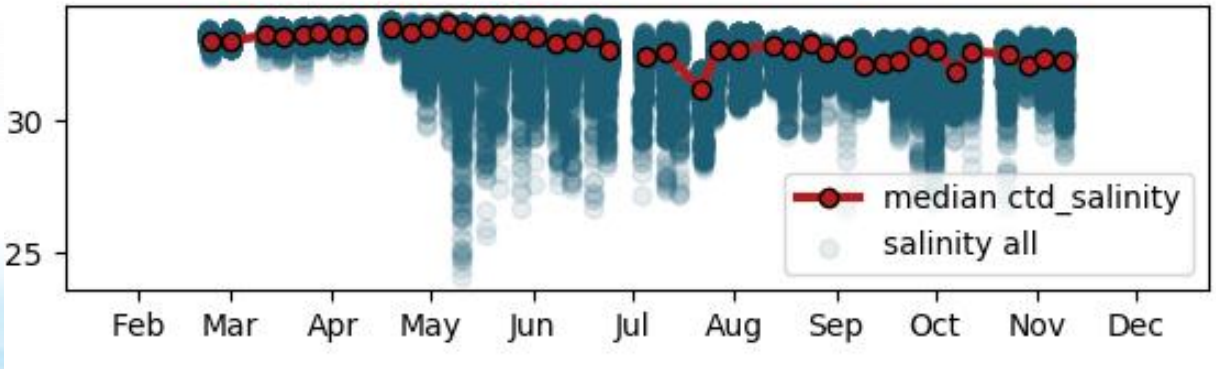
A new approach to automatic quality control and API access to FerryBox data: application usage by assimilating data into ocean models

Trond Kristiansen, Kai Sørensen, Pierre Jaccard, Elizaveta Protsenko, Zofia Rudjord, Ann Kristin Sperrevik, and Andrew King



NIVA - Ferrybox streaming of data

- Real-time streaming of data from Ferrybox to NIVA database
- Automatic quality control of data
- Access to data available through Python API
- Data can be assimilated into ocean circulation models using the API to improve weather predictions

NIVA - Quality Control

- **Quality control** is essential to building a successful business that delivers products that meet or exceed customers' expectations.

- Data quality is maintaining and assuring the **accuracy** and consistency of **data** over its entire life-cycle. **Data** quality means that the **data** is **accurate** and reliable.

