

## FluxEngine: A Flexible Processing System for Calculating Atmosphere–Ocean Carbon Dioxide Gas Fluxes and Climatologies

JAMIE D. SHUTLER,<sup>\*</sup> PETER E. LAND,<sup>+</sup> JEAN-FRANCOIS PIOLLE,<sup>#</sup> DAVID K. WOOLF,<sup>@</sup>  
LONNEKE GODDIJN-MURPHY,<sup>&</sup> FREDERIC PAUL,<sup>#</sup> FANNY GIRARD-ARDHUIN,<sup>#</sup>  
BERTRAND CHAPRON,<sup>#</sup> AND CRAIG J. DONLON<sup>\*\*</sup>

<sup>\*</sup> *University of Exeter, Penryn, Cornwall, United Kingdom*

<sup>+</sup> *Plymouth Marine Laboratory, Plymouth, Devon, United Kingdom*

<sup>#</sup> *Laboratoire d’Océanographie Physique et Spatiale (LOPS), IUEM, F-29280, Ifremer, University of Brest, CNRS, IRD, Brest, France*

<sup>@</sup> *Heriot-Watt University, Edinburgh, United Kingdom*

<sup>&</sup> *Environmental Research Institute, North Highland College, University of the Highlands and Islands, Thurso, Caithness, United Kingdom*

<sup>\*\*</sup> *European Space Agency, Noordwijk, Netherlands*

(Manuscript received 21 October 2014, in final form 22 June 2015)

### ABSTRACT

The air–sea flux of greenhouse gases [e.g., carbon dioxide (CO<sub>2</sub>)] is a critical part of the climate system and a major factor in the biogeochemical development of the oceans. More accurate and higher-resolution calculations of these gas fluxes are required if researchers are to fully understand and predict future climate. Satellite Earth observation is able to provide large spatial-scale datasets that can be used to study gas fluxes. However, the large storage requirements needed to host such data can restrict its use by the scientific community. Fortunately, the development of cloud computing can provide a solution. This paper describes an open-source air–sea CO<sub>2</sub> flux processing toolbox called the “FluxEngine,” designed for use on a cloud-computing infrastructure. The toolbox allows users to easily generate global and regional air–sea CO<sub>2</sub> flux data from model, in situ, and Earth observation data, and its air–sea gas flux calculation is user configurable. Its current installation on the Nephelae Cloud allows users to easily exploit more than 8 TB of climate-quality Earth observation data for the derivation of gas fluxes. The resultant netCDF data output files contain >20 data layers containing the various stages of the flux calculation along with process indicator layers to aid interpretation of the data. This paper describes the toolbox design, which verifies the air–sea CO<sub>2</sub> flux calculations; demonstrates the use of the tools for studying global and shelf sea air–sea fluxes; and describes future developments.

### 1. Introduction

The climate of Earth is sensitive to the radiative impact of a number of gases and different types of particles in the atmosphere. The atmospheric concentration of many important gases and particles is sensitive to the air–sea transfer of volatile compounds. These gases can also play a substantial role in the biogeochemistry of the oceans. It is therefore important to quantify

contemporary air–sea fluxes of gases and also to provide the understanding necessary to project possible future changes in these fluxes. The air–sea fluxes of gases can in some cases be inferred indirectly, but most flux estimates depend on a calculation using a standard bulk air–sea gas transfer (e.g., as defined within Takahashi et al. 2009, hereafter T09). For each gas, this calculation depends upon measurements of the gas concentration in the surface ocean and the lower atmosphere and upon “transfer coefficients” that describe the “rate constants” for transfer across the sea surface. The simplest calculation requires only a single transfer coefficient, the gas transfer velocity.

 Denotes Open Access content.

*Corresponding author address:* Jamie D. Shutler, Department of Geography, University of Exeter, Peter Lanyon Building, Treliether Road, Penryn, Cornwall TR10 9FE, United Kingdom.  
E-mail: j.d.shutler@exeter.ac.uk



This article is licensed under a [Creative Commons Attribution 4.0 license](https://creativecommons.org/licenses/by/4.0/).

Greenhouse gases are those that can absorb and emit infrared radiation. Of these gases, CO<sub>2</sub> is one of the most studied and systematically observed. Increasing levels of atmospheric CO<sub>2</sub>, caused by the burning of fossil fuels and biomass, are of growing concern due to their impact on the global climate system. Understanding the pathways, sources, sinks, and impact of CO<sub>2</sub> on the earth's climate system is essential for monitoring climate and predicting future scenarios. The global ocean is thought to annually absorb ~25% of anthropogenic CO<sub>2</sub> emissions (Le Quéré et al. 2015, 2014), and it constitutes the only true net sink for anthropogenic CO<sub>2</sub> over the last 200 years (Sabine et al. 2004). The North Atlantic sink in particular has been shown to be highly variable (Watson et al. 2009) and the mechanisms driving this variability are not well understood. Therefore, isolating and reducing the uncertainties in the estimates of the oceanic sink of CO<sub>2</sub> is a crucial goal of climate science (Le Quéré et al. 2009).

In the last decade there has been an explosion in the availability of large (>1 TB) high-quality, well-characterized, and often multisensor cross-calibrated Earth observation (EO) datasets. For example, the European Space Agency (ESA) GlobWave project (<http://globwave.ifremer.fr/>) produced a 20-plus-year time series of global coverage multisensor cross-calibrated wave and wind data. Similar efforts in the United States resulted in the National Aeronautics and Space Administration (NASA) project Making Earth System Data Records for Use in Research Environments (MEaSUREs), which produced a 13-plus-year time series of global coverage multisensor surface biology datasets (Maritorena et al. 2010). These successes and the classification by the Group on Earth Observations (GEO) of a number of parameters discernable from space as essential climate variables prompted ESA to start their Climate Change Initiative (CCI) projects. There are currently 14 different CCI projects. For those interested in air–sea gas fluxes, arguably the most interesting of these projects is the Sea Surface Temperature CCI project, which has provided a 20-plus-year time series of global coverage multisensor sea surface skin and subskin temperature data (Merchant et al. 2012). Unfortunately, the resources required to download and exploit these large global and decadal datasets can limit their exploitation by the scientific community. The development of cloud technologies and storage provides a solution. Cloud computing can be defined as interconnected computing resources that can be easily scaled up (grown) or down (shrunk) while maintaining its capability or function, rather like a “cloud” in the atmosphere. These systems have a number of key features, such as a high level of redundancy (e.g., servers can be

removed or upgraded without users noticing) and their scalable nature (e.g., they use standard hardware and software, allowing low-cost expansion of a cloud). Through cloud computing it becomes possible for users to easily remotely access and process large volumes of data and then simply download the results to their local desktop computers or laptops.

#### a. The OceanFlux Greenhouse Gases project

The OceanFlux Greenhouse Gases project was funded by the European Space Agency in 2011 to encourage the use of satellite Earth observation data for studying air–sea gas fluxes. To achieve this, the objectives of the project included the development and validation of novel gas flux Earth observation algorithms (Godijn-Murphy et al. 2013, 2012) and scientific analyses (Land et al. 2013). Another objective was to provide datasets and processing tools that can be used by the scientific community. Accordingly, the gas flux data processing tools, collectively named the “FluxEngine,” are described in this paper. The FluxEngine allows users to configure the flux parameterization and to select their chosen input data, and then it generates the resulting monthly global flux datasets. A plethora of climate study quality EO, in situ, and model data are available as input to the toolbox and to aid the interpretation of the resultant flux data (see Table 1). The outputs from the system are standard netCDF datasets that can be easily read into a number of third-party scientific software packages. The primary gas of interest for the OceanFlux Greenhouse Gases project was CO<sub>2</sub>. Therefore, the FluxEngine has been developed to aid the study of the air–sea flux of CO<sub>2</sub>, although as described later in the paper, the toolbox can also be used to support the study of other gases, such as N<sub>2</sub>O and dimethyl sulfide (DMS).

#### b. Air–sea flux calculations

The flux of CO<sub>2</sub> between the atmosphere and the ocean (air–sea) is controlled by wind speed; sea state; sea surface temperature; surface processes, including any biological activity; and the difference in CO<sub>2</sub> fugacity between the ocean and the atmosphere. The air-to-sea flux of CO<sub>2</sub> ( $F$ , g m<sup>-2</sup> s<sup>-1</sup>) is calculated using the gas transfer velocity  $k$  (m s<sup>-1</sup>), and the difference in CO<sub>2</sub> concentration (g m<sup>-3</sup>) between the base [CO<sub>2AQW</sub>] and the top [CO<sub>2AQ0</sub>] of a thin (~10–250 μm) mass boundary layer at the sea surface:

$$F = k([CO_{2AQW}] - [CO_{2AQ0}]). \quad (1)$$

The concentration of CO<sub>2</sub> in seawater is the product of its solubility  $\alpha$  (g m<sup>-3</sup> μatm<sup>-1</sup>) and its fugacity  $fCO_2$  (μatm). As gas solubility is a function of salinity and

TABLE 1. The EO, in situ, model, and climatology (<sup>C</sup>) data available to use in the FluxEngine. ARC is the Along-Track Scanning Radiometer (ATSR) Reprocessing for Climate. OSTIA denotes Operational Sea Surface Temperature and Sea Ice Analysis. GDS is the GHRSSST Data Specification. OC CCI is the Ocean Color CCI.

