A flexible processing system for the calculation of air-sea gas fluxes: a dynamic tool for the community

Jamie Shutler, Jean-Francois Piollé, Peter Land, David Woolf, Lonneke Goddijn-Murphy, Frederic Paul, Fanny Girard-Ardhuin,Craig Donlon, Bertrand Chapron

oceanflux ghg
support to science element

OceanFlux GHG is funded by: esa and affiliated to: solas

eri environmental research institute
PML Plymouth Marine Laboratory
HERIOT WATT UNIVERSITY
Ifremer
National Oceanography Centre NATURAL ENVIRONMENT RESEARCH COUNCIL
stse support to science element

Introduction

- Keen for Earth observation data to be exploited for SOLAS studies
- Issues:
  - Simplifying access to these data (e.g. standard file types and tools)
  - Ability to handle large data sets

- Flexible system needed for project uncertainty and scientific analyses
- Simple to provide community access.
  - Greater transparency and traceability for publications

- Users can create own climatologies, generate net fluxes and re-grid data
- Data can be ingested into standard software packages and tools (e.g. Matlab, IDL, Excel).

Cloud computing

Cloud computing: inter-connected 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).

Features:

- Redundancy
  - Servers can be removed or upgraded without users noticing
- Scale-able
  - Uses standard hardware and software
- Simple backup and restoration
- No specific skills required by user
- Speed
  - Processing is close to data (no bottleneck)
- System is tailored
  - No reliance on physical hardware e.g. use of virtual servers
- Maximises use of resources
  - Dynamic re-allocation of resources as and when needed

Cloud computing

Cloud computing: inter-connected 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).

Features:

- Redundancy
  - Servers can be removed or upgraded without users noticing
- Scale-able
  - Uses standard hardware and software
- Simple backup and restoration
- No specific skills required by user
- Speed
  - Processing is close to data (no bottleneck)
- System is tailored
  - No reliance on physical hardware e.g. use of virtual servers
- Maximises use of resources
  - Dynamic re-allocation of resources as and when needed

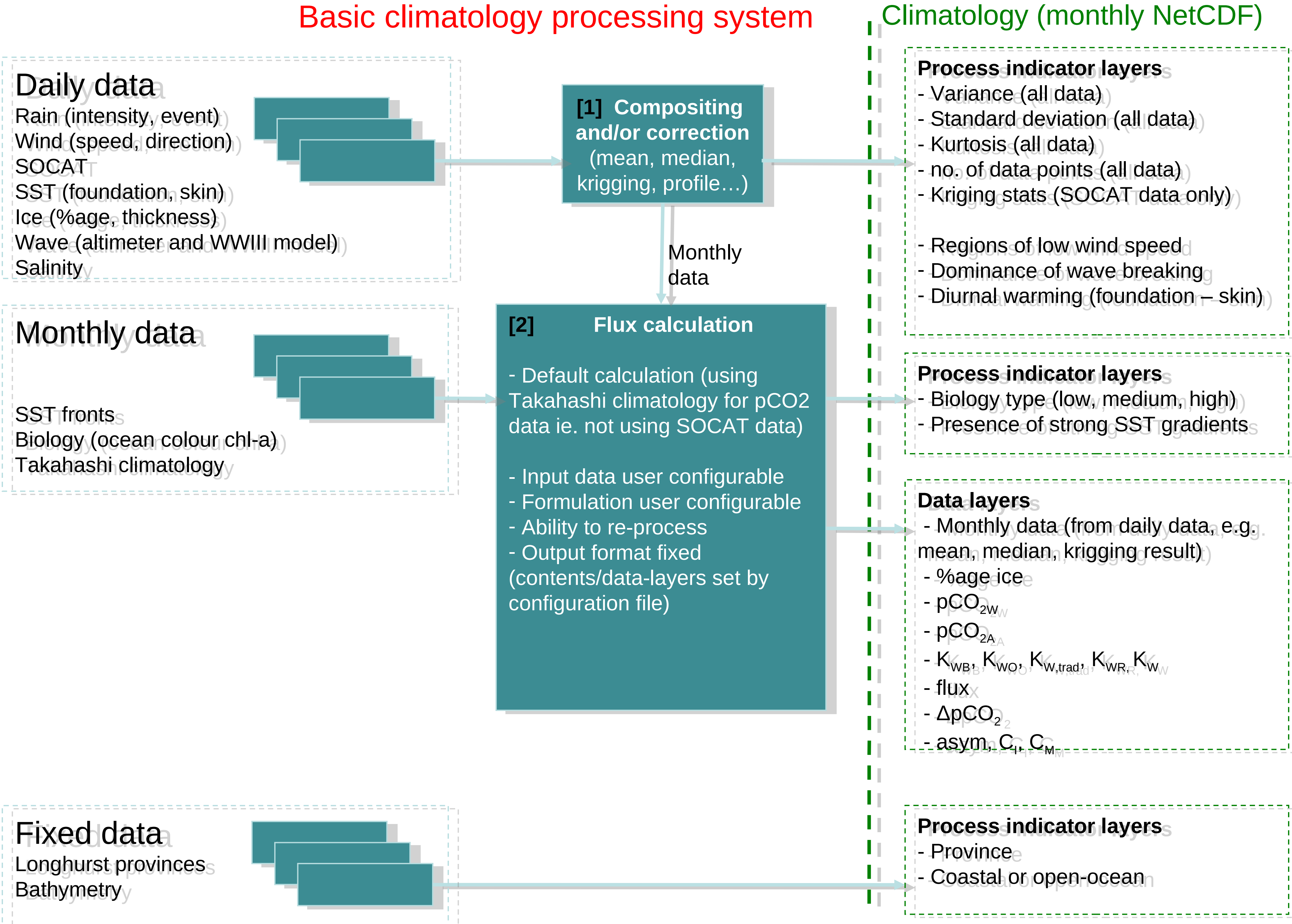
Nephelae

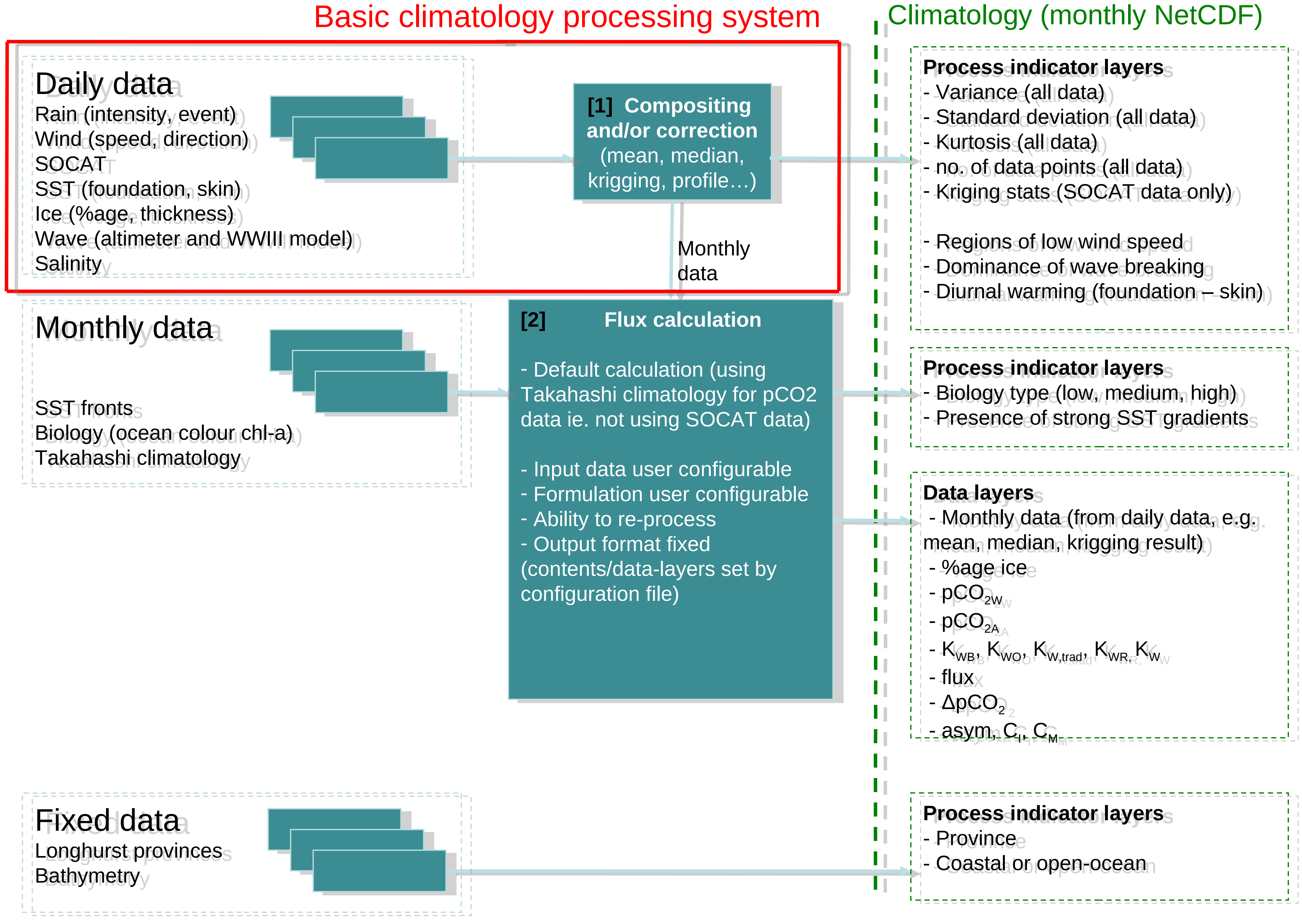
600 processing nodes.
1.5 Peta bytes of storage

