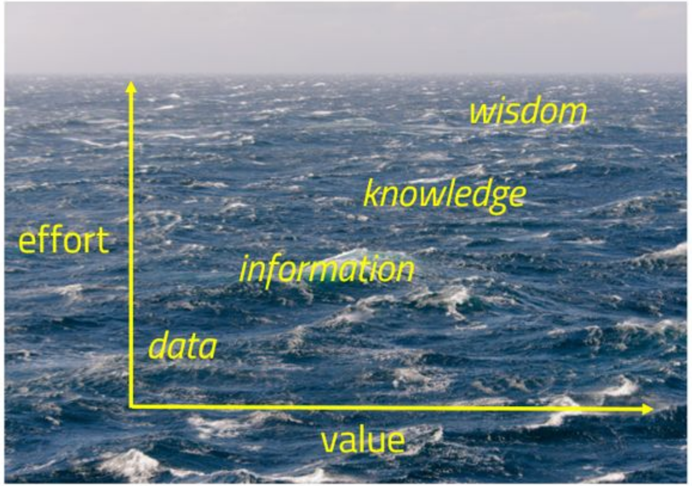
A unified data model to facilitate operational outcomes

Geoff Wake, OceanComm

Data requires context to become useful



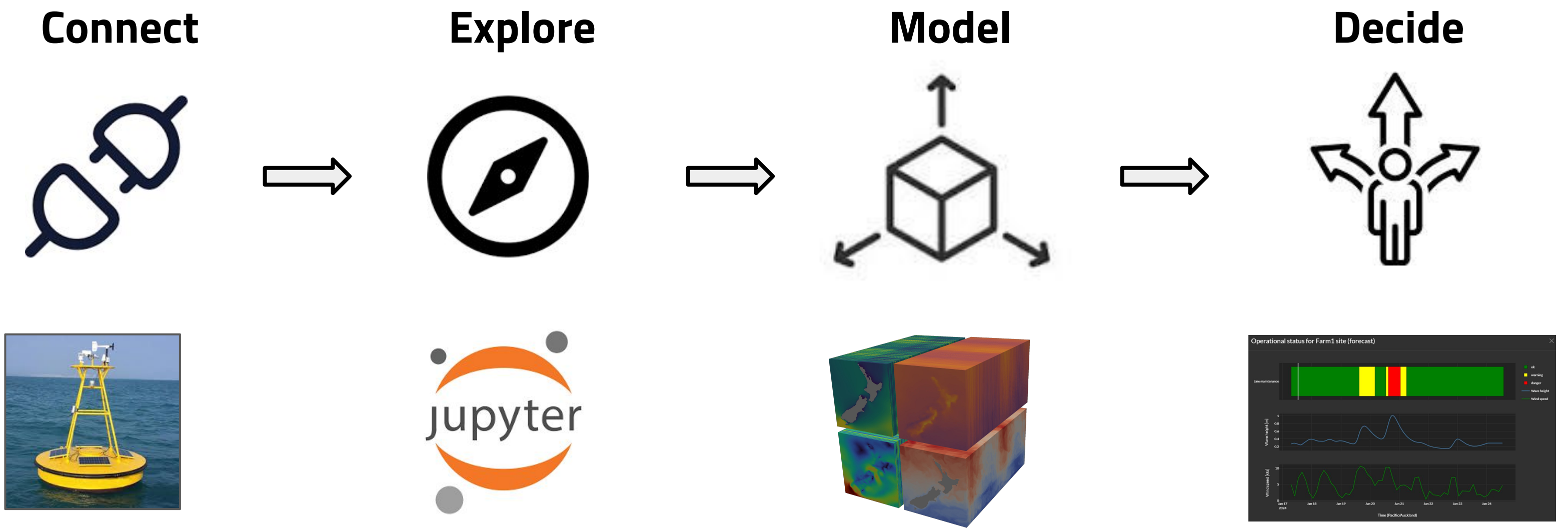
context that can only be understood by humans

↑

data that can only be read by machines

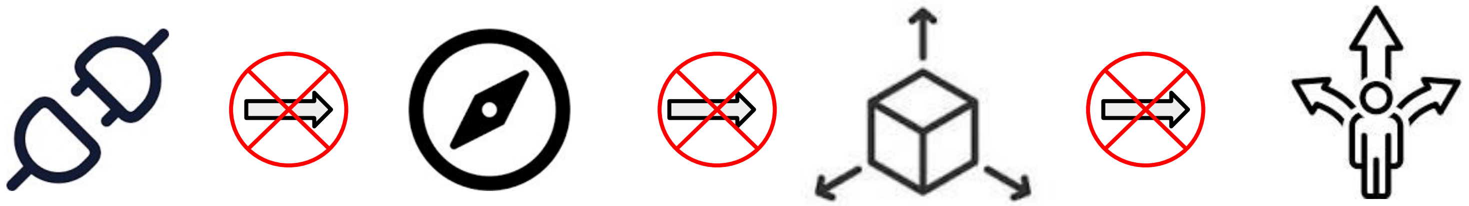
Well structured, machine accessible data is needed to leverage the emerging generative AI opportunities.

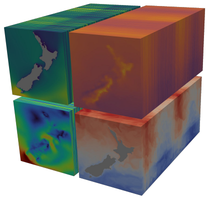
Our collective mission is transforming data into decisions

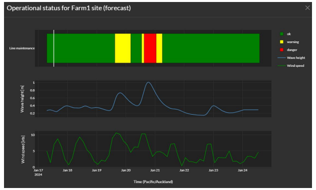


Our collective mission is transforming data into decisions

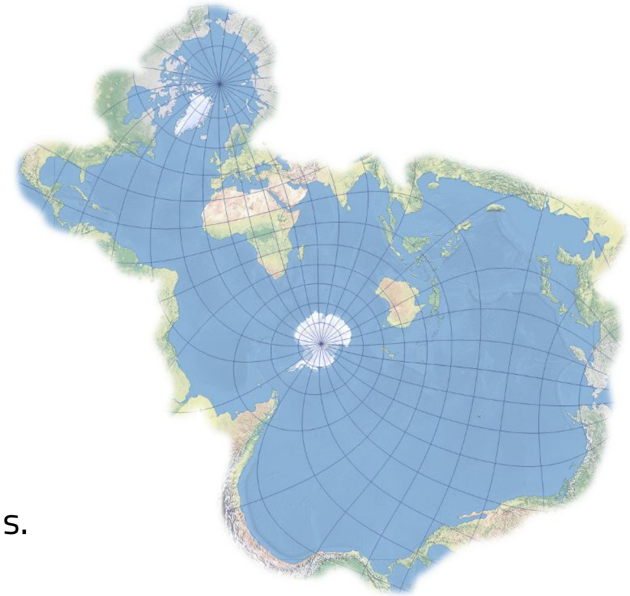
Connect **Explore** **Model** **Decide**







There are oceans of data!

