**Title**: Mapping threats to marine birds to inform conservation and management in Atlantic Canada

**Authors**: David Beauchesne1\*, Kevin Cazelles1, Stephanie Avery-Gomm2, April Hedd3, Philippe Muise4, Allison Patterson5, Robert Ronconi5, Katharine Studholme6, Sarah Wong6

###### 

**Author affiliations**:

1inSileco, Québec, Canada

2Environment and Climate Change Canada, Ottawa, ON, Canada

3Environment and Climate Change Canada, Mount Pearl, Newfoundland and Labrador, Canada

4Canada Centre for Mapping and Earth Observation, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada

5Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada

**Abstract**

Anthropogenic activities pose significant threats to marine ecosystems, with oil pollution, vessel traffic, and artificial light contributing to environmental risks across Canada’s Exclusive Economic Zone (EEZ) and adjacent waters. This study presents a comprehensive dataset of spatiotemporal threat layers, integrating a wide range of primary data sources to inform conservation, environmental response planning, and marine spatial management.

The dataset comprises geospatial layers developed from various primary datasets, including open access and proprietary data shared for this project. These layers quantify petroleum pollution, oil and gas infrastructure, vessel activity, and artificial light exposure across the annual cycle. Data sources include the Integrated Satellite Tracking of Pollution (ISTOP), National Aerial Surveillance Program (NASP), National Environmental Emergency Centre (NEEC), and Automatic Identification System (AIS) shipping data. Additionally, VIIRS satellite observations capture artificial light and flaring activity.

These spatiotemporal threat layers were developed using a reproducible workflows that integrate historical incident data and geospatial modeling. Petroleum pollution layers, for instance, evaluate the density of oiling events, while vessel disturbance layers quantify the intensity of ship traffic and associated risks. All datasets were processed into a 1-$km^2$ resolution study grid covering Canada’s EEZ, using the WGS84 coordinate reference system (EPSG:4326). The resulting data products are stored as Cloud Optimized GeoTIFFs (COGs), with comprehensive metadata documenting workflows and processing details. Openly accessible, this dataset supports diverse applications, including evaluating cumulative environmental impacts, informing environmental response planning, supporting recovery strategies for at-risk species, and enhancing marine spatial planning to mitigate threats. For example, this work provides a foundation for evaluating cumulative effects on vulnerable species, such as the Leach’s Storm-Petrel, which regularly forages beyond Canada’s jurisdiction.

By providing an integrated and high-resolution review of key anthropogenic stressors, this dataset bridges critical knowledge gaps in ocean management. It enables data-driven decision making for conservation practitioners, policymakers and researchers working to mitigate environmental threats and ensure the sustainable ocean management of marine resources within and beyond Canadian waters.

##### 

*Key words/phrases:* Annual cycle, Light, Petroleum activities, Petroleum pollution, Marine birds, Marine threats, Shipping

\* E-mail:Open Research: Data is available in Zenodo at https://doi.org/TO/ADD. Code is available in Zenodo at https://doi.org/TO/ADD.

\*Corresponding Author E-mail: david.beauchesne@insileco.io

# Class I. Data Set Descriptors

## A. Data set identity

Threats to marine birds to inform conservation and management in Atlantic Canada

## B. Data set identification code

For this project, each dataset identification code (file name) combines:

* Dataset Category: The folder name indicating the thematic category and version (e.g., offshore\_petroleum\_activity-1.0.0 or shipping\_night\_light\_intensity\_density-1.0.0).
* Dataset Descriptor: File names specifying temporal (e.g., month/year) and subcategory details (e.g., vessel type or spill type), where applicable.
* File Format: Cloud Optimized GeoTIFFs (.tif).

## C. Data set description

### Originators

* David Beauchesne
  + david.beauchesne@insileco.io
  + inSileco, Québec, Canada
* Kevin Cazelles
  + kevin.cazelles@insileco.io
  + inSileco, Québec, Canada
* Stephanie Avery-Gomm
  + Stephanie.Avery-Gomm@ec.gc.ca
  + Environment and Climate Change Canada, Ottawa, ON, Canada
* April Hedd
  + April.Hedd@ec.gc.ca
  + Environment and Climate Change Canada, Mount Pearl, Newfoundland and Labrador, Canada
* Philippe Muise
  + Philippe.Muise@ec.gc.ca
  + Canada Centre for Mapping and Earth Observation, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada
* Allison Patterson
  + Allison.Patterson@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Robert Ronconi
  + Robert.Ronconi@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Katharine Studholme
  + Katharine.Studholme@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Sarah Wong
  + sarah.wong2@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada

### Abstract

Anthropogenic activities such as oil pollution, vessel traffic, and artificial light pose significant threats to marine ecosystems across Canada’s Exclusive Economic Zone (EEZ) and adjacent waters. This dataset integrates secondary-use geospatial layers from multiple primary data sources to support conservation, environmental response planning, and marine spatial management in the northwest Atlantic. The dataset spans the Canadian EEZ and adjacent waters to X, Y, providing annual and/or monthly threat layers from 2006–2023 at a 1-km² resolution (WGS84, EPSG:4326). It quantifies petroleum pollution, oil and gas infrastructure, vessel activity, and artificial light exposure. Data sources include the Integrated Satellite Tracking of Pollution (ISTOP), National Aerial Surveillance Program (NASP), National Environmental Emergency Centre (NEEC), Automatic Identification System (AIS) shipping data, and VIIRS satellite observations. Proprietary and public datasets were harmonized through reproducible workflows to ensure consistency and accessibility. Threat layers were developed using geospatial modeling and analysis. Petroleum pollution layers highlight oiling event density, vessel disturbance layers quantify ship traffic by type and time of day, and artificial light layers summarize exposure in critical marine habitats. All data products are stored as Cloud Optimized GeoTIFFs (COGs) with comprehensive metadata. This dataset enables assessments of cumulative environmental impacts, supports risk management for at-risk species, and informs marine spatial planning. For example, it provides insight into cumulative stressors affecting vulnerable species like the globally threatened Leach’s Storm-Petrel, a nocturnally active pelagic seabird. By integrating diverse datasets into an accessible format, this dataset enhances data-driven decision-making for conservation practitioners, policymakers, and researchers working toward sustainable ocean management.

## D. Key words/phrases:

Northwest Atlantic, Atlantic Canada, Annual cycle, Light, Petroleum activities, Petroleum pollution, Marine birds, Marine threats, Shipping

# Class II. Research origin descriptors

## A. Overall project description

### 1. Identity

Threats to marine birds to inform conservation and management in Atlantic Canada

### 2. Originators

* David Beauchesne
  + david.beauchesne@insileco.io
  + inSileco, Québec, Canada
* Kevin Cazelles
  + kevin.cazelles@insileco.io
  + inSileco, Québec, Canada
* Stephanie Avery-Gomm
  + Stephanie.Avery-Gomm@ec.gc.ca
  + Environment and Climate Change Canada, Ottawa, ON, Canada
* April Hedd
  + April.Hedd@ec.gc.ca
  + Environment and Climate Change Canada, Mount Pearl, Newfoundland and Labrador, Canada
* Philippe Muise
  + Philippe.Muise@ec.gc.ca
  + Canada Centre for Mapping and Earth Observation, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada
* Allison Patterson
  + Allison.Patterson@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Robert Ronconi
  + Robert.Ronconi@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Katharine Studholme
  + Katharine.Studholme@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada
* Sarah Wong
  + sarah.wong2@ec.gc.ca
  + Environment and Climate Change Canada: Dartmouth, Nova Scotia, Canada

### 3. Period of study

The study uses data from 2006 to 2023. The number of years varies according to each primary dataset used, as reported below. Data integration and threat layer development commenced in 2022, with final outputs and reproducible workflows completed in January 2025.

### 4. Objectives

This project aimed to improve the understanding of the spatiotemporal distribution of anthropogenic threats in the northwest Atlantic. The study area extends beyond Canada’s Exclusive Economic Zone (EEZ) to encompass the waters regularly used by seabirds breeding in Atlantic Canada. By integrating geospatial data on petroleum pollution, vessel activity, and artificial light, this project provides actionable insights to inform conservation strategies, recovery plans for at-risk species, and marine spatial planning initiatives under Canada’s Ocean Protection Plan (OPP) and Marine Spatial Planning (MSP) frameworks.

### 5. Abstract

See Class 1, section C.

### 6. Sources of funding

Not applicable.

## B. Specific subproject description

### 1. Site description

The study focuses on Canada’s Exclusive Economic Zone (EEZ), including the Atlantic Ocean, the Gulf of St. Lawrence, and adjacent international waters. The study area encompasses diverse marine habitats critical to migratory seabirds, spanning over 5 million km2. The area features high levels of industrial and maritime activity, making it a key region for assessing the cumulative effects of anthropogenic threats on marine ecosystems. Data layers were analyzed within a standardized 1-km2 resolution grid.

