International Journal of Greenhouse Gas Control 64 (2017) 99–112

Contents lists available at ScienceDirect

International Journal of Greenhouse Gas Control

journal homepage: www.elsevier.com/locate/ijggc

Monitoring of offshore geological carbon storage integrity: Implications of natural variability in the marine system and the assessment of anomaly detection criteria

Jerry Blackford*, Yuri Artioli, James Clark, Lee de Mora

Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, United Kingdom

ARTICLE INFO

Keywords:
Offshore geological storage
Marine
Monitoring
Baseline variability
Biogeochemistry
Carbonate system

ABSTRACT

The design of efficient monitoring programmes required for the assurance of offshore geological storage requires an understanding of the variability and heterogeneity of marine carbonate chemistry. In the absence of sufficient observational data and for extrapolation both spatially and seasonally, models have a significant role to play. In this study a previously evaluated hydrodynamic-biogeochemical model is used to characterise carbonate chemistry, in particular pH heterogeneity in the vicinity of the sea floor. Using three contrasting regions, the seasonal and short term variability are analysed and criteria that could be considered as indicators of anomalous carbonate chemistry identified. These criteria are then tested by imposing a number of randomised DIC perturbations on the model data, representing a comprehensive range of leakage scenarios. In conclusion optimal criteria and general rules for developing monitoring strategies are identified. Detection criteria will be site specific and vary seasonally and monitoring may be more efficient at periods of low dynamics. Analysis suggests that by using high frequency, sub-hourly monitoring anomalies as small as 0.01 of a pH unit or less may be successfully discriminated from natural variability – thereby allowing detection of small leaks or at distance from a leakage source. Conversely assurance of no leakage would be profound. Detection at deeper sites is likely to be more efficient than at shallow sites where the near bed system is closely coupled to surface processes. Although this study is based on North Sea target sites for geological storage, the model and the general conclusions are relevant to the majority of offshore storage sites lying on the continental shelf.

1. Introduction

The effectiveness of carbon dioxide capture and storage (CCS) as a greenhouse gas emissions reduction strategy (IPCC, 2005; IEA GHG, 2008) depends in part on a rigorous demonstration of storage integrity. Regulations governing CCS vary from country to country, but in general require the storage site operator to monitor the deep geological storage complex for leakage and perform an environmental impact assessment of any plausible leakage event at the surface (Dixon et al., 2015). Further, there may be a requirement to quantify leakage, should it occur, with a view to carbon accounting (IPCC, 2006). Primary monitoring of storage reservoirs will utilise seismic techniques capable of imaging CO_2 through a sedimentary overburden of the order of a kilometre thick. However such techniques have limitations in that the detection threshold may be of the order of 10^3t CO_2 and are expensive to perform (Jenkins et al., 2015). Monitoring for emissions at the surface (land or sea floor) therefore provides an important secondary monitoring strategy which can be deployed more frequently and rapidly to detect or respond to any anomalies which might indicate leakage (Blackford et al., 2015). Such surface monitoring is also necessary for environmental impact assessment and may have the potential to quantify CO_2 flow. Critically, a comprehensive monitoring program can provide assurance that no leakage is occurring, as is expected for appropriately sited and operated storage programs.

Globally many potential storage reservoirs are located offshore, for example in NW Europe, China, Japan, Korea, Australia, Brazil and the United States. Consequently the ability to detect or discount anomalous CO_2 emissions from offshore storage requires marine deployments of suitable instrumentation at or near the sea floor.

Offshore storage sites are predominantly coastally located, frequently under continental shelves with overlying water depths between 10 and 250 m (Bradshaw and Dance, 2005). Monitoring systems are likely to need a combination of fixed sea floor landers, situated near known risk points (e.g. the injection point) and autonomous underwater vehicles patrolling wider areas (Blackford et al., 2015). Deployment of sea floor instrumentation at these depths is routine, and

* Corresponding author.
E-mail address: jcb@pml.ac.uk (J. Blackford).

http://dx.doi.org/10.1016/j.ijggc.2017.06.020
Received 20 December 2016; Received in revised form 15 May 2017; Accepted 29 June 2017

methods of transferring data to land in near real time established. Nevertheless, developing instrumentation that can efficiently cover the area of the storage complex, which may be in excess of 100 km^2, requires some technological development. Further an important consideration is what to measure? Research suggests that leakage events are complex and resulting signals heterogeneous in time and space. In particular tidal mixing may cause plumes of high CO_2 water to circulate a release point (Blackford et al., 2013, 2014; Maeda et al., 2015), following a tidal ellipse, and bubble plumes may be intermittent due to both the tidal cycle and evolution of sub surface pathways (Blackford et al., 2014; Bergès et al., 2015; Shitashima et al., 2015). Further the morphology of individual bubble plumes can cause considerable small scale heterogeneity in chemical signatures (Atamanchuk et al., 2015). It is likely therefore, that sensors mounted on fixed platforms or autonomous vehicles would be exposed to continuous oscillations between normal and release plume conditions. Current recommendations are for the combined deployment of at least two types of sensor: chemical, sampling for changes in carbonate chemistry (e.g. pH or pCO_2) and other associated variables and acoustic, using either sonar or hydrophones to locate bubble plumes (Blackford et al., 2015). Further, it has been proposed that monitoring strategies should be hierarchical in approach, initially confined to identifying anomalies (or the lack thereof) only, thus maximising efficiency. Only if an anomaly is identified should more detailed surveys be performed to confirm, attribute and assess the phenomenon (Blackford et al., 2015; Shitashima et al., 2013; Romanak et al., 2012).

Economically efficient and reliable monitoring strategies are the goal for both site operators and regulators. This translates to minimising the deployment of instrument platforms, while maximising detection range, spatial and temporal coverage and accuracy. Alongside sensor development, a crucial component is knowing when a measured signal should be judged anomalous and worthy of further, more expensive, scrutiny. Understanding natural variability and heterogeneity in relation to likely signals of leakage is therefore vital.

This paper addresses the natural variability of marine chemistry relevant to CO_2 (carbonate chemistry), using the geological storage rich North Sea as an example to illustrate how seasonal and spatial heterogeneity affect detectability and the selection of appropriate anomaly detection criteria. Despite this region being one of the most intensely sampled in the marine world, there is a significant lack of observations characterising the carbonate system at or near the sea floor. Further, existing observations are generally not targeted at revealing variability at temporal and spatial scales relevant to leakage, i.e. over some 10 s of meters or the tidal cycle. For this reason we turn to coupled hydrodynamic biogeochemical models which provide comprehensive spatial and temporal fields of marine chemistry. Although such models are never perfect representations of reality, evaluation against available observations provides a degree of confidence in outputs.

The reaction kinetics of CO_2 in seawater are well known (Zeebe and Wolf-Gladrow, 2001; Dickson, 2010). Adding CO_2 to seawater leads to an increase in bicarbonate ions (HCO_3^-) and hydrogen ions, (H^+, measured as a decrease in pH) and a decrease in carbonate ions (CO_3^{2-}) according to the following equations.

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow HCO_3^- + H^+$$

$$CO_3^{2-} + H^+ \rightarrow HCO_3^-$$

The reaction kinetics are controlled by temperature, pressure, salinity and alkalinity (the capacity of seawater to neutralise acid). Total dissolved CO_2 in seawater ($CO_2 + H_2CO_3 + HCO_3^- + CO_3^{2-}$) is known as Dissolved Inorganic Carbon (DIC). pH ($-\log_{10}$ [H^+], although see Zeebe and Wolf-Gladrow, 2001, for a formal definition) or pCO_2 (the partial pressure of CO_2 in seawater) are the most accessible parameters for routine, automated measurement. This paper takes pH (seawater scale) as the parameter of choice; qualitatively the same outcomes would apply if pCO_2 were considered.

In marine systems the concentration of DIC and/or the resulting pH and pCO_2 vary according to a number of processes:

- Biological uptake of DIC occurs via photosynthesis and calcification (the formation of calcium carbonate shells etc.). Release of DIC occurs via respiration and dissolution of carbonate structures. These processes are decoupled in time and space and can be especially dynamic in coastal waters, rich in nutrients. There is variability on diurnal and seasonal cycles as well as in response to stochastic weather events and other external influences.
- External inputs, e.g. advection of oceanic or riverine water and rainfall with varying DIC, or alkalinity contents. Variability on ocean boundaries is dominantly seasonal while atmospheric and terrestrial influences are also driven by stochastic weather events.
- Changes in temperature associated with the annual cycle and the mixing or separation (stratification) of different water bodies. The primary signal is seasonal, along with gradients which may be vertical, horizontal (across fronts) or latitudinal.
- Exchange across the air-sea interface serves to equilibrate atmospheric and seawater partial pressures over monthly timescales, so both out gassing and uptake of CO_2 can occur depending on the sum of biological and physical processes. Over annual to decadal timescales the continued emission of anthropogenic CO_2 to the atmosphere is responsible for a gradual net uptake of CO_2 by the oceans, resulting in an increase in DIC and pCO_2 and a decline in ocean pH (ocean acidification).

There is sufficient observational evidence to demonstrate significant spatial, annual and short term variability of the marine carbonate system, including pH (Hofmann et al., 2011; Thomas et al., 2005, 2007; Bates, 2007) which reveal annual ranges of $\sim$0.1 pH units in the open ocean to ranges of 1.0 pH units in near shore/estuarine systems. In the North Sea, away from the immediate coast the range of pH is 0.2–0.3 pH units, over an annual cycle (Clargo et al., 2015).

If leakage occurred at the sea floor the resulting pH change would scale with the release rate, but will dissipate with distance from the leak point. For a small leak a clearly anomalous pH signal may be restricted to a few metres from source (Dewar et al., 2013), consequently monitoring over an entire storage complex with limited resolution may well depend on identifying signals at some distance from the source that are similar to natural variability.

In this paper a previously evaluated marine system model of the North West European Shelf (Artioli et al., 2012; Blackford and Gilbert, 2007) is used to produce a three dimensional, 30 year time-series of marine pH (Section 2). The pH range and it's variability over relevant spatial and temporal scales are quantified along with a selection of criteria that could be useful indicators of anomalies (Section 3). These criteria are tested by applying a comprehensive range of pseudo leak signals to the model output, assessing how successful detection varies with the magnitude of the anomaly, its timing and the frequency of monitoring observations (Section 4). Finally the application of the outcomes to global offshore storage initiatives is discussed (Section 5).

