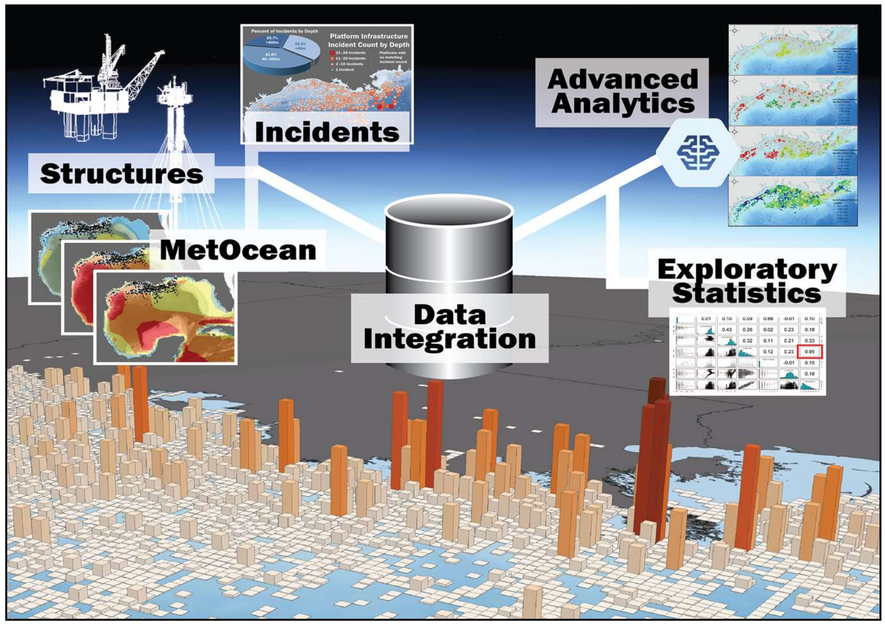


Evaluating Offshore Infrastructure Integrity

29 April 2021

Office of Fossil Energy

DOE/NETL-2021/2643

Disclaimer

Cover Illustration: The study took an integrated approach to modeling infrastructure integrity. The cover illustration provides a graphical overview of the process taken to combine data, exploratory and predictive analysis to estimate offshore shore integrity.

Suggested Citation: Nelson, J.; Dyer, A.; Romeo, L.; Wenzlick, M.; Zaengle, D.; Duran, R.; Sabbatino, M.; Wingo, P.; Barkhurst, A.; Rose, K.; Bauer, J. *Evaluating Offshore Infrastructure Integrity*; DOE/NETL-2021/2643; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Albany, OR, 2021; p 70. DOI: 10.2172/1780656.

An electronic version of this report can be found at:

https://edx.netl.doe.gov/offshore

Evaluating Offshore Infrastructure Integrity

Jake Nelson[1,3], Alec Dyer[1,4], Lucy Romeo[1,5], Madison Wenzlick[1,4], Dakota Zaengle[1,4], Rodrigo Duran[1,6], Michael Sabbatino[1,4], Patrick Wingo[1,4], Aaron Barkhurst[7], Kelly Rose[1], Jennifer Bauer[1]

[1] U.S. Department of Energy, National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321

[3] Oak Ridge Institute for Science and Education, 1450 Queen Avenue SW, Albany, OR 97321

[4] NETL Support Contractor, 1450 Queen Avenue SW, Albany, OR 97321

[5] NETL Support Contractor, 3610 Collins Ferry Road, Morgantown, WV 26507, USA

[6] Theiss Research, 7411 Eads Avenue, La Jolla, CA 92037

[7] Matric, 430 Drummond St. #2, Morgantown, WV 26505

DOE/NETL-2021/2643

29 April 2021

NETL Contacts:

Jennifer Bauer, Principal Investigator

Lucy Romeo, Co-Principal Investigator

Kelly Rose, Technical Portfolio Lead

Bryan Morreale, Executive Director, Research & Innovation Center

This page intentionally left blank.

Table of Contents

List of Figures

List of Tables

Acronyms, Abbreviations, and Symbols

Term	Description
API	American Petroleum Institute
ANN	Artificial neural network
BOEM	Bureau of Energy Management
BP	British Petroleum
BOEM	Bureau of Ocean Energy Management
BSEE	Bureau of Safety and Environmental Enforcement
CAIS	Caisson
CP	Cathodic protection
CSIL	Cumulative Spatial Impact Layer™
CT	Complaint tower
DHSG	Deepwater Horizon Study Group
DOE	Department of Energy
DWH	Deepwater Horizon
EDX	Energy Data eXchange®
EIA	Energy Information Administration
EOR	Enhanced oil recovery
ESDA	Exploratory spatial data analysis
FPSO	Floating Production, Storage, and Offloading Systems
GBC	Gradient boosting classifier
GoM	Gulf of Mexico
GWR	Geographically-weighted regression
ISO	International Standards Organization
KPI	Key Performance Indicators
LSTM	Long short-term memory (networks)
MetOcean	Meteorological and Oceanographic
ML	Machine learning
MMS	Minerals Management Service
MODU	Mobile Offshore Drilling Unit
MOPU	Mobile Offshore Production Unit
MTLP	Mini Tension-leg Platform
NETL	National Energy Technology Laboratory

Acronyms, Abbreviations, Symbols (cont.)

Term	Description
NN	Neural network
OCR	Optical character recognition
ONRR	Office of Natural Resource Revenue
ORM	Offshore Risk Modeling suite
pH	Potential of hydrogen
QA/QC	Quality assurance quality control
SEMI	Semi-submersible
SIM	Structural Integrity Management
SSTMP	Subsea Template Manifold
SWOT	Strengths, Weakness, Opportunities, and Threats
TLP	Tension-leg platform
U.S.	United States of America
USCG	U.S. Coast Guard
VIF	Variance inflation factor
WP	Well protector

Acknowledgments

This work was completed as part of National Energy Technology Laboratory (NETL) research for the U.S. Department of Energy's (DOE) Complementary Research Program under Section 999 of the Energy Policy Act of 2005. Parts of this technical effort were performed in support of the National Energy Technology Laboratory's ongoing research under the Offshore Unconventional Resources – DE FE-1022409 by NETL's Research and Innovation Center, including work performed by Leidos Research Support Team staff under the (RSS contract 89243318CFE000003). The authors wish to acknowledge Roy Long (NETL Strategic Center for Natural Gas and Oil) and Elena Melchert (DOE Office of Fossil Energy) for programmatic guidance, direction, and support.

This page intentionally left blank.

ABSTRACT

Drilling in the offshore environment involves a complex network of infrastructure including pipelines, platforms, rigs, subsea installations, ports, and terminals. Government and industry partners have developed this network over many decades and it remains a critical part of the United States (U.S.) energy portfolio. Many of the major components of this system have been designed with a 20- to 30-year lifespan, yet consistent and growing energy demands support the need to extend the design life of existing infrastructure or repurpose it for secondary needs (i.e. enhanced oil recovery, carbon storage, and new wells). As a result, a growing portion of the offshore infrastructure in the U.S. is approaching or has exceeded its original design life. A critical step in ensuring the continued safe and effective operation of offshore infrastructure is developing a comprehensive understanding of the state of offshore infrastructure and the factors that effect it.

The purpose of this project is to assess the current state of existing infrastructure and identify the factors involved in infrastructure degradation through the development and application of big data analytics, machine learning, and advanced spatio-temporal analysis. The project leverages existing data at NETL and combines it with new information on offshore oil and gas structures and the ambient offshore environment in an effort to identify patterns associated with infrastructure integrity. Building on the identified trends and patterns, this project incorporates exploratory analytics and spatial analysis tools in conjunction with machine learning and statistical models to characterize the condition of existing platforms in the offshore environment and predict their risk of failure.

1. INTRODUCTION

Hydrocarbon exploration and production in the offshore environment remains an important part of the United States (U.S.) energy portfolio. Offshore production activities require a dense network of complex infrastructure that includes platforms, pipelines, rigs, support vessels, and myriad other components that work in tandem to facilitate the exploration, extraction, and transportation of hydrocarbons from underwater reservoirs. While some of these components receive routine maintenance and replacement, inventory surveys suggest that the majority of active infrastructure in the more developed production regions were installed more than two decades ago (Stacey et al., 2008), far exceeding the intended design life. Steps can be taken to manage this aging infrastructure which is critical as both government and industry experts are increasingly looking towards methods and technologies that can extend the infrastructures design life. These efforts are aimed at minimizing cost while maximizing the production potential of existing reservoirs and drilling technology (i.e. horizontal drilling). In addition, existing infrastructure can also be repurposed for alternative uses including enhanced oil recovery (EOR) that add additional value to the existing structures and reservoirs. That said, any efforts made towards extending infrastructure design life must be matched with approaches and innovations that effectively reduce the possibility of deleterious events. This cross-fertilization ensures that continued operation is done with an eye towards safe production that maximizes efficiency.

With increased access to data and information on the offshore production environment, along with information directly related to operational efficiency, there are increased opportunities to build a comprehensive understanding of the current state of infrastructure (i.e. platforms). In this work, “current state” refers to the current operational capacities of the infrastructure given the age of the infrastructure and exposure to ambient environmental conditions. Given that a comprehensive dataset on the current state of infrastructure is not available, a significant data collection effort is required along with the development and application of spatio-temporal methods and models. Specifically, a data inventory of the age of the infrastructure, the intended design life and the location of each platform is critical for understanding how variations in the operating environment and structural characteristics influence the lifespan of the infrastructure.

This is to say, as offshore infrastructure ages the integrity of the infrastructure is called into question. The infrastructure is subject to various stressors overtime which can decrease the longevity of the equipment. When referring to integrity it is important to be specific as the concept is multifaceted. To the extent that it is applied in this research, integrity refers to the degree in which a system is impaired. An infrastructure system of high integrity is one where the systems of components that make up the macrostructures of the offshore network are operating in a way that do not impede the performance of the system as a whole. Conversely, a low integrity system is one where changes in the materials within the macrostructures impede the system from performing at its optimum.

Impediments to offshore integrity can come from many sources. These include corrosion, structural stress from the environment, age, or simply fatigue from consistent use of the equipment. Although all of these sources can decrease integrity, some may contribute more prominently to system degradation than others. In an effort to prevent catastrophic failure, this work aims at identifying the factors related to decreases in infrastructure integrity and evaluate their overall impact on the infrastructure’s lifespan. By identifying these factors, the results of this and subsequent work can be used to alert decision makers, planners, and other stakeholders to locations where the potential for failure is higher relative to other locations in the offshore environment.

There are two objectives for this study which were carried out in two phases. First, this study sought to develop an integrity assessment framework and apply it to offshore infrastructure in the Gulf of Mexico (GoM). Second, this study worked to fill knowledge- and technology-gaps surrounding the current state of offshore infrastructure through analyses that identify the antecedents of possible infrastructure failure based on the results of the integrity evaluation.

Where integrity is concerned, the first phase of the analysis developed a conceptual framework for measuring infrastructure integrity. Data was collected to match the framework while inferential statistics were used to operationalize the framework. Throughout this first phase, the framework was continually updated to reflect the availability and type of data. The second phase consisted of the implementation and testing of the conceptual framework using the newly amassed data, exploratory and predictive analysis methods. Specifically, the antecedents to changes in integrity and lifespan were evaluated using geostatistical and machine learning (ML) techniques.

With the ability to relate these findings back to a specific location (i.e. structure or lease block), this project serves as an important step towards enhanced risk mitigation for offshore oil production and exploration. Furthermore, the development of datasets and analysis to achieve the project objectives will aid in building a more robust understanding of the current state of offshore infrastructure. This will ultimately increase the resilience and efficiency of offshore hydrocarbon production systems. Although the study focuses specifically on the infrastructure in the GoM, the conceptual framework and analysis are generalizable to diverse settings.

1.1 AGING ASSETS: OFFSHORE OIL AND GAS INFRASTRUCTURE

"A*ging is not about how old your equipment is; it is about what you know about its condition, and how that is changing over time" (Nabavian and Morshed, 2010)*.

According to the International Standards Organization (ISO), design life is defined as "the assumed period for which a structure is to be used for its intended purpose with anticipated maintenance but without substantial repair from aging processes being necessary" (ISO, 2015). As far as the offshore oil and gas infrastructure is concerned, there is a (general) expected design life of 20 to 30 years, but this number greatly depends on the operating environment, the frequency of dry-dock repair, and maintenance of drilling equipment. A report in 2010 estimated that out of the roughly 6,500 platforms in operation worldwide, 30% of them have been in operation for over 20 years, far exceeding their original design life (Nabavian and Morshed, 2010). This finding is corroborated in a second report citing data on infrastructure in the North Sea, which suggests that a significant number of offshore platforms have exceeded their original design life of about 25 years (Stacey et al., 2008). Across many regions in the world, the general trend of offshore infrastructure exceeding its design life is becoming strikingly apparent (Ersdal and Selnes, 2010; Solland et al., 2011). Interestingly, no explicit ties between infrastructure age and failure have been made. However, age may serve as a proxy for several other mechanisms that negatively affect integrity over time.

For example, a 2016 report on offshore oil spill occurrence classifies infrastructure age within equipment failure and corrosion as causes of spills. The report also points to increases in incidents resulting from weather or natural causes and external factors which increase over time

(BOEM, 2016). Relatedly, a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis conducted by the National Oil Spill Detection Response Agency in Nigeria identified over-aged pipelines and production infrastructure as possible causes of spill incidence (Rim-Rukeh, 2015). Moreover, Burgherr (2007) found that tankers less than 10 years of age were related to fewer spills than tankers that were over 20 years of age. Although this is not directly tied to installed infrastructure, it does lend some support to a potential correlation between age and infrastructure failure.

In response to these findings, researchers and practitioners have turned their attention toward several related areas of research. One is aimed at life cycle management and maintenance strategies to increase operating capabilities while decreasing fatigue and undue stress on the equipment (Moan, 2018; Soom et al., 2018). The Structural Integrity Management (SIM) process focuses on a lifecycle process that ensures the fitness of fixed offshore platforms (O'Connor et al., 2005) which has evolved significantly over the last 25 years and has been applied to platform management strategies worldwide (Guédé, 2019; Nichols and Khan, 2015). Other life-cycle management strategies include the work by Aeran et al. (2017), which focused on corrosion models and the work by Tygesen et al. (2019), who used Bayesian ML to update offshore maintenance schedules.

1.2 FACTORS INVOLVED IN INFRASTRUCTURE DEGRADATION

An essential component for characterizing the current state of offshore infrastructure is the evaluation of factors known or thought to be involved in the decline of infrastructure integrity. This includes meteorological and oceanographic (MetOcean) conditions, structural materials, and incidents (loss of well control, collisions, fires, and equipment failures), among others. Taken together, these factors can influence the expected design life and lead to catastrophic failures if not mitigated. The following details the most influential factors that have been found to contribute to the structural integrity of offshore infrastructure.

1.2.1 Ocean and Atmospheric Conditions

Knowledge of MetOcean conditions are critical to industries operating in the offshore environment. Ambient conditions such as wave height, ocean-current speed, and wind speed—known as ambient loadings—are major considerations when determining criteria for offshore platform design and for safe operating conditions. An under-designed offshore structure may result in loss of life, injury, and considerable structural damage. Other MetOcean variables, such as those related to corrosion rates, may have less of an immediate effect on offshore structures, yet may also cause catastrophic failure over longer periods of time.

The effects from wave, current, and wind actions are known hazards to offshore infrastructure (ISO, 2011). The environmental loadings caused by environmental conditions can cause immediate damage during storms (including complete destruction), or progressively fatigue a structure, leading to increased risk of failure and incidents (Moan, 2018; Sharp et al., 2015). For example, in extreme wave conditions water may cause additional damage or fatigue to a structure through slamming and in-deck loading (Moan, 2018). For offshore jacket structures, the structural loading effects from waves are increased if wave height reaches deck level (Guédé, 2019) and waves from storms and ocean currents, as well as the GoM Loop Current and associated eddies are important considerations for the design of tension leg platforms (BSEE, 2018). Although waves pose the most immediate effect, high wind speeds can also increase fatigue, albeit gradually, to structural components (Sharp et al., 2015).

The effects of the environment are well known, and as a result an abundance of standards and regulations are in place to ensure the robustness of offshore infrastructure (ISO, 2011). However, establishing these regulations has been a process of trial and error. Past standards and regulations have required revisions when extreme MetOcean conditions have caused catastrophic damage (e.g. Berek et al., 2007). In 2005 alone, hurricanes Katrina and Rita destroyed 115 platforms and damaged 52 others (Zhang, 2017). Indeed, many offshore platform incidents have occurred in regions that are prone to the most extreme MetOcean events (see Figures 1). An additional concern is the less-understood cumulative effect of these loadings on offshore structures that accumulate over years and decades. Again, this underpins the increasing importance of understanding platform age and the variables effecting design life.

There exists an extensive amount of research surrounding the effects of MetOcean conditions and corrosion rates. Mechanical properties such as strength, ductility, and impact strength will degrade gradually which can lead to failure under certain environmental conditions (Bhandari et al., 2015). While the material used for offshore structures vary, in practice most structures are made of steel and steel-reinforced concrete, making steel the main focus of corrosion studies (Melchers, 2016). Some of the ambient conditions that are known to affect corrosion by seawater include: salinity, dissolved oxygen concentration, temperature, potential of hydrogen (pH), carbonate solubility, pollutants and biological growth, bacteria, pressure, wave action, and water velocity (Matsushima, 2011; Melchers, 2016). Others have identified the influence of water temperature on steel corrosion, noting an increased corrosion rate with higher water temperatures (Bhandari et al., 2015; Nunez, 2007). These findings support the notion that structures in warm, coastal oxygenated waters will have higher corrosion rates than deep water structures (Guedes Soares et al., 2011).