Available input data

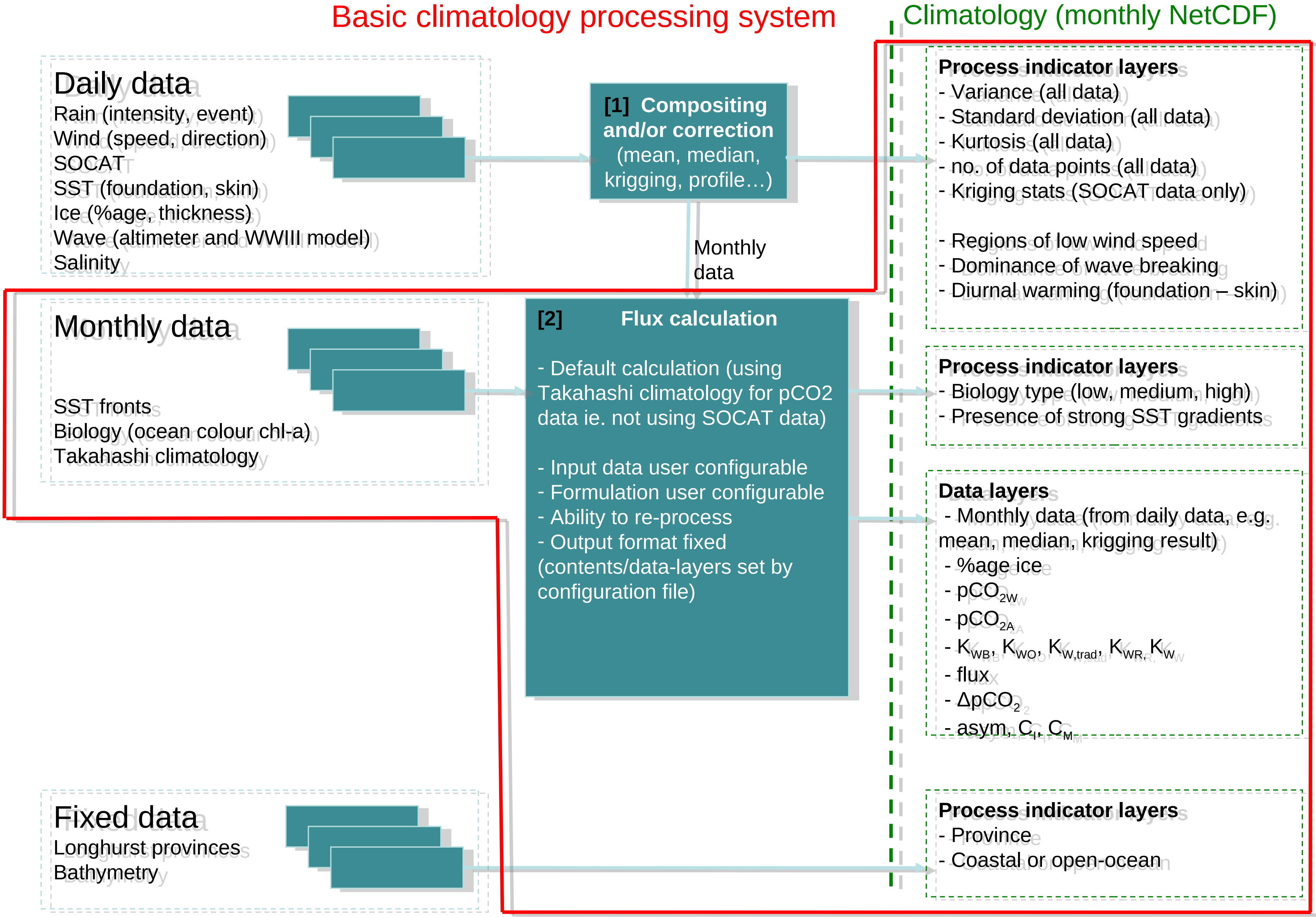
- All input data are 1º x 1º global monthly composites.

Parameter	Dataset (current number of datasets)	Uncertainty	Years
SST	NOAA AVHRR, ESA CCI, GHRSST datasets (4)	yes	1992-2010
U10	ESA GlobWave archive (2+)	yes	1992-2010
Hs	ESA GlobWave archive (1+)	yes	1992-2010
pCO2/fCO2	LDEO Takahashi 2002, LDEO Takahashi 2009, OceanFlux/SOCATv1.5, OceanFlux/SOCATv2 (4)	majority	2000 2010
Rain	GPCP, TRMM, SSMI (4)	yes	1992-2012
Chl-a	ESA GlobColour, ESA Ocean Colour CCI (2)	yes	1997-2011


Basic climatology processing system
Climatology (monthly NetCDF)
Daily data
Rain (intensity, event)
Wind (speed, direction)
SOCAT
SST (foundation, skin)
Ice (%age, thickness)
Wave (altimeter and WWIII model)
Salinity
[1] Compositing and/or correction
(mean, median, kriging, profile…)
Monthly data
Monthly data
SST fronts
Biology (ocean colour chl-a)
Takahashi climatology
[2] Flux calculation
- Default calculation (using Takahashi climatology for pCO2 data ie. not using SOCAT data)
- Input data user configurable
- Formulation user configurable
- Ability to re-process
- Output format fixed (contents/data-layers set by configuration file)
Fixed data
Longhurst provinces
Bathymetry
Process indicator layers
- Variance (all data)
- Standard deviation (all data)
- Kurtosis (all data)
- no. of data points (all data)
- Kriging stats (SOCAT data only)
- Regions of low wind speed
- Dominance of wave breaking
- Diurnal warming (foundation – skin)
Process indicator layers
- Biology type (low, medium, high)
- Presence of strong SST gradients
Data layers
- Monthly data (from daily data, e.g. mean, median, krigging result)
- %age ice
- pCO_{2W}
- pCO_{2A}
- K_{WB}, K_{WO}, $K_{W,trad}$, K_{WR}, K_W
- flux
- ΔpCO_2
- asym, C_I, C_M
Process indicator layers
- Province
- Coastal or open-ocean



Basic climatology processing system
Climatology (monthly NetCDF)
Daily data
Rain (intensity, event)
Wind (speed, direction)
SOCAT
SST (foundation, skin)
Ice (%age, thickness)
Wave (altimeter and WWIII model)
Salinity
[1] Compositing and/or correction (mean, median, krigging, profile…)
Monthly data
Monthly data
SST fronts
Biology (ocean colour chl-a)
Takahashi climatology
[2] Flux calculation
- Default calculation (using Takahashi climatology for pCO2 data ie. not using SOCAT data)
- Input data user configurable
- Formulation user configurable
- Ability to re-process
- Output format fixed (contents/data-layers set by configuration file)
Fixed data
Longhurst provinces
Bathymetry
Process indicator layers
- Variance (all data)
- Standard deviation (all data)
- Kurtosis (all data)
- no. of data points (all data)
- Kriging stats (SOCAT data only)
- Regions of low wind speed
- Dominance of wave breaking
- Diurnal warming (foundation – skin)
Process indicator layers
- Biology type (low, medium, high)
- Presence of strong SST gradients
Data layers
- Monthly data (from daily data, e.g. mean, median, krigging result)
- %age ice
- pCO_{2W}
- pCO_{2A}
- K_{WB}, K_{WO}, $K_{W,trad}$, K_{WR}, K_W
- flux
- ΔpCO_2
- asym, C_l, C_M
Process indicator layers
- Province
- Coastal or open-ocean


Basic climatology processing system
Climatology (monthly NetCDF)
Daily data
Rain (intensity, event)
Wind (speed, direction)
SOCAT
SST (foundation, skin)
Ice (%age, thickness)
Wave (altimeter and WWIII model)
Salinity
[1] Compositing and/or correction (mean, median, krigging, profile…)
Monthly data
Monthly data
SST fronts
Biology (ocean colour chl-a)
Takahashi climatology
[2] Flux calculation
- Default calculation (using Takahashi climatology for pCO2 data ie. not using SOCAT data)
- Input data user configurable
- Formulation user configurable
- Ability to re-process
- Output format fixed (contents/data-layers set by configuration file)
Fixed data
Longhurst provinces
Bathymetry
Process indicator layers
- Variance (all data)
- Standard deviation (all data)
- Kurtosis (all data)
- no. of data points (all data)
- Kriging stats (SOCAT data only)
- Regions of low wind speed
- Dominance of wave breaking
- Diurnal warming (foundation – skin)
Process indicator layers
- Biology type (low, medium, high)
- Presence of strong SST gradients
Data layers
- Monthly data (from daily data, e.g. mean, median, krigging result)
- %age ice
- pCO_{2W}
- pCO_{2A}
- K_{WB}, K_{WO}, $K_{W,trad}$, K_{WR}, K_W
- flux
- ΔpCO_2
- asym, C_I, C_M
Process indicator layers
- Province
- Coastal or open-ocean