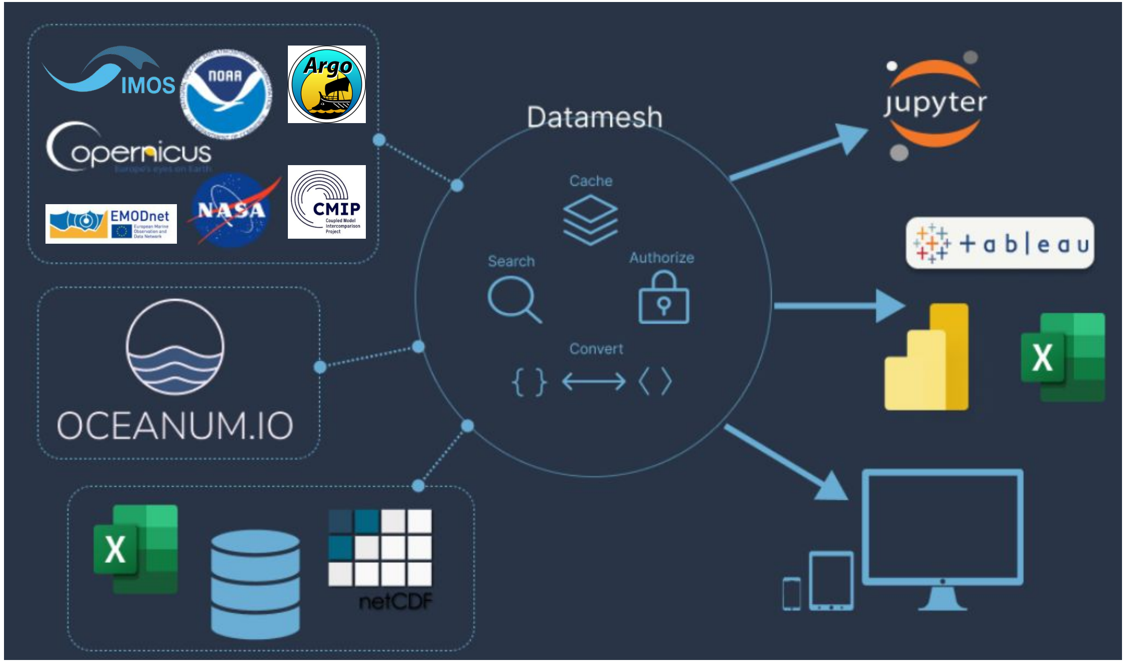


- But it's often hard to find and even more difficult to access.
- It might be poorly described or in a weird format.
- People typically spend a lot of time just getting it ready for analysis.
- Most data is not in a suitable format for AI /ML applications.
- Creating operational outcomes from real time data requires a lot of data engineering.

OCEANUM Datamesh was built to solve many of these challenges leverage a wealth of modern data science tools for use in operational oceanography.

Datamesh - it's not a 'lake'

Everything registered in Datamesh is fitted to a unified data model.



- Secure
- Searchable
- Extendable
- Flexible metadata
- Spatial / temporal
- Data sovereignty
- Parallel processing
- Format conversions
- User interface
- Machine interface
- Fully federated
- Storage options
- Authorization options

Practical examples

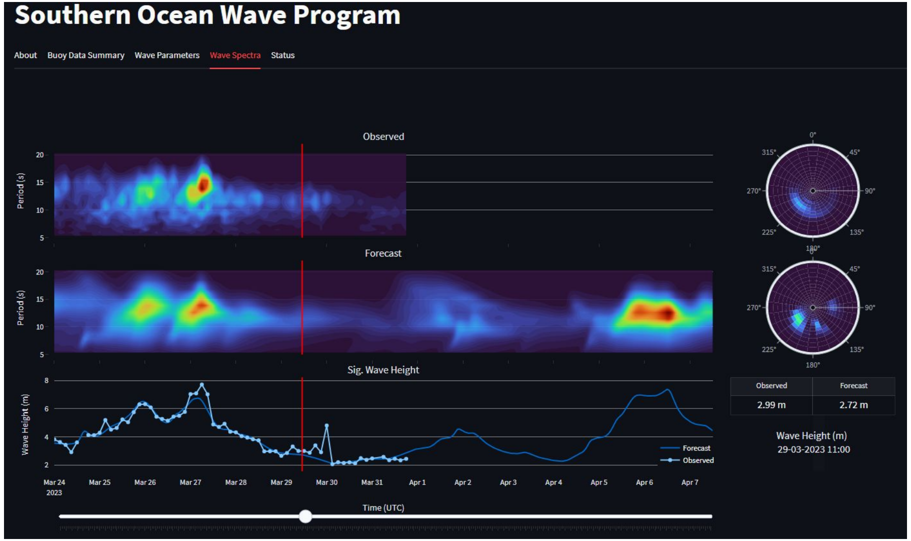
Display real time wave spectra

Use the API to automate, for example:

- Connect to real time wave buoy spectra
- Connect to forecast spectra
- Reprocess both sets of spectra
- Make spectrograms and co-plot all data
- Deploy into operations

