide threshold of 15,000 cells l⁻¹ remain to be used (Table 6.2). An examination of Figure 7.18 provides sufficient information to answer the questions of the five statistical assessments: 1) Prevalence and odds of red tides; 2) Overall accuracy; 3) Specificity, sensitivity, false negatives and false positives (Table 7.4); 4) Predictive values (Table 7.5); and 5) Probabilities (Table 7.6).

**7.5.1. First statistical assessment**

Even though the number of total valid match-ups of the *Operational* method (n=328) was smaller than to the other two algorithms (n=371), their red tide prevalence
was about the same: ~40% (Operational: 130/328; Bio-optical and Empirical: 145/371). However, the odds of a red tide were somewhat different: 78% (130/167) for the Operational, and 64% (145/226) for the others.

### 7.5.2. Second statistical assessment

The success rate at identifying correct cases (red tides plus non-blooms together) of the three algorithms was quite similar. The overall accuracy of the Operational is ~61% of (201 correct samples of a total of 328); the Bio-optical technique ~70% (258 correct samples of a total of 371), and the Empirical approach ~65% (243 correct samples of a total of 371).

### 7.5.3. Third statistical assessment (Table 7.4)

Focusing on the sensitivity of the algorithms, the Empirical approach had the best red tide detection rate (124 of 145 samples; 85.5%), followed by the Bio-optical with 76.5% (111 of 145 samples), while the Operational method had an extremely low success rate of correct red tide identification (34 of 130 samples; 26%).

The false negatives of the Empirical approach were very low, only 14.5% (24 of 145 samples). The Bio-optical also exhibited a low rate (34 of 145 samples; 23.5%). The Operational method provided a high number of false negatives (96 of 130 samples; 74%).

On the other hand, focusing on the non-bloom match-up pairs, the Operational method had the highest specificity (167 of 198; 84%), which generated a very low assemblage of false positives (31 of 198 samples; 16%). The Bio-optical technique
Specificity was 65% (147 of 226 samples) and false positives were 35% (79 of 226 samples). The lowest observed specificity was for the *Empirical* approach, which had almost 50:50 of correct detection (119 of 226 samples) and false positives (107 of 226 samples).

<table>
<thead>
<tr>
<th><strong>Operational method</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DETECTED CASES</td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td><strong>Operational method</strong></td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td>SUM</td>
<td><strong>130 (100%)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Bio-optical technique</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DETECTED CASES</td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td><strong>Bio-optical technique</strong></td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td>SUM</td>
<td><strong>145 (100%)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Empirical approach</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DETECTED CASES</td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td><strong>Empirical approach</strong></td>
<td><strong>True (Red tide)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td>SUM</td>
<td><strong>145 (100%)</strong></td>
</tr>
</tbody>
</table>
7.5.4. Fourth statistical assessment (Table 7.5)

Despite the reassuring sensitivity and specificity results, the positive predictive values of the algorithms were moderate and about the same (~55%), but the negative predictive value of the *Operational* method was the lowest (~60%), while the other two algorithms ranged ~80%.

### Table 7.5: Predictive values (Original algorithms)

<table>
<thead>
<tr>
<th>Method</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>False (Non-bloom)</td>
<td></td>
</tr>
<tr>
<td><strong>Operational method</strong></td>
<td>34 (52%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31 (48%)</td>
<td>65 (100%)</td>
</tr>
<tr>
<td><strong>Operational method</strong></td>
<td>96 (36.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>167 (63.5%)</td>
<td>263 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>False (Non-bloom)</td>
<td></td>
</tr>
<tr>
<td><strong>Bio-optical technique</strong></td>
<td>111 (58.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>79 (41.5%)</td>
<td>190 (100%)</td>
</tr>
<tr>
<td><strong>Bio-optical technique</strong></td>
<td>34 (19%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>147 (81%)</td>
<td>181 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>False (Non-bloom)</td>
<td></td>
</tr>
<tr>
<td><strong>Empirical approach</strong></td>
<td>124 (54%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>107 (46%)</td>
<td>231 (100%)</td>
</tr>
<tr>
<td><strong>Empirical approach</strong></td>
<td>21 (15%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>119 (85%)</td>
<td>140 (100%)</td>
</tr>
</tbody>
</table>
Since the inverse of the positive predictive value is couple with the positive predictive value, it was the about the same (~45%). The inverse of the negative predictive value of the *Operational* method was the highest (~40%), while the other two algorithms ranged ~20%.

### 7.5.5. Fifth statistical assessment (Table 7.6)

The likelihood of an event occurring was also extracted from an examination of Figure 7.18. The probability of detecting red tides is about the same (~30%) for the *Bio-optical* and the *Empirical* algorithms, while the *Operational* method has only ~10% of chance of detecting a red tide. The probability of misclassifying a red tide sample (i.e., false negatives) is about 30% for the *Operational*, while a lower incidence is founded for the *Bio-optical* (~10%), and an even lower inaccuracy at detecting red tide samples is exhibited by the *Empirical* (~5%).

The likelihood of having non-bloom samples correctly flagged is ~50% for the *Operational*, ~40% for the *Bio-optical* and ~30% for the *Empirical*. The *Operational* method misclassifies non-bloom cases (i.e., false positives) about 10% of the times; the *Bio-optical* ~20%; and the *Empirical* ~30%.

### 7.6. Satellite-based algorithms optimization

The optimization of both algorithms (i.e., *Bio-optical* and *Empirical*) created new red tide and non-bloom success identification rates when the calculated threshold slope lines (Morel, 1988) were multiplied by factors varying from 0.33 to 3.0. The percentage of correct identification of non-bloom samples is high when the factors are set low
(Figure 7.19 top row). However, the lower the factor is, the worse is the percentage of correct identification of red tide samples (Figure 7.19 mid row). As expressed by the top row of Figure 7.20 (Bio-optical) and Figure 7.21 (Empirical), both algorithms successfully detect more than 90% of non-blooms, but only about 5% of the red tide samples were accurately detected.

<table>
<thead>
<tr>
<th>Operational method PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operational method results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.292</td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bio-optical technique PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bio-optical technique results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Empirical approach PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Empirical approach results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
As the factors are incremented increased, the accuracy of detecting non-bloom decreases, while the red tide increases (Figure 7.19). With the highest factor (3.0), the non-bloom accuracy was degraded (Figure 7.19 top row), but a great number of red tide samples were correctly identified (Figure 7.19 middle row). Figure 7.20 (second row the from top) depicts this case for the Bio-optical technique where only ~42% of non-bloom were accurately detected, and up to ~86% of the red tides were detected. While Figure 7.21 (second row the from top) shows the same pattern but for the Empirical method, where only ~10% of non-bloom were detected and 100% of the red tide were flagged.

The vertical dashed lines on Figure 7.19 (top and mid rows) are the factor 1.0 and represent the original calculated thresholds slope lines (Morel, 1988). The indicated percentages on these plots (horizontal dotted lines) are the original success identification rate of each algorithm in detecting non-bloom and red tide. The original detection is also shown in the third row from the top of Figure 7.20 (Bio-optical) and Figure 7.21 (Empirical), which are the same of those used on the trial steps of section 7.5.

The factor where the sensitivity and specificity were concurrently maximized, and the number of false negatives and false positives were minimized, provided the best-optimized stage (vertical solid lines on plot of the bottom row of Figure 7.19). For the Bio-optical technique the optimal factor encountered was 0.87, and for the Empirical method, 0.82. The accuracy for both algorithms at this optimal stage was ~70% (horizontal dotted lines on plot of the bottom row of Figure 7.19). The new separation level deciding if samples are to be classified as red tides or as non-blooms are shown as the solid lines in the plots of the bottom row of Figure 7.20 and Figure 7.21. This level is the original threshold slope line (Morel, 1988) multiplied by the best-optimized factor.
Figure 7.19: Percentage of correct identification as a function of Morel (1988) threshold slope lines multiplied by factors from 0.33 to 3.0. Red tide (top); non-bloom (center); Optimized (bottom). Bio-optical: (left); Empirical: (right). Vertical dashed line is the factor 1.0 leading to the original Morel (1988) threshold slope line. Vertical dashed lines are the best-optimized factors Long horizontal dotted line is the expected 70% optimized percentage of correct identification.
Figure 7.20: Optimization results for the Bio-optical technique.
Figure 7.21: Optimization results for the *Empirical* approach.
Figure 7.20 and Figure 7.21 (bottom and third row from the top) show the increase in the non-bloom detection accuracy cause by the optimization procedure. These figures also show the decrease in the accuracy of correct identifying red tide samples.

