



Western Michigan University
ScholarWorks at WMU

Dissertations

Graduate College

4-2019

Developing Early Warning Systems for Debris Flows and Harmful Algal Blooms

Sita Karki

Western Michigan University, karksita@isu.edu

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Geology Commons

Recommended Citation

Karki, Sita, "Developing Early Warning Systems for Debris Flows and Harmful Algal Blooms" (2019).
Dissertations. 3417.

<https://scholarworks.wmich.edu/dissertations/3417>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



DEVELOPING EARLY WARNING SYSTEMS FOR DEBRIS FLOWS
AND HARMFUL ALGAL BLOOMS

by

Sita Karki

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Geological and Environmental Sciences
Western Michigan University
April 2019

Doctoral Committee:

Mohamed Sultan, Ph.D., Chair
Alan E. Kehew, Ph.D.
Matt Reeves, Ph.D.
Richard Becker, Ph.D.
Racha El Kadiri, Ph.D.

Copyright by
Sita Karki
2019

ACKNOWLEDGMENTS

There were many seen and unseen hands behind the successful completion of this work. I offer my heartfelt gratitude to all the individuals who supported me in numerous ways possible during my five years in Kalamazoo. My first reason to come to Michigan was the Michigan Geological Survey, and I am foremost thankful to John Yellich and Dr. Alan Kehew for helping me transition to graduate studies. I am forever grateful for all the assistance and guidance they offered during my time with survey and afterwards as a graduate student. Their support provided the strong foundation for all my achievements as a graduate student.

I offer sincere gratitude to my advisor Dr. Mohamed Sultan, without him I would not have been able to complete this research and achieve other accomplishments that followed it. I am extremely grateful for his unwavering trust when I needed it the most. I would like to remember all my friends at the Earth Science Remote Sensing Laboratory who were there to encourage and support me during the ups and downs of graduate studies. I am forever thankful to my parents for the unconditional love they poured over me and for their support for all the decisions I made throughout the years. At last, I would like to thank someone very special who came the last but provided the meaning to my work.

Sita Karki

DEVELOPING EARLY WARNING SYSTEMS FOR DEBRIS FLOWS AND HARMFUL ALGAL BLOOMS

Sita Karki, Ph.D.

Western Michigan University, 2019

This study focused on developing early warning systems for two types of geohazards using methods that heavily rely on remote sensing data. The first investigation attempted to develop a prototype version of an early warning system for landslide development, whereas the second focused on harmful algal bloom prediction.

Construction of intensity-duration (ID) thresholds, and early warning and nowcasting systems for landslides (EWNSL) are hampered by the paucity of temporal and spatial archival data. This work represents significant steps towards the development of prototype EWNSL to forecast and nowcast landslides over Faifa Mountains in the Red Sea Hills. The developed methodologies rely on temporal, readily available, archival Google Earth and Sentinel-1A imagery, precipitation measurements, and limited field data to construct an ID threshold for Faifa. Adopted procedures entailed the generation of an ID threshold to identify the intensity and duration of precipitation events that cause landslides in the Faifa Mountains, and the generation of pixel-based ID curves to identify locations where movement is likely to occur. Spectral and morphologic variations in temporal Google Earth imagery following precipitation events were used to identify landslide-producing storms and to generate the Faifa ID threshold ($I = 4.89 \cdot D - 0.65$). Backscatter coefficient variations in radar imagery were used to generate pixel-based ID curves and to identify locations where mass movements are likely to occur following

landslide-producing storms. These methodologies accurately distinguished landslide-producing storms from non-landslide producing ones and identified the locations of these landslides with an accuracy of 60%.

Over the past two decades, persistent occurrences of harmful algal blooms (HAB; *Karenia brevis*) have been reported in Charlotte County, southwestern Florida. I developed data-driven models that rely on spatiotemporal remote sensing and field data to identify factors controlling HAB propagation, provide a same-day distribution (nowcasting), and forecast their occurrences up to three days in advance. I constructed multivariate regression models using historical HAB occurrences (213 events reported from January 2010 to October 2017) compiled by the Florida Fish and Wildlife Conservation Commission and validated the models against a subset (20%) of the historical events. The models were designed to capture the onset of the HABs instead of those that developed days earlier and continued thereafter. A prototype of an early warning system was developed through a threefold exercise. The first step involved the automatic downloading and processing of daily Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua products using SeaDAS ocean color processing software to extract temporal and spatial variations of remote sensing-based variables over the study area. The second step involved the development of a multivariate regression model for same-day mapping of HABs and similar subsequent models for forecasting HAB occurrences one, two, and three days in advance. Eleven remote sensing variables and two non-remote sensing variables were used as inputs for the generated models. In the third and final step, model outputs (same-day and forecasted distribution of HABs) were posted automatically on a web map. Our findings include: (1) the variables most indicative of the timing of

bloom propagation are bathymetry, euphotic depth, wind direction, sea surface temperature (SST), ocean chlorophyll three-band algorithm for MODIS [chlorophyll-a OC3M] and distance from the river mouth, and (2) the model predictions were 90% successful for same-day mapping and 65%, 72% and 71% for the one-, two- and three-day advance predictions, respectively. The adopted methodologies are reliable at a local scale, dependent on readily available remote sensing data, and cost-effective and thus could potentially be used to map and forecast algal bloom occurrences in data-scarce regions.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES.....	v
LIST OF FIGURES	vi
CHAPTER	
1. INTRODUCTION.....	1
2. PROJECT I.....	3
2.1 Introduction.....	3
2.1.1 Study Area.....	7
2.2 Methods.....	10
2.2.1 Generation of the ID threshold for the Faifa Mountains.....	13
2.2.2 Selection, calibration, and pre-processing of radar images	20
2.2.3 Generation of backscatter difference images	22
2.2.4 Refinement and validation of the model.....	23
2.2.5 Pixel-based adaptation of Faifa ID threshold as a predictive tool	33
2.2.6 Validation of ID threshold.....	35
2.3 Results and Discussion.....	37
2.4 Conclusion	40
2.5 References.....	42

Table of Contents—Continued

3. PROJECT II.....	50
3.1 Introduction.....	50
3.2 Methods	54
3.2.1 Step 1: Data collection and processing.....	55
3.2.2 Step 2: Development of prediction model.....	66
3.2.3 Step 3: Prediction.....	68
3.3 Results and Discussion.....	69
3.4 Conclusion	78
3.5 References	79
4. CONCLUSION	97

LIST OF TABLES

1. Intensity and duration of the precipitation events used for the construction of the Faifa intensity-duration curve.....	16
2. Field observations collected (February 26 to March 7, 2016) for the assessment of radar-based distribution of active landslides, the areas exceeding 1 standard deviation on difference images.....	30
3. Overview of the inputs, outputs, processor and relevant parameters applied to extract level 2 products.....	57
4. Selected variables with their relative abilities to explain variability.....	70
5. Multivariate regression coefficients for each variable in predicting blooms for same-day mapping, and one-, two- and three-day advanced predictions.....	71

LIST OF FIGURES

1. Location of the study area.....	7
2. Maps showing the study area... ..	9
3. Flow chart summarizing the developed methodology	12
4. Intensity-duration scatter plot generated from precipitation events.	20
5. Pictures showing the landslides proximal to roads	24
6. Google Earth Imagery showing the spatial refinement procedures.... ..	26
7. Google Earth imagery showing the distribution of debris flows.....	29
8. Demonstration of pixel-based ID curves	35
9. Demonstration of the prediction result.....	37
10. Figure showing the study area in Charlotte County in Florida.....	53
11. Three-step workflow established for harmful algal blooms forecasting.....	55
12. Maps showing the mean values for the significant variables.....	66
13. Assessment of the accuracy of the generated multivariate models	72

CHAPTER 1

INTRODUCTION

Natural disasters are common in many parts of the world and it causes loss of property and life. In this study, I attempted to address two types of natural hazards, landslides and algal blooms. Specifically, I wanted to develop methodologies by which the negative impacts of such events could be reduced. I accomplished that by first studying the historical events and then used that understanding to predict the future occurrences (temporarily and spatially) of these geohazards. I selected two such geohazards for which the knowledge from the past can be used to develop the predictive framework. The two geohazards chosen for this study were landslides in the Faifa Mountains in Jazan province of Saudi Arabia and the harmful algal blooms in the coastal waters of Charlotte County in Florida.

The development of predictive models demands the use of extensive historical field data. Unfortunately, in both of our study areas, the wealth of relevant field datasets that is needed to develop such forecasting systems does not exist. To compensate for the lack of relevant datasets, I used a wide range of remote sensing data sets. There are many benefits of using remote sensing data. These include: acquisition of data with the same observational parameters (data consistency), wide geographic coverage, and cost-effectiveness (compared to field data collection). Throughout my studies, I demonstrate that although field data is an integral part of research for validation and testing purpose, remote sensing data can be used to compensate for the lack of field data to develop predictive models.

I studied landslides in Faifa Mountains to develop the prototype version of early warning system in Saudi Arabia. For this study, I used the combination of Google Earth imagery and radar data (Sentinel-1A) to acquire information pertaining to historical landslides data. Similarly, I used rainfall data from satellite to compensate for the lack of consistent rain gauge rainfall data. For the study on harmful algal blooms, I used the daily acquisition of satellite data collected by Moderate Resolution Imaging Spectroradiometer (MODIS) to extract temporal variables (factors) that cause the onset of algal blooms in the Charlotte bay. In this research work I develop two semi-automated modules for two early warning systems that heavily rely of readily available remote sensing data and limited field data; the two techniques provide replicable models to be applied in similar settings elsewhere with minimal customization.

CHAPTER 2

PROJECT I

2.1 Introduction

Mountainous areas worldwide with steep slopes, high precipitation, and limited vegetative cover often experience landslides. Two main types of landslides are often reported from Faifa Mountains in Saudi Arabia. The first are debris flows that occur when water-saturated soils (largely from weathered bedrock and fragmented rock) move down mountainsides, get channelled into streams, pick up objects along their paths, and deposit their thick load down the valley slopes (Iverson, 1997). The second type results from failure along preexisting fracture planes (Lowell, 1990) that occurs when the following conditions are met (Norrish and Wyllie, 1996): (1) the strike of the planar discontinuity is similar (within 20°) to the strike of the slope face, (2) the dip of the planar discontinuity is less than that of the slope face and oriented in the same general direction, (3) the dip of the planar discontinuity is greater than the angle of the friction of the surface, and (4) the friction angle of the rock material is partially controlled by the size and shape of the grains exposed on the fracture surface and by the mass of the block above the planar discontinuity (Alharbi et al., 2014). These two types of landslides pose a substantial threat to human life and property in mountainous areas, especially in populated regions that are witnessing unplanned urbanization.

One such area is the Faifa Mountains (area: 119 km^2) in the Jazan Province of Saudi Arabia (Fig. 1a). The population density in Faifa is high ($\sim 35,000$ inhabitants in 137 km^2 ; MMRA, 2017), receives relatively high precipitation (mean annual

precipitation [MAP]: 252 mm/year; Fig. 1b) compared to the remaining parts of Saudi Arabia (83 mm), has steep slopes (up to 65°, Fig. 2b), and witnesses frequent landslide occurrences (1–4 events/year during the study period [2007 to 2017]). Unfortunately, monitoring programs for landslide occurrences (date and time) and conditions (intensity and duration of precipitation) are very limited in Faifa. If such programs existed, they would have generated the archival landslide data needed for the construction of intensity-duration (ID) relationships and for the development of an early warning and nowcasting system for landslides (EWNSL). The paucity of such data in Faifa is largely caused by its rugged nature and its limited road network coverage; both factors rendered many areas inaccessible and hindered the development of monitoring systems. Despite the absence of organized monitoring programs for landslides in the study area, a few were recorded by the Saudi Geological Survey (SGS) in the past few years. Our field observations, and those collected by the SGS in the study area and in its surroundings (Youssef et al., 2014) revealed that debris flows are by far the most prominent landslide type in the study area. To compensate for the deficiencies in field data in Faifa, I complemented the available field data with observations extracted from readily available temporal remote sensing data. These include high-resolution Google Earth images (spatial resolution: 15 m to 15 cm), Sentinel-1A Ground Range Detected (GRD level-1 Synthetic Aperture Radar (SAR) data; spatial resolution: 10 m), satellite-based precipitation data including Tropical Rainfall Measuring Mission (TRMM, 3-hourly_3B42 v7, spatial resolution: 0.25°x 0.25°, ~ 30 km in Faifa) and Global Precipitation Measurement (GPM) IMERG Final Precipitation L3 Half Hourly (V05, spatial resolution: 0.1° x 0.1°, ~ 12 km in Faifa), and a high-resolution digital elevation

model (TanDEM-X DEM; spatial resolution: 12 m). Although GPM provided higher temporal resolution (half-hourly) data compared to TRMM (3-hourly) data, both sensors provided rainfall intensity data in mm/h allowing continuous measurements of rainfall with consistent observational parameters and acceptable (~67%) to high (87%) correlation at pixel and basin scale respectively (Tang et al., 2016). Field observations were collected (period: 2014 to 2016) from the areas that witnessed landslides following precipitation events. In this manuscript, I develop procedures that take advantage of readily available remotely acquired data, and limited field data, to develop ID thresholds for the study area, procedures that could potentially be applied to similar areas with limited field data.

A recent review by Segoni et al. (2018) shows that the majority of the ID-threshold development studies were conducted over well-monitored areas where landslide-related archival data are available from reports, surveys, field works sets (e.g., Burtin et al. 2009; Erener and Düzgün, 2013; Staley et al., 2013; Lagomarsino et al., 2015; Vallet et al., 2016; Piciullo et al., 2017), or even from automatic systems (Battistini et al., 2017). In these areas, several advanced thresholding techniques (e.g., geotechnical process-based, empirical, and rainfall measurement-based) are common (Guzzetti et al., 2007). Unfortunately, Faifa lacks the historical landslides archives and rain gauge measurements to implement any of these thresholding techniques. Using observation from limited field data and satellite-based data sets (e.g., intensity and duration of precipitation, location of landslides), I developed rainfall-based ID thresholds. In this respect, our approach does not require extensive archival field data sets to generate ID thresholds and thus, the approach could potentially be applied in

many of the world's mountainous locations lacking adequate archival field data.

The majority of the ID thresholds that were constructed for various landslide types (e.g., shallow landslides and debris flow [Caine, 1980; Innes, 1983, Crosta and Frattini, 2001; Aleotti, 2004; Jakob et al., 2012]; soil slips [Clarizia et al., 1996], and postfire debris flow [Cannon and Gartner, 2005; Cannon et al., 2011]) provide the magnitude and intensity of rainfall that triggers landslides, but not the locations where they are likely to occur. I generated a unified ID threshold for the Faifa Mountains to identify the landslide triggering precipitation events and pixel-based thresholds to identify locations where landslides are likely to occur. The pixel-based threshold is adjusted to the response of the individual pixels to historical rainfall events.

Several advances in rainfall thresholding techniques were developed to account for the role of antecedent rainfall conditions preceding landslide development (e.g., Kim et al., 2014, Hong et al., 2017). Others consider software applications that rely on rain gauge records, extensive historical data catalogue or advance statistical analyses (e.g., Lagomarsino et al., 2015; Peruccacci et al., 2017; Rossi et al., 2017b). Such techniques cannot be applied in the Faifa area due to the absence of such measurements, instead I adopt the minimum thresholding technique that was successfully applied in several studies (e.g., Caine, 1980; Larsen and Simon, 1993; Cannon et al., 2008; Brunetti et al., 2010; Berti et al., 2012). I acknowledge that if and once such data sets become available for Faifa, the ID thresholds need to be updated to enhance their performance (Rosi et al., 2015).

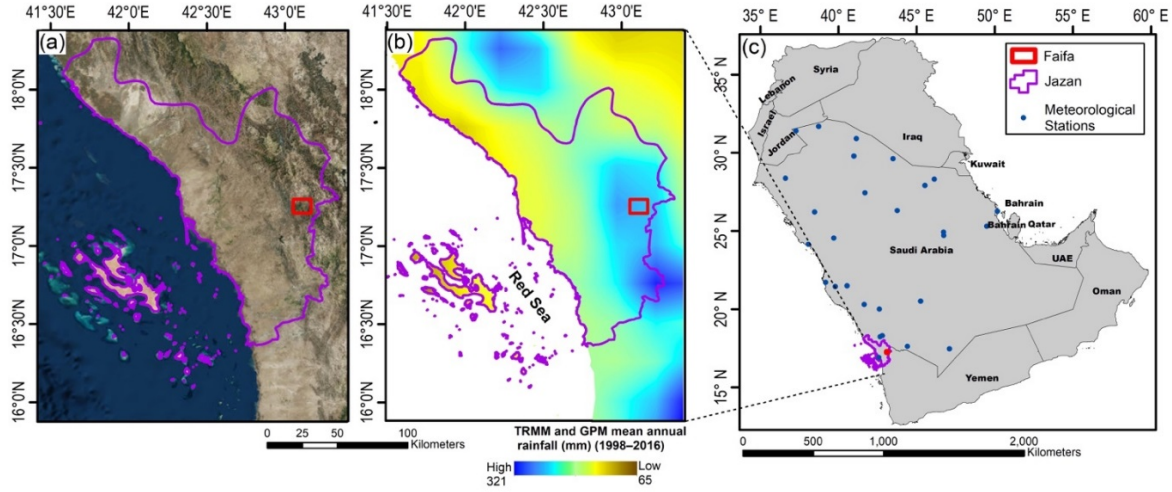


Figure 1. Location of the study area. (a) Faifa Mountains within the Jazan province. (b) Mean annual precipitation (MAP; 1998–2016) extracted from TRMM (1998–2014, 3B43 level-3, monthly, spatial resolution: $0.25^{\circ} \times 0.25^{\circ}$) and GPM (2014–2016, IMERG level-3, monthly, spatial resolution: $0.1^{\circ} \times 0.1^{\circ}$) showing the higher regional rainfall around Faifa Mountains (MAP in Faifa: 252 mm/year) in the southeast part of the Jazan Province. (c) Distribution map of the meteorological stations in the Arabian Peninsula (Mashat and Basset, 2011) in Saudi Arabia.

2.1.1 Study area

The study area (119 km²) lies within the Red Sea Hills, and covers an area (17.20° N to 17.29° N and from 43.05° E to 43.16° E) proximal to the Saudi-Yemeni border (Figs. 1a and 1c). The elevation is high (range: 259 to 1817 meters above mean sea level [m.a.m.s.l]) compared to the surrounding lowlands, the topography is steep (slope as high as 67° ; Figs. 2a and 2b), and vegetation is extensive over the mountains but sparse in the surrounding lowlands, as shown in the normalized difference vegetation index (NDVI) map (Fig. 2c). The Faifa region is located within the north to

northeast trending Tayyah tectonic belt that consists of a complex of metamorphosed volcanic and pyroclastic rocks of basaltic, andesitic, and clastic metasedimentary rocks (Greenwood et al., 1983) that were generated and accreted in an island arc setting some 800 to 900 Ma (Stoeser and Camp, 1985). The Faifa Mountains are predominantly composed of highly foliated, deformed, and weathered rocks of variable compositions including granite gneiss, amphibolite schist, phyllite, quartzite, biotite and sericitic schist that are intruded by a massive intergranular syenite (Schmidt et al., 1973; Greenwood, 1979; Greenwood et al., 1983; Alharbi et al., 2014). The area is highly dissected by north-south, northwest-southeast, and east-west trending fault and fracture systems (Fairer, 1985; Alharbi et al., 2014; Fig. 2d). The presence of highly weathered, foliated, and deformed rocks, together with the high elevations, steep slopes and sparse vegetation makes this area prone to landslides even under modest precipitation intensities.

