Machine learning reveals regime shifts in future ocean carbon dioxide fluxes inter-annual variability

Check for updates

Damien Couespel [1] ✉, Jerry Tjiputra [1], Klaus Johannsen[1], Pradeebane Vaittinada Ayar [1] & Bjørnar Jensen [1]

The inter-annual variability of global ocean air-sea CO_2 fluxes are non-negligible, modulates the global warming signal, and yet it is poorly represented in Earth System Models (ESMs). ESMs are highly sophisticated and computationally demanding, making it challenging to perform dedicated experiments to investigate the key drivers of the CO_2 flux variability across spatial and temporal scales. Machine learning methods can objectively and systematically explore large datasets, ensuring physically meaningful results. Here, we show that a kernel ridge regression can reconstruct the present and future CO_2 flux variability in five ESMs. Surface concentration of dissolved inorganic carbon (DIC) and alkalinity emerge as the critical drivers, but the former is projected to play a lesser role in the future due to decreasing vertical gradient. Our results demonstrate a new approach to efficiently interpret the massive datasets produced by ESMs, and offer guidance into future model development to better constrain the CO_2 flux.

The ocean takes up roughly 25% of the total human-induced carbon emissions per year[1], thereby mitigating the consequences of the anthropogenic perturbation on the Earth system and its climate. This estimate is based on both, global ocean biogeochemistry models and observation-based data products[2]. While model-based estimates rely on simulations of ocean circulation and carbon biogeochemistry [e.g.[3]], the observation-based estimations rely on pCO₂-products that use machine learning to combine in-situ observation, remote sensing and reanalysis products [e.g.[4],]. Although estimates of present-day ocean carbon uptake rate and its long-term trends agree reasonably well using either tool, discrepancies in the spatio-temporal variability of the air-sea CO_2 flux (CO_2 flux, hereafter) remain, and these biases even appear to increase over time[5]. However, our ability to accurately quantify the magnitude of the ocean carbon sink and the variability of the CO_2 flux across multiple timescales is crucial to project the future evolution of the Earth's climate and to improve our ability to robustly detect long-term anthropogenic climate change[6,7].

The rate at which CO_2 is taken up by the ocean correlates on longer timescales with the fairly steady increase of atmospheric CO_2 concentration[8], but is modulated by inter-annual variability of the CO_2 flux (IAV, hereafter), regionally and globally[4,9,10]. Year-to-year down to decadal variations in the IAV of the CO_2 flux may be driven by modes of atmospheric variability such as the El Niño Southern Oscillation (ENSO), the

Southern Annular Mode (SAM) or the North Atlantic Oscillation (NAO) that modulate the ocean circulation, the sea surface temperature, the surface winds and the biology and consequently the CO_2 flux[9,11,12]. Modelling and observational studies have achieved substantial progress in quantifying IAV and identifying regional drivers over the historical and preindustrial periods (see literature review in[2,13,14]), however, part of the IAV remains unexplained and the dominant drivers in some areas have yet to be clearly identified.

Traditionally, variability of the global ocean carbon sink was attributed to equatorial Pacific Ocean variability of the CO_2 flux associated with ENSO[10,15,16] and to variability in the Southern Ocean linked to the SAM[4,12,17–19]. In these regions, the CO_2 flux IAV is driven by the upwelling of DIC-rich waters. In the equatorial Pacific, the upwelling is modulated by the ENSO phases[11,12,20,21]. In the Southern Ocean, oscillations in the position and strength of the westerly winds (related to the SAM,[18,22–24]) modulate the strength of the overturning and thus the upwelling[25]. However, recent findings suggest that the CO_2 flux pattern in the Southern Ocean are more sensitive to local atmospheric variability[17,26].

Besides the equatorial Pacific and the Southern Ocean, substantial IAV of the CO_2 flux has been described in the North Atlantic associated with the NAO[27–29] and the Atlantic Multidecadal Oscillation (AMO)[30], in the tropical Atlantic in response to Atlantic Niño[31], and in the Pacific Ocean linked to the Pacific Decadal Oscillation (PDO)[32]. However, here, models tend to

[1]NORCE Norwegian Research Centre AS, Bjerknes Centre for Climate Research, Bergen, Norway.
✉e-mail: daco@norceresearch.no

disagree on the dominant drivers, and multiple driving mechanisms are often at play[2,12,33]. Our understanding of the IAV remains incomplete, especially in these regions, where climate modes explain less than 20% of the large-scale variance in oceanic surface pCO_2 and CO_2 flux[2,30,32].

The CO_2 flux IAV is substantial and modulates the global warming signal, which makes it difficult to identify the latter[10,34–36]. While global ocean biogeochemistry models tend to underestimate IAV[5,37], a signal that could even amplify in the future due to a higher Revelle factor[20,33], sparse and inhomogeneous sampling leads to biased IAV estimates from pCO_2 products as well[38,39].

Currently, there is no study that explores the drivers of the CO_2 flux IAV in the global ocean, applying multiple future scenarios from the latest generation Earth System Models (ESMs) ensemble. Studies that aim to identify mechanistic drivers of CO_2 flux variability are generally limited to a single model, on the historical/preindustrial period or on specific regions, making it difficult to draw a consistent synthesis[12,32,40,41]. Previous works have largely used the Taylor expansion approach to decompose the perturbation of pCO_2 into the sum of the perturbations of DIC, alkalinity, temperature and salinity (e.g., refs. 33,42–44). This approach has two constraints: (1) it relies on a linear decomposition around a mean state, and (2) it is restricted only to the four aforementioned drivers of pCO_2. The potential pCO_2 drivers are more numerous (e.g. nutrient, export production, depth of the mixed layer), which altogether interact in a non-linear way. Moreover, the CO_2 flux also depends on the surface wind, the CO_2 solubility and the sea-ice coverage. Earlier study[22] extend the Taylor expansion to wind and sea-ice, computing the sensitivity parameters (i.e. the partial derivatives) specifically for their model. The sensitivity parameters can also be computed using a set of sensitivity simulations, as done in some studies for pCO_2[45], or a dedicated adjoint model[46]. However, these approaches are currently less feasible using multiple highly complex Earth system models. Our new approach allows us to identify a larger set of drivers, assessing their robustness across ESMs and scenarios and coherently determining the specificities of individual ESMs.

Simulating multiple climate scenarios with multiple ESMs and analysing them to examine CO_2 flux IAV require considerable computational and human resources. ESMs are state-of-the-art tools for investigating past and present drivers of the CO_2 flux and projecting future changes[12,47]. ESMs simulate sophisticated interactions among different earth system components, which often vary considerably in level of complexity and process representations[48–50]. Setting-up multi-model simulations only to investigate the role of each bio-physical driver of the CO_2 flux is a prohibitively costly exercise. Further, the identification of the underlying mechanisms is particularly complicated due to the multiple number of processes integrated in the models and the non-linear interactions between them, not to mention the particularly large amounts of data generated by ESMs simulations. This situation calls for more feasible solutions.

Recently, machine learning techniques have been introduced to objectively, systematically and efficiently explore large volumes of data from ESMs and observational systems[51,52]. They are capable of reconstructing complex and non-linear variables such as marine primary production[53]. Going beyond the "black box" use of machine learning algorithms, explainable and interpretable machine learning approaches are particularly promising as they enable to assess the physical relevance of the results. For instance, unsupervised clustering has been used to reveal regimes of the ocean dynamics and ocean carbon budget that are consistent with the classical theoretical framework[54–56]. Qualitatively correct mechanisms controlling phytoplankton growth can be inferred from the machine learning reconstruction by examining the contributions of the different predictors, at least for relatively basic relationships, and providing that the time scales considered are appropriate[57]. In line with classical interpretation of the impact of global warming on primary production from biogeochemical models, the analysis of time series from the Sargasso Sea (BATS) with Genetic Programming revealed strong statistical relationship between warming, stratification and primary production decrease[58]. Quantifying the relative importances of the predictors for the reconstruction of pCO_2 reveals

that mixed layer depth is a major contributor of the seasonal cycle in the mid and high latitude and that anomalies of sea surface temperature causes important adjustments to the predicted pCO_2 in the equatorial Pacific[59].

The main objectives of this study are (1) to show that the present and future CO_2 flux IAV can be reconstructed from a set of predictors, (2) to identify the essential regional drivers of these reconstructions and (3) to determine how the identified drivers and mechanisms are projected to change in the future inferring new knowledge on the CO_2 flux IAV. The novelty of our work is the application of a machine learning approach as an analysis tool on a suite of multi-model simulations. In particular, it is an alternative approach to the Taylor expansion commonly used to identify pCO_2 drivers, but here it allows for more drivers to be evaluated, coherently across ESMs and scenarios, for the CO_2 flux.