### 2. Experimental or sampling design

This project integrates diverse primary datasets to generate spatiotemporal threat layers for Canada’s Exclusive Economic Zone (EEZ) and adjacent waters. The overarching process involves three key phases: 1) data acquisition, 2) data processing and 3) threat layer preparation:

1. **Data Acquisition:** Primary datasets were sourced from a combination of publicly available repositories and proprietary data shared by collaborating organizations. Sources include Canadian government programs (e.g., ISTOP, NASP, NEEC), satellite observations (e.g., VIIRS for light and flaring data), and Automatic Identification System (AIS) for vessel tracking. Data formats varied widely, encompassing vectorized polygons, raster files, tabular data, and time-series observations. Metadata was documented for each dataset (more below).
2. **Data Processing:** All datasets underwent a standardized processing pipeline to harmonize formats, spatial projections, and temporal resolutions. Key steps included:
   * **Spatial Filtering:** Clipping data to the defined study area (EEZ) using a bounding box and ensuring alignment with the WGS84 coordinate reference system (EPSG:4326).
   * **Temporal Aggregation:** Aggregating data to monthly or annual time steps, depending on the dataset’s temporal resolution and application needs.
   * **Data Cleaning:** Removing invalid or missing entries, applying quality assurance filters, and standardizing attribute fields.
   * **Reformatting:** Converting data into standardized and non-proprietary file formats (e.g., Cloud Optimized GeoTIFFs (COGs) for raster data, GeoPackages for vector data).
3. **Threat Layer Preparation:** Threat layers were prepared by integrating individual or multiple primary datasets and by applying spatial modeling and analysis techniques. Each threat layer was mapped onto a 1-km2 resolution study grid to ensure uniform spatial representation. Threat layers were then rasterized to quantify intensity or density metrics (e.g., ship traffic density, light pollution radiance). This approach allowed for the integration of multiple data sources while maintaining spatial and temporal consistency. All threat layers were exported as Cloud Optimized GeoTIFFs (COGs).

### 3. Research methods

#### Data Sources Overview

##### Marine oil pollution incidents

###### Integrated Satellite Tracking of Pollution (ISTOP)

The [Integrated Satellite Tracking of Pollution (ISTOP)](https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions/oil-pollution-monitoring-overview/satellite-spills-overview.html) program (Environment and Climate Change Canada 2024) utilizes satellite imagery to monitor and detect oil spills in coastal waters. Since 2006, daily satellite analyses have been conducted to identify and report both illegal and accidental marine oil pollution. These analyses detect sea-surface anomalies, which are vectorized into polygons representing affected areas. The data obtained span 2006 to 2023 and include 637 recorded incidents. Access to the data is restricted and they are currently stored on a secure Google Cloud Storage bucket accessible through an authentication key and maintained by inSileco.

* **Source**: Environment Canada Integrated Satellite Tracking of Pollution (ISTOP) Program
* **Accessibility**: Restricted
* **Data Type**: Vectorized Polygons
* **Coverage**: North American coastal waters, 2006–2023
* **Processing Script**: prc\_istop.R
* **Output File**: istop.gpkg

###### National Aerial Surveillance Program (NASP)

The [National Aerial Surveillance Program (NASP)](https://tc.canada.ca/en/programs/national-aerial-surveillance-program) (Transport Canada 2024) is a Canada-wide aerial surveillance initiative designed to detect pollution in order to safeguard marine environments, endangered marine life, and promote safe maritime transport along Canada’s coastlines. The program covers pollution detections in the Atlantic region from April 2012 to March 2023, and includes 3578 incidents. Access to the data is restricted and they are currently stored on a secure Google Cloud Storage bucket accessible through an authentication key and maintained by inSileco.

* **Source**: Transport Canada National Aerial Surveillance Program (NASP)
* **Accessibility**: Restricted
* **Data Type**: CSV
* **Coverage**: Atlantic Region, Canada, 2012–2023
* **Processing Script**: prc\_nasp.R
* **Output File**: nasp.gpkg

###### National Environmental Emergency Centre (NEEC)

The [National Environmental Emergency Centre (NEEC)](https://www.canada.ca/en/environment-climate-change/services/environmental-emergencies-program/national-centre.html) (National Environmental Emergency Centre 2024), a program under Environment and Climate Change Canada, maintains an inventory of environmental emergency incidents across Canada. Via the Canadian Wildlife Service, the NEEC contributed data for incidents impacting waterbodies in the Atlantic region, including Québec, from 2016 to 2023, comprising 9570 recorded incidents. Access to the data is restricted and they are currently stored on a secure Google Cloud Storage bucket accessible through an authentication key and maintained by inSileco.

* **Source**: National Environmental Emergency Centre (NEEC)
* **Accessibility**: Restricted
* **Data Type**: CSV
* **Coverage**: Atlantic Region, Canada (including Québec), 2016–2023
* **Processing Script**: prc\_neec.R
* **Output Files**: neec.csv and substances.csv

##### Offshore petroleum activities

###### Canada-Newfoundland and Labrador Offshore Energy Regulator (C-NLOER)

The [Canada-Newfoundland and Labrador Offshore Energy Regulator (C-NLOER)](https://www.cnlopb.ca/), known prior to January 31, 2025 as the Canada-Newfoundland and Labrador Offshore Petroleum Board (Canada-Newfoundland and Labrador Offshore Petroleum Board 2024), oversees and provides data on offshore energy activities in Newfoundland. The dataset used for this project includes multiple files covering land tenure sectors of the Scheduled Land Tenure System, calls for bids, exploration licenses, significant discovery licenses, production licenses, and well locations and detailed information summaries ([Table 1](#tbl-offshore-petroleum-nfl)), providing a comprehensive view of the offshore petroleum landscape in Newfoundland. The dataset includes both spatial and tabular data. Key variables encompass license IDs, well IDs, geographic coordinates, vessel types, exploration phases, and parcel identifiers. Data files are provided in various formats, including shapefiles and Excel files, and have been processed to create spatial representations of offshore petroleum activity sites. For the purpose of this project, we only retained, if available, the classification, status, start and end dates when processing the data. The different categories of offshore petroleum activities are detailed in [Table 2](#tbl-classifications). Each entry in the integrated dataset was given a unique identifier. Individual entries with dual classification (e.g. exploration and delineation) were duplicated, each retaining the same id to avoid overestimating the number of entries in the dataset. As such, the number of unique entries corresponds to the number of unique identifiers rather than the number of rows in the dataset. For data with well status, we simplified status categories for the processed data, as detailed in [Table 3](#tbl-status). How date ranges were established for each entry is detailed in [Table 1](#tbl-offshore-petroleum-nfl).

* **Source**: Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB)
* **Accessibility**: Open, Creative Commons Attribution License
* **Data Type**: Shapefiles, XLSX
* **Coverage**: Offshore regions of Newfoundland, Canada
* **Processing Script**: prc\_offshore\_petroleum\_nfl.R
* **Output File**: offshore\_petroleum\_nfl.gpkg