### 7.6.1. Re-computation of the statistical assessments after the optimization

The optimized accuracies of the algorithms were used to compute the statistical assessments once again. From Figure 7.20 and Figure 7.21 the success detection showed that the optimized sensitivity and specificity of both optimized algorithms was ~70%, generating false negatives and false positives ~30% (Table 7.7).

Table 7.7: Sensitivity, specificity, false negatives, false positives

( Optimized algorithms)

<table>
<thead>
<tr>
<th>Bio-optical technique DETECTED CASES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>False (Non-bloom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-optical technique results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>104 (71.7%)</td>
<td>64 (28.3%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>41 (28.3%)</td>
<td>162 (71.7%)</td>
</tr>
<tr>
<td>SUM</td>
<td>145 (100%)</td>
<td>226 (100%)</td>
</tr>
<tr>
<td>Empirical approach DETECTED CASES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical approach results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>102 (70.3%)</td>
<td>74 (32.7%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>43 (29.7%)</td>
<td>152 (67.3%)</td>
</tr>
<tr>
<td>SUM</td>
<td>145 (100%)</td>
<td>226 (100%)</td>
</tr>
</tbody>
</table>
Both algorithms had in the second statistical assessments about 70% of success at identifying correct cases. The *Bio-optical* technique flagged correctly 266 samples of a total of 371 (71%), while the *Empirical* method had a total correct rate of ~68% (254 correct samples of a total of 371).

The optimizations did not change their red tide prevalence or the odds of a red tide since the number of red tide and non-bloom samples was defined *a priori*, and had nothing to do with the change in the detectable characteristics of the algorithms.

Table 7.8 presents the results of optimized predictive values, which are all equivalent for both algorithms. Their positive predictive values are about the same (~60%), so is the inverse of their positive predictive values (~40%); while the negative predictive values are also similar (~80%), and the inverse of the negative predictive values are still low (~20%).

<table>
<thead>
<tr>
<th><strong>Bio-optical technique</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
<th><strong>SUM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PREDICTIVE VALUES</strong></td>
<td><strong>True (Red tide)</strong></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td>Bio-optical technique results</td>
<td>104 (62%)</td>
<td>64 (38%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>41 (20%)</td>
<td>162 (80%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Empirical approach</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
<th><strong>SUM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PREDICTIVE VALUES</strong></td>
<td><strong>True (Red tide)</strong></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td>Empirical approach results</td>
<td>102 (58%)</td>
<td>74 (42%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>43 (22%)</td>
<td>152 (78%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The probability (Table 7.9) of detecting red tides after optimized was almost (~30%) for both algorithms, and their probability of misclassifying a red tide sample (i.e., false negatives) ~10%. The optimized likelihood of having non-bloom samples flagged was of ~40% and the misclassification of non-bloom (i.e., false positives) ~20%.

Table 7.9: Probabilities (Optimized algorithms)

<table>
<thead>
<tr>
<th>Bio-optical technique</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
<td>False (Non-bloom)</td>
</tr>
<tr>
<td>Bio-optical technique results</td>
<td>0.280</td>
<td>0.172</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.111</td>
<td>0.437</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Empirical approach</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
<td>False (Non-bloom)</td>
</tr>
<tr>
<td>Empirical approach results</td>
<td>0.275</td>
<td>0.199</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.116</td>
<td>0.410</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**7.7. The combined Hybrid scheme**

The main consequence of the optimization procedure was a great opportunity to combine the ability of these two improved algorithms into the Hybrid scheme. The Table 7.10 depicts the frequency of the two scheme levels: higher predictability and higher confidence.
Table 7.10: The Hybrid scheme.

<table>
<thead>
<tr>
<th>1st HYBRID SCHEME: HIGH PREDICTABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red tide</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>False negative</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>False positive</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd HYBRID SCHEME: HIGH CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red tide</td>
</tr>
<tr>
<td>False negative</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Non-bloom</td>
</tr>
<tr>
<td>False positive</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

7.7.1. Hybrid scheme statistical assessment

The red tide prevalence or the odds of a red tide does not changed since the number of red tide and non-bloom samples were defined a priori. The statistics of the Hybrid scheme were computed for both levels (i.e., high predictability and high confidence).

The high predictability scheme had a rate of success at identifying correct samples of ~80% (293 correct samples of a total of 371); while the rate of success at identifying correct samples of the high confidence was 61% (227 correct samples of a total of 371).
The high predictability had ~80% of sensitivity and specificity, and ~20% of false negatives and false positives; while the high confidence scheme had sensitivity and specificity of ~60%, and about 40% of false negatives and false positives (Table 7.11).

Table 7.11: Sensitivity, specificity, false negatives, false positives (Hybrid algorithm).

<table>
<thead>
<tr>
<th>3rd statistical assessment: DETECTED CASES</th>
<th>In situ observation of the presence or absence of red tide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Red tide)</td>
</tr>
<tr>
<td></td>
<td>False (Non-bloom)</td>
</tr>
<tr>
<td>1st level of the Hybrid scheme:</td>
<td></td>
</tr>
<tr>
<td>High predictability</td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>116 (80%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>49 (22%)</td>
</tr>
<tr>
<td>SUM</td>
<td>145 (100%)</td>
</tr>
<tr>
<td>2nd level of the Hybrid scheme:</td>
<td></td>
</tr>
<tr>
<td>High confidence</td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>90 (62%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>89 (39%)</td>
</tr>
<tr>
<td>SUM</td>
<td>145 (100%)</td>
</tr>
</tbody>
</table>

Table 7.12 presents the results of the predictive values of the two Hybrid scheme levels. The positive predictive value of the 1st scheme level, the positive predictive value were 70% and its invert, 30%; while the negative predictive value was 86%, and its invert 14%. In the case of the 2nd scheme level and its invert, were 50:50; while the negative predictive value was 71%, and its invert 29%.
Table 7.12: Predictive values (*Hybrid* algorithm).

The *Hybrid* scheme probabilities (Table 7.13) of detecting red tides was ~31% for the 1<sup>st</sup> level and ~24% for the 2<sup>nd</sup> level. The probability of misclassifying a red tide sample (*i.e.*, false negatives) was ~8% for the 1<sup>st</sup> level, and almost double of that for the 2<sup>nd</sup> level (~15%). The likelihood of having non-bloom samples flagged was of ~50% for the 1<sup>st</sup> level, and ~37% for the 2<sup>nd</sup> level. The misclassification of non-bloom (*i.e.*, false positives) was 13% for the 1<sup>st</sup> level, and ~24% for the 2<sup>nd</sup> level.

<table>
<thead>
<tr>
<th><strong>4&lt;sup&gt;th&lt;/sup&gt; statistics assessment:</strong></th>
<th><strong>In situ observation of the presence or absence of red tide</strong></th>
<th><strong>SUM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PREDICTIVE VALUES</strong></td>
<td><strong>True (Red tide)</strong></td>
<td><strong>False (Non-bloom)</strong></td>
</tr>
<tr>
<td><strong>1&lt;sup&gt;st&lt;/sup&gt; level of the <em>Hybrid</em> scheme:</strong></td>
<td><strong>High predictability</strong></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>116 (70%)</td>
<td>49 (30%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>29 (14%)</td>
<td>177 (86%)</td>
</tr>
<tr>
<td><strong>2&lt;sup&gt;nd&lt;/sup&gt; level of the <em>Hybrid</em> scheme:</strong></td>
<td><strong>High confidence</strong></td>
<td></td>
</tr>
<tr>
<td>True (Red tide)</td>
<td>90 (50%)</td>
<td>89 (50%)</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>55 (29%)</td>
<td>137 (71%)</td>
</tr>
<tr>
<td><strong>SUM</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.13: Probabilities (*Hybrid* algorithm).