Features of climatology processing system

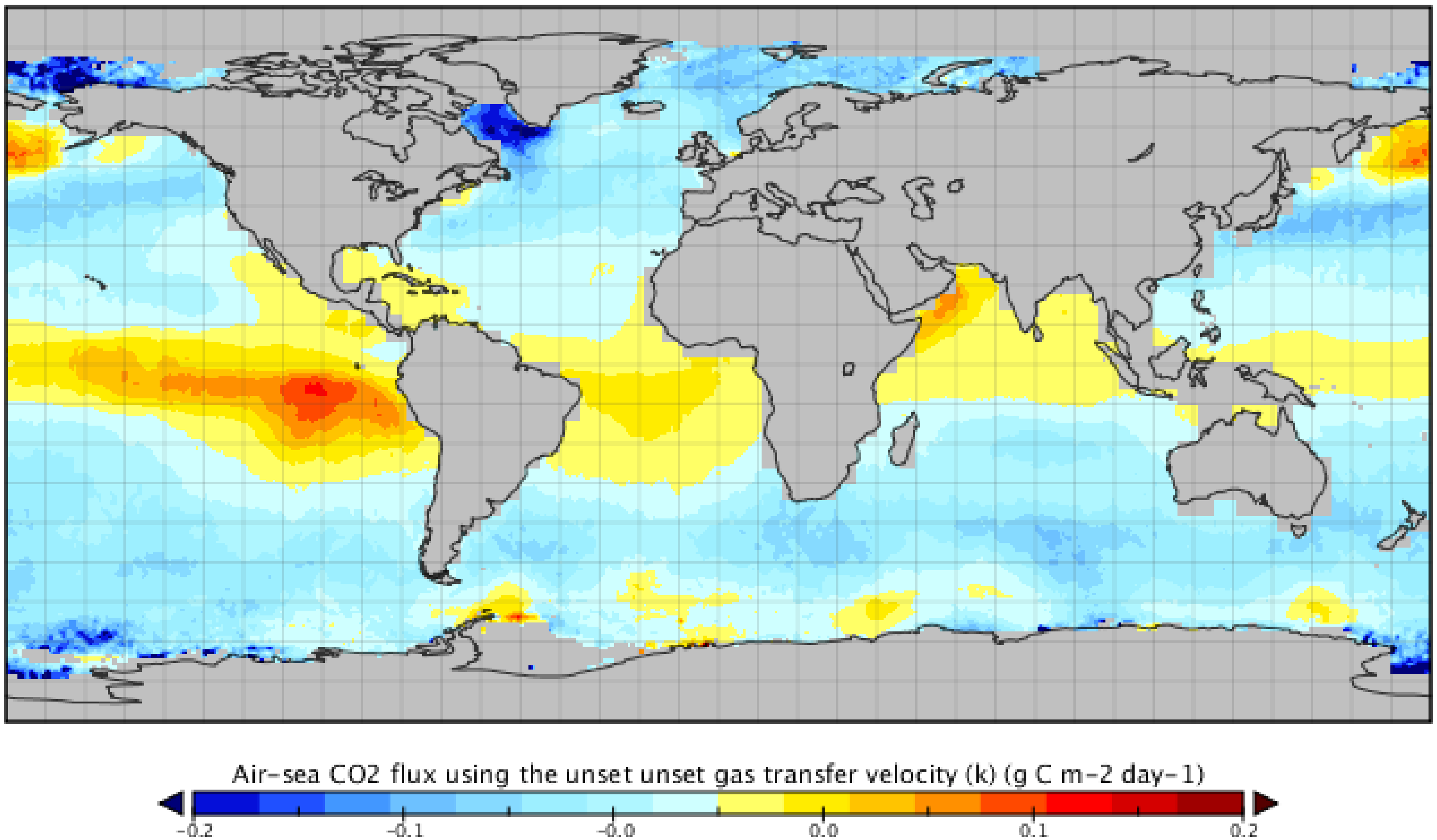
- All datasets are online and pre-processed into monthly composites.
- Monthly composites include mean, median, σ^2 and 2^{nd} - 4^{th} order moments.
- 1 Year global climatology takes 40 mins to generate.
- Ability to disable process indicator layers for speed increase
  - 1 year takes ~20 mins to generate
  - 10 year global run = 3.5 hours.

- Flux calculation is user configurable.
  - e.g user configurable wind based k relationships, optional handling of vertical thermal and haline gradients.
- Ability to inject noise or biases into input datasets
  - e.g. inject random noise based on known uncertainties of individual input data.

- Additional tools:
  - Net flux tool
  - Takahashi style grid re-sampling tool
- Python and Perl using standard libraries. ie no licenses are required and portable (>4000 lines of code !)