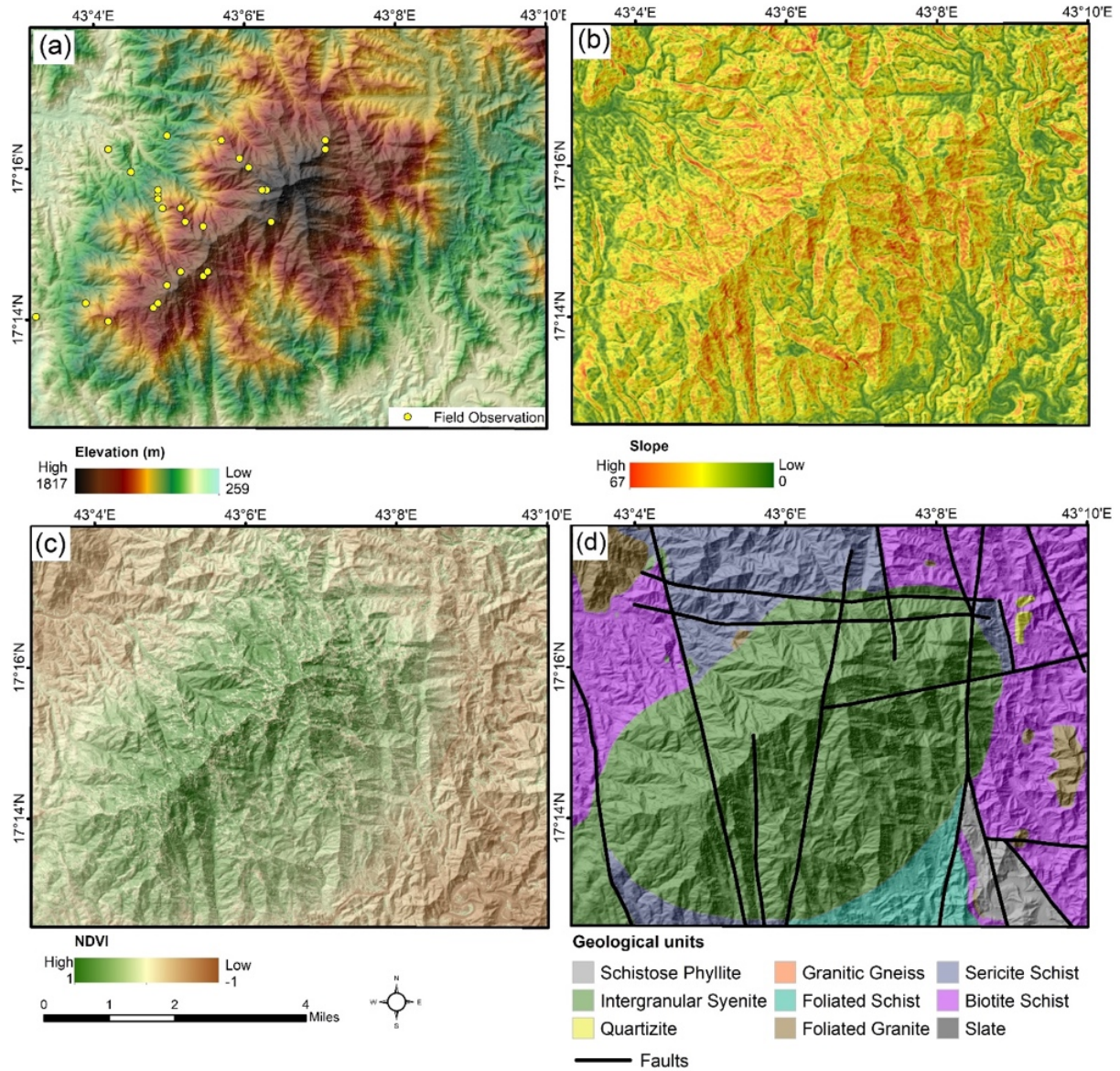


Figure 2. Maps showing the study area. (a) Elevation map showing locations where field observations were collected for the period extending from February 26th to March 7th, 2016. (b) Slope map generated from TanDEM-X DEM. (c) NDVI map generated from Sentinel-2 data (date of acquisition: 2017-06-05). (d) Geologic map for the Faifa Mountains (after Fairer, 1985 and Alharbi et al., 2014).

Analysis of TRMM (1998- 2014) and GPM (2014-2016) measurements for the period 1998 to 2016 revealed sparse precipitation over the Saudi Arabian landscape (MAP: 83 mm/year), relatively higher precipitation over the Red Sea Hills in western Saudi Arabia, which receive relatively higher MAP of 108 mm/year. These analyses involved spectral resampling of GPM data to match the TRMM resolution followed by the zonal and pixelwise averaging over the indicated time span. Comparison of TRMM to the resampled and averaged GPM measurements over the study area revealed highly correlated (>85%) values during the period of overlap (March to September of 2014). A progressive increase in overall rainfall over Faifa was noted over the past six years (MAP: 2010–2016: 315 mm/year; 1998–2009: 227 mm/year) with the wettest year on record in 2016 (total annual rainfall: 450 mm). Two systems of wind regimes are responsible for the rainfall over Faifa: (1) northerly and northwesterly winter cyclonic regimes from the Mediterranean, and (2) summer monsoons from the Arabian Gulf and the Indian Ocean (Alsharhan et al., 2001).

2.2 Methods

The methodology I developed entailed two main steps. The ID curve for the Faifa Mountains was first generated to identify storms that caused landslides (temporal analysis; Sect. 2.1), and then pixel-based ID curves were constructed to identify the locations where movement is likely to occur (spatial analysis, Sect. 2.2–2.4). The latter step involved: (1) selection, calibration, and pre-processing of radar images; (2) generation of backscatter coefficient difference images as a measure of surface roughness change due to precipitation-induced landslides; and (3) development,

refinement, and validation of the model to identify pixels susceptible to movement under user-defined precipitation conditions.

Data used for the study include: (1) Temporal Google Earth imagery, (2) Sentinel-1A radar imagery; (3) TRMM and GPM; (4) field observation of landslide (debris flow and failure along fracture plane) locations; and (5) the TanDEM-X DEM. A detailed flow chart is presented in Fig. 3.

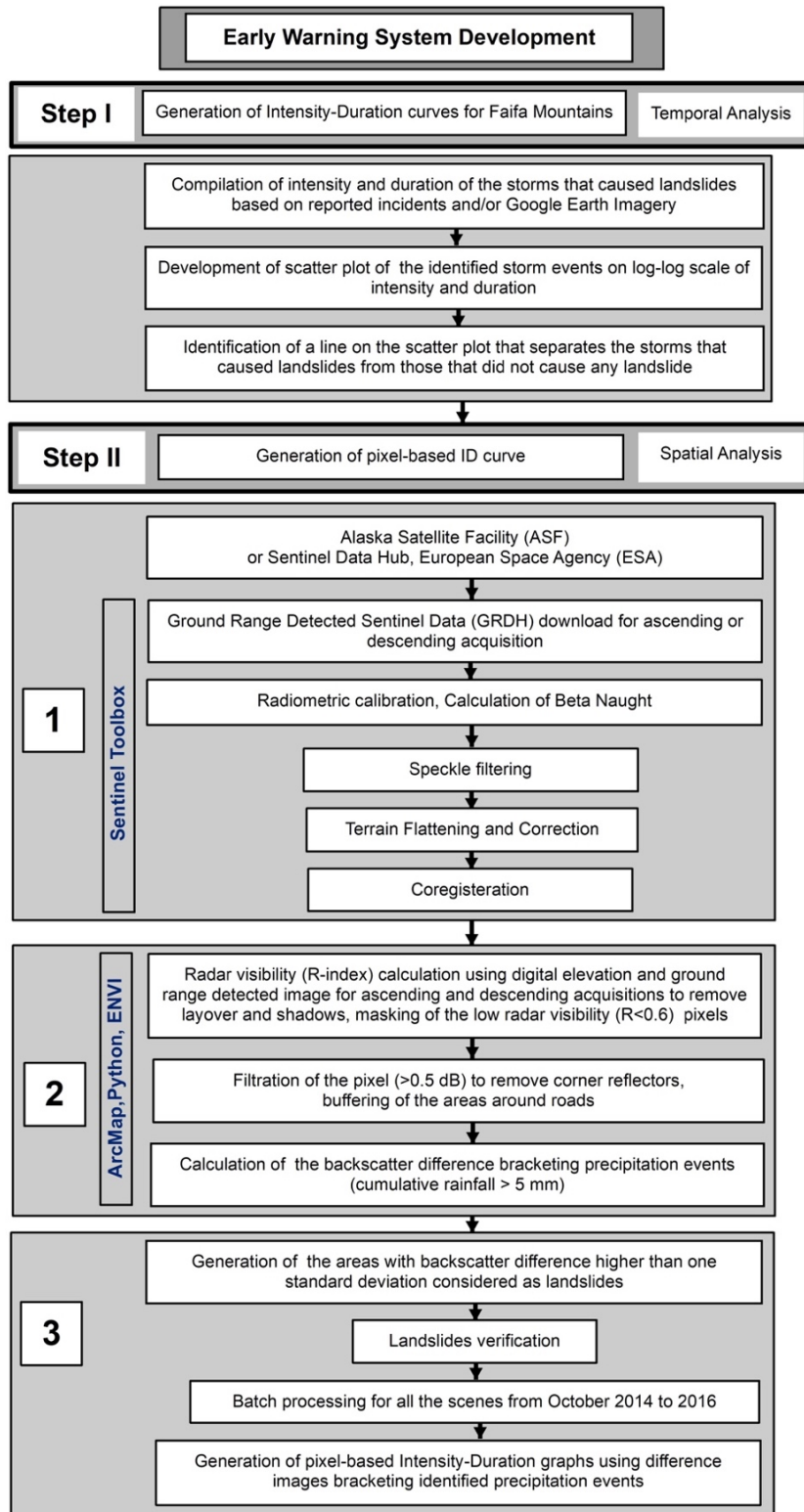


Figure 3. Flow chart summarizing the developed methodology that could serve as important steps towards the construction of an EWNSL. The developed procedures

involved the analysis of temporal Google Earth images, Sentinel-1A radar scenes, and TRMM and GPM rainfall data. Analysis involved two main steps: generation of an ID curve for the Faifa Mountains to identify storms that are likely to produce landslides (landslide-producing storms), and generation of pixel-based ID curves to identify the locations where movement is likely to occur during landslide-producing storms. Step II involved: (1) selection, calibration, and pre-processing of radar images, (2) generation of backscatter coefficient difference images as a measure of surface roughness change due to precipitation-induced landslides and, (3) development, refinement, and validation of the model to identify pixels susceptible to movement under user-defined precipitation conditions. The downloaded scenes were processed using the ESA's Sentinel Toolbox software, ENVI, ArcMap, and Python.

2.2.1 Generation of the ID threshold for the Faifa Mountains

Unfortunately, the distribution of rain gauges is inadequate in the study area (Fig. 1c). There is only one station within Jazan province, and three more stations in its surroundings (Fig. 1c). I utilized the GPM half-hourly (spatial resolution: $0.1^\circ \times 0.1^\circ$) and TRMM 3-hourly (spatial resolution: $0.25^\circ \times 0.25^\circ$) data to extract the intensity and duration of rainfall that caused landslides throughout the period 2007 to 2016 (Table 1 and Fig. 4). In generating the ID threshold for Faifa, I used the peak intensity values, in other words the shorter sections of the precipitation event with the higher intensity were selected. Landslides triggered by the same storm at different locations were assigned the same intensity and duration values. Although semiautomated procedures have been successfully used to extract intensity and duration of landslide-producing

precipitation events (e.g., Segoni et al., 2014; Rosi et al., 2016) I adopted a manual approach given the coarse spatial and temporal resolution of satellite data and the limited inventory of historical landslide data over Faifa Mountains. Rainfall events of less than 1 mm/h were omitted given that no landslides were reported from the area at these low rainfall rates and TRMM could mistakenly identify fog for low rainfall events (<1 mm/h; Milewski et al., 2009). Altogether, 131 precipitation events were extracted from TRMM and GPM data throughout this period, of which 19 events were identified as landslide-producing storms (Table 1). These storms were identified using spectral and morphologic variations associated with landslide development, variations detected in the field and/or extracted visually from pairs of Google Earth images bracketing large precipitation events. Google Earth images were favoured over other readily available visible near-infrared (VNIR) satellite data sets given their high spatial resolution (15 m to 15 cm) and long temporal coverage for the study area (2007 to present). An area that witnessed landslides will be covered by spectrally dark vegetation on the Google Earth image preceding the landslide and by spectrally bright rocks and sediments on the image acquired after the landslide development. In many cases the latter image, not the former image, shows a major scar in the source area (onset of landslide) that gives way to more linear scars in vegetation along the landslide path. It is worth noting that different parts of the Faifa Mountains have differing numbers of Google Earth image acquisitions, amount of coverages, and resolutions.

Starting in October 2014, the SGS initiated a program to field-verify reports of landslide occurrences. Field observations were conducted by our research team

following the December 25th, 2015 landslide-producing storm and by the SGS researchers throughout the period from October 2014 to October 2016. Our collective field investigations revealed extensive landslides following the events on December 25, 2015, April 13, 2016, April 29, 2016, August 1, 2016, and August 25, 2016 events. During a number of these storms, landslides were reported from the mountainous areas proximal to, but outside of, the study area (e.g., Youssef et al., 2014).

Using detected storm-induced spectral and morphologic variations in pairs of archival temporal Google Earth imagery, two additional landslide-producing events (November 18 and December 1, 2014) were extracted; these were apparently not reported to, or verified by, the SGS researchers during their field campaign due to their location in inaccessible areas. Using the same techniques (storm-induced spectral and morphologic variations) 16 storms were detected in the period (2007 through 2014) preceding the SGS field campaign (2014–2016). Given the paucity of Google Earth images (18 images in 10 years), a number of precipitation events are likely to have occurred between consecutive Google Earth images. If landslides were detected within the period covered by the consecutive Google Earth image acquisitions, it was assumed that the largest of these storms caused the observed landslides. In the case of some of the identified precipitation events, the rainfall intensity and duration varied from one part of Faifa to another which resulted into the inclusion of more than one landslide-producing storm events between the same set of Google Earth images. In doing so, a few landslide-producing storms and those that did not cause landslides were not identified. The latter type of storms were identified during the field campaign period (Table 1).

Table 1. Intensity and duration of the precipitation events used for the construction of the Faifa ID curve. Landslide-producing storms were verified through field observations and by examining spectral and morphologic variations in pairs of Google Earth archival images bracketing significant storm events.

	Intensity (mm/h)	Duration (h)	Storm Date	Land- slides	Google Earth Imagery Dates
1	3.17	3.00	2007- 06-02	Yes	Google Earth (2007-12-30 and 2007-03-01)
2	4.83	3.00	2008- 10-11	Yes	Google Earth (2010-04-19 and 2007-12-30)
3	5.34	3.00	2008- 10-24	Yes	Google Earth (2010-04-19 and 2007-12-30)
4	2.58	3.00	2010- 07-29	Yes	Google Earth (2010-10-28 and 2010-05-10)
5	2.69	6.00	2010- 07-11	Yes	Google Earth (2010-10-28 and 2010-05-10)
6	3.62	3.00	2010-	Yes	Google Earth

Table 1 – continued

			08-25		(2010-10-28 and 2010-05-10)
7	1.85	6.00	2011- 07-31	Yes	Google Earth (2012-03-05 and 2010-10-28)
8	1.86	9.00	2011- 08-27	Yes	Google Earth (2010-10-28 and 2012-03-05)
9	3.29	3.00	2011- 08-28	Yes	Google Earth (2012-03-05 and 2010-10-28)
10	2.91	3.00	2012- 06-21	Yes	Google Earth (2013-04-14 and 2012-03-05)
11	1.63	6.00	2013- 07-22	Yes	Google Earth (2013-10-11 and 2013-04-14)
12	1.04	12.00	2014- 05-18	Yes	Google Earth (2014-12-24 and 2014-01-06)
13	3.70	2.00	2014-	Yes	Google Earth

Table 1 – continued

			11-18		(2014-12-24 and 2014-05-23)
14	5.77	4.00	2014- 12-01	Yes	Google Earth (2014-12-24 and 2014-10-21)
15	2.42	1.00	2015- 03-22	No	Field Visit
16	4.67	0.50	2015- 06-02	No	Field Visit
17	1.77	0.50	2015- 06-20	No	Field Visit
18	1.37	0.50	2015- 07-31	No	Field Visit
19	2.39	1.00	2015- 08-25	No	Field Visit
20	3.07	1.50	2015- 09-14	No	Field Visit
21	2.62	2.00	2015- 11-05	No	Field Visit
22	7.91	2.00	2015- 12-25	Yes	Field visit

Table 1 - continued

23	2.85	0.50	2016-03-25	No	Field Visit
24	5.02	6.50	2016-04-13	Yes	Field Visit
25	4.76	1.50	2016-04-29	Yes	Field Visit
26	2.82	2.00	2016-06-02	No	Field Visit
27	1.64	1.00	2016-06-15	No	Field Visit
28	8.85	12.00	2016-08-01	Yes	Field Visit
29	6.40	3.00	2016-08-25	Yes	Field Visit
30	2.93	2.50	2016-09-30	No	Field Visit

The data presented in Table 1 were plotted to extract the ID threshold for the Faifa Mountains. Landslide-producing storms were represented in Fig. 4 by solid circles (red and black), and the non-landslide producing storms by open circles. The solid black circles are for field-verified landslide-producing storms, and the red circles are for landslide-producing events extracted from Google Earth images. The figure shows the

extracted ID curve (red line; equation: $I = 4.89 \cdot D^{-0.65}$) that provides the best visual separation between the landslide-producing (solid circles above red line) and non-producing (open circles below red line) precipitation events. Given the limited number of storms that were identified throughout the investigated period I believe that the adopted approach for defining the ID threshold is adequate at this stage.

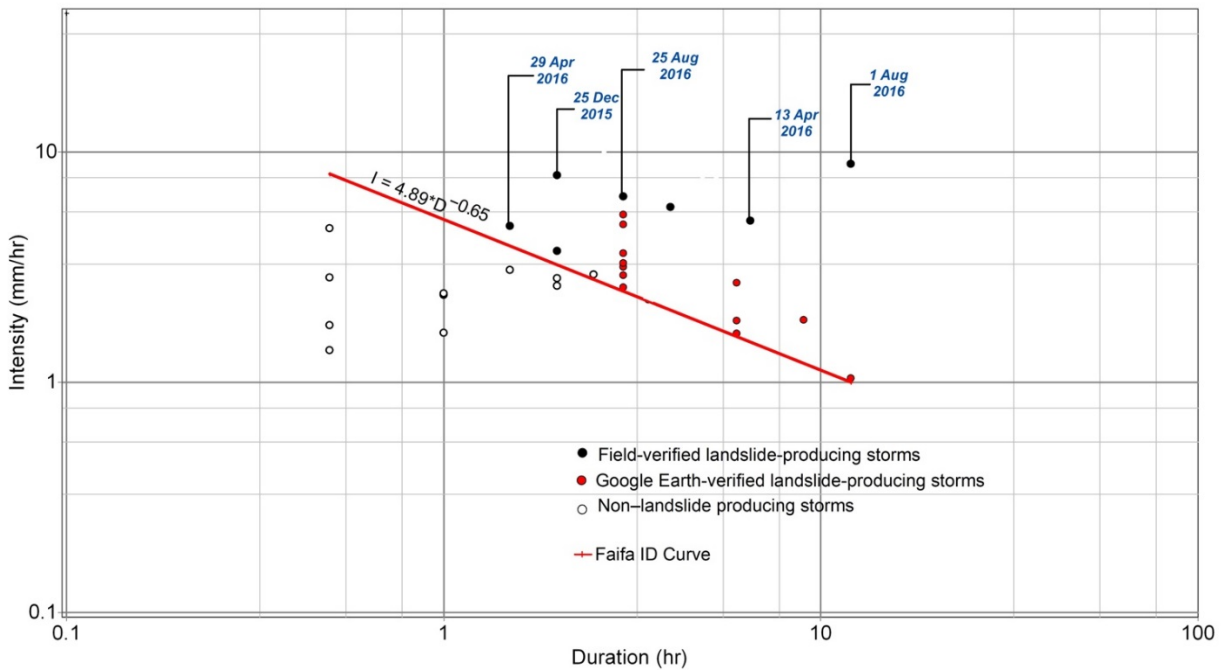


Figure 4. ID scatter plot generated from landslide-producing storm events (solid circles) and non-producing precipitation events (open circles) during the period 2007 to 2016. Faifa ID curve (equation: $I = 4.89 \cdot D^{-0.65}$; duration threshold [x]: 0.5 to 12 h) separates the landslide-producing events from non-producing events.

2.2.2 Selection, calibration, and pre-processing of radar images

The radar backscatter differences were used to determine the location of the landslide for the storm that caused the landslides. Sentinel-1A radar scenes were downloaded for ascending and descending acquisition modes from the Sentinel Data

Hub (<https://scihub.copernicus.eu/dhus/#/home>), a download platform for the European Space Agency (ESA), for the period between October 2014 and October 2016. The scenes can also be downloaded from Alaska Satellite Facility's (ASF) website (<https://vertex.daac.asf.alaska.edu/>). The scenes acquired immediately after (1 day or less) the rainfall were not used in the generation of backscatter coefficient difference images to avoid differences in backscatter due to precipitation-related change in moisture content. The pre-processing steps that were applied to the downloaded scenes included radiometric calibration and calculation of beta naught (β^0 , the radar brightness coefficient), speckle filtering, terrain flattening and correction, and image co-registration. Ascending and descending scenes for the same area provide different degrees of visibility, and depending on the orientation and complexity of the topography (van Zyl et al., 1993), one acquisition mode may provide better visibility than the other.

Standard processing procedures for SAR scenes were applied together with additional filtrations to remove backscatter anomalies that could be confused with our target. Ground range detected (GRD) level 1 images were downloaded and radiometrically calibrated using ESA's Sentinel Toolbox following the basic processing steps established by Veci (2016). The level-1 GRD products are focused SAR data that has been detected, multi-looked and projected to ground range using an Earth ellipsoid model (Small and Schubert, 2008). The GRD images were used to calculate the β^0 (Small, 2011), a measure of radar backscatter energy in decibels (dB; Raney et al., 1994), for both ascending and descending modes. The existing granular noise that degrades the quality of SAR data, known as speckle, was minimized in the extracted

radar backscatter coefficient images using the Lee Filter (window size: 3×3 ; Lee, 1983; Lee et al., 2009) and high-resolution DEM (TanDEM-X DEM; resolution: 12.5 m). The Terrain Flattening and the Range Doppler Terrain Correction (Small, 2011) was applied to the speckle filtered scenes to correct for radiometric biases introduced by the rugged topography of the study area. Each of the processed scenes was co-registered (sub-pixel co-registration) to the previously acquired one in Sentinel Toolbox (Press et al., 1992). Following the generation of the backscatter images, the scenes were cropped to the extent Faifa area to facilitate the execution of the steps that follow. The details of the processes have been provided in Fig. 3 (step II).

