Supplementary Materials for:
Tracer-based Rapid Anthropogenic Carbon Estimation (TRACE)

Brendan R. Carter[1], Jörg Schwinger[3], Rolf Sonnerup[2], Andrea J. Fassbender[1], Jonathan D. Sharp[2,1], Larissa M. Dias[2,1]

[1]NOAA/OAR Pacific Marine Environmental Laboratory, Seattle, WA, USA

[2]Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, WA, 98115

[3]Norwegian Research Center, University of Bergen, Nygårdsgaten 112, 5008 Bergen, Norway

1 Supplementary Text S1: Example application that would benefit from TRACEv1

The Empirical Seawater Property Estimation Routines (ESPERs) (Carter et al., 2021a) provide an example of an application that needs computationally efficient and time-varying C_{anth} estimates in the absence of transient tracer measurements. A subset of these routines is trained from reference-year-adjusted DIC and produce DIC and pH estimates that account for both the reference year conditions and the C_{anth} accumulation since the reference year. The C_{anth} estimation strategy currently employed in these routines is both computationally inefficient when returning a small number of values (requiring up to a minute for the first C_{anth} estimate) and becomes highly unreliable for generating plausible estimates beyond a narrow date range. The accuracy of the C_{anth} adjustments used by ESPER is also unassessed and presumed to be low.

2 Supplementary Test S2: Construction of the neural network committees

TRACEv1 relies on five sets of neutral networks that are formulated using identical architecture to the ESPER_NN (Empirical Seawater Property Estimation Routines with Neural Networks) routines of Carter et al. (2021a). Each "neural network" is, more accurately, two committees of four feed-forward neural-networks each with up to two hidden layers, with the following number of neurons per layer: [20 -], [15 5], [10 10], and [5 15]. Neural network training is accomplished with a Levenberg-Marquardt backpropagation (using the train.m MATLAB function) with default settings: 70% of the data are used for training and 15%, each, used for testing and validation. One committee is used in the Atlantic and the Arctic, with a second committee being used in the rest of the ocean. When transitioning between the two committees in the South Atlantic and near the Bering Strait, a weighted average is taken between the committee results with the transitions in the weights occurring linearly over ~8° of latitude.

All neural networks have only a single output. The results from all members of the appropriate committee are averaged for the final neural network result. All neural networks are trained with the following predictor information: user-provided longitude is converted within the routines to the cosine of the longitude expressed in positive °E after subtracting 20 °E (to align the minimum sensitivity with the narrow band of ocean south of Africa) and the sine of the Longitude expressed in positive °E minus 20 °E (to align the minimum sensitivity with the transition between the Indian and Pacific Oceans), *S* provided as a unitless number, and temperature is expressed in °C which is converted by TRACEv1 routines to potential temperature. While *S* is predominantly available in global measurement repositories as a unitless number, supplying TRACEv1 with the modern convention of absolute salinities expressed as parts per thousand of the mass of seawater will not meaningfully change the estimated C_{anth}. The rationales behind these neural network choices are similar to those given for the nearly-identical choices used by Carter et al. (2021) and the related choices used earlier by Bittig et al. (2018) for the CANYON-B routines that inspired ESPER_NN.

We note that the neural network for α is fit to $\log_{10}|\alpha|$ instead of directly to α because small errors in small values of α have a comparable impact on the TTD age distribution and C_{anth} estimates to larger errors in large values of α.

3 Supplementary Text S3: Preformed property estimate errors are likely a small component of Canth estimate errors

The TRACEv1 uncertainty tolerance for the TA^0 estimate is large: an error of 10 µmol kg^{-1} in TA^0 corresponds to TRACEv1 C_{anth} estimate error of only ~0.4 µmol kg^{-1}. Considering that the neural network reconstruction error for the TA^0 gridded training data is modest (bias=0.0 µmol kg^{-1}, ±1s = 5 µmol kg^{-1}) compared to the uncertainty in the underlying TA^0 estimates of 9.4 µmol kg^{-1} (Carter et al., 2021b), and the sensitivities and reconstruction errors are significantly smaller still for Si^0 and P^0, the contributions from preformed property estimates to C_{anth} uncertainty are therefore thought to be small relative to other contributions (as was also found by He et al. (2018)). This contribution is also implicitly incorporated into our model-derived uncertainty estimate when we use the neural networks trained on the Carter et al. (2021b) preformed property estimates—made for the real-world ocean— for GOBM model output with unknown and presumedly-different preformed properties unique to the GOBM circulation and biogeochemistry. These uncertainties are therefore also implicitly included in the TRACEv1 uncertainty estimates.

