S.2 Methods

S.2.1 Quality control

An overview of the quality control methods used for the different time-series has been included in **Table S1**. The quality of the DIC and TA analyses is ensured by either using Certified Reference Material (CRM) provided by Prof. A. Dickson, UCSD, USA, and/or taking part in QUASIMEME intercalibration exercises.

The CRMs, which are glass bottles filled with seawater with known and certified concentration of DIC and TA, are analysed using the same instruments as the ordinary seawater samples. The results are used to determine the quality of the analyses. CRMs have been available since the 1990s (Dickson et al., 2007).

In 2021, QUASIMEME launched the AQ15 intercalibration exercise for TA and DIC and provided in each round 3 different types of seawater (high and low salinity seawater). About 25 labs, most of them successfully, participated in the first round. After this initial round, QUASIMEME set up a survey to further optimise this intercalibration exercise. Future workshops and different platforms and forums will help labs to strength the quality of their analysis and will lead more and more to harmonisation of methods and stronger data sets.

For pH data from the French SOMLIT stations, the quality control is performed by using CRMs (provided by Prof. A. Dickson, UCSD, USA) where reference pH values are calculated from certified DIC and TA values. Alternatively, the pH analyses can be quality checked by frequently measuring synthetic seawater (provided b Prof. A. Dickson, University of California, San Diego, USA) with known pH (TRIS buffer). The French SOMLIT datasets are also quality controlled by taking part in annually organized inter-laboratory tests at national level where most of the variables, including pH, are assessed.

The Belgian pH sensor data are daily calibrated with commercial NIST certified pH buffers followed by measurement of a commercial NIST certified buffer of pH=8. Furthermore, the Belgian pH sensors are checked using homemade TRIS buffers as described by Paulsen and Dickson (2020).

The Dutch pH data (NIOZ-Rijkswaterstaat) are measured spectrophotometrically with purified meta-cresol purple dye. Quality control is performed by using seawater CRMs and tris buffer (both from Prof A. Dickson, UCSD, USA) and by comparison with pH calculated from DIC and TA measurements on duplicate samples. Pre-2018 Dutch data (Rijkswaterstaat) were measured at sea with an electrode calibrated in standard NBS buffers.

pCO_2 analyses of discrete water samples are calibrated towards reference gasses with known amount of pCO_2.

S.2.2 Carbonate system calculations

For time-series datasets where pH was not measured directly, pH was calculated from DIC and TA or DIC and pCO_2 when TA was not available (see **Table S1**). Calculations were made using CO2sys_v2.5.xls (Pelletier et al., 2007) with K1 and K2 equilibrium constants from Lueker et al. (2000), $KHSO_4$ constant from Dickson (1990), KHF constant from Perez and Fraga (1987), and borate concentration from Lee et al. (2010). All pH values are given on total scale. An additional output from this calculation was also Ω (calcium carbonate saturation state).

S.2.3 Comparison of time-series

Time-series data for pH, Ω, temperature and salinity were averaged over season (Winter = December, January, February; Spring = March, April, May; Summer = June, July, August; Autumn = September, October, December). Using Minitab 18 (https://www.minitab.com/en-us/products/minitab/), the data were then analysed using Time Series Decomposition, which separates a time-series into linear trend, seasonal, and error components (Methods and formulas for Decomposition - Minitab). We used a seasonal length of 4 and a Multiplicative model type with model components of trend plus seasonal. We then used regression analysis to assess the linear trend through time in the seasonal (original) data as well as the 'deseasonalised' data.

S.2.4 Synthesis products and models

S.2.4.1 OceanSODA-ETHZ v2021 data (surface pH and Ω_{Arag}) for OSPAR regions

The details for the calculation of pH and Ω_{Arag} are given in Gregor and Gruber (2021). Here, we outline some important elements of the way in which the trends were calculated.

The OS-ETHZ dataset gives maps of the full marine carbonate system from 1985-2020 by estimating surface pCO_2 and Total Alkalinity (TA) from satellite and reanalysis model outputs. Phyton version of the speciation software CO2SYS is used to solve the full marine carbonate system from these two variables (details in Gregor and Gruber, 2021).

The updated v2021 dataset (https://www.ncei.noaa.gov/access/metadata/landingpage/bin/iso?id=gov.noaa.nodc:0220059) provides propagated uncertainties for several variables, including pH and Ω_{Arag}. Note that the values differ from the OS-ETHZ-v2020 values that were published in Gregor and Gruber (2021) since the Arctic was not included in the previous version.