2.2.3 Generation of backscatter difference images

Following the identification of precipitation events over Faifa, backscatter difference images were generated between scenes bracketing the identified precipitation events. The initial analysis of these difference images revealed that corner reflectors and areas of low visibility can produce a response similar to that of landslides and hence procedures were developed to identify and mask out these areas. The generation of the backscatter difference images involved a number of steps: (1) calculation of radar visibility and removal of low visibility areas, (2) identification and removal of corner reflectors, and (3) generation of backscatter difference images.

Ascending and descending scenes for the same area provide different degrees of visibility depending on the topography and satellite orientation. (Notti et al., 2014). A radar visibility index (R; Notti et al., 2014) image was used to identify and mask out areas of low visibility in both the ascending and descending backscatter images. The R

index is a function of local variables (slope, aspect, incidence angle, layover, and shadow), and satellite geometry (line of sight azimuth). Using high-resolution DEMs, digital images were computed for each of those variables which were then used to generate R index images for ascending and descending geometries. These R index products were applicable for all backscatter scenes of same geometry and its values range from 0 (low visibility) to 1 (high visibility). Pixels with R values below a threshold of 0.6 were found to be spatially correlated with areas affected by overlays and by shadowing, and were masked out. The distribution of pixels with backscatter coefficients exceeding 0.5 dB were found to correlate with that of buildings, construction areas, vehicles, and parking spaces. Such features can act as corner reflectors and produce high radar returns by reflecting waves towards the source. Pixels with backscatter coefficients exceeding 0.5 dB (corner reflectors) were masked out. The filtered backscatter images were used to generate backscatter difference images between pairs of consecutive backscatter scenes, and only those pairs bracketing storm events were considered for further analysis.

2.2.4 Refinement and validation of the model

The refinement and validation of the model involved: (1) spatial refinement and standard deviation (SD) image generation, (2) field verification, and (3) batch processing of scenes. The population density in and around the road networks is high, and so are the risks for human and property losses if landslides occur in their vicinity. The construction of roads can trigger debris flows, especially in cases when roads intersect steep slopes (Fig. 5a) or terraces constructed on these steep slopes (Fig. 5b),

ephemeral valleys, and fracture planes dipping towards the road (Fig. 5c; Alharbi et al., 2014).



Figure 5. Landslides proximal to, possibly triggered by, road construction and intensified by rainfall. (a) Debris flow caused by failure on steep slopes intersected by roads. (b) Debris flow caused by failure of terraces constructed on steep slopes. (c) Landslide caused by failure on fracture planes dipping towards the road.

The distribution of historical debris flows in the study area was investigated to identify areas susceptible to debris flow. As described earlier, areas that witnessed recent debris flows are characterized by spectrally bright rocks and sediments, a major scar in the vegetation within the source area (onset of landslide) that gives way to more linear scars in the vegetation along the landslide path. As years go by, spectral and morphologic features indicative of debris flows can get obscured by encroaching vegetation making it more difficult to identify the older debris flows. Many of the historical debris flows were found on steep slopes, along first order streams, above and proximal to the main roads as shown in Fig. 6. Using these three criteria, areas susceptible to debris flows were identified by: (1) extracting stream networks using a

stream delineation algorithm (Tarboton et al., 1991) in ArcGIS 10.5 over the steep slopes ($>30^\circ$) and capturing first order streams, using a small flow accumulation value (10 pixels), (2) assigning a buffer zone (width: 20 m) around the extracted streams to delineate the areas that are likely to be triggered by runoff during and following rainfall events, and (3) assigning a buffer zone (width: 100 m) around the roads. The use of the latter criterion allows the identification of areas susceptible to failure along pre-existing fractures as well since our field observations showed that the majority of such failures were triggered by road construction. The selected width of the buffer zones was determined by examining the proximity of the historical landslides to roads and extracted streams.

The selection of the buffered zones for further investigation served two purposes: (1) targeting areas of high risk, and (2) capturing the backscatter variations that are related to landslides, variations that could have been confused with those caused by factors other than landslides (e.g., change in vegetation intensity or vegetative cover) if the entire area was considered. Figure 6 shows several landslides within areas identified as being susceptible to landslides using the three above-mentioned criteria.

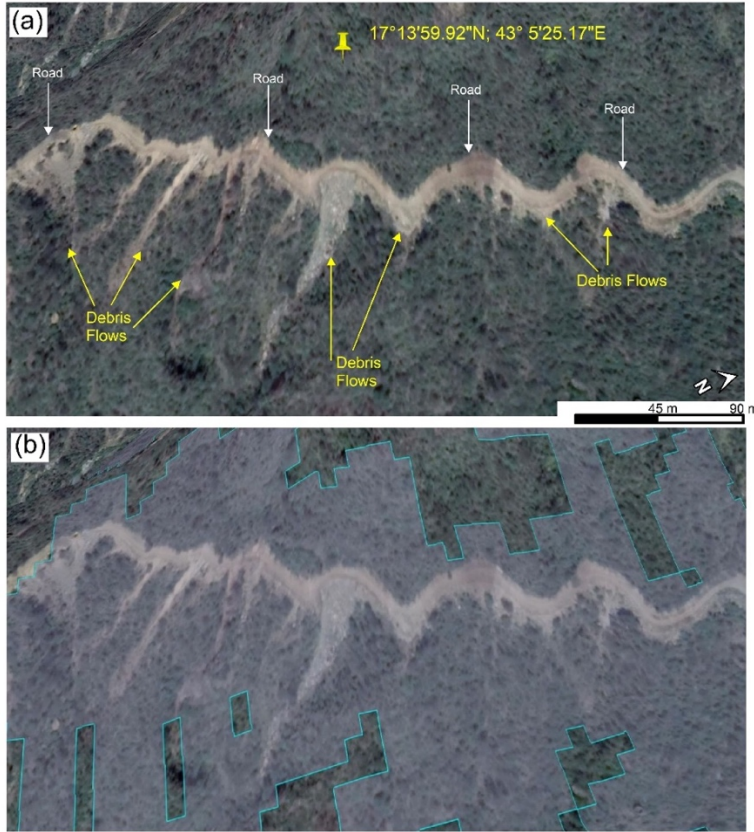


Figure 6. Google Earth Imagery showing the spatial refinement procedures. (a) Areas showing debris flows within areas characterized by steep slopes ($>30^\circ$), and proximity to roads (<100 m) and to the first order streams (<20 m). (b) The areas that satisfy these three conditions are outlined by the shaded polygon.

The spatial refinement was followed by filtration to detect spectral anomalies in the resultant image. A standard deviation–classified image was generated for a backscatter difference image that bracketed the December event (images acquired on December 24, 2015 and February 12, 2016). The differences in backscatter were classified into four groups: area of $\leq 1SD$ (no change), $>1SD$ to $\leq 2SD$ (lowest change), $>2SD$ to $\leq 3SD$ (medium change), and $>3SD$ (highest change). Within the buffered areas

on any of the backscatter difference images, the largest variations are expected to correspond to areas that witnessed landslide-related changes in roughness. Field observations following the December 25, 2015 precipitation event (15 mm) were conducted (February 26 to March 7, 2016) to test this assumption. The investigation proved the examination of the variations in spatially refined and spectrally filtered backscatter difference images, and the effectiveness of the applied filtering techniques in omitting the false positives.

Altogether I visited 27 sites in Faifa during our field investigation (Fig. 2a). It was found that the distribution of areas with $\leq 1\text{SD}$ variations on the extracted difference images did not correspond to any of the observed landslides and are here attributed to temporal variations in vegetation, minor roughness changes, and possibly sub-pixel errors in co-registration. Areas exceeding 1SD on the difference images (Fig. 7; clusters of red, yellow, and green pixels representing highest, medium, lowest changes, respectively) in the backscatter difference image corresponded to landslide locations and showed evidence for recent redistribution of boulders and sediments in the field. If these conditions were met, a landslide was considered as being verified. Altogether, 90%, 60% and, 86% of the pixels exceeding 3σ , 2σ , and 1σ were located within contiguous areas identified as being locations of landslides. The remaining areas were mostly random distribution of individual pixels resulting from corner reflectors or artefacts due to inadequate speckle filtering. Out of the 15 landslides that were identified with the proposed method, 14 were verified in the field. There was a false positive where the pixels exceeding 1SD corresponded to road construction-related changes (Table 2: site 26). Field investigations of 12 sites verified that spatial

refinement and filtration techniques was successful in filtering out 9 of the 12 false positives resulting from corner reflectors (e.g., building, constructions), but mistakenly removed an active debris (Table 2: site 1) and structurally stabilized fracture plane (Table 2: site 19 and 20).

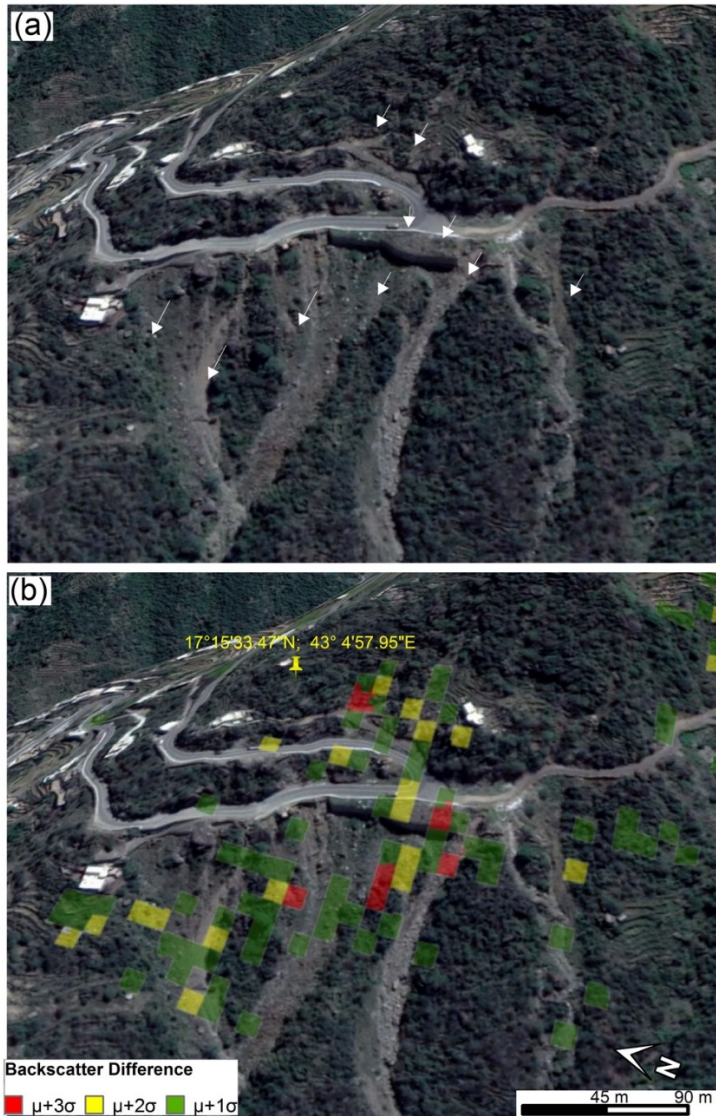


Figure 7. (a) Google Earth imagery showing the distribution of debris flows (identified by white arrows). (b) Backscatter difference image for two descending scenes bracketing (acquisition dates: December 24, 2015 and February 10, 2016) a precipitation event on December 25, 2015 showing correspondence of areas of low to negligible variations ($\leq 1SD$) with vegetation and areas exceeding 1SD (clusters of red, yellow, and green pixels) to debris flow locations that showed evidence for recent redistribution of boulders and cobbles in the field.

Table 2. Field observations collected (February 26 to March 7, 2016) for the assessment of radar-based distribution of active landslides, the areas exceeding 1 standard deviation (1SD) on difference images. Locations shown in Fig. 2a.

	Difference image	Remarks	Long. (°E)	Lat. (°N)
1		<i>Filtered out active debris flow; false positive</i>	43.054	17.234
2	3SD, 2SD	Verified active debris flow	43.065	17.237
3	3SD, 2SD	Verified active debris flow	43.070	17.233
4		<i>Filtered out road construction</i>	43.080	17.236
5		<i>Filtered out terraces, bare soil, no vegetation</i>	43.081	17.237
6	3SD, 2SD	Verified active debris flow, recently mitigated at intersection with road	43.083	17.241
7	2SD	Verified active debris	43.086	17.244

Table 2 – continued

		flow		
8		<i>Filtered out terraces</i>	43.091	17.243
9	2SD	Verified active debris flow; locals reported activity during rainfall	43.092	17.244
10	2SD	Verified active debris flow used to dispose construction material	43.106	17.255
11	3SD, 2SD	Verified active debris flow, recently mitigated at intersection with road	43.105	17.262
12		<i>Filtered out buildings</i>	43.118	17.271
13	2SD	Verified active debris flow	43.118	17.273
14		<i>Filtered out road construction</i>	43.070	17.271
15		<i>Filtered out road construction</i>	43.075	17.266
16	2SD	Debris flow related to terraces	43.081	17.261

Table 2 – continued

17	2SD	Verified debris flow recently mitigated proximal to road	43.081	17.262
18	3SD, 2SD	Verified active debris flow	43.081	17.260
19		<i>Filtered out shotcrete to stabilize the fracture planes; false positive</i>	43.087	17.255
20		<i>Filtered out shotcrete to stabilize the fracture planes; false positive</i>	43.086	17.258
21	3SD, 2SD	Verified active debris flow	43.082	17.258
22	2SD	Verified failure along fracture plane dipping towards the road	43.091	17.254
23		<i>Filtered out buildings</i>	43.104	17.262
24		<i>Filtered out terraces</i>	43.101	17.267
25	3SD, 2SD	Verified active debris flow bordering a terrace	43.099	17.269

Table 2 – continued

26	3SD, 2SD	Construction related debris flows downhill from the road; false negative	43.095	17.273
27		<i>Filtered out construction along the road</i>	43.083	17.274

2.2.5 Pixel-based adaptation of Faifa ID threshold as a predictive tool

The ID curve for any pixel should separate landslide-producing events (backscatter difference > 1SD) from non-landslide producing (backscatter difference \leq 1SD) precipitation events. On these graphs, landslide-producing events plot above the curve, and the non-producing events plot below it.

A pixel-based debris detection system was developed by adopting the slope of the extracted Faifa ID curve. The assignment of the ID curves to the individual pixels will depend on the relative stability of the individual pixel. The less stable pixels, such as those on steep slopes, are expected to experience movement in response to weak, moderate, and extreme storm events, whereas the more stable pixels will move during the extreme events only. Figure 8 demonstrates the ID curve for a more stable pixel. Curve A represents the ID curve for more stable location as it showed evidence for landslide-related movement (>1SD on the radar backscatter difference image) in response to five bigger events (December 1, 2014, December 25, 2015, April 13, 2016,

August 1, 2016, and August 25, 2016) but no movement ($<1SD$ on the backscatter difference image) following the April 29, 2016 and November 18, 2014 storms. Curve A has the slope of the Faifa ID threshold, but a different intercept. Thus, knowing the historical response of each individual pixel to these storms, each pixel was assigned an ID curve whose slope is similar to that of the Faifa curve. In other words, the pixel-based ID curve uses the historical landslide response of a pixel to estimate the intensity and duration of the precipitation that would cause landslide in the future. Any event that plots above the pixel ID curve would produce landslides at that location, whereas the one that plots below would not produce a landslide. The pixel-based ID curve helps to predict the location that will witness landslides under any future storm event. With the current limitation of data sets, only few upward translations of Faifa ID curve is possible. As the inventory of landslide grows, I expect that the placements of pixel-based ID thresholds will get progressively refined hence representing more realistic view of the stability of individual pixels. The expanded inventory will also enable the application of advanced thresholding techniques.

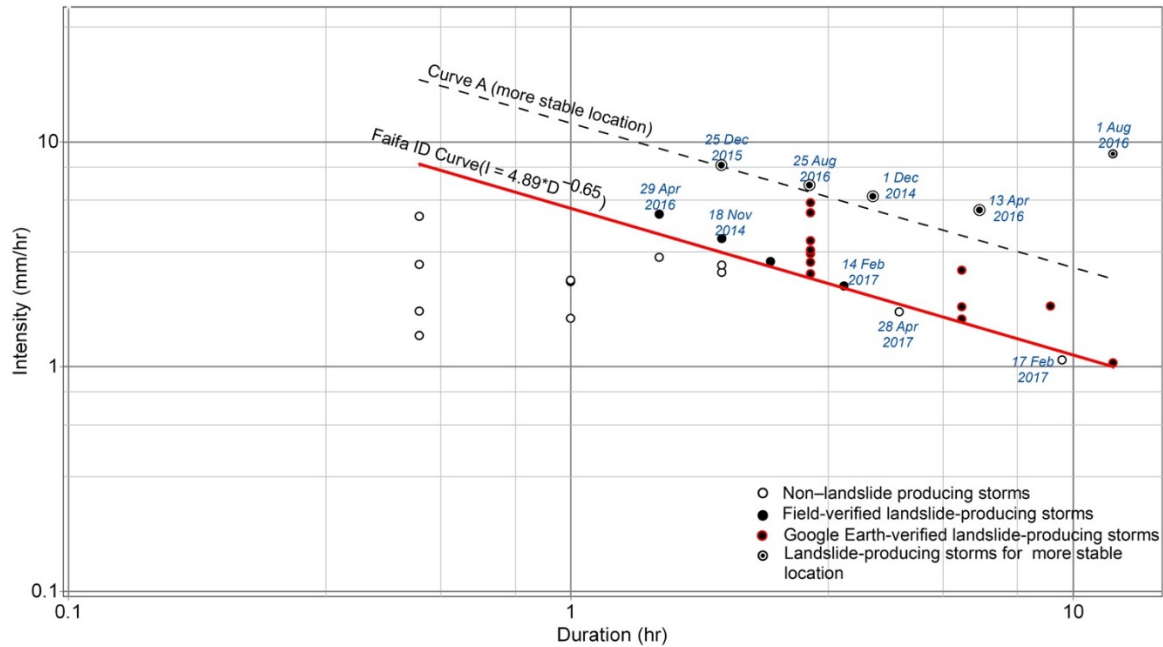


Figure 8. Demonstration of pixel-based ID curves. Curve “A” is a curve for a more stable pixel that witnessed landslides in response to five storms (December 1, 2014, December 25, 2015, April 13, 2016, August 1, 2016, and August 25, 2016). The ID curve for more stable location is parallel to the Faifa ID curve that separates landslide-producing storm events from landslide non-producing storm events from Fig. 4.

2.2.6 Validation of ID threshold

Three precipitation events larger than the threshold (1 mm/h for 1 h) were recorded during the period from November 2016 to April 2017. These occurred on February 14 (intensity: 2.28 mm/h; duration: 3.5 h), February 17 (intensity: 1.07 mm/h; duration: 9.5 h), and April 28 (intensity: 1.75 mm/h; duration: 4.5 h) of 2017. The event on February 14 plotted above the Faifa ID curve, whereas those on February 17 and April 28 plotted below the curve. Landslides were reported following the February 14 event, but not for the two other storm events, an observation that

supports the validity of the extracted ID curve for Faifa.

Using the precipitation intensity and duration for the February 14 storm, and the extracted pixel ID curves, I generated a map showing the areas (three or more pixels) that are likely to witness movement under the specified precipitation conditions. I visually inspected these areas on the Google Earth images that were acquired before (October 2, 2016) and after (March 29, 2017) the February 14 storm as shown in Fig. 9. Spectral and morphological variations indicative of landslides were detected on the March 29 Google Earth image. Specifically, 13 landslides were predicted, out of which 6 were verified by inspecting the March 29 image, an accuracy of 60%. Similarly, out of 7 locations where no landslides were predicted, one location witnessed a landslide. I suspect that the high number of false positives (7 locations) is largely an artifact of the adopted method of landslide verification. The morphological variations observed on Google Earth images and indicative of landslides are effective in detecting the large, but not the small, landslides. The higher number of false positives (7 locations) than the false negative (one location) suggests that pixel-based ID curve is significantly reducing the number of false negatives. With limited number of post study storm events, the entire Faifa area currently can accommodate only few adjustments to the ID curve. With inclusion of more storms and accumulation of archival data, pixel-based ID curve is expected to represent unique historical signature of the landslide records. Thus, over time, the number of false positive is expected to decrease as more and more areas would have their thresholds based on its relative stability instead of minimum threshold established for the Faifa region. At this stage, the reported accuracy is reasonable for developing a prototype EWNSL given: (1) the uncertainties associated

with extracting the Faifa and pixel-based ID curves, and (2) the fact that landslides in a particular area tend, in some cases, to stabilize the location and reduce the chances of landslide recurrence in the same area.