2. Model methodology

2.1. The model system

The model system dynamically simulates the spatial and temporal evolution of DIC over the North West European Shelf and includes all of the processes that significantly alter DIC as outlined above. The model system has a long development history and has previously been used to examine carbonate chemistry in the region (Blackford and Gilbert, 2007; Wakelin et al., 2012; Artioli et al., 2012, 2014) and CCS leakage scenarios (Blackford et al., 2008). The model system (Fig. 1) comprises of a coupling between a 3D hydrodynamic model (POLCOMS, Holt and James, 2001), the ERSEM model of marine ecosystems (Baretta et al.,

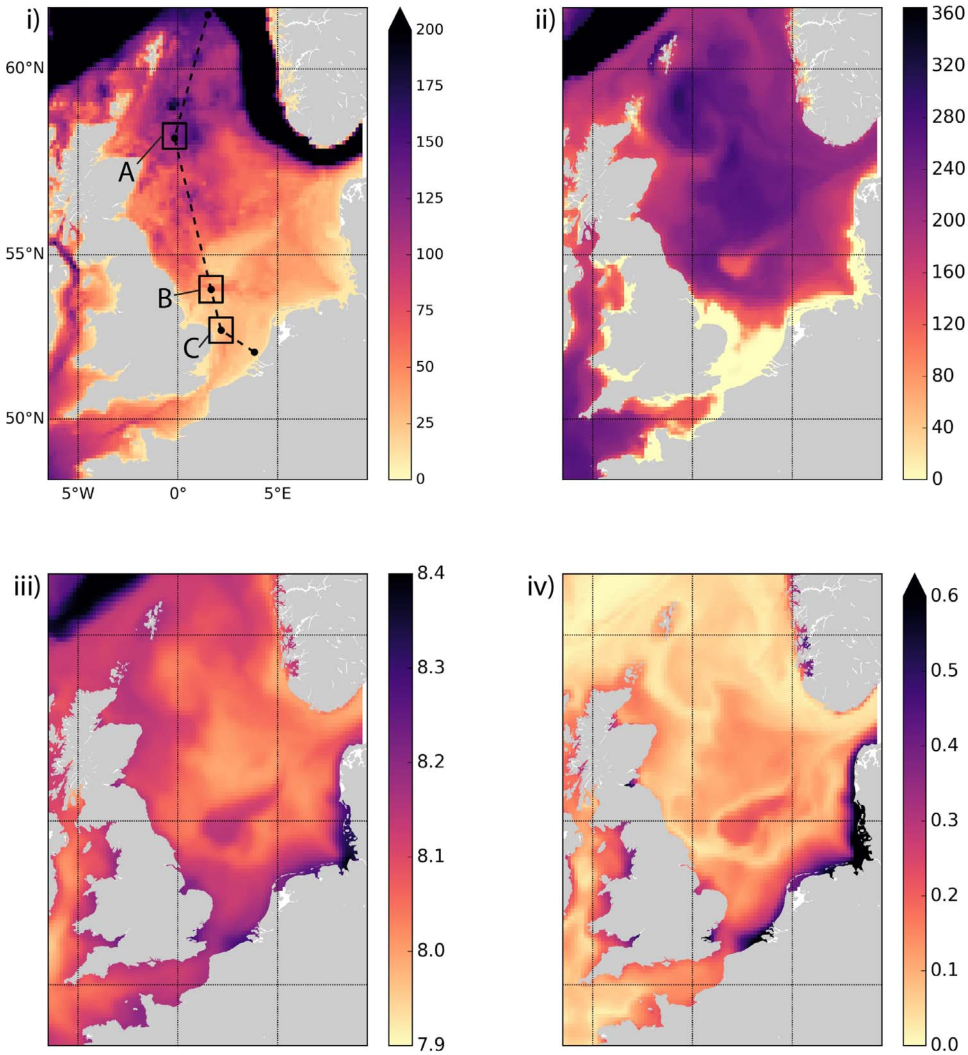

Fig. 1. i) Bathymetry of the North West European Shelf Seas, projected onto the model grid (m). Locations of the localities used in this study are shown. A) Seasonally stratified region, B) Intermittently stratified region, C) Permanently mixed region. The transect used to discuss spatial patterns is also illustrated. ii) Broad physical characteristics of the region illustrated by the frequency (days per year) of vertical thermal stratification. Model derived maps of: iii) pH mean over the annual cycle, iv) annual pH range.

1995; Blackford et al., 2004; Butenschön et al., 2016) and a carbonate system model (Blackford and Gilbert, 2007). The hydrodynamic model uses a rectangular grid of approximately 12 km in the horizontal with 40 vertical layers, with layer depth varying depending on the local depth of the water column. The ecosystem model describes a semi-complex community of primary producers and their grazers, and fully resolves the carbon cycle, including photosynthetic and respiratory process in both the pelagic and benthic systems. The carbonate chemistry module uses the internationally agreed protocols for dissociation constants (Dickson, 2010). The model system is forced by atmospheric forcing (wind, cloud cover, surface heat flux) extracted from the ECMWF ERA40 reanalysis product (Uppala et al., 2005) and inputs (nutrients, DIC, alkalinity) from the principle river systems draining into the North Sea. Open boundary conditions are taken from global reanalysis models and the model simulated for the period 1975–2004.

2.2. Evaluation

The model system has been subjected to rigorous evaluation throughout its development history, focussing on particular model elements e.g.: hydrodynamics (Holt et al., 2005); biochemistry (Allen et al., 2007; Holt et al., 2012); zooplankton (Lewis et al., 2006) and the carbonate system (Artioli et al., 2012). As is typical for models of this nature, the physical model components demonstrate the best correlation with observations, followed by biogeochemistry with higher trophic levels lagging (Radach and Moll, 2006). The models ability to reproduce the fundamental components of the carbonate system DIC and alkalinity is reasonable, the evaluation against pH (seawater scale) less so, the latter biased by the lack of seasonality in available pH observations. An analysis of error shows that these are mostly associated with coastal waters influenced by major river plumes, where sufficient data to characterise river flow in the model is lacking. Other mismatches are associated with peaks in primary production; small errors in timing or position of these features in the model leads to large deviations in direct like to like evaluation. Nevertheless the model is accepted to qualitatively represent the NW European shelf system accurately, with reasonable quantitative skill, making it appropriate for this study of the general trends and dynamics of the carbonate system (RMSE = 0.04, Bias = −0.0008 for pH, Artioli et al., 2012). In particular the specific regions chosen for detailed analysis in this study are not associated with systematic or significant error in carbonate chemistry (Artioli et al., 2012). The atmospheric data have a 6-hourly resolution and are linearly interpolated on to the 20 min time step used by the model. As such it is likely that diel scale variability of pH is under-estimated by the model, although the dynamics of the near-bed region are not strongly coupled to surface forcing.

2.3. Analysis approach

In addition to quantifying near sea floor carbonate chemistry and its drivers across the NW European Shelf, three localities coincident with potential geological storage formations are utilised for more detailed analysis (Fig. 1i). Each possesses contrasting oceanographic and biogeochemical conditions and hence distinct carbonate chemistry dynamics. Further a transect connecting each locality, extending northwards to the extent of identified potential storage sites in the UK sector of the North Sea (Pale Blue Dot, 2016) and south to the approximate location of a proposed Dutch storage project (ROAD, Read et al., 2014) is used to demonstrate the range of influences on and general trends of carbonate chemistry in the North Sea region.

The localities are:

A **Seasonally stratified**, northern North Sea, water depth of 120 m. Storage centred at 58.00°N, 0.35°W, and an approximation of the suggested Goldeneye storage site.

B **Intermittently stratified**, mid North Sea, water depth of 38 m. Storage centred at 54.22°N, 1.00°E, and an approximation of the Endurance storage site previously identified for the White Rose capture project.

C **Permanently mixed**, southern North Sea, with a water depth of 28 m. Nominally storage centred at 53.00°N 2.50°E, an approximation of the Hewett and southern Bunter Closure potential storage sites.

Reflecting general operating procedures for automated underwater vehicles, (which are the likely option for monitoring operations, Blackford et al., 2015), and the dynamics of CO_2 plumes (Dewar et al., 2013), model data has been taken from 5 m above the sea floor. For the purposes of baseline characterisation each locality is defined as a region spanning approximately 84 × 84 km (7 × 7 model grid points) centred on the location above. Although this area is larger than a typical storage complex, it is chosen to fully encompass regional variability reflecting the model's spatial skill (Saux-Picart et al., 2012). In addition to analysing daily mean pH over the full 30 year simulation, two contrasting years, 1995, 2002 have been subject to more detailed analysis based on high frequency (20 min interval) pH predictions. Founded on the natural variability of the system, a set of detection criteria, firstly utilising absolute pH values and secondly rates of change of pH over various temporal intervals are proposed. Finally a comprehensive set of pseudo leak perturbations are imposed on the time series and detection success evaluated based on quantifying both true positive and false positive signals.

3. Natural variability of pH across the NW European Shelf

3.1. Overview of spatial pH variability

The model analysis of North Sea seafloor pH (Fig. 1) shows distinct spatial heterogeneity. Broadly the region can be described by three domains, riverine influenced, mixed and seasonally stratified. Beyond this basic characterisation which is largely depth dependent (Fig. 1-I, ii), finer scale variability is a feature of the region, driven by topography and circulation patterns (Figs. 1 and 2). In brief:

- **Coastal–riverine regions** associated with continental Europe and major UK estuaries are typified by a relatively high mean pH (~8.2, Fig. 1-iii) and a large annual pH range (0.5–1.0 pH units, Fig. 1-iv). There are two major drivers of this carbonate chemistry, directly via variable riverine inputs of DIC and alkalinity and indirectly via riverine nutrients driving primary production and the uptake and subsequent respiration of CO_2 within the food web. In these very shallow and relatively turbulent systems the water column is homogenous and light levels sometimes sufficient to permit primary production throughout the water column.
- **Mixed regions**, offshore and especially in the southern North Sea are typified by a pH mean around 8.1 (Fig. 1-iii) and a range typically of 0.3 pH units over the annual cycle (Fig. 1-iv). These relatively shallow systems are hydrodynamically mixed with high biological activity, such that sea floor carbonate chemistry is directly affected by autotrophic processes occurring nearer the surface with respiration occurring throughout the water column.
- **Stratified regions**, away from coastal influences and in deeper water have lower mean pH (~8.0, Fig. 1-iii) and annual ranges of 0.1–0.2 pH units (Fig. 1-iv). In these systems thermal stratification during the summer isolates the sea floor from autotrophic processes, partially trapping respired CO_2.

Across the North Sea, annual pH minima follow the trend in mean pH. Minima tend to occur during Nov–Dec in stratified systems and during Jan–Feb in mixed/coastal systems.