Results
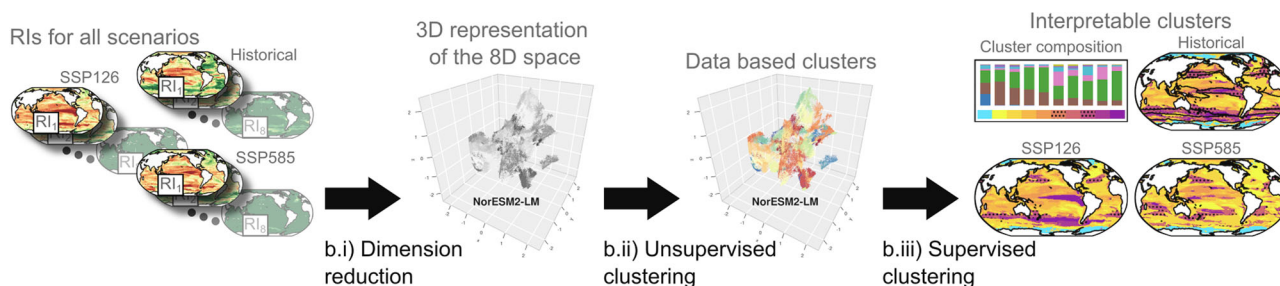### CO_2 flux reconstruction
In order to establish a linkage between CO_2 flux and its drivers without performing many model simulations, a reconstruction method is developed. Here, we use existing monthly outputs from 20 CMIP6 simulations[60]: 4 scenarios (preindustrial, historical, SSP126 and SSP585) performed with 5 ESMs (ACCESS-ESM1-5, CESM2, IPSL-CM6A-LR, MPI-ESM1-2-LR, NorESM2-LM). The entire time period of the simulations are used, i.e. respectively 200 years, 165 years (1850-2014) and 86 years (2015-2100) for the preindustrial, historical and global warming scenarios. In addition to the target CO_2 flux, 8 predictors of surface ocean properties have been selected: dissolved inorganic carbon (DIC), alkalinity (ALK), temperature (SST), salinity (SSS), phosphate (PO_4), surface wind speed (sfcWind), sea ice concentration (siconc) and atmospheric CO_2 concentration (atmCO2). The predictors' selection is guided by the variables commonly used in the calculation of the CO_2 flux, but also by the data availability. Mixed layer depth and biological export production have also been considered. Although they are not used in the models for calculating CO_2 flux, they are key variables[43,61] and mixed layer depth is an important predictor for building pCO_2-products[59,62]. However, a test with NorESM2-LM outputs shows that the reconstruction is not improved by including them and that their relative importances are very low (Fig. S17). These ESMs' outputs (both predictors and targets) are fed to the kernel ridge regression algorithm (hereafter regression) that returns a function F, which can reconstruct the full temporal variability of the CO_2 flux from the 8 predictors in each grid point and each simulation (step a.i in Fig. 1).

The regression succeeds in reconstructing most of the CO_2 flux IAV. The ability of the function F (output of the regression) in reproducing the original CO_2 flux IAV is assessed using the coefficient of determination, or R^2 score, between the reconstructed and original signal, both de-trended and de-seasonalized (step a.ii in Fig. 1). This score measures the proportion of the CO_2 flux IAV (measure by the variance) that is reproduced by the reconstruction. It should be pointed out that since deseasonalisation is achieved by removing the climatological seasonal cycle for each simulation, the amplification of the seasonal cycle induced by global warming[63] will be accounted for as an amplification of the IAV. The inter-model minimum score in the different simulations is very close to one in the major part of the ocean (Fig. 2, see also Fig. S1 for scores of individual model and scenario). In the preindustrial scenario, the average score for all ESMs in all grid-points is 0.94. 80% of the ocean grid-points have an inter-model minimum score above 0.81 and 0.82 in the preindustrial and historical simulations, respectively. In the global warming scenarios, the score is slightly lower, with 80% of the ocean having a minimum score above 0.72 and 0.68 in the SSP126 and SSP585, respectively. The score used here is relatively restrictive, even though the score is low (e.g., 0.41 in a grid point in the Southern Ocean, Fig. 2h), the correlation between the original and reconstructed time-series is higher (e.g. 0.66 in the same grid point, see also Fig. S18) indicating that the dominant part of the IAV is well reconstructed. Regionally, the lowest scores are found in the high latitudes, north (south) of 40°N(S), with the lowest values in the Southern Ocean. Our investigations suggest that the variability in the Southern Ocean is more complex and richer than in other

a) CO₂ flux reconstruction and drivers characterization with machine learning
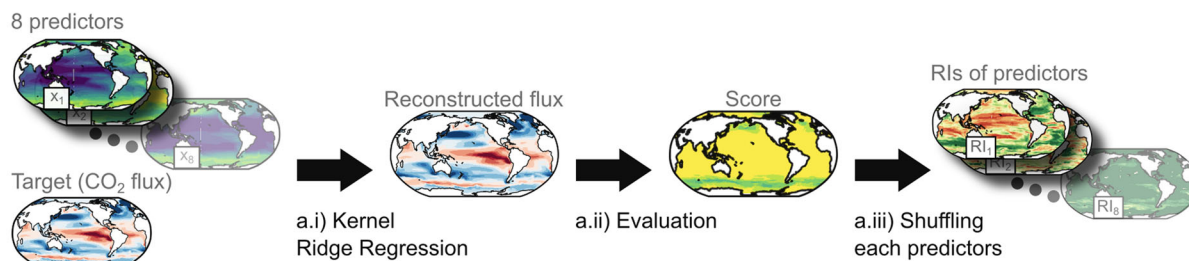
b) Spatio-temporal characterization of the drivers through clustering

Fig. 1 | Schematic workflow of CO_2 flux reconstruction and spatio-temporal characterisation of the drivers. Steps **a**.i to **a**.iii are applied to each scenario run with each ESM, while steps **b**.i to **b**.iii are applied to each ESM, gathering the results from the four scenarios. (step **a**.i) Determine a function F that reconstructs the CO_2 flux IAV based on 8 predictors and the actual air-sea CO_2 applying a kernel ridge regression. (step **a**.ii) Estimate the ability of F to reconstruct the CO_2 flux IAV from the 8 predictors. (step **a**.iii) Estimate the relative importance (RI) of each predictor. Each grid-point is now described by 8 RIs. (step **b**.i) Reduce the grid-point description from 8 to 3 dimensions to speed up the clustering analysis. (step **b**.ii)

Perform an unsupervised model-based clustering using an Expectation-Maximization algorithm. (step **b**.iii) Apply a tree-based classification to reduce the number of cluster to 10 per ESM. Finally, the results are back-projected on latitude and longitude for each scenario to get the clusters maps. Steps **a**.i to **a**.iii are repeated 50 times following a bootstrapping approach. The results (score, RIs) shown are derived from the average on the bootstrap ensemble. The standard deviation of the bootstrap ensemble measure the uncertainty of the kernel ridge regression. The full sequence of the analysis procedure is detailed in the "Methods" section.

regions (Figs. S14 and S15). Spectral analysis of specific grid points with low score in the Southern Ocean and subpolar North Pacific shows strong sub-seasonal variations, which generally are not present in regions with higher score, as the Subtropical and equatorial Pacific (Fig. S15, first column). In addition, the variability of the predictors is mainly dominated by the seasonal cycle and does not have strong sub-seasonal variations (Fig. S15). Under the future global warming scenarios, grid points with an important drop in the score are the ones where variability gets even richer (Fig. S16c, i). Altogether, this suggests that the kernel ridge regression is less optimum in reproducing variability that includes several equally dominant modes occurring at different time scales, especially when this variability is not present in the predictors. To further demonstrate that the regression is successful at reproducing the dominant IAV signals, Fig. 2b, d, f and h show the time series of the IAV of the original and reconstructed CO_2 flux in one grid point of the Southern Ocean where the score is particularly low (pre-industrial: 0.81, historical: 0.75, SSP126: 0.54, SSP585: 0.41). The time series is shown for the model with the lowest score. Even in this worst case, the reconstruction well reproduces the main CO_2 flux IAV.

#### Key predictors of contemporary CO₂ flux IAV
The positive outcome of the regression in reconstructing the CO_2 flux IAV does not necessarily mean that this variability emerges from physically meaningful interactions between the different predictors. The "normalised permutation feature importance" measures the predictors' relative importances in the reconstruction and helps to interpret the regression results (step a.iii in Fig. 1). For each model and each scenario, we obtain 8 maps, one per predictor, indicating their relative importances in the reconstruction (Figs. S2–S6). Preliminary work convinced us that the kernel ridge regression was able to reproduce the seasonal variability with physically meaningful interactions between the predictors. As expected from the literature[64],

temperature is the most important predictor in the subtropics while DIC is key in the higher latitudes (Fig. S20).

To aggregate the relative importances' information, ease the interpretation and assess the physical meaning of the reconstruction at the global scale, we use an unsupervised clustering method, followed by a tree-based classification (steps b.i, b.ii and b.iii in Fig. 1). For each ESM, grouped into preindustrial, historical, SSP126 and SSP585 scenarios, the spatial clustering reveals regions sharing the same relative importances of the predictors. The relative importance of the sea ice concentration in the ice-covered region[65,66] is rather obvious. Thus, grid points with the relative importance of sea-ice concentration >0.1 are grouped into a single cluster, the sea-ice cluster. On the other grid points, clustering and classification are performed. The total number of clusters is restricted to 9, to balance between an ease of interpretation and a fairly detailed description. Figure 3 shows the composition of the different clusters and their geographical distribution. Since ALK and DIC are persistently identified as the most important predictors, the 9 clusters are sorted according to the ratio between the relative importances of ALK and DIC. In yellowish clusters, ALK is relatively important, while in purplish clusters DIC is relatively important. Because the results from the preindustrial scenario are visually very similar to the historical scenario (Fig. S19), they are not further discussed. Note also that, by design, the simulations used here are driven by atmospheric CO_2 concentrations increasing quite smoothly in time and with no feedback from the ocean. Thus, the almost negligible importance of atmospheric CO_2 for IAV is to be expected.

