Detection and quantification of CO₂ seepage in seawater using the stoichiometric \( C_{\text{seep}} \) method: Results from a recent subsea CO₂ release experiment in the North Sea

Abdirahman M. Omar\(^{a,\ast}\), Maribel I. García-Ibáñez\(^{a,1}\), Allison Schaap\(^{d}\), Anna Oleynik\(^{b}\), Mario Esposito\(^{c}\), Emil Jeansson\(^{a}\), Socratis Loucaides\(^{d}\), Helmut Thomas\(^{e}\), Guttorm Alendal\(^{b}\)

\(^{a}\) NORCE, Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway
\(^{b}\) Department of Mathematics, University of Bergen, Norway
\(^{c}\) GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
\(^{d}\) National Oceanography Centre, Southampton, UK
\(^{e}\) Institute of Carbon Cycles, Helmholtz Center Geesthacht, Geesthacht, Germany

A R T I C L E   I N F O

Keywords:
Subsea CO₂ release
Seafloor CO₂ seepage monitoring
CO₂ seepage detection
Baseline C establishment

A B S T R A C T

Carbon Capture and Storage (CCS) is a potential significant mitigation strategy to combat climate change and ocean acidification. The technology is well understood but its current implementation must be scaled up nearly by a hundredfold to become an effective tool that helps meet mitigation targets. Regulations require monitoring and verification at storage sites, and reliable monitoring strategies for detection and quantification of seepage of the stored carbon need to be developed. The \( C_{\text{seep}} \) method was developed for reliable determination of CO₂ seepage signal in seawater by estimating and filtering out natural variations in dissolved inorganic carbon (C). In this work, we analysed data from the first-ever subsea CO₂ release experiment performed in the north-western North Sea by the EU STEMM (DT) to reliably differentiate released \( \text{CO}_2 \) signal from natural variability and quantify released \( \text{CO}_2 \) dissolved in the sampled seawater. DT values were around 20 % of the natural C variations indicating high sensitivity of the method. Moreover, with the availability of DT value, the identification of released \( \text{CO}_2 \) required no pre-knowledge of seepage occurrence, but we used additional available information to assess the confidence of the results. Overall, the \( C_{\text{seep}} \) method features high sensitivity, automation suitability, and represents a powerful future monitoring tool both for large and confined marine areas.

1. Introduction

Carbon Capture and Storage (CCS) is a potential significant mitigation strategy to combat climate change and ocean acidification in climate change scenarios (e.g., van Vuuren et al., 2011; IPCC, 2013; Fuss et al., 2014). Three out of four 1.5 °C-consistent emission pathways published by the IPCC include CCS as well as bioenergy CCS (BECCS) as mitigation options (IPCC, 2018). However, the scale of current CCS implementation is insufficient to impact the global climate, and thousands of large-scale CCS facilities need to be deployed by 2040 to meet mitigation targets (Global CCS Institute, 2019). Moreover, the technology is in early stage of research, development, and demonstration (e.g., Hammond, 2018) and involves uncertainties and risks (e.g., Damen et al., 2006) that might raise ethical and governance issues, as is the case for other so-called climate geoengineering techniques (Lawrence et al., 2018).

Many potential CCS storage reservoirs are located below the seafloor (Blackford et al., 2017), and a number of subsea storage demonstration projects are in operation (IEAGHG, 2008; Jenkins, 2020). However, before subsea CO₂ storage can be carried out at industrial scales, potential ecological consequences and adverse environmental impacts of any unintended CO₂ seepages need to be identified. Therefore,
monitoring and verification is required at storage sites by national and international regulations (e.g., U.S. EPA, 2010; European Commission, 2011; Dixon and Romanak, 2015), and reliable monitoring strategies for detection and quantification of seepage need to be developed.

Primary monitoring of storage reservoirs is based on seismic techniques imaging CO$_2$ through the overburden (Jenkins, 2020). However, the detection threshold of such techniques may be of the order of 10$^3$ t CO$_2$ (Jenkins, 2020), and they are prone to interpretation uncertainty (Schaaf and Bond, 2019). Therefore, possible seepage at low levels may not be detected by seismic techniques, and monitoring for seepage at the seafloor provides an important secondary strategy. The modelling work of Blackford et al. (2020) reported that release events can be detected at the seafloor at thresholds well below levels that would compromise storage performance or significantly damage the environment.

When CO$_2$ seeping through the seafloor dissolves into seawater, it goes through a series of reactions that ultimately increase the concentration of dissolved inorganic carbon (DIC; hereafter denoted as C for simplicity). Thus, the excess C dissolved in the seawater ($C_{\text{seep}}$) should, in principle, be readily quantified from the resulting change in C ($\Delta C$). In practice, however, this is complicated by the fact that the dissolution of seepage CO$_2$ may be occurring on top of simultaneous C changes that result from natural processes, such as the formation/mineralization of organic matter and/or calcium carbonate (CaCO$_3$), mixing between water masses, and uptake of atmospheric CO$_2$. Thus, any natural parts of $\Delta C$ must be accounted for before the CO$_2$ seepage signal can be identified and $C_{\text{seep}}$ accurately determined.

The challenge of distinguishing subsea CO$_2$ seepage signal from natural variability can be overcome using different techniques. Uchimoto et al. (2017; 2018) proposed a seepage CO$_2$ detection method based on site-specific covariance between partial pressure of CO$_2$ in seawater (pCO$_2$) and dissolved oxygen (DO). Blackford et al. (2017) proposed site-specific anomaly detection criteria that recognises unnatural rates of change in seawater CO$_2$ concentrations. Gundersen et al. (2020) demonstrated the use of machine learning techniques for detecting signals from a seep using data from seep simulations. The above techniques are concentration-based, meaning that they rely on measured concentrations to identify seepage signals without analysing the processes that govern the natural variations. Romanak et al. (2012) discussed the limitations of concentration-based monitoring and recommended process-based techniques, which identify drivers of the natural variability and estimate their magnitude. Specifically, process-based methods are able to address the sensitivity of seepage signal detection criteria to changes in the natural processes due to, for instance, climatic and/or ecosystem variations.

The EU project STEMM–CCS (https://www.stemm-ccs.eu) carried out a first of its kind CO$_2$ release experiment designed to imitate CO$_2$ seepage from an offshore storage site (Fig. 1). The details of different aspects of this unique experiment have been described elsewhere (e.g., Connelly et al., 2019; Dean et al., 2020; Esposito et al., 2021; Flohr et al., 2021). The experiment was carried out at the Goldeneye site (58°N, 0.4°W) in the North Sea where a controlled mixture of CO$_2$ and tracer gases was released into the sediments at 3 m below the seafloor. The suitability of existing and new methods for the monitoring, detection and assessment of potential environmental impacts of the CO$_2$ release were demonstrate/evaluated. Before the release experiment, the STEMM–CCS project conducted a number of cruises around the Gold- eneye site in order to establish environmental and ecological baselines. The project also developed/further-developed novel chemical sensors.
which have been utilized for data acquisition during both the baseline gathering and during the release experiment.

In this study, we use data from the STEMM–CCS release experiment to demonstrate the ability of the recently developed, process-based, C\text{seep} method (Botnen et al., 2015; Omar et al., 2018) to (i) distinguish CO\text{2} seepage signal from the natural variability, and (ii) quantify the excess seepage CO\text{2} dissolved in the seawater, i.e., C\text{seep}. The C\text{seep} method (section 2) estimates the magnitude of the natural variations (ΔC\text{nat}) and filters them out to facilitate the detection of CO\text{2} seepage signal. We demonstrate the aptness of the method to establish a baseline C concentration (C_{b}) and a detection threshold (DT) (section 3). We discuss the benefits of detecting seepage signals through the C\text{seep} method (section 4), which is a promising marine monitoring tool: its implementation is flexible, it is suitable for automation, it can provide large area coverage, and it requires no pre-knowledge of CO\text{2} seepage occurrence (sections 5 and 6).