Parameter	Type	No. of sensors	Name	Input data version/source	Year range
SST skin	EO	Multiple	ESA ARC	v1.1.1 (Merchant et al. 2012)	1992–2010
SST foundation	EO	Multiple	OSTIA	1.0 (GDS v1.7) (Donlon et al. 2012)	2008–10
SST fronts	EO <sup>C</sup>	Multiple	Met Office data	v01, provided by the Met Office	2010 <sup>C</sup>
$U_{10}$	EO	Multiple	ESA GlobWave	v1.0, <a href="http://globwave.ifremer.fr/">http://globwave.ifremer.fr/</a>	1992–2010
Hs	EO	Multiple	ESA GlobWave	v1.0 <a href="http://globwave.ifremer.fr/">http://globwave.ifremer.fr/</a>	1992–2010
$\sigma_0$	EO	Multiple	ESA GlobWave	v1.0, <a href="http://globwave.ifremer.fr/">http://globwave.ifremer.fr/</a>	1992–2010
Rn	EO/in situ	Multiple	GPCP	v2.2 (Huffman et al. 2012)	1981–2010
Rn	EO	Single	TRMM	v4.2	1998–2011
Rn	EO	Single	SSM/I	f16, f17	2007–10
Chl- <i>a</i>	EO	Multiple	ESA OC CCI	v0.95, <a href="http://www.esa-oceancolour-cci.org">http://www.esa-oceancolour-cci.org</a>	1997–2011
Chl- <i>a</i>	EO	Multiple	ESA GlobColour	v1.9, <a href="http://www.globcolour.org">http://www.globcolour.org</a>	1997–2008
$p\text{CO}_2$	In situ <sup>C</sup>	Multiple	LDEO Takahashi	(T09)	2000 <sup>C</sup>
			climatology (T09)		
$p\text{CO}_2$	In situ <sup>C</sup>	Multiple	LDEO Takahashi	(Takahashi et al. 2002)	1999 <sup>C</sup>
			climatology (T02)		
$X_{[\text{CO}_2]}$	In situ/model	Multiple		(T09)	2000 <sup>C</sup>
$S$	In situ <sup>C</sup>	Multiple	LDEO Takahashi	(T09)	2000 <sup>C</sup>
			climatology (T09)		
$P$	Model	—	ECMWF	ERA-40 reanalysis and operational version, <a href="http://www.ecmwf.int">http://www.ecmwf.int</a>	2007–10
$P$	Model	—	NCEP CSFR	Reanalysis 1979–2010, <a href="http://cfs.ncep.noaa.gov/cfsr">http://cfs.ncep.noaa.gov/cfsr</a>	1979–2009
Ice	EO	Single	SSM/I	v2.1, <a href="http://cersat.ifremer.fr">http://cersat.ifremer.fr</a>	1991–2012

temperature, it varies across the aqueous boundary layer. Hence, Eq. (1) now becomes

$$F = k(\alpha_W f\text{CO}_{2W} - \alpha_S f\text{CO}_{2A}), \quad (2)$$

where the subscripts denote values in water ( $W$ ), at the sea–air interface ( $S$ ), and in air ( $A$ ).

The  $\text{CO}_2$  concentration (and thus fugacity) is normally measured a few meters below the sea surface rather than at the surface. Variations in temperature at the sea surface (such as diurnal warming) will affect the fugacity via the carbonate reaction. For simplicity we can substitute partial pressure for fugacity because their values differ by <0.5% over the temperature range of interest (McGillis and Wanninkhof 2006). Equation (2) can therefore be alternatively represented as

$$F = k(\alpha_W p\text{CO}_{2W} - \alpha_S p\text{CO}_{2A}). \quad (3)$$

It is also popular for Eqs. (2) and (3) to be collapsed into formulations that ignore the differences between the two solubilities, and just use the waterside solubility  $\alpha_W$  for both halves of the equation, resulting in

$$F = k\alpha_W(f\text{CO}_{2W} - f\text{CO}_{2A}) \quad (4)$$

or

$$F = k\alpha_W(p\text{CO}_{2W} - p\text{CO}_{2A}), \quad (5)$$

and this formulation is often referred to as the “bulk” parameterization. The airside partial pressure of  $\text{CO}_2$ ,  $p\text{CO}_{2A}$  ( $\mu\text{atm}$ ) can be calculated using the concentration of  $\text{CO}_2$  in dry air and air pressure using

$$p\text{CO}_2 = 0.001X_{[\text{CO}_2]}(P - p\text{H}_2\text{O}). \quad (6)$$

Term  $X_{[\text{CO}_2]}$  is the molar fraction of  $\text{CO}_2$  in the dry atmosphere (expressed as the zonal mean in T09 and the FluxEngine),  $P$  is the air pressure (mb, expressed as a daily mean in the FluxEngine), and  $p\text{H}_2\text{O}$  is the saturation vapor pressure (mb), which is defined in terms of sea surface temperature and salinity (Weiss and Price 1980) by

$$p\text{H}_2\text{O} = 1013.25 \exp[24.4543 - 67.4509(100/\text{SST}_K) - 4.8489 \ln(T/\text{SST}_K) - 0.000544S], \quad (7)$$

where  $\text{SST}_K$  is the sea surface temperature dataset of interest (K) and  $S$  is the salinity. For  $p\text{CO}_{2W}$  the FluxEngine relies upon in situ  $p\text{CO}_{2W}$  measurements (e.g., data from a buoy or ship) or an in situ–derived climatology of  $p\text{CO}_{2W}$  data (e.g., T09).

## 2. The development of the FluxEngine

The following sections describe the design of the FluxEngine, the methods used for calculating the air–sea

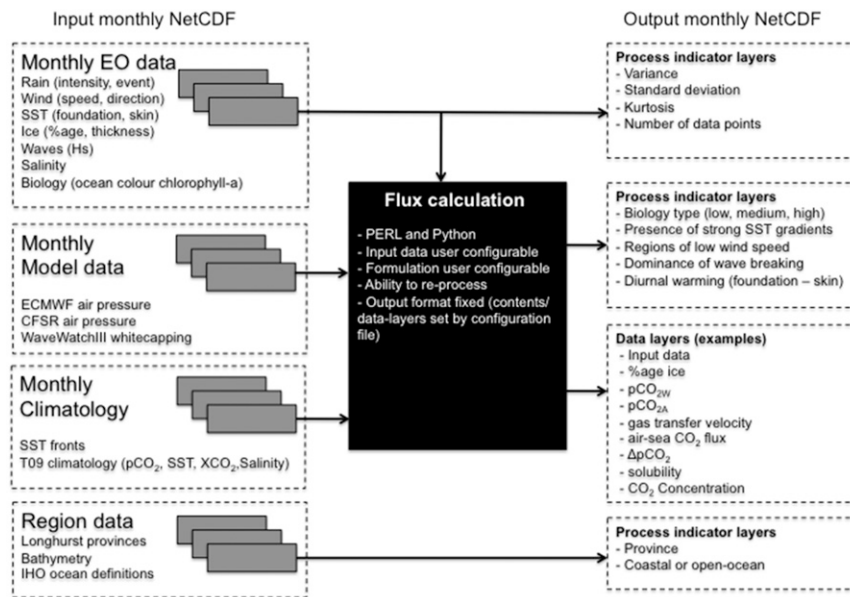


FIG. 1. Conceptual overview of the FluxEngine air-sea gas flux processing approach. Description of (left) the input data and (right) the netCDF output content. The user-configurable air-sea CO<sub>2</sub> flux calculation is represented by (middle) the main component.

fluxes, the input datasets that are available, and the methods used to verify the implementation.

#### a. The FluxEngine design, input data, and implementation

Figure 1 shows a diagram of the main parts of the FluxEngine, conceptually showing the input data and the contents of the output files. The input data are on the left of the diagram, with the calculation in the middle and an overview of the contents of each output file shown on the right. The air-sea fluxes are calculated using monthly composite data and the generation of these input data is described below. The output file is a netCDF4 Climate Format (CF) 1.6 compliant file that contains >20 data layers. The data layers within each output file include the different components of the gas flux calculation, statistics of the input datasets (e.g., variance of the wind speed), and process indicator layers to aid interpretation of the fluxes. The process indicator layers include fixed masks (e.g., land, ocean basins, open ocean, and coastal classification), climatological data (e.g., persistent SST fronts), and other modeling or Earth observation datasets useful for interpreting the fluxes (e.g., chlorophyll-*a* concentrations and model-generated estimates of wave whitecapping). All output from the toolbox consists of monthly global coverage 1° × 1° spatial resolution data (360 × 180 arrays). Therefore, generation of gas fluxes for a complete year (12 months) requires 12 sets of data inputs (one set per

month) and the FluxEngine then produces 12 output files. The air-sea flux calculation utility contains internal integrity checks for all of the output data layers. These integrity checks highlight if any data layers contain data outside of a predefined expected range. If a data value for a data layer falls outside its specified valid data range, then a count is added to the corresponding data element position in one of the process indicator data layers.

The inputs to the flux calculation (box 2 in Fig. 1) are monthly composite data. Table 1 shows all of the monthly datasets that are currently available for the FluxEngine to use as input. Where required, these monthly netCDF composite data were generated using the original daily data. For those datasets that were generated, each monthly composite file contains the mean (first-order moment), median, standard deviation, and the second-, third-, and fourth-order moments as calculated using one calendar month of data.

The FluxEngine was developed using license-free software tools and libraries. The flux calculation utility and supporting utilities use Python [version 2.7 (v2.7)], NumPy (v1.7.1), and the SciPy toolbox (v0.13.0). This source code comprises more than 6000 lines of code.

#### b. The computing platform

The FluxEngine is currently installed on the Centre ERS d'Archivage et de Traitement (CERSAT) Nephelae Cloud. This Linux-based cloud-computing platform provides a petascale storage capacity and distributed

processing over more than 600 computing nodes. It was specifically designed by CERSAT for massive archive processing, for applications including data mining and multidecadal Earth observation data reprocessing. As with all cloud computing, it offers facilities for simple backup and restoration, high-speed data processing (i.e., the processing nodes are close to data, so there is no potential for input/output bottlenecks), the system can be tailored to a specific job (i.e., there is no reliance on physical hardware as it uses virtual servers), no specific skills are required to use it, and it maximizes the use of resources through dynamic reallocation. Data processing runs submitted by the user are scheduled and run on the cloud-computing nodes using a system developed by CERSAT called “GoGo list.” This is simply a wrapper enabling processing jobs to be executed and monitored on the cloud, and the use of GoGo list is completely invisible to the user. It must be noted though that the main reason for using the Nephelae Cloud here is to provide potential users with access to a large and continually growing satellite Earth observation dataset. The FluxEngine can also be installed and used on a desktop or laptop computer with no loss of performance.