Example output

Global regular grid $1^{o} \times 1^{o}$ NetCDF 3.0, CF 1.6

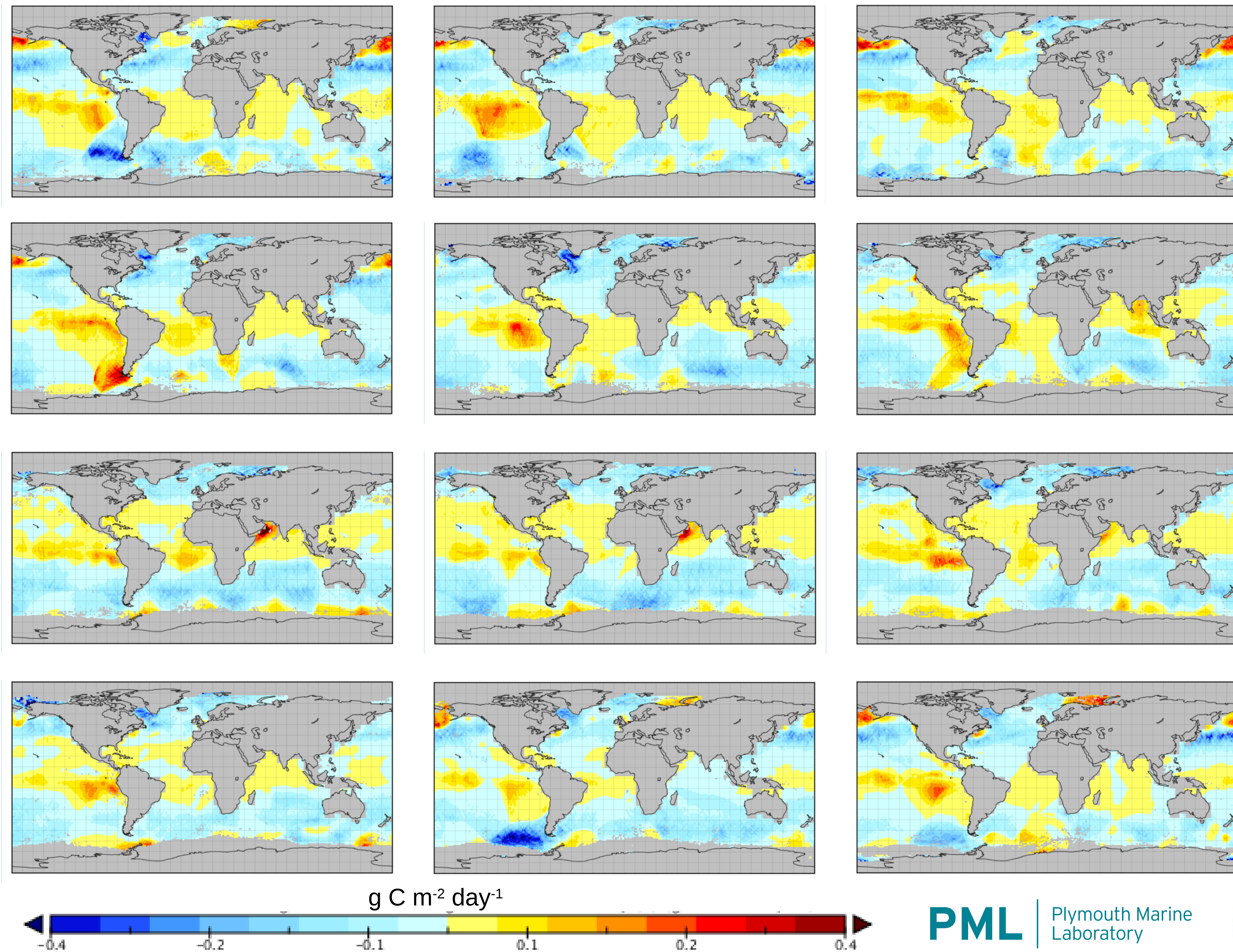


Example daily mean flux

PML Plymouth Marine Laboratory

oceanflux ghg support to science element

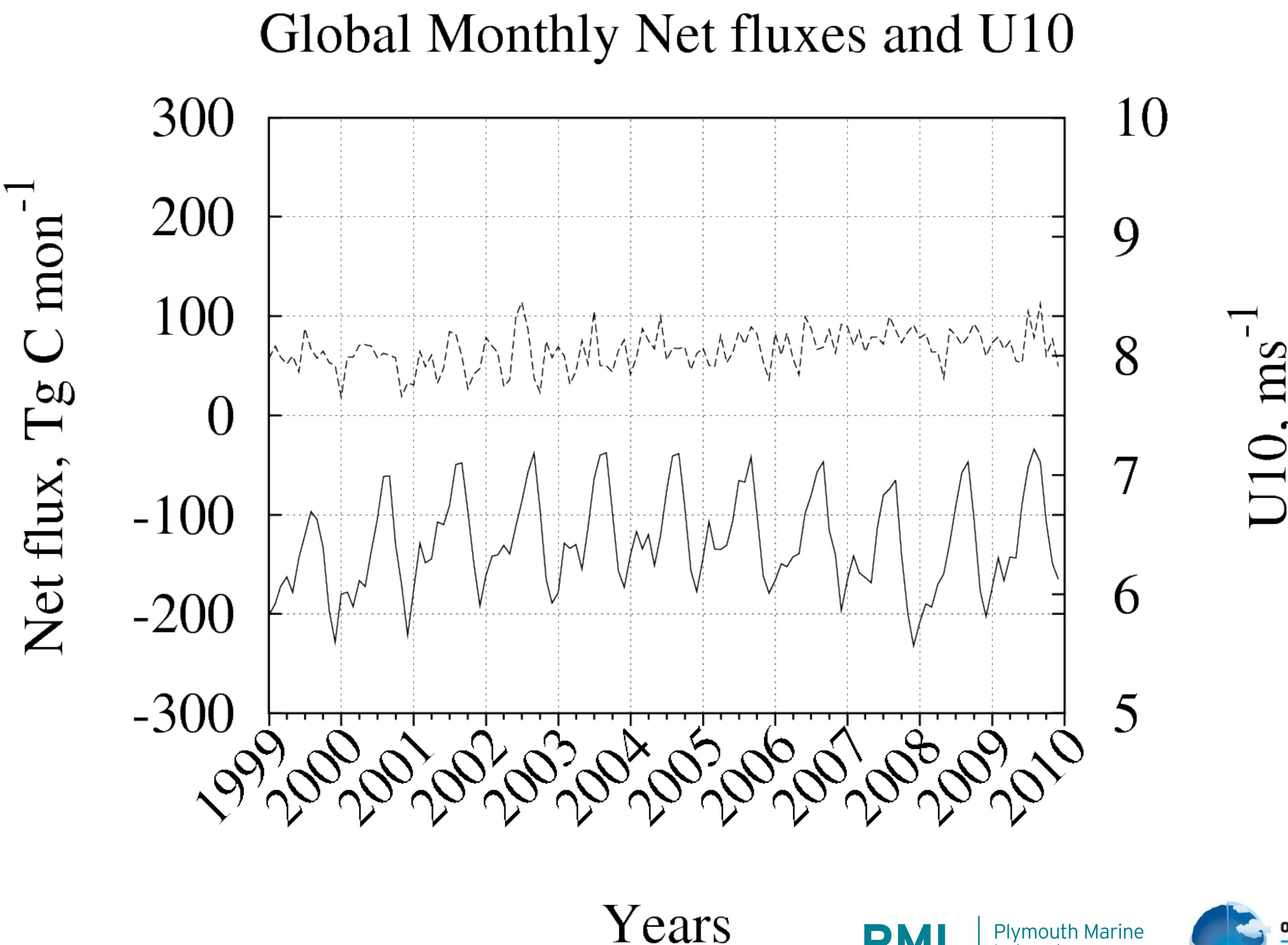
Example 2010



PML Plymouth Marine Laboratory

oceanflux ghg
support to science element

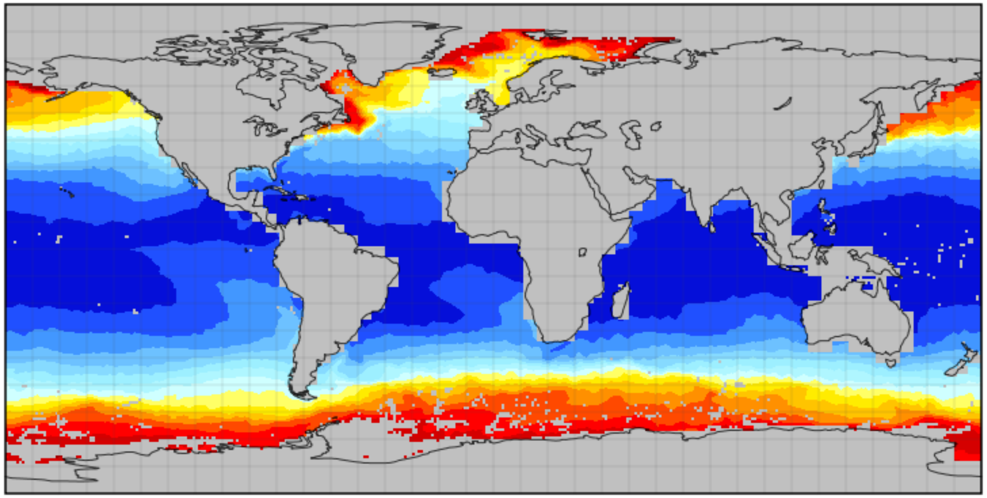
Example global time series



PML Plymouth Marine Laboratory

oceanflux ghg support to science element

Example output

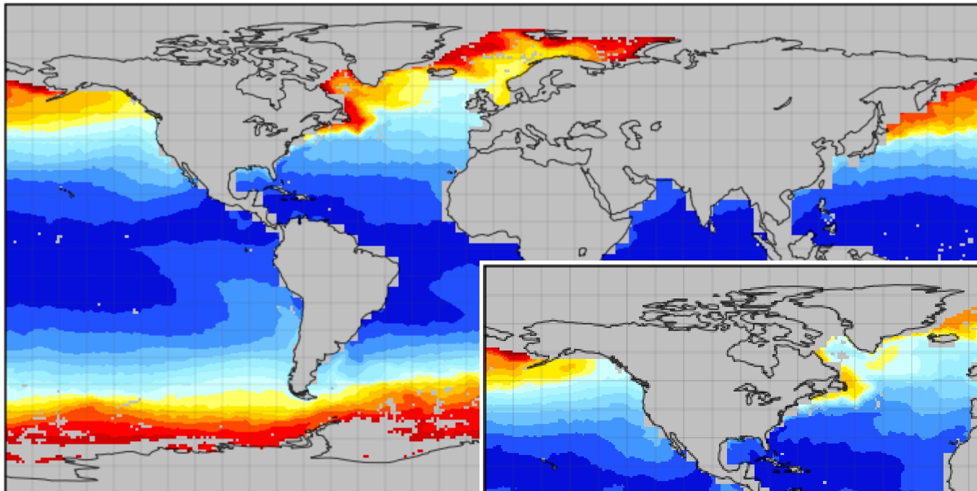


Concentration at the interface (C_I)

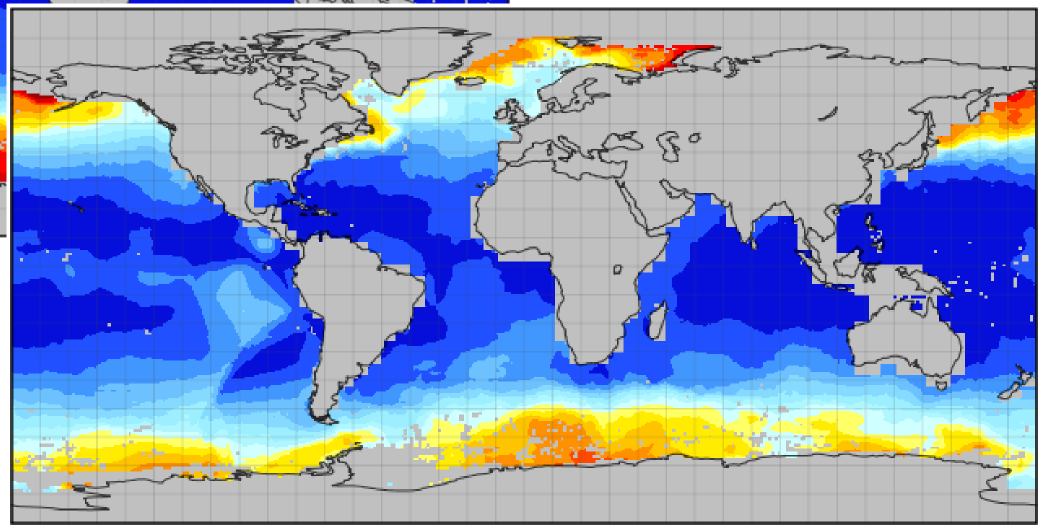
PML | Plymouth Marine Laboratory

oceanflux ghg support to science element

Example output



Concentration at the interface (C_I)

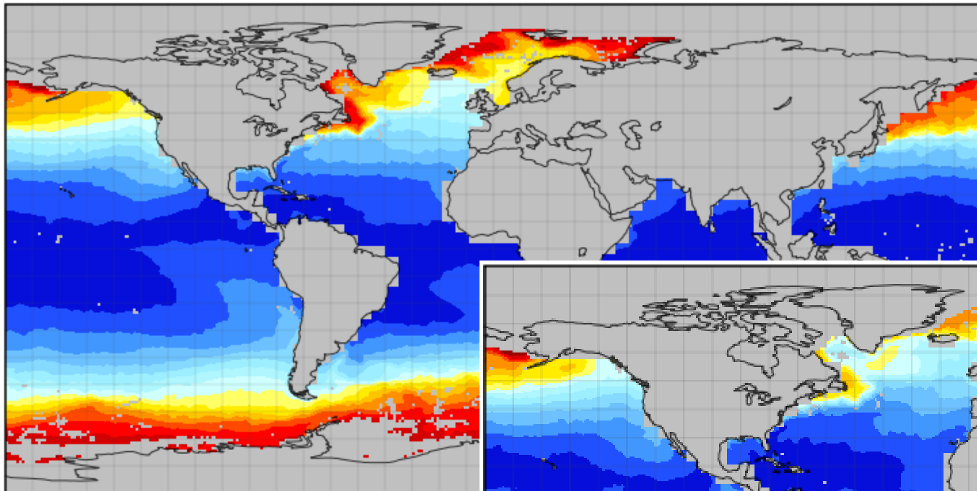


Concentration at the MBL (C_M)

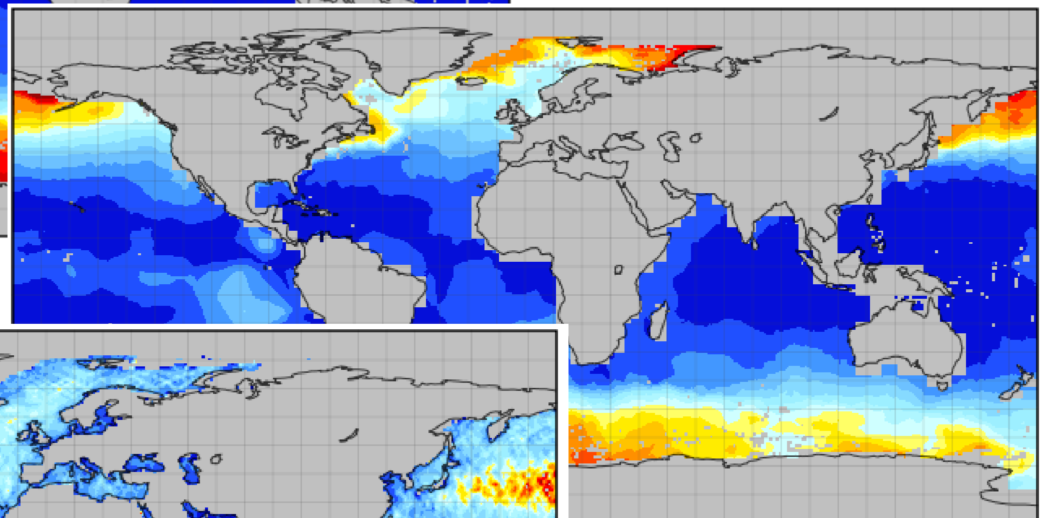
PML Plymouth Marine Laboratory

oceanflux ghg support to science element

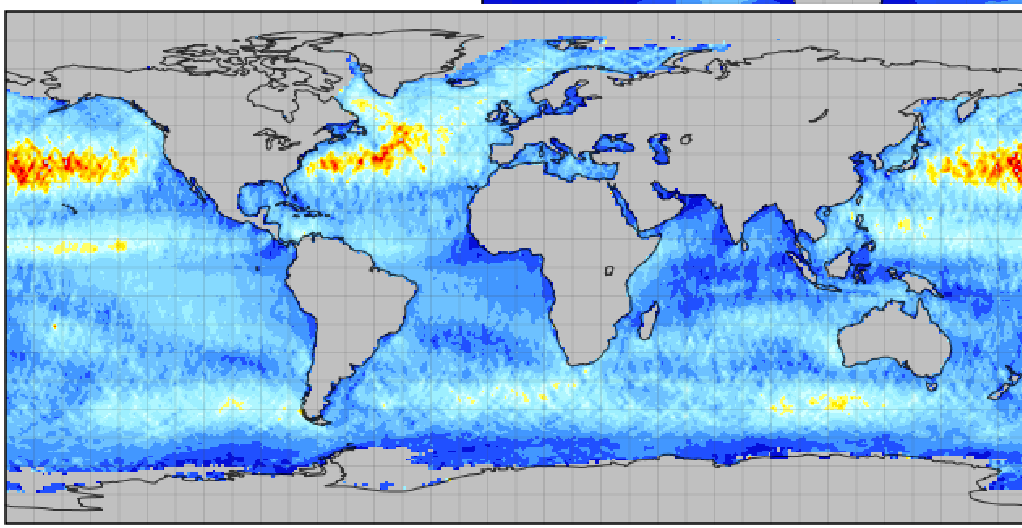
Example output



Concentration at the interface (C_I)