South-North transects of near-bed pH, illustrated for model year 2002 in Fig. 2i, further indicate the heterogeneity of regional carbonate chemistry and Fig. 2ii, iii and iv illustrate the primary drivers of this heterogeneity. There is a particularly strong positive correlation between pH and temperature, however this is not a direct effect (the direct impact of temperature on pH, driven by modification of the ionic speciation of the carbonate system is negative), but due to a number of related processes, primarily mediated via the temperature effect on stratification. Within biologically active marine systems, increased seafloor temperature increases biological activity and generally signifies a lack of stratification. Moving south to north, large seasonal ranges and elevated summer pH are associated with biologically active, shallow, well-mixed, terrestrially-driven coastal regions in the south.The transition to reduced seasonality occurs at approximately 450 km along the transect where the system deepens becoming seasonally stratified and the seafloor consistently heterotrophic. At 600 km the transect crosses the Dogger Bank, a shallower region which reduces stratification and the system briefly reverts to a seasonally autotrophic mixed type. From ~700 km the system reverts to seasonal stratification with reduced pH and variability. Between 1050 and 1250 km the transect crosses a region generally influenced by the Fair Isle current, a shelf sea derived water mass which acts to promote mixing. Beyond 1250 km the system becomes influenced by Atlantic inflow and is seasonally stratified; variability in this region is generally associated with large scale circulation features and sea floor topography.

3.2. Decadal, interannual and seasonal variability of pH

Temporal variability in pH near the sea floor is complex and distinct between the three exemplar regions (Fig. 3, Table 1). In the southern mixed region the dominant signal is the seasonality (Fig. 3c; primary production dominating in the spring – summer, respiration dominating during autumn-winter). This follows from the connectivity between the sunlit surface and the near sea floor found in permanently mixed systems. Because the system is well mixed there is relatively less variability within the region (Fig. 3c, shaded region, Table 1 last column). By contrast in the seasonally stratified region (Fig. 3a), the sea floor becomes periodically disconnected from the productive surface waters and is far more dominated by respiration. Without the signal of primary production near the sea floor, the seasonal range is reduced. However variability within the region is comparatively large, due to the presence of two semi-distinct season regimes. At the intermittently stratified region the seasonal pH range is larger as the near bed waters are influenced by primary production while mixed and respiration during stratified periods. Inter-annual variability is a feature of all regions (Fig. 3) but has a relatively larger influence on pH in the stratified system, compared with the weaker seasonal signal. Inter-annual variability is hard to predict, being a product of several linked, nonlinear processes such as the Atlantic Multi-decadal Oscillation, the North Atlantic Oscillation, regional currents, riverine inputs and local weather patterns (e.g. Nye et al., 2014; Harris et al., 2014; Skogen et al., 2007; Gypens et al., 2011).

All regions show a long-term general decline in pH (Fig. 3), of the order of 0.001–0.002 pH units yr^{-1} (consistent with that reported in

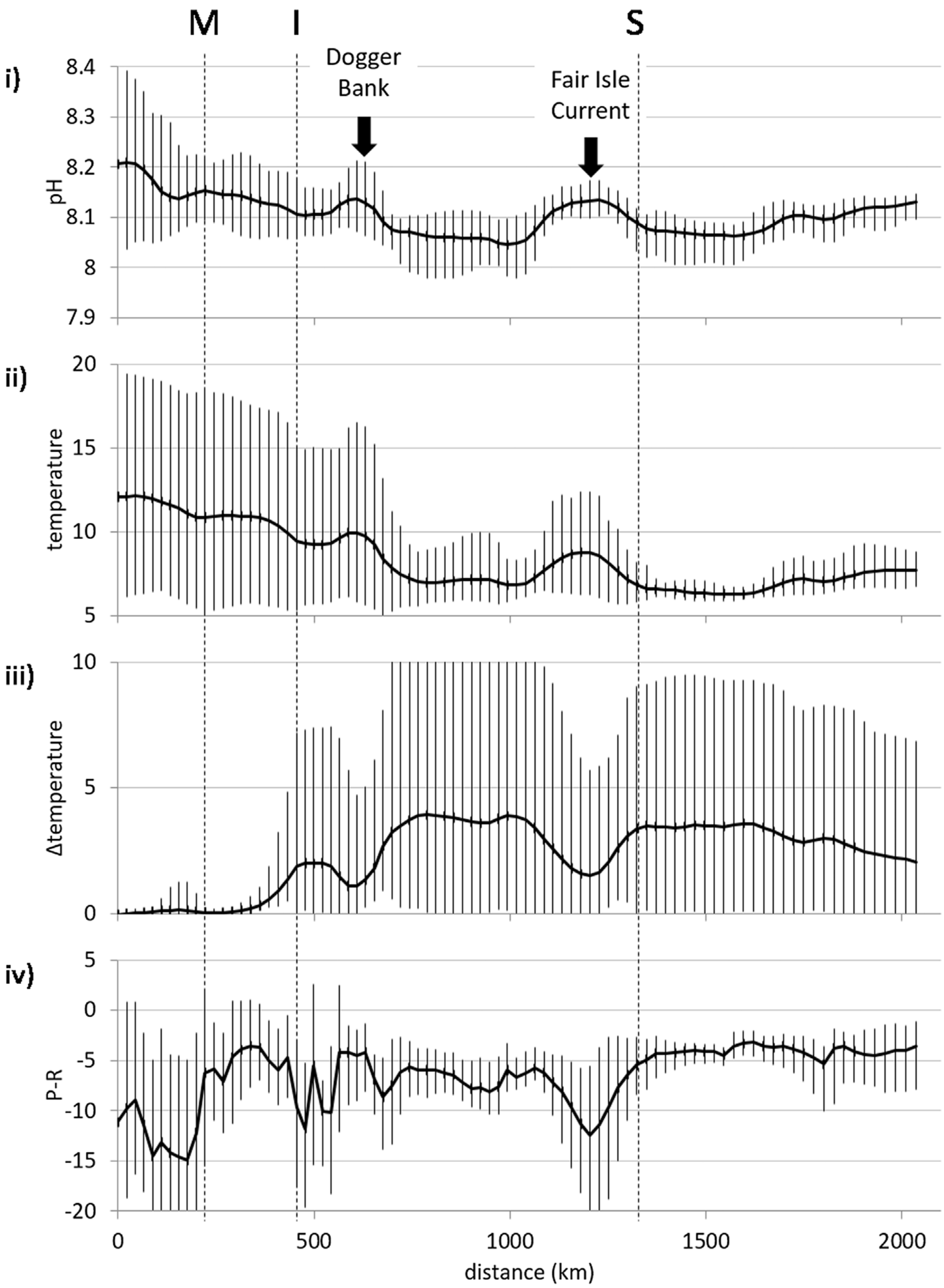

Fig. 2. Spatial variability of pH and key drivers along the South-North transect illustrated in Fig. 1i, showing the annual mean and seasonal spread for simulation year 2002. (Patterns for other years are qualitatively similar). The study regions, mixed (M), intermittent (I), Stratified (S), are marked by vertical lines, the x-axis denotes distance from the southern point of the transect. From top to bottom the figures show i) pH, ii) temperature, iii) the temperature difference between sea floor and sea surface and iv) net biological CO_2 production (primary production minus consumption). i, ii and iv are integrated over the bottom five metres, ii & iii in degrees Celsius and iv in mmol C m^{-3} d^{-1}.

Bates et al., 2014). This is driven by increasing atmospheric concentrations of carbon dioxide, a phenomena known as ocean acidification. There is some variability in the slope of the decline between regions and decades due to interplay with biological processes (e.g. Clargo et al., 2015), which are themselves subject to direct and indirect pressures from climate change, ocean acidification, changing nutrient loads and other drivers.

3.3. Rates of change of pH

As well as examining absolute values of pH, it is also instructive to consider rates of change in either time or space. The relatively coarse horizontal resolution of the model precludes a detailed spatial analysis. Across the domain, the mean difference in daily averaged pH between neighbouring sea floor model grid cells is of the order of 0.007–0.008 pH units, with a maximum of the order of 0.09 units. The larger spatial discontinuities are driven by relatively steep changes in topography (e.g. the Dogger Bank) which influence stratification, boundaries between current structures (e.g. the Fair Isle Current edge), proximity to riverine influence and boundaries between regions of different biological productivity (Fig. 2). Apart from topography such spatial discontinuities show significant temporal variability. Considering temporal dynamics, Fig. 4 illustrates the frequency with which the change in daily mean pH for each grid cell exceeds thresholds of 0.1, 0.01 and 0.001 pH units between successive days. The model suggests that day to day changes in pH exceeding 0.1 pH units are rare (occurring less than once a year on average, Fig. 4-i) and confined to the German Bight and some coastal regions. A day to day pH change of 0.001 units occurs on the majority of days across most of the region (Fig. 4-iii), while a change of 0.01 units occurs between ~1 and 100 days per year depending on location (Fig. 4-ii).

Significant short term variability is driven by diurnal processes, in particular the day/night light-driven biological productivity and the physical tidal cycle. An analysis of within day pH range for two

