

STRATUS 20 Data Report

Yu Gao

June 12, 2025

Contents

1	Introduction	2
1.1	Deployment Summary	2
2	Pre-Processing	2
2.1	Data Loading and Initial Processing	2
2.2	Time Correction	2
3	Truncation and Visualization	6
3.1	Data Truncation	6
3.2	Re-calculate Salinity	7
3.3	Data Export	8
4	Quality Control and Deployment Catalog	8
4.1	Spike Removal	9
4.2	Difference Statistics	9
4.3	Human in the Loop (HITL) Quality Control	10
4.4	Data Export	12
5	Merged Data Set	13
5.1	Metadata and Selection Strategy	13
5.2	Overlap Analysis	14
6	Documentation and Archiving	17
7	Conclusion	17

1 Introduction

This document details the preprocessing, quality control, and merging steps for the deep temperature and salinity (T/S) data from stratus 20. This report is supplemented by the Python code used in the analysis.

1.1 Deployment Summary

- **Deployment Period:** From 01/25/2022 12:00 to 2023/03/02 16:06
- **Instruments Used:** SBE37 10600 and SBE37 10601
- **Merged Data:** SBE37 10600 was selected for the merged dataset
- **Overlap with Previous Deployment (stratus 19):** yes

2 Pre-Processing

Detailed steps described during the pre-processing of stratus 20 data include:

2.1 Data Loading and Initial Processing

- Read .mat files containing SBE37 data using custom functions.
- Convert MATLAB data to xarray datasets for use in Python.
- Process and update metadata.

2.2 Time Correction

- Adjust time data to account for MATLAB's year zero issue.
- Convert time units to 'days since 0001-01-01'.
- Subtract 365 days from converted times to correct for MATLAB's leap year assumption.
- Time base Interpolation (if necessary): Data are interpolated from the initial timing mark to the final timing mark, so this needs to be done before truncation.

Potential timing errors are addressed by referring to the "recover.xls" spreadsheet on buoy5, under the "subsurface" tab, to compare spike start and end times with the apparent time in the instrument data. If the offset is greater than one sample interval (+/- 5 min for Stratus deployments), it indicates misalignment in time, and the data is linearly interpolated to a new time base to synchronize instruments.

In reviewing the temperature data against recorded spike times for stratus 20, it is observed that the differences are consistently within 5 minutes. Specifically, for the deployment phase, the significant temperature drop commences at 01/25/2022 12:00, closely aligning with the recorded spike time of 01/25/2022 12:00. Similarly, during the recovery phase, a notable temperature drop starts at 2023/03/02 16:06, which slightly follows the recorded spike time of 2023/03/02 16:06.

Time Correction Decision: Based on these findings, interpolation was not performed for stratus 20 as the temperature data aligns well with the recorded times, suggesting minimal time offset. See Figs.1 and 2 for details.

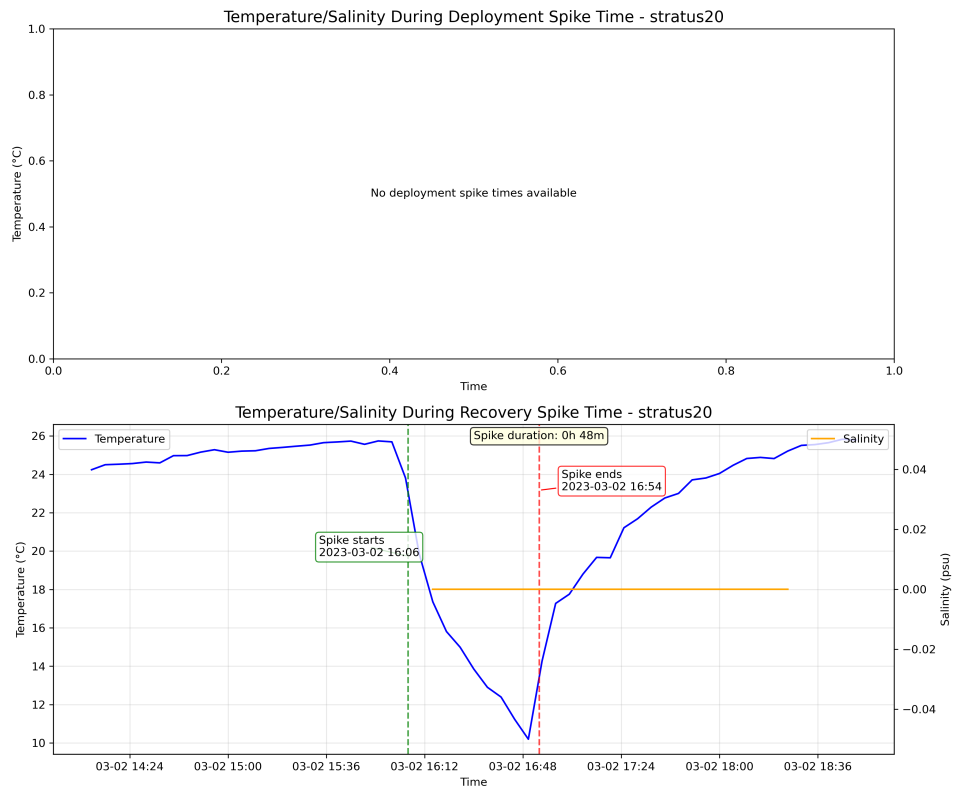


Figure 1: Deployment and recovery spike times for instrument SBE37 10600 of stratus 20 showing temperature changes during deployment and recovery operations. The vertical dashed lines indicate the recorded spike times (01/25/2022 12:00 for deployment and 2023/03/02 16:06 for recovery).