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Description of licenses, call for bids, and sectors for the petroleum industry in Newfoundland and Labrador.   | Type | Description | Period | | --- | --- | --- | | Active Exploration Licenses | Offshore areas in Newfoundland and Labrador that are currently licensed for petroleum exploration. An Exploration License (EL) grants rights to conduct exploratory drilling and other activities aimed at discovering petroleum resources within a designated offshore area. These licenses are typically issued for a fixed period, encouraging exploration activities that assess the resource potential in Newfoundland and Labrador’s offshore regions. | The beginning is the effective date (effective) and the end is the second exploration period specified (period2exp). | | Active Significant Discovery Licenses | Offshore areas within Newfoundland and Labrador where significant petroleum discoveries have been identified and are actively licensed. An Active Significant Discovery License (SDL) allows the holder to retain rights over a discovery area with proven, potentially commercial quantities of petroleum, preserving the right to future exploration and development. | The beginning is the date of first exploration activity (spud\_date) and the termination of the last renewal phase (term\_date). | | Active Production Licenses | Offshore areas in Newfoundland and Labrador where petroleum resources are actively being produced. A Production License (PL) authorizes the holder to extract petroleum within a specified offshore area, following proven resource discoveries. This license allows companies to conduct production activities while adhering to regulatory requirements for resource management, environmental protection, and safety. | The beginning is the date of first exploration activity (spud\_date) and the termination of the last renewal phase (term\_date). | | Production Installations | Locations of offshore production installations used in petroleum extraction activities in Newfoundland and Labrador’s offshore areas. A Production Installation refers to infrastructure such as platforms, rigs, or other facilities designed to extract, process, and transport petroleum resources from beneath the seafloor. These installations play a critical role in supporting production operations within licensed areas. | The beginning is the date of first exploration activity (spud\_date) and the termination of the last renewal phase (term\_date). | | Delineation Wells | Locations of delineation wells drilled in offshore areas of Newfoundland and Labrador. A Delineation Well is drilled to determine the extent and characteristics of a known petroleum discovery, helping to assess the size, quality, and commercial viability of the resource. These wells are essential in refining the understanding of hydrocarbon deposits discovered by exploration wells. | The beginning is the date of first exploration activity (spud\_date) and the termination of the last renewal phase (term\_date). | | Development Wells | Locations of development wells in offshore areas of Newfoundland and Labrador. Development Wells are drilled within proven petroleum fields to facilitate resource extraction. These wells are strategically placed to optimize production from known hydrocarbon deposits, making them critical for efficient and sustainable petroleum extraction. | The beginning is the date when the well first began producing oil (first\_oil); there is no end date specified. | | Dual Classified Wells | Offshore wells in Newfoundland and Labrador that have been classified under more than one category during their lifecycle. Dual Classified Wells may serve multiple purposes, such as transitioning from an exploration role to development or delineation, depending on the results of drilling and assessments. This dual classification reflects the adaptive approach taken in offshore petroleum exploration and production. | The beginning is the date when the date on which information about the well or license was publicly posted (date\_posted) and the end is the specified call for bids closing year (cfb\_clse\_dte) and assumed to be on January 1st to allow for date formatting. | | Exploration Wells | Locations of offshore wells in Newfoundland and Labrador that are designated for exploration. Exploration Wells are drilled to investigate potential petroleum deposits in unexplored or underexplored areas. These wells are the first step in assessing hydrocarbon presence and play a crucial role in determining the viability of further development. | The beginning is the date when the date on which information about the well or license was publicly posted (date\_posted) and the end is the specified call for bids closing year (cfb\_clse\_dte) and assumed to be on January 1st to allow for date formatting. | | Call for Bids - Eastern Newfoundland NL23-CFB01 | The Call for Bids NL23-CFB01 pertains to the Eastern Newfoundland Region and was issued by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) in 2023. This call offered 28 parcels, totaling approximately 7,222,551 hectares, for exploration licensing. Interested parties were invited to submit sealed bids by November 1, 2023. However, the C-NLOPB [announced](https://www.cnlopb.ca/news/c-nlopb-announces-results-of-calls-for-bids-in-the-eastern-newfoundland-and-south-eastern-newfoundland-regions/) that no bids were received for this call. | The beginning is the date when the date on which information about the well or license was publicly posted (date\_posted) and the end is the specified call for bids closing year (cfb\_clse\_dte) and assumed to be on January 1st to allow for date formatting. | | Call for Bids - South Eastern Newfoundland NL23-CFB02 | The Call for Bids NL23-CFB02 was issued by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) in 2023, offering 19 parcels totaling approximately 4,982,275 hectares in the South Eastern Newfoundland Region for exploration licensing. Interested parties were invited to submit sealed bids by November 1, 2023. However, the C-NLOPB [announced](https://www.cnlopb.ca/news/c-nlopb-announces-results-of-calls-for-bids-in-the-eastern-newfoundland-and-south-eastern-newfoundland-regions/) that no bids were received for this call. | The beginning is the date when the date on which information about the well or license was publicly posted (date\_posted) and the end is the specified call for bids closing year (cfb\_clse\_dte) and assumed to be on January 1st to allow for date formatting. | | Call for Bids - Eastern Newfoundland NL24-CFB01 | The Call for Bids NL24-CFB01 was issued by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) on April 29, 2024, inviting bids for exploration licenses in the Eastern Newfoundland Region. This call comprised 41 parcels, covering a total of 10,287,196 hectares. Interested parties were required to submit sealed bids by November 6, 2024. On November 6, 2024, the C-NLOPB [announced](https://www.cnlopb.ca/news/c-nlopb-announces-results-of-calls-for-bids-in-the-eastern-newfoundland-region/) that no bids were received for Call for Bids NL24-CFB01. | The beginning is the date is the opening date of the call for bids (cfb\_open\_dte) and the end is the closing date of the call for bids (cfb\_clse\_dte). | | Sectors - Labrador South NL02-LS | The Labrador South Region (Sector NL02-LS) is an offshore area managed by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) as part of its Scheduled Land Tenure System. | The beginning is the date is the opening date of the call for bids (cfb\_open\_dte) and the end is the closing date of the call for bids (cfb\_clse\_dte). | | Sectors - Eastern Newfoundland NL06-EN | The Eastern Newfoundland Region (Sector NL06-EN) is an offshore area designated by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) under its Scheduled Land Tenure System. | The beginning is the date is the opening date of the call for bids (cfb\_open\_dte) and the end is the closing date of the call for bids (cfb\_clse\_dte). | | Sectors - North Eastern Newfoundland NL01-NEN | The North Eastern Newfoundland Region (Sector NL01-NEN) is an offshore area designated by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) under its Scheduled Land Tenure System. | The beginning is the date on which the sector’s licensing or regulatory conditions officially took effect (effective); there is no end date specified. | | Sectors - Southern Newfoundland NL01-SN | The Southern Newfoundland Region (Sector NL01-SN) is an offshore area designated by the Canada-Newfoundland and Labrador Offshore Petroleum Board (C-NLOPB) under its Scheduled Land Tenure System. | The beginning is the date on which the sector’s licensing or regulatory conditions officially took effect (effective); there is no end date specified. | | Well Info Summary | Contains detailed information on wells drilled within offshore petroleum exploration and production areas. This dataset is designed to offer an inventory of wells, capturing essential data for regulatory oversight, resource management, and operational history. The dataset includes information on well names, operators, drilling dates, statuses, and geographic coordinates, among other fields. | The beginning corresponds to the date when drilling for the well began (spud\_date) and the end is the date on which drilling or operational activities for the well concluded (well\_termination\_date). | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Description of classifications for the petroleum industry in Newfoundland   | Classification | Description | | --- | --- | | Sectors | Defined geographic areas designated for potential exploration and development activities, often used in the regulatory framework to organize land tenure. | | Call for Bids | Competitive bidding rounds where companies submit proposals to obtain exploration or development rights for specified areas within sectors. | | Exploration | Initial phase focused on discovering hydrocarbons, involving activities like seismic surveys and drilling exploration wells. | | Significant Discovery | Identification of a commercially viable hydrocarbon discovery during exploration, associated with a Significant Discovery License (SDL) to preserve rights to the area. | | Delineation | Appraisal phase to define the size, extent, and quality of a discovered reservoir, refining understanding of the resource’s potential. | | Development | Preparation phase involving construction of infrastructure, drilling of production wells, and installation of facilities needed to bring hydrocarbons to the surface. | | Production | Active phase of extracting hydrocarbons, processing them, and transporting them for sale, generating revenue from the resource. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Simplified well status classifications with descriptions   | Well Status | Description | | --- | --- | | Abandoned | Wells where activities have been permanently stopped. | | Suspended | Wells with temporarily halted activities that may resume. | | Active Production | Wells currently involved in production or injection activities. | | Drilling | Wells still in the drilling phase. | | Disposal | Wells used for waste or byproduct disposal. | | Closed | Wells officially closed, possibly for regulatory reasons. | | Off Station | Wells or rigs temporarily off their regular location. | | NA | No information available. | |

###### Canada-Nova Scotia Energy Regulator (CNSOER)

The [Canada-Nova Scotia Energy Regulator (CNSOER), known prior to January 31, 2025 as the Canada-Nova Scotia Offshore Petroleum Board (CNSOPB)](https://www.cnsopb.ns.ca/), (Canada-Nova Scotia Offshore Petroleum Board 2024) oversees and provides data relevant to offshore energy activities in Nova Scotia, including petroleum development. The dataset used for this project includes multiple files covering significant discovery areas (NA coordinates), significant discovery licenses (NA coordinates), production licenses (NA coordinates), and the latest call for bids (NA polygons). The data encompasses information on various petroleum exploration and production activities. Key variables include license IDs, well IDs, geographic coordinates, exploration phases, and parcel identifiers. Data files include both Excel sheets (XLS, XLSX) and shapefiles, and they have been processed to create spatially explicit representations of petroleum activity sites. For this project, we only retained the classification of each entry. The classification of petroleum offshore activities follows the same categories as the one presented in [Table 2](#tbl-classifications). Unlike the data for Newfoundland and Labrador (see [Section 2.2.3.1.2.1](#sec-offshore_petroleum_nfl)), there was no information on well status or the time period covered.

