
PIs: Lucy Romeo\textsuperscript{1,2} and Jennifer Bauer\textsuperscript{1}
Key Personnel: Aaron Barkhurst\textsuperscript{3}, Rodrigo Duran\textsuperscript{4}, Alec Dyer\textsuperscript{1,2}, Jake Nelson\textsuperscript{5}, Kelly Rose\textsuperscript{1}, Michael Sabbatino\textsuperscript{1,2}, Madison Wenzlick\textsuperscript{1,2}, Patrick Wingo\textsuperscript{1,2}, Dakota Zaengle\textsuperscript{1,2}

1. National Energy Technology Laboratory, Morgantown WV and Albany OR
2. NETL Support Staff, Morgantown WV and Albany OR
3. Mid-Atlantic Technology, Research & Innovation Center, Morgantown WV
4. Theiss Research, San Diego CA
5. Oak Ridge Institute of Science and Technology, Auburn AL

Offshore Unconventional Resources
Project Number DE FE-1022409
AIIM: Advanced Infrastructure Integrity Modeling
Applying smart, data-driven analytics to identify infrastructure opportunities and risks

Assessing Current and Future Infrastructure Hazards

- Built and cross-validated machine learning (ML) and advanced analytical models forecasting platform remaining lifespan & risk
- Develop, applied, and released AIIM framework data and results through NETL Common Operating Platform (COP)

Smart Models to Optimize Use or Reuse of Production & Transport Infrastructure

- Building off Task 3, analyze permanent, semi-permanent, high-risk infrastructure (platforms, pipelines)
- Build a smart modeling tool to evaluate long-term integrity
- Assess environmental & operational impacts for more environmentally prudent planning

Task 3: Completed, June 2021
Task 10: New Start, April 2021

A few definitions to set the stage...

- **Artificial Intelligence (AI)**
  - “Programmed” intelligence

- **Machine Learning (ML)**
  - **Supervised ML**, the machine is trained, taught
  - **Unsupervised ML**, the machine learns on its own (Google’s cat video experiment)

- **Big Data**
  - Large volumes, variety, variability, velocity of data

- **Big Data Computing**
  - Computing engineering & systems to handle big data
What are the needs?

- We are asking **more** from aging infrastructure
- Offshore operations can be hazardous
  - Environment
  - Operational wear and tear

Values Delivered

- Identify and address hazards
- Mitigate risk and cost
- Inform life extension & resource optimization strategies

Meteorological & Oceanographic (MetOcean) impacts

Integrity risks

Offshore drilling regulator warns of bolt failures in Gulf of Mexico

Need to prevent hazards & support response planning
Data-Driven Approach & Driving Insights

Leveraging **Big Data & Big Data Computing** of the whole to inform the local

**Data representing the natural-engineered offshore system**

- Structural platform records
- Incident reports
- Metocean* data
- Geohazard data
- Production information

**Predicting Integrity**

1. What is the remaining lifespan?
2. Likelihood of future risk?

Structure & Incident Records

Offshore platform structure types:

Sources include BOEMRE, BSEE, USCG, MMS

Evaluating infrastructure integrity based on:

- Location
- Use
- Operating condition
- Incident history

- Current velocity
- Wind speeds
- Wave height
- Storms
- External corrosion

Metocean Big Data

Data example: Density of storm occurrences by wind speed

>51,000 layers (>130 GB)

Sources include NOAA, US Navy, World Ocean Database, and external models

Exploratory Stats & Variable Analyses

- **Understanding the data**
- Found relationships between age, incident severity, structural complexity, and environment
- Key variable selection improved understanding of operational platform integrity

Potential plateau in the number of times a platform can sustain extreme hurricane conditions (Ordinary Least Squares regression)

Strong relationship (0.95) between hurricanes and removal age (Pearson's Correlation)

Machine Learning (ML) & Advanced Algorithms

Predicting Lifespan

1. Gradient Boosted Regression Tree
   Dyer et al. in review
2. Artificial Neural Network (ANN)
   Nelson et al. in review
3. Geographically Weighted Regression
   Dyer et al. in review

Risk Likelihood

• ANN Regression

Analyses through ML & advanced statistical models

Production information

Structural- & weather-related incidents

Publicly available structural information

Metocean variables

Wind Speed
Wave Height
Current Velocity

Geohazard data

Running **multiple models** allows us to better understand and **internally validate results**

Values Delivered
- Helps to locally inform use/reuse strategies
- Inform maintenance and monitoring decisions

Dyer et al., in review

Which attributes (features) had predicting power?