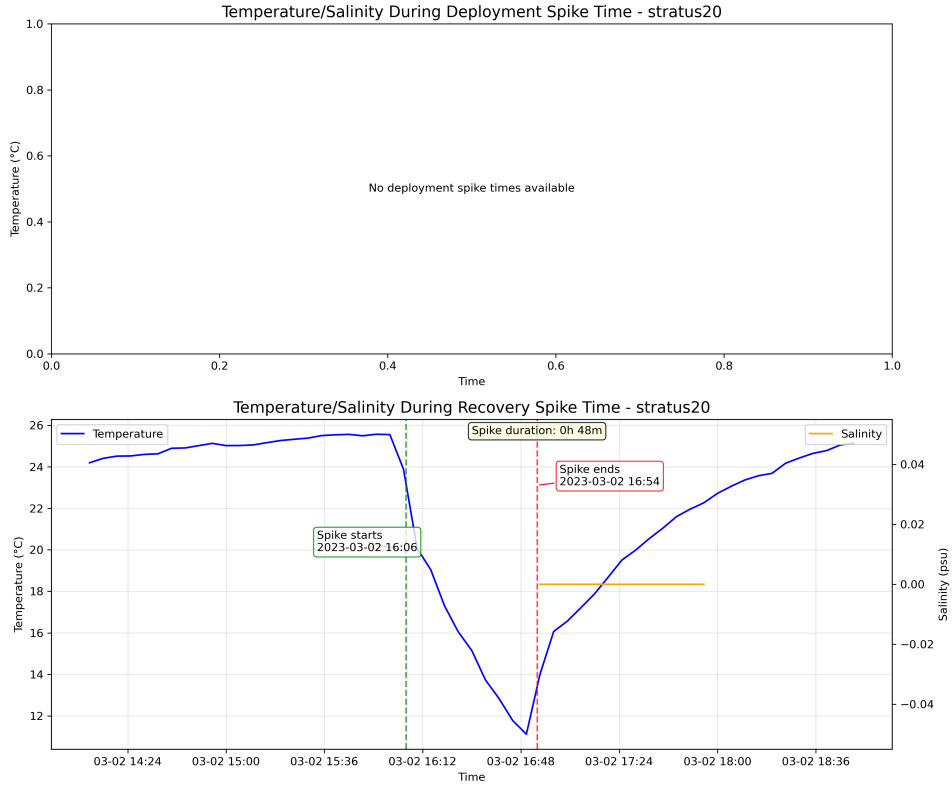


Figure 2: Deployment and recovery spike times for instrument SBE37 10601 of stratus 20 showing temperature changes during deployment and recovery operations. The vertical dashed lines indicate the recorded spike times (01/25/2022 12:00 for deployment and 2023/03/02 16:06 for recovery).

The deployment and recovery phases are visualized in Figs.3 and 4. The figures show the temperature changes during deployment and recovery operations for both instruments. The vertical dashed lines indicate the anchor over time (anchor dropped) and release fired times (mooring released), which are used to truncate the data.

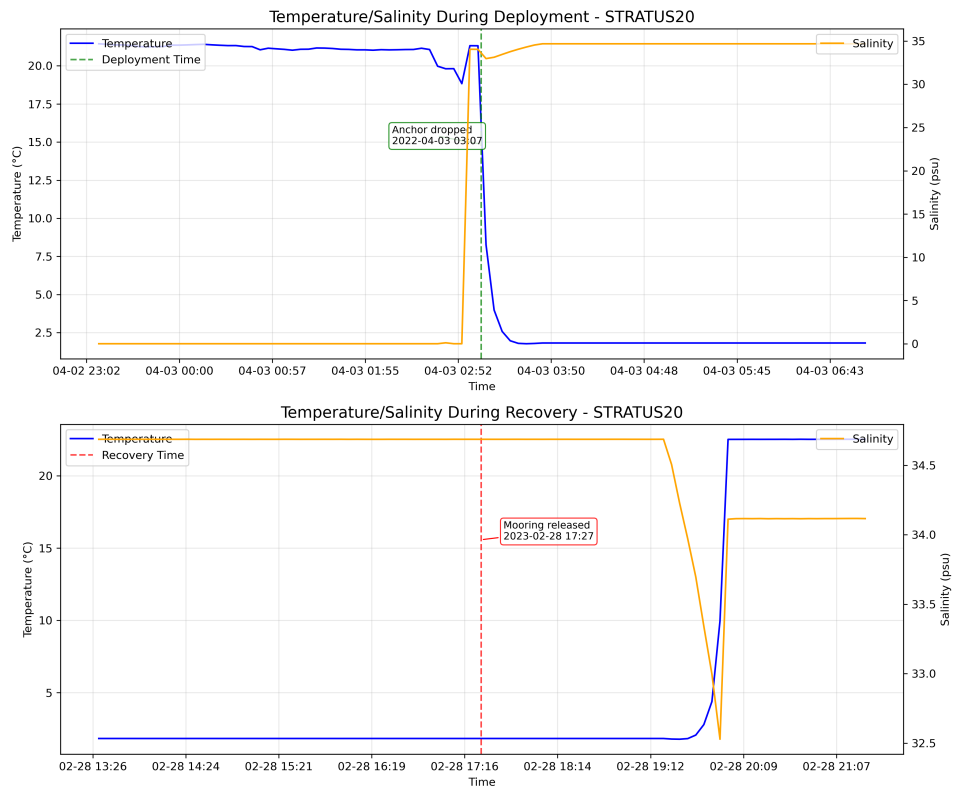


Figure 3: Detailed view of temperature changes during deployment and recovery phases for stratus 20, showing instrument SBE37 10600.