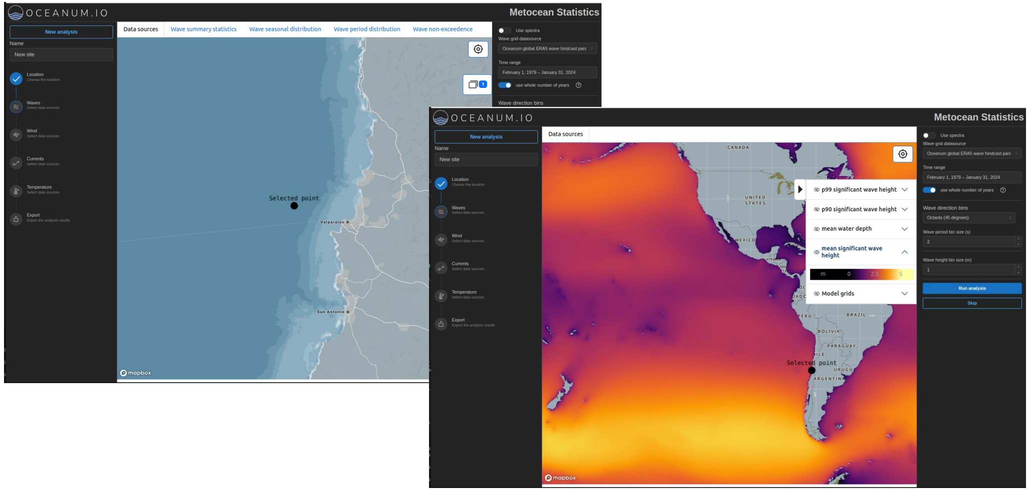
Functions to plot the data

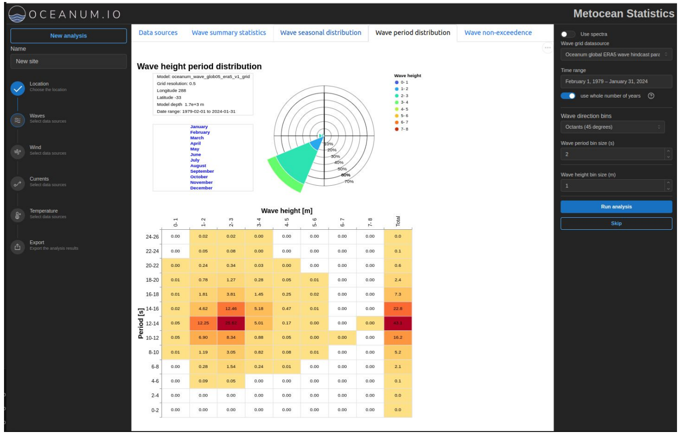
```
In [47]:
def plot_map(df, datasource_id, ax):
    """Plot map showing the location at the NDBC datasource."""

    # Bathy contour
    c = cali.depth.plot.contour(ax=ax, levels=[300, 3000], colors="0.0", linewidths=0.5, transform=TRANSFORM)
    c.clabel(fmt="%0.0fm", fontsize=6)

    # Buoy location
    ax.plot(xb, yb, "o", color="#187bcd", alpha=0.3, markersize=5, transform=TRANSFORM)
    ax.plot(df.lon_obs, df.lat_obs, "o", color="#187bcd", markersize=5, alpha=0.7, transform=TRANSFORM, label="NDBC")
    ax.plot(df.lon_hds, df.lat_hds, "o", color="#d84418", markersize=5, alpha=0.7, transform=TRANSFORM, label="Hindcast")
    ax.legend(loc=1)

    # Show features
    ax.add_feature(cfeature.LAND.with_scale("10m"))
    geopandas.GeoSeries(ds_hds.geom).plot(ax=ax, facecolor="none", edgecolor="0.3", alpha=0.5, transform=TRANSFORM)

    # Axis settings
```

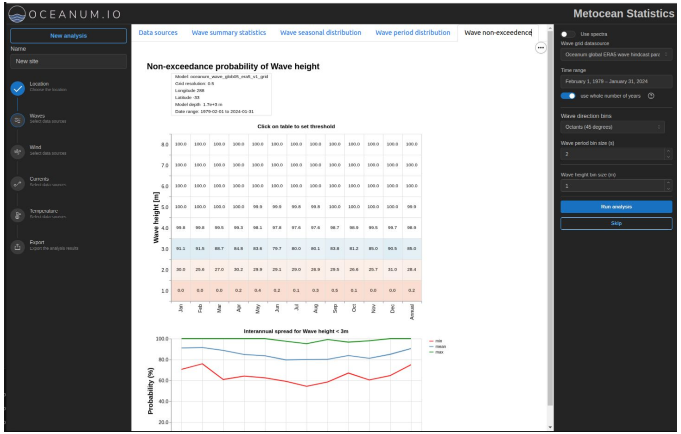


Rapidly create metocean statistics

Create metocean statistics from curated hindcasts

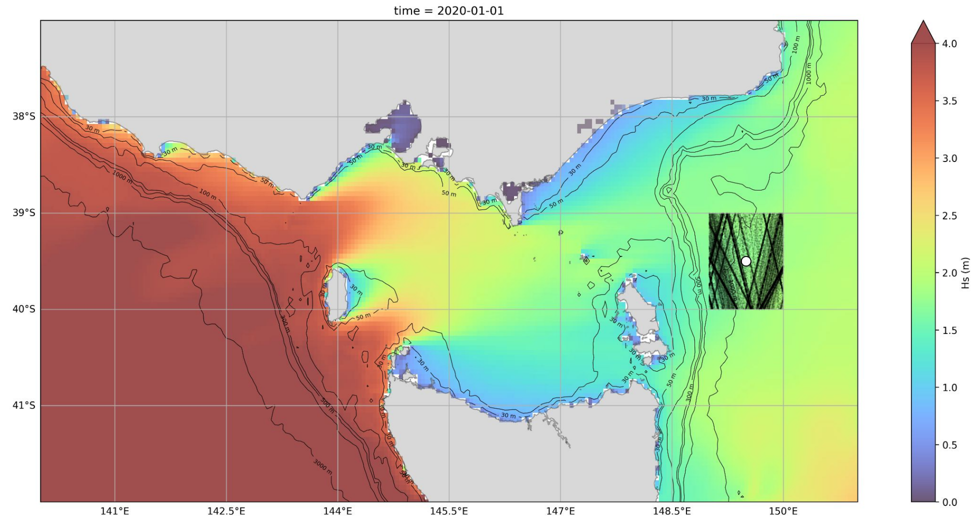


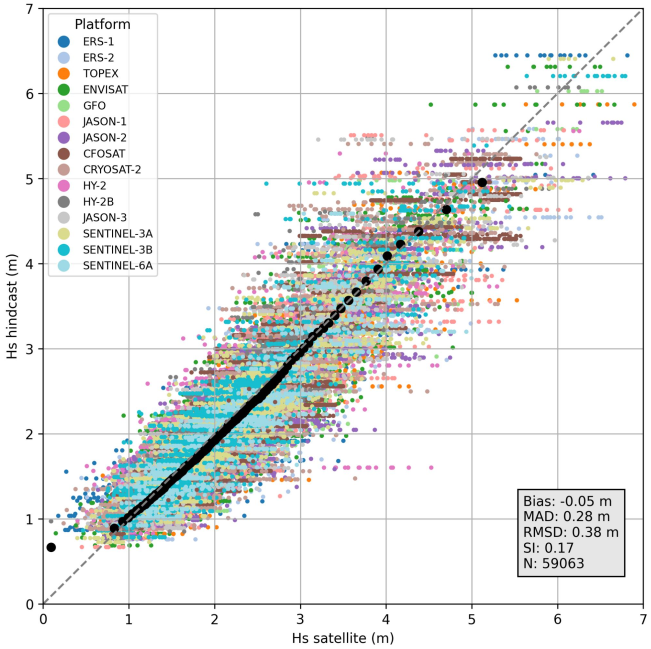




Validate hindcasts in a few seconds

- Bass Strait 1 km wave hindcast
- [Ribas & Young (2019)] 33-year altimeter
- Run via the App or [Jupyter Notebook]





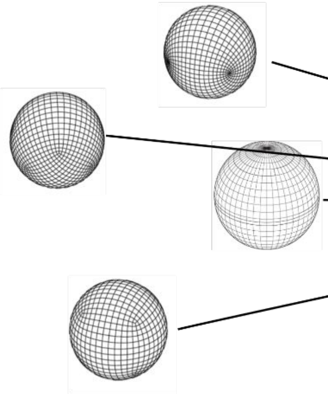
Interrogate CMIP6

The CMIP6 climate projections are massive datasets with many different structures and formats.

We register key CMIP6 experiments as raw data for dynamical downscaling or as bundles of collocated key variables, with easy access for non-expert users.