While the effects of water velocity are not fully understood, several researchers have identified a positive correlation between water velocity and the rate of pitting corrosion in marine environments (Bhandari et al., 2015; Guedes Soares et al., 2011; Melchers, 2005). Importantly, pitting corrosion is the type of corrosion more likely to cause catastrophic failure; however, a majority of rigs have a cathodic protection (CP) system and other forms of protection to defend against this type of corrosion (Sharp et al., 2015). That said, corrosion remains a prominent threat to the integrity of offshore systems. It is also worth noting that a large portion of the corrosion studies that establish quantitative relations between ambient conditions and corrosion are done under laboratory conditions; the results do not immediately translate to in-situ ambient conditions supporting the need to explore the relationship between corrosive ocean conditions, offshore integrity, and platform lifespan.

1.2.2 Incidents

Incidents are defined as situations in which personnel working on the platform are injured or a part of the platform itself has been damaged. Over the past 15 years there has been an increase in available incident data following changes in reporting regulations that mandated the reporting of both serious and potentially serious incidents by the Minerals Management Service (MMS, 2006). Muehlenbachs et al. (2013) indicate that incidents fall into several types of categories related to causes, effects, and resulting damages (Table 1). The type of incident can fall into one or more of these categories, and the causes and damages as well may include one or many of these descriptors. In other words, incidents can occur due to day-to-day activities, for example during drilling and workover events, or from irregular events such as lightning strikes and ship collisions. The effects of such incidents, and even the incidents themselves (fires, explosions, collisions, and dropped objects on offshore structures) are an important feature of structural management systems and may even suggest latent measures of overall integrity (Moan, 2018; Sharp et al., 2015).

Table 1: Breakdown of the Possible Types, Causes, and Damages from Incidents

Types of Incidents	Causes of Incidents	Resulting Damages
Blowout	Completion equipment	Cranes
Vessel collision	Equipment failure	Structural damage
Fire	Development or production operations	Overboard drilling fluid
Explosion	Human error	
Injury/Fatality	Slip, trip or fall	
Pollution	Weather	

Reported incidents may also include a loss of well control or a well kick that may result in the release of hydrocarbons to the surrounding environment (a blowout). The most common activity associated with a blowout is drilling, especially during the exploration phase (Kaiser and Pulsipher, 2007). When blowouts occur, oil and gas emanating from the well or riser pipes can cause fires or explosions, leading to potentially serious structural damages depending on the intensity (ISO, 2011; Sharp et al., 2015). An example of a historical blowout is Deepwater Horizon (DWH) (Graham et al., 2012).

1.2.3 Physical Conditions

The age of a platform can be used as a proxy to understand the physical condition of a structure. As one would expect, the effects of fatigue and corrosion become greater over time, leading to higher instances of degradation unless these effects are actively managed (Muehlenbachs et al., 2013; Stacey et al., 2008). With design codes becoming stricter over the years, installation year has been found to be a strong predictor of structural integrity (Moan, 2018; Stacey et al., 2008). The older the installation date, the less integrity. As more information has been gathered, newer design codes are adapted to consider the increased intensity of weather events, including higher wave heights and stronger wind speeds. This has materialized in taller deck heights, reinforced welding requirements, and redundant safety systems. Guédé (2019) has identified several additional physical conditions of fixed offshore platforms that affect the susceptibility of failure including structural configuration, and foundation system, and also confirmed the year of design as an indicator of integrity. Although these were specific to fixed platforms, some may also apply to floating structures. In particular, the existence of damaged components, corroded components, corrosion protection system, marine growth, and physical or environmental loadings may all be indicative of floating platform integrity (Guédé, 2019; Sharp et al., 2015).

Regarding the location of a structure, Muehlenbachs et al. (2013) found a relationship between the operational water depth of a structure and the number of self-reported incidents between 1996 and 2010. These results contradicted the findings of Jablonowski (2007) who found that water depths deeper than 121.9 m (400 ft) are insignificant in predicting the likelihood of an incident on drilling rigs from 1990 to 1998. Still, Shultz and Fischbeck (1999) found a negative relationship between water depth and accidents between 1986 and 1995, confirming the findings in Muehlenbachs et al. (2013). Deeper water depths have also been found to have long term corrosion affects but are subject to other parameters including oxygen concentration, temperature, metal type, pollution, salinity, and water velocity, to name a few (Bhandari et al., 2015). These findings support the investigation of both platform age and depth in the analysis of integrity.

1.2.4 Organization Factors

Organizational responsibility and accountability in managing offshore structures plays a role in the safe operation of each structure. Consider the 2010 DWH incident; post spill, a myriad of factors involved in the demise of the offshore platform came to light. Apart from the faulty blowout preventer, a rushed and poorly sealed cement job, and the failure to cap and contain the released hydrocarbon, several reports pointed to "a multi-decade history of organizational malfunctions and shortsightedness" (Deepwater Horizon Study Group (DHSG), 2011). Across the oil and gas industry writ large, a culture of "trip-and-fall" reactionary compliance had developed and it was not until something occurred that companies sought to achieve regulatory compliance. Concerning the DWH disaster, British Petroleum (BP) had pursued similar decision-making strategies, opting toward saving time and money, rather than heeding warning signs of decreasing integrity. The DHSG ultimately identified the failure of a system for checks and balances between regulators and industry as responsible for the disaster. When it came to identify and taking action to ensure the safe operation of offshore oil and gas exploration and production, regulatory oversight was simply not there (DHSG, 2011).

The cascade of poor decision-making related to the DWH operations was not unique to BP or the DWH case. For years, industry and federal regulators had been working in tandem to create a risk prone, rather than averse, organizational environment. This has been pointed out by many public administration, organizational behavior, and risk analysis scholars (Bozeman, 2011; Reader and O'Connor, 2014). Since the DWH additional regulatory checks and balances have been established to improve safety. For example, the MMS, who at the time of the DWH were the sole governing body for offshore oil and gas operations, was broken up into three distinct departments – the Bureau of Ocean Energy Management (BOEM), the Bureau of Safety and Environmental Enforcement (BSEE), and the Office of Natural Resources Revenue (ONRR). Each of these entities was given different mandates surrounding the oversight of offshore oil and gas activity. With this came an increased semblance of checks and balances which, going forward, should help regain a more conscious and risk averse approach to oil and gas development. Still, incidents and spills remain a part of offshore oil and gas operations. Whether it is due to operating in extreme conditions, or due to remnants of a risk-prone safety culture, it is important to consider organizational factors as an influential part of infrastructure integrity.

1.3 ESTIMATING INFRASTRUCTURE INTEGRITY: EXISTING MODELS AND METHODS

In addition to the codes and regulations set by industry (American Petroleum Institute (API)) and federal agencies (BOEM, BSEE) to ensure the safe operation of offshore oil and gas activity (API, 2014; Visser, 2011), researchers are using novel approaches and technological advancements to develop strategies aimed at incident prevention. Many of these methods are shared across stakeholder groups that focus on structural management plans, design specifications, operational data, and routine inspections.

In the U.S., there are various classification organizations (i.e. API and ISO) that develop and set safety and operational codes for structures in offshore environments. These standards must be met at various stages of a structure's life including "design criteria formulated in terms of serviceability and safety limit states, considering payloads, environmental, and accidental loads" and "life cycle feature, with strong links between design, and inspection, monitoring, maintenance, and repair" (Moan, 2018). Stacey et al. (2008) have categorized codes into several areas aimed at extending the life of offshore infrastructure. These include assessment issues, fatigue life extension, corrosion protection, and inspection, maintenance, and survey. The standards create structural redundancies and optimized inspection plans that are necessary to reduce the operational risks posed by operating in harsh offshore environments. While many of these standards provide a model for a risk-based approach to structural management, in some cases "no detail is given on how to implement those methods" (Guédé, 2019).

Contrasting industry standards, researchers are incorporating performance indicators, qualitative and quantitative assessments, and state-of-the-art modeling methods to create management programs, risk rankings, and routine inspection schedules and techniques (Guédé, 2019; Jablonowski, 2007; Sharp et al., 2015; Tygesen et a., 2019; Yang and Frangopol, 2018). For example, Guédé (2019) utilizes API guidelines to develop analysis methods for global and local risk assessments for fixed offshore structures. These methods use design specifications, present conditions, modifications, and loading exposure to assess risk. Independent variables include design year, last inspection, damaged members, topside weight change, and wave-in-deck exposure. Additionally, Sharp et al. (2015) developed Key Performance Indicators (KPIs) for both fixed and mobile units that identify hazards and chain of events to incidents using a hazard analysis method. The approach leverages design specifications, inspection, and MetOcean data.

The availability of computationally diverse ML algorithms offers new capabilities for the petroleum industry to predict the lifespan and structural integrity of offshore oil and gas infrastructure. There is a wide selection of relevant ML algorithms including random forests (Ho, 1995), support vector machines (Cortes and Vapnik, 1995), artificial neural networks (Jain et al., 1996), and decision trees (Quinlan, 1986). Specific to offshore infrastructure, ML models have helped in predicting stress factors from various environmental forces. These include Gaussian process models to improve predictions for the effects of fluid forces from waves and currents on structural integrity (Worden et al., 2017) and applying artificial neural networks (ANN) to characterize complicated physical subsurface stresses from drilling activities and their affiliated effects on infrastructure (Elkatatny et al., 2017; Onalo et al., 2018).

Recently, in conjunction with the concerns of aging infrastructure, these models have been applied to understanding the lifespan of offshore infrastructure. Almedallah and Walsh, (2019a) evaluated asset lifespan by combining data science and econometrics to determine the feasibility of maintaining an old system or installing a new one. Furthermore, studies have combined the different factors affecting drilling and production into their computational models. Almedallah and Walsh, (2019b) highlight the importance of incorporating well and facility data into optimization models that minimize development and production costs. Similarly, Tygesen et al. (2019) assessed the maintenance schedules of offshore infrastructure by developing a Bayesian neural network model that updates fatigue predictions in real time using structure and sea condition observations.

The analytical approach to lifespan assessment of offshore structures proposed in this paper builds off several of these studies by incorporating structural characteristics along with historic MetOcean, structural- and weather-related incident data into unique ML and statistical models.

1.4 CURRENT STATE OF INFRASTRUCTURE IN THE GULF OF MEXICO

Following a low in September 2011, offshore oil production and exploration has maintained a steady year-over-year increase in number of barrels produced. As of March 2019, daily oil production in the GoM was 1.9 million barrels, an increase of about 300,000 barrels per day compared to March 2018 (EIA, 2019). According to publicly available platform data from BOEM (2019), there are roughly 2,000 exploratory or production platforms operating in the GoM and only 50 of those were installed after September 2011. This indicates, at least in part, that most of the increase in production can be attributed to older wells—some of which are approaching 50 years in operation. As a whole, the average age of currently installed platforms in the GoM is just over 34 years (BOEM, 2019), on par with the ages of offshore infrastructure reported in other areas of the world (Animah and Shafiee, 2016). Assuming that the average design life of the infrastructure in the GoM is also similar to other areas of the world, it would mean that a portion of the operating infrastructure has surpassed its design life.

Rigs and platforms are the two main types of infrastructure operating in the GoM. BSEE defines *rigs* as “a major component that is added atop an offshore structure or platform, or is a part of a ‘self-contained’ mobile drilling unit such as a drillship, semisubmersible, jackup, or mobile offshore drilling unit (MODU)” (Hawkins, 2019). Similarly, a *platform* is “a raised offshore structure that can support many different functions for offshore oil and gas operations. Platforms typically have one or more wells and can either be manned or unmanned” (Hawkins, 2019). Within each of these categories are several different types of rigs and platforms (Table 2).

Table 2: Descriptive Statistics for the Platforms in the GoM (BOEM, 2019)

Platform Type	Current Count	Removed Count	Operating Depth
Fixed	Fixed: 1,320 Caisson (CAIS): 408 Subsea Template (SSTMP): 0 Well Protector (WP): 5 Compliant Tower (CT): 3	Fixed: 2,461 CAIS: 2,173 Subsea Template (SSTMP): 1 Well Protector (WP): 637 Compliant Tower (CT): 0	<= 1,540 m
Mobile Offshore Production Unit (MOPU)	MOPU: 1 Floating Production, Storage, and Offloading Systems (FPSO): 2 Spar: 18 Tension-leg Platform (TLP): 14 Mini TLP (MTLP): 4 Semi-Submersible (SEMI): 11	MOPU: 3 Floating Production, Storage, and Offloading Systems (FPSO): 0 Spar: 1 Tension-leg Platform (TLP): 0 Mini TLP (MTLP): 1 Semi-Submersible (SEMI): 2	>500 m
Unknown	228	0	varies

There are at least nine different platform and rig types operating in the GoM (Table 2). The majority of the platforms are fixed and situated in water depths of less than or equal to 1,540 m (5,050 ft). The deepest operating platforms are semi-submersible, operating in depths of about 3,000 m (9,800 ft) and located up to 200 mi. from shore. Platforms operating at the greatest depths and distances from shore are not as common, yet they account for a large portion of the total offshore oil production in the GoM (Figure 1) (EIA, 2019).

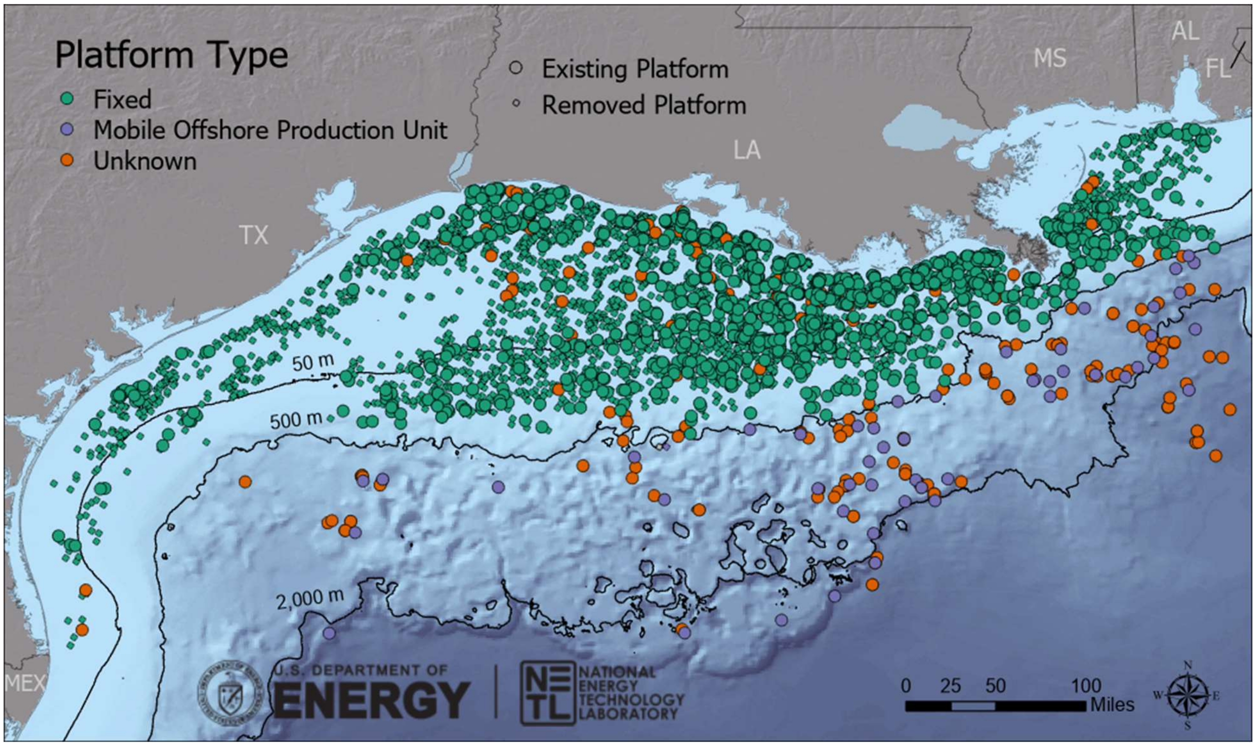


Figure 1: Map of the locations of platforms in the northern Gulf of Mexico symbolized by structure type including fixed, mobile offshore production units (MOPU), and unknown. [1]

Table 3 provides information on the age at which platforms in the GoM are removed from their installed location. This does not necessarily mean they are decommissioned. In some cases, the removed platforms are moved to other locations to continue their operations elsewhere. Some may receive a workover after removal, some may not. The average age at which the platforms are removed is 6.5 years and 20.6 years for MOPU and fixed platforms, respectively.

[1] Platform data was acquired by BOEM (2019) and is current as of December 2019. Platforms with a removal date are considered removed, otherwise are currently in use as of the data acquisition date.

Table 3: Age at Removal Statistics for Removed Platforms by Overall Structure Type

	Fixed	MOPU
Count	5,271	8
Average Age at Removal (years)	20.6	6.5
Standard Deviation	13.4	2.5
Min	0	2.6
Max	71.6	10.5

2. METHODS

The approach for analyzing the integrity of offshore platforms in the GoM proceeded in two phases. Phase 1 was exploratory with a focus on identifying possible data sources for the analysis of infrastructure integrity and subsequently evaluating ways the data could be collected, processed, and incorporated. The initial data discovery was guided by previous work on infrastructure integrity (Section 1.2). Data includes offshore rig and platform incidents (BSEE, 2019a), platform structural and location records (BSEE, 2019b), and MetOcean data (Appendix B). Figure 2 (Phase 1) details the steps taken to compile the data into a usable database for analyses. After collection, Phase 1 also focused on exploratory analysis to build an initial understanding of the relationship between the removal age of a platform and other environmental and structural correlates that were identified as integral to the integrity of offshore infrastructure.