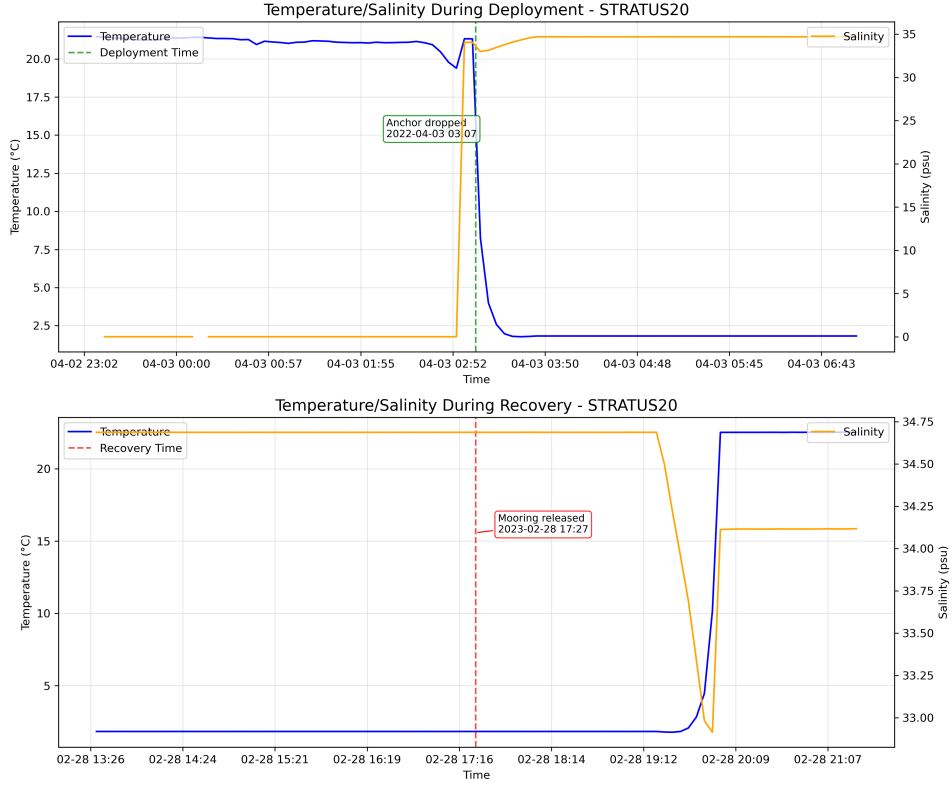


Figure 4: Detailed view of temperature changes during deployment and recovery phases for instrument SBE37 10601 of stratus 20.

3 Truncation and Visualization

3.1 Data Truncation

- Read anchor over time and release fired time from metadata.
- Truncated dataset starts 4 hours after anchor over time and ends at release fired time.
- Plot and inspect temperature time series for both SBE37 instruments.
- Ensure that truncation has successfully excluded data from before deployment and after recovery.
- The time-corrected (if needed) and truncated data files are saved after this stage of processing. These define the "baseline" files for each deployment, before quality control but after time correction and truncation.

The truncated data (Fig.5) shows the expected time window of good data after truncation, starting 4 hours after anchor deployment time and ending at the release fired time.

