The Atmospheric Potential Oxygen forward Model Intercomparison Project (APO-MIP1): Evaluating simulated atmospheric transport of air-sea gas exchange tracers and APO flux products

Yuming Jin[1,2], Britton B. Stephens[1], Matthew C. Long[3], Naveen Chandra[4], Frédéric Chevallier[5], Joram J.D. Hooghiem[6], Ingrid T. Luijkx[6], Shamil Maksyutov[7], Eric J. Morgan[8], Yosuke Niwa[7], Prabir K. Patra[4,9], Christian Rödenbeck[10], Jesse Vance[11]

1. Earth Observing Laboratory, NSF National Center for Atmospheric Research, Boulder, CO 80301, USA

2. Advanced Study Program, NSF National Center for Atmospheric Research, Boulder, CO 80301

3. [C]Worthy, Boulder, CO 80302, USA

4. Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, 236-0001, Japan

5. Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, F-91198, France

6. Wageningen University, Environmental Sciences Group, Wageningen, 6700AA, The Netherlands

7. Earth System Division, National Institute for Environmental Studies, Tsukuba, 305-8506, Japan

8. Geosciences Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA

9. Seto Inland Sea Carbon Neutral Research Center, Hiroshima University, Higashi-Hiroshima, 739-8529, Japan

10. Max Planck Institute for Biogeochemistry, Jena, 07745, Germany

11. Ebb Carbon, San Carlos, CA 94070, USA

Correspondence: Yuming Jin (yumingjin@ucar.edu)

Abstract

Atmospheric Potential Oxygen (APO, defined as $O_2 + 1.1 \times CO_2$) is primarily a tracer of ocean biogeochemistry and fossil fuel burning. APO exhibits strong seasonal variability at mid-to-high latitudes, driven mainly by seasonal air-sea O_2 exchange. We present results from the first version of the Atmospheric Potential Oxygen forward Model Intercomparison Project (APO-MIP1), which forward transports three air-sea APO flux products in eight atmospheric transport models or model variants, aiming to evaluate atmospheric transport and flux representations by comparing simulations against surface station, airborne, and shipboard observations of APO. We find significant spread and bias in APO simulations at eastern Pacific surface stations, indicating inconsistencies in representing vertical and coastal atmospheric mixing. A framework using airborne APO observations demonstrates that most atmospheric transport models (ATMs) participating in APO-MIP1 overestimate tracer diffusive mixing across moist isentropes (i.e., diabatic mixing) in mid-latitudes. This framework also enables us to isolate ATM-related biases in simulated APO distributions using independent mixing constraints derived from moist static energy budgets from reanalysis, thereby allowing us to assess large-scale features in air-sea APO flux products. Furthermore, shipboard observations show that ATMs are unable to reproduce seasonal APO gradients over Drake Passage and near Palmer Station, Antarctica, which could arise from uncertainties in APO fluxes or model transport. The transport simulations and flux products from APO-MIP1 provide valuable resources for developing new APO flux inversions and evaluating ocean biogeochemical processes.

Short Summary

We carry out a comprehensive atmospheric transport model (ATM) intercomparison project. This project aims to evaluate errors in ATMs and three air-sea O_2 exchange products by comparing model simulations with observations collected from surface stations, ships, and aircraft. We also present a model evaluation framework to independently quantify transport-related and flux-related biases that contribute to model-observation discrepancies in atmospheric tracer distributions.

1. Introduction

Atmospheric potential oxygen (APO), defined as the weighted sum of O_2 and CO_2 concentration ($APO \approx O_2 + 1.1\ CO_2$), is an important tracer of fossil fuel burning and ocean biogeochemical processes (Stephens et al., 1998). APO is intended to be unaffected by terrestrial photosynthesis and respiration due to the cancellation of O_2 and CO_2 exchange at an approximate O_2:C ratio of -1.1 (Severinghaus, 1995). APO exhibits a large seasonal cycle driven mainly by air-sea O_2 exchange due to upper ocean biological activities, deep water ventilation, and thermally induced O_2 solubility changes. Seasonal APO variability is also slightly affected by the air-sea exchange of CO_2 and N_2 (Manning and Keeling, 2006). APO is decreasing in the atmosphere due to fossil fuel combustion, which acts as an O_2 sink and CO_2 source with a more negative O_2:CO_2 ratio (global mean ~ -1.4) compared to the assumed -1.1 ratio from terrestrial processes. Although fossil fuel combustion contributes to an annual interhemispheric gradient that has lower APO in the Northern Hemisphere, it has only a minor effect on the seasonal cycle globally (Keeling & Manning, 2014).

APO measurements provide critical constraints on seasonal air-sea O_2 fluxes, which have been used to estimate air-sea gas exchange rates and ocean net community production (NCP), and to benchmark marine NCP in Earth system models (Naegler et al., 2007; Nevison et al., 2018, 2012, 2015, 2016). APO has been used for improved partitioning of ocean and land carbon sinks (Friedlingstein et al., 2025; Manning and Keeling, 2006), to constrain ocean heat uptake and meridional heat transport (Resplandy et al., 2016, 2019), and to quantify fossil fuel emissions (Pickers et al., 2022; Rödenbeck et al., 2023). APO measurements are available at surface stations (e.g., Adcock et al., 2023; Battle et al., 2006; Goto et al., 2017; Keeling & Manning, 2014; Manning & Keeling, 2006; Nguyen et al., 2022; Tohjima et al., 2019), on ship transects (e.g., Ishidoya et al., 2016; Pickers et al., 2017; Stephens et al., 2003; Thompson et al., 2007; Tohjima et al., 2012, 2015, 2024), and from aircraft (e.g., Bent, 2014; Ishidoya et al., 2012; Jin et al., 2023; Langenfelds, 2002; Morgan et al., 2021; Stephens et al., 2018, 2021).

Global-scale air-sea APO fluxes have been estimated from APO measurements and an ATM within a Bayesian inversion framework (Rödenbeck et al., 2008). ATMs are also used to forward transport APO fluxes simulated from ocean biogeochemistry models (Carroll et al., 2020; Yeager et al., 2022) and surface ocean dissolved oxygen (DO) measurements (Garcia and Keeling, 2001;

Najjar and Keeling, 2000) to compare with atmospheric observations, providing a basis for model and flux product evaluation (Jin et al., 2023; Keeling et al., 1998; Stephens et al., 1998). However, using atmospheric data to evaluate flux products and to derive fluxes through inversion is fundamentally limited by biases in ATMs, particularly in their representation of vertical transport and diabatic mixing (Jin et al., 2024; Naegler et al., 2007; Nevison et al., 2008; Schuh et al., 2019; Schuh and Jacobson, 2023; Stephens et al., 2007). The systematic uncertainties in transport modeling limit inversions of APO, CO_2, and other greenhouse gases, underscoring the need for independent transport bias assessments to advance global carbon budget constraints.

To address uncertainty in ATMs for studying large-scale tracer atmospheric transport and the corresponding surface fluxes, several community model intercomparison (TransCom) projects have been established for various tracers including CO_2 (Baker et al., 2006; Gurney et al., 2003, 2004; Law et al., 2008; Patra et al., 2008), N_2O (Thompson et al., 2014), SF_6 (Denning et al., 1999), SF_6 and CH_4 jointly (Patra et al., 2011), as well as an age of air tracer (Krol et al., 2018). Blaine (2005) coordinated a TransCom O_2 experiment to compare model simulations of the O_2 seasonal cycle across the Scripps O_2 network. While this experiment provided valuable initial insights into ATM performance in simulating atmospheric O_2 from ocean fluxes, substantial advances in ATMs and more data collected also from aircraft and ships since then motivate an updated intercomparison study with more extensive model-data comparisons and analyses. More recently, CO_2 inversion intercomparisons have been coordinated through the OCO-2 MIP (Crowell et al., 2019; Peiro et al., 2022; Byrne et al., 2023) and the Global Carbon Project (e.g., Friedlingstein et al., 2025). These experiments reveal substantial spread in forward tracer (e.g., CO_2) atmospheric distribution and inverted surface fluxes, driven by different ATMs and inversion setups. The spread in forward transport simulations stems from multiple factors, including the choice of wind fields from various reanalysis products or online simulation, regridding fine resolution meteorological data to coarse model grids, the advection scheme that governs large-scale mixing, and parameterized sub-grid processes, such as boundary layer mixing and deep convection. Despite the complexity of different transport pathways, long-lived tracers (e.g., CO_2 and O_2) at mid-latitudes tend to show tracer distributions that are aligned with moist potential temperature (θ_e) surfaces. This is because θ_e surfaces are preferential surfaces for mixing, leading to rapid along-θ_e mixing and slow cross-θ_e mixing (Bailey et al., 2019; Jin et al., 2021; Miyazaki et al., 2008; Parazoo et al., 2011).

It is a critical challenge to accurately quantify the rate-limiting cross-θ_e mixing time-scales, which are largely driven by diabatic processes including moist convection and radiative cooling. Here, we define "diabatic mixing rates" as diffusivities that are inversely related to cross-θ_e mixing time-scales. These mixing rates are important for determining the large-scale tracer distribution in ATMs. Jin et al. (2024) established a framework to calculate cross-θ_e mixing rates from ATMs and moist static energy (MSE) budgets from reanalysis based on a mass-indexed isentropic coordinate called $M_{\theta e}$ (Jin et al., 2021). This framework also allows cross-θ_e tracer gradients from airborne observations to provide independent constraints on diabatic mixing. Jin et al. (2024) tested four ATMs used in CO_2 inversions, showing that these models tend to have too fast mixing in the mid-latitudes of the Southern Hemisphere in the austral summer. The too fast mixing is also confirmed by the fact that models simulate smaller CO_2 gradients compared to airborne observations, which is an independent constraint on the mixing rate. The mixing rate constraint and CO_2 gradient constraint also have implications for biases in the inverse model estimates, indicating a too large summer-time Southern Ocean (SO) CO_2 sink. This framework provides a system for independently evaluating transport simulations and flux estimates.

Previous TransCom experiments focused primarily on tracers that only have significant sources and sinks over the land, and large seasonal flux cycles tied to the northern terrestrial biosphere. In contrast, APO is a tracer of surface ocean exchange with the largest seasonal variability observed over mid-to-high latitude oceans in both hemispheres. APO offers a distinct perspective for studying atmospheric mixing within and above the marine boundary layer, the long-range tracer transport into and out of the remote Southern Hemisphere, and the ability for inverting tracer flux over the SO from atmospheric measurements.

Here we use output from the APO-MIP1 (Stephens et al., 2025), which generated a suite of forward ATM simulations of APO and its components (air-sea O_2, CO_2, and N_2 flux, and fossil fuel CO_2 emission and O_2 uptake) from different source fields. This effort was initially motivated by a need to support the calibration of hemispheric-scale seasonal air-sea APO flux estimates from spatially and temporally sparse observations from airborne campaigns (e.g., Jin et al., 2023), stations, and ships. Here we focus on the other goals of APO-MIP1 which were to use atmospheric APO observations to characterize errors in ATMs and APO flux products.

In Section 2, we describe APO measurements from surface stations, aircraft, and ships, and the experimental design of APO-MIP1. In Section 3.1, we evaluate simulations against observations, revealing large model spread and errors at eastern Pacific surface stations due to mixing uncertainties, while airborne column-average data show smaller cross-ATMs variability and errors. In Section 3.2, we analyze diabatic mixing rates, demonstrating that ATMs generally overestimate mid-latitude mixing in both hemispheres, allowing us to separate transport and flux-related biases. In Section 3.3, we examine simulations of shipboard data around Drake Passage and the Antarctic Peninsula, revealing that current ATMs and flux products underestimate meridional gradients in APO seasonal amplitude from 53-65°S. The models also fail to capture the APO contrast between Palmer Station flask samples and nearby in-situ ship data due to limitations in representing local topographic flows with coarse-resolution ATMs. In Section 3.4, we discuss the broader implications of our analysis for developing methods to identify processes that introduce transport biases and for improving atmospheric transport modeling.

2. Materials and Methods

2.1 Definition of APO

APO in the unit of per meg (see Keeling et al., 1998) is calculated from atmospheric observations of relative changes in the O_2/N_2 ratio (per meg) and CO_2 mole fraction (ppm) according to Stephens et al. (1998) as

$$APO \approx \delta(O_2/N_2) \; + \; \frac{1.1}{X_{O_2}}(CO_2 - 350), \qquad (1)$$

with

$$\delta(O_2/N_2) = \left(\frac{(\frac{O_2}{N_2})_{sample}}{(\frac{O_2}{N_2})_{reference}} - 1\right) \cdot 10^6. \qquad (2)$$

The factor 1.1 represents the approximate exchange ratio of O_2 to CO_2 in terrestrial biospheric processes (Severinghaus, 1995). We note that this ratio generally varies from 1.01 to 1.14 in aboveground carbon pools across different temporal and spatial scales (Gallagher et al., 2017; Hockaday et al., 2009; Keeling, 1988; Worrall et al., 2013). This ratio also exhibits diurnal change and varies between respiration and photosynthesis in biosphere-atmosphere O_2 and CO_2 exchanges (Faassen et al., 2023, 2024). With our focus on seasonal variations, we use 1.1 as representative of the O_2 to CO_2 exchange ratio during seasonal growth and decay of terrestrial biota. A sensitivity test in Jin et al. (2023) showed that varying this ratio by ± 0.05 only leads to ±5.1% changes in hemispheric average APO. The impact on APO seasonal cycle amplitude (SCA) is ±1.44% and ±0.41% in the Northern and Southern Hemisphere, respectively. X_{O_2} (0.2094) is the reference dry-air mole fraction of O_2 used in the definition of the O_2 scale of the Scripps O_2 Program (Keeling et al., 2020). $\delta(O_2/N_2)$ is expressed in units of per meg, while CO_2 is converted from ppm units to per meg units by subtracting a reference value of 350 ppm and then dividing by X_{O_2}. APO observations are typically expressed in per meg units, but they can be converted to ppm equivalent units by multiplying by X_{O_2}.

2.2 Atmospheric measurements

The APO-MIP1 (Stephens et al., 2025) required model output sampled to match a collection of surface station, airborne, and shipboard observations, and also accepted optional output at additional locations, at higher time resolution, and for full 3-D fields, as shown in Tables S1-2. Here we evaluate model APO simulations using observation data collected at 10 surface stations, on 10 airborne campaigns from three projects, and one repeated shipboard transect from 50 cruises. We show sampling locations, and horizontal flight and ship tracks in Fig. 1. We use surface station APO measurements (2009 to 2018) from 10 sampling sites mainly in the Pacific from the Scripps O_2 Program surface flask network (Keeling & Manning, 2014; Manning & Keeling, 2006). The airborne measurements (Stephens et al., 2021) were made on the NSF NCAR GV aircraft during the HIAPER Pole-to-Pole Observation project from 2009 to 2011 (HIPPO, Wofsy, 2011) and the O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study in 2016 (ORCAS, Stephens et al., 2018), and from the NASA DC-8 aircraft during the Atmospheric Tomography Mission from 2016-2018 (ATom, Thompson et al., 2022). Shipboard measurements were made on transects crossing the Drake Passage by the NSF ARSV Laurence M. Gould from 2012-2017 (Stephens, 2025). Details of surface station, airborne, and shipboard APO measurements are provided in Appendix A.

As the primary focus of this study is the APO seasonal cycle and its latitudinal distribution, we remove interannual trends from the observational data. For surface station and airborne measurements, we remove the long-term trend by subtracting a deseasonalized cubic spline fit (smoothing parameter of 0.8) derived from the global mean APO time series using Scripps O_2 Program data following Hamme & Keeling (2008). For the ship data, we apply a similar detrending procedure but use only South Pole Observatory (SPO) data to derive the long-term trend.

2.3 Components of APO in the atmosphere and prescribed surface fluxes

APO exhibits seasonal variations primarily driven by air-sea exchange (F_{APO}^{ocn}), which comprises three components: air-sea exchange of O_2 ($F_{O_2}^{ocn}$), CO_2 ($F_{CO_2}^{ocn}$), and N_2 ($F_{N_2}^{ocn}$). Additionally, APO is influenced by fossil fuel emission of CO_2 ($F_{CO_2}^{ff}$) and consumption of O_2 ($F_{O_2}^{ff}$), which together combine to form a sink for APO due to fossil fuel burning (F_{APO}^{ff}). Fluxes are defined as positive to the atmosphere.

In this study, we primarily simulate APO by performing forward transport of these individual flux components in ATMs, except one inverse model flux product that provides net F_{APO}^{ocn} directly. We combined these components to calculate the net atmospheric APO anomalies in units of per meg as

$$\delta APO = \delta APO^{ocn} + \delta APO^{ff}, \tag{3}$$

with

$$\delta APO^{ocn} = \frac{1}{X_{O_2}} \cdot \Delta O_2^{ocn} - \frac{1}{X_{N_2}} \cdot \Delta N_2^{ocn} + \frac{1.1}{X_{O_2}} \cdot \Delta CO_2^{ocn}, \tag{4}$$

and

$$\delta APO^{ff} = \frac{1}{X_{O_2}} \cdot \Delta O_2^{ff} + \frac{1.1}{X_{O_2}} \cdot \Delta CO_2^{ff}. \tag{5}$$

where ΔO_2^{ocn}, ΔN_2^{ocn}, ΔCO_2^{ocn}, ΔO_2^{ff}, and ΔCO_2^{ff} represents the atmospheric fields in units of deviations in ppm of each flux component ($F_{O_2}^{ocn}$, $F_{CO_2}^{ocn}$, $F_{N_2}^{ocn}$, $F_{O_2}^{ff}$, and $F_{CO_2}^{ff}$) that is forward transport in the ATMs (Stephens et al., 1998). The δ sign denotes tracers in units of per meg.

We utilize three distinct ocean APO flux products: (1) the Jena product, which directly provides F_{APO}^{ocn} from an atmospheric APO inversion framework that assimilates surface station measurements (Rödenbeck et al., 2008); 2) the CESM product, an Earth System Model simulation with prognostic ocean biogeochemistry (Yeager et al., 2022; Long et al., 2021a) that generates separate flux components ($F_{O_2}^{ocn}$ and $F_{CO_2}^{ocn}$) ; and 3) the DISS product, which provides separate observation-based flux components incorporates surface ocean dissolved oxygen measurements (Garcia & Keeling, 2001; Resplandy et al., 2016) and pCO_2 data (Jersild et al., 2017; Landschützer et al., 2016). $F_{N_2}^{ocn}$ for CESM and DISS is estimated by scaling ocean heat fluxes from CESM and ERA-5, respectively, using the relationship of Keeling et al. (1993). For fossil fuel contributions, we employ the OCO2MIP product for CO_2 emissions (Basu & Nassar, 2021) and the GridFED database for coupled O_2 and CO_2 fluxes from fossil fuel combustion (Jones et al., 2021). Details of each product are provided in Appendix B. All flux fields were linearly interpolated from their original temporal and spatial resolution to 1° longitude × 1° latitude with daily temporal resolution from 1986 to 2020. When flux data were unavailable in the earlier portion of this time period (Jena and OCO2MIP), we set the corresponding fluxes to zero. Participating modelers were requested to simulate at least from 2009 to 2018, following three years of spin up from 2006 to 2008, and optionally longer (Table 1). In addition to Jena, which is simulated directly, we construct the two ΔAPO^{ocn} products using Eq. 4 and two ΔAPO^{ff} products using Eq. 5, as described in Appendix B. Fig. 2 illustrates the seasonal and latitudinal flux patterns of these three ocean APO flux products and the fossil fuel APO flux from GridFed, which serves as our primary fossil fuel flux dataset in this study.

2.4 Atmospheric tracer transport models

We simulate each component of APO in the atmosphere using the flux fields described in Section 2.3, and eight ATMs (see Table 1). All tracer atmospheric fields are modeled as tracer deviations against an arbitrary background with concentrations in ppm dry air mole fraction (as for CO_2). These tracer mole fractions are later converted to deviations in units of per meg after subtracting the model-specific arbitrary reference according to Eq. 4. We describe key model parameters and setups below.

2.4.1 CAM-SD

The Community Atmosphere Model (CAM) version 6.0 is the atmospheric component of CESM2 (Danabasoglu et al., 2020). The version used here is run online with specified dynamics (SD), wherein the model is constrained with MERRA-2 reanalysis, and uncoupled from the other climate system components. Temperature and horizontal winds (u and v) are nudged to MERRA-2, 8 times per day, with a normalized strength coefficient of 0.25. Shallow convection is parameterized following the Cloud-Layers Unified by Binormals framework (CLUBB, Golaz et al., 2002), and deep convection is parameterized following Zhang & McFarlane (1995). CAM has not been used for tracer inversions, but has been evaluated extensively for its dynamical properties (e.g., Bailey et al., 2019; Kay et al., 2012)

2.4.2 CAMS_LMDZ

CAMS_LMDZ refers here to the offline transport model from the Atmospheric General Circulation Model of Laboratoire de Météorologie Dynamique, called LMDz. LMDz is the atmospheric component of the Earth System Model of Institut Pierre-Simon-Laplace (IPSL). It is also used to drive the offline model CAMS_LMDz, in which case its horizontal winds are nudged to those of the ERA5 reanalysis wind fields (Hersbach et al., 2020). From the computer code of LMDz, CAMS-LMDz only keeps the transport subroutines for advection (Hourdin & Armengaud, 1999), deep convection (Emanuel, 1991), thermals (Rio & Hourdin, 2008), and boundary-layer turbulence (Hourdin et al., 2006). All other processes are replaced by an archive of relevant meteorological variables (air mass fluxes, exchange coefficients, temperature, etc.) built with the full LMDz model at the target spatial resolution, thereby allowing relatively small computing time and resources for the offline model. LMDz ensures the physical consistency of the archive of meteorological variables. The meteorological variables are stored as 3-hourly averages. CAMS_LMDZ has been regularly participating in OCO-2 MIP (Byrne et al., 2023) and TransCom intercomparison studies.

2.4.3 CTE_TM5

TM5 is a tracer transport model used for simulating atmospheric trace gas chemistry and transport (Krol et al., 2005). We refer to it as CTE_TM5 because the model was run with the CarbonTracker-Europe (CTE) shell, but this does not alter the TM5 physics and chemistry. TM5 advection is computed using the slopes advection scheme (Russell & Lerner, 1981) and in this work it is driven by ERA-5 reanalysis wind fields, making it an offline model. The convection is computed from the convective entrainment and detrainment fluxes from the ERA-5 reanalysis. Free tropospheric diffusion is computed using the formulation by Louis (1979). Diffusion in the boundary layer is computed using the parametrization by Holtslag & Boville (1993), where the diurnal variability in the boundary layer height is computed using Vogelezang and Holtslag (1996). TM5 is widely used in inversions and regularly participates in MIPs, for different tracers at different model resolutions and driven with different wind reanalysis products (for example, Byrne et al., 2023; Friedlingstein et al., 2025; Gaubert et al., 2019; Krol et al., 2018).