Concentration at the MBL (C_M)

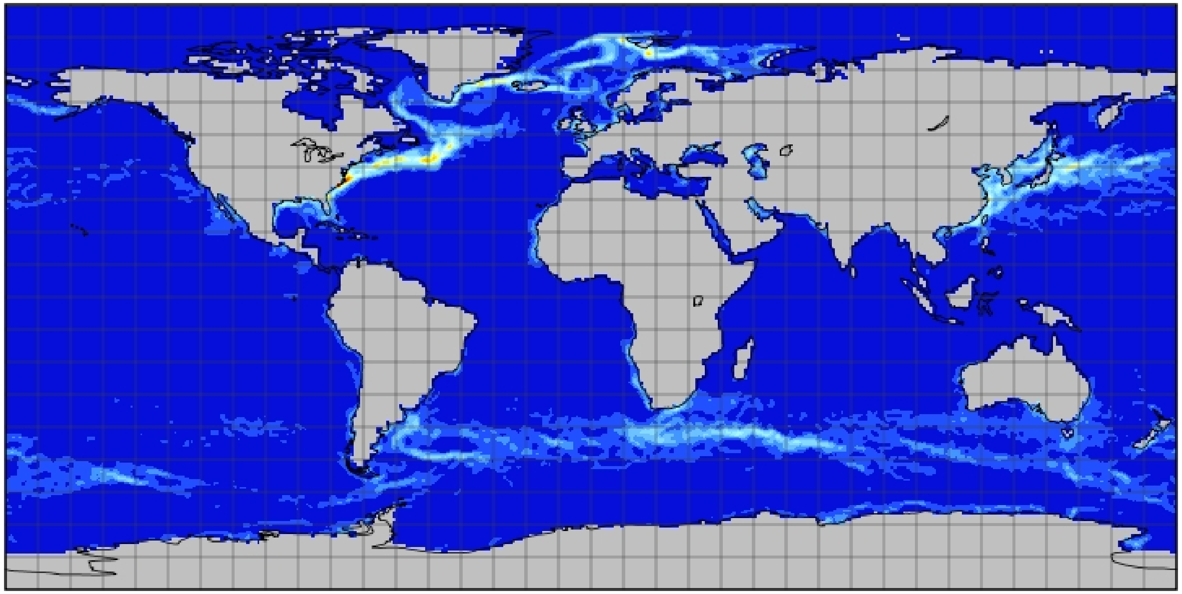


Gas transfer velocity (k)

PML Plymouth Marine Laboratory

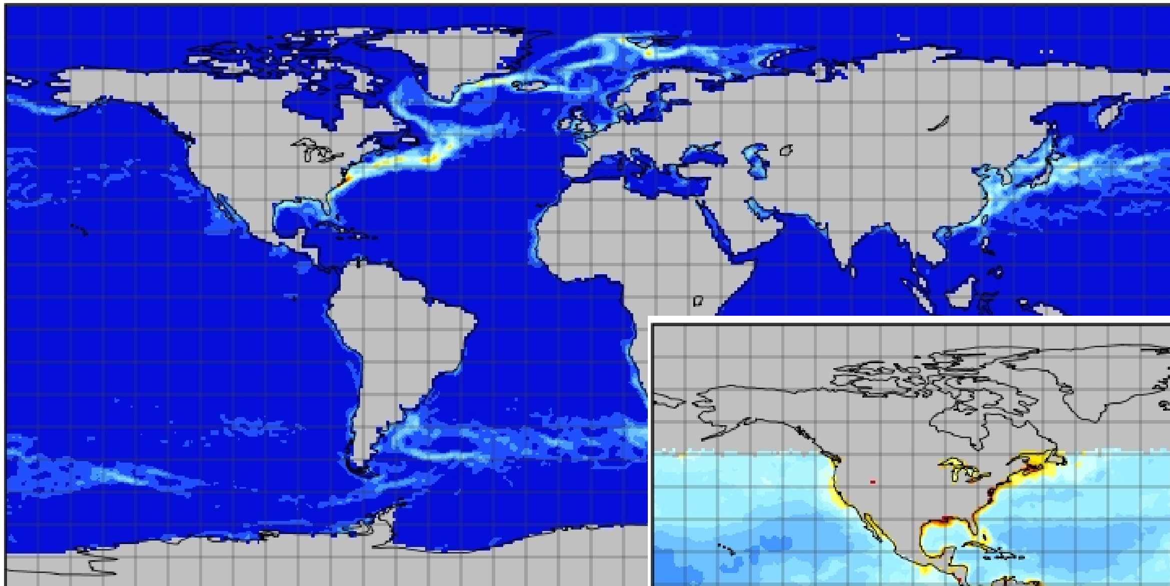
oceanflux ghg support to science element

Example output – process indicator layers

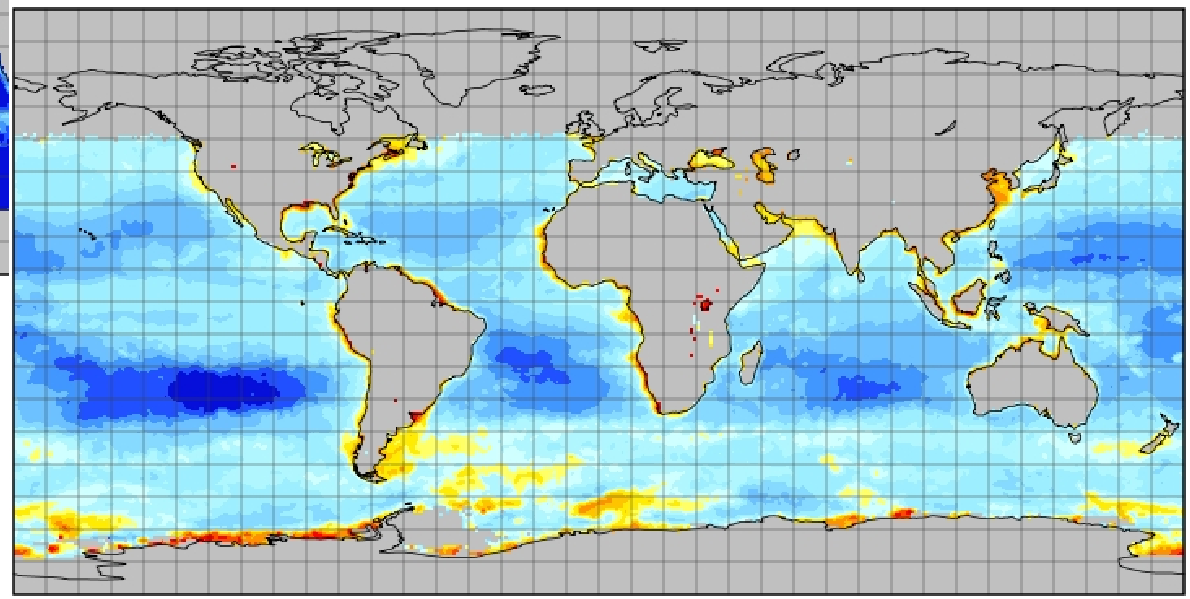


Persistent SST fronts (GHRSST)

Example output – process indicator layers

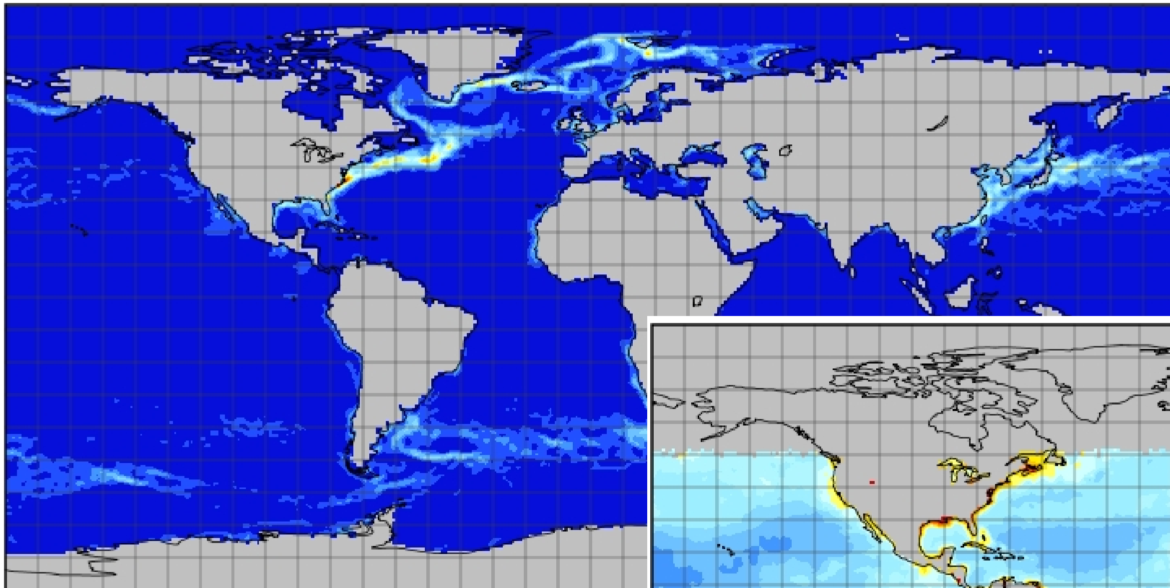


Persistent SST fronts (GHRSST)