NIVA - Quality Control

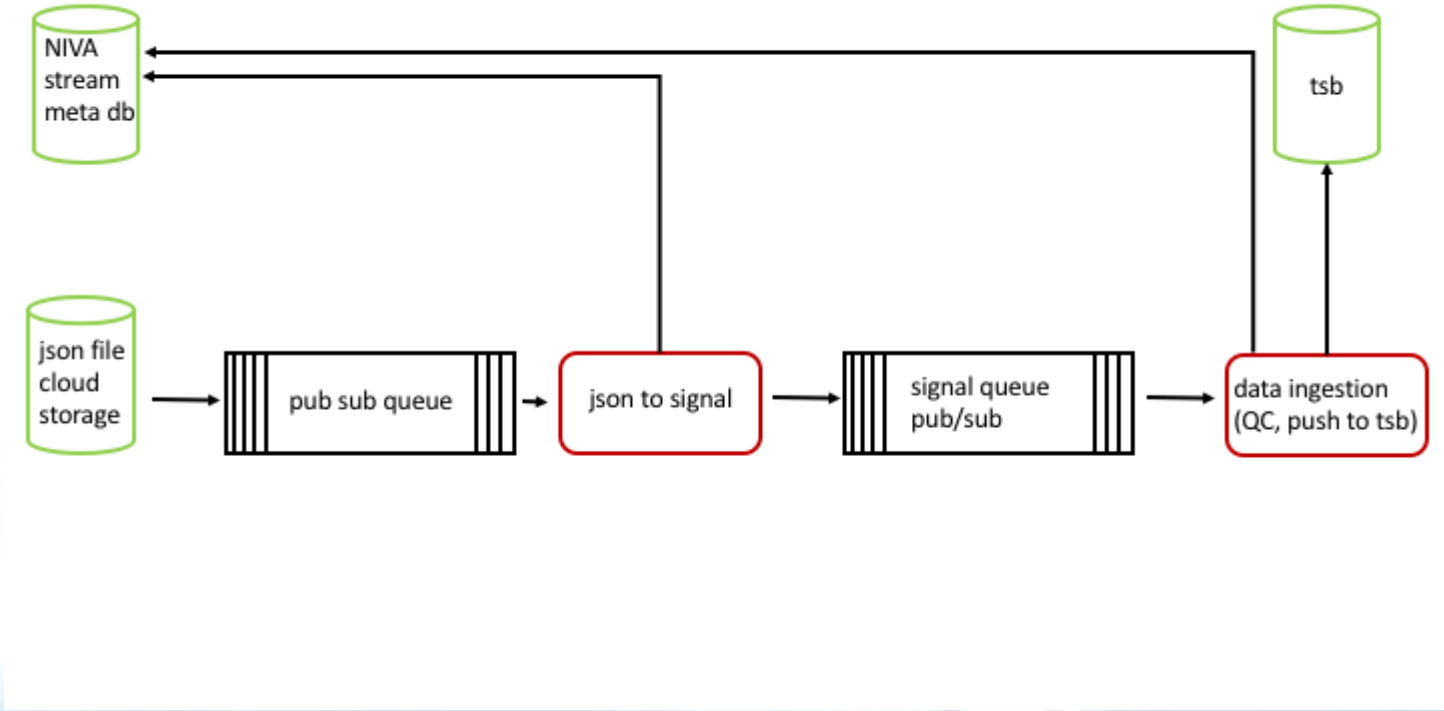
- NIVA built the QC on documented approaches suggested by the worlds largest observational program: **Copernicus** (www.copernicus.eu). This the recommended international standard.
- QC system running on Google Cloud ensuring high-availability

Example data

code	date	time	lat.	long.	temp.
FA	07.12.2018	13:25:39	59.8571	10.6093	9.050
FA	07.12.2018	13:26:39	59.8535	10.6023	8.864
FA	07.12.2018	13:27:39	59.8499	10.5952	8.865
FA	07.12.2018	13:28:40	59.8461	10.5886	8.479
FA	07.12.2018	13:29:40	59.8419	10.5830	8.302
FA	07.12.2018	13:30:40	59.8372	10.5785	8.132
FA	07.12.2018	13:31:40	59.8321	10.5753	7.986
.....					

columns: measurements
rows: signals

Core data streaming part



QC flags

- Each measurement undergoes QC tests specified in metadata
- Each QC test results in a flag (-1/0/1)
- Data source may also provide quality flags
- The overall flag (-1/0/1) is derived based on individual flags