* **Source**: Canada-Nova Scotia Offshore Petroleum Board (CNSOPB)
* **Accessibility**: Open
* **Data Type**: XLS, XLSX, Shapefiles
* **Coverage**: Offshore regions of Nova Scotia, Canada, 2009–2024
* **Processing** Script: prc\_offshore\_petroleum\_ns.R
* **Output** File: offshore\_petroleum\_ns.gpkg

##### Automatic Identification System (AIS) Vessel Data

The [Automatic Identification System (AIS)](https://tc.canada.ca/en/marine-transportation/navigation-aids/automatic-identification-system-ais) is a global transportation safety system that uses satellite- and terrestrial-based ship tracking systems to collect identity and dynamic positional data from vessels. The International Convention for the Safety of Life at Sea (SOLAS), Chapter 5, Regulation 19.2, sets out requirements for the use of AIS on ships, though individual flag states can exempt certain vessels. In Canada, vessels must operate an AIS if they are: a) 500 gross tonnage or more, b) 300 gross tonnage or more engaged on an international voyage (except fishing vessels), c) 150 gross tonnage or more carrying more than 12 passengers and engaged on an international voyage, or c) every vessel engaged on a voyage beyond sheltered waters other than those subject to criteria a-b. These AIS data are used to enhance maritime safety, monitor environmental impacts, and facilitate shipping analyses.

The dataset used for this project, provided by Transport Canada (Transport Canada 2023), focuses on 2023 vessel movements across Atlantic Canada and adjacent waters. The AIS data provides dynamic vessel information, including position, date and time, speed, and course over ground, transmitted through terrestrial, satellite, and tier-based receivers ([Table 6](#tbl-shipping-ais-attr)). The AIS dataset we received includes both terrestrial and satellite AIS messages, fused to minimize spatial and temporal gaps and stored as compressed parquet files. The data shared also contains pre-processed tracklines as a geodatabase. However, we used the raw AIS positional data to recreate tracklines for the purposes of this project to allow for more temporal granularity, such as identifying vessel activity at night. Access to the data is restricted and they are currently stored on a secure Google Cloud Storage bucket accessible through an authentication key and maintained by inSileco.

* **Source**: Transport Canada
* **Accessibility**: Restricted
* **Data Type**: Parquet, Geodatabase
* **Coverage**: Atlantic Canada, 2023
* **Processing Script**: prc\_shipping\_ais\_tracklines.R
* **Output File**: shipping\_ais\_tracklines.gpkg

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: Columns (attributes) associated with the positional AIS parquet files (table from Transport Canada (2023)).   | Column\_Name | Description | Information | Data\_Type | | --- | --- | --- | --- | | MMSI | Maritime Mobile Service Identity | 9-digits (MMSIs with valid ship country codes are from 201000000 to 775999999). All ship MMSIs use the format M1I2D3X4X5X6X7X8X9, the first three digits represent the Maritime Identification Digits (MID) and they represent the flag/nationality of the ship. The following digits (X4-X9) can be any numerical value from 0 to 9. | Integer | | Latitude | Latitude of the positional message | Latitude recorded in decimal degrees with an accuracy of approximately 10 metres (-90 to 90). | Float | | Longitude | Longitude of the positional message | Longitude recorded in decimal degrees with an accuracy of approximately 10 metres (-180 to 180). | Float | | SOG | Speed over ground of vessel | The speed of the ship with respect to the ground, recorded in knots (0 to 102.2 are valid; 102.3 is invalid). | Float | | COG | Course over ground of vessel | The course of the ship, recorded in degrees (0-359.9 are valid; 360 is invalid). | Float | | Heading | True heading of vessel | The true heading of the ship, recorded in degrees (0-359 are valid; 360 is invalid). | Integer | | YMD\_HMS | Time of positional message in UTC | The UTC time stamp of when the positional AIS message was sent from the ship (YYYY-MM-DD hh:mm:ss). | Datetime | | Source | Data source/message type of AIS message | The source (e.g., terrestrial, dynamic, satellite) and class (A or B) of the AIS message sent (e.g., TAIS\_A, TAIS\_B, SAIS\_A). | String | |

To create meaningful tracklines representing vessel movements over space and time, we followed the processing steps described in Veinot et al. (2023), with additional steps to distinguish vessel activity at night versus during the day:

1. **Data Extraction**
   * AIS data was extracted from compressed archive files and loaded into memory using efficient tools such as the arrow (https://CRAN.R-project.org/package=arrow) and vroom (https://CRAN.R-project.org/package=vroom) R packages. This ensured scalability given the dataset size, which often exceeds 10GB annually.
2. **Initial Data Cleaning**
   * Raw AIS data was filtered to remove invalid records:
     + Only positions with valid **MMSI** values (201000000–775999999) were retained.
     + **SOG** values were restricted to the range 0.1–100 knots to remove implausible speeds.
3. **Day/Night Classification**
   * Using the suncalc R package (https://cran.r-project.org/web/packages/suncalc/index.html), each AIS position was classified as daytime or nighttime. The classification was based on local sunrise and sunset times at each position, ensuring alignment with the UTC timestamp.
4. **Trackline Generation**
   * Tracklines were created by:
     1. Sorting AIS data by **MMSI** and timestamp.
     2. Calculating distances between consecutive positions using geodesic measurements.
     3. Splitting movements into separate tracks when:
        + The distance exceeded 50 nautical miles.
        + The time interval between positions exceeded 300 minutes.
     4. Further segmenting tracklines to differentiate between daytime and nighttime navigation.
5. **Trackline Attributes**
   * Additional attributes were calculated for each trackline to facilitate further analysis:
     + **Average and Maximum Speeds**: Derived from AIS speed data.
     + **Number of Positions**: Total AIS messages contributing to the trackline.
     + **Elapsed Hours**: Duration of the trackline.
     + **Track Length**: Computed in kilometers using geodesic measurements.
     + **Daytime Classification**: Indicates whether the trackline occurred during the day, night, or a mix of both.
6. **Post-Processing**
   * Erratic tracklines were flagged and removed using heuristic thresholds:
     + Extremely short tracks (<100 meters).
     + Unrealistic average speeds (<0.3 knots or >100 km/h).
     + Speeds exceeding vessel-type-specific thresholds:
       - 30 km/h for pleasure vessels and special ships.
       - 35 km/h for fishing and tug vessels.
       - 40 km/h for all other vessel types.
   * Tracklines identified as spoofing were excluded from further analyses.
7. **Output**
   * The cleaned tracklines were exported as monthly parquet files (.parquet). Each file contains the processed AIS points with vessel information and the necessary information to differentiate individual tracks and whether tracks are during daytime or nighttime. Each point includes relevant attributes, such as speed, duration, and daytime classification, enabling detailed analyses of vessel movements across Canadian waters ([Table 7](#tbl-shipping-ais-points)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: Columns (attributes) associated with the AIS positional points processed from the raw parquet files.   | Column Name | Description | Information | Data Type | | --- | --- | --- | --- | | mmsi | Maritime Mobile Service Identity | 9-digit identifier assigned to a vessel | Numeric | | latitude | Latitude of the AIS positional point in decimal degrees | Recorded in WGS 84 (EPSG:4326) | Numeric | | longitude | Longitude of the AIS positional point in decimal degrees | Recorded in WGS 84 (EPSG:4326) | Numeric | | sog | Speed over ground at the AIS positional point | Measured in knots converted to kilometers per hour | Numeric | | ymd\_hms | Timestamp of the AIS positional point in UTC | Recorded in POSIXct format with timezone UTC | POSIXct | | date | Date corresponding to the AIS positional point | Derived from the timestamp for easier grouping and filtering | Date | | day\_or\_night | Classification of the AIS positional point as day or night | Derived using sunrise and sunset times from solar calculations | Character | | dist\_miles | Distance (in miles) between consecutive AIS positional points | Calculated using geodetic distance between successive points | Numeric | | time\_diff | Time difference (in minutes) between consecutive AIS positional points | Difference between timestamps of successive points | Numeric | | new\_track | Flag indicating whether a new trackline is initiated at the point | Determined based on distance or time thresholds for new tracklines | Logical | | track\_id | Unique identifier for each trackline within a ship’s movement | Generated using the cumulative sum of new\_track flags | Integer | | day\_night\_segment | Unique identifier for segments of tracklines split by day or night | Calculated to split tracklines into day and night segments | Integer | |

##### Visible Infrared Imaging Radiometer Suite

The Visible Infrared Imaging Radiometer Suite (VIIRS) is a key instrument aboard the National Oceanic and Atmospheric Administration’s (NOAA), Suomi National Polar-orbiting Partnership (Suomi NPP) and NOAA-20 satellites. VIIRS collects visible and infrared imagery and radiometric measurements, supporting applications such as environmental monitoring, weather forecasting, and maritime surveillance, including boat detection.