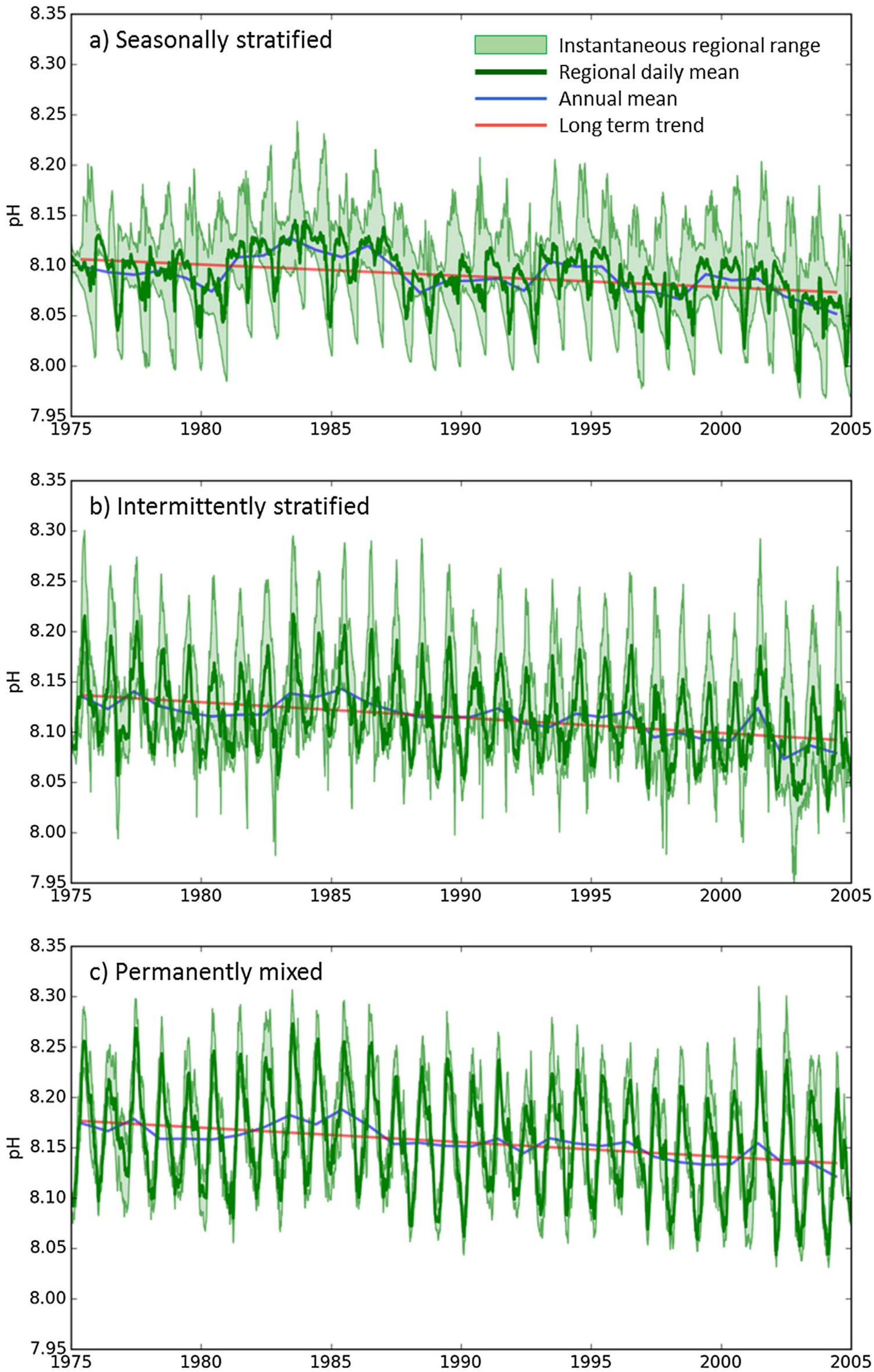

Fig. 3. Thirty year time series of modelled pH values at 5 m above the sea floor for each region. The shaded area represents the instantaneous range of data within each analysis box, the thick line represents the daily mean pH, the thinner variable line represents annual means while the straight line illustrates the linear trend over the simulation period.

contrasting years, utilising model outputs at 20 min intervals (Fig. 5) reveals that short-term variability also has distinct spatial, seasonal and inter-annual heterogeneity. In all of the study regions, the within day pH range can exceed 0.025 units. In the stratified region, the highest diurnal ranges occur between July and January, at the intermittent region the largest diurnal ranges occur during summer and autumn, while in the mixed region pH is dynamic year round. The largest instances of short term variability are associated with episodic events, such as the advection of areas of biological production across a region, the onset or cession of phytoplankton blooms or the onset or breakdown of physical mixing. The timing and drivers of these events are complex and variable and often influenced by weather scale phenomena. As a consequence, there is distinct inter-annual variability in all regions.

4. Development of criteria to signify anomalies

4.1. Absolute thresholds as detection criteria

Given that a release of CO_2 would decrease local pH, quantifying the natural pH minima provides an obvious and simple criterion by which

Table 1
pH metrics for each region.

Region	Mean depth (m)	Mean pH	Trend y^{-1}	Mean annual pH range	STD entire data set	Mean of daily STD
Seasonally stratified	120.4	8.090	−0.0011	0.187	0.038	0.025
Intermittently stratified	37.5	8.114	−0.0015	0.237	0.043	0.021
Permanently mixed	27.5	8.155	−0.0014	0.207	0.049	0.015

to identify an anomaly, or to indicate the potential for ecological impact. Fig. 6 presents, as a climatology (mean over all years, de-trended with respect to OA), the range of pH metrics within each region, thereby reflecting variability within each forty-nine grid cell region and between years. Each locality displays a unique pattern, in the stratified region a slow decline in pH minima occurs from spring to mid-December, increasing sharply over winter. The intermediate region shows a sharp decline over late summer to early autumn with a slow increase during spring-summer. The mixed region shows a significant increase during the first half of the year, followed by a similarly paced decline over the second half of the year. As a result, thresholds that could be taken as strongly indicative of an anomaly vary between

- pH 7.99 (Dec–Jan) and pH8.06 (March) at the stratified region,
- pH 7.99 (Oct) and pH 8.10 (Jul) at the intermittent region and
- pH 8.06 (Feb) and pH 8.20 (Jun) mixed region.

Given that detection may depend on discriminating small signals at some distance from an unknown leak location, the potential for anomalous perturbations that do not move the observed pH below the climatological minima suggests a fundamental weakness of using absolute thresholds as detection rather than impact criteria. It might be expected that while false positives (natural variability miss-interpreted as anomalies) would be minimised, false negatives (missed anomalies) could be overlooked. Some increase in detection rate could theoretically be achieved by using the climatological mean minima, rather than the absolute minima, or some percentile between the two, as illustrated in Fig. 6 and tested below.

4.2. Dynamic criteria for identifying anomalies

Deployed instrumentation has the potential to sample at intervals of the order of 1 min. Fig. 7 illustrates the relative frequency of various pH changes over a range of short term intervals. Consistent with the analysis presented in Fig. 5, day to day changes of 0.001 pH unit are very common at all regions. However when the sampling interval is reduced to the model time step of 20 min, natural changes of 0.001 pH become significantly rarer occurring in < 5% of instances. Natural changes of 0.005 pH units do not occur over intervals of less than ~40 min, while changes of 0.01 pH units do not feature if the sampling interval is less than 2 h. Fluctuations of 0.025 pH units are unlikely to have a natural origin if sampling is more frequent than daily. In general and unsurprisingly the magnitude of pH fluctuations increases with sampling interval, however a discontinuity is apparent, such that natural fluctuations over a twelve hour period are less than those that occur over a six hour period. This is a physical effect whereby the tidal oscillation has performed a complete cycle, bringing a given water mass close to its initial position, thereby minimising advective effects compared with sampling intervals that are a-synchronous with the tidal cycle. Consequently it is hypothesised that high frequency observations may be optimal in terms of discriminating small anomalies as well as minimising false positives due to natural variability.

4.3. Detection criteria evaluation

The efficacy of a selection of detection criteria outlined above has been quantified by assessing the true positive and false positive detection rates for a range of imposed perturbations on modelled pH as illustrated by Receiver Operating Characteristic (ROC) plots in Fig. 8. The perturbations tested excluded very large changes in pH that would be trivial to detect, concentrating on a range of detection challenges spanning from pH changes of the order of instrument sensitivity that may be associated with small leaks or the periphery of CO_2 plumes to pH changes that are significant but within annual variability. In all, eight levels of perturbation ranging from −0.00103 to −0.12720 pH units (Table 2i) are applied. In setting up the detection tests there is, for each criterion, one test per time step to assess if natural variability triggers that criterion (a false positive) and eight tests per time step to

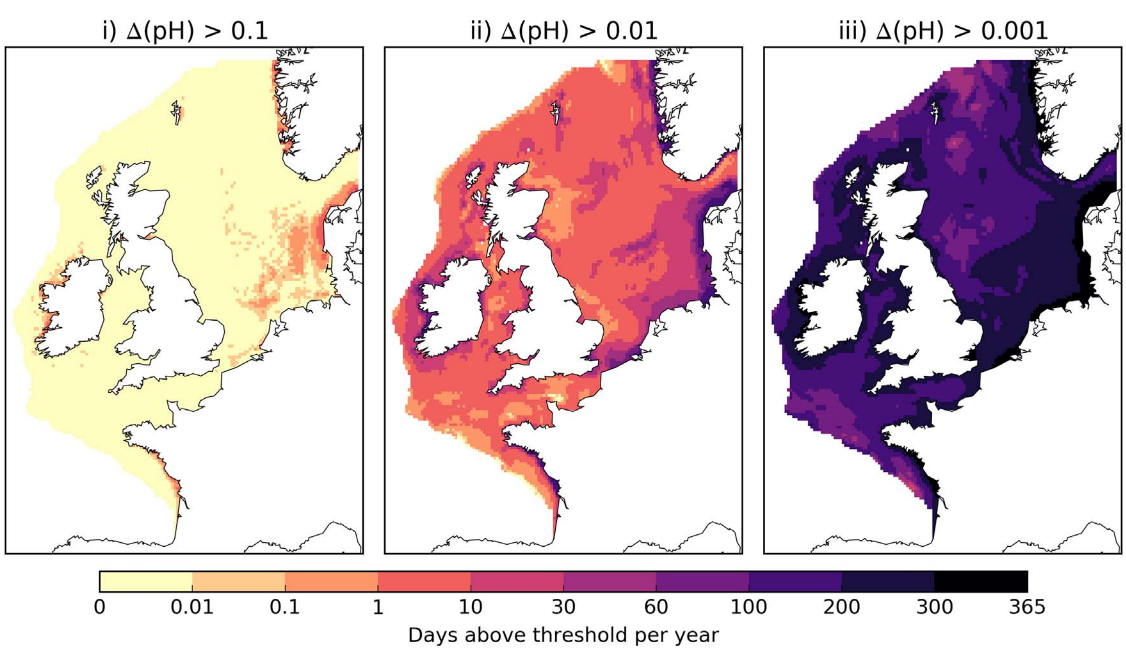

Fig. 4. The frequency, in terms of days per year, of day to day changes in pH exceeding given thresholds, i) 0.1 pH units, ii) 0.01 pH units, iii) 0.001 pH units.

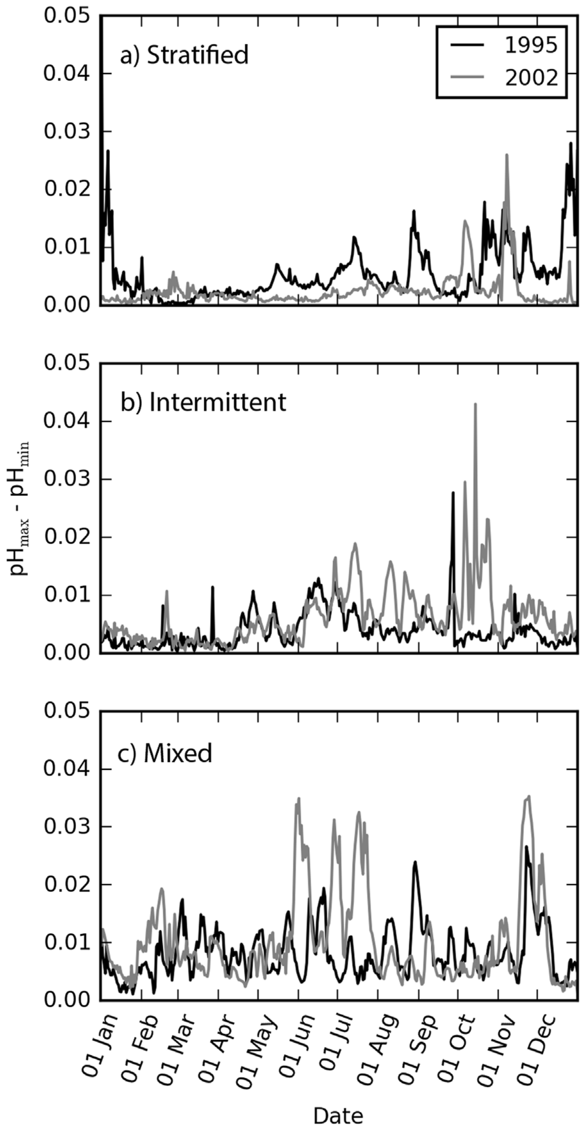

Fig. 5. Daily pH ranges from two contrasting years from each study region, a) seasonally stratified, b) intermittently stratified, and c) permanently mixed. The ranges are calculated as maximum minus minimum pH within each calendar day, irrespective of when these occur.

assess if the criterion is triggered by the perturbation. The two classes of detection criteria tested are outlined in the following sections.