QC library

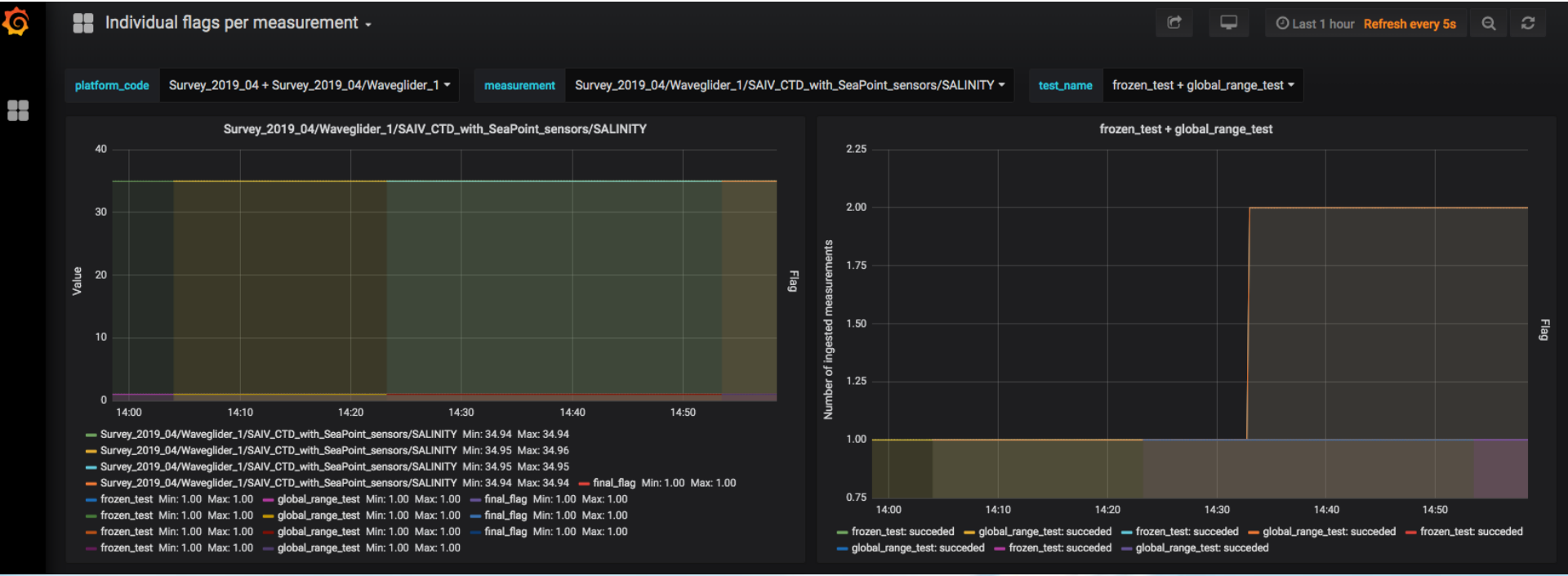
- Separate python package: https://github.com/NIVANorge/qclib
- Contains implementation of documented tests (Copernicus)
- Source control identifes which tests and what version of code were applied to any given published dataset (reproducable from raw data)

Test	requirements
Missing value	1 sample
Frozen test	4+ samples
Local/Global range	1 sample
Spike	3+ samples, symmetric in time