Variables were analyzed through exploratory statistics and spatial regression. This provided insights in the factors that were statistically related to the conceptualization of lifespan—age at removal. Data and results of the analysis were then released to an internal Offshore Analytical Platform for visualization.

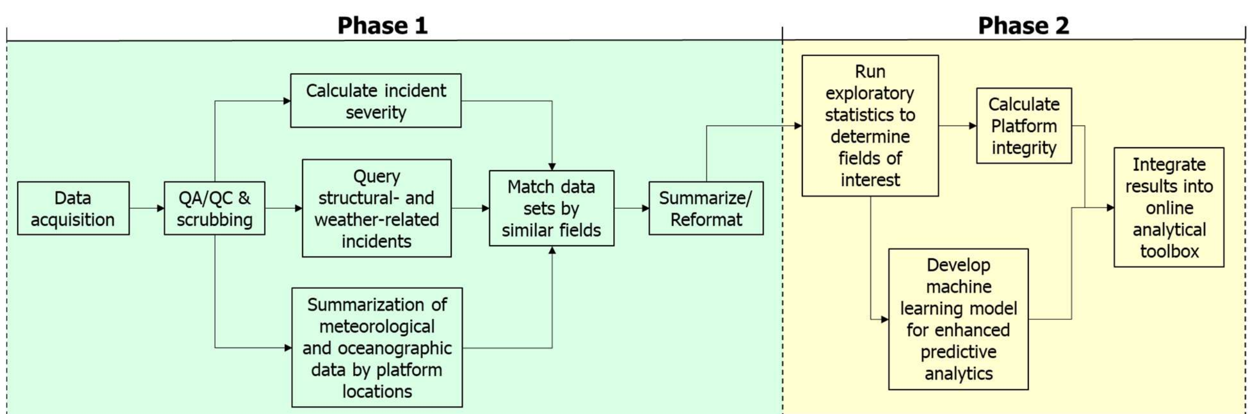


Figure 2: Workflow for processing, combining, and analyzing platform data in the two phase approach.

The Phase 2 leveraged the data and findings from Phase 1, introduced additional data to fill any identified gaps, and used several analyses to predict age at removal and risk likelihood. Specifically, Phase 2 implemented predictive ML models and geostatistics. These methods included gradient boosting classification (GBC), ANN, and geographically-weighted regression (GWR). In addition, Phase 2included the refinement of the lifespan variable, aimed at building a more explicit connection between the information provided in the incident dataset and the project's overall goal of characterizing offshore infrastructure integrity. The refinement of the lifespan variable also aimed at creating a more robust dataset for use by the ML models. Work on improving the models is ongoing as the team continues to collect and prepare new sources of data related to offshore infrastructure integrity.

2.1 DATA ACQUISITION AND PROCESSING

The research team identified several publicly available data sources describing offshore infrastructure, platform, and rig incidents from BOEM and BSEE. This data was acquired and regularly updated using customized Python scripts (Appendix A). As for environmental variables, the research team consulted with subject matter experts to identify and acquire MetOcean data (Appendix B). All datasets went through a rigorous quality assurance and quality control (QA/QC) procedure to check for errors in formatting, spelling, and record redundancy. All data was standardized (structure names in infrastructure and incident data tables) for compilation and included in the exploratory and advanced statistical and spatial analytics. Any data collection and processing that occurred during the Phase 2 of the study was aimed at filling data gaps that were identified during Phase 1. A version of this dataset will be released through NETL's Energy Data eXchange website (EDX) at the end of Phase 2 for future use (Romeo et al., 2021).

2.1.1 Infrastructure Records

Information on platform and rig complexes were acquired through BSEE (2019b). Data were available as a series of related tables that required joining. The tables "Platform Masters", "Platform Structures", and "Platform Locations" were joined based on the key fields "Complex ID" and "Structure Number".[2] The "Platform Masters" table contained 6,927 records (December 2019) and included information on flagged status (i.e. drilling, abandoned, production), equipment counts, and lease block. The "Platform Structures" table contained 7,074 records (December 2019) and included fields for installation, revision and, if applicable, removal date, authority information, and structure type (i.e. fixed, tension-leg platform). The "Platform Locations" table, made up of 7,293 records (December 2019), included latitude and longitude for each record, providing for spatially explicit analyses. All formatting was performed using Python scripts that have been designed to automatically update NETL's resources as new data became available.

The resulting table contained 7,293 records with information on structure name, production status, equipment counts, locational information, water depth, installation, and if applicable, removal dates. This table was divided into two separate tables for platforms and rigs. Phase 1 of this project was specifically focused on platforms. Future work will focus on rigs. The resulting platform table contained 7,293 records. In addition to the flat file, it was also converted into a spatial format (shapefile), which is displayed by structural type in Figure 1.

2.1.2 Incident Data

In addition to platform information, BSEE also maintains a database of incident records for production platforms, drilling rigs, and pipelines operating in U.S. waters. Incidents are defined to include all serious accidents, fatalities, injuries, explosions, and fires that take place on the offshore infrastructure. As previously mentioned, reporting requirements for incidents changed in 2006 when BSEE began to require operators to report all incidents that had the potential to be

[2] These tables can be found on the BSEE data portal and are named as such.

serious. This included situations that caused damage to the facility, worker injuries requiring days away from work, and also damage to the property that exceeded $25,000. Following 2006, incident reporting also began to include weather-related damage to infrastructure. A breakdown for the number of incidents associated with each category of platform is detailed in Figure 3.

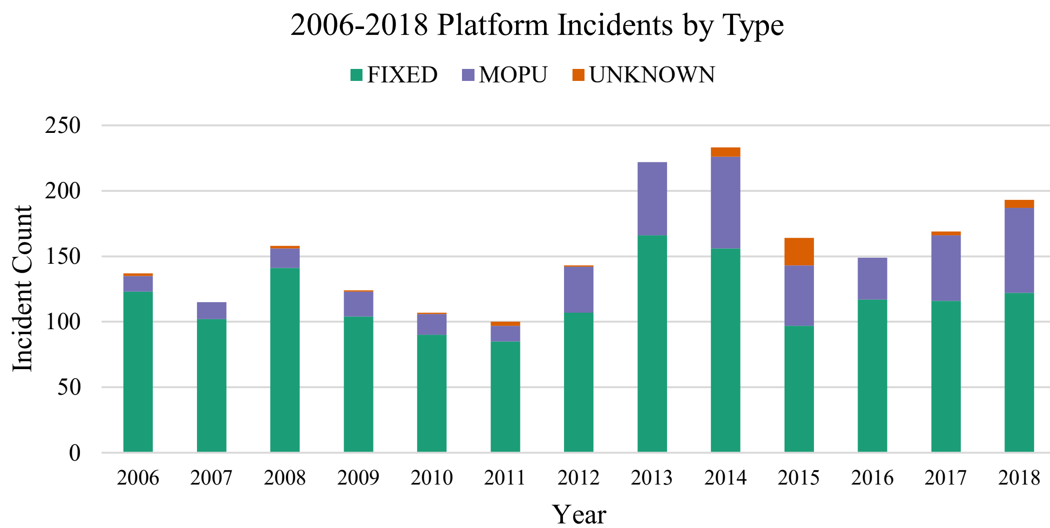


Figure 3: Number of incidents from 2006–2018 per type of platform.

Incident reports from 2006 through 2013 were acquired through from BSEE's oil and gas operational reports and incidents ranging from 2013 through 2018 were acquired as data tables from BSEE (Appendix A). The latter included self-reported incidents that occur on offshore oil and gas platforms in the GoM. Incident records from all sources were merged based on similar attributes. Additional attributes included structure name and type, lease block, damage cost, fatalities, injuries, loss of well control, fire and explosions, collisions, and whether or not a spill occurred. Incidents ranged from false alarms to deadly explosions, while other incidents resulted in no structural damage. The combined incident table currently contains 3,908 incident records for both platforms and rigs, of which 2,598 are structural- and weather-related (December 2019).

2.1.2.1 Incident Classification

To determine whether an incident may affect structural integrity, the analysis began by classifying them into three groups—structural, weather-related, and human-related—based on impact and cause. The incident dataset contained rudimentary indications of what and who was involved, but minimal information on what ultimately led to the incident. Therefore, a set protocol was developed and applied to semi-automatically classify each incident based on the description, identified keywords, and any prior categorization (human-related). Structural and weather-related incidents were used in the integrity analyses, while human-caused incidents were only included if the incident may have been structural or weather-related.

To ensure the precision of the classifications, a short test-case was completed using 50 randomly selected incidents that were manually classified separately by four researchers. Each researcher recorded key terms that directly corresponded to whether an incident was structural, weather-related, or human-related. Once complete, the classifications were compared among researchers and any discrepancies were noted. Findings from the comparison were used to build a coding protocol for the remainder of the unclassified incidents. The keywords and phrases generated from the test-case were compiled into a lookup table for automatic classification. For example, “wind” or “large swell” indicated weather, while “injured person” or “slip” were human-related terms. The resulting classification rules and term library were written into a Python script and used to classify the remaining incidents.

The script classified incidents by creating a list of all the structural, weather, human, and “no damage” key terms. At this initial stage, all incidents are assumed to be structural (i.e. the structural category receives a 1 and the other categories receive a 0 in the dataset). Then the script progressed iteratively through each incident while scanning the incident descriptions for key terms related to human, weather, and no damage in the lookup table. In the case where the script encountered structural key words along with key words related to the other categories, the script alerted the team member who manually determined the final classification based on the coding protocol below[3]. This process continued until all incidents had been classified. The protocol for the classification of incidents were as follows:

1. The structural category is defined as whether or not the incident **caused damage** to the structure or equipment
2. The weather and human categories are defined as whether or not the incident was **caused by** weather or human error
3. When an incident involves an injury but no equipment damage or impact to operations, the incident is classified as human-related
4. Human-related cannot be assumed if not explicitly stated (i.e. if it was an incident involving a crane and not explicitly called out as a human-made an error, it was not assumed human-caused)
5. All incidents with a cost associated are structural
6. If no equipment damage is stated explicitly and there is no weather-related damage then an incident can be all 0, and will not be included in the analysis
7. If there is a fire and the incident does not explicitly state that there was no damage, it is classified as a 1 for structural (and other categories if applicable)
8. Fire with no damage is classified as only human-related if it was caused by a human (scaffolding left in front of exhaust vent)
9. Lightning causing a fire is classified as both structural and weather related

Following this process, 90% of the classifications were structural, 10% weather-related, and 22% human-related, keeping in mind that classifications are not mutually exclusive. In total, 2,655 incident records were included in the analyses.

[3] Classifications are not mutually exclusive.

2.1.2.2 Severity Index

The incident descriptions made clear that not all incidents were created equal. Some were more severe than others which required the development of a severity index. The team initially summarized severity on a per incident basis. Structural and weather-related incidents were given a severity score using the variables posited to relate back to structural integrity. Specifically, the variables included fire explosion ranked by category (i.e. Catastrophic = 5, Major = 4, Minor = 3, Incidental = 2), presence of explosion or fire (0/1), loss of well control (0/1), oil spill (0/1), or H_2S released (0/1), and if there was equipment involved or equipment failure noted (0/1). Severity values were further informed by whether a collision had occurred, whether there was any property or external damage, the property damage cost (Major or over $25,000 = 3, Minor or less than or equal to $25,000 = 2), and whether the incident required muster or required facility shut-in. The initial score ranged in value from 0 to 4.82. To account for multiple incidents on a single platform, severity scores were summarized and standardized by platform for a final value between 0 and 1.

2.1.2.3 Incident-to-Infrastructure Matching

Incident data was structured as a flat file with no explicit spatial information. Moreover, there was no unique identifier linking the incidents to specific platforms within the BSEE and BOEM data (Appendix A). The team took a unique approach to matching by developing a Python script (also adaptable for incident-to-rig matching operations) that matched incidents to platform based on structure name, type (i.e. fixed, CAIS, MOPU), date of the incident, and, if applicable, the removal date. Matches were then validated at the lease block level. In total, this method matched 1,702 structural and weather-related incidents (64%) to 425 platforms. Of the 2,598 structural and weather-related incidents, 36% were left unmatched, possibly due to missing data, data entry errors, and inconsistent formatting.

The matched data was formatted in two ways. First, the most widely applied dataset contained a single record per platform, with summarized incident attribute counts. It included information on each platforms age (calculated as years from installation to present or installation to removal), equipment counts (as a proxy for complexity), locational information (latitude and longitude coordinates and lease block), incident count, and standardized incident severity scores (as shown in Figures 4 and 5). Second, the dataset was formatted by incident record with platform information associated to each incident. This dataset contained more descriptive information for each incident along with platform location and structural information.

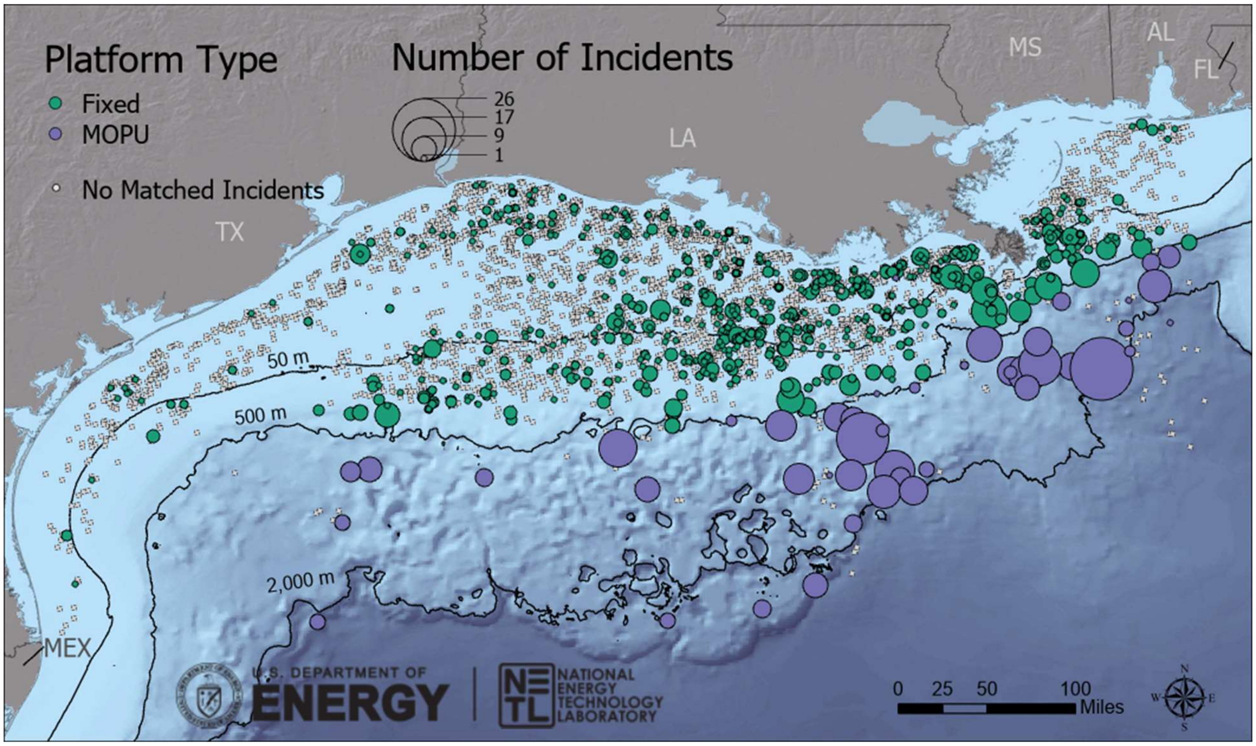


Figure 4: Number of structural- and weather-related incidents per platform record, colored by type. [4]

[4] Incident records acquired from BSEE (2019a) and matched to platform records for visualization and analytics.

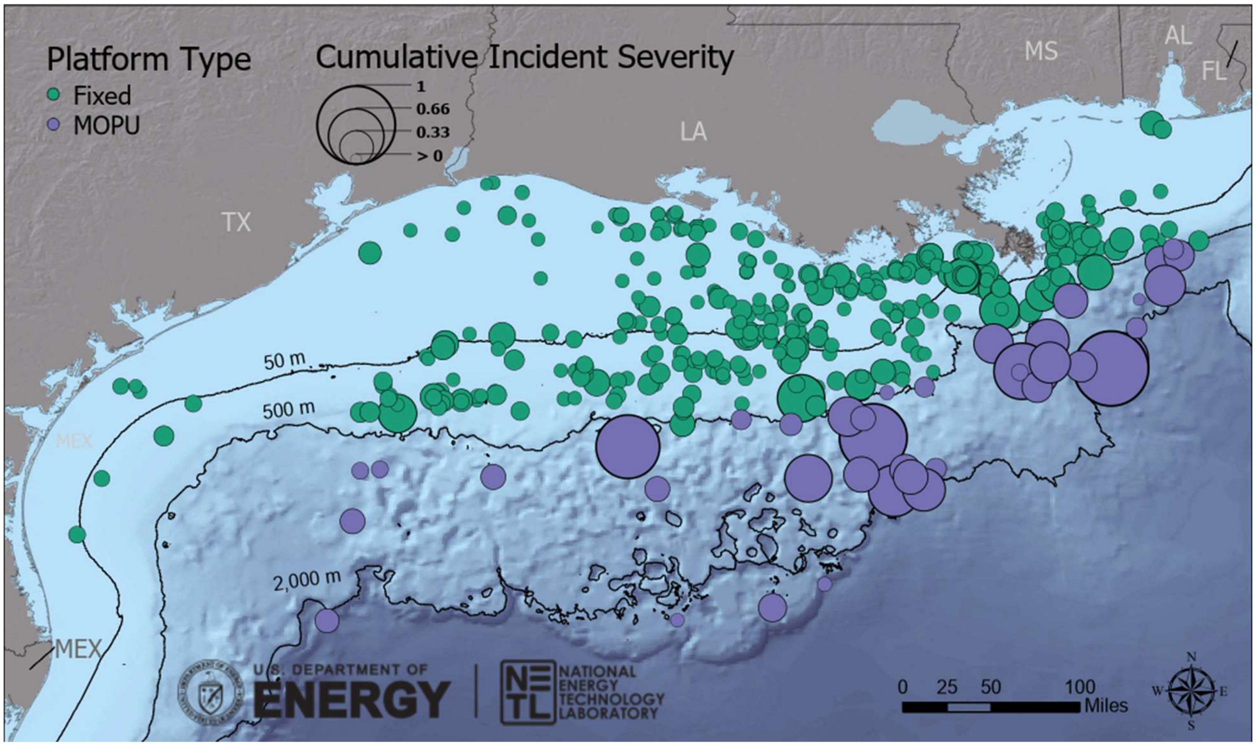


Figure 5: Normalized cumulative incident severity by platform and type.[5]

2.1.3 <u>Metocean Data</u>

An extensive collection of MetOcean data was acquired (Appendix B), including datasets for the main sources of ambient loadings (wind, wave, and currents), data associated with corrosion in the GoM, and storm data which includes information on global hurricanes. Figure 6 illustrates the spatial extent and temporal changes of sea-surface velocity as an example.

