Author Contributions:
Conceptualization: Siyuan Wang
Data curation: Siyuan Wang
Formal analysis: Siyuan Wang
Investigation: Siyuan Wang
Methodology: Siyuan Wang
Project administration: Siyuan Wang
Resources: Siyuan Wang
Software: Siyuan Wang
Supervision: Siyuan Wang
Validation: Siyuan Wang
Visualization: Siyuan Wang
(continued)

Ocean Biogeochemistry Control on the Marine Emissions of Brominated Very Short-Lived Ozone-Depleting Substances: A Machine-Learning Approach

Siyuan Wang[1,2], Douglas Kinnison[2], Stephen A. Montzka[3], Eric C. Apel[2], Rebecca S. Hornbrook[2], Alan J. Hills[2], Donald R. Blake[4], Barbara Barletta[4], Simone Meinardi[4], Colm Sweeney[3], Fred Moore[3], Matthew Long[5], Alfonso Saiz-Lopez[6], Rafael Pedro Fernandez[6,7], Simone Tilmes[2,5], Louisa K. Emmons[2], and Jean-François Lamarque[5]

[1]Advanced Study Program, National Center for Atmospheric Research, Boulder, CO, USA, [2]Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [3]Global Monitoring Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [4]Department of Chemistry, University of California, Irvine, CA, USA, [5]Climate and Global Dynamics, National Center for Atmospheric Research, Boulder, CO, USA, [6]Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano, CSIC, Madrid, Spain, [7]National Research Council (CONICET), FCEN-UNCuyo, UNT-FRM, Mendoza, Argentina

Abstract Halogenated very short lived substances (VSLS) affect the ozone budget in the atmosphere. Brominated VSLS are naturally emitted from the ocean, and current oceanic emission inventories vary dramatically. We present a new global oceanic emission inventory of Br-VSLS (bromoform and dibromomethane), considering the physical forcing in the ocean and the atmosphere, as well as the ocean biogeochemistry control. A data-oriented machine-learning emulator was developed to couple the air-sea exchange with the ocean biogeochemistry. The predicted surface seawater concentrations and the surface atmospheric mixing ratios of Br-VSLS are evaluated with long-term, global-scale observations; and the predicted vertical distributions of Br-VSLS are compared to the global airborne observations in both boreal summer and winter. The global marine emissions of bromoform and dibromomethane are estimated to be 385 and 54 Gg Br per year, respectively. The new oceanic emission inventory of Br-VSLS is more skillful than the widely used top-down approaches for representing the seasonal/spatial variations and the annual means of atmospheric concentrations. The new approach improves the model predictability for the coupled Earth system model and can be used as a basis for investigating the past and future ocean emissions and feedbacks under climate change. This model framework can be used to calculate the bidirectional oceanic fluxes for other compounds of interest.

Plain Language Summary Halogen atoms released from the man-made, long-lived ozone-are the major cause of the stratospheric ozone depletion. Recent studies found that natural bromine-containing very short lived substances are of particular importance for the ozone radiative forcing in the lower stratosphere. These bromine-containing short-lived ozone-depleting substances are naturally produced from phytoplankton in seawater and released into the atmosphere. The past decade has seen increased applications of machine-learning techniques in climate-related research. In this work, we use a data-oriented machine-learning algorithm to calculate the production of bromine-containing short-lived substances in the ocean, representing a fairly accurate and computationally efficient solution for addressing future climate predictions.

1. Introduction

Brominated very short lived substances (Br-VSLS), including bromoform ($CHBr_3$) and dibromomethane (CH_2Br_2), are important precursors of reactive bromine species in the atmosphere. The multiphase chemistry involving reactive bromine species affects ozone and mercury in both the troposphere and the stratosphere (Saiz-Lopez et al., 2012; Saiz-Lopez & von Glasow, 2012; Schmidt et al., 2016; Sherwen et al., 2017; Wales et al., 2018; Wang et al., 2015; Wang, McNamara, et al., 2019) and exert a particularly strong ozone

Writing - original draft: Siyuan Wang
Writing – review & editing: Siyuan Wang

radiative impact in the lower stratosphere (Engel et al., 2019; Hossaini et al., 2015; Navarro et al., 2015; Salawitch et al., 2005). Br-VSLS are naturally produced in seawater, and the ocean is thought to be the dominant global source through air-sea exchange (Butler et al., 2007; Carpenter & Liss, 2000; Quack & Wallace, 2003; Yokouchi et al., 2017). The natural synthesis mechanisms of Br-VSLS in seawater remain poorly understood but are mainly related to phytoplankton and photosynthetic processes in the seawater (Moore et al., 1996; Lin & Manley, 2012; Liu, Yvon-Lewis, Thornton, Butler, et al., 2013; Liu et al., 2015; Shibazaki et al., 2016; Hughes & Sun, 2016).

Current oceanic emission inventories of Br-VSLS vary dramatically. For example, the global annual oceanic emissions of $CHBr_3$ range from 72–238 Gg Br per year in bottom-up inventories (Lennartz et al., 2015; Stemmler et al., 2015; Ziska et al., 2013) to 425–840 Gg Br per year in the top-down (observationally derived) inventories (Butler et al., 2007; Liang et al., 2010; Ordóñez et al., 2012; Quack & Wallace, 2003; Warwick et al., 2006). These inventories show particularly large discrepancies in the high-latitude regions (Hossaini et al., 2013, 2016). The top-down inventories, although providing reasonable constraints on the stratospheric source gas injection (Engel et al., 2019; Hossaini et al., 2016), may have limited predictability of future climate variations and feedbacks, as the emission fluxes (derived from atmospheric observations) have not been designed to provide information on how these fluxes would respond to changes in local conditions or external forcing. Ordóñez et al. (2012) derived global marine Br-VSLS emissions for chemistry-climate models by scaling the observed fluxes to the chlorophyll-a observations (climatology from the SeaWiFS satellite), which is an assumed proxy for the ocean biogeochemical production of these compounds. However, in situ observations found no robust relationships between the Br-VSLS and chlorophyll (Carpenter et al., 2009; Liu, Yvon-Lewis, Thornton, Campbell, et al., 2013). Efforts have been made to develop bottom-up emission inventories, to better quantify the oceanic emissions in the coupled Earth system. Lennartz and colleagues investigated the global marine emissions of Br-VSLS (Lennartz et al., 2015) using a modular Earth system model (Jöckel et al., 2005) with an online air-sea exchange module (Pozzer et al., 2006), based on the observationally derived global distributions of the surface seawater Br-VSLS (Ziska et al., 2013). This coupled framework is a key step toward a more skillful representation and improved predictability of interactive emissions within chemistry-climate models. The surface seawater concentrations, primarily determined by the ocean biogeochemical and the physical processes in the seawater, set the foundation for the air-sea exchange processes. Ziska et al. (2013) conducted pioneer work in producing a global product of the surface seawater distributions of Br-VSLS for use in global models. However, the objective mapping approach used (Ziska et al., 2013) did not adequately represent the complex ocean biogeochemistry control and hence was unable to fully resolve the seasonal or spatial variations in the natural seawater. Stemmler et al. (2015) developed a state-of-the-art three-dimensional ocean biogeochemistry model to simulate the explicit sources and sinks of $CHBr_3$ in the ocean, yet the potentially important $CHBr_3$ production in coastal waters as well as from sea ice algae (Carpenter et al., 2007) are not represented. Additionally, to our knowledge, CH_2Br_2 coupled air-sea emissions were not simulated (Stemmler et al., 2015).

Machine-learning and artificial intelligence have proven to be a valuable tool in the Earth system science. To name a few inspiring applications, Sherwen et al. (2019) and Roshan and DeVries (2017) leveraged existing satellite or gridded in situ observations and developed new products for model use. Recent years have seen "online" applications as well. For example, machine-learning "emulators," trained by process-level models (often computationally expensive) or observations, may be used to replace the (sometimes highly uncertain) parameterizations (e.g., planetary boundary layer schemes and cloud microphysics schemes) in Earth system models (Sobhani et al., 2018). Although common algorithms (e.g., random forest) may be used in the online (e.g., Sobhani et al., 2018) and offline applications (e.g., Sherwen et al., 2019), the online approach has advantages: The machine-learning emulator incorporated into the Earth system model (online) is coupled to other physical and chemical processes within the Earth system model, therefore will respond to changes in local conditions or external forcing, and hence may reveal insights into the feedback mechanisms. In the aforementioned examples (Roshan & DeVries, 2017; Sherwen et al., 2019), climatologies (usually monthly) of gridded observations and satellite products are used; therefore, the temporal variations beyond the climatologies (e.g., decadal, interannual, and daily) or feedback mechanisms cannot be resolved. Moreover, the offline approaches, trained by present-day observations, may be prone to systematic biases when used for future/past projections. The online approaches may be less affected by such biases (if thoroughly evaluated), since the same variables are used for both training and predicting; therefore, the

biases would be bypassed. Moreover, the online emulator can be flexible in terms of temporal (e.g., daily and weekly) and spatial resolution, depending on the temporal and spatial coverage of the in situ observation and the model resolution.

The future trend of the marine emissions of these ozone-depleting substances remains unclear. A recent study examined the future marine emissions of Br-VSLS under different climate change scenarios, assuming constant surface seawater concentrations of Br-VSLS in the future (Ziska et al., 2017). However, anthropogenically forced climate change may lead to considerable changes in the ocean biogeochemical states and primary productivity (Krumhardt et al., 2017; Long et al., 2016). To our knowledge, there has been no attempt to implement an approach to understand the influence of future changes in the ocean biogeochemical states and the productivity on the marine emissions of Br-VSLS. A trained machine-learning emulator representing the ocean biogeochemistry control, coupled to an Earth system model, may provide insights into this issue.

In this work, we present a data-oriented machine-learning emulator to calculate the global surface seawater distributions of $CHBr_3$ and CH_2Br_2 based on ocean biogeochemistry. The calculated present-day monthly climatologies of the surface seawater concentrations of $CHBr_3$ and CH_2Br_2 are evaluated with observations (HalOcAt data set: Halocarbons in the Ocean and Atmosphere; see Ziska et al., 2013) and used to drive the air-sea exchange processes in the global chemistry-climate model, the Community Atmospheric Model with chemistry (CAM-chem). An online air-sea exchange framework (Wang, McNamara, et al., 2019) was developed for CAM-chem, which is used in this work to investigate the marine emissions of Br-VSLS in a 10-year period (2005–2015). The modeled surface atmospheric concentrations of $CHBr_3$ and CH_2Br_2 are compared to ground-based observations from the National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) global monitoring network (Montzka et al., 2003). The new bottom-up oceanic emission inventories of Br-VSLS provide improved predictability for the seasonal/spatial variations of these short-lived ozone-depleting substances over the widely used top-down inventories. Moreover, the modeled vertical distributions of $CHBr_3$ and CH_2Br_2 are compared to the airborne observations obtained from the first and the second deployments of the National Aeronautics and Space Administration (NASA) Atmospheric Tomography Mission (Wofsy et al., 2018). The global marine emissions of $CHBr_3$ and CH_2Br_2 are compared with previous studies.

2. Methods

CAM-chem (Lamarque et al., 2012; Tilmes et al., 2015) is the atmospheric component of the National Center for Atmospheric Research (NCAR) Community Earth System Model (CESM). In this work, CAM-chem (from the publicly released CESM2.1.1) is nudged (Tilmes et al., 2015) to NASA MERRA2 meteorology fields with a horizontal resolution of 0.9° latitude × 1.25° longitude and 56 levels (surface to 3 hPa). The chemistry scheme is based on that in the CESM2.1.1, which includes a detailed representation of tropospheric and stratospheric chemistry (same as in Tilmes et al., 2015). In addition, the air-sea exchange module (see section 2.1 for details) and the reactive halogen chemistry (Fernandez et al., 2014; Navarro et al., 2015; Saiz-Lopez et al., 2012), are implemented on top of the publicly released CESM2.1.1. Anthropogenic emissions are from the Coupled Model Intercomparison Project Phase 6 (Hoesly et al., 2018), and emissions from vegetation are calculated using the Model of Emissions of Gases and Aerosols from Nature (Guenther et al., 2012). The model is spun-up for 5 years (2000–2004), and the modeled results in 2005–2015 are compared to ground-based observations around the globe. In addition, model simulations from another global chemical transport model, TOMCAT (Chipperfield, 2006), using a number of other oceanic VSLS emission inventories (Liang et al., 2010; Ordóñez et al., 2012; Warwick et al., 2006; Ziska et al., 2013), are included in this work for comparison. The TOMCAT simulations (see section 4 for details) are from Hossaini et al. (2013).