4 Supplementary Text S4: Information about code speed optimization

TRACEv1 takes ~0.06 seconds per estimate on a personal laptop. Increasing the pH convergence tolerance in the CO2SYS calculation speeds the code by reducing the iterations required to converge on a final value while only changing the TRACEv1 C_{anth} reconstruction RMSE at the 5th significant digit. Recognizing that the carbonate chemistry of seawater is non-linear whereas DIC mixes conservatively, earlier versions of the code were designed to calculate C_{anth} values for each of the atmospheric CO_2 concentrations specific to each of the many ages that comprise a TTD age distribution before combining the distributions into a single fraction-weighted mean C_{anth} estimate. However, this approach generated estimates that were indistinguishable from the values obtained from the adopted method (within uncertainties) and required invoking the carbonate chemistry solver hundreds of additional times per calculation. This significantly slowed the calculation without meaningfully improving fidelity.

5 Supplementary Text S5: Model and OCIM information extraction methods

The age (A) values are linearly interpolated by latitude, longitude, and depth from the 3-D OCIM grid with nearest neighbor extrapolation using "scatteredInterpolant.m." For this and subsequent gridded-value interpolations, the values with a 0° E < longitude < 20° E are copied to 360° E < longitude < 380° E and the values with a 340° E < longitude < 360° E are copied to -20° E<longitude< 0° E. In addition, depths are divided by 25 m and latitudes multiplied by 4 such that properties tend to vary more equally for a given change in the various coordinates. These measures are often essential for plausibly interpolating scattered 3-D data, but usually only provide a minor improvement when interpolating readily interpolated gridded values like the OCIM A estimates and the GOBM output.

In order to subsample the GOBMs in a fashion that is consistent with the available real-world observations, modeled values of annually-averaged simulated CFC-11, CFC-12, and SF_6 concentrations; temperature; salinity; and DIC are linearly interpolated from the 3-dimensional model grid to the sparse locations and times at which all three transient tracer measurements of interest are available in the GLODAPv2.2023 data product. Rather than a 4-dimensional interpolation with time, the GOBM output from each year is used for 3-dimensional interpolations at locations measured in the same year. The GOBM output comes from the NorESM RECCAP2 simulation suite (Müller, 2023). We use output from simulation "A" to estimate simulated measurements. Simulation A aims to represent a variable forcing as represented by atmospheric reanalysis products along with continued C_{anth} accumulation in the atmosphere and ocean. Simulated C_{anth} is then defined as the difference between this simulation and the "D" simulation that has identical physical forcing but fixed preindustrial atmospheric C_{anth} (and therefore no marine C_{anth} accumulation). As a final step, the model transient tracer output is converted to partial pressure equivalents using the pCFCfromCFC.m function distributed alongside the TRACEv1 code (though never invoked by it), and simulated concentrations are converted to substance amount contents by dividing by the density of the seawater calculated from the modeled S and T. For the reconstruction error analysis, annually-averaged output is extracted from 1980 and 2014 c.e.

6 Supplementary Text S6: Model reconstruction estimate of TRACEv1 estimate uncertainty

The α values fit to the NorESM output can be used directly—without invoking the neural networks used by TRACEv1 to remap α—to estimate C_{anth} at the locations and times where the α values were fit to simulated transient tracer distributions and the interpolated OCIM age distribution. These reconstructions show strong agreement with NorESM C_{anth} distributions, with bias (±root-mean-squared, RMS) errors of −1.5 (±4.5) µmol kg^{-1} (Table S1). Interestingly, these errors decrease slightly—to −1.2 (±4.1) µmol kg^{-1}—when the TRACEv1_Validation neural networks are used to reconstruct the α values at these same locations. This improvement is likely found because the statistical smoothing across space (along a cruise) and time (between cruises) inherent to the neural network estimates is helpful for the C_{anth} reconstruction.