###### VIIRS Boat Detection (VBD) Dataset

The [VIIRS Boat Detection (VBD) Dataset](https://eogdata.mines.edu/products/vbd/) (Earth Observation Group, Payne Institute for Public Policy 2024a) provides nightly global data on boat detections based on radiance observations captured by the Visible Infrared Imaging Radiometer Suite (VIIRS). This dataset is widely used for maritime surveillance, tracking vessel activity, and monitoring illegal fishing. The raw data includes lighted boat detections with associated radiance, confidence metrics, and spatial coordinates ([Table 8](#tbl-viirs-vbd-attr)). The data was acquired for the years 2020–2023 and pre-processed to facilitate further analysis of vessel patterns within the Canadian Exclusive Economic Zone (EEZ) and adjacent waters. The processed data is stored as a Geopackage and includes filtered, high-confidence detections with key attributes for radiance intensity and spatial location.

* **Source**: Earth Observation Group, Payne Institute for Public Policy
* **Accessibility**: Public domain with registration
* **Data Type**: CSV files
* **Coverage**: Global, 2020–2023
* **Processing Script**: prc\_viirs\_boat\_detection.R
* **Output File**: viirs\_boat\_detection.gpkg

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 8: Columns (attributes) associated with the VIIRS Boat Detection dataset used for this project.   | ColumnName | Description | Data\_Type | | --- | --- | --- | | Date\_Mscan | Date and time of the observation in UTC | POSIXct | | Lat\_DNB | Latitude of the detected boat in decimal degrees | Numeric | | Lon\_DNB | Longitude of the detected boat in decimal degrees | Numeric | | Rad\_DNB | Radiance intensity detected at the location (nW/cm²/sr) | Numeric | | QF\_Detect | Quality flag indicating confidence in the detection (1 = high confidence) | Integer | |

To prepare the VBD dataset for further analyses, the following pre-processing steps were applied:

* Raw .csv.gz files were downloaded and processed using the arrow package, which allows for efficient reading of compressed CSV files.
* High-confidence detections were retained by filtering for rows with:
  + **QF\_Detect = 1**: Ensures only high-confidence detections are included.
  + **Rad\_DNB > 0**: Removes detections with non-positive radiance values.
* Optionally, filters for **SMI** (Signal-to-Mean Index) and **SHI** (Signal-to-High Index) can be applied for stricter data quality criteria (e.g., **SMI > 0.1**, **SHI > 0.75**).
* The filtered dataset was converted to a spatial object with spatial projection **EPSG:4326** using the sf package, enabling geospatial analyses and integration with other datasets.
* Observation timestamps were standardized to UTC using the lubridate package to ensure consistent temporal referencing across datasets.
* The pre-processed dataset was saved as a Geopackage (viirs\_boat\_detection.gpkg)

While filters based on the **Signal-to-Mean Index (SMI)** and **Signal-to-High Index (SHI)** were not directly applied in the current processing pipeline, Ffuture iterations of the pipeline could incorporate these metrics for scenarios requiring stricter detection thresholds, such as high-confidence studies of vessel activity in heavily lit areas.

* **SMI**: Measures the signal strength relative to the mean radiance value in the surrounding area. Higher values indicate stronger, more distinct detections.
* **SHI**: Measures the signal strength relative to a high radiance threshold, helping to exclude detections influenced by bright land-based lights or other high-radiance sources.

###### VIIRS Night Fire (VNF) Dataset

The [VIIRS Night Fire (VNF) Dataset](https://eogdata.mines.edu/products/vnf/) (Elvidge et al. 2013b, Earth Observation Group, Payne Institute for Public Policy 2024b) provides nightly and annual global thermal anomaly detections captured by the Visible Infrared Imaging Radiometer Suite (VIIRS). This dataset is commonly used for monitoring gas flares, wildfires, industrial emissions, and other heat sources. The raw nightly data includes processed radiance-derived fields (e.g., temperature and radiant heat indices), quality flags for detection confidence, spatial coordinates, and emitter-specific information. [Table 9](#tbl-viirs-vnf-attr) presents a subset of available attributes for the nightly data that were selected for their relevance to this project.

In addition to the nightly VNF dataset, the annual VNF data from 2017 to 2023 was harvested. This dataset provides aggregated yearly thermal anomaly detections, including gas flares, and includes fields for average temperature, detection frequency, and estimated burnable carbon mass (BCM), among others (see [Table 10](#tbl-viirs-vnf-annual-attr)).

* **Source**: Earth Observation Group, Payne Institute for Public Policy
* **Accessibility**: Restricted with credentials
* **Data Type**: Nightly CSV files and annual GeoPackage files
* **Coverage**: Global
* **Processing Scripts**: prc\_viirs\_night\_fire.R, prc\_viirs\_night\_fire\_annual.R
* **Output Files**: viirs\_night\_fire.gpkg, viirs\_night\_fire\_annual.gpkg

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 9: Columns (attributes) associated with the nightly VIIRS Night Fire dataset used for this project.   | Column\_Name | Description | Data\_Type | | --- | --- | --- | | Date\_Mscan | Date and time of the observation in UTC | POSIXct | | Lat\_GMTCO | Latitude of the detected thermal anomaly in decimal degrees | Numeric | | Lon\_GMTCO | Longitude of the detected thermal anomaly in decimal degrees | Numeric | | Temp\_primary | Primary temperature of the emitting source (Kelvin) | Numeric | | ESF\_primary | Emission scaling factor for the primary source | Numeric | | RHI\_primary | Radiant heat index for the primary source | Numeric | | Area\_primary | Estimated area of the primary emitting source (square meters) | Numeric | | Temp\_secondary | Secondary temperature of the emitting source (Kelvin) | Numeric | | ESF\_secondary | Emission scaling factor for the secondary source | Numeric | | RHI\_secondary | Radiant heat index for the secondary source | Numeric | | id\_iremitter | Unique identifier for the potential emitter | Character | | Type\_iremitter | Type of the emitter (e.g., industrial, flaring) | Character | | Category\_iremitter | Category of the emitter (e.g., petroleum, thermal anomaly) | Character | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 10: Columns (attributes) associated with the annual VIIRS Night Fire dataset used for this project.   | Column\_Name | Description | Data\_Type | | --- | --- | --- | | avg\_temp\_k | Average temperature of the emitting source (Kelvin) | Numeric | | detection\_frequency | Detection frequency as a proportion of clear-sky observations | Numeric | | clear\_obs | Number of clear-sky observations available for the source | Numeric | | type | Type of the thermal anomaly (e.g., flare, fire, etc.) | Character | | bcm | Estimated burnable carbon mass (BCM) | Numeric | | year | Year of the aggregated data | Integer | | geom | Spatial point geometry (latitude and longitude) | sfc\_POINT | |

To prepare the VNF dataset for further analyses, the following pre-processing steps were applied, with information on flares coming from Elvidge et al. (2016) and Zhizhin et al. (2021):