Further, in v2021, ΔpCO_2 is the target value for the regression, compared to pCO_2 for v2020, which showed trends closer to the CMEMS-FFNNv2 output. In this OSPAR report, we used the original ensemble members to calculate the uncertainties of the trends of pH and Ω_{Arag}. There are 18 individual ensemble members of varying pCO_2, where nine of the members are estimated with feed-forward neural networks (FFNN) and nine are estimated with gradient boosted regression trees (GBDT). Note that the variability of FFNNs is much larger than that of GBDTs, thus the uncertainties given here are lower than those for the CMEMS approach.

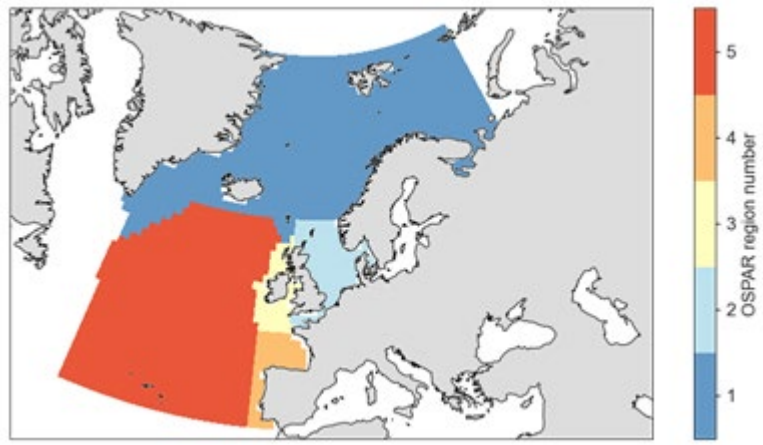

***Figure S1**: OSPAR regions and corresponding colours for time series plots in Figure S2.*

We calculate trends based on each individual member, and then use the standard deviation of the ensemble members as the "uncertainty" estimate. Trends are calculated on a per-pixel-basis. For regional averages, pixels are averaged using an area-weighted average. A note on uncertainties: the uncertainties are based on the ensemble member standard deviation. Thus, the uncertainty depends on the intrinsic variability of the regression method used to predict pCO_2 / TA. FFNN approaches typically have a much larger variability compared to tree-based approaches (i.e., GBDT) and linear regression approaches (which return the exact output every time). It is thus a tricky topic to tackle. In the past we've used inter-method σ rather than ensemble σ.