When applying the TRACEv1_Validation_NorESM to the full model grid from 1980 and 2014, the errors become −0.6 (±4.4) µmol kg^{-1}. The overall inventory estimates (given in Table S1) are within 15% of the simulated values, though we caution that this is not a robust result with only 2 realizations from 1 model simulation set. We note that the average year of the viable combinations of transient tracer measurements used to fit versions of TRACE (notably those that include the recent SF_6 measurements) is 2012.4 c.e., so the smaller average absolute bias in 2014 could perhaps be explained by the measurements providing a stronger constraint for the ventilation history in this later period. If correct, this would imply that new measurements remain useful for quantifying variability in the true ventilation history of the ocean interior, and that steady state circulation estimates (and thus projections) can only yield so much reconstruction fidelity. An alternative explanation is that the inventory error is relatively consistent while the inventory is smaller in the previous period. Based on these observations and limitations of a reconstruction of only a single model simulation, we assign an (±1σ) ±4.4 µmol kg^{-1} C_{anth} uncertainty to TRACEv1 estimates, that grows to ±15 % of any estimate when this percentage is larger than ±4.4 µmol kg^{-1}. This uncertainty estimate is combined with the Monte Carlo uncertainty estimate (Supplementary Text S7) using a formula provided in the main text.

Table S1. TRACEv1_validation_NorESM misfits for reconstructions of the NorESM C_{anth} distributions. Misfits are quantified directly from simulated transient tracer measurements interpolated from full 3-D gridded output (meas. direct α) to the locations and times of measurements in the real ocean, from these same values when the α has been reconstructed at these same locations using a neural network (meas. with NN), and for the full 3D model domain again using the NN to remap α (full grid with NN). These last reconstructions are also used to calculate an error in the reconstructed full ocean C_{anth} inventory in 1980 and 2014 given in PgC (and as a percentage of the total inventory in parentheses).

Quantity	Bias or error	RMSE	Units
meas., direct α	-1.5	4.5	µmol C_{anth} kg^{-1}
meas. with NN	-1.2	4.1	µmol C_{anth} kg^{-1}
full grid, with NN (u_{MR})	-0.6	4.4	µmol C_{anth} kg^{-1}
Inventory in 1980 (modeled: 110 PgC)	-11 (-10%)	-	PgC (% of total)
Inventory in 2014 (modeled: 197 PgC)	-7 (-3%)	-	PgC (% of total)

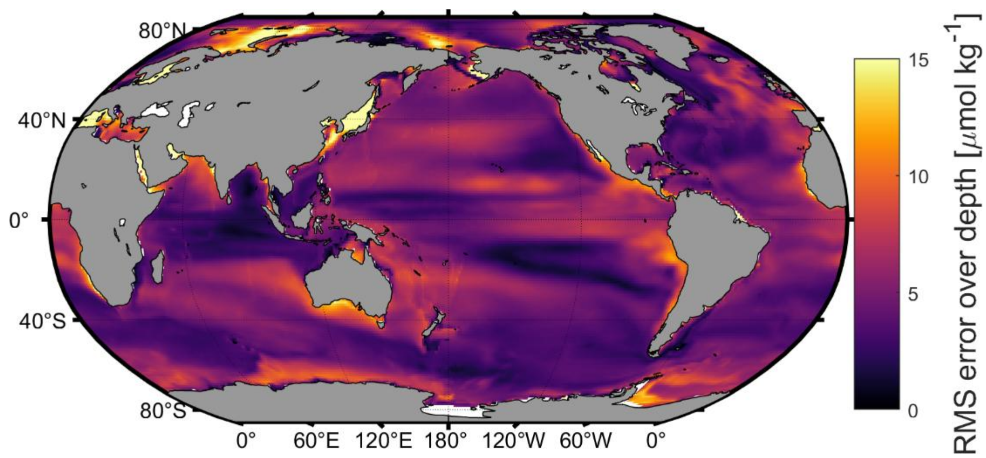

Figure S1. Root mean squared C_{anth} error mapped regionally (for all estimates across depth) in NorESM model year 2014.5 c.e.