The historical scenario allows us to assess the physical interpretation of the reconstruction by comparing the geographical cluster distribution and composition with the current knowledge of CO_2 flux IAV drivers (left column in Fig. 3). Firstly, the major oceanographic regions such as the subtropical and subpolar gyres, the Southern Ocean and the equatorial band

Fig. 2 | Ability of the kernel ridge regression in reproducing the original CO₂ flux inter-annual variability. Inter-model minimum score of the kernel ridge regression in (**a**) the preindustrial, (**c**) the historical, (**e**) the SSP126 and (**g**) the SSP585 scenarios. **b, d, f, h** show 5 years of the de-trended and de-seasonalized CO₂ flux time-series from the ESM with the lowest score (black lines) and from the kernel ridge regression reconstruction (dashed gold lines). Grey shading shows the standard deviation of the reconstructed CO₂ flux derived using boot-strapping. Shown are the 5 last years of the simulations from a single grid point in the Southern Ocean (grey point in the maps). In a), red dashed lines indicate regions with special focus in the analysis: the North Atlantic (10˚N to 70˚N), the equatorial Pacific (10˚S to 10˚N) and the Southern Ocean (south of 40˚S).
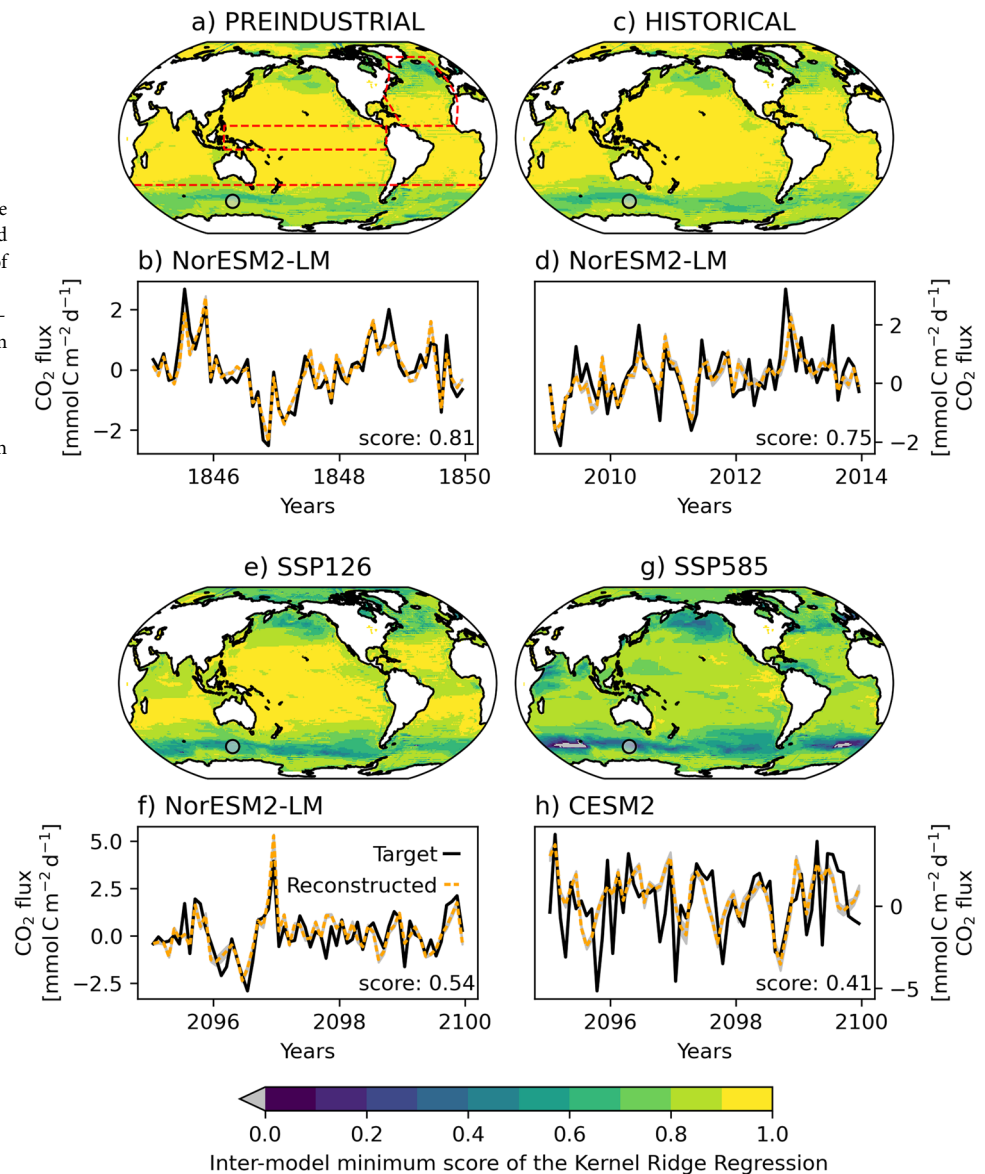
emerge from the analysis. The regions emerging from our analysis roughly matches the biogeochemical provinces previously defined based on other criteria (ref. 67, black contour in Fig. 3). Of special interest and an encouraging result is that both the equatorial Pacific and the Southern Ocean are associated with similar clusters, in which DIC is the most important predictor. This agrees with the fact that the CO₂ flux IAV in these two regions is driven by the variation in the upwelling of DIC-rich water masses[9,11,12,22–25]. A study applying a Taylor expansion for identifying the drivers of pCO₂ IAV in ESMs on the historical period reasonably agrees with our results, notably highlighting the strong contribution of DIC in the Southern Ocean and the equatorial versus its weak contribution in the subtropical gyres [Fig. 2 of ref. 33]. However, the Taylor expansion approach reveals a strong contribution from temperature and a weak contribution from alkalinity, while our results tend to emphasise the dominant role of alkalinity. The Taylor expansion approach has also been used with obser-vational data to investigate the drivers of 30-year trends (and not specifically on IAV) in surface pH and CO₂ flux over the same biogeochemical provinces[68]. A deeper and regionally detailed analysis would be feasible with a higher number of cluster, restricting the clustering to a specific region or even directly using the relative importance maps. Such a detailed analysis is beyond the scope of this study, as we aim to describe the large-scale regional

drivers of CO₂ flux IAV. Nonetheless, to get a finer understanding, we focus on key CO₂ flux regions with the strongest IAV[2,12,14]: the Southern Ocean (south of 40˚S), the Equatorial Pacific 10˚S to 10˚N) and the North Atlantic (10˚N to 70˚N, see red dashed lines in Fig. 2a).

A closer look at the Southern Ocean shows some zonal differences between the Pacific, Atlantic and Indian sectors, specially for the ACCESS-ESM1-5, CESM2, and NorESM2-LM models. In the Pacific sector, DIC has a relatively high importance, while in the two other sectors its importance is weakened. Such zonal differences in the dominant drivers have already been suggested in former studies: in the Pacific sector, the CO₂ flux is inter-mittently driven by temperature or non-thermal processes (such as mixed layer deepening or biology) while in the Atlantic and Indian sector, it is mainly driven by non-thermal processes[14,17,69]. There are also different behaviour between ESMs. For instance, in CESM2, the contribution of DIC is a bit weaker than the respective contribution in MPI-ESM1-2-LR (Figs. S3b and S5b), in line with findings for the former version of these models [Fig. 2e of ref. 33]. It should be noted that ESMs are known to under-represent decadal variations[12,14]. A previous study[69] suggested that decadal variability is due to the upwelling of DIC rich waters while the inter-annual variability is rather controlled by biology, wind and SST. Moreover, we note that the Southern Ocean is a region with a relatively low score for the IAV
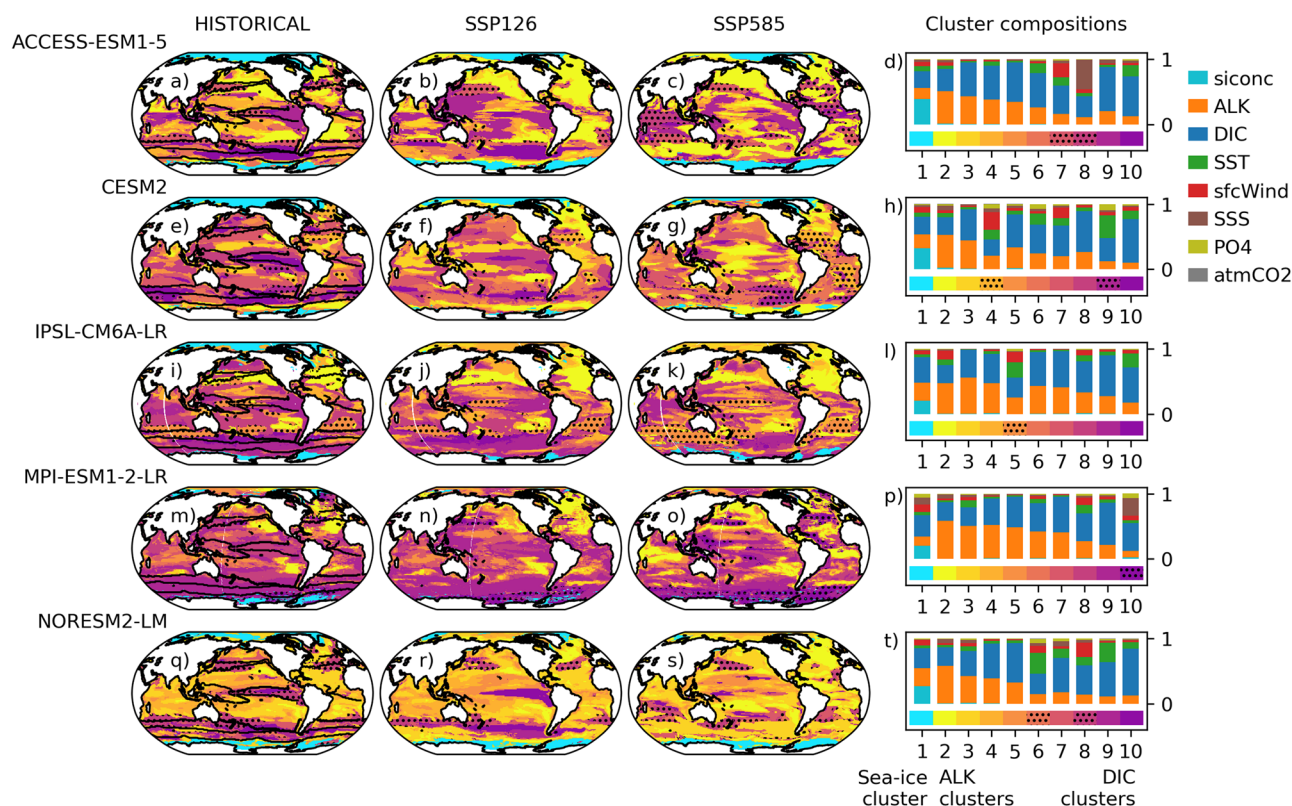
Fig. 3 | The relative importances of the predictors of the CO₂ flux inter-annual variability across Earth System Models and scenarios. The first 3 columns show the maps of relative predictor importance clusters for each scenario (historical in **a, e, i, m, q**, SSP126 in **b, f, j, n, r** and SSP585 in panels **c, g, k, o, s**). The last column (**d, h, l, p, t**) shows the clusters' composition. The results for each ESM are displayed in rows. Cluster composition (vertical bars) shows the relative importances of the predictors averaged over the grid points assigned to each of the 10 clusters. Cluster 1 gathers the grid points with high relative importance of sea-ice concentration (>0.1)

and is thereby named sea-ice cluster. Cluster 2 to 10 are defined with the unsupervised clustering method and a tree based classification (see Fig. 1 and "Methods" section). They are sorted according to the decreasing ratio between the relative importances of ALK and DIC: cluster 2 is dominated by ALK, while cluster 10 is dominated by DIC. The stippling highlights clusters with a low sum of the ALK and DIC relative importances (<0.6). The black contour lines show the biomes define in ref. 67. Cluster maps for the preindustrial scenario are very similar to the historical scenario and thus omitted.