1. Data Extraction
   * Raw CSV files were processed using arrow::read\_csv\_arrow() for efficient import.
   * Only necessary columns were selected during import to minimize memory usage.
2. Quality Filtering
   * Data quality was ensured by applying bitwise filtering to the QF\_Detect field.
   * Flags corresponding to high-confidence detections and specific radiance thresholds were retained:
     + **High Radiance Thresholds**: Retained flags ensuring significant thermal anomalies in bands M07–M16.
     + **Thermal Anomalies**: Retained detections flagged as local maxima or cluster detections.
3. Spatial Filtering
   * The dataset was filtered to detections within a bounding box approximating the Canadian Exclusive Economic Zone (EEZ):
     + **Longitude**: -141° to -50°.
     + **Latitude**: 40° to 85°.
4. Processed Radiance Fields
   * Radiance-derived fields, including Temp\_primary, RHI\_primary, and ESF\_primary, were used for analyzing thermal anomalies.
   * Secondary radiance-derived fields (Temp\_secondary, RHI\_secondary, ESF\_secondary) were retained for comprehensive heat calculations.
5. Total Radiant Heat Calculation
   * A total\_heat field was computed as the sum of RHI\_primary and RHI\_secondary, representing the combined radiant heat output of both primary and secondary sources.
6. Flare Identification
   * A temperature threshold of **1450 K** was applied to Temp\_primary to flag likely gas flares.
   * This threshold is based on established studies identifying gas flares as high-temperature sources.
7. Temporal Standardization
   * Timestamps were standardized to UTC using the lubridate package to ensure temporal consistency across datasets.
8. Output
   * The pre-processed dataset was saved as a GeoPackage (viirs\_night\_fire.gpkg).
   * The GeoPackage contains spatial coordinates in the EPSG:4326 projection, radiance fields, emitter information, and retained quality flags.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 11: Retained quality flags for VIIRS Nightfire data processing.   | Flag | Description | Rationale | | --- | --- | --- | | 1 | M10 radiance above threshold | Ensures M10 radiance is significant for thermal analysis. | | 2 | M10 local maximum (bow-tie corrected) | Captures prominent local maxima in M10, reducing noise from bow-tie duplication. | | 4 | M7 radiance above threshold | Identifies potential detections based on M7 radiance thresholds. | | 8 | M8 radiance above threshold | Ensures inclusion of detections with high M8 radiance, often indicative of thermal events. | | 16 | M12 radiance outside scatterplot limits | Captures M12 detections outside scatterplot-defined radiance limits, indicating anomalies. | | 32 | M13 radiance outside scatterplot limits | Captures M13 detections outside scatterplot-defined radiance limits, indicating anomalies. | | 64 | DNB local maximum within 3 km of M10 | Identifies significant DNB maxima close to thermal detections for contextual analysis. | | 128 | M10 cluster detection | Highlights clusters of M10 detections, often related to persistent sources. | | 65536 | M11 cluster detection | Highlights clusters of M11 detections for thermal anomaly identification. | | 2048 | M10 tophat-filtered | Filters M10 detections using a tophat approach to remove spurious signals. | | 4096 | M12 tophat-filtered | Filters M12 detections using a tophat approach to remove spurious signals. | | 8192 | M13 tophat-filtered | Filters M13 detections using a tophat approach to remove spurious signals. | | 262144 | M11 tophat-filtered | Filters M11 detections using a tophat approach to remove spurious signals. | | 524288 | M14 local thermal anomaly | Identifies strong thermal anomalies in M14, essential for flare detection. | |

###### VIIRS Nighttime Light (VNL) Dataset

The [VIIRS Nighttime Light (VNL) Dataset](https://eogdata.mines.edu/products/vnl/) (Elvidge et al. 2013a, n.d., Earth Observation Group, Payne Institute for Public Policy 2024c) provides global observations of nighttime light intensity captured by the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi NPP (npp) and NOAA-20 (j01) satellites. This dataset is widely used for analyzing human activity, urbanization, and environmental monitoring. The monthly VNL data includes radiance measures aggregated over time, with options for stray-light corrected data and various file types, including cloud-free composites and coverage metrics.

For this analysis, the file type avg\_rade9h was selected, which represents the average radiance values over the specified temporal period, expressed in nanowatts per square centimeter per steradian (nW/cm²/sr). This file type is particularly suited for analyzing overall light intensity while excluding temporary anomalies. The data were obtained using the vcm configuration, which excludes stray light, ensuring higher data accuracy in regions with minimal artificial light. Additionally, the data was sourced from the j01 platform, representing the NOAA-20 satellite, part of the VIIRS suite, which provides enhanced coverage and resolution compared to earlier satellites.

The raw raster files were downloaded in GeoTIFF format and processed to focus on the Canadian Exclusive Economic Zone (EEZ) and adjacent waters. Processing involved loading the rasters, cropping them to a bounding box defined by the coordinates (-141, 40, -50, 85), and exporting each monthly raster as a Cloud Optimized GeoTIFF (COG). This format allows for efficient storage, faster data access, and better interoperability with modern geospatial tools, particularly for web-based or cloud-hosted applications. The processed COGs retain the original resolution and include internal overviews for optimized rendering.

* **Source**: Earth Observation Group, Payne Institute for Public Policy
* **Accessibility**: Restricted access, credentials required
* **Data Type**: GeoTIFF files (.tif.gz)
* **Coverage**: Global
* **Temporal Resolution**: Monthly
* **Processing Script**: prc\_viirs\_night\_light.R
* **Output File**: viirs\_night\_light.tif

#### Threat layers

##### Marine pollution incidents

The objective was to identify areas with relatively higher frequencies of past petroleum release incidents over a defined period. Such areas may also be at higher risk of future incidents, assuming high incident frequencies may be associated with factors such as high vessel traffic or fueling locations. Ideally, a more detailed approach would be used which incorporates spill diffusion modelling based on factors such as total volume, substance type, and local or regional currents to estimate the likely spread of each event. However, this level of detail was beyond the scope of the current study. Given the retrospective nature of the dataset and its focus on historical rather than predictive risk, we chose a simplified approach. Each recorded or probable incident was treated as an individual, discrete event in the dataset to characterize exposure to petroleum-related incidents.

This threat layer was constructed using the [ISTOP](@sec-istop), [NEEC](@sec-neec), and [NASP](@sec-nasp) datasets. Each dataset was processed to create spatially and temporally explicit threat layers:

* **ISTOP**: Categories 1A (possible oil with a clearly associated target), 1B (possible oil with a target within 50 km), and 2 (possible oil without an identifiable source) were retained. The ISTOP data already included polygons representing areas of exposure, which were directly utilized as pollution events in the analysis.
* **NASP**: Records with missing coordinates were removed. We inferred exposure areas for each NASP incident, initially represented as a point with latitude/longitude coordinates and spill volumes in liters, by categorizing spill volumes and applying a variable buffer around each point based on the reported volume ([Table 12](#tbl-spillvolumes)).
* **NEEC**: This dataset underwent extensive filtering, including the removal of records with empty coordinates, selection of the highest quantity category for multi-substance spills, exclusion of terms irrelevant to the marine environment (e.g., “mvi,” “furnace oil”), and filtering out observations beyond 1 km of the coastline. NEEC incidents were also initially represented as points and were processed into polygons similarly to the NASP dataset (Table 12), with additional classification by substance type based on chemical state and the presence of an oil sheen.

Each dataset was clipped to the boundary of the area of interest. Temporal information was standardized, with incident dates grouped by month. The datasets were rasterized using the study grid, with each grid cell representing the count of overlapping incidents for a given month divided by the number of years covered by each dataset. This resulted in incident density units of number of events per year for each calendar month. For the NEEC dataset, threat layers were created for each substance classification (e.g., liquid with oil sheen, liquid without oil sheen, unknown). We also created monthly rasters combining all the NEEC, ISTOP, and NASP datasets together. The resulting rasters provide spatially explicit representations of petroleum pollution incidents, highlighting areas of greater historical exposure.

* **Processing Scripts**:
  + ana\_petroleum\_pollution\_incidents\_istop.R
  + ana\_petroleum\_pollution\_incidents\_nasp.R
  + ana\_petroleum\_pollution\_incidents\_neec.R
  + ana\_petroleum\_pollution\_incidents.R
* **Output Files**: Rasterized layers stored in workspace/data/analyzed/petroleum\_pollution\_incidents-1.0.0/.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 12: Volume categories and buffer size selected to transform incident as point locations to area of exposure from   | Volume (L) | Buffer size (m) | | --- | --- | | [0, 100] | 250 | | [100, 1000] | 500 | | [1000, 10000] | 1000 | | [10000+ [ | 2000 | |

##### Offshore petroleum activities

###### Installations, wells, and lease areas

The offshore petroleum activity threat layers were developed to enable assessment of spatial exposure to exploration, resource delineation, and production activities, as well as potential future activity, within the defined study area. This analysis integrates spatial data on petroleum industry operations from Newfoundland [Section 2.2.3.1.2.1](#sec-offshore_petroleum_nfl) and Nova Scotia [Section 2.2.3.1.2.2](#sec-offshore_petroleum_ns). Data preparation involved processing point (installations, wells) and polygon features (lease areas, sectors) from the harvested datasets. Point geometries were buffered by 500 meters to approximate their spatial influence, while polygon geometries were used as-is. It must be noted that the buffer size was selected only to represent the spatial presence of offshore petroleum activities rather than the risk that they pose in case of accidental oil spills. The combined dataset was clipped to the boundary of the area of interest and grouped by classification ([Table 2](#tbl-classifications)) and status categories ([Table 3](#tbl-status)). Spatial features were grouped into discrete threat layers using union operations to consolidate overlapping geometries and identify areas where activities occurred as presence-absence. To quantify exposure, the threat layers were rasterized using the study grid. Each cell in the raster represents the count of overlapping features for a given combination of classification and status. This process generated a series of spatially explicit raster layers, which were exported as geospatial files for further use.

* **Processing Script**: ana\_offshore\_petroleum\_activity.R
* **Output File(s)**: Rasterized layers stored in workspace/data/analyzed/offshore\_petroleum\_activity-1.0.0/.