We can examine the regional distribution of RMS (computed for all estimates vertically) residuals globally (Figure S1). The largest errors are found in the Arctic and in marginal seas including the Baltic, Mediterranean, and the East Sea/Sea of Japan. In most of these areas there are few training data, and similar neural networks have noted reconstruction errors in such locations previously (Carter et al., 2021a). This provides a caution against using TRACEv1 in enclosed basins that lack the full suite of transient tracer measurements. In addition, there is an indication that RMS reconstruction error increases near coastlines and especially in areas of strong upwelling. This feature is even more noticeable at the ocean surface where there seems to be a pronounced over-estimation of C_{anth} in upwelling regimes such as the California Current, the Eastern Tropical Pacific, the Peru Current, and the Benguela Current (Figure S2a). This implies that the neural network fit of α is not well-capturing the shift in the C_{anth} content that accompanies changes in S and T when upwelled water with lower C_{anth} is brought to the surface. Larger reconstruction errors in dynamic coastal environments are a common problem for C_{anth} estimates where uncertainties of 50% have been reported (Feely et al., 2016). The TRACEv1 reconstruction uncertainty seems to be closer to ~30% in these areas (Figure 2b). TRACEv1 uncertainty estimates do not account for this regionally-enhanced uncertainty. It is likely that a broader transient tracer measurement training product and iteration on the fitting terms for the neural networks could allow better resolution of these dynamic features in future versions of TRACE.