Specific data for ambient loading included current velocity (m/s), wind speed (m/s), significant wave height of combined wind waves and swell (m), mean wave period (s), primary wave direction (°), and wave power (kW/m) all at the water surface, or the standard 10 m above sea level in the case of wind. The mean, median, minimum, maximum, 25th percentile, 75th percentile, and 90th percentile was calculated for each MetOcean variable at each platform location. Two of the MetOcean variables required additional processing: the wind and current speed, which is the magnitude of the u- and v-components, and the wave power, which is calculated using wave height (h) and wave period (p) (Herbich, 2000):

$$wave\ power \qquad 0.5h^2 \times p$$

[5] Cumulative incident severity calculated using structure- and weather-related incidents from BSEE (2019a) and is normalized.

In addition to the MetOcean variables, storm data were processed and spatially summarized using the NETL's Cumulative Spatial Impact Layer™ (CSIL) tool to reflect the number of times each platform was potentially impacted by a tropical storm or hurricane (Romeo et al., 2019a; Romeo et al., 2019b). Separate summaries were made for each hurricane category (1–5) as well as tropical storms. In order to determine whether a storm impacted a platform, first the radius of each storm center needed to be estimated. Following Mei et al., (2013), the track distances were pulled from the storm data and applied to each storm track. An interaction between storm and platform was recorded if a platform fell within the radius of the storm. When calculating the number of days, the mean time between storm observations (0.23 days) was considered (i.e. storms had to last longer than 0.23 days in order to increase the number of days a platform was affected by a storm). In addition to storm radii, the storm track data also included wave height for a given radii, maximum sustained wind speed, and maximum reported wind gust. The count, minimum, maximum, and mean for each of those variables were summarized for each platform.

It is important for the time series analysis to cover the full temporal range of each platform's lifespan; however, the current availability of MetOcean data is not adequate to cover the full platform temporal range (1942 to 2020). Specific MetOcean datasets were picked to best cover this temporal range. Currently, the wind and wave-related data covers a range from 1979 to 2019, the storm data ranges from 1842 to 2019, and the surface current data ranges from 1993 to 2019. When available, MetOcean statistics were calculated for the time period between the install and removal dates of each platform. If a platform did not have a removal date, it was assumed that the platform was still active and the data was processed to the most current MetOcean date. After all of the MetOcean and hurricane data was summarized for each platform, the statistics were matched to the infrastructure-incident dataset to create a summary table of the physical characteristics, incident history, and MetOcean conditions for each platform record.

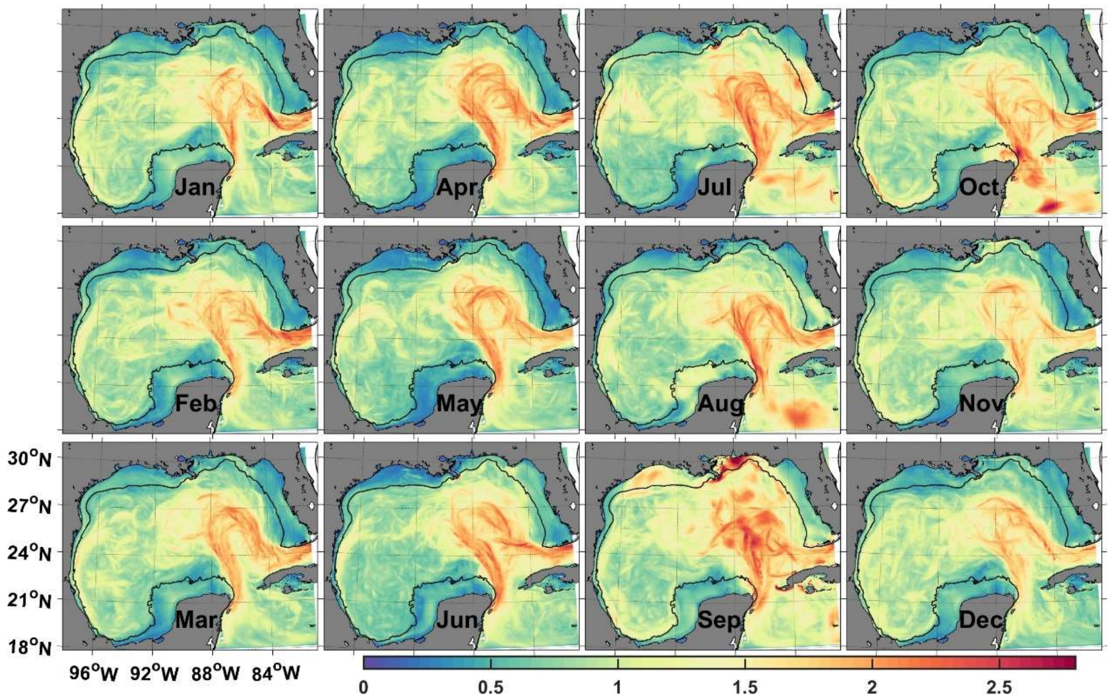


Figure 6: Maximum sea-surface velocity magnitude (m/s) from a data-assimilating ocean model (HyCOM GoM) from twelve years of data (2003–2014).

2.1.4 Corrosion Data

Ocean ambient conditions influence the rate at which the materials of a structure will corrode. In particular, current velocity, temperature, salinity, dissolved oxygen, and microbial activity are known to affect the corrosion rate of metals. A full list of the variables used to model corrosion are included in Appendix C. Nitrate, phosphate, and silicate were used here as a proxy for microbial activity and biological growth since nutrients usually imply living organisms. The research team found that the effects of corrosion directly impact the integrity of an offshore structure over long periods of time. Therefore, it is hypothesized that adding corrosion data into the models will increase performance and contribute to a more robust result. There are two methods used to incorporate corrosion information into the models.

The first method used corrosion equations based from the literature. The most basic equation for modeling corrosion is the "power-law" model (Melchers, 2016):

$$c = 0.14 \times t^{0.75}$$

where t is the number of years and c is the corrosion value. The corrosion value can be calculated for each platform record, which will be the corrosion variable to add into the models.

The second method follows the approach taken for wind, wave, and current velocity data, which is to include statistics for ambient conditions per platform location directly into the model. The advantage of this approach is that it will test to see if the models can indirectly detect a signal of corrosion and its hypothesized effect on the dependent variables (integrity measures).

2.1.4.1 Dependent Variable

This analysis considers platform age as the dependent variable of interest. Platforms with longer lifespans are generally considered to have higher integrity (or maintained operational integrity over a longer period of time), while structures with shorter lifespans are assumed to have experienced conditions that degraded their integrity at a higher rate, resulting in removal at a younger age. The variable was calculated as the difference of the installation year and removal year. As such, only platforms that have been decommissioned or removed are included in the initial assessment.

2.1.4.2 Explanatory Variables

As detailed in the previous sections, the selection of explanatory variables was based on a literature review followed by a cataloging of data availability. Of particular interest was the structural factors found across the platforms (Table 4), the wind, wave, current, and storm information derived from MetOcean data (Appendix B), and data that may contribute to corrosion (Appendix C).

Table 4: Explanatory Variables Related to the Platform Characteristics, the Range of Each Variable's Value, and the Data Type

Variable	Type of Variable	Range of Values
Structure Type	Categorical	Fixed, CAIS, SSMT, WP, CT, MOPU, MODU, FPSO, TLP, MTLP, Spar, SEMI, unknown
Water depth	Continuous	1–10,000 (ft)
Incident count	Continuous	0–28
Overall incident severity	Continuous	0–1
Average age of platforms during incidents	Continuous	0.2–63.5
Age of platform at last recorded incident	Continuous	0.3–63.5
Rig count	Continuous	0–1
Crane count	Continuous	0–5
Slant slot count	Continuous	0–11
Slot Count	Continuous	0–62
Slot drill count	Continuous	0–60

2.2 DATA ANALYSIS

Phase 1 of the data analysis evaluated potential correlation and spatio-temporal trends among the dependent variable (2.1.5.1) and explanatory variables (2.1.5.2) through statistical and spatial analytics and visualization. Explanatory variables were also examined for cross-correlation to determine any covariance and the possibility of variance inflation.

Leveraging lessons learned from Phase 1, Phase 2 employed ML algorithms and predictive analytics to model the expected lifespan of platforms and identify potential areas or platforms of greater risk.

2.2.1 Phase 1 – Exploratory Analysis

Inferential statistics in the form of correlational analyses identified whether any significant patterns existed between the independent and dependent variables. Considering that different platform structures have different purposes, platforms were analyzed by their type (fixed, mobile, and other, Figure 1). The fixed category contains several sub-types including CAIS, SSTMP, WP, and CT. Similarly, the mobile category included MOPU, MODU, FPSO, Spar, TLP, MTLP, and SEMI as sub-categories. Unknown structure types included all platform records where structure type was not provided. In terms of when these platforms are removed, MOPU tend to be more mobile than fixed platforms which contributes to their lower age at removal when compared to fixed platforms (Figure 7).

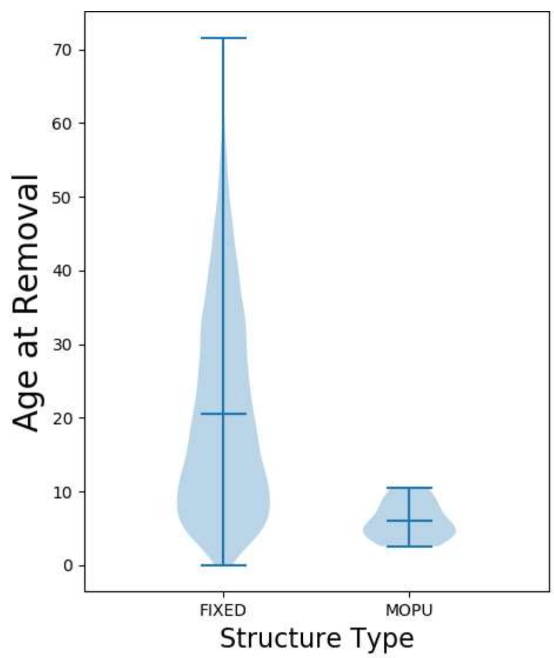


Figure 7: Violin plots of platform age at removal by overall structure type using known removed platforms data.

2.2.1.1 Correlation Analysis and Initial Risk Index Generation

Statistical tests were used to evaluate which platform and incident variables were correlated in a statistically significant way to the platform's lifespan. This began by identifying whether there were statistically significant differences between the different platform types using T-Tests on age at removal (removed platforms). Then, a covariance matrix was used to test the explanatory variables (Table 4) for independence. This was done using each structure type and then for the dataset as a whole (Figures 8–11). High covariance between some of the explanatory variables led to the removal of some variables from further analytics. The remaining variables were evaluated against age at removal using several statistical tests. The specific test that was applied to each variable varied based on key test assumptions (i.e. variable types, ranges).

Numerical Data: The Kendall rank correlation test was used to evaluate the similarities in paired data of numerical values (Abdi, 2007). This test determines the strength of the association between the values for each tested pair of variables. In this instance, it was used to compare several of the explanatory variables (water depth, incident severity, average age of incident(s), last age of incident, crane count, slot count, slant slot count, slot drill count, classified wave height, and classified wind speed) with age at removal, stratified by type. Each of the tested variables had a continuous scale and followed a monotonic relationship, meeting the assumptions of the test. Statistical significance was determined by the tau value and sign (positive or negative relationship).

Categorical Data (2 Categories): In addition to measuring the statistical significance of structure type by lifespan, rig count and slant slot count were evaluated against age at removal using T-Tests and the analysis was stratified by structure type. Both rig count and slant slot count had two categories. The T-Test was used to compare the means between the observations within the categories to determine whether the variation was statistically significant (Table 4) (Kim, 2015). Statistical significance was determined by a p-value ≤ 0.05.

Categorical Data (3+ Categories): The Kruskall-Wallis Test is a non-parametric test commonly used for comparing a categorical variable to a numerical variable (McDonald, 2009). This test is appropriate for data that are not normally distributed. Thus, it was an appropriate approach for comparing crane count, slot count, slot drill count, slant slot count, classified current speed, classified wave height, and classified windspeed against age at removal. These explanatory variables had more than two categories and therefore this test was used to determine whether there was a significant difference in median age at removal between the categories. Statistical significance was determined by a p-value ≤ 0.05.

Risk Index: Statistically significant results from each of the tests were then used to create a risk index, which was applied to the remaining existing platforms. This index represented integrity, where lower risk indicates higher integrity, and lower integrity indicates higher risk. The first step in calculating the integrity index is to normalize the platform age by subtracting the average age at removal by type t from the platform i current age:

$$Age_{it} = CurrentAge_i - Avg(RemovalAge_t)$$

The age (Age_{it}) value was then scaled depending on a positive or negative result. If Age_{it} was negative it was divided by the minimum age of the platforms of type t and multiplied by negative one. If the value was positive it was divided by the max age of platform type t:

$$AgeNorm_{it} = f(Age_{it})\begin{cases} \dfrac{Age_{it}}{min} * -1, & if Age_{it} < 0 \\ \dfrac{Age_{it}}{max}, & if Age_{it} > 0 \end{cases}$$

Next, the risk values considered the crane count, slot count, and slot drill count for each platform i. This was determined by the total number installed on the platform (j) and then scaled by the platform type (t):

$$CraneCount_{ti} = f(t,j)\begin{cases} j * .75, & if\ t = CAIS \\ j * -1, & if\ t = FIXED \\ j * -.5, & if\ t = WP \\ j * .25, & if\ t = oth \end{cases}$$

$$SlotCount_{ti} = f(t,j)\begin{cases} j * .75, & if\ t = CAI \\ j * -1, & if\ t = FIX \\ j * .25, & if\ t = WP \\ j * .25, & if\ t = other \end{cases}$$

$$SlotDrillCount_{ti} = f(t,j)\begin{cases} i * -1, & if\ t = CAIS \\ i * -1, & if\ t = FIXED \\ i * -.75, & if\ t = WP \\ i * -.25, & if\ t = other \end{cases}$$

The risk related to incidents was determined as the sum of the scaled (0 – 1) incident severity and scaled incident count. Lastly, the risk associated with oceanographic conditions was determined with the classified values for current velocity, wave height, and wind speed, and the scaled depth values (0 – 1).

$$IncidentIndex_i = \sum_{i=1} IncSeverity_i, IncCount_i$$

$$AmbientIndex_i = \sum_{i=1} CurrVel, WaveHeight, WindSpeed, Depth_i$$

The formal risk index for each platform was then calculated as follows:

$$Risk_{it} = \sum_{i=1} AgeNorm_{it}, CraneCount_{it}, SlotCount_{it}, SlotDrillCount_{it}, IncidentIndex_i, AmbientIndex_i$$

2.2.1.2 Exploratory Spatial Regression

Exploratory spatial data analysis (ESDA) is the process of investigating data to understand patterns and associations between variables that are explicitly tied to the geography of the data. Most spatial data exhibit some pattern of correlation related to where they are located. Features that are closer together in geographic space tend to be more closely related than features that are further apart (Tobler, 1970). This characteristic of spatial data is effectively termed as spatial autocorrelation.

Failing to consider autocorrelation can have a detrimental effect on the interpretation of the modeling results. Specifically, it can alter both the significance and magnitude of the effect of the independent variables. That said, the Phase 1 analysis used spatial regression as an exploratory analysis tool to identify whether spatial autocorrelation was present in the dataset and to begin to understand the magnitude of the effect that the explanatory variables had on integrity. Spatial autoregressive models are a type of exploratory spatial regression and are particularly well suited for controlling for the effect of space on the dependent variables. The spatial autoregressive model takes the form:

$$y_i = \beta_o + X\beta + \rho w_i y_i + \epsilon_i$$

where y_i is the dependent variable, β_o is the intercept, X is an n x k matrix of independent variables, β is a k x 1 vector of coefficients, and $\rho w_i y_i$ is the spatial weights matrix reflecting the lagged dependent variable.

Following the correlational analysis, several independent variables were considered, including the incident severity, current speed, wave height, and dummy variables representing structure types WP, CAIS, and other, with fixed platforms as the reference group. In total, four models were estimated—two ordinary least square models that included tests for spatial autocorrelation (Moran's I) and variance inflation factors (multicollinearity among independent variables), as well as two spatial lag models with the additional independent variables of platform depth, incident count, wind speed, and distance to shore (Table 4).