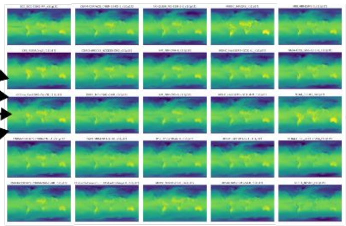
CMIP6

Many centers, many models, many grids, many experiments, many variables distributed as independent datasets by ESGF and Pangeo



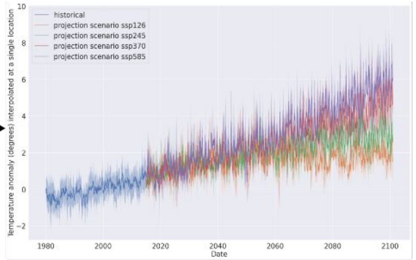
Oceanum CMIP6 API
+
datamesh driver

Interpolated datacube collating many realisations and variables for a given experiment

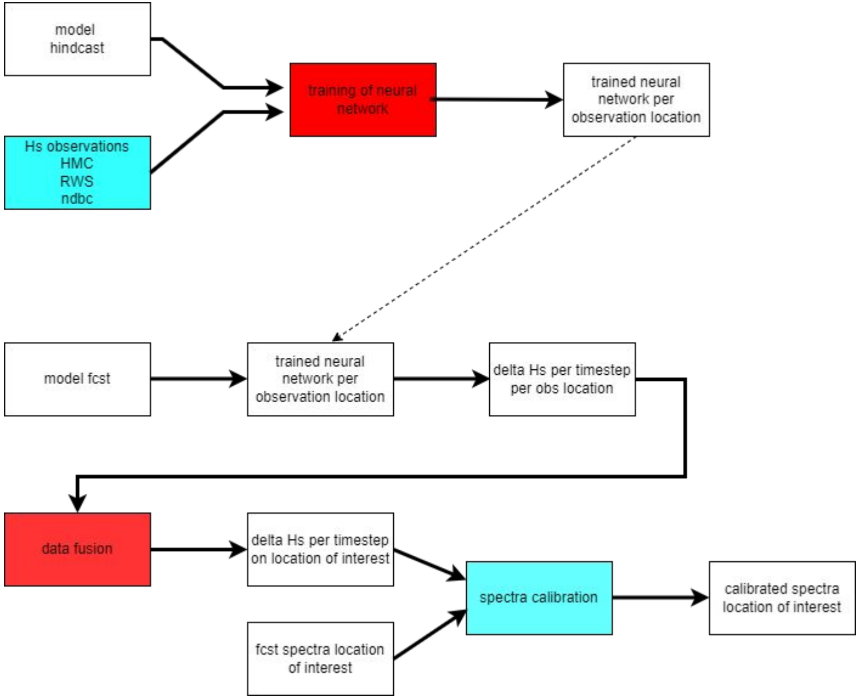


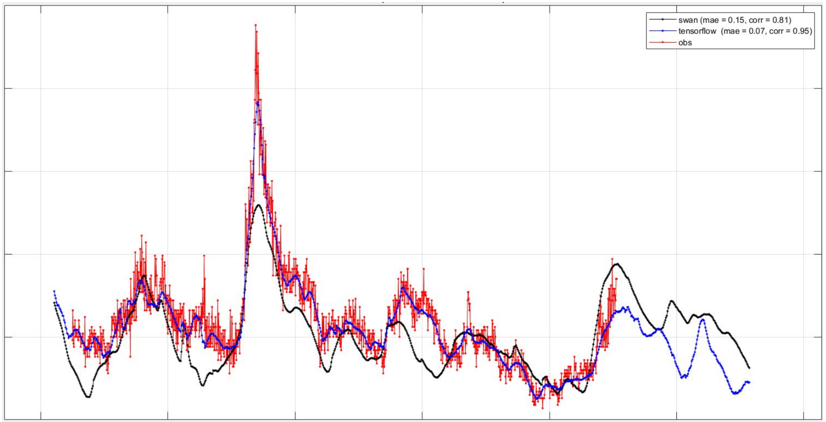
Oceanum datamesh queries
+
a few lines of python code

Example output: Here we present a plot of surface temperature from 4 CO_2 emission scenarios (2015-2100), plus the historical period from 1980 to 2015. These can be created in a few minutes.



Add value with ML and data fusion





Combine historical hindcasts with observations for improved forecasts

Ask specific questions of data

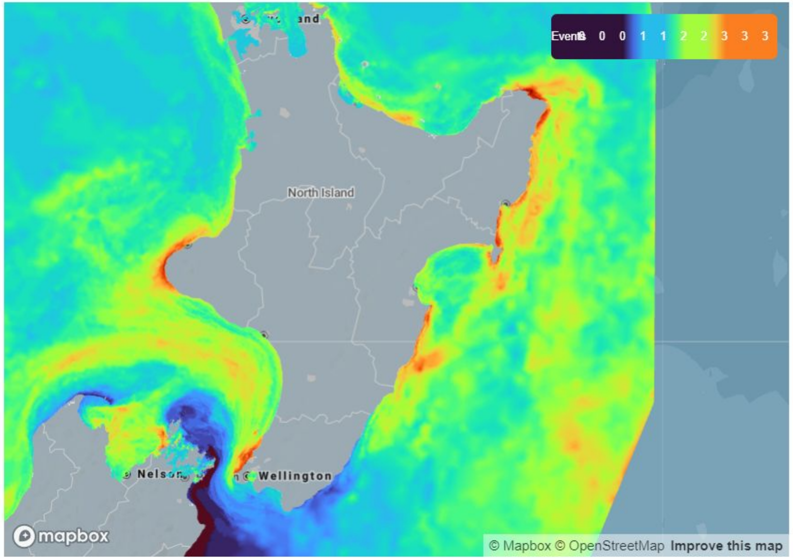
all seasons

Events longer than: 7 days

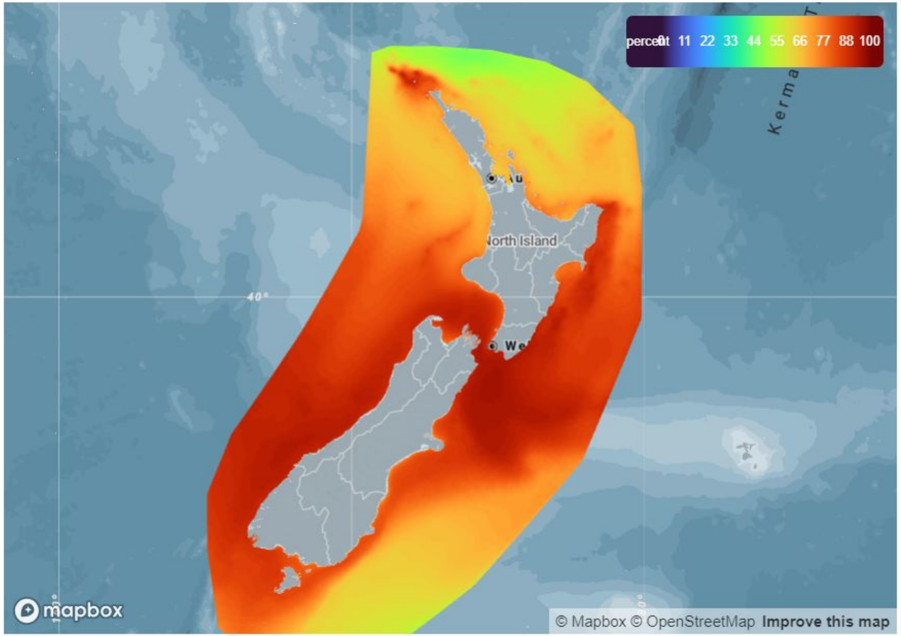
Values: above

Temperature: 18 deg.