### c. Configurable options

The main flux calculation within the FluxEngine is user configurable through the use of a plain text ASCII configuration file. Within the configuration a user can choose the input datasets, the flux calculation model [choose between Eqs. (4) and (9)], and the gas transfer velocity parameterization. A range of different gas transfer parameterizations are available (e.g., McGillis et al. 2001; Nightingale et al. 2000; Wanninkhof 1992), including those based on sea state and surface roughness (Fangohr and Woolf 2007; Goddijn-Murphy et al. 2012). Through a generic formulation (see the appendix), it is also possible for users to use their own wind speed-based parameterization. Through the configuration file, the user can choose to inject random noise (normally distributed with specified mean and standard deviation) to any of the main input datasets (SST,  $U_{10}$ , Hs,  $p\text{CO}_2$ , or  $f\text{CO}_2$ ). In the same way, the user can choose to add a bias offset value to any of the main input datasets. This functionality allows the impact of known input data uncertainties (i.e., root-mean-square error and bias) to be propagated through to the final flux datasets, allowing the impact of these uncertainties to be quantified in terms of the gas fluxes. Example configuration files are included with the open source software.

### d. Additional software tools

Three additional tools exist within the FluxEngine toolbox. These enable the calculation of integrated net

fluxes, regridding of the output data, and cruise or buoy in situ data to be used as inputs. The integrated net flux tool can provide gross and net fluxes, mean values for each dataset within the output data, and estimates of the open ocean fluxes and the flux contribution from any missing data (such as that from coastal, shelf, and enclosed seas). The regridding tool converts the  $1^\circ \times 1^\circ$  data to the  $5^\circ \times 4^\circ$  grid used by the T09 climatological dataset. This allows users to easily compare the FluxEngine output with that of the T09 climatology. The in situ netCDF conversion tool converts sparse in situ data into a spatially and temporally binned or gridded ( $1^\circ \times 1^\circ$  grid) format. For example, the tool allows in situ  $p\text{CO}_2$  data to be used as input to the FluxEngine. Further details on all of the toolbox software components can be found in the appendix of this paper.

### 3. Data quality and verification of the calculations

The widely used T09 climatology dataset provides an ideal benchmark for verifying the operation and output of the FluxEngine, as it contains both the air–sea flux estimates and the input values used to calculate these fluxes. Therefore, the T09 air–sea flux data were first used to verify the FluxEngine integrated net flux tool, and then the main T09 input fields were used to verify the combination of the FluxEngine flux calculation and the integrated net flux tool.

To verify the integrated net flux tool, the year 2000 T09 climatology data were linearly interpolated to a  $1^\circ \times 1^\circ$  grid and then provided as the input to the tool. The integrated net flux tool used the T09 ice normalization (see the appendix). The resultant net air–sea  $\text{CO}_2$  flux for the open ocean region was  $-1.39 \text{ GtCyr}^{-1}$  with an additional  $-0.17 \text{ GtCyr}^{-1}$  attributed to missing data, large lakes, the Mediterranean Sea, and coastal and shelf seas, giving a total of  $-1.56 \text{ GtCyr}^{-1}$ . The difference between this global open ocean result (as calculated using the linearly interpolated data and the net flux tool) and that stated in the original publication is  $<1\%$ .

The T09 climatology data of SST,  $X\text{CO}_2$ ,  $U_{10}$ ,  $p\text{CO}_{2w}$ , air pressure, and percentage ice cover (linearly interpolated to a  $1^\circ \times 1^\circ$  grid) were then used as input to the FluxEngine. The main flux utility was configured to use Eq. (5) to calculate the air–sea fluxes. Figure 2a shows the resultant daily mean air–sea flux map. The projection and scale bar have been chosen to allow easy comparison with Fig. 13 in T09. The resultant flux outputs were then compared with the corresponding linearly interpolated T09 fluxes. The following is true for all monthly outputs: the  $p\text{CO}_{2w}$  data were identical to six decimal places; 2% of the  $p\text{CO}_{2A}$  data elements

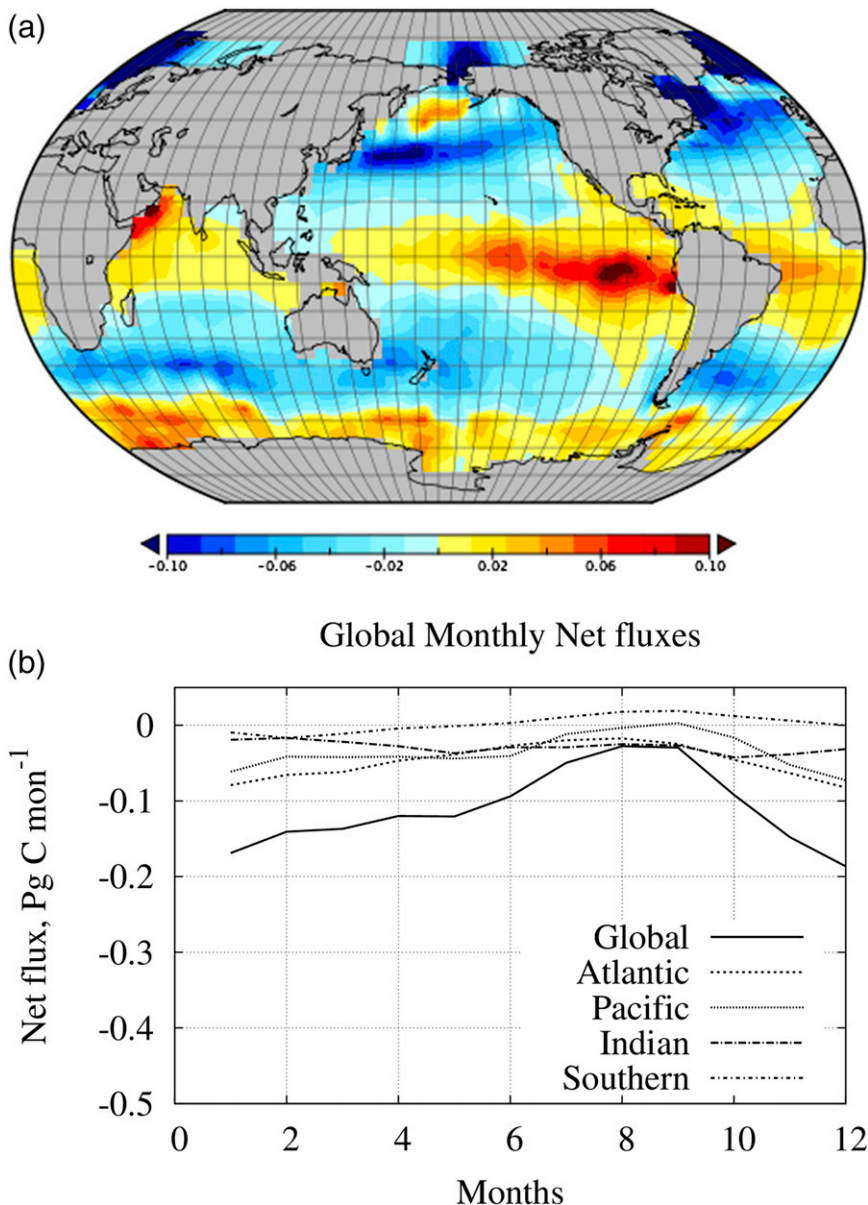


FIG. 2. Air–sea CO<sub>2</sub> fluxes for the year 2000 derived from using the T09 data for all inputs. (a) Annually averaged mean air–sea CO<sub>2</sub> flux per day ( $\text{gC m}^{-2} \text{day}^{-1}$ ). Regions of red and yellow are where CO<sub>2</sub> is leaving the ocean (outgassing), and regions of blue are where the ocean is absorbing CO<sub>2</sub> (ingassing). (b) Monthly air–sea CO<sub>2</sub> fluxes ( $\text{PgC month}^{-1}$ ). The projection for (a) has been chosen to allow for comparison with Fig. 13 in T09.

were  $>\pm 0.01\%$  different; 2% of the  $\alpha$  data elements were  $>\pm 3\%$  different; and 2% of the  $k$  data elements were  $>\pm 2\%$  different. These small differences within the calculations collectively results in 5%–12% (dependent upon the month; see Fig. 3) of the air–sea flux  $F$  data elements being  $>\pm 5\%$  different. The differences in these data fields between the output and that of the original linearly interpolated T09 dataset are likely to be a combination of (i) minor rounding differences in the

flux calculations (e.g., due to differences in precision between the FluxEngine calculations and those of the original T09 publication) and (ii) interpolation issues at boundaries. Figure 3 shows that the majority of the differences between the two air–sea flux datasets correspond to the sea ice boundary at high latitudes. Since we are comparing a linearly interpolated air–sea flux (originally on a  $4^\circ \times 5^\circ$  grid) with the output of  $1^\circ \times 1^\circ$  calculation (where all inputs are also  $1^\circ \times 1^\circ$ ), differences at

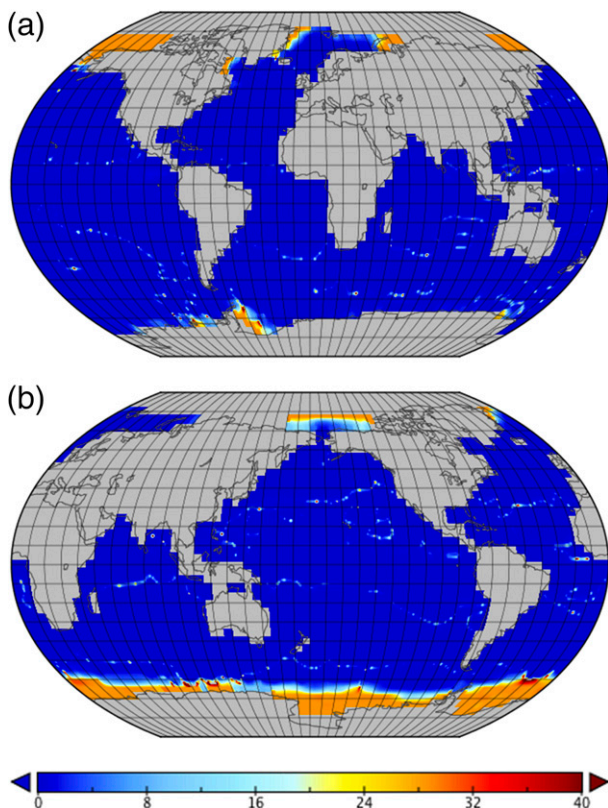


FIG. 3. Percentage differences in monthly air–sea  $\text{CO}_2$  flux between the linearly interpolated T09 air–sea  $\text{CO}_2$  flux data and the equivalent air–sea  $\text{CO}_2$  flux calculated using the linearly interpolated T09 data as input for (a) February (5% of differences explained in the text) and (b) October (12% of differences explained in the text).

these boundaries are expected. Using the integrated net flux utility on the output, the annual net integrated flux for 2000 was  $-1.33 \text{ GtCyr}^{-1}$  with an additional value of  $-0.16 \text{ GtCyr}^{-1}$  contribution from missing data, large lakes, the Mediterranean Sea, and coastal and shelf seas, giving a total of  $-1.49 \text{ GtCyr}^{-1}$ . This final result is within 5% of that derived from the original  $1^\circ \times 1^\circ$  linearly interpolated T09 data. Figure 2b shows the annual mean air–sea  $\text{CO}_2$  flux and the monthly net integrated fluxes for the four main oceanic basins.