1.1. Variables of the seawater CO\text{2} system

The dissolution of CO\text{2} in seawater and subsequent equilibrium reactions form a complex chemical system (section 2), often referred to as the seawater CO\text{2} system, equilibrium properties of which are well characterised (e.g., Zeebe and Wolf-Gladrow, 2001). The seawater CO\text{2} system can be characterised by the following four measurable master variables (Dickson, 2010):

Total alkalinity (TA), which is related to the charge balance in seawater:

\[
TA = [\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + \text{[B(OH)]}_3^- + [\text{OH}]^- + [\text{HPO}_4^{2-}]^- + 2[\text{PO}_4^{3-}]^- + [\text{SiO(OH)}_3^+] + [\text{HS}^-] + 2[\text{S}^2^-] + [\text{NH}_4^-]^- + [\text{H}_2\text{SiO}_4]^- + [\text{HF}]^- + [\text{H}_2\text{PO}_4]^- \quad (1)
\]

where brackets denote concentrations. The bicarbonate and carbonate ions ([HCO\text{3}^-] and [CO\text{3}^{2-}]) contribute about 96% of the seawater TA.

Dissolved inorganic carbon (C), which is the sum of the concentrations of all forms of inorganic carbon dissolved in seawater:

\[
C = [\text{CO}_3^-] + [\text{HCO}_3^-] + [\text{CO}_2^-] \quad (2)
\]

where [CO\text{2}^-] represent the sum of the concentrations of CO\text{2} in aqueous solution.

Partial pressure of CO\text{2} (pCO\text{2}) in an air sample in equilibrium with the seawater sample:

\[
p\text{CO}_2 = x\text{CO}_2 \times p \quad (3)
\]

where xCO\text{2} is the mole fraction of CO2 in the gas phase and p is the total pressure.

pH\text{H}(henceforth pH for simplicity) which represents the total hydrogen ion concentration in seawater:

\[
\text{pH} = -\log[H^+] \quad (4)
\]

The above master variables can be measured using a wide variety of techniques (Dickson, 2010) following standard operating procedures (Dickson et al., 2007). TA and C are traditionally determined by acidimetric titration and, currently, discrete sampling and benchtop instrumentation give the highest accuracy.

By measuring any two of the above-mentioned four master variables (along with temperature, salinity, pressure and the knowledge of other non-CO\text{2} acid-base systems in seawater), it is possible to calculate the remaining variables e.g., by using the CO2SYS software (van Heuven et al., 2011). Additionally, by selecting the appropriate two parameters for measurements, the uncertainties in the computed parameters can be held at the same order of magnitude as their experimental errors (Millero, 2007). Sensors that can autonomously measure pCO\text{2} and pH in situ at high frequency are commercially available and show satisfactory precision and accuracy (e.g., Bushinsky et al., 2019). Paired measurements of pH and pCO\text{2}, however, are the least favoured since the computed TA and C values carry the highest uncertainty. Therefore, a major focus of the STEMM–CCS project was to develop new Lab-on-chip (LOC)-based sensor technologies to enable direct high-resolution in situ measurements of pH, TA and C. LOC-based sensors enable the automation of high performance wet-chemical analytical techniques offering high quality measurements usually only possible using benchtop laboratory instrumentation (Beaton et al., 2012; Clinton-Bailey et al., 2017; Rérölle et al., 2013).

Although measuring two of the four master variables gives the basis for a complete determination of the seawater CO\text{2} system, a process-based interpretation of the variability included in the data requires ancillary variables that are indicators for the governing processes. Examples of such variables include nutrients as indicators of biological activity, and salinity as an indicator of mixing between different water masses. Autonomous LOC sensors for the determination of nutrients (Beaton et al., 2012; Clinton-Bailey et al., 2017), and commercial off-the-shelf sensors (Sea-Bird Scientific) for temperature and salinity measurements were deployed on a number of platforms during the CO\text{2} release experiment to collect high-resolution in situ observations. Thus, the STEMM–CCS project demonstrated that high resolution data needed for the application of the C\text{seep} methodology can be acquired by using cost-effective in situ sensors (section 4.1, Table 3).

2. Conceptual description of the C\text{seep} method

When seepage CO\text{2} dissolves in seawater, it reacts with the water forming carbonic acid (H\text{2}CO\text{3}), which rapidly dissociates into bicarbonate ions (Eq. 5), which in turn can also dissociate into carbonate ions (Eq. 6) according to the following equilibrium reactions:

\[
\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{H}_2\text{CO}_3 \leftrightarrow \text{HCO}_3^- + \text{H}^+ \quad (5)
\]

\[
\text{HCO}_3^- \leftrightarrow \text{CO}_3^{2-} + \text{H}^+ \quad (6).
\]

In terms of measurable variables, the net result of the above reactions is an increase in C and a decrease in pH. Thus, any seepage CO\text{2} dissolved in seawater should, in principle, be readily quantified from the resulting concentration changes. However, C in seawater is also influenced by natural processes, such as photosynthesis/respiration, biosynthesis/dissolution of CaCO\text{3} and changes in temperature and salinity. Therefore, the approach of the C\text{seep} method is to quantify the natural variability and filter it out to facilitate the detection of CO\text{2} seepage dissolved in seawater. To achieve this, the method conveniently focuses on changes in C because this variable tracks concentration changes in all of the inorganic carbon species dissolved in seawater (Eq. 2) regardless of whether the changes are due to natural processes or to CO\text{2} seepage. Additionally, C mixes conservatively in the absence of biological activity.

The amount of C in seawater is set largely set by equilibrium reactions whilst natural processes produce deviations through various sinks and sources of seawater constituents. Therefore, the C\text{seep} method assumes that there is an essentially invariant theoretical baseline seawater C (C_{b}), which is dictated by the equilibrium history and physical properties of the water parcel, and over which fluctuations created by natural processes (ΔC\text{nat}) and/or CO\text{2} seepage (C\text{seep}) are superimposed so that:

\[
C = C_{b} + \Delta C_{\text{nat}} + C_{\text{seep}} \quad (7)
\]

where C is the measured background concentration and C_{b} is assumed to be invariant. The terms on the right hand-side of Eq. 7 are determined in two steps (see also Fig. 4): 1

1 Acquisition of background measurements of C (and ancillary parameters) that are influenced by ΔC_{nat} only and not by CO_{2} seepage,
Table 1
Overview of the data used to establish the baseline C (C_{bas}) around the Goldeneye site (Fig. 3).

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Measured parameters</th>
<th>Accuracy</th>
<th>Sampling mode</th>
<th>Sampling period</th>
<th>Sample depth</th>
<th>Data originator(s) (contact person)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TA (μmol kg^{-1})</td>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T (°C)</td>
<td>0.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PO_4 (μmol kg^{-1})</td>
<td>5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NO_3 (μmol kg^{-1})</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C (μmol kg^{-1})</td>
<td>0.08%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TA (μmol kg^{-1})</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS518 and POS527</td>
<td>T (°C)</td>
<td>C</td>
<td>Bottle data</td>
<td>Aug 2018 and Oct 2017</td>
<td>Deepest sample, 10–20 m above the sea floor</td>
<td>STEMM-CS (M. Esposito)</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PO_4 (μmol kg^{-1})</td>
<td>5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NO_3 (μmol kg^{-1})</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TA (μmol kg^{-1})</td>
<td>0.12%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PO_4 (μmol kg^{-1})</td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCI80BL</td>
<td>pH</td>
<td>0.01</td>
<td>Sensor data</td>
<td>Apr-May 2019</td>
<td>126 m</td>
<td>STEMM-CS (A. Schaap)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.4%&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current (direction and speed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ C_{bas} = C^b - \Delta C_{bas} \] (8)

where the superscript ‘b’ stands for background whereas the subscript ‘Δ’ stands for baseline.

2. Application of the model of step (1) to C measurements from the site-monitoring to determine the sum of C_b and C_{seep} (if any seepage occurred) according to:

\[ C_{seep} = C^M - \Delta C_{mix} - C_b \] (9)

where the superscript ‘M’ stands for site-monitoring.

Note the important distinction between \( C^b \) and \( C_b \). The former is fluctuating due changes in \( \Delta C_{mix} \) and, thus, depends on time and location whereas the latter is theoretically invariant. In practice, however, the errors in calculating \( \Delta C_{mix}^b \) affect the values of \( C_b \) values so that they include some variability. This will be referred to as the “remnant variability”. Similarly, errors in calculating \( \Delta C_{mix}^M \) and \( C_b \) affect \( C_{seep} \). In the following section we detail the determination of \( C_b \), and analyse its remnant variability and its effects on \( C_{seep} \) determination.