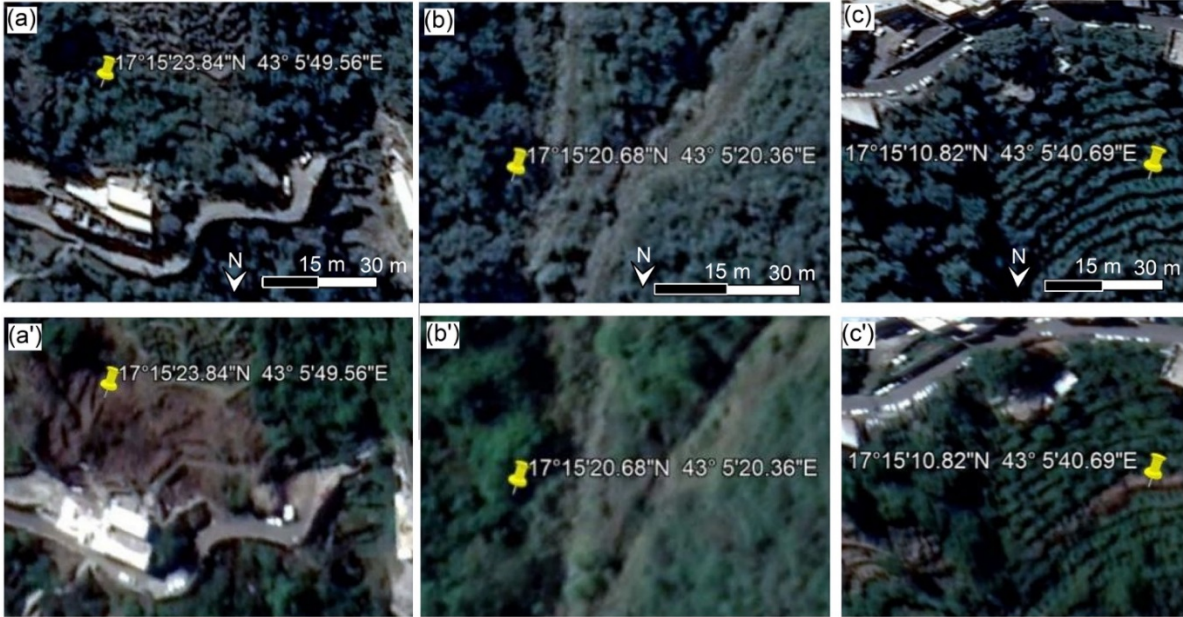


Figure 9: Demonstration of the prediction result for storm that occurred on February 14, 2017. (a) Area that witnessed landslide and was predicted (true positive). (b) Area that did not witness landslide but was predicted (false positive). (e) Area predicted as stable but witnessed landslide (false negative).

2.3 Results and Discussion

Our ability to predict the landslide-producing storms and the locations of these landslides will depend largely on the accuracy of the extracted/adopted ID curves. The production of the Faifa ID curve was based on precipitation data for 30 storms, approximately 63% of which produced landslides. Precipitation was extracted from the earlier coarse TRMM (3 h; $0.25^\circ \times 0.25^\circ$) and later from the finer resolution GPM (1 h;

0.1° × 0.1°) data; field observations and temporal Google Earth images were used to identify which of these storms produced landslides. The temporal coarseness of the precipitation data, especially 3-hourly TRMM data, provides the uncertainty in the precise placement of the ID curve. In upcoming years, additional data points (landslide producing and non-producing storms), especially those extracted from GPM or rain gauge data with higher spatial and temporal resolution will be used to refine the initial Faifa ID curve.

Refining the Faifa ID curve will enhance the accuracy of the pixel-based ID curves as well, given that they are assumed to be parallel to the Faifa ID curve. In the construction of these curves, I was constrained by the following limitations in data availability and processing technique: (1) the limited number of storms (18) that occurred throughout the time period (October 30, 2014 to October 31, 2016) during which radar images were available; (2) coarse spatial resolution of the DEM (12.5 m TanDEM-X) and radar data (multi-looked using DEM, 12.5 m) that made it difficult to identify landslides of limited size (<25 m or the size of two pixels); (3) the discontinuous acquisition of Sentinel-1A data (ascending and descending modes) that interrupted the monitoring of landslides in response to storm events; (4) sub-pixel errors in co-registration of radar data and anomalous backscatter spikes originating from buildings and construction activities that produced radar responses similar to landslide-related radar response and were not filtered; (5) drastic changes in the slope and/or vegetation in a particular pixel that impaired the functionality of several pixel-based ID curves; (6) frequent rainfall with short duration could have gone undetected given the coarse temporal resolution of the satellite-based precipitation data; and (7)

limited field investigations and reliance on Google Earth imagery did not provide enough information to develop robust thresholding technique.

In coming years, the pixel-based ID curves I developed will be refined by: (1) acquiring high spatial and temporal resolution precipitation data, (2) identifying additional landslide-producing storms to augment the existing database and update the existing pixel-based ID curves, (3) applying additional filtration techniques (e.g., coherence threshold filters to reduce false positives), and (4) developing an urban mask to exclude radar responses from corner reflectors that could be confused with those from landslides. I will also explore refining our methodologies to account for the impact of antecedent precipitation on landslide development (e.g., Chen et al., 2015). To date, the application of ID thresholds for landslide hazard assessment is widespread in early warning systems at local and regional scales (e.g., Peruccacci et al., 2017; Rossi et al., 2017a), yet over the past few years there has been increasing recognition of the role of hydrology in landslide initiation, a factor that is not fully incorporated in the ID threshold analysis. The intensity and duration of rainfall during which a landslide occurs are not the only triggers for landslides; the rainfall events (antecedent rainfall) that preceded the landslide-causing precipitation are triggers, as well (Kim et al., 2014, Hong et al., 2017). It has been shown that the antecedent and peak rainfall play important role in triggering landslides in general, but debris flow development is more related to rainfall peak than antecedent rainfall (Chen et al., 2015). The proposed procedures could be used in the absence of extensive field-based data sets although it does not account for the role of antecedent rainfall preceding such landslide-causing storm events. Given that the overwhelming majority of our landslides are debris flows,

I do not anticipate that such refinement will largely affect our findings.

The adopted methodologies and suggested refinements represent significant steps towards the development of a prototype EWNSL. To better achieve this goal, the following additional automated steps have to be accomplished. Near real-time measurements of precipitation should be collected from the rain gauge network over the study area to avoid the delays associated with posting satellite-based precipitation (GPM: 3 to 6 hours). Temporal precipitation distributions can be derived from the acquired rain gauge measurements and used as inputs to our developed modules. Currently, our methodology identifies vulnerable areas based on user-defined precipitation intensity and duration. Once the nowcasting system is established, as rainfall data is collected, it will be fed automatically into the EWNSL to identify the areas likely to witness landslides at any time. The precipitation at any location could be assumed to continue for a pre-determined time period (e.g., 1 hour) and the model outputs under such assumptions could be used to predict the areas that are likely to witness landslides in that pre-determined time period. The predictive model outputs could be posted in near-real time on a web-based GIS, giving the authorities and citizens in threatened areas enough time to vacate these locations.

2.4 Conclusion

I developed a predictive system that shows whether a storm with a particular intensity and duration can cause landslides in the Faifa Mountains. For the identified landslide-producing storms, the developed methodologies will also select areas that are likely to witness landslide development. The extracted ID curve for the Faifa is used for

the former and the extracted pixel-based ID curves for the latter.

The current usage of satellite-derived rainfall data limits the accurate determination of intensity as it is acquired periodically (3-hourly for TRMM and half-hourly for GPM). This is also reflected in the ID threshold plot with the placement of rainfall events in regular intervals. The proposed methodology can be refined and recalibrated to include data from rain gauge stations which will decrease the uncertainties associated with the duration and time used to compute the intensity of the rainfall. There are uncertainties associated with manual delineation of the threshold which will be gradually reduced with the addition of more rainfall observations. During the identification of the landslide-causing storms through Google Earth observation, the assumption about the biggest rainfall being responsible for the landslide was made. This assumption might have excluded some rainfall events from our observation causing the upward shift of the ID curve for many locations.

The methodologies advanced here are robust and cost-effective procedures that could be readily applied to many data-deficient locations worldwide. The proposed methodology relies heavily on readily available satellite data and thus could be applicable to many of the world's mountainous locations. The developed methodologies and rigorous refinements represent significant steps towards the development an EWNSL if precipitation forecasts become available. The proposed procedures for the development of ID curves should not be considered as alternatives to the well-developed field-based ID relationships and to recently introduced advances in such applications, but could be used in absence of such field-based datasets.

2.5 References

- Aleotti, P.: A warning system for rainfall-induced shallow failures, *Eng. Geol.*, 73, 247–265, <http://dx.doi.org/10.1016/j.enggeo.2004.01.007>, 2004.
- Alharbi, T., Sultan, M., Alsefry, S., Elkadiri, R., Ahmed, M., Chase, R., Milewski, A., Abu Abdullah, M., Emil, M., and Chounaird, K.: An assessment of landslide distribution in the Faifa area, Saudi Arabia, using remote sensing and GIS techniques, *Nat. Hazards Earth Syst. Sci.*, 14, 1553–1564, <https://doi.org/10.5194/nhess-14-1553-2014>, 2014.
- Alsharhan, A. S., Rizk, Z. A., Nairn, A. E. M., Bakhit, D. W., and Alhajari, S. A.: Hydrogeology of an arid region: the Arabian Gulf and adjoining areas, Elsevier, Amsterdam, Netherlands, 2001.
- Battistini, A., Rosi, A., Segoni, S., Lagomarsino, D., Catani, F. and Casagli, N.: Validation of landslide hazard models using a semantic engine on online news, *Appl. Geogr.*, 82, 59–65, <https://doi.org/10.1016/j.apgeog.2017.03.003>, 2017.
- Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A., and Pizziolo, M.: Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach, *J. Geophys. Res.*, 117, F04006, <https://doi.org/10.1029/2012JF002367>, 2012.
- Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., and Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, *Nat. Hazards Earth Syst. Sci.*, 10, 447–458, <https://doi.org/10.5194/nhess-10-447-2010>, 2010.
- Burtin, A., Bollinger, L., Cattin, R., Vergne, J., and Nabelek, J. L.: Spatiotemporal sequence of Himalayan debris flow from analysis of high-frequency seismic noise, *J.*

- Geophys. Res.-Earth Surf.,114, F04009, <https://doi.org/10.1029/2008JF001198>, 2009.
- Caine, N.: The rainfall intensity-duration control of shallow landslides and debris flows, *Geografiska Annaler*, 62, 23–27, <https://doi.org/10.2307/520449>, 1980.
- Cannon, S. H., Boldt, E., Laber, J., Kean, J., and Staley, D.: Rainfall intensity–duration thresholds for postfire debris-flow emergency response planning, *Nat. Hazards*, 59, 209–236, <https://doi.org/10.1007/s11069-011-9747-2>, 2011.
- Cannon, S. H., and Gartner, J. E.: Wildfire-related debris flow from a hazards perspective, in: *Debris flow hazards and related phenomena*, edited by: Jakob, M., and Hungr, O., Springer, Berlin, 363–385, http://dx.doi.org/10.1007/3-540-27129-5_15, 2005.
- Cannon, S. H., Gartner, J. E., Wilson, R., Bowers, J., and Laber, J.: Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California, *Geomorphology*, 96, 250–269, <https://doi.org/10.1016/j.geomorph.2007.03.019>, 2008.
- Chen, C. W., Saito, H., and Oguchi, T.: Rainfall intensity–duration conditions for mass movements in Taiwan, *Prog. in Earth and Planet. Sci.*, 2, <https://doi.org/10.1186/s40645-015-0049-2>, 2015.
- Clarizia, M., Gullà, G., and Sorbino, G.: Sui meccanismi di innesco dei soil slip, in: *International conference Prevention of hydrogeological hazards: the role of scientific research*, 1, 585–597 (in Italian), 1996.
- Crosta, G. B., and Frattini, P.: Rainfall thresholds for triggering soil slips and debris flow, in: *Mediterranean storms*, edited by: Mugnai, A., Guzzetti, F., and Roth, G.,

- Proceedings of the 2nd EGS Plinius Conference on Mediterranean Storms, Siena, Italy, 463–487, 2001.
- Erener, A., and Düzgün, H. B. S.: A regional scale quantitative risk assessment for landslides: case of Kumluca watershed in Bartın, Turkey, *Landslides*, 10, 55–73, <https://doi.org/10.1007/s10346-012-0317-9>, 2013.
- Fairer, G. M.: Explanatory notes to the geologic map of the Wadi Baysh Quadrangle, Deputy Ministry for Mineral Resources, Ministry of Petroleum and Mineral Resources, Kingdom of Saudi Arabia, Denver, Colorado, USA, 1–23, 1985.
- Greenwood, W. R., Stoesser, D. B., Fleck, R. J., and Stacey J. S.: Late Proterozoic island-arc complexes and tectonic belts in the southern part of the Arabian shield Kingdom of Saudi Arabia, *US Geol. Surv. Open File Rep.*, 1983.
- Greenwood, W. R.: Geology of the An Nimas quadrangle sheet 19/42 C, Kingdom of Saudi Arabia, with a section on aeromagnetic studies by G.E. Andersen, Saudi Arabian Directorate General of Mineral Resources Geologic Map, GM-37, 33, scale 1:100,000, 1979.
- Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, *Meteorol. Atmos. Phys.*, 98, 239–267, <https://doi.org/10.1007/s00703-007-0262-7>, 2007.
- Hong, M., Kim, J., and Jeong, S.: Rainfall intensity-duration thresholds for landslide prediction in South Korea by considering the effects of antecedent rainfall, *Landslides*, 15, 523–534, <https://doi.org/10.1007/s10346-017-0892-x>, 2017.
- Innes, J. L.: Debris flows, *Prog. Phys. Geogr.*, 7, 469–501, <https://doi.org/10.1177/030913338300700401>, 1983.

- Iverson, R. M.: The physics of debris flows, *Rev. Geophys.*, 35, 245–296,
<https://doi.org/10.1029/97RG00426>, 1997.
- Jakob, M., Owen, T., and Simpson, T.: A regional real-time debris flow warning system for the District of North Vancouver, Canada, *Landslides*, 9, 165–178,
<https://doi.org/10.1007/s10346-011-0282-8>, 2012.
- Kim, H. G., Lee, D. K., Park, C., Kil, S., Son, Y., and Park, J. H.: Evaluating landslide hazards using RCP 4.5 and 8.5 scenarios, *Environ. Earth Sci.*, 73, 1385–1400,
<https://doi.org/10.1007/s12665-014-3775-7>, 2014.
- Lagomarsino, D., Segoni, S., Rosi, A., Rossi, G., Battistini, A., Catani, F., and Casagli, N.: Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting, *Nat. Hazards Earth Syst. Sci.*, 15, 2413–2423,
<https://doi.org/10.5194/nhess-15-2413-2015>, 2015.
- Larsen, M. C., and Simon, A.: A rainfall intensity–duration threshold for landslides in a humid-tropical environment, Puerto Rico, *Geograf. Ann. A.*, 75, 13–23,
<https://doi.org/10.1080/04353676.1993.11880379>, 1993.
- Lee, J. S.: Digital image smoothing and the sigma filter, *Comput. Vis. Graph. Image Process.*, 24, 255–269, [https://doi.org/10.1016/0734-189X\(83\)90047-6](https://doi.org/10.1016/0734-189X(83)90047-6), 1983.
- Lee, J. S., Wen, J. H., Ainsworth, T. L., Chen, K. S., and Chen, A. J.: Improved sigma filter for speckle filtering of SAR imagery, *IEEE Trans. Geosci. Remote Sens.* 47,
<https://doi.org/10.1109/TGRS.2008.2002881>, 2009.
- Lowell, S.: The K M Mountain landslide near Skamokawa, Washington Geologic Newsletter 18:3–7, 1990.
- Mashat, A., and Basset, H. A.: Analysis of rainfall over Saudi Arabia, *Met. Env. & Arid Land*

- Agric. Sci., 22, 59–78, 2011.
- Milewski, A., Sultan, M., Jayaprakash, S. M., Balekai, R., and Becker, R.: RESDEM, a tool for integrating temporal remote sensing data for use in hydrogeologic investigations, *Comput. Geosci.*, 35, 2001–2010, <https://doi.org/10.1016/j.cageo.2009.02.010>, 2009.
- Ministry of Municipal and Rural Affairs (MMRA), Faifa Municipality.
<https://momra.gov.sa/>, last access: 16 August 2017.
- Norrish, N. I., and Wyllie, D. C.: Rock slope stability analysis, in: *Landslides: investigation and mitigation*, Special Report 247, edited by: Turner, A. K., and Schuster, R. L., Transportation Research Board, Board, Washington D.C., 391–425, 1996.
- Notti, D., Herrera, G., Bianchini, S., Meisina, C., García-Davalillo, J. C., and Zucca, F.: A methodology for improving landslide PSI data analysis, *Int. J. Remote Sens.*, 35, 2186–2214, 2014.
- Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., and Guzzetti, F.: Rainfall thresholds for possible landslide occurrence in Italy, *Geomorphology*, 290, 39–57, <https://doi.org/10.1016/j.geomorph.2017.03.031>, 2017.
- Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., and Calvillo, M.: Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides, *Landslides*, 14, 995–1008, <https://doi.org/10.1007/s10346-016-0750-2>, 2017.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: *The art of scientific computing. Numerical Recipes in C*, 2nd edn., Cambridge University Press, Cambridge, United Kingdom, 1992.

- Raney, K., Freeman, A., Hawkins, B., and Bamler, R.: A plea for radar brightness, in: Proc. IEEE Geosci. Remote Sensing Symposium, Pasadena, California, 1090–1092, 1994.
- Rosi, A., Lagomarsino, D., Rossi, G., Segoni, S., Battistini, A., and Casagli, N.: Updating EWS rainfall thresholds for the triggering of landslides, *Nat. Hazards*, 78, 297–308, <https://doi.org/10.1007/s11069-015-1717-7>, 2015.
- Rosi, A., Peternel, T., Jemec-Auflic, M., Komac, M., Segoni, S., and Casagli, N.: Rainfall thresholds for rainfall-induced landslides in Slovenia, *Landslides*, 13, 1571–1577, <https://doi.org/10.1007/s10346-016-0733-3>, 2016.
- Rossi, M., Kirschbaum, D., Valigi, D., Mondini, A. C., and Guzzetti, F.: Comparison of satellite rainfall estimates and rain gauge measurements in Italy, and impact on landslide modelling, *Climate*, 5, 90, <https://doi.org/10.3390/cli5040090>, 2017a.
- Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M. T., Peruccacci, S., and Guzzetti, F.: Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data, *Geomorphology*, 285, 16–27, <https://doi.org/10.1016/j.geomorph.2017.02.001>, 2017b.
- Schmidt, D. L., Hadley, D. G., Greenwood, W. R., Gonzalez, L., Coleman, R. G., and Brown, G. F.: Stratigraphy and tectonism of Saudi Arabia, Saudi Arabian Directorate General of Mineral Resources bulletin, 8, 13, 1973.
- Segoni, S., Piciullo, L., and Gariano, S. L.: A review of the recent literature on rainfall thresholds for landslide occurrence, *Landslides*, 15, 1483–1501, <https://doi.org/10.1007/s10346-018-0966-4>, 2018.

- Segoni, S., Rossi, G., Rosi, A., and Catani, F.: Landslides triggered by rainfall: A semi-automated procedure to define consistent intensity–duration thresholds, *Comput and Geosci*, 63, 123–131, <https://doi.org/10.1016/j.cageo.2013.10.009>, 2014.
- Small, D.: Flattening gamma: radiometric terrain correction for SAR imagery, *IEEE Trans. Geosci. Remote Sens.*, 49, 3081–3093, <https://doi.org/10.1109/TGRS.2011.2120616>, 2011.
- Small, D., and Schubert, A.: Guide to ASAR geocoding, RSL, University of Zuerich, 36, 2008.
- Staley, D. M., Kean, J. W., Cannon, S. H., Schmidt, K. M., and Laber, J. L.: Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California, *Landslides*, 10, 547–562, <https://doi.org/10.1007/s10346-012-0341-9>, 2013.
- Stoeser, D. B., and Camp, V. E.: Pan-African microplate accretion of the Arabian shield, *Geol. Soc. Am. Bull.*, 96, 817–826, [https://doi.org/10.1130/0016-7606\(1985\)96%3C817:PMAOTA%3E2.0.CO;2](https://doi.org/10.1130/0016-7606(1985)96%3C817:PMAOTA%3E2.0.CO;2), 1985.
- Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., and Hong, Y.: Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: is day-1 IMERG a good successor for TMPA 3B42V7?, *J. Hydrometeorol.*, 17, 121–137, <https://doi.org/10.1175/JHM-D-15-0059.1>, 2016.
- Tarboton, D. G., Bras, R. L., and Rodriguez-Iturbe, I.: On the extraction of channel networks from digital elevation data, *Hydrol. Process.*, 5, 81–100, <https://doi.org/10.1002/hyp.3360050107>, 1991.
- Youssef A. M., Al-kathery, M., Pradhan, B., and El-sahly, T.: Debris flow impact assessment along the Al-Raith Road, Kingdom of Saudi Arabia, using remote

sensing data and field investigations, *Geomat. Nat. Haz. Risk*, 1–19,

<https://doi.org/10.1080/19475705.2014.933130>, 2014.

Vallet, A., Varron, D., Bertrand, C., Fabbri, O., and Mudry, J.: A multi-dimensional statistical rainfall threshold for deep landslides based on groundwater recharge and support vector machines, *Nat. Hazards*, 84, 821–849, <https://doi.org/10.1007/s11069-016-2453-3>, 2016.

van Zyl, J. J., Chapman, B. D., Dubois, P., and Shi, J.: The effect of topography on SAR calibration, *IEEE Trans. Geosci. Remote Sens.*, 21, 5, 1036–1043, <http://doi.org/10.1109/36.263774>, 1993.

Veci, L.: SAR basics tutorial, Sentinel-1 Toolbox. European Space Agency. <http://step.esa.int/docs/tutorials/S1TBX%20SAR%20Basics%20Tutorial.pdf>, last access: 01 August 2016.