#### 4. Scientific application

This section illustrates how the FluxEngine and resultant data can be used to study global and regional air–sea gas fluxes.

##### a. Global analyses and the Southern Ocean

A time series global analysis was performed to demonstrate the use of the FluxEngine. The linearly interpolated (to  $1^\circ \times 1^\circ$ ) T09 climatology data of  $X_{[\text{CO}_2]}$ ,

$p\text{CO}_{2w}$  (and associated SST), and Earth observation SST (SST skin) and  $U_{10}$  data, NCEP Climate Forecast System Reanalysis (CFRS) air pressure, and SSM/I percentage ice cover were used as the input to the FluxEngine (please see Table 1 for dataset specifics).

When the FluxEngine is configured to use the T09  $p\text{CO}_{2w}$  data and the chosen SST dataset is not from T09 (either due to studying a different year than the original T09 year 2000, or due to the selection of a different SST dataset), the  $p\text{CO}_{2w}$  data need to be reanalyzed to be consistent with this new SST dataset. Therefore, following previous studies (Fangohr et al. 2008; Kettle and Merchant 2005; Land et al. 2013), the FluxEngine re-analyses the T09  $p\text{CO}_{2w}$  data to correct them to the chosen SST datasets using the relationship provided by T09:

$$p\text{CO}_{2w} = p\text{CO}_{2w} \exp\{[0.0433(\text{SST}_C - T_C)] - 4.35 \times 10^{-5}(\text{SST}_C^2 - T_C^2)\}, \quad (8)$$

where  $T_C$  is the original temperature dataset ( $^\circ\text{C}$ ) and the subscript  $\text{SST}_C$  denotes the sea surface temperature ( $^\circ\text{C}$ ).

The FluxEngine was run for years 1995–2009 using Eq. (5) for the flux calculations and the gas transfer velocity of Nightingale et al. (2000). To generate an estimate of the uncertainties in the flux estimate due to the known uncertainties in the  $U_{10}$ ,  $p\text{CO}_{2w}$ , and SST input data, the FluxEngine time series processing was repeated with injected random noise enabled. The system assumes a  $1.5 \mu\text{atm yr}^{-1}$  change in the atmospheric and oceanic partial pressures of  $\text{CO}_2$  relative to the year 2000 of the original T09 data. Therefore, the partial pressure difference is imposed to have no interannual variation or trend, and therefore differences between years are due to other factors. Integrated net fluxes were then calculated using the T09 ice normalization. As an example output, Fig. 4 shows the February and August monthly air–sea  $\text{CO}_2$  flux maps for year 2000 and Fig. 5 shows the resulting year 2000 daily mean air–sea  $\text{CO}_2$  flux. The integrated net flux for year 2000 was  $-1.46 \text{ PgCyr}^{-1}$ . The annual global integrated net flux across all years falls in the range of  $-1.63$  to  $-1.00 \text{ PgCyr}^{-1}$  with a mean and standard deviation (interannual variability) of  $-1.22 \pm 0.21 \text{ PgCyr}^{-1}$ . These results are within the annual climatological global estimate estimated by T09 and fall within the range of recent estimates (Le Quéré et al. 2014). The time series of air–sea fluxes for the global oceans and the four main oceanic regions (Atlantic, Pacific, Southern, and Indian) can be seen in Fig. 6, and the International Hydrographic Office (IHO) oceanic region definitions are shown in Fig. 5c. The gray-shaded areas in Fig. 6 are the

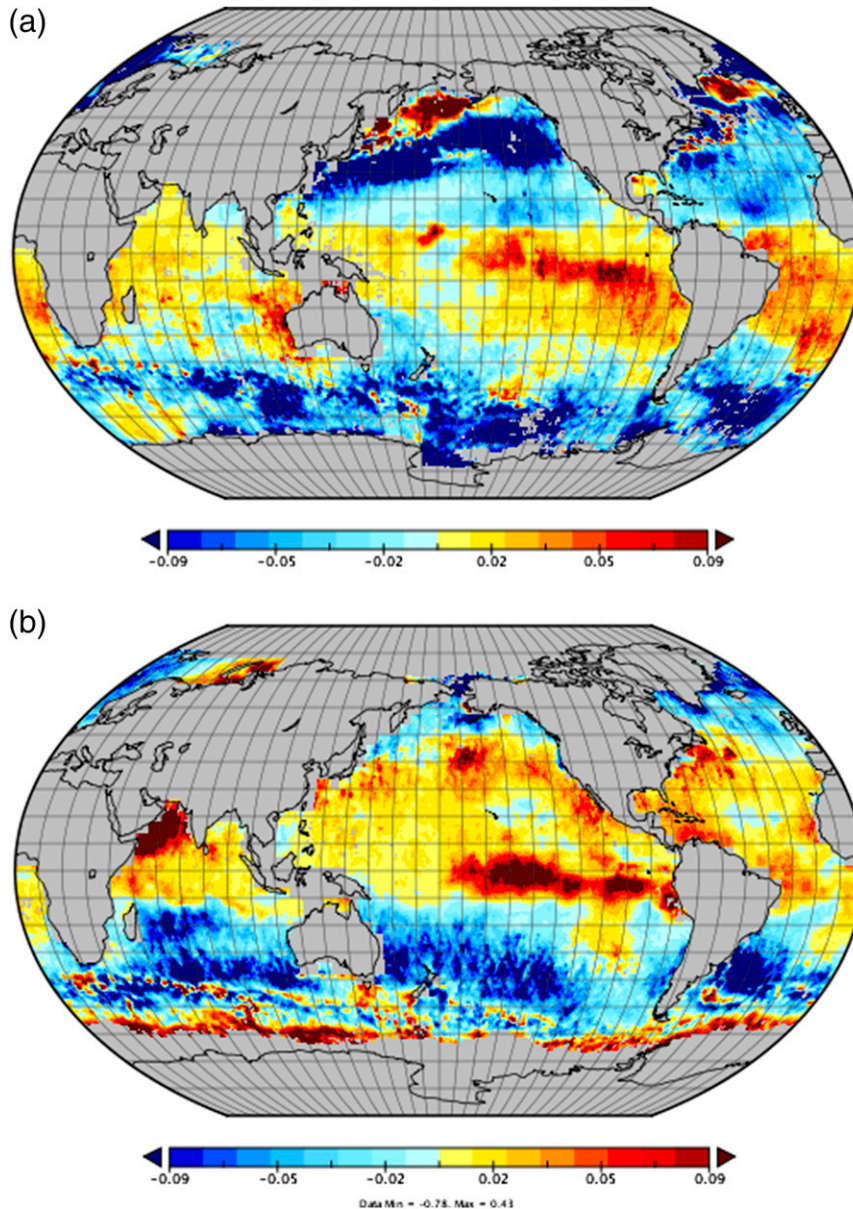


FIG. 4. Example monthly air-sea  $\text{CO}_2$  fluxes ( $\text{gCm}^{-2}\text{day}^{-1}$ ) for (a) February 2000 and (b) August 2000. The scale and projection have been chosen to allow easy comparison with Fig. 15 in T09.

uncertainties in the global air-sea flux estimates based on the known uncertainties of the  $U_{10}$ ,  $p\text{CO}_{2W}$ , and SST input data. The largest variability in the oceanic sink in the time series is during periods of large ENSO variation, 1997–2000 (strong positive ENSO followed by a strong negative ENSO; i.e., El Niño followed by La Niña) and 2007–08 (positive ENSO followed by negative ENSO), and Fig. 6 suggests that a large part of this variability is driven by the air-sea  $\text{CO}_2$  flux in the Pacific Ocean. The mean and standard deviation (interannual

variability) of the air-sea fluxes for the Atlantic, Pacific, Indian, and Southern Oceans were  $-0.42 \pm 0.05$ ,  $-0.28 \pm 0.15$ ,  $-0.30 \pm 0.05$ ,  $-0.04 \pm 0.02 \text{ PgC yr}^{-1}$ , respectively. These values are all comparable to their equivalent values in T09, although it must be noted that the ocean definitions in T09 vary slightly from those of the IHO. Clearly, the Southern Ocean contributes a relatively small amount to the net global flux, and the standard deviation shows that the Southern Ocean flux has a low temporal variability in comparison to the other