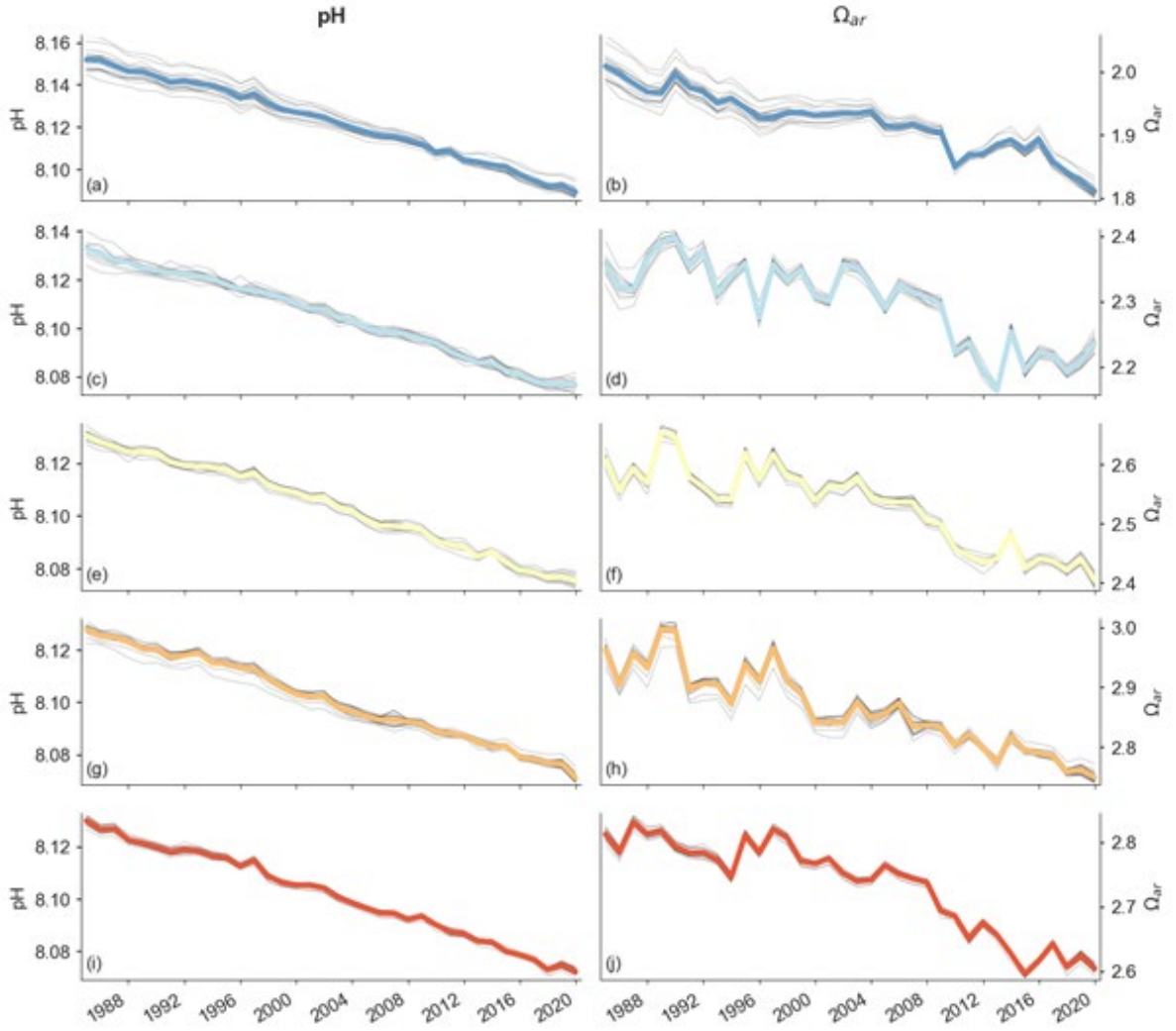

***Figure S2**: Time series of annually averaged pH (left) and Ω_{Arag} (right) for OSPAR regions. The colour of the thick lines corresponds to the colour in Figure S1. The thin lines represent the 18 ensemble members used to calculate the mean pH and Ω_{Arag}. Data was weighted by pixel area in the calculation of the mean value.*

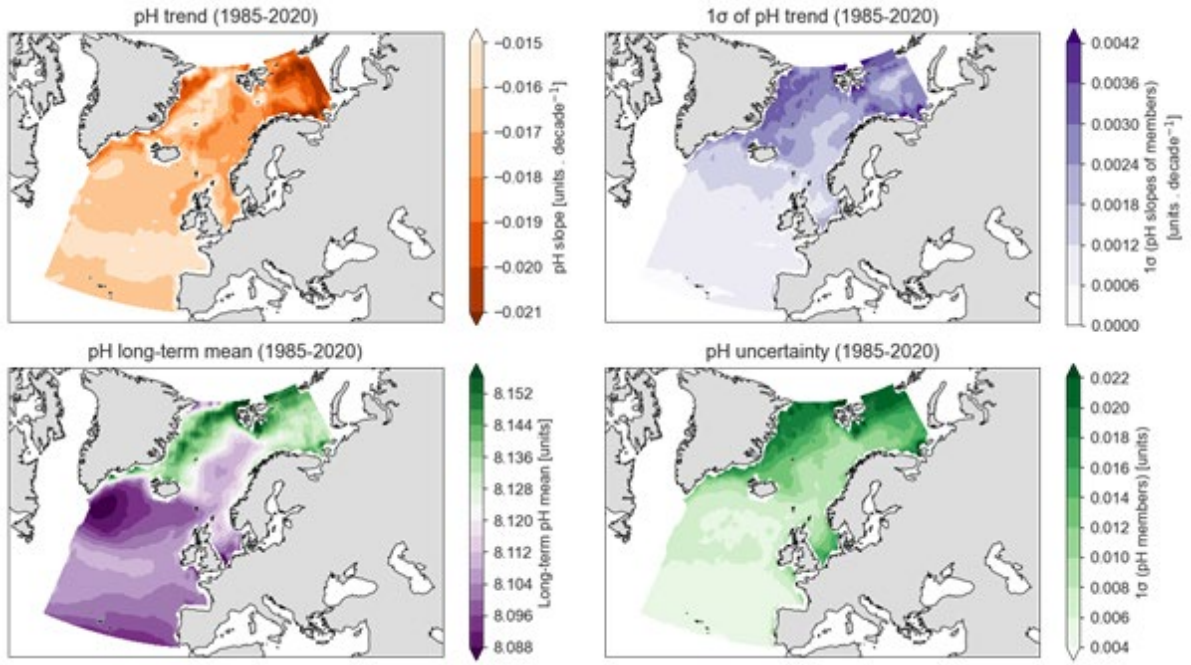

***Figure S3**: (a) the decadal mean slope of pH from 1985-2020; (b) the standard deviation of the slopes of the ensemble member used to calculate the slope in (a); (c) the long-term average of pH from 1985-2020.*