Comparison within STRATUS20 Data
Location: (22.4536°S, 85.6657°W) | Depth: 4222m

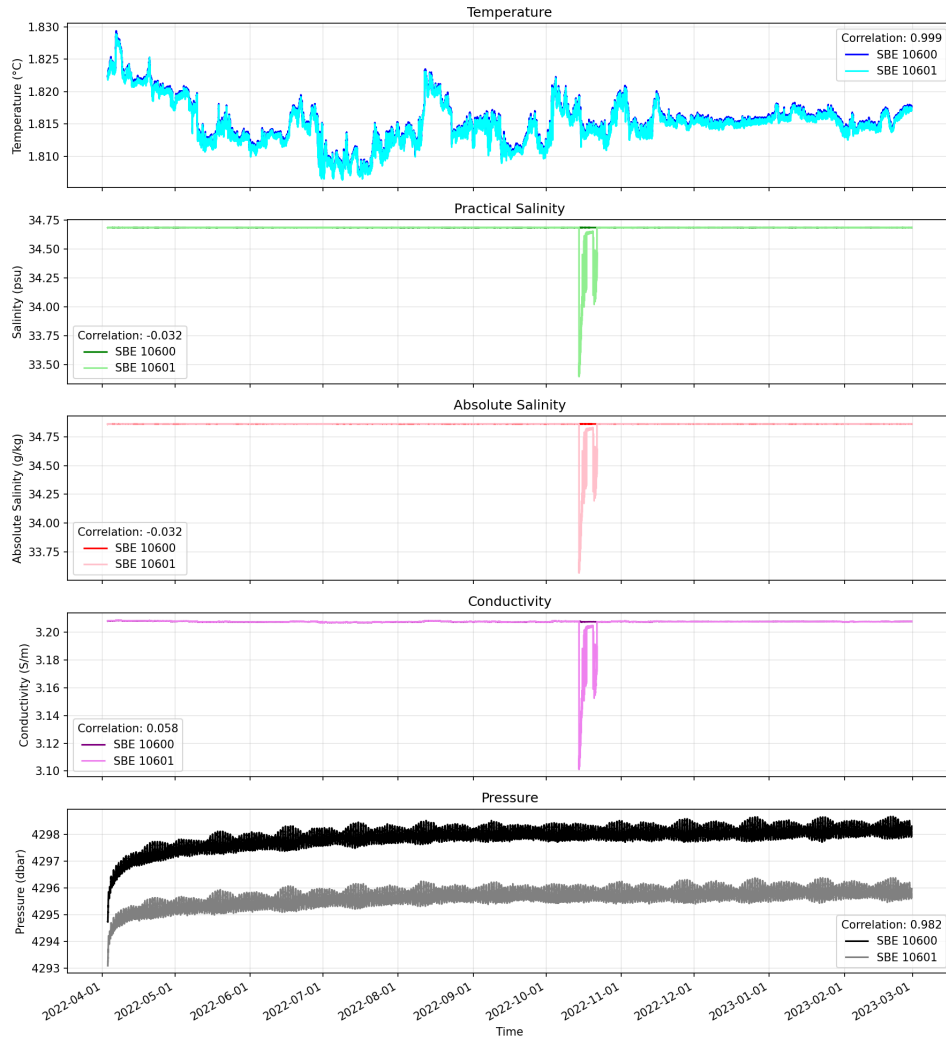


Figure 5: Deployment and recovery timing analysis for stratus 20, comparing instruments SBE37 10600 and SBE37 10601. This figure shows the expected time window of good data after truncation. Truncated dataset starts 4 hours after anchor over time and ends at release fired time.

3.2 Re-calculate Salinity

Pressure data is available for this deployment but may have drifted significantly in the SBE37 instruments, which can lead to inaccurate salinity calculations. Therefore, we recalculate salinity using a fixed pressure value based on the instrument depth. For stratus 20, salinity was recalculated using the following procedure:

1. **Pressure Calculation:** Instead of using the potentially drifting pressure sensor data, we calculate a fixed pressure value based on instrument depth:
 - Calculate pressure using the GSW function `p_from_z`, which converts geometric height to pressure

- Use the formula $z = -1 \times \text{instrument_depth}$ (negative because depth is positive downward)
- Use latitude from the anchor survey for the calculation

2. Practical Salinity Calculation:

- Convert conductivity from S/m to mS/cm by multiplying with a factor of 10
- Apply the TEOS-10 (Thermodynamic Equation Of Seawater - 2010) algorithm via `gsw.SP_from_C`
- Use measured temperature and the calculated fixed pressure (rather than varying measured pressure)
- Store the computed values as `sea_water_practical_salinity` with units 'psu'

3. Absolute Salinity Calculation:

- Use the TEOS-10 algorithm `gsw.SA_from_SP` to convert practical salinity to absolute salinity
- Incorporate geographical dependencies using latitude and longitude from the anchor survey
- Store the computed values as `sea_water_absolute_salinity` with units 'g/kg'

4. Post-Processing:

- The original salinity variables (`sal` and `abssal`) are removed from the output datasets

This method ensures consistent salinity calculations and eliminates potential errors from pressure sensor drift, which can be substantial at deep ocean deployments. Using a fixed pressure calculated from the known instrument depth produces more reliable salinity values.

3.3 Data Export

- Ensure all required variables (`sea_water_temperature`, `sea_water_electrical_conductivity`, `sea_water_practical_salinity`, `sea_water_absolute_salinity`, `sea_water_pressure`) exist in the dataset.
- Fill missing data or create non-existent variables with -99999 values and encode them as missing values.
- Carry forward time series for all five variables even if not present, by setting missing values to -99999.
- Save the time-corrected, truncated datasets as NetCDF files for further processing.
- Files are saved in the directory `/data/truncated/stratus20` as `stratus20_10600_truncated.nc` and `stratus20_10601_truncated.nc`.

4 Quality Control and Deployment Catalog

The following quality control measures are implemented:

4.1 Spike Removal

Spike detection and removal are set up in function `src/qc_function.py`. The spike removal function identifies and removes spikes from time series data using a rolling mean and standard deviation. It calculates these statistics within a specified window size around each data point. Points that deviate from the rolling mean by more than 3 times the standard deviation are considered spikes and replaced with missing values (-99999). For stratus 20 processing, the window size was 12 points and the threshold was 3 times the standard deviation.

See Figs.6 and 7 for examples of the original and cleaned plots for temperature, salinity, absolute salinity, conductivity, and pressure for both instruments. These serve as a basis for collaborative review and further data correction.

The spikes removed from the datasets are shown in the table below.

Table 1: Spike removal statistics for stratus 20

Variable	SN 10601 (%)	SN 10600 (%)
sea_water_absolute_salinity	0.6066	0.4274
sea_water_electrical_conductivity	0.2315	0.2106
sea_water_practical_salinity	0.4222	0.4274
sea_water_pressure	0.0031	0.0031
sea_water_temperature	0.0042	0.0073

4.2 Difference Statistics

When both instruments are available and functional, the mean and standard deviation of the difference between sensors and of individual sensors are computed. These statistics provide benchmarks for accuracy and precision, and are critical for quality control decisions. The difference statistics are computed after spike removal to avoid contamination by known bad points.

The difference statistics are shown in tables below.