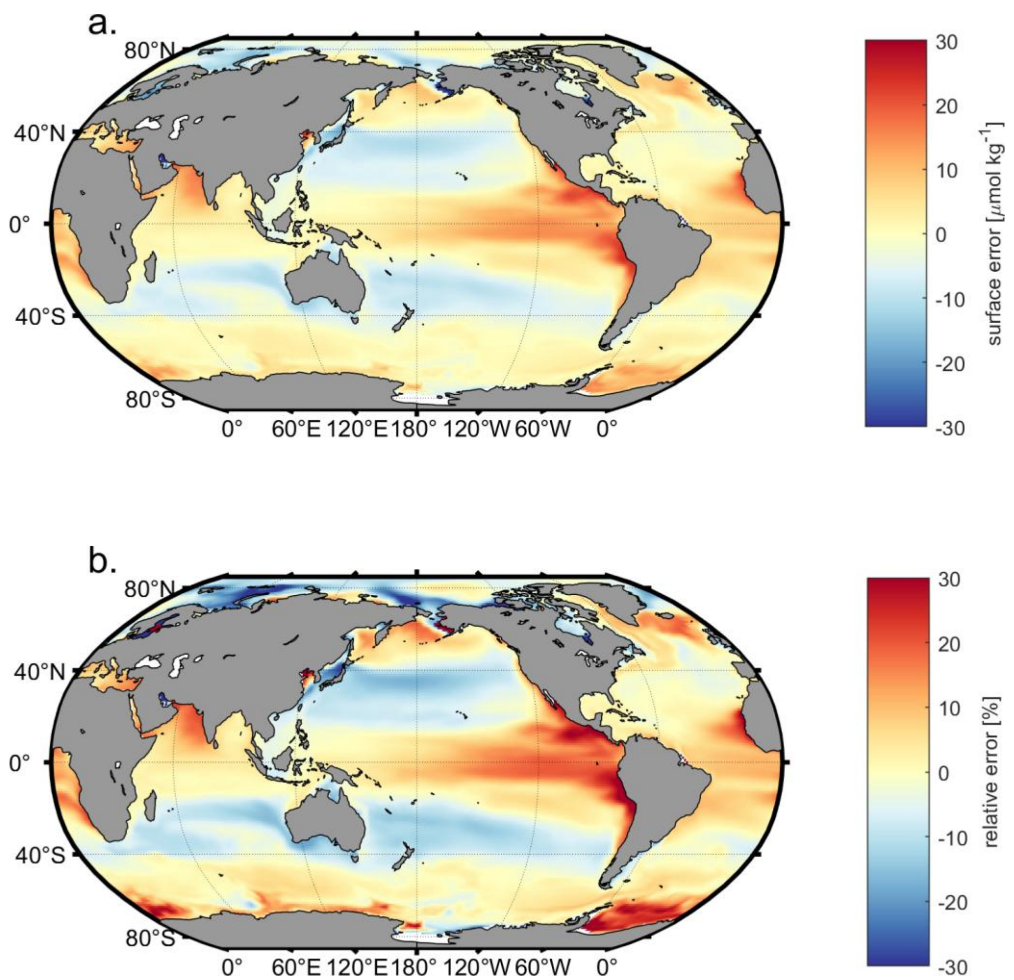

Figure S2. Map of the (**a**) surface C_{anth} estimate error and (**b**) its value expressed as a percentage of the estimate.

7 Supplementary Text S7: Monte Carlo Analysis

The Monte Carlo analysis uses 10 versions of TRACE, TRACEv1_MonteCarlo_1 through TRACEv1_MonteCarlo_10, which were trained using data with two sets of perturbations applied. For each cruise in GLODAPv2.2023, relative perturbations were selected from a normally distributed population with a mean of 0 and a standard deviation of 2% for CFCs and 3% for SF_6 and added to the measurements. However, when 2% of the CFC-11 or CFC-12 measurement is less than 0.005 pmol kg^{-1}, or when 3% of the SF_6 measurement is less than 0.05 fmol kg^{-1}, then we select random measurement perturbations from populations with standard deviations equal to the larger percentages implied by these two minimum-measurement error thresholds (up to 100% of the values for the smallest concentrations). This perturbation represents systematic cruise-wide measurement uncertainty sources. We similarly perturb each individual transient tracer measurement in the data product by a second unique perturbation selected for that measurement from the same population of offsets. This perturbation represents measurement-specific uncertainties.

The Monte Carlo analysis reveals that the applied measurement uncertainties generate an additional ±2 μmol kg^{-1} of RMS C_{anth} estimate variability and a small (−0.2 μmol kg^{-1}) C_{anth} bias. On a global scale, this analysis induces an RMS global inventory estimate variability of ±0.6% from transient tracer measurement uncertainties and increases the overall inventory by 1.5%. Interestingly, the measurement uncertainties induce a larger +0.8 μmol kg^{-1} C_{anth} bias if the OCIM *A* estimates are omitted from the fitting. Omitting this constraint results in an inventory bias of +3.4 and +6.4 PgC in 1980 and 2014, respectively. The positive biases induced by transient tracer measurement errors are likely a consequence of the non-linearities between transient tracer concentration histories, apparent age, and atmospheric C_{anth} accumulation.

8 Supplementary Text S8: Age estimate comparisons

TRACEv1 is used to reproduce the *A* data product of Jeannson et al. (2021) and it shows approximately comparable ages for the subset of the estimates that have more than 30 μmol kg^{-1} C_{anth} (Figure S3a). TRACEv1 ages are, if anything, on average younger than the TTD ages of Jeannson et al. (2021) obtained from SF_6 by an average of 2.5 years for this subset. The largest differences are found in the deep Atlantic and Pacific where—like the OCIM *A* values—the TRACEv1 ages are younger in the Atlantic and older in the Pacific. TRACEv1 *A* estimates are generally comparable to the OCIM *A* estimates, though they are younger on average, particularly for the water masses where the C_{anth} is estimated to be between ~10 and 30 μmol kg^{-1} (Fig. S3c).