Depth: 0 m

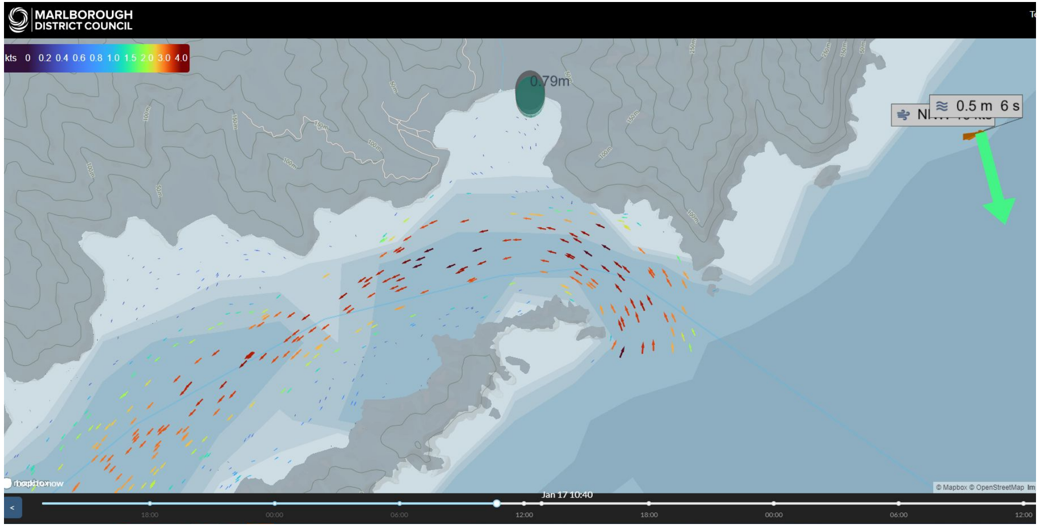


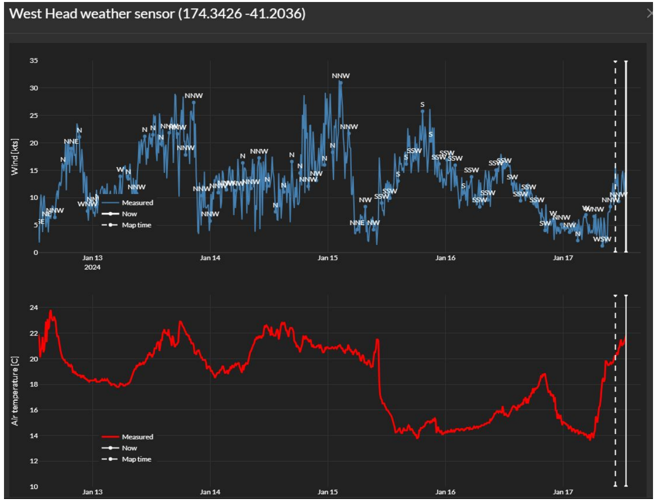
Number of events per year

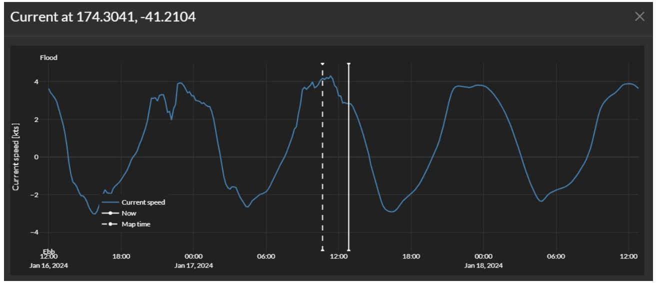


Salmon growth rate

Make real time navigational decisions



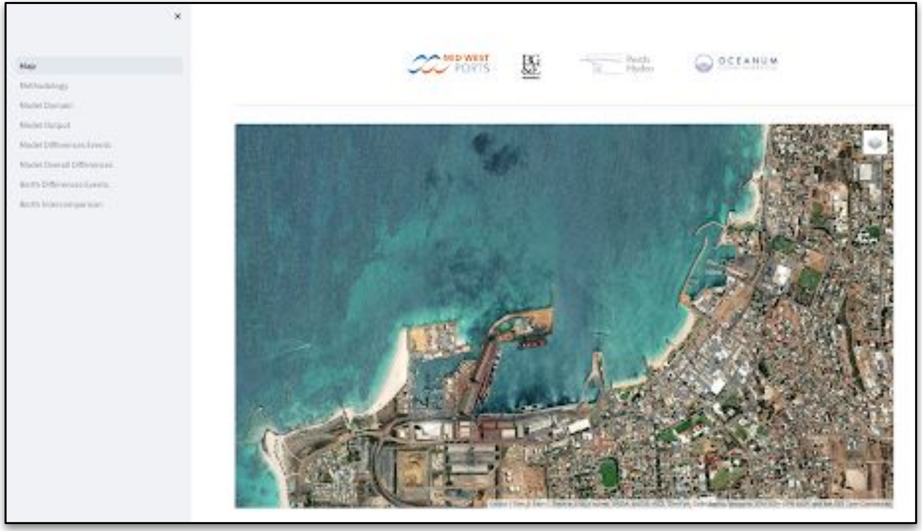


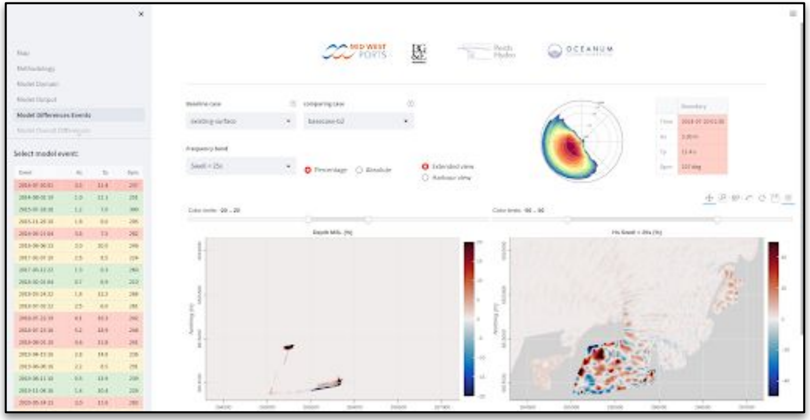


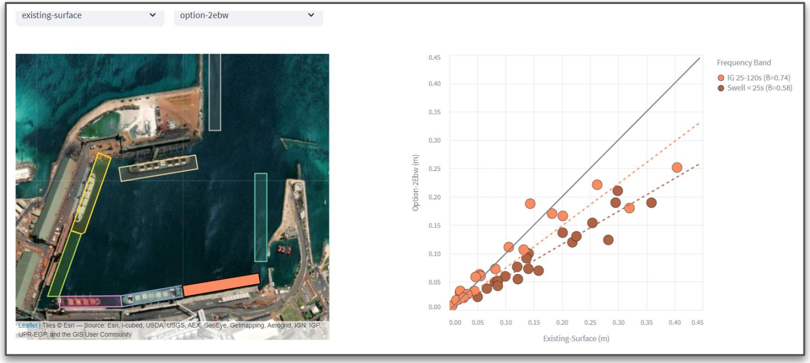
The complex tidal and non-tidal flows in the Tory Channel are resolved by a hybrid ML model using real time water level data.