###### Flaring

The offshore petroleum flaring threat layers were created using the VIIRS Night Fire (VNF) dataset ([Section 2.2.3.1.5.2](#sec-viirs_night_fire)), with separate analyses for monthly and annual data. These layers quantify spatial and temporal variations in flaring activity associated with offshore petroleum platforms, providing insight into their environmental footprint and potential risk to seabirds sensitive to light pollution.

**Annual Layers**

The annual VNF dataset was used to generate rasterized layers summarizing flaring activity by year. Each annual dataset contains spatial information on flaring platform locations, including the average temperature of flares (avg\_temp\_k) and frequency of flaring activity. The following steps were applied to create the annual threat layers:

1. **Data Preparation**: The annual VNF dataset was filtered to retain only records within the area of interest.
2. **Aggregation**: Flaring activity was grouped by year to calculate annual summaries of total heat output (total\_heat) at each platform location.
3. **Rasterization**: The annual flaring data was rasterized using the study grid, with each grid cell containing the sum of total\_heat values for all platforms in that cell for a given year.
4. **Export**: Rasterized layers were exported as geospatial files, with one raster file generated for each year.

The resulting layers provide annual snapshots of flaring activity for comparison across years.

* **Processing Script**: ana\_offshore\_petroleum\_platform\_annual.R
* **Output Files**: Rasterized annual layers stored in workspace/data/analyzed/offshore\_petroleum\_platform\_annual-1.0.0/.

**Monthly Layers**

The monthly VNF dataset was used to generate high-resolution temporal layers, summarizing flaring activity at a monthly level. These layers provide finer granularity in monitoring platform activity. Only data for the year 2023 are available. The following steps were applied to create the monthly threat layers:

1. **Data Preparation**: The monthly VNF dataset was filtered to retain only records flagged as likely\_flare and located within the area of interest.
2. **Temporal Grouping**: Flaring activity was grouped by month, with the total\_heat value aggregated for each month.
3. **Rasterization**: For each month, flaring data was rasterized onto the study grid, with cell values representing the sum of total\_heat values for all platforms within that cell.
4. **Export**: Rasterized layers were exported as geospatial files, with one raster file generated for each month.

The monthly layers allow for detailed temporal analyses of flaring activity, allowing comparison with seasonally variable distributions of aerofauna.

* **Processing Script**: ana\_offshore\_petroleum\_platform\_monthly.R
* **Output Files**: Rasterized monthly layers stored in workspace/data/analyzed/offshore\_petroleum\_platform\_monthly-1.0.0/.

##### Marine vessel traffic

###### Disturbance and pollution risk

The vessel traffic threat layers quantify the monthly density of different vessel types within the area of interest, providing insight into the spatial and temporal variability of shipping, fishing, and other vessel activities and their potential environmental impacts. This analysis utilizes Automatic Identification System (AIS) data from 2023 ([Section 2.2.3.1.4](#sec-shipping_ais)). The following processing steps were used to create these threat layers, but refer to [Section 2.2.3.1.4](#sec-shipping_ais) for more information on AIS data processing:

1. **Data Preparation**:
   * AIS data for ship locations and movements were ingested in parquet format and pre-processed to ensure compatibility with the study grid.
   * Auxiliary information on vessel types was included by merging with a reference dataset containing vessel classification (ntype).
2. **Spatial Filtering**:
   * Data points were filtered to retain only those within the area of interest, as defined by its bounding box.
   * Vessel movements were further processed to create continuous tracklines by connecting sequential AIS points for individual vessels.
3. **Temporal Grouping**:
   * Vessel tracklines were grouped by month, allowing for the calculation of monthly ship traffic densities.
4. **Density Calculation**:
   * Ship densities were calculated by rasterizing the trackline data onto the study grid. For each grid cell, the number of unique vessel occurrences was normalized by the number of days in the corresponding month to obtain a density assessment in number of ships per month in each grid cell.
5. **Vessel Type-Specific Analysis**:
   * Ship densities were categorized by vessel type (ntype), creating separate layers for different vessel classes.
6. **Export**:
   * Rasterized ship density layers were exported as geospatial files, with separate files generated for each month and vessel type.

* **Processing Script**: ana\_shipping\_intensity\_density.R
* **Output Files**: Monthly ship density layers stored in workspace/data/analyzed/shipping\_intensity\_density-1.0.0/.

###### Ship-based light pollution

The boat-based light pollution threat layers quantify the spatial and temporal distribution of artificial light emitted by boats at night. Two distinct data sources were utilized to create these layers: Automatic Identification System (AIS) data filtered for nighttime activities and VIIRS Boat Detection (VBD) data.

**Ship-based light pollution using AIS data**

This layer builds on the methods described in section [Section 2.2.3.2.5.1](#sec-shipping_threat) but focuses specifically on nighttime vessel activities. The following key differences in processing were applied:

1. **Nighttime Filtering**: AIS data were filtered to retain only ship positions recorded during nighttime, as determined by the day\_or\_night field in the processed AIS dataset (see [Section 2.2.3.1.4](#sec-shipping_ais)).
2. **Trackline Creation**: Tracklines were constructed using only nighttime data points, maintaining the methodology described previously.
3. **Density Calculation**: The density of nighttime shipping activities was calculated and normalized by the number of nights in each month, resulting in monthly raster layers providing the number of ships per month in each grid cell.

These layers highlight areas with intense nighttime activity for a range of vessel types, reflecting potential light pollution from ship-based operations.

* **Processing Script**: ana\_shipping\_night\_light\_density.R
* **Output Files**: Rasterized monthly layers stored in workspace/data/analyzed/shipping\_night\_light\_intensity\_density-1.0.0/.

**Ship-based light pollution using VBD data**

The second approach utilized VIIRS Boat Detection (VBD) data ([Section 2.2.3.1.5.1](#sec-viirs_boat_detection)), which directly measures the radiance of lights detected from boats at night. The following steps were applied:

1. **Data Preparation**: The VBD dataset was filtered to include only observations within the area of interest.
2. **Temporal Grouping**: Observations were grouped by month, with radiance values aggregated for each month.
3. **Rasterization**: Monthly radiance values were rasterized onto the study grid. Each grid cell contains the sum of radiance values for all detected boats within that cell across all years, divided by the number of years of data harvested, resulting in average total radiance per year in each grid cell.
4. **Masking and Export**: Rasterized layers were masked to the area of interest and exported as monthly geospatial files.

This layer provides a direct measurement of ship light emissions, offering a complementary perspective to the AIS-derived night lights layer.

* **Processing Script**: ana\_ship\_light\_detection.R
* **Output Files**: Rasterized monthly layers stored in workspace/data/analyzed/ship\_light\_detection-1.0.0/.

##### Coastal light pollution

The coastal light pollution threat layers were developed to quantify the intensity and spatial and temporal distribution of coastal light that may be affecting light-sensitive fauna in the nearshore environment. This analysis used the 2023 VIIRS Monthly Nighttime Light dataset ([Section 2.2.3.1.5.3](#sec-viirs_nightlight)), which provides high-resolution observations of nighttime light intensity. The data was analyzed using the following steps:

1. **Data Preparation**: Each monthly VIIRS nighttime light raster file was resampled to the study grid.
2. **Coastal Masking**: A coastal mask was applied to focus the analysis on regions near the coastline. The mask was created by:
   * Downloading and merging coastline data for Canada, the United States, and Greenland.
   * Buffering the coastlines by 1 km offshore and 5 km inland.
   * Masking the resampled nighttime light data.
3. **Temporal Aggregation**: Each monthly raster file was processed individually to extract nighttime light intensity values.
4. **Raster Export**: Processed raster files were exported as geospatial files, with one raster generated for each month in the dataset.

The night lights threat layer provides a spatially explicit representation of artificial light pollution along the coastlines, with monthly granularity.

* **Processing Script**: ana\_night\_light\_monthly.R
* **Output Files**: Rasterized monthly layers stored in workspace/data/analyzed/night\_light\_monthly-1.0.0/.

### 4. Project personnel:

* David Beauchesne
  + david.beauchesne@insileco.io
  + inSileco, Québec, Canada
* Kevin Cazelles
  + kevin.cazelles@insileco.io
  + inSileco, Québec, Canada

# Class III. Data set status and accessibility

## A. Status

1. Latest update:
2. Latest archive date: current
3. Metadata status: current
4. Data verification: current

## B. Accessibility

1. Storage location and medium: the complete dataset of threat layers is available on Zenodo at https://doi.org/TO/ADD.
2. Contact persons: David Beauchesne, email: david.beauchesne@insileco.io
3. Copyright restrictions: CC BY 4.0
4. Proprietary restrictions:
   1. Release date: When manuscript is accepted
   2. Citation: When using the dataset, please cite this article to recognize the effort involved in harvesting the primary data and preparing the threat layers.
   3. Disclaimer(s): The threat layers and methods presented in this project represent the best available data and methodologies known to us at the time of its development. They were specifically designed to align with the scope and objectives of this project, focusing on generating broad, spatiotemporal assessments of anthropogenic threats within Canada’s Exclusive Economic Zone (EEZ) and adjacent waters. The taxa for which these layers constitute threats are primarily marine birds, although threats may be broadly applicable to all marine taxa. While these layers provide valuable insights, they may not be suitable for all applications. Users are advised to critically evaluate their specific objectives and compare them with the scope, resolution, and limitations of this project to determine whether these threat layers meet their needs. Alternative datasets or methods may be more appropriate for generating highly detailed or application-specific threat assessments.
5. Costs: Costs associated with acquiring data (may vary by size of data request, desired medium, etc.)