2.1. Online Air-Sea Interface for Soluble Species

The oceanic fluxes of $CHBr_3$ and CH_2Br_2 are calculated using the Online Air-Sea Interface for Soluble Species (OASISS) developed for CESM2 CAM-chem (Wang, Hornbrook, et al., 2019), which is based on the widely used two-layer model framework (Johnson, 2010; Liss & Slater, 1974). In brief, the air-sea exchange is described by the air-side and water-side transfer velocities (k_{air} and k_{water}). k_{air} is based on the NOAA COARE algorithm (Jeffery et al., 2010), with the addition of the still air diffusive flux adjustment
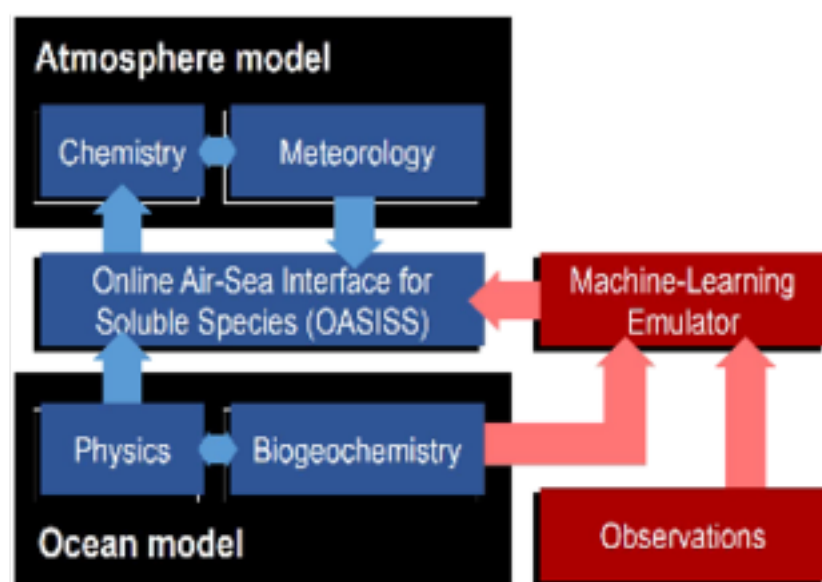
Figure 1. Schematic diagram of the new bottom-up oceanic emission inventory developed for CESM CAM-chem, with the online air-sea interface for soluble species (OASISS) and the data-oriented machine-learning emulator. CESM = Community Earth System Model; CAM-chem = Community Atmospheric Model with chemistry; OASISS = Online Air-Sea Interface for Soluble Species.

(Mackay & Yeun, 1983). k_{water} is based on Nightingale et al. (2000). This OASISS model framework (Figure 1) is fully coupled with CAM-chem, calculating the bidirectional fluxes of trace gases considering the local physical state of the ocean (sea surface temperature, salinity, and waves/bubbles) and the atmosphere (temperature, pressure, and wind). The surface seawater concentrations of $CHBr_3$ and CH_2Br_2 (used to drive the air-sea exchange) are predicted by a machine-learning algorithm based on observed seawater concentrations and other ancillary parameters (section 2.2).

2.2. Decision Trees and Random Forest

This machine-learning emulator utilizes a random forest regression algorithm provided in the Python Scikit-learn package (Pedregosa et al., 2011). Supporting information Figure S1 gives a simple illustration of the principle of the random forest. The training data set (section 2.3) consists of a number of data records; each data record contains a dependent variable (in this case, either $CHBr_3$ or CH_3Br_2) and a number of ancillary variables. A random forest is the ensemble of a number of decision trees. A decision tree consists of a "root node," a number of "split nodes," and finally a number of "leaf nodes." At each node (root/split/leaf), the parent data set is randomly split into two subsets. In this work, the mean square error (MSE) is used as the error metric to evaluate the split: Each possible split results in a reduction of MSE, and the best split is made in a way that the reduction of MSE is maximized. The splitting process is continued until a minimum number of data records are left in a node, or a maximum depth of a tree is reached, and the final nodes are referred to as leaf nodes. The mean of all values of the dependent variable in a leaf node is the prediction of this particular tree following a particular path, unless a leaf node contains only one data record, in which case the value of the dependent variable is the prediction of this tree. The combination of the ancillary variables in a data record determines how it will "travel through" a decision tree. The random forest prediction for a particular data record is the mean value of the predictions made by all decision trees. In order to avoid overfitting, the random forest algorithm in this work used "bootstrap aggregation" or "bagging" approach (Pedregosa et al., 2011), in which the training data are randomly sampled with replacement from all data labeled as training, thus increasing the randomness and diversity (Breiman, 2001). To further avoid overfitting, 20% of the entire data set is randomly sampled and used for validation (not involved in the training).

2.3. Surface Seawater Concentrations Predicted by Machine-Learning Algorithm

In this work, the monthly global surface seawater concentrations of $CHBr_3$ and CH_2Br_2 are predicted by a trained machine-learning emulator which mimics the ocean biogeochemical control on the synthesis of these compounds in the seawater. The machine-learning training data set consists of (i) surface seawater concentrations of $CHBr_3$ and CH_2Br_2 observations in the HalOcAt data set, which spans from 1987 to 2011 (Ziska et al., 2013); (ii) ocean depth (indicating the open ocean and coastal regions; ~0.08° horizontal resolution); (iii) ancillary ocean physical and biogeochemical variables (monthly means; Table S1). The ancillary ocean biogeochemical variables (see next section for details) are from the fully coupled control run in the NCAR CESM Large Ensemble data set (Kay et al., 2014). All ancillary biogeochemical variables are sampled based on the time (year-month) and location of each HalOcAt observation. Due to the sparse availability of the surface seawater observations (less than 5,000, Figures 2 and 4), our current approach is unable to resolve the short-term variations (e.g., <1 month). Stemmler et al. (2015) indicated that the $CHBr_3$ residence time in the ocean is 200–300 days; therefore, our choice of temporal resolution for the machine learning (1 month) may not cause considerable structural biases. If more surface seawater observations are available in the future, our approach can certainly accommodate the need for higher temporal resolution (e.g., weekly). Once trained, the machine-learning algorithm uses the same modeled ocean biogeochemical variables for prediction at all model grid points. For testing and evaluation purposes, the machine-learning emulator is currently running offline in this work. As a component of the Earth system model (CESM), CAM-chem is capable of operating in coupled configurations, for example, with the active sea ice model, ocean model, and marine ecosystem model through a coupler. In such coupled configurations, the trained machine-learning emulator may directly retrieve the variables from the marine ecosystem model via the coupler, providing "in-house" predictability, which is essential for the future and past climate projections.

Recent studies have used satellite products and gridded observations to investigate the oceanic influence of organohalogens (Ordóñez et al., 2012; Sherwen et al., 2019), providing remarkable insights into the ocean biogeochemistry control on the organohalogen sources and chemistry, especially in the tropical regions. There are a few limitations to the applications using satellite or gridded observations: (i) As discussed earlier, the satellite and gridded observations have limited temporal resolution (e.g., monthly). (ii) The satellite products may have (relatively) limited coverage in the high-latitude regions, for example, the Southern Ocean in boreal summer, due to the satellite orbital patterns and the impacts of cloudy and highly reflective surface (e.g., sea ice) conditions. This is particularly relevant for the Br-VSLS, since current oceanic Br-VSLS inventories show large discrepancies in the Southern Ocean (Hossaini et al., 2013). (iii) The future/past climate projections may not be easily justified if using the satellite and gridded observations. (iv) The satellite and gridded observations may not cover the full complexity of the physical and biogeochemical processes affecting the natural synthesis of Br-VSLS in the ocean. For example, the chlorophyll a product from the SeaWiFS satellite observations (Hu et al., 2012) do not separate the different chlorophyll content from different phytoplankton groups nor provide insights into the growth/nutrient limitations. In this work, we propose an alternative approach and use the upper ocean biogeochemistry variables from the CESM Parallel Ocean Program (POP) and the Biogeochemical Elemental Cycling (BEC) model (Moore et al., 2013) for the machine-learning emulator. We consider this alternative approach has the following advantages: (i) As discussed earlier, the machine-learning emulator based on the CESM POP/BEC may accommodate the need for higher temporal resolution (once more seawater observations of Br-VSLS become available). (ii) The current configuration of the machine-learning emulator provides improved spatial coverage in the high-latitude regions compared to the satellite products, especially the Southern Ocean in boreal summer. The Southern Ocean represents a large fraction of the global ocean and a diversity of biogeochemical regimes and is particularly sensitive to climate change (Stephens et al., 2018). The Southern Ocean also features a variety of unique conditions affecting the air-sea exchange processes, such as the waves and bubbles at the air-sea interface driven by the strong surface winds, and the coupled ocean-sea ice-atmosphere interactions (Fernandez et al., 2019). Given the long tropospheric lifetimes (a few weeks or longer) of Br-VSLS (Hossaini et al., 2016), the Southern Ocean-emitted $CHBr_3$ and CH_2Br_2 might contribute to the tropospheric Br-VSLS burden and affect their seasonal variations on the hemispheric scale (Abrahamsson et al., 2018). (iii) A variety of physical and biogeochemical variables are incorporated into the machine learning, providing the complexity that is needed to capture the spatial and temporal variations of Br-VSLS in the ocean. For
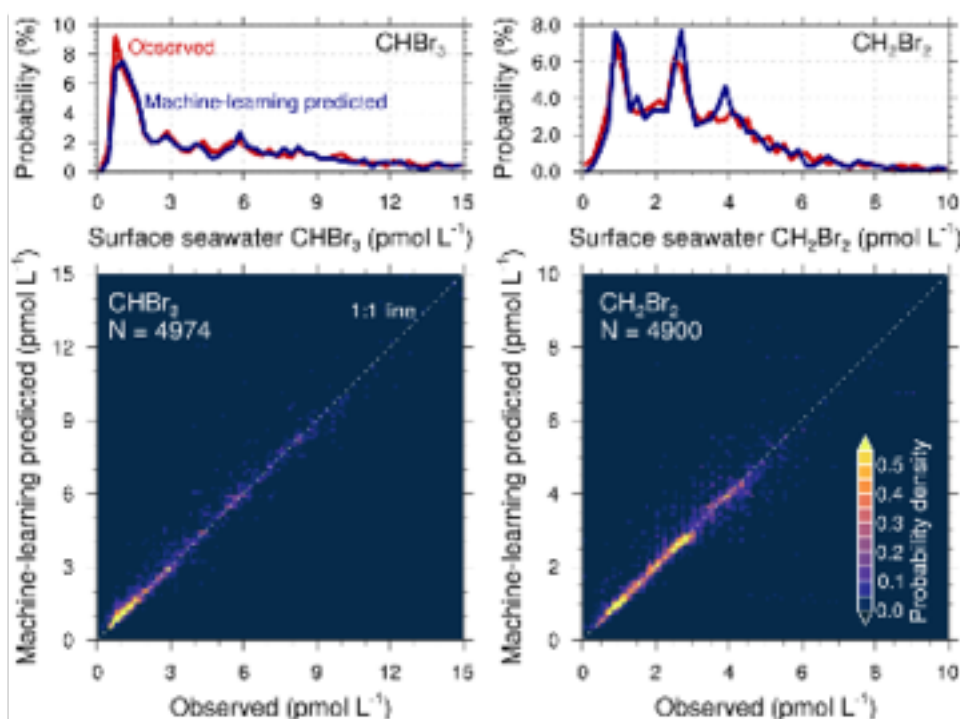
Figure 2. (top) One-dimensional and (bottom) two-dimensional probability distributions of the observed (HalOcAt) and machine-learning predicted surface seawater concentrations of (left) $CHBr_3$ and (right) CH_2Br_2. Data points in both the open ocean and the coastal regions (ocean depth shallower than 200 m) are included in this analysis. HalOcAt data set = Halocarbons in the Ocean and Atmosphere.

example, different types of variables are used in the machine-learning emulator, including nutrients (iron, phosphate, nitrate, etc.), different chlorophyll content (diatom, diazotroph, and small phytoplankton), salinity, oxygen (dissolved oxygen and surface oxygen flux), and radiation. (iv) These CESM POP/BEC variables are used in both training and the prediction, providing the inherent consistency for the CESM framework that is important for future/past climate projections. The CESM POP/BEC models have been extensively studied and proven to capture the large-scale features as revealed in the observations related to the nutrients and productivity (Danabasoglu et al., 2011; Doney et al., 2009; Krumhardt et al., 2017; Long et al., 2016; Moore et al., 2013; Moore & Braucher, 2008). More details are provided in the supporting information. We later show that this training-prediction workflow yields overall quite reasonable agreement with observations including those in the Southern Ocean.