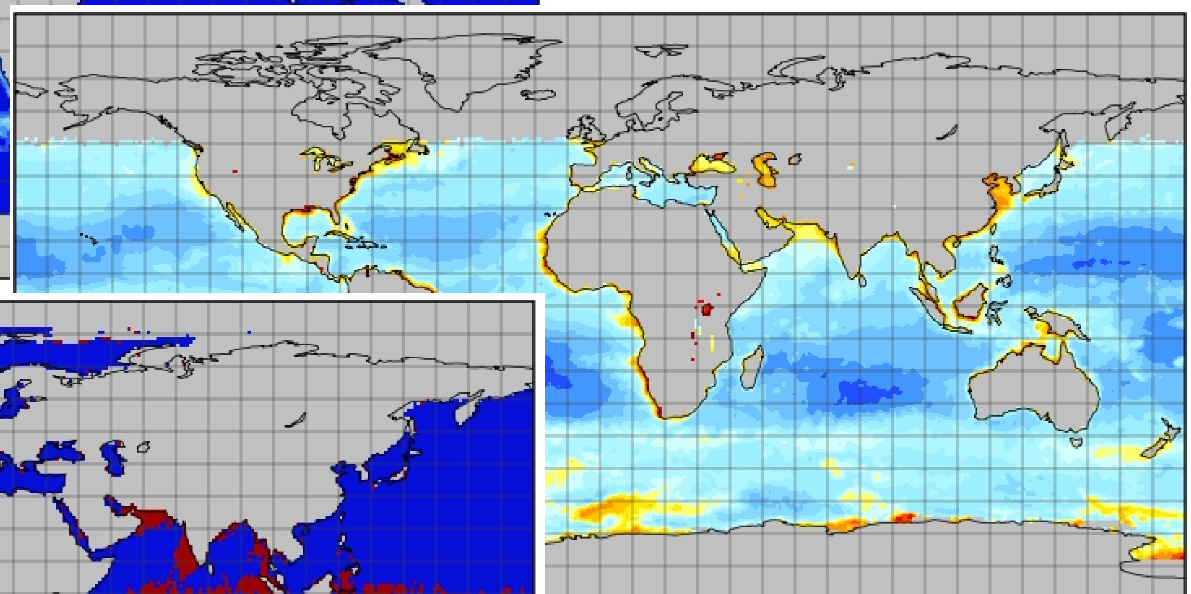


Surface chl-a (ESA CCI)

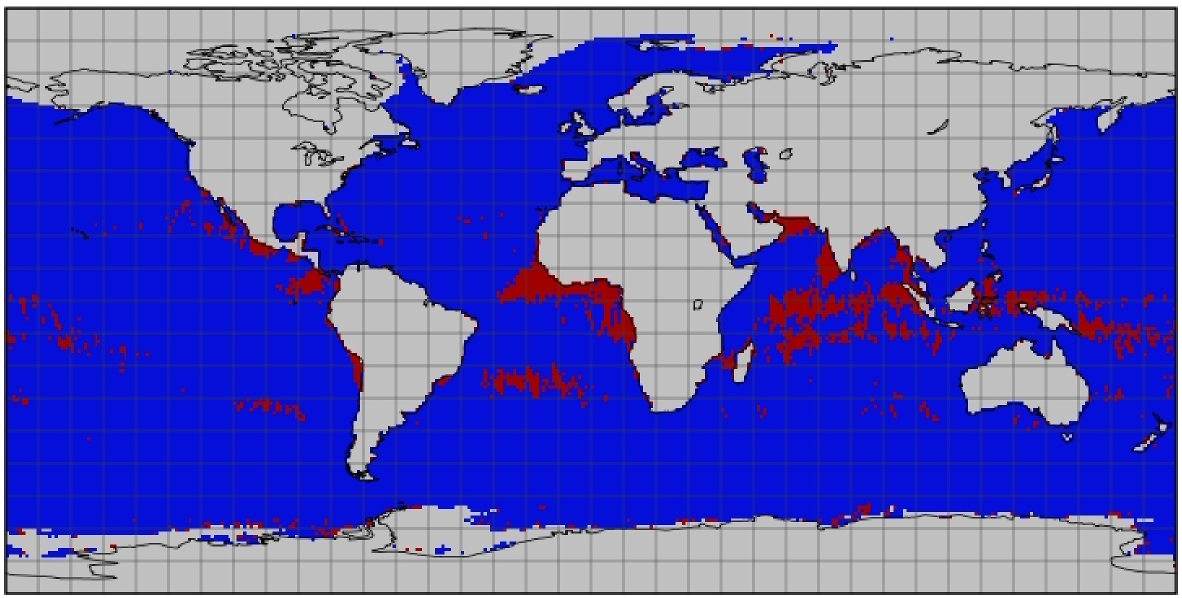
Example output – process indicator layers



Persistent SST fronts (GHRSST)



Surface chl-a (ESA CCI)



Regions of low wind

PML | Plymouth Marine Laboratory

oceanflux ghg
support to science element

Verification and testing

- Software extensively tested over a period of months.
- Quality layers highlight any regions which fail quality criteria as set out in the OceanFlux GHG Technical Specification (TS).

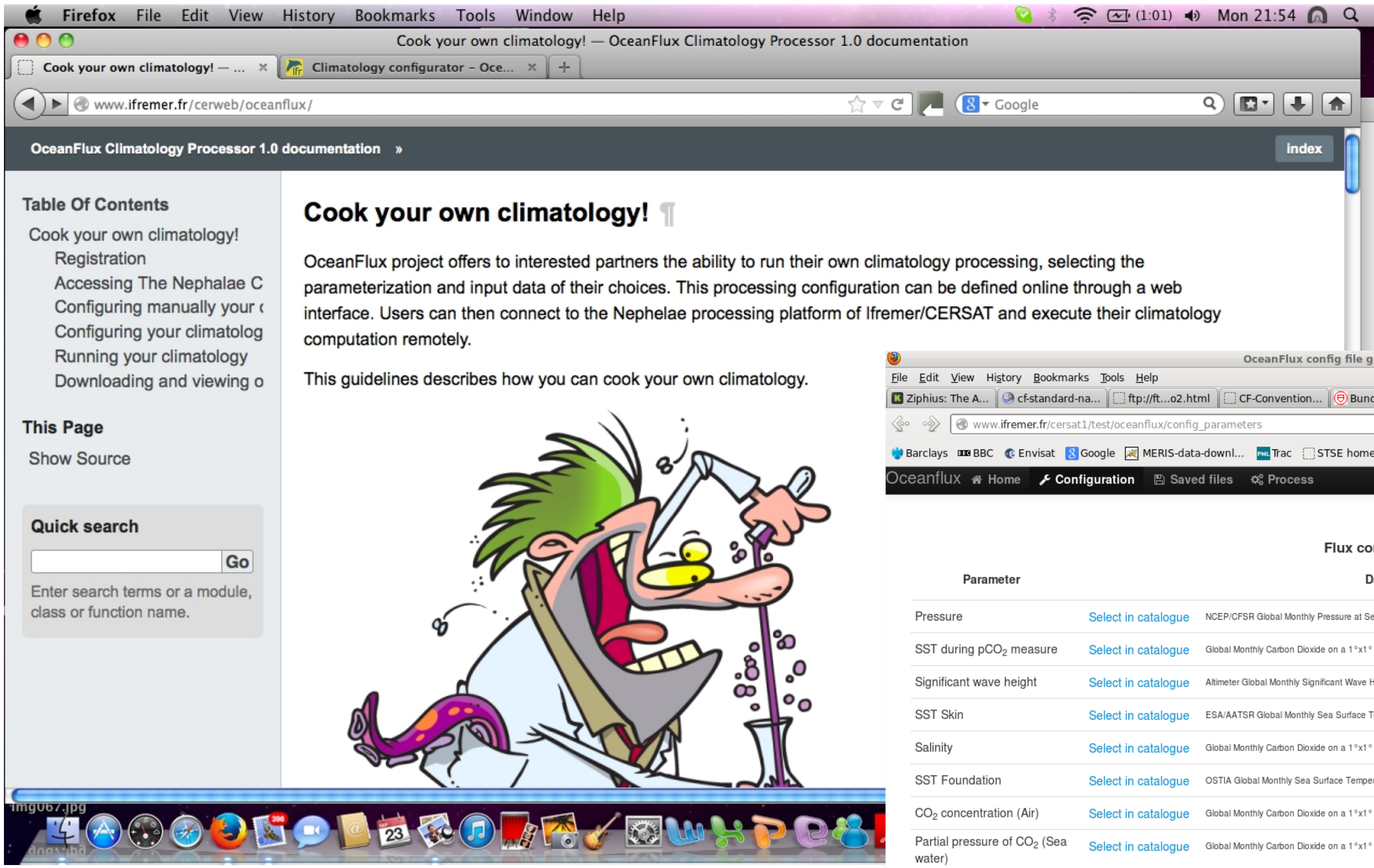
- Software verified using Takahashi climatology data (SST, XCO_2, U_{10}, pCO_{2w}, air pressure, ice) as input (at $1^o \times 1^o$):
  - kSW06 < 2 % difference
  - pCO_{2a} < 0.25 % difference
  - pH_2O < 1.4 % difference
  - K_o < 2 % difference
  - Annual global net flux 1.34 Gt C yr^{-1} (ref: Takahashi 1.4 ± 0.7 Gt C yr^{-1})