2.2.2 Phase 2 – Predictive Analytics

While there are many predictive algorithms that could be applied for the goals of this project, the team developed a GBC model, ANN model, and Long Short-Term Memory (LSTM) neural network (NN) for two purposes. First, these models enable us to explore the tradeoffs between statistical ML models and neural networks. Second, this multi-faced approach serves as a robustness check of the results.

The GBC and ANN models were used to predict the proxy variable for integrity, *Age at Removal,* which was recast into removal age classes of 0–11 years, 11–20 years, 20–30 years, 30–42 years, or 42–72 years. Both classification models were designed to predict the age bin a platform should fall into given the associated features and their relationship to age at removal. The GBC uses variables from the training of the model that are identified through an iterative feature evaluation technique. The ANN uses all features except individual incident presence absence labels (Appendix C and D). These models have computational characteristics that benefit this classification problem, but also have characteristics that introduce uncertainty. In order to assess the best approach moving forward, these two models are compared using multiple evaluation metrics to determine which algorithm is better fit for this classification problem.

In addition to the above ML approaches, an additional spatial-analytic predictive model was developed and employed. GWR is a relatively new technique that is sensitive to spatial non-stationarity across the dependent and independent variables. The GWR model considers local spatial variation in the data and estimates model coefficients for each observation. Moreover, a predictive model can be developed using a train-test approach after creating a GWR model. Phase 1 provided the necessary information for variable selection and the GWR model was developed using distance to shore, Category 4 hurricane days, max wave power (log), max wave period, and max wave height.

Finally, a LSTM NN model is currently under development to predict risk likelihood and estimate incident impact on lifespan. Unique to LSTM NN models is the use of time series predictions (Malhotra et al., 2015; Pham et al., 2017). For this project, a LSTM NN model is beneficial because it takes as input both categorical and continuous features. With this information, LSTM NN models are able to capture incidents overtime and their relationship to the combination of environmental and structural conditions that converged at the time of the incident. However, with the size of the current dataset and number of input features available, the LSTM NN model is unable to adequately learn and make predictions. The model can be improved with a larger number of features and records or, alternatively, a model may be created similar to the ANN mentioned above with the addition of incident dates and infrastructure info at those dates to predict risk under current circumstances.

2.2.2.1 Evaluating Model Performance

While the GBC and ANN algorithms have a different architecture, there are still common ML methods that can be used to evaluate the performance of each classification model and compare the two. Comparing each model's ability to predict each class provides a deeper understanding of how well certain age at removal year ranges can be predicted.

Here, the percent of correct predictions on the training and testing sets for both models are used to evaluate which model has an overall higher accuracy at predicting the age at removal for a platform. The evaluation metrics *precision*, *recall*, and *F1* score are used to determine how well each model can predict the age at removal class. The precision metric indicates the ability of the classifier to not mislabel a class, the recall metric indicates the ability of a classifier to find all correct labels, and the F1 score is a weight of the precision and recall metrics.

As far as the GWR model is concerned, it is a completely statistical approach to prediction. Drawing from linear relationships between variables and the spatial patterns present in the data, the model predicts values at unknown locations using the coefficient values from a training dataset. To evaluate model performance, an R^2 can be calculated between the y (observed) and y-hat (predicted) values for the out-of-sample prediction. In addition, a correlation coefficient between the observed and predicted values is also computed.

Each variable in the model will have a certain amount of predictive power when determining the age at removal. This is important to understand as it can indicate which variables are strong and poor predictors of integrity within a given study area. Importantly, not all offshore environments are the same, and the results from one system may not carry over to another. To gain a better understanding of a specific operating environment, decision makers need to know which variables are contributing the most to the degradation of infrastructure integrity. They can then monitor these specific variables and make more informed decisions regarding the deployment of resources, mitigation efforts, or increased monitoring.

For the GBC model, a Gini number is used to identify the variables with the most predictive power. For the ANN model, Feature Permutation Importance (Breiman, 2001) as a measure of the change in cross-entropy loss when the feature is randomly permuted from the model is used to determine the most influential features. Finally, the variables in the GWR model can be standardized and the magnitude of the standardized variables can be used to determine importance.

3. OBSERVATIONS AND RESULTS

This work was aimed at two objectives. First, it sought to assess the integrity of existing offshore infrastructure by collecting data on the offshore environment, associating structural and environmental characteristics to infrastructure, and analyzing the relationships between those variables and the proxy variable used to measure integrity. Second, this work set out to identify areas or specific infrastructure that are at-risk of potential failure given the identified relationships between infrastructure integrity and the associated structural and environmental variables. These two objectives were met using a two-phase approach. Phase 1 explored the relationship between each explanatory variable and their individual contributions to integrity using several inferential statistical techniques. Phase 2 developed and tested methodological approaches capable of predicting the date that platforms were removed.

3.1 CORRELATION ANALYSIS

Correlation tables were prepared for each of the explanatory variables and for each structure type. In addition, Figure 8 details the correlation matrix for each of the explanatory variables over all structures. Within the correlation plots, red values indicate a negative correlation and blue values indicate a positive correlation. The hue of the color represents the strength of the correlation.

Across each of the platforms, there are some clear patterns that emerge. The strongest correlations for all structure types are between slot count, slant slot count, and age at last incident (Figure 8). Slot count and slot drill count are not surprisingly correlated and slant slot count and average age at incident are highly correlated (.88). The number of incidents is strongly correlated with cumulative incident severity (.91) because it was used in the creation of the severity variable. The number of incidents is moderately correlated with water depth (.53), indicating that there is some association between deeper platforms having more incidents.

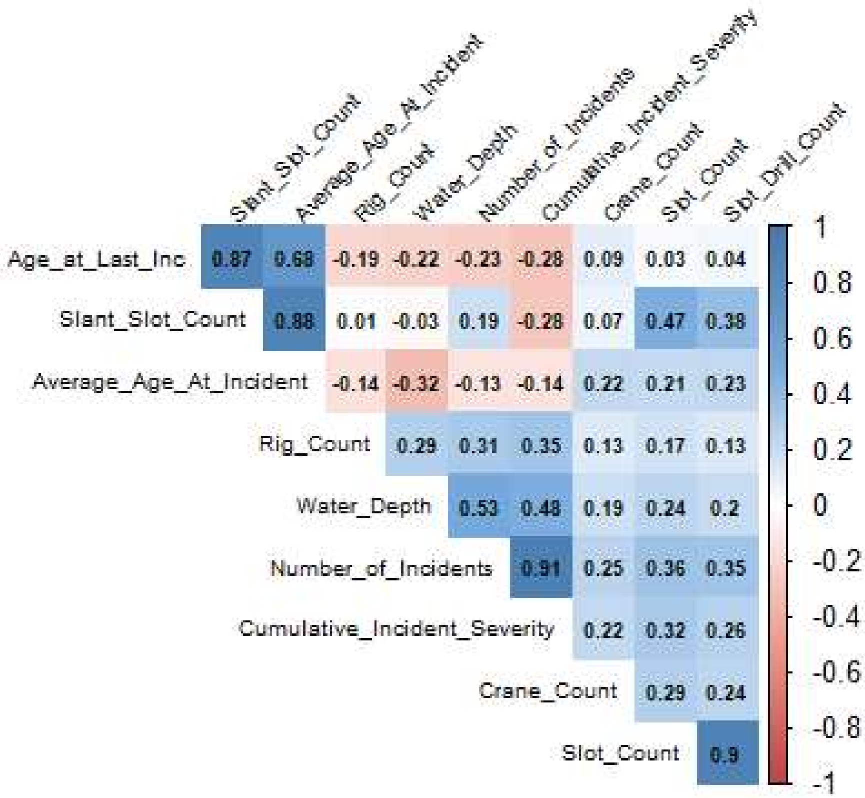


Figure 8: Correlation of explanatory variables for all structure types.

Recall that fixed structure types are most abundant in the offshore GoM and have a correlation pattern that is similar, yet distinct, from the pattern depicted when all structure types are considered (Figure 9). Slot count and slant slot count remain highly correlated (.93), and so too does cumulative incident severity and number of incidents. The correlation between water depth and number of incidents has dropped (from .53 to .47), but this is to be expected given the geographic concentration of fixed structures in shallow waters. Interestingly, there is a substantial drop in correlation between rig count and number of incidents when only considering fixed platforms (from .31 to .16).

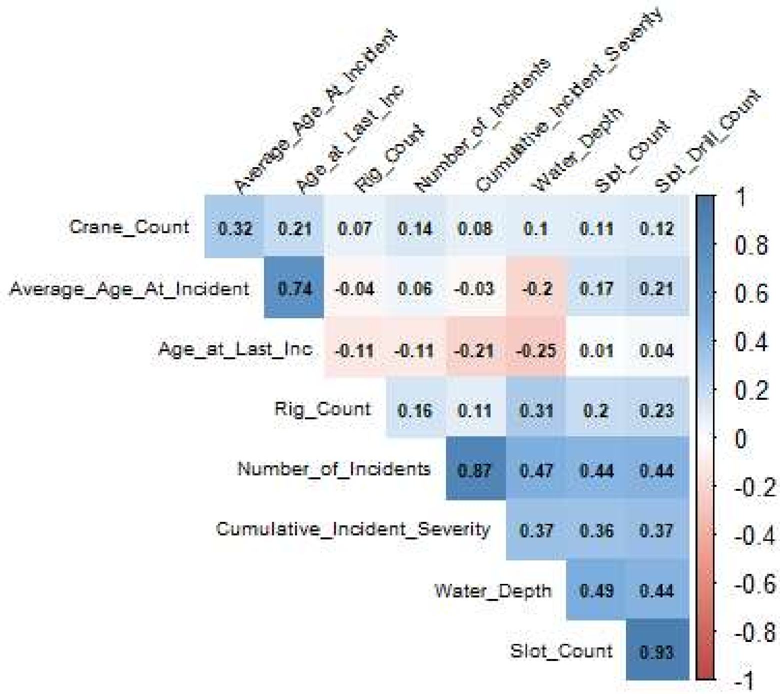


Figure 9: Correlation matrix of explanatory variables for fixed platform structure type.

MOPUs generally operate in deeper water and are more complex structures, resulting in a different correlation pattern than fixed or all structures combined (Figure 10). In a similar manner to fixed platforms, the correlation between water depth and number of incidents is significantly lower when only considering MOPU (.11). Where water depth was positively correlated to slot drill count for fixed structures (Figure 9) and for all structures (Figure 8), it displays a moderate negative correlation for MOPU (-.34). For number of incidents, rig count, crane count, and slot count are all positively correlated and positive—perhaps indicating that a more complicated structure is associated with more incidents.

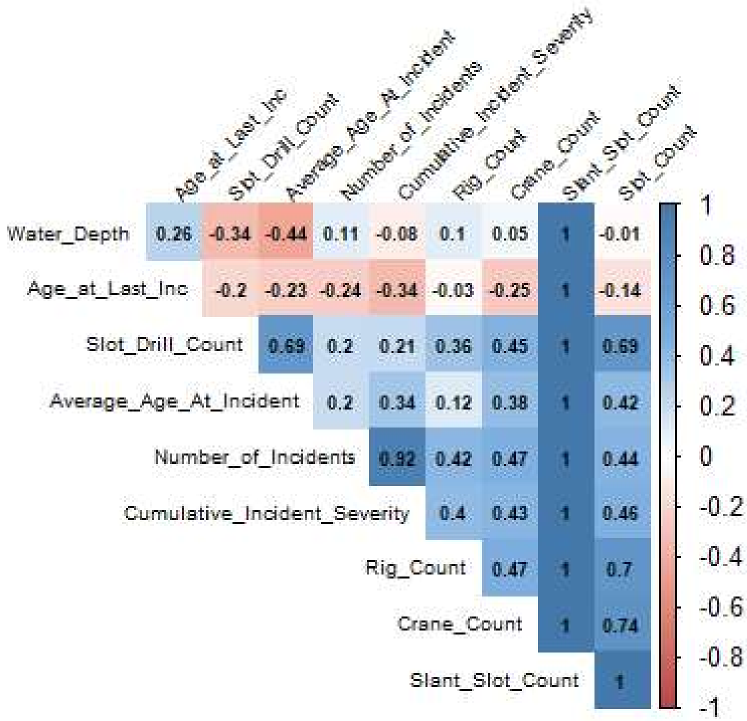


Figure 10: Correlation matrix of explanatory variables for MOPUs structure type.

Due to the unknown structure type, there were very few structural correlates that could be examined (Figure 11). Information on rig count, crane count, and water depth, were examined along with number of incidents. Interestingly, unlike the other structures, these variables were negatively correlated with number of incidents.

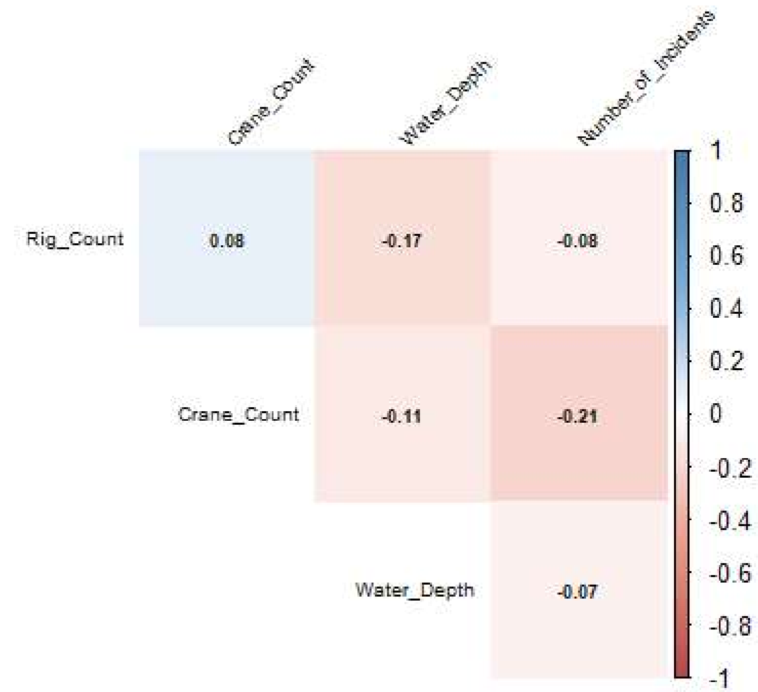


Figure 11: Correlation matrix of explanatory variables for unknown structure type.

3.2 SPATIAL REGRESSION

Table 5 shows the magnitude and direction for each coefficient. The first two models do not account for spatial autocorrelation but did test for it. The test statistic (Moran's I) and VIF (variance inflation factor) are both significant, indicating strong spatial autocorrelation across the dependent variable and high multicollinearity among some of the explanatory variables. The second OLS model drops several of the correlated variables and with that a significant drop in the VIF.

Model 3 includes controls for spatial autocorrelation and indicates a positive and significant effect of severity of incident on age at removal. This is not an expected result as it was assumed that incident severity would decrease lifespan (contribute negatively to integrity). The coefficient on wave height is significant and negative indicating that as wave height increases, platforms have a shorter lifespan. The model explains 33% of the variance.

Model 4 also accounts for spatial autocorrelation and indicates a small improvement in explained variance with the addition of variables for platform type using fixed as the reference category (R^2 = .37). Severity and wave height retain their direction and significance. All platform variables are significant and negative indicating a younger age at removal compared to their fixed-type counterparts.

After controlling for spatial effects, the results indicate that several of the explanatory variables of interest exert a significant independent effect on the age at removal. With the exception of the coefficient on severity, the independent variables act on the age at removal in the expected direction. In particular, as current speed and wave height increase, a significant decrease in the age at removal can be expected.

Table 5: Coefficients and Significance Table Estimating Age at Removal for Platforms that Have Already Been Removed

	Model 1 (OLS)	Model 2 (OLS)	Model 3 (SL)	Model 4 (SL)
Severity	-5.1	88.99**	77.07**	68.43**
Current velocity at surface	-8.75	2.54	0.302	-2.96
Wave Height	-5.3	-8.05**	-3.25**	-8.09**
WP				-1.07*
CAIS				-6.09**
Other				-13.57**
Log(depth)	2.92**			
Incident Count	6.27**			
Wind Speed	-11.05**			
Distance to shore	-0.03**			
Lag_Age			.629**	.618**
Constant	25.33**	23.77**	8.99**	14.03**
Moran's I	46.5**	47.32**		
R-squared	0.05	0.03	0.33	0.37
VIF	34.56	3.82		
P-value : .01, .05*, .10**				

3.3 SPATIAL DISTRIBUTION OF RISK

The risk index for each platform calcualted using the equations in 2.2.1.1, are plotted in Figure 12. The resulting risk index values ranged from -0.86 to 4.37 where lower values indicate higher integrity (lower risk of failure), and higher values indicate lower integrity (higher risk of failure). Existing platfrom risk indices were then sorted into 5 classes using Jenks nautral breaks optimization (Jenks and Caspall, 1971), which minimized the average deviation from the mean per class. Platforms with lower risk rankings (green squares) are clustered in shallow waters along the shelf, with notable clusters appearing along Louisiana's coastline. Platforms classified as higher risk, shown in red, are generally located further offshore and in deeper waters. Out of the 2,089 existing platforms displayed in Figure 12, 37.9% (n =791) were classified as having low to low-medium (light green) risk and 22.5 (n = 469) were clasified as having medium-high (orange) to high risk.