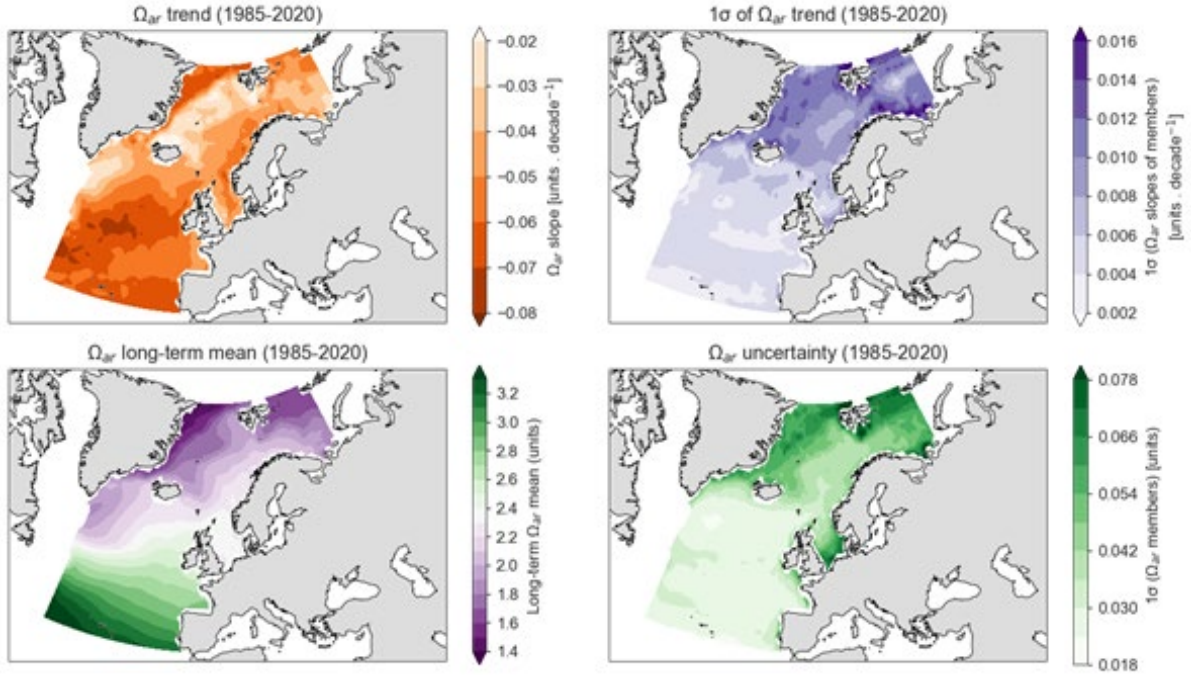

***Figure S4**: (a) the decadal mean slope of Ω_{Arag} from 1985-2020; (b) the standard deviation of the slopes of the ensemble member used to calculate the slope in (a); (c) the long-term average of Ω_{Arag} from 1985-2020.*

S.2.4.2 CMEMS-LSCE-FFNNv2 data (surface pH and Ω_{Arag}) over OSPAR regions

Here we report on annual trends and uncertainties of surface ocean pH and Ω_{Arag} over OSPAR regions for the period 1985-2020. Trends and uncertainties are derived from monthly reconstructions of these fields using the CMEMS-LSCE-FFNN model.

These two monthly fields were computed from reconstructed surface partial pressure of CO_2-spco$_2$- and surface ocean alkalinity using the CO2sys speciation software (Van Heuven et al., 2011; Lewis and Wallace, 1998). Time and space varying fields of best estimates of spco2 and model uncertainties are derived from an ensemble of 100 feed forward neural network models mapping the monthly gridded SOCATv2021 data. The monthly alkalinity fields were obtained from a multivariate linear regression with salinity, temperature, dissolved silica and nitrate as independent variables (LIAR, Carter et al., 2016; 2018). For each month t and each grid box ij, model best estimates ($\mu_{ij,t}$) are defined as the ensemble mean of the 100 model outputs of pH and Ω_{Arag}. Uncertainties ($\sigma_{ij,t}$) of the reconstructed monthly pH and Ω_{Arag} are defined as the total uncertainties combining the ensemble dispersion of the pH or Ω_{Arag} fields and the error propagation following Orr et al. (2018). See details in Chau et al. (2021)

Data: Data reconstructed by the CMEMS-LSCE-FFNN model and distributed for the assessment over OSPAR regions cover latitudes below 80°N (Figure S5). Small (for instance, locally over the eastern Greenland) or missing data coverage is linked to seasonal sea-ice cover and/or missing data of predictors taken account in our reconstruction.