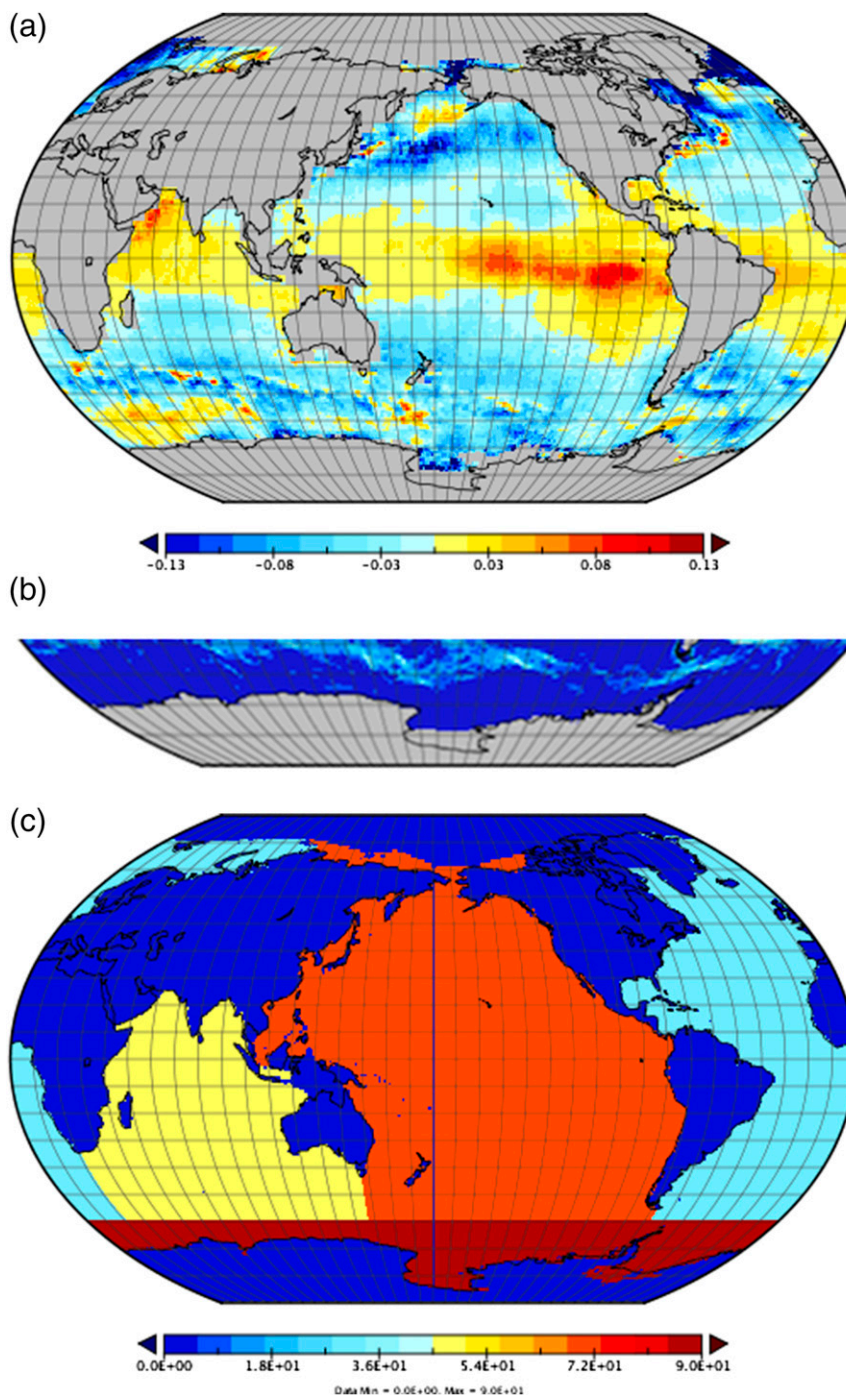


FIG. 5. Mean air-sea fluxes and process indicator layers. (a) Global annually averaged mean air-sea CO<sub>2</sub> flux per day ( $\text{gC m}^{-2} \text{day}^{-1}$ ) for 2000 as derived from using year 2000 EO data and the T09 pCO<sub>2</sub> climatology. (b) Persistent oceanic currents in Southern Hemisphere waters. (c) Oceanic basin descriptions as defined by the IHO.

oceanic basins. It is interesting to note that the T09 definition of the Southern Ocean extends farther north by  $10^\circ$  than the IHO definition and that this additional area encompasses regions of persistent ocean currents

(Fig. 5b). Recalculating the net air-sea CO<sub>2</sub> flux using the T09 Southern Ocean definition reduces the estimate of the temporal variability further, producing a net air-sea flux of  $-0.04 \pm 0.01 \text{ Pg yr}^{-1}$ .

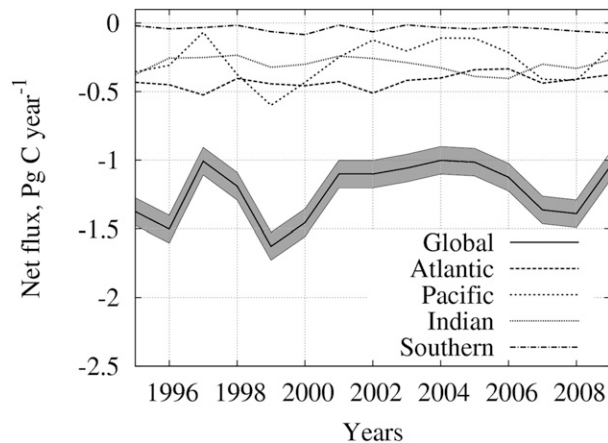


FIG. 6. Time series of global and regional mean air-sea  $\text{CO}_2$  fluxes ( $\text{PgC yr}^{-1}$ ). Regions are defined by the IHO oceanic basin descriptions as shown in Fig. 5c. The gray-shaded area represents the uncertainty in the global air-sea fluxes due to the known uncertainties in the input data.

### b. European shelf seas

Despite their relatively small area, accounting for just  $\sim 5\%$  of the World Ocean's surface, shelf seas play an important part in the global carbon cycle and in buffering human impacts on marine systems. These regions have a disproportionately large role in primary and new production, remineralization, and the sedimentation of organic matter [Chen et al. (2013) and the references therein]. The high biological activity in these regions can result in considerable drawdown of atmospheric  $\text{CO}_2$ , with the potential for the carbon to be exported to the deep ocean. A recent study (Chen and Borges 2009) has estimated that 29% of the global air-to-sea  $\text{CO}_2$  flux occurs in shelf seas. The North Sea is considered to be a sink for atmospheric  $\text{CO}_2$  (e.g., Frigstad et al. 2011; Thomas et al. 2004), but the ability of the entire northwestern European shelf to act as a sink of  $\text{CO}_2$ , and the variability of this sink, are areas of active research. Assuming negligible net burial rates of carbon in shelf sediments (de Haas et al. 2002), the net off-shelf carbon export will equal the region's net air-sea  $\text{CO}_2$  exchange. Therefore, estimating the net air-sea  $\text{CO}_2$  exchange in the European shelf seas can help quantify the carbon export from this shelf sea.

The global time series data described in section 4a were used to study the air-sea  $\text{CO}_2$  fluxes in the European shelf seas. Four different bathymetry-based definitions of the northwestern European shelf seas [ $<1000$ ,  $<500$ ,  $<200$ , and  $<200$  m plus the Norwegian Trench (NT)] were generated using Python and the General Bathymetric Chart of the Oceans, GEBCO\_08 grid, and were used for calculating the net flux. Figure 7

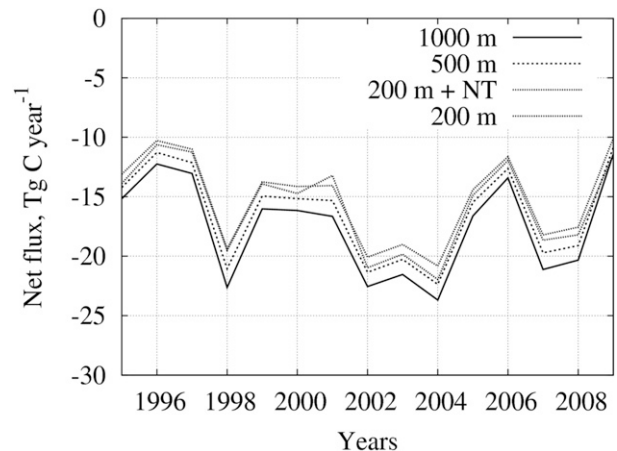


FIG. 7. Time series of mean air-sea  $\text{CO}_2$  fluxes for the northwestern European shelf sea ( $\text{TgC yr}^{-1}$ ). Each line represents a different region mask used to calculate the integrated net fluxes based on bathymetry and in one case the addition of the NT, where the depth values imply that the mask includes all areas that are less than the specified depth; i.e., the 1000-m mask contains all areas where the bathymetry is  $<1000$  m in depth.

shows the estimated net sink for each year and each region definition. The northwest European shelf sea-integrated net flux across all years and definitions falls in the range of  $-10.1$  to  $-23.7 \text{ TgC yr}^{-1}$  (Table 2). The limits of this range are set by using the 200 m ( $-10.1 \text{ TgC yr}^{-1}$ ) and 1000 m ( $-23.7 \text{ TgC yr}^{-1}$ ) bathymetry masks. The air-sea fluxes generated using the 200- and 1000-m masks differ by 13%–14%, assuming the 200-m flux as the reference. Despite the relatively coarse spatial resolution of these data, these estimates are comparable to previous in situ-based studies. A recent review and assimilation of the published literature on shelf sea and coastal air-sea fluxes from in situ data (Chen et al. 2013) estimates the northwest European shelf net air-sea  $\text{CO}_2$  flux to be  $-13.88 \text{ TgC yr}^{-1}$  (for an unknown year), which they estimate to be  $\sim 4\%$  of the global net flux due to estuaries and shelves. A recent modeling study (Wakelin et al. 2012) estimates the European shelf average long-term net air-sea  $\text{CO}_2$  flux to be  $-39.6 \text{ TgC yr}^{-1}$  based on a 16-yr average from a hydrodynamic-ecosystem model. There are differences in the definition of the northwestern European shelf between all of these studies, and the results in Table 2 illustrate that a precise definition is required. One possible reason for the differences between the model estimates and those presented here is likely to be the relatively coarse near-surface vertical resolution of the model. Each model surface grid cell will represent a volume of water that is typically between 0.1 and 2 m deep, dependent upon the underlying bathymetry (Shutler et al. 2011). This means that the model is

TABLE 2. Net fluxes of CO<sub>2</sub> for the European shelf seas calculated using a range of different shelf sea definitions and all years within the 1995–2009 time series. Negative values indicate a flux into the ocean (a sink).

