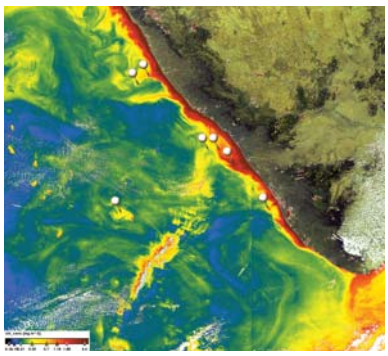
The CoastColour Validation Dataset

CoastColour is a LOICZ affiliated project

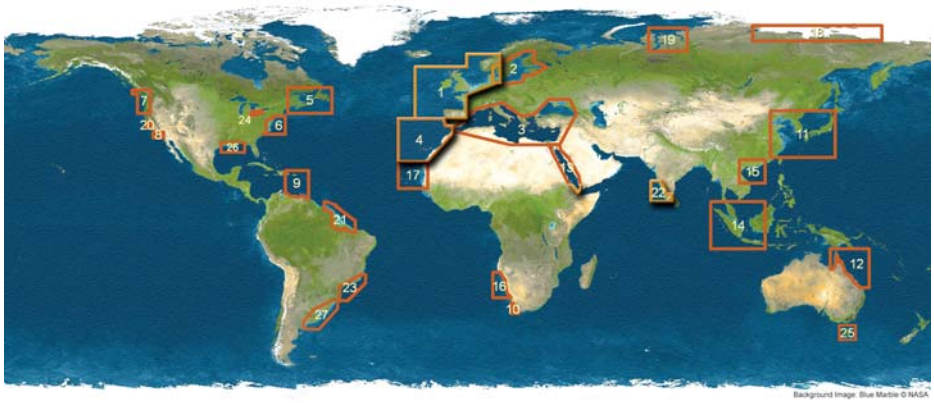
The Validation Dataset on Calvalus

To ensure a fast validation of new algorithms, a defined CoastColour validation dataset (VDS) is permanently stored on the processing cluster Calvalus. The size of the CoastColour validation dataset has been limited to a number of three testsites, to guarantee that the time required for processing and validation stays within a certain limit. For each of these sites all available FRS data for the years 2005-2009 are processed to all CC product levels, CC L1P, CC L2R and CC L2W.
Despite the reduction to three sites, the size of the output data set reaches approximately 4.9 TB. The processing time is approximately 60 hours.
Every time a new algorithm has been developed and implemented, processing of the VDS starts from the beginning and the output products are provided to the corresponding project partners for further validation and analysis.

The three CoastColour sites that are part of the VDS are:
- Site 1 (North Sea, English Channel, Bay of Biscay, Celtic Sea)
- Site 4 (Morocco)
- Site 22 (South India)



Southwest coast of India; 10.02.2005, orbit: 15418
Composition L1 RGB and chlorophyll concentration

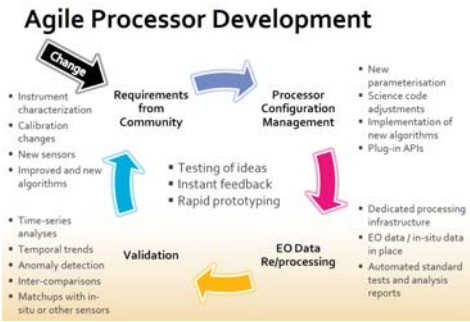


Background Image: Blue Marble © NASA

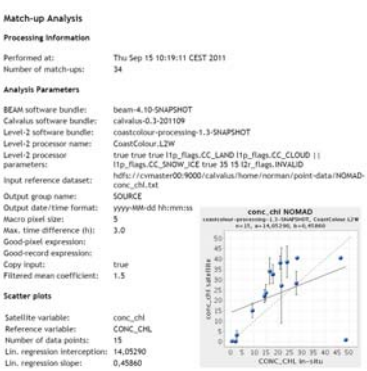
About Calvalus

The technology has been designed for processing of very large amounts of data. Calvalus gains its performance from massive paralellisation of tasks combined with a Distributed File System (DSF) and the data-locality paradigm.

The Calvalus system has been developed under the ESA LET-SME programme. The system is based on the MapReduce programming model (MR) combined with a Distributed File System and comprises a cluster of 20 commodity computers with a total disk capacity of 112 TB. The processing system software is based on Apache Hadoop - an open-source implementation of MR and DSF in Java.
The Calvalus system allows users to efficiently perform cal/val and EO data processing functions, thus allowing an agile product development and fast improvement cycles. The match up data are collected and match-up and trend analysis reports are generated.



Calvalus as Validation Tool



Match-up Analysis

Processing Information

Performed at:	Thu Sep 15 10:19:11 CEST 2011
Number of match-ups:	34

Analysis Parameters

BEAM software bundle:	beam-4.10-SNAPSHOT
Calvalus software bundle:	calvalus-0.3-201109
Level-2 software bundle:	coastcolour-processing-1.3-SNAPSHOT
Level-2 processor name:	CoastColour.L2W
Level-2 processor parameters:	true true true l1p_flags.CC_LAND l1p_flags.CC_CLOUD | | l1p_flags.CC_SNOW_ICE true 35 15 l2r_flags.INVALID
Input reference dataset:	hdfs://cvmaster00:9000/calvalus/home/norman/point-data/NOMAD-conc_chl.txt
Output group name:	SOURCE
Output date/time format:	yyyy-MM-dd hh:mm:ss
Macro pixel size:	5
Max. time difference (h):	3.0
Good-pixel expression:	
Good-record expression:	
Copy input:	true
Filtered mean coefficient:	1.5

Scatter plots

Satellite variable:	conc_chl
Reference variable:	CONC_CHL
Number of data points:	15
Lin. regression interception:	14.05290
Lin. regression slope:	0,45860

Calvalus offers the possibility to run a semi-automatic validation of algorithm performance, using a set of match-up pixels and the corresponding in-situ data. The in-situ data include meta-information like geolocation and acquisition time which are also used for the identification of corresponding match-up pixels. The Match-up definition for CoastColour includes pixels within a time difference of +/-3 hours and if the surrounding 5x5 pixels are flagged as water and valid for processing. A macro pixel of 5x5 pixels around the match-up pixel is processed and builds the output.
The in-situ data that are stored on the Calvalus system are organised in csv files and contain measurements of selected parameters, which will be validated in the following step.
Output of the validation are scatterplots for each validation parameter with short reports containing meta-information about the algorithm and the processing. The output data combining in-situ data and match-ups are provided as csv files.

	Parameter	Definition
IOPs	iop_bb_spm_443	Backscattering of suspended particulate matter at 443 nm [m^-1]
	iop_a_total_443	Total absorption coefficient of all water constituents at 443 nm [m^-1]
	iop_a_ys_443	Yellow substance absorption coefficient at 443 nm [m^-1]
	iop_a_pig_443	Pigment absorption coefficient at 443 nm
	kd_490	Downwelling irradiance attenuation coefficient at 490 nm [m^-1].
Conc	conc_tsm	Total suspended matter dry weight concentration [g*m^-3]
	conc_chl	Chlorophyll concentration [mg*m^-3]
Radiance	norm_refl_#	Normalised water leaving radiance reflectance at xxx nm [sr^-1]

Selected parameters for the validation cycle

Coordinated by Dr. Carsten Brockmann, Brockmann Consult · Email: carsten.brockmann@brockmann-consult.de