<table>
<thead>
<tr>
<th>5&lt;sup&gt;th&lt;/sup&gt; statistics assessment: PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; level of the <em>Hybrid</em> scheme: High predictability</td>
<td>True (Red tide)</td>
<td>0.313</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.078</td>
<td><strong>0.477</strong></td>
</tr>
<tr>
<td><strong>SUM</strong></td>
<td></td>
<td><strong>1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5&lt;sup&gt;th&lt;/sup&gt; statistics assessment: PROBABILITIES</th>
<th>In situ observation of the presence or absence of red tide</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; level of the <em>Hybrid</em> scheme: High confidence</td>
<td>True (Red tide)</td>
<td>0.243</td>
</tr>
<tr>
<td>False (Non-bloom)</td>
<td>0.148</td>
<td><strong>0.369</strong></td>
</tr>
<tr>
<td><strong>SUM</strong></td>
<td></td>
<td><strong>1</strong></td>
</tr>
</tbody>
</table>
CHAPTER 8

DISCUSSION

The successful replication of the results of the Bio-optical technique (Cannizzaro, 2004; Cannizzaro et al., 2008) and the application of the simpler, alternative algorithm, the Empirical approach introduced by this work, showed an initial compatibility between the results of the two algorithms. In reality, they are quite similar, since besides using chlorophyll to flag *K. brevis* blooms, the Bio-optical utilizes the particulate backscatter at 550nm (\(b_{bp(550)}\)), while the Empirical makes use of the water-leaving radiance at 550nm (\(L_w(550)\)). Indeed, \(b_{bp(550)}\) is a satellite product derived from \(L_w(550)\) retrievals (Carder et al., 1999). Also, at 550 nm the backscatter is the major attenuation factor because absorption is small (Cannizzaro, 2004), so the backscatter (not only for the particulate, but for all classes of optical active constituents) is closely related to the water leaving radiance. But of the two, the backscatter is a more highly derived variable and therefore subject to greater uncertainties.

Besides improving the confidence of red tide identification with a multi-algorithm application, this work supports the concept of multi-sensor analyses, since comparable results were found using data from SeaWiFS and MODIS. The successful use of these two sensors in the same ecological analysis is in accordance with expectations regarding their similar characteristics (Blondeau-Patissier et al., 2004; Darecki and Stramski, 2004; Maritorena and Siegel, 2005; Zibordi et al., 2006). Thus, the consistent results in this
multi-sensor investigation implies a potential approach to improve temporal sampling using visible sensors, since it increases the chances of obtaining images without cloud contamination. It also represents an additional enhancement to monitoring systems, and additional improvement at detecting red tides.

Given the promising results of this multi-algorithm/multi-sensor application, an observation should be stressed relating the MODIS imagery usage. The Terra instrument was used, but this instrument is not as well suited to ocean color applications as the MODIS on Aqua (Barnes et al., 2003; Hoge et al., 2003; Franz et al., 2000). The MODIS mission serves as a prototype, representing a potential linkage between the ongoing and the next generation of ocean color satellite instruments (IOCCG, 2007).

However, additional constraints related with the detection of *K. brevis* blooms may apply even if using MODIS-Aqua data, or any other well-calibrated sensor (e.g., SeaWiFS). This is mainly related to the fundamental basis of a satellite remotely sensed dataset. The satellite ocean color sensors are limited in their sensitivity and spectral bands (Esaias et al., 1998; Wang, 1999; Guenther et al., 2002; Gower and Borstad, 2004), and this may lead them to not correctly identifying the presence or absence of red tides in all cases (Millie et al., 1997; McRoyer-Etheridge and Roesler, 1999; Schofield et al., 1999).

In near-shore applications, errors in the satellite retrieval may occur by interference produced by reflection of the bottom (Gould and Armone, 1998; Lee et al., 1999; 2001; Ohde and Mueller, 2001; Ohde and Siegel, 2001; D'Sa et al., 2002; Cannizzaro and Carder, 2006). Imperfect atmospheric corrections are another serious problem for these sensors, especially in coastal waters (Gordon, 1978; 1997; Gordon and Wang, 1994; Siegel et al., 2000). Applications in these regions face serious limitations since the
presence of other in-water constituents, such suspended solids and CDOM that vary independently of phytoplankton, cause erroneous estimations of absorbing aerosols (Hu et al., 2000; 2006a; Morel and Bélanger, 2006), causing more light at the top of the atmosphere to be attributed to the atmosphere itself, decreasing the water leaving radiance (Gordon et al., 1997; Schollaert et al., 2003). This being so, the use of standard global chlorophyll band-ratio algorithms (e.g., OC3M for MODIS; O’Reilly et al., 2000), are susceptible to overestimating values (Ruddick et al., 2000; Stumpf, 2001; Hu et al., 2003a; 2003b; 2005a). Therefore, since the Carder semi-analytic algorithm (Carder et al., 1999; 2004) is designed for coastal applications, future investigations using chlorophyll derived from the Carder algorithm could address whether more accurate retrievals would be achieved than obtained with the OC3M retrievals used here.

The half pixel size (i.e., 500 m) buffer zone applied around the Florida coast might have reduced the number of stations influenced by terrigenous sediment input. However, it may have led to the use of few inappropriate valid match-ups contaminated by land effects. To minimize residual along-scan effects on the satellite retrievals from near field scattering contamination by land, and even bright targets (e.g., clouds or sun glint), distances of three pixels (i.e., 3 km) from those targets would probably improve the elimination of contaminated satellite data (Franz et al., 2005).

Challenges to the detection of red tides from spacecraft data were implicit in the nature of the field measurements used here. While through the course of this investigation the in situ database provided by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute was used to determine the accuracies of the satellite-based algorithms, these samples were not collected for this reason. This in
in situ database is an effort to sample the Florida Red Tide. Thus the in situ sampling is itself biased towards the occurrence of *K. brevis*. In addition, several *Karenia* congenerics co-exist in the Gulf of Mexico, *e.g.*, *K. brevis*, *K. papilionacea*, *K. mikimotoi*, *K. selliformis*, and *K. “mexican hat”* (FWRI, 2002). Over the study period (2002~2006) half of the samples corresponded to *K. brevis*, the most toxic, not only among *Karenia* species but also among the nearly 40 toxic species found throughout the Gulf of Mexico (Kusek et al., 1999). Thus, the focus of this research was chosen to be *K. brevis* as the closure of shellfish beds is based on *K. brevis* abundance (*i.e.*, 5x10^3 cells l^-1; FDA, 2005). However, any similar satellite investigation into the detection of the Florida Red Tide, using the genus *Karenia* instead, may not find the same results of those presented here.

Two fundamental pieces of information needed for validating satellite data in shallow zones were not provided for most in situ database entries: the sampling time and the bathymetry at the site. The absence of the former made it impossible to know the elapsed time between satellite overpasses and field measurements. Thus, it required the use of a temporal window of 24h, which is more relaxed than the usual NASA’s standard validation near-coincident threshold of ± 3 hours (Bailey et al., 2000). The wider time window is not too critical given the relative slow e-folding time of the growth rate of *K. brevis*, but in regions of horizontal gradients in the *K. brevis* concentration, the effects of advection by near-surface currents will have added additional noise to the analysis. To cover for the lack of the bathymetry, the ETOPO2 database (Geodas, 2007) was used to determine local water depth. Since the Continental Shelf of the West Florida presents a slope that varies gently, having no major features and not many abrupt isobath variations (Li and Weisberg, 1999), an interpolation procedure (*e.g.*, linear, bilinear, polynomial,
etc.) could provide more accurate depth values at the *in situ* stations exact location, possibly bringing some of the eliminated stations back into play. Additionally, while detecting the absence and presence of *K. brevis* blooms using satellite measurements, point sources (*i.e.*, *in situ* cell abundances) were used in comparison with satellite data. Since information about bloom patchiness was not provided, and because MODIS-Aqua ground resolution at nadir is 1 km², the *in situ* sampling may not always supply sufficient information representative of the spatial averaging implicit in the signal sensed by the satellite.

To increase confidence in the match-up analysis, the exclusion criteria resulted in a considerable data loss, introducing possible biases from the reduction in the number of samples. On the other hand, it eliminated possible sources of other errors, and enhanced the accuracy of this study. The product of the match-up analysis (*i.e.*, valid *in situ*-satellite pairs) is taken as the best that can be done with available data for assessing the accuracy of detecting *K. brevis* blooms with remote sensed data. Despite all restrictions associated with field measurements, the overall use of this database served as a reasonable training dataset to assess the performance of the three selected satellite-based algorithms, *i.e.*, the simple *Empirical* approach, the *Bio-optical* technique, and NOAA’s official *Operational* method (Stumpf et al., 2003b; Tomlinson et al., 2004). Excepting the North-FL and South-FL in the Winter-Spring, where not many non-bloom match-up pairs were present, all other regions and periods had representative non-bloom cases. But this did not hold true for the valid red tide match-ups within any other region and period other the Central West Florida Shelf (Figure 1.2) in the Summer-Fall. The results non-heterogeneous spatial and temporal distribution in the valid match-ups permitted
verification of the hypothesis only within this region and period. This bias observed on the pairing of valid \textit{in situ}-satellite pairs was not expected but it is not surprising, and may result from two factors. Firstly, because the field effort has its focus on sampling red tides, this caused an irregular measurement domain in time and space. Secondly, the seasonal nature of the phenomenon itself means it tends to occur during the Summer-Fall in the Central West Florida Shelf.