4.3.1. Absolute thresholds

The absolute thresholds tested consisted of seasonally varying pH minima for each region as derived in Section 4.1 and included the:

- daily minima of the forty-nine grid cells
- the percentile demarking ninety percent of modelled minima
- percentile demarking eighty percent of modelled minima

…as derived from the de-trended seasonal climatology (Fig. 6). The annual cycle of daily mean pH for each of the forty-nine grid cells centered on each site, for the years 1995 and 2002 were chosen as baseline data. A false positive was scored every time the daily mean pH dipped below any of the threshold criteria for the given day, giving a maximum possible 17,885 (49*365) false positive chances at each region for each year. The perturbations detailed in Table 2i were applied individually to each day, for each of the forty-nine annual cycles within each region for each year, giving 143,080 (49*365*8) potential true positives for each region and year. A true positive was scored each time the perturbed pH dipped below one of the given thresholds.

The resulting true positive/false positive rates are plotted in Fig. 8a. Primarily there is a reasonably linear trade off such that increased detection success comes at the expense of a concomitant increase in the false positive rate. The skill of this set of criteria is uniformly low. Nevertheless there are some interesting nuances within the results. The distinct difference between 1995 and 2002 results illustrates the challenge that inter-annual variability poses to criteria based on climatological analysis. Clearly and unsurprisingly using the minimum criteria reduces both true and false positive rates, compared to the 90% and 80% criteria. The stratified region also scores lower true and false positives than the other regions. This stems from the higher variability around the mean seen in this region (Table 1) implying that the minimum criteria are more distinct from the mean regional behavior than at the other regions. It is hypothesised that absolute thresholds may have better utility as indicators of impact potential than as detection criteria.

4.3.2. Dynamic thresholds

Given that a higher observational frequency generally reduces the difference in pH between successive observations (Section 4.2), defining criteria as an abnormal pH change over a given time interval has potential to offer an alternative approach to anomaly detection. In this section, the relative success of a set of dynamic criteria (a decrease of 0.001, 0.003, 0.01, 0.031, 0.1 pH units, Table 2ii; coupled with a set of sampling intervals, 20, 40 min, 1, 6, 12, 24 h, Table 2iii, giving 30 different criteria) is evaluated. Thus the question is, which combination of pH change and sampling interval can maximise detection of perturbations whilst restricting false positives to an acceptable minimum? Using the modelled annual cycle of pH, resolved to the model time step of 20 min, as a baseline, each of the eight perturbations previously used (Table 2i) were applied at each time step, repeated for each of the three sites and both years. This results in a total of 210240 (8 × 72 × 365) potentially detectable events per simulation year. Each individual perturbation was compared to the non-perturbed pH value at selected sample intervals over the previous 24 h (Table 2iii) resulting in a ΔpH value. Each ΔpH perturbation was then compared to the set of dynamic detection criteria (Table 2ii). When a resulting ΔpH exceeded a given criterion (i.e. more negative) a true positive was scored for that combination of time, perturbation, sampling interval and criterion. The analysis was repeated with unperturbed pH data such that when the natural change in pH between any given sampling interval exceeded a particular detection criterion a false positive was recorded for that combination of time, sampling interval and detection criterion. The precise results of these detection assessments are somewhat dependent on the relative values chosen for the set of criteria and the set of perturbations. The detection criteria were chosen such that the most sensitive (−0.001 pH) was less than the smallest imposed anomaly (−0.00103 pH), thereby ensuring a theoretical possibility of 100% detection, (in the absence of natural variability). The least sensitive detection criterion of Δ−0.1 pH is theoretically only capable of discriminating the largest of the eight perturbations (−0.12720 pH) giving a maximum theoretical success rate of 0.125 (Table 2ii).

Fig. 8ii shows the relative success of each combination of detection criterion and sampling interval, amalgamating results for the set of perturbations, regions, time and year. Each detection criterion performs well, approximating to its maximum theoretical success rate (Table 2ii), as based on the distribution of perturbations relative to each criterion.

Self-evidently only the smallest criterion, Δ−0.001 pH, can achieve near 100% successful detection, in practice scoring between 88 and 95% depending on sampling interval. The underperformance occurs when pH is naturally increasing over the sampling interval, offsetting the effect of an imposed pH decrease such that the resulting change falls below the criterion threshold.

The other criteria show variable responses of the true positive rate

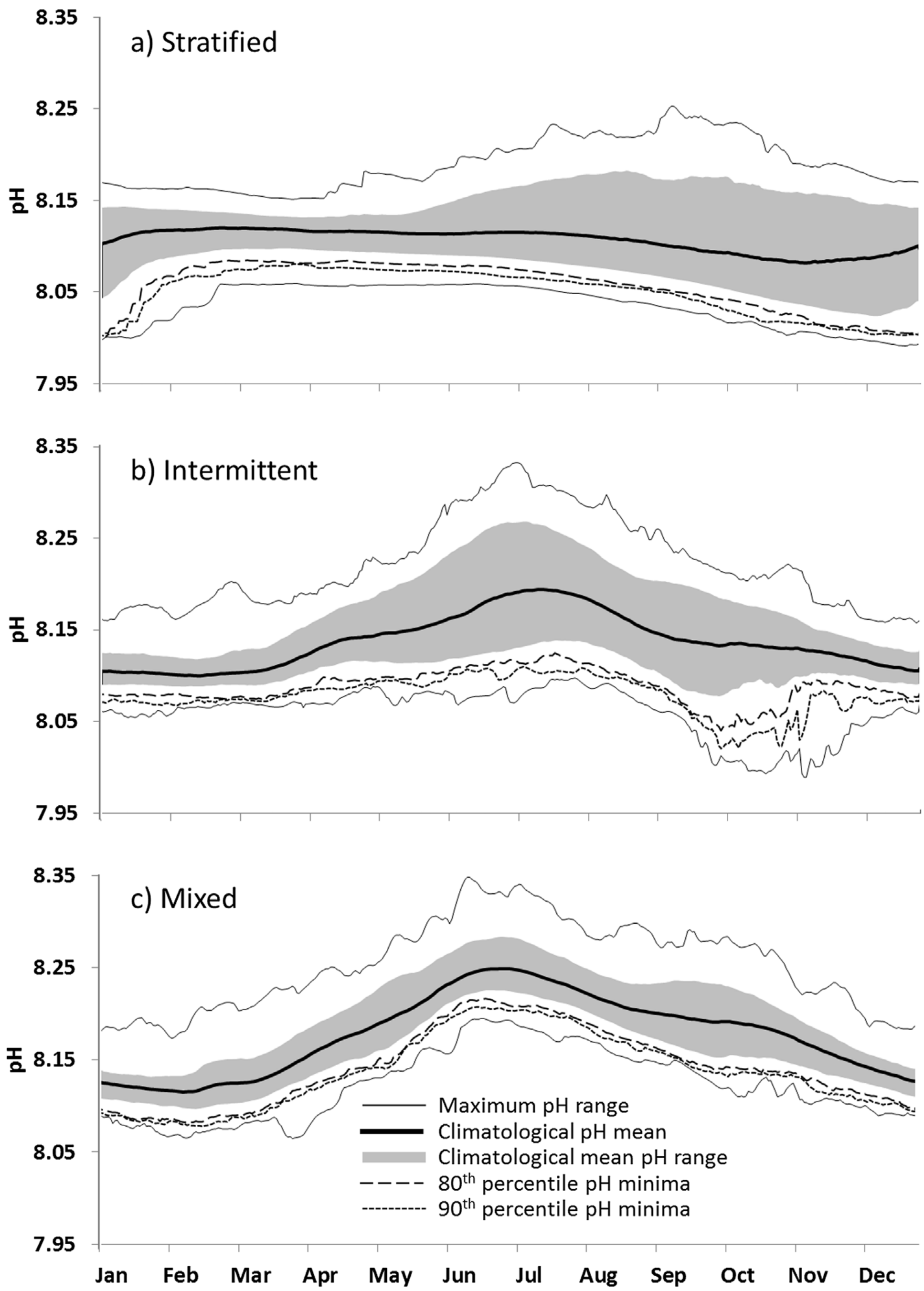

Fig. 6. Seasonal climatology of pH using de-trended data from the 30 year simulation. Respectively a) stratified, b) intermittent, c) mixed. The shaded area indicates the range of pH between the mean maximum and mean minimum of the 30 year data set, the bounding lines represent the maximum and minimum pH recorded and the dashed lines the 80th and 90th percentile of seasonal minima. The bold central line indicates the seasonal mean pH.

to sampling interval − this however is an artifact of the relative distributions of the perturbation set and the criteria set. A natural underlying decrease in pH between sampling points in combination with a perturbation would serve to increase the magnitude of the observed ΔpH. Consequently on occasion a perturbation can be pushed over a larger (more negative) detection criteria, hence a longer sampling interval can propel detection success over the theoretical maximum. This is only observed where the chosen perturbations are slightly less than the chosen detection criteria (as for −0.003 pH and −0.01 pH but not for −0.031 pH). While the false positive rates are relatively small for the larger magnitude detection criteria, increasing the sampling interval significantly increases the false positive rate for the smaller detection criteria, such that one quarter to one third of observations with a detection criterion of Δ pH = −0.001 and a sampling interval of more than 1 h would register false positives. It can be seen that sampling at 6 h intervals produces a worse result than sampling at 12 h intervals. This follows from the tidal mixing cycle, such that water masses are likely to have a more similar distribution if sampled in synchronization with the tidal cycle.