reconstruction (Fig. 2). In the Antarctic, the sea-ice cluster cover mostly the Ross and Weddell Seas as well as some of the Antarctic margins for ACCESS-ESM1-5, MPI-ESM1-2-LR and NorESM2-LM. The sea-ice cluster does not extend until the northern boundary of the ice biomes from ref. 67 and corresponds roughly to the sea-ice covered regions [ref. 65, Figs. S2 and S3]. Yet, the differences between ESMs do not appear to be related to the differences in sea-ice extent area. Deep convection and the induced input of inorganic carbon from the deep ocean has been identified as an important driver of IAV in these regions[12].

In the North Atlantic, the general pattern mostly shared between the 5 models is ALK-dominated clusters in the subtropical and subpolar gyres, separated by DIC-dominated clusters in the inter-gyre (Gulf Stream and its eastward extension across the Atlantic). This is particularly clear in NorESM2-LM and ACCESS-ESM1-5 while IPSL-CM6A-LR does not show the inter-gyre separation. The CO₂ flux IAV in the North Atlantic is known to result from the balance between multiple drivers: biology[70,71], temperature[10,30,40,41,70–73], alkalinity[73,74], vertical mixing of DIC[12,30,41,71,72] and horizontal transport of DIC and/or heat[41,71,74] without a clear consensus on the dominant driver. This complexity in the balance is reflected in our analysis, as the sum of the relative importances of ALK and DIC is quite low in the region, meaning that the other predictors are potentially important too (dotted clusters in Figs. 3 and S7). Moreover, SST also emerges with a relatively high importance, in line with the studies mentioned above.

In the equatorial Pacific, the dominant driver of the CO₂ flux IAV is known to be the upwelling of cold and DIC-rich waters, with the intensity of the upwelling being modulated by ENSO[11,20]. In line with this, DIC-dominated clusters are present in the equatorial Pacific. A closer look reveals

that SST and surface winds have a higher relative importance in this region compare to the others (Figs. S2–S6, panels j, n). Temperature has previously been identified as a relatively important factor in offsetting the variability due to upwelling, notably in the eastern part of the equatorial Pacific[9,12,75]. As for the Southern Ocean, there are various behaviour between ESMs. As identified in their former model versions [Fig. 2e of ref. 33], CESM2/MPI-ESM1-2-LR exhibits a stronger/weaker contribution of DIC in the equatorial Pacific (Figs. S3b and S5b).

Key predictors of future CO₂ flux IAV
The analysis of the relative importance of the predictors and the comparison with the existing knowledge of the variability of the CO₂ flux suggests that this reconstruction has sound physical meaning. This gives us confidence that the analysis of relative importance can also help us to generate reliable knowledge on the projected CO₂ flux IAV drivers and provide guidance for future analyses. This section focuses on analysis of how the drivers of CO₂ flux IAV in key regions evolve under two future scenarios.

In the Arctic Ocean, the sea-ice cluster is consistently projected by the models to shrink in the future as the sea-ice cover declines under global warming scenarios. In the SSP585 scenario, notably, the Arctic Ocean becomes ice-free in the summer from 2050 onwards[66]. Along the Antarctic, despite a similar decrease in sea-ice cover in all global warming simulations[65], the sea-ice cluster does not obviously shrink in our analysis. It shrinks for MPI-ESM1-2-LR, but remains the same or expands for others (ACCESS-ESM1-5, CESM2, IPSL-CM6A-LR and NorESM2-LM).

In the rest of the ocean, the most noticeable change between the historical and global warming scenarios is the change in the relative

importances of DIC and ALK with a shift from DIC-dominated clusters to ALK-dominated clusters (from purple to yellow in Fig. 3). Depending on the model and scenario, the sum of the relative importance of DIC and ALK slightly increases or decreases but remains high in all scenarios (Fig. S7), indicating that these two predictors remain as the most important drivers for the CO_2 flux IAV. However, the relatively constant sum is due to the decreasing DIC relative importance while that of ALK increases (Fig. S8). This result is more obvious in the SSP585 global warming scenario and in the previously highlighted regions, the equatorial Pacific, the Southern Ocean and the North Atlantic. Earlier studies[20,75] focusing on ENSO and the equatorial Pacific also conclude that alkalinity has a greater influence than DIC in the SSP585 scenario. A closer look (Figs. S2–S6) reveals noticeable changes in the relative importances of other predictors, regionally. The relative importance of SSS, visibly negligible in the historical and SSP126 scenarios, grows under the SSP585 scenario (e.g. in the Southern Ocean) in four out of five models. This suggests that salinity changes could play an increasing role in driving CO_2 flux IAV in some regions. The relative importance of PO_4 and surface winds also noticeably change under the future scenarios, giving potential indications about the actual mechanisms behind the changes in the drivers of the CO_2 flux IAV, as detailed in the next paragraph.

The decline in the relative importance of DIC in the CO_2 flux IAV under the global warming scenarios is plausibly connected to a reduction in the DIC IAV (Fig. S9 and ref. 33). In particular, once the variability related to precipitation, which affects ALK as much as DIC, is removed, the DIC IAV decreases more (or increases less) than that of ALK in the vast majority of the ocean and notably in the equatorial Pacific, the Southern Ocean, and the North Atlantic (see Figs. S9, S10, and S11). Ref. 75 suggest that because precipitation-induced changes in DIC are buffered by changes in the CO_2 flux, the impact of precipitation is stronger on ALK than on DIC. As summarised in ref. 33, there are several reasons for the decline in DIC IAV. (1) The shoaling of mixed layer depth in global warming scenarios[76] may lead to a weakened variability associated with the mixing of DIC-rich deep waters. (2) Increase or decrease in primary production variability driven by several mechanisms can lead to an increase or decrease in DIC IAV. For instance, mixed layer shoaling and stratification increase can reduce nutrient inputs and, consequently, primary production variability[76,77]. However, mixed layer shoaling in light-limited areas, such as the Southern Ocean, can alleviate this constraint and increase primary production variability[78]. The complex interplay between temperature-induced changes in phytoplanktons growth or zooplankton grazing may also contribute to primary production variability[50]. The growing importance of primary production for future CO_2 flux has been pointed out in the equatorial Pacific[75] and the Southern Ocean[79,80]. In our analysis, the higher relative importance of PO_4, which fuels the primary production, in the Southern Ocean under the SSP585 scenario is consistent with their findings (Figs. S2 to S6). (3) Changes in atmospheric circulation may modulate the upwelling of DIC-rich water. The weaker Walker circulation in the equatorial Pacific may lead to a weaker upwelling during La Niña conditions, and thus reduce the amplitude of the upwelling variability[81,82]. In line with that, our analysis reveals a decline in the relative importance of surface wind in the equatorial Pacific, in the future scenarios, consistent between models (Figs. S2 to S6). Winds are also expected to get stronger in future scenarios [Fig. 4.26 of ref. 83], which could increase the relative importance of wind speed through changes in gas exchange velocities[84]. However, our results do not indicate an increase in wind speed relative importances, suggesting that the increasing wind speed is still less important than DIC and ALK variability in driving the IAV.

The ocean invasion by anthropogenic carbon is another reason for the decline in DIC IAV and its relative importance in the future CO_2 flux IAV. Since anthropogenic carbon mostly resides in the upper ocean, the total DIC concentration there also increases and the vertical gradient of DIC is reduced while the alkalinity gradient tends to sharpen (Fig. S12). Smoother gradient implies a decrease in variability induced by vertical exchanges between surface and subsurface[20]. Such flattening in the vertical DIC gradient occurs nearly everywhere in the ocean across different ESMs, as does

the decline in the relative importance of DIC. Hence, future increase in anthropogenic carbon invasion could play a major role in reducing the importance of DIC in the inter-annual variability of future CO_2 flux in upwelling and ventilation regions where vertical DIC gradient play a strong role in modulating surface DIC variations.

Conclusions and perspectives

The first finding of this study is that the present and future CO_2 flux IAV, resulting from the non-linear interactions between many processes, can be reconstructed with a kernel ridge regression from a limited number of predictors. The two major predictors for this reconstruction are DIC and alkalinity. Depending on the scenario and the region, some second-order predictors also emerge. Furthermore, the clustering approach, used to identify regions sharing the same characteristics, provides a synthesis of the reconstruction analysis. It enables us to assess the physical interpretation of the reconstructions, which are consistent with previous studies across different ocean domain under the contemporary period. We find that the influence of the DIC on CO_2 flux IAV decreases under climate change scenarios, supporting the insights of prior studies[20,75]. This diminution may be connected to the global attenuation of vertical DIC gradients in the water column, among other processes[20,33,79]. Our work depicts the first successful attempt of applying a machine learning technique on a large dataset generated by multiple models and across different scenarios, to analyse the CO_2 flux IAV and its drivers in different ocean domains.

