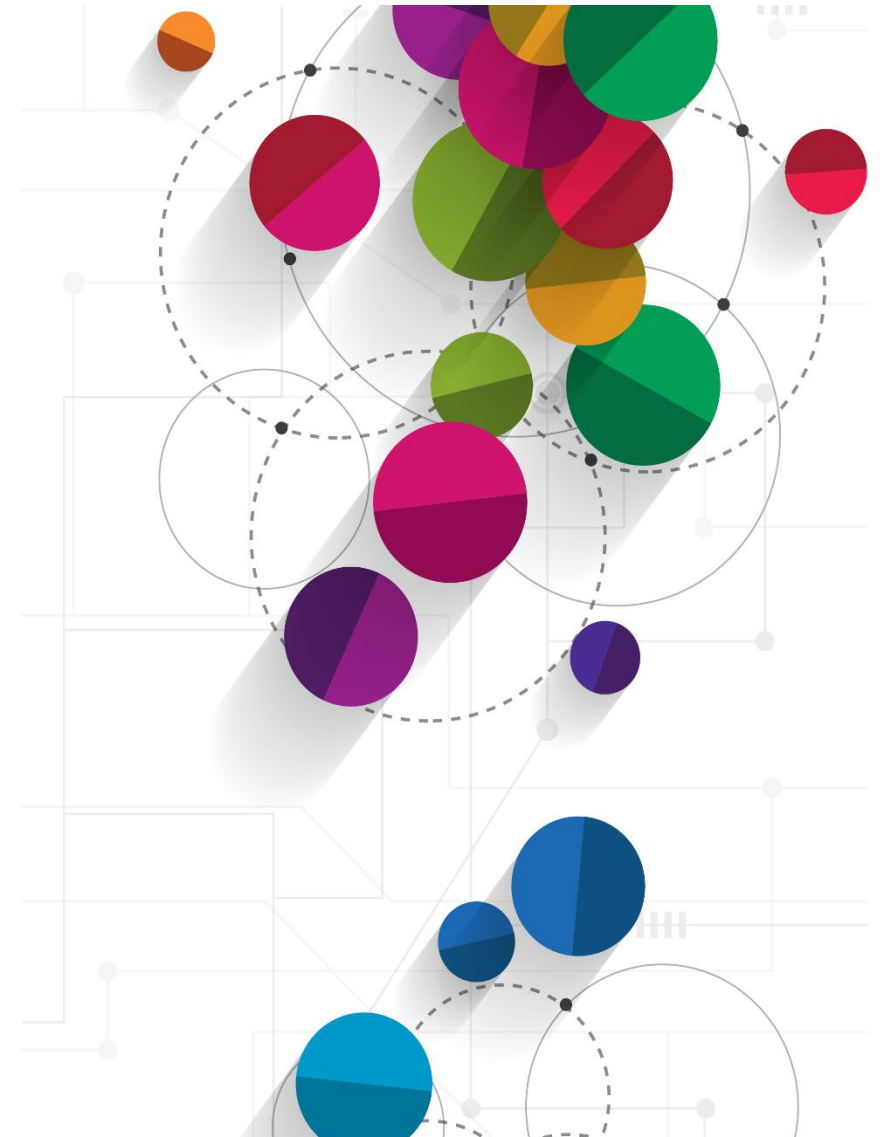




Machine Learning Models for Ocean Data

Quality Control and Anomaly Detection in CTD Data





What is CTD Data?

- Large quantities of oceanographic data are collected as Conductivity, Temperature, Depth (CTD) profiles
 - A CTD profile is a sequence of sensor scans taken across a depth range as the equipment is lowered through the water
 - Minimally, conductivity, temperature and pressure measures are directly collected through the sensors
 - Other measures, such as salinity, can later be derived from the collected data
- Analysis of CTD data supports the departmental mandate to apply sound science for the protection of oceans and aquatic ecosystems from negative impacts
 - CTD data represents a foundational source of physical oceanographic data contributing to national and international programs



Source: Fisheries and Oceans Canada Youtube channel. "The Rosette"



Building a Suite of CTD Machine Learning Tools

- The Office of the Chief Data Steward is developing a suite of machine learning models to assist in the processing and study of CTD data
- Two use cases are currently being developed for use in production:
 1. Quality control of CTD profile scans
 2. Detection of anomalous CTD profiles
- This work is being done in collaboration with the Pacific Region Ocean Sciences Division