Variable	SN 10601		SN 10600	
	Mean	Std Dev	Mean	Std Dev
sea_water_temperature	1.81504	0.00337	1.81531	0.00338
sea_water_practical_salinity	34.67974	0.08648	34.68738	0.02414
sea_water_absolute_salinity	34.85415	0.08242	34.86167	0.03381
sea_water_electrical_conductivity	3.20708	0.00634	3.20774	0.00027
sea_water_pressure	4295.62222	0.31088	4297.87009	0.38278

Table 2: Statistics for individual sensors on stratus 20

Variable	Mean Diff	Std Diff	QC Threshold
sea_water_temperature	-0.00027	0.00477	0.01432
sea_water_practical_salinity	-0.00764	0.08979	0.26936
sea_water_absolute_salinity	-0.00751	0.08909	0.26727
sea_water_electrical_conductivity	-0.00066	0.00634	0.01903
sea_water_pressure	-2.24788	0.49312	1.47936

Table 3: Statistics for difference between sensors on stratus 20

Note: These statistics are also stored as metadata in the NetCDF files with the following attributes:

- `std_single_sensor`: Standard deviation of the time series from a single sensor
- `std_sensor_diff`: Standard deviation of the difference between two co-located sensors
- `mean_sensor_diff`: Mean of the difference between two co-located sensors

Note: NaN values that exist in the difference statistics are because the data were not available for that time period.

4.3 Human in the Loop (HITL) Quality Control

A catalog of results for each deployment was created, featuring original and QC plots for all measured variables.

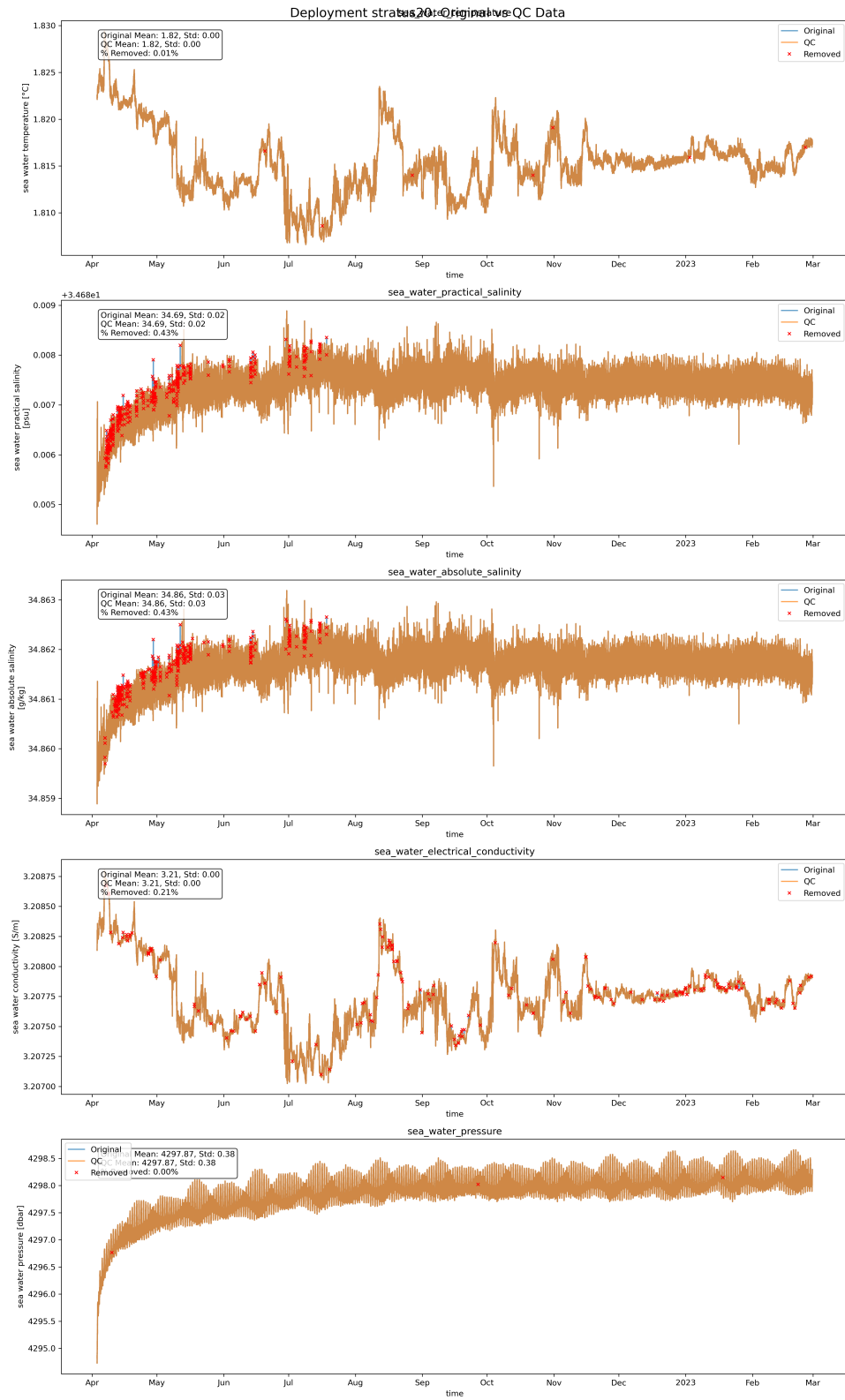


Figure 6: Original and clean plots of temperature, salinity, absolute salinity, conductivity, and pressure for instrument SBE37 10600 of stratus 20.