- Do they make sense?
- Do they agree between the models?
- How do they differ?

**GBRT (10 out of 23)**
1. Category 1 hurricane yearly mean
2. Category 3 hurricane yearly mean
3. Category 4 hurricane yearly mean
4. Min. central pressure, 25th percentile
5. Min. central pressure, std. dev.
6. Tropical storm yearly mean
7. Wave direction, 90th percentile
8. Mean meridional wind
9. Category 2 hurricane yearly mean
10. Category 1 hurricane max. days yearly

**ANN (10 out of 792)**
1. Category 1 hurricane yearly mean
2. Category 3 hurricane yearly mean
3. Category 4 hurricane yearly mean
4. Tropical storm yearly mean
5. Max. reported wind gust, 25th percentile
6. Category 2 hurricane yearly mean
7. Area code
8. Category 1 hurricane max. days yearly
9. Category 4 hurricane yearly max.
10. Zonal wind minimum

Example: *Using the whole to inform the local*

**B Fixed Platform**
Installed 1972 (~49 years old)

Both models found platform to be past the predicted removal age.

Poor Risk Communication, Inadequate Maintenance Behind Fatal Gulf of Mexico Platform Incident

OE Staff - March 12, 2021

Production since 1972

4 reported incidents
272 storms
11 tropical storms
53 hurricanes, including 2 C5

Oil Gas Water

Production since 1972

Models can adapt to answer a range of questions

Currently models are forecasting:
• Remaining lifespan (age at removal)
• Future Risk (incident severity)

Attribute breakdown within comprehensive dataset:

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>54%</td>
</tr>
<tr>
<td>Incident</td>
<td>18%</td>
</tr>
<tr>
<td>Structural</td>
<td>18%</td>
</tr>
<tr>
<td>Metocean</td>
<td>19%</td>
</tr>
<tr>
<td>Geohazard</td>
<td>3%</td>
</tr>
</tbody>
</table>

Modeling Framework

Given the right input data and training...
...models can predict other variables

Potential Outputs
• Environmental load preparedness
• Decommissioning
• Incidents
• Maintenance needs
• Production potential

Comprehensive dataset
Supplemental information
Pre-processing, training and scoring iterations
Altered hyperparameters
Learning something new
Outputs

Increasing Access to Data & Models
Through NETL’s Common Operating Platform

- Integrate ML, big data, and analytical outputs with online platform
- Platform securely released through Energy Data eXchange®
- Leverage award-winning Offshore Risk Modeling Suite
- Real-time analyses to better understand offshore infrastructure integrity

https://edx.netl.doe.gov/offshore
Synergy Opportunities
Internal collaboration and external interests

Potential additional external interests

Lessons Learned from Task 3

- **Big data, big data computing, and AI/ML** for solving complex problems
- **ML models** predicting existing platform longevity at **95–97% accuracy**
- **Full system analytics** enables us to **identify regional trends and explain our results locally**

**Next Steps – Task 10**

- Build off Task 3 and fill in identified information gaps
- Apply **time series** and **causal analytics**
- Develop **smart tool** for **rapid evaluation of infrastructure integrity**

[Image of ocean floor with infrastructure models and labels like FIXED, JACK-UP, SPAR, TLP, Fault, Reservoir, Multivariate Ambient, natural seeps, Mudslides, Wind, and Energy Data Exchange and Joule Watt logos.]