2.4.4 TM3

TM3 (Heimann and Körner, 2003) is an offline atmospheric tracer transport model, in the present runs driven by meteorological fields from the NCEP reanalysis (Kalnay et al., 1996). It was run here on a spatial resolution of 5 degrees longitude, about 3.8 degrees latitude, and 19 vertical layers. The advection uses the slopes scheme (Russell and Lerner, 1981), which is the same as in TM5. Boundary layer mixing is parameterized according to Louis (1979). Vertical mixing due to sub-gridscale cumulus clouds is calculated using the mass flux scheme of Tiedke (1989). TM3 is the ATM used in Jena APO inversion (Rödenbeck et al., 2008), which is one of the flux products used in this study.

2.4.5 MIROC4-ACTM

MIROC4-ACTM is a new generation Model for Interdisciplinary Research on Climate (MIROC, version 4.0; Watanabe et al., 2008) atmospheric general circulation model (AGCM)-based chemistry-transport model (ACTM; Patra et al., 2018) . This AGCM is evolved from the Center for Climate System Research, University of Tokyo (CCSR) / National Institute for Environmental Studies (NIES) / Frontier Research Center for Global Change, JAMSTEC (FRCGC) AGCM version 5.7b (Numaguti et al., 1997). The MIROC4 AGCM propagates only explicitly resolved gravity waves into the stratosphere through the implementation of a hybrid vertical coordinate system compared to its predecessor AGCM5.7b. The MIROC4 AGCM online-simulated horizontal winds and temperature are nudged to the Japanese 55-year Reanalysis (JRA-55) at 6-hourly time intervals (Kobayashi et al., 2015). MIROC4-ACTM produces “age-of-air” up to about 5 years in the tropical upper stratosphere (~1 hPa) and about 6 years in the polar middle stratosphere (~10 hPa), in agreement with observational estimates. The convective transport and inter-hemispheric transport of tracers in the model are validated using ^{222}Radon and sulphur hexafluoride (SF_6), respectively (Patra et al., 2018).

2.4.6 NICAM-TM_gl5 and NICAM-TM_gl6

NICAM-TM is an atmospheric transport model based on the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) (Niwa et al., 2011; Satoh et al., 2014). In this study, we used the offline mode of NICAM-TM, which uses air mass fluxes, vertical diffusion coefficients and other meteorological variables; those data are calculated in advance by an online calculation of NICAM, in which horizontal winds are nudged toward the JRA-55 data. In NICAM, the air mass fluxes are calculated consistently with the continuity equation while conserving tracer masses, which do not require any numerical mass fixing (Niwa et al., 2011). For APO-MIP1, two horizontal resolutions were used: “glevel-5” (gl5) and “glevel-6” (gl6), whose mean grid intervals are 223 and 112 km, respectively. The number of the vertical model layers is 40 and the top of the model domain is at approximately 45 km. The vertical diffusion coefficients are calculated with the MYNN (Mellor and Yamada, 1974; Nakanishi and Niino, 2004) Level 2 scheme (Noda et al., 2010). The cumulus parameterization scheme used in NICAM-TM is

Chikira & Sugiyama (2010). Model performance for atmospheric constituent transport can be found in Niwa et al. (2011, 2012).

2.4.7 NIES

NIES-TM-FLEXPART is a coupled transport model combining Eulerian (NIES-TM) and Lagrangian (FLEXPART) models. It is a transport modeling component of the variational flux inverse modeling system NIES-TM-FLEXPART-Variational (NTFVAR, Maksyutov et al., 2021). The NIES Transport Model (NIES-TM) is an offline model, originally developed in the 1990s (Maksyutov et al., 2008). In this study, the NIES-TM v.21 is used, which improves SF_6 transport and tropopause height over the former v.08.1 (Belikov et al., 2013), as evaluated in Krol et al. (2018), due to (a) using ERA5 hourly wind data, including vertical wind on model coordinates, on 137 model levels and a 0.625° grid for preparation of the 4-hourly average mass fluxes on 42 hybrid-pressure levels, (b) transporting first-order moments (Russell & Lerner, 1981; Van Leer, 1977) for advection, (c) applying penetrative convection rate and turbulent diffusivity supplied by the ERA5 reanalysis (Hersbach et al., 2020). The version v.21 is the same as used in the OCO-2 MIP (Byrne et al., 2023). NIES-TM is coupled with the Lagrangian model FLEXPART (Stohl et al., 2005) to provide refinement to the near field transport during the last 3 days prior to the observation event as presented by (Belikov et al., 2016). FLEXPART model v.8.0 is driven by 6-hourly JRA-55 winds, interpolated to 40 hybrid pressure levels and 1.25°x1.25° resolution. The surface flux footprints are produced by FLEXPART at 1°x1° resolution and daily time step.

Table 1. Participating ATMs and model parameters.

Abbreviation	Model System	Grid (latitude × longitude × levels)	Meteorology	Run start, valid period	Reference(s)
CAM-SD	Community Atmospheric Model	0.9° × 1.25° × 56	MERRA-2	1986, 1989-2019	Danabasoglu et al., 2020
CAMS_ LMDZ	Copernicus	1.875° × 3.75° × 39	ERA5	1986,	Chevallier, 2013;

	Atmosphere Monitoring Service			1991-2020	Chevallier et al., 2005, 2010
CTE_TM5	CarbonTracker Europe	1° × 1° × 25	ERA5	2000, 2003-2020	Luijkx et al., 2017
Jena_TM3	TM3	4° × 5° × 19	NCEP	1986, 1989-2020	Heimann & Körner, 2003
MIROC4-ACTM	MIROC4-ACTM	2.8° × 2.8° × 67	JRA-55	1986, 1991-2020	Chandra et al., 2022; Patra et al., 2018
NICAM-TM_gl5	NICAM-based Transport Model	~223 km × 40	JRA-55	1986, 1989-2020	Niwa et al., 2011, 2017
NICAM-TM_gl6		~112 km × 40			
NIES	NIES-TM-FLEXPART	3.75° × 3.75° × 42 (NIES-TM); 1° × 1° × 40 (FLEXPART)	JRA-55	2000, 2003-2020	Belikov et al., 2016; Maksyutov et al., 2021

2.5 Outputs from transport models

For each ATM, we required simulations for all species sampled to match with the observation locations and times in a subset of the full ObsPack CO_2 files GLOBALVIEWplus v7.0 ObsPack (Schuldt et al., 2021), excluding the model spin-up period. This subset corresponds to existing APO observations that are analyzed in this study from Scripps O_2 Program surface stations, NSF NCAR airborne observations, and NSF NCAR and AIST/JMA shipboard programs. The full list of these records is in Table S1. We note that, while the HIPPO, ORCAS, ATom, and Gould ObsPack files contain CO_2 observations from different instruments, their 10-sec sampling times align with the NSF NCAR APO measurements, except during calibration periods for either instrument.

We also received optional output, which includes the full set of ObsPack files, 3-D atmospheric fields, meteorological variables, additional ship data, and output at additional fixed sites (Table S2). Further details are provided in the APO-MIP1 protocol available at Stephens et al. (2025). We obtained output matching the full set of ObsPack files from four ATMs, which will be useful for future network design. We obtained daily mean 3-D gridded concentration fields from six ATMs. These fields support the calculation of diabatic mixing rates, which we use to evaluate ATMs and the flux products, following the method of Jin et al. (2024). Details are in Section 3.2. We also received hourly (from two versions of NICAM) or 3-hourly (from NIES) output for an extensive list of sites with past or ongoing APO measurements, and co-located samples for ship sampling programs of NIES VOS, AIST R/V Mirai, and UEA Cap San Lorenzo (Hamburg Süd) from three models. These data are not analyzed in this study, but are made available at Stephens et al. (2025).

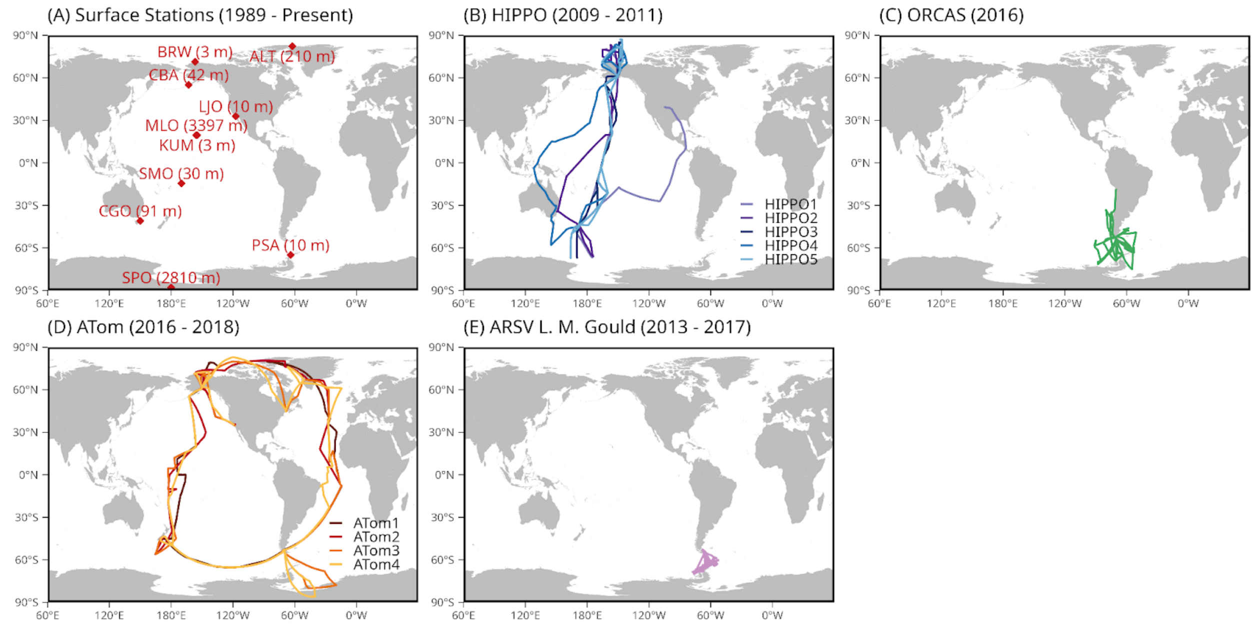


Figure 1: Geographic distribution of APO observations used in this study: (A) Scripps O_2 Program surface stations (red diamonds) with station codes and inlet elevation in meters above sea level; (B) HIPPO (1 to 5) airborne campaign horizontal flight tracks covering the Pacific Ocean; (C) ORCAS aircraft measurements concentrated in the Drake passage; (D) ATom (1 to 4) airborne campaign horizontal flight tracks covering the Pacific and Atlantic Oceans; and (E) Ship-based measurements from the RV *Laurence M. Gould* operating in the Drake passage.

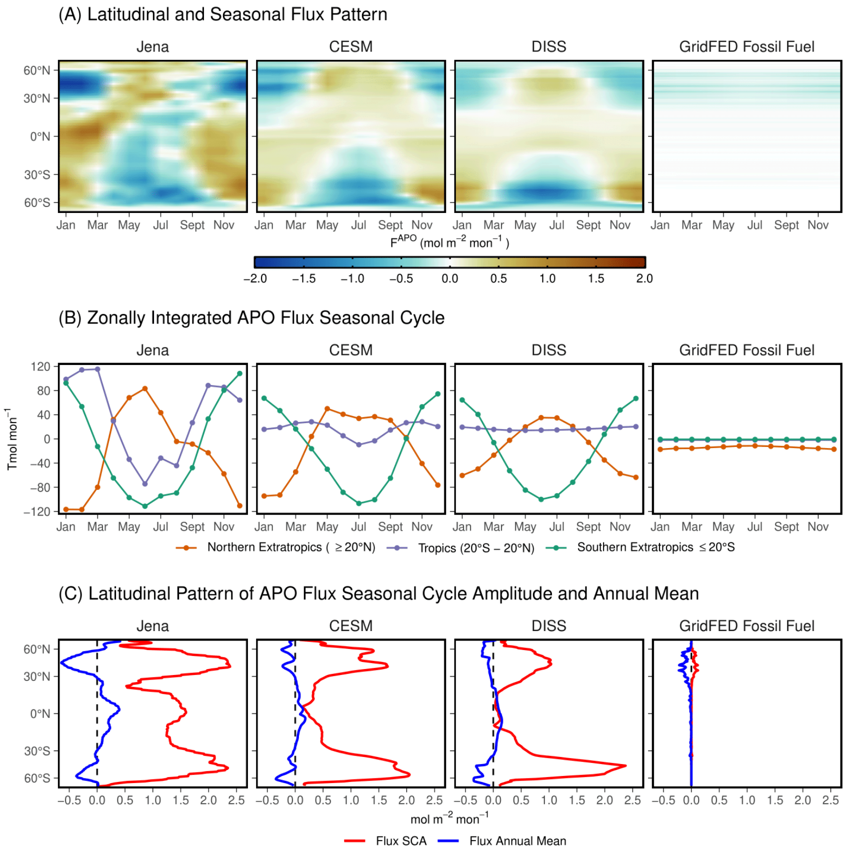


Figure 2: Comparison of APO flux patterns from the three air-sea flux products (Jena, CESM, and DISS) and fossil fuel emissions (GridFed), averaged from 2009 to 2018. (a) Hovmöller diagrams showing the spatiotemporal distribution of APO fluxes (mol m^{-2} mon^{-1}) as a function of latitude and month. (b) Seasonal cycles of zonally integrated fluxes for three latitude bands: Northern Extratropics (≥20°N, orange), Tropics (20°S-20°N, lavender), and Southern Extratropics (<20°S, green). (c) Latitudinal profiles of flux seasonal cycle amplitude (SCA, red) and annual mean flux (blue). For the annual mean profiles (blue lines in panel C), only the latitudinal gradients should be interpreted, as the global means may contain biases in the ocean flux products, which are not the focus of this paper.

3. Results and discussion

3.1 APO model-observation comparisons at surface stations and along aircraft flight tracks

3.1.1 APO seasonal and latitudinal variations at surface stations

We show observations and model simulations of APO seasonal cycles at 10 surface stations of the Scripps O_2 program network in Fig. 3. We present annual mean values, seasonal cycle amplitudes (SCA), and phase from both observations and model simulations at these surface stations in Fig. 4, with model-observation differences shown as colors. Observations show clear meridional gradients in APO annual means (Fig. 5A), with higher values in the Southern Hemisphere than Northern Hemisphere, and a southern tropical “bulge” evident at SMO and in the airborne data centered on 15°S (Battle et al., 2006; Gruber et al., 2001; Stephens et al., 1998). The APO SCA shows higher values in the high latitudes of both hemispheres, with larger amplitudes in the Southern Hemisphere compared to the Northern Hemisphere, yet reaches its maximum at the northern mid-latitude station Cold Bay (CBA) (Fig. 4B). The seasonal phase exhibits an approximately 6-month difference between hemispheres, while remaining relatively uniform within each hemisphere (Fig. 4C).

The higher annual mean APO in the Southern Hemisphere and the southern tropical “bulge” is a result of southward O_2 and CO_2 transport by the oceans, further amplified by net APO uptake in the Northern Hemisphere from fossil fuel burning (Keeling & Manning, 2014; Stephens et al., 1998). The larger APO SCA in mid- to high-latitudes reflects more pronounced seasonal flux cycles resulting from larger marine net primary production (NPP) and sea surface temperature changes in these regions. The thermal and biological effects on APO SCA are further enhanced at eastern Pacific coastal sites (e.g., LJO and CBA), where the shallow marine boundary layer traps high-APO air masses during summer. The 180-days phase difference between the two hemispheres is a result of different seasonal heating and cooling, as well as the biological cycle.

3.1.2 Biases in APO-MIP1 simulations at surface stations

APO-MIP1 simulations of APO annual means and seasonal cycles at surface stations broadly agree with observations (Figs. 3-4). Simulations driven by CESM fluxes show the best agreement with observed APO features. For annual mean spatial patterns (Fig. 4A), CESM- and DISS-driven simulations show comparable performance in representing the southern tropical “bulge” and north-south gradient in annual means, while significantly outperforming simulations using the Jena flux model in northern stations. The main limitation of simulations using CESM fluxes is an overestimation of annual mean APO values across Pacific sites in the Southern Hemisphere, and an underestimation at LJO. Simulations using DISS fluxes also underestimate the annual mean APO at LJO.

APO SCA is well represented in simulations driven by CESM flux, but the SCA at LJO is significantly underestimated in all ATMs except CAM-SD. The underestimation is caused by an overly weak summer-time APO peak (Fig. 3), which also leads to the small annual mean presented above. Simulations using DISS flux generally underestimate SCA, especially in the high latitudes. Simulations using Jena flux, however, generally overestimate the SCA in the mid- to high-latitudes. We find largest SCA biases and cross-ATMs spread at LJO and CBA when using the Jena flux. The biases and model spread are closely related to underrepresentation in ATMs, and will be discussed in the next section. We note that the model biases and spread observed at surface stations are smaller than those reported in the previous TransCom-O_2 experiment (Blaine, 2005), indicating improved atmospheric transport modeling.

Phase simulations using CESM flux are consistent with observations at most stations, except at two northern low-latitude stations, KUM and MLO, where we find too late seasonal minimum day by up to two weeks. Simulations using DISS flux show even larger biases, with earlier seasonal minimum days at all southern and northern low-latitude stations.

3.1.3 Impact of ATM mixing biases

We find APO-MIP1 simulations have large model spread and biases at two northern mid-latitudes stations, LJO and CBA (Fig. 3), especially simulations using Jena fluxes. We note that the interdependence of transport models and fluxes in inversions can be seen for the Jena flux product simulations at LJO (Figs. 3-4). As expected, we see good agreement with observations for the Jena flux product transported by the same model used in the Jena APO inversion (Jena_TM3). However, all other ATMs overestimate summertime APO, and consequently SCA, for the Jena flux product at LJO, CBA, and BRW. All other ATMs also simulate too negative wintertime APO at LJO. These biases suggest a stronger regional APO source in the Jena flux product that could have resulted from too rapid dilution of surface flux signals at LJO in both summer and winter.

Surface station simulations using CESM flux (Figs. 3-4) also reveal elevated model spread and observation deviations at LJO and CBA. At LJO, all ATMs underestimate summertime APO, and consequently SCA, implying too weak upwind outgassing fluxes. The relative magnitude of simulated summer-time peaks for CESM at LJO and CBA maintains a consistent pattern across different flux products, with CAM-SD consistently showing the highest values and Jena_TM3 the lowest, regardless of the flux product used, suggesting consistent biases in the ATMs.

This substantial cross-ATMs variability highlights the challenges in accurately representing complex atmospheric vertical transport processes in regions where strong temperature inversions and stratocumulus clouds significantly influence vertical mixing (Naegler et al., 2007; Nevison et al., 2008). The Jena flux product, derived from an inversion that assimilates these station data, relies on the TM3 tracer transport model (Rödenbeck et al., 2008). Previous studies indicate that TM3 consistently overestimates vertical mixing over the Eastern Pacific, leading to larger inverted seasonal fluxes to match station observations (Jin et al., 2023; Naegler et al., 2007). Our analysis suggests that in comparison to Jena_TM3, vertical mixing is weaker in the two versions of NICAM, CAM-SD, MIROC4-ACTM, and CTE_TM5, which show larger summer-time APO anomalies at LJO and CBA. This pattern is consistent across the three flux products considered.

The larger model spread at northern coastal sites (e.g., LJO, CBA, and BRW) also highlights the limitations of current coarse-resolution ATMs in representing horizontal coastal flows and land-sea breezes. At LJO, samples are collected only during steady west wind (from the ocean) conditions (Keeling et al., 1998). However, ATMs failed to capture the actual small-scale atmospheric conditions associated with on-shore winds during episodic storm systems, which leads to significant underestimation of oceanic influence (Keeling et al., 1998). APO, as a tracer of air-sea gas exchange, is particularly sensitive to the dilution effects in coarse-resolution models.

3.1.4 APO seasonal and latitudinal variations along flight tracks and biases in APO-MIP1

We present zonal averages of APO annual means, SCA, and seasonal minimum days derived from airborne data, grouped into 10-degree latitude and 100-mbar bands in Fig. 5A-C (full seasonal cycles in Fig. S1). We further calculate these three metrics as column-average (black) and at 900-mbar (blue) in Fig. 5D-F, where we also compare them with surface station data (shown as red points). The airborne data show patterns similar to those seen at surface stations but provide detailed vertical structures. The vertical profiles consistently show larger SCA at low altitudes, indicating that the main drivers of SCA are near the surface, while annual means and seasonal phases remain uniform across altitudes. Airborne column averages show increasing SCA and decreasing annual means from low to high latitudes, with similar SCA and annual mean values north of 50°N (Fig. 5D-E), whereas station observations show peaks in the mid-latitudes (LJO and CBA) due to high-APO air masses being trapped below the summer marine boundary layer. This trapping effect is also evident in airborne data interpolated to 900-mbar.

We also calculate APO annual means, SCA, and phases using aircraft simulations from APO-MIP1 (full seasonal cycles in Fig. S1) and compare simulated and observed column averages (1000-400 mbar average) in Fig. 6, with biases in column averages and vertical profiles shown in Figs. S2 and S3-5, respectively. Airborne observation-model comparisons complement those using surface station data. We find similar model biases to those seen in surface data, for example, larger SCA at northern high latitudes with the Jena flux product and smaller SCA at high latitudes with the DISS flux product. The airborne data also reveal three key biases that are not resolved at surface stations. Observations suggest a consistent near-zero annual mean APO in the Southern Hemisphere (south of 30°S), with a spike between 40° and 50°S. However, all three flux products show gradually decreasing annual mean APO south of 30°S, with CESM and DISS flux products showing a smaller spike in magnitude between 40° and 50°S. Simulations using CESM and DISS flux products show larger annual mean values in the northern mid-latitudes (40 - 60°N). Additionally, simulations using the Jena flux product in the low northern latitudes show a seasonal minimum day similar to the Southern Hemisphere phase. This bias is caused by low-latitude flux features in the Jena inversion that largely replicate the Southern Hemisphere cycle, likely due to limited observational constraints in this region (Jin et al., 2023).

Our analysis demonstrates that global airborne measurements provide distinct advantages over station data for evaluating large-scale flux patterns due to the reduced sensitivity of column averages to boundary-layer ATM transport uncertainties. While surface stations show substantial cross-model spread in simulated APO (Figs. 3-4), column-averaged airborne simulations (Fig. 6) reveal remarkable consistency across ATMs when driven by the same flux product. This consistency suggests that column-averaged measurements effectively integrate over local transport features that often dominate surface observations. Here we establish CESM as the most realistic flux product among the three products. The better agreement between observations and CESM-driven simulations provides a more reliable baseline for isolating and quantifying transport-related discrepancies in individual ATMs.