Thus, the population of valid samples within the Central West Florida Shelf in Summer-Fall is assumed to represent a true environmental variability of this region and period. This population size is comparable to other remote sensing validation studies found in the literature. The total number of valid match-ups was 371 throughout almost 4 years for years, leading to \textasciitilde100 samples average per year, and representing 11.5\% of the initial sampling database for this region and period (\(n=3420\)). Hyde et al. (2007), while validating SeaWiFS chlorophyll in Massachusetts Bay, had 426 (\textasciitilde6.3\%) valid match-ups (\(n=6773\)) over \textasciitilde7 years period (\textasciitilde60 per year). In an evaluation of the \textasciitilde9 years of satellite chlorophyll in the Mediterranean Sea, Volpe et al. (2007) had 440 of 1144, reaching an elevated \textasciitilde40\% of valid match-ups, but an average of \textasciitilde50 samples per year. However, the analysis performed on these two studies used multi-pixel boxes (5x5 and 3x3, respectively) centered on the field station. While this multi-pixel box match-up permits an increased number of valid pairs (by accepting samples having the \textit{in situ} station collocated pixel covered by cloud, but adjacent pixels are cloud-free), it also introduces uncertainties through horizontal gradients, \textit{i.e.}, bloom patchiness (Hu et al., 2001; Bailey and Werdell, 2006).
Through the process of defining the accuracy of the algorithms in the Central West Florida Shelf (Figure 1.2) during the Summer-Fall, the *in situ* dataset was divided into two subsets based on the threshold of 15,000 cells l$^{-1}$ (*i.e.*, red tide and non-bloom; Table 6.2). All statements hereon are related with this threshold and this region and period, unless otherwise stated. The first statistical test revealed that the red tide prevalence was not the ideal as desired (*i.e.*, 50:50, meaning an even number of red tide and non-blooms samples), but quite close ~40%. High abundances of *K. brevis* are not as easy to come across as are low *K. brevis* concentrations, in another words, red tides are fewer in comparison to non-blooms. So, this ~40% can be considered an elevated value, resulting from the targeted nature of the *in situ* sampling. However, since the used *in situ* database, is not based on a blind random sampling, but it is target to find red tides, this number represents the prevalence of the red tide signatures in the database, but not in the real world. In the case of low prevalence, false positives, as opposed to false negatives, are more frequent in occurrence (Bayer’s theorem; Sokal and Rohlf, 1995). This was indeed the case for the *Bio-optical* and *Empirical* algorithms that had fewer false negatives than false positives, but not for the *Operational*.

In fact, the third statistical test demonstrated that the *Operational* method was the most effective (84%) in identifying the non-bloom cases (high specificity), but the least accurate at detecting red tides (26%; weak sensitivity). On the contrary, the *Empirical* approach was the best to flag red tides (~85%; high sensitivity), but the worst in identifying non-bloom samples correctly, close to 50% (low specificity). From the point of view of a resource manager, or a scientist studying red tides, or even someone who wants to know which beach to go, none of the results of two algorithms are desirable. At
an extreme an algorithm classifying the entire Gulf of Mexico as non-bloom would have
the highest success at correct identification (i.e., high specificity), with relatively few
false negatives, since red tides are not such common occurrences (e.g., top row on Figure
7.17 and 7.18). On the other hand, an algorithm flagging the whole Gulf of Mexico as red
tides would correctly identify all red tides in the satellite data (i.e., high sensitivity), but
would produce many false positives (e.g., second row from the top on Figure 7.17 and
7.18). But neither of these extreme algorithms would provide any useful information,
even though according to some metrics they would exhibit high levels of success. Of the
three algorithms studies here, the Bio-optical technique produced more even sensitivities
and specificities ratio than the other two algorithms.

Even though the sensitivities and specificities were different for the three
algorithms, their overall accuracy was not so different (second statistical test). Thus the
sole use of the overall accuracy to express the success of the algorithm could mislead the
user (e.g., resource manger, scientist, general public), by not being a valid representation
of the real accuracy at predicting correct field conditions (i.e., absence or presence of red
tides).

The fourth statistical test reveals that the three algorithms have similar median red
tide predictability (~55%), i.e., the algorithms are only flagging correctly about half of
the times red tides are present. A user would not be confident with such positive
predictive values. The negative predictive value is reliable for the Bio-optical and for the
Empirical algorithms (above 80%), but this does not serve the purpose of the applications
of the algorithms, which is to detect the presence and the absence of red tides.
Regarding the probabilities of an event to occur (fifth statistical test), the *Operational* method had the best likelihood of identifying non-bloom samples, but this is not the original intention of the algorithm. An overall evaluation of the statistical tests illustrated that for detecting red tides, this method showed to be the poorest results of the three. This method does not look into the specific optical property of the water related to *K. brevis* blooms, but into chlorophyll anomalies only, and a reference filed (*i.e.*, mean) to derive the anomaly. Also, if blooms persist long enough to contribute to the signal of the calculated mean, the use of this biased mean will not allow the algorithm to correctly flag this particular bloom. Erroneous reference fields will produce anomalies that are not useful at detecting new blooms. However, the main feature of this method is the use of previous environmental information such as the knowledge of pre-existing bloom (Fisher et al., 2005). Unfortunately, in the present research this ancillary information was not included, nor was a case-to-case discrimination performed to determine if an existent bloom was biasing the generation of the anomaly. Instead, the results presented here are from a plain match-up analysis. The success of this method at detecting new blooms is pointed out by Hu et al. (2008), who evaluated a similar *in situ*-satellite pairing analysis, but using SeaWiFS data.

The *Empirical* approach and the *Bio-optical* technique showed similar event probabilities (fifth statistical test). On general, the statistical results of these two algorithms can not be used to determine weather to differentiate one algorithm is better than the other, because the five tests are to be evaluated in a wide-ranging mode. Their predictability is alike, but the *Empirical* has a very good sensitivity and it is much simpler to implement. The *Bio-optical* has a better sensitivity-specificity balance. This observed
Bio-optical statistical performance is in accordance with the expectations of Cannizzaro (2004) and Cannizzaro et al. (2008). But Hu et al. (2008) evaluated this technique on its original basis, over a longer period using SeaWiFS data, and concluded that it is not effective at detecting red tides. However Hu et al. (2008) included samples south of 25.75oS, and as it is shown in this study that the red tide detection in this region breaks down and more than 50% of false negatives were observed (Figures 7.12 and 7.13), within the limitations imposed by the low number of samples.

After the optimization performed on the Bio-optical and Empirical algorithms, the statistical tests were re-commutated and acceptable levels of ~70% of red tide and non-bloom detection accuracy were achieved. The predictive values were also improved, reaching ~60%. Because the clustering of the red tide subset data overlaps those of non-blooms (Figure 7.15: center and bottom rows; 7.17 and 7.18) this optimization procedure was not able to reach better levels, but it resulted in improvements in relation to the original set up of the algorithms. The Operational method was not optimized because the red tide and non-bloom overlap was almost total (Figure 7.15 top).

The optimized algorithms were combined into a single hybridized scheme using the their abilities not only to improve red tide predictability (1st scheme), but also to find higher confidence in the outcomes (2nd scheme). This Hybrid scheme provided a new and powerful-monitoring tool to detect K. brevis blooms in the Central West Florida Shelf during the blooming season (i.e., Summer-Fall) more accurate than the individual algorithms.

The 1st Hybrid scheme produces results of outstanding sensitivity (80%) and specificity (78%), plus good predictive values, and its application would bring benefits to
monitoring systems. By identifying the cases where either of the two algorithms correctly detects a red tide increases the chances of flagging red tides, and reduces false negatives. The same is applicable for the non-bloom cases. Through better guidance of field surveys many unanswered questions could be addressed, such as those about the red tide origin, maintenance and termination trigger mechanism. When one algorithm does not flag a red tide, the likelihood of this being flagged by the other is high, and vice-versa. Someone who simply wants to go to the beach will have a good chance to go to a place without red tides, since few false negatives are created by this scheme. Even though the threshold stated here for detecting red tides with satellite measurements is above of those to close shellfish beds (15,000 compared to 5,000 cell l⁻¹), this threshold is more realistic than the one used by the Operational method (100,000 cell l⁻¹). Satellite monitoring might never be able to provide the necessary information to result in the of shellfish beds, but the application of this Hybrid scheme could also be used to direct the resource managers attention to regions likely to be subject red tides, which might help to improve the process leading to the closure of beds. Shellfish bed owners do not want their beds to be closed without good reasons (i.e., without an actual red tide), but they would probably prefer to have them closed to avoid an outbreak of shellfish poisoning, rather then have false negatives situations where contaminated shellfish enter the food chain.