Mask definition	Net integrated CO <sub>2</sub> flux (TgC yr <sup>-1</sup> )		
	Min	Max	Mean ± std dev
<200-m depth	-20.82	-10.09	-15.17 ± 3.70
<200-m depth + Norwegian Trench	-21.95	-10.63	-15.67 ± 3.83
<500-m depth	-22.36	-10.87	-16.39 ± 3.92
<1000-m depth	-23.69	-11.43	-17.51 ± 4.14

unlikely to be able to resolve any near-surface temperature gradients. In contrast the FluxEngine-derived fluxes presented here use satellite SST skin measurements that are observations of the temperature within a thin layer (500 μm) at the waterside of the air–sea interface (Donlon et al. 2002). The concentration of CO<sub>2</sub> is highly temperature dependent, so the lack of near-surface vertical resolution within the model could have a large impact on the calculated CO<sub>2</sub> air–sea gas fluxes.

### c. Underway in situ data

To demonstrate the flexibility of the system, the air–sea fluxes for a research cruise in the central equatorial Atlantic were calculated using the FluxEngine. In situ *f*CO<sub>2</sub> and associated SST data were downloaded from the community Surface Ocean CO<sub>2</sub> Atlas (SOCAT) website (v1.5, tropical Atlantic group) and then pre-processed into the format required by the FluxEngine (using the in situ tool; see the appendix). The FluxEngine was then used to calculate the air–sea CO<sub>2</sub> fluxes using the in situ *f*CO<sub>2</sub> data and the temporally and spatially corresponding EO data (the EO data sources are the same as used in sections 4a and 4b). Figure 8a shows the resultant gridded *f*CO<sub>2</sub> data for all SOCAT, v1.5, tropical Atlantic in situ *f*CO<sub>2</sub> data, and Fig. 8b shows the gridded in situ *f*CO<sub>2</sub> data from a single cruise (Bakker 2014) in the tropical Atlantic. Figure 8c shows Earth observation–derived mean gas transfer velocity for October 2000, and Fig. 8d shows the resultant air–sea CO<sub>2</sub> fluxes from using Fig. 8b as the input *f*CO<sub>2</sub> data. The fluxes can be seen to vary along the cruise track between a source (positive) and a sink (negative) of CO<sub>2</sub>. The missing gas transfer velocity data (causing a hiatus in the estimates of air–sea fluxes) are due to missing data in the Earth observation sea surface temperature dataset.

### d. Benchmarking

The FluxEngine was run on a single cloud node (2.4-GHz Intel Xeon, 3 GB of memory), and the time taken to process a single-year dataset was determined. All input datasets were quality filtered and converted to a suitable

netCDF format in advance of this analysis. A 1-yr climatology of fluxes at 1° × 1° spatial resolution took 40 min (total process time) to complete. Disabling the process indicator layer output reduced this time to 25 min. Running a 5-yr time series (with the indicator layers off) on the single cloud node took 2 h [i.e., ~25 × 5 yr (60 min)<sup>-1</sup> total process time], whereas repeating the same 5-yr time series processing across five nodes using GoGo list took 25 min (as each year was executed on an independent node and so executed simultaneously). Calculating the integrated net fluxes from the 1-yr climatology netCDF files took 10 min (total process time). So, the total end-to-end time for calculating a 1-yr climatology and then calculating the integrated net fluxes using one node (with all indicator layers turned on) was 50 min.

## 5. Future developments

The FluxEngine has been developed to allow the study of CO<sub>2</sub> and the current version uses the T09 climatology *p*CO<sub>2W</sub> data for the waterside component of the calculation. To increase the versatility of the system, work is ongoing to extend the toolbox to use the community SOCAT datasets (e.g., Pfeil et al. 2013; Bakker et al. 2014). There are many other climatically important gases, including nitrous oxide (N<sub>2</sub>O) and methane (CH<sub>4</sub>). Partial support for these gases exists within the toolbox, as N<sub>2</sub>O, CH<sub>4</sub>, and CO<sub>2</sub> are all poorly soluble gases, so their gas transfer velocity parameterizations can be considered as interchangeable with that for CO<sub>2</sub>. Therefore, the toolbox can be used to generate maps of gas transfer velocity to enable air–sea fluxes of N<sub>2</sub>O and CH<sub>4</sub> to be studied. Similarly, *k* for the gas DMS can be considered to be the direct component (i.e., nonbubble component) of a *k* CO<sub>2</sub> parameterization, and the toolbox already includes two methods for deriving this direct component. The generic nature of the gas calculation and parameterization lends itself to being extended to determine air–sea fluxes for other gases. For each additional gas, a gas transfer velocity parameterization and an in situ dataset or climatology of in-water concentrations, partial pressures or fugacity, and suitable solubility equations are required. After CO<sub>2</sub> the next largest gas climatology (in terms of the number of in situ data points) is the Lana et al. (2010) DMS climatology, and so future extensions of the FluxEngine are likely to include the addition of a DMS capability through exploiting published work (e.g., Goddijn-Murphy et al. 2012; Johnson 2010; Lana et al. 2011). This will involve exploiting (or linking in) the open source code of Johnson (2010), which can be used to calculate gas transfer velocities. The full capability to calculate air–sea fluxes for N<sub>2</sub>O, CH<sub>4</sub>, and other gases is not currently possible as the in-water data

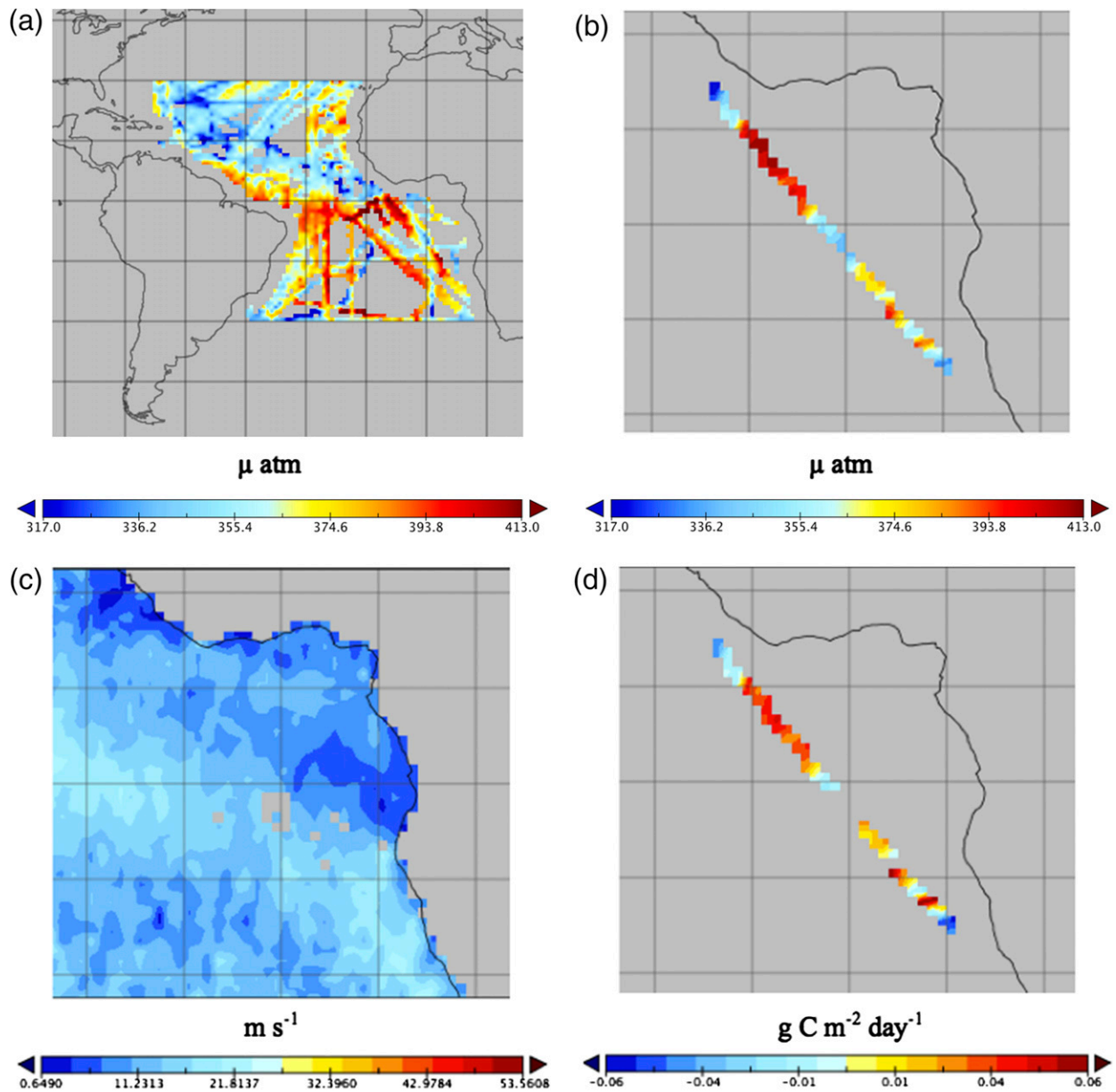


FIG. 8. SOCAT  $f\text{CO}_2$  data processed using the FluxEngine: (a) all of the SOCAT v1.5 tropical Atlantic  $f\text{CO}_2$  data binned into a common grid, (b) the SOCAT v1.5 tropical Atlantic  $f\text{CO}_2$  data for one cruise (Bakker 2014) in October 2000 (west of the African coast), (c) the EO-derived gas transfer velocity for October 2000, and (d) the resultant air-sea  $\text{CO}_2$  fluxes corresponding to the  $f\text{CO}_2$  data in (b).

collections for these gases are still in their infancy, but efforts have begun to collate such datasets, for example, the Marine Methane and Nitrous Oxide (MEMENTO) database (Bange 2006).