Given that short term variability is seasonally and spatially heterogeneous (Fig. 5), a pertinent question is if monitoring success (defined as maximising true positives and minimising false positives) can be improved by choosing site specific time periods where natural variability or underlying pH dynamics facilitate detection. Further by identifying optimal monitoring time-windows, is it possible to reduce sampling intervals to save costs? Subjectively defining monitoring success as better than 95% of perturbations detected with less than one false detection per day, Fig. 9 defines the periods and associated sampling strategies that deliver acceptable detection success for each site. Inconsistency between the two chosen years suggests that due to interannual variability there may not be a consistently optimal time-window for sampling at any site. The stratified region has the best

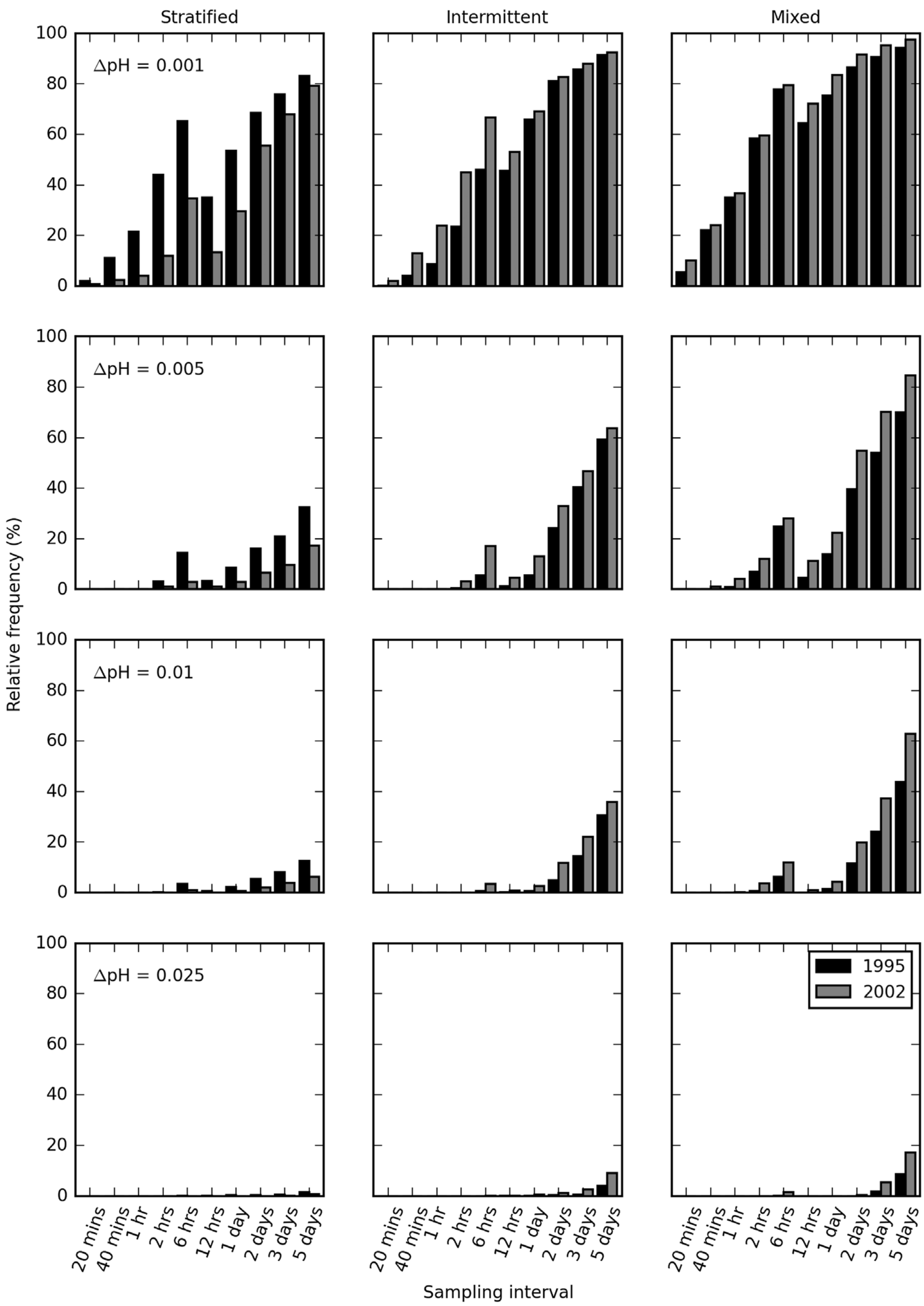

Fig. 7. Relative frequency of changes in pH exceeding a variety of thresholds (top to bottom: 0.001, 0.005, 0.01, 0.025 pH units) between different time intervals (x-axis, 20 min to 5 days) for two contrasting years for each region (left to right: Stratified, Intermittent, Mixed).

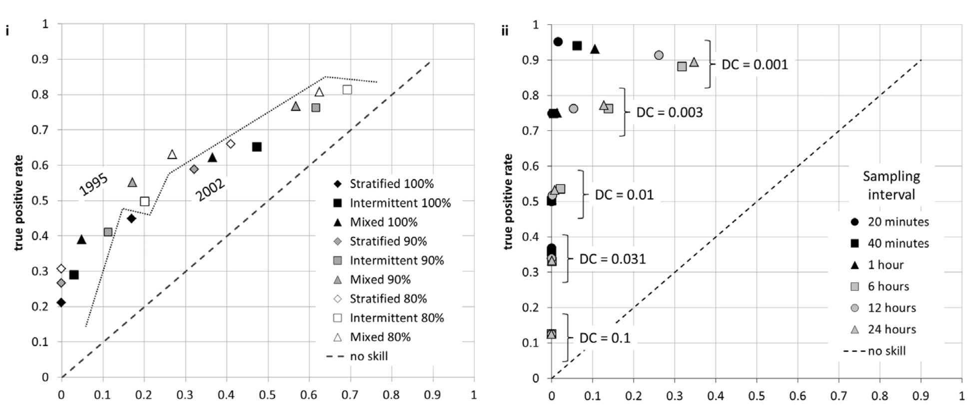

Fig. 8. Receiver operating characteristic plots showing the true positive and false positive scores for the detection criteria tested. i) Absolute thresholds, the three regions represented by different symbol shapes, the three threshold criteria represented by different shading. 2002 results are encircled, 1995 results are not. ii) Dynamic thresholds, showing each combination of detection criterion as labelled and sampling interval delineated by shape and shade.

Table 2
i) pH perturbations used to test detection criteria. ii) Dynamic detection criteria tested with (in brackets) the theoretical maximum detection success rate based on the perturbation data set. iii) The subset of sample intervals analysed for dynamic detection criteria.

i Perturbation set: Imposed decrease in pH	ii Detection criteria set: ΔpH (maximum detection success rate)	iii Sample intervals presented
−0.00103	−0.00100 (100%)	20 min
−0.00206	−0.00316 (75%)	40 min
−0.00409	−0.01000 (50%)	1 h
−0.00814	−0.03162 (37.3%)	6 h
−0.01618	−0.10000 (12.5%)	12 h
−0.03217		24 h
−0.06397		
−0.12720		

spread of optimal monitoring periods, with high frequency sampling delivering potentially acceptable results over much of the annual cycle and lower frequency sampling (hourly) delivering acceptable results during some seasons (Fig. 9a). However for the intermittent and mixed regions (Fig. 9b, c) the seasonal windows for successful monitoring are significantly reduced and the requirement for high frequency monitoring significantly increased. Improved detection is uniformly associated with low background variability in pH (Fig. 5).

5. Implications for operational CCS

The premise remains that well operated and appropriately monitored CCS will be secure in the very long term. Nonetheless a comprehensive and scientifically sound monitoring programme will contribute significantly to public confidence and acceptance of CCS as a bridge to a low carbon global economy. The balance of operational versus regulatory feasibility requires that costs for such monitoring need to be minimised without impacting detection fidelity. For this reason the focus here is on anomaly detection only. If an anomaly is detected, additional monitoring will be required for confirmation and if necessary quantification, in which case additional samples for DIC and alkalinity may be required. This work shows that in the absence of comprehensive baseline information on the natural variability of parameters relating to carbon leakage, models can provide sufficient resource with which to explore various strategies for monitoring, by providing biogeochemically consistent, four dimensional data sets. However caution is advisable, because of the models low horizontal resolution compared with the scale of leakage events and relatively low frequency model forcing (e.g. surface wind speed, heat flux and cloud cover) it is more likely that the model underestimates heterogeneity and temporal variability, therefore overestimating the potential success of monitoring. While the sea floor will be relatively immune to vagaries in surface forcing, we lack sufficient high frequency bottom water data with which to properly evaluate the model findings.

A clear outcome of this analysis is the dissimilarity between different regions, characterised by different hydrodynamic and to a lesser extent biological regimes. Results suggest that at a deeper, seasonally stratified site similar to Goldeneye, anomaly detection may require less deployments and less data; less false positives will occur and smaller signals could be discriminated compared with a shallower fully mixed site similar to the southern North Sea. Consequently monitoring at some sites will inevitably be less expensive than at others. It is a strong recommendation that a bespoke baseline survey and analysis of natural variability should be obtained for each target storage site in order to maximise the efficiency of monitoring, and given that observations are more expensive than models, that a combination of modelling evaluated by in situ data has promise in identifying the most efficient approach to monitoring for any particular region. This study also shows that there is distinct seasonality in the natural baseline producing variation in the optimal criteria and sampling strategies depending on

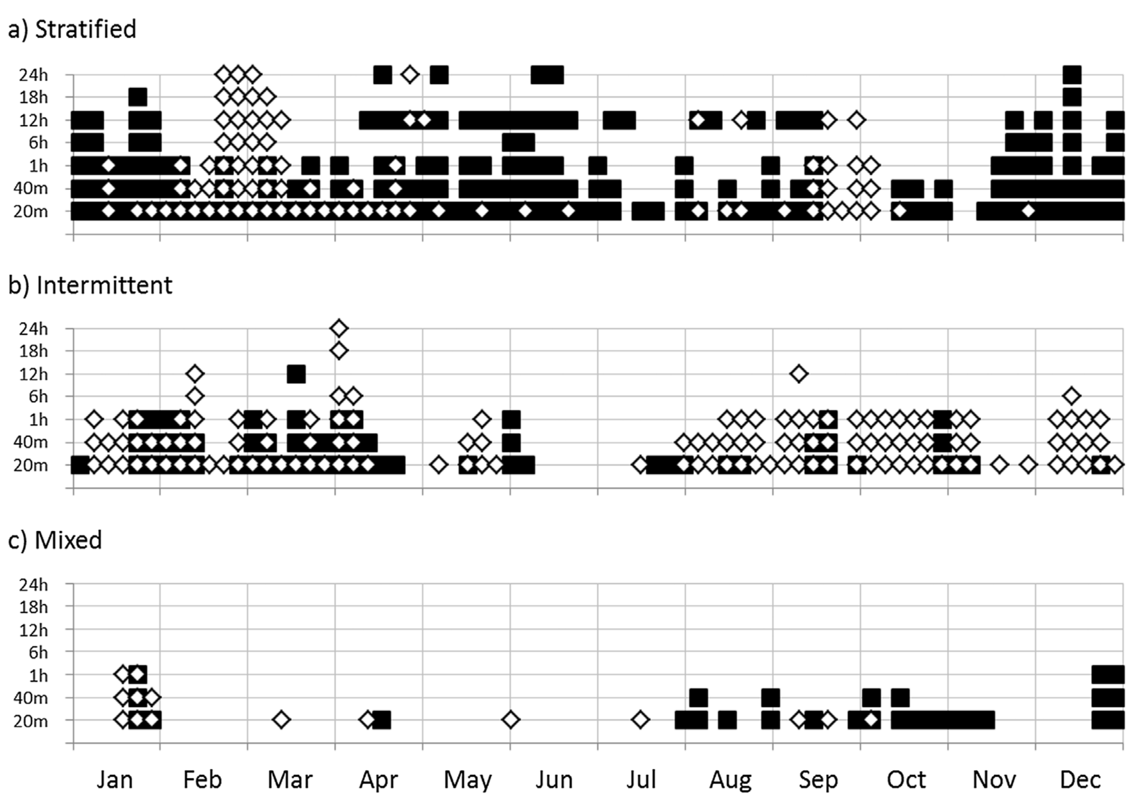

Fig. 9. Seasonal analysis of detection success. Each symbol represents a time period when detection is deemed successful, according to the criteria of exceeding 95% detection of perturbations and triggering no more than 1 false positive per day. Data has been binned into five day periods for clarity and is analysed for each of the subset of sampling intervals (y-axis). White diamonds represent 1995 and black squares represent 2002. a) stratified region, b) intermittent region, c) mixed region. This analysis utilises the −0.001 pH unit detection criteria.