CHAPTER 3

PROJECT II

3.1 Introduction

An increase in agricultural activities introduces nutrients into water bodies and may adversely affect the biodiversity and habitats of aquatic ecosystems. One of the major sources of such nutrients are nitrogen-based fertilizers (Glibert et al., 2006) that are widely used to increase agricultural productivity. These nonpoint sources of nitrogen through fertilization were found to be the predominant sources of overall nitrogen quantities in the Gulf of Mexico (Howarth et al., 1996), where the study area (Charlotte County) is located. The introduction of nutrients increases the productivity of aquatic systems and enhances the growth of harmful algal blooms (HABs) which, in turn, produce toxins causing detrimental health effects (Landsberg, 2002) to humans and ecosystems (Fleming et al., 2011). *Karenia brevis* (*K. brevis*), formerly known as *Gymnodinium breve* and *Ptychodiscus brevis*, is the most predominant HAB species in the Gulf of Mexico (Steidinger et al., 1998; Amin et al., 2009; Thyng et al., 2013), and its adverse socioeconomic impacts on the region have been investigated in previous studies (Evans and Jones, 2001). These impacts include but are not limited to adverse effects to human health, marine life, tourism, and recreational activities (Evans and Jones, 2001; Landsberg, 2002; Fleming et al., 2011).

Earlier efforts to map or forecast HAB occurrences examined the distribution of HABs in relation to a wide range of causal parameters, such as wind-driven water exchanges (Raine et al., 2010), temperature (Cha et al., 2014), relative abundance of

protozoans that feed on algae (e.g., *Mesodinium* species) (Harred et al., 2014), cell distribution through oceanic currents (McGillicuddy et al, 2005), and hydrodynamic variables (e.g., current pathways, rate and volume of flow, upwelling and downwelling pulses) (Cusack et al., 2016) . Such parameters were subsequently used to conduct same-day mappings of bloom occurrences, to model onsets of blooms (Stumpf et al., 2008; Aleynik et al., 2016; Gillibrand et al., 2016) and to forecast seasonal algal bloom occurrences (McGillicuddy et al, 2005). These investigations and mapping efforts provided the basis for the development of early warning systems based on (i) solid-phase adsorption toxin tracking (Turrell et al., 2007), (ii) real-time field monitoring of chlorophyll and dissolved oxygen (Lee et al., 2005), and (iii) Moderate Resolution Imaging Spectroradiometer (MODIS)-derived fluorescence data to detect and monitor algal blooms (Hu et al., 2005; Carvalho et al., 2010; Al Shehhi et al., 2013). The latter (fluorescence) was found to be sensitive to chlorophyll-a concentrations (Neville and Gower, 1977; Pan et al., 1989; Fischer and Kronfeld, 1990; Hoge et al., 2003). The development and operation of the overwhelming majority of these monitoring and forecasting systems require continuous current and archival field data (e.g., nutrient concentration in surface water). Unfortunately, such datasets are not present for many of the coastal areas where HAB monitoring and/or forecasting systems are needed. This study addresses this potential problem. Although our methodology does require continuous records of present and archival data, it instead utilizes readily available, global remote sensing datasets in the public domain. Additionally, limited field data, where available, are utilized.

Earlier studies that utilized remote sensing datasets in identifying and mapping

the distribution of HABs focused on a limited number of ecological variables. Examples include utilization of a single ecological variable (e.g., chlorophyll-a concentration) (Hu et al., 2005; Balch et al., 1989; Stumpf et al., 2003; Gower et al., 2004; Tomlinson et al., 2009; Zhao et al., 2010), two variables such as chlorophyll-a concentration and sea surface temperature (SST) (Tang et al., 1999; Raine et al., 2001; Chang et al., 2003; Stumpf and Tomlinson, 2005; Ahn et al., 2006; Tang et al., 2006; Sarangi and Mohammed et al., 2011) or chlorophyll and primary productivity (Vargo et al., 1987), and three variables (chlorophyll, SST and wind) (Tang et al., 2002; Tang et al., 2004). A review article by Shen et al. (2012) indicated that most of the remote sensing-based detection techniques of HABs were restricted to three parameters or less and these limited number of parameters do not fully constrain ecosystem model parameters (Shen et al., 2012; Blondeau-Patissier et al., 2014). Although more robust field-based HAB detection and early warning systems are in place in some areas (Cusack et al., 2016), those systems are absent in many other locations where there is a need to monitor and predict HAB occurrences. Their absence could be related to the extensive resources needed to construct and maintain monitoring networks, to support the continuous sampling and analysis (geochemical, biological, and physical) of the investigated water bodies. In this study, I develop methodologies that utilize a large number of remote sensing-based water quality parameters together with optical properties that are extracted from readily available remote sensing datasets to map HAB occurrences and predict their distribution.

The study area is in the Charlotte County, Florida; it incorporates the county's coastal areas (width: 15 to 30 km) and nearby estuaries (Fig. 10). Like many other

coastal areas within the Gulf of Mexico, the study area has been subjected to persistent HAB outbreaks that pose serious environmental challenges to the county's tourism and fishery industries (Evans and Jones, 2001). Unfortunately, the study area lacks continuous field-based monitoring of water quality as it is challenging to cover large geographic areas with limited resources (Clark et al., 2017). The primary goals of this study involved identifying the factor(s) controlling HAB occurrences in the study area, developing same-day mapping and predictive models for HAB occurrences by utilizing daily remote sensing data, disseminating our findings, and automating the process.

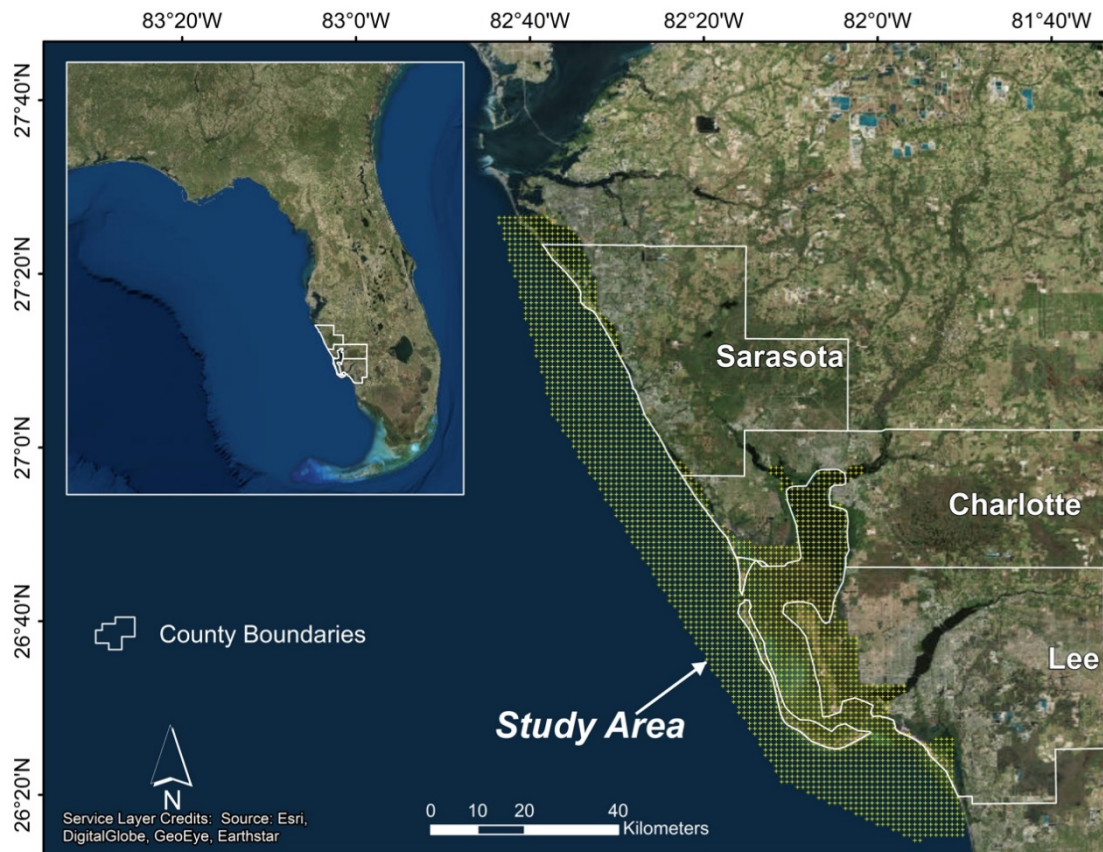


Figure 10. Figure showing the study area, which covers coastal waters (width: 15–30 km) surrounding Charlotte County in Florida. The study area also covers the brackish water within the estuarine systems where freshwater and seawater mix.

3.2 Methods

I accomplished the goals described above by developing multivariate linear regression statistical models, distributing our findings via a web-based interface and utilizing a geographic information system (GIS) framework for automation purposes. Data-driven models that rely on historical remote sensing and corresponding field data were developed to identify factors controlling the algal blooms and to forecast their occurrences. An inventory was compiled for the reported (dates and locations) HABs in the coastal waters surrounding Charlotte County by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWRI: <http://myfwc.com/research/redtide/monitoring/database/>), and a database was generated for remote sensing datasets that were acquired during the reported HAB occurrences. The compiled satellite and field data covered the period between January 2010 and October 2017 in which 213 HAB events were reported. The workflow involved three major steps: (1) downloading and processing of daily MODIS data; (2) developing multivariate regression models based on historical HAB occurrences; and (3) using the model for same-day mapping and forecasting HAB, automating the process, and publishing the findings (Fig. 11).

There were some basic assumptions I made during our study to facilitate the use of remote sensing to study harmful algal bloom. I assumed that algal colonies are stationary and are not moving across the pixels or along the water column.

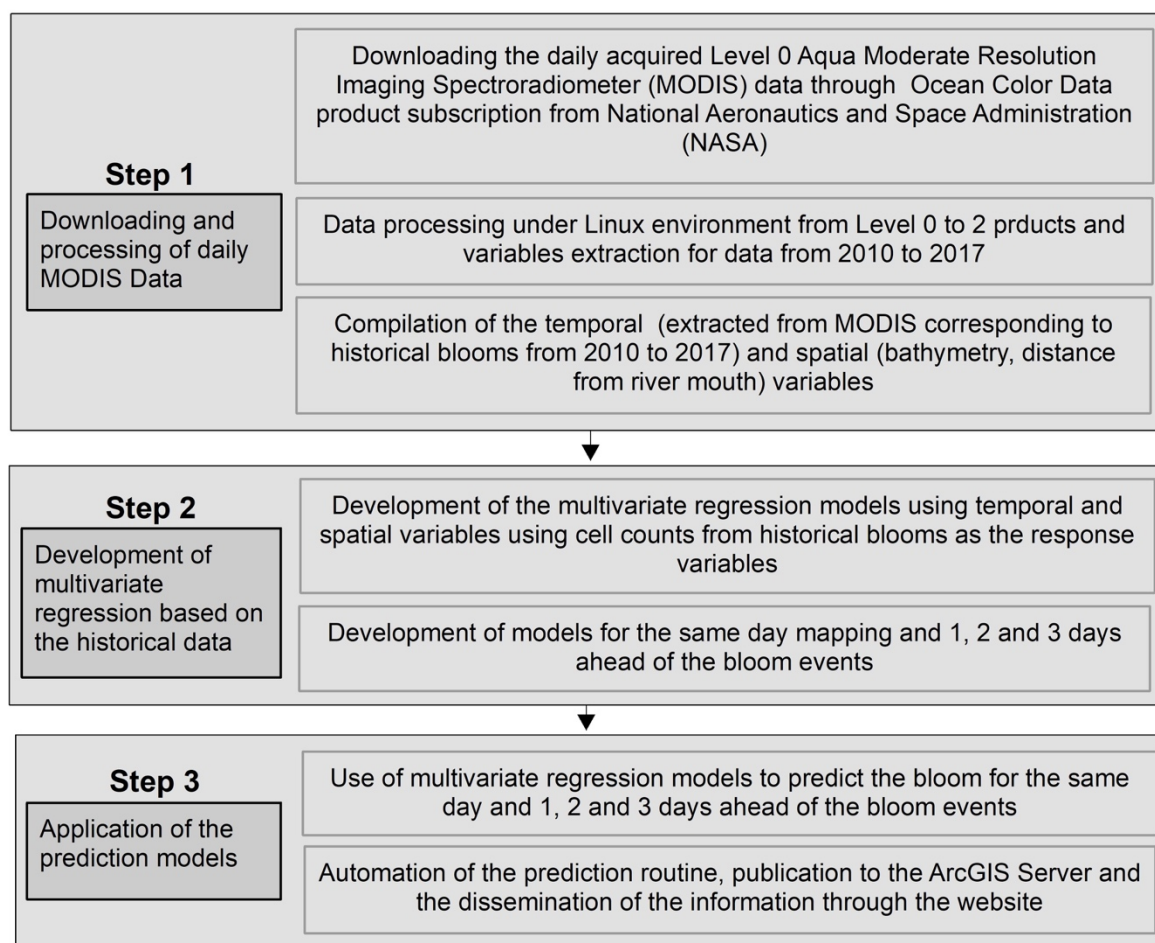


Figure 11. Three-step workflow established for harmful algal blooms mapping and forecasting.

3.2.1 Step 1: Data collection and processing

The first step involved the identification of temporal ocean color products and spatial variables that could control, or correlate with, the distribution of algal blooms in general and/or the HAB in the study area (in our case *K. brevis*). The selection of these variables was largely based on reported findings from similar settings elsewhere and, to a lesser extent, on our observations.

This step involved automatic downloading and processing of daily ocean color

data products acquired by the National Aeronautics and Space Administration (NASA) MODIS Aqua satellite. NASA's ocean color processing website (<https://oceancolor.gsfc.nasa.gov/>) provides an option for periodical data download for specified regions via a free data subscription service. I specified southwestern Florida as a region of interest, Aqua MODIS as a source of data and daily data as a download option. The automatic data download was scheduled using the task scheduling programs available within the Linux environment. The downloaded Level 0 data was processed to Level 1 and later to Level 2 using SeaDAS (NASA, Greenbelt, Maryland, USA, version 7.4) Ocean Color Science Software (OCSSW). Level 1 data has the radiometric and geometric calibrations applied and the ocean data products were extracted during the Level 2 processing. The applications of these calibrations correct for differences in acquisition geometry for the scenes although minor variation in ocean color products is unavoidable (Barnes and Hu, 2016). The OCSSW software was used to extract relevant temporal variables as shown in Table 3. The table shows the input, output, processor and parameters specified in the command line operator in Linux environment to enable unattended data extraction. A total of 13 ocean color products were extracted from the downloaded MODIS products. These products include: euphotic depth, ocean chlorophyll three-band algorithm for MODIS (chlorophyll-a OC3M), chlorophyll-a Generalized Inherent Optical Property (GIOP), chlorophyll-a Garver-Siegel- Maritorena (GSM), fluorescence line height (FLH), a diffuse attenuation coefficient for downwelling irradiance at 490 nm (Kd₄₉₀), particulate backscattering coefficient at 547 nm (bbp_{547_giop}), turbidity index, sea surface temperature (SST), wind direction, wind speed, chromophoric dissolved

organic matter (CDOM) index (Morel and Gentili, 2009) and Secchi disk depth (Zsd morel, based on Morel version) (Tyler, 1968; Morel et al., 2007). Additional spatially relevant variables were considered, as well. Our preliminary inspection of these products revealed large and rapid variations in chlorophyll-a content, SST, the attenuation coefficient, and euphotic depth in proximity to the shoreline and to the freshwater outlets (river mouth; Fig. 12), thus suggesting that bathymetry and distance from the river mouth should be incorporated in the model's development. Uniformly spaced grid points were used to extract the values from products of different resolutions and subsequent processing was done on the same grid to achieve computational efficiency.

Table 3. Overview of the inputs, outputs, processor and relevant parameters applied in SeaDAS OCSSW to extract level 2 products.

Data Input	Data Output	Processor	Relevant Parameters	Task
Level 0	Level 1A	modis_L1A.py	Default	Sensor calibration
Level 0	Level 1B	modis_L1B.py	Default	File conversion
Level 1A	GEO file	modis_GEO.py	Default	File conversion
Level 1A and 1B	Level 2	l2gen	Product Selector: Radiances/Reflectances (Rrs); Calibration option: Standard processing; mode: forward processing; resolution: 1k resolution including aggregated 250 and 500 land bands; Gas option: 1-Ozone, 2-CO ₂ , 4: NO ₂ , 8-H ₂ O;	Reflectance calculation

Table 3 - continued

			Glint option: 1-standard glint correction	
Level 2		l2mapgen	Products: Zeu_morel (euphotic depth), Zsd_morel (Secchi disk depth), cdom_index, chlor_a, chl_gsm, Kd_490 (diffuse attenuation coefficient), SST, chl_giop, FLH, wind speed, wind angle, bbp_547_giop (particulate backscattering coefficient), tindx_morel (turbidity index); Flag use: flags to be masked; mask: default mask to land, cloud and glint; Atmospheric Correction: 1 (on)	Level 2 product generation

The collected ocean color products were later checked for consistency and significance. Discontinuous data were not considered. For example, the data for CDOM index was found to be discontinuous and patchy over the investigated period (2010 to 2017) and was thus omitted from the list of potential variables considered for model development.

An exploratory stepwise linear regression was conducted to identify the determinant and significant variables, as well as the optimum combination of the variables. Spatial Statistics extension in ArcGIS together with Minitab software were used for these analyses. The significance (0.05 or 95% of confidence level) of the variables was investigated using the p-value and R-square value and the combination of variables with the lowest Akaike information criterion (AIC) value (Akaike, 1974) was considered to represent the best model fit or performance. Variables that were found to be highly correlated (redundant) and insignificant were omitted. The

variables that contributed to the multicollinearity (redundant variables) were identified using the extracted Variance Inflation Factor (VIF) (O'Brien, 2007) values. A variable with a VIF value exceeding 7.5 was considered redundant with the second highest VIF value. In cases where multiple variables were identified as being redundant, the significant variables were retained and the insignificant ones were omitted. Using water clarity measurements as an example, Secchi disk depth was found to be redundant with euphotic depth, and the former was found to be less significant and was dropped. Following the omission of redundant variables, the multivariate regression was run again to make sure the R-square value and model's significance did not decrease. The overall target of this iterative exercise was to obtain the highest R-square value with a minimal number of significant variables. Only 13 of the initial 15 variables were considered for model construction. The spatial and temporal variables included in the model are explained below.

(a) Euphotic Depth (m)

The euphotic depth represents the depth at which 1% of the light incident on the ocean's surface can reach (Ryther, 1956; Kirk, 1994; Morel et al., 2007). This depth provides a measure of the depth where light penetrates, nutrients and algae diminish, and productivity decreases (Behrenfeld and Falkowski, 1997). Water bodies with low euphotic depths generally have a high nutrient content, are more productive and eutrophic (Behrenfeld et al., 2005), and provide favorable conditions for HAB development (Anderson et al., 2002). The euphotic depth was calculated using the technique documented in a previous study (Morel et al., 2007).

(b) Wind Direction (Degrees) and Wind Speed (m/s)

The wind direction and speed can affect the distribution of algal blooms in three major ways: (1) prevailing wind directions create ocean currents and water exchanges that transport HAB cells (Edwards et al., 1996; Raine et al., 2010) and biotoxins (Raine et al., 2010); (2) wind and bathymetry guide the location of nutrient upwelling facilitating the concentration of the algae (Cusack et al., 2016); and (3) winds can transfer the aerosols (Al Shehhi et al., 2013) promoting the growth of toxic phytoplankton (Al Shehhi et al., 2012). The wind direction and wind speed were calculated using a reflectance model based on the Cox-Munk wave-slope distribution (Cox and Munk, 1954).

(c) Chlorophyll-a (mg/m³)

The concentration of chlorophyll-a provides direct measurements of the growth of the algae in aquatic environments (Baban, 1996). Three different types of algorithms were used to compute the chlorophyll-a content: chlorophyll-a OC3M [ocean chlorophyll three-band algorithm for MODIS, (O'Reilly et al., 2000)], chlorophyll-a GSM [Garver-Siegel-Maritorena, (Maritorena et al., 2002)] and chlorophyll-a GIOP [Generalized Inherent Optical Property, (Franz and Werdell, 2010)]. These algorithms use different sets of reflectance bands to estimate phytoplankton biomass (Tilstone et al., 2013) and these have been validated with field observations in different parts of the world (Campbell and Feng, 2005; Komick et al., 2009; Hattab et al., 2013; Lah et al., 2013). An increased chlorophyll-a concentration has been taken as a strong indicator of

HAB distribution (Tang et al., 2006; Wei et al., 2008), and chlorophyll-a OC3M data has been used for detecting HAB along the west coast of Florida (Carvalho et al., 2010). Three types of chlorophyll-a measurements were considered in this study as they were found to be correlated with algal cell count during the exploratory multivariate regression. These products were not redundant to each other suggesting that the algorithms were either picking up unique spectral signatures exhibited by chlorophyll-a in the optically complex estuarine environment, or it may be the results of uncertainties in the algorithms (Shang et al., 2014).

(d) Diffuse Attenuation Coefficient

The Diffuse Attenuation Coefficient for downwelling irradiance at 490 nm (K_d_{490} ; m^{-1}) measures the attenuation of the light (blue to green) for turbid water (Austin and Petzold, 1981; Lee et al., 2005). A study in the Bohai Sea (Chen et al., 2014) showed that the attenuation coefficient can be used as a proxy for the growth of phytoplankton in turbid coastal waters given that the blue to green light attenuation positively correlates with scattering particles (e.g., HABs). A high correlation between chlorophyll-a concentration and diffuse attenuation coefficient was observed under harmful red tide conditions in the Persian Gulf using MODIS data (Ghanea et al., 2016). In another study done in the coastal waters of India, HABs were detected using satellite derived chlorophyll-a and diffuse attenuation coefficient images and were also validated through in situ measurement (Sarangi and Mohammed, 2011). The diffuse attenuation coefficient was calculated using the technique described in a previous study (Lee et al., 2005).