3. Establishing \( C_b \) for the Goldeneye site

As mentioned in the previous section, the main purpose of establishing \( C_b \) is to facilitate reliable detection of seepage CO\(_2\) signal and quantification of \( C_{seep} \) by comparing baseline and monitoring data. However, baseline determination has its own value. For instance, by analysing \( C_b \) values obtained from Eq. 8 we can check our understanding of the natural variability of C in the study area. A good understanding of the natural variability of C results in a successful characterization of \( \Delta C_{bas} \), which, in turn, results in confident \( C_b \) values. Moreover, by determining \( C_b \) over large spatial and temporal scales we can analyse the variability in \( C_b \) in relation to sampling location/time. This, in turn, allows us to examine over what spatial and temporal scales we can expect an invariant \( C_b \). Furthermore, by computing the difference between \( C_b \) values obtained for different stations/times we yield an estimate on the uncertainty in \( C_{seep} \), which is a measure of the detection threshold for seepage CO\(_2\) signal. All these aspects are detailed in following subsections.

3.1. Baseline data

For the determination of \( C_b \) (Eq. 8) we used water column measurements from the area at and around the Goldeneye site (Fig. 3). An overview of the datasets is shown in Table 1, and it comprises publicly available historical data as well as new data gathered within the STEMM–CCS project. The historical dataset was obtained during multi-year (2001–2002, 2005, 2008, and 2011) basin-wide cruises using discrete sampling and shipboard analyses (Crago et al., 2015; Thomas, 2002). We used a 2-by-2.5-degree subset of this historical cruise dataset to characterize the large-scale spatiotemporal variability around the Goldeneye site. The new dataset (2017–2019) comprises both cruise data and high-frequency in situ sensors data. The cruise data were gathered onboard RV Poseidon using samples drawn from Niskin bottles mounted on a CDT-rosette (Fig. 2A). These water samples were analysed onboard using benchtop instrumentations (Fig. 2B and D) as described in Esposito et al. (2021). We used the data acquired within the geographical area of 0.35°S–2.6°S, 0.3°W–5.2°W, between April and May 2019, measuring at a two-hourly sampling frequency. These data will be henceforth referred to as the JCI80BL dataset.

Our analyses focus on the deep samples (up to 10–20 m above the seafloor) where the CO\(_2\) seepage signals are expected to appear first.

3.2. Determination of \( C_b \) through modelling the natural variability

The processes assumed to govern the natural variability of the seawater CO\(_2\) system around the Goldeneye site are summarized in Fig. 4. They include (i) biological processes (\( \Delta C_{bio} \)) that comprise photosynthesis/respiration (organic matter cycling; \( \Delta C_{ase} \)) and formation/dissolution of CaCO\(_3\) (CaCO\(_3\) cycling; \( \Delta C_{CaCO3} \)), (ii) mixing of water masses (\( \Delta C_{mix} \)), and (iii) air-sea gas exchange (\( \Delta C_{omc} \)) including oceanic...
uptake of anthropogenic CO\textsubscript{2}. In the context of this study, the oceanic uptake of anthropogenic CO\textsubscript{2} from the atmosphere is considered a natural process, i.e., a process that is not influenced by the subsea seepage. Therefore, the $\Delta C_{\text{nat}}$ term in Eq. 8 consists of contributions driven by biology ($\Delta C_{\text{bio}}$), air-sea gas exchange ($\Delta C_{\text{ase}}$), and mixing ($\Delta C_{\text{mix}}$). Furthermore, as indicated in Fig. 4, the contributions of $\Delta C_{\text{nat}}$ are parameterized through a combination of changes in measured variables relative to some arbitrary reference values (Table 2), stoichiometric ratios, and empirical constants.

The details of how the contributions $\Delta C_{\text{bio}}$, $\Delta C_{\text{ase}}$, $\Delta C_{\text{mix}}$, and $\Delta C_{\text{nat}}$ are computed from the model parameters are given in Appendix A. Once the above contributions are determined, we obtain an estimate of $\Delta C_{\text{nat}}$ ($\approx \Delta C_{\text{bio}} + \Delta C_{\text{ase}} + \Delta C_{\text{mix}} + \Delta C_{\text{nat}}$), and then evaluate Eq. 8 to estimate $C_b$.

The mean values and standard deviations for $\Delta C_{\text{bio}}$, $\Delta C_{\text{ase}}$, $\Delta C_{\text{mix}}$, and $\Delta C_{\text{nat}}$ are shown in Table 2. In terms of mean values, the most influential contributions are, in decreasing order, $\Delta C_{\text{ase}}$, $\Delta C_{\text{bio}}$, $\Delta C_{\text{mix}}$, and $\Delta C_{\text{nat}}$. In terms of contributions to $\Delta C_{\text{nat}}$ variability, the contribution order changes to $\Delta C_{\text{ase}}$, $\Delta C_{\text{mix}}$, $\Delta C_{\text{bio}}$, and $\Delta C_{\text{nat}}$ in decreasing order. For details of the above contributions see Fig. A1 in Appendix A.

The variability in $C_b$, $\Delta C_{\text{nat}}$, and $C_b$ is compared in Fig. 5. $C_b$ varies on average between 2090 and 2180 $\mu$mol kg$^{-1}$, with mean and standard deviation values of 2160 $\pm$ 17 $\mu$mol kg$^{-1}$ (Fig. 5A), confirming the highly dynamic CO\textsubscript{2} system parameters known for shelf regions (e.g., DeGrandpre et al., 1997). Changes observed at any given station arise from temporal changes occurring on the sub-seasonal to interannual time scales. The historical data captures the seasonal to interannual variability, whereas the sub-seasonal changes are captured by the JCI180BL data, which were acquired over 22 days (29 Apr–21 May). Spatial changes, on the other hand, are observed between different stations and their magnitude is comparable to the temporal changes. Another noticeable feature is the lower $C_b$ values observed in the coastal stations (Fig. 5A; stations 49, 50, 55, and 75, station locations shown Fig. 3), which are due to the combination of lower salinity and higher primary production (Fig. A1B). The average amplitude of the variability in $\Delta C_{\text{nat}}$ (85 $\mu$mol kg$^{-1}$, Fig. 5B) closely matches that in $C_b$ (90 $\mu$mol kg$^{-1}$, Fig. 5A). Therefore, the average variability in $C_b$ ($= C_b - \Delta C_{\text{nat}}$) is substantially reduced (35 $\mu$mol kg$^{-1}$, Fig. 5C) being 60% less variable than $C_b$. Furthermore, $C_b$ reduces not only the station-to-station variations in $C_b$ (shorter boxes in Fig. 5C compared to those in Fig. 5A) but also the variability within each station (difference between individual boxes Fig. 5C compared to those in Fig. 5A). This is because the variations in individual boxes of $C_b$ and $\Delta C_{\text{nat}}$ partially cancel out.

In summary, Fig. 5 demonstrates that the application of the $C_{\text{step}}$ method substantially reduces the dependency of $C_b$ on time and location, although it does not completely eliminate the spatiotemporal variability. Furthermore, the spatial extent of the study area does not seem to explain the remnant variability in $C_b$. This can be seen in Fig. 5C, where the range in $C_b$ values computed for station 56, JCI180BL, POS518, and POS527 brackets the entire remnant variability despite these data being acquired in quite close locations (Fig. 3).
Fig. 3. Study area in the North Sea centred around the Goldeneye site (58° N, 0.4° W, rectangle). Also shown are depth contours (in m) and the locations sampled during historical cruises (asterisks) and baseline cruises (rectangle). Upper inset: detailed positions and depth contours of STEMM–CCS seabed baseline lander (JC180BLL green star), which is about 2 km east of station 56 of the historical cruises (asterisk), and sampling locations of the STEMM–CCS baseline cruises (POS518, triangles; POS527, circles). Bottom inset: detailed map of the area around the CO₂ release experiment site and the sampling locations during monitoring the released CO₂ by the POS534 cruise (dots = sampling through pumping; circles = sampling from Niskin bottles) and by the JC180BLLL lander (blue star). The latter was deployed ~2.6 m south of the centre of the CO₂ bubble streams. Also shown is the position of the JC180BLL baseline lander (green star), which was deployed 375 m southeast of the CO₂ release point.