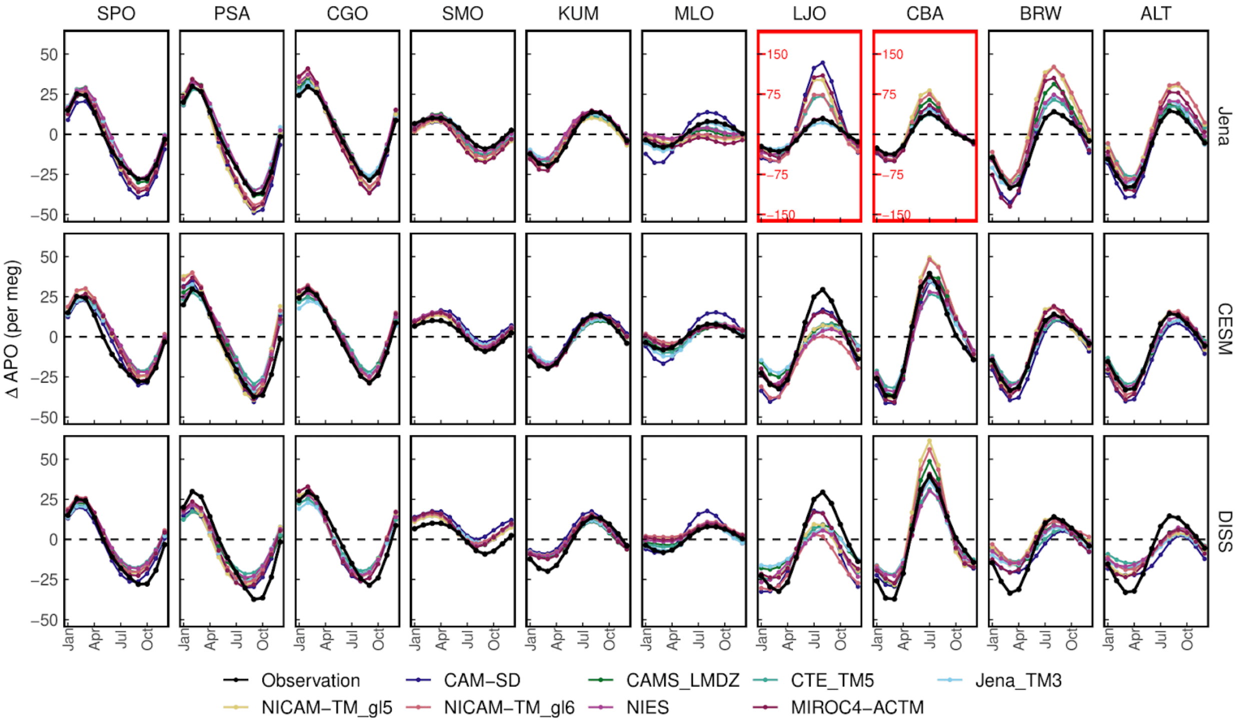
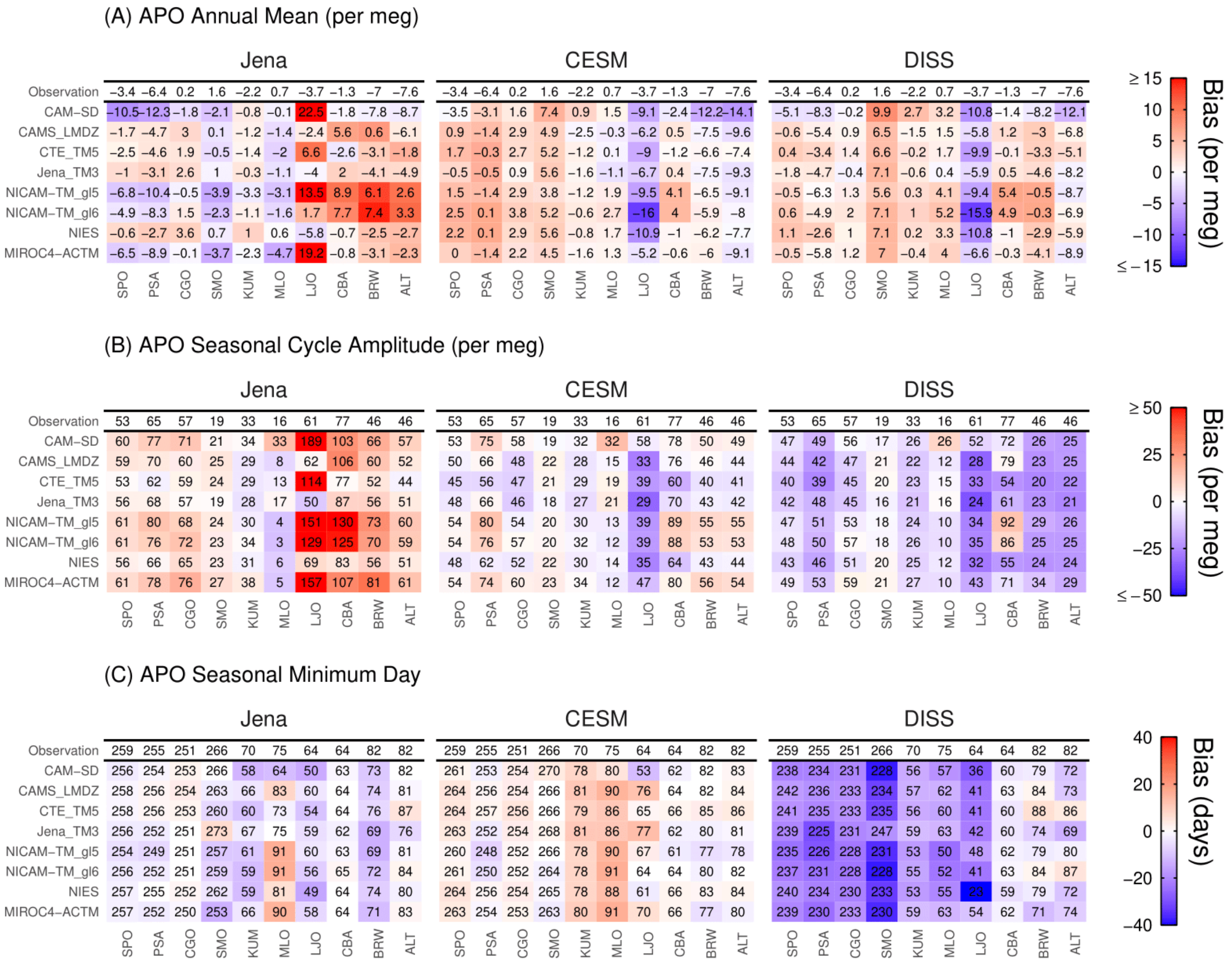


Figure 3: Comparison of simulated and observed APO seasonal cycles at 10 surface stations (Fig. 1A), organized from southern high-latitudes (left) to northern high-latitudes (right). In each panel, the black line represents observations, while colored lines show simulations from different transport models. Each row of panels corresponds to the three different flux products (Jena, CESM, and DISS). In each panel, the y-axis shows APO anomalies in per meg units, and the x-axis shows months from January to December. We note that, for LJO and CBA simulations using the Jena fluxes, a different y-axis range (three times larger) is used compared to the other panels. Observations and model simulations at each station are first detrended using a multiple-station weighted average trend. We calculate monthly mean seasonal APO from 2009 to 2018 for both observations and model simulations.



(A) APO Annual Mean (per meg)

Jena

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	−3.4	−6.4	0.2	1.6	−2.2	0.7	−3.7	−1.3	−7	−7.6
CAM−SD	−10.5	−12.3	−1.8	−2.1	−0.8	−0.1	22.5	−1.8	−7.8	−8.7
CAMS_LMDZ	−1.7	−4.7	3	0.1	−1.2	−1.4	−2.4	5.6	0.6	−6.1
CTE_TM5	−2.5	−4.6	1.9	−0.5	−1.4	−2	6.6	−2.6	−3.1	−1.8
Jena_TM3	−1	−3.1	2.6	1	−0.3	−1.1	−4	2	−4.1	−4.9
NICAM−TM_gl5	−6.8	−10.4	−0.5	−3.9	−3.3	−3.1	13.5	8.9	6.1	2.6
NICAM−TM_gl6	−4.9	−8.3	1.5	−2.3	−1.1	−1.6	1.7	7.7	7.4	3.3
NIES	−0.6	−2.7	3.6	0.7	1	0.6	−5.8	−0.7	−2.5	−2.7
MIROC4−ACTM	−6.5	−8.9	−0.1	−3.7	−2.3	−4.7	19.2	−0.8	−3.1	−2.3

CESM

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	−3.4	−6.4	0.2	1.6	−2.2	0.7	−3.7	−1.3	−7	−7.6
CAM−SD	−3.5	−3.1	1.6	7.4	0.9	1.5	−9.1	−2.4	−12.2	−14.1
CAMS_LMDZ	0.9	−1.4	2.9	4.9	−2.5	−0.3	−6.2	0.5	−7.5	−9.6
CTE_TM5	1.7	−0.3	2.7	5.2	−1.2	0.1	−9	−1.2	−6.6	−7.4
Jena_TM3	−0.5	−0.5	0.9	5.6	−1.6	−1.1	−6.7	0.4	−7.5	−9.3
NICAM−TM_gl5	1.5	−1.4	2.9	3.8	−1.2	1.9	−9.5	4.1	−6.5	−9.1
NICAM−TM_gl6	2.5	0.1	3.8	5.2	−0.6	2.7	−16	4	−5.9	−8
NIES	2.2	0.1	2.9	5.6	−0.8	1.7	−10.9	−1	−6.2	−7.7
MIROC4−ACTM	0	−1.4	2.2	4.5	−1.6	1.3	−5.2	−0.6	−6	−9.1

DISS

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	−3.4	−6.4	0.2	1.6	−2.2	0.7	−3.7	−1.3	−7	−7.6
CAM−SD	−5.1	−8.3	−0.2	9.9	2.7	3.2	−10.8	−1.4	−8.2	−12.1
CAMS_LMDZ	−0.6	−5.4	0.9	6.5	−1.5	1.5	−5.8	1.2	−3	−6.8
CTE_TM5	0.4	−3.4	1.4	6.6	−0.2	1.7	−9.9	−0.1	−3.3	−5.1
Jena_TM3	−1.8	−4.7	−0.4	7.1	−0.6	0.4	−5.9	0.5	−4.6	−8.2
NICAM−TM_gl5	−0.5	−6.3	1.3	5.6	0.3	4.1	−9.4	5.4	−0.5	−8.7
NICAM−TM_gl6	0.6	−4.9	2	7.1	1	5.2	−15.9	4.9	−0.3	−6.9
NIES	1.1	−2.6	1	7.1	0.2	3.3	−10.8	−1	−2.9	−5.9
MIROC4−ACTM	−0.5	−5.8	1.2	7	−0.4	4	−6.6	−0.3	−4.1	−8.9

(B) APO Seasonal Cycle Amplitude (per meg)

Jena

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	53	65	57	19	33	16	61	77	46	46
CAM−SD	60	77	71	21	34	33	189	103	66	57
CAMS_LMDZ	59	70	60	25	29	8	62	106	60	52
CTE_TM5	53	62	59	24	29	13	114	77	52	44
Jena_TM3	56	68	57	19	28	17	50	87	56	51
NICAM−TM_gl5	61	80	68	24	30	4	151	130	73	60
NICAM−TM_gl6	61	76	72	23	34	3	129	125	70	59
NIES	56	66	65	23	31	6	69	83	56	51
MIROC4−ACTM	61	78	76	27	38	5	157	107	81	61

CESM

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	53	65	57	19	33	16	61	77	46	46
CAM−SD	53	75	58	19	32	32	58	78	50	49
CAMS_LMDZ	50	66	48	22	28	15	33	76	46	44
CTE_TM5	45	56	47	21	29	19	39	60	40	41
Jena_TM3	48	66	46	18	27	21	29	70	43	42
NICAM−TM_gl5	54	80	54	20	30	13	39	89	55	55
NICAM−TM_gl6	54	76	57	20	32	12	39	88	53	53
NIES	48	62	52	22	30	14	35	64	43	44
MIROC4−ACTM	54	74	60	23	34	12	47	80	56	54

DISS

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	53	65	57	19	33	16	61	77	46	46
CAM−SD	47	49	56	17	26	26	52	72	26	25
CAMS_LMDZ	44	42	47	21	22	12	28	79	23	25
CTE_TM5	40	39	45	20	23	15	33	54	20	22
Jena_TM3	42	48	45	16	21	16	24	61	23	21
NICAM−TM_gl5	47	51	53	18	24	10	34	92	29	26
NICAM−TM_gl6	48	50	57	18	26	10	35	86	25	25
NIES	43	46	51	20	25	12	32	55	24	24
MIROC4−ACTM	49	53	59	21	27	10	43	71	34	29

(C) APO Seasonal Minimum Day

Jena

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	259	255	251	266	70	75	64	64	82	82
CAM−SD	256	254	253	266	58	64	50	63	73	82
CAMS_LMDZ	258	256	254	263	66	83	60	64	74	81
CTE_TM5	258	256	253	260	60	73	54	64	76	87
Jena_TM3	256	252	251	273	67	75	59	62	69	76
NICAM−TM_gl5	254	249	251	257	61	91	60	63	69	81
NICAM−TM_gl6	256	252	251	259	59	91	56	65	72	84
NIES	257	255	252	262	59	81	49	64	74	80
MIROC4−ACTM	257	252	250	253	66	90	58	64	71	83

CESM

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	259	255	251	266	70	75	64	64	82	82
CAM−SD	261	253	254	270	78	80	53	62	82	83
CAMS_LMDZ	264	256	254	266	81	90	76	64	82	84
CTE_TM5	264	257	256	266	79	86	65	66	85	86
Jena_TM3	263	252	254	268	81	86	77	62	80	81
NICAM−TM_gl5	260	248	252	266	78	90	67	61	77	78
NICAM−TM_gl6	261	250	252	264	78	91	64	64	80	82
NIES	264	256	254	265	78	88	61	66	83	84
MIROC4−ACTM	263	254	253	263	80	91	70	66	77	80

DISS

	SPO	PSA	CGO	SMO	KUM	MLO	LJO	CBA	BRW	ALT
Observation	259	255	251	266	70	75	64	64	82	82
CAM−SD	238	234	231	228	56	57	36	60	79	72
CAMS_LMDZ	242	236	233	234	57	62	41	63	84	73
CTE_TM5	241	235	233	235	56	60	41	60	88	86
Jena_TM3	239	225	231	247	59	63	42	60	74	69
NICAM−TM_gl5	235	226	228	231	53	50	48	62	79	80
NICAM−TM_gl6	237	231	228	228	55	52	41	63	84	87
NIES	240	234	230	233	53	55	23	59	79	72
MIROC4−ACTM	239	230	233	230	59	63	54	62	71	74

Figure 4: Evaluation of APO (A) annual mean relative to a multi-station global mean, (B) seasonal cycle amplitude, and (C) seasonal minimum day across surface stations using different flux-transport model combinations. For each panel, results are organized by flux products (JENA, CESM, DISS) in columns and transport models in rows, with observations on the top. The metrics are printed in black, with background colors indicating biases relative to observations. Positive bias is shown in red, and negative bias is shown in blue. Stations from left to right are organized by latitudes from south to north.

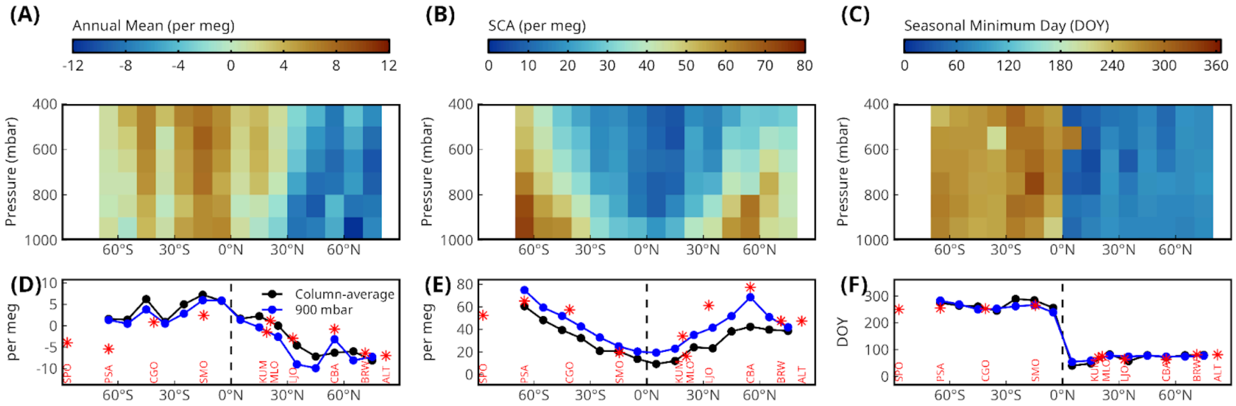


Figure 5: APO annual means (A and D), SCA (B and E), and seasonal minimum day (C and F) derived from airborne observations. In A-C, we show latitude-pressure distributions, with data binned into 10 deg latitude by 100-mbar pressure boxes. In D-F, we show 1000-400 mbar column-averaged (black) and 900-mbar interpolated (blue) values, and also surface station observations (2009 to 2018). Annual mean is derived from a two-harmonic fit with constant offset, where the global multi-station trend has been subtracted to detrend the airborne observations and center the values around zero globally. SCA is calculated as the peak-to-trough amplitude of the two-harmonic fit, and seasonal minimum day is calculated as the day of seasonal trough of the two-harmonic fit.

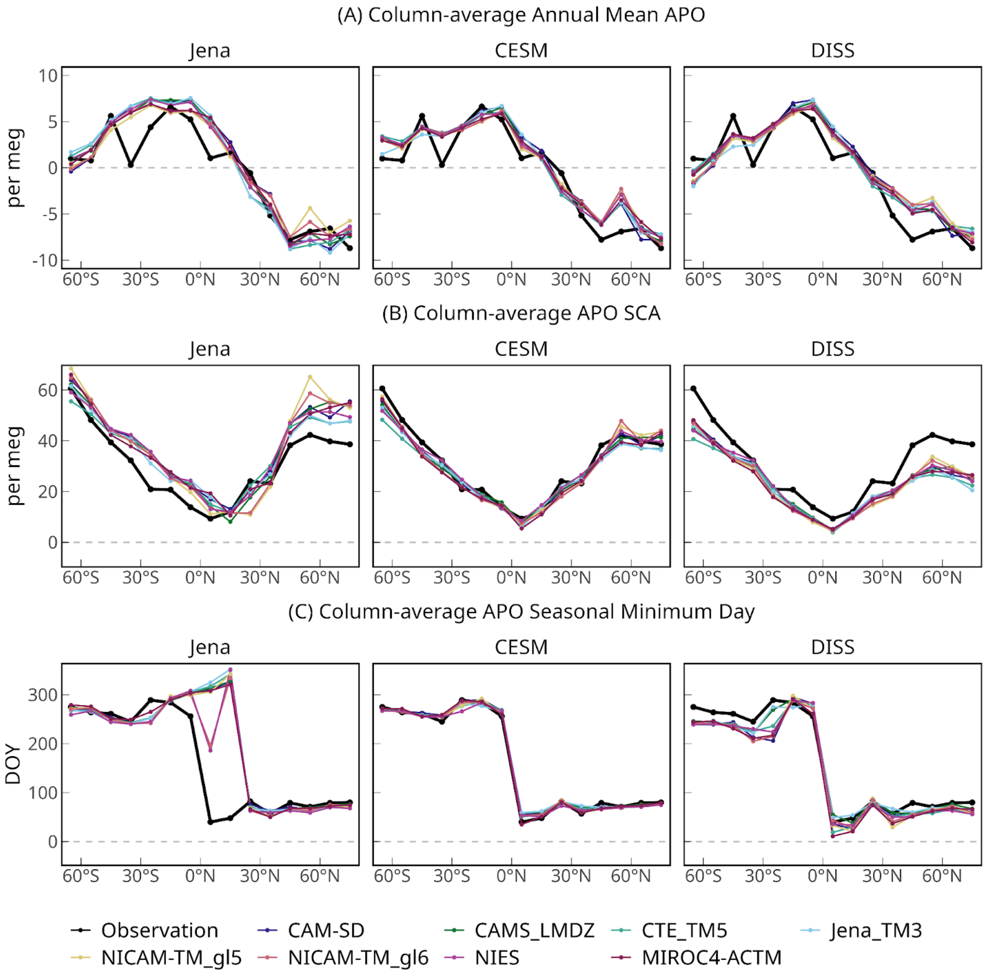


Figure 6: Comparison of column-average (1000-400 mbar) APO features across latitude from aircraft observations and model simulations using three different flux products (Jena, CESM, and DISS). The figure is organized into three sets of panels showing (A) annual mean APO relative to a multi-station global mean, (B) SCA, and (C) seasonal minimum day. For each feature, we show latitudinal distributions of observations (black lines) and model simulations (colored lines). We note that the global mean value has been subtracted from the annual mean values (A) at each latitude to highlight spatial patterns. We show the column-average (400-1000 mbar) seasonal cycles of observed and simulated APO for each 10° latitude band in Fig. S1.

3.2. Evaluation of diabatic mixing rates diagnosed from transport models

In this section, we evaluate the mixing timescale across mid-latitude moist isentropes of each ATM using the framework developed in Jin et al. (2024). This framework was applied to identify biases in four ATMs in the mid-latitude Southern Hemisphere using two independent constraints: (1) diagnosed diabatic mixing rates, and (2) cross-isentrope CO_2 gradients. Here we extend the framework to use APO gradients, to include two more reanalysis products, and the analysis in the Northern Hemisphere. We evaluate six of the eight ATMs participating in APO-MIP1 that provide 3-D atmospheric fields (CAM-SD, CTE_TM5, Jena_TM3, NICAM-TM_gl5, NICAM-TM_gl6, and MIROC4-ACTM), which are required to diagnose diabatic mixing rates. Diabatic mixing rates and APO gradients are diagnosed based on the mass-indexed isentropic coordinate $M_{\theta e}$, which was first introduced by Jin et al. (2021). For each pair of transport models and flux products, we resolve cross-$M_{\theta e}$ diabatic mixing rates and cross-$M_{\theta e}$ APO gradients in the mid-latitudes of both hemispheres. We use observation-based diabatic mixing constraints diagnosed from four meteorological reanalyses, and observed APO gradient constraints calculated from three airborne campaigns. The detailed methodology for calculating $M_{\theta e}$ surfaces, diabatic mixing rates, and cross-$M_{\theta e}$ APO gradients is provided in Appendix C.