Make real time navigational decisions

Use interactive tools to present complex model results

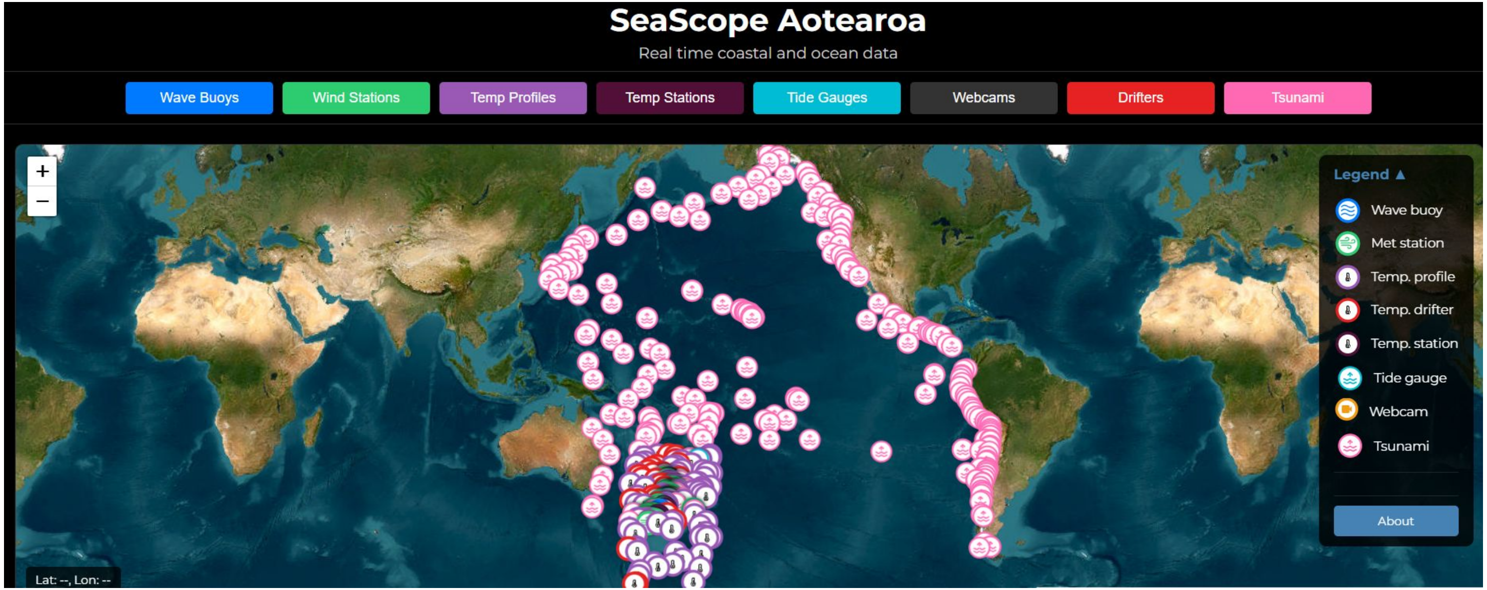






Effects of port modifications on the harbour tranquility can be analysed by the range of stakeholders.

Seascope Aotearoa - an open access community project for real time data

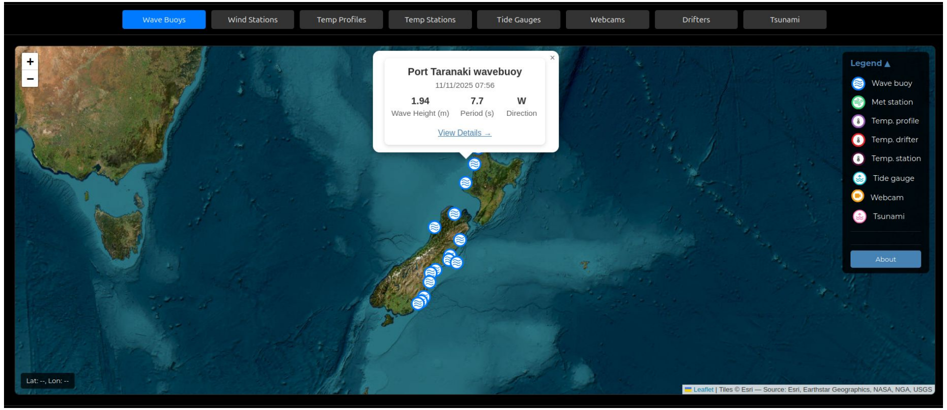


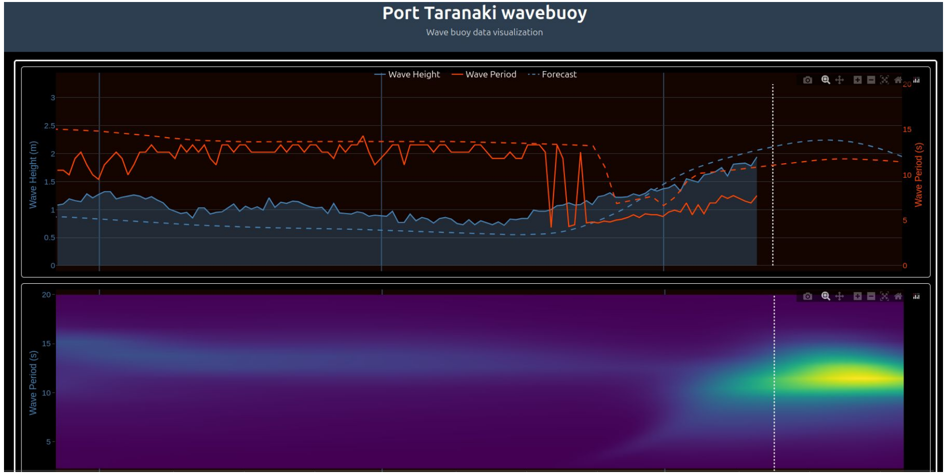
Connecting all the open data for waves, winds, temperatures, tides, tsunami and more

Seascope Aotearoa - waves

Real time wave data:

- Spotter, Triaxys, Obscape
- Co-plotted with nowcast / forecast
- Interactive spectrograms
- Provided by ports, councils and projects