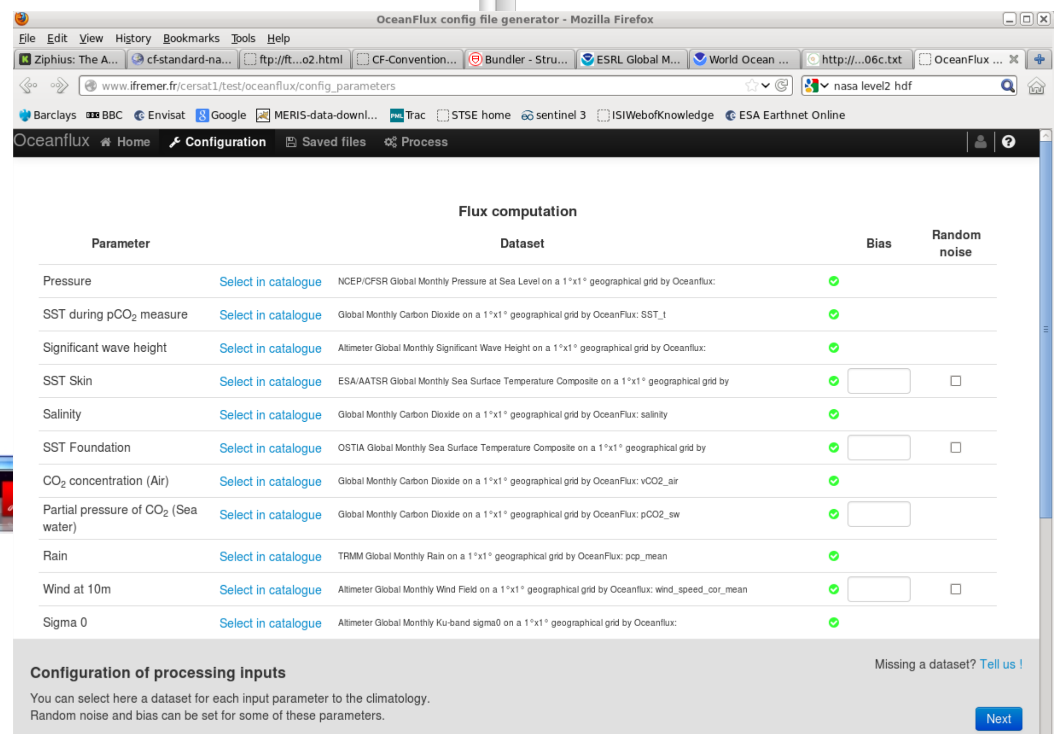
- Development version of software and output formats have been evaluated by an external independent third party company.
- NetCDF3 output follow CF 1.6 guidelines