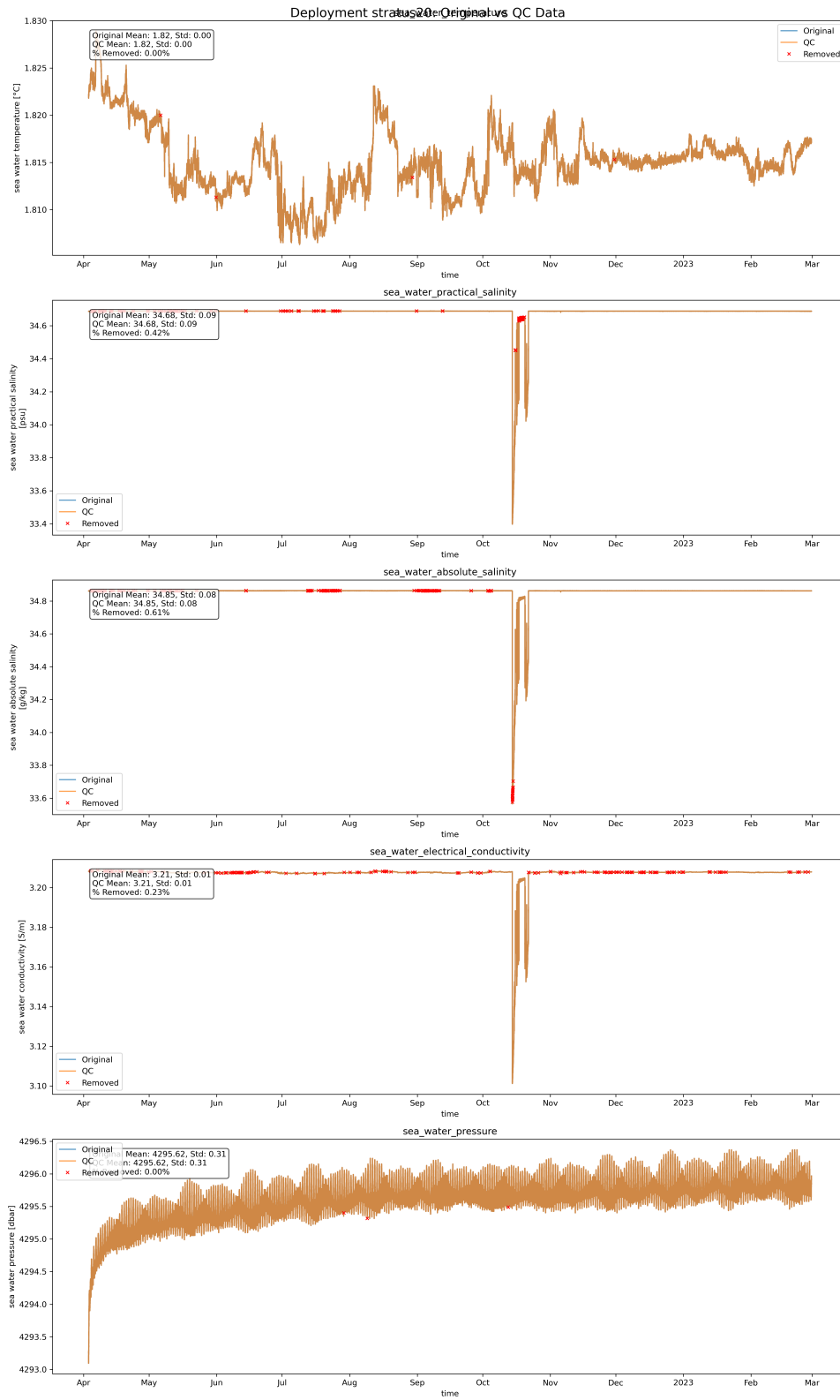


Figure 7: Original and clean plots of temperature, salinity, absolute salinity, conductivity, and pressure for instrument SBE37 10601 of stratus 20.

4.4 Data Export

- Export the cleaned datasets as NetCDF files for further processing.

- Files are saved in the directory `/data/cleaned/stratus20` as `stratus20_10600_cleaned.nc` and `stratus20_10601_cleaned.nc`
- Each file contains the following variables:
 - `sea_water_temperature` (°C)
 - `sea_water_electrical_conductivity` (S/m)
 - `sea_water_practical_salinity` (PSU)
 - `sea_water_absolute_salinity` (g/kg)
 - `sea_water_pressure` (dbar)

5 Merged Data Set

5.1 Metadata and Selection Strategy

- All necessary metadata are incorporated from individual deployments and retained in the merged file.
- Selection strategy for merging:
 - If both sensors for a deployment have full records and are of good quality, a Human-In-The-Loop (HITL) evaluation is done for the transitions at the start and end of the deployment, and data records with the least discrepancy between deployments are chosen to be merged.
 - If one sensor is obviously bad or does not have a complete record, the other sensor is used.
 - For stratus 20, SBE37 10600 was selected for the merged dataset.
 - If the sampling rate of the two time series is different, the time series with the higher sampling rate is resampled to match the lower sampling rate.