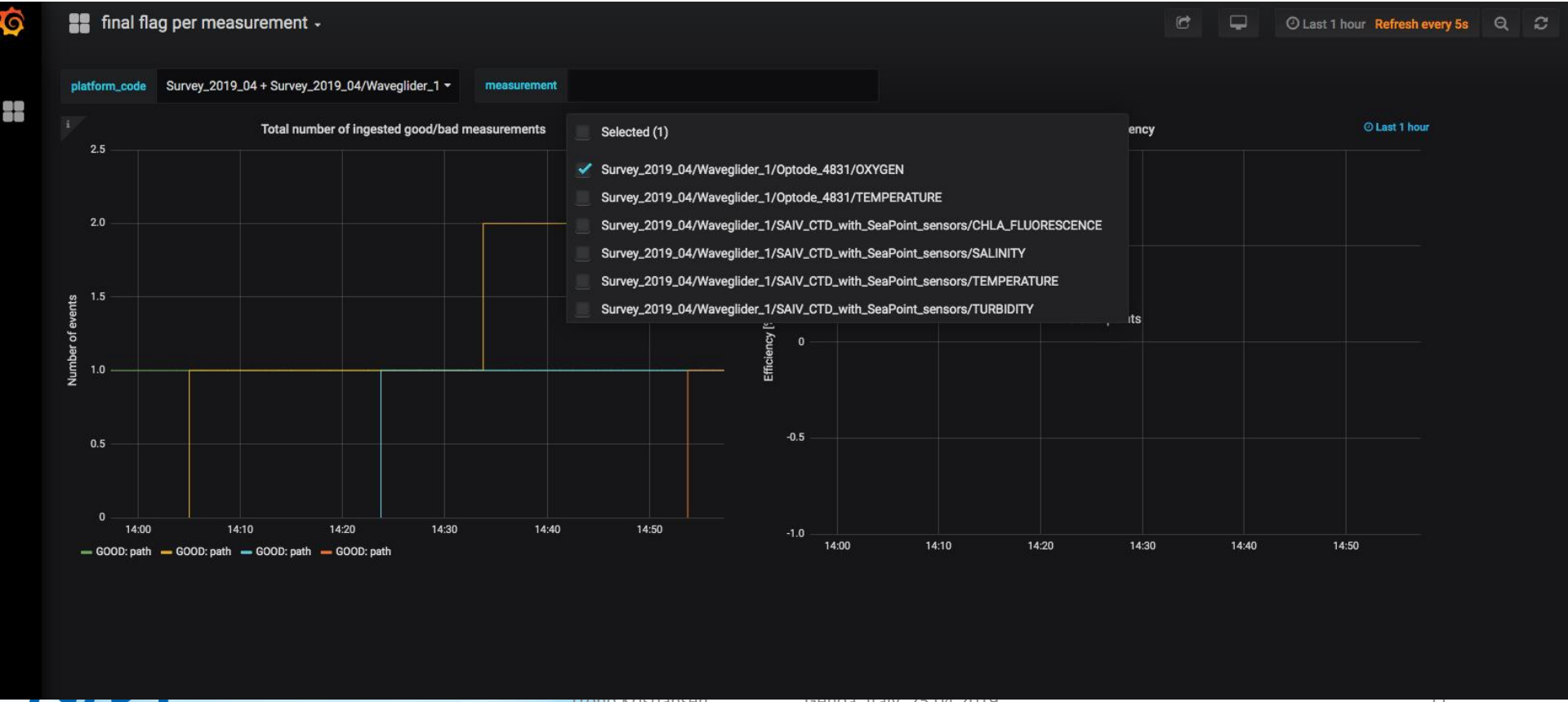
Push QC flag according to standards

	Meaning	Comment
0	No QC was performed	-
1	Good data	All real-time QC tests passed.
2	Probably good data	-
3	Bad data that are potentially correctable	These data are not to be used without scientific correction.
4	Bad data	Data have failed one or more of the tests.
5	Value changed	Data may be recovered after transmission error.
6	Not used	-
7	Nominal value	Data were not observed but reported (e.g.an instrument target depth)
8	Interpolated value	Missing data may be interpolated from neighbouring data in space or time.
9	Missing value	The value is missing

