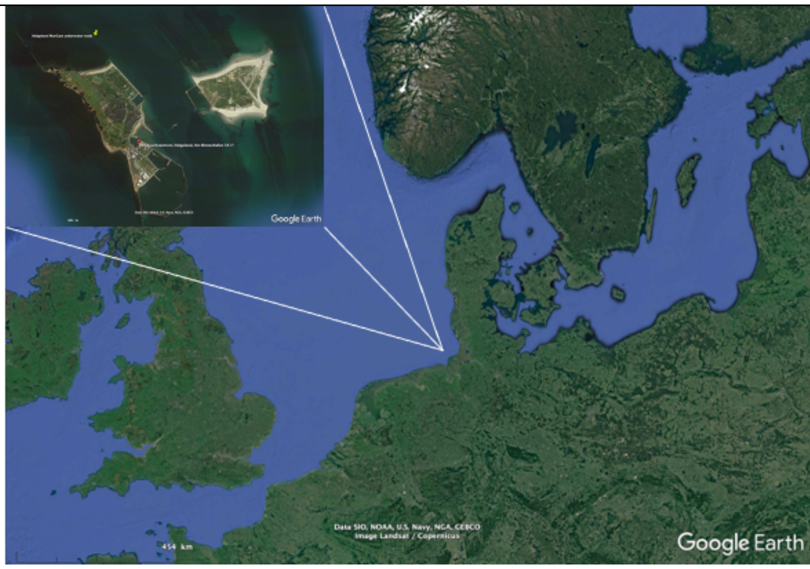
Metadata for file:	Helgoland_MarGate_Underwaterobservatory_2022_Hydrography.csv
Abstract:	Oceanographic data (temperature, salinity, oxygen saturation, chlorophyll A and turbidity) were sampled in 2022 year-round in the underwater experimental area MarGate off Helgoland at N 54° 11.594 / E 07° 52.762 (WGS84) in a water depth of 10 m +/- tide. All sensors were attached to the COSYNA underwater observatory and remotely controlled and supervised. Sampling frequency for all parameters was 1Hz. Data were transferred in real-time to the AWI data science centre where the raw data are stored in the O_2A database (https://spaces.awi.de/display/DM). For each parameter, multiple different sensors were available and were operated either parallel or sequentially over year (see section "**Sensors"** below**)**. Shown data are data products calculated from raw data from the available sensors using a four-step plausibility and quality control procedure, explained in detail in section "**Data handling and processing steps" below.** The data are collected in the framework of "COSYNA", "ACROSS" and "MOSES" aiming at providing basic hydrographical data of the target area for the use in physical, biochemical and biological projects with respect to environmental changes based on climate change and anthropogenic disturbances as well as providing high-resolution hydrographic data for a functional understanding of the effects of short-term and long-term changes in the local hydrography on biological processes in shallow coastal marine systems. Additionally, the dataset is used as permanent oceanographic reference station for the MOSES project "Hydrological extremes" as well as for a joint project of AWI, MARUM, HZG, BSH and others on data quality and stability of marine sensors in shallow coastal waters.
General description of the sensor setup:	The dataset contains temperature, salinity, oxygen saturation, chlorophyll A and turbidity data from the Helgoland MarGate-underwater observatory from the year 2022 in a temporal resolution of 1 hour. The cabled observatory is located in 10m water depth and comprises single or multiple sensors for a specific parameter (see https://www.awi.de/en/science/biosciences/shelf-sea-system-ecology/main-research-focus/cosyna/underwater-node-helgoland.html)
Location & Map:	The Helgoland MarGate-underwater observatory is located at N 54° 11.594 / E 07° 52.762 (WGS84).

	
Platform description:	Helgoland MarGate-underwater observatory is an experimental field station (id: heluwobs) operated jointly by the AWI and the Helmholtz Centre Geesthacht (Germany). The station hosts multiple sensors and sensor units in the southern North Sea ecosystem close to Helgoland Island. The station comprises an underwater node system about 400 m North of the island. The system is fully remote controlled and accessible from all over the world by VPN-connection. The attached lander systems are equipped with (redundant) multiple sensors and are designed to also host sensors from external cooperation partners. The underwater observatory provides power and Gbit internet connection at 10 independent wet-matable underwater plugs so that sensors and sensor-units can be exchanged by divers e.g. for maintenance. The underwater observatory additionally hosts a fully remote-controlled vertical profiling system for a variety of different sensors which perform daily or hourly vertical profiles from 10m water depth to the surface. All data from the observatory are archived at the AWI data management system O2A (Observation to Archive) and are available as quality-controlled data with 1-hour temporal resolution at PANGAEA (www.pangaea.de) or with higher temporal resolution upon request to philipp.fischer@awi.de
Sensors:	**Temperature: 3 sensors have been used to calculate temperature.** 1. Manufacturer Sea-Bird Electronics; Model SBE 38 Digital Oceanographic Thermometer 2. Manufacturer Teledyne RD Instruments; Model Workhorse Monitor 3. Manufacturer Sea and Sun Technology; Model CTD 90/ Merz Pt 100/1509 **Salinity: 2 sensors have been used to calculate salinity.**