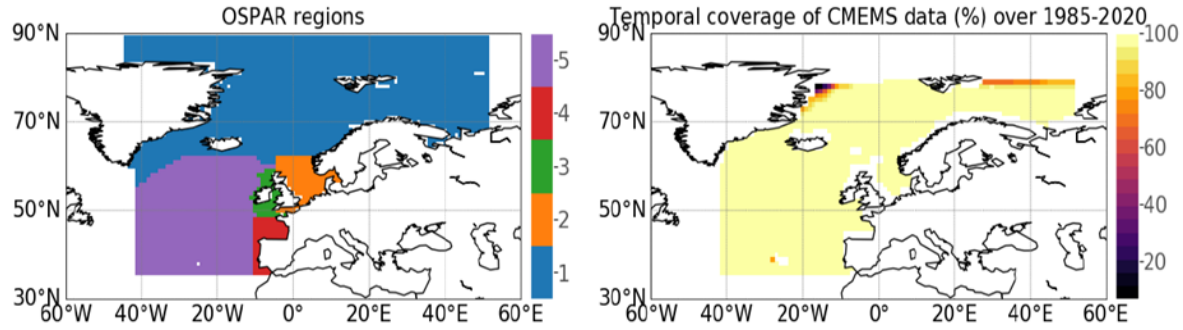

Figure S5*: (Left) OSPAR regions mask regridded at 1x1 resolution (original OSPAR mask at 0.25 resolution created by Luke Gregor, last access: 05/10/2021). (Right) Percentage of the total number of data reproduced by the CMEMS-LSCEFFNNv2 model over the 36-year period.*

A. Trend and uncertainty maps computed from CMEMS-LSCE-FFNNv2 products

Method: Computation of pH [Ω_{Arag}] trends and uncertainty at each grid point in the OSPAR regions.

With the best estimate $\mu_{ij,t}$ and model uncertainty $\sigma_{ij,t}$ of the monthly pH and Ω_{Arag} fields, we can regenerate the corresponding 100-member ensembles by assuming Gaussian distribution.

As the first step, we compute yearly means from the 100-member ensemble of the reconstructed monthly pH field. Denote $\{\mu_{ij}\}$ and $\{\sigma_{ij}\}$ as a linear trend and its uncertainty computed from a batch of 100 yearly mean pH [Ω_{Arag}] timeseries ($x_{t,ij,n}$) at each 1°x1°-grid cell ij, where t is now a time index of a year in the period from 1985 to 2020 and n indicates a member in the 100-ensemble. These two quantities (μ_{ij} and σ_{ij}) are estimated as a slope and its residual standard deviation derived from linear least-squares regression on the 100 timeseries.

The *stats.linregress* function in the *scipy* python package is used to fit a linear least-squares regression between two sets of data (i.e., $x_{t,ij,n}$ against t values) for each grid box ij. An illustration of linear fits on the yearly mean pH data at particular locations is shown in Figure S6. Blue points represent $pH_{t,ij,n}$ data in the 100-ensemble of yearly mean pH and orange lines stand for their linear fits. Values $\mu_{ij} \pm \sigma_{ij}$ displayed on the figure legend are estimates of a trend and its uncertainty.

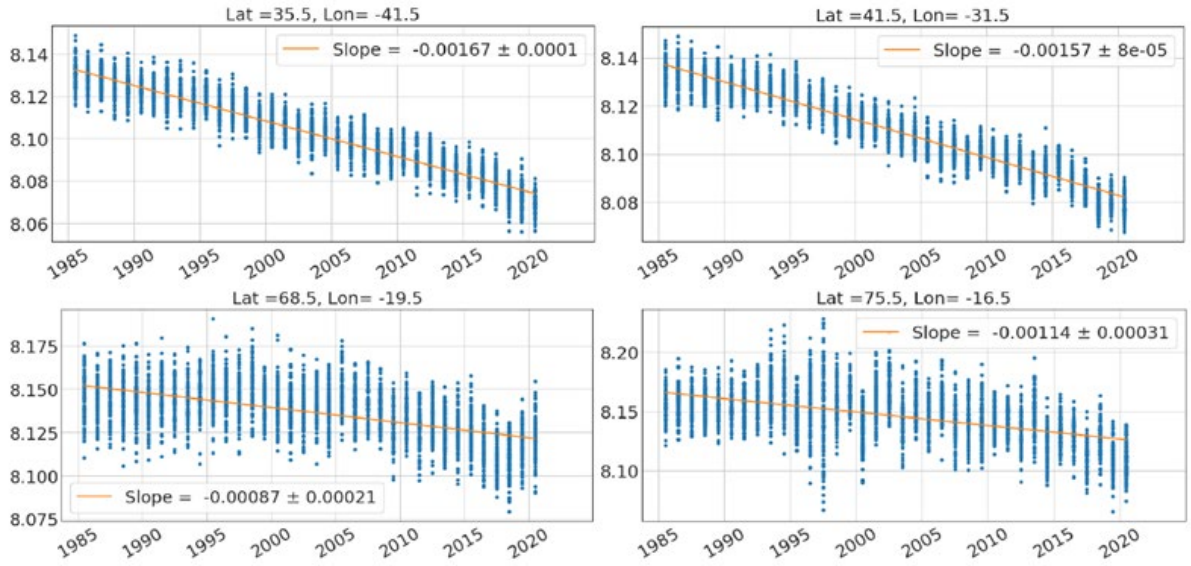

***Figure S6**: Annual trend and uncertainty of pH (-/yr) estimates derived from the 100-member ensemble of model outputs for different locations.*

This computation was applied for all grid points in the OSPAR regions, results are shown in Figures S7 and S8 (bottom plots) for pH and Ω_{Arag}, respectively.