The training process of the random forest algorithm is as follows: 80% of all valid data (as flagged by the HalOcAt data set), as well as the ancillary variables, are randomly sampled as the *training data set*, and the remaining is used as *testing data set*. The training data set is used to construct the random forest, and the testing data set is used to evaluate the performance of the training (i.e., Pearson coefficient better than ~0.8 and the linear fit slope between ~0.9 and ~1.1). The aforementioned bootstrap approach (Pedregosa et al., 2011) was used for sampling to avoid overfitting. Some of the hyperparameters of the algorithm may be tuned to improve the performance. In this work, the random forest hyperparameterization configuration is as follows: number of decision trees: 2,000; criterion (error metric to evaluate the split): MSE; the maximum depth of the tree is not restricted (so that the nodes will be split until all leaf nodes contain only one sample). The algorithm may return a predictor importance rating, which is the percentage of time that a predictor (independent/ancillary variable) is used to split the tree. The ancillary variables with low importance rating may be eliminated from further training and predictions. The importance rating is given in Table S1. Lastly, the monthly fields (multiyear means between 2005 and 2015) of these variables are used in the trained algorithm to predict the surface seawater concentrations of Br-VSLS as the inputs of the air-sea exchange module.
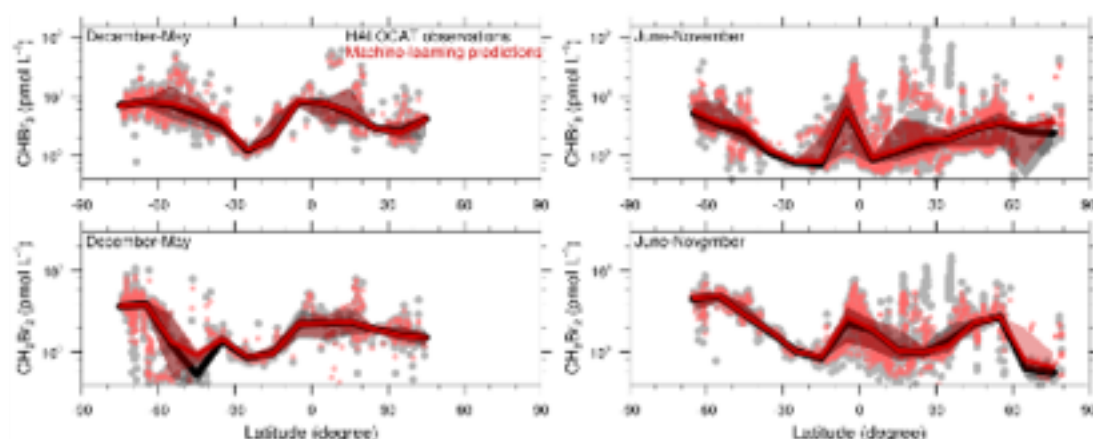
Figure 3. Latitude dependencies of the predicted surface seawater $CHBr_3$ in the open oceans compared to the HalOcAt data set. Gray and pink dots represent the HalOcAt observations and the machine-learning predictions, respectively. Black and red lines indicate the median profiles of the HalOcAt observations and the machine-learning predictions, respectively. The 25–75% percentiles of the HalOcAt observations and the machine-learning predictions are shown with gray and pink shadings, respectively. The medians and the quartiles of the observations are reasonably well captured by the machine learning. HalOcAt data set = Halocarbons in the Ocean and Atmosphere.

2.4. Atmospheric Observations Used for Model Comparison

In this work, surface $CHBr_3$ and CH_2Br_2 measurements from 14 stations around the globe (NOAA/ESRL global monitoring network Montzka et al., 2011) are used to evaluate the modeled seasonal and spatial variations of the $CHBr_3$ and CH_2Br_2. In brief, whole air samples were collected approximately weekly into paired steel or glass flasks, prior to being analyzed at NOAA/ESRL in Boulder, Colorado (Montzka et al., 2011). In addition, the airborne $CHBr_3$ and CH_2Br_2 measurements from the NASA ATom field campaign (ATom-1: July–August 2016 and ATom-2: January–February 2017). During ATom-1 and ATom-2, the heavily instrumented NASA DC-8 aircraft transected the lengths of the Pacific and Atlantic Oceans during two seasons (Figure S2), constantly profiling from the surface to the upper troposphere/lower stratosphere. During ATom-1 and ATom-2, $CHBr_3$ and CH_2Br_2 were simultaneously measured by three techniques: the NCAR Trace Organic Gas Analyzer (TOGA; Apel et al., 2003, 2015), the University of California, Irvine (UCI) Whole Air Sampler (WAS; Blake et al., 2003), and the NOAA Programmable Flask Package (PFP) whole air sampler. NCAR TOGA is a fast online gas chromatograph/mass spectrometer system, with an analyzing cycle of ~2 min. The UCI WAS and NOAA PFP are both offline whole air samples, and the samples were analyzed afterward in the respective laboratories. The model outputs were sampled along the flight track for comparison.

3. Predicted Surface Seawater Concentrations

In this section, the machine-learning predicted surface seawater concentrations of $CHBr_3$ and CH_2Br_2 are compared to the HalOcAt observation data set. The statistical distributions of the observed $CHBr_3$ and CH_2Br_2 are both well captured by the machine-learning emulator (Figure 2). Overall, the observed latitude dependencies of $CHBr_3$ and CH_2Br_2 are both well captured by the machine-learning algorithm (Figure 3).

The mean absolute percentage error ($MAPE \equiv \frac{100}{n}\sum_{i=1}^{n}|\frac{MOD_i - OBS_i}{OBS_i}|$, where OBS_i and MOD_i represent the individual observations and modeled results, respectively, and n is the number of samples) in the open ocean are 12% and 21% for $CHBr_3$ and CH_2Br_2, respectively. In the coastal regions (ocean depth $\leq$ 200 m), the mean absolute percentage error (MAPE) for $CHBr_3$ and CH_2Br_2 are 37% and 46%, respectively. The root-mean-

square errors ($RMSE \equiv \sqrt{\frac{\sum_{i=1}^{N}(MOD_i - OBS_i)^2}{N}}$) in the open ocean are 4.9 and 0.8 pmol/L for $CHBr_3$ and CH_2Br_2, respectively; in the coastal regions the RMSE are 33 and 3.1 pmol/L for $CHBr_3$ and CH_2Br_2,
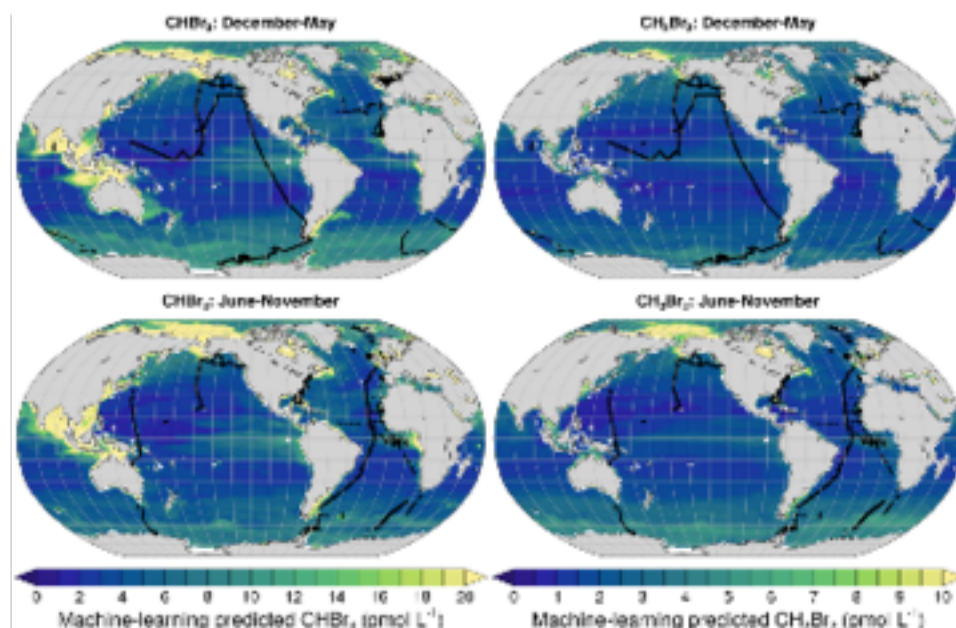
Figure 4. December–May and June–November means of surface seawater concentrations of $CHBr_3$ and CH_2Br_2. Black dots represent the locations where observations (HalOcAt) were available. HalOcAt data set = Halocarbons in the Ocean and Atmosphere.

respectively. It remains challenging to fully resolve the coastal features in current global chemistry-climate models. For instance, in a state-of-the-art ocean VSLS model, a fixed concentration was prescribed in waters shallower than 200 m, and the coastal sources (macroalgae with tide-dependent bromoform production, released from benthic algae and seagrass) are not explicitly represented (Stemmler et al., 2015). Similarly, Ordóñez et al. applied a scaling factor to account for the enhanced VSLS emissions in the coastal regions to reproduce the observations at the coastal sites (Ordóñez et al., 2012). We will show later that coastal regions contribute to a substantial fraction to the global total Br-VSLS emissions.

Figure 4 shows the December–May and June–November means of the surface seawater predicted $CHBr_3$ and CH_2Br_2. In general, the surface seawater $CHBr_3$ and CH_2Br_2 are enhanced in coastal regions, as well as the productive upwelling zones in the tropics, where the nutrient-rich surface waters support the phytoplankton growth or the enhanced vertical mixing with the maximum production of Br-VSLS below the (shallow) mixed layer as revealed from previous studies (Liu et al., 2013). Similarly, elevated surface seawater $CHBr_3$ and CH_2Br_2 are also found in the Southern Ocean in December–May, which may be related to the upwelling in the Antarctic divergence zone. In contrast, lower surface seawater Br-VSLS is found in the ocean gyres due possibly to the limited nutrients. Overall, the observationally trained machine-learning algorithm captures the large-scale features and the seasonality of the observed surface seawater Br-VSLS. Advanced chemistry-climate models with higher horizontal resolution or regional refinement capability may help to improve the model bias in the coastal regions. In the meantime, long-term (multiseasonal) observations will likely improve the performance of the machine-learning emulator in the coastal regions.

4. Surface Atmospheric Concentrations Compared to Surface Observations

The machine-learning predicted surface seawater concentrations of Br-VSLS are used to drive the air-sea exchange of Br-VSLS (OASISS), and the CAM-chem predicted monthly mean surface atmospheric mixing ratios are compared to the surface observations obtained at 14 stations from the NOAA/ESRL global network. This model configuration is hereafter denoted as *CAM-chem OASISS/Machine-learning*. The CAM-chem simulation with prescribed oceanic VSLS emissions scaled to satellite chlorophyll (Ordóñez et al., 2012) is also shown, which is denoted as *CAM-chem Ordóñez-2012*. The model configuration of

CAM-chem Ordóñez-2012 has been extensively used for VSLS studies (Fernandez et al., 2014; Navarro et al., 2015; Saiz-Lopez et al., 2012) and is otherwise identical to CAM-chem OASISS/Machine-learning. Meanwhile, previous TOMCAT model predictions from Hossaini et al. (2013) using a full set of oceanic emission inventories (Liang et al., 2010; Ordóñez et al., 2012; Warwick et al., 2006; Ziska et al., 2013), here-after denoted as *TOMCAT Ordóñez-2012, TOMCAT Liang-2010, TOMCAT Ziska-2013,* and *TOMCAT Warwick-2011,* are compared to the CAM-chem simulations. Note that these TOMCAT simulations are for a different period (1997–2011) than the CAM-chem simulations and the NOAA observations presented in this work (2005–2015). The interannual variability of ocean emissions of VSLS is usually assumed to be small (Hossaini et al., 2013, 2016). The modeled and observed monthly mean surface $CHBr_3$ and CH_2Br_2 mixing ratios were determined for selected locations (Figures 5 and 6):

In the Arctic, CAM-chem OASISS/Machine-learning captures the seasonal variations of observed $CHBr_3$ in Alert (Canada) and Summit (Greenland) but not in Barrow (Alaska, USA). In Alert and Barrow, the observed surface $CHBr_3$ mixing ratios were not well captured by any of the models discussed in this work, especially in winter, possibly due to the sea ice emissions not considered in the models. Performance-wise, in Alert (Canada), TOMCAT Liang-2010 yields the best agreement with observations, followed by CAM-chem OASISS/Machine-learning. In Summit (Greenland), CAM-chem OASISS/Machine-learning yields the best agreement.

In the coastal midlatitudes, the observed seasonal variations in Mace Head (Ireland) and Trinidad Head (USA) are not captured by any of the previous models, possibly due to the unique local (coastal) conditions that deviate from the global background (Yokouchi et al., 2017). By contrast, CAM-chem OASISS/Machine-learning yields the best agreement with observations in these coastal midlatitude locations.

In the continental midlatitudes, the observed seasonal variations in Harvard Forest (USA), Niwot Ridge (USA), and Wisconsin (USA) are captured by CAM-chem OASISS/Machine-learning, as well as CAM-chem Ordóñez-2012 and TOMCAT Ordóñez-2012. TOMCAT Liang-2010 also produces quite reasonable agreement (within the variations of the observations) in Harvard forest (USA) and Niwot Ridge (USA) but overestimates surface $CHBr_3$ measurements in Wisconsin (USA).

In the tropics, TOMCAT Ziska-2013 and TOMCAT Liang-2010 both yield overall very good agreement in Mauna Loa (USA), Cape Kumukahi (USA), and Cape Matatula (American Samoa), while CAM-chem Ordóñez-2012 and TOMCAT Ordóñez-2012 tend to overestimate the surface $CHBr_3$ observations in these tropical locations. The surface $CHBr_3$ modeled by CAM-chem OASISS/Machine-learning in this work is also in good agreement with observations in Mauna Loa (USA) but is lower (by about half a part per thousand) than that observed in Cape Kumukahi (USA). In Cape Matatula (American Samoa), however, CAM-chem OASISS/Machine-learning predicts a pronounced annual maximum in May–July, while observed surface $CHBr_3$ in the same period is only slightly enhanced.