Fig. 4. Box diagram showing the required data and parameters for modelling the contributions of natural C variability (ΔC_nat). Also indicated are the outputs of the mathematical model for baseline and monitoring data, and that C_seep is determined as the difference of these outputs. In the model parameters, subscripts r denotes reference values, and Δ-values are computed as the difference: measured minus reference. PA: is potential alkalinity i.e. TA from which the influence of organic matter is eliminated (Eq. A3, Appendix) dPA/dS: is change in PA resulting from salinity changes; PA₀: is the PA concentration at S = 0, i.e., in freshwater; dC/dt: is the annual C increase due to oceanic uptake of CO₂ from the atmosphere.
Table 2
The first part of the table shows arbitrary reference values to which all measured data in this work have been adjusted to. Mean values and standard deviations of the contributions of \( \Delta C_{\text{nat}} \) are shown in the second part (see Appendix A for a boxplot of \( \Delta C_{\text{nat}}, \Delta C_{\text{CO}_2}, \Delta C_{\text{mix}}, \) and \( \Delta C_{\text{seep}} \)).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reference value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>35.1</td>
<td></td>
</tr>
<tr>
<td>( PO_4 ) ((\mu\text{mol kg}^{-1}))</td>
<td>0.62</td>
<td>Mean values of the historical data</td>
</tr>
<tr>
<td>PA ((\mu\text{mol kg}^{-1}))</td>
<td>2232</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>2010</td>
<td>Close to median of the sampling years, 2011.</td>
</tr>
</tbody>
</table>

Table 3
Overview of the released–CO\(_2\) monitoring data used to compute \( C_{\text{seep}} \) values at the Goldeneye site (Fig. 3).

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Measured parameters</th>
<th>Accuracy</th>
<th>Sampling mode</th>
<th>Sampling period</th>
<th>Sample depth</th>
<th>Data originator(s) (contact person)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P POSS34P and</td>
<td>C ((\mu\text{mol kg}^{-1}))</td>
<td>0.08 %</td>
<td>Niskin bottles or continuous pumping</td>
<td>May 10–20, 2019</td>
<td>115–120 m</td>
<td>STEMM-CCS project (M. Esposito)</td>
</tr>
<tr>
<td>POSS34B</td>
<td>TA ((\mu\text{mol kg}^{-1}))</td>
<td>0.1 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>(T) (({}^\circ\text{C}))</td>
<td>0.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PO_4) ((\mu\text{mol kg}^{-1}))</td>
<td>0.3 %</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TA) ((\mu\text{mol kg}^{-1}))</td>
<td>0.05 %</td>
<td>9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T) ((\circ\text{C}))</td>
<td>0.12 %</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>(PH)</td>
<td>0.3 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PO_4) ((\mu\text{mol kg}^{-1}))</td>
<td>10 %</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PH)</td>
<td>0.4 %**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( C = \text{dissolved inorganic carbon}; \ TA = \text{total alkalinity}; \ S = \text{salinity}; \ T = \text{temperature}; \ PO_4 = \text{phosphate}; \) and \( NO_3 = \text{nitrates}.\)

3.3. Evaluation of the uncertainty in \( C_{\text{seep}} \) and detection threshold

The remnant variation in \( C_b \) (Fig. 3C) is due to imperfection of the model for \( \Delta C_{\text{seep}} \) (Eqs. A1–A8) and errors in the model input, i.e., errors in the measured variables (Table 1) and in the estimates of \( dC/dt (1.2 \pm 0.3 \text{\(\mu\text{mol kg}^{-1} \text{yr}^{-1}\))}, \ R_{CP} (117 \pm 16), \) dPA/dS (67.5 ± 7.4 ppm), and \( PA_0 (-67 \pm 262 \text{\(\mu\text{mol kg}^{-1}\))}. We used the historical data to estimate the total error introduced into \( C_{\text{seep}} \) (E_{\text{seep}}) by the above uncertainties and due to the observed station-to-station differences in \( C_b \) (Fig. 5C). To achieve this, we first arbitrarily considered station 57 as a baseline station and the remaining stations as monitoring stations. Since there was no seepage in the area when the historical data were acquired (2001–2011) every station should, in principle, have the same \( C_b \) as station 57 so that \( C_{\text{seep}} = 0. However, the aforementioned uncertainties and remnant variability in \( C_b \) will introduce an error such that:

\[
C_{\text{seep}} = C_b^x - C_b + E_{\text{seep}}
\]

where \( x \) is any of the station numbers shown in Fig. 3 different from 57.

To estimate \( E_{\text{seep}} \) we used Monte Carlo (MC) simulations by

\[
C_b = C_b^{17} - C_b^1 + E_{\text{seep}}
\]

where \( x \) is any of the station numbers shown in Fig. 3 different from 57.
evaluating the underlying model (Eqs. A1–A8; Fig. 4) 50,000 times, i.e., 1000 simulations for each of 50 unique model inputs. We chose MC simulations as an easier alternative for the traditional error propagation by differentiation and summations (Anderson, 1976) which would be rather tedious to program. During the simulations individual uncertainties (with magnitudes as given above and in Table 1) were randomly sampled from normal distributions and added to model inputs which were at their mean values.

The distribution of $E_{\text{Cseep}}$ values driven by the individual uncertainties is well approximated by a normal distribution with a mean and standard deviation of $0.7 \pm 7.6 \mu\text{mol kg}^{-1}$ (Fig. 6; red curve and magenta bars). For comparison, we estimated the $E_{\text{Cseep}}$ values driven by the remnant variability in the historical $C_b$ values (Fig. 5C, for station numbers 48–50, 55–57, and 72–75). To achieve this, we evaluated the underlying model by randomly sampling model inputs from normal distributions with means equal to the observed data and uncertainty set to zero. The resulting $E_{\text{Cseep}}$ distribution is well approximated by a normal distribution with a mean and standard deviation of $-1.0 \pm 11.4 \mu\text{mol kg}^{-1}$ (Fig. 6; black curve and cyan bars).

From the above comparison we deduce that remnant variability in the historical $C_b$ values (Fig. 5C), which results from all kinds of model imperfections, produces $E_{\text{Cseep}}$ values with 1-STD of $\pm 11.4 \mu\text{mol kg}^{-1}$ of which 67% (7.6/11.4) are due to individual uncertainties in the model inputs.

We also used the Monte Carlo simulation described above to assess the influence of the individual uncertainties in the distribution of $C_b$. This time we simulated $C_b$ instead of $E_{\text{Cseep}}$ by holding all model inputs at their mean values and adding to them randomly sampled individual uncertainties. The resulting $C_b$ values were well approximated by a normal distribution with a mean and standard deviation of $2142 \pm 4.7 \mu\text{mol kg}^{-1}$. The remnant variability observed in the historical $C_b$ values (Fig. 5C; for station numbers 48–50, 55–57, and 72–75) could be approximated by a normal distribution with mean and standard deviation of $2140 \pm 10.7 \mu\text{mol kg}^{-1}$. The individual uncertainties in the model inputs, therefore, account for 44% (4.7/10.7) of the remnant variability in historical $C_b$ value and the rest must be ascribed to processes/factors not resolved by our model for $\Delta C_{\text{nat}}$. Two possible candidates of such factors include inter-laboratory differences in $C$ measurements, which can be up to $10 \mu\text{mol kg}^{-1}$ (Bockmon and Dickson, 2015), and differences in sampling procedures. This may introduce systematic errors so that the uncertainty in $C$ increases from 1% (Table 2) to maximum 5% i.e. $\sim 10 \mu\text{mol kg}^{-1}$. In fact, we were able to reproduce the whole remnant variability observed in $C_b$ (1-STD = $\pm 11 \mu\text{mol kg}^{-1}$; Fig. 5C) by repeating the Monte Carlo simulations with a $C_b$ uncertainty of 5%.

From this analysis, we conclude that the processes included in our underlying model (Fig. 4) are adequate to explain the natural $C$ fluctuations in the study area, and the observed remnant variability in $C_b$ arises
mainly from uncertainties in model inputs.

Based on the above results, we define the detection threshold (DT), over which $C_{\text{seep}}$ values will be interpreted as CO$_2$ seepage signal, for two situations. If $E_{\text{seep}}$ is affected by only individual uncertainties in model inputs then a DT = 15.2 μmol kg$^{-1}$ (mean $E_{\text{seep}}$ plus two standard deviations from the red curve in Fig. 6) will be used. On the other hand, when $E_{\text{seep}}$ is expected to be influenced by both systematic errors and individual uncertainties in model inputs then DT = 24.8 μmol kg$^{-1}$ (mean $E_{\text{seep}}$ plus two standard deviations from the black curve in Fig. 6) will be used. The effect of changing DT to 1-STD or 3-STD is discussed in section 3.3.