5.2 Overlap Analysis

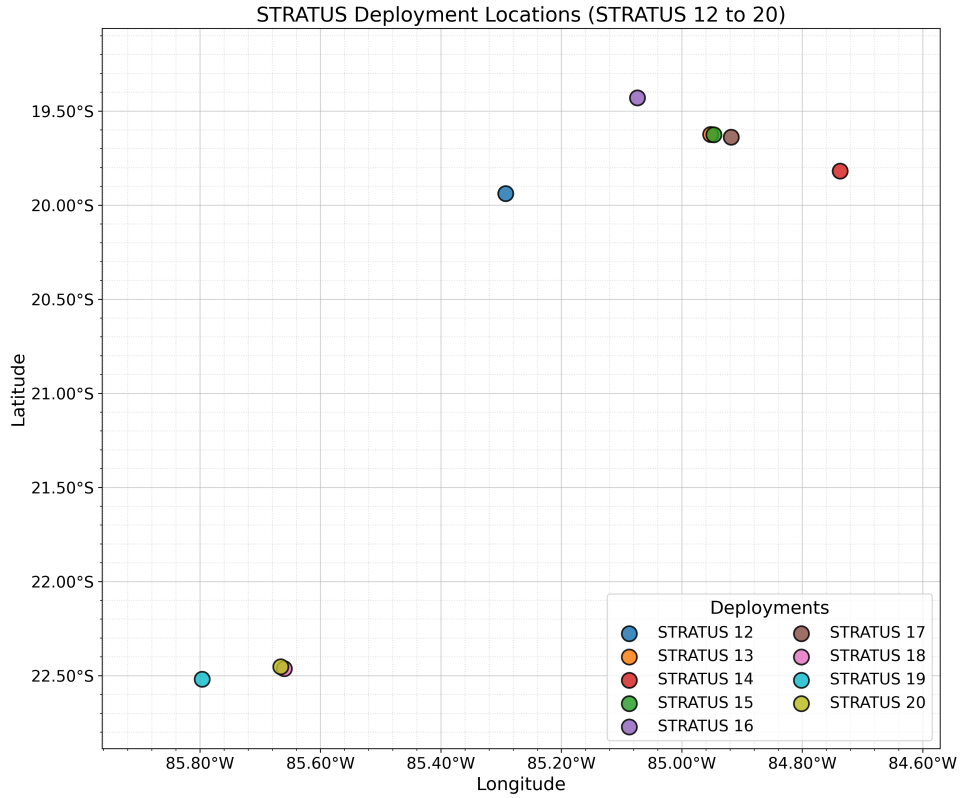


Figure 8: Deployment locations for Stratus moorings from Stratus 19 to Stratus 20. The map shows the anchor positions determined during deployment surveys.

The distance between STRATUS 19 and STRATUS 20 deployments is 15.30 kilometers (8.26 nautical miles).

Table 4: Deployment coordinates

Deployment	Latitude	Longitude
STRATUS 19	22.5193°S	85.7965°W
STRATUS 20	22.4536°S	85.6657°W

Figure 8 shows the positions of the Stratus 19 and 20 moorings, separated by approximately the distance mentioned above.

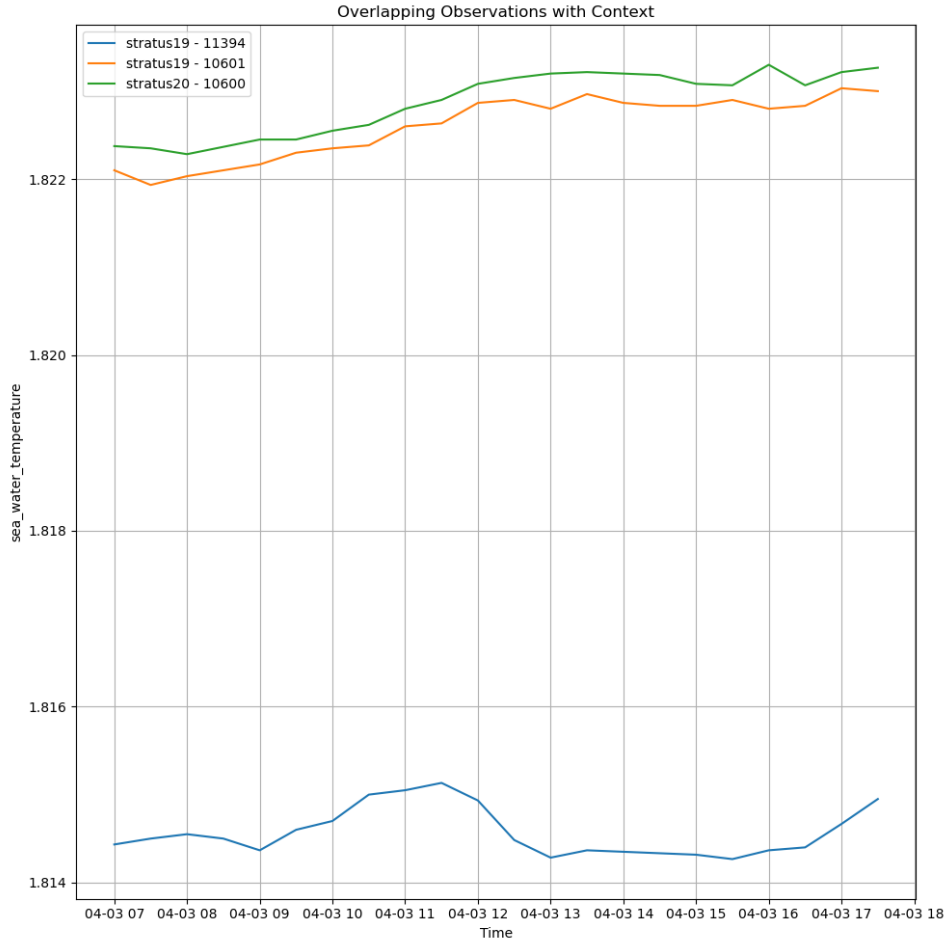


Figure 9: Visualization of overlapping observations between stratus 19 and stratus 20. The vertical dashed lines indicate: start of new deployment, optimal merge point, and end of previous deployment.