the time of year. Although it is hard to predict precise optimal sampling periods based on this study, avoiding highly dynamic periods such as the spring bloom or autumnal overturning based on short term operational modelling would reduce the chance of false positives. Opportunistic monitoring deployments based on short-term forecasts of low dynamics may be a strategy worth considering. Of course, intermittent monitoring does not provide the assurance of continuous monitoring and could theoretically allow release to occur for several months before detection. However if continuous monitoring was prohibitively expensive, especially for the entirety of large storage complex areas, it would be important to optimise an intermittent monitoring strategy. Full scale geological monitoring of a storage complex is expensive and likely to be deployed at annual frequencies or less (e.g. Chadwick, 2010).

This study shows that identifying parameter ranges and extremes (in this case pH minima), even with sub-seasonal discrimination, does not provide an operationally useful criteria for anomaly detection, although they have much relevance to impact assessment. While the absolute criteria of pH minima provide a relatively simple metric, it is quite possible that given the large range of natural pH, a leak event may not cause a change in pH sufficient to be distinct from this range, especially in the case of a small leak, or at the periphery of larger events. There is little point in conducting an expensive biological impact assessment if the carbonate system has not been perturbed outside of its normal variability for a significant amount of time (Lessin et al., 2016). Nuancing absolute thresholds by considering the distribution of minima within each region, for example adopting the mean of the minima, or a percentile, in place of the absolute minima improves detectability slightly but at the expense of unacceptable levels of false positives.

This analysis does suggest that monitoring for an unusual change in pH over relatively short intervals may deliver a highly sensitive method by which to detect anomalies. This approach benefits from the mobility associated with both target and sensor platform. Within tidally influenced environments, modelling of the dispersion of CO_2 plumes suggests that the plumes will be highly mobile, following the local tidal ellipse (Blackford et al., 2013; Phelps et al., 2015). Monitoring deployments are likely to utilise either fixed landers near particular risk points or autonomous underwater vehicles (AUVs) for larger areas (Blackford et al., 2015). Whether via tidal circulation of CO_2 plumes, AUV transit or a combination of both it is likely that deployed sensors would experience frequent oscillations between the natural and perturbed state as plumes were advected over sensors or AUVs traversed plumes. Based on the minimum time-step of the model system, twenty minute sampling frequencies are shown to give reasonable detection of small perturbations, however operational sampling frequencies can be higher, which could improve detection certainty or offset model underestimations of natural variability, if shown to exist. In practice sampling frequency is limited by the ability to store and transfer data as well as the battery life of the sensor and platform. If for operational reasons only low frequency sampling is possible, then from a baseline consideration, sampling at semi-diurnal frequencies may minimise false positives, however from a detection point of view it may also minimise the chance of registering oscillations between the natural and perturbed state. Previous studies (Greenwood et al., 2015; Hvidevold et al., 2015) have shown that appropriate choice of location of limited fixed sensors is not immediately intuitive, with respect to potential leak location. This study suggests that heterogeneous natural variability may be another factor to consider when planning the location of such sensors. Similarly when planning the path of AUVs, traversing across, rather than with the tidal ellipse should increase the opportunity to detect anomalous oscillations.

This analysis also suggests that false positives may be very difficult to eradicate entirely, unless a relatively high anomaly threshold is deemed acceptable by regulatory authorities. One of the yet to be resolved challenges is to optimise the balance between detection sensitivity, the potential for false negatives (at worst rare but potentially very expensive) and the occurrence of false positives (more common and moderately expensive). There is also a challenge to understand how to respond to a detected anomaly and at what stage to trigger a costly decision to deploy additional sensors or sampling. A single small anomaly may only necessitate a more in depth assessment of the monitoring data and ambient conditions, rather than an automatic additional deployment. However an ability to divert AUV based systems to resample suspect regions would be a worthwhile capability.

The primary recommendation of this study is for a programme of baseline observations comprising of high frequency near sea floor measurements of the pH (along with temperature, salinity, pCO_2) with both a seasonal and spatial component. Lower frequency supporting determinations of a basic biogeochemical parameter set, such as oxygen, chlorophyll and nutrients would enable evaluation of the drivers of pH change and improved model evaluation. It would be valuable to assess if significantly different results arise within one metre of the sea floor (i.e. the likely height of a fixed sensor) compared to around five metres (corresponding with the operational safe height of AUVs, Wynn et al., 2014). Deployment of AUV based baseline surveys would allow for an analysis of local spatial heterogeneity, which is not captured by the model. Thus, while this study has focussed on analysis of single time series, spatial correlation between sensors on mobile and fixed platforms, operating as part of an array, will also be important for improving sensitivity of detection (as shown by Greenwood et al., 2015). Surveys at hydrodynamically contrasting sites would be valuable as the model predicts fundamentally different dynamics according to water depth and stratification. Seasonality is also an important dimension. Targeting surveys at periods when low natural variability is predicted may be more efficient than attempting a full seasonal discrimination.

If criteria are based on historical observations then the potential for inter-annual variability and long term drift must also be accounted for. Apart from the signal induced by atmospheric CO_2 increase, of the order of 0.001 pH units per year, other factors such as changes in buffering capacity (Thomas et al., 2007) and riverine run off due to terrestrial nutrient management (Provoost et al., 2010; Gypens et al., 2009) can induce regional changes that are significantly more rapid. Hence, over decadal scales, monitoring criteria will need to be adjusted to account for decadal scale change. Consequently combining data from short-term, high resolution, CCS targeted surveys with longer, term multi-purpose, marine research data and climate-targeted initiatives (GOA-ON, Newton et al., 2015) via models is most likely to be cost effective.

Although this study is predicated on the North Sea, the majority of offshore geological carbon storage will be situated under the continental shelf below water bodies with a similar range of hydrodynamic characteristics to the sites studied here. On a qualitative basis it is likely that the principles outlined here would apply to other storage sites, although site specific models and observations would be needed to fine tune local monitoring strategies.

In conclusion, this study suggests that detection of anomalies based on pH fluctuations as small as -0.001 pH units may be achievable, although model under-estimation of natural variability may imply that fluctuations of, say, -0.01 pH units are a more reliable indicator. This paper demonstrates that it is possible to utilise the current accuracy of available sensors (of the order of 0.01 pH units) to detect small and locally environmentally harmless anomalies that could indicate unforeseen release of CO_2. Further, future improvements in sensor accuracy could enhance detection fidelity. Of course detecting a small anomaly does not begin to quantify a release rate or impact – such a signal could indicate a small release close by or a large release some kilometres distant. Further survey work would be required to locate, attribute, quantify and assess a release.

Acknowledgements

We gratefully acknowledge the funding provided by NERC for the QICS project under grant number NE/H013962/1. In addition this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654462.