In the Southern Hemisphere, the surface observations of $CHBr_3$ show interesting seasonal variations. In Cape Grim (Australia), slightly enhanced surface $CHBr_3$ are observed in December–March, while all models predict an annual maximum in April–August. At the Palmer Station (Antarctica), slightly enhanced surface $CHBr_3$ are found from January to June (with large variations in January–February), and CAM-chem OASISS/Machine-learning predicts an annual maximum in May–July. All other models predict a consistent June–August maximum at the Palmer Station. At the South Pole, however, the observed surface $CHBr_3$ peaks in July–October, which is reasonably well captured by CAM-chem Ordóñez-2012, TOMCAT Ordóñez-2012, and TOMCAT Liang-2010. CAM-chem OASISS/Machine-learning predicts an annual maximum April–September at the South Pole, 2–3 months earlier than observations.

It should be noted that the TOMCAT simulation period (1997–2011) from Hossaini et al. (2013) is not the same as the studied period in this work (2005–2015), although the observed $CHBr_3$ at these stations in 2005–2015 is quite similar to that in 1997–2011 (Hossaini et al., 2013). The modeled atmospheric $CHBr_3$ may be affected by photolysis, hydroxyl radicals, and transport, which may well differ between CAM-chem in this work and TOMCAT. The inter-model discrepancy may be partially glimpsed by examining the difference between CAM-chem Ordóñez-2012 and TOMCAT Ordóñez-2012: The difference is quite small, except in Trinidad Head (USA) and Harvard Forest (USA) where the discrepancy is close to 1 ppt (Figure 5).
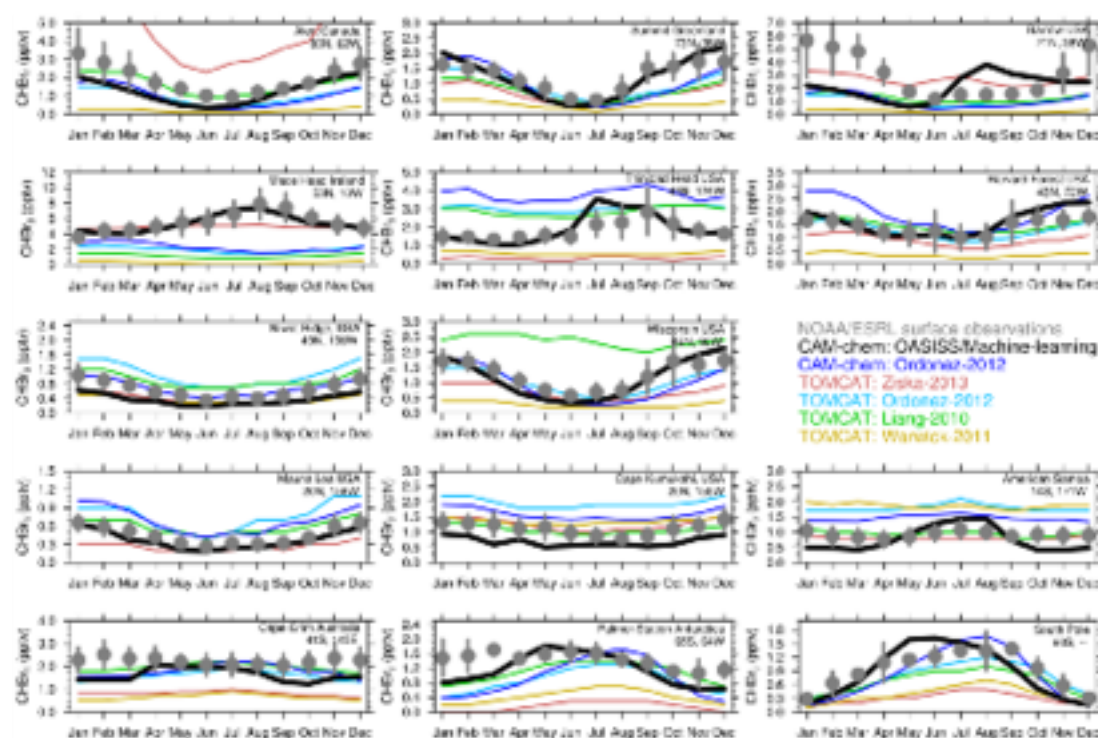
Figure 5. Comparison of observed monthly mean CHBr$_3$ mixing ratio at 14 NOAA/ESRL ground stations with output from this work, as well as previous modeling results using different oceanic emission inventories from Hossaini et al. (2013). NOAA/ESRL = National Oceanic and Atmospheric Administration/Earth System Research Laboratory; CAM-chem = Community Atmospheric Model with chemistry.

Surface CH$_2$Br$_2$ mixing ratios, however, do not show a pronounced seasonal and spatial variations as CHBr$_3$, due to its longer tropospheric lifetime (Figure 6). Figure 6 shows the modeled and observed monthly mean surface CH$_2$Br$_2$ mixing ratios in different locations. In general, CAM-chem OASISS/Machine-learning captures the observed levels and seasonal variations of CH$_2$Br$_2$ in 12 out of 14 sites; in Barrow (USA) and Trinidad Head (USA) CAM-chem OASISS/Machine-learning predicts a pronounced summer maximum (similar is seen for CHBr$_3$) which is not seen in observations. The performance of other models varies with locations. Notably, in Summit (Greenland), Alert (Canada), Niwot Ridge (USA), Wisconsin (USA), Mauna Loa (USA), and Cape Kumukahi (USA), TOMCAT simulations show little seasonal variations, while CAM-chem simulations capture the summer minimum of observed CH$_2$Br$_2$.

The overall performance of different models can be evaluated using the Taylor diagram (Figure 7). In this analysis, monthly mean surface observations and the modeling results for all locations are aggregated. The observed annual mean CHBr$_3$ is well captured by the top-down inventories such as Ordóñez et al. (2012) and Liang et al. (2010), as model simulations using these top-down inventories cluster around the 1:1 reference line in the Taylor diagram (Figure 7). But the seasonal variations may not be fully resolved in these top-down inventories (as discussed previously), which is indicated by moderate Pearson coefficient (0.2–0.4) in the Taylor diagram. Ziska et al. (2013), a bottom-up inventory, better captures the seasonal variations compared to the top-down inventories, as indicated by an improved Pearson coefficient (~0.7). The newly developed bottom-up inventory in this work (CAM-chem OASISS/Machine-learning) tends to underestimate the annual mean CHBr$_3$ (by ~23% globally), but the seasonal variations of CHBr$_3$ are very well captured (Pearson coefficient: ~0.88). As for CH$_2$Br$_2$, all models discussed in this work are capable of reproducing the observed mean CH$_2$Br$_2$ levels within ~30%. CAM-chem model tends to better resolve the observed seasonal variations of CH$_2$Br$_2$ (Pearson coefficient 0.5–0.7) than TOMCAT (Pearson coefficient < 0.2).
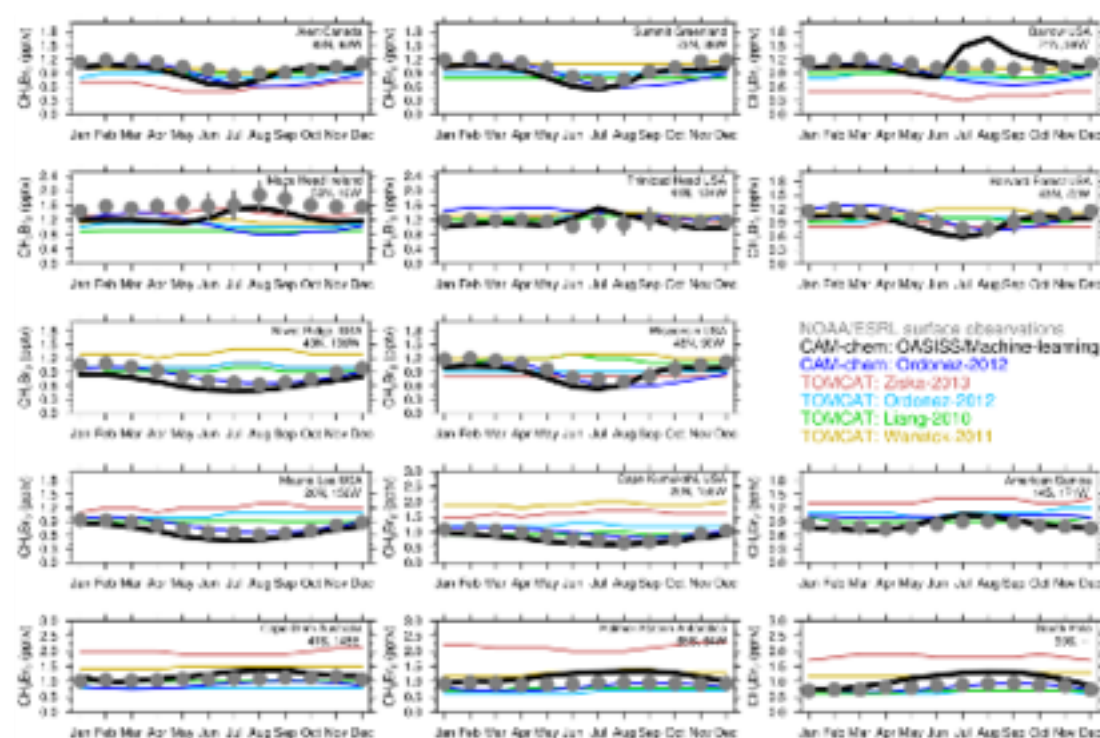
Figure 6. Comparison of observed monthly mean CH_2Br_2 mixing ratio at 14 NOAA/ESRL ground stations with output from this work, as well as previous modeling results using different oceanic emission inventories from Hossaini et al. (2013). NOAA/ESRL = National Oceanic and Atmospheric Administration/Earth System Research Laboratory; CAM-chem = Community Atmospheric Model with chemistry.

5. Comparison of the Global Oceanic Emissions

Figure 8 compares the newly developed bottom-up oceanic emission inventory of $CHBr_3$ (with online air-sea exchange and machine-learning predicted surface seawater concentrations) to a successful top-down emission inventory from Ordóñez et al. (2012). The Ordóñez-2012 inventory shows little seasonal variations, partially leading to the lower Pearson coefficient in the Taylor diagram (Figure 7). A recent study highlighted the importance of seasonally resolved oceanic emissions for the stratospheric injection of Br-VSLS (Fiehn et al., 2018). The new emission inventory (Figure 8) shows clear seasonal variations, leading to improved seasonal variations globally (Figure 7). The Ordóñez-2012 inventory also has fixed values in the subtropical, midlatitude, and high-latitude oceans. We noted that the new bottom-up oceanic emission inventory of $CHBr_3$ shows high sea-to-air flux over the Indian Ocean, South China Sea, and Java Sea (2–3 $\times 10^7$ molecules·cm^{-2}·s^{-1}), which is comparable to the oceanic $CHBr_3$ flux reported over the South China Sea (2.5 $\pm$ 2.9 $\times 10^7$ molecules·cm^{-2}·s^{-1}; Fuhlbrügge et al., 2016), but toward the upper end of that over the Indian Ocean (1.5 $\pm$ 2.0 $\times 10^7$ molecules·cm^{-2}·s^{-1}; Fiehn et al., 2017). The Ordóñez-2012 inventory, however, shows approximately 1.5–2.0 $\times 10^7$ molecules·cm^{-2}·s^{-1}, in this region, comparable to that reported in Fiehn et al. (2017), but is toward the lower end of that in Fuhlbrügge et al. (2016). The reason that the new inventory in this work shows high $CHBr_3$ flux is that one cruise study in the Bay of Bengal (6.87–11.96°N, 88.07–88.26°E) reported high $CHBr_3$ levels in the surface noncoastal waters (46 $\pm$ 10 pmol/L; Yamamoto et al., 2001); as a result the machine-learning algorithm predicted high surface seawater concentrations in that region as well (Figure 4). This is the only surface seawater measurement available in the HalOcAt data set in that region (Figure 4). Figure 9 shows the same comparison for CH_2Br_2. Similarly, the Ordóñez-2012 inventory shows little seasonal variations and has no oceanic fluxes for CH_2Br_2 in the majority of the Southern Hemisphere. The new inventory in this work also shows high oceanic CH_2Br_2 fluxes over the Indian Ocean and South China Sea (0.6–0.9 $\times 10^7$ molecules·cm^{-2}·s^{-1}), very close to that reported over
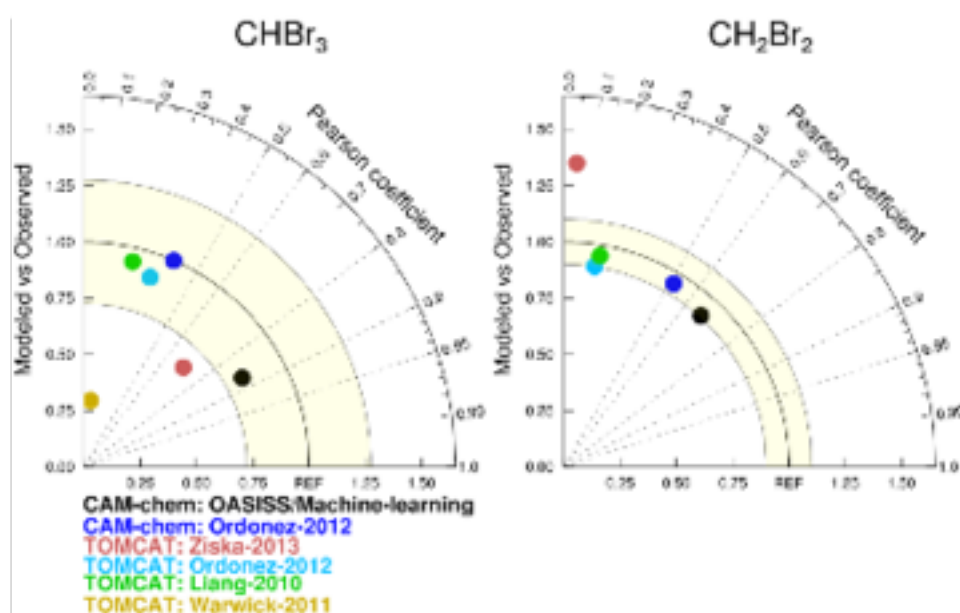
Figure 7. Taylor diagrams showing the modeled surface atmospheric $CHBr_3$ and CH_2Br_2 from different models compared to the surface observations, across all 14 stations. The yellow shading indicates the standard deviation of all observations, and the Pearson coefficient represents the combined temporal/spatial variation. CAM-chem = Community Atmospheric Model with chemistry.