The most obvious caveat of our work is its reliance on the performance of the regression in reconstructing a physically meaningful CO_2 flux IAV. Results in areas where the regression is not as successful should be interpreted with caution, since part of the IAV is not satisfactorily reproduced. Inferring the processes driving the CO_2 flux IAV from the predictors' relative importances is a challenging exercise, as variations in several predictors may interact with each others. For example, cooling at the surface will increase solubility (and thus CO_2 uptake) and may also deepen the mixed layer, increasing surface DIC (and thus decreasing CO_2 uptake). Plus, mixed layer deepening may bring nutrient to the surface, boosting DIC biological consumption (and thus increasing CO_2 uptake). The interaction between cooling, mixed layer deepening and biology would emerge in the relative importances of SST, DIC and PO_4, yet it is difficult to infer these interactions from the relative importances alone. Moreover, our method shares the caveats from many statistical approaches. The reliability depends on the a priori knowledge we have on the drivers of the CO_2 flux IAV and the conclusions on the actual dynamic causal relationships between the CO_2 flux IAV and its drivers and variability are based on correlations. Nevertheless, our work opens up a number of promising perspectives.

Foremost, this work provides guidance for future studies on CO_2 flux IAV. Indeed, our results suggest that alkalinity and DIC are the two most influential drivers for the CO_2 flux IAV. This implies that the poor representation of CO_2 flux IAV by ESMs may be linked to a poor representation of DIC and alkalinity. For instance, CMIP6 ESMs tend to underestimate alkalinity in the upper ocean whilst overestimate in the interior[85]. Besides, since models are equilibrated to a fixed preindustrial atmospheric CO_2 concentration, the respective DIC concentration is biased as well. Notably, our work reveals a potentially increasing role of alkalinity in future CO_2 flux IAV. In regions where CO_2 flux IAV is strong, such as the equatorial Pacific, bias in alkalinity could have strong implications on the future projections[20]. In the wake of prior studies[85], this calls for a better representation of alkalinity in ESMs. Our findings further indicates that accurate representation of vertical gradients are important for projecting the CO_2 flux IAV. The development of future observation strategies should consider long-term measurement of vertical profiles of DIC and alkalinity over inter-annual to decadal timescale (i.e., include contrasting climate regimes of El Niño vs La Niña in the equatorial Pacific, positive vs negative SAM and NAO in the Southern Ocean and North Atlantic, etc.), complemented regionally by temperature, wind or phosphate, to better constrain the importance of each driver for CO_2 flux IAV. This work also provides a new tool for exploring the CO_2 flux IAV by efficiently examining the massive volume of datasets

created by the CMIP exercise. Together with that, by narrowing down the list of variables driving the CO_2 flux IAV, this work can assist in the development of further analysis or ad-hoc simulations to better understand the mechanisms controlling the variability of the CO_2 flux. Future studies could consider extending the predictors list to export production, vertical velocity, surface current as well as salinity-normalised DIC and alkalinity to shed more light on the role of freshwater fluxes and surface ocean circulation. A follow-up study examining in detail the drivers of DIC (and ALK) variability would be valuable. The different physical and biogeochemical processes, such as upwelling, mixing, advection, primary production, freshwater fluxes, CO_2 fluxes, will have to be individually quantified in the different regions and ESMs. Most of them are currently not provided by ESMs and, thus, could be considered in the future CMIP exercise. Finally, by demonstrating that the CO_2 flux IAV can be reconstructed from a limited number of variables using a statistical model, this work is a further step in the development of a new generation of Earth System Models combining physical and statistical models[52].

Methods

Earth Systems Models, scenarios and data
In this work, we use a kernel ridge regression algorithm together with CMIP6 ESMs' outputs (Coupled Model Intercomparison Project Phase 6, ref. [60]). The variables are the air-sea CO_2 flux (CO_2 flux hereafter), as the target, and 8 other variables, as the predictors: dissolved inorganic carbon (DIC), alkalinity (ALK), temperature (SST), salinity (SSS), phosphate (PO_4), surface wind speed (sfcWind), sea-ice concentration (siconc) and atmospheric CO_2 concentration (atmCO2). The variable selection is constrained by data availability at the time the study was conducted, but also based on variables usually used for CO_2 flux calculations in ESMs. It should be noted that DIC and ALK are not salinity normalised. Monthly outputs, and surface fields are used, from 5 ESMs (ACCESS-ESM1-5, CESM2, IPSL-CM6A-LR, NorESM2-LM, MPI-ESM1-2-LR) following 4 scenarios (preindustrial, historical, SSP126 and SSP585), i.e., $5 \times 4 = 20$ simulations in total. The SSP126 and SSP585 scenarios are in the lower and upper range of global warming scenarios, respectively. For all ESMs but CESM2, the r1i1p1f1 variant is used for the historical and global warming scenarios. For CESM2, the r10i1p1f1 variant is used.

Kernel ridge regression and relative importances of predictors
We use a non-parametric regression to model the single response variable (the CO_2 flux) as a function of the eight predictors. Machine learning models (later referred to as ML-models) are derived for each geographic location separately, i.e., the ML-models aim at capturing the temporal behaviour only. The ocean surface is resolved by 1x1-degree, 2x2-degree, or 4x4-degree grids (axes parallel to longitude/latitude). Investigations not documented here have shown that the dependencies are smooth. Since the time series were short (either 1032, 1980 or 2400 data points, representing the number of months in the future scenarios, historical, and preindustrial control simulations), it was possible to use (non-scalable) kernel methods, like kernel ridge regression. Data scientific investigations have shown that smooth regression schemes were superior to less regular ones. We, therefore, chose the kernel ridge regression with the smooth RBF kernel, discarding the possibilities of e.g. kernel ridge regression with Laplacian kernel, k-nearest neighbours regression, or Random Forest regression, see also ref. [86]. The relative importances of the predictors were assessed as normalised permutation feature importances[87]. To assess uncertainties of the statistical quantities (train and test accuracy, feature importances), we employ simple non-parametric bootstrapping with 50 resamples[88]. All assessments are based on a modified R^2 score, focusing on the inter-annual variability.

In the sequel, we first define the modified R^2 score. After that, we discuss our use of the kernel ridge regression, followed by a discussion of permutation feature importance. Finally, we give details on the bootstrap procedure used.

The modified R² score
We focus on inter-annual variabilities in our investigations. To this end, we split the signal into the trend, the seasonal

signal, and the residual, and the latter represents the inter-annual variabilities. Let $n > 0$ and $\mathbf{t}, \mathbf{x} \in \mathbb{R}^n$. Then we denote by $\mathbf{s} = (\mathbf{t}, \mathbf{x})$ a CO_2 flux time series of length n. Let:

$$\mathbf{s}^{\text{trend}} := \text{spline}(\mathbf{t}, \mathbf{x}, k = 3), \tag{1}$$

$$\mathbf{s}^{\text{seasonal}} := (\mathbf{t}, \mathbf{x}^{\text{se}}), x_i^{\text{se}} := \text{avg}\{x_{i+12k}^{\text{trend}}, |i + 12k \in [0, n[, k \in \mathbb{Z}\}, \tag{2}$$

$$\mathbf{s}^{\text{residual}} := \mathbf{s} - \mathbf{s}^{\text{trend}} - \mathbf{s}^{\text{seasonal}} \tag{3}$$

where *spline* is a spline interpolation of degree three ($k = 3$, i.e. cubic spline), evaluated at $\mathbf{t}$. We use the implementation *UnivariateSpline* from *scipy*[89]. The resulting trend is very smooth (see Fig. S13) ensuring that $\mathbf{s}^{\text{residual}}$ retains the inter-annual variability signal, from year-to-year to decadal variation. For simplicity of writing, we use an intuitive notion for the components of $\mathbf{t}$ and $\mathbf{x}$, as well as for addition and subtraction of time series. The modified R^2 score for a variable $\mathbf{s}$ (here the CO_2 flux time series from one EMS simulation) approximated by $\tilde{\mathbf{s}}$ (here the CO_2 flux time series reconstructed by the regression) is defined as:

$$R^{2,\text{residual}}(\mathbf{s}, \tilde{\mathbf{s}}) := R^2(\mathbf{s} - \mathbf{s}^{\text{trend}} - \mathbf{s}^{\text{seasonal}}, \tilde{\mathbf{s}} - \mathbf{s}^{\text{trend}} - \mathbf{s}^{\text{seasonal}}), \tag{4}$$

where $R^2(\mathbf{s}, \tilde{\mathbf{s}})$ is used as intuitive writing for $R^2(\mathbf{x}, \tilde{\mathbf{x}})$ and similar for $R^{2,\text{residual}}$. In essence, $R^{2,\text{residual}}$ measures the explained variance of the residual (i.e. inter-annual) signal $\mathbf{x}^{\text{residual}}$.

The kernel ridge regression
We use the kernel ridge regression, which combines the ridge regression with the kernel trick. The ridge regression is a linear least squares fitting with l^2-norm regularisation. The regularisation is important in the case of highly correlated input variables, which cannot be excluded for the regression problems we are dealing with. For non-linear kernels, the regression scheme becomes a non-parametric, non-linear regression. We employ the 'rbf'-kernel, which establishes an approximation in an infinitely high dimensional smooth space. The method has two parameters, α and γ. α is a regularisation parameter and needs to be determined experimentally (hyperparameter fitting) to obtain the best regression results. The γ is a squared inverse length scale and can be derived from the data. For details, we refer to ref. [86].