We show the climatological monthly mean diabatic mixing rates of two $M_{\theta e}$ surfaces in the Southern Hemisphere in Fig. 7, as well as schematics of the geographic distribution of the two $M_{\theta e}$ surfaces. For each ATM, mixing rates in Fig. 7 are calculated from APO and averaged over three realizations diagnosed from using three flux products. The reanalysis mixing rates are calculated from moist static energy (MSE) budget and shown as average and 1σ spread over the four reanalysis products. The six ATMs and the reanalyses show diabatic mixing rates with clear seasonal cycles, suggesting more rapid mixing across isentropes in the austral winter than summer. ATMs generally overestimate diabatic mixing rates, especially in the summer and winter, when there are large cross-$M_{\theta e}$ APO gradients that lead to well-defined mixing rates. Among the six ATMs, CTE_TM5 and Jena_TM3 show too rapid mixing that is biased high in all seasons. The other four ATMs align better with reanalysis, but still show significant overestimation for most of the year. MIROC4-ACTM shows the best performance. These findings align with Jin et al. (2024), which previously identified that the southern hemisphere summer-time mixing rates are overestimated in ATMs used for CO_2 inversions, with consistent results for the three ATMs (MIROC4-ACTM, Jena_TM3, and CTE_TM5) being used in both studies.

We find that biases in diagnosed diabatic mixing rates correlate with biases in cross-$M_{\theta e}$ APO gradients in each season, with stronger diabatic mixing leading to smaller APO gradients (Fig. 8). Fig. 8 shows the ATM-diagnosed diabatic mixing rates and simulated APO gradients (points) across six transport models and three flux products at two $M_{\theta e}$ surfaces (30 and 45×10^{16} kg $M_{\theta e}$) for three selected 2-month periods in the Southern Hemisphere. The points suggest clear linear relationships between diagnosed mixing rates and simulated APO gradients for each flux product (shown as fit lines for each flux product). The linear relationships persist across all seasons and $M_{\theta e}$ surfaces, though with varying slopes depending on the underlying fluxes (Fig. 8). ATMs generally underestimate cross-$M_{\theta e}$ absolute APO gradients (i.e., a closer to zero gradient) at both $M_{\theta e}$ surfaces, corresponding to the overestimation of diabatic mixing rates in these models. For each flux product, biases in cross-$M_{\theta e}$ APO gradients are always larger in fast mixing ATMs (e.g., Jena_TM3 and CTE_TM5) compared to slow mixing ATMs (e.g., two versions of NICAM-TM, MIROC4-ACTM, and CAM-SD), with MIROC4-ACTM showing the best agreement. For each transport model, the simulated gradient shows clear spread across different flux products. The largest spread occurs in austral winter and spring (Fig. 8C-D), when simulations with the DISS fluxes show much larger gradients compared to CESM or Jena fluxes. We note that the direct comparison of simulated and observed gradients for individual models is complicated by the interplay of ATM biases and flux product biases.

To evaluate flux products independently of transport model biases, we leverage both diabatic mixing rates and APO gradients. For each flux product, the intersection between the mixing rate-gradient linear fit and the MSE-diagnosed mixing rate indicates the expected APO gradient with realistic mixing characteristics. Therefore, we can evaluate large-scale flux features in the flux products by comparing this expected gradient to the observed gradient. Our analysis in Fig. 8 suggests that CESM is the most realistic flux product in the mid-latitude Southern Hemisphere in all seasons. The expected CESM gradients (intersections of thin blue line and vertical gray band) fall within the observation uncertainty range in all seasons and surfaces except austral summer at the 30×10^{16} kg $M_{\theta e}$ surface (Fig. 8A), which suggests a slight underestimation of uptake in the CESM product. The expected gradients of the Jena flux product also generally fall within the observation uncertainty range, but shows an even larger underestimation in Fig. 8A. The expected gradients of the DISS flux product have large biases in the mid-latitude Southern Hemisphere. The expected gradient is significantly larger in the austral winter (Fig. 8C-D), and significantly smaller at the 30 × 10^{16} kg $M_{\theta e}$ surface in austral summer (Fig. 8A) and austral spring (Fig. 8E), suggesting seasonal biases in the flux pattern.

Biases in expected gradients relative to observed gradients result from errors in the magnitude and spatial distribution of air-sea APO flux, specifically the difference in flux magnitudes between regions north and south of the target $M_{\theta e}$ surface. For instance, a positive expected gradient bias during austral summer at the 30 × 10^{16} kg $M_{\theta e}$ surface (Fig. 8A) in the DISS product could stem from underestimated outgassing in high southern latitudes, excessive outgassing in lower latitudes, or both. In addition, a flux product could produce realistic expected gradients despite underestimating absolute fluxes both north and south of the $M_{\theta e}$ surface if the difference remains correct. Resolving these inherent ambiguities requires additional observational constraints from surface stations, ships, and aircraft, which we addressed in Section 3.1.

While the focus of Jin et al. (2024) was on the mid-latitude Southern Hemisphere, we extend our analysis of the mid-latitude diabatic mixing rates to the Northern Hemisphere at the 45 × 10^{16} kg $M_{\theta e}$ surface (Fig. 9). ATMs also generally overestimate diabatic mixing rates in the Northern Hemisphere, except during summer (JJA). Whereas MSE-diagnosed mixing rates peak in northern summer, ATM-diagnosed mixing rates have their seasonal minimum at this time. We note that APO gradients in ATMs are close to zero during JJA, leading to poorly defined diabatic mixing rates. We carry out the same transport model and flux product analyses in the Northern Hemisphere in January to March (Fig. 10A) and August to October (Fig. 10B). MIROC4-ACTM still demonstrates the closest agreement with reanalysis data in both seasons, and CTE_TM5 shows the largest mixing rate bias. We note that TM3 and TM5 are based on similar parameterization schemes, but TM3 outperforms TM5. In both seasons, the expected gradients inferred from CESM flux align with the airborne observations, while Jena and DISS overestimate and underestimate expected gradients, respectively.

Our attempt to diagnose mixing rates in ATMs in the Northern Hemisphere mid-latitudes using ocean tracers alone is partly limited by the predominantly land surface. We find both summer and winter peaks in seasonal diabatic mixing rates in the northern mid-latitudes, driven by strong convection. Over land, convection peaks in summer due to strong surface heating that creates unstable atmospheric conditions. Over the ocean, however, convection peaks in winter due to larger air-sea temperature differences. Our ATM-diagnosed mixing rates in the Northern

Hemisphere may not capture the summer peak because atmospheric mixing processes over land may not be adequately reflected in transport of air-sea APO flux signals, which occurs initially over the ocean. This limitation is particularly significant in the Northern Hemisphere, where zonal mixing is slower (2-4 weeks) due to topographic blocking and stationary wave patterns. We plan to diagnose the land and ocean contrast in atmospheric diabatic mixing in the next APO-MIP1 by also forward transporting land tracers (e.g., CO_2 sources/sinks from the land biosphere). Our method is more robust in the Southern Hemisphere mid-latitudes due to faster zonal mixing (1-2 weeks) and the predominantly ocean surface. We also note that the distinct thermal capacities of land and ocean in the Northern Hemisphere create more complex surface $M_{\theta e}$ outcrops with larger latitudinal shifts across seasons (Jin et al., 2021), as shown in Fig. 9C. We, however, account for these shifts in our analysis.

Our analysis reveals that the ATM-diagnosed diabatic mixing rate primarily reflects an intrinsic characteristic of the transport model, at least in the Southern Hemisphere, showing little sensitivity to the underlying flux pattern, tracers, and land-ocean differences, particularly in models with smaller mixing rates (i.e., two versions of NICAM-TM, MIROC4-ACTM, and CAM-SD). These four models demonstrate consistent mixing rates across different flux products (Figs. 8 and 10). This consistency is further supported by our analysis of diagnosed mixing rates for individual APO components (ΔO_2^{ocn}, ΔN_2^{ocn}, ΔCO_2^{ocn}, ΔO_2^{ff}, and ΔCO_2^{ff}) transported by ATMs with smaller mixing rates, which yields similar mixing rates despite these tracers having distinct signs, seasonal patterns, and magnitudes (Fig. S6). However, ATMs with faster mixing rate (e.g., Jena_TM3 and CTE_TM5) show large variability both across flux products (Fig. 9-10) and across tracers (Fig. S6). Notably, these two models exhibit approximately 50% slower diagnosed mixing rates for the fossil fuel CO_2 tracer (ΔCO_2^{ff}) compared to the other ocean flux tracers in the austral summer at the 30×10^{16} kg $M_{\theta e}$ surface. We note that the fossil fuel CO_2 tracer has its main source in the Northern Hemisphere, and its mixing at the mid-latitude Southern Hemisphere preferentially occurs in the upper troposphere. In contrast, the air-sea flux tracers have significant sources/sinks over the Southern Ocean with rapid cross-isentrope mixing preferentially in the lower troposphere. This behavior suggests that these models simulate distinctly different mixing patterns between the planetary boundary layer (0-2 km) and the free troposphere. Specifically, these models appear to have excessive vertical mixing in the boundary layer while maintaining more realistic transport in the free troposphere. Our method, however, assumes a constant cross-$M_{\theta e}$ diabatic mixing rate over the entire $M_{\theta e}$ surface. The excessive boundary layer mixing causes the diagnosed mixing rates in these models to be overly sensitive to the specific vertical distribution of air-sea APO flux components.

Our evaluation of ATMs using simulations from APO-MIP1 advances the original framework of Jin et al. (2024) in two key aspects. First, we expand the experimental design by increasing the number of participating ATMs to six and employing three different flux fields with each ATM, generating 18 model realizations. This comprehensive matrix of simulations enables a more systematic evaluation of both transport and flux-related biases. We demonstrate how atmospheric tracer observations can be leveraged to independently evaluate and distinguish between biases in surface fluxes and atmospheric transport models. Second, we enhance the robustness of our MSE-diagnosed mixing rate calculations by incorporating two additional reanalysis products and computing mixing rates at the native high resolution of each reanalysis, rather than averaging to a coarser grid before the calculation. One limitation in our method is that we only use $M_{\theta e}$ calculated from MERRA-2 for each of the transport models rather than using $M_{\theta e}$ calculated from the individual transport model, which in principle can be done by interpolating the temperature and humidity from parent reanalysis to the ATM grid. This limitation would lead to slight inconsistency between the actual $M_{\theta e}$ in the model and the value we assigned to it. However, the differences between $M_{\theta e}$ calculated from different reanalyses remain small and our method ensures consistency in geography of each $M_{\theta e}$ surface (Jin et al., 2021).

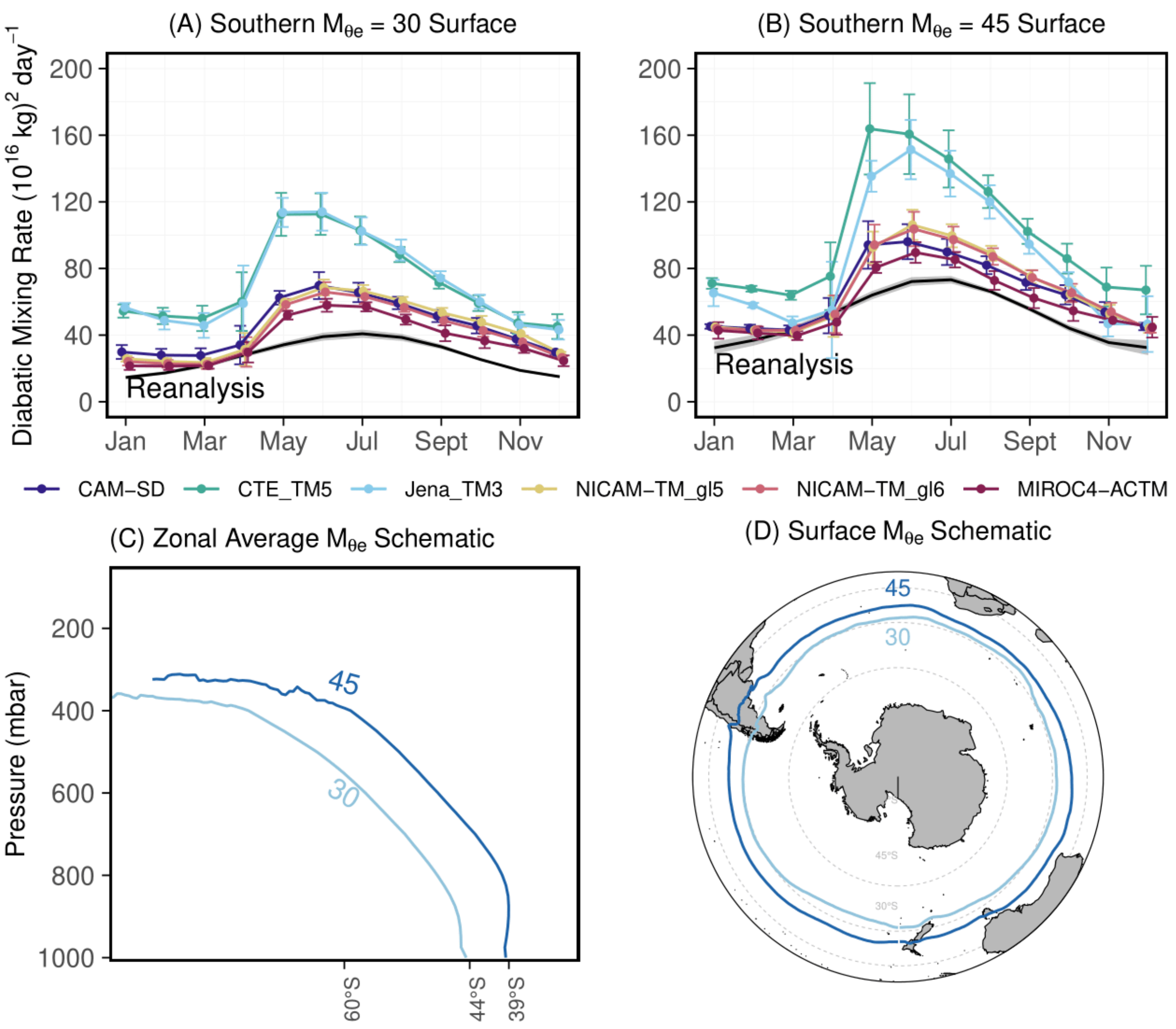


Figure 7: Climatological monthly diabatic mixing rates across the (A) 30 and (B) 45 (10^{16} kg) $M_{\theta e}$ surfaces in the Southern Hemisphere. ATM-diagnosed mixing rates are derived from six ATMs in APO-MIP1 that provide 3-D APO fields. Error bars represent the 1σ spread across the 30 and 45 × 10^{16} kg $M_{\theta e}$ of three flux products used here. Black lines represent MSE-diagnosed mixing rates as the average of four reanalysis MSE budgets, while the gray shaded regions represent the 1-sigma spread. (C) Schematic showing latitude-pressure distribution of troposphere zonal annual average $M_{\theta e}$, and (D) annual average near-surface $M_{\theta e}$ contours of the 30 and 45 (10^{16} kg) surfaces, computed from MERRA-2 reanalysis for the year 2009. These two $M_{\theta e}$ surfaces have very small seasonal meridional variability.

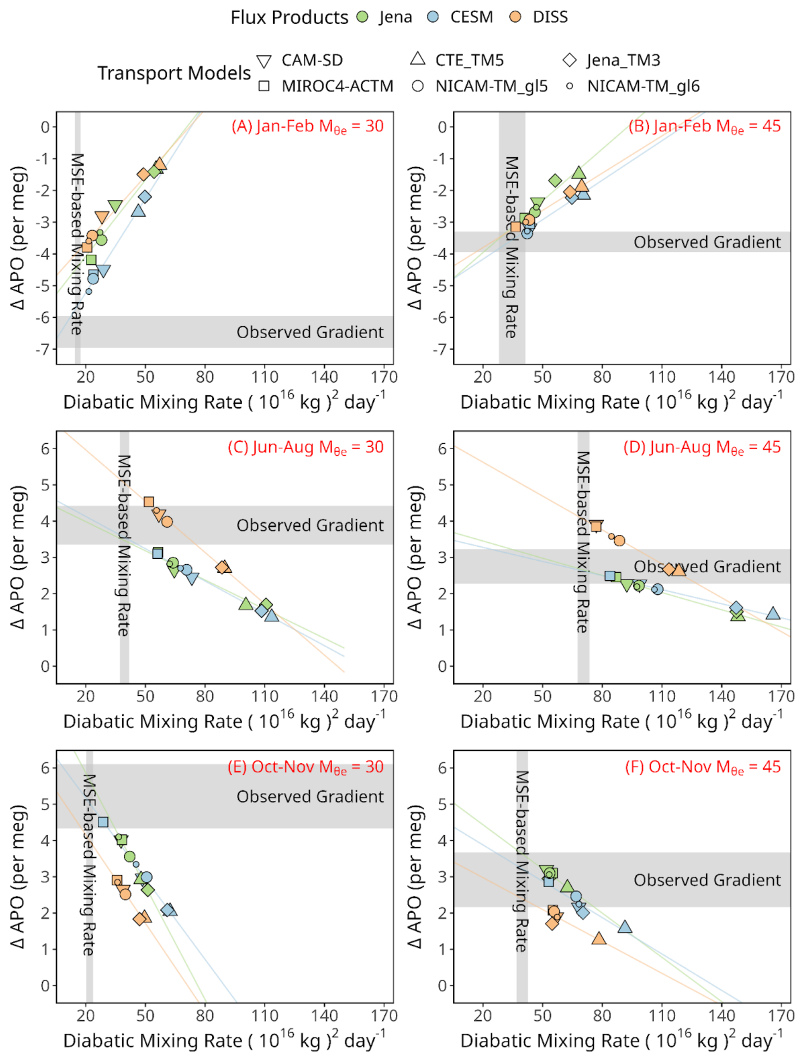


Figure 8: Using MSE-based diabatic mixing rates and airborne observations of cross-isentrope APO gradients to evaluate ATMs and flux models. Each panel compares model-diagnosed diabatic mixing rates (x-axis) and cross-$M_{\theta e}$ APO gradients (y-axis) at the 30 × 10^{16} kg $M_{\theta e}$ surface (A, C, E, ~44°S surface outcrop) and at 45 × 10^{16} kg $M_{\theta e}$ (B, D, F, ~39°S surface outcrop). Results are shown for three seasonal periods: Jan-Feb (A-B), Jun-Aug (C-D), and Oct-Nov (E-F) based on available airborne campaigns. Points represent individual model simulations, with colors indicating flux products (Jena, CESM, DISS) and symbols denoting different ATMs. Vertical gray bands show the 1σ range of MSE-based mixing rates derived from four reanalysis products. Horizontal gray bands indicate the 1σ range of observed APO gradients after spatial and temporal bias correction. Colored lines show linear fits of mixing rates and APO gradients for each flux product across different transport models.

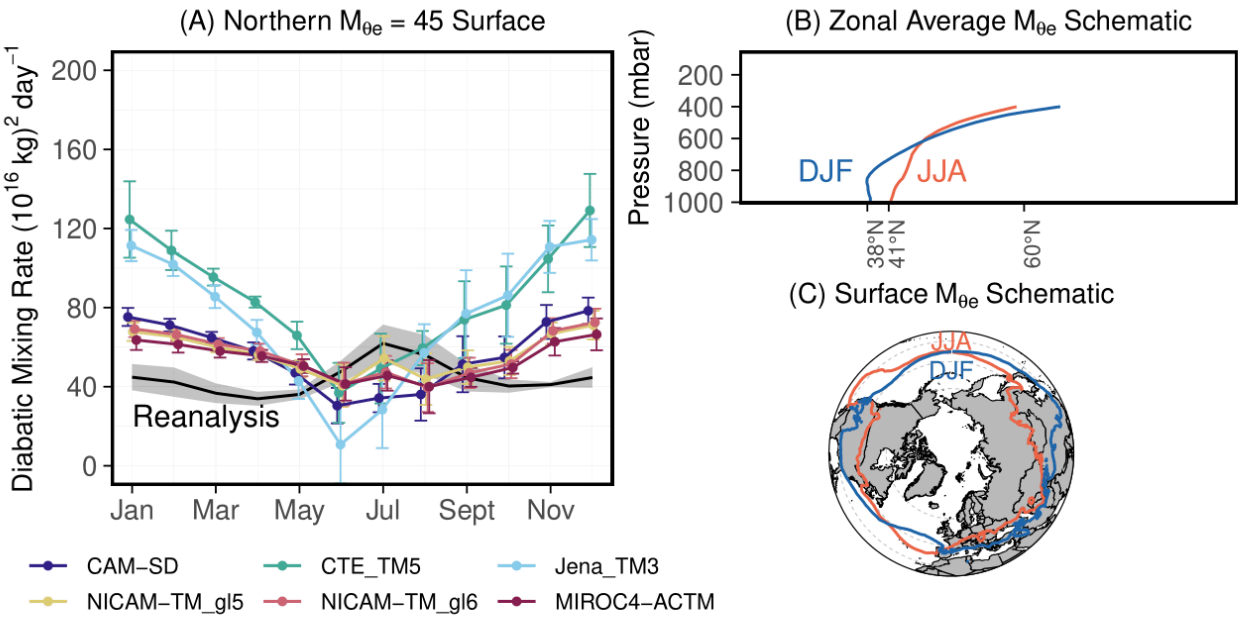


Figure 9: (A) Similar to Fig. 7, but showing climatological monthly diabatic mixing rates across the 45 (10^{16} kg) $M_{\theta e}$ surface in the Northern Hemisphere. We note that JJA diabatic mixing rates in ATMs are poorly constrained due to close-to-zero cross-$M_{\theta e}$ APO gradients. (B) Latitude-pressure distribution of zonal average 45×10^{16} kg $M_{\theta e}$ surfaces during boreal summer (JJA) and winter (DJF). The two $M_{\theta e}$ surfaces end at the tropopause, which is higher in the summer in the mid-latitudes. (C) Corresponding Earth surface outcrops of the JJA and DJF 45×10^{16} kg $M_{\theta e}$ surfaces. Unlike in the Southern Hemisphere where seasonal meridional variations in $M_{\theta e}$ surfaces are small, the Northern Hemisphere shows pronounced seasonal shifts due to different land/ocean heating and cooling cycles.