the South China Sea ($0.7 \pm 0.6 \times 10^7$ molecules·cm^{-2}·s^{-1}; Fuhlbrügge et al., 2016), but is a factor of ~2 lower than that reported over the Indian Ocean ($1.6 \pm 3.3 \times 10^7$ molecules·cm^{-2}·s^{-1}; Fiehn et al., 2017). The Ordóñez-2012 inventory shows less than 0.5×10^7 molecules·cm^{-2}·s^{-1} in the entire region, lower than that reported in Fuhlbrügge et al. (2016) or Fiehn et al. (2017). More surface seawater observations in this region will certainly improve the performance of the machine-learning emulator, as well as the bottom-up oceanic emission inventory for $CHBr_3$ and CH_2Br_2.

The global oceanic emissions of $CHBr_3$ and CH_2Br_2 estimated in this work and previous studies are summarized in Table 1. The global marine emission of $CHBr_3$ estimated in this work is 385 Gg Br per year, toward the lower end of the range predicted by the top-down estimates (381–840 Gg Br per year) but higher than other recently developed bottom-up inventories (70–200 Gg Br per year; Table 1). The global oceanic emission of CH_2Br_2 is estimated to be 54 Gg Br per year, which is lower than the top-down estimates (57–280 Gg Br per year) but comparable to the recent bottom-up inventories (62–78 Gg Br per year). In particular, Stemmler et al. (2015) calculated the air-sea fluxes of $CHBr_3$ using the atmospheric concentration of $CHBr_3$ (observations mapped to the global scale) from Ziska et al. (2013) and surface seawater concentration of $CHBr_3$ predicted by a state-of-the-art ocean biogeochemistry model, suggesting that the majority of the Southern Ocean is a net sink of $CHBr_3$ in boreal summer. In this work, the majority of the Southern Ocean is found to be a net sink of $CHBr_3$ in boreal summer, consistent with Ziska et al. (2013) and Stemmler et al. (2015), but the ocean uptake of $CHBr_3$ simulated in this work (CAM-chem OASISS/Machine-learning) is weaker than that predicted in Ziska et al. (2013) and Stemmler et al. (2015). The machine-learning emulator is capable of capturing the enhanced surface $CHBr_3$ in the marginal sea ice regions in the Southern Ocean (with limited radiation but enriched in nutrients), as indicated by the good agreement in the Southern Ocean (Figure 3). Yet, the observed surface atmospheric $CHBr_3$ (and CH_2Br_2) abundances and seasonality in the Southern Hemisphere locations remain not fully explained by models (Figure 5 and 6). Recent studies have reported episodic $CHBr_3$ emissions from the sea ice in Antarctica (Abrahamsson et al., 2018), much higher than the long-term monthly mean $CHBr_3$ observed at the Palmer Station or the South Pole in the same season. The sea ice biogeochemical cycles remain poorly understood, and the VSLS emissions from sea ice are not included in our model configuration. Given the relatively long atmospheric lifetimes of these VSLS in the high latitude regions (mostly during austral winter), the Southern Ocean and sea ice emissions of VSLS may
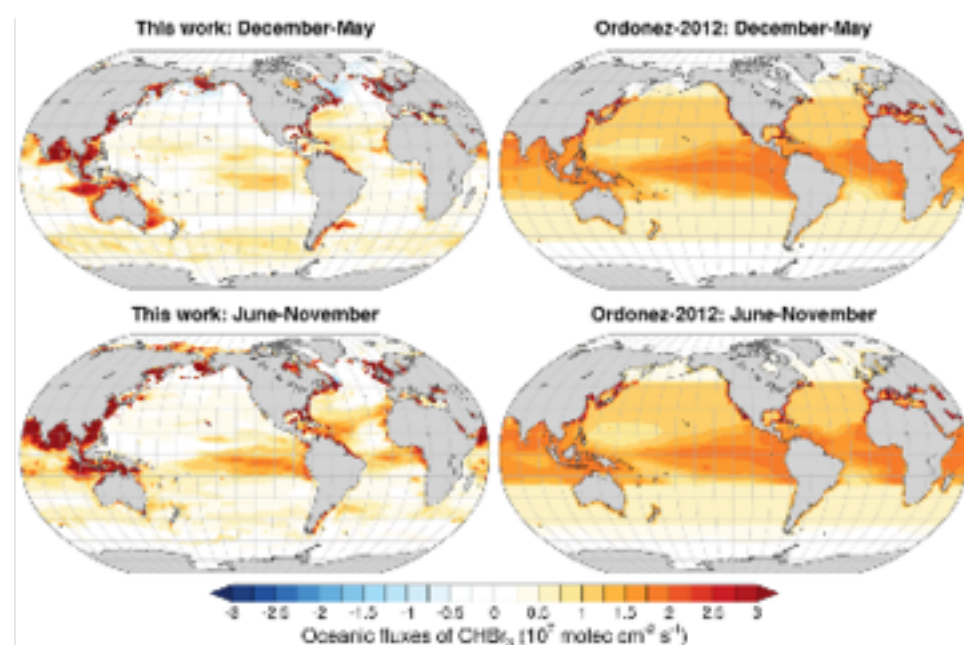
Figure 8. December–May and June–November averages of the oceanic emission fluxes of $CHBr_3$ predicted in this work (left) compared to a top-down inventory from Ordóñez et al. (2012; right). Warmer colors indicate upward fluxes (i.e., the ocean is net emitting), while colder colors indicate downward fluxes (i.e., the ocean is a net sink).

have broad impacts on the hemispheric scale (Abrahamsson et al., 2018). Future Br-VSLS studies that target the Southern Ocean, in particular, during boreal summer months could help resolve this discrepancy. We find that coastal regions (200 m or shallower) contribute 26% and 16% to the global marine emissions of $CHBr_3$ and CH_2Br_2, respectively. The coastal contributions derived in this work may be subject to substantial uncertainties (e.g., MAPE for the surface seawater $CHBr_3$ and CH_2Br_2 are 37% and 46%, respectively; the uncertainties associated with the air-sea exchange processes are commonly quoted as a factor of 2 or so; Johnson, 2010). Our estimates of the coastal contributions in this work are consistent with a recent study focusing on the western Pacific (Butler et al., 2018) but are lower than that in Stemmler et al. (2015), which may be partially due to the different model configurations: The ocean emissions in our model (CAM-chem OASISS/Machine-learning) are fully coupled with atmospheric transport and chemistry, while Stemmler et al. (2015) calculated the sea-to-air fluxes offline (using objectively mapped air concentrations from Ziska et al., 2013).

6. Vertical Distributions and Model Comparison

In this section, the CAM-chem modeled vertical distributions of $CHBr_3$ and CH_2Br_2 are compared to the airborne measurements using NCAR TOGA, UCI WAS, and NOAA PFP during ATom-1 and ATom-2 (NOAA PFP $CHBr_3$ was not available for ATom-1). Flight tracks are given in Figure S2 (we focus on research flights over the oceans only). Two CAM-chem simulations with both OASISS/ML and the Ordóñez-2012 inventory are discussed, and the results are binned spatially (Northern Hemisphere, tropical, Southern Hemisphere, and Southern Ocean, over both the Pacific and the Atlantic; see Figure S2). In general, elevated $CHBr_3$ were reported in the marine boundary layer by all three techniques (Figure 10), implying ocean is a net source. However, the vertical distribution of $CHBr_3$ and CH_2Br_2 shows considerable spatial and seasonal variations. Interestingly, the $CHBr_3$ reported by these techniques sometimes shows quite significant discrepancies.

Northern Hemisphere during ATom-1. Both NCAR TOGA and UCI WAS reported ~0.7 ppt (median) $CHBr_3$ in the marine boundary layer over the Pacific, consistent with CAM-chem simulation with OASISS/ML
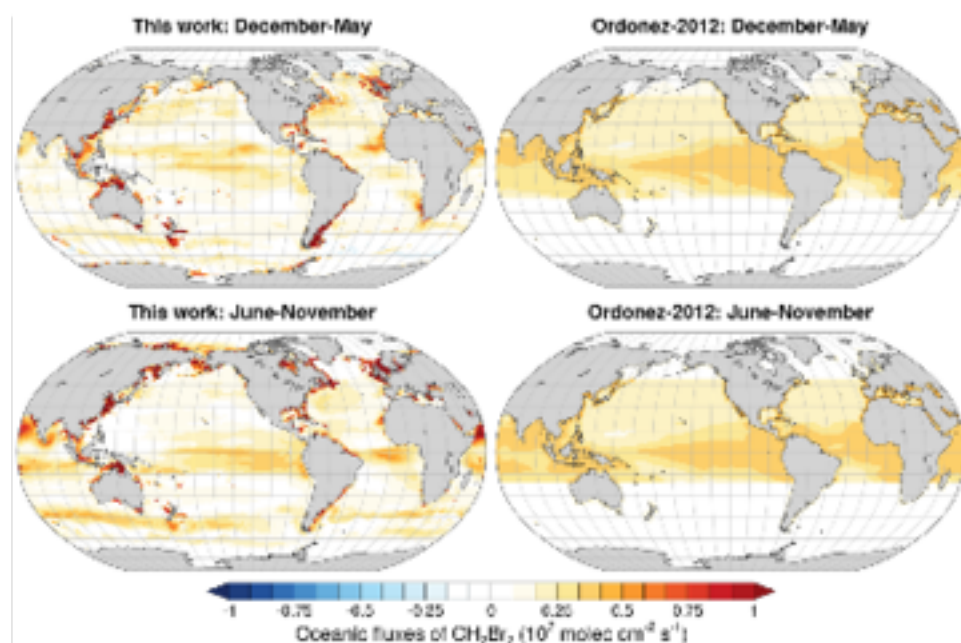
Figure 9. December–May and June–November averages of the oceanic emission fluxes of CH_2Br_2 predicted in this work (left) compared to that from Ordóñez et al. (2012; right). Warmer colors indicate upward fluxes (i.e., the ocean is net emitting), while colder colors indicate downward fluxes (i.e., the ocean is a net sink).

(Figure 10a). The simulation with Ordóñez-2012 tends to overestimate $CHBr_3$ in the Northern Hemisphere over the Pacific (Figure 10a). On the Atlantic side, measured $CHBr_3$ reached 1.2–1.4 ppt (median) in the marine boundary layer, which is well captured by the simulation using the simulation with Ordóñez-2012 (Figure 11b). The simulation using OASISS/ML is only slightly lower than observed in the marine boundary layer, still within the ranges of the observations (Figure 10b). Both simulations tend to underestimate $CHBr_3$ in the free troposphere in the Northern Hemisphere over both the Pacific and the Atlantic (Figures 10a and 10b), which may be attributed to deficiencies in representing the vertical uplift in the model.