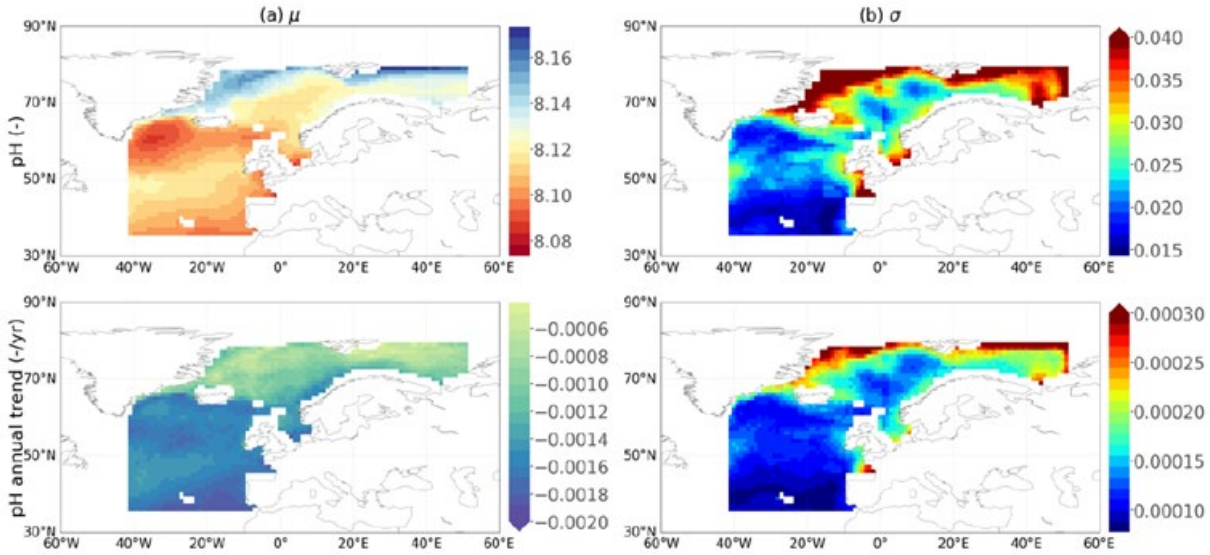

***Figure S7**: Temporal mean (top) and annual trend (bottom) of pH over 1985-2020: a) best estimate ($\mu=\{\mu_{ij}\}$) and b) model uncertainty ($\sigma=\{\sigma_{ij}\}$).*

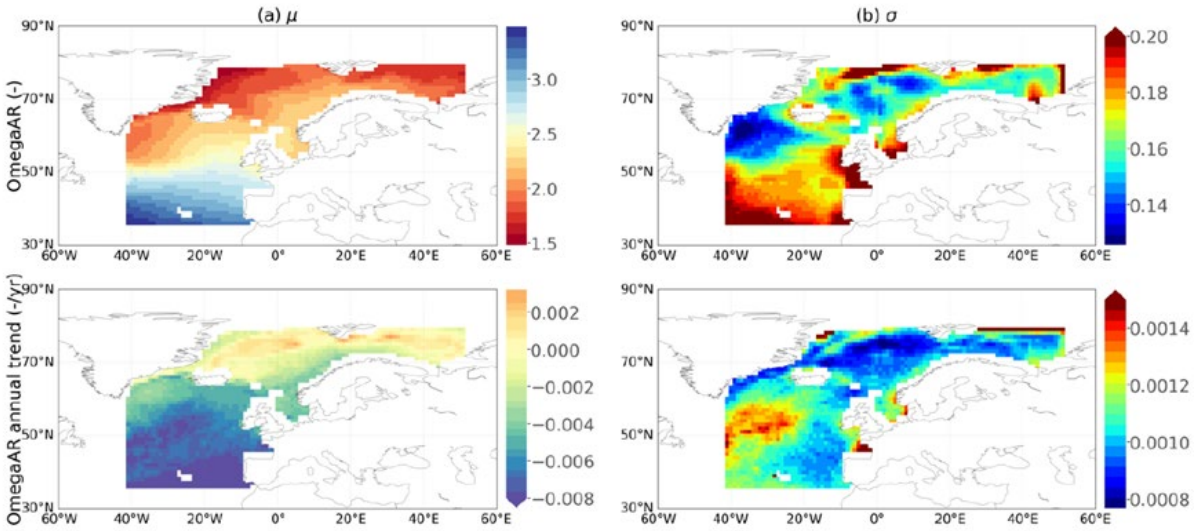

Figure S8*: Temporal mean (top) and annual trend (bottom) of Ω_{Arag} over 1985-2020: a) best estimate ($\mu=\{\mu_{ij}\}$) and b) model uncertainty ($\sigma=\{\sigma_{ij}\}$).*