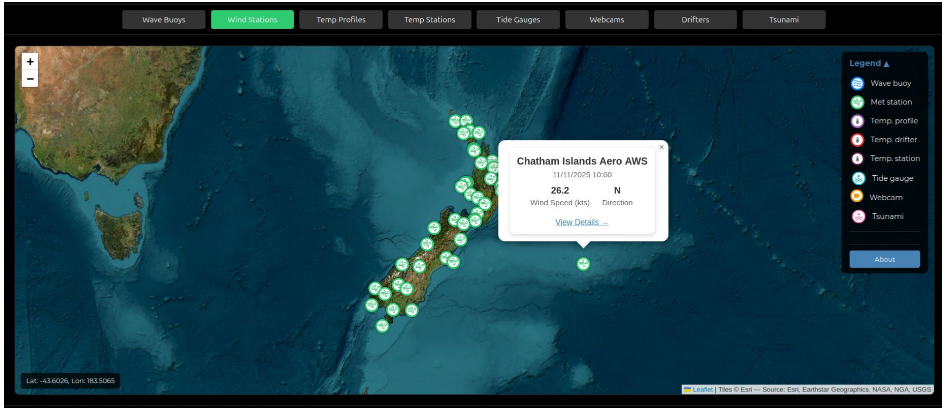


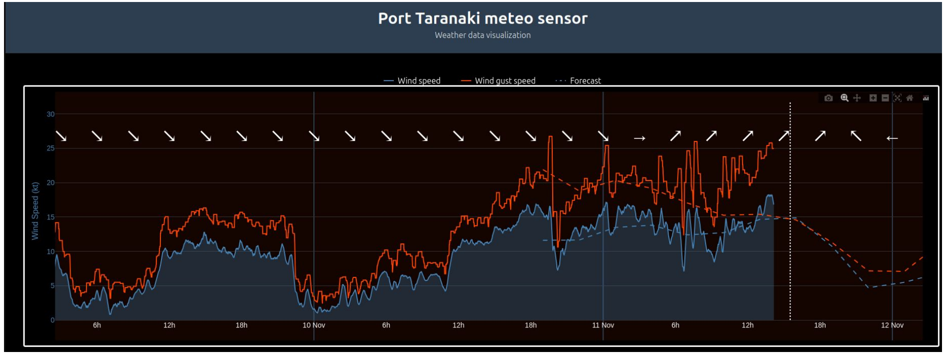


Seascope Aotearoa - winds

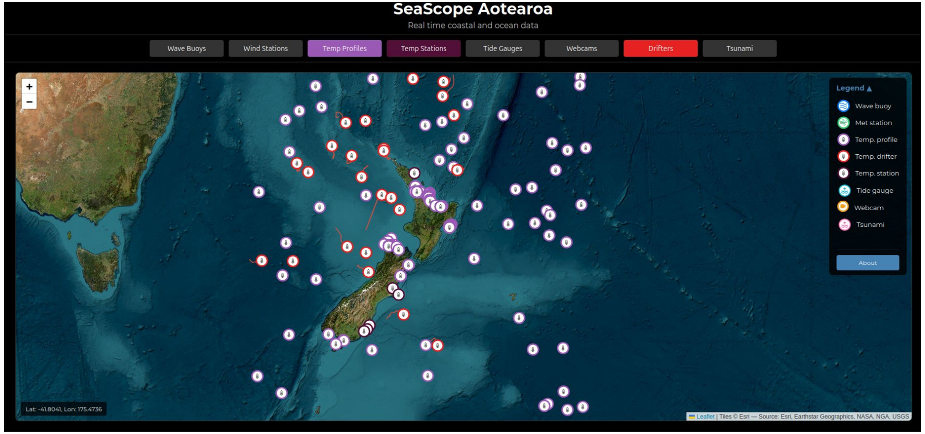
Real time wind data:

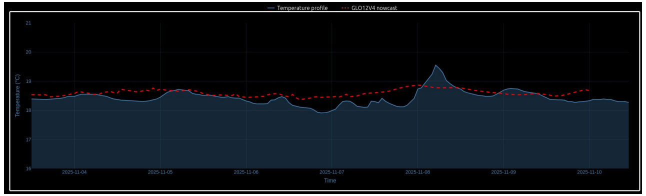
- Various frequencies
- Co-plotted with nowcast / forecast
- Interactive graphs
- Provided by ports, councils and projects