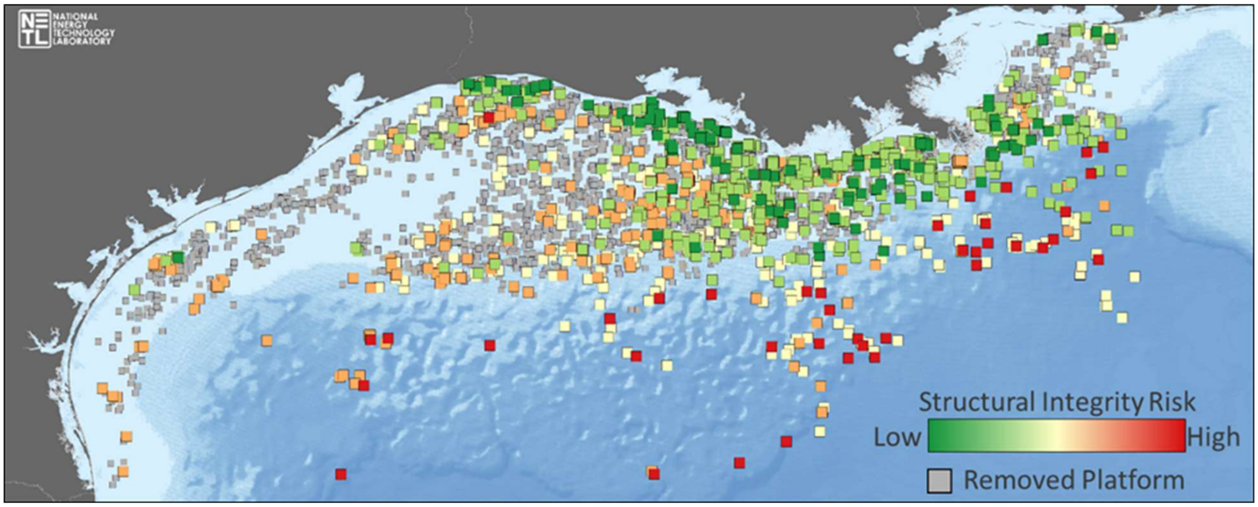


Figure 12: displays the risk index applied to existing platforms, based on significant relationships identified while statistically analzing removed platform data.

3.4 PREDICTIVE ANALYTICS RESULTS

3.4.1 Machine Learning Models

Table 6 shows the current prediction accuracy of the GBC and ANN models using the training and test sets. Again, these measures are used to evaluate each model's overall performance and test for the presence of bias or variance. The GBC model has a training set accuracy of 100% and a testing set accuracy of 85.6%. The ANN model has a training set accuracy of 90.1% and a testing set accuracy of 84.3%. The GBC model has a higher accuracy overall on the test set by 1.3%. The higher accuracy for the training set for the GBC model (14.4%) and ANN model (5.8%) indicate predictions with a fairly large degree of variance and suggest overfitting of the training data. It is also possible that there is avoidable bias present in both models; however, this cannot be confirmed because the human-level accuracy is not known.

Table 6: Training Set and Testing Set Accuracy Scores for the GBC and ANN Models

Model	Training Set Accuracy	Testing Set Accuracy
GBC	100%	85.6%
ANN	90.1%	84.3%

The values of precision, recall, and the F1 score in Table 7 are used as evaluative metrics for each age at removal class. The GBC model performed slightly better in all metrics for the 20–30 year, 30–42 year, and 42–72 year classes. In general, the difference in evaluation metric values for each model and class are somewhat negligible, with an average difference of only 2.35%. In comparison to the other classes, both of the model's predictive performance for the 11–20 year and 20–30 year ranges were lower. The models are best at predicting the 0–11 year and 42–72 year ranges.

Table 7: Classification Reports for the GBC and ANN Models.[6]

Class	0–11 years	11–20 years	20–30 years	30–42 years	42–72 years
Precision (ANN)	**0.924**	0.747	0.818	0.847	0.941
Precision (GBC)	0.913	**0.786**	**0.821**	**0.855**	**0.948**
Recall (ANN)	0.843	**0.867**	0.779	0.868	0.851
Recall (GBC)	**0.902**	0.811	**0.786**	**0.897**	**0.901**
F1 score (ANN)	0.882	**0.803**	0.798	0.857	0.894
F1 score (GBC)	**0.908**	0.798	**0.803**	**0.875**	**0.924**

Table 8 lists the top features for both ML models based on the Gini number for the GBC model and the feature permutation importance for the ANN model. In general, both models found the storm data to be important variables for predicting age at removal. The GBC model identified total hurricane counts, underwater completion count, maximum sustained wind speed, rig count, and Category 3 hurricane count yearly max as most important. From the ANN model, Category 1 hurricane count, yearly mean, Category 1 hurricane sum, longitude, Category 3 hurricane count yearly mean, and Category 3 hurricane sum are most important. Even though different hurricane-related variables were found to be important, the storm variables related to Category 1 and 3 hurricanes were generally the most important. The top 5 variables for the ANN model are location-specific, meaning that these variables change based on geographic location. Only 3 of the top 5 variables for the GBC are location-specific while the other two, underwater completion count and rig count, are specific to the platform characteristics.

Table 8: Top 5 Most Important Features of the GBC Model-Selected Variables Based on the Gini Index of the ANN Input Variables Based Off of Cross-Entropy Loss

Rank	GBC	ANN
1	Total hurricane counts	Category 1 hurricane count yearly mean
2	Underwater completion count	Category 1 hurricane sum
3	Maximum sustained wind speed	Longitude
4	Rig count	Category 3 hurricane count yearly mean
5	Category 3 hurricane count yearly max	Category 3 hurricane sum

[6] The results are for the prediction of age at removal classes for the testing data and include the precision, recall, F1 score, and support metrics.

3.4.2 Geographically Weighted Regression

The GWR model performed well. Across all platforms, the model accounts for nearly 90% of the variance (R^2 = .895) using distance to shore, Category 4 hurricane days, max wave power (log), max wave period, and max wave height. On a platform-wise basis, the R^2 ranges from .51–.97. With the exception of Category 4 hurricane days, the coefficient values range from negative and positive (Table 9). This is an indication of spatial non-stationarity which is further illustrated after plotting the coefficient values for each of the variables of interest for each platform (Appendix E). Depending on the location, the effect of each coefficient on age at removal will change—sometimes being associated with a decrease in removal age and other times being associated with an increase in removal age.

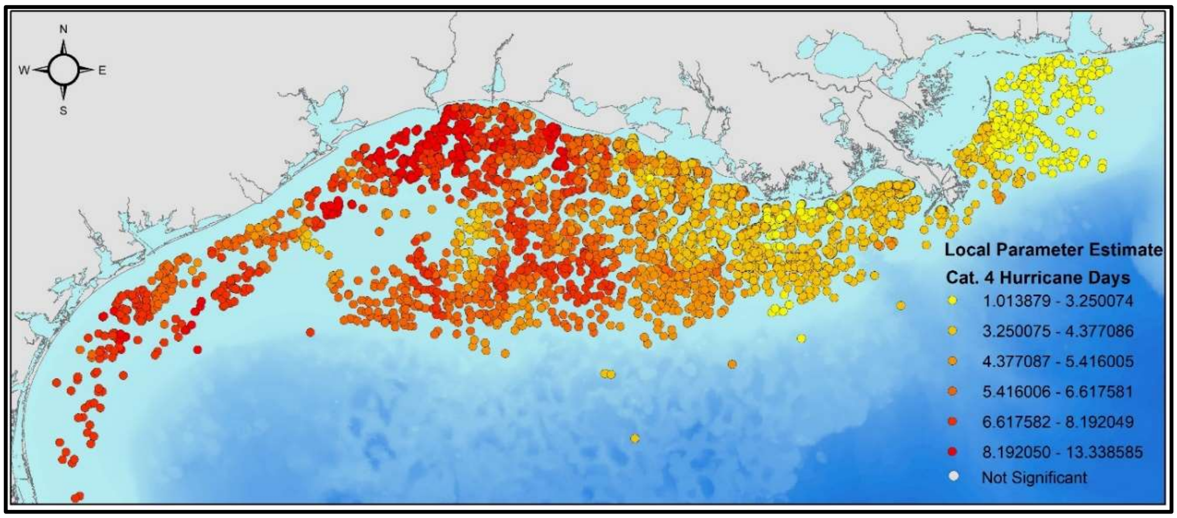


Figure 13: Local parameter estimates for the Category 4 hurricane variable. The relationship changes as one moves from east to west.

Where Category 4 hurricane days is concerned, the positive coefficient indicates that a one unit increase in Category 4 hurricane days is associated with an increase in age at removal on average. This is somewhat expected, given that older platforms will experience more hurricanes simply because they were installed earlier. However, the increase in age at removal varies over all platforms. There are clear pockets of platforms that respond differently to the number of Category 4 hurricane days (Figure 13). More specifically, platforms in the western GoM have a larger positive relationship with hurricane 4 days than platforms in the eastern GoM. This is likely a reflection of the severity and frequency of hurricanes across the GoM. Platforms around the Mississippi Delta are more likely to experience more severe hurricanes at a higher frequency than platforms in the eastern GoM which can contribute to an earlier removal age.

Table 9: GWR Model Statistics for Each of the Calculated Coefficient Values

GWR Gaussian Model					
Variable	**Mean**	**STD**	**Min**	**Median**	**Max**
Intercept	-10.155	27.699	-158.655	-1.832	110.483
Distance to Shore	-0.049	0.301	-1.529	-0.025	1.591
Cat 4 Hurr. Days	5.136	1.757	1.011	4.671	12.432
Max Wave Power (log)	5.056	10.418	-49.395	3.63	103.644
Max Wave Period	0.335	1.108	-2.847	0.295	5.951
Max Wave Height	-3.34	6.894	-104.782	-2.492	20.271
				R2	**0.895**
				AIC	**28,160.32**

The GWR model took a similar approach to prediction as the ML models. Specifically, a training dataset was used to calculate the spatial effects and coefficient estimates for each variable at each platform location. Then, the model was applied to an out-of-sample test dataset to predict age at removal. The R^2 for the observed and predicted values was .895 (Figure 14) and the correlation coefficient was .92.

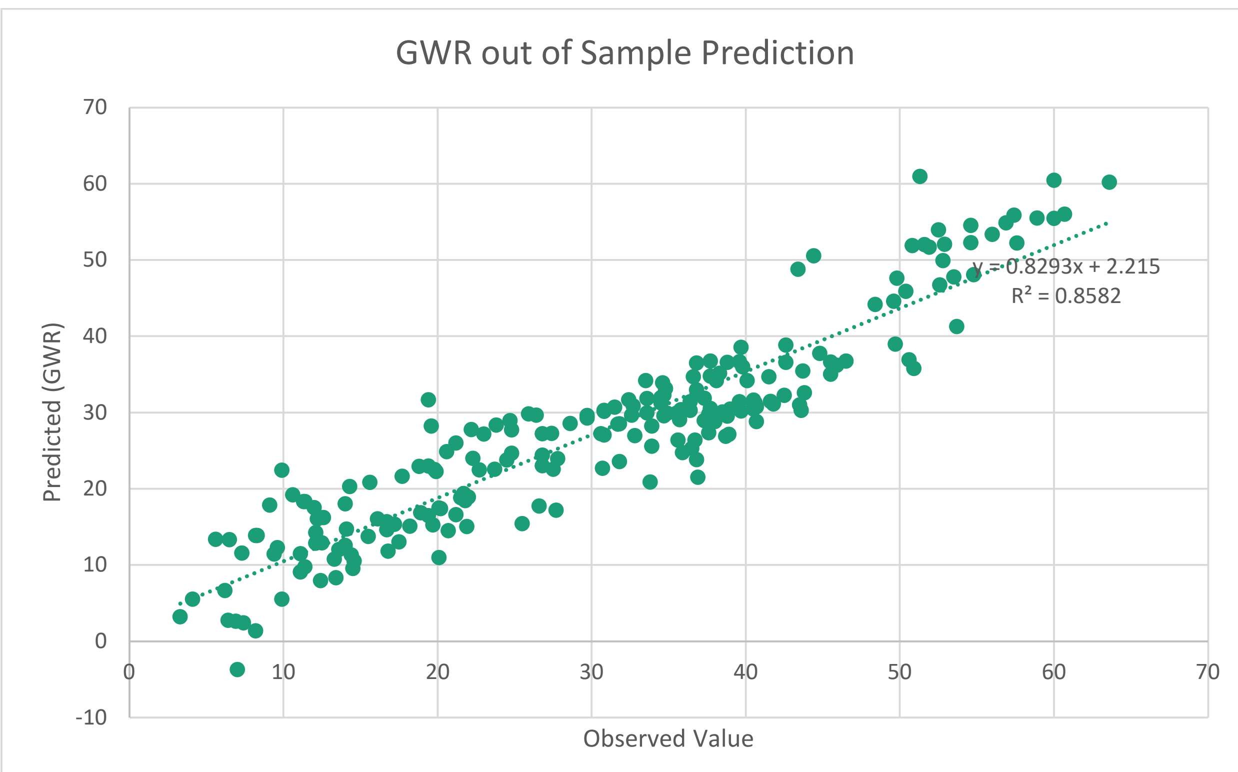


Figure 14: Out of sample prediction for the GWR model using a test-train data split.

4. DISCUSSION

Much of the infrastructure in the offshore environment is approaching or has already passed its original design life. As infrastructure ages it becomes more susceptible to failure, and this increases the older it gets. In order to maintain the highest degree of safety in offshore operations as possible, it is important to continue developing novel approaches to guide safety interdiction efforts and enhance decision making surrounding the mitigation of infrastructure failure. This research set out to meet these needs through a concerted data collection effort, statistical analysis, and the application of predictive analytics. The developed approaches continue to be enhanced, but have proven to be robust in their ability to predict the age at which a platform is removed.

4.1 EXPLORATORY DATA ANALYSIS

The results from the exploratory data analysis identified strong correlations between several of the predictor variables. Many of these were specific to the structural components and included slot count, slant slot count, slot drill count, rig count, and crane count. Furthermore, these correlations varied by structure. As a whole, there were far more fixed structures operating in the GoM and the use of a full infrastructure dataset that includes all structure types may obfuscate the relationships between MOPU and the factors involved in those structures ages. However, it is important to recognize that MOPU structure types are inherently mobile. When modeling age at removal, the fact that these structures can more easily be removed may bias the results. That being said, future model iterations may need to consider structure integrity by type. For now, the lack of data on important variables of interest (i.e. incidents) prevents that analysis from coming to fruition.

Incident data were critical for the conceptualization of integrity and will likely remain an important variable going forward. It is important to note that the correlations between structural, environmental, and incident variables showed promise. The development and integration of the incident data was difficult and was heavily dependent on the team's ability to connect ancillary information across multiple sets and sources to match incidents to specific platforms and rigs.

The exploratory spatial regression model also relied on incident data, among other things, and took the correlation analysis one step further by focusing on the combined effects of several variables of interest. Unfortunately, the model was plagued by multicollinearity which forced a step back to remove several correlated variables. Still, the regression results were telling and corroborated the evidence presented through the correlation analysis—that platform types and subtypes are a significant predictor of age at removal. Specifically, all platforms that are not fixed are removed earlier than their fixed counterparts. The spatial regression also identified a significant negative relationship with wave height. This suggests that structures exposed to taller waves are, on average, removed earlier. This is an important finding as it aligns with findings from previous studies (Guédé, 2019) and supplied support for its inclusion in later model iterations (GWR).

4.2 PREDICTIVE ANALYTICS

The ML algorithms were powerful tools for evaluating the integrity of the existing offshore platforms. Using the structural characteristics, historic incidents, and MetOcean conditions throughout a platform's life proved to be sufficient to predict the age at removal range of a platform. These results are in line with those found in statistical tests. Performance for GBC and

ANN models, at 85.6% and 84.3% respectively, is satisfactory considering how complex these offshore infrastructure systems are.

While the prediction accuracies are suitable for the goal of this task, there is still room for improvement. Further work can be done to reduce the variance (overfitting) and also increase the prediction accuracy for both models. Additional data can be fed into the models to make them more robust. For example, a more comprehensive MetOcean database with better temporal representation or additional incident records could be included. This may allow the algorithms to better model the effects of physical and environmental conditions on infrastructure lifespan. A complete analysis on these efforts will be necessary to fully develop understanding of the power of ML algorithms to predict the lifespan of offshore platforms. In turn, this may also inform the ability to predict the lifespan of other types of offshore infrastructure (pipelines, wells, risers).

The GWR model identified a clear distinction between eastern and western platforms. Figure 13 presented this relationship with respect to the estimated coefficient values for the total number of days a platform endured a Category 4 hurricane throughout its lifespan. This pattern is also exemplified when plotting the coefficient values for the other parameter estimates, although the relationship is not as strong (Appendix F). For the Category 4 hurricane variable, the coefficients increase as one moves from the east to the west. Furthermore, this relationship remains positive across the GoM. In general, this means that platforms have a positive association with hurricanes which, at first blush, is not intuitive. Hurricanes should obviously reduce the lifespan of a platform if it is assumed that hurricanes exacerbate rates of integrity through structural stress. Another interpretation is that structures in the eastern GoM have a shorter lifespan as a result of their interaction with Category 4 hurricanes. It could very well be that Category 4 hurricanes impact structures in the eastern GoM more frequently and more directly, such that it is harder to fully recover or make the necessary repairs before the next hurricane occurs. Moreover, the wear-and-tear from hurricanes could be a factor in earlier removals.