The toolbox currently uses climatological salinity data from the T09 climatology. Recent advancements in satellite Earth observation have seen the launch of two sensors that can measure surface salinity from space. These are NASA's *Aquarius* and ESA's Soil Moisture Ocean Salinity (SMOS) missions, and future work will

enable SMOS salinity data to be used within the toolbox. A web interface is also being developed that will enable users to create configuration files, execute processing, and download the resulting output.

## 6. Conclusions

A flexible air-sea  $\text{CO}_2$  data processing toolbox called the FluxEngine has been developed and presented. The flux calculation itself is user configurable, and the

outputs have been extensively evaluated and compared with reference datasets. No specialist knowledge is required to use the toolbox, and it is based on standard software tools and packages that require no licenses. It is currently installed and running on the Nephelae Cloud at the Institut Français de Recherche pour l'Exploitation de la Mer (Ifremer), where >8 TB of climate-quality data can be used as input to the flux calculations. The use of cloud-computing approaches means that the data processing is scalable, and this feature is completely transparent to the user. Here we have used the toolbox to estimate the 15-yr-average net air–sea flux of CO<sub>2</sub> for the global oceans (including shelf seas and coastal zones), the four main oceanic basins, and the European shelf seas. We have shown how subtle differences in the definitions of the European shelf seas can cause differences of >10% in the calculated annual net fluxes. Similarly differences in the Southern Ocean definition can have an impact on the calculated air–sea flux. We therefore urge the scientific community to use a common set of oceanic region definitions, to allow the outputs from differing studies to be easily compared and contrasted. The FluxEngine provides a mechanism for this and its open source nature allows the scientific community to freely exploit the toolbox. It is hoped that the FluxEngine will help to improve the transparency and traceability of results from air–sea gas flux studies. The FluxEngine was originally developed for the ESA OceanFlux Greenhouse Gases project, and it is currently being used to produce air–sea gas flux climatologies and to study method and data uncertainties. Users can access the version of the toolbox installed on the Nephelae Cloud on the OceanFlux Greenhouse Gases project website (<http://www.oceanflux-ghg.org>); alternatively, the open-source software is available in GitHub (<https://github.com/oceanflux-ghg/FluxEngine>).

*Acknowledgments.* This work was funded by the European Space Agency (ESA) Support to Science Element (STSE) through the OceanFlux Greenhouse Gases project (Contract 4000104762/11/I-AM) and the U.K. NERC Carbon and Nutrient Dynamics and Fluxes over Shelf Systems (CANDYFLOSS) project (Contract NE/K002058/1). The Surface Ocean CO<sub>2</sub> Atlas (SOCAT) is an international effort, supported by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS), and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), to deliver a uniformly quality-controlled surface ocean CO<sub>2</sub> database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT.

## APPENDIX

### The FluxEngine Tools, Data, and Generic Gas Transfer Parameterization

#### a. Tools

Table A1 lists the software utilities that are available in the FluxEngine toolbox. Specific details on some of the tools are given in the following sections.

##### 1) OFLUXGHG-FLUX-BUDGETS.PY

The utility calculates integrated net fluxes ( $F_{IN}$ ) over a given region from the monthly mean flux and ice cover data as follows. The input netCDF files are assumed to contain monthly mean net flux  $F$ , ice cover, gas transfer velocity, in-water CO<sub>2</sub> concentration, and interfacial CO<sub>2</sub> concentration. A high-spatial-resolution land mask is required and a region definition file if nonglobal regions are to be analyzed. The four main oceanic regions (Atlantic, Pacific, Indian, and Southern Oceans), as defined by the IHO, are provided within the monthly netCDF output (within the process indicator data layers), so these can be used as the region definitions for this utility if desired. A high-spatial-resolution land mask is also provided within the toolbox and its use is described below.

The method for the integrated net flux tool is now described. At each pixel or data element, the integrated net flux is initially calculated from  $F$  and the pixel's total area, which is calculated assuming the earth to be an oblate spheroid. Next, we need to account for the effect of ice, and two methods of ice normalization are available. The first is from T09, which specifies that if ice cover within a data element is <10%, then it has a negligible effect on the integrated net flux and so the ice cover value is assumed to be 0%. Data elements with ice cover >90% are set to 90% to account for leads, polynyas, etc. The net flux for each data element where the ice cover is >0% is then reduced linearly by the percentage of ice cover value. The second ice normalization method (Loose et al. 2009) posits that the integrated net flux from partially ice-covered regions is greater than would be expected from proportionality with ice-free fraction. Here, the integrated net flux is proportional to (ice-free fraction)<sup>0.4</sup>. This implies that the integrated net flux from a pixel with 90% ice cover (the maximum assumed by T09) is actually responsible for 40% of the ice-free integrated net flux. This results in a quadrupling of the estimated net flux from regions of fast ice over that assumed by the T09 method. Users are able to choose which method they prefer to use.

The data are at a relatively coarse spatial resolution of 1° × 1° (which at the equator is ~111 km × ~111 km). Therefore, to calculate the contribution of

each data element to the regional or global integrated net flux, we need to know the proportion of ocean (whether ice covered or not) actually contained within each data element. For most elements this will be 1 (open ocean or sea ice) or 0 (land), but it may be intermediate in data elements that cover region boundaries (e.g., either oceanic boundaries or coastal regions). The oceanic (nonland) proportion of the pixel is multiplied by the ice-corrected integrated net flux to give the contribution to the regional or global integrated net flux.

Where  $F$  data are missing (termed the missing integrated net flux), a first-order correction is made in order not to underestimate the regional net flux. As  $F$  is calculated from remotely sensed SST and wind speed, areas with significant ice cover, persistent cloud cover, or some coastal regions are likely to have missing  $F$ . We sum the ice-corrected ocean area of such data elements contained within the region of interest and use the areas with measured  $F$  to calculate a regional average  $F$ , which is then multiplied by the missing area to give an estimate of the integrated net flux from the missing regions. This is added to the integrated net flux to give an estimate of the total regional integrated net flux. Note that the flux from individual regions may not exactly sum to the global region flux using this method, since in one case the regional mean flux is used to estimate the missing flux, while in the other case the global mean flux is used.

The net flux tool also provides gross fluxes and the average values for all spatially varying variables within the netCDF files. The gas transfer velocity  $k$ , in-water  $\text{CO}_2$  concentration ( $[\text{CO}_{2\text{AQW}}]$ ), and interfacial  $\text{CO}_2$  concentration ( $[\text{CO}_{2\text{AQ0}}]$ ) are used to calculate the upward and downward integrated gross fluxes. The upward gross flux,  $F_{\text{UG}}$ , is defined as  $k[\text{CO}_{2\text{AQW}}]$  and the downward gross flux,  $F_{\text{DG}}$ , is defined as  $k[\text{CO}_{2\text{AQ0}}]$ . Missing data are treated in the same way for these calculations as for integrated net flux.

When using the net flux utility, each selected region has its own output data file. Data are output for each month of each year for which input data are supplied, and the annual totals are provided for each year. Net flux outputs include  $F_{\text{IN}}$  (integrated net flux based on calendar days), missing integrated net flux, and integrated net flux assuming a 30.5-day month, along with similar values for  $F_{\text{UG}}$  and  $F_{\text{DG}}$ .

## 2) RESAMPLE\_NETCDF.PY

This tool calculates the mean for each  $5^\circ \times 4^\circ$  grid cell using the corresponding  $1^\circ \times 1^\circ$  grid cells in the input data. No correction due to variations in area (between

TABLE A1. The utilities available within the FluxEngine toolbox.

Tool name	Purpose
offluxghg-run-climatology.pl	Main program to calculate monthly air-sea fluxes
offluxghg-flux-calc.py	Main air-sea flux calculation
offluxghg-flux-budgets.py	Calculates integrated net fluxes
resample_netcdf.py	Resamples the $1^\circ \times 1^\circ$ data to a $5^\circ \times 4^\circ$ grid
txt2ncdf.py	Convert in situ data to a netCDF file

the  $5^\circ \times 4^\circ$  and the  $1^\circ \times 1^\circ$  grids) due to interactions with land is considered and any missing (masked)  $1^\circ \times 1^\circ$  data are not used in the calculation. If all data in the  $1^\circ \times 1^\circ$  grid cells are masked (missing values), then the output  $5^\circ \times 4^\circ$  grid cell is also masked. The tool can output the regridded fields as netCDF or as a single ASCII comma-separated variable (CSV) file. The output file will simply reflect the contents of the input netCDF with each data array (2D dataset) replaced by its  $5^\circ \times 4^\circ$  equivalent, while the optional CSV file contains columns corresponding to latitude, longitude, and all 2D datasets found in the input file.

## 3) TXT2NCDF.PY

This utility assumes that the input data are in CSV format, with headings that include “latitude,” “longitude,” and optionally “date” or “date/time” on the first line, and corresponding values on the subsequent lines. Latitude and longitude must be in decimal degrees, while date or date/time must follow the format DD/MM/(YY)YY [hh:mm:ss], where parentheses indicate options and DD is the two-digit day number within the month, MM is the two-digit month number (starting at 01) and (YY)YY is the two- or four-digit year. The resulting output data are in netCDF format, with latitude and longitude limits copied from a reference global netCDF file that is passed as one of the inputs. The output netCDF includes a dataset “count” containing the number of observations found in each grid cell. The user can optionally specify a startTime and endTime (in the same range of formats as date) and then the tool will only count observations on or after startTime and/or before endTime. If the end time is only a date with no time specified, it is assumed to be inclusive; that is, the end time is the end of the specified day. Hence, using the options of startTime “01/01/08” and endTime “31/01/08” would only count observations in January 2008. Any data are binned into each grid cell and no interpolation of the data is performed. The “time” value in the netCDF is set as follows: If any data contain date/time information, then the time is set to the midpoint of the

range of valid times (i.e., times corresponding to data that have been counted). Otherwise, if startTime and end-Time are both set, time is set to the midpoint of these. In all other cases, time is taken from the reference netCDF.