4. Detection and quantification of dissolved released – CO$_2$

4.1. Monitoring data of the CO$_2$ release

For the computation of $C_{\text{seep}}$ (Eq. 9), we used two $C_{\text{seep}}$ datasets collected during the monitoring of the CO$_2$ release (Table 1). The first is a high-frequency dataset acquired by LOC sensors mounted on a lander (Fig. 7A; Flohr et al., 2021) situated on the seafloor about 2.6 m to the south of the centre of the CO$_2$ bubble plumes. The data obtained by these sensors, which were deployed by RV James Cook to monitor the CO$_2$ bubble plume in the benthic boundary layer (BBL), will be henceforth referred to as the JC180BBL dataset. The sensors on this lander had two inlets for seawater, one at 17 cm above the seafloor and one at 87 cm above the seafloor. The sensor alternated between these two inlets and each measurement took 10 min. Here we used data sampled from both the inlets.

The second dataset is data gathered during a cruise with RV Poseidon using towed video-CTD (Fig. 7B) with an attached water pump to continuously monitor parameters in the water column around and above the CO$_2$ release site (Fig. 3, lower inset, dots). This technique has been described in detail in Linke et al. (2015) and Schmidt et al. (2015). Briefly, the 1-inch inlet tubing connected to an underwater pump was attached to the lower part of the video-CTD frame and the water was continuously pumped at a flow rate of about 30 m$^{-1}$ to sample from specific water depths. A three-way tube connector was installed to split the flow and allow for discrete water sample collection and later on-board analysis.

These observations were made during a full tidal cycle on May 13–14, 16–17, and 19–20, 2019 (Schmidt, 2019). Water column measurements sampled with Niskin bottles were also conducted during the RV Poseidon cruise (POSS34; Fig. 3, lower inset, circles), both above the CO$_2$ release site (May 13–14, 16–17, and 19–20) and few kilometres away from the site (May 10–13). The cruise dataset will be referred to as POSS34P or POSS34B, where P and B denote sampling through pumping or through Niskin bottles, respectively. For this dataset, our analyses focus on the samples taken from 109 m or deeper, i.e., maximum ≈10 m above the seafloor where the strongest signal of the released–CO$_2$ is expected.

For both datasets, we include data acquired before the start of the CO$_2$ release to be able to assess if the application of the $C_{\text{seep}}$ method results false positives, i.e., identify false released–CO$_2$ signals.

4.2. $C_{\text{seep}}$ computed from high-frequency sensor data

$C_{\text{seep}}$ was determined according to Eq. 9, using the high-frequency CO$_2$ release monitoring dataset (JC180BBL), $\Delta C_{\text{nat}}$ computed for the same dataset, and an average $C_0$ value determined from the high-frequency baseline dataset (JC180BL). During the computations of $\Delta C_{\text{nat}}$ and $C_0$, some data quality and paucities were circumvented as detailed in Appendix 8. Since JC180BBL and JC180BL datasets were acquired simultaneously (but at different locations; Fig. 3) using the same sensor technology, we applied DT$_{\text{senor}}$ = 16.6 μmol kg$^{-1}$ as discussed in section 3.3. In other words, only $C_{\text{seep}}$ values larger than DT$_{\text{senor}}$ were confidently interpreted as a signal of released–CO$_2$.

The temporal development of $C_{\text{seep}}$ together with those of the released–CO$_2$ rate and water movement (velocity) are shown in Fig. 8. Notably, prior to the onset of the CO$_2$ release, e.g., May 8–9, the $C_{\text{seep}}$ values are close to zero (2.4 ± 2.3 μmol kg$^{-1}$) as expected since no released–CO$_2$ was dissolved in the seawater. $C_{\text{seep}}$ values that are consistently above DT$_{\text{senor}}$ start to appear on May 15, when the released–CO$_2$ rate was ≥ 15 kg d$^{-1}$. After that date, the CO$_2$ release rate was even higher and elevated/extreme $C_{\text{seep}}$ values ranging from 40 to 160 μmol kg$^{-1}$ regularly appeared and their peak values generally scaled with rate of the CO$_2$ release. Moreover, the elevated $C_{\text{seep}}$ values occurred whenever the current direction was southwards, i.e., aligned with the direction of the sensor location relative to the CO$_2$ bubble stream through the seafloor (Fig. 3, caption). On the basis of the above observations, we conclude that the $C_{\text{seep}}$ method is able to detect the released–CO$_2$ signal in addition to quantifying the concentration of excess carbon dissolved in the sampled seawater. Moreover, the method...
unequivocally detects the released $\text{CO}_2$ for moderate-to-high $\text{CO}_2$ release rates $> 15 \text{ kg d}^{-1}$. For low $\text{CO}_2$ release rate $\leq 5 \text{ kg d}^{-1}$, $C_{\text{seep}}$ values were essentially indistinguishable from zero. The detection of the released $\text{CO}_2$ signals is based on the $\text{DT}_{\text{cruise}}$ criterion. However, the fact that the identified signals are in line with other parameters such as the onset of the $\text{CO}_2$ release, current direction, and the strength of the $\text{CO}_2$ release rate lends confidence to the detected signals.

4.3. $C_{\text{seep}}$ computed from the cruise data

To evaluate Eq. 9 for the cruise observations, we used the mean $C_0$ value obtained for the POS527 data (2157 $\mu\text{mol kg}^{-1}$), while $\Delta C_{\text{nat}}$ values were computed from the monitoring datasets POS534P and POS534B. The baseline and monitoring datasets were acquired on different cruises (Tables 1 and 3), but we assumed no systematic errors in the measurements and applied a $\text{DT}_{\text{cruise}} = 16.6 \mu\text{mol kg}^{-1}$ (section 3.3). The temporal development of the resulting $C_{\text{seep}}$ values together with those of the $\text{CO}_2$ release rate and the distance between the sampling point and the $\text{CO}_2$ bubble stream at the seafloor are shown in Fig. 9.

For the POS534B data (sampled through Niskin bottles), no distinct $C_{\text{seep}}$ maxima were observed and all values were within the $\text{DT}_{\text{cruise}}$. For POS534P (sampled through pumping), on the other hand, $C_{\text{seep}}$ values above the threshold were observed on the night shift of May 16–17 (Fig. 9B). There was also a distinct local maximum of $C_{\text{seep}}$ on the night shift of May 19–20, but this did not rise over the threshold. Based on the $\text{DT}_{\text{cruise}}$ criterion, the application of the $C_{\text{seep}}$ method on this dataset identified a $\text{CO}_2$ release signal only on May 16–17. However, RV Poseidon was investigating the released $\text{CO}_2$ plume from late May 13 (Fig. 9C), and the local $C_{\text{seep}}$ maximum on May 19–20 was probably due to released $\text{CO}_2$. In any case, it is interesting to note that this local maximum occurred when the $\text{CO}_2$ release rate was three times higher than May 17. It seems, therefore, that the cruise data did not capture the $\text{CO}_2$ release signal as good as the sensor data (Fig. 8) even for high $\text{CO}_2$ release rates. Consequently, false negatives might occur more often for the cruise data. Possible causes for the seemingly lower sensitivity of the cruise data are discussed in section 5.2.

There is also one $C_{\text{seep}}$ value (-18.8 $\mu\text{mol kg}^{-1}$) that exceeds the negative side of $\text{DT}_{\text{cruise}}$. This is interpreted as an outlier and indicates that a $\text{DT}_{\text{cruise}} = 19 \mu\text{mol kg}^{-1}$ is more suitable for this data probably because data acquisition on different cruises introduced additional errors in the model inputs. Note, however, that a higher $\text{DT}_{\text{cruise}}$ value would increase the risk of false negatives even more.

Finally, the above identification of released $\text{CO}_2$ signal was based solely on $\text{DT}_{\text{cruise}}$, which depends on the remnant variability in $C_0$. This means that with the availability of a DT value, the application of the $C_{\text{seep}}$ method does not require any pre-knowledge of whether there is a seepage occurring nor its eventual whereabouts. However, any additional information available about the seepage situation is very useful to assess the confidence of the results.

5. Discussion

5.1. Implementation flexibility

The modular nature of the $C_{\text{seep}}$ method gives the possibility of implementing it at different complexities depending on the purpose of the computations. Some examples of this are provided below.