PML Plymouth Marine Laboratory

oceanflux ghg support to science element

Web based control



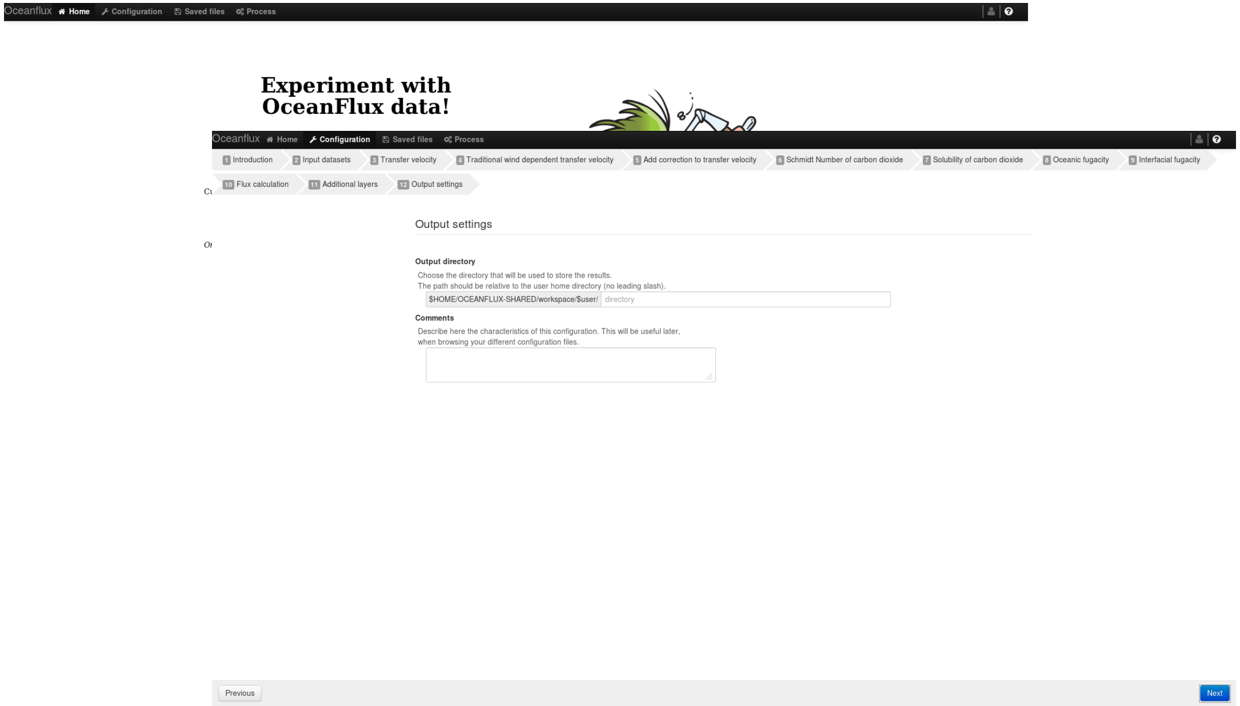


PML Plymouth Marine Laboratory

oceanflux ghg support to science element

Web based control

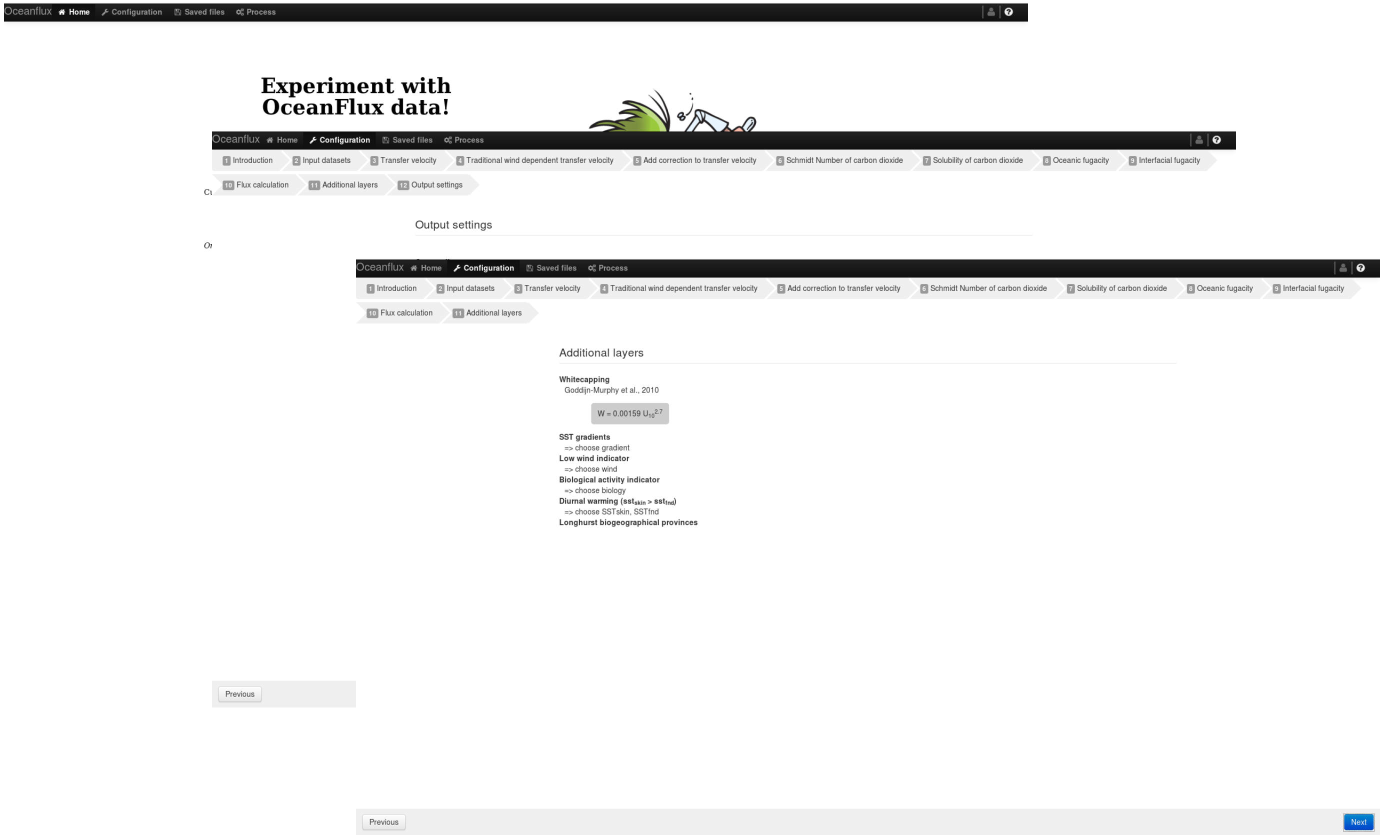
Web based control



PML Plymouth Marine Laboratory

oceanflux ghg
support to science element

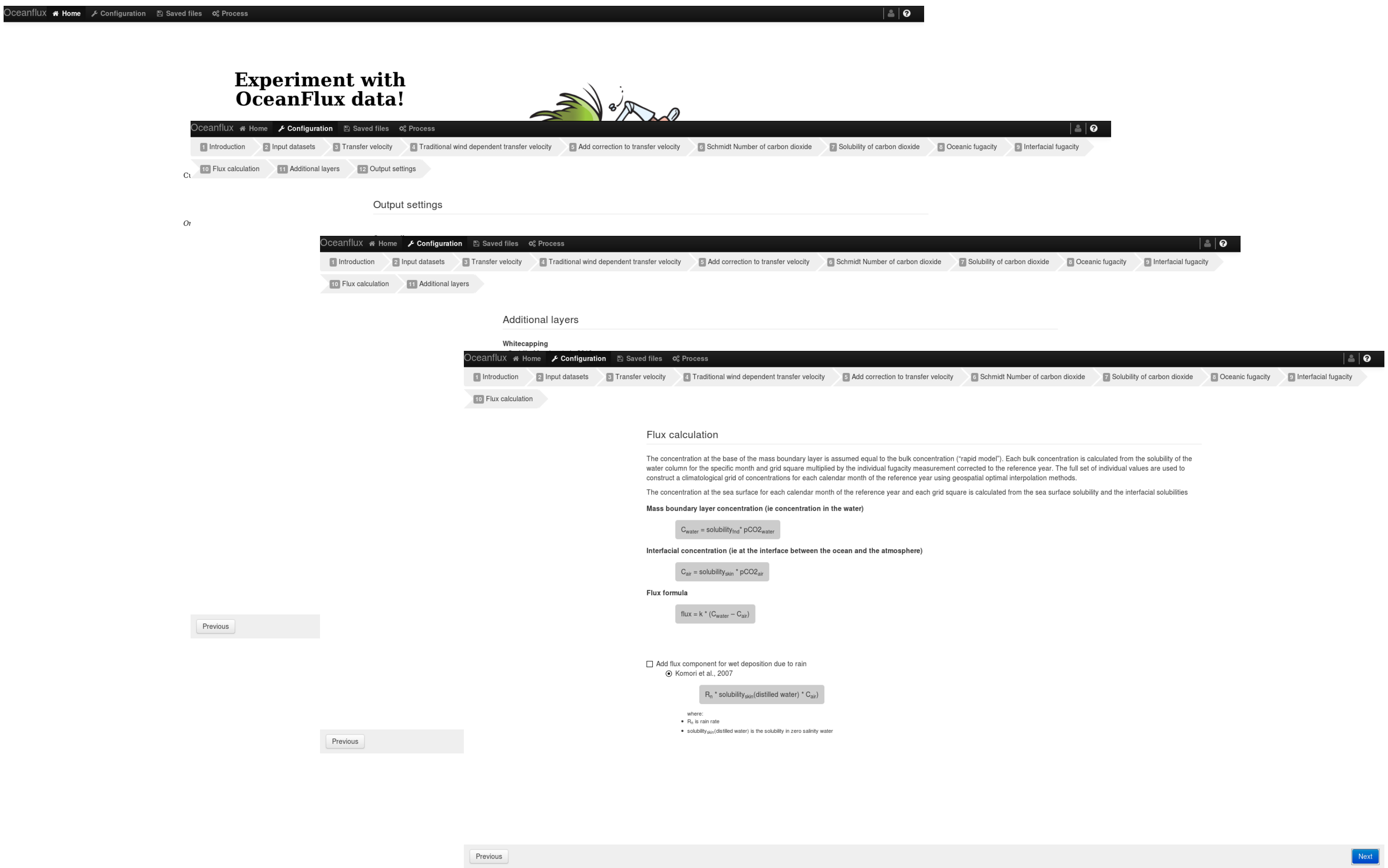
Web based control



PML | Plymouth Marine Laboratory

oceanflux ghg
support to science element

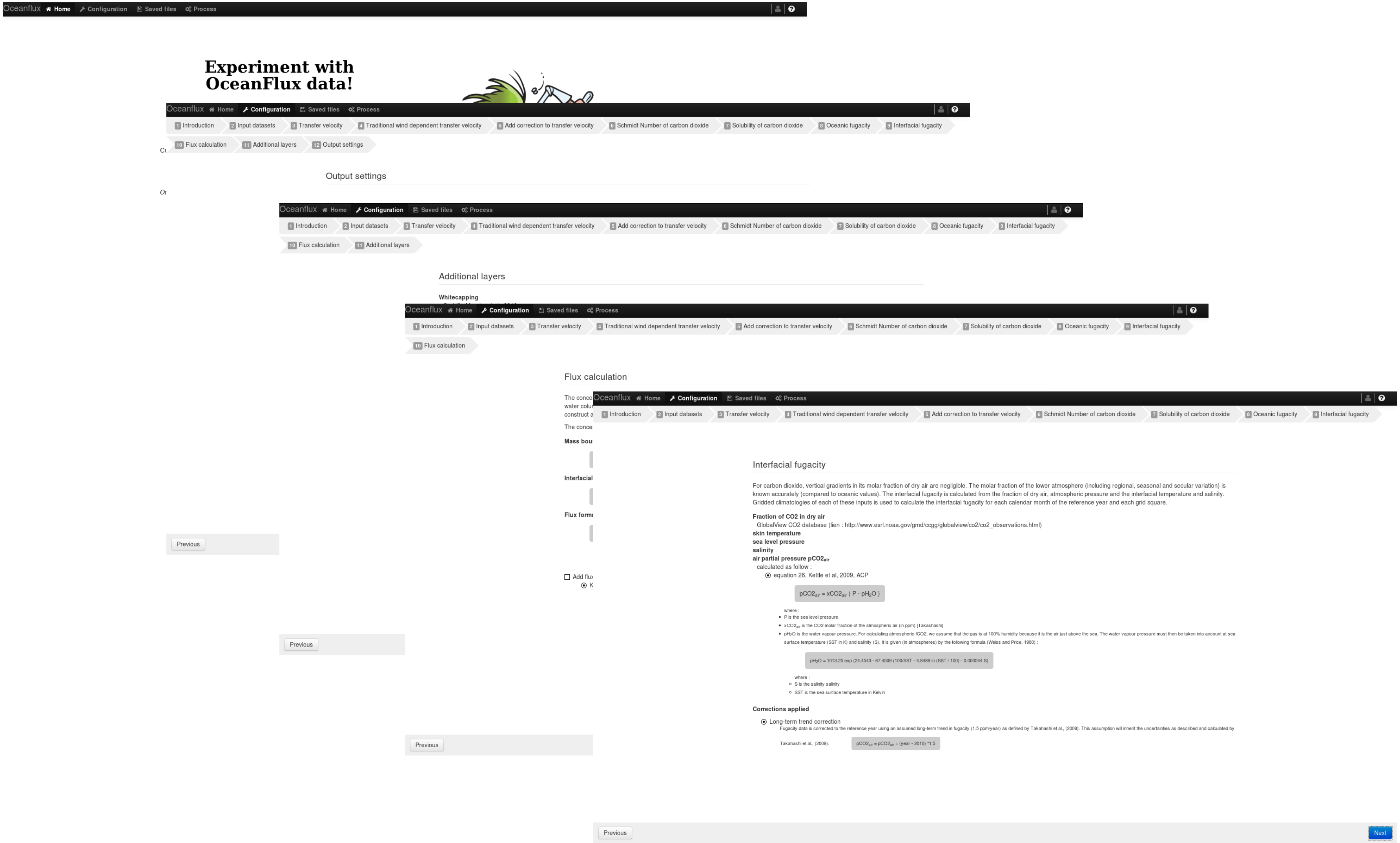
Web based control



PML Plymouth Marine Laboratory

oceanflux ghg support to science element

Web based control

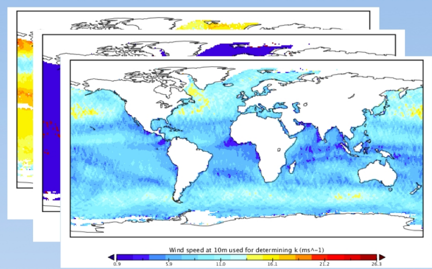


PML | Plymouth Marine Laboratory

oceanflux ghg
support to science element

And finally…. a brochure

INPUT



Monthly input Data

Climate quality data
SST (ESA Sea Surface Temperature CCI)
Wind and waves (ESA GlobWave)
Biology (ESA Ocean Colour CCI)

Additional Earth observation data
Wind, waves, rain, sea ice

In situ data
SOCAT fCO_2
(corrected to 2010 and interpolated)

Model data
Wave WWIII
ECMWF

COMBINING DATA

oceanflux ghg
support to science element

AIR-SEA CO_2 FLUX CLIMATOLOGY

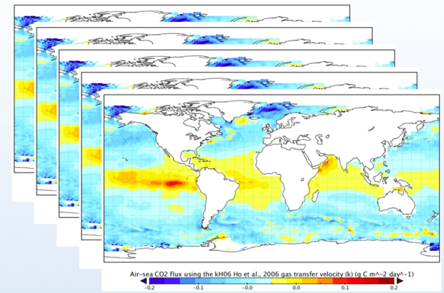
Flux calculation

Woolf et al., 2013 (Biogeosciences).
Validated and published algorithms

Processing system
Optional use
Open access
User configurable
Web interface for configuration
Choose your flux parameterisation
Selection of input data
Cloud data processing

Tools
All written in Python
For re-gridding to Takahashi grid
For net flux calculation

CLIMATOLOGY



1° × 1° global NetCDF

Monthly data layers
- air-sea CO_2 flux
- C_I (interfacial conc. of CO_2)
- C_M (conc. of CO_2 at base of MBL)
- K_{wb}, K_{wo}, $K_{w,trad}$, K_{wr}, K_w
- ΔpCO_2
- %age ice cover
+ others

Process indicator layers
- Interpolation stats (SOCAT data only)
- Regions of low wind speed
- Dominance of wave breaking
- Biology type (low, medium, high)
- Persistent SST fronts
- Coastal or open ocean
- Longhurst Provinces

Potential users
SOLAS scientists
Climate modellers
(e.g. data assimilation, model evaluation)

PDF on project website

PML Plymouth Marine Laboratory

oceanflux ghg
support to science element

Conclusions

- Highly flexible community tool to calculate global air-sea CO_2 fluxes.
- Exploits Cloud based computing
- Large amount of *climate quality data* and processing capability available.
  - 8 Terra bytes available and/or has been pre-processed
  - 500 Terra bytes of EO data available
  - 600 processing cores
- Flexible solution
  - e.g. very simple to add additional datasets or run long time series.
- Output fluxes have been verified.
- System exploited for a number of OceanFlux GHG investigations and analyses.

Future:

Expand to handle other gases ?

Shutler, J. D., Piolle, J. F., Land, P. E., Woolf, D. K., Goddijn-Murphy, L., Paul, F., *et al*. (in-prep) Air-sea gas flux data processing system using Cloud computing: a flexible and dynamic tool for the community, *to be submitted to Journal of Atmospheric and Oceanic Technology.*