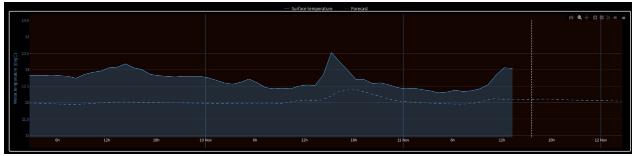




Seascope Aotearoa - temperature

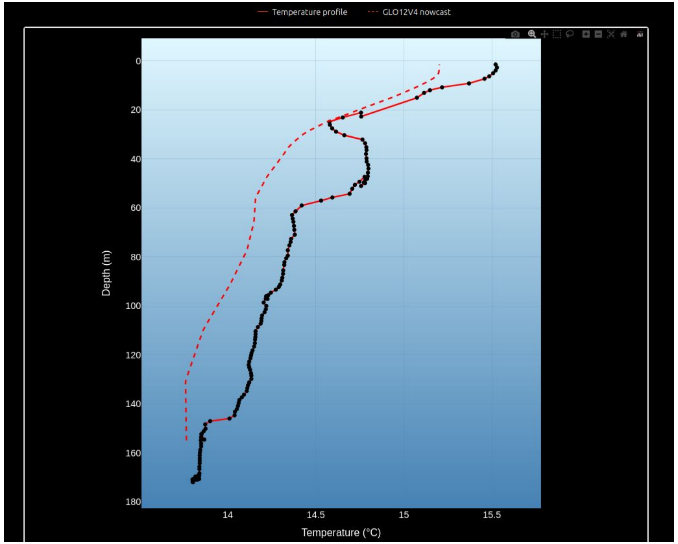




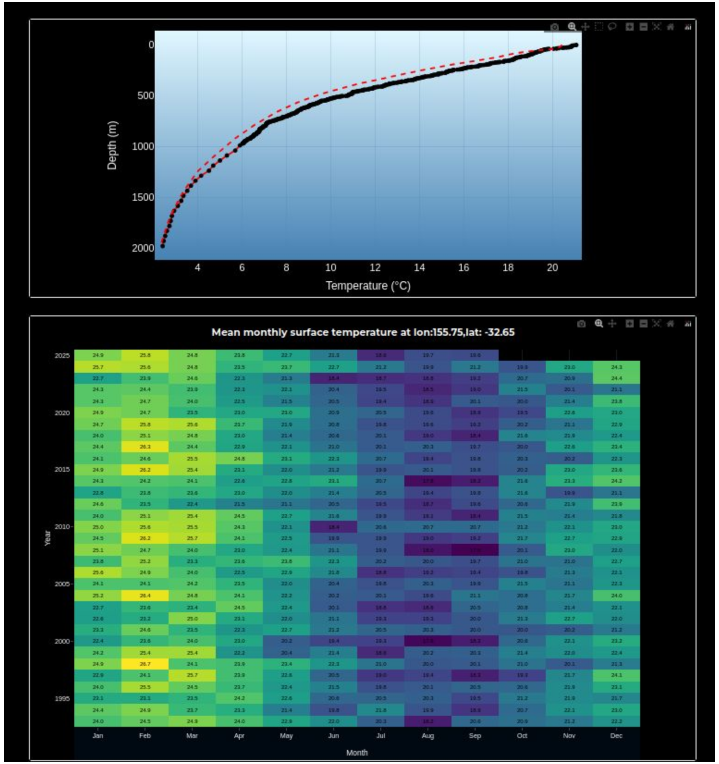


- ARGO, Drifters and FVON.
- Co-plotted with nowcast / forecast

Seascope Aotearoa - temperature

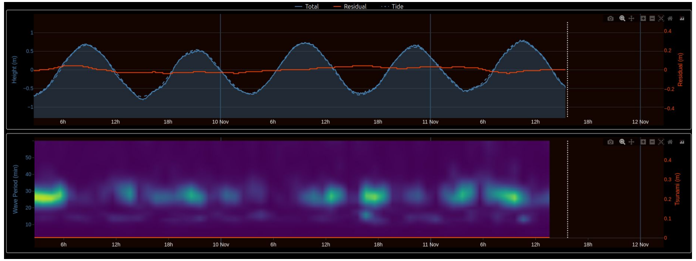


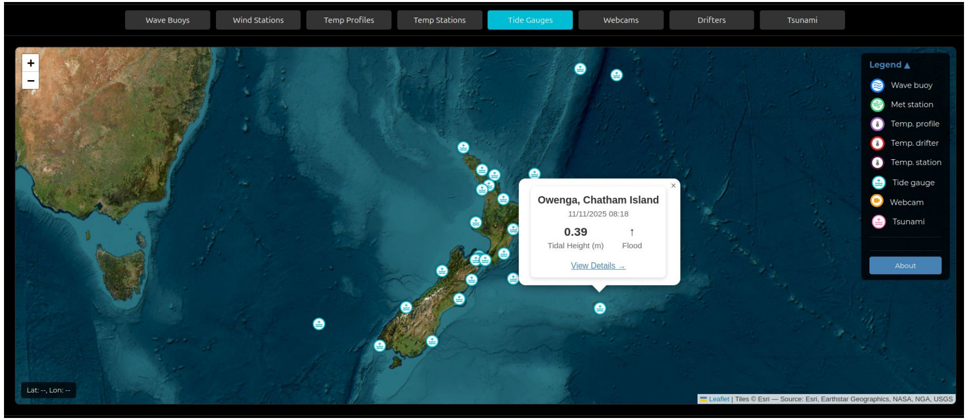
Compare measured and modelled profile



Monthly climatologies

Seascope Aotearoa - tides





Tidal Analysis

Tidal Constituents

Name	Amplitude (m)	Phase (degrees)	Description
M2	0.47	211.7	Principal lunar semidiurnal constituent
N2	0.12	245.2	Lunar elliptic semidiurnal constituent
O1	0.03	143.0	Principal lunar diurnal constituent
K1	0.02	94.7	Lunisolar diurnal constituent
NU2	0.02	237.6	Lunar elliptic semidiurnal constituent
S2	0.02	8.3	Principal solar semidiurnal constituent
MU2	0.02	264.3	Lunar elliptic semidiurnal constituent
2N2	0.02	272.6	Shallow water compound of N2
MF	0.02	185.0	Lunisolar fortnightly constituent
MSM	0.01	168.2	Mean Solar
K2	0.01	325.6	Lunisolar semidiurnal constituent

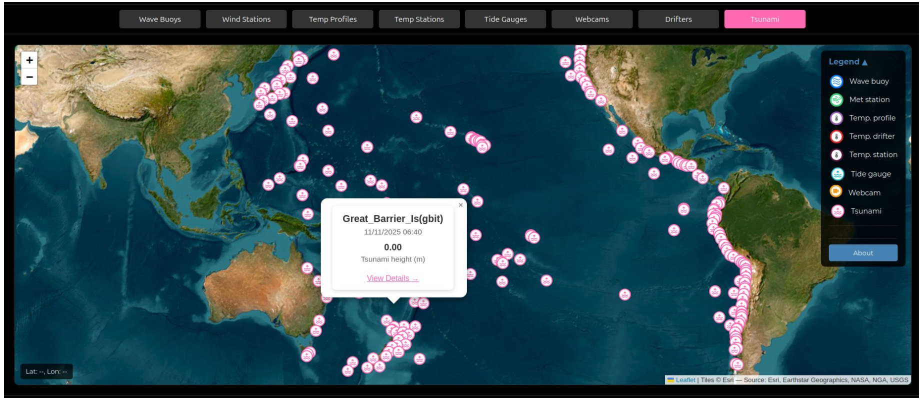
Tidal Datums

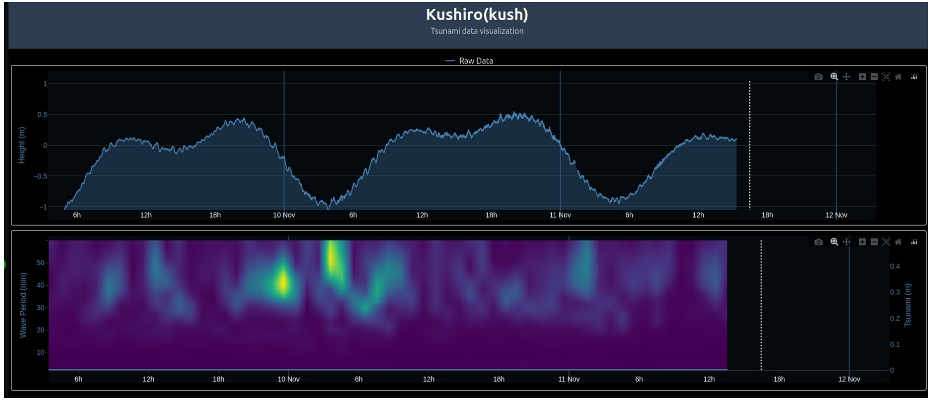
Name	Formula	Value (m)
Mean High Water	MSL + Σ(Semidiurnal + Diurnal)	0.74
Mean High Water Spring	MSL + M2 + S2	0.49
Mean High Water Neap	MSL + |M2 - S2|	0.45
Mean Sea Level	Reference level (0m)	0.00
Mean Low Water Neap	MSL - |M2 - S2|	-0.45
Mean Low Water Spring	MSL - M2 - S2	-0.49
Mean Low Water	MSL - Σ(Semidiurnal + Diurnal)	-0.74

Seascope Aotearoa - tsunami

Real time tsunami gauge data:

- 1-min data updates
- >300 stations within the Pacific
- Plotting raw and processed data
- Spectrograms to show periodicity
- ZDC analysis to show max tsunami height observed over the previous hour





Conclusion

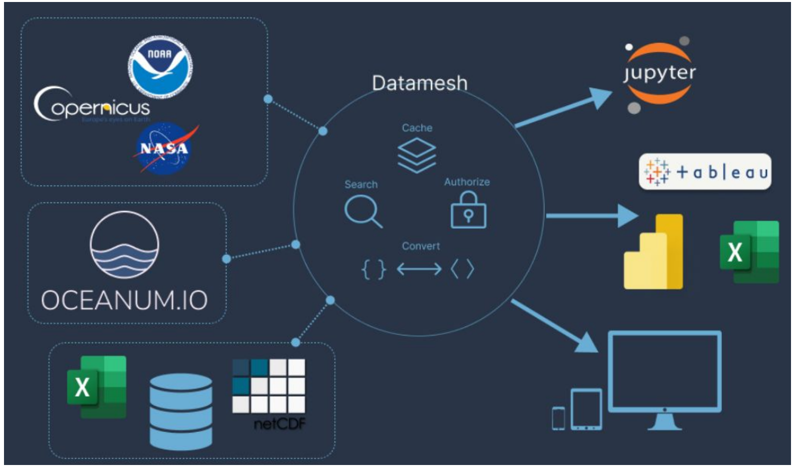
Conclusion

The emerging use cases for AI require accessible, well-structured data.

A solution has been developed to unify disparate types of data and facilitate rapid development and innovation in operational oceanography.

We are enabling users in industry, government, military, university, projects, small businesses and applied research.

The Datamesh is future-focussed on the expected demands for the coming years.



Thanks for your time