References

Allen, J.I., Holt, J.T., Blackford, J., Proctor, R., 2007. Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: part 2. Chlorophyll-*a*, nutrients and SPM. J. Mar. Syst. 68, 381–404.
Artioli, Y., Blackford, J.C., Butenschön, M., Holt, J.T., Wakelin, S.L., Allen, J.I., 2012. The carbonate system of the NW European shelf: sensitivity and model validation. J. Mar. Syst. 102–104 1–13.
Artioli, Y., Blackford, J.C., Nondal, G., Bellerby, R.G.J., Wakelin, S.L., Holt, J.T., Butenschon, M., Allen, J.I., 2014. Heterogeneity of impacts of high CO2 on the north western European shelf. Biogeosciences 11, 601–612.
Atamanchuk, D., Tengberg, A., Aleynik, D., Fietzek, P., Shitashima, K., Lichtschlag, A., Hall, P.O.J., Stahl, H., 2015. Detection of CO_2 leakage from a simulated sub-seabed storage site using three different types of $p$$CO_2$ sensors. Int. J. Greenh. Gas Control 38, 121–134.
Baretta, J.W., Ebenhöh, W., Ruardij, P., 1995. The European regional seas ecosystem model, a complex marine ecosystem model. Neth. J. Sea Res. 33, 233–246.
Bates, N.R., Astor, Y.M., Church, M.J., Currie, K., Dore, J.E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., Santana-Casiano, J.M., 2014. A time-series view of changing ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification. Oceanography 27 (1), 126–141.
Bates, N.R., 2007. Interannual variability of the oceanic CO_2 sink in the subtropical gyre of the North Atlantic Ocean over the last 2 decades. J. Geophys. Res. 112, C09013.
Bergès, B.J.P., Leighton, T.G., White, P.R., 2015. Passive acoustic quantification of gas fluxes during controlled gas release experiments. Int. J. Greenh. Gas Control 38, 64–79.
Blackford, J.C., Gilbert, F.J., 2007. pH variability and CO_2 induced acidification in the North Sea. J. Mar. Syst. 64, 229–241.
Blackford, J.C., Allen, J.I., Gilbert, F.J., 2004. Ecosystem dynamics at six contrasting sites: a generic modelling study. J. Mar. Syst. 52, 191–215.
Blackford, J.C., Jones, N., Proctor, R., Holt, J., 2008. Regional scale impacts of distinct CO_2 additions in the North Sea. Mar. Pollut. Bull. 56, 1461–1468.
Blackford, J.C., Torres, R., Cazanave, P., Artioli, Y., 2013. Modelling dispersion of CO_2 plumes in sea water as an aid to monitoring and understanding ecological impact. Energy Procedia 37, 3379–3386.
Blackford, J.C., Stahl, H., Bull, J.M., Bergès, B.J.P., Cevatoglu, M., Lichtschlag, A., Connelly, D.P., James, R.H., Kita, J., Long, D., Naylor, M., Shitashima, K., Smith, D., Taylor, P., Wright, I., Akhurst, M., Chen, B., Gernon, T.M., Hauton, C., Hayashi, M., Kaieda, H., Leighton, T.G., Sato, T., Sayer, M.D.J., Suzumura, M., Tait, K., Vardy, M.E., White, P.R., Widdicombe, S., 2014. Detection and impacts of leakage from sub-seafloor deep geological carbon dioxide storage. Nat. Clim. Change 4, 1011–1016.
Blackford, J., Bull, J.M., Cevatoglu, M., Connelly, D., Hauton, C., James, R.H., Lichtschlag, A., Stahl, H., Widdicombe, S., Wright, I.C., 2015. Marine baseline and monitoring strategies for carbon dioxide capture and storage (CCS). Int. J. Greenh. Gas Control 38, 221–229.
Bradshaw, J., Dance, T., 2005. Mapping geological storage prospectivity of CO_2 for the world's sedimentary basins and regional source to sink matching. In: Rubin, E.S., Keith, D.W., Gilboy, C.F., Wilson, M., Morris, T., Gale, J., Thambimuthu, K. (Eds.), Greenhouse Gas Control Technologies 7 (Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies (GHGT-7), 5–9 September 2004, Vancouver, Canada). Elsevier, pp. 583–592.
Butenschön, M., Clark, J., Aldridge, J.N., Allen, J.I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., de Mora, L., Polimene, L., Sailley, S., Stephens, N., Torres, R., 2016. ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. Geosci. Model Dev. 9, 1293–1339.
Chadwick, R.A., 2010. Measurement and monitoring technologies for verification of carbon dioxide (CO2) storage in underground reservoirs. In: Maroto-Valer, M.M. (Ed.), Developments and Innovation in Carbon Dioxide (CO2) Capture and Storage Technology. Woodhead Publishing, pp. 203–239.
Clargo, N.M., Salt, L.A., Thomas, H., de Baar, H.J.W., 2015. Rapid increase of observed DIC and pCO2 in the surface waters of the North Sea in the 2001–2011 decade ascribed to climate change superimposed by biological processes. Mar. Chem. 177, 566–581.
Dewar, M., Wei, W., McNeil, D., Chen, B., 2013. Small scale modelling of the physiochemical impacts of CO_2 leaked from sub-seabed reservoirs or pipelines within the North Sea and surrounding waters. Mar. Pollut. Bull. 73, 504–515.
Dickson, A.G., 2010. The carbon dioxide system in seawater: equilibrium chemistry and measurements. In: Riebesell, U., Fabry, V.J., Hansson, L., Gattuso, J.P. (Eds.), Guide to Best Practices for Ocean Acidification Research and Data Report. Publications Office of the European Union, Luxembourg.
Dixon, T., McCoy, S.T., Havercroft, I., 2015. Legal and regulatory developments on CCS. Int. J. Greenh. Gas Control 40, 431–448.
Greenwood, J., Craig, P., Hardman-Mountford, N., 2015. Coastal monitoring strategy for geochemical detection of fugitive CO_2 seeps from the seabed. Int. J. Greenh. Gas Control 39, 74–78.
Gypens, N., Borges, A.V., Lancelot, C., 2009. Effect of eutrophication on air-sea CO2 fluxes in the coastal Southern North Sea: a model study of the past 50 years. Global Change Biol. 15, 1040–1056.
Harris, V., Edwards, M., Olhede, S.C., 2014. Multidecadal Atlantic climate variability and its impact on marine pelagic communities. J. Mar. Syst. 133, 55–69.
Hofmann, G.E., Smith, J.E., Johnson, K.S., Send, U., Levin, L.A., Micheli, F., Paytan, A., Price, N.N., Peterson, B., Takeshita, Y., Matson, P.G., Crook, E.D., Kroeker, K.J., Gambi, M.C., Rivest, E.B., Frieder, C.A., Yu, P.C., Martz, T.R., 2011. High-frequency dynamics of ocean pH: a multi-ecosystem comparison. PLoS One 6 (12), e28983.
Holt, J.T., James, I.D., 2001. An s-coordinate model of the north west European continental shelf: part 1. Model description and density structure. J. Geophys. Res. 106 (C7), 14015–14034.
Holt, J.T., Allen, J.I., Proctor, R., Gilbert, F.G., 2005. Error quantification of a high resolution coupled hydrodynamic-ecosystem coastal ocean model: part 1. Model overview and hydrodynamics. J. Mar. Syst. 57, 167–188.
Holt, J., Butenschön, M., Wakelin, S.L., Artioli, Y., Allen, J.I., 2012. Oceanic controls on the primary production of the northwest European continental shelf: model experiments under recent past conditions and a potential future scenario. Biogeosciences 9, 97–117.
Hvidevold, H.K., Alendal, G., Johannessen, T., Ali, A., Mannseth, T., Avlesen, H., 2015. Layout of CCS monitoring infrastructure with highest probability of detecting a footprint of a CO2 leak in a varying marine environment. Int. J. Greenh. Gas Control 37, 274–279.
IEA, Greenhouse Gas R & D programme (IEA GHG), 2008. Assessment of Sub-sea ecosystem impacts, 2008/8.
IPCC, 2005. Special report on carbon dioxide capture and storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
IPCC, 2006. Guidelines for National Greenhouse Gas Inventories. Vol. 2 Energy, Chapter 5, Carbon Dioxide Transport, Injection and Geological Storage. IGES, Japan.
Jenkins, C., Chadwick, A., Hovorka, S.D., 2015. The state of the art in monitoring and verification-ten years on. Int. J. Greenh. Gas Control 40, 312–349.
Lessin, G., Artioli, Y., Queirós, A.M., Widdicombe, S., Blackford, J.C., 2016. Modelling impacts and recovery in benthic communities exposed to localised high CO_2. Mar. Pollut. Bull. 109, 267–280.
Lewis, K., Allen, J.I., Richardson, A.J., Holt, J.T., 2006. Error quantification of a high resolution coupled hydrodynamic-ecosystem coastal-ocean model: part3, validation with continuous plankton recorder data. J. Mar. Syst. 63, 209–224.
Maeda, Y., Shitashima, K., Sakamoto, A., 2015. Mapping observations using AUV and numerical simulations of leaked CO_2 diffusion in sub-seabed CO_2 release experiment at Ardmucknish Bay. Int. J. Greenh. Gas Control 38, 143–152.
Newton, J.A., Feely, R.A., Jewett, E.B., Williamson, P., Mathis, J., 2015. Global Ocean Acidification Observing Network: Requirements and Governance Plan, second edition. GOA-ON. http://www.goa-on.org/docs/GOA-ON_plan_print.pdf.
Nye, J.A., Baker, M.R., Bell, R., Kenny, A., Kilbourne, K.H., Friedland, K.D., Martino, E., Stachura, M.M., Van Houtan, K.S., Wood, R., 2014. Ecosystem effects of the Atlantic multi-decadal oscillation. J. Mar. Syst. 133, 103–116.
Pale Blue Dot Energy, Axis Well Technology, Costain, 2016. Progressing Development of the UK's Strategic Carbon Dioxide Storage Resource. ETI. https://s3-eu-west-1.amazonaws.com/assets.eti.co.uk/legacyUploads/2016/04/D16-10113ETIS-WP6-Report-Publishable-Summary.pdf.
Phelps, J.J.C., Blackford, J.C., Holt, J.T., Polton, J.A., 2015. Modelling large-scale CO_2 leakages in the North Sea. Int. J. Greenh. Gas Control 38, 210–220.
Provoost, P., van Heuven, S., Soetaert, K., Laane, R.W.P.M., Middelburg, J.J., 2010. Seasonal and long-term changes in pH in the Dutch coastal zone. Biogeosciences 7, 3869–3878.
Radach, G., Moll, A., 2006. Review of three-dimensional ecological modeling related to the North Sea shelf system. Part II: model validation and data needs. Oceanogr. Mar. Biol. 44, 1–60.
Read, A., Tillema, O., Ros, M., Jonker, T., Hylkema, H., 2014. Update on the ROAD project and lessons learnt. Energy Procedia 63, 6079–6095.
Romanak, K.D., Bennett, P.C., Yang, C.B., Horvorka, S.D., 2012. Process-based approach to CO_2 leakage detection by vadose zone gas monitoring at geologic CO_2 storage sites. Geophys. Res. Lett. 39, L15405.
Saux-Picart, S., Butenschön, M., Shutler, J.D., 2012. Wavelet-based spatial comparison technique for analysing and evaluating two-dimensional geophysical model fields. Geosci. Model Dev. 5, 223–230.
Shitashima, K., Maeda, Y., Ohsuki, T., 2013. Development of detection and monitoring techniques of CO_2 leakage from seafloor in sub-seabed CO_2 storage. Appl.Geochem. 30, 114–124.
Shitashima, K., Maeda, Y., Sakamoto, A., 2015. Detection and monitoring of leaked CO2 through sediment: water column and atmosphere in a sub-seabed CCS experiment. Int. J. Greenh. Gas Control 38, 135–142.
Skogen, M.D., Budgell, W.P., Rey, F., 2007. Interannual variability in Nordic seas primary production. ICES J. Mar. Sci. 64, 889–898.
Thomas, H., Bozec, Y., Elkalay, K., de Baar, H.J.W., Borges, A.V., Schiettecatte, L.-S., 2005. Controls of the surfacewater partial pressure of CO_2 in the North Sea. Biogeosciences 2, 323–334.
Thomas, H., Friederike Prowe, A.E., van Heuven, S., Bozec, Y., de Baar, H.J.W., Schiettecatte, L.-S., Suykens, K., Koné, M., Borges, A.V., Lima, I.D., Doney, S.C., 2007. Rapid decline of the CO_2 buffering capacity in the North Sea and implications for the North Atlantic Ocean. Global Biogeochem. Cycl. 21, GB4001.
Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M.,

Berg, L.V.D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. Q. J. R. Meteorol. Soc. 131, 2961–3012.

Wakelin, S.L., Holt, J.T., Blackford, J.C., Allen, J.I., Butenschön, M., Artioli, Y., 2012. Modelling the carbon fluxes of the northwest European continental shelf: validation and budgets. J. Geophys. Res. 117, C05020.

Wynn, R.B., Huvenne, V.A.I., Le Bas, T.P., Murton, B.J., Connelly, D.P., Bett, B.J., Ruhl, H.A., Morris, K.J., Peakall, J., Parsons, D.R., Sumner, E.J., Darby, S.E., Dorrell, R.M., Hunt, J.E., 2014. Autonomous Underwater Vehicles (AUVs): their past, present and future contributions to the advancement of marine geoscience. Mar. Geol. 352, 451–468.

Zeebe, R.E., Wolf-Gladrow, D.A., 2001. CO_2 in seawater: equilibrium, kinetics and isotopes. Elsevier Oceanogr. Ser. 65, 346.