The 2nd Hybrid scheme results in high confidence of detection when both of the two algorithms flag a red tide. But this scheme would cause more false negatives, since not full recognition could be credited to each of the outcomes, i.e., when one algorithm flags a red tide and the other does not. A full study of the optimization procedure to utilizing this Hybrid scheme, and others using more algorithms would be a suitable topic for future
research. Also, a spatial coherence analysis should be performed to extend the results of the preliminary study of the replication of the *Bio-optical* technique to the optimized algorithms and the *Hybrid* scheme.
CHAPTER 9

CONCLUSIONS

The results obtained by this research have proved that the initial hypothesis that blooms of *K. brevis* can be identified in satellite ocean color images is valid. It has been shown that the MODIS-Aqua imagery (2002~2006) satisfactorily reveals these optical patterns. It has been demonstrated for the first time that this satellite dataset has utility at identifying *K. brevis* blooms of the West Florida coast over an extended period of time. The advantage of using remote sensing data to study oceanographic and biological surface features is that satellites retrievals provide a synoptic view of the phenomenon, unlike that available through *in situ* sampling (Werdell et al., 2003).

An extensive match-up analysis was successfully applied to derive a simpler, alternative algorithm to detect *K. brevis* blooms with satellite imagery: the *Empirical* approach. The pairs of match-ups were also effectively utilized to determine the ability of this satellite-based algorithm to identify red tides, as well as of two other published algorithms: the *Bio-Optical* technique (Cannizzaro, 2004; Cannizzaro et al., 2008) which showed to be useful to separate *K. brevis* blooms from non-blooming waters using shipboard radiometric measurements; and the *Operational* method (Stumpf et al., 2003b; Tomlinson et al., 2004) which has been implemented by the NOAA CoastWatch program to issue routine Chl anomaly alert bulletins in the Gulf of Mexico (HabFS, 2007).
The training of the algorithms was accomplished by utilizing an in situ database courtesy of the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute. The use of this database provided positive identification results, however, the valid pairs produced by the match-up analysis were constrained by geographical and seasonal aspects. The data density was not sufficient to permit a study of the time vary signal. Conclusions away from the Central West Florida Shelf (Figure 1.2) were limited by the relatively small number of in situ-satellite match-up pairs. The strong seasonality of red tide events, reflected in the sampling effort, also restricted conclusions to the blooming season, i.e., the Summer-Fall period.

The Operational method was found to be good identifying noon-bloom samples but that did not hold true for its skills at detecting red tides. This result was obtaining using not only the original cell threshold of 100,000 cells l\(^{-1}\) and also with the more realistic 15,000 cells l\(^{-1}\). The other two algorithms were only tested with the more realistic threshold, and the sensitivity of the Empirical method was high (~85%) but its specificity was lower (~50%). The Bio-optical technique had more even sensitivity-specificity rates. The positive predictive values of the three algorithms were not much different, ranging around 55%.

The Empirical and Bio-optical algorithms were both optimized establishing new thresholds, with red tide detection and non-bloom identification simultaneously maximized to ~70% for both algorithms. Further improvement will follow by elucidating of some of the unknowns and the use of an independent dataset to verify these results. The combined use of the optimized algorithms in sequence, represented by the Hybrid scheme, provided another new monitoring tool with better characteristics (e.g.,
predictability and confidence) for characterizing and understanding of blooms of *K. brevis* in the Central West Florida Shelf during the blooming season. The linkage between the results of a multi-algorithm investigation on a single analysis intent to ameliorate the remotely-sensed detection of *K. brevis* blooms, improving field monitoring direction to areas that should receive special attention.
Bibliography

A wide-ranging satellite-related bibliography, a comprehensive list of references regarding the harmful algal bloom topic is presented on here.


Brand, L.E., Compton, A.. 2007. Long-term increase in Karenia brevis abundance along the Southwest Florida Coast. Harmful Algae, 6, 232-252.


FDA (Florida Department of Agriculture and Consumer Services). National Shellfish Sanitation Program: Guide for the Control of Molluscan Shellfish.


Fleming L.E., Jerez E., Stephan W., Cassedy A., Bean J.A., Reich A., Kirkpatrick B., Backer L.,


Hemmert, W. H.. 1980?. The public health implications of Gymnodinium Breve red tides, a review of the literature and recent events.


McLeroy-Etheridge, S.L., Roesler, C.S.. 1999? . Are the inherent optical properties of phytoplankton responsible for the distinct ocean colors observed during harmful algal blooms?


for remote sensing. Limnology and Oceanography, 44 (3), 618-627.


Yoder, J.A., Kennelly, M.. 2006. What have we learned about ocean variability from satellite ocean color imagers?. Oceanography, 19(1), 152-171.


APPENDIX 1

HEALTH EFFECTS CAUSED BY HARMFUL ALGAL BLOOMS

Phytoplankton toxins can adversely affect public well-being in two ways: by inhalation (Backer et al., 2003; Kirkpatrick et al., 2006) and by ingestion of tainted seafood (Kirkpatrick et al., 2004). A third hazard may arise by direct contact with the toxin, but there have been too few cases to establish the consequences (Kusek et al., 1999). Moreover, associated with the serious public health threats, a constant risk of mass marine wildlife kills also exists which can contribute to detrimental degradation of the coastal aesthetics (Van Dolah, 2000).

Some believe that brevetoxins produced by *K. brevis* are not released into the environment while cells are alive (i.e., endotoxins), but whenever cells break apart (e.g., lyse) the toxin is released into the water. Once in the water column, the toxin can be exported via bubbles and or contaminated sea-spray, and depending upon local environmental conditions (such as ocean currents, breaking surf waves or onshore winds), humans and other mammals onshore may be exposed to the odorless aerosolized brevetoxins (Asai et al., 1982; Quirino et al., 2004; Coblentz and Naar, 2007; Milian et al., 2007). The common symptoms range from itchiness, skin rashes, profuse watery, burning or stinging eyes, dizziness, tunnel vision, conjunctival irritation, throat/nasal irritation, copious catarrhal exudates and rhinorrhea, shortness of breath, difficulty to take a deep breath, acute reversible upper respiratory syndrome (dry, choking cough and
bronchoconstriction), and in individuals with asthma, a worsening of their asthma (the reader is referred to Backer et al. (2003), Cheng et al. (2005), Fleming et al. (2005a; 2007b) for pertinent related bibliography).

Brevetoxin is tasteless and thus difficult to detect in food, and hard to eliminate through any food preparation procedure, and cooking cannot neutralize tainted seafood since this toxin is temperature stable (Kirkpatrick et al. 2004; Van Dolah, 2000). After ingesting brevetoxin-contaminated seafood, consumers have the risk of acquiring Neurotoxic Shellfish Poisoning (NSP; Poli et al., 2000). NSP is a temporary illness that can be characterized gastrointestinal problems, including nausea, diarrhea and abdominal pain. It can also cause neurological paresthesias, as well as vertigo and lack of coordination. Acute symptoms include respiratory depression, bradycardia, headache, dilated pupils, and convulsions (Bronson, 2002). No human fatalities have yet been directly attributed to brevetoxins (Landsberg, 2002), but brevetoxin levels during *K. brevis* blooms can reach levels that can be potentially fatal to humans (Baden et al., 1984).

Public health awareness concerning Gulf of Mexico shellfish consumption began in the late 1940s after one of the worst and largest (in both longevity and geographical distribution) *K. brevis* blooms on record (Gunter et al., 1948). In 1972, the Florida Department of Agriculture and Consumer Services began requiring all U.S.A. Gulf of Mexico states to close shellfish beds due to the NSP threat. This closure happens whenever *K. brevis* concentrations exceed of 5x10^3 cells l^{-1} near or in harvesting areas (FDA, 2005). Beds are reopened after levels Mousse Bioassay are less than 20 units (Bronson, 2002).
In general, the toxin produced by the HABs species often bioaccumulate in the food web as they pass to higher trophic levels (Turner and Tester, 1997; Steidinger et al., 1998; Tester et al., 2000). Compendiums of investigation have traced the toxic pathway from the causative primary producers to small herbivorous consumer vectors (Chattopadhyay et al., 2002; Lester et al., 2008), like phytophagous fish (Lefebvre et al., 2007) and crustaceans, such as copepods (Avery et al., 2007) or crabs (Garcia et al., 2007). Considerable research has gone into understanding the toxicological effect on commercially important species of filter-feeding bivalve shellfish (Roberts et al., 1979), such as clams (Reich et al., 2007), oysters (Plakas et al., 2007), and mussels (Sutherland and Silver, 2007).