Table 1
Global Total Marine Emissions of $CHBr_3$ and $CH2Br_2$ Estimated in This Work and Previous Studies

Global total marine emissions (Gg Br/year)	$CHBr_3$	CH_2Br_2	Note
This work	385	54	Bottom-up (I)
Quack and Wallace (2003)	822	—	Top-down (II)
Warwick et al. (2006)	381–571	104	Top-down (II)
Butler et al. (2007)	840	280	Top-down (II)
Liang et al. (2010)	425	57	Top-down (II)
Ordóñez et al. (2012)	507	62	Top-down (III)
Ziska et al. (2013)	120–200	62–78	Bottom-up (IV)
Lennartz et al. (2015)	226	58	Bottom-up (V)
Stemmler et al. (2015)	72	—	Bottom-up (VI)

Note. (I) = air-sea exchange is coupled (online) with atmospheric chemistry and dynamics, and the surface seawater concentrations (to drive the air-sea exchange) are predicted by machine learning; (II) = estimated based on global observations; (III) = marine emissions of very short lived substances scaled to satellite-chlorophyll. (IV) = oceanic fluxes calculated offline based on surface seawater concentrations and surface atmospheric concentrations, both fitted from observations (objective mapping); (V) = air-sea exchange is coupled (online) with atmospheric chemistry and dynamics, using seawater concentrations from Ziska et al. (2013; objective mapping); (VI) = oceanic fluxes calculated offline based on surface seawater concentrations modeled using an ocean biogeochemistry model and surface atmospheric concentrations from Ziska et al. (2013).
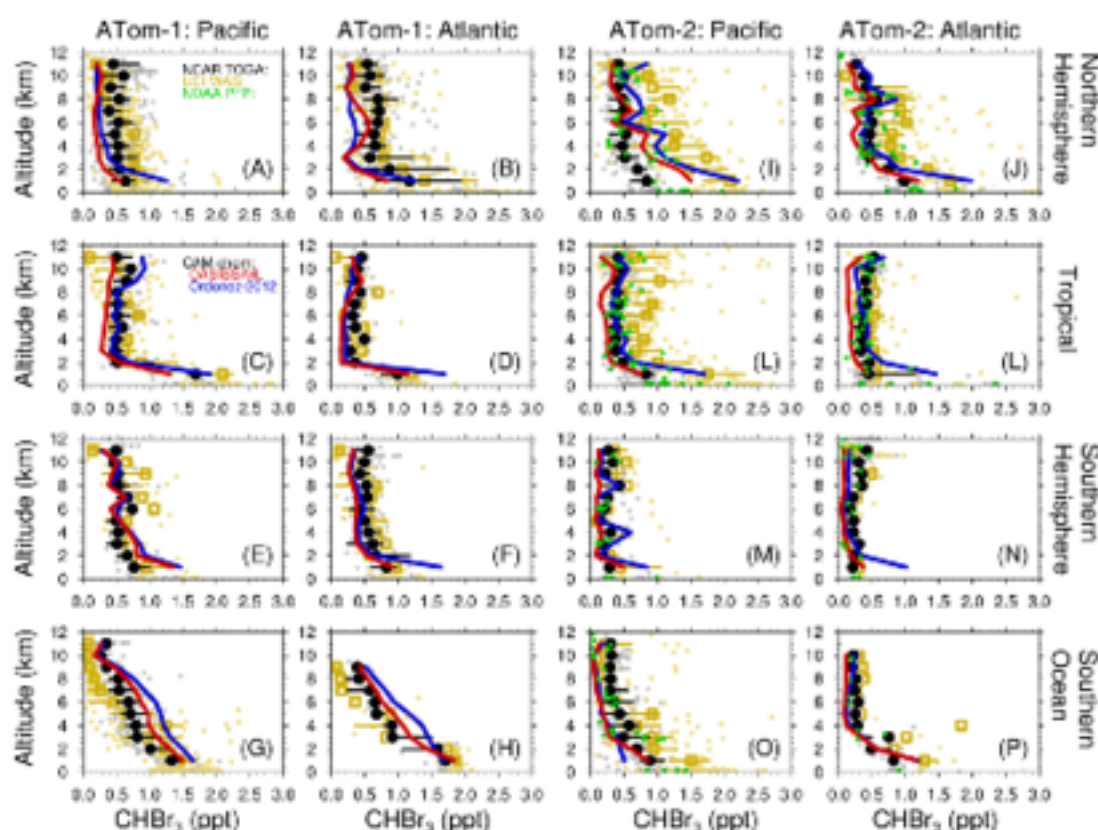
Figure 10. Vertical distributions of CHBr₃ measured using NCAR TOGA (gray: raw data; black: median/quartiles), UCI WAS (light yellow dots: raw data; gold boxes: median/quartiles), and NOAA PFP (green: raw data; medians not calculated due to relatively small sample sizes), as well as the CAM-chem modeled results using OASISS/Machine-learning (ML) and the Ordóñez-2012 inventory, during ATom-1 (A–H) and ATom-2 (I–P). The flight tracks and the spatial domains are defined in Figure S2. The modeling results were sampled along the flight tracks. NOAA PFP CHBr₃ measurements were not available for ATom-1 at this time. NCAR TOGA = National Center for Atmospheric Research Trace Organic Gas Analyzer; UCI WAS = University of California, Irvine Whole Air Sampler; NOAA PFP = National Oceanic and Atmospheric Administration Programmable Flask Package; CAM-chem = Community Atmospheric Model with chemistry; OASISS = Online Air-Sea Interface for Soluble Species.

Tropics during ATom-1. Observed CHBr₃ show strong vertical gradient in the marine boundary layer, and the vertical profile shapes were very well captured by the simulations using both inventories (Figure 10c). On the Pacific side, the simulation with the Ordóñez-2012 inventory shows better agreement with observations, while the simulation with OASISS/ML slightly underestimate CHBr₃ in the marine boundary layer (Figure 10c). On the Atlantic side, however, the observed CHBr₃ in the marine boundary layer is well captured by the simulation using OASISS/ML and is slightly overestimated by Ordóñez-2012 (Figure 10c). Both simulations captured the CHBr₃ levels in the tropical free troposphere over the Pacific and the Atlantic (Figure 10c).

Southern Hemisphere during Atom-1. Enhanced CHBr₃ was observed in the marine boundary layer, with weaker vertical gradient compared to the tropics, implying weaker oceanic emissions (Figures 10e and 10f). Both simulations overestimate CHBr₃ in the marine boundary layer on the Pacific side (Figure 10e). On the Atlantic side, however, CHBr₃ mixing ratios in the marine boundary layer were well captured by the simulation using OASISS/ML but was underestimated by the Ordóñez-2012 inventory (Figure 10f).

Southern Ocean during ATom1. Elevated CHBr₃ was reported by both NCAR TOGA and UCI WAS, which was gradually decreased with increasing altitude, implying that the Southern Ocean is a net source of CHBr₃ during boreal summer and possibly prolonged tropospheric lifetime of CHBr₃ (Figures 10g and 10h). The
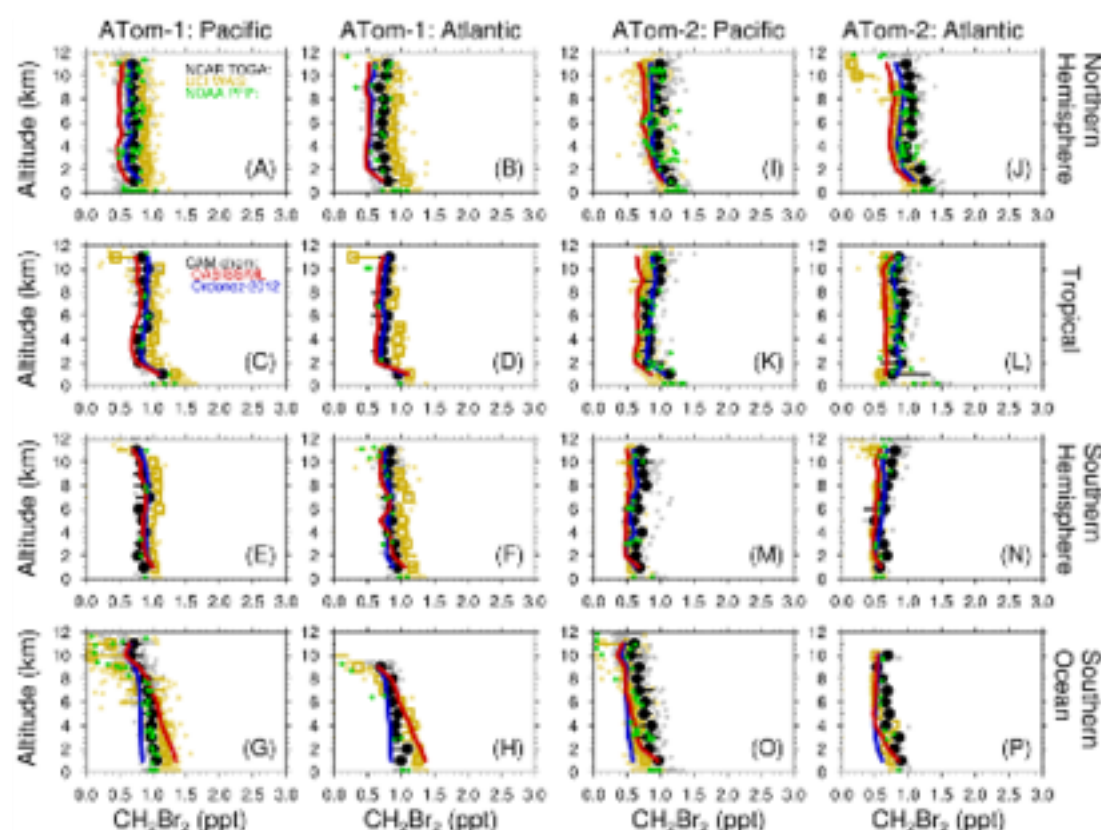
Figure 11. Vertical distributions of CHBr₃ measured using NCAR TOGA (gray: raw data; black: median/quartiles), UCI WAS (light yellow dots: raw data; gold boxes: median/quartiles), and NOAA PFP (green: raw data; medians not calculated due to relatively small sample sizes), as well as the CAM-chem modeled results using OASISS/Machine-learning and the Ordóñez-2012 inventory, during ATom-1 (A–H) and ATom-2 (I–P). The modeling results were sampled along the flight tracks. NCAR TOGA = National Center for Atmospheric Research Trace Organic Gas Analyzer; UCI WAS = University of California, Irvine Whole Air Sampler; NOAA PFP = National Oceanic and Atmospheric Administration Programmable Flask Package; CAM-chem = Community Atmospheric Model with chemistry; OASISS = Online Air-Sea Interface for Soluble Species.

observed vertical distributions of CHBr₃ over the Southern Ocean were reasonably well captured by both simulations, and the simulation using OASISS/ML shows slightly better agreement (Figures 10g and 10h).

Northern Hemisphere during ATom-2. Observed CHBr₃ decreased with increasing altitude. However, CHBr₃ reported by NCAR TOGA and UCI WAS show a factor of ~2 difference on both the Pacific and the Atlantic side, while NOAA PFP measurements show similar range as NCAR TOGA (Figures 10i and 10j). Both simulations captured the vertical profile shapes reasonably well, and the discrepancies between these two simulations were generally smaller than that between the three measurement techniques.

Tropics during ATom-2. Elevated CHBr₃ was reported in the marine boundary layer on the Pacific side but not so much on the Atlantic side. Again the three sets of measurement techniques report nearly a factor of 2 difference in CHBr₃ over the tropical Pacific (Figure 10k), but much better agreement was observed over the tropical Atlantic (Figure 10l). Interestingly, over the tropical Pacific, the simulation using OASISS/ML was in good agreement with the NCAR TOGA and NOAA PFP measurements, but the simulation using Ordóñez-2012 was in good agreement with the UCI WAS measurements (Figure 10k). Over the tropical Atlantic, the simulation using OASISS/ML was in reasonable agreement with both observations in the marine-boundary layer but showed a noticeable low bias in the free troposphere, possibly due to the uncertainties in the vertical transport. The Ordóñez-2012 inventory tends to overestimate CHBr₃ in the marine-boundary layer (Figure 10l).

Table 2
Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) of the CAM-chem Modeled $CHBr_3$ and CH_2Br_2, Using Either OASISS/Machine-Learning (ML) or Ordóñez-2012 Inventory, Compared to NCAR TOGA, UCI WAS, and NOAA PFP Measurements During ATom-1 and ATom-2

Species	CAM-chem inventory	NCAR TOGA		UCI WAS		NOAA PFP	
		RMSE (ppt)	MAPE (%)	RMSE (ppt)	MAPE (%)	RMSE (ppt)	MAPE (%)
$CHBr_3$	OASISS/ML	0.34	40	0.51	136	0.22	28
	Ordóñez-2012	0.58	71	0.51	124	0.52	76
CH_2Br_2	OASISS/ML	0.30	44	0.27	39	0.22	25
	Ordóñez-2012	0.37	59	0.24	35	0.18	21

Note. CAM-chem = Community Atmospheric Model with chemistry; OASISS = Online Air-Sea Interface for Soluble Species; NCAR TOGA = National Center for Atmospheric Research Trace Organic Gas Analyzer; UCI WAS = University of California, Irvine Whole Air Sampler; NOAA PFP = National Oceanic and Atmospheric Administration Programmable Flask Package.