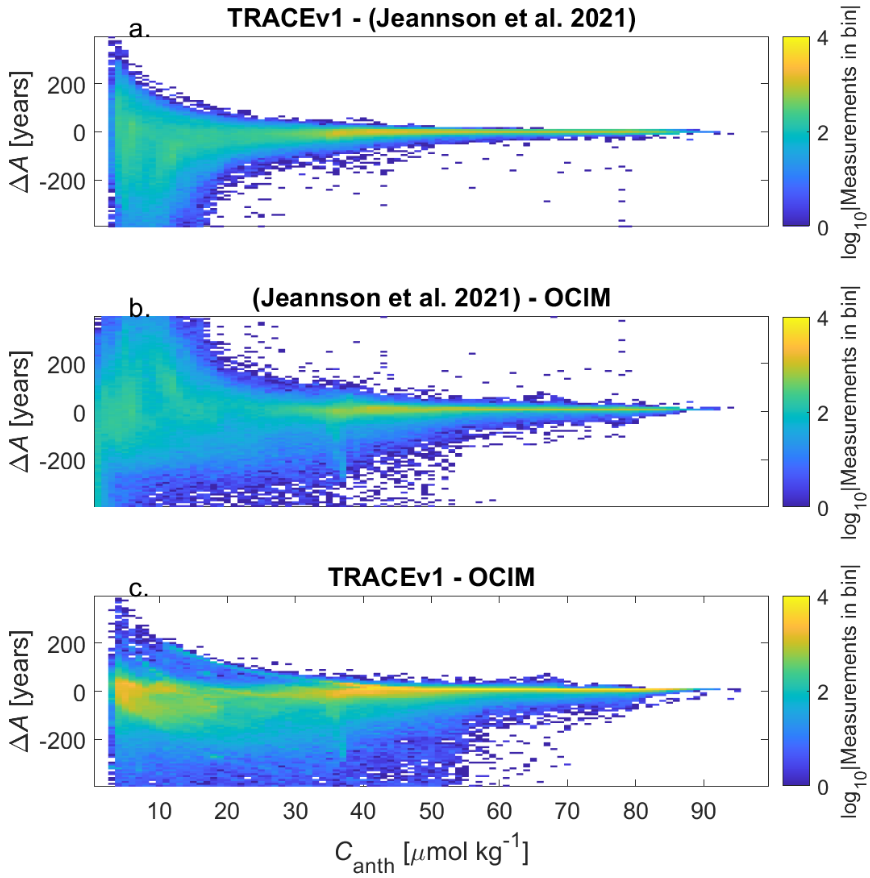

Figure S3. Two dimensional histograms in color comparing the distributions of age estimates from various approaches. There is comparatively strong coherence between estimates in the well-ventilated waters with significant C_{anth} (i.e., the warmest colors indicating bins with high data density are near ΔA=0). For waters with less C_{anth} the OCIM tends to have the oldest A estimates and TRACEv1 tends to have the youngest.

9 Supplementary Text S9: Comparison to the GLODAPv2 gridded product

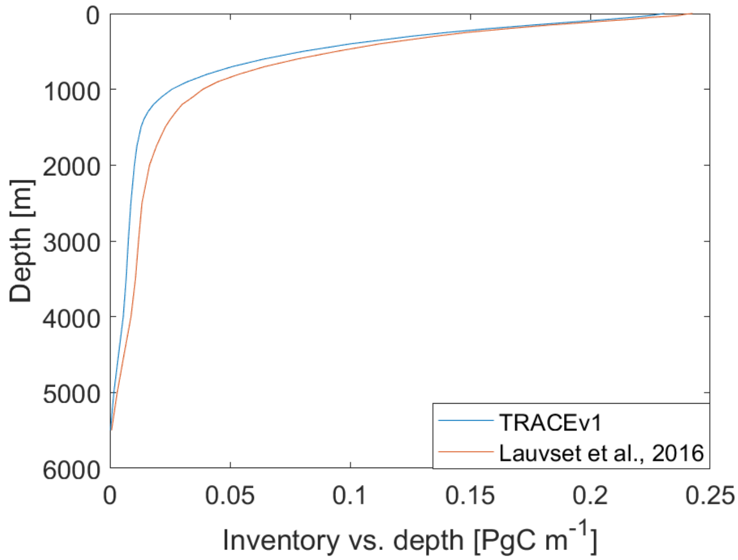

Figure S4. Inventory vs. depth in 2002 from the TRACEv1 and GLODAPv2 gridded (Lauvset et al., 2016) products.