In this work, the aim was to detect $\text{CO}_2$ seepage signal in addition to quantifying the concentration of excess $\text{CO}_2$ dissolved in the sampled seawater. Thus, we implemented the method in its most rigorous form in four modules by (i) setting up a model that includes as many as possible processes known to influence $C$ concentration in the study area (section 2); (ii) establishing a baseline $C$ through minimising the natural variability (section 3.2); (iii) determining a detection threshold (section 3.3); and (iv) quantifying $C_{\text{seep}}$ by evaluating Eq. 9 using $C_0$ and $\Delta C_{\text{nat}}$ values computed from baseline and monitoring data, respectively (section 4). This implementation can be most relevant during the monitoring phase of CCS projects when both identification of anomalous concentration and quantification of the dissolved $\text{CO}_2$ are necessary.

If only determination of $C_0$ is needed, for instance, during site characterisation phase of CCS projects when baseline gathering and understanding/identification of governing processes are the focus, then only the first and second modules of the method need to be implemented. In other cases, the determination of the detection threshold in addition to $C_0$ might be necessary. This can be important for modelling purposes for example. In fact, the recent works of Hvidevold et al. (2015,
In the current underlying model, and with such elevated DT value the signal would be captured only during the periods with higher CO₂ release rate (> 80 kgC d⁻¹, Fig. 8). This calls for improved estimate of the long-term C trend in order to assure reliable detection of signals from lower seepage rates as well. To minimise potential error in the long-term trend, baseline data should be up-to-date. This is not necessarily a cumbersome task since any monitoring data that do not reveal CO₂ seepage can be used to update the baseline.

5.2. Importance of distance to the seepage source, sampling frequency, and measurement accuracy

An interesting finding from the results presented in section 4 is that although the experimental cruise data measured with benchtop instrumentation (POSS534) is more accurate than the in situ sensor data (JC180BBL) (Tables 1 and 3), the latter captured the CO₂ release signal more clearly. For instance, the Cseep peak values from JC180BBL were more regular, in phase with the current velocity, and scaled with the CO₂ release rate. A comprehensive analysis of this issue is beyond the scope of this work, but we note two factors that seem to be important. The first is the relative locations of the seepage source and the sensor, with respect to the currents, which is a key factor to take into account. In particular, the sensor data showed a strong vertical gradient. The Cseep values were based on samples taken through two inlets: at 17 cm and 87 cm above the seafloor (section 4.1) with an average depth of 119.5 m. At the lower inlet Cseep values above DTsensor were between 18 and 170 μmol kg⁻¹ whereas for the upper inlet the corresponding Cseep values were much lower, 17–74 μmol kg⁻¹ (Fig. 8B). This indicates that released—CO₂ signal was nearly double at the lower inlet compared to 70 cm higher. The average sampling depth of POS534P data that were acquired at the experimental site was 116.4 m (114.9–117.4 m) and, thus, around 3 m shallower than the JC180BBL data. This is probably the main reason for the lower Cseep values computed for the former dataset (Fig. 9B versus Fig. 8B). Nevertheless, it is worth mentioning that Cseep values for the POS534P data on 19–20 May were lower than those on 16–17 May (Fig. 9B). This is despite the facts that on the 19th–20th the CO₂ release rate was higher and the average sampling depth was a metre lower. On both days the hydrodynamic conditions and ship position were very real marine monitoring situations, which can be more complicated and challenging than the current release experiment. During monitoring of CCS sites one might not know if seepage is occurring or not, whereas during the experiment we knew where and when to look for a seepage signal. Furthermore, during the experiment, the baseline and monitoring data were acquired simultaneously and at different (but close) locations. Therefore, the difference between the baseline and monitoring C data resulting from natural variability, i.e., ΔCnat, was small and mostly contributed to by ΔCconc. Consequently, as soon as a signal of high C concentration was observed one could intuitively assume (with some confidence) it stems from the CO₂ release. In real monitoring situations, however, spatial and temporal difference between the baseline and monitoring data can be much larger. This, in turn, will produce larger ΔCnat, which might be contributed to by several processes and even conceal the seepage CO₂ signal. Therefore, in real marine monitoring situations, it is essential to model all of the significant natural processes that influence C, such as in the complete implementation of the Cseep method.

Finally, we note the trade-off involving in the above-implied flexibility regarding the timing of baseline data collection relative to the monitoring data. In section 4, we suggested DTsensor larger than DTsensor was more suitable because baseline and monitoring were collected nearly simultaneously for the sensor data, while for the cruises data collection took place in different years. Furthermore, sensitivity calculations showed that if we were to calculate Cseep from JC180BBL using the average Cconc value obtained from the historical cruise data collected in May 2002 then a DTsensor ≈ 30 μmol kg⁻¹ would be required. This indicates some of the long-term natural C changes were not captured by the current underlying model, and with such elevated DT value the released—CO₂ signal would be captured only during the periods with higher CO₂ release rate (> 80 kgC d⁻¹, Fig. 8). This calls for improved estimate of the long-term C trend in order to assure reliable detection of signals from lower seepage rates as well. To minimise potential error in the long-term trend, baseline data should be up-to-date. This is not necessarily a cumbersome task since any monitoring data that do not reveal CO₂ seepage can be used to update the baseline.

Fig. 10. Concentration of salinity normalized C (nC) plotted against that of nitrate (NO₃) using data from baseline cruises (POSS18, POSS27, and historical cruises; grey) and the CO₂ release sensor data JC180BBL (blue). Samples contaminated by the released—CO₂ in the JC180BBL data clearly deviate from the slope of the CN stoichiometry of the data from the baseline cruises.
similar. This suggests, at least for the cruise data, distance to the source alone does not explain the strength of the signal.

The other important factor is the sampling frequency and its effect can be appreciated from the observation that we could not identify any release CO$_2$ signal in the Niskin bottle data (POSS34B) although its analytical method had the best resolution of all the datasets. Therefore, given similar hydrographic settings and distance to the CO$_2$ seepage source, high-frequency data suits better for capturing the quickly changing seepage signal.

Measurement precision and accuracy are also prerequisites for accurate C$_{seep}$ values as illustrated by the following final example. The contribution of CaCO$_3$ precipitation/dissolution to the natural C variations around the Goldeneye site is generally small (Table 2 and Fig. A1C). However, we cannot rule out the possibility that its importance became significant during the CO$_2$ release experiment due to potential CaCO$_3$ dissolution in the sediment. If such a dissolution took place and if reaction products entered into the water column in significant amounts, then TA would increase in the water column. This would, in turn, increase $\Delta$CaCO$_3$ values in the monitoring data, which would make a positive contribution to C$_{seep}$. The experimental TA data from the water column (POSS34, JC18088B) did not show any systematic TA increase in connection with the CO$_2$ release. However, the effect of CaCO$_3$ dissolution on TA might have not been captured due to noise in the sensor TA data (Appendix B, and Fig. B1B) and low resolution of the cruise data. Thus, TA sensor data with high precision/accuracy is needed to confidently monitor the importance of $\Delta$CaCO$_3$.

For up-to-date information on commercially available sensors that can deliver the data necessary for C$_{seep}$ computation, we refer the reader to the International Carbon Coordination Project (IOCCP) which promotes, among others, the development of measurement technology for marine biogeochemistry. On the IOCCP website (http://www.ioccp.org) there is a directory of commercially available hardware (sensors and instruments), references listing documents on standards, best practices and user guides, etc.

5.3. Future outlooks for reducing the C$_b$ variability

Several ways to improve modelling $\Delta$c$_{calc}$ need to be tested in the future to reduce the remnant variability in C$_b$ and keep DT as low as possible. For instance, improvement of the uncertainties in dPA/dS and PA$_0$ maybe be sought by assuming several water mass end-members instead of the two used in the current analysis (section A.3). Also, using equilibrium C value computed from parameters not impacted by CO$_2$ seepage such as S, TA (obtained from S, as in Eq. A5), T, and atmospheric pCO$_2$ using the CO2SYS package (van Heuven et al., 2011) has the potential to improve the long-term trend. This also minimises the concern of potential contamination in C$_b$ by seepage CO$_2$. Lastly, site-specific C$_b$ value may be taken from outputs of well-calibrated regional biogeochemical models. This is specially desirable in situations when it is difficult to acquire new CO$_2$ seepage-free baseline data. Observational data from open databases, such as GLDAP v2 (Olsen et al., 2016) and SOCAT (https://www.socat.info), will be an asset to calibrate the models.