### b. Input data currently available

Table 1 gives an overview of the available input datasets.

### c. Generic gas transfer parameterization

Several published gas transfer velocity parameterizations (e.g., as used in T09) are of the form

$$k_w = aU^2(\text{Sc}/660)^{-1/2}. \quad (\text{A1})$$

The toolbox provides a more generic polynomial expression that enables a large range of different wind speed-based gas transfer velocity parameterizations to be used (Nightingale et al. 2000; Wanninkhof et al. 2009), for example,

$$k_w = (\text{Sc}/660)^{-1/2}(a_0 + a_1U + a_2U^2 + a_3U^3 + a_4U^4). \quad (\text{A2})$$

This parameterization allows users to exploit their own wind speed-based gas transfer relationship.

## REFERENCES

- Bakker, D., 2014: Underway physical oceanography and carbon dioxide measurements during POLARSTERN cruise ANT-XVIII/1. School of Environmental Sciences, University of East Anglia, accessed 17 March 2016, doi:10.1594/PANGAEA.812023.
- , and Coauthors, 2014: An update to the Surface Ocean CO<sub>2</sub> Atlas (SOCAT version 2). *Earth Syst. Sci. Data*, **6**, 69–90, doi:10.5194/essd-6-69-2014.
- Bange, H. W., 2006: The importance of oceanic nitrous oxide emissions. *Atmos. Environ.*, **40**, 198–199, doi:10.1016/j.atmosenv.2005.09.030.
- Chen, C.-T. A., and A. V. Borges, 2009: Reconciling opposing views on carbon cycling in the coastal ocean: Continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO<sub>2</sub>. *Deep-Sea Res. II*, **56**, 578–590, doi:10.1016/j.dsr2.2009.01.001.
- , T.-H. Huang, Y.-C. Chen, Y. Bai, X. He, and Y. Kang, 2013: Air–sea exchanges of CO<sub>2</sub> in the world’s coastal seas. *Biogeosciences*, **10**, 6509–6544, doi:10.5194/bg-10-6509-2013.
- de Haas, H., T. C. E. van Weering, and H. de Stigter, 2002: Organic carbon in shelf seas: Sinks or sources, processes and products. *Cont. Shelf Res.*, **22**, 691–717, doi:10.1016/S0278-4343(01)00093-0.
- Donlon, C. J., P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369, doi:10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2.
- , M. Martin, J. D. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158, doi:10.1016/j.rse.2010.10.017.
- Fangohr, S., and D. K. Woolf, 2007: Application of new parameterizations of gas transfer velocity and their impact on regional and global CO<sub>2</sub> budgets. *J. Mar. Syst.*, **66**, 195–203, doi:10.1016/j.jmarsys.2006.01.012.
- , —, C. D. Jeffery, and I. S. Robinson, 2008: Calculating long-term global air–sea flux of carbon dioxide using scatterometer, passive microwave, and model re-analysis wind data. *J. Geophys. Res.*, **113**, C09032, doi:10.1029/2005JC003376.
- Frigstad, H., T. Andersen, D. O. Hessen, L.-J. Naustvoll, T. M. Johnsen, and R. G. J. Bellerby, 2011: Seasonal variation in marine C:N:P stoichiometry: Can the composition of seston explain stable Redfield ratios? *Biogeosciences*, **8**, 2917–2933, doi:10.5194/bg-8-2917-2011.
- Goddijn-Murphy, L., D. K. Woolf, and C. A. Marandino, 2012: Space-based retrievals of air–sea gas transfer velocity using altimeters: Calibration for dimethyl sulfide. *J. Geophys. Res.*, **117**, C08028, doi:10.1029/2011JC007535.
- , —, C. Bertrand, and P. Queffelec, 2013: Improvements to estimating the air–sea gas transfer velocity by using dual-frequency, altimeter backscatter. *Remote Sens. Environ.*, **139**, 1–5, doi:10.1016/j.rse.2013.07.026.
- Huffman, G. J., D. T. Bolvin, and R. F. Adler, 2012: Global Precipitation Climatology Project monthly product v2.2. National Climatic Data Center, accessed 17 March 2016. [Available online at <http://www.ncdc.noaa.gov/wdc/wdcamet-ncdc.html>.]
- Johnson, M. T., 2010: A numerical scheme to calculate temperature and salinity dependent air–water transfer velocities for any gas. *Ocean Sci.*, **6**, 913–932, doi:10.5194/os-6-913-2010.
- Kettle, H., and C. J. Merchant, 2005: Systematic errors in global air–sea CO<sub>2</sub> flux caused by temporal averaging of sea-level pressure. *Atmos. Chem. Phys.*, **5**, 1459–1466, doi:10.5194/acp-5-1459-2005.
- Lana, A., and Coauthors, 2011: An updated climatology of surface dimethylsulfide concentrations and emission fluxes in the global ocean. *Global Biogeochem. Cycles*, **25**, GB1004, doi:10.1029/2010GB003850.
- Land, P. E., J. D. Shutler, R. D. Cowling, D. K. Woolf, P. Walker, H. S. Findlay, R. C. Upstill-Goddard, and C. J. Donlon, 2013: Climate change impacts on air–sea fluxes of CO<sub>2</sub> in three Arctic seas: As sensitivity study using Earth observation. *Biogeosciences*, **10**, 8109–8128, doi:10.5194/bg-10-8109-2013.
- Le Quéré, C., and Coauthors, 2009: Trends in the sources and sinks of carbon dioxide. *Nat. Geosci.*, **2**, 831–836, doi:10.1038/ngeo689.
- , and Coauthors, 2014: Global carbon budget 2013. *Earth Syst. Sci. Data*, **6**, 235–263, doi:10.5194/essd-6-689-2013.
- , and Coauthors, 2015: Global carbon budget 2014. *Earth Syst. Sci. Data*, **7**, 47–85, doi:10.5194/essd-7-47-2015.
- Loose, B., W. R. McGillis, P. Schlosser, D. Perovich, and T. Takahashi, 2009: Effects of freezing, growth, and ice cover on gas transport processes in laboratory seawater experiments. *Geophys. Res. Lett.*, **36**, L05603, doi:10.1029/2008GL036318.
- Maritorea, S., O. H. F. d’Andon, M. Antoine, and D. A. Siegel, 2010: Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sens. Environ.*, **114**, 1791–1804, doi:10.1016/j.rse.2010.04.002.
- McGillis, W. R., and R. Wanninkhof, 2006: Aqueous CO<sub>2</sub> gradients for air–sea flux estimates. *Mar. Chem.*, **98**, 100–108, doi:10.1016/j.marchem.2005.09.003.
- , J. B. Edson, J. E. Hare, and C. W. Fairall, 2001: Direct covariance air–sea CO<sub>2</sub> fluxes. *J. Geophys. Res.*, **106**, 16 729–16 745, doi:10.1029/2000JC000506.

- Merchant, C. J., and Coauthors, 2012: A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. *J. Geophys. Res.*, **117**, C12013, doi:10.1029/2012JC008400.
- Nightingale, P., G. Malin, C. Law, A. J. Watson, P. L. Liss, M. Liddicoat, J. Boutin, and R. U. Goddard, 2000: In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Global Biogeochem. Cycles*, **14**, 373–387, doi:10.1029/1999GB900091.
- Pfeil, B., and Coauthors, 2013: A uniform, quality controlled Surface Ocean CO<sub>2</sub> Atlas (SOCAT). *Earth Syst. Sci. Data*, **5**, 125–143, doi:10.5194/essd-5-125-2013.
- Sabine, C. L., and Coauthors, 2004: The oceanic sink for anthropogenic CO<sub>2</sub>. *Science*, **305**, 367–371, doi:10.1126/science.1097403.
- Shutler, J. D., and Coauthors, 2011: Evaluating the ability of a hydrodynamic ecosystem model to capture inter- and intra-annual spatial characteristics of chlorophyll-*a* in the north east Atlantic. *J. Mar. Syst.*, **88**, 169–182, doi:10.1016/j.jmarsys.2011.03.013.
- Takahashi, T., and Coauthors, 2002: Global air–sea CO<sub>2</sub> flux based on climatological surface ocean *p*CO<sub>2</sub> and seasonal biological and temperature effects. *Deep-Sea Res. II*, **49**, 1601–1622, doi:10.1016/S0967-0645(02)00003-6.
- , and Coauthors, 2009: Climatological mean and decadal change in surface ocean *p*CO<sub>2</sub> and net sea–air CO<sub>2</sub> flux over the global oceans. *Deep-Sea Res. II*, **56**, 554–577, doi:10.1016/j.dsr2.2008.12.009.
- Thomas, H., Y. Bozec, K. Elkalay, and H. J. W. de Baar, 2004: Enhanced open ocean storage of CO<sub>2</sub> from shelf sea pumping. *Science*, **304**, 1005–1008, doi:10.1126/science.1095491.
- Wakelin, S., J. T. Holt, J. Blackford, J. I. Allen, M. Butenschön, and Y. Artioli, 2012: Modeling the carbon fluxes of the northwest European continental shelf: Validation and budgets. *J. Geophys. Res.*, **117**, 05020, doi:10.1029/2011JC007402.
- Wanninkhof, R., 1992: Relationship between wind speed and gas exchange over the ocean. *J. Geophys. Res.*, **97**, 7373–7382, doi:10.1029/92JC00188.
- , W. E. Asher, D. T. Ho, C. Sweeney, and W. R. McGillis, 2009: Advances in quantifying air-sea gas exchange and environmental forcing. *Annu. Rev. Mar. Sci.*, **1**, 213–244, doi:10.1146/annurev.marine.010908.163742.
- Watson, A. J., and Coauthors, 2009: Tracking the variable North Atlantic sink for atmospheric CO<sub>2</sub>. *Science*, **326**, 1391–1393, doi:10.1126/science.1177394.
- Weiss, R. F., and B. A. Price, 1980: Nitrous oxide solubility in water and seawater. *Mar. Chem.*, **8**, 347–359, doi:10.1016/0304-4203(80)90024-9.