Although the mean value for max wave power and max wave period is positive in Table 9, the values range from negative to positive at the level of the individual platform. Plotting the values (Appendix F) for those variables reveal several clusters where these values are associated with negative coefficient estimates for the platforms indicating a decrease in age at removal. In particular, the platforms in the eastern GoM are expected to be removed earlier than the platforms in the western GoM for these variables. Furthermore, the pattern for wave height is even more revealing, showing that wave height is associated with a decrease in removal age for most of the significant coefficient values in the GoM. That is, higher waves decrease the age at removal. Although this is true for most of the GoM, there are several clusters of platforms in the near-shore environment where wave height is associated with a positive removal age. These model results suggest that in the north-central GoM platforms are removed earlier than platforms in the western GoM, perhaps due to more harsh environmental conditions. It should be noted that extreme waves are not typically seen in shallow coastal waters.

Interestingly, all of the predictive models picked up on some form of hurricane variable as an important factor in determining age at removal. The ML models have the benefit of using as much information as possible when making predictions whereas the GWR model is subject to the same deficiencies as ordinary least square regression techniques. That is to say, GWR must be selective in the variables used in the model. Yet, the fact that each model capitalized on some form of hurricane days is telling from a predictive standpoint underpinning the importance of further exploration of this variable in future model development. One notable caveat concerning the hurricane variables is that they are highly correlated with one another. As a result, the use of all hurricane variables—while useful from a predictive standpoint—is not necessarily adding a substantial amount of *new* information from a planning or policy point of view. There is clearly a relationship between removal age and hurricane interaction, but this relationship appears to be similar across all variables. Thus, it may be fruitful to explore which of the hurricane variables is the most useful and then remove the others. Additionally, the spatial structure of hurricane response can be explored. For example, typically the strongest winds and the highest waves are found in the northeast quadrant of a hurricane. In addition to the stronger winds this region of a hurricane may favor wave growth when the hurricane moves at comparable speed relative to waves, causing the fetch region to increase (dynamic fetch).

Another interesting finding was that the ANN picked up on longitude as one of the top 5 strongest predictors of age at removal. The GWR corroborates this result to a large extent with its identification of significant east to west variation when considering the relationship between factors (variables) and their removal age. Moreover, although longitude is not noted in Table 9 for the GBC, it was routinely identified as a strong predictor. Whether this pattern is an indication of lower integrity for eastern GoM structures, or whether it reflects the spatial structure of MetOcean extreme events, is yet to be fully understood, but this evidence does provide some useful insights into where integrity may be compromised at an early stage of a platform's lifespan.

There is still some work to be done regarding the predictive power and associated relationships between the MetOcean variables and the age at removal variable. Neither GBC, ANN, or GWR identified the same MetOcean variables to be the strongest predictors. Where the GWR is concerned, this is partly due to the threat of multicollinearity. Many of the MetOcean variables showed a moderate to high degree of correlation, so the selection was based on minimizing the possibility of variance inflation. As a result, only three MetOcean variables were included in the GWR model.

As for the ANN and GBC, the ANN identified only hurricane variables as the strongest predictive factors while the GBC was more diverse. Specifically, the GBC identified the most diverse set of predictors that spanned both structural (rig count and underwater completion count), environmental (max sustained wind speed), and hurricane related variables. This is possibly due to the GBC pre-processing step to remove highly correlated predictors which reduced the amount of collinearity and allowed for more variance among the set of predictors. Taken together, these results underpin the importance of considering a constellation of theoretically viable factors in determining when a platform is removed. Although there is strong predictive power with hurricanes, there are clearly other factors involved that may add more substantive explanatory information to the models by offering more concrete guidance on what factors effect integrity.

5. CONCLUSIONS

The purpose of this project was to develop and apply big data analytics, ML, and advanced spatio-temporal analysis to assess the current state of existing infrastructure. In this first phase of research, the team uncovered several important factors related to the integrity of offshore infrastructure. This included both structural and environmental correlates, as well as other information on tropical storms and hurricanes. Furthermore, this work supports the continued exploration of integrity from several vantage points. Here, three predictive models with varying strengths and weaknesses were developed. Although different, these models work together to corroborate findings and support the re-evaluation of contradictory information. This happened in several instances which has made this investigation into offshore integrity more robust. Several important strides were made through this work, that will support a deeper understanding of offshore integrity and the factors that effect it.

First, to the best of the research team's knowledge there is no publicly available dataset related to offshore infrastructure integrity. This is likely due to the fact that the integrity of individual structures is proprietary information. The costs of collecting the data using advanced monitoring is perhaps one reason for this deficiency. Moreover, a dataset that spanned all GoM structures is even less likely. To that extent, this report has outlined a possible method for investigating offshore integrity across a broad geographic range using a unique set of variables that are hypothesized to relate to integrity. What's more, this method is not unique to the GoM and could theoretically be deployed to any offshore (or even onshore) location in the world. With this information in hand, the state of offshore infrastructure can be made more transparent, enabling more proactive approaches to impact mitigation, response planning, and infrastructure management.

Second, the correlation analysis was used to investigate the relationship between the components associated with each structure and a measure of integrity *visa vis* incidents and incident severity. Multiple structural factors were significantly correlated with integrity; however, the correlation analysis also revealed that several of these structural components were correlated with one another. These results informed later models and identified pathways toward more parsimonious model formulations.

Third, the predictive models developed in this work show a respectable degree of accuracy when it comes to predicting the age at which platforms are removed. Admittedly, a lot of that predictive power comes from only a few variables, several of which are related to hurricanes and storms. Conceptually, the interaction between storms and age at removal makes sense; more storms put higher stress on structural components which leads to earlier removal. However, it would also be true that as structures age they will inherently experience more storms. At this point it cannot be determined from this analysis with certainty whether tropical storms are *causing* the platforms to be removed. If in the model tropical storms are contributing to the removal, it would be prudent to understand the significance of that role in relation to other potential factors that contribute to the deterioration of integrity.

When it comes to next steps, there are many. The continued identification of data that can be leveraged to explain integrity is perhaps the most prudent. This includes the ongoing work of translating historical structure incidents into a useable data format and adding well and geohazard information to the analysis for a more system-wide approach. The current models rely heavily on a small subset of factors that may or may not be *causing* a decrease in integrity.

Future recommended research includes causal relations, for example with statistical tests such as Granger Causality (e.g. Runge et al., 2019), or through information theory estimates (San Liang, 2014, 2015). Much of this future work will rely on the ability to collect and integrate new data and sources of information, but it will also depend on the ability to enhance the measure of integrity should there be a need in the future.

This page intentionally left blank.

6. REFERENCES

Abdi, H. The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics*; Sage: Thousand Oaks, CA, 2007; p. 508–510.

Aeran, A.; Siriwardane, S. C.; Mikkelsen, O.; Langen, I. A framework to assess structural integrity of ageing offshore jacket structures for life extension. *Marine Structures* **2017,** *56*, 237–259. https://doi.org/https://doi.org/10.1016/j.marstruc.2017.08.002

Almedallah, M. K.; Walsh, S. D. C. A numerical method to optimize use of existing assets in offshore natural gas and oil field developments. *Journal of Natural Gas Science and Engineering* **2019a,** *67*, 43–55. https://doi.org/https://doi.org/10.1016/j.jngse.2019.04.012

Almedallah, M. K.; Walsh, S. D. C. Integrated well-path and surface-facility optimization for shallow-water oil and gas field developments. *Journal of Petroleum Science and Engineering* **2019b,** *174*, 859–871. https://doi.org/https://doi.org/10.1016/j.petrol.2018.11.025

Animah, I.; Shafiee, M. Development of a condition index matrix to support technical feasibility of life extension in the offshore oil and gas industry. International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, 2016; p. 150–158.

API. API Recommended Practice 2A-WSD: Planning, Designing, and Constructing Fixed Offshore Platforms—Working Stress Design. American Petroleum Institute, Washington, DC, 2014.

Berek, E. P.; Cooper, C. K.; Driver, D. B.; Heideman, J. C.; Mitchell, D. A.; Stear, J. D.; Vogel, M. J. Development of Revised Gulf of Mexico Metocean Hurricane Conditions for Reference by API Recommended Practices. Offshore Technology Conference, 2007; p. 13. https://doi.org/10.4043/18903-MS

Bhandari, J.; Abbassi, R.; Garaniya, V.; Khan, F. Risk analysis of deepwater drilling operations using Bayesian network. *Journal of Loss Prevention in the Process Industries* **2015,** *38*, 11–23.

BOEM. 2016 Update of Occurrence Rates for Offshore Oil Spills. Bureau of Ocean Energy Management, Arlington, VA, 2016.

BOEM. Offshore Statistics by Water Depth. Bureau of Ocean Energy Management, Leasing Data Center. https://www.data.boem.gov/Leasing/OffshoreStatsbyWD/Default.aspx (accessed June 6, 2019).

Bozeman, B. The 2010 BP Gulf of Mexico oil spill: Implications for theory of organizational disaster. *Technology in Society* **2011,** *33*, 244–252.

Breiman, L. Random Forests. *Machine Learning* **2001,** *45*, 5–32. https://doi.org/10.1023/A:1010933404324

BSEE. *Integrity Management Process of Tension Leg Platforms*; BSEE Project Number: E17PC00018; Bureau of Safety and Environmental Enforcement: Sterling, VA, 2018.

BSEE. Offshore Incident Statistics. Bureau of Safety and Environmental Enforcement, 2019a. https://www.bsee.gov/stats-facts/offshore-incident-statistics (accessed Feb 21, 2020).

BSEE. Platform/Rig Information. Bureau of Safety and Environmental Enforcement, 2019b. https://www.data.bsee.gov/Main/Platform.aspx

Burgherr, P. In-depth analysis of accidental oil spills from tankers in the context of global spill trends from all sources. *Journal of Hazardous Materials* **2007,** *140*, 245–256. https://doi.org/https://doi.org/10.1016/j.jhazmat.2006.07.030

Chadwell, J. Incidents Associated with Oil and Gas Operations: Outer Continental Shelf 1995-1996; 1996. https://www.bsee.gov/sites/bsee.gov/files/incident-summaries/safety/incidentsassociatedwithoilandgasoperationsocs95-96-pdf.pdf

Chadwell, J.; Blundon, C.; Anderson, C. Incidents Associated with Oil and Gas Operations: Outer Continental Shelf 1997; 1998. https://www.bsee.gov/sites/bsee.gov/files/incident-summaries/safety/finalocs97-pdf.pdf

Chadwell, J.; Blundon, C.; Anderson, C.; Cacho, M. *Incidents Associated with Oil and Gas Operations: Outer Continental Shelf 1998*; Washington, DC, 2000.

Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995,** *20*, 273–297. https://doi.org/10.1007/BF00994018

Cotton, U. Accidents Associated With Oil and Gas Operations: Outer Continental Shelf 1991-1994; 1994. https://www.bsee.gov/sites/bsee.gov/files/incident-statisticssummaries-fatalities/exploration-and-production/ocsincidents1991to1994-pdf.pdf

DHSG. *Final Report on the Investigation of the Macondo Well Blowout;* Deepwater Horizon Study Group: Berkley, 2011.

EIA. *Federal Offshore--Gulf of Mexico Field Production of Crude Oil*; U.S. Energy Information Administration, 2019. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=mcrfp3fm2&f=m (accessed June 17, 2019).

Elkatatny, S. M.; Tariq, Z.; Mahmoud, M. A.; Al-AbdulJabbar, A. Optimization of Rate of Penetration using Artificial Intelligent Techniques. 51st U.S. Rock Mechanics/Geomechanics Symposium, 2017; p. 8. https://doi.org/

Enverus. 2020. https://www.enverus.com/industry/exploration-and-production/

Enverus DrillingInfo. 2020. DrillingInfo API, Producing Entities, Producing Entity Details, Injection Entity Details, and Well Rollups. https://app.drillinginfo.com/direct/#/

Ersdal, G.; Selnes, P. O. Life extension of ageing petroleum facilities offshore. Society of Petroleum Engineers International Conference, 2010.

Garcia, H. E.; Boyer, T. P.; Locarnini, R. A.; Antonov, J. I.; Mishonov, A. V.; Baranova, O. K., … Levitus, S. World Ocean Atlas 2019, Dissolved oxygen, Apparent Oxygen Utilization, and Oxygen Saturation; Mishonov, A., Technical, Ed.; NOAA Atlas NESDIS, 2019a; Vol 3 p. 38.

Garcia, H. E.; Locarnini, R. A.; Boyer, T. P.; Antonov, J. I.; Baranova, O. K.; Zweng, M. M.; … Levitus, S. World Ocean Atlas 2019, Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate); Mishonov, A., Technical Ed.; NOAA Atlas NESDIS, 2019b; Vol 4, p. 35.

Graham, B.; Reilly, W. K.; Beinecke, F.; Boesch, D. F.; Garcia, T. D.; Murray, C. A.; Ulmer, F. Deep water: The Gulf Oil disaster and the future of offshore drilling; Report to the President; Washington, DC, 2012.

Guédé, F. Risk-based structural integrity management for offshore jacket platforms. *Marine Structures* **2019,** *63*, 444–461.

Guedes Soares, C.; Garbatov, Y.; Zayed, A. Effect of environmental factors on steel plate corrosion under marine immersion conditions. *Corrosion Engineering, Science and Technology* **2011,** *46*, 524–541.

Hawkins, E. BSEE Personal communication. Albany, OR, 2019.

Herbich, J. B. *Handbook of Coastal Engineering*; 2000.

Ho, T. K. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995; Vol 1, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Hoover, M. J. *Incidents Associated with Oil and Gas Operations: Outer Continental Shelf 2000*; OCS Report, MMS 2002-016; 2002. https://www.bsee.gov/sites/bsee.gov/files/reports/reports/accidentreport2000-march25-pdf.pdf

ISO. General principles on reliability for structures. International Standards Organization, 2015.

ISO. Petroleum and Natural Gas Industries -- Fixed Steel Offshore Structures; International Standards Organization, 2011.

Jablonowski, C. J. Employing detection controlled models in health and environmental risk assessment: A case in offshore oil drilling. *Human and Ecological Risk Assessment: An International Journal* **2007,** *13*, 986–1013.

Jain, A. K.; Mao, J.; Mohiuddin, K. M. Artificial neural networks: A tutorial. *Computer* **1996,** *29*, 31–44.

Jenks, G. F.; Caspall, F. C. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers* **1971,** *61*, 217–244.

Kaiser, M. J.; Pulsipher, A. G. A review of the oil and gas sector in Kazakhstan. *Energy Policy* **2007,** *35*, 1300–1314.

Kim, T. K. T test as a parametric statistic. *Korean Journal of Anesthesiology* **2015,** *68*, 540.

Locarnini, R. A.; Mishonov, A. V.; Baranova, O. K.; Boyer, T. P.; Zweng, M. M.; Garcia, H. E.; … Smolyar, I. World Ocean Atlas 2019, Volume 1: Temperature. A. Mishonov Technical Ed.; NOAA Atlas NESDIS, 2019; Vol 1, pp. 52.

Malhotra, P.; Vig, L.; Shroff, G.; Agarwal, P. Long short term memory networks for anomaly detection in time series. *Proceedings* **2015,** *89*, 89–94. Presses universitaires de Louvain.

Matsushima, I. Carbon steel - Corrosion by seawater. In Uhlig's Corrosion Handbook (Third); Revie, R., Ed.; John Wiley and Sons: Hoboken, NJ, 2011.

McDonald, J. H. *Handbook of Biological Statistics*; Sparky House Publishing: Baltimore, MD, 2009; Vol. 2.

Mei, W.; Primeau, F.; McWilliams, J.C.; Pasquero. C. Sea surface height evidence for long-term warming effects of tropical cyclones on the ocean. *Proceedings of the National Academy of Sciences* **2013,** *38*, 15207–15210.

Melchers, R. E. Principles of marine corrosion. In *Springer Handbook of Ocean Engineering*, 1st ed.; Dhanak, M. R., Xiros, N. I., Eds.; 2016; pp. 111–126. https://doi.org/10.1007/978-3-319-16649-0

Melchers**,** R. E. The effect of corrosion on the structural reliability of steel offshore structures. *Corrosion Science* **2005,** *47*, 2391–2410.

MMS. Accidents Associated with Oil and Gas Operations: Outer Continental Shelf 1956-1990. Minerals Management Service, 1990. https://www.bsee.gov/sites/bsee.gov/files/reports/blowout-prevention/ocsincidents1956to1990-pdf.pdf

MMS. Oil and Gas and Sulphur Operations in the Outer Continental Shelf—Incident Reporting Requirements. Minerals Management Service, 2006.

Moan, T. Life cycle structural integrity management of offshore structures. *Structure and Infrastructure Engineering* **2018,** *14*, 911–927. https://doi.org/10.1080/15732479.2018.1438478

Muehlenbachs, L.; Cohen, M. A.; Gerarden, T. The impact of water depth on safety and environmental performance in offshore oil and gas production. *Energy Policy* **2013,** *55*, 699–705.

Nabavian**,** N.; Morshed, A. Extending Life of Fixed Offshore Installations by Integrity Management: A Structural Overview (No. 138386). Abu Dhabi; 2010.

Nichols, N. W.; Khan, R. Structural integrity management system (SIMS) implementation within PETRONAS' operations. *Journal of Marine Engineering & Technology* **2015,** *14*, 61–69. https://doi.org/10.1080/20464177.2015.1096612

Nunez, M. Prevention of Metal Corrosion: New Research; Nova Publishers, 2007.

O'Connor, P. E.; Bucknell, J. R.; DeFrance, S. J.; Westlake, H. S.; Westlake, F. J. Structural Integrity Management (SIM) of Offshore Facilities. Offshore Technology Conference, 2005; p. 11. https://doi.org/10.4043/17545-MS

Onalo, D.; Adedigba, S.; Khan, F.; James, L. A.; Butt, S. Data driven model for sonic well log prediction. *Journal of Petroleum Science and Engineering* **2018,** *170*, 1022–1037.

Pham, T.; Tran, T.; Phung, D.; Venkatesh, S. Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of Biomedical Informatics* **2017,** *69*, 218–229.