Throughout the world’s oceans, HABs have littered thousands of kilometers of shorelines with tons of dead fish (Van Dolah, 2000). This situation may worsen with the spread of the increase of aquaculture and mariculture commerce (Botes et al., 2003). Also, piscivorous predators, carnivorous fish, and many other marine organisms like sea turtles (Gunter et al., 1948; Stumpf 2001), and scavenging seabirds (Forrester et al., 1977; Shumway et al. 2003) have been found moribund or dead, due to HABs. In addition, evidence of the effects of toxins can be seen in large-scale marine mammals (Flewelling et al., 2005), which are susceptible to reoccurring multiple die-offs. These include manatees (Bossart et al., 1998; Landsberg and Steidinger, 1998; Trainer and Baden, 1999), dolphins (Landsberg, 2002; Radwan, et al., 2007), seals (McCauley et al., 2007), sea lions (Scholin et al., 2000), and whales (Durbin et al., 2002; Fire et al., 2007). Even land mammals are vulnerable if exposed to HABs, e.g., bats (Pybur et al., 1986), dogs
(Ellis et al., 1978; Nehring, 1993), cattle (Cronberg et al., 1999), horses and poultry (Dillenberg and Dehnel, 1960).
APPENDIX 2

DISTRIBUTION OF HARMFUL ALGAL BLOOMS

Over the past decades, the subject of relating the oceans to human health has gained public attention from media and society at large (Hemmert, 1980; Anderson, 1995; Kusek et al., 1999; Fleming and Laws, 2006; Flaming et al., 2006). As an example of the relationship between oceans and human health, HABs have received increasing scientific attention (Landsberg, 2002; Sellner et al., 2003; Stewart et al., 2007). Of particular interest is whether the increased observations of HABs over time are indicative of a trend of increasing incidence of HAB events worldwide (Anderson, 1989; Smayda, 1990; Hallograeff, 1993; Gilbert et al. 2005a). Better documentation, monitoring and surveillance are among the most obvious reasons for the increasing frequency of these reports (Van Dolah, 2000). But an expanded range could also be explained by increased shipping ballast water transport and subsequent introduction of non-indigenous species to other areas (Gollasch et al., 2000), influences produced by changes in water column stratification (Smayda, 2002; Gentien et al., 2005), changing (particularly increasing) ocean temperatures (Smayda, 2007), upwelling (Kudela et al., 2005; Lanerolle et al., 2006), eutrophication (Gilbert et al. 2005b; Lester et al., 2001), weather conditions such as wind (Kanoshna et al., 2003; Janowitz et al., 2008), and the effects of light and salinity (Nagasoe et al., 2005; Magaña and Villareal, 2006; Brown et al., 2006). Other possible
causes for an increased trend include speculations such as estuarine iron stocks, re-
suspension of spores, ocean circulation, convergence, fronts, rainfall, storm events, short-
term environmental perturbation, local transient climatic conditions, and El Niño events (Roesler and McLeroy-Etheridge, 1998; Kusek et al., 1999; Van Dolah, 2000; Magaña et al., 2003; Kirkpatrick et al., 2004).

The mechanism for toxic bloom formation is indeed not well understood. In Florida specifically, global warming (Tester et al., 1993), hurricanes (Hu et al., 2006b), the Loop Current (Murphy et al., 1975; Haddad and Carver, 1979), riverine input (Stumpf et al., 2008), and related co-occurrence of *Trichodesmium* blooms with atmospheric transport of aeolian Saharan dust (Walsh and Steidinger, 2001; Walsh et al., 2006) have been hypothesized as possible explanations for *K. brevis* blooms. Some believe that the CWFS HABs usually initiate offshore (Steidinger and Hadadd, 1981; Tester and Steindinger, 1997), indicating a natural source. Other evidence suggests that the real initiation mechanism is strongly related with anthropogenic pressure from the shore (Brand and Compton, 2007). The belief held by some regarding the recent increase of the HABs is not supported by results of Walsh and Steidinger (2001), which found no difference in the occurrence of the Florida Red Tide between 1957 and 1992. However, others, such as Brand and Compton (2007), state that more than a 13-fold increase on Florida Red Tide events was found during the same period (1957~1992), using the same data as Walsh and Steidinger (2001).
APPENDIX 3

MONITORING OF HARMFUL ALGAL BLOOMS

To date, there has been an extensive effort to improve detection, monitoring and forecasting of HABs, focusing on documenting their location, magnitude and transport (Anderson, 1997; Tester and Steidinger, 1997; Sacau-Cuadrado et al., 2003). Early warning systems (Lee et al., 2005a) and autonomous prototype instruments (Robbins et al., 2006; Doucette et al., 2007; Kirkpatrick et al., 2007) have all shown promise in adding valuable information to the detection and monitoring of HABs. Some Asian countries that have massive aquaculture industries maintain active HAB monitoring programs (Lee et al., 2005a), e.g., Singapore (Liew et al., 1999), Korea (Palanisamy et al., 2005; Ahn et al., 2006), Vietnam (Tang et al., 2004), and China (Chen et al., 2003; Qi et al., 2004; Gao and Song, 2005) which alone derives ~70% of the total world's aquaculture production (Pittenger et al., 2007). In Europe, some agencies also provide systematic coastal alert programs (Cracknell et al., 2001; Vila et al., 2001). A large global HAB program, GeoHAB, had a branch-monitoring program, including the CWFS (Ecology and Oceanography of Harmful Algal Blooms - EcoHAB Florida: 1998 - 2001).

In the US, the states of Washington (Olympic Region Harmful Algal Bloom - ORHAB), Maine, Alabama, and Florida all conduct routine sampling. The lower Great Lakes also have a HAB monitoring system (Monitoring and Event Response for Harmful
Algal Blooms - MERHAB). Other Gulf states do response monitoring (i.e., whenever blooms are reported). In addition to these local programs, the National Oceanic and Atmospheric Administration (NOAA) maintains a satellite-based HAB detection system in the Gulf of Mexico (HabFS, 2007). A Floridian effort to provide outreach and education started in 1998 when the University of Miami’s Florida Poison Information Center started a 24 hours 365 day per year toll-free (1888-232-8635) Aquatic Toxin Hotline (Fleming et al., 2007; 2007a). The Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWRI, 2002) also gathers and distributes a historical database of field sampling efforts for *K* *brevis*.

The biggest problem with these monitoring efforts is that most traditional monitoring programs are primarily based on discrete field sampling, which leads to large spatial and temporal biases. Up to the 1970s, most oceanographic studies were based on conventional ship-based surveys with the limitation of relying on favorable weather conditions, and having discrete data stations creating datasets with considerable spatial and temporal biases due to sparse non-uniform coverage. Moreover, ship operations are extremely time-consuming, and incur high costs in man-hours and ship-time (Schofield and Glenn, 2004).

Other attempts, such as observations from moorings, drifting buoys and modeling, have added high-quality contributions to the classical environmental research (Prandle, 2000). However, many questions that have arisen about HABs remain unanswered due to inadequate data sources. Indeed, an effective and remote HAB identification method could contribute to greater understanding of the origin, maintenance and termination
trigger mechanism (Babin et al., 2005; Milroy et al., 2008), as well as to ameliorate economic impacts (Habas and Gilbert, 1975; Hoagland et al., 2002; 2007; Tester, 2007).
APPENDIX 4

DETECTING HARMFUL ALGAL BLOOMS FROM SPACE

At its inception, satellite remote sensing proved to be a successful and valuable complementary tool for operationally taking radiometric measurements of the Earth’s surfaces, and of monitoring trends and processes (Johannessen et al., 2000). Satellite imagery offers relatively low cost data to routinely map large areas (millions of km² per satellite scene) and thereby provides the potential for long-term, regular, repeated, and high-frequency synoptic sampling and with the possibility of daily revisits. Particularly in the marine realm, near-real-time satellite imagery can enhance ocean field surveys by locating regions that should receive focused sampling attention (Smyth et al, 2003), reducing timelines to find surface features such as upwelling (Carvalho, 2002), eddies (Johannessen et al., 1989; Souza, 2005), fronts (Takahashi and Kawamura, 2005), oil slicks (Bentz et al., 2003; 2004; 2005), river plumes (Jolliff et al., 2003; Hu et al., 2005b), coastal run-off (Hu et al., 2004b), and red tides (Hu et al., 2005a). Finding these surface features requires extensive field coverage not easily obtained through surveys that are based solely on ship observations (Tester et al., 1991; Tang et al., 1998; 2003a; 2003b). In fact, the accurate ocean visualization and the general scene interpretation brought by the use of satellite imagery added functional and strategic guidelines to researchers and resource administrators. The synoptic aspect of the remote sensing usage could be
particularly useful in detecting red tides before they become a severe problem. Other methods may not detect HABs until shoreline algae populations are in full bloom. Therefore it is more difficult to mitigate the negative impacts on human health and the economy (Roesler and McLeary-Etheridge, 1998).