(e) Turbidity Index

The turbidity index provides a measure for the clarity of the water through the scattering of light caused by suspended particles (Austin, 1973; Davies-Colley and Smith, 2007). Spatial and temporal variations of turbidity in water bodies has been successfully used to identify phytoplankton blooms (May et al, 2003; Cloern, 1987). Although a turbidity index is not a direct indicator of HAB occurrences, it has been successfully used to estimate the severity of a HAB once it was independently detected (Kahru et al., 2004). The turbidity index was calculated using procedures described in a previous study (Morel and Belanger, 2006).

(f) Particulate Backscattering Coefficient at 547 nm

This is the backscattering coefficient of particles at 547 nm. The backscattering coefficient as determined by satellite sensors and in situ measurements has been used in the past to identify the distribution of HABs. A research study (Zhao et al., 2013) employed satellite-based and underwater glider measurements of the backscattering coefficient at 547 nm to detect *K. brevis* blooms in the Gulf of Mexico and verified their findings by in situ observations. A backscattering coefficient at 551 nm extracted from a Visible Infrared Imaging Radiometer Suite (VIIRS) sensor, which is analogous to the MODIS backscattering coefficient at 547 nm, was used in conjunction with fluorescence data to detect the *K. brevis* bloom at the West Florida shelf (El-habashi et al., 2016). In the same area, in situ measurements of the backscattering coefficient at 551 nm and chlorophyll-a data were successfully used to detect a *K. brevis* bloom (Cannizzaro et al.,

2009). The backscatter coefficient of particles at 547 nm was calculated using an algorithm available in the literature (Lee et al., 2002; Gordon, 1998).

(g) Sea Surface Temperature (°C)

SST influences phytoplankton productivity in multiple ways: (i) individual biological species including algal blooms thrive under different and specific temperature regimes, and (ii) the availability and solubility of many biochemical materials needed for their growth and development is temperature dependent (Edwards et al., 1996; Goldman and Carpenter, 1974). Many studies have shown a correlation between SST and algal bloom distributions in the Mediterranean Basin (Bricaud et al., 2002), Kuwait Bay (Elkadiri et al., 2016; Glibert et al., 2002) and on a global scale (Hallegraeff, 2010). The productivity of *K. brevis* increases in the fall and early spring at the west Florida shelf primarily because of the ideal temperature conditions during these times (Hu et al., 2005). Increased SST was found to be conducive to HAB development in the coastal waters of Oman (Sarma et al., 2013) and in Gulf of Mexico (Errera et al., 2014).

(h) Fluorescence Line Height (FLH)

Fluorescence line height (FLH) provides the relative measure of radiance leaving the sea surface in the chlorophyll fluorescence emission band (Xing et al., 2007). It has been successfully used in the detection of chlorophyll-a in several studies (Neville and Gower, 1977; Pan et al, 1989; Hoge et al., 2003) including the one in southwestern Florida ((Hu et al., 2005). A review of previous studies shows a positive

correlation between the MODIS-derived fluorescence and chlorophyll-a concentration in ocean waters with algal blooms (Xing et al., 2007; Behrenfeld et al., 2009). More recently, an in situ FLH measurement was done in conjunction with the backscattering coefficient to map the distribution of *K. brevis* in the Gulf of Mexico (Zhao et al., 2013). Similarly, FLH derived from VIIRS was used to detect *K. brevis* blooms at the West Florida Shelf (El-habashi et al., 2016).

Although other pigments (chlorophyll-b, chlorophyll-c, phycoerythrin and carotenoids) are common in HAB, chlorophyll-a estimate is the first choice in oceanography because of the practical reasons (Wang et al., 2017; Hoepffner and Sathyendranath, 1991). It has been difficult to attain the detection limit of phycoerythrin using MODIS, instead there has been more efforts on the absorption bands at 495 nm and 545 nm (Hoge et al., 1999). MODIS provides fluorescence band (676 nm) to derive FLH primarily for HAB detection (Shen et al., 2012; Bernard et al., 2014). FLH has been successfully used to detect *K. brevis* bloom in the Gulf of Mexico (Hu et al., 2005; Tomlinson et al., 2009). The concentration of *K. brevis* was found to have direct correlation with FLH in the Charlotte harbor in Florida (Hu et al., 2005).

(i) Bathymetry (m)

Shelf properties, including bottom topography, influence the distribution of HAB in many ways (Figueiras et al., 2006). For example, water stratification, which is controlled in part by bottom topography, inhibits productivity (Lozier et al., 2011), whereas the vertical mixing and added nutrient supply in shallow waters can enhance the primary productivity in coastal ecosystems (Lozier et al., 2011). Our study site, and

the continental shelf systems and coastal areas in general, are considered to be vulnerable to HAB occurrences due to the accumulation of biomass (Seegers et al., 2015). Bathymetric data acquired from United States Geological Survey (<https://coastal.er.usgs.gov/flash/bathy-entireFLSH.html>) were used as one of the spatial variables.

(j) Distance from the River Mouth (m)

Riverine organics are major sources of nutrients for the West Florida Shelf of the Gulf of Mexico (Del Castillo et al., 2001). The riverine discharge provides high nutrient loads (Del Castillo et al., 2000) that largely control the phytoplankton population and eutrophication around the river discharge locations and adjoining estuarine systems (Heil et al., 2007; Pinckney et al., 2001). The distance from the mouth of the river was computed using the Euclidean Distance function in ArcGIS (Environmental Systems Research Institute, Redlands, California, USA, version 10.5).

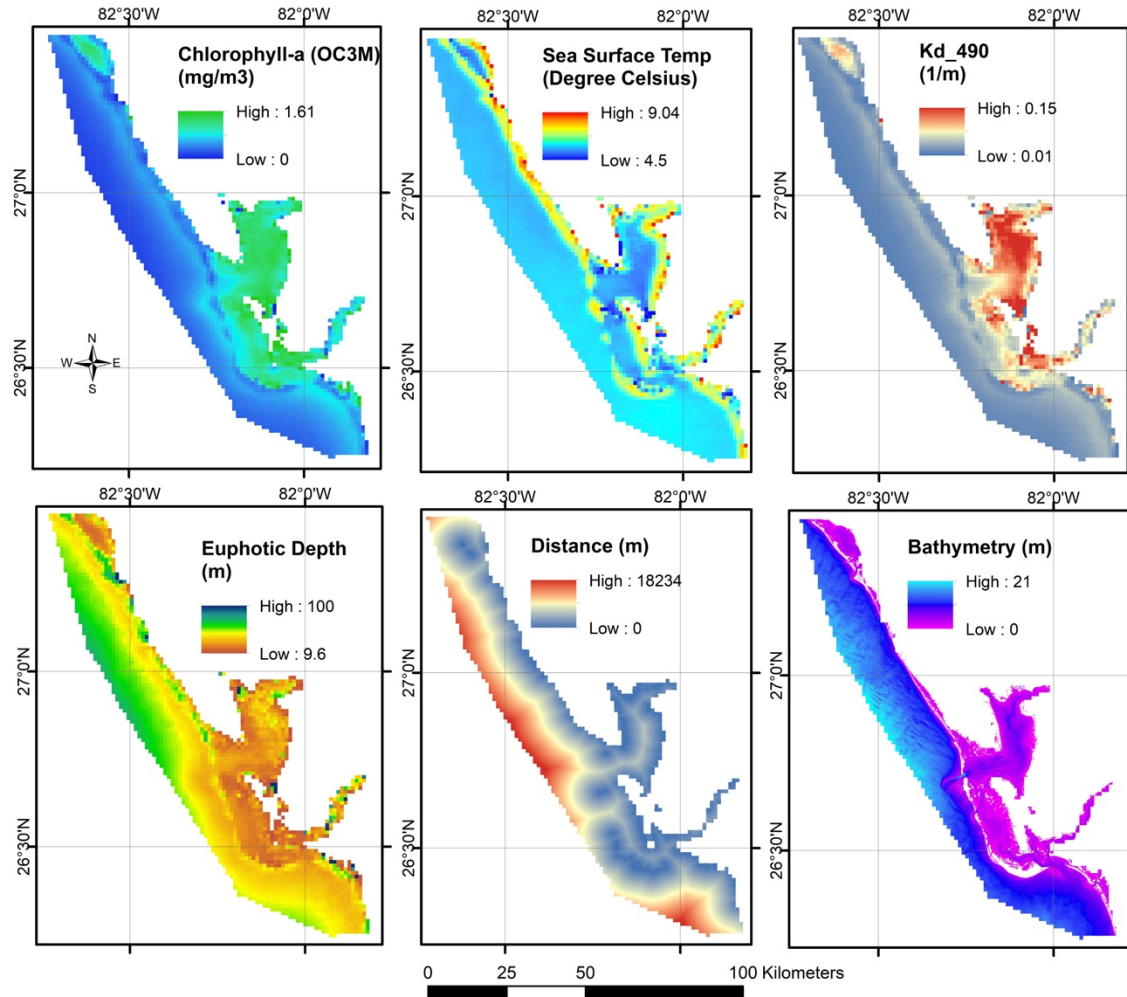


Figure 12. Mean values for the significant variables including chlorophyll-a (OC3M), SST, diffuse attenuation coefficient, and euphotic depth calculated from MODIS products acquired throughout the period 2010 to 2017 over the study area. The distance from river mouth and bathymetry data are also shown.

3.2.2 Step 2: Development of prediction model

The logarithm of *K. brevis* cell counts (base 10) in samples as analyzed by FWRI was used as the response variable because the growth of the algae takes place exponentially. Measurements were largely performed in response to reported *K. brevis*

blooms around Charlotte County, Lee County to the south, and Sarasota County to the north. The cell count responses were lumped into three groups: (i) no bloom (cell count ≤ 300 per/L), (ii) low concentration (cell count > 300 and $< 10,000$ per L) and (iii) high concentration (cell count $\geq 10,000$ per L). I adopted the threshold values used by the Harmful Algal Bloom Observation System (<https://habsos.noaa.gov/>) in categorizing cell count to facilitate comparisons with NOAA's observations. During the investigated period (2010 to 2017), 128 blooms were reported with cell counts higher than 300 per L. Each of the input variables was normalized to the -1 to $+1$ range because the inputs displayed large variations in range and magnitude. Such variations, if not accounted for, could affect model outputs. For each inventoried location, I extracted the values of the normalized input variables.

Four linear regression models were constructed: (i) same-day; (ii) one day in advance; (iii) two days in advance; and (iv) three days in advance. For each of the models, data were divided into training (80%) and testing (20%) datasets. The training data were used to develop the regression, and the accuracy assessment was done on the testing datasets. For the same-day model, the regression was conducted on 80% of the reported HABs occurrences (102 unique-day bloom events) against the variables on bloom days (days when blooms were reported). For the one day in advance model, the regression was conducted for the response versus the variables acquired one day in advance of the bloom day. Similarly, for the two-day and three-day in advance models, the response was regressed against the variables acquired two and three days in advance, respectively. Each of the four models had its individual datasets (response and variables) for regression and validation. Predictive models (ii, iii, and iv) were

designed to capture the onset of the HABs in contrast to those that developed days earlier and continued in the following days. To this end, a bloom reported on day_n was excluded from the one-day advance dataset if another was reported in the same location in day_{n-1}. Similarly, a bloom reported on day_n was excluded from the two-day dataset if another was reported in the same location in day_{n-1} or day_{n-2} and from the three-day dataset if a bloom was reported on day_{n-1}, day_{n-2} or day_{n-3}. The multivariate regression model was developed for each group of the data. For any new satellite data for any specific day, a respective regression was used to predict the HAB on the same-day and one, two and three days in advance of the potential HAB occurrence.

3.2.3 Step 3: Prediction

The generated regression equations were utilized for same-day mapping and one-, two- and three-day advance predictions of HAB. The regression models were developed for three bloom lag periods and applied to the collective set of variables. The results (same-day mapping and one-, two- and three-day advance prediction of HAB) are published on our website (<http://www.esrs.wmich.edu/webmap/bloom/>) using the ArcGIS server and ArcGIS API for JavaScript. The MODIS data for every day is acquired at ~4 pm, is made available for download on NASA's website at ~5 pm, and is processed for HAB occurrences and published on our website at ~10 pm. This process was coded in Python 2.7 to allow the program to run automatically at the same time every day.

3.3 Results and Discussion

The prediction was done in two phases: (a) nowcasting and (b) one, two and three days in advance forecasting. The model outputs are provided in Tables 4 and 5. Table 4 lists the selected variables together with their respective contributions to the response variable for the same-day and the one-, two- and three-day predictions. Table 5 provides the multivariate regression coefficients for each of the selected variables for the same-day and the one-, two- and three-day predictions. The sign (+/-) in front of the coefficient for each variable indicates the nature (positive/negative) of the relationship between the variable in question and the response. The assessment of the performance of the four models is presented in Fig. 13. The accuracy of the same day was the highest (90.5%), and the accuracies of the one-day, two-day, and three-day prediction models were assessed at 65.6%, 72.1%, and 71.9%, respectively. The prediction accuracies were calculated based on the three categories of the cell count that were pre-established instead of using binary value (presence and absence of the bloom) as an indicator of the success. In order to obtain stringent prediction criteria, I integrated locational accuracy as a part of verification process as I was using cell count data with spatial information provided by FWRI.

Table 4. Selected variables and their respective contributions to the response variable for same-day nowcasting, and one-, two- and three-day predictions.

	Same-Day Nowcasting	Forecasting		
		One Day in Advance	Two Days in Advance	Three Days in Advance
1	Bathymetry (35.9%)	Bathymetry (16.1%)	Euphotic Depth (25%)	Euphotic Depth (16.6%)
2	Euphotic depth (22.1%)	SST (15.5%)	Chlorophyll-a (OC3M) (14.2%)	Distance to river mouth (16.1%)
3	Wind direction (7.1%)	Wind direction (13.4%)	Distance to river mouth (14%)	Chlorophyll-a (OC3M) (15.1%)
4	Chlorophyll-a (OC3M) (6.7%)	Chlorophyll-a (OC3M) (10.3%)	Diffuse attenuation coefficient (Kd_490) (8.9%)	Wind direction (10%)
5	Wind speed (5.8%)	Diffuse attenuation coefficient (Kd_490) (9.9%)	SST (7.7%)	SST (9.3%)
6	Distance to river mouth (5.5%)	Distance to river mouth (9.1%)	Wind direction (6.4%)	Chlorophyll-a (GSM) (7.9%)
7	Chlorophyll-a (GIOP) (3.4%)	Wind speed (7.6%)	Fluorescence line height (5.4%)	Turbidity Index (7%)
8	Fluorescence line height (3.2%)	Turbidity index (7.1%)	Turbidity Index (5.4%)	Particulate backscattering coefficient (bbp_547_giop) (4.6%)
9	Diffuse attenuation coefficient (Kd_490) (3.1%)	Particulate backscattering coefficient (bbp_547_giop) (5.2%)	Bathymetry (4.8%)	Fluorescence line height (4.5%)
10	Chlorophyll-a (GSM) (2.4%)	Chlorophyll-a (GSM) (3.2)	Chlorophyll-a (GSM) (3.3%)	Wind speed (3%)
11	Turbidity index (2.4%)	Euphotic depth (1.9%)	Chlorophyll-a (GIOP) (2.4%)	Bathymetry (2.8%)

Table 4 - continued

12	Particulate backscattering coefficient (bbp_547_giop) (1.4%)	Chlorophyll-a (GIOP) (0.5%)	Wind speed (1.5%)	Chlorophyll-a (GIOP) (1.9%)
13	SST (0.8%)	Fluorescence line height (0.2%)	Particulate backscattering coefficient (bbp_547_giop) (0.7%)	Diffuse attenuation coefficient (Kd_490) (1.3%)

Table 5. Multivariate regression coefficients for each variable in predicting HABs for same-day mapping, and one-, two- and three-day advanced predictions.

Variables	Coefficients			
	Same-Day	One Day in Advance	Two Days in Advance	Three Days in Advance
Bathymetry (m)	-0.2662	0.0609	0.0874	0.1326
Euphotic Depth (m)	0.0296	0.0022	0.0231	0.0189
Wind Direction (degrees)	216.5790	270.1179	162.3239	287.9521
Chlorophyll-a (OC3M) (mg/m ³)	0.5120	0.5418	0.9005	1.0869
Wind Speed (m/s)	-250.5957	214.0178	53.2687	119.2459
Distance to Mouth of River (m)	-0.1593	0.0001	0.0001	0.0001
Chlorophyll-a GIOP (mg/m ³)	0.2617	0.0044	0.1549	-0.1380
Normalized Fluorescence Line Height (mWcm ⁻² um ⁻¹ sr ⁻¹)	-395.2367	0.0001	-551.1232	-514.3536
Diffuse Attenuation Coefficient (m ⁻¹)	2.6613	5.2936	6.2945	1.0121
Chlorophyll-a GSM (mg/m ³)	0.1845	0.1417	0.2116	0.5693
Turbidity Index	73.0929	143.0333	135.8686	202.1670
Particulate Backscattering Coefficient (m ⁻¹)	4287.9827	-12,551.1376	-1854.9512	-3779.3400
SST (°C)	-0.0188	-0.2164	-0.1514	-0.2002
Intercept	2.6629	1.1849	-0.1563	-0.5590

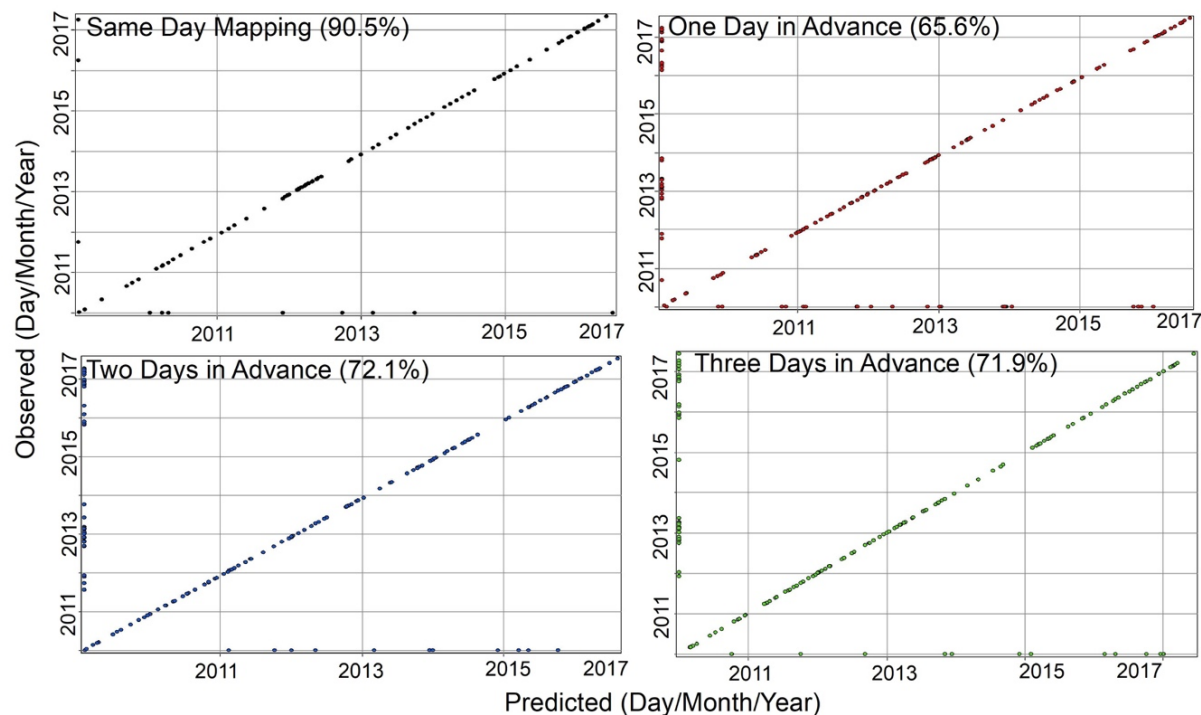


Figure 13. Assessment of the accuracy of the generated multivariate models (same-day mapping and one-, two- and three-day advanced predictions). The points on the diagonal line represent the bloom events that were observed and also predicted. The points on the vertical axis represents the algal bloom events that were observed but not predicted by the model. The points on the horizontal axis are the algal bloom events that were predicted but not observed. The accuracy of each prediction models are given in parenthesis.

The information provided in Tables 4 and 5 can be used to interpret the nature of the relationship between HAB occurrences and the individual variables. The information can also be used to determine the directionality (negative or positive) of the relationship and to evaluate the comparative contributions of the individual variables to the response variable. For same-day mapping (or nowcasting), the

bathymetry, euphotic depth, wind direction, chlorophyll-a (OC3M, and wind speed were found to have a 78% contribution to the response variable as presented in Table 4. For the one-day forecasting, bathymetry, SST, wind direction, chlorophyll-a (OC3M), and diffuse attenuation coefficient (KD_490) were found to have a 65% contribution to the response variable. For the two-day forecasting, euphotic depth, chlorophyll-a (OC3M), distance to river mouth, diffuse attenuation coefficient (KD_490) and SST were found to have 69% contribution to the response variables. The euphotic depth, distance to river mouth, chlorophyll-a (OC3M), wind direction and SST had a 67% contribution to the response variable for the three-day forecasting.