# Class IV. Data Structural Descriptors

## A. Data Set File

1. Identity:
   * Threat Layers: Separate folders for each thematic threat layer (e.g., shipping\_night\_light\_intensity\_density, petroleum\_pollution) containing processed geospatial raster files in Cloud Optimized GeoTIFF (COG) format. File names indicate the specific type, temporal coverage, and subcategories (e.g., fisheries\_intensity\_longline\_2008\_10.tif, shipping\_night\_light\_intensity\_density\_2023\_5\_container.tif).
   * Raw Data: Processed datasets (e.g., istop.gpkg, nasp.gpkg) and tabular records (e.g., neec.csv) provided in GeoPackage, CSV, or Parquet format. Note that raw data are not part of the archived dataset, but referenced here for the reader only. The codebase allows access to openly available data, but not to proprietary data shared exclusively for this project.
   * Metadata and Scripts: Processing scripts (e.g., prc\_fisheries\_intensity.R) and metadata files provided in plain text or RMarkdown format.
2. Size:
   * Geospatial threat layers: ~300 MB total, with individual raster files ranging from 10 MB to 500 MB, depending on resolution and coverage.
   * Raw datasets: ~110 GB across multiple formats, including GeoPackages, CSVs, and Parquet files (not part of the archived dataset).
   * Metadata and scripts: ~1 MB, covering all workflows and metadata documentation.
3. Format and Storage Mode:
   * All threat layers were exported as Cloud Optimized GeoTIFFs (COGs)
4. Header Information:
   * Raster files: No headers.
5. Alphanumeric Attributes:
   * None
6. Special Characters/Fields:
   * None
7. Authentication Procedures:
   * Files include MD5 checksums for integrity validation during transfer.
   * File access for proprietary data (e.g., AIS) requires secure authentication via API keys.
   * No authentication required to access threat layers.

## B. Variable Information

1. Variable Identity:
   * Raster: Intensity or density values for threat layers (e.g., light\_radiance, vessel\_density).
2. Variable Definition:
   * Threat layers: Quantify the intensity or presence of threats (e.g., petroleum incidents, vessel traffic).
3. Units of Measurement:
   * Radiance (nW/cm²/sr for VIIRS datasets), density (vessels/km²), and pollution volume (litres for NEEC dataset), presence/absence.
4. Data Type:
   * Storage Type: Cloud Optimized GeoTIFFS (COGs)
   * Variable Codes: No explicit coding in rasters. Explicit names given to stored files.
   * Range for Numeric Values: Variable between dataset (e.g., radiance: 0–500 nW/cm²/sr, vessel density: 0–50/km²).
   * Missing Value Codes: NA
   * Precision: Raster precision varies between datasets.
5. Data Format:
   * Not applicable

## C. Data Anomalies

* Missing Data: Certain datasets (e.g., NEEC incidents) may have incomplete spatial or temporal coverage.
* Anomalous Data: Noise in satellite datasets (e.g., VIIRS light pollution) filtered using confidence thresholds.
* Calibration Errors: Historical incident records may lack consistent reporting methods.

# Class V. Supplemental descriptors

## A. Data acquisition

1. Data forms or acquisition methods: Not applicable
2. Location of completed data forms: Not applicable
3. Data entry verification procedures: Not applicable

## B. Quality assurance/quality control procedures

See above in section II.B.3. Research methods

## C. Related materials

Not applicable

## D. Computer programs and data-processing algorithms

The open-source software R 4.2.3 was used for all analyses (Team 2023) and the code to reproduce the codebase is available on Zenodo (https://doi.org/TO/ADD) and on GitHub (https://github.com/inSilecoInc/ceacws).

## E. Archiving

1. Archival procedures: The complete dataset is available on Zenodo at https://doi.org/TO/ADD
2. Redundant archival sites: None

## F. Publications and results

None

## G. History of data set usage

1. Data request history: none
2. Data set update history: Dataset modified on YYYY/MM/DD
3. Review history: None
4. Questions and comments from secondary users: Please direct any questions or comments to David Beauchesne (david.beauchesne@insileco.io)

# Literature Citations

Bureau of Ocean Energy Management. 2024. [United states offshore wind data - BOEM](https://www.boem.gov/renewable-energy/boem-renewable-energy-geodatabase). Bureau of Ocean Energy Management.

Canada-Newfoundland and Labrador Offshore Petroleum Board. 2024. Offshore petroleum data for newfoundland. C-NLOPB.

Canada-Nova Scotia Offshore Petroleum Board. 2024. Offshore petroleum data for nova scotia. CNSOPB.

Earth Observation Group, Payne Institute for Public Policy. 2024a. [VIIRS Boat Detection (VBD) Dataset](https://eogdata.mines.edu/products/vbd/). Earth Observation Group, Payne Institute for Public Policy.

Earth Observation Group, Payne Institute for Public Policy. 2024b. [VIIRS Night Fire (VNF) Dataset](https://eogdata.mines.edu/products/vnf/). Earth Observation Group, Payne Institute for Public Policy.

Earth Observation Group, Payne Institute for Public Policy. 2024c. [VIIRS Night Light (VNL) Dataset](https://eogdata.mines.edu/products/vnl/). Earth Observation Group, Payne Institute for Public Policy.

Elvidge, C. D., K. E. Baugh, M. Zhizhin, and F.-C. Hsu. 2013a. Why VIIRS Data Are Superior to DMSP for Mapping Nighttime Lights. Asia-Pacific Advanced Network 35:62.

Elvidge, C. D., M. Zhizhin, K. Baugh, F.-C. Hsu, and T. Ghosh. 2016. [Methods for Global Survey of Natural Gas Flaring from Visible Infrared Imaging Radiometer Suite Data](https://doi.org/10.3390/en9010014). Energies 9:14.

Elvidge, C. D., M. Zhizhin, T. Ghosh, and F.-C. Hsu. (n.d.). Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. Remote Sensing.

Elvidge, C. D., M. Zhizhin, F.-C. Hsu, and K. E. Baugh. 2013b. [VIIRS Nightfire: Satellite Pyrometry at Night](https://doi.org/10.3390/rs5094423). Remote Sensing 5:4423–4449.

Environment and Climate Change Canada. 2024. Integrated satellite tracking of pollution (ISTOP) program. Dataset, Environment and Climate Change Canada.

Impact Assessment Agency of Canada. 2024a. [Focus area for the regional assessment of offshore wind development in newfoundland and labrador](https://open.canada.ca/data/en/dataset/81560d1e-8394-4b51-8212-2bbe4d8a5ea2). Government of Canada.

Impact Assessment Agency of Canada. 2024b. [Study area for the regional assessment of offshore wind development in nova scotia](https://open.canada.ca/data/en/dataset/051c7034-c13b-49a9-9d4a-99b0b594d157). Government of Canada.

National Environmental Emergency Centre. 2024. Atlantic (including Quebec) notifications impacting or potentially impacting a waterbody with select substances, 2016-2023. Extracted by Philippe Muise 2024-01-29. Environment and Climate Change Canada.

Team, R. C. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Transport Canada. 2024. National Aerial Surveillance Program (NASP). Data extraction for period between April 2012 to March 2023. Dataset, Transport Canada.

Transports Canada. 2023. Satellite navigation data from automatic identification system (AIS) in atlantic canada, 2023. Data provided by transports canada. Transports Canada.

Veinot, T., A. Nicoll, K. Rozalska, and S. Coffen-Smout. 2023. [Vessel Density Mapping of 2019 Automatic Identification System (AIS) Data in the Northwest Atlantic. Can. Tech. Rep. Fish. Aquat. Sci. 3520: Vi + 29 p.](https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/41105163.pdf) Pages vi + 29 p. Department of Fisheries and Oceans.

Zhizhin, M., A. Matveev, T. Ghosh, F.-C. Hsu, M. Howells, and C. Elvidge. 2021. [Measuring Gas Flaring in Russia with Multispectral VIIRS Nightfire](https://doi.org/10.3390/rs13163078). Remote Sensing 13:3078.