B. Average data for each month and for each region, summaries of the trends for each region

Method: With the best estimate $\mu_{ij,t}$ and model uncertainty $\sigma_{ij,t}$ of the monthly pH and Ω_{Arag} fields, we can regenerate the corresponding 100-member ensembles ($x_{t,ij,n}$) by assuming Gaussian distribution. ij denotes an index of a 1x1°-grid box, t is a time index of a month in the period from 1985 to 2020, and n indicates a member in the 100-ensemble.

The monthly area-averaged pH [Ω_{Arag}] best estimate ($\mu_{r,t}$) and model uncertainty ($\sigma_{r,t}$) over each region r are computed as follows

$$\mu_{r,t} = \Sigma_n (\Sigma_{ij} x_{t,ij,n} A_{ij}) / (100\ \Sigma_{ij} A_{ij});$$

$$\sigma_{r,t} = \{\Sigma_n [\Sigma_{ij} (x_{t,ij,n} - x_{t,ij})^2\ A_{ij}\ ] / (100\ \Sigma_{ij} A_{ij})\}^{1/2};$$

where A_{ij} is the area of the ij^{th} 1x1°-grid box and $x_{t,ij} = \Sigma_n x_{t,ij,n} / 100$.

This computation was applied for all the 5 OSPAR regions (Figure 3.4), results are shown in Figures S9 and S10 for pH and Ω_{Arag}, respectively. The slope best estimate and its uncertainty for each region are computed as exemplified for a particular location presented in Section 3.1.