Quality Control of CTD Profile Scans



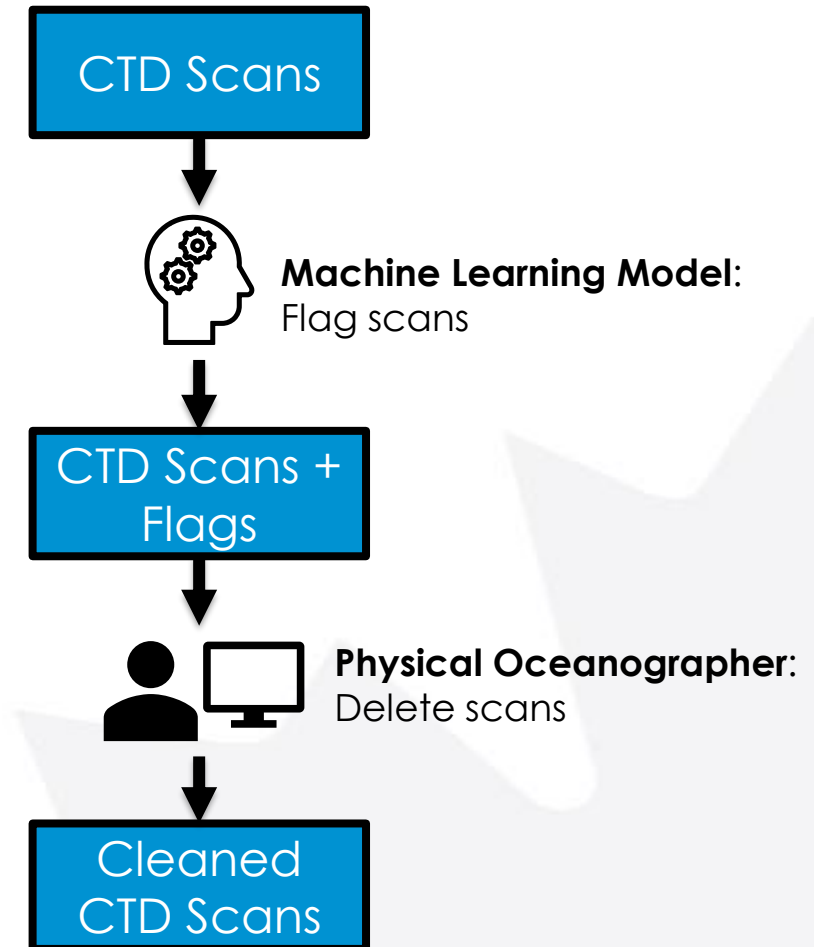
Problem: CTD data must undergo rigorous quality control, however, human verification of the data is slow and extremely time-consuming



Use Case Goal: Apply machine learning models to alleviate manual burden from oceanographers and help to speed up the CTD quality control process



Business Value: Flagged scans allow the oceanographer to quickly focus attention on crucial areas, reducing the time and effort required to delete scans





Detection of Anomalous CTD Profiles



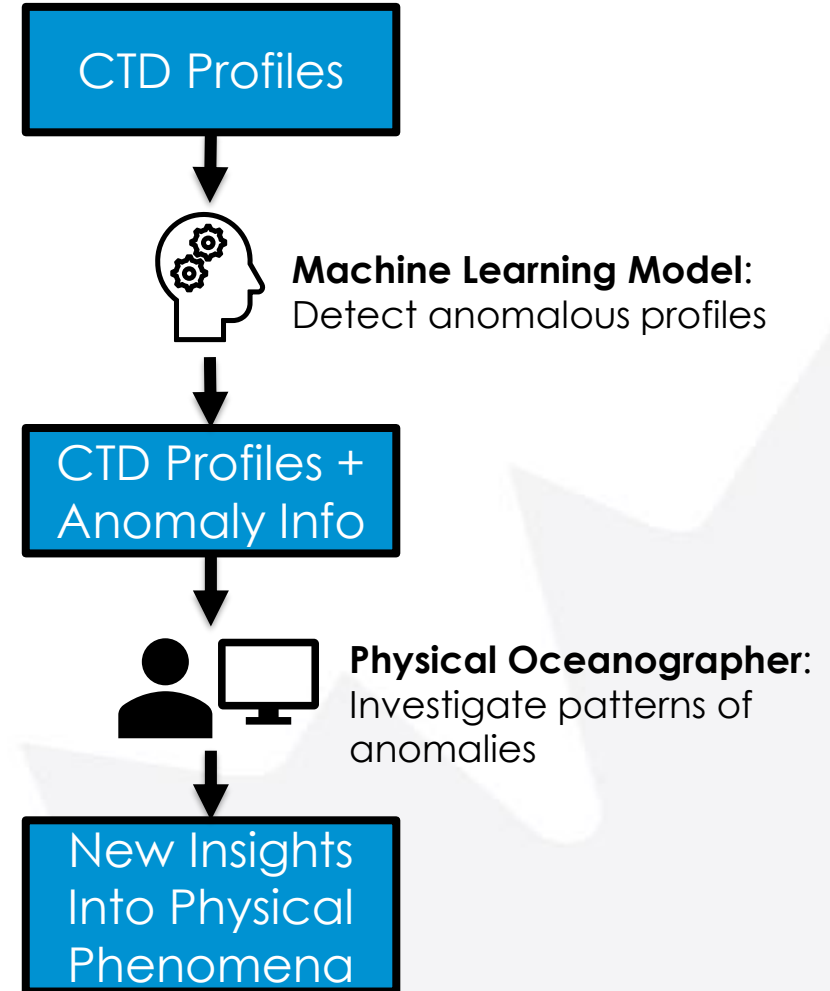
Problem: Analysis of CTD data can be complex and the data features required to detect patterns may go beyond what humans and standard analytical models can reasonably examine



Use Case Goal: Apply machine learning models to detect anomalous CTD profiles, highlighting meaningful physical phenomena for closer investigation



Business Value: Anomaly detection models combine large amounts of data too detailed for human inspection and allow for the identification of anomalies of an unspecified nature, filling in gaps where other more specific models may overlook significant changes in data





Use Case Roadmap

Use Case Planning

- Use case formulation
- Consultations with client
- Problem and solution domain definition

Field Test

- Model deployment
- Business process integration
- User testing

Proof of Concept

- Data cleaning/processing
- Feature engineering
- Model evaluation/tuning

Production

- Production environment development
- Full-scale deployment
- Monitoring and support

The CTD quality control use case is now in preparation for production

The CTD anomaly detection use case is now beginning field testing



Office of the Chief Data Steward / Bureau de l'Intendance principale des données

Strategic Policy
Politiques Stratégiques

DFO.OCDS-BIPD.MPO@dfo-mpo.gc.ca