A 1:1 correspondence in the ranking and contributions to the response variable in the four models should not be expected given that the variables could have varying lag time effects on HAB development. For example, a study (Egerton et al., 2014) found a positive correlation between algal bloom events and nitrate and ammonium concentrations as early as five days prior to the bloom. Similarly, a study (Tian et al., 2018) found a positive correlation between HAB occurrences and temperature and aerosols particle distribution, which are the air-borne sources of phosphate, iron and trace elements in the East China Sea. Higher concentrations of phosphorous and iron above the threshold did not correlate with the HAB events because these are limiting nutrients for HAB growth. The increase in concentration of nitrogen, however, correlated with the HAB concentration. The lag time between the spike in the nitrogen concentration in the aerosols and HAB event was two days. Similarly, in the coastal waters of Charlotte County, one can attribute the high ranking of some variables (distance to river mouth, and chlorophyll-a) and their relatively high contributions to

the response variable in our two- and three-day predictive models to the presence of a two day lag time for nutrients in rivers to reach the coastal waters and induce HABs.

Inspection of Tables 4 and 5 reveals differences in the rankings and contributions to the variabilities in the response variable among four solutions. However, there are multiple variables that appear to be important ($\geq 5\%$) for three or more of the four solutions. These include chlorophyll-a, Euphotic depth, SST, wind direction, Chlorophyll-a OC3M, distance to river mouth and turbidity index. Other variables appear to be consistently less important ($< 5\%$) in 3 or more of the model outputs. These include bathymetry, wind speed, chlorophyll-a GIOP, fluorescence line height, diffuse attenuation coefficient (K_d_{490}), chlorophyll-a GSM and particulate backscattering coefficient (bbp_{547_giop}).

Additional spatial and temporal relations are inferred from Tables 3 and 4. Shallow bathymetry seems to be an important factor for same-day predictions. The association of HABs with shallow bathymetry was inferred from the -ve sign of the coefficient for the bathymetry variable for the same-day prediction. Similarly, the association of HABs with increasing euphotic depth and turbidity is inferred from the +ve sign for the coefficient for these two variables (euphotic depth and turbidity index) in each of the four models. The chlorophyll-a (OC3M) content and wind direction show a positive correlation with bloom occurrences for all lag times, but the contribution of the individual variables to the response variable and their ranking varies for the investigated models. SST seems to be less important on the day of the bloom compared to the one-day, two-days and three-days advanced predictions. Blooms occur at cooler SST as indicated by the -ve coefficients. For same-day predictions, the shorter the

distance from the river mouth, the more likely HABs will develop as evidenced by the –ve coefficient value (Table 5). I suspect that the above-mentioned observations (rankings, contribution to the response variable, and the spatial and temporal relations) are largely of local nature, and thus comparisons with findings from published works are not straightforward. Moreover, such comparisons are also hampered by the paucity of similar applications that involve a large number of variables.

Despite the local and empirical nature of regression models, I favored this method over the analytical and semi-analytical solutions given the lack of continuous field-based datasets that are required for the application of these analytical approaches. The multivariate regression method was also favored over other statistical approaches (e.g., artificial neural networks, principal component analysis) that do not provide insights into the nature and the contributions of the factors controlling HAB occurrences. Although other techniques may provide better one time prediction than multivariate regression, this is the only technique supported in ArcGIS server environment that allows publication of results in the present form via web-based GIS. This is the reason why an operational algal bloom early warning system are nonexistent that utilize other prediction techniques (e.g., machine learning, hydrogeological modeling, principal component analysis) and are based on satellite-derived variables on a daily basis.

There are several limitations with the applied methodology. There can be differences between the time a bloom was reported and the time it was captured by the satellite imagery. The vertical and lateral movement of the algae can also occur in the

water column due to changes in temperature, stratification and other factors. In field controlled and natural environments, a previous study (Evens et al., 2001) showed a decrease in fluorescence before reaching the maximum value under natural photosynthetically available radiation, while another study (Kamykowski et al., 1998) reported diurnal variations in algae populations in the surface and in the mid-column. Similar diurnal variations were reported for *K. brevis* in the West Florida Shelf (Vargo, 2009). Our approach does not consider this, but it assumes that the algae lie on the surface and are stationary. With the current temporal and spatial resolution of MODIS, diurnal physiological and ecological variations are not captured in the analysis. This could also be the reason why the contributions of the independent variables to the response variable fluctuates in different lag days (Table 4) in our model leading to low prediction accuracy one day prior to the bloom compared to other days as shown in Fig. 13. Because of these practical limitations, physiological responses of the algae cannot be understood using the satellite data alone.

The developed models apply constant lag times of one, two or three days for all of the variables. Ideal models should instead apply lag times that produce an optimum target response. These lag times will undoubtedly differ from one variable to another, and such an application will enhance the predictability of the model. Unfortunately, this enhancement is difficult to achieve as that it requires a continuous acquisition of satellite imagery over an extended period, and such daily acquisitions are halted on cloudy days. Due to the weather dependency of the temporal variables, it is difficult to develop models that include multiple variables with varying time lags. With the archive of MODIS data processed from 2010 to 2017, I was only able to extract 110 samples

with continuous data for three consecutive days. This number dropped to 57, 29 and 21 for four, five and six consecutive days respectively. Thus, although it would be extremely helpful to study the onset of bloom using the data acquired on different lag days, with the current use of MODIS data, it is extremely difficult to implement that technique. The acquisition of data regardless of whether it was a cloudy or non-cloudy day with alternative platforms such as Unmanned Aerial Vehicles (UAV) could enable the continuous acquisition of remote sensing data. If this happens, the onset of blooms in the Charlotte bay could be studied more rigorously and that would offer increased predictability.

The daily acquisition of MODIS imagery has offered tremendous benefits to this research, but the coarse temporal resolution has limited our capacity to adequately and optimally investigate the spatial trends and the factors that contribute to the onset of HABs. With the coarse spatial resolution of MODIS derived variables (e.g., SST: 1000 m; turbidity index: 500), the predictability of HAB detection is limited. The option to downscale satellite derived data is also impractical due to the lack of real time measurements at a finer scale. In this context, I am utilizing proxies (e.g., euphotic depth and turbidity indices) that can account for the nutrient content in the aquatic system despite its resolution. The predictability of the developed models could be improved with the inclusion of daily measurements of nitrate and phosphate in the water that can be used in the model. Together with the nutrient data, precipitation and stream flow data from nearby streams can be included as additional variables in the model, which could assist in understanding of, and accounting for, the seasonality of bloom occurrences. The inclusion of these variables in our model will not only enhance

our abilities to HABs in Charlotte bay, but also assist in understanding the factors causing bloom propagation in the study area. The current problem of discontinuous and coarse resolution of the MODIS data could be addressed locally if it was replaced by UAVs datasets. If I was to use UAV-generated datasets, the investigated area will have to be narrowed down to the immediate coastal waters to reduce the operational costs and to simplify the logistics involved in permitting flights. UAVs acquire high-resolution images devoid of atmospheric influences. These new data acquisition systems could increase the accuracy, predictability and replicability of our model in Charlotte County and elsewhere in the world. These methods would enable the construction of robust models that account for varying lag times and produce high-resolution (spatial and temporal) prediction. Although, the current MODIS spectral resolution is not perfect for HAB detection, it allows daily prediction until other option such as UAV is pursued in this area. The adopted empirical methodology could be applicable to many other coastal areas worldwide, yet it is to be expected that different sets of regression relationships should be developed for the individual areas to represent the local conditions that affect HAB occurrences.

3.4 Conclusion

The study focused on developing an early warning system for *K. brevis*-related HABs off the coast of South Florida. I used historical field HAB data from 2010 to 2017 to develop a multivariate regression and determine the contributions of the significant variables to the response variable under different prediction scenarios. The prediction system involved the same-day nowcasting method and forecasting for one, two and

three days in advance of the onset of the bloom. The same-day nowcasting provided 90% accuracy, whereas the one, two and three days in advance forecasting provided 65%, 72% and 71% accuracies, respectively. The investigation took advantage of ocean color data to develop methodologies and procedures that may enhance decision-making processes, improve citizens' quality of life, and strengthen the local economy. Even though this project focuses on the *K. brevis* related HAB in Charlotte County and its surrounding neighbours, the model can be replicated for other species and can be applied in other areas. The prediction system can be utilized to plan uses of coastal waters for recreational purposes and other environmental services. Monitoring the extent and intensity of HABs could be used to improve the environmental and socioeconomic status of this area and develop long-term environmental programs and policies. This monitoring and early warning system for HABs could provide benefits, in Charlotte County to the public, policy makers, and the scientific community and could assist local agencies in developing solutions and plans to mitigate HABs.

3.5 References

- Ahn, Y. H., Shanmugam, P., Ryu, J. H., and Jeong, J.: Satellite detection of harmful algal bloom occurrences in Korean waters, *Harmful Algae*, 5, 213–231, <https://doi.org/10.1016/j.hal.2005.07.007>, 2006.
- Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, 716–723, <https://doi.org/10.1109/TAC.1974.1100705>, 1974.
- Al Shehhi, M. R., Gherboudj, I., and Ghedira, H.: Temporal-spatial analysis of chlorophyll

- concentration associated with dust and wind characteristics in the Arabian Gulf, In Proceedings of the OCEANS-Yeosu, Yeosu, pp. 1–6, <https://doi.org/10.1109/OCEANS-Yeosu.2012.6263404>, 2012.
- Al Shehhi, M. R., Gherboudj, I., Zhao, J., Mezhoud, N., and Ghedira, H.: Evaluating the performance of MODIS FLH ocean color algorithm in detecting the Harmful Algae Blooms in the Arabian Gulf and the Gulf of Oman, In OCEANS 2013 MTS/IEEE—San Diego, An Ocean in Common: San Diego, CA, USA, pp. 1–7, 2013.
- Aleynik, D., Dale, A. C., Porter, M., and Davidson, K.: A high resolution hydrodynamic model system suitable for novel harmful algal bloom modelling in areas of complex coastline and topography, *Harmful Algae*, 53, 102–117, <https://doi.org/10.1016/j.hal.2015.11.012>, 2016.
- Amin, R., Zhou, J., Gilerson, A., Gross, B., Moshart, F., and Ahmed, S.: Novel optical techniques for detecting and classifying toxic dinoflagellate *Karenia brevis* blooms using satellite imagery, *Opt. Express*, 17, 9126–9144, <https://doi.org/10.1364/OE.17.009126>, 2009.
- Anderson, D. M., Glibert, P. M., and Burkholder, J. M.: Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences, *Estuar. Coast*, 25, 704–726, <https://doi.org/10.1007/BF02804901>, 2002.
- Austin, R. W.: Problems in Measuring Turbidity as a Water Quality Parameter, U.S. EPA Seminar on Methodology for Monitoring the Marine Environment: Seattle, WA, USA, October 1973.
- Austin, R. W., and Petzold, T. J.: The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner, In *Oceanography from Space*,

- Marine Sciences, 1st ed., Gower, J.F.R., Ed., Springer: Boston, MA, USA, Volume 13, pp. 239–256, ISBN 978-1-4613-3317-3, 1981.
- Baban, S. M. J.: Trophic classification and ecosystem checking of lakes using remotely sensed information, *Hydrol. Sci. J.*, 41, 939–957, <https://doi.org/10.1080/02626669609491560>, 1996.
- Balch, W. M., Eppley, R. W., Abbott, M. R., and Reid, F. M. H.: Bias in satellite-derived pigment measurements due to coccolithophores and dinoflagellates, *J. Plankton Res.*, 11, 575–581, <https://doi.org/10.1093/plankt/11.3.575>, 1989.
- Barnes, B. B., and Hu, C.: Dependence of satellite ocean color data products on viewing angles: A comparison between SeaWiFS, MODIS, and VIIRS, *Remote Sens. Environ.*, 175, 120–129, <https://doi.org/10.1016/j.rse.2015.12.048>, 2016.
- Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M.: Carbon-based ocean productivity and phytoplankton physiology from space. *Glob. Biogeochem. Cycles*, 19, GB1006, <https://doi.org/10.1029/2004GB002299>, 2005.
- Behrenfeld, M. J., and Falkowski, P. G.: A consumer's guide to phytoplankton primary productivity models, *Limnol. Oceanogr.*, 42, 1479– 1491, <https://doi.org/10.4319/lo.1997.42.7.1479>, 1997.
- Behrenfeld, M. J., Westberry, T.K., Boss, E. S., O'Malley, R. T., Siegel, D. A., Wiggert, J. D., Franz, B. A., McClain, C. R., Feldman, G. C., Doney, S. C., et al.: Satellite-detected fluorescence reveals global physiology of ocean phytoplankton, *Biogeosciences*, 6, 779–794, <https://doi.org/10.5194/bg-6-779-2009>, 2009.
- Bernard, S., Pitcher, G., Evers-King, H., Robertson, L., Matthews, M., Rabagliati, A., and Balt, C.: Ocean colour remote sensing of harmful algal blooms in the Benguela

- system, In Remote Sensing of the African Seas, Barale, V., Gade, M., Eds., Springer: Dordrecht, pp.185–203, ISBN 978-94-017-8007-0, 2014.
- Blondeau-Patissier, D., Gower, J. F. R., Dekker, A. G., Phinn, S. R., and Brando, V.E.: A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans, Prog. Oceanogr., 123, 123–144, <https://doi.org/10.1016/j.pocean.2013.12.008>, 2014.
- Bricaud, A., Bosc, E., and Antoine, D.: Algal biomass and sea surface temperature in the Mediterranean basin: Intercomparison of data from various satellite sensors, and implications for primary production estimates, Remote Sens. Environ., 81, 163–178, [https://doi.org/10.1016/S0034-4257\(01\)00335-2](https://doi.org/10.1016/S0034-4257(01)00335-2), 2002.
- Campbell, J. W., and Feng, H.: The empirical chlorophyll algorithm for MODIS: Testing the OC3M algorithm using NOMAD data, In Ocean Color Bio-Optical Algorithm Mini-workshop, New England Center, USA, 27–29 September, University of New Hampshire: Dunham, NY, USA, pp. 1–9, 2005.
- Cannizzaro, J. P., Hu, C., English, D. C., Carder, K. L., Heil, C. A., and Müller-Karger, F. E.: Detection of *Karenia brevis* blooms on the west Florida shelf using in situ backscattering and fluorescence data, Harmful Algae, 8, 898–909, <https://doi.org/10.1016/j.hal.2009.05.001>, 2009.
- Carvalho, G. A., Minnett, P. J., Fleming, L. E., Banzon, V. F., and Baringer, W.: Satellite remote sensing of harmful algal blooms: A new multi-algorithm method for detecting the Florida Red Tide (*Karenia brevis*), Harmful Algae, 9, 440–448, <https://doi.org/10.1016/j.hal.2010.02.002>, 2010.

- Cha, Y., Park, S. S., Kim, K., Byeon, M., and Stow, C. A.: Probabilistic prediction of cyanobacteria abundance in a Korean reservoir using a Bayesian Poisson model, *Water Resour. Res.*, 50, 2518–2532, <https://doi.org/10.1002/2013WR014372>, 2014.
- Chang, F. H., Uddstrom, M., Richardson, K., Pinkerton, M., and Beauchamp, T.: Feasibility of monitoring of major HAB events in New Zealand using satellite remote ocean color and SST images, In *Proceedings of the Workshop on Red Tide Monitoring in Asian Coastal Water*, Tokyo, Japan, 10–12 March, pp. 106–108, 2003.
- Chen, J., Cui, T., Tang, J., and Song, Q.: Remote sensing of diffuse attenuation coefficient using MODIS imagery of turbid coastal waters: A case study in Bohai Sea, *Remote Sens. Environ.*, 140, 78–93, <https://doi.org/10.1016/j.rse.2013.08.031>, 2014.
- Clark, J. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., Johnston, J. M., Ignatius, A., Myer, M. H., Loftin, K. A., Werdell, P. J., and Stumpf, R.P.: Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources, *Ecol. Indic.*, 80, 84–95, <https://doi.org/10.1016/j.ecolind.2017.04.046>, 2017.
- Cloern, J. E.: Turbidity as a control on phytoplankton biomass and productivity in estuaries, *Cont. Shelf Res.*, 7, 1367–1381, [https://doi.org/10.1016/0278-4343\(87\)90042-2](https://doi.org/10.1016/0278-4343(87)90042-2), 1987.
- Cox, C., and Munk, W.: The measurements of the roughness of the sea surface from photographs of the sun's glitter, *J. Opt. Soc. Am.*, 44, 838–850, <https://doi.org/10.1364/JOSA.44.000838>, 1954.
- Cusack, C., Dabrowski, T., Lyons, K., Berry, A., Westbrook, G., Salas, R., Duffy, C., Nolan, G.,

- and Silke, J.: Harmful Algal Bloom forecast system for SW Ireland. Part II: Are operational oceanographic models useful in a HAB warning system? *Harmful Algae*, 53, 86–101, <https://doi.org/10.1016/j.hal.2015.11.013>, 2016.
- Davies-Colley, R. J., and Smith, D. G.: Turbidity, suspended sediment, and water clarity: A review, *J. Am. Water Resour. Assoc.*, 37, 1085–1101, <https://doi.org/10.1111/j.1752-1688.2001.tb03624.x>, 2007.
- Del Castillo, C. E., Coble, P. G., Conmy, R. N., Müller-Karger, F. E., Vanderbloemen, L., and Vargo, G.A.: Multispectral in situ measurements of organic matter and chlorophyll fluorescence in seawater: Documenting the intrusion of the Mississippi River plume in the West Florida Shelf, *Limnol. Oceanogr.*, 46, 1836–1843, <https://doi.org/10.4319/lo.2001.46.7.1836>, 2001.
- Del Castillo, C. E., Gilbes, F., Coble, P.G., and Muller-Karger, F. E.: On the dispersal of riverine colored dissolved organic matter over the West Florida shelf., *Limnol. Oceanogr.*, 45 1425–1432, <https://doi.org/10.4319/lo.2000.45.6.1425>, 2000.
- Edwards, A., Jones, K., Graham, J. M., Griffiths, C. R., MacDougall, N., Patching, J., Richard, J. M., and Raine, R.: Transient coastal upwelling and water circulation in Bantry Bay, a ria on the SW coast of Ireland, *Estuar. Coast. Shelf Sci.*, 42, 213–230, <https://doi.org/10.1006/ecss.1996.0016>, 1996.
- Egerton, T. A., Morse, R. E., Marshall, H. G., and Mulholland, M. R.: Emergence of algal blooms: The effects of short-term variability in water quality on phytoplankton abundance, diversity, and community composition in a tidal estuary, *Microorganisms*, 2, 33–57, <https://doi.org/10.3390/microorganisms2010033>, 2014.

- El-habashi, A., Ioannou, I., Tomlinson, M. C., Stumpf, R. P., and Ahmed, S.: Satellite retrievals of *Karenia brevis* harmful algal blooms in the west Florida shelf using neural networks and comparisons with other techniques, *Remote Sens.*, 8, 377, <https://doi.org/10.3390/rs8050377>, 2016.
- Elkadiri, R., Manche, C., Sultan, M., Al-Dousari, A., Uddin, S., Chouinard, K., and Abotalib, A. Z.: Development of a coupled spatiotemporal algal bloom model for coastal areas: A remote sensing and data mining-based approach, *IEEE J. Sel. Top. Appl. Earth Observ.*, 9, 5159–5171, <https://doi.org/10.1109/JSTARS.2016.2555898>, 2016.
- Errera, R. M., Yvon-Lewis, S., Kessler, J. D., and Campbell, L.: Responses of the dinoflagellate *Karenia brevis* to climate change: pCO₂ and sea surface temperatures, *Harmful Algae*, 37, 110–116, <https://doi.org/10.1016/j.hal.2014.05.012>, 2014.
- Evans, G., and Jones, L.: Economic Impact of the 2000 Red Tide on Galveston County, A Case Study: Austin, TX, USA, 2001.
- Evens, T. J., Kirkpatrick, G. J., Millie, G. F., Chapman, D. J., and Schofield, O. M. E.: Photophysiological responses of the toxic red-tide dinoflagellate *Gymnodinium breve* (Dinophyceae) under natural sunlight, *J. Plankton Res.*, 23, 1177–1193, <https://doi.org/10.1093/plankt/23.11.1177>, 2001.
- Figueiras, F. G., Pitcher, G. C., and Estrada, M.: Harmful Algal Bloom dynamics in relation to physical processes, In *Ecology of Harmful Algae, Ecological Studies*, Graneli, E., Turner, J.T., Eds., Springer-Verlag: Berlin, Germany, Volume 189, pp.127–138, ISBN 978-3-540-32209-2, 2006.