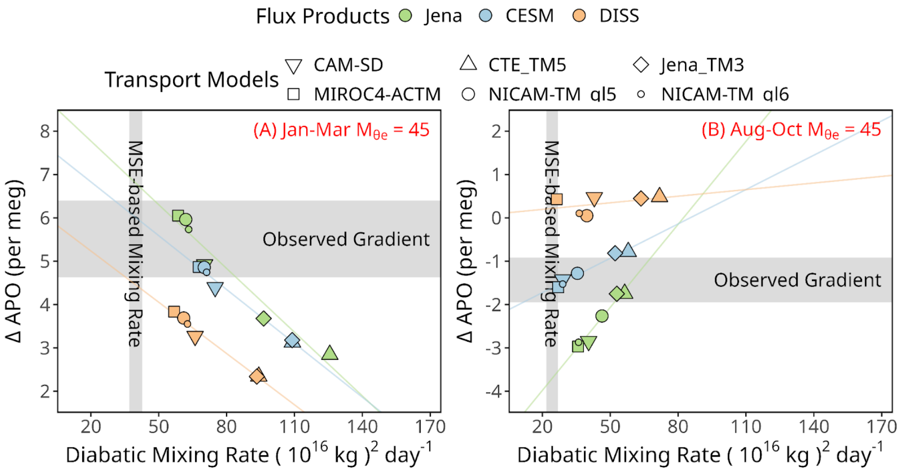


Figure 10: Similar to Fig. 8, but showing diabatic mixing rates and cross-$M_{\theta e}$ APO gradients in the Northern Hemisphere late winter / early spring (A) and late-summer / early fall (B) of the 45 × 10^{16} kg $M_{\theta e}$ surface. We choose January to March and August to October due to sufficient aircraft sampling and maximum cross-$M_{\theta e}$ APO gradients in these months.

3.3. Shipboard model-observation comparison over the Drake Passage

The APO-MIP1 simulations could not reproduce latitudinal variations in APO seasonal cycle amplitude observed from shipboard measurements from 53 to 65°S over the Drake Passage and adjacent to Tierra del Fuego and the Antarctic Peninsula. Observations reveal a strong meridional SCA gradient (-2.1 per meg deg^{-1}, with deg positive northward), with SCA increasing sharply towards higher southern latitudes (Fig. 11). Model simulations substantially underestimate this latitudinal gradient (Fig. 11), showing weaker slopes averaged across ATMs of -1.2 (Jena), -0.5 (CESM), and 0.8 (DISS) per meg deg^{-1}. Notably, these gradients remain generally consistent across different ATMs for each flux product (±0.26, ±0.13, and ±0.29 per meg deg^{-1}, respectively), suggesting this may predominantly be a result of zonal-scale latitudinal biases in flux seasonality. Underrepresentation of enhanced summertime productivity along the coast of the Antarctic Peninsula in flux products could also play a role. However, the Gould typically only transits waters with elevated chlorophyll south of approximately 62°S while the gradient biases appear further north. Furthermore, seasonally, the SCA biases are caused more by underestimation of the winter/spring drawdown in APO at high latitudes, rather than the smaller underestimation of summertime APO enhancement (Figs. S10-11). For CESM, this bias could originate from incomplete process representation in the ocean biogeochemistry model and the underestimation of winter mixed-layer depths in the Pacific sector of the Southern Ocean, which has historically been a problem for Earth System Models (Sallée et al., 2013). The Jena flux product provides the closest match to the observed SCA gradient. However, several limitations remain, which likely stem from the coarse spatial resolution, limited atmospheric observational constraints over the Southern Ocean, and underrepresentation of mixing patterns around the PSA station (see details below and in SI). The DISS flux product is biased due to its underlying assumptions and sparse observational constraints, as discussed in Jin et al. (2023).

Across ATMs, we find systematic differences of up to ±20% in simulated mean SCA for the entire ship transects over the Drake Passage, independent of the input flux field, with CTE_TM5 consistently producing the smallest SCA and NICAM-TM_gl5 showing the largest. These differences across ATMs are likely caused by differences in marine boundary-layer ventilation in the models. Near-surface mixing over the Southern Ocean is challenging to model, owing to complex boundary-layer structure, strong wind shear, frequent storm systems, SST variations, and poorly represented clouds (Hyder et al., 2018; Knight et al., 2024; Lang et al., 2018; Truong et al., 2020). The coarse-resolution models used here may struggle to capture such phenomena, and the resulting variations in the concentration or dilution of flux signals near the surface drives differences in mean APO SCA. The systematic spread also likely reflects biases in the representation of large-scale diabatic mixing over the high southern latitudes. Models with strong diabatic mixing rates, such as TM5, tend to dilute the meridional gradient of seasonal amplitude through excessive mixing with lower-latitude air masses that have smaller SCAs, resulting in reduced amplitudes at high southern latitudes.

We find that observed SCA at PSA (64.5°S) from SIO flask measurements (~ 70 per meg, averaged from 2012 to 2017) is significantly smaller than nearby ship data from 64°S to 65°S (~ 80 per meg). However, model simulations suggest similar values for both locations. The shipboard measurements are closely tied to the SIO O_2 calibration scale, and any remaining scale differences would be unlikely to affect the seasonal APO SCA. Rather, the observed SCA difference occurs because SIO flask samples collected at PSA predominantly sample descending air masses from the east that have passed over Anvers Island and the Antarctic Peninsula, with peaks above 2000 m (characterized by small APO SCA), whereas the ship samples marine boundary layer air including that over highly productive ocean regions (large APO SCA). As shown in Figs. S7-9, the SIO flasks are collected from the Terra Lab, on the east side of the station, with a wind selection criteria of 5-205°. Even while docked at Palmer (left-most points in Fig. 11), the Gould measurements show elevated SCA compared to PSA flask samples, because the pier is located to the west of the station with samples filtered to exclude air influenced by the station (Figs. S7-9). None of the ATMs, regardless of the flux product used, could reconstruct this feature, even though the models were sampled at the flask collection times. This difference is consistent with that seen between 900-mbar airborne samples and PSA flasks (Fig. 5E). The systematic bias points to the lack of resolution or physics that would be necessary, in either the reanalysis products or the ATMs, to accurately capture fine-scale circulation patterns, particularly the distinct air mass origins affecting ship versus station measurements. We note that the Jena flux product has been optimized to match seasonal APO cycles at Cape Grim Observatory (41°S) and at PSA (64.5°S), which may be the reason for its better performance on the SCA latitudinal gradient. It may do even better if the shipboard data were used in the inversion or if the effective sampling altitude of the SIO flasks at PSA were better accounted for.

Our analysis underscores the need for improvements in both ocean biogeochemistry models and ATMs. Future ocean process model developments should include improving accuracy of winter mixed-layer depths and higher-resolution ocean models with enhanced process representation to capture the fine-scale productivity patterns in the Southern Ocean. Additionally, current atmospheric transport models require improved resolution and physics to better represent the complex circulation patterns characteristic of coastal regions.

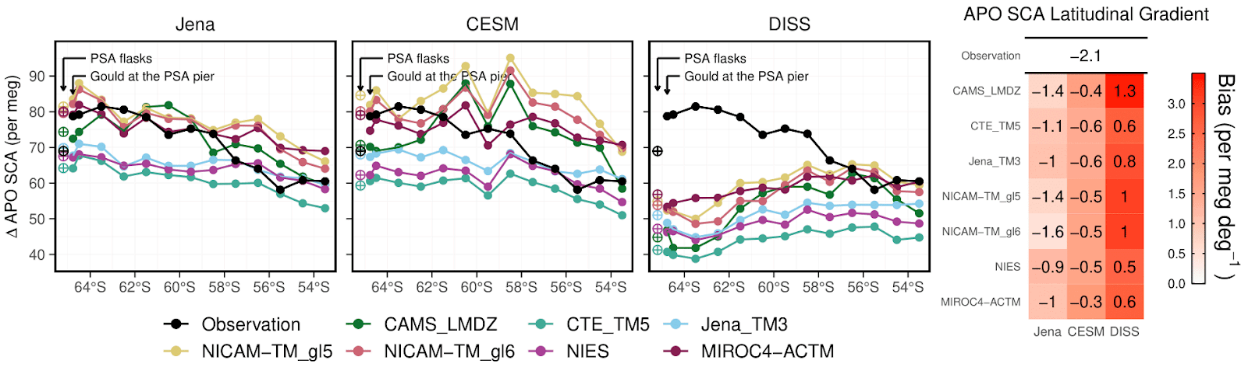


Figure 11: Latitudinal distribution of APO SCA across the Drake Passage region (53°S-65°S) derived from ship observations and model simulations. We calculate SCA by grouping observations and model simulations into 1 deg latitude bands, shown as points. Model results are color-coded by ATM and organized by flux products in separate panels. The full seasonal cycles of observed and simulated APO of these latitude bands are shown in Fig. S10. We also show SCA observed and simulated for the PSA flask record as open crossed circles (~64.5°S, shifted 0.7° south for visibility), and for ship data while the Gould is docked at or close to the PSA pier (left-most points, calculated by selecting data from 64.82°S to 64.72°S and 64.1°W to 64.0°W). The right-most three bands (53°S to 55°S) are typically downwind of Tierra del Fuego (Figs. S7-9). Both observational and model data for each latitude band or at PSA were detrended using corresponding cubic smooth spline fits from SPO. SCA was calculated using two-harmonic fits. The rightmost panel shows the SCA latitudinal gradients (per meg deg^{-1}) from 53°S to 65°S, with red shading indicating model biases relative to observations. The gradient is calculated as linear fits of SCA from 53°S to 65°S for each ATM and flux product pair, and the observations. We exclude CAM-SD in this analysis because the ship data simulation is only available from 2012 to 2015 (i.e., missing 2016 to 2017 data).

3.4. Implications for APO and CO_2 inversions and ATM development

Our study motivates a community effort to conduct APO inversions. Estimates of spatial and temporal variations in APO fluxes can improve our understanding of ocean biogeochemical processes and heat transport, and support verification of fossil-fuel emission estimates (Pickers et al., 2022; Rödenbeck et al., 2023). Currently, only one global-scale APO inversion product from Jena CarboScope (Rödenbeck et al., 2008) exists. This product shows excessive seasonal flux amplitudes (Fig. 2) in the southern low-latitudes (~ 30 to 0°S) and northern mid-latitudes (~ 30 to 60°N) relative to the other two flux products, which show better consistency with aircraft observations in their forward transport simulations (Fig. 6). These biases in Jena APO inversion partly result from limitations in the TM3 model, which exhibits excessive vertical mixing, particularly in the eastern North Pacific, too rapid diabatic mixing in the southern mid-latitudes, and underrepresentation of monsoon dynamics primarily due to coarse resolution (Jin et al., 2023). The large spread and biases in ATMs shown in this study highlight the importance of developing APO inversions using different ATMs and methodologies, as this will improve our ability to fully assess methodological uncertainties and potential biases in inverted air-sea APO flux estimates.

We encourage future inversion efforts to also assimilate column-mean data from airborne campaigns, in addition to sparse surface stations, especially for studying climatological seasonal fluxes. Our study finds that forward simulations from ATMs generally show large spread at northeastern Pacific sites, particularly at LJO and CBA (Fig. 2), where simulations are sensitive to model representation of the marine boundary layer and vertical mixing. The Scripps APO observation network consists mainly of stations along a Pacific transect close to the primary oceanic sources and sinks. Given this limited spatial coverage and our findings of significant vertical mixing biases (e.g., at CBA and LJO) and local wind-direction biases (e.g., at LJO and PSA) in ATMs at the station level, APO inversions that rely solely on these surface observations may be subject to large representation errors. Airborne data, however, provide larger surface footprints and column average metrics that are much less sensitive to vertical mixing biases. Our analysis shows that ATMs are generally consistent with each other in simulating large-scale annual and seasonal column-mean features along flight tracks (Fig. 6). Thus, inversions configured to assimilate airborne column-mean observations would be promising. Further improvement could also be achieved by incorporating shipboard observations to expand zonal coverage, such as from the Gould, across the Atlantic (Pickers et al., 2017), and in the Western Pacific (Tohjima et al., 2012). The study of Jin et al. (2023) used a different configuration of the Jena inversion that also assimilated Japanese ship-based observations across the western Pacific (Tohjima et al., 2012) from 40°S to 50°N. Forward transport of APO fluxes in that configuration aligns better with station and airborne data compared to the configuration used in this study, particularly in reducing the SCA bias in the tropics, suggesting better flux representations.

Biases in diabatic mixing diagnosed from ATMs (Section 3.2) imply that CO_2 inversions using these ATMs are also likely biased. A previous study showed that summer-time Southern Ocean CO_2 estimates from inversion products are correlated with corresponding simulated summer-time cross-isentrope CO_2 gradients in inversions (Long et al., 2021b). The simulated gradients are shown to be biased too small due to too rapid diabatic mixing bias in ATMs leading to an overestimation of Southern Ocean CO_2 uptake in the summer (Jin et al., 2024). It is likely that biases in ATMs also contribute to the large spread found in OCO-2 MIP and Global Carbon Project (GCP) inversion ensembles (Byrne et al., 2023; Crowell et al., 2019; Friedlingstein et al., 2025; Peiro et al., 2022). We identify several priority areas for understanding biases in ATMs, particularly the inconsistency between diabatic mixing rates diagnosed from the MSE budgets of parent reanalysis and the tracer fields of coarser resolution ATMs identified here. These inconsistencies likely stem from several potential sources: (1) regridding of original reanalyses to the coarser resolution of the ATM grid, (2) for online GCMs using nudging, incomplete matching of the input meteorology, and (3) for offline models, recalculation or parameterization of convective mass fluxes in the coarser ATM. The first potential source of error from regridding could be evaluated by comparing MSE-based diabatic mixing rates from the parent and regridded fields as long as all components of MSE were included in the regridding. The second potential source of error from nudging could be evaluated by comparing MSE-based diabatic mixing rates from the regridded parent model and the nudged online simulation. Finally, the third potential source of error from recalculating or parameterizing vertical mass fluxes could be evaluated by comparing the MSE-based diabatic mixing rates from the regridded parent model and the tracer-based mixing rates from the ATM. It is notable that diabatic mixing rates diagnosed from two online models, MIROC4-ACTM and CAM-SD, which do not require regridding, are generally consistent with observations, with MIROC4-ACTM showing the best performance among all models (Figs. 7-10).

An important consideration is that the real atmosphere mixes MSE and tracers at different spatial and temporal scales. In the Northern Hemisphere, APO fluxes initially mix vertically over oceans, while strong CO_2 fluxes initially mix vertically over land. In contrast, MSE fluxes mix initially over both land and ocean. Due to the large land area in the Northern Hemisphere, the zonal mixing time scale is much longer (~ 2-4 weeks) so that diabatic mixing rates diagnosed from APO or CO_2 tracers could differ from each other and from those diagnosed from MSE tracers. In the Southern Hemisphere mid-latitudes, these potential differences are much smaller due to the predominance of ocean and rapid zonal mixing (~ 1-2 weeks). In general, the timescales for diabatic mixing are longer than the timescales of zonal mixing, which support our approach of using tracer fluxes over both ocean and land to evaluate zonal-mean diabatic mixing. Future work should also develop metrics for quantifying along-isentrope (adiabatic) transport to complement our understanding of tracer mixing across isentropes. The timescales of adiabatic mixing influences tracer gradients along isentropic surfaces, which in turn affects diabatic mixing differently in the upper versus lower troposphere. It is also necessary to examine the sensitivity of mixing rates to model resolution, particularly vertical levels at the interface between the boundary layer and free troposphere, and boundary layer schemes. These ATM improvements are essential for enhancing both forward simulations and inverse estimates of surface fluxes.

4. Summary and Outlook

We conducted the Atmospheric Potential Oxygen forward Model Intercomparison Project (APO-MIP1) to generate forward simulations of APO and its components using different flux products and eight ATMs. This effort provides model APO simulations at surface stations, along aircraft flight paths, and on ships that can be directly compared with observations. Additionally, we provide 3-D APO fields from six of the eight ATMs. We use simulations from APO-MIP1 to evaluate eight ATMs and three flux products by comparing simulations against observations from surface stations, aircraft, and ships.

We find that model simulations of APO seasonal cycles using a given flux product show considerable summer-time spread at northern surface stations, particularly at two eastern Pacific stations, LJO and CBA (Fig. 3). The bias stems from challenges in accurately representing complex atmospheric vertical transport processes, marine boundary layer mixing, and coastal horizontal mixing in these regions. These findings highlight the limitations of current APO inversions that rely on a single ATM (i.e., TM3 used in Jena APO inversion) and sparse surface observations. However, model simulations of column-average APO resolved from sampling aircraft tracks are consistent across different ATMs, emphasizing the importance of airborne measurements for constraining large-scale flux features.

Using airborne observations and a moist-isentropic coordinate framework, we demonstrate that most ATMs overestimate diabatic mixing rates in the mid-latitudes of both hemispheres when compared to mixing rates derived from energy budgets of reanalyses. Among all ATMs used here, Jena_TM3 and CTE_TM5 show the largest biases. These constraints also enable us to separate flux biases from transport-related biases, allowing independent evaluation of flux models, which show that the CESM flux product is the best among the three flux products used in this study. This prognostic model outperforms two observation based products because of sparse atmospheric and surface observations, limitations in ATM used in atmospheric inversion, and because seasonal APO fluxes are driven by physical and biological processes that CESM represents well.

We encourage the broader community to develop new APO inversions, which could provide independent constraints on ocean biogeochemical processes and improve our understanding of the ocean carbon sink. Model simulations from APO-MIP1 can be used in other applications, including the calibration of methods for estimating seasonal air-sea APO fluxes from global atmospheric observations (e.g., Jin et al., 2023), constraining the representation of regional to global marine production in Earth system models (e.g., Nevison et al., 2012, 2015, 2018), and for understanding ESM biases in seasonal air-sea CO_2 exchange related to both thermal and non-thermal forcings. The transport simulations can also support the evaluation of long-term trends in O_2:CO_2 ratios over the Southern Ocean based on surface station gradients, useful for assessing biogeochemical responses to climate change.

We expect APO-MIP1 to continue evolving as an active collaboration examining atmospheric tracer transport and air-sea O_2 flux estimates. The current implementation excluded the air-sea CO_2 component and long-term flux trends from the Jena flux product, and does not include interannual and long-term flux trends in the DISS flux product, making these simulations unsuitable for interpreting interannual to long-term air-sea O_2 fluxes features. Thus, we only analyze APO seasonal cycles and meridional gradients here. The next phase of APO-MIP1 will address these limitations by incorporating updated inversion flux fields based on a larger set of atmospheric APO observations and including interannual variability. We will expand the scope by including terrestrial O_2 flux fields for O_2-specific analyses and seasonal-only component fluxes to investigate rectifier effects. The seasonal rectifier effect refers to the creation of non-zero annual mean atmospheric concentration gradients at surface stations even with balanced seasonal O_2 fluxes. This occurs when fluxes correlate with seasonal variations in atmospheric mixing. For example, strong summer O_2 outgassing combined with shallow PBL heights concentrates APO near the surface, while higher winter PBL dilutes the O_2 uptake signal, resulting in observed annual mean APO gradients even when the annual mean flux is zero. Additionally, we plan to update air-sea O_2 fluxes derived from surface ocean dissolved oxygen measurements by replacing Garcia and Keeling (2001) with fluxes calculated from recent machine learning interpolation of dissolved oxygen products (Gouretski et al., 2024; Ito et al., 2024; Sharp et al., 2023). We encourage broader participation from diverse modeling groups in the next phase of APO-MIP1.

Appendix A: Surface station, airborne, and shipboard APO measurements.

The surface station APO observations from the Scripps O_2 program have been described in (Keeling et al., 1998). Briefly, flask triplicates have been collected at biweekly to monthly frequency during clean background air conditions at a network of sites for over three decades, and returned to Scripps for analysis using interferometric and mass-spectrometric techniques. Here we use monthly data that was averaged from roughly bi-weekly data. The flask measurements are first adjusted to the middle of each month, parallel to the mean seasonal cycle for that station, before averaging. The APO-MIP1 output for these stations was reported matching the ObsPack CO_2 files from the Scripps O_2 Program, to take advantage of the established ObsPack format. These CO_2 measurements correspond to the same flask air on which O_2 is measured. The model output is treated in the same way as the observations to generate monthly means.

Airborne APO measurements from HIPPO, ORCAS, and ATom campaigns were made in situ with the NSF NCAR Airborne Oxygen Instrument (AO2), using a vacuum-ultraviolet absorption technique to measure O_2 and a single-cell infrared gas analyzer to measure CO_2 (Stephens et al., 2021). AO2 produces measurements every 2.5 s, which are averaged to 10 sec frequency for merging with other aircraft data. To correct for flight-specific sampling offsets, the in situ AO2 data were adjusted to agree with flask measurements collected during each flight using the NSF

NCAR / Scripps Medusa flask sampler on a flight-by-flight average basis (Jin et al., 2023; Stephens et al., 2021).

HIPPO and ATom had nearly pole-to-pole coverage, and from near surface (150 - 300 m) to above the tropopause. HIPPO consisted of five campaigns between 2009 and 2011, and most data were collected above the Pacific. ATom consisted of four campaigns between 2016 and 2018, and each campaign had a Pacific transect and an Atlantic transect. ORCAS was a 6-week campaign with dense temporal sampling over the Drake Passage and ocean areas adjacent to the tip of South America and the Antarctic Peninsula. The APO-MIP1 output for these aircraft measurements was reported matching the ObsPack CO_2 files for each campaign. These data are also at 10 sec frequency but correspond to different instruments with different calibration intervals. To match the observed and model time series, we mask observations when model output is not available, and vice versa. We also exclude any stratospheric data, with the stratosphere defined as water vapor concentrations below 50 ppm and either ozone concentrations exceeding 150 ppb, or detrended N_2O levels (normalized to 2009) below 319 ppb (Jin et al., 2021). Water vapor and ozone were measured by the NOAA UAS Chromatograph for Atmospheric Trace Species instrument (Hintsa et al., 2021). N_2O was measured by the Harvard Quantum Cascade Laser System instrument (Santoni et al., 2014). We filter the airborne data to exclude continental or urban boundary-layer air sampled while landing, taking off, or conducting missed approaches at airports (Jin et al., 2021).

Shipboard APO measurements from the ARSV *L. M. Gould* were made in situ during over 90 transects of Drake Passage on 50 cruises between 2012 and 2017 using a fuel-cell method for O_2 and a two-cell non-dispersive infrared gas analyzer for CO_2. The instrumentation was similar to a previously developed tower system (Stephens et al., 2003), but adapted and optimized for shipboard use. The instrument produces measurements at 1 min frequency. The cruises occurred in all months of the year but are more sparse during austral winter. The Gould operated almost exclusively between Punta Arenas, Chile and Palmer Station, Antarctica, in support of resupplying and transferring personnel to Palmer Station. The cruises span from 53° to 65°S in all months, and extend as far as 70°S during summer months. The APO-MIP1 output for the Gould was reported matching the ObsPack CO_2 file from the NOAA underway pCO_2 system. This system measures atmospheric CO_2 for 15 min every two hours. To match the observed and model time series, we first calculate hourly means for each and then mask observations when model output is not available, and vice versa.