Southern Hemisphere during ATom-2. The observed $CHBr_3$ did not show strong vertical gradient over the Southern Hemisphere (Figures 10m and 10n). On both the Pacific and the Atlantic side, the observed vertical distribution of $CHBr_3$ was reasonably well captured by the simulation using OASISS/ML, and Ordóñez-2012 overestimate $CHBr_3$ in the marine boundary layer by ~50% (Figures 10m and 10n).

Southern Ocean during ATom2. The observed $CHBr_3$ gradually decreased with increasing altitude over the Southern Ocean, and the two instruments showed once again a factor of ~2 discrepancy in the lower troposphere (Figures 10o and 10p). The observed vertical distribution of $CHBr_3$ was better captured by the simulation using OASISS/ML on the Pacific side (Figure 10o). Both simulations were consistent on the Atlantic side (Figure 10p).

CH_2Br_2, however, does not show as strong vertical and spatial variations as $CHBr_3$ (Figure 11), due to its weaker oceanic emissions and longer tropospheric lifetime. During ATom-1, the vertical profile shapes of CH_2Br_2 were reasonably well captured by both simulations (Figures 11a–11h). The simulation with Ordóñez-2012 inventory showed better agreement with observations except for over the Southern Ocean (Figures 11a–11f). Over the Southern Ocean, the simulation using Ordóñez-2012 inventory slightly underestimated the observed CH_2Br_2, yet the simulation using OASISS/ML showed a slightly overestimation (Figures 11g and 11h). During ATom-2, the observed vertical distributions of CH_2Br_2 in all regions except for the Southern Ocean were reasonably well captured by both simulations as well, although the simulation using OASISS/ML slightly underestimate the observed CH_2Br_2 (Figures 11i–11n). Over the Southern Ocean, however, the simulation using OASISS/ML showed better agreement with the observations, while the Ordóñez-2012 inventory underestimated CH_2Br_2 observations (Figure 11p). Note that the Ordóñez-2012 inventory has virtually no CH_2Br_2 emissions over the majority of the Southern Ocean.

Table 2 summarizes the overall model performance compared to the three sets of observations during ATom-1 and ATom-2. For $CHBr_3$, CAM-chem simulation using OASISS/ML shows better agreement with NCAR TOGA and NOAA PFP, yet CAM-chem using Ordóñez-2012 compares slightly better with the UCI WAS measurements. For CH_2Br_2, CAM-chem simulation using OASISS/ML again shows better agreement with NCAR TOGA, but CAM-chem using Ordóñez-2012 shows slightly improved agreement with UCI WAS and NOAA PFP.

7. Conclusions and Remarks

In this work, we present new online bottom-up oceanic emission inventories for $CHBr_3$ and CH_2Br_2 developed for the NCAR CESM2, powered by a data-oriented machine-learning emulator representing the ocean biogeochemistry control. The machine-learning emulator, trained by observations, predicted surface seawater concentrations of $CHBr_3$ and CH_2Br_2 that are in reasonable agreement with the HalOcAt data set, a long-term global observation compilation (1987–2011). The predicted monthly mean surface seawater concentration fields are then used to drive the air-sea exchange, considering the local physical state of the ocean and the atmosphere. The air-sea exchange is fully coupled with the atmospheric chemistry and dynamics. The modeled surface atmospheric $CHBr_3$ and CH_2Br_2 are evaluated with ground-based observations from the NOAA/ESRL ground-based monitoring network, which are more sensitive to nearby oceanic

emissions and local conditions. Overall, the observed mean surface atmospheric concentrations, as well as the seasonal and spatial variations of $CHBr_3$ and CH_2Br_2, are reasonably captured by the new model framework. Additionally, the modeled vertical distributions of $CHBr_3$ and CH_2Br_2, using both the new online bottom-up oceanic emission inventory (this work) and a successful top-down inventory (Ordóñez et al., 2012), are also compared to the global airborne measurements using three measurement techniques (NCAR TOGA, UCI WAS, and NOAA PFP) in both boreal summer and winter (ATom-1 and ATom-2). The performance of the new online bottom-up inventory (this work) is comparable to the Ordóñez-2012 inventory, with better agreement in the Southern Hemisphere and the Southern Ocean. Note that the three sets of observations sometimes show substantial discrepancies especially during ATom-2, which are larger than the differences between the CAM-chem modeling results using different inventories, and further investigation is warranted. The global annual oceanic Br-VSLS emissions estimated in this work are generally comparable to previous estimates but show considerable discrepancies with other recent bottom-up approaches in the Southern Ocean in boreal summer, which are not fully understood. It is notable that the new online bottom-up oceanic inventory (this work) predicts quite high emissions of $CHBr_3$ over the Indian Ocean, the Bay of Bengal, and the South China Sea, partially due to the lack of surface seawater measurements of $CHBr_3$ in the machine-learning training data set (HalOcAt).

The new oceanic emission framework is in general more skillful in reproducing the seasonal and vertical distributions of $CHBr_3$ and CH_2Br_2 compared to the widely used top-down approaches. The ocean biogeochemistry control on the synthesis of the Br-VSLS in the seawater is captured by the machine-learning emulator, and the performance can be further improved once more seawater concentration observations become available, in particular, long-term (multiseasonal) observations in coastal and near sea ice regions, as well as in the western Pacific. This new framework, calibrated by present-day observations (surface seawater and atmosphere), can be used to investigate the future and past oceanic emissions of Br-VSLS under different climate scenarios. Last but not least, we show that this machine-learning emulator provides a fairly accurate, while computationally inexpensive, alternative (compared to a detailed marine ecosystem model) for chemistry-climate models for representing the air-sea exchange of other climate-relevant trace gases, such as dimethyl sulfide and nitrous oxide.

References

Abrahamsson, K., Granfors, A., Ahnoff, M., Cuevas, C. A., & Saiz-Lopez, A. (2018). Organic bromine compounds produced in sea ice in Antarctic winter. *Nature Communications*, 9(1), 5291. https://doi.org/10.1038/s41467-018-07062-8

Apel, E. C., Hills, A. J., Lueb, R., Zindel, S., Eisele, S., & Riemer, D. D. (2003). A fast-GC/MS system to measure C_2 to C_4 carbonyls and methanol aboard aircraft. *Journal of Geophysical Research*, 108(D20), 8794. https://doi.org/10.1029/2002JD003199

Apel, E. C., Hornbrook, R. S., Hills, A. J., Blake, N. J., Barth, M. C., Weinheimer, A., et al. (2015). Upper tropospheric ozone production from lightning NO_x-impacted convection: Smoke ingestion case study from the DC3 campaign. *Journal of Geophysical Research: Atmospheres*, 120, 2505–2523. https://doi.org/10.1002/2014JD022121

Blake, N. J., Blake, D. R., Simpson, I., Meinardie, S., Swanson, A. L., Lopez, J. P., et al. (2003). NMHCs and halocarbons in Asian continental outflow during the Transport and Chemical Evolution over the Pacific (TRACE-P) field campaign: Comparison With PEM-West B. *Journal of Geophysical Research*, 108(D20), 8806. https://doi.org/10.1029/2002JD003367

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Butler, J. H., King, D. B., Lobert, J. M., Montzka, S. A., Yvon-Lewis, S. A., Hall, B. D., et al. (2007). Oceanic distributions and emissions of short-lived halocarbons. *Global Biogeochemical Cycles*, 21, GB1023. https://doi.org/10.1029/2006GB002732

Butler, R., Palmer, P. I., Feng, L., Andrews, S. J., Atlas, E. L., Carpenter, L. J., et al. (2018). Quantifying the vertical transport of $CHBr_3$ and CH_2Br_2 over the western Pacific. *Atmospheric Chemistry and Physics*, 18(17), 13135–13153. https://doi.org/10.5194/acp-18-13135-2018

Carpenter, L. J., Jones, C. E., Dunk, R. M., Hornsby, K. E., & Woeltjen, J. (2009). Air-sea fluxes of biogenic bromine from the tropical and North Atlantic Ocean. *Atmospheric Chemistry and Physics*, 9(5), 1805–1816. https://doi.org/10.5194/acp-9-1805-2009

Carpenter, L. J., & Liss, P. S. (2000). On temperate sources of bromoform and other reactive organic bromine gases. *Journal of Geophysical Research*, 105(D16), 20,539–20,547. https://doi.org/10.1029/2000JD900242

Carpenter, L. J., Wevill, D. J., Palmer, C. J., & Michels, J. (2007). Depth profiles of volatile iodine and bromine-containing halocarbons in coastal Antarctic waters. *Marine Chemistry*, 103(3), 227–236. https://doi.org/10.1016/j.marchem.2006.08.003

Chipperfield, M. P. (2006). New version of the TOMCAT/SLIMCAT offline chemical transport model: Intercomparison of stratospheric tracer experiments. *Quarterly Journal of the Royal Meteorological Society*, 132(617), 1179–1203. https://doi.org/10.1256/qj.05.51

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., et al. (2011). The CCSM4 ocean component. *Journal of Climate*, 25(5), 1361–1389. https://doi.org/10.1175/JCLI-D-11-00091.1

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., et al. (2009). Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems*, 76(1-2), 95–112. https://doi.org/10.1016/j.jmarsys.2008.05.015

Engel, A., Rigby, M., Burkholder, J., Fernandez, R. P., Froidevaux, L., Hall, B., et al. (2019). Update on ozone-depleting substances (ODSs) and other gases of interest to the Montreal Protocol. *Scientific Assessment of Ozone Depletion: 2018*, 1–87.

Fernandez, R. P., Carmona-Balea, A., Cuevas, C. A., Barrera, J. A., Kinnison, D. E., Lamarque, J. F., et al. (2019). Modeling the sources and chemistry of polar tropospheric halogens (Cl, Br, and I) using the CAM-Chem global chemistry-climate model. *Journal of Advances in Modeling Earth Systems*, *11*, 2259–2289. https://doi.org/10.1029/2019MS001655

Fernandez, R. P., Salawitch, R. J., Kinnison, D. E., Lamarque, J. P., & Saiz-Lopez, A. (2014). Bromine partitioning in the tropical tropopause layer: Implications for stratospheric injection. *Atmospheric Chemistry and Physics*, *14*(24), 13,391–13,410. https://doi.org/10.5194/acp-14-13391-2014

Fiehn, A., Quack, B., Hepach, H., Fuhlbrügge, S., Tegtmeier, S., Toohey, M., et al. (2017). Delivery of halogenated very short-lived substances from the west Indian Ocean to the stratosphere during the Asian summer monsoon. *Atmospheric Chemistry and Physics*, *17*(11), 6723–6741. https://doi.org/10.5194/acp-17-6723-2017

Fiehn, A., Quack, B., Stemmler, I., Ziska, F., & Krüger, K. (2018). Importance of seasonally resolved oceanic emissions for bromoform delivery from the tropical Indian Ocean and west Pacific to the stratosphere. *Atmospheric Chemistry and Physics*, *18*(16), 11,973–11,990. https://doi.org/10.5194/acp-18-11973-2018

Fuhlbrügge, S., Quack, B., Tegtmeier, S., Atlas, E., Hepach, H., Shi, Q., et al. (2016). The contribution of oceanic halocarbons to marine and fine tropospheric air over the tropical West Pacific. *Atmospheric Chemistry and Physics*, *16*(12), 7569–7585. https://doi.org/10.5194/acp-16-7569-2016

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*(6), 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Hossaini, R., Chipperfield, M. P., Montzka, S. A., Rap, A., Dhomse, S., & Feng, W. (2015). Efficiency of short-lived halogens at influencing climate through depletion of stratospheric ozone. *Nature Geoscience*, *8*(3), 186–190. https://doi.org/10.1038/ngeo2363, http://www.nature.com/journal/v8/n3/abs/ngeo2363.html#supplementary-information

Hossaini, R., Mantle, H., Chipperfield, M. P., Montzka, S. A., Hamer, P., Ziska, F., et al. (2013). Evaluating global emission inventories of biogenic bromocarbons. *Atmospheric Chemistry and Physics*, *13*(23), 11,819–11,838. https://doi.org/10.5194/acp-13-11819-2013

Hossaini, R., Patra, P. K., Leeson, A. A., Krysztofiak, G., Abraham, N. L., Andrews, S. J., et al. (2016). A multi-model intercomparison of halogenated very short-lived substances (TransCom-VSLS): Linking oceanic emissions and tropospheric transport for a reconciled estimate of the stratospheric source gas injection of bromine. *Atmospheric Chemistry and Physics*, *16*(14), 9163–9187. https://doi.org/10.5194/acp-16-9163-2016

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research*, *117*, C01011. https://doi.org/10.1029/2011JC007395

Hughes, C., & Sun, S. (2016). Light and brominating activity in two species of marine diatom. *Marine Chemistry*, *181*, 1–9. https://doi.org/10.1016/j.marchem.2016.02.003

Jeffery, C. D., Robinson, I. S., & Woolf, D. K. (2010). Tuning a physically-based model of the air-sea gas transfer velocity. *Ocean Modelling*, *31*(1), 28–35. https://doi.org/10.1016/j.ocemod.2009.09.001