There is an overlap between stratus 19 and stratus 20 deployments. We chose SBE37 10600 for merging because it has better quality data.

The merged dataset is visualized in Fig.10.

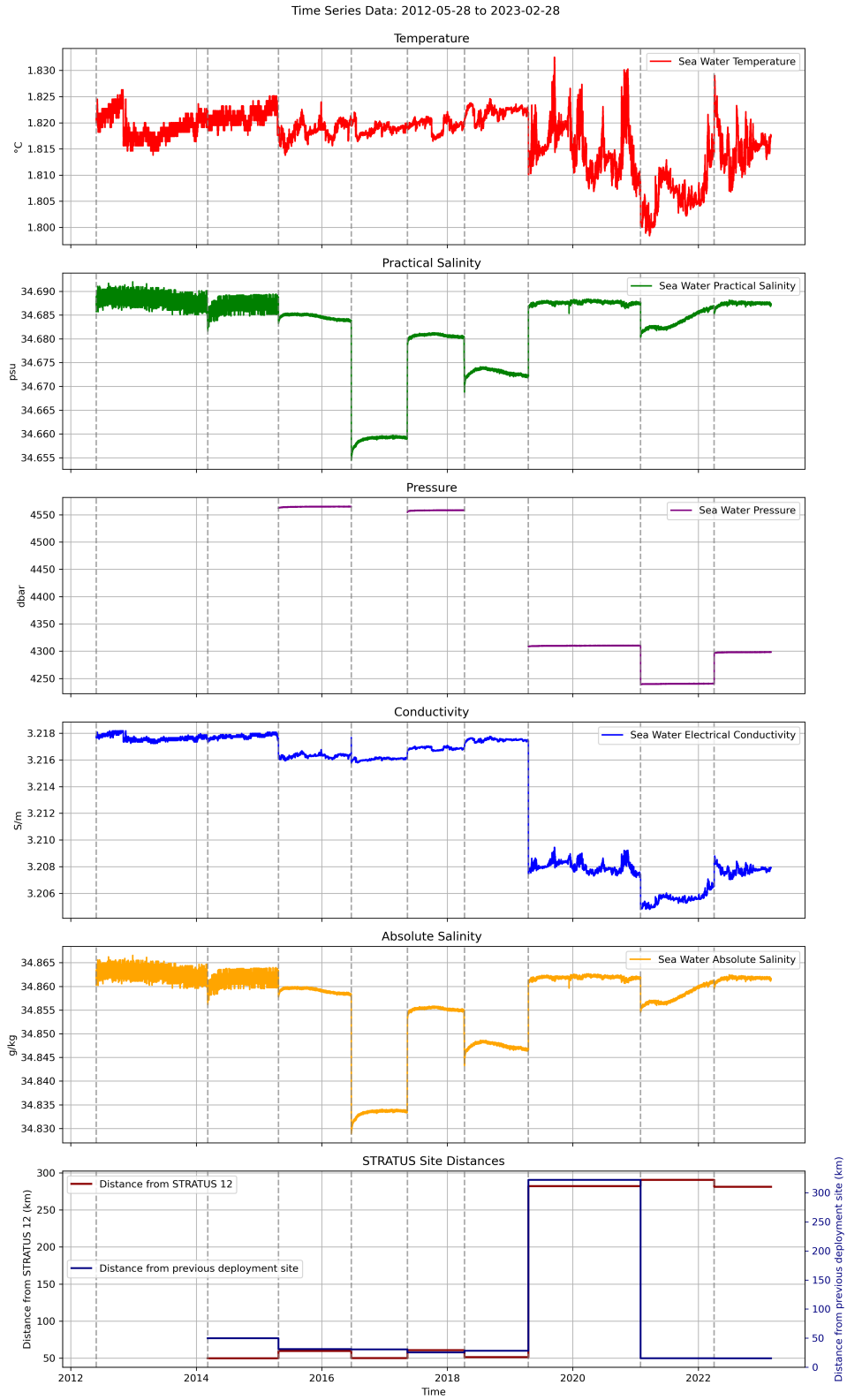


Figure 10: Merged dataset of stratus 12 to stratus 20 showing the transition at the merge point. This dataset combines data from the selected instrument (SBE37 11394) from stratus 19 and SBE37 10600 from stratus 20. This dataset incorporates all previous merges in the time series. The gray dashed line indicates the merge point between the two datasets.

6 Documentation and Archiving

Finalized data sets are documented extensively and prepared for inclusion on the project website. The metadata is updated to be compliant with the NetCDF Climate and Forecast (CF) conventions, and OceanSITES data manual.

7 Conclusion

This report documents the preprocessing, quality control, and analysis steps for the deep temperature and salinity data from stratus 20. The data has been processed through several stages:

- Time correction and data truncation to isolate the deployment period
- Quality control including spike removal and difference statistics
- Recalculation of salinity using the TEOS-10 algorithm at the instrument depth with observed temperature and conductivity
- Merging with adjacent deployment data

The resulting cleaned and merged dataset provides a valuable record of deep ocean temperature and salinity at the stratus site, contributing to long-term climate monitoring efforts.