The resolved APO annual mean and seasonal cycles have negligible measurement uncertainty compared to model spread because we average data over long time series for stations and over large spatial domains for aircraft and ships, effectively reducing the already small short-term instrument imprecision.

Appendix B: APO flux products

B.1. Air-sea APO flux products

The first air-sea APO flux product (Jena) is air-sea APO flux from the Jena CarboScope APO Inversion (version ID: apo99X_v2021), which is available directly as F_{APO}^{ocn} (update of Rödenbeck et al., 2008). In this inversion, the posterior fluxes (variable name: apoflux_ocean) were optimized to best match observed APO at 9 stations in the Scripps O_2 Program surface network (Manning and Keeling, 2006) and at 2 stations from the National Institute for Environmental Studies (Tohjima et al., 2012). The prior air-sea CO_2 flux was not included in the forward simulations here. We note that the exclusion of prior air-sea CO_2 flux has only minimal impact on the simulated APO seasonal cycle and north-to-south annual gradient but reduces the tropical “bulge” of annual mean by approximately 1 per meg and results in close to zero long-term APO trend. The Jena product is available from 1999 to 2020 originally with spatial resolution of 2° latitude × 2.5° longitude at daily intervals, converted to 1° × 1°. The Jena inversion used the TM3 transport model, which is also one of the models participating in APO-MIP1. In the case of TM3 forward transport simulation, the Jena inversion posterior fluxes have been re-run forward through the ATM, and thus this combination of fluxes and transport should agree well at the surface stations used for inversion optimization.

The second air-sea APO flux product (CESM) uses air-sea O_2, CO_2, and N_2 flux components from the Community Earth System Model (CESM2) Forced Ocean-Sea-Ice (FOSI) simulation (Yeager et al., 2022), which is forced by atmospheric fields from JRA55-do reanalysis (Tsujino et al., 2018) and prognostic ocean biogeochemistry using the Marine Biogeochemistry Library

(MARBL, Long et al., 2021a). The model directly produces $F_{O_2}^{ocn}$ and $F_{CO_2}^{ocn}$, while $F_{N_2}^{ocn}$ is calculated by scaling the ocean heat flux (Q, W m^{-2}) output using the relationship from Keeling and Shertz (1992) following

$$F_{N_2}^{ocn} = -\frac{1}{1.3} \cdot \frac{dS}{dT} \cdot \frac{Q}{C_p}, \qquad \text{(B1)}$$

where dS/dT (mol kg^{-1} C^{-1}) is the temperature derivative of solubility using solubility coefficients from Hamme & Emerson (2004). C_p represents the specific heat capacity of seawater, which is assumed to be 3993 J kg^{-1} C^{-1}. The factor of 1/1.3 is to adjust the seasonal amplitude due to the temporal lag between tracer flux and heat flux, as proposed by Jin et al. (2007).

These three CESM flux components have a resolution of 1° latitude × 1° longitude grid with the North Pole displaced to Greenland. All fields are available from 1958 to 2020, but we only use fluxes from 1986 to 2020. $F_{O_2}^{ocn}$ and $F_{CO_2}^{ocn}$ are output from the model at daily resolution, whereas $F_{N_2}^{ocn}$ is calculated from monthly model heat fluxes then interpolated to daily resolution. This version of CESM was designed to initialize a seasonal-to-multiyear large ensemble (SMYLE) of coupled simulations for evaluating predictability. It is forced by observed meteorology starting in 1958, at which point it branches off of a FOSI configuration using JRA55-do atmospheric fields as surface boundary conditions (Yeager et al., 2022). The FOSI simulation consists of six consecutive cycles of 1958-2018 forcing, with the sixth cycle (used for SMYLE) extended through 2020. Annual mean heat fluxes from this configuration show a small cooling drift over the historical period, and thus the inferred annual mean and long-term trend of O_2 and N_2 flux should not be interpreted as realistic.

The third air-sea APO flux product (DISS) uses bottom-up air-sea O_2 and CO_2 flux estimates derived primarily from dissolved gas measurements. $F_{O_2}^{ocn}$ consists of a seasonal component calculated from the dissolved O_2 measurement based climatology of Garcia & Keeling (2001), with seasonal amplitude scaled by 0.82 according to Naegler et al. (2006), and an annual mean component from the ocean inversion of Resplandy et al. (2016) for 21 regions using transport from MITgcm-ECCO. Bent (2014) reported that the 0.82 scaling factor significantly improved agreement between GK flux and HIPPO observations, based on simulations using one ATM (a different MIROC4-ACTM configuration). However, our results show that applying this 0.82 scaling factor actually leads to an underestimation of modelled column-mean APO SCA when comparing with the combined HIPPO, ORCAS, and ATom observations at high latitudes in both hemispheres. The seasonal component (1.125° × 1.125° × monthly) was linearly regridded to 1° × 1° × daily resolution. For the annual mean component, the original regional values (21 regions) were spatially interpolated to 1° × 1° resolution while conserving the total sum within each region, then temporally interpolated to daily values. We use $F^{ocn}_{CO_2}$ from the machine learning interpolation of pCO$_2$ based air-sea CO$_2$ fluxes (Jersild et al., 2017; Landschützer et al., 2016). The version of this product that we used provides fluxes from 1982 to 2020, with resolution of 1° latitude × 1° longitude × monthly, which we interpolated to daily. We use Eq. B1 to calculate $F^{ocn}_{N_2}$ with heat fluxes from ERA5 reanalyses (Hersbach et al., 2020), which is available from 1979 onwards, with resolution of 0.25° latitude × 0.25° longitude × monthly. Sea-surface temperature (SST) estimates required to calculate dS/dT (Eq. B1) are from World Ocean Atlas (WOA) v2018 with resolution of 1° latitude × 1° longitude × monthly. SST is available as a 1981 to 2010 climatology but we use it repeatedly for 1986 to 2020.

B.2. Fossil fuel APO uptake products

We used two products for F^{ff}_{APO}. The first product (GridFED) uses fossil CO$_2$ emission and O$_2$ uptake fluxes from Jones et al. (2021), downloaded from Jones et al. (2022). This product is available from 1959 to 2020, with resolution of 0.1° latitude × 0.1° longitude × monthly, which we interpolate to daily.

The second product (OCO2MIP) use $F^{ff}_{CO_2}$ as prepared for the OCO-2 Model Intercomparison Project (MIP) version 10, downloaded from Basu & Nassar (2021), with resolution of 1° latitude x 1° longitude x hourly. This $F^{ff}_{CO_2}$ product uses fossil fuel CO$_2$ emission from ODIAC (Oda et al., 2018) for 2000 to 2019. For 2020, the flux was scaled from 2019 using the ratio of 2020 to

2019 global emissions reported by Liu et al. (2020). $F^{ff}_{O_2}$ is not available from this product, but we scale the atmospheric field of ΔCO^{ff}_2 by a factor of -1.4 to estimate ΔO^{ff}_2 (Keeling, 1988; Steinbach et al., 2011). We primarily use GridFED, except for CAMS_LMDZ where we use OCO2MIP instead, because $F^{ff}_{O_2}$ from GridFED is missing for years after 2015. The differences between these two products are negligible compared to the magnitude of ocean-driven APO variations, for the seasonal metrics considered here.

Appendix C: Calculation of $M_{\theta e}$, cross-$M_{\theta e}$ diabatic mixing rates and APO gradients

The mass-indexed moist isentropic coordinate $M_{\theta e}$ is defined as the total dry air mass under a specific moist isentropic surface (θ_e) in the troposphere of a given hemisphere. Surfaces of constant $M_{\theta e}$ are parallel to surfaces of constant θ_e but the relationship changes with season, as the atmosphere warms and cools. $M_{\theta e}$ surfaces have air mass (10^{16} kg) as the unit, and are adjusted to conserve dry air mass below the surface at any instant in time. $M_{\theta e}$ is calculated as a function of θ_e and time following

$$M_{\theta e}(x, t) \ = \ \sum M_x(t) | \theta_{e_x} < \theta_e, \qquad \text{(C1)}$$

where x indicates an individual grid cell of the atmospheric field, $M_x(t)$ is the dry air mass of each grid cell x at time t, and θ_{e_x} is the equivalent potential temperature of the grid cell. For a given θ_e threshold, the corresponding $M_{\theta e}$ value is calculated by integrating the air mass of all grid cells with θ_e value smaller than the threshold. We only integrate air mass in the troposphere, which is defined here as potential vorticity unit (PVU) smaller than 2. At each time step, this calculation yields a unique value of $M_{\theta e}$ for each value of θ_e as well as a 3-D field of atmospheric $M_{\theta e}$. Following the spatial pattern of θ_e, $M_{\theta e}$ values generally increase from low to high altitudes and from poles to equator. We generate daily $M_{\theta e}$ fields using four different reanalysis products (MERRA-2, JRA-55, JRA-3Q, and ERA5) at their native resolution, avoiding potential information loss from grid interpolation (Gelaro et al., 2017; Hersbach et al., 2020; Kobayashi et al., 2015; Kosaka et al., 2024).

The calculation of diabatic mixing rates in ATMs is based on a box model approach, which uses $M_{\theta e}$ as boundaries. A schematic of the box model is available as Fig. 1 of Jin et al. (2024). The box model invokes tracer air mass balance, which recognizes tracer inventory change (M_i, Tmol) of each $M_{\theta e}$ box equal to the sum of surface fluxes (F_i, Tmol day^{-1}) and the diabatic transport between boxes ($T_{i,i+1}$, Tmol day^{-1}, positive poleward). The transport term is considered as a diffusive system, which is parameterized as the product of diabatic mixing rate across the $M_{\theta e}$ boundary ($D_{i,i+1}$, $(10^{16}$ kg$)^2$ day^{-1}) and the tracer concentration (χ_{i+1}, Tmol tracer per kg air mass) gradient between two boxes. The full mass balance follows

$$\frac{\partial M_i}{\partial t} = \begin{cases} F_i + T_{i,i+1} & \text{if } i = 1 \\ F_i + T_{i,i+1} - T_{i-1,i} & \text{if } i > 1 \end{cases}, \tag{C2}$$

with

$$T_{i,i+1} = D_{i,i+1} \cdot \frac{\chi_{i+1} - \chi_i}{\Delta M_{\theta e}}. \tag{C3}$$

In these equations, i is the number label of the box and is set to be 1 at the highest latitude, $\Delta M_{\theta e}$ is the distance in $M_{\theta e}$ coordinates between box centers, which for evenly spaced boxes as used here, is the same as the total air mass of each box. In this study, we set the range of each $M_{\theta e}$ box to be 15 × 10^{16} kg air mass, and therefore $\Delta M_{\theta e}$ equals the same value. The diabatic mixing rate (D) can be expressed as

$$D_{i,i+1}(t) = \frac{\left[\sum_{i'=1}^{i'=i} \left(\frac{dM_{i'}(t)}{dt} - F_{i'}(t)\right)\right]}{[\chi_{i+1}(t) - \chi_i(t)]} \cdot \Delta M_{\theta_e}. \tag{C4}$$

This method effectively reconstructs large-scale tracer transport features (T) in ATMs, as demonstrated in Jin et al. (2024). We note that the diabatic mixing rate is a property of the corresponding $M_{\theta e}$ and is theoretically insensitive to the choice of box sizes. We calculate climatological monthly average (2009 to 2018) diabatic mixing rates for each of the six transport models using the 3-D APO fields from transporting each of the three flux products (Figs. 7 and 9). To assign $M_{\theta e}$ at the model grid locations and times for each ATM, we always use $M_{\theta e}$ from

MERRA-2 interpolated to the ATM grid, to ensure spatial consistency. Using other reanalyses only leads to small (< 5%) differences in ATM-diagnosed diabatic mixing rates (Jin et al., 2024).

Independent observational constraints on ATM-diagnosed mixing rates are calculated from moist static energy (MSE) budgets of four meteorological reanalyses (Figs. 7 and 9). MSE is a measure of static energy that is conserved in adiabatic ascent/descent and during latent heat release due to condensation, and naturally aligns with surfaces of θ_e or $M_{\theta e}$. This diagnostic approach offers more robust mixing rate estimates than tracer-based methods in part because MSE maintains consistent, non-zero gradients at each reanalysis time step, unlike chemical tracers. Additionally, MSE-based mixing rates are directly diagnosed from reanalysis on the original grid, avoiding potential artifacts introduced when these fields are interpolated to coarser transport model grids, and any recalculation of vertical mass fluxes and subgrid-scale mixing parameterizations in ATMs.

The MSE-diagnosed mixing rate calculation adapts our tracer box model framework. In this adaptation, we replace tracer inventory (M_i, Tmol) by MSE (S_i, J), replace surface tracer flux (F_i, Tmol day^{-1}) by surface heat flux (Q_i, J day^{-1}), and add an additional term to account for atmospheric radiative energy balance (R_i, J day^{-1}), following

$$D_{i,i+1}(t) = \frac{\left[\sum_{i'=1}^{i'=i}\left(\frac{dS_{i'}(t)}{dt} - Q_{i'}(t) - R_{i'}(t)\right)\right]}{[\chi_{i+1}(t) - \chi_i(t)]} \cdot \Delta M_{\theta_e} \quad \text{(C5)}$$
.

We note that the gradient on the denominator in Eq. C5 represents the MSE density gradient (J per kg air mass) across the $M_{\theta e}$ surface. The calculation of these terms requires air temperature, specific humidity, surface heat flux, including surface sensible and latent heat flux, and radiative imbalance from reanalysis. Further details on the process to diagnose mixing rate from both ATMs and reanalyses can be found in Jin et al. (2024).

The cross-$M_{\theta e}$ APO gradient was calculated using data grouped into two adjacent boxes in the $M_{\theta e}$ space, with box centers spanning 15 × 10^{16} kg air mass across the target surface boundary. For each box, we calculate the average APO concentration by trapezoidal integration of detrended APO as a function of $M_{\theta e}$ and dividing by the $M_{\theta e}$ range (Jin et al., 2021). We carry out the calculation for each airborne campaign, using the observations, model flight track output, and

3-D model fields. Flight-track estimated cross-$M_{\theta e}$ APO gradients are not directly comparable to simulated gradients from full 3-D fields, due to spatial and temporal coverage biases in airborne observations. We correct for both biases in the APO airborne observations and model flight track output (detailed in Supplement Text S1).

Code and Data Availability

The 10 components of air-sea APO flux and fossil fuel APO uptake products, and the output of ATM forward transport simulations of these 10 components, including ATM samples at surface stations, ship transects, aircraft measurements, and 3-D atmospheric fields, are available at https://doi.org/10.5065/f3pw-a676 (Stephens et al., 2025). APO observations at surface stations from the Scripps O_2 network are available at https://doi.org/10.6075/J0WS8RJR (Keeling, 2019). All HIPPO 10-s merge data are available from Wofsy, 2017. Here we use updated HIPPO AO2 data from (Stephens et al., 2021a, 2021b, 2021c, 2021d, 2021e). All ORCAS 10-s merge data are available at Stephens (2017). Here we use updated ORCAS AO2 data from Stephens et al. (2021f). All ATom 10-s merge data are available at https://doi.org/10.3334/ ORNLDAAC/1925 (Wofsy, 2021), including the version of AO2 data used here. O_2 and CO_2 measurements from ARSV Gould are available at https://doi.org/10.26023/FDDD-PC3X-4M0X (Stephens, 2025). Note that airborne O_2/N_2 data are all on the Scripps O_2 Program SIO2017 O_2/N_2 scale defined on March 16, 2020, surface station data are on the SIO2023 O_2/N_2 scale defined on August 30, 2024, and shipboard data are on the SIO2023 O_2/N_2 scale defined on August 30, 2024. Airborne CO_2 measurements are on the WMO X2007 CO_2 scale, while station and shipboard CO_2 data are on the WMO X2019 CO_2 scale. The use of different scales has only minor impacts on interpreting APO seasonal cycles and latitudinal gradients. Code used to produce input flux files and to post-process submitted ObsPack files is available at https://doi.org/10.5065/f3pw-a676 (Stephens et al., 2025).

Acknowledgments

We would like to acknowledge the efforts of the full HIPPO, ORCAS, and ATom science teams and the pilots and crew of the NSF NCAR GV and NASA DC-8, as well as the NSF NCAR and NASA project managers, field support staff, and logistics experts. Atmospheric O_2 measurements on HIPPO were supported by NSF grants ATM-0628519 and ATM-0628388. ORCAS was supported by NSF grants PLR-1501993, PLR-1502301, PLR-1501997, and PLR-1501292. Atmospheric O_2 measurements on ATom 1 were supported by NSF grants AGS-1547626 and AGS-1547797. Atmospheric O_2 measurements on ATom 2-4 were supported by NSF AGS-1623745 and AGS-1623748. The recent atmospheric measurements of the Scripps O_2 program have been supported via funding from the NSF and the National Oceanographic and Atmospheric Administration (NOAA) under Grants OPP-1922922, OPP-2329254, and NA20OAR4320278, respectively. The atmospheric O_2 measurements from ARSV Laurence M. Gould were supported by NSF grants ANT-0944761, PLR-1341425, and PLR-1543511. For sharing O_3, N_2O, and H_2O measurements, we thank Jim Elkins, Eric Hintsa, and Fred Moore for ATom-1 N_2O data; Ru-Shan Gao and Ryan Spackman for HIPPO O_3 data; Ilann Bourgeois, Jeff Peischl, Tom Ryerson, and Chelsea Thompson for ATom O_3 data; Stuart Beaton, Minghui Diao, and Mark Zondlo for HIPPO and ORCAS H_2O data; and Glenn Diskin and Joshua DiGangi for ATom H_2O data. YJ would like to acknowledge the Advanced Study Program Postdoctoral Fellowship in the NSF National Center for Atmospheric Research. This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977. The work of FC was granted access to the HPC resources of CCRT under the allocation CEA/DRF, and of TGCC under the allocation A0130102201 made by GENCI. NC and PKP are supported by the Environment Research and Technology Development Fund (grant no. JPMEERF24S12205) and Arctic Challenge for Sustainability II (ArCS-II) project (grant no. JPMXD1420318865). YN is supported by JSPS KAKENHI (grant no. JP22H05006, JP80282151) and the Environment Research and Technology Development Fund (grant no. JPMEERF24S12210). IL and JH were supported by the Netherlands Organisation for Scientific Research (grant no. VI.Vidi.213.143 and NWO-2023.003).

Author Contributions

YJ and BS carried out the research and wrote the paper with input from all co-authors. YJ, BS, and MC designed the research. MC prepared input fluxes for the transport models. BS provided airborne and shipboard observation data. EM provided surface station and airborne observation data. YJ, FC, NC, JH, IL, SM, YN, PP, CR, and JV provided forward transport model simulations. All authors contributed to reviewing and editing the text.

Competing Interests

The contact author has declared that none of the authors has any competing interests.

References

Adcock, K. E., Pickers, P. A., Manning, A. C., Forster, G. L., Fleming, L. S., Barningham, T., Wilson, P. A., Kozlova, E. A., Hewitt, M., Etchells, A. J., and Macdonald, A. J.: 12 years of continuous atmospheric O_2, CO_2 and APO data from Weybourne Atmospheric Observatory in the United Kingdom, Earth Syst. Sci. Data, 15, 5183–5206, https://doi.org/10.5194/essd-15-5183-2023, 2023.

Bailey, A., Singh, H. K. A., and Nusbaumer, J.: Evaluating a Moist Isentropic Framework for Poleward Moisture Transport: Implications for Water Isotopes Over Antarctica, Geophys. Res. Lett., 46, 7819–7827, https://doi.org/10.1029/2019GL082965, 2019.

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003, Glob. Biogeochem. Cycles, 20, https://doi.org/10.1029/2004GB002439, 2006.

Basu, S. and Nassar, R.: Fossil Fuel CO_2 Emissions for the OCO2 Model Intercomparison Project (MIP) (2020.1), https://doi.org/10.5281/zenodo.4776925, 2021.

Battle, M., Fletcher, S. M., Bender, M. L., Keeling, R. F., Manning, A. C., Gruber, N., Tans, P. P., Hendricks, M. B., Ho, D. T., Simonds, C., Mika, R., and Paplawsky, B.: Atmospheric potential oxygen: New observations and their implications for some atmospheric and oceanic models, Glob. Biogeochem. Cycles, 20, 2005GB002534, https://doi.org/10.1029/2005GB002534, 2006.

Belikov, D. A., Maksyutov, S., Krol, M., Fraser, A., Rigby, M., Bian, H., Agusti-Panareda, A., Bergmann, D., Bousquet, P., Cameron-Smith, P., Chipperfield, M. P., Fortems-Cheiney, A., Gloor, E., Haynes, K., Hess, P., Houweling, S., Kawa, S. R., Law, R. M., Loh, Z., Meng, L., Palmer, P. I., Patra, P. K., Prinn, R. G., Saito, R., and Wilson, C.: Off-line algorithm for calculation of vertical tracer transport in the troposphere due to deep convection, Atmospheric Chem. Phys., 13, 1093–1114, https://doi.org/10.5194/acp-13-1093-2013, 2013.

Belikov, D. A., Maksyutov, S., Yaremchuk, A., Ganshin, A., Kaminski, T., Blessing, S., Sasakawa, M., Gomez-Pelaez, A. J., and Starchenko, A.: Adjoint of the global Eulerian–Lagrangian coupled atmospheric transport model (A-GELCA v1.0): development and validation, Geosci. Model Dev., 9, 749–764, https://doi.org/10.5194/gmd-9-749-2016, 2016.

Bent, J.: Airborne Oxygen Measurements over the Southern Ocean as an Integrated Constraint of Seasonal Biogeochemical Processes, University of California, San Diego, 2014.

Blaine, T.: Continuous Measurements of Atmospheric Ar/N_2 as a Tracer of Air-Sea Heat Flux: Models, Methods, and Data, University of California, San Diego, 2005.

Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., Crisp, D., Crowell, S., Deng, F., Deng, Z., Deutscher, N. M., Dubey, M. K., Feng, S., García, O. E., Griffith, D. W. T., Herkommer, B., Hu, L., Jacobson, A. R., Janardanan, R., Jeong, S., Johnson, M. S., Jones, D. B. A., Kivi, R., Liu, J., Liu, Z., Maksyutov, S., Miller, J. B., Miller, S. M., Morino, I., Notholt, J., Oda, T., O'Dell, C. W., Oh, Y.-S., Ohyama, H., Patra, P. K., Peiro, H., Petri, C., Philip, S., Pollard, D. F., Poulter, B., Remaud, M., Schuh, A., Sha, M. K., Shiomi, K., Strong, K., Sweeney, C., Té, Y., Tian, H., Velazco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake, Earth Syst. Sci. Data, 15, 963–1004, https://doi.org/10.5194/essd-15-963-2023, 2023.

Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Landschützer, P., Lauderdale, J. M., Liu, J., Manizza, M., Naviaux, J. D., Rödenbeck, C., Schimel, D. S., Van Der Stocken, T., and Zhang, H.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean pCO_2 and Air-Sea CO_2 Flux, J. Adv. Model. Earth Syst., 12, e2019MS001888, https://doi.org/10.1029/2019MS001888, 2020.

Chandra, N., Patra, P. K., Niwa, Y., Ito, A., Iida, Y., Goto, D., Morimoto, S., Kondo, M., Takigawa, M., Hajima, T., and Watanabe, M.: Estimated regional CO_2 flux and uncertainty based on an ensemble of atmospheric CO_2 inversions, Atmospheric Chem. Phys., 22, 9215–9243, https://doi.org/10.5194/acp-22-9215-2022, 2022.

Chevallier, F.: On the parallelization of atmospheric inversions of CO_2 surface fluxes within a variational framework, Geosci. Model Dev., 6, 783–790, https://doi.org/10.5194/gmd-6-783-2013, 2013.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. -M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res. Atmospheres, 110, 2005JD006390, https://doi.org/10.1029/2005JD006390, 2005.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, D21307, https://doi.org/10.1029/2010JD013887, 2010.

Chikira, M. and Sugiyama, M.: A Cumulus Parameterization with State-Dependent Entrainment Rate. Part I: Description and Sensitivity to Temperature and Humidity Profiles, J. Atmospheric Sci., https://doi.org/10.1175/2010JAS3316.1, 2010.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmospheric Chem. Phys., 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Syst., 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.

Denning, A. S., Holzer, M., Gurney, K. R., Heimann, M., Law, R. M., Rayner, P. J., Fung, I. Y., Fan, S.-M., Taguchi, S., Friedlingstein, P., Balkanski, Y., Taylor, J., Maiss, M., and Levin, I.: Three-dimensional transport and concentration of SF_6 A model intercomparison study (TransCom 2), Tellus B Chem. Phys. Meteorol., 51, 266–297, https://doi.org/10.3402/tellusb.v51i2.16286, 1999.

Emanuel, K. A.: A Scheme for Representing Cumulus Convection in Large-Scale Models, J. Atmospheric Sci., 48, 2313–2329, https://doi.org/10.1175/1520-0469(1991)048<2313:ASFRCC>2.0.CO;2, 1991.

Faassen, K. A. P., Nguyen, L. N. T., Broekema, E. R., Kers, B. A. M., Mammarella, I., Vesala, T., Pickers, P. A., Manning, A. C., Vilà-Guerau de Arellano, J., Meijer, H. A. J., Peters, W., and Luijkx, I. T.: Diurnal variability of atmospheric O_2, CO_2, and their exchange ratio above a boreal forest in southern Finland, Atmospheric Chem. Phys., 23, 851–876, https://doi.org/10.5194/acp-23-851-2023, 2023.

Faassen, K. A. P., Vilà-Guerau de Arellano, J., González-Armas, R., Heusinkveld, B. G., Mammarella, I., Peters, W., and Luijkx, I. T.: Separating above-canopy CO_2 and O_2 measurements into their atmospheric and biospheric signatures, Biogeosciences, 21, 3015–3039, https://doi.org/10.5194/bg-21-3015-2024, 2024.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Arneth, A., Arora, V., Bates, N. R., Becker, M., Bellouin, N., Berghoff, C. F., Bittig, H. C., Bopp, L., Cadule, P., Campbell, K., Chamberlain, M. A., Chandra, N., Chevallier, F., Chini, L. P., Colligan, T., Decayeux, J., Djeutchouang, L. M., Dou, X., Duran Rojas, C., Enyo, K., Evans, W., Fay, A. R., Feely, R. A., Ford, D. J., Foster, A., Gasser, T., Gehlen, M., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A. K., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Kato, E., Keeling, R. F., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Lan, X., Lauvset, S. K., Lefèvre, N., Liu, Z., Liu, J., Ma, L., Maksyutov, S., Marland, G., Mayot, N., McGuire, P. C., Metzl, N., Monacci, N. M., Morgan, E. J., Nakaoka, S.-I., Neill, C., Niwa, Y., Nützel, T., Olivier, L., Ono, T., Palmer, P. I., Pierrot, D., Qin, Z., Resplandy, L., Roobaert, A., Rosan, T. M., Rödenbeck, C., Schwinger, J., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Steinhoff, T., et al.: Global Carbon Budget 2024, Earth Syst. Sci. Data, 17, 965–1039, https://doi.org/10.5194/essd-17-965-2025, 2025.

Gallagher, M. E., Liljestrand, F. L., Hockaday, W. C., and Masiello, C. A.: Plant species, not climate, controls aboveground biomass O_2:CO_2 exchange ratios in deciduous and coniferous ecosystems, J. Geophys. Res. Biogeosciences, 122, 2314–2324, https://doi.org/10.1002/2017JG003847, 2017.

Garcia, H. E. and Keeling, R. F.: On the global oxygen anomaly and air-sea flux, J. Geophys. Res. Oceans, 106, 31155–31166, https://doi.org/10.1029/1999JC000200, 2001.

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., Patra, P. K., Peters, W., Rödenbeck, C., Saeki, T., Schimel, D., Van Der Laan-Luijkx, I., Wofsy, S., and Yin, Y.: Global atmospheric CO_2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate, Biogeosciences, 16, 117–134, https://doi.org/10.5194/bg-16-117-2019, 2019.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Clim., 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-Based Model for Boundary Layer Clouds. Part I: Method and Model Description, J. Atmospheric Sci., 59, 3540–3551, 2002.

Goto, D., Morimoto, S., Aoki, S., Patra, P. K., and Nakazawa, T.: Seasonal and short-term variations in atmospheric potential oxygen at Ny-Ålesund, Svalbard, Tellus B Chem. Phys. Meteorol., 69, 1311767, https://doi.org/10.1080/16000889.2017.1311767, 2017.

Gouretski, V., Cheng, L., Du, J., Xing, X., Chai, F., and Tan, Z.: A consistent ocean oxygen profile dataset with new quality control and bias assessment, Earth Syst. Sci. Data, 16, 5503–5530, https://doi.org/10.5194/essd-16-5503-2024, 2024.

Gruber, N., Gloor, M., Fan, S., and Sarmiento, J. L.: Air-sea flux of oxygen estimated from bulk data: Implications For the marine and atmospheric oxygen cycles, Glob. Biogeochem. Cycles, 15, 783–803, https://doi.org/10.1029/2000GB001302, 2001.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, Tellus B Chem. Phys. Meteorol., 55, 555–579, https://doi.org/10.3402/tellusb.v55i2.16728, 2003.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, Glob. Biogeochem. Cycles, 18, https://doi.org/10.1029/2003GB002111, 2004.

Hamme, R. C. and Emerson, S. R.: The solubility of neon, nitrogen and argon in distilled water and seawater, Deep Sea Res. Part Oceanogr. Res. Pap., 51, 1517–1528, https://doi.org/10.1016/j.dsr.2004.06.009, 2004.

Hamme, R. C. and Keeling, R. F.: Ocean ventilation as a driver of interannual variability in atmospheric potential oxygen, Tellus B Chem. Phys. Meteorol., 60, 706–717, https://doi.org/10.1111/j.1600-0889.2008.00376.x, 2008.

Heimann, M. and Körner, S.: The global atmospheric tracer model TM3: Model description and user’s manual Release 3.8a, 2003.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hintsa, E. J., Moore, F. L., Hurst, D. F., Dutton, G. S., Hall, B. D., Nance, J. D., Miller, B. R., Montzka, S. A., Wolton, L. P., McClure-Begley, A., Elkins, J. W., Hall, E. G., Jordan, A. F., Rollins, A. W., Thornberry, T. D., Watts, L. A., Thompson, C. R., Peischl, J., Bourgeois, I., Ryerson, T. B., Daube, B. C., Gonzalez Ramos, Y., Commane, R., Santoni, G. W., Pittman, J. V., Wofsy, S. C., Kort, E., Diskin, G. S., and Bui, T. P.: UAS Chromatograph for Atmospheric Trace Species (UCATS) – a versatile instrument for trace gas measurements on airborne platforms, Atmospheric Meas. Tech., 14, 6795–6819, https://doi.org/10.5194/amt-14-6795-2021, 2021.

Hockaday, W. C., Masiello, C. A., Randerson, J. T., Smernik, R. J., Baldock, J. A., Chadwick, O. A., and Harden, J. W.: Measurement of soil carbon oxidation state and oxidative ratio by ^{13}C nuclear magnetic resonance, J. Geophys. Res. Biogeosciences, 114, 2008JG000803, https://doi.org/10.1029/2008JG000803, 2009.

Holtslag, A. a. M. and Boville, B. A.: Local Versus Nonlocal Boundary-Layer Diffusion in a Global Climate Model, J. Clim., 6, 1825–1842, https://doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2, 1993.

Hourdin, F. and Armengaud, A.: The Use of Finite-Volume Methods for Atmospheric Advection of Trace Species. Part I: Test of Various Formulations in a General Circulation Model, Mon. Weather Rev., 127, 822–837, https://doi.org/10.1175/1520-0493(1999)127<0822:TUOFVM>2.0.CO;2, 1999.

Hourdin, F., Talagrand, O., and Idelkadi, A.: Eulerian backtracking of atmospheric tracers. II: Numerical aspects, Q. J. R. Meteorol. Soc., 132, 585–603, https://doi.org/10.1256/qj.03.198.B, 2006.

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., Wood, R. A., Meijers, A. J. S., Mulcahy, J., Field, P., Furtado, K., Bodas-Salcedo, A., Williams, K. D., Copsey, D., Josey, S. A., Liu, C., Roberts, C. D., Sanchez, C., Ridley, J., Thorpe, L., Hardiman, S. C., Mayer, M., Berry, D. I., and Belcher, S. E.: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors, Nat. Commun., 9, 3625, https://doi.org/10.1038/s41467-018-05634-2, 2018.

Ishidoya, S., Morimoto, S., Aoki, S., Taguchi, S., Goto, D., Murayama, S., and Nakazawa, T.: Oceanic and terrestrial biospheric CO_2 uptake estimated from atmospheric potential oxygen observed at Ny-Ålesund, Svalbard, and Syowa, Antarctica, Tellus B Chem. Phys. Meteorol., 64, 18924, https://doi.org/10.3402/tellusb.v64i0.18924, 2012.

Ishidoya, S., Uchida, H., Sasano, D., Kosugi, N., Taguchi, S., Ishii, M., Morimoto, S., Tohjima, Y., Nishino, S., Murayama, S., Aoki, S., Ishijima, K., Fujita, R., Goto, D., and Nakazawa, T.: Ship-based observations of atmospheric potential oxygen and regional air–sea O_2 flux in the northern North Pacific and the Arctic Ocean, Tellus B Chem. Phys. Meteorol., 68, 29972, https://doi.org/10.3402/tellusb.v68.29972, 2016.

Ito, T., Cervania, A., Cross, K., Ainchwar, S., and Delawalla, S.: Mapping Dissolved Oxygen Concentrations by Combining Shipboard and Argo Observations Using Machine Learning Algorithms, J. Geophys. Res. Mach. Learn. Comput., 1, e2024JH000272, https://doi.org/10.1029/2024JH000272, 2024.

Jersild, A., Landschützer, P., Gruber, N., and Bakker, D. C. E.: An observation-based global monthly gridded sea surface pCO_2 and air-sea CO_2 flux product from 1982 onward and its monthly climatology (NCEI Accession 0160558), 2017.

Jin, X., Najjar, R. G., Louanchi, F., and Doney, S. C.: A modeling study of the seasonal oxygen budget of the global ocean, J. Geophys. Res. Oceans, 112, 2006JC003731, https://doi.org/10.1029/2006JC003731, 2007.

Jin, Y., Keeling, R. F., Morgan, E. J., Ray, E., Parazoo, N. C., and Stephens, B. B.: A mass-weighted isentropic coordinate for mapping chemical tracers and computing atmospheric inventories, Atmospheric Chem. Phys., 21, 217–238, https://doi.org/10.5194/acp-21-217-2021, 2021.

Jin, Y., Stephens, B. B., Keeling, R. F., Morgan, E. J., Rödenbeck, C., Patra, P. K., and Long, M. C.: Seasonal Tropospheric Distribution and Air-Sea Fluxes of Atmospheric Potential Oxygen From Global Airborne Observations, Glob. Biogeochem. Cycles, 37, e2023GB007827, https://doi.org/10.1029/2023GB007827, 2023.

Jin, Y., Keeling, R. F., Stephens, B. B., Long, M. C., Patra, P. K., Rödenbeck, C., Morgan, E. J., Kort, E. A., and Sweeney, C.: Improved atmospheric constraints on Southern Ocean CO_2 exchange, Proc. Natl. Acad. Sci., 121, e2309333121, https://doi.org/10.1073/pnas.2309333121, 2024.

Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F., and Le Quéré, C.: Gridded fossil CO_2 emissions and related O_2 combustion consistent with national inventories 1959–2018, Sci. Data, 8, https://doi.org/10.1038/s41597-020-00779-6, 2021.

Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Dou, X., Liu, Z., Pickers, P., Ciais, P., Patra, P. K., Chevallier, F., and Le Quéré, C.: Gridded fossil CO_2 emissions and related O_2 combustion consistent with national inventories 1959-2020 (GCP-GridFEDv2021.3), https://doi.org/10.5281/zenodo.5956612, 2022.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteorol. Soc., 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., Gettelman, A., Eaton, B., Boyle, J., Marchand, R., and Ackerman, T. P.: Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators, J. Clim., https://doi.org/10.1175/JCLI-D-11-00469.1, 2012.

Keeling, R.: Development of an Interferometric Oxygen Analyzer for Precise Measurement of the Atmospheric O_2 Mole Fraction, Harvard University, 1988.

Keeling, R. F.: Scripps O_2 Program Data. UC San Diego Library Digital Collections, https://doi.org/10.6075/J0WS8RJRs, 2019.

Keeling, R. F. and Manning, A. C.: Studies of Recent Changes in Atmospheric O_2 Content, in: Treatise on Geochemistry, Elsevier, 385–404, https://doi.org/10.1016/B978-0-08-095975-7.00420-4, 2014.

Keeling, R. F. and Shertz, S. R.: Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, Nature, 358, 723–727, https://doi.org/10.1038/358723a0, 1992.

Keeling, R. F., Najjar, R. P., Bender, M. L., and Tans, P. P.: What atmospheric oxygen measurements can tell us about the global carbon cycle, Glob. Biogeochem. Cycles, 7, 37–67, https://doi.org/10.1029/92GB02733, 1993.

Keeling, R. F., Manning, A. C., McEvoy, E. M., and Shertz, S. R.: Methods for measuring changes in atmospheric O_2 concentration and their application in southern hemisphere air, J.

Geophys. Res. Atmospheres, 103, 3381–3397, https://doi.org/10.1029/97JD02537, 1998.

Keeling, R. F., Walker, S. J., and Paplawsky, W.: Span Sensitivity of the Scripps Interferometric Oxygen Analyzer, 2020.

Knight, C. L., Mallet, M. D., Alexander, S. P., Fraser, A. D., Protat, A., and McFarquhar, G. M.: Cloud Properties and Boundary Layer Stability Above Southern Ocean Sea Ice and Coastal Antarctica, J. Geophys. Res. Atmospheres, 129, e2022JD038280, https://doi.org/10.1029/2022JD038280, 2024.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Jpn. Ser II, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T. Y., Tokuhiro, T., Sato, Y., Matsushita, Y., and Onogi, K.: The JRA-3Q Reanalysis, J. Meteorol. Soc. Jpn. Ser II, 102, 49–109, https://doi.org/10.2151/jmsj.2024-004, 2024.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmospheric Chem. Phys., 5, 417–432, https://doi.org/10.5194/acp-5-417-2005, 2005.

Krol, M., De Bruine, M., Killaars, L., Ouwersloot, H., Pozzer, A., Yin, Y., Chevallier, F., Bousquet, P., Patra, P., Belikov, D., Maksyutov, S., Dhomse, S., Feng, W., and Chipperfield, M. P.: Age of air as a diagnostic for transport timescales in global models, Geosci. Model Dev., 11, 3109–3130, https://doi.org/10.5194/gmd-11-3109-2018, 2018.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Glob. Biogeochem. Cycles, 30, 1396–1417, https://doi.org/10.1002/2015GB005359, 2016.

Lang, F., Huang, Y., Siems, S. T., and Manton, M. J.: Characteristics of the Marine Atmospheric Boundary Layer Over the Southern Ocean in Response to the Synoptic Forcing, J. Geophys. Res. Atmospheres, 123, 7799–7820, https://doi.org/10.1029/2018JD028700, 2018.

Langenfelds, R. L.: Studies of the global carbon cycle using atmospheric oxygen and associated tracers, University of Tasmania, 2002.

Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S.-J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Patra, P. K., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Takigawa, M., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO_2: Experimental overview and diurnal cycle results for 2002, Glob. Biogeochem. Cycles, 22, https://doi.org/10.1029/2007GB003050, 2008.

Liu, Z., Ciais, P., Deng, Z., Davis, S. J., Zheng, B., Wang, Y., Cui, D., Zhu, B., Dou, X., Ke, P., Sun, T., Guo, R., Zhong, H., Boucher, O., Bréon, F.-M., Lu, C., Guo, R., Xue, J., Boucher, E., Tanaka, K., and Chevallier, F.: Carbon Monitor, a near-real-time daily dataset of global CO_2 emission from fossil fuel and cement production, Sci. Data, 7, 1–12, https://doi.org/10.1038/s41597-020-00708-7, 2020.

Long, M. C., Moore, J. K., Lindsay, K., Levy, M., Doney, S. C., Luo, J. Y., Krumhardt, K. M., Letscher, R. T., Grover, M., and Sylvester, Z. T.: Simulations With the Marine Biogeochemistry Library (MARBL), J. Adv. Model. Earth Syst., 13, e2021MS002647, https://doi.org/10.1029/2021MS002647, 2021a.

Long, M. C., Stephens, B. B., McKain, K., Sweeney, C., Keeling, R. F., Kort, E. A., Morgan, E. J., Bent, J. D., Chandra, N., Chevallier, F., Commane, R., Daube, B. C., Krummel, P. B., Loh, Z., Luijkx, I. T., Munro, D., Patra, P., Peters, W., Ramonet, M., Rödenbeck, C., Stavert, A., Tans, P., and Wofsy, S. C.: Strong Southern Ocean carbon uptake evident in airborne observations, Science, 374, 1275–1280, https://doi.org/10.1126/science.abi4355, 2021b.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Bound.-Layer Meteorol., 17, 187–202, https://doi.org/10.1007/BF00117978, 1979.

Luijkx, I. T., Velde, I. R., Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Zhang, H. F., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., and Peters, W.: The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001–2015, Geosci. Model Dev., 10, 2785–2800, https://doi.org/10.5194/gmd-10-2785-2017, 2017.

Maksyutov, S., Patra, P., Onishi, R., Saeki, T., and Nakazawa, T.: NIES/FRCGC Global Atmospheric Tracer Transport Model: Description, Validation, and Surface Sources and Sinks Inversion, J. Earth Simulator, 9, 3–18, https://doi.org/10.32131/jes.9.3, 2008.

Maksyutov, S., Oda, T., Saito, M., Janardanan, R., Belikov, D., Kaiser, J. W., Zhuravlev, R., Ganshin, A., Valsala, V. K., Andrews, A., Chmura, L., Dlugokencky, E., Haszpra, L., Langenfelds, R. L., Machida, T., Nakazawa, T., Ramonet, M., Sweeney, C., and Worthy, D.: Technical note: A high-resolution inverse modelling technique for estimating surface CO_2 fluxes based on the NIES-TM–FLEXPART coupled transport model and its adjoint, Atmospheric Chem. Phys., 21, 1245–1266, https://doi.org/10.5194/acp-21-1245-2021, 2021.

Manning, A. C. and Keeling, R. F.: Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network, Tellus B Chem. Phys. Meteorol., 58, 95, https://doi.org/10.1111/j.1600-0889.2006.00175.x, 2006.

Mellor, G. L. and Yamada, T.: A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers, J. Atmospheric Sci., https://doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2, 1974.

Miyazaki, K., Patra, P. K., Takigawa, M., Iwasaki, T., and Nakazawa, T.: Global-scale transport of carbon dioxide in the troposphere, J. Geophys. Res. Atmospheres, 113, 2007JD009557, https://doi.org/10.1029/2007JD009557, 2008.

Morgan, E. J., Manizza, M., Keeling, R. F., Resplandy, L., Mikaloff-Fletcher, S. E., Nevison, C. D., Jin, Y., Bent, J. D., Aumont, O., Doney, S. C., Dunne, J. P., John, J., Lima, I. D., Long, M. C., and Rodgers, K. B.: An Atmospheric Constraint on the Seasonal Air-Sea Exchange of Oxygen and Heat in the Extratropics, J. Geophys. Res. Oceans, 126, e2021JC017510, https://doi.org/10.1029/2021JC017510, 2021.

Naegler, T., Ciais, P., Rodgers, K., and Levin, I.: Excess radiocarbon constraints on air-sea gas exchange and the uptake of CO2 by the oceans, Geophys. Res. Lett., 33, https://doi.org/10.1029/2005GL025408, 2006.

Naegler, T., Ciais, P., Orr, J. C., Aumont, O., and Rödenbeck, C.: On evaluating ocean models with atmospheric potential oxygen, Tellus B Chem. Phys. Meteorol., 59, https://doi.org/10.1111/j.1600-0889.2006.00197.x, 2007.

Najjar, R. G. and Keeling, R. F.: Mean annual cycle of the air-sea oxygen flux: A global view, Glob. Biogeochem. Cycles, 14, 573–584, https://doi.org/10.1029/1999GB900086, 2000.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification, Bound.-Layer Meteorol., 112, 1–31, https://doi.org/10.1023/B:BOUN.0000020164.04146.98, 2004.

Nevison, C., Munro, D., Lovenduski, N., Cassar, N., Keeling, R., Krummel, P., and Tjiputra, J.: Net Community Production in the Southern Ocean: Insights From Comparing Atmospheric Potential Oxygen to Satellite Ocean Color Algorithms and Ocean Models, Geophys. Res. Lett., 45, 10549–10559, https://doi.org/10.1029/2018GL079575, 2018.

Nevison, C. D., Mahowald, N. M., Doney, S. C., Lima, I. D., and Cassar, N.: Impact of variable air-sea O_2 and CO_2 fluxes on atmospheric potential oxygen (APO) and land-ocean carbon sink partitioning, Biogeosciences, 5, 875–889, https://doi.org/10.5194/bg-5-875-2008, 2008.