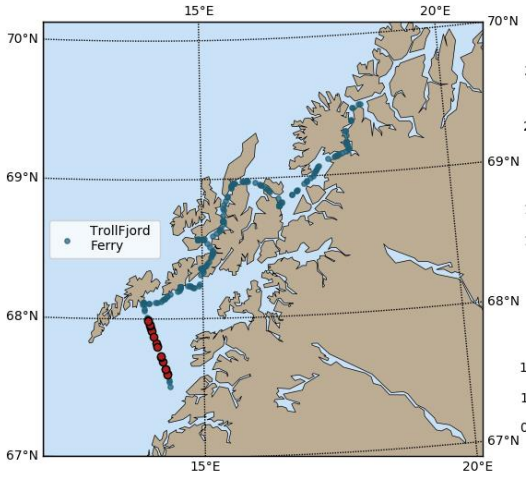
QC dashboard



QC dashboard

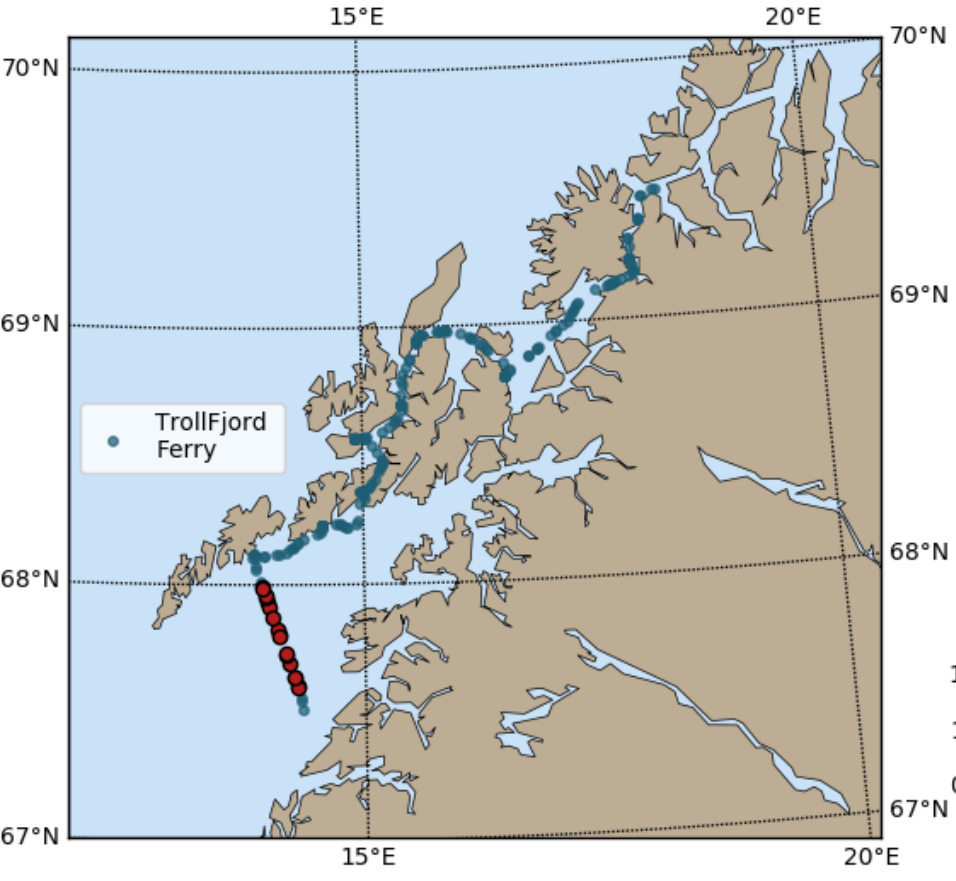


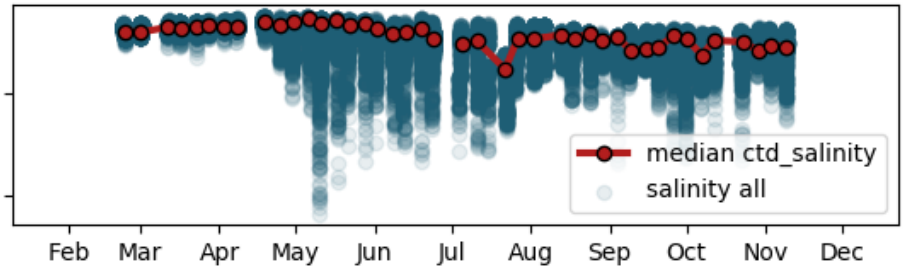
Trollfjord Ferrybox

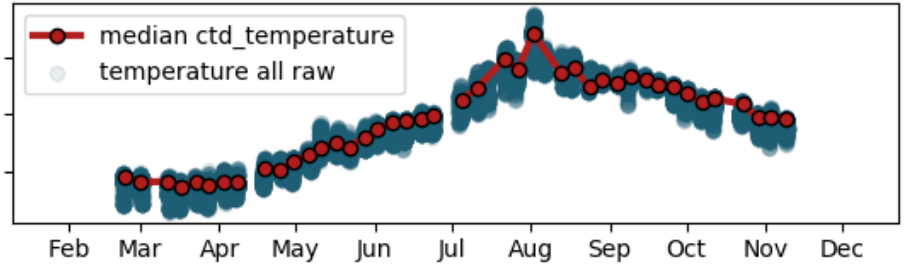


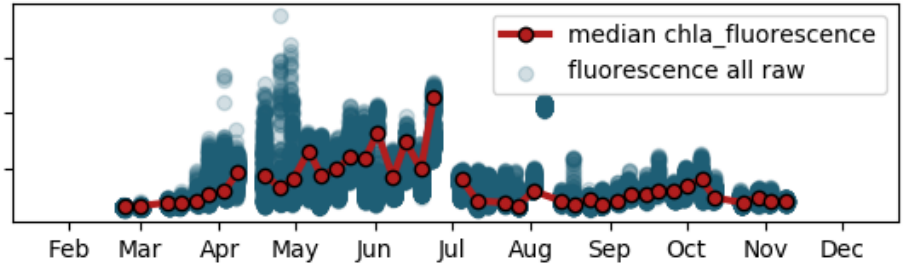
- **Hurtigruten Trollfjord :** ship data are streamed operationally to NIVA server, quality controlled, and new data made available to users every **5 minutes**
- Operates in coastal Lofoten-Vesterålen and sails through the area 2 times per week.
- This area is extremely important as a fisheries habitat, spawning ground and nursery
- Combined observations and models provide background for management decisions on ecosystem status