6. Summary and concluding remarks

In this work we analysed baseline data from historical and recent cruises as well as monitoring data from the first-ever CO$_2$ release experiment in the north-western North Sea, carried out as part of the EU project STEMM–CCS. We successfully demonstrated the ability of the C$_{seep}$ method to (i) adequately predict natural C variations around the Goldeneye site, over seasonal to interannual time scales; (ii) establish a process-based baseline concentration (C$_b$) with minimal variability; (iii) determine CO$_2$ seepage detection threshold and reliably differentiate released–CO$_2$ from natural variability; and (iv) quantify released–CO$_2$ dissolved in the sampled seawater (C$_{seep}$) with concentrations above a pre-defined threshold.

In addition to the above mentioned utilities, we also showed that the C$_{seep}$ method can be implemented in modules depending on the purpose, and with the availability of detection threshold it does not require prior knowledge of CO$_2$ seepage. Furthermore, the detection threshold used in this study ranged 16.6–19.0 $\mu$mol kg$^{-1}$, which was 18–20 % of the observed natural variability in the study area (Fig. 5), illustrating the high sensitivity of the method. Therefore, the C$_{seep}$ method can be used to monitor large areas by employing pre-defined detection thresholds, C$_b$ value(s), and mobile platforms, e.g., autonomous underwater vehicles equipped with fast-response sensors.

Based on the above results, we conclude that the C$_{seep}$ method features high sensitivity, automation suitability, and with a pre-defined detection threshold, it requires no pre-knowledge of seepage occurrence. Thus, it represents a powerful future monitoring tool both for large and confined marine areas.

Current CCS monitoring requirements comprise of five steps (Dixon and Romanak, 2015): background (baseline) measurements, assessment of CO$_2$ storage performance in the reservoir, detection of leakage, and (if leakage is detected, suspected or alleged) quantification of leakage and impact assessment. A particular strength of the C$_{seep}$ method is that, if used as a water column monitoring tool, it can address several of the above requirements. Other methods were also tested during the STEMM–CCS project to detect, attribute, and quantify the CO$_2$ release. For instance, natural tracers that are inherent to the injected CO$_2$ ($\delta^{13}$C$_{CO2}$, $\delta^{18}$O$_{CO2}$) and artificial tracers added to it (Kr, SF$_6$, C$_3$F$_8$) were tested to identify anomalies originating from the released-CO$_2$ rather than natural sources (Flohr et al., 2021). Studies on the presentations and comparisons of the performance of the different methods are underway and expected to enter the scientific literature soon.

Authors’ contribution

Abdirahman M. Omar: conceptualization, methodology, data curation, coding, formal analysis, writing – original draft including visualization.

Maribel I. García-Ibáñez: methodology, coding, writing – reviewing and editing.

Allison Schaap: sensor data curation, writing – reviewing, editing and language.

Anna Oleynik: methodology, coding, writing – reviewing and editing.

Mario Esposito: cruise data curation, writing – reviewing and editing.

Emil Jeansson: data analyses, writing – reviewing and editing.

Socratis Loucaides: sensor data curation, writing – reviewing and editing.

Helmuth Thomas: historical data curation, writing – reviewing and editing.

Guttorm Alendal: writing – contextualization, reviewing, and editing.

Declaration of Competing Interest

None.

Acknowledgement

This work has received funding from the European Union’s Horizon 2020 research and innovation programme (contract No. 654462, STEMM–CCS), the ACT programme (Accelerating CCS Technologies, Horizon2020 project No 294766, the ACTom project), and the Research Council of Norway (Project No. 254711, BayMoDe). H. Thomas acknowledges support by the German Academic Exchange Service (DAAD, #57429828) from funds of the German Federal Ministry of Education and Research (BMBF). We acknowledge Martin Arundell, Statthys Papatdimirotiu, Matthew Mowlem, John Walk, and Chris Cardwell at the
National Oceanography Centre for the development of the in-situ pH sensors used during the release experiment. We would like to thank the Captains and the crew on RV Poseidon for their excellent support during the three expeditions. Thanks also to André Mutzberg, Dominik Jasinski and Maria Martínez-Cabanas at GEOMAR for their help with sampling and measuring nutrients and carbonate parameters. We are also grateful for two anonymous referees who’s comments and suggestions improved the manuscript.

Appendix A. Contributions of natural C variability (ΔCnat)

In the following subsections, we describe the details on how the components of ΔCnat (ΔComc, ΔCaCO3, Δmix, and Δase) are computed (Fig. 4). It must be noted that here we are not interested in quantifying ΔCnat integrated over a given period (e.g., one year). Rather, we are interested in quantifying how much (the contribution) each component of ΔCnat caused the measured C concentration to deviate from the theoretical, nominally constant value of C0 that should exist for a given reference value of S, phosphate (PO4), PA and Δase (Table 2).

A.1 Biology-driven variability

A.1.1 Contribution of organic matter cycling

The formation of organic matter through photosynthesis takes up C and nutrients from seawater, whereas the decay of organic matter through respiration releases C and nutrients back to seawater. Moreover, changes in C and nutrients during photosynthesis and respiration take place in stoichiometric ratios (e.g., Anderson and Sarmiento, 1994; Redfield, 1934; Redfield et al., 1963). Therefore, we quantified the change in C related to photosynthesis and respiration (ΔComc) using observed changes in phosphate (PO4) relative to the reference value PO4ref (Table 2) and the stoichiometric ratio between C and PO4 (RCP) according to:

$$\Delta C_{\text{omc}} = R_C(PO_4 - PO_4^{\text{ref}})$$

(A1)

Fig. A1. Modelled contributions of the natural C changes (ΔCnat) arising from the following individual processes: (A) organic matter cycling (ΔComc < 0: production, ΔComc > 0: respiration relative to the reference PO4 in Table 2); (B) salinity changes, i.e. mixing between water masses (ΔCmix); (C) calcium carbonate cycling (ΔCaCO3 < 0: formation, ΔCaCO3 > 0: dissolution relative to the reference PA in Table 2); and (D) ΔCnat, i.e., the sum of the above contributions plus ΔCase. For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are crosses.
We used RCP value of 117 ± 16, which brackets the different values reported in the literature (Redfield, 1934; Anderson and Sarmiento, 1994; Körtzinger et al., 2001).

The resulting values for ΔC_{omic} are shown in Fig. A1 A and these range from -40 to +40 μmol kg⁻¹. Generally, the shallow near-coast stations (49, 50, and 55; Fig. 3) show net organic matter production relative to the reference state (Table 2), while the rest of the stations show either net respiration or a balance between respiration and photosynthesis.

A.1.2 Contribution of calcium carbonate (CaCO₃) cycling

In situ formation (dissolution) of CaCO₃ structures remove (release) CO₂ from (to) seawater and therefore decrease (increase) TA and C according to Eqs. 1 and 2. The resulting TA and C changes occur in a stoichiometric ratio of 2:1 (Zeebe and Wolf-Gladrow, 2001). Consequently, TA data can be used to (i) monitor the in situ CaCO₃ changes, and (ii) quantify the resulting C change (ΔC_{CaCO₃}) using (e.g., Brewer, 1978):

$$\Delta C_{\text{CaCO}_3} = 0.5 \ast (\Delta \text{PA} - \Delta \text{PA}^{\text{ref}})$$

where PA is the potential TA computed from the measured TA and nitrate (NO₃) using

$$\text{PA} = \text{TA} + 1.36^*\text{NO}_3$$

In Eq. A2, nPA is salinity normalized PA (Friis et al., 2003):

$$n\text{PA} = \frac{(\text{PA} - \text{PA}_0)S^{\text{ref}}}{S} + \text{PA}_0$$

PA^{\text{ref}} is the theoretical PA predicted for the reference salinity (S^{ref}) (Table 2). To determine PA^{ref} we regressed computed PA with salinity and the best fit equation was:

$$\text{PA} = \frac{d\text{PA}}{dS} S + \text{PA}_0$$

where the slope dPA/dS = 67.5 ± 7.4 μmol kg⁻¹ and the intercept PA₀ = -46 ± 261 μmol kg⁻¹. Note, the fact that our observational data do not include TA values for freshwater and/or brackish water results in the high uncertainty of the intercept. Eq. A5 and S^{ref} = 35.1 were used to compute PA^{ref} (Table 2). Moreover, the predicted PA value at zero salinity (PA₀) was determined from the intercept of Eq. A5.