The satellite ocean color dataset encompasses almost thirty years (Robinson and Mitchelson, 1983; Muller-Karger et al., 2005; McClain et al., 2006; Yoder and Kennelly, 2006) enabling temporal variability studies of marine ecosystem of short-term (e.g., seasonal; Gregg and Conkright, 2001) and long-term (e.g., decadal; Gregg and Conkright, 2002). These types of satellite sensors were designed to provide global estimates of phytoplankton biomass by using the spectrally-resolved backscattered light coming out of the sea to give information about the biological state of the upper ocean layer by a proxy of phytoplankton biomass, \(i.e.,\) chlorophyll-a concentration at the surface layer plus associated phaeopigments (denoted hereafter as Chl; \(i.e.,\) Gordon et al., 1983; Carder et al., 1999; 2004; O’Reilly et al., 2000). Another set of bio-optical attributes retrieved by remotely sensed measurements of the color of the ocean is the seawater inherent optical properties (Hoge et al., 2001).

The capability of retrieving such key parameters from space potentially addresses a large range of environmental issues of great interest to the scientific community and for the management of resources. Many studies have explored the spatial and temporal variability of phytoplankton concentration using ocean color observations acquired from space (\(e.g.,\) Muller-Karger et al., 1991; Tang et al., 1998; Liew et al., 1999; Kampel, 2003; Kampel, 2005). These data have been instrumental in elucidating the physical constraints on algal blooms and their biogeography. In addition, the fishery industry also
has begun using these data to locate fishing grounds (Mano et al., 2003). To date, the effort to make use of more than one visible sensor in the same ecological analysis has been focused almost exclusively on calibration, validation or merging data (Kwiatkowska and Fargio, 2002; Blondeau-Patissier et al., 2004; Darecki and Stramski, 2004; Barbini et al., 2005; Maritorena and Siegel, 2005; Zibordi et al., 2006). Only few have directly tackled hydrographic studies (Denman and Abbott, 1994; Gregg and Conkright, 2002; Dall’Olmo et al., 2005; Zhang et al., 2006a), and none has yet evaluated HABs with a long-term dataset using a multi-sensor approach.

An ocean color sensor was first employed to investigate *K. brevis* by Mueller (1979), using an aircraft-based instrument. Subsequently, Steidinger and Hadadd (1981) and Haddad (1982) were among the first to show the potential use of a satellite ocean color sensor to study HABs in the CWFS. Since then, many authors have tried to use ocean color measurements acquired from space to differentiate between *K. brevis* blooms and non-bloom waters (Cullen et al., 1997; Kahru and Mitchell, 1998; Kirkpatrick et al., 2000), while others have attempted to relate satellite-Chl patterns with possible HABs (Vargo et al., 1987; Tyler & Stumpf, 1989; Schofield et al., 1999; Lavander and Groom, 2001; Holt et al., 2003; Ahn and Shanmugam, 2006).

A variety of remote sensing methods have been used to track HABs from space, such as patterns of sea surface temperature (SST; Gower, 1994; Huang and Lou, 2001) and synthetic aperture radar (SAR; Svejkovsky and Shandley, 2001). Tester and Stumpf (1998) presented a description of the ocean color satellite sensors capable of monitoring HABs. Recently, some studies have presented promising ocean color algorithms for satellite measurements for detecting the Florida Red Tide, e.g., Thomas (2000)
mentioned the use of annual Chl anomalies; Cannizzaro et al. (2002a; 2002b) introduced an algal taxonomic approach; Cannizzaro (2004) and Cannizzaro et al. (2008) developed a technique based on Chl and inherent optical properties of the water; and the utility of a parameter related to photosynthetic efficiency (i.e., fluorescence line height) has been shown by Hu et al. (2005a) and Cheng et al. (2008). Of the existing methods, so far, only anomalous Chl patterns (Stumpf et al., 2003b; Tomlinson et al., 2004) have been operationally implemented by the NOAA’s HAB Forecasting System (HabFS, 2007) to locate possible HABs in the Gulf of Mexico.

Despite the benefits of using passive ocean color measurements from sensors onboard satellites, there are weaknesses (Freitas and Noernberg, 2007), especially in locating features in coastal shallow water regions where HABs are often located. A major problem is how to relate the spatial sampling resolution of the in situ measurement with the satellite pixel size. The ocean color satellite ground resolution usually falls between 175~6000 meters at satellite nadir (IOCCG, 2007), and may not allow smaller, patchy blooms to be resolved or tracked (Franks, 1997b). Also, although many algal blooms tend to concentrate near or at the air-sea interface (including K. brevis with its positive phototactic nature; Heil, 1986; Heil et al., 2008), some blooms may occur at depths below the detecting capability of the ocean color sensors. The satellite-retrieved Chl may suffer from uncertainties in regions with high concentrations of suspended sediments (Darecki and Stramski, 2004; Hu et al., 2000; 2006a; Hyde et al., 2007), and can be influenced by bottom reflection (Gould and Arone, 1998; Lee et al., 2001; Ohde and Mueller, 2001; Ohde and Siegel, 2001; D’Sa et al., 2002). The loss of data due to specular reflection (i.e., glint) of the Sun on the ocean surface (Wang and Bailey, 2001) and the lack of data
caused by the space between swaths are additional limitations of this type of dataset. Unfortunately, to date the current generation of hyper spectral satellite sensors is not yet practical for observations within the framework of detecting HABs in an operational basis. Thus the available remote sensing sensors with adequate characteristics to study HABs only acquire information in discrete spectral channels limited by atmospheric spectral window bands and on-board data storage capacity.

Unfortunately to date, satellite visible sensors are not yet hyper spectral, thus only acquiring information in discrete spectral channels limited by atmospheric spectral window bands and on-board data storage capacity.

But the major problem of the ocean color dataset comes from the fact that measurements taken in the visible part of the electromagnetic spectrum are strongly dependent on the unpredictable occurrence of clear skies. Cloud cover effectively blocks the retrieval of the ocean surface information, which may mean that usable images may be limited to one or two per week per sensor. Satellite-acquired information near or at the outer edges of clouds may also need to be discarded (Gordon, 1998). The complex bio-optical signature of turbid regions also presents challenges to ocean color remote sensing (Bricaud et al., 2002; D’Ortenzio et al., 2002; Hu et al., 2003c; Hyde et al. 2007), as do the imperfection of the correction involved in removing the large atmospheric signal reaching the satellite sensor (Gordon, 1978; 1997; Gordon and Clark, 1981; Gordon and Wang, 1994; Siegel et al., 2000). Moreover, satellite-monitoring systems are not yet able to replace in situ methods. Ship cruises can investigate depth-dependent process, and the continuous information provided by Eulerian mooring sampling and Lagrangian drifting buoys contribute enormously to oceanographic knowledge. Therefore, the different
information content of *in situ* observations and satellites imagery should be used to complement each other (Smith et al., 1987).

Because Chl is present in all sorts of plants besides HABs, a single, satellite-derived Chl spectral signature cannot be used alone as a unique tracer to permit confident classification among different phytoplankton species (Garver et al., 1994). It is possible to work around this problem by invoking spectral features (Canizzarro et al., 2002a; 2002b; Canizzarro, 2004), and then, the remote classification of distinct algal communities may be accomplished (Schofield et al., 1999; 2006). Satellite ocean color measurements may have the potential to provide algal taxonomic discrimination if particular criteria are met. The present requirement is that the observed phytoplankton species should comprise a considerable fraction of surface biomass (preferentially represented by a monospecific bloom), and an exclusive set of bio-optical signatures should be established (Schofield et al., 1999; 2006; cf. Limnology and Oceanography 1989: vol. 34-8 and Journal of Geophysical Research 1995: vol. 100-C7). This has been accomplished for the blue-green algae *Trichodesmium* (Borstad et al., 1992; Subramaniam et al., 1999a; 1999b; 2002) and for *coccolithophores*, which are easily seen by satellite data due to their characteristic calcite shells (Balch et al., 1991; Brown and Yoder, 1994; Cokacar et al., 2004).