Trollfjord Ferrybox

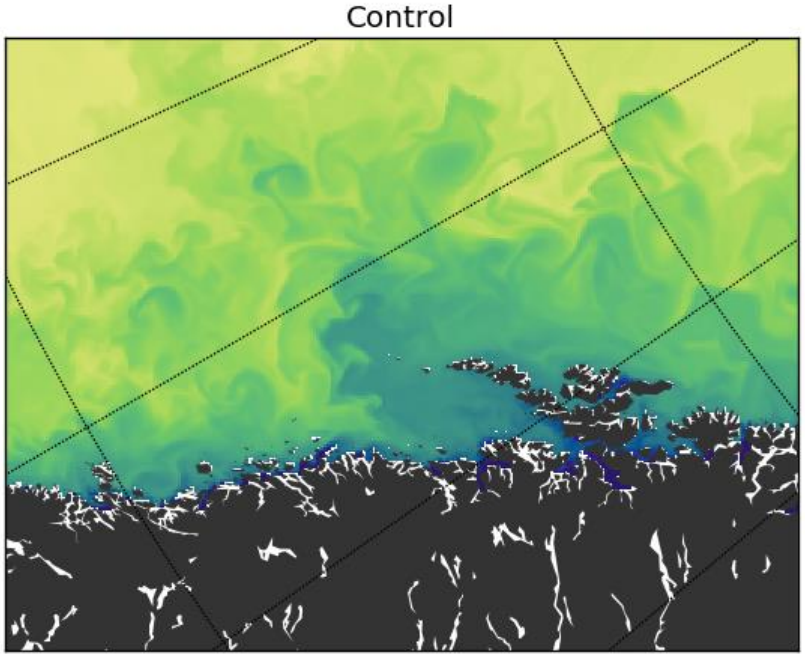


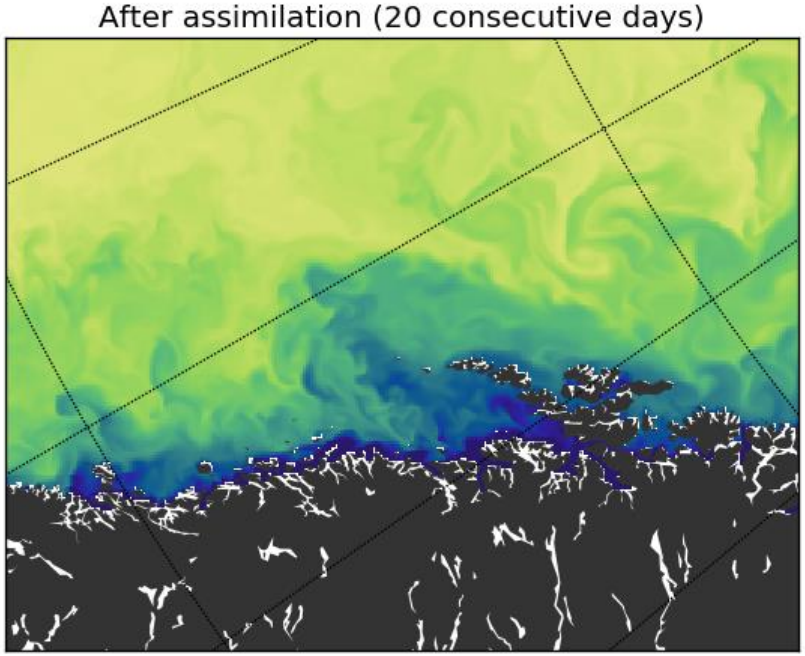






Assimilation into ocean circulation model





Summary

- Real-time streaming, QC, and availability of data from Ferrybox through API
- Easy to add new FB routes to system
- Assimilating FB data into ocean circulation models improves predictions of ocean temperature, currents, and salinity particularly along coastlines
- **Ferrybox data can improve weather predictions!**