Task 10: Smart Models to Optimize Use or Reuse of Production & Transport Infrastructure

**Accomplishments**
- Kick-off, April 2021
- Completed literature review of >40 additional publications
- Acquired >1.2TB of data
- Began planning evaluation methods to include *causality testing* and *time series analytics*

**Lessons Learned**
- Big data ≠ better data
- Correlation ≠ causation
- Advanced models must be *explainable*

**Challenges**
- Historic data trapped in reports
- Access to maintenance and inspection records
- Limited equipment information

Disclaimer & Acknowledgement

• Disclaimer: This presentation was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed therein do not necessarily state or reflect those of the United States Government or any agency thereof.

• Acknowledgement: 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.
Questions?

Lucy.Romeo@netl.doe.gov
Jennifer.Bauer@netl.doe.gov

Thank you!
Applied **select** variables to **prevent multicollinearity**
- Singular value decomposition (SVD)
  1. Incident rates
  2. Phosphorus
  3. Salinity
  4. Slot drill count
  5. Surface magnitude
  6. Max. reported wind gust
  7. Category 5 hurricane rates
  8. Water depth

- Model R² ranged spatially from 0.17 – 0.91
- Found age at removal has **spatial non-stationarity**

Nelson et al., in review

Benefits of a Multi-model Approach

1. Evaluated available data
2. Identified key parameters
3. Assessed multiple approaches
4. Compare results for internal model validation

**Gradient Boosted Regression Tree**
- **Stagewise additive ML model** that optimizes model loss by adding weak learners (i.e. decision trees)
- **Strengths**
  - Handles outliers and able to perform when data is missing
- **Limitations**
  - Difficult to scale model
  - Requires careful tuning of parameters

**Artificial Neural Network**
- **Deep learning ML model** where data are passed through multiple layers that learn weights and biases, which are combined and transformed to learn the phenomena being modeled.
- **Strengths**
  - Handles complex and non-linear relationships
  - Handles categorical and continuous variables
  - Can be used for classification and regression
- **Limitations**
  - Easy to overfit
  - Difficult to interpret
  - Many parameters to tune

**Geographically Weighted Regression**
- **Spatial model** that fits regression equations to each feature (platform), based on dependent and explanatory variables of other local features
- **Strengths**
  - Captures local variation in spatial data
  - Can be used for prediction
- **Limitations**
  - Computational overhead increases with data volume
  - Local collinearity must be considered
  - Assumes linear relationships

*Models are only as good as the data going into them*

Cautions for big data, ML driven analytics

- **Correlation does not equal causation**
  - Just because you have an analysis doesn't mean the results are meaningful
- **Uncertainty is critical**
  - Capture, reduce if possible, represent, utilize, quantify
  - Not all data are equal or valuable

- Analytical methods must be appropriate for the **goal**
- Analytical methods must be appropriate for the **data**
- Analysis must be made in **context** of data collection

**The Parable of Google Flu: Traps in Big Data Analysis**

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (3, 4), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can predict x has become commonplace (5–7) and is often put in sharp contrast to the more traditional methods and hypotheses that underlie traditional epidemiology, but does not usually come with the same level of scrutiny and validation.

NETL Offshore

Common Operating Platform Viewer

The Common Operating Platform (COP) provides access to data and tools in a virtualized, EDX hosted environment.

Notice to Users

Disclaimer of Liability
Disclaimer of Endorsement

OK
Benefit to the Program

- **Tasks 3 and 10** supports the development of data, methods, models, and tools to support the evaluation of the current state of offshore oil and gas infrastructure in the U.S. Gulf of Mexico.

- **Task 3 (completed)** resulted in a novel analytical framework utilizing big data, big data computing, ML models, and an advanced geographic model to forecast the remaining lifespan and future risk of platforms. Resulting data and model outputs were hosted onto the newly developed NETL COP, also built as a result from this project.