The resulting values for ΔC_{CaCO₃} were generally less than ±10 μmol kg⁻¹ (Fig. A1C) and were insignificant compared ΔC_{omic} (Fig. A1A).

A.2 Air-sea gas exchange-driven variability

At the air-sea interface, CO₂ is exchanged between atmosphere and ocean. The direction of the flux is dictated by the CO₂ partial pressure difference between surface seawater and the overlying air (ΔpCO₂ = pCO₂_{2sea} - pCO₂_{2air}). The CO₂ flux is from (to) the atmosphere towards (from) the ocean if ΔpCO₂ > 0 (ΔpCO₂ < 0). Furthermore, the strength of the flux scales with wind speed (e.g., Sweeney et al., 2007). This process influences C in two ways. First, fluctuations in ΔpCO₂ and/or in wind speed cause fluctuations in C in the surface water and some of this signal could be transported to the deeper waters. However, given that the northern North Sea is seasonally stratified, this fluctuation has only a strong impact on the surface water and we thus neglect its effect in this work. Second, since we are using data covering a long period (18 years), accumulation of anthropogenic CO₂ (C_{ant}) is expected to steadily increase C over time. The rate of this increase could not be discerned from the low-resolution data we used here. However, Omar et al. (2019) analysed a decade of underway surface data in the northern North Sea and reported an annual C increase of 1.2 ± 0.3 μmol kg⁻¹ yr⁻¹. The addition of C_{ant} does not necessarily take place locally in the North Sea, but can be carried by advection of water masses into the region, e.g., with the inflow of North Atlantic Water and, therefore, should be expected to impact also the deep waters. Thus, we used the estimate of Omar et al. (2019) to adjust all C data to the year 2010 according to:

$$\Delta C_{\text{ant}} = C_{\text{ant}} = 1.2 \ast (\text{yr} - 2010)$$

where yr is the year of observation.

For the historical C values, ΔC_{ant} values ranged between -10.8 and 1.2 μmol kg⁻¹. For the C data acquired in 2017, 2018, and 2019 ΔC_{ant} values were 8.4, 9.6, and 10.8 μmol kg⁻¹, respectively.

A.3 Mixing-driven variability

Mixing of water masses changes the distribution of C and TA in the water column. Here, we consider the PA–S linear relationship in Eq. A5 to be the result of mixing between low-salinity Coastal Water and high-salinity North Atlantic Water. Other water mass end-members maybe contributing to the observed S relationship, but the effect of this is accounted for through the substantial uncertainties in the regression coefficients of Eq. A5.

Measured C is not a conservative parameter because of the sinks/sources driven by ΔC_{bio} and ΔC_{ase}. However, once the fluctuations arising from ΔC_{bio} and ΔC_{ase} are accounted for we assume that the resulting modified C mix conservatively just like PA (above). Therefore, we computed the contribution of salinity changes (ΔC_{mix}) using the difference between measured and reference salinity and the slope of Eq. A5:

$$\Delta C_{\text{mix}} = \frac{d\text{PA}}{dS} (S - S^{\text{ref}})$$

Finally, C_b was calculated as

$$\Delta C_{\text{mix}} = \frac{d\text{PA}}{dS} (S - S^{\text{ref}})$$
\[ C_b = C - (\Delta C_{baso} + \Delta C_{sea} + \Delta C_{mix}) = C - \Delta C_{nat} \]  
\hspace{1cm} (A8)

where the terms in parenthesis are by definition \( \Delta C_{nat} \). As can be seen from Fig. A1B, values of \( \Delta C_{mix} \) range between -40 and 20 \( \mu \text{mol kg}^{-1} \) and are lower for the coastal water stations. This means that higher \( C \) adjustments were necessary for the low-salinity coastal water to be brought to \( C \) levels corresponding to reference samples with \( S = 35.1 \).

Appendix B. Circumvention of data coverage and quality limitations

The monitoring and baseline datasets JC180BBL and JC180BL, respectively, included some differences in temporal coverage and data quality as depicted in Fig. B1. The following limitations are notable from the graphs:

First, salinity was measured only at the JC180BL lander (Fig. B1A). We circumvented this limitation by using the same salinity values for both sites, i.e. we implicitly assumed that there was no significant systematic salinity difference between the two sites, which are < 0.5 km apart (Fig. 3). A comparison between the salinity of the deepest samples in the POSS34P dataset (Fig. B1A, black dots) and those measured at the JC180BBL lander (Fig. B1A, navy dots), when \textit{RV Poseidon} was up to 1 km away from the lander, showed practically indistinct salinity values supporting the above assumption. For example, the largest observed deviation of 0.05 would produce a 3.4 \( \mu \text{mol kg}^{-1} \) difference in PA (according to Eq. A5) and 1.7 \( \mu \text{mol kg}^{-1} \) difference in \( C_b \) (according to Eq. A7). Both these changes are way lower than other uncertainties involved in the determination of the \( C_{seep} \) (section 3.3).

Second, TA values from the JC180BBL dataset were highly noisy (Fig. B1B, navy dots). This could partly result from regular disturbance of sediment sampling activity conducted with a ROV moving around and inserting/removing sediment-based instruments quite close to the sensors for TA and PO4 (below). This could have caused pore waters with higher TA and PO4 to mix into the water column occasionally. In any case, we dealt with the noise by using the average value of TA (2333 \( \mu \text{mol kg}^{-1} \)) measured at JC180BL. This is a reasonable approximation given that (i) mean values of the TA values from the two landers are essentially indistinguishable (2309 ± 56 \( \mu \text{mol kg}^{-1} \) vs. 2333 ± 5 \( \mu \text{mol kg}^{-1} \)), and (ii) TA is mainly a function of \( S \) and we used the same \( S \) values for both sites (above).

Third, PO4 values from the JC180BBL dataset are noisier than their JC180BL counterparts (Fig. B1D) probably due to the same reasons as in the TA data (above). Furthermore, from Fig. B2 we noticed that while all cruise data falls around a line defining the stoichiometric relationship known to exist between PO4 and NO3 (e.g., Redfield, 1934; Redfield et al., 1963; Anderson and Sarmiento, 1994), the sensor data deviates from this relationship, especially for high PO4 values. We also noticed that, according to the cruise data, PO4 values higher than 1 \( \mu \text{mol kg}^{-1} \) are not typical for the study area. Thus, we considered any sensor PO4 values higher than 0.8 \( \mu \text{mol kg}^{-1} \) as suspicious. This also implies that some of PO4 variability of the sensor data is noise. Thus, for the computation of \( C_b \) (section 3.2) in the JC180BL data we used the average PO4 (0.6 \( \mu \text{mol kg}^{-1} \)) based on values \( \leq 0.8 \mu \text{mol kg}^{-1} \).

Fourth, we also noted that NO3 values from the JC180BBL dataset are systematically lower than those from the JC180BL by around 1 \( \mu \text{mol kg}^{-1} \) (Fig. B1C). On the other hand, PO4 values from the JC180BBL are higher than those from JC180BL. We considered the less noisy JC180BBL NO3 measurements to be more trustworthy. Therefore, for the computation of \( \Delta C_{nat} \) from JC180BBL data (section 4.2) we used a PO4 value of 0.41 \( \mu \text{mol kg}^{-1} \) calculated from the average NO3 concentration (3.74 \( \mu \text{mol kg}^{-1} \)) using the relationship defined by the cruise data (Fig. B2).

![Fig. B1. Four days long example of: measured (A) \( S \), (B) TA, (C) NO3, (D) PO4, and (E) computed \( C \). JC180BL data points are shown in red, JC180BBL in navy blue. Also shown are measured deep water \( S \) for POSS34P (A, black) when \textit{RV Poseidon} was 0.6 to 1.6 km away from the JC180BBL lander.](image-url)
Fig. B2. Nitrate (NO₃) vs Phosphate (PO₄) from JC180BL (red), JC180BBL (blue), POSS18, POSS27, POSS34 (black), and historical cruises (grey). Blue dots demonstrate that NO₃ and PO₄ are decoupled for the sensor data. The solid line indicates the best fit for all cruise data, NO₃ = 18.3(±1)*PO₄ – 3.84(± 0.7).

References