Nevison, C. D., Keeling, R. F., Kahru, M., Manizza, M., Mitchell, B. G., and Cassar, N.: Estimating net community production in the Southern Ocean based on atmospheric potential oxygen and satellite ocean color data, Glob. Biogeochem. Cycles, 26, https://doi.org/10.1029/2011GB004040, 2012.

Nevison, C. D., Manizza, M., Keeling, R. F., Kahru, M., Bopp, L., Dunne, J., Tiputra, J., Ilyina, T., and Mitchell, B. G.: Evaluating the ocean biogeochemical components of Earth system models using atmospheric potential oxygen and ocean color data, Biogeosciences, 12, 193–208, https://doi.org/10.5194/bg-12-193-2015, 2015.

Nevison, C. D., Manizza, M., Keeling, R. F., Stephens, B. B., Bent, J. D., Dunne, J., Ilyina, T., Long, M., Resplandy, L., Tjiputra, J., and Yukimoto, S.: Evaluating CMIP5 ocean biogeochemistry and Southern Ocean carbon uptake using atmospheric potential oxygen: Present-day performance and future projection, Geophys. Res. Lett., 43, 2077–2085, https://doi.org/10.1002/2015GL067584, 2016.

Nguyen, L. N. T., Meijer, H. A. J., van Leeuwen, C., Kers, B. A. M., Scheeren, H. A., Jones, A. E., Brough, N., Barningham, T., Pickers, P. A., Manning, A. C., and Luijkx, I. T.: Two decades of flask observations of atmospheric $\delta(O_2/N_2)$, CO_2, and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland), and 3 years from Halley station (Antarctica), Earth Syst. Sci. Data, 14, 991–1014, https://doi.org/10.5194/essd-14-991-2022, 2022.

Niwa, Y., Tomita, H., Satoh, M., and Imasu, R.: A Three-Dimensional Icosahedral Grid Advection Scheme Preserving Monotonicity and Consistency with Continuity for Atmospheric Tracer Transport, J. Meteorol. Soc. Jpn. Ser II, 89, 255–268, https://doi.org/10.2151/jmsj.2011-306, 2011.

Niwa, Y., Machida, T., Sawa, Y., Matsueda, H., Schuck, T. J., Brenninkmeijer, C. A. M., Imasu, R., and Satoh, M.: Imposing strong constraints on tropical terrestrial CO_2 fluxes using passenger aircraft based measurements, J. Geophys. Res. Atmospheres, 117, https://doi.org/10.1029/2012JD017474, 2012.

Niwa, Y., Tomita, H., Satoh, M., Imasu, R., Sawa, Y., Tsuboi, K., Matsueda, H., Machida, T., Sasakawa, M., Belan, B., and Saigusa, N.: A 4D-Var inversion system based on the icosahedral grid model (NICAM-TM 4D-Var v1.0) – Part 1: Offline forward and adjoint transport models, Geosci. Model Dev., 10, 1157–1174, https://doi.org/10.5194/gmd-10-1157-2017, 2017.

Noda, A. T., Oouchi, K., Satoh, M., Tomita, H., Iga, S., and Tsushima, Y.: Importance of the subgrid-scale turbulent moist process: Cloud distribution in global cloud-resolving simulations, Atmospheric Res., 96, 208–217, https://doi.org/10.1016/j.atmosres.2009.05.007, 2010.

Numaguti, A., Takahashi, M., Nakajima, T., and Sumi, A.: Description of CCSR/NIES Atmospheric General Circulation Model., CGER's Supercomput. Monogr. Rep., 3 (Ch 1), National Institute for Environmental Studies, Tsukuba, Japan., 1997.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2 version 2016 (ODIAC2016): a global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst. Sci. Data, 10, 87–107, https://doi.org/10.5194/essd-10-87-2018, 2018.

Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., Randall, D. A., Kawa, S. R., Pauluis, O., and Doney, S. C.: Moist synoptic transport of CO_2 along the mid-latitude storm track, Geophys. Res. Lett., 38, 2011GL047238, https://doi.org/10.1029/2011GL047238, 2011.

Patra, P. K., Law, R. M., Peters, W., Rödenbeck, C., Takigawa, M., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S.-J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO_2: Analysis of synoptic-scale variations for the period 2002–2003, Glob. Biogeochem. Cycles, 22, https://doi.org/10.1029/2007GB003081, 2008.

Patra, P. K., Houweling, S., Krol, M., Bousquet, P., Belikov, D., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Corbin, K., Fortems-Cheiney, A., Fraser, A., Gloor, E., Hess, P., Ito, A., Kawa, S. R., Law, R. M., Loh, Z., Maksyutov, S., Meng, L., Palmer, P. I., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C.: TransCom model simulations of CH_4 and related species: linking transport, surface flux and chemical loss with CH_4 variability in the troposphere and lower stratosphere, Atmospheric Chem. Phys., 11, 12813–12837, https://doi.org/10.5194/acp-11-12813-2011, 2011.

Patra, P. K., Takigawa, M., Watanabe, S., Chandra, N., Ishijima, K., and Yamashita, Y.: Improved Chemical Tracer Simulation by MIROC4.0-based Atmospheric Chemistry-Transport Model (MIROC4-ACTM), SOLA, 14, 91–96, https://doi.org/10.2151/sola.2018-016, 2018.

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7, Atmospheric Chem. Phys., 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

Pickers, P. A., Manning, A. C., Sturges, W. T., Le Quéré, C., Mikaloff Fletcher, S. E., Wilson, P. A., and Etchells, A. J.: In situ measurements of atmospheric O_2 and CO_2 reveal an unexpected O_2 signal over the tropical Atlantic Ocean, Glob. Biogeochem. Cycles, 31, 1289–1305, https://doi.org/10.1002/2017GB005631, 2017.

Pickers, P. A., Manning, A. C., Le Quéré, C., Forster, G. L., Luijkx, I. T., Gerbig, C., Fleming, L. S., and Sturges, W. T.: Novel quantification of regional fossil fuel CO_2 reductions during COVID-19 lockdowns using atmospheric oxygen measurements, Sci. Adv., 8, eabl9250, https://doi.org/10.1126/sciadv.abl9250, 2022.

Resplandy, L., Keeling, R. F., Stephens, B. B., Bent, J. D., Jacobson, A., Rödenbeck, C., and Khatiwala, S.: Constraints on oceanic meridional heat transport from combined measurements of oxygen and carbon, Clim. Dyn., 47, 3335–3357, https://doi.org/10.1007/s00382-016-3029-3, 2016.

Resplandy, L., Keeling, R. F., Eddebbar, Y., Brooks, M., Wang, R., Bopp, L., Long, M. C., Dunne, J. P., Koeve, W., and Oschlies, A.: Quantification of ocean heat uptake from changes in atmospheric O_2 and CO_2 composition, Sci. Rep., 9, 20244, https://doi.org/10.1038/s41598-019-56490-z, 2019.

Rio, C. and Hourdin, F.: A Thermal Plume Model for the Convective Boundary Layer: Representation of Cumulus Clouds, J. Atmospheric Sci., https://doi.org/10.1175/2007JAS2256.1, 2008.

Rödenbeck, C., Quéré, C. L., Heimann, M., and Keeling, R. F.: Interannual variability in oceanic biogeochemical processes inferred by inversion of atmospheric O_2/N_2 and CO_2 data, Tellus B Chem. Phys. Meteorol., 60, 685–705, https://doi.org/10.1111/j.1600-0889.2008.00375.x, 2008.

Rödenbeck, C., Adcock, K. E., Eritt, M., Gachkivskyi, M., Gerbig, C., Hammer, S., Jordan, A., Keeling, R. F., Levin, I., Maier, F., Manning, A. C., Moossen, H., Munassar, S., Pickers, P. A., Rothe, M., Tohjima, Y., and Zaehle, S.: The suitability of atmospheric oxygen measurements to constrain western European fossil-fuel CO_2 emissions and their trends, Atmospheric Chem. Phys., 23, 15767–15782, https://doi.org/10.5194/acp-23-15767-2023, 2023.

Russell, G. L. and Lerner, J. A.: A New Finite-Differencing Scheme for the Tracer Transport Equation, J. Appl. Meteorol., 20, 1483–1498, https://doi.org/10.1175/1520-0450(1981)020<1483:ANFDSF>2.0.CO;2, 1981.

Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., and Wang, Z.: Assessment of Southern Ocean mixed-layer depths in CMIP5 models: Historical bias and forcing response, J. Geophys. Res. Oceans, 118, 1845–1862, https://doi.org/10.1002/jgrc.20157, 2013.

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., McManus, J. B., Peischl, J., Ryerson, T. B., Holloway, J. S., Andrews, A. E., Sweeney, C., Hall, B., Hintsa, E. J., Moore, F. L., Elkins, J. W., Hurst, D. F., Stephens, B. B., Bent, J., and Wofsy, S. C.: Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite – CO_2, CH_4, N_2O, and CO – during the CalNex and HIPPO campaigns, Atmospheric Meas. Tech., 7, 1509–1526, https://doi.org/10.5194/amt-7-1509-2014, 2014.

Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., Noda, A. T., Yamada, Y., Goto, D., Sawada, M., Miyoshi, T., Niwa, Y., Hara, M., Ohno, T., Iga, S., Arakawa, T., Inoue, T., and Kubokawa, H.: The Non-hydrostatic Icosahedral Atmospheric Model: description and development, Prog. Earth Planet. Sci., 1, 18, https://doi.org/10.1186/s40645-014-0018-1, 2014.

Schuh, A. E. and Jacobson, A. R.: Uncertainty in parameterized convection remains a key obstacle for estimating surface fluxes of carbon dioxide, Atmospheric Chem. Phys., 23, 6285–6297, https://doi.org/10.5194/acp-23-6285-2023, 2023.

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., Chevallier, F., Crowell, S., Davis, K. J., Deng, F., Denning, S., Feng, L., Jones, D., Liu, J., and Palmer, P. I.: Quantifying the Impact of Atmospheric Transport Uncertainty on CO_2 Surface Flux Estimates, Glob. Biogeochem. Cycles, 33, 484–500, https://doi.org/10.1029/2018GB006086, 2019.

Schuldt, K. N., Mund, J., Luijkx, I. T., Aalto, T., Abshire, J. B., Aikin, K., Andrews, A., Aoki, S., Apadula, F., Baier, B., Bakwin, P., Bartyzel, J., Bentz, G., Bergamaschi, P., Beyersdorf, A., Biermann, T., Biraud, S. C., Boenisch, H., Bowling, D., Brailsford, G., Chen, G., Chen, H., Chmura, L., Clark, S., Climadat, S., Colomb, A., Commane, R., Conil, S., Cox, A., Cristofanelli, P., Cuevas, E., Curcoll, R., Daube, B., Davis, K., De Mazière, M., De Wekker, S., Della Coletta, J., Delmotte, M., DiGangi, J. P., Dlugokencky, E., Elkins, J. W., Emmenegger, L., Fang, S., Fischer, M. L., Forster, G., Frumau, A., Galkowski, M., Gatti, L. V., Gehrlein, T., Gerbig, C., Gheusi, F., Gloor, E., Gomez-Trueba, V., Goto, D., Griffis, T., Hammer, S., Hanson, C., Haszpra, L., Hatakka, J., Heimann, M., Heliasz, M., Hensen, A., Hermanssen, O., Hintsa, E., Holst, J., Ivakhov, V., Jaffe, D., Joubert, W., Karion, A., Kawa, S. R., Kazan, V., Keeling, R., Keronen, P., Kolari, P., Kominkova, K., Kort, E., Kozlova, E., Krummel, P., Kubistin, D., Labuschagne, C., Lam, D. H., Langenfelds, R., Laurent, O., Laurila, T., Lauvaux, T., Lavric, J., Law, B., Lee, O. S., Lee, J., Lehner, I., Leppert, R., Leuenberger, M., Levin, I., Levula, J., Lin, J., Lindauer, M., Loh, Z., Lopez, M., Machida, T., et al.: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2020; obspack_co2_1_GLOBALVIEWplus_v7.0_2021-08-18, , https://doi.org/10.25925/20210801, 2021.

Severinghaus, J.: Studies of the terrestrial O_2 and carbon cycles in sand dune gases and in Biosphere 2, Columbia University, 1995.

Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., and Dunne, J. P.: GOBAI-O_2 : temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly 2 decades, Earth Syst. Sci. Data, 15, 4481–4518, https://doi.org/10.5194/essd-15-4481-2023, 2023.

Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H.: The CO_2 release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, Atmospheric Chem. Phys., 11, 6855–6870, https://doi.org/10.5194/acp-11-6855-2011, 2011.

Stephens, B.: ORCAS Merge Products. Version 1.0. UCAR/NCAR - Earth Observing Laboratory, https://doi.org/10.5065/D6SB445X., 2017.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: HIPPO-1 airborne oxygen instrument. Version 2.0. UCAR/NCAR—Earth Observing Laboratory., https://doi.org/10.5065/D6J38QVV, 2021a.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: HIPPO-2 airborne oxygen instrument. Version 2.0. UCAR/NCAR—Earth Observing Laboratory., https://doi.org/10.5065/D65Q4TF0, 2021b.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: HIPPO-3 airborne oxygen instrument. Version 2.0. UCAR/NCAR—Earth Observing Laboratory., https://doi.org/10.5065/D67H1GXJ, 2021c.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: HIPPO-4 airborne oxygen instrument. Version 2.0. UCAR/NCAR—Earth Observing Laboratory., https://doi.org/10.5065/D679431D, 2021d.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: HIPPO-5 airborne oxygen instrument. Version 2.0. UCAR/NCAR—Earth Observing Laboratory., https://doi.org/10.5065/D6WW7G0D, 2021e.

Stephens, B., Keeling, R., Bent, J., Watt, A., Shertz, S., and Paplawsky, W.: ORCAS Airborne Oxygen Instrument. Version 2.0. UCAR/NCAR - Earth Observing Laboratory., https://doi.org/10.5065/D6N29VC6, 2021f.

Stephens, B., Long, M., Jin, Y., Chandra, N., Chevallier, F., Hooghiem, J., Luijkx, I., Maksyutov, S., Morgan, E., Niwa, Y., Patra, P., Rodenbeck, C., and Vance, J.: Atmospheric Potential Oxygen forward Model Intercomparison Project (APO-MIP), https://doi.org/10.5065/F3PW-A676, 2025.

Stephens, B. B.: ARSV Laurence M. Gould Atmospheric O_2 and CO_2 Measurements. Version

1.0. UCAR/NCAR - Earth Observing Laboratory, https://doi.org/10.26023/FDDD-PC3X-4M0X, 2025.

Stephens, B. B., Keeling, R. F., Heimann, M., Six, K. D., Murnane, R., and Caldeira, K.: Testing global ocean carbon cycle models using measurements of atmospheric O_2 and CO_2 concentration, Glob. Biogeochem. Cycles, 12, 213–230, https://doi.org/10.1029/97GB03500, 1998.

Stephens, B. B., Keeling, R. F., and Paplawsky, W. J.: Shipboard measurements of atmospheric oxygen using a vacuum-ultraviolet absorption technique, Tellus B Chem. Phys. Meteorol., 75B, 857–878, https://doi.org/10.3402/tellusb.v55i4.16386, 2003.

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J., and Denning, A. S.: Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO_2, Science, 316, 1732–1735, https://doi.org/10.1126/science.1137004, 2007.

Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., Atlas, E. L., Beaton, S., Bent, J. D., Blake, N. J., Bresch, J. F., Casey, J., Daube, B. C., Diao, M., Diaz, E., Dierssen, H., Donets, V., Gao, B.-C., Gierach, M., Green, R., Haag, J., Hayman, M., Hills, A. J., Hoecker-Martínez, M. S., Honomichl, S. B., Hornbrook, R. S., Jensen, J. B., Li, R.-R., McCubbin, I., McKain, K., Morgan, E. J., Nolte, S., Powers, J. G., Rainwater, B., Randolph, K., Reeves, M., Schauffler, S. M., Smith, K., Smith, M., Stith, J., Stossmeister, G., Toohey, D. W., and Watt, A. S.: The O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study, Bull. Am. Meteorol. Soc., 99, 381–402, https://doi.org/10.1175/BAMS-D-16-0206.1, 2018.

Stephens, B. B., Morgan, E. J., Bent, J. D., Keeling, R. F., Watt, A. S., Shertz, S. R., and Daube, B. C.: Airborne measurements of oxygen concentration from the surface to the lower stratosphere and pole to pole, Atmospheric Meas. Tech., 14, 2543–2574, https://doi.org/10.5194/amt-14-2543-2021, 2021g.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmospheric Chem. Phys., 5, 2461–2474, https://doi.org/10.5194/acp-5-2461-2005, 2005.

Thompson, C., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W., Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R., Wang, S., Allen, H. M., Asher, E., Bian, H., Blake, D., Bourgeois, I., Budney, J., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J. D., Daube, B., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Fiore, A. M., Flynn, C. M., Guo, H., Hall, S. R., Hannun, R. A., Hills, A., Hintsa, E. J., Hodzic, A., Hornbrook, R. S., Huey, L. G., Jimenez, J. L., Keeling, R. F., Kim, M. J., Kupc, A., Lacey, F., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A., Nguyen, L., Gonzalez, Y., Rollins, A., Rosenlof, K., Sargent, M., Schill, G., Schwarz, J. P., Clair, J. M. St., Steenrod, S. D., Stephens, B. B., Strahan, S. E., Strode, S. A., Sweeney, C., Thames, A. B., Ullmann, K., Wagner, N., Weber, R., Weinzierl, B., Wennberg, P. O., Williamson, C. J., Wolfe, G. M., and Zeng, L.: The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere, Bull. Am. Meteorol. Soc., 103, E761–E790, https://doi.org/10.1175/BAMS-D-20-0315.1, 2022.

Thompson, R., Manning, A. C., Lowe, D. C., and Weatherburn, D. C.: A ship-based methodology for high precision atmospheric oxygen measurements and its application in the Southern Ocean region, Tellus B Chem. Phys. Meteorol., 59, 643, https://doi.org/10.1111/j.1600-0889.2007.00292.x, 2007.

Thompson, R. L., Patra, P. K., Ishijima, K., Saikawa, E., Corazza, M., Karstens, U., Wilson, C., Bergamaschi, P., Dlugokencky, E., Sweeney, C., Prinn, R. G., Weiss, R. F., O'Doherty, S., Fraser, P. J., Steele, L. P., Krummel, P. B., Saunois, M., Chipperfield, M., and Bousquet, P.: TransCom N_2O model inter-comparison – Part 1: Assessing the influence of transport and surface fluxes on tropospheric N_2O variability, Atmospheric Chem. Phys., 14, 4349–4368, https://doi.org/10.5194/acp-14-4349-2014, 2014.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Weather Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Tohjima, Y., Minejima, C., Mukai, H., Machida, T., Yamagishi, H., and Nojiri, Y.: Analysis of seasonality and annual mean distribution of atmospheric potential oxygen (APO) in the Pacific region, Glob. Biogeochem. Cycles, 26, 2011GB004110, https://doi.org/10.1029/2011GB004110, 2012.

Tohjima, Y., Terao, Y., Mukai, H., Machida, T., Nojiri, Y., and Maksyutov, S.: ENSO-related variability in latitudinal distribution of annual mean atmospheric potential oxygen (APO) in the equatorial Western Pacific, Tellus B Chem. Phys. Meteorol., 67, 25869, https://doi.org/10.3402/tellusb.v67.25869, 2015.

Tohjima, Y., Mukai, H., Machida, T., Hoshina, Y., and Nakaoka, S.-I.: Global carbon budgets estimated from atmospheric O_2/N_2 and CO_2 observations in the western Pacific region over a 15-year period, Atmospheric Chem. Phys., 19, 9269–9285, https://doi.org/10.5194/acp-19-9269-2019, 2019.

Tohjima, Y., Shirai, T., Ishizawa, M., Mukai, H., Machida, T., Sasakawa, M., Terao, Y., Tsuboi, K., Takao, S., and Nakaoka, S.: Observed APO Seasonal Cycle in the Pacific: Estimation of Autumn O_2 Oceanic Emissions, Glob. Biogeochem. Cycles, 38, e2024GB008230, https://doi.org/10.1029/2024GB008230, 2024.

Truong, S. C. H., Huang, Y., Lang, F., Messmer, M., Simmonds, I., Siems, S. T., and Manton, M. J.: A Climatology of the Marine Atmospheric Boundary Layer Over the Southern Ocean From Four Field Campaigns During 2016–2018, J. Geophys. Res. Atmospheres, 125, e2020JD033214, https://doi.org/10.1029/2020JD033214, 2020.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Boeira Dias, F., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., Josey, S. A., Kobayashi, C., Kobayashi, S., Komuro, Y., Large, W. G., Le Sommer, J., Marsland, S. J., Masina, S., Scheinert, M., Tomita, H., Valdivieso, M., and Yamazaki, D.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), Ocean Model., 130, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002, 2018.

Van Leer, B.: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, J. Comput. Phys., 23, 276–299, https://doi.org/10.1016/0021-9991(77)90095-X, 1977.

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, Bound.-Layer Meteorol., 81, 245–269, https://doi.org/10.1007/BF02430331, 1996.

Watanabe, S., Miura, H., Sekiguchi, M., Nagashima, T., Sudo, K., Emori, S., and Kawamiya, M.: Development of an Atmospheric General Circulation Model for Integrated Earth System Modeling on the Earth Simulator, J. Earth Simulator, 9, 27–35, https://doi.org/10.32131/jes.9.27, 2008.

Wofsy, S.: ATom: Merged atmospheric chemistry, trace gases, and aerosols, version 2 (version 2.0). ORNL Distributed Active Archive Center., https://doi.org/10.3334/ORNLDAAC/1925, 2021.

Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Philos. Trans. R. Soc. A, 369, 2073–2086, https://doi.org/10.1098/rsta.2010.0313, 2011.

Wofsy, S. C.: HIPPO merged 10-second meteorology, atmospheric chemistry, and aerosol data. Version 1.0. UCAR/NCAR - Earth Observing Laboratory., https://doi.org/10.3334/CDIAC/HIPPO_010, 2017.

Worrall, F., Clay, G. D., Masiello, C. A., and Mynheer, G.: Estimating the oxidative ratio of the global terrestrial biosphere carbon, Biogeochemistry, 115, 23–32, https://doi.org/10.1007/s10533-013-9877-6, 2013.

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., Molina, M. J., Krumhardt, K., Mogen, S., Lindsay, K., Lombardozzi, D., Wieder, W., Kim, W. M., Richter, J. H., Long, M., Danabasoglu, G., Bailey, D., Holland, M., Lovenduski, N., Strand, W. G., and King, T.: The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2, Geosci. Model Dev., 15, 6451–6493, https://doi.org/10.5194/gmd-15-6451-2022, 2022.

Zhang, G. J. and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model, Atmosphere-Ocean, 33, 407–446, https://doi.org/10.1080/07055900.1995.9649539, 1995.