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., & Lelieveld, J. (2005). Technical Note: The Modular Earth Submodel System (MESSy)—A new approach towards Earth system modeling. *Atmospheric Chemistry and Physics*, *5*(2), 433–444. https://doi.org/10.5194/acp-5-433-2005

Johnson, M. T. (2010). A numerical scheme to calculate temperature and salinity dependent air-water transfer velocities for any gas. *Ocean Science*, *6*(4), 913–932. https://doi.org/10.5194/os-6-913-2010

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2014). The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Krumhardt, K. M., Lovenduski, N. S., Long, M. C., & Lindsay, K. (2017). Avoidable impacts of ocean warming on marine primary production: Insights from the CESM ensembles. *Global Biogeochemical Cycles*, *31*, 114–133. https://doi.org/10.1002/2016GB005528

Lamarque, J. F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., et al. (2012). CAM-chem: Description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. *Geoscientific Model Development*, *5*(2), 369–411. https://doi.org/10.5194/gmd-5-369-2012

Lennartz, S. T., Krysztofiak, G., Marandino, C. A., Sinnhuber, B. M., Tegtmeier, S., Ziska, F., et al. (2015). Modelling marine emissions and atmospheric distributions of halocarbons and dimethyl sulfide: The influence of prescribed water concentration vs. prescribed emissions. *Atmospheric Chemistry and Physics*, *15*(20), 11,753–11,772. https://doi.org/10.5194/acp-15-11753-2015

Liang, Q., Stolarski, R. S., Kawa, S. R., Nielsen, J. E., Douglass, A. R., Rodriguez, J. M., et al. (2010). Finding the missing stratospheric Br$_y$: A global modeling study of CHBr$_3$ and CH$_2$Br$_2$. *Atmospheric Chemistry and Physics*, *10*(5), 2269–2286. https://doi.org/10.5194/acp-10-2269-2010

Lim, C. Y., & Manley, S. L. (2012). Bromoform production from seawater treated with bromoperoxidase. *Limnology and Oceanography*, *57*(6), 1857–1866. https://doi.org/10.4319/lo.2012.57.6.1857

Liss, P. S., & Slater, P. G. (1974). Flux of gases across the air-sea interface. *Nature*, *247*(5438), 181–184. https://doi.org/10.1038/247181a0

Liu, Y., Thornton, D. C. O., Bianchi, T. S., Arnold, W. A., Shields, M. R., Chen, J., & Yvon-Lewis, S. A. (2015). Dissolved organic matter composition drives the marine production of brominated very short-lived substances. *Environmental Science & Technology*. https://doi.org/10.1021/es506464k

Liu, Y., Yvon-Lewis, S. A., Thornton, D. C. O., Butler, J. H., Bianchi, T. S., Campbell, L., et al. (2013). Spatial and temporal distributions of bromoform and dibromomethane in the Atlantic Ocean and their relationship with photosynthetic biomass. *Journal of Geophysical Research: Oceans*, *118*, 3950–3965. https://doi.org/10.1002/jgrc.20299

Liu, Y., Yvon-Lewis, S. A., Thornton, D. C. O., Campbell, L., & Bianchi, T. S. (2013). Spatial distribution of brominated very short-lived substances in the eastern Pacific. *Journal of Geophysical Research: Oceans*, *118*, 2318–2328. https://doi.org/10.1002/jgrc.20183

Long, M. C., Deutsch, C., & Ito, T. (2016). Finding forced trends in oceanic oxygen. *Global Biogeochemical Cycles*, *30*, 381–397. https://doi.org/10.1002/2015GB005310

Mackay, D., & Yeun, A. T. K. (1983). Mass transfer coefficient correlations for volatilization of organic solutes from water. *Environmental Science & Technology*, *17*(4), 211–217. https://doi.org/10.1021/es00110a006

Montzka, S. A., Butler, J. H., Hall, B. D., Mondeel, D. J., & Elkins, J. W. (2003). A decline in tropospheric organic bromine. *Geophysical Research Letters*, *30*(15), 1826. https://doi.org/10.1029/2003GL017745

Montzka, S. A., Krol, M., Dlugokencky, E., Hall, B., Jöckel, P., & Lelieveld, J. (2011). Small interannual variability of global atmospheric hydroxyl. *Science*, *331*(6013), 67–69. https://doi.org/10.1126/science.1197640

Moon, J. K., & Braucher, O. (2008). Sedimentary and mineral dust sources of dissolved iron to the world ocean, 36.

Moon, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine ecosystem dynamics and biogeochemical cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 scenarios. *Journal of Climate*, *26*(23), 9291–9312. https://doi.org/10.1175/JCLI-D-12-00566.1

Moore, R. M., Webb, M., Tokarczyk, R., & Wever, R. (1996). Bromoperoxidase and iodoperoxidase enzymes and production of halogenated methanes in marine diatom cultures. *Journal of Geophysical Research*, *101*(C9), 20,899–20,908. https://doi.org/10.1029/96JC01248

Navarro, M. A., Atlas, E. L., Saiz-Lopez, A., Rodriguez-Lloveras, X., Kinnison, D. E., Lamarque, J.-F., et al. (2015). Airborne measurements of organic bromine compounds in the Pacific tropical tropopause layer. *Proceedings of the National Academy of Sciences*, *112*(45), 13,789–13,793. https://doi.org/10.1073/pnas.1511463112

Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., et al. (2000). In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Global Biogeochemical Cycles*, *14*(1), 373–387. https://doi.org/10.1029/1999GB900091

Ordóñez, C., Lamarque, J. F., Tilmes, S., Kinnison, D. E., Atlas, E. L., Blake, D. R., et al. (2012). Bromine and iodine chemistry in a global chemistry-climate model: Description and evaluation of very short-lived oceanic sources. *Atmospheric Chemistry and Physics*, *12*(3), 1423–1447. https://doi.org/10.5194/acp-12-1423-2012

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pozzer, A., Jöckel, P., Sander, R., Williams, J., Ganzeveld, L., & Lelieveld, J. (2006). Technical note: The MESSy-submodel AIRSEA calculating the air-sea exchange of chemical species. *Atmospheric Chemistry and Physics*, *6*(12), 5435–5444. https://doi.org/10.5194/acp-6-5435-2006

Quack, B., & Wallace, D. W. R. (2003). Air-sea flux of bromoform: Controls, rates, and implications. *Global Biogeochemical Cycles*, *17*(1), 1023. https://doi.org/10.1029/2002GB001890

Roshan, S., & DeVries, T. (2017). Efficient dissolved organic carbon production and export in the oligotrophic ocean. *Nature Communications*, *8*(1), 1–8. https://doi.org/10.1038/s41467-017-02227-3

Saiz-Lopez, A., Lamarque, J. F., Kinnison, D. E., Tilmes, S., Ordóñez, C., Orlando, J. J., et al. (2012). Estimating the climate significance of halogen-driven ozone loss in the tropical marine troposphere. *Atmospheric Chemistry and Physics*, *12*(9), 3939–3949. https://doi.org/10.5194/acp-12-3939-2012

Saiz-Lopez, A., & von Glasow, R. (2012). Reactive halogen chemistry in the troposphere. *Chemical Society Reviews*, *41*(19), 6448–6472. https://doi.org/10.1039/C2CS35208G

Salawitch, R. J., Weisenstein, D. K., Kovalenko, L. J., Sioris, C. E., Wennberg, P. O., Chance, K., et al. (2005). Sensitivity of ozone to bromine in the lower stratosphere. *Geophysical Research Letters*, *32*, L05811. https://doi.org/10.1029/2004GL021504

Schmidt, J. A., Jacob, D. J., Horowitz, H. M., Hu, L., Sherwen, T., Evans, M. J., et al. (2016). Modeling the observed tropospheric BrO background: Importance of multiphase chemistry and implications for ozone, OH, and mercury. *Journal of Geophysical Research: Atmospheres*, *121*, 11,819–11,835. https://doi.org/10.1002/2015JD024229

Sherwen, T., Chance, R. J., Tinel, L., Ellis, D., Evans, M. J., & Carpenter, L. J. (2019). A machine learning based global sea-surface iodide distribution. *Earth System Science Data Discussions*, 1–40. https://doi.org/10.5194/essd-2019-40

Sherwen, T., Evans, M. J., Carpenter, L. J., Schmidt, J. A., & Mickley, L. J. (2017). Halogen chemistry reduces tropospheric O₃ radiative forcing. *Atmospheric Chemistry and Physics*, *17*(2), 1557–1569. https://doi.org/10.5194/acp-17-1557-2017

Shibazaki, A., Ambiru, S., Kurihara, M., Tamegai, H., & Hashimoto, S. (2016). Phytoplankton as a temperate marine source of brominated methanes. *Marine Chemistry*, *181*, 44–50. https://doi.org/10.1016/j.marchem.2016.03.004

Sobhani, N., Loft, R., Gettelman, A., Chen, C.-C., & Gagne, D. J., II (2018). Using machine learning to emulate critical cloud microphysical processes. *AGU Fall Meeting Abstracts*, 13. Retrieved from http://adsabs.harvard.edu/abs/2018AGUFMIN11C0686S

Stemmler, I., Hense, I., & Quack, B. (2015). Marine sources of bromoform in the global open ocean—Global patterns and emissions. *Biogeosciences*, *12*(6), 1967–1981. https://doi.org/10.5194/bg-12-1967-2015

Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., et al. (2018). The O₂/N₂ ratio and CO₂ Airborne Southern Ocean Study. *Bulletin of the American Meteorological Society*, *99*(2), 381–402. https://doi.org/10.1175/BAMS-D-16-0206.1

Tilmes, S., Lamarque, J. F., Emmons, L. K., Kinnison, D. E., Ma, P. L., Liu, X., et al. (2015). Description and evaluation of tropospheric chemistry and aerosols in the Community Earth System Model (CESM1.2). *Geoscientific Model Development*, *8*(5), 1395–1426. https://doi.org/10.5194/gmd-8-1395-2015

Wales, P. A., Salawitch, R. J., Nicely, J. M., Anderson, D. C., Canty, T. P., Baidar, S., et al. (2018). Stratospheric injection of brominated very short-lived substances: Aircraft observations in the western Pacific and representation in global models. *Journal of Geophysical Research: Atmospheres*, *123*, 5690–5719. https://doi.org/10.1029/2017JD027978

Wang, S., Hornbrook, R. S., Hills, A., Emmons, L. K., Tilmes, S., Lamarque, J. F., et al. (2019). Atmospheric acetaldehyde: Importance of air-sea exchange and a missing source in the remote troposphere. *Geophysical Research Letters*, *46*, 5601–5613. https://doi.org/10.1029/2019GL082034

Wang, S., McNamara, S. M., Moore, C. W., Obrist, D., Steffen, A., Shepson, P. B., et al. (2019). Direct detection of atmospheric atomic bromine leading to mercury and ozone depletion. *Proceedings of the National Academy of Sciences*, *116*(29), 14,479–14,484. https://doi.org/10.1073/pnas.1900613116

Wang, S., Schmidt, J. A., Baidar, S., Coburn, S., Dix, B., Koenig, T. K., et al. (2015). Active and widespread halogen chemistry in the tropical and subtropical free troposphere. *Proceedings of the National Academy of Sciences*, *112*(30), 9281–9286. https://doi.org/10.1073/pnas.1505142112

Warwick, N. J., Pyle, J. A., Carver, G. D., Yang, X., Savage, N. H., O'Connor, F. M., & Cox, R. A. (2006). Global modelling of biogenic bromocarbons. *Journal of Geophysical Research*, *111*, D24305. https://doi.org/10.1029/2006JD007264

Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E., Asher, E. C., Barletta, B., et al. (2018). ATom: Merged atmospheric chemistry, trace gases, and aerosols. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1581

Yamamoto, H., Yokouchi, Y., Otsuki, A., & Itoh, H. (2001). Depth profiles of volatile halogenated hydrocarbons in seawater in the Bay of Bengal. *Chemosphere*, *45*(3), 371–377. https://doi.org/10.1016/S0045-6535(00)00541-5

Yokouchi, Y., Saito, T., Zeng, J., Mukai, H., & Montzka, S. (2017). Seasonal variation of bromocarbons at Hateruma Island, Japan: Implications for global sources. *Journal of Atmospheric Chemistry*, *74*(2), 171–185. https://doi.org/10.1007/s10874-016-9333-9

Ziska, F., Quack, B., Abrahamsson, K., Archer, S. D., Atlas, E., Bell, T., et al. (2013). Global sea-to-air flux climatology for bromoform, dibromomethane and methyl iodide. *Atmospheric Chemistry and Physics*, 13(17), 8915–8934. https://doi.org/10.5194/acp-13-8915-2013

Ziska, F., Quack, B., Tegtmeier, S., Stemmler, I., & Krüger, K. (2017). Future emissions of marine halogenated very-short-lived substances under climate change. *Journal of Atmospheric Chemistry*, 74(2), 245–260. https://doi.org/10.1007/s10874-016-9355-3