Quinlan, J. R. Induction of decision trees. *Machine Learning* **1986,** *1*, 81–106. https://doi.org/10.1007/BF00116251

Reader, T. W.; O'Connor, P. The Deepwater Horizon explosion: non-technical skills, safety culture, and system complexity. *Journal of Risk Research* **2014,** *17*, 405–424. https://doi.org/10.1080/13669877.2013.815652

Rim-Rukeh, A. Oil spill management in Nigeria: SWOT analysis of the joint investigation visit (JIV) process. *Journal of Environmental Protection* **2015,** *6*, 259.

Romeo, L.; Dyer, A.; Duran, R.; Wenzlick, M.; Nelson, J.; Sabbatino, M.; Wingo, P.; Bauer, J.; Rose, K. *Offshore Platform, Production, Incident, MetOcean, and Geohazard Spatial Dataset*. In preparation, 2021.

Romeo, L.; Nelson, J.; Wingo, P.; Bauer, J.; Justman, D.; Rose, K. Cumulative spatial impact layers: A novel multivariate spatio-temporal analytical summarization tool. *Transactions in GIS* **2019a,** *23*, 908–936. https://doi.org/10.1111/tgis.12558

Romeo, L.; Wingo, P.; Nelson, J.; Bauer, J.; Rose, K. Cumulative Spatial Impact Layers, 2019b. https://edx.netl.doe.gov/dataset/cumulative-spatial-impact-layers, DOI: 10.18141/1491843

Rose, K.; Bauer, J. R.; Mark-Moser, M. A systematic, science-driven approach for predicting subsurface properties. *Interpretation*. **2020,** *8*, T167-81.

Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; … Muñoz-Marí, J. (Inferring causation from time series in Earth system sciences. *Nature Communications* **2019,** *10*, 1–13.

San Liang, X. Normalizing the causality between time series. *Physical Review E* **2015,** *92*, 22126.

San Liang, X. Unraveling the cause-effect relation between time series. *Physical Review E* **2014,** *90*, 52150.

Sharp, J.; Ersdal, G.; Galbraith, D. Meaningful and Leading Structural Integrity KPIs. SPE Offshore Europe Conference and Exhibition. Society of Petroleum Engineers, 2015.

Shultz, J. R.; Fischbeck, P. Predicting Risks Associated with Offshore Production Facilities: Neural Network, Statistical, and Expert Opinion Models. SPE/EPA Exploration and Production Environmental Conference. Society of Petroleum Engineers, 1999.

Solland, G.; Sigurdsson, G.; Ghosal, A. Life extension and assessment of existing offshore structures. SPE Project and Facilities Challenges Conference at METS. Society of Petroleum Engineers, 2011.

Soom, E. M.; Husain, M. K. A.; Zaki, N. I. M.; Nor, M. N. K. M.; Najafian, G. Lifetime Extension of Ageing Offshore Structures by Global Ultimate Strength Assessment (GUSA). *Malaysian Journal of Civil Engineering* **2018,** *30*.

Stacey, A.; Birkinshaw, M.; Sharp, J. V. Life Extension Issues for Ageing Offshore Installations 2008; pp. 199–215. http://dx.doi.org/10.1115/OMAE2008-57411

Tobler, W. R. A computer movie simulating urban growth in the Detroit region. *Economic Geography* **1970,** *46*, 234–240.

Tygesen, U. T.; Worden, K.; Rogers, T.; Manson, G.; Cross, E. J. State-of-the-Art and Future Directions for Predictive Modelling of Offshore Structure Dynamics Using Machine Learning BT - Dynamics of Civil Structures; Pakzad, S., Ed.; Cham: Springer International Publishing, 2019; Vol 2.

Visser, R. C. Offshore Accidents, Regulations and Industry Standards. SPE Western North American Region Meeting 2011, p. 9. https://doi.org/10.2118/144011-MS

Worden, K.; Rogers, T.; Cross, E. J. Identification of Nonlinear Wave Forces Using Gaussian Process NARX Models BT - Nonlinear Dynamics; Kerschen, G., Ed.; Cham: Springer International Publishing, 2017; Vol 1.

Yang, D. Y.; Frangopol, D. M. Evidence-based framework for real-time life-cycle management of fatigue-critical details of structures. *Structure and Infrastructure Engineering* **2018,** *14*, 509–522.

Zhang, S. When a Hurricane Hits an Offshore Oil Platform. The Atlantic, 1. https://www.theatlantic.com/science/archive/2017/08/harvey-offshore-platform-oil-gas/537960/ (access Aug 2017).

Zweng, M. M.; Reagan, J. R.; Antonov, J. I.; Locarnini, R. A.; Mishonov, A. V.; Boyer, T. P. … Seidov, D. World Ocean Atlas 2019; Mishonov, A., Technical Ed.; NOAA Atlas NESDIS 2019; pp. 50. Vol 2.

APPENDIX A: INFRASTRUCTURE AND INCIDENT DATA USED IN THE ANALYSIS

Table includes information on source, and temporal resolution per dataset.

Type	Name	Source	Attributes	Temporal Resolution
Infrastructure	Platform and Rig Information	BSEE	Structure type, status, location, equipment, dates	Records from 1942–present Updated monthly, last acquired December 9, 2019
Incidents	Incident Stats and Summaries Archive	BSEE		Annual reports from 2006 – 2013 Last updated January 17, 2020
Incidents	USCG Incident Investigation Report Database	USCG (https://cgmix.uscg.mil/IIR/IIRSearch.aspx)		Reports from 2002 to present Updated monthly, last acquired March 2, 2020

This page intentionally left blank.

APPENDIX B: METOCEAN DATA BY VARIABLE WITH INFORMATION ON SPATIAL RESOLUTION, EXTENT, TEMPORAL RESOLUTION, AND SOURCE

Variable	Spatial Resolution and Extent	Temporal Resolution and Range
Ocean currents at the sea surface and at depth from HyCOM (U.S. Navy operational ocean model).	Horizontal ~4 km, depth=0M GoM	Every 3 hours, 2003–2019
IBTrACS–Worldwide storms	Varies Global coverage	1842–2019, resolution varies
WAVEWATCH III® 30-year Hindcast–Peak Wave Period, Wave Period, mean wave direction, significant wave height and the wind used to force the model	~16 km GoM	Every 3 hours, 1979–2009
WAVEWATCH III® Production Hindcast–Peak Wave Period, Wave Period, mean wave direction, significant wave height and the wind used to force the model	~16 km GoM	Every 3 hours, 2005–2019

This page intentionally left blank.

APPENDIX C: VARIABLES USED TO MODEL AMBIENT CONDITIONS THAT HAVE BEEN FOUND TO CONTRIBUTE TO THE CORROSION OF METAL IN AN OFFSHORE ENVIRONMENT

Variable	Horizontal Resolution	Time Coverage Start	Time Coverage End	Units	Reference
Nitrate	1 degree	1900-01-01	2017-12-31	Micro moles per kilogram of sea water	(Garcia et al., 2019b)
Dissolved oxygen	1 degree	1900-01-01	2017-12-31	Micro moles per kilogram of sea water	(Garcia et al., 2019a)
Phosphate	1 degree	1900-01-01	2017-12-31	Micro moles per kilogram of sea water	(Garcia et al., 2019b)
Silicate	1 degree	1900-01-01	2017-12-31	Micro moles per kilogram of sea water	(Garcia et al., 2019b)
Salinity	0.25 degree	1955-01-01	2017-12-31	unitless	(Zweng et al., 2019)
Temperature	0.25 degree	1955-01-01	2017-12-31	Degrees Celsius	(Locarnini et al., 2019)

This page intentionally left blank.

APPENDIX D: VARIABLES PICKED BY THE GBC MODEL FOR TRAINING THAT HAVE A SIGNIFICANT IMPACT ON THE PREDICTIVE POWER

The parentheses indicate the categorical value of the field converted to a binary variable.

Latitude	Abandon Flag (N)	LACT meter flag (N)
Longitude	Condensate production flag (N)	Maximum sustained wind speed max
Crane count	Injection Code (G)	Maximum sustained wind speed average
Slot drill count	Heliport Flag (N)	Oxygen mean 65% depth
District code	Production equipment flag (N)	Oxygen mean 25% depth
Distance to shore	Power source type (Diesel)	Oxygen mean surface
Rig count	Compressor Flag (N)	Oxygen std. dev. 65% depth
Water depth	Commingling production flag (N)	Oxygen std. dev. 25% depth
Deck count	Number of incidents	Phosphorus mean 65% depth
Underwater completion count	Cumulative incident severity	Salinity mean 65% depth
MMS company number	Average age at incident	Salinity std. dev. 65% depth
Bed count	Total hurricane impacts	Salinity std. dev. 25% depth
Structure type code (CAIS)	Tropical storm days max yearly	Salinity std. dev. Surface
Structure type code (WP)	Category 1 hurricane max yearly	Silicate std. dev. Surface
Authority status (PROD)	Category 2 hurricane max yearly	Temperature mean 65% depth
Authority status (TERMIN)	Category 3 hurricane max yearly	Temperature mean 25% depth
	Wave count	Temperature std. dev. surface

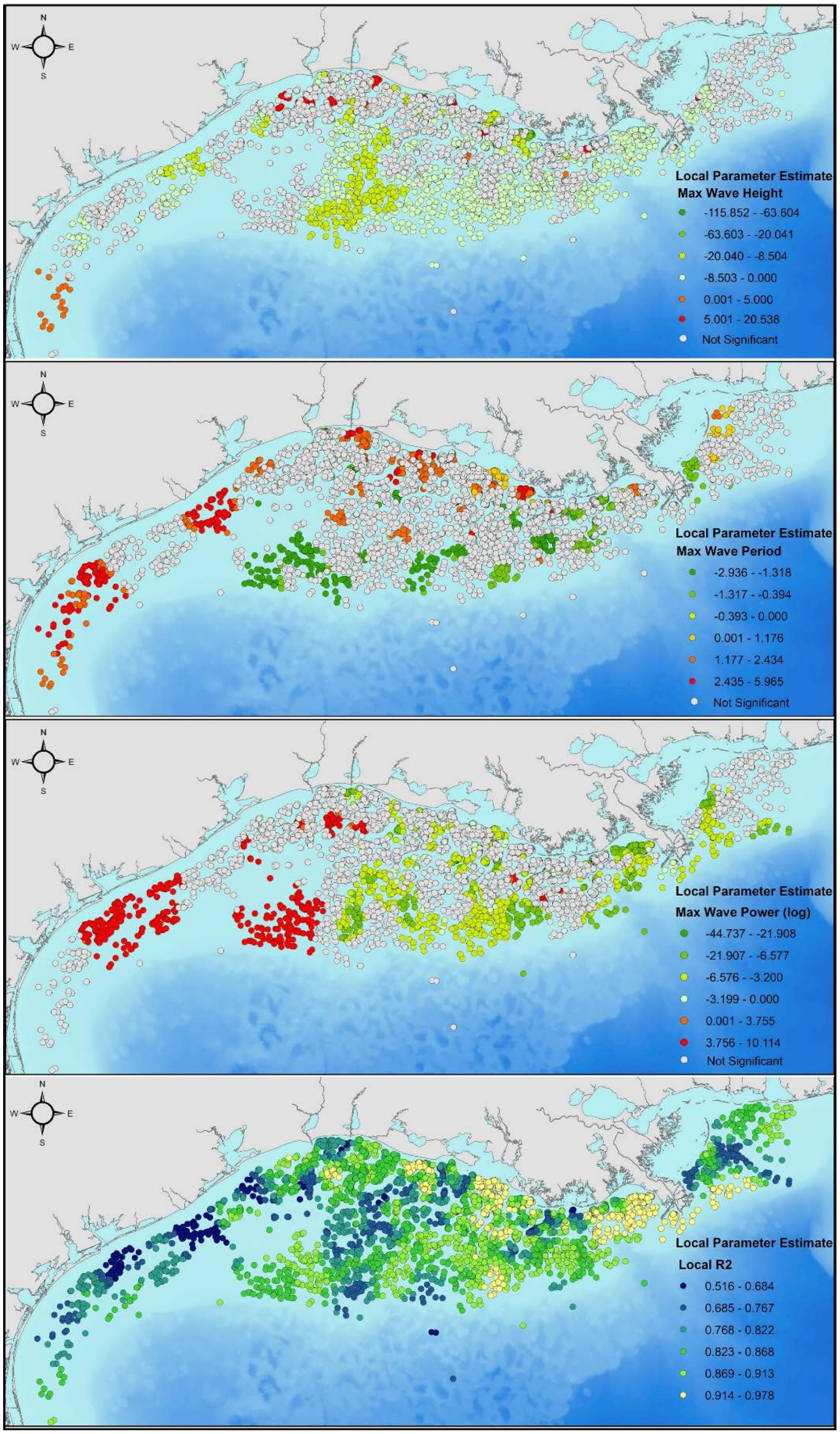
This page intentionally left blank.

APPENDIX E: VARIABLES USED IN THE ANN MODEL FOR TRAINING

LATITUDE	LONGITUDE	CraneCount
SlotDrillCount	DISTRICTCODE	Severity_CurrentScale
Avg_Age_at_Inc	DistanceToShore	NumberofIncidents
SlotCount	RigCount	SlantSlotCount
WaterDepth	DeckCount	SatelliteCompletionCount
UnderwaterCompletionCount	MmsCompanyNum	BedCount
TotalStorms	CatNoneCount	TropicalSum
TropicalMinYearly	TropicalMaxYearly	TropicalMeanYearly
TropicalDaysSum	TropicalDaysMinYearly	TropicalDaysMaxYearly
TropicalDaysMeanYearly	C1Sum	C1MinYearly
C1MaxYearly	C1MeanYearly	C1DaysSum
C1DaysMinYearly	C1DaysMaxYearly	C1DaysMeanYearly
C2Sum	C2MinYearly	C2MaxYearly
C2MeanYearly	C2DaysSum	C2DaysMinYearly
C2DaysMaxYearly	C2DaysMeanYearly	C3Sum
C3MinYearly	C3MaxYearly	C3MeanYearly
C3DaysSum	C3DaysMinYearly	C3DaysMaxYearly
C3DaysMeanYearly	C4Sum	C4MinYearly
C4MaxYearly	C4MeanYearly	C4DaysSum
C4DaysMinYearly	C4DaysMaxYearly	C4DaysMeanYearly
C5Sum	C5MinYearly	C5MaxYearly
C5MeanYearly	C5DaysSum	C5DaysMinYearly
C5DaysMaxYearly	C5DaysMeanYearly	WaveHeightCount
WaveHeightNoneCount	WaveHeightSum	WaveHeightMin
WaveHeightMax	WaveHeightAverage	MSWSCount
MSWSNoneCount	MSWSSum	MSWSMin
MSWSMax	MSWSAverage	MRWGCount
MRWGNoneCount	MRWGSum	MRWGMin
MRWGMax	MRWGAverage	Nitrate_Mean65PercDepth
Nitrate_Mean25PercDepth	Nirate_MeanSurface	Nitrate_StdDev65PercDepth
Nitrate_StdDev25PercDepth	Nitrate_StdDevSurface	Oxygen_Mean65PercDepth
Oxygen_Mean25PercDepth	Oxygen_MeanSurface	Oxygen_StdDev65PercDepth

Oxygen_StdDev25PercDepth	Oxygen_StdDevSurface	Phosphorus_Mean65PercDepth
Phosphorus_Mean25PercDepth	Phosphorus_MeanSurface	Phosphorus_StdDev65PercDepth
Phosphorus_StdDev25PercDepth	Phosphorus_StdDevSurface	Salinity_Mean65PercDepth
Salinity_Mean25PercDepth	Salinity_MeanSurface	Salinity_StdDev65PercDepth
Salinity_StdDev25PercDepth	Salinity_StdDevSurface	Silicate_Mean65PercDepth
Silicate_Mean25PercDepth	Silicate_MeanSurface	Silicate_StdDev65PercDepth
Silicate_StdDev25PercDepth	Silicate_StdDevSurface	Temp_Mean65PercDepth
Temp_Mean25PercDepth	Temp_MeanSurface	Temp_StdDev65PercDepth
Temp_StdDev25PercDepth	Temp_StdDevSurface	StructureTypeCode
MajorStructureFlag	AuthorityStatus	AbandonFlag
AlocMtrFlag	Attended8HrFlag	CondnProdFlag
DrillingFlag	FiredVesselFl	GasProdFlag
GasFlaringFlag	Manned24HrFl	MajCmplxFlag
LactMtrFlag	InjectionCode	HeliportFlag
WorkoverFlag	WaterProdFlag	TankGaugeFlag
SulfurProdFlag	StoreTankFlag	QtrType
ProdEqmtFlag	ProductionFlag	PowerSourceType
PowerGenFlag	OilProdFlag	GasSaleMtrFlag
CompressorFlag	ComglProdFlag	

APPENDIX F: LOCAL PARAMETER ESTIMATES FOR THE GWR MODEL VARIABLES OF INTEREST ALONG WITH THE LOCAL R2 VALUES (BOTTOM)



This page intentionally left blank.

Brian J. Anderson
Director
National Energy Technology Laboratory
U.S. Department of Energy

John Wimer
Deputy Director
Science & Technology Strategic Plans & Programs
National Energy Technology Laboratory
U.S. Department of Energy

Roy Long
Offshore Technology Manager
Science & Technology Strategic Plans & Programs
National Energy Technology Laboratory
U.S. Department of Energy

Elena Melchert
Director
Upstream Research Division Office of Oil and Gas
Office of Fossil Energy
U.S. Department of Energy

Bryan Morreale
Executive Director
Research & Innovation Center
National Energy Technology Laboratory
U.S. Department of Energy