We use the implementation of the kernel ridge regression provided as the module *KernelRidge*[90]. The predictors and the response variable are normalised to zero mean and unit variance. The two parameters are set to:

$$\alpha = 2 \cdot 10^{-4} \tag{5}$$

$$\gamma = h^{-2}, h = 6 \, p_{80}(d), d := \{\| X_i - X_j \|_2 | i, j \in [0, n[, i \neq j\}, \tag{6}$$

where $X_i \in \mathbb{R}^8, i \in [0, n[$ denotes the vector of the (normalised) values of all predictor variables at time t_i and $p_{80}(Q)$ denotes the 80-th percentile of a set d of real numbers. The choices for α and γ have been determined experimentally. Note that the value for γ is relatively small, indicating a rather slowly varying behaviour of the response variable. As the approximation is good and better if compared with non-smooth regression schemes (like random forest or kernel ridge regression with Laplacian kernel), we can also conclude that the response variable is also smooth depending on the predictors. The value of α is small, indicating that regularisation due to highly correlated input variables is of minor importance.

Permutation feature importance
We determine the importances of the predictor variables in modelling the response variable using permutation feature importance. We use a modified version of permutation feature importance due to our focus on inter-annual variability. To explain our procedure, we first recall the standard definition of permutation feature importance. Let $\mathbf{X} := (\mathbf{x_1}, \ldots, \mathbf{x_8})$ be the collection of the eight values

vectors of the eight predictor variables and $\mathbf{X_i} := (\mathbf{x_1}, \ldots, \tilde{\mathbf{x}}_i, \ldots, \mathbf{x_8})$ the same collection, where the i-th variable shuffled, indicated by $\tilde{\mathbf{x}}_i$. Then, the permutation feature importance of the i-th predictor variable, derived from R^2 score, is defined by:

$$\mathit{fi}_i := R^2(\mathbf{y}, f(\mathbf{X})) - R^2(\mathbf{y}, f(\mathbf{X_i})) \tag{7}$$

where $f(\cdot)$ denotes the regression function obtained, in our case, from the kernel ridge regression modelling $\mathbf{y}$ by $\mathbf{X}$. To estimate the importance of predictor variables in modelling the inter-annual variabilities, we modify the shuffling of the data $\mathbf{x_i}$ and apply the modified R^2-score, leading to our definition of the absolute feature importance RI_i^a. Let $i \in [0, 8[$ then:

$$\tilde{\mathbf{x}}_i := \text{shuffle}(\mathbf{x_i^{residual}}) + \mathbf{x_i^{trend}} + \mathbf{x_i^{seasonal}}, \tag{8}$$

$$\mathbf{X_i} := (\mathbf{x_1}, \ldots, \tilde{\mathbf{x}}_i, \ldots, \mathbf{x_8}), \tag{9}$$

$$\mathit{fi}_i := R^{2,\text{residual}}(\mathbf{y}, f(\mathbf{X})) - R^{2,\text{residual}}(\mathbf{y}, f(\mathbf{X_i})). \tag{10}$$

The relative importance of the i-th predictor RI_i is defined as the normalisation (scaling) of the absolute feature importance (fi_i) to have a unit sum.

Bootstrapping. We use simple non-parametric bootstrapping with 50 resamples of each time series. Accordingly, 50 regressions are carried out per time series, and every time series is split into a train and a test set, where the size of the test set is ~36.8% of the length of the time series, and the training set contains about 63.2% unique values / 36.8% duplicates. Consequently, across the bootstrap samples, each point in time is, on average, $50 \times 63.2\% = 31.6$ times part of the training set and $50 \times 36.8\% = 18.4$ times part of the test set. This split is used to estimate the average training prediction and its standard deviation, as well as the average test prediction and its standard deviation. The estimation of the average $R^{2,\text{residual}}$ score and its standard deviation is obtained by the simple non-parametric bootstrapping as well, and similarly, the estimation of the relative feature importance.

Clustering and classification
Dimension reduction : t-stochastic neighbour embedding (t-SNE).
Defining similar regions from 8-dimension space is not necessarily straightforward with classical clustering techniques such as K-means, especially when the 8 relative importances (RI) of the predictors have (a priori) no reason to be Gaussian[56,91]. Originally created for high-dimensional data visualisation[92], t-SNE algorithm provides the intuition of the data feature proximity in high-dimensional space in a lower-dimensional representation, i.e., a dimension reduction preserving the topology (local and global structure) of the high-dimensional data. In our case, first, the grid-points with sea-ice concentration RI greater than 0.1 are removed. Then, t-SNE maps each 6-dimensional (atmCO2 is excluded because of the very small RI as well as siconc) object onto a point in 3D phase space and ensures a high probability of similar objects remaining close in both the high- and low-dimensional space. Given a set of N 6-dimensional elements (grid-points) $x_1, \cdots, x_N$, the t-SNE algorithm performs a reduction by minimising the Kullback-Leibler (KL) divergence[93]. The KL divergence is a measure of the dissimilarity of two probability distribution functions, which in this study assess how well the statistical properties are preserved from high to low dimension. For two objects x_i and x_j the high-dimensional space and y_i and y_j their low-dimensional counterparts, t-SNE defines the pairwise similarity probabilities, p_{ij} as:

$$p_{ij} = \frac{\exp(-||x_i - x_j||^2 / 2\sigma_i^2)}{\sum_{k \neq l} \exp(-||x_k - x_l||^2 / 2\sigma_i^2)} \tag{11}$$

σ_i is the variance of the Gaussian that is centred on data-point x_i. A Student t-distribution with one degree of freedom is used as the heavy-tailed distribution in the low-dimensional map. Using this distribution, the joint probabilities q_{ij} are defined as:

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq l}(1 + ||y_k - y_l||^2)^{-1}} \tag{12}$$

Thus, the Kullback-Leibler divergence between a joint probability distribution, P, in the high-dimensional space and a joint probability distribution, Q, in the low-dimensional space is:

$$KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{13}$$

The t-SNE is performed using the implementation of R package 'Rtsne'[94] available in the free software R[95].

Model-based clustering. Our approach consists in clustering all the grid-points y in the low-dimensional space for each ESM regrouping preindustrial, historical, SSP126 and SSP585 scenarios. It is based on the Expectation-Maximization (EM) algorithm[96]. This EM clustering algorithm fits a multivariate probability density function (pdf), f, to our data in order to classify the grid-points. The multivariate pdf is a weighted sum of K Gaussian pdfs f_k $(k = 1, \cdots, K)$ or Gaussian Mixture Model[97,98] (GMM):

$$f(y) = \sum_{k=1}^{K} \pi_k f_k(y; \alpha_k) \tag{14}$$

where α_k contains the parameters (means μ_k and covariance matrix Σ_k) of f_k and π_k is the mixture ratio corresponding to the prior probability that y (i.e., a grid-point value) belongs to f_k. The parameters α_k and π_k $(k = 1, \ldots, K)$ of the GMM are to be estimated. The estimation is performed using the R package 'Mclust'[99].

The parameters of the GMM are the means μ_k, covariance matrix Σ_k, and mixture ratio π_k, describing the K Gaussian distributions. Their estimation is performed through iterative the Expectation Maximization (EM) algorithm by maximising the likelihood[100]. GMM parameters are initialised by the result of a hierarchical model-based agglomerative clustering. Thus, local maxima are avoided when optimising the likelihood function (e.g., ref. 99). EM is based on the possibility to calculate π when knowing α and vice-versa, thus optimising both. After the initialisation, the Expectation-step (or E-step) estimates the posterior probability τ_k^i (update of π_k^i) that the grid-point y_g belongs to f_k with the current parameter estimates (at iteration i):

$$\tau_k^i(y_g) = \frac{\pi_k^i f_k(y_g, \alpha_k^i)}{\sum_{k=1}^{K} \pi_k^i f_k(y_g, \alpha_k^i)} \tag{15}$$

Then, the Maximization-step (or M-step) uses the posterior probabilities to improve the estimates of GMM parameters (iteration $i + 1$):

$$\pi_k^{i+1} = \frac{1}{n} \sum_{g=1}^{n} \tau_k^i(y_g) \tag{16}$$

$$\mu_k^{i+1} = \frac{1}{n \pi_k^{i+1}} \sum_{g=1}^{n} x_g \tau_k^i(y_g) \tag{17}$$

$$\Sigma_k^{i+1} = \frac{1}{n \pi_k^{i+1}} \sum_{g=1}^{n} \tau_k^i(y_g) \left(y_g - \mu_k^{i+1}\right)' \left(y_g - \mu_k^{i+1}\right) \tag{18}$$

where n is the number of grid-point. The algorithm repeats the E- and M-steps iteratively until termination when model parameters converge and the maximum likelihood is reached (convergence of the log-likelihood function) or after a maximum number of iterations.

Finally, each cluster C_k of grid-points is defined based on the Gaussian pdfs, according to the principle of posterior maximum:

$$C_k = \{y; \pi_k f_k(y; \alpha_k) \geq \pi_j f_j(y; \alpha_j), \forall j = 1, \ldots K\} \quad (19)$$

In other words, each grid-point is assigned to the cluster for which the probability of belonging is maximum. The freedom of EM in the definition of the biomes depends on the number K of clusters and on the constraints applied to the covariance matrices.

Different values of K (from $K = 1$ to $K = 125$) have been evaluated through the Bayesian Information Criterion[101] (BIC). Optimising the BIC achieves a compromise between a good likelihood and a reasonable number of parameters.

The Bayesian Information Criterion (BIC) is a criterion for model selection that helps to prevent overfitting by introducing penalty terms for the complexity of the model (number of parameters). In our case, minimising the BIC achieves a good compromise between keeping the model simple and a good representation of the observed data.

$$BIC(K) = p \log(n) - 2 \log(L) \quad (20)$$

where K is the number of clusters, L the likelihood of the parameterised mixture model, p the number of parameters to estimate, and n the size of the sample (around 200 thousand grid-points).