- Fischer, J., and Kronfeld, U.: Sun-stimulated chlorophyll fluorescence: 1. Influence of oceanic properties, *Int. J. Remote Sens.*, 11, 2125–2147, <https://doi.org/10.1080/01431169008955166>, 1990.
- Fleming, L. E., Kirkpatrick, B., Backer, L. C., Walsh, C. J., Nierenberg, K., Clark, J., Reich, A., Hollenbeck, J., Benson, J., Cheng, Y. S., et al.: Review of Florida red tide and human health effects, *Harmful Algae*, 10, 224–233, <https://doi.org/10.1016/j.hal.2010.08.006>, 2011.
- Franz, B. A., and Werdell, P. J.: A generalized framework for modeling of inherent optical properties in ocean remote sensing applications, In *Proceedings of Ocean Optics*, Anchorage, AK, USA, 27 September–1 October 2010.
- Ghanea, M., Moradi, M., and Kabiri, K.: A novel method for characterizing harmful algal blooms in the Persian Gulf using MODIS measurements, *Adv. Space Res.*, 58, 1348–1361, <https://doi.org/10.1016/j.asr.2016.06.005>, 2016.
- Gillibrand, P. A., Siemering, B., Miller, P. I., and Davidson, K.: Individual-based modelling of the development and transport of a *Karenia mikimotoi* bloom on the North-West European continental shelf, *Harmful Algae*, 53, 118–134, <https://doi.org/10.1016/j.hal.2015.11.011>, 2016.
- Glibert, P. M., Harrison, J., Heil, C., and Seitzinger, S.: Escalating worldwide use of urea—A global change contributing to coastal eutrophication, *Biogeochemistry*, 77, 441–463, <https://doi.org/10.1007/s10533-005-3070-5>, 2006.
- Glibert, P. M., Landsberg, J. H., Evans, J. J., Al-Sarawi, M. A., Faraj, M., Al-Jarallah, M. A., Haywood, A., Ibrahim, S., Klesius, P., Powell, C., et al.: A fish kill of massive proportion in Kuwait Bay, Arabian Gulf, 2001: The roles of bacterial disease, harmful

- algae, and eutrophication, *Harmful Algae*, 1, 215–231,
[https://doi.org/10.1016/S1568-9883\(02\)00013-6](https://doi.org/10.1016/S1568-9883(02)00013-6), 2002.
- Goldman, J. C., and Carpenter, E. J.: A kinetic approach to the effect of temperature on algal growth, *Limnol. Oceanogr.*, 5, 756–766,
<https://doi.org/10.4319/lo.1974.19.5.0756>, 1974.
- Gordon, H. R.: In-orbit calibration strategy for ocean color sensors. *Remote Sens.*, 63, 265–278, [https://doi.org/10.1016/S0034-4257\(97\)00163-6](https://doi.org/10.1016/S0034-4257(97)00163-6), 1998.
- Gower, J. F. R., Brown, L., and Borstad, G.A.: Observations of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor, *Can. J. Remote Sens.*, 30, 17–25, <https://doi.org/10.5589/m03-048>, 2004.
- Hallegraeff, G. M.: Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge, *J. Phycol.*, 46, 220–235, <https://doi.org/10.1111/j.1529-8817.2010.00815.x>, 2010.
- Harred, L. B., and Campbell, L.: Predicting harmful algal blooms: A case study with *Dinophysis ovum* in the Gulf of Mexico, *J. Plankton Res.*, 36, 1434–1445, <https://doi.org/10.1093/plankt/fbu070>, 2014.
- Hattab, T., Jamet, C., Sammari, C., and Lahbib, S.: Validation of chlorophyll-a concentration maps from Aqua MODIS over the Gulf of Gabes (Tunisia): Comparison between MedOC3 and OC3M bio-optical algorithms, *Int. J. Remote Sens.*, 34, 7163–7177, <https://doi.org/10.1080/01431161.2013.815820>, 2013.
- Heil, C.A., Revilla, M., Glibert, P.M., and Mrasko, S.: Nutrient quality drives differential phytoplankton community composition on the southwest Florida shelf. *Limnol. Oceanogr.*, 52, 1067–1078, <https://doi.org/10.4319/lo.2007.52.3.1067>, 2007.

- Hoepffner, N., and Sathyendranath, S.: Effect of pigment composition on absorption properties of phytoplankton. *Mar. Ecol. Prog. Ser.*, 73, 11–23, <http://doi.org/10.3354/meps073011>, 1991.
- Hoge, F. E., Lyon, P. E., Swift, R. N., Yungel, J. K., Abbott, M. R., Letelier, R. M., and Esaias, W. E.: Validation of Terra-MODIS phytoplankton chlorophyll fluorescence line height. I. Initial airborne lidar results, *Appl. Opt.*, 42, 2767–2771, <https://doi.org/10.1364/AO.42.002767>, 2003.
- Hoge, F. E., Wright, C. W., Lyon, P. E., Swift, R. N., and Yungel, J. K.: Satellite retrieval of the absorption coefficient of phytoplankton phycoerythrin pigment: Theory and feasibility status, *Appl. Opt.*, 38, 7431–7441, <https://doi.org/10.1364/AO.38.007431>, 1999.
- Howarth, R. W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., Downing, J. A., Elmgren, R., Caraco, N., Jordan, T., et al.: Regional nitrogen budgets and riverine inputs of N and P for the drainages to the North Atlantic Ocean: Natural and human influences, *Biogeochemistry*, 35, 75–139, 1996.
- Hu, C., Muller-Karger, F. E., Taylor, C., Carder, K. L., Kelble, C., Johns, E., and Heil, C. A.: Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters, *Remote Sens. Environ.*, 97, 311–321, <https://doi.org/10.1016/j.rse.2005.05.013>, 2005.
- Kahru, M., Michell, B. G., Diaz, A., and Miura, M.: MODIS detects a devastating algal bloom in Paracas Bay, Peru, *Eos Trans. AGU*, 85, 465–472, <https://doi.org/10.1029/2004EO450002>, 2004.
- Kamykowski, D., Millegan, E. J., and Reed, R.E.: Biochemical relationships with the

- orientation of the autotrophic dinoflagellate *Gymnodinium breve* under nutrient replete conditions, Mar. Ecol. Prog. Ser., 167, 105–117, 1998.
- Kirk, J. T. O.: Light and Photosynthesis in Aquatic Ecosystems, 2nd ed., Cambridge University Press: New York, NY, USA, ISBN 0521459664, 1994.
- Komick, M., Costa, M., and Gower, J.: Bio-optical algorithm evaluation for MODIS for western Canada coastal waters: An exploratory approach using in situ reflectance, Remote Sens. Environ., 113, 794–804, <https://doi.org/10.1016/j.rse.2008.12.005>, 2009.
- Lah, N. Z. A., Reba, M. N. M., and Siswanto, E.: An improved MODIS standard chlorophyll-a algorithm for Malacca Straits water, In IOP Conference Series: Earth and Environmental Science, Kuching, Sarawak, Malaysia, 26–29 August, <https://doi.org/10.1088/1755-1315/18/1/012113>, 2013.
- Landsberg, J. H.: The effects of harmful algal blooms on aquatic organisms, Rev. Fish. Sci. Aquac., 10, 113–390, <https://doi.org/10.1080/20026491051695>, 2002.
- Lee, J. H. W., Hodgkiss, I. J., Wong, K. T. M., and Lam, I. H. Y.: Real time observations of coastal algal blooms by an early warning system, Estuar. Coast Shelf Sci., 65, 172–190, <https://doi.org/10.1016/j.ecss.2005.06.005>, 2005.
- Lee, Z., Carder, K., and Arnone, R.: Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters, Appl. Opt., 41, 5755–5772, <https://doi.org/10.1364/AO.41.005755>, 2002.
- Lee, Z. P., Du, K. P., and Arnone, R.: A model for the diffuse attenuation coefficient of downwelling irradiance, J. Geophys. Res., 110, C02016, <https://doi.org/10.1029/2004JC002275>, 2005.

- Lozier, M. S., Dave, A. C., Palter, J. B., Gerber, L. M., and Barber, R. T.: On the relationship between stratification and primary productivity in the North Atlantic, *Geophys. Res. Lett.*, 38, L18609, <https://doi.org/10.1029/2011GL049414>, 2011.
- Maritorena, S., Siegel, D. A., Peterson, A. R.: Optimization of a semi- analytical ocean color model for global-scale applications, *Appl. Opt.*, 41, 2705–2714, <https://doi.org/10.1364/AO.41.002705>, 2002.
- May, C. L., Koseff, J. R., Lucas, L., Cloern, J. E., and Schoellhamer, D.H.: Effects of spatial and temporal variability of turbidity on phytoplankton blooms, *Mar. Ecol. Prog. Ser.*, 254, 111–128, <https://doi.org/10.3354/meps254111>, 2003.
- McGillicuddy, D. J., Anderson, D. M., Lynch, D. R., and Townsend, D. W.: Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: Results from a physical-biological model, *Deep Sea Res. Part II Top. Stud. Oceanogr.*, 52, 2698–2714, <https://doi.org/10.1016/j.dsr2.2005.06.021>, 2005.
- Morel, A., and Belanger, S.: Improved detection of turbid waters from ocean color sensors information, *Remote Sens. Environ.*, 102, 237–249, <https://doi.org/10.1016/j.rse.2006.01.022>, 2006.
- Morel, A., and Gentili, B.: A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data, *Remote Sens. Environ.*, 113, 998–1011, <https://doi.org/10.1016/j.rse.2009.01.008>, 2009.
- Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., and Franz, B. A.: Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach, *Remote Sens.*

- Environ., 111, 69–88, <https://doi.org/10.1016/j.rse.2007.03.012>, 2007.
- Neville, R. A., and Gower, J. F. R.: Passive remote sensing of phytoplankton via chlorophyll fluorescence, J. Geophys. Res., 82, 3487–3493, <https://doi.org/10.1029/JC082i024p03487>, 1977.
- O’Brien, R. M. A.: Caution regarding rules of thumb for Variance Inflation Factors. Qual. Quant., 41, 673–690, <https://doi.org/10.1007/s11135-006-9018-6>, 2007.
- O’Reilly, J. E., Maritorena, S., Siegel, D., O’Brien, M. C., Toole, D., Mitchell, B. G., Kahru, M., Chavez, F. P., Strutton, P., Cota, G., et al.: Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4, In SeaWiFS Postlaunch Technical Report Series, SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3, Hooker, S.B., Firestone, E.R., Eds., NASA Goddard Space Flight Center: Greenbelt, MD, USA, Volume 11, pp. 9–23, 2000.
- Pan, D. L., Gower, J. F. R. and Lin, S. R.: A study of band selection for fluorescence remote sensing of ocean chlorophyll-a, Oceanol. Limnol. Sin., 20, 564–570, 1989.
- Pinckney, J. L., Paerl, H. W., Tester, P., and Richardson, T. L.: The role of nutrient loading and eutrophication in estuarine ecology, Environ. Health Perspect., 109, 699–706, 2001.
- Raine, R., Boyle, O., O’Higgins, T., White, M., Patching, J., Cahill, B., and McMahon, T.: A satellite and field portrait of a *Karenia mikimotoi* bloom off the south coast of Ireland, August 1998. Hydrobiologia, 465, 187–193, <https://doi.org/10.1023/A:1014524420705>, 2001.
- Raine, R., McDermott, G., Silke, J., Lyons, K., Nolan, G., and Cusack, C.: A simple short range model for the prediction of harmful algal events in the bays of

- southwestern Ireland, *J. Mar. Syst.*, 83, 150–157,
<https://doi.org/10.1016/j.jmarsys.2010.05.001>, 2010.
- Ryther, J. H.: Photosynthesis in the Ocean as a Function of Light Intensity, *Limnol. Oceanogr.*, 1, 61–70, <https://doi.org/10.4319/lo.1956.1.1.0061>, 1956.
- Sarangi, R. K., and Mohammed, G.: Seasonal algal bloom and water quality around the coastal Kerala during southwest monsoon using in situ and satellite data, *Indian J. Geo-Mar. Sci.*, 40, 356–369, 2011.
- Sarma, Y. V. B., Al-Hashmi, K., and Smith, S. L.: Sea surface warming and its implications for harmful algal blooms off Oman, *J. Mar. Sci.*, 3, 65–71,
<http://dx.doi.org/10.5376/ijms.2013.03.0008>, 2013.
- Seegers, B. N., Birch, J. M., Marin, R., Scholin, C.A., Caron, D. A., Seubert, E. L., Howard, M. D., Robertson, G. L., and Jones, B. H.: Subsurface seeding of surface harmful algal blooms observed through the integration of autonomous gliders, moored environmental sample processors, and satellite remote sensing in southern California, *Limnol. Oceanogr.*, 60, 754–764, <https://doi.org/10.1002/lno.10082>, 2015.
- Shang, S. L., Dong, Q., Hu, C. M., Lin, G., Li, Y. H., and Shang, S. P.: On the consistency of MODIS chlorophyll a products in the northern South China Sea, *Biogeosciences*, 11, 269–280, <https://doi.org/10.5194/bg-11-269-2014>, 2014.
- Shen, L., Xu, H., and Guo, X.: Satellite remote sensing of harmful algal blooms (HABs) and a potential synthesized framework, *Sensors*, 12, 7778–7803,
<https://doi.org/10.3390/s120607778>, 2012.
- Steidinger, K. A., Vargo, G. A., Tester, P. A., and Tomas, C. R.: Bloom dynamics and

- physiology of *Gymnodinium breve* with emphasis on the Gulf of Mexico. In
Physiological Ecology of Harmful Algal Blooms, 1st ed., Anderson, D.M., Cambella,
A.D., Hallegraeff, G.M., Eds., Springer-Verlag: Heidelberg, Germany, pp. 133–154,
ISBN 3540641173, 1998.
- Stumpf, R. P., Culver, M. E., Tester, P. A., Tomlinson, M., Kirkpatrick, G. J., Pederson, B. A.,
Truby, E., Ransibrahmanakul, V., and Soracco, M.: Monitoring *Karenia brevis*
blooms in the Gulf of Mexico using satellite ocean color imagery and other data,
Harmful Algae, 2, 147–160, [https://doi.org/10.1016/S1568-9883\(02\)00083-5](https://doi.org/10.1016/S1568-9883(02)00083-5),
2003.
- Stumpf, R. P., Litaker, R. W., Lanerolle, L., and Tester, P. A.: Hydrodynamic accumulation
of *Karenia* off the West Coast of Florida, Cont. Shelf Res., 28, 189–213,
<https://doi.org/10.1016/j.csr.2007.04.017>, 2008.
- Stumpf, R. P., and Tomlinson, M. C.: Use of remote sensing in monitoring and forecasting
of harmful algal blooms, Proc. SPIE, 5885, 148–151,
<https://doi.org/10.1117/12.614376>, 2005.
- Tang, D. L., Di, B. P., Wei, G., Ni, I. H., Oh, I. S., and Wang, S. F.: Spatial, seasonal and
species variations of harmful algal blooms in the South Yellow Sea and East China
Sea, Hydrobiologia, 568, 245–253, <https://doi.org/10.1007/s10750-006-0108-1>,
2006.
- Tang, D. L., Kawamura, H., Doan-Nhu, H., and Takahashi, W.: Remote sensing
oceanography of a harmful algal bloom (HAB) off the coast of southeastern
Vietnam, J. Geophys. Res., 109, 1–7, <https://doi.org/10.1029/2003JC002045>,
2004.

- Tang, D. L., Kawamura, H., and Luis, A. J.: Short-term variability of phytoplankton blooms associated with a cold eddy on the North-western Arabian Sea, *Remote Sens. Environ.*, 81, 82–89, [https://doi.org/10.1016/S0034-4257\(01\)00334-0](https://doi.org/10.1016/S0034-4257(01)00334-0), 2002.
- Tang, D.L., Kawamura, H., Oh, I. S., and Baker, J.: Satellite evidence of harmful algal blooms and related oceanographic features in the Bohai Sea during autumn 1998. *Adv. Space Res.*, 37, 681–689, <https://doi.org/10.1016/j.asr.2005.04.045>, 2006.
- Tang, D. L., Ni, I. H., Kester, D. R., and Müller-Karger, F. E.: Remote sensing observation of winter phytoplankton blooms southwest of the Luzon Strait in the South China Sea, *Mar. Ecol. Prog. Ser.*, 191, 43–51, <https://doi.org/10.3354/meps191043>, 1999.
- Thyng, K. M., Hetland, R. D., Ogle, M. T., Zhang, X., Chen, F., and Campbell, L.: Origins of *Karenia brevis* harmful algal blooms along the Texas coast, *Limnol. Oceanogr. Fluids Environ.*, 3, 269–278, <https://doi.org/10.1215/21573689-2417719>, 2013.
- Tian, R., Chen, J., Sun, X., Li, D., Liu, C., and Weng, H.: Algae explosive growth mechanism enabling weather-like forecast of harmful algal blooms, *Sci. Rep.*, 8, 9923, <https://doi.org/10.1038/s41598-018-28104-7>, 2018.
- Tilstone, G. H., Lotliker, A. A., Miller, P. I., Ashraf, P. M., Kumar, T. S., Suresh, T., Ragavan, B., and Menon, H. B.: Assessment of MODIS-Aqua chlorophyll-a algorithms in coastal and shelf waters of the eastern Arabian Sea, *Cont. Shelf Res.*, 65, 14–26, <https://doi.org/10.1016/j.csr.2013.06.003>, 2013.
- Tomlinson, M. C., Wynne, T. T., and Stumpf, R. P.: An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia Brevis*, *Remote Sens. Environ.*, 113, 598–609,

- <https://doi.org/10.1016/j.rse.2008.11.003>, 2009.
- Turrell, E., Stobo, L., Lacaze, J. P., Bresnan, E., and Gowland, D.: Development of an ‘early warning system’ for harmful algal blooms using solid-phase adsorption toxin tracking (SPATT), In Proceedings of the OCEANS 2007—Europe, Aberdeen, UK, 18–21 June 2007, pp. 1–6, <https://doi.org/10.1109/OCEANSE.2007.4302436>, 2007.
- Tyler, J. E.: The Secchi disk, *Limnol. Oceanogr*, 13, 1–6, <https://doi.org/10.4319/lo.1968.13.1.0001>, 1968.
- Vargo, G.A.: A brief summary of the physiology and ecology of *Karenia brevis* Davis (G. Hansen and Moestrup comb. nov.) red tides on the West Florida Shelf and of hypotheses posed for their initiation, growth, maintenance, and termination, *Harmful Algae*, 8, 573–584, <https://doi.org/10.1016/j.hal.2008.11.002>, 2009.
- Vargo, G. A., Carder, K. L., Gregg, W., Shanley, E., Hell, C., Steidinger, K. A., and Haddad, K.D.: The potential contribution of primary production by red tides to the west Florida shelf ecosystem, *Limnol. Oceanogr.*, 32, 762–767, <https://doi.org/10.4319/lo.1987.32.3.0762>, 1987.
- Wang, G., Lee, Z., and Mouw, C.: Multi-spectral remote sensing of phytoplankton pigment absorption properties in cyanobacteria bloom waters: A regional example in the western basin of Lake Erie, *Remote Sens.*, 9, 1309, <https://doi.org/10.3390/rs9121309>, 2017.
- Wei, G., Tang, D. L., and Wang, S.: Distribution of chlorophyll and harmful algal blooms (HABs): A review on space based studies in the coastal environments of Chinese marginal seas, *Adv Space Res.*, 41, 12–19,

<https://doi.org/10.1016/j.asr.2007.01.037>, 2008.

Xing, X., Zhao, D., Liu, Y., Yang, J., Xiu, P., and Wang, L.: An overview of remote sensing of chlorophyll fluorescence, *Ocean Sci. J.*, 42, 49–59,

<https://doi.org/10.1007/BF03020910>, 2007.

Zhao, D. Z., Xing, X. G., Liu, Y. G., Yang, J. H., and Wang, L.: The relationship of chlorophyll-a concentration with the reflectance peak near 700 nm in algae-dominated waters and sensitivity of fluorescence algorithms for detecting algal bloom, *Int. J. Remote Sens.*, 31, 39–48, <https://doi.org/10.1080/01431160902882512>, 2010.

Zhao, J., Hu, C., Lenes, J. M., Weisberg, R. H., Lembke, C., English, D., Wolny, J., Zheng, L., Walsh, J. J., and Kirkpatrick, G.: Three-dimensional structure of a *Karenia brevis* bloom: Observations from gliders, satellites, and field measurements, *Harmful Algae*, 29, 22–30, <https://doi.org/10.1016/j.hal.2013.07.004>, 2013.

CHAPTER 4

CONCLUSION

In this study, I demonstrated the potential of using readily available remote sensing data and limited field data to develop semi-automated predictive systems geohazards (landslides and algal blooms). The accuracy of both systems was tested and was proven to be satisfactory although the potential for improvement still exist with the inclusion of additional relevant variables and/or extending the temporal window of the applied data sets (variables). I developed a prototype version of an early warning system for landslides over the Faifa Mountains using a wide range of remote sensing products. Likewise, I developed an operational early warning system for harmful algal blooms in the coastal waters of Charlotte county in Florida. In both studies, I successfully demonstrated that remote sensing data can be leveraged to generate early warning systems, even if only limited field data is available. Having said that, the availability of field data was an integral part of the research as it was used to validate our findings. In both the studies, it became evident that with the inclusion of additional relevant variables and expanding the temporal window of the variables the generated model can become more robust.

Currently the early warning system for landslides is only a prototype version and once the real-time rainfall data becomes available, the system can be implemented. With the inclusion of more data, the current version of ID curve will get more refined. Once this happens, the predictability of the ID curve is expected to increase with the reduction of false positive incidents. In the case of harmful algal blooms, once the field data for nutrients becomes available, the model can be retrained to include those

variables. With the inclusion of nutrients and other potential variables such as precipitation and stream flow, the predictability of the model is expected to improve.

In conclusion, this research demonstrates the utility of remote sensing resources in in developing predictive models and for compensating for the lack of extensive field data that has been previously required for developing such models. In the end, the importance of field data cannot be ignored as it is essential for the validation of model outputs. Both of these early warning systems are replicable and can be easily implemented in the areas witnessing similar geohazards elsewhere.