- **Task 10 (new start)** will result in a multivariate, novel, and intelligent approach encompassing the full offshore natural-engineered system is needed to inform local use and reuse optimization strategies, minimize cost, and mitigate operational and environmental impacts. The application of big data, big data computing, causality testing, and ML will uncover key insights necessary for the safe and efficient offshore energy operations that are currently missing from existing approaches. Results from this project will improve predictive abilities to inform energy exploration and production strategies, prevent risk and promote safety, and help limit the already small FE offshore footprint while expanding energy resources.
Task 10 Project Overview

Goals and Objectives

- Funded by DOE as part of Offshore Unconventional Resources DE FE-1022409
- RSS Contract researchers
- Ongoing performance dates 2021-2023
- Project Participants
  - Federal: Jennifer Bauer (co-PI), Kelly Rose
  - LRST: Lucy Romeo (co-PI), Alec Dyer, Michael Sabbatino, Madison Wenzlick, Patrick Wingo, and Dakota Zaengle
  - Theiss Research: Rodrigo Duran, Ph.D.
Task 3
Assessing Current & Future Infrastructure Hazards

**Leads:** Jennifer Bauer, Lucy Romeo  
**Team:** Aaron Barkhurst, Rodrigo Duran, Alec Dyer, Jake Nelson, Michael Sabbatino, Madison Wenzlick, Patrick Wingo, Dakota Zaengle, and Kelly Rose

Task 10
Smart Models to Optimize Use or Reuse of Production & Transport Infrastructure

**Leads:** Jennifer Bauer, Lucy Romeo  
**Team:** Rodrigo Duran, Alec Dyer, Michael Sabbatino, Madison Wenzlick, Patrick Wingo, Dakota Zaengle, and Kelly Rose
<table>
<thead>
<tr>
<th>Number</th>
<th>Expected Completion Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.A</td>
<td>06/21</td>
<td>Complete a literature review on structural and external factors that influence offshore platform infrastructure operational capabilities, as well as past analytical techniques used to measure structural lifespan.</td>
</tr>
<tr>
<td>10.B</td>
<td>11/21</td>
<td>Acquire big, disparate data representing structural information, incident records, metocean variables, geohazard data, and well information. Integrate data into a comprehensive spatial dataset.</td>
</tr>
<tr>
<td>10.C</td>
<td>01/22</td>
<td>Complete preliminary analytics, including causality testing, on dataset to identify initial trends and patterns, as well as knowledge gaps. Identify potential advanced analytical techniques (machine learning [ML], geographic, statistical) for evaluating the current state of infrastructure for use and reuse optimization.</td>
</tr>
<tr>
<td>10.D</td>
<td>02/22</td>
<td>Outline technical report or manuscript covering data and methods for evaluating the current state of platform infrastructure for use and reuse optimization.</td>
</tr>
<tr>
<td>10.E</td>
<td>03/22</td>
<td>Determine if the preliminary analytics and current state of knowledge confirm that models can effectively analyze the integrity of offshore production infrastructure.</td>
</tr>
<tr>
<td>10.F</td>
<td>09/22</td>
<td>Develop analytical framework for a novel data-driven tool that uses smart tool logic to evaluate infrastructure.</td>
</tr>
<tr>
<td>10.G</td>
<td>10/22</td>
<td>Acquire data representing environmental and operational risk relating to exploration and production activities in the U.S. GOM.</td>
</tr>
<tr>
<td>10.H</td>
<td>02/23</td>
<td>Analyze potential environmental and operational impacts using existing tools from the ORM suite.</td>
</tr>
<tr>
<td>10.I</td>
<td>08/23</td>
<td>Provide demonstrable applications of novel tool for evaluating the current state of platform infrastructure, streamlined with environmental and operational impact assessments.</td>
</tr>
<tr>
<td>10.K</td>
<td>02/24</td>
<td>Submit technical report or manuscript for publication.</td>
</tr>
</tbody>
</table>
Research Products from Offshore Tasks 3 and 10

• Publications

• Data
Presentations from Offshore Tasks 3 and 10


- Rose, K., 2019. An Ounce of Prevention is Worth a Pound of Response, NETL’s Big Data Technologies for Offshore Spill Prevention. Addressing the nation's energy needs through technology innovation – 2019 carbon capture, utilization, storage, and oil and gas technologies integrated review meeting, August 26-30, 2019