Tree based classification. Depending on the ESM, the optimal number of clusters varies from 80 to 105 which is too numerous to be interpreted. One way to regroup them to be interpreted is to build a decision tree, that is, a classification method in the form of a tree structure separating a dataset into smaller and smaller subsets. Starting from a known partition (obtained from the GMM clustering), the method predicts the class to which a grid-point belongs, following some decision rules based on the predictors' relative importances. For this purpose, R package "rpart"[102] has been used, based mainly on ref. 103 and on binary trees: each node has at most two children. Each new separation according to a decision rule between the nodes have been performed via maximal impurity reduction, with the use of the Gini index as an impurity function. That means that the tree tries to build nodes containing as few clusters as possible from the reference partition. The trees are grown to a certain level of complexity, then pruned to nine terminal nodes. The pruning is done by setting the complexity parameter, measuring the "cost" of adding another explanatory variable among the 8 possible ones in the decision tree. For more technical details, see ref. 102.

Data availability
CMIP6 outputs are available from the Earth System Grid Federation (ESGF) portals (e.g. https://esgf-node.ipsl.upmc.fr). The inputs and outputs of the kernel ridge regression are available at https://doi.org/10.11582/2023.00017[104].

Code availability
The code for the kernel ridge regression analysis are available at https://doi.org/10.11582/2023.00017[104]. The codes for the clustering analysis are described at https://cran.r-project.org/web/packages/Rtsne/index.html [last access 7-dec-2022] (t-SNE), https://cran.r-project.org/web/packages/mclust/index.html [last access 7-dec-2022] (model-based clustering GMM) and https://cran.r-project.org/web/packages/rpart/index.html [last access 7-dec-2022] (classification tree). The code for producing the figure is available at https://github.com/damiencouespel/article_kr_regression_co2_flux_cmip6_diagnostics[105].

References
1. Friedlingstein, P. et al. Global Carbon Budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005 (2022).
2. McKinley, G. A., Fay, A. R., Lovenduski, N. S. & Pilcher, D. J. Natural variability and anthropogenic trends in the ocean carbon sink. *Ann. Rev. Mar. Sci.* **9**, 125–150 (2017).
3. Schwinger, J. et al. Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1). *Geosci. Model Dev.* **9**, 2589–2622 (2016).
4. Landschützer, P., Gruber, N., Bakker, D. C. E. & Schuster, U. Recent variability of the global ocean carbon sink. *Glob. Biogeochem. Cycles* **28**, 927–949 (2014).
5. Hauck, J. et al. Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Front. Mar. Sci.* **7**, 571720 (2020).
6. McKinley, G. A., Bennington, V., Meinshausen, M. & Nicholls, Z. Modern air-sea flux distributions reduce uncertainty in the future ocean carbon sink. *Environ. Res. Lett.* **18**, 044011 (2023).
7. Tjiputra, J. F., Negrel, J. & Olsen, A. Early detection of anthropogenic climate change signals in the ocean interior. *Sci. Rep.* **13**, 3006 (2023).
8. Fay, A. R. & McKinley, G. A. Global trends in surface ocean pCO_2 from in situ data. *Glob. Biogeochem. Cycles* **27**, 541–557 (2013).
9. Landschützer, P., Ilyina, T. & Lovenduski, N. S. Detecting regional modes of variability in observation-based surface ocean pCO_2. *Geophys. Res. Lett.* **46**, 2670–2679 (2019).
10. Landschützer, P., Gruber, N. & Bakker, D. C. E. Decadal variations and trends of the global ocean carbon sink. *Glob. Biogeochem. Cycles* **30**, 1396–1417 (2016).
11. Betts, R. A. et al. ENSO and the Carbon Cycle. In *El Niño Southern Oscillation in a Changing Climate*. Ch. 20, 453–470 (American Geophysical Union (AGU), 2020).
12. Resplandy, L., Séférian, R. & Bopp, L. Natural variability of CO_2 and O_2 fluxes: What can we learn from centuries-long climate models simulations? *J. Geophys. Res. Oceans* **120**, 384–404 (2015).
13. Gruber, N. et al. Trends and variability in the ocean carbon sink. *Nat. Rev. Earth Environ.* **4**, 119–134 (2023).
14. Gruber, N., Landschützer, P. & Lovenduski, N. S. The variable southern ocean carbon sink. *Annu. Rev. Mar. Sci.* **11**, 159–186 (2019).
15. Rödenbeck, C. et al. Interannual sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences* **11**, 4599–4613 (2014).
16. Le Quéré, C., Orr, J. C., Monfray, P., Aumont, O. & Madec, G. Interannual variability of the oceanic sink of CO_2 from 1979 through 1997. *Glob. Biogeochem. Cycles* **14**, 1247–1265 (2000).
17. Landschützer, P. et al. The reinvigoration of the Southern Ocean carbon sink. *Science* **349**, 1221–1224 (2015).
18. Le Quéré, C. et al. Saturation of the southern ocean CO_2 sink due to recent climate change. *Science* https://doi.org/10.1126/science.1136188 (2007).
19. McKinley, G. A., Rödenbeck, C., Gloor, M., Houweling, S. & Heimann, M. Pacific dominance to global air-sea CO_2 flux variability: A novel atmospheric inversion agrees with ocean models. *Geophys. Res. Lett.* https://doi.org/10.1029/2004GL021069 (2004).
20. Vaittinada Ayar, P. et al. Contrasting projections of the ENSO-driven CO_2 flux variability in the equatorial Pacific under high-warming scenario. *Earth Syst. Dyn.* **13**, 1097–1118 (2022).
21. Wong, S. C.-K., McKinley, G. A. & Seager, R. Equatorial Pacific pCO_2 Interannual Variability in CMIP6 Models. *J. Geophys. Res. Biogeosciences* **127**, e2022JG007243 (2022)

22. Lovenduski, N. S., Gruber, N., Doney, S. C. & Lima, I. D. Enhanced CO2 outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode. *Glob. Biogeochem. Cycles* https://doi.org/10.1029/2006GB002900 (2007).

23. Lenton, A. & Matear, R. J. Role of the Southern Annular Mode (SAM) in Southern Ocean CO2 uptake. *Glob. Biogeochem. Cycles* https://doi.org/10.1029/2006GB002714 (2007).

24. Wetzel, P., Winguth, A. & Maier-Reimer, E. Sea-to-air CO2 flux from 1948 to 2003: a model study. *Glob. Biogeochem. Cycles* https://doi.org/10.1029/2004GB002339 (2005).

25. DeVries, T., Holzer, M. & Primeau, F. Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature* **542**, 215–218 (2017).

26. Keppler, L. & Landschützer, P. Regional wind variability modulates the southern ocean carbon sink. *Sci. Rep.* **9**, 7384 (2019).

27. Ullman, D. J., McKinley, G. A., Bennington, V. & Dutkiewicz, S. Trends in the North Atlantic carbon sink: 1992–2006. *Glob. Biogeochem. Cycles* https://doi.org/10.1029/2008GB003383 (2009).

28. Thomas, H. et al. Changes in the North Atlantic Oscillation influence CO2 uptake in the North Atlantic over the past 2 decades. *Glob. Biogeochem. Cycles* https://doi.org/10.1029/2007GB003167 (2008).

29. Schuster, U. & Watson, A. J. A variable and decreasing sink for atmospheric CO2 in the North Atlantic. *J. Geophys. Res. Oceans* https://doi.org/10.1029/2006JC003941 (2007).

30. Breeden, M. L. & McKinley, G. A. Climate impacts on multidecadal pCO_2 variability in the North Atlantic: 1948–2009. *Biogeosciences* **13**, 3387–3396 (2016).

31. Koseki, S., Tjiputra, J., Fransner, F., Crespo, L. R. & Keenlyside, N. S. Disentangling the impact of Atlantic Niño on sea-air CO2 flux. *Nat. Commun.* **14**, 3649 (2023).

32. McKinley, G. A. et al. North Pacific carbon cycle response to climate variability on seasonal to decadal timescales. *J. Geophys. Res. Oceans* https://doi.org/10.1029/2005JC003173 (2006).

33. Gallego, M. A., Timmermann, A., Friedrich, T. & Zeebe, R. E. Anthropogenic Intensification of Surface Ocean Interannual pCO2 Variability. *Geophys. Res. Lett.* **47**, e2020GL087104 (2020).

34. Li, H. & Ilyina, T. Current and future decadal trends in the oceanic carbon uptake are dominated by internal variability. *Geophys. Res. Lett.* **45**, 916–925 (2018).

35. McKinley, G. A. et al. Timescales for detection of trends in the ocean carbon sink. *Nature* **530**, 469–472 (2016).

36. Lovenduski, N. S., McKinley, G. A., Fay, A. R., Lindsay, K. & Long, M. C. Partitioning uncertainty in ocean carbon uptake projections: Internal variability, emission scenario, and model structure. *Glob. Biogeochem. Cycles* **30**, 1276–1287 (2016).

37. DeVries, T. et al. Decadal trends in the ocean carbon sink. *Proc. Natl Acad. Sci. USA* **116**, 11646–11651 (2019).

38. Hauck, J. et al. Sparse observations induce large biases in estimates of the global ocean CO2 sink: An ocean model subsampling experiment. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **381**, 20220063 (2023).

39. Gloege, L. et al. Quantifying errors in observationally based estimates of ocean carbon sink variability. *Glob. Biogeochem. Cycles* **35**, e2020GB006788 (2021).

40. Séférian, R., Bopp, L., Swingedouw, D. & Servonnat, J. Dynamical and biogeochemical control on the decadal variability of ocean carbon fluxes. *Earth Syst. Dyn.* **4**, 109–127 (2013).

41. Tjiputra, J. F., Olsen, A., Assmann, K., Pfeil, B. & Heinze, C. A model study of the seasonal and long–term North Atlantic surface pCO_2 variability. *Biogeosciences* **9**, 907–923 (2012).

42. Jin, C., Zhou, T. & Chen, X. Can CMIP5 earth system models reproduce the interannual variability of air–sea CO2 fluxes over the tropical Pacific Ocean? *J. Clim.* **32**, 2261–2275 (2019).

43. Doney, S. C. et al. Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea CO2 fluxes: physical climate and atmospheric dust. *Deep Sea Res. Part II: Top. Stud. Oceanogr.* **56**, 640–655 (2009).

44. Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W. & Sutherland, S. C. Seasonal variation of CO2 and nutrients in the high-latitude surface oceans: a comparative study. *Glob. Biogeochem. Cycles* **7**, 843–878 (1993).

45. Schwinger, J. et al. Nonlinearity of ocean carbon cycle feedbacks in CMIP5 earth system models. *J. Clim.* **27**, 3869–3888 (2014).

46. Tjiputra, J. F. & Winguth, A. M. E. Sensitivity of sea-to-air CO_2 flux to ecosystem parameters from an adjoint model. *Biogeosciences* **5**, 615–630 (2008).

47. Arora, V. K. et al. Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences* **17**, 4173–4222 (2020).

48. Bahl, A., Gnanadesikan, A. & Pradal, M.-a. S. Scaling global warming impacts on ocean ecosystems: Lessons from a suite of earth system models. *Front. Mar. Sci.* **7**, 698 (2020).

49. Séférian, R. et al. Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Curr. Clim. Change Rep.* **6**, 95–119 (2020).

50. Laufkötter, C. et al. Drivers and uncertainties of future global marine primary production in marine ecosystem models. *Biogeosciences* **12**, 6955–6984 (2015).

51. Sonnewald, M. et al. Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environ. Res. Lett.* **16**, 073008 (2021).

52. Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Machine Intell.* **3**, 667–674 (2021).

53. Martinez, E. et al. Reconstructing global chlorophyll-a variations using a non-linear statistical approach. *Front. Mar. Sci.* **7**, 464 (2020).

54. Krasting, J. P., De Palma, M., Sonnewald, M., Dunne, J. P. & John, J. G. Regional sensitivity patterns of Arctic Ocean acidification revealed with machine learning. *Commun. Earth Environ.* **3**, 1–11 (2022).

55. Jones, D. C. & Ito, T. Gaussian mixture modeling describes the geography of the surface ocean carbon budget. In *Proceedings of the 9th International Workshop on Climate Informatics: CI2019* (eds. Brajard, J., Charantonis, A., Chen, C., & Runge, J.) 6 (2019).

56. Sonnewald, M., Wunsch, C. & Heimbach, P. Unsupervised learning reveals geography of global ocean dynamical regions. *Earth Space Sci.* **6**, 784–794 (2019).

57. Holder, C. & Gnanadesikan, A. Can machine learning extract the mechanisms controlling phytoplankton growth from large-scale observations? – A proof-of-concept study. *Biogeosciences* **18**, 1941–1970 (2021).

58. D'Alelio, D. et al. Machine learning identifies a strong association between warming and reduced primary productivity in an oligotrophic ocean gyre. *Sci. Rep.* **10**, 3287 (2020).

59. Bennington, V., Galjanic, T. & McKinley, G. A. Explicit physical knowledge in machine learning for ocean carbon flux reconstruction: the pCO2-residual method. *J. Adv. Model. Earth Syst.* **14**, e2021MS002960 (2022).

60. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

61. Jones, D. C., Ito, T., Takano, Y. & Hsu, W.-C. Spatial and seasonal variability of the air-sea equilibration timescale of carbon dioxide. *Glob. Biogeochem. Cycles* **28**, 1163–1178 (2014).

62. Zhong, G. et al. Reconstruction of global surface ocean pCO_2 using region-specific predictors based on a stepwise FFNN regression algorithm. *Biogeosciences* **19**, 845–859 (2022).

63. Gallego, M. A., Timmermann, A., Friedrich, T. & Zeebe, R. E. Drivers of future seasonal cycle changes in oceanic pCO_2. *Biogeosciences* **15**, 5315–5327 (2018).

64. Rodgers, K. B. et al. Seasonal variability of the surface ocean carbon cycle: a synthesis. *Glob. Biogeochem. Cycles* **37**, e2023GB007798 (2023).

65. Roach, L. A. et al. Antarctic sea ice area in CMIP6. *Geophys. Res. Lett.* **47**, e2019GL086729 (2020).

66. Notz, D. & Community, S. Arctic sea ice in CMIP6. *Geophys. Res. Lett.* **47**, e2019GL086749 (2020).

67. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: mean and temporal variability. *Earth Syst. Sci. Data* **6**, 273–284 (2014).

68. Lauvset, S. K., Gruber, N., Landschützer, P., Olsen, A. & Tjiputra, J. Trends and drivers in global surface ocean pH over the past 3 decades. *Biogeosciences* **12**, 1285–1298 (2015).

69. Gregor, L., Kok, S. & Monteiro, P. M. S. Interannual drivers of the seasonal cycle of CO_2 in the Southern Ocean. *Biogeosciences* **15**, 2361–2378 (2018).

70. Henson, S. A. et al. Controls on Open-Ocean North Atlantic ΔpCO2 at Seasonal and Interannual Time Scales Are Different. *Geophys. Res. Lett.* **45**, 9067–9076 (2018).

71. Corbière, A., Metzl, N., Réverdin, G., Brunet, C. & Takahashi, T. Interannual and decadal variability of the oceanic carbon sink in the North Atlantic subpolar gyre. *Tellus Ser. B-Chem. Phys. Meteorol.* **59**, 168–178 (2007).

72. Fröb, F. et al. Wintertime f CO 2 variability in the subpolar north atlantic since 2004. *Geophys. Res. Lett.* **46**, 1580–1590 (2019).

73. Halloran, P. R. et al. The mechanisms of North Atlantic CO2 uptake in a large Earth System Model ensemble. *Biogeosciences* **12**, 4497–4508 (2015).

74. Metzl, N. et al. Recent acceleration of the sea surface fCO2 growth rate in the North Atlantic subpolar gyre (1993–2008) revealed by winter observations. *Glob. Biogeochem. Cycles* **24**, GB4004 (2010).

75. Liao, E., Resplandy, L., Liu, J. & Bowman, K. W. Future weakening of the ENSO ocean carbon buffer under anthropogenic forcing. *Geophys. Res. Lett.* **48**, e2021GL094021 (2021).

76. Kwiatkowski, L. et al. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences* **17**, 3439–3470 (2020).

77. Steinacher, M. et al. Projected 21st century decrease in marine productivity: a multi-model analysis. *Biogeosciences* **7**, 979–1005 (2010).

78. Doney, S. C. Oceanography: plankton in a warmer world. *Nature* **444**, 695–696 (2006).

79. Hauck, J. et al. On the Southern Ocean CO 2uptake and the role of the biological carbon pump in the 21st century. *Glob. Biogeochem. Cycles* **29**, 1451–1470 (2015).

80. Hauck, J. & Völker, C. Rising atmospheric CO2 leads to large impact of biology on Southern Ocean CO2 uptake via changes of the Revelle factor. *Geophys. Res. Lett.* **42**, 1459–1464 (2015).

81. Zhao, X. & Allen, R. J. Strengthening of the Walker Circulation in recent decades and the role of natural sea surface temperature variability. *Environ. Res. Commun.* **1**, 021003 (2019).

82. Vecchi, G. A. et al. Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature* **441**, 73–76 (2006).

83. Lee, J.-Y. et al. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 553–672 (eds. Masson-Delmotte, V. et al.) (Cambridge University Press, Cambridge, 2021).

84. Wanninkhof, R. & Triñanes, J. The impact of changing wind speeds on gas transfer and its effect on global air-sea CO2 fluxes. *Glob. Biogeochem. Cycles* **31**, 961–974 (2017).

85. Planchat, A. et al. The representation of alkalinity and the carbonate pump from CMIP5 to CMIP6 Earth system models and implications for the carbon cycle. *Biogeosciences* **20**, 1195–1257 (2023).

86. Kung, S. Y. *Kernel Methods and Machine Learning* (Cambridge University Press, 2014).

87. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

88. Efron, B. & Tibshirani, R. J. An Introduction to the Bootstrap. No. 57 in *Monographs on Statistics and Applied Probability* (Chapman & Hall/ CRC, 1993).

89. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261–272 (2020).

90. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

91. Maze, G. et al. Coherent heat patterns revealed by unsupervised classification of argo temperature profiles in the north atlantic ocean. *Prog. Oceanogr.* **151**, 275–292 (2017).

92. van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

93. Ghosh, S. et al. Letters to the editor. *Am. Stat.* **41**, 338–341 (1987).

94. Krijthe, J. H. *Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation*. https://github.com/jkrijthe/Rtsne. R package version 0.15. (2015).

95. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020). https://www.R-project.org/.

96. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc.: Ser. B (Methodol.)* **39**, 1–22 (1977).

97. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185**, 71–110 (1894).

98. Peel, D. & McLachlan, G. J. Robust mixture modelling using the t distribution. *Stat. Comput.* **10**, 339–348 (2000).

99. Scrucca, L. & Raftery, A. E. Improved initialisation of model-based clustering using gaussian hierarchical partitions. *Adv. Data Anal. Classif.* **9**, 447–460 (2015).

100. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **97**, 611–631 (2002).

101. Schwarz, G. et al. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).

102. Therneau, T. & Atkinson, B. *rpart: Recursive Partitioning and Regression Trees* https://CRAN.R-project.org/package=rpart. R package version 4.1-15. (2019).

103. Breiman, L., Friedman, J., Olshen, R. & Stone, C. *Classification and Regression Trees*. p. 368 (The Wadsworth and Brooks-Cole statistics-probability series, The Wadsworth statistics/probability series, Wadsworth and Brooks, 1984).

104. Johannsen, K. & NORCE Norwegian Research Center AS. Regime shifts in future ocean CO2 fluxes revealed through machine learning [Dataset]. Norstore. https://doi.org/10.11582/2023.00017 (2023).

105. Couespel, D. Machine learning reveals regime shifts in future ocean carbon dioxide fluxes inter-annual variability: Jupyter notebooks [Software]. Zenodo. https://doi.org/10.5281/zenodo.10514490. Zenodo (2024).

Acknowledgements

Author contributions

Competing interests

The authors declare no competing interests.

Additional information