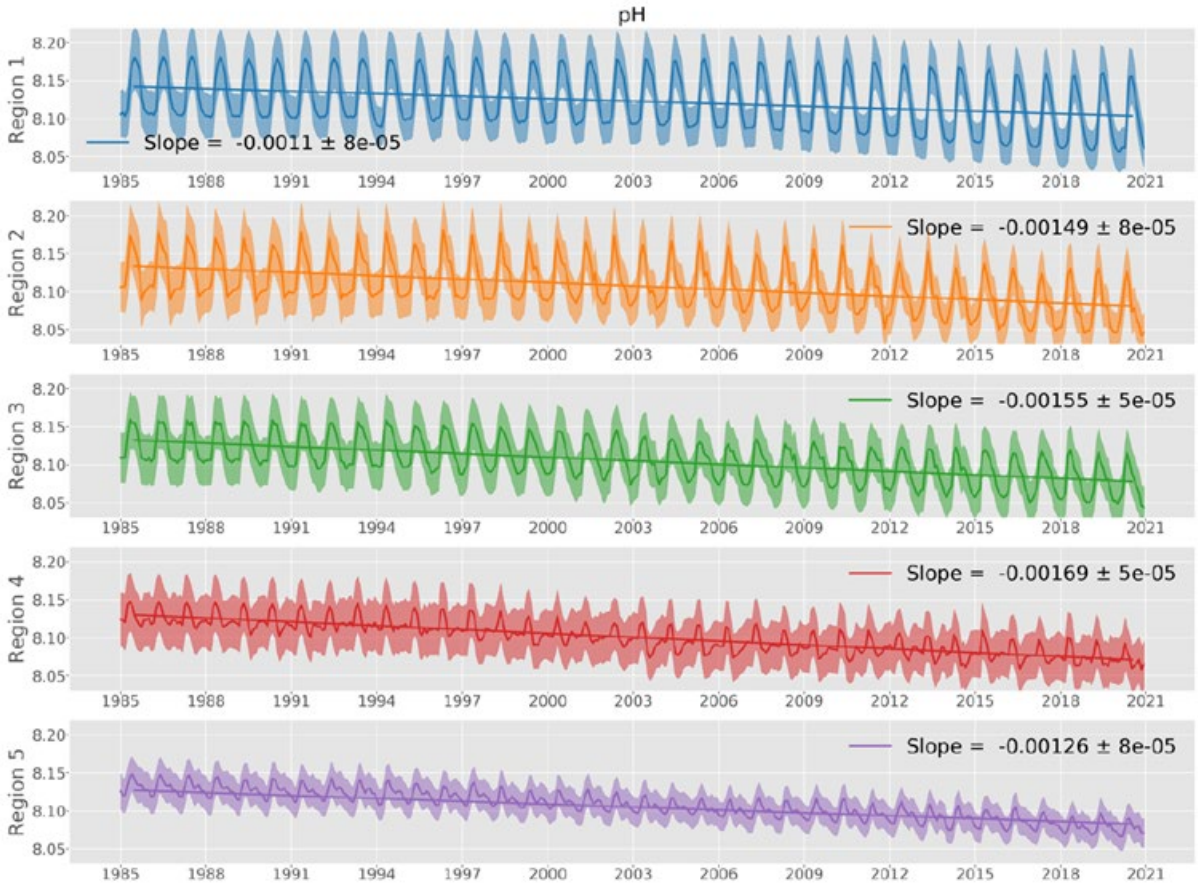

***Figure S9**: Monthly area-averaged pH best estimate (curve) and 68%-model spread, i.e. $\mu_{r,t} \pm \sigma_{r,t}$, (shaded area) over each OSPAR region. Trend slope and associated uncertainty estimates are shown in the legend.*

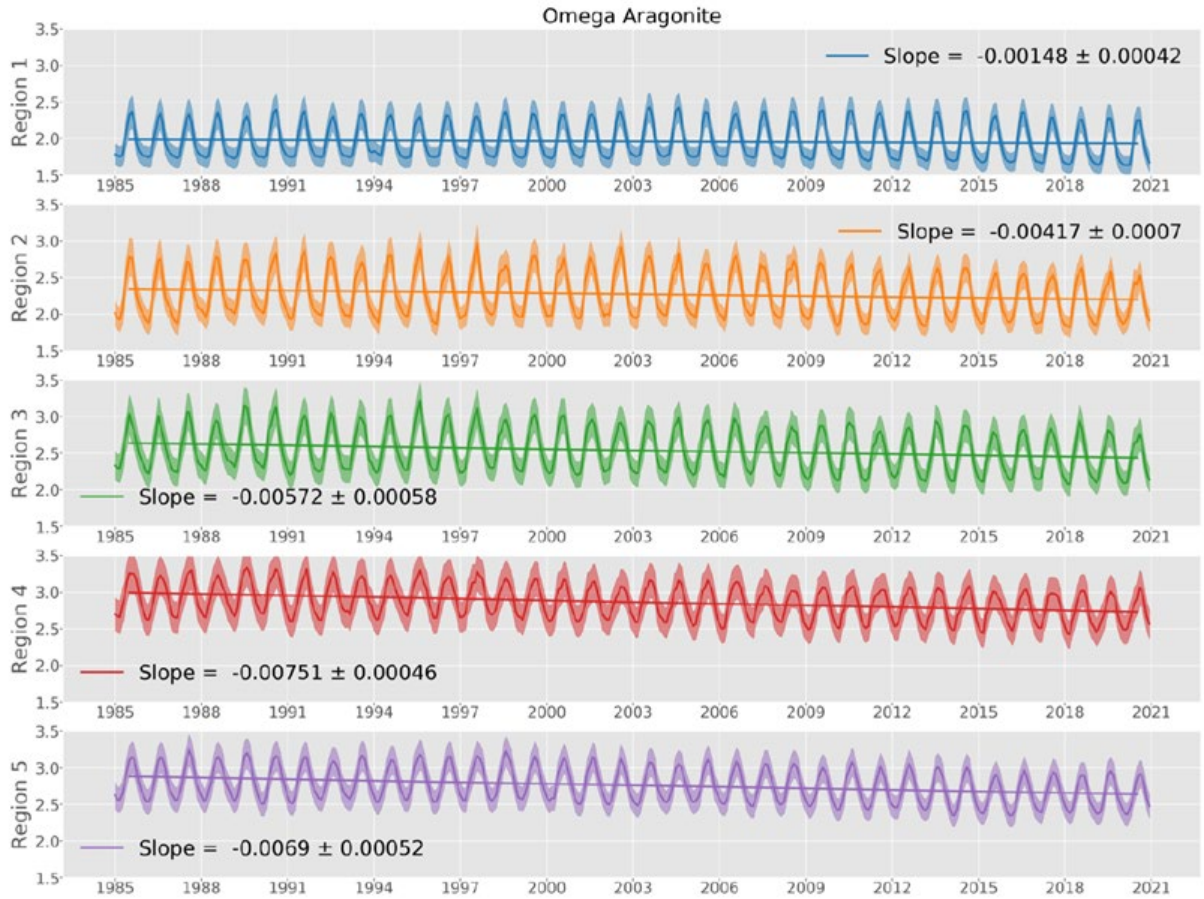

***Figure S10**: Monthly area-averaged Ω_{Arag} best estimate (curve) and 68%-model spread, i.e. $\mu_{r,t} \pm \sigma_{r,t}$, (shaded area) over each OSPAR region. Trend slope and associated uncertainty estimates are shown in the legend.*