Ultimately, the paucity of specific algorithms may limit the use of the remote sensing techniques in the analysis of HABs. The lack of long-term investigations studying HABs with satellite data is another strong motive for continued study.
APPENDIX 5

CONCEPTUAL ASPECTS OF OCEAN COLOR REMOTE SENSING

Inherent optical properties of the water can be partitioned into classes of optically active constituents (Doerffer and Fischer, 1994): Colored Dissolved Organic Matter (CDOM; also referred to as yellow substance, gelbstoff or gilvin; Kahru and Mitchell, 2001; Kirkpatrick et al., 2003); phytoplankton (ph; Lee and Carder, 2004); and non-chlorophyllous particulate matter suspended in the water, or simply detritus (d). These last two optically active constituents can be referred to as total particles \((p = ph + d)\). A fourth optically active constituent is represented by the water molecules themselves \((w)\), which is constant and known (Pope and Fry, 1997; Smith and Baker, 1981). CDOM does not scatter light, but does absorb it (Brando et al., 2006). Because CDOM and detritus absorb light in a similar manner, they are usually grouped together (Muller-Karger et al., 2005). Hence, the coefficients \(a(\lambda)\) and \(b(\lambda)\) can be calculated as the sum of their individual optically active constituents classes as follows:

\[
a(\lambda) = a_w(\lambda) + a_p(\lambda) + a_{\text{CDOM}}(\lambda)
\]

\[
b(\lambda) = b_w(\lambda) + b_p(\lambda)
\]

Given that the concentrations of optically active constituents exhibit vast temporal and spatial variability (Gould and Arnone, 1997; Arnone et al., 2004), especially
considering coastal and estuarine ecosystems, so does the color of the ocean \( i.e., R_{rs(\lambda)} \) which also varies in time and space due to unique light absorbance \( c(\lambda) = a(\lambda) + b(\lambda) \). The oceanographic community usually classifies water bodies according to their optically active constituent (Jerlov, 1951; 1976). A simple category, represented by regions away from coastal, terrigenous source is referred to as Case-1 water (Loisel and Morel, 1998). In this instance, the color of the ocean is mainly determined by phytoplankton, with the other inherent optical properties of the water co-varying. A more complex condition, \( i.e., \) Case-2 water (Morel and Prieur, 1977), has independent optically active constituents concentrations and the water color is strongly influenced by agents other than phytoplankton, such as the non-biogenic optically active constituents.

The highly complex optical nature of coastal environments contributes to the intricacy of the optically active constituents, which add numerous challenges to the ocean color remote sensing signal retrieval (Morel and Bélanger, 2006). For instance, in turbid zones \( i.e., \) Case-2 water the interpretation of the optical signal sensed by satellites may include significant uncertainties (Darecki and Stramski, 2004; Hu et al., 2000; 2006a). Although numerous studies have focused on overcoming the limitations imposed by the shallow coastal proximity worldwide (Darecki and Stramski, 2004; Maritorena and Siegel, 2005; Zibordi et al., 2006; Hyde et al. 2007), sediment-rich water, such as that found on the West Florida Shelf, still causes satellite Chl retrievals to contain significant errors (Hu et al., 2003a; 2003b; 2005a). Therefore, any remote sensing study aiming at differentiating HABs from other constituents \( i.e., \) other non-toxic blooming species or suspended sediment) becomes very complex and might not be straightforward to resolve (Garver et al., 1994).
The underwater light field in optically shallow waters has a vital relevance for any electromagnetic radiation analysis (Bricaud and Morel, 1986; Cannizzaro and Carder, 2006). Secchi disks have traditionally been used to estimate water transparency (Thiemann and Kaufmann, 2002; Kratzer et al., 2003; Beenton, 1999) but this method can undergo human interpretation errors (Brando et al., 2006). For satellite purposes, the effective penetration depth (EPD; or the first attenuation length; Mueller, 2000) is more appropriate to measure light availability as a function of depth. For a homogeneous ocean, the EPD, as defined by Gordon and McCluney (1975), corresponds to the upper layer from which 90% of the $L_w(\lambda)$ emerges. Gordon and McCluney demonstrated that the EPD is equivalent to the depth where the surface incident irradiance falls $e^{-1}$ (or the inversion of the downwelling diffuse attenuation coefficient; $k^{-1}$). They found maximum EPDs of $\sim1.5$ m (at $\lambda \sim600$ nm) for Case-2 water, and $\sim55$ m (at $\lambda \sim475$ nm) for Case-1 water.

Whenever EPDs are deeper than the physical depth, bottom reflection may significantly contribute to the satellite signal (Lee et al., 1999a; 1999b), and the use of standard Chl band-ratio algorithms (e.g., Gordon et al., 1983; Carder et al., 1999; 2004; O’Reilly et al., 2000) may be contaminated by strong bottom reflectance (Gould and Arnone, 1998; Lee et al., 2001; Ohde and Mueller, 2001; Ohde and Siegel, 2001; D’Sa et al., 2002). Bailey and Werdell (2006) outlined the steps of the National Aeronautics and Space Administration’s (NASA) standard processes for validating satellite ocean color products. To ensure that bottom reflection is not contributing to the satellite signal being validated, Bailey and Werdell showed that the light penetration depth limit into the water column should be expressed in terms of Apparent Optical Depth ($AOD = 1.3 / K$ at $\lambda$...
~490 nm) instead of EPD (i.e., $1 / K_{490}$), since AOD is a more conservative criterion than EPD. The use of the AOD is cautiously justified to avoid any possible miscalculation in satellite retrievals of $K_{490}$. 
Gustavo de Araújo Carvalho was born in the city of Rio de Janeiro, Brazil, around thirty one years ago. A year after he was brought to Earth, his parents Sandra Barbosa and Francisco Salgado gave him a younger brother, Guilherme. They both received elementary and secondary education at Colégio Santa Úrsula, Rio de Janeiro, Brazil. In 1996 Gustavo entered the Department of Oceanography at the Universidade do Estado do Rio de Janeiro (UERJ) for a bachelor’s degree in Oceanography, while his brother studied Psychology at the same University.

During his undergraduate studies at UERJ he took many specialized courses on water quality, coastal oceanography, GIS, and remote sensing related topics. Involved in several internships, he had the opportunity to gain field experience in a number of scientific programs, especially those related with water quality of the Coastal-Urban Ecosystems of Rio de Janeiro where he worked together with experts as Dr. David Zee and Carla Sabino. As a certified scuba diver, he was one of the divers responsible for the maintenance of a tide gauge and a current meter located 15 meters below sea level in Ilha Grande, south of Rio de Janeiro city. Enjoying being at sea, he gained some at-sea experience while onboard various short and long scientific cruises.

As an undergraduate, he was employed as a junior technician in a well-known Brazilian remote sensing company, now called Oceansat-peg Fugro S.A.. In this post, in addition to being in charge of an HRPT satellite receiving station, he was also responsible for the operational data processing and analysis, focused on fishing forecasts, oil spills detection, and determination of surface current patterns. His undergraduate research
thesis utilized satellite-based sea surface temperature data from the AVHRR sensor to create indices to classify the intensity of the Cabo Frio upwelling system (23°S, 42°W). These indices were then correlated with local wind fields derived from measurements of the QuikScat satellite to determine the direction and persistence of the wind necessary to upwell the South Atlantic Central Water. In this research, which was guided by two outstanding professionals, Dr. Alexandre Cabral (Managing Director, Oceansat-peg Fugro) and Dr. Marcos Fernandez (Research Professor, UERJ), he was awarded as one of the best presentations at the XI Brazilian Remote Sensing Symposium, in September, 2003.

In August 2004 he was admitted to the Rosenstiel School of Marine and Atmospheric Science of the University of Miami’s, within the Division of Meteorology and Physical Oceanography, to work with Dr. Edward J. Kearns. About a year later, the sad news that Dr. Kearns joined the National Park Service was pleasurably compensated by the opportunity to study under the guidance of Dr. Peter J. Minnett. Gustavo’s research was focused on the optical properties of the water related to harmful algal blooms (Karenia brevis) off the West Florida coast. The conclusions of his investigation, using for the first time in this region an extensive dataset from the MODIS-Aqua sensor for detecting these types of blooms, are presented here for the award of his Master of Science degree in Meteorology and Physical Oceanography.

Permanent Address: 4600 Rickenbacker Causeway - MPO, Miami, Florida, 33149