	1. Manufacturer Sea and Sun Technology; Model CTD 90 / ADM 7 pole electrode cell. 2. Manufacturer Sea and Sun Technology; Model CTD 90 / ADM 7 pole electrode cell. **Oxygen saturation: 2 sensors have been used to calculate oxygen saturation.** 1. Manufacturer Sea and Sun Technology; Model CTD 90 / Aanderaa 4175C. 2. Manufacturer Sea and Sun Technology; Model CTD 90 / Aanderaa 4175C **Chlorophyll a: 2 sensors have been used to calculate chlorophyll A concentration.** 1. Manufacturer Sea and Sun Technology; Model CTD 90 / Cyclops7 Fluorometer. 2. Manufacturer Sea and Sun Technology; Model CTD 90 / Cyclops7 Fluorometer. **Turbidity: 2 sensors have been used to calculate turbidity.** 1. Manufacturer Sea and Sun Technology; Model CTD 90 / Seapoint turbidity meter. 2. Manufacturer Sea and Sun Technology; Model CTD 90 / Seapoint turbidity meter.
Data handling and processing steps:	Data are generally sampled with 1 Hz per single sensor and saved as raw data files every hour on the AWI-O2A file server. Data are then processed in a four-step routine: **Single sensor routines:** 1. Step: Data of each sensor are median averaged per minute to eliminate spikes and outliers within minute intervals. 2. Step: Plausibility checks are applied on all step 1 median values using the ARGO plausibility routines step 6, 7, 9, 11 and 13 (see http://www.argodatamgt.org/content/download/341/2650/file/argo-quality-control-manual-V2.8.pdf) 3. Step: The data are visually inspected on a monthly basis using an interactive R-Shiny application. Datapoint plausibility classification from step 2 (good, probably good, probably bad and bad) has to be

	confirmed based on expert knowledge. Further datapoints can be manually classified as probably good, probably bad or bad. 4. Step: According to the selected temporal resolution (1h), the arithmetic means of the step 3 plausibility checked data is done. -> For all subsequent calculation steps, only level 1 data (good data) are used. **Multiple sensor routines:** 5. Step: From all available sensors for the respective parameter, the sensor showing the least obvious outliers and the lowest temporal drift (see step 3) is used as lead sensor. 6. Step: Missing sensor data from the lead sensor are imputed (as far as possible) using the imputation routine "Amelia" from the R package (Honaker J, King G, Blackwell M (2011). "Amelia II: A Program for Missing Data." *Journal of Statistical Software*, **45**(7), 1–47. http://www.jstatsoft.org/v45/i07/) 7. Step: Using multiple linear regression, a model for predicting the lead sensor values is set up using the available auxiliar sensors. This model is used to calculate predicted lead sensor values using the available auxiliary sensor data. 8. Step: Lead sensor and predicted lead sensor data are analysed with respect to their goodness of fit by computer added analysis. Residuals of fitted to lead sensor values are calculated and visually inspected. Lead sensor values and associated predicted lead sensor values with a numerical difference of more than 3 x studentized standard deviation of the lead sensor are suggested to be classified as probably bad. 9. Step: Based on step 8 lead sensor values and the associated predicted lead sensor values kernel density estimates (for details see https://vita.had.co.nz/papers/density-estimation.pdf) are calculated. A kernel weight value of 1 is used for the lead sensor value. For the associated predicted lead sensor value, a kernel weight value of 0.7 is used. Additionally, 90% confidence limits (-> the real average value is located within the upper and lower confidence limit with a probability of 90%) are calculated for the mean value based on kernel density bandwidth calculation. In case that only a single sensor is available for a parameter, no further calculations are done in this step.
File structure and measuring units:	Each file has a single header line with the column labels. The first column is the datetime value (UTC). The second column is the kernel density mean value of the respective parameter (e.g. temperature). The third column is the 90% confidence of the mean value (conf_90).