The published C_{anth} distribution from the gridded GLODAPv2 data product (Lauvset et al., 2016) allows us to examine the source of the largest disagreement in greater detail. Most of the disagreement is found at depth, with the gridded GLODAPv2 product showing greater anthropogenic carbon inventories below 500 m (Figure S4). There are several differences between TRACEv1 and the TTD approach of Lauvset et al. (2016) that could account for the differences in the inferred inventories in 2002: the inclusion of the "CO_2 disequilibrium" adjustment in TRACEv1 (which lowers C_{anth} estimates except following atmospheric xCO_2 declines in some SSP projections), the use of OCIM A and the inclusion of SF_6 as additional TTD fitting constraints by TRACEv1, differences between the shape of the TTDs fit, and the use of transient steady state by Lauvset et al. (2016) to adjust C_{anth} values to their 2002 values before gridding them to the GLODAPv2 grid vs. the use of neural networks with TRACEv1 to generate TTDs and C_{anth} values for the GLODAPv2 grid. Omission of the OCIM ages has modest impact on the calculated inventories (<+1 PgC in the model validation and +3.4 to +6.4 PgC from the Monte Carlo analysis), but the combination remains insufficient to account for the difference, so there must be another source of disagreement. Waugh et al. (2006) revised their TTD estimates, which are similar to the estimates from (Lauvset et al. 2016), lower by 20% after observing a consistent overestimate from their TTD approach when it was applied in a model environment with known C_{anth}. The disagreement was particularly strong in the Southern Ocean, and we note that this is one location where it can be unclear whether small transient tracer contents correspond to a large amount of old water mixed with a small amount of recent ventilation or a comparatively large fraction of water that was ventilated in the era when atmospheric transient tracer concentrations were first becoming measurable. As the rise in atmospheric CO_2 began well before the rise in transient tracer concentrations (Fig. 4a), these two interpretations result in different calculated C_{anth} values. The use of SF_6, with a distinct atmospheric growth history from CFC-11 and CFC-12, as an additional constraint by TRACEv1 presumably helps with this disambiguation somewhat. It does not appear that TRACEv1 A estimates are significantly older than the ages obtained from other transient tracer

TTD based estimates, as might be implied by TRACEv1 C_{anth} inventory being smaller than the inventory from Lauvset et al. (2016) (Supplementary Text S8). As there are multiple possible explanations, the primary source of the disagreement is unclear.

10 Supplementary References

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO_2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks, Frontiers in Marine Science, 5, 328, https://doi.org/10.3389/fmars.2018.00328, 2018.

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y. Y., Álvarez, M., Wanninkhof, R., Feely, R. A., and Barbero, L.: New and updated global empirical seawater property estimation routines, Limnology and Oceanography: Methods, https://doi.org/10.1002/LOM3.10461, 2021a.

Carter, B. R., Feely, R. A., Lauvset, S. K., Olsen, A., DeVries, T., and Sonnerup, R.: Preformed Properties for Marine Organic Matter and Carbonate Mineral Cycling Quantification, Global Biogeochemical Cycles, 35, e2020GB006623, https://doi.org/10.1029/2020GB006623, 2021b.

Feely, R. A., Alin, S. R., Carter, B. R., Bednar??ek, N., Hales, B., Chan, F., Hill, T. M., Gaylord, B., Sanford, E., Byrne, R. H., Sabine, C. L., Greeley, D., and Juranek, L.: Chemical and biological impacts of ocean acidification along the west coast of North America, Estuarine, Coastal and Shelf Science, 183, https://doi.org/10.1016/j.ecss.2016.08.043, 2016.

He, Y.-C., Tjiputra, J., Langehaug, H. R., Jeansson, E., Gao, Y., Schwinger, J., and Olsen, A.: A Model-Based Evaluation of the Inverse Gaussian Transit-Time Distribution Method for Inferring Anthropogenic Carbon Storage in the Ocean, Journal of Geophysical Research: Oceans, 123, 1777–1800, https://doi.org/10.1002/2017JC013504, 2018.

Lauvset, S. K., Key, R. M., Olsen, A., Heuven, S. van, Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2, Earth System Science Data, 8, 325–340, https://doi.org/10.5194/ESSD-8-325-2016, 2016.

Müller, J. D.: RECCAP2-ocean data collection, https://doi.org/10.5281/zenodo.7990823, 2023.

Waugh, D. W., Hall, T. M., Mcneil, B. I., Key, R., and Matear, R. J.: Anthropogenic CO2 in the oceans estimated using transit time distributions, Tellus B: Chemical and Physical Meteorology, 58, 376–389, https://doi.org/10.1111/j.1